Generation and chromosome mapping of expressed sequence tags (ESTs) from a human infant thymus

Mee-Yul Hwang, Yoon-Joong Kang, Young-Ho Kim, Steve W. Scherer, Lap-Chee Tsui, and Uik Sohn

Abstract: In an effort to identify novel genes that are expressed differentially in an infant thymus, we constructed an oligo-d(T) primed cDNA library from a human infant thymus followed by single-run partial sequencing to generate expressed sequence tags (ESTs). Characterization of more than 1400 sequences enabled us to convert human thymus transcripts into 1223 useful ESTs. These ESTs consisted of 613 (50.1%) showing homology to known human genes, 51 (4.2%) matching to genes from other species, 289 (23.6%) matching ESTs of unknown functions, and 182 (14.9%) being novel transcripts. The expression profile of an infant thymus features a high number of genes related to cell division–DNA synthesis and gene–protein expression, indicating the active growth stage of an infant thymus. To identify the chromosomal localization of 43 thymus ESTs, PCR-based mapping was performed using a human–rodent somatic cell hybrid or radiation hybrid mapping panel. The results indicated that several novel genes were determined to be located in the vicinity of previously mapped disease loci; histidinemia loci, plasminogen Tochigi disease loci, Ehlers-Danlos syndrome, hypertriglyceridemia, thyroid resistance locus, ocular albinism, galactosemia, porphyria variegata, Charcot-Marie-tooth disease, FEOM (fibrosis of extraocular muscles), Prader-Willi syndrome.

Key words: cDNA, expression profile, radiation hybrid mapping, disease locus.

Résumé : Afin d'identifier de nouveaux gènes qui sont exprimés de manière différentielle dans le thymus d'un nouveau-né, une banque d'ADNc provenant du thymus d'un nouveau-né et dont la synthèse a été initiée avec une amorce oligo-d(T) a été produite. Ces ADNc ont ensuite fait l'objet d'une seule réaction de séquençage de façon à générer une collection de séquences exprimées (EST). Une caractérisation de plus de 1400 séquences a permis de convertir des transcrits humains en 1223 EST utiles. Ces EST comprenaient 613 (50,1%) séquences montrant une homologie à des gènes humains connus, 51 (4,2%) étaient homologues à des gènes connus chez d'autres espèces, 289 (23,6%) étaient apparentés à des EST de fonction inconnue et 182 (14,9%) correspondaient à de nouvelles séquences. Le profil d'expression d'un thymus de nouveau-né montre un nombre élevé d'expression de gènes impliqués dans la division cellulaire–synthèse de l'ADN et dans l'expression des gènes–protéines, ce qui est le reflet du stade de croissance active du thymus chez le nouveau-né. Afin de déterminer la position chromosomique de 43 EST du thymus, une cartographie a été réalisée à l'aide d'une technique PCR employée sur des hybrides somatiques ou une collection d'hybrides résultant d'irradiation. Les résultats indiquent que plusieurs nouveaux gènes sont localisés à proximité de loci de maladies : des loci d'histidinémie, de la maladie du plasminogène Tochigi, du syndrome d'Ehlers-Danlos, de l'hypertriglycéridémie, de la résistance thyroïdienne, de l'albinisme oculaire, de la galactosémie, du porphyra variegata, de la maladie dentaire Charcot-Marie, de la FEOM (fibrose des muscles extraoculaires) et du syndrome de Prader-Willi.

Mots clés : ADNc, profil d'expression, cartographie à l'aide d'hybrides résultant d'irradiation, locus de maladie.

[Traduit par la Rédaction]

Corresponding Editor: R.S. Singh.

Received July 8, 1998. Accepted November 11, 1998.

M.-Y. Hwang, Y.-J. Kang, and U. Sohn.[1] Department of Genetic Engineering, College of Natural Sciences, Kyungpook National University, Taegu, 702–701, Korea.
Y.-H. Kim. Department of Microbiology, College of Natural Sciences, Kyungpook National University, Taegu, 702–701, Korea.
S.W. Scherer and L.-C. Tsui. Department of Genetics, The Hospital For Sick Children, Toronto, Ontario, M5G 1X8, Canada.

[1]Author to whom all correspondence should be addressed (e-mail: usohn@bh.kyungpook.ac.kr).

Introduction

Since ESTs have been used as a starting point for the functional and structural analysis of the entire human genome, many efforts to identify all the human genes have focused on studies of expressed sequence tags (ESTs) generated by single-run partial cDNA sequencing (Kirkness 1996). This large-scale sequencing has not only provided comprehensive information on the expression pattern of a variety of human tissues and cells (Adams et al. 1993 *a*, 1993*b*; Affara et al. 1994; Choi et al. 1995; Frigerio et al. 1995; Hwang et al. 1995; Kawamoto et al. 1996; Liew et al. 1994; Lee et al. 1995; Okubo et al. 1995; Sudo et al. 1994; Tanaka et al. 1996) but has also generated gene-based sequence-tagged sites for rapid large scale PCR mapping (Berry et al. 1995). The utility of EST mapping has been shown by the uncovering of new insights into genome organization, evolution, and expression, as well as by the facilitating of the positional cloning of human genes linked to diseases (Boguski and Schuler 1995; Collins et al. 1995).

The thymus is a central lymphoid organ in the upper chest which is responsible for the development of immunocompetent cells from immature T cells. In 1995, the first report appeared on the generation and chromosomal localization of human adult thymus ESTs from a normalized cDNA library (Lamerdin et al. 1995). In addition, the TIGR collection also reported over 1100 ESTs from primary or substracted cDNA libraries of human adult normal and tumor thymus tissues (Adams et al. 1995). However, there has not been any report on a gene expression profile reflecting the specialized functions and physiological characteristics of the human infant thymus. Accordingly, we describe here the generation and analysis of 1223 ESTs from a human infant thymus, and the chromosome mapping of 43 ESTs using PCR-based mapping strategy.

Methods

Construction of cDNA library and sequencing

Total RNA was isolated from a 14 mo old human thymus according to the guanidine thiocyanate method (Sambrook et al. 1989). Poly (A)+ RNA was obtained from total RNA by an oligo d(T)-cellulose column. cDNAs were synthesized using a cDNA synthesis kit (Amersham, U.K.). The cDNAs were anchored with a *Not*I–*Eco*RI adaptor and digested by the *Eco*RI restriction enzyme, and subsequently inserted into an *Eco*RI-digested λZAPII vector and packaged utilizing a Gigapack gold II packaging system (Stratagen, La Jolla, Cal.). The cDNA library was in vivo excised with a R408 helper phage and transfected into an *E. Coli* strain, XL1-Blue. For sequencing, DNA templates were prepared by the alkaline lysis method (Sambrook et al. 1989). Sequencing reactions were performed with a Sequenase v. 2.0 kit (Amersham, UK).

Database analysis

Sequences obtained by the single-run partial sequencing were subject to a BLAST search (Altschul et al. 1990) against nonredundant nucleotide and protein databases, as well as dbEST. When the *P*-value of BLAST reports was less than 1.0×10^{-5}, the sequences were considered as database-matched clones.

Chromosome mapping

PCR primers were designed by the PRIMER program available on the Web site (http://www-genome.wi.mit.edu). PCR reactions were performed with specific primers against a NIGMS (Coriell Institute for Medical Research, Camden, N.J.) human–rodent somatic cell hybrid mapping panel 2 (Drwinga et al. 1993) and against a GENEBRIDGE 4 RH panel (Research Genetics of Huntsville, Alta.). All PCR reactions were performed in a 50 μL volume containing 25 ng of template DNA and 2.5 U of AmpliTaq Gold DNA polymerase (Perkin-Elmer Cetus, Conn.). Cycling parameters were 30 s at 94°C, 30 s at 55–60°C, and 1 min at 72°C for 35 cycles, followed by an additional 7 min at 72°C in a DNA thermal cycler (Perkin-Elmer Cetus). The radiation hybrid scoring data resulting from the PCR screenings were statistically analyzed to localize unknown genes on the framework map using the RHMAPPER program at the Whitehead/MIT center for Genome Research (http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl).

Table 1. Classification of human infant thymus ESTs.

Classification	Percent	No. of clones
Human matched	50.1	613
Non-human matched	4.2	51
EST	23.6	289
rRNA	4.7	58
Alu repeat	2	24
MtDNA	0.5	6
Unknown	14.9	182
Total	100	1223

Results and discussion

Random sequencing and database analysis

A bidirectional cDNA library was constructed and a total of 1223 ESTs were generated by single-run partial sequencing of over 1400 randomly selected cDNA clones. The average insert size was 0.7 kb and the average length of sequence was 220 bp, which is considered to be useful ESTs. Hillier et al. (1996) reported that the highest quality portion of the EST sequence was between 100 and 300 bases. We classified the results of the database comparisons according to materials and methods. A summary of the database searches for the 1223 sequences is shown in Table 1. Of the 1223 clones, 613 (50.1%) showed significant similarities to human genes and 51 (4.2%) to non-human genes. These 664 clones matching known genes represented 416 unique genes. The other database-matched clones contained 289 (23.6%) ESTs of unknown functions, 24 (2.0%) *Alu*-repeats, 58 (4.7%) rRNA, and 6 (0.5%) mitochondrial DNA. The frequencies of *Alu*-like repeats and mitochondrial DNA are significantly lower than those from previous reports (Sudo et al. 1994; Choi et al. 1995). Of the 1223 cDNA clones, 182 (14.9%) were found to be novel transcripts not showing any significant homology to known genes.

The 664 ESTs showing significant homology to known genes were classified into 7 groups according to their functions (Adams et al. 1995); cell signaling–communication, gene–protein expression, cell division–DNA synthesis, cell structure–motility, cell–organism defense and homeostasis, metabolism, and unclassified (Appendix 1). Among these 7 groups, the gene–protein expression group contained the greatest percentage (34.6%) of total putative identifications from the infant thymus ESTs. In order to investigate the changes of the gene expression in relation to the aging of the

thymus, we compared the expression profile of infant thymus ESTs with that of adult thymus ESTs (Adams et al. 1995). Compared to the adult thymus ESTs, the infant thymus ESTs contained a much higher percentage of ESTs that are involved in gene–protein expression and cell division–DNA synthesis (Fig.1). These differences in the EST profiles between infant and adult thymus' may reflect the active growth state of the infant thymus and involution of thymic tissue with aging. Despite the fact that it is an immune-related organ, the infant thymus appeared to express a proportion of ESTs for cell–organism defense similar to other organs (Adams et al. 1995) and twice as small as that of the adult thymus. This is likely to indicate that the relatively high levels of ESTs of gene–protein expression and cell division–DNA synthesis in the infant thymus are more closely related to rapid growth than immune function of thymus.

Regarding the EST frequency, eEF-1 a was the most frequently represented in the infant thymus followed by thymosin beta 4. eEF-1 a was abundantly expressed in cancer, infant tissue, and fetal tissue (Frigerio et al. 1995; Adams et al. 1993*b*; Hwang et al. 1995; Sudo et al. 1994). Ribosomal proteins also showed high redundancies. Abundant expression of ribosomal proteins was observed in several tissues and differentially expressed in different tissues (Kawamoto et al. 1996; Choi et al. 1995; Tanaka et al. 1996; Frigerio et al. 1995). Consequently, these data coincided with the expression profile of infant thymus ESTs.

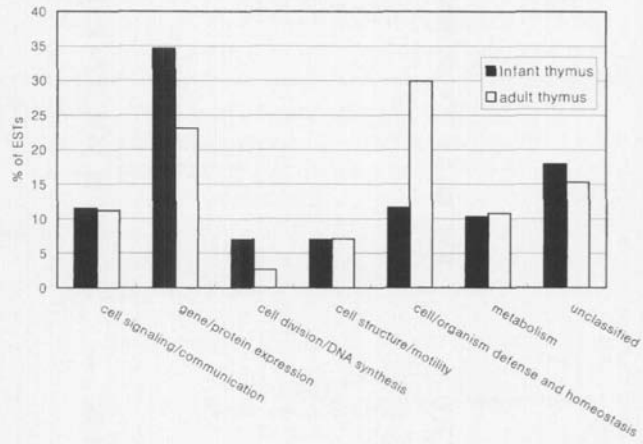
We found that some EST sequences corresponding to putative ribosomal proteins (RPs), RTH108 (RP L7a), KTH086 (RP S3), JTH138 (RP S27), and CJC060 (RP S4), showed striking similarities or identities with PLA-X, Fte-1 (v-fos transformation factor), metallopanstimulin, and the SCAR protein, respectively. Interestingly, the sequence of RTH100 (RP S11) was identical to the antisense sequence of isocitrate dehydrogenase. In 1996, Chan et al. reported that rat ribosomal protein L10 showed a 99% and 100% amino acid identity with a human QM and mouse QM, respectively, and suggested that their data was a presumptive example of the extraribosomal function of the ribosomal protein (Wool 1996).

When we assembled ESTs of unknown functions into a redundant EST group with a TIGR Assembler (Sutton et al. 1995), the group (ThyGR-15) had the highest redundancy (n = 7), which was equal to that of ribosomal L30. Thus, we can assume that the ThyGR-15 group can be considered as a moderately expressed transcript in a human infant thymus.

Chromosome mapping of unknown ESTs

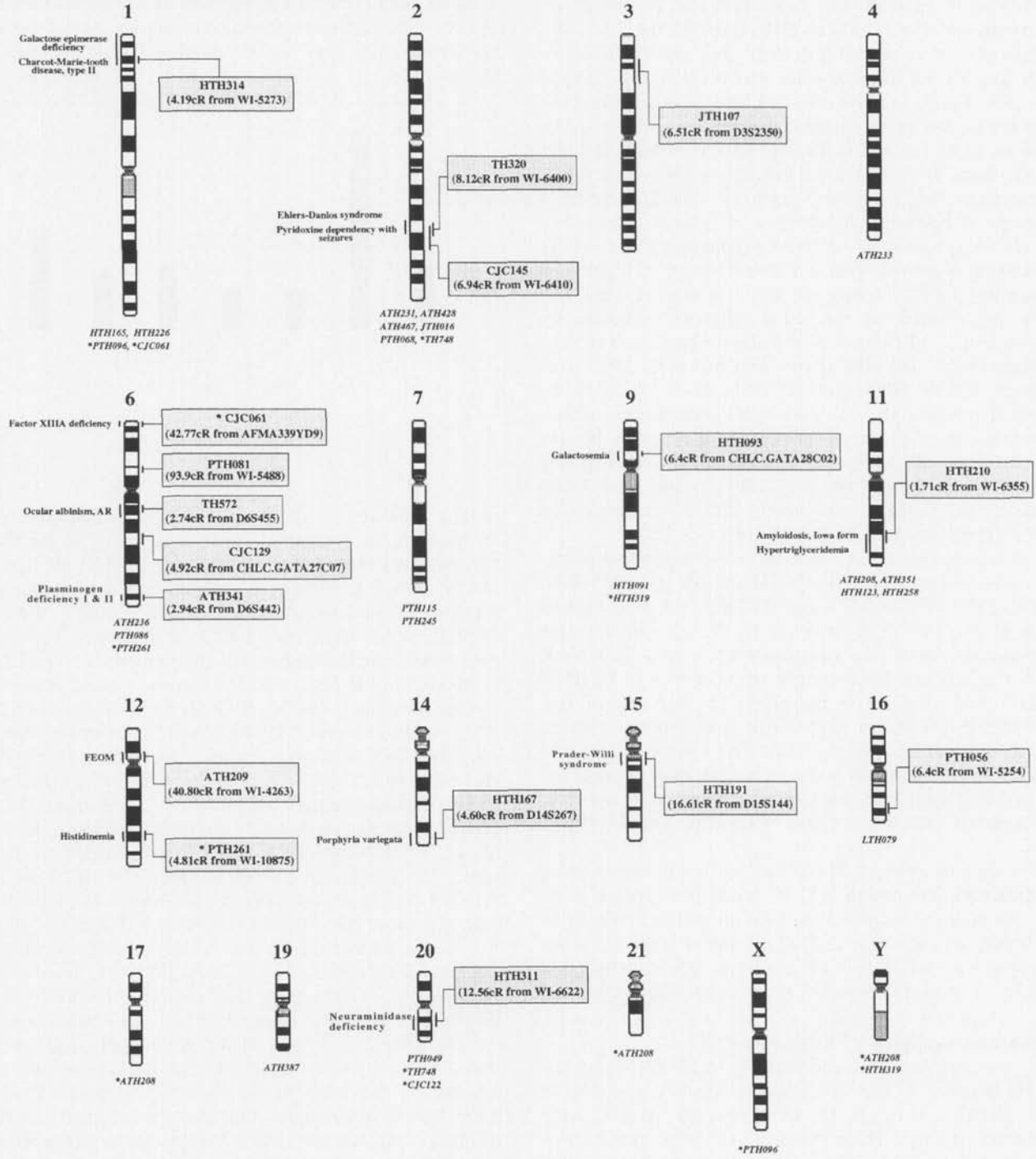
Forty-three clones were successfully localized onto a human chromosome as the first stage to identify gene-linked genetic disorders (Fig. 2). Of 43 clones, 35 (81.8%) were mapped onto a single chromosome, 7 (15.9%) onto 2 chromosomes, and 1 (2.2%) onto 4 chromosomes. Chromosome 6 was most predominantly hit by the infant thymus ESTs. To obtain a more refined map of novel ESTs, 17 clones were screened against a GENEBRIDGE 4 radiation hybrid panel using a PCR-based mapping method and then mapped through the statistical analysis of radiation hybrid mapping data. To survey the ESTs linked to disease loci, the RH maps of thymus ESTs were combined with the cytogenetic map available at the National Center for Biotechnology Information (NCBI).

Fig. 1. Comparison of expression profile of infant thymus ESTs with that of adult thymus ESTs. The 664 ESTs showing significant homology to known genes were classified into 6 groups based on putative functional categories (Appendix 1). After the adult thymus ESTs were extracted from the Genome Directory, the expression profile was compared by calculating the number of the genes of each group of both infant and adult thymus ESTs.



JTH107 placed 6.51 cR from D6S2350, which corresponds to 3p24 on the cytogenetic map, was linked to the thyroid hormone resistance locus. PTH261 placed 4.81 cR from WI-10875 (12q23), which is related to the histidinemia; ATH341 2.94 cR from D6S442 (6q26) related to the region of plasminogen deficiency I & II, plasminogen Tochigi disease, and dysplasmingogenic thrombophilla. TH572 was mapped 2.74 cR from D6S455 (6q14.1) on chromosome 6 (ocular albinism); TH320, 8.12 cR from WI-6400 (2q31.2–32.1) on chromosome 2 (Ehlers-Danlos syndrome, fibromuscular dysplasia of arteries, familiar aneurysm); HTH093, 6.4 cR from CHLCC GATAA 28C02 (9p13.2–13.3) on chromosome 9 (galactosemia); HTH210, 1.71 cR from WI-6355 (11q23.3) on chromosome 11 (hypertriglyceridemia, amyloidosis); HTH167, 4.6 cR from D14S267 (14q32) on chromosome 14 (porphyria variegata); HTH314, 4.19 cR from WI5273 (1p35) on chromosome 1 (Charcot-Marie-tooth disease, galactose epimerase deficiency); PTH056, 6.4 cR from WI-5254 on chromosome 16; ATH209, 40.8 cR from WI-4263 on chromosome 12 (FEOM), HTH191, 16.61cR from D15S144 on chromosome 15 (Prader-Willi syndrome), and HTH311, 12.56 cR from WI-6622 on chromosome 20 (neuraminidase deficiency, diabetes mellitus). The combination of EST sequences and map locations are currently in the process of providing numerous new candidate genes for those human pathologies that can be mapped genetically (Boguski and Schuler 1995). Several novel genes from the infant thymus were mapped on to disease loci, including the histidinemia loci, the plasminogen Tochigi disease loci, the Ehlers-Danlos syndrome, the hypertriglyceridemia, etc. Consequently, the radiation hybrid mapping of unknown ESTs from an infant thymus could facilitate the discovery of novel genes involved in genetic diseases or immune disorders. In summary, our cDNA approach enabled us to convert transcripts of a human infant thymus into 1223 expressed sequence tags which featured a high level of genes related to

Fig. 2. Chromosomal localization of human infant thymus ESTs. The shadow boxes represent the locations of ESTs obtained by radiation hybrid mapping. Diseases linked to these locations are indicated on the left side of each chromosome. When an EST is located on two or more chromosomes, it is represented by an asterisk. Chromosome assignments using a somatic cell hybrid panel are indicated on the bottom of each chromosome with an italic letter.



the active cell cycling of thymocytes. The infant thymus ESTs can also be used as useful mapping markers for radiation hybrid mapping in addition to providing the expression information of active genes in the human infant thymus.

Acknowledgements

This work was supported by grants from the Ministry of Education, the Ministry of Science and Technology of Korea, and the Canadian Genome Analysis and Technology Program. L.-C.Tsui is a Senior Scientist of the Medical Re-

search Council of Canada, holds the Sellers Chair in Cystic Fibrosis Research, and is a Howard Hughes International Scholar. We would like to thank J.H. Lee, H.C. Lee, J.P. Jeon, J.C. Choi, D.S. Park, and W.K. Cho for their assistance in sequencing.

References

Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C., and Venter, J.C. 1993*a*. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nature Genet. **4**: 373–380.

Adams, M.D., Kerlavage, A.R., Fields, C., and Venter J.C. 1993*b*. 3,400 new expressed sequence tags identify diversity of transcripts in human brain. Nature Genet. **4**: 256–267.

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Kelley, J.M. Jr., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palanques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C. 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature, **377**: 3–17.

Affara, N.A., Bentley, E., Davey, P., Pelmear, A., and Jones, M.H. 1994. The identification of novel gene sequences of the human adult testis. Genomics, **22**: 205–210.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403–410.

Berry, R., Stevens, T.J., Walter, N.A.R., Wilcox, A.S., Rubano, T., Hopkins, J.A., Weber, J., Goold, R., Soares, M., And Sikela, J.M. 1995. Gene-based sequence-tagged sites (STSs) as the basis for a human gene map. Nature Genet. **10**: 415–423.

Boguski, M.S., and Schuler, G.D. 1995. ESTablishing a human transcript map. Nature Genet. **10**: 319–371.

Chan, Y.L., Diaz, J.J., Denoroy, L., Madjar, J.J., and Wool, I.G. 1996. The primary structure of rat ribosomal protein L10: Relationship to a jun binding protein and to a putative Wilms' tumor suppressor. Biochem. Biophy. Res. Commun. **225**: 952–956.

Choi, S.S., Yun, J.W., Choi, E.K., Cho, Y.G., Sung, Y.C., and Shin, H.S. 1995. Construction of a gene expression profile of a human fetal liver by single-pass cDNA sequencing. Mamm. Genome, **6**: 653–657.

Collins, F.S. 1995. Positional cloning moves from perditional to traditional. Nature Genet. **9**: 347–350.

Drwinga, H.L., Toji, L.H., Kim, C.H., Greene, A.E., and Mullivor, R.A. 1993. NIGMS human/rodent somatic cell hybrid mapping panel 1 and 2. Genomics, **16**: 311–314.

Frigerio, J.M., Berthezene, P., Garrido, P., Ortiz, E., Barthellemy, S., Vasseur, S., Sastre, B., Seleznieff, I., Dagorn, J.C., and Iovanna, J.L. 1995. Analysis of 2166 clones from a human colorectal cancer cDNA library by partial sequencing. Hum. Mol. Genet. **4**: 37–43.

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Meg, J., Trevaskis, E., Underwood, K., Wohldman, P., Waterston, R., Wilson, R., and Marra, M. 1996. Generation and Analysis of 280,000 Human expressed sequence tags. Genome Res. **6**: 807–828.

Hwang, D.M., Fung, Y.W., Wang, R.X., Laurenssen, C.M., Ng, S.H., Lam, W.Y., Tsui, K.W., Fung, K.P., Waye, M., Lee, C.Y., and Liew, C.C. 1995. Analysis of expressed sequence tags from a fetal human heart cDNA library. Genomics, **30**: 293–298.

Kawamoto, S., Matsumoto, Y., Mizuno, K., Okubo, K., and Matsubara, K. 1996. Expression profiles of active genes in human and mouse livers. Gene, **174**: 151–158.

Kirkness, E.F. 1996. Assessment of human gene diversity and expression pattern using expressed sequence tags. Essays in Biochem. **31**: 1–9.

Lamerdin, J.E., Athwal, R.S., Kansara, M.S., Sandhu, A.R., Patanjali, S.R., Weissman, S.M., and Carrano, A.V. 1995. Chromosomal localization and expressed sequence tag generation of clones from a normalized human adult thymus cDNA library. Genome Res. **5**: 359–367.

Lee, N.H., Weinstock, K.G., Kirkness, E.F., Earle-Hughes, J.A., Fuldner, R.A., Marmaros, S., Glodek, A., Gocayne, J.D., Adams, M.D., Kerlavage, A.R., Fraser, C.M., and Venter, J.C. 1995. Comparative expressed-sequence-tag analysis of differential gene expressed profiles in PC-12 cells before and after nerve growth factor treatment. Proc. Natl. Acad. Sci. U.S.A. **92**: 8303–8307.

Liew, C.C., Hwang, D.M., Fung, Y.W., Laurenssen, C., Cukerman, E., Tsui, S., and Lee, C.Y. 1994. A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. Proc. Natl. Acad. Sci. U.S.A. **91**: 10 645 – 10 649.

Okubo, K., Itoh, K., Fukushima, A., Yoshii, J., and Matsubara, K. 1995. Monitering cell physilogy by expression profiles and discovering cell type-specific genes by compiled expression profiles. Genomics, **30**: 178–186.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular cloning: A laboratory manunual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sudo, K., Chinen, K., and Nakamura, Y. 1994. 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library. Genomics, **24**: 276–279.

Sutton, G.G, White, O., Adams, M.D., and Kerlavage, A.R. 1995. TIGR Assembler: A new tool for assembling large shotgun sequencing projects. Genome Sci. & Tech. **1**: 9–19.

Tanaka, T., Ogiwara, A., Uchiyama, I., Takegi, T., Yazaki, Y., and Nakamura, Y. 1996. Construction of a normalized directionally clones cDNA library from adult heart and analysis of 3040 clones by partial sequencing. Genomics, **35**: 231–235.

Wool. I.G. 1996. Extraribosomal functions of ribosomal proteins. Trends Biochem. Sci. **21**: 164–165.

Appendix 1. Infant thymus ESTs matched to known genes in the database.

Cell Division (6.8%)
BTG1 gene (2)
BTG2 gene
Cellular oncogene c-fos
Cyclin D3 (2)
Cyclin protein
Deoxynucleotidyltransferase
DNA synthesis inhibitor
Histone H1B, H2A, H2B, H4 genes
Histone H2A, F (2)
Histone H2A.2
Histone H3.3 (4)
Histone H3.3, B
HMG (High mobility group) 2
Inhibitor of apoptosis protein 2
MA-3(apoptosis-related gene) (6)
Nonhistone protein HMG1 mRNA
Notch3
Nucleosome assemble protein 2 (2)
Oncogene ets-2
Oncogene PTI-1
Prothymosin alpha (6)
Proto oncogene c-sis
Proto oncogene Ets-1
SET nuclear phosphoprotein (3)
Tis11d
TRE5 sequence (tre oncogene)

Cell Signaling/Communication (11.4%)
14–3–3 eta chain
14–3–3, beta
14–3–3, epsilon
ADP-ribosylation factor 1 (2)
Androgen receptor associated protein 24
Aurora-related kinase 2 (ARK 2)
B-cell growth factor
Cadherin
Calcyclin/PRA
Calmodulin (CALM1) gene
Calmodulin dependent protein phophatase
Calmodulin homologue (NB-1)
Calmodulin-I mRNA (2)
Calmodulin-like pseudogene
cdc2/CDC28 like protein kinase
Coupling protein G(s) alpha subunit
Cytokine receptor CRFB4
ERK activator kinase
Glucocorticoid receptor
gpISG20
Grb3–3
GTP binding protein beta subunit-like protein
GTP-binding protein (rhoA) (2)
hTGR
ICLN protein
IL2 receptor gamma
Inerferone-gamma IEF
Laminin receptor (4)
MAGHO mRNA
Oxytocin receptor
Parathyroid hormone
PI-3 kinase
Plexin 2 (3)
Pre-B cell stimulating factor homologue (2)
Protein kinase (rho-associated)
Protein kinase 63 kDa related to rat ERK3
Protein kinase, TTK
Protein tyrosine phosphatase (3)
Rap1B mRNA
Ras-related C3 botulinum toxin substrate(rac)
Renin
Rho GDP dissociation inhibitor 2
Rod outer segment membrane protein 1 (Rom1)
Serine/threonine kinase receptor 2
Stathmin (2)
Thymosin beta 10 (2)
Thymosin beta 4 (13)
Tumor necrosis factor (4)

Cell Structure/Mobility (7.1%)
Actin, beta (5)
Actin, cytoplasmic
Actin, gamma (2)
Ankyrin binding glycoprotein-1 (4)
Capping protein, alpha subunit 1 (4)
CENP-F kinetochore protein mRNA
Clathrin coat assembly protein-like
Cofilin (2)
COP(36 kDa coatomer), epsilon
Cytokeratin
Dynamitin (2)
Elastin (4)
Fibrinogen gamma chain
Heterochromatin protein
Keratin (2)
Kid (kinesin-like DNA binding protein)
Kinectin (motor kinesin binding protein)
Laminin B1
Matrin 3
Moesin-related mRNA sequence
Moitotic kinesin like protein
Myosin light chain, non-muscle, alkali
Profilin 1
RIG-like 14–1 mRNA
SPARC/osteonectin
Tubulin alpha (3)
Tubulin beta (2)

Cell/Organism Defense (11.8%)
27 kDa heat shock protein (2)
28 kDa heat shock protein
APEX nuclease
Beta-2 microglobulin (2)
CAMPATH-1 (3)
CD1 antigen R2, MHC-related (6)
Chitinase precursor
Coagulation factor IX gene
Complement C2 precursor
Complement C3
Copper transport protein HAH1
Csa-19 (4)
dnaJ like protein
Ferritin heavy chain (2)
Ferritin light chain
Heat shock cognate 70 kDa protein
Heat shock cognate 89 kDa protein (2)
Heat shock protein 90
HLA classI locus C
HLA classII SB beta-chain
HLA-A2 classI antigen (3)
HLA-B27
HLA-C 1
HLA-DQA2
HLA-DR, alpha, heavy chain
HLA-Dw12 DQ beta
HLA-E
HLA-SB beta
HLA-xrp gene
hMSH6 gene, exon 6–10
Ig associated invariant gamma chain
Ig gamma 1 chain C region
Ig J chain gene
Ig kappa chain VK-1
Ig kappa light chain
Ig lambda light chain (3)
Ig mu heavy chain C-region
RAG-1 (7)
RAG-2 (2)
RAG-3
rGSTK1–1
RING3 gene
T-cell receptor beta chain (D-J-C) (2)
T-cell receptor beta chain loci
T-cell receptor beta-chain (C region) (3)
T-cell receptor, T3 epsilon
Transferrin receptor
VH4–34 (isotype switch)

Gene/Protein Expression (34.6%)
Acidic ribosomal phosphoprotein P0
Acidic ribosomal phosphoprotein P1
Acidic ribosomal phosphoprotein P2
Antisecretory factor 1
Bat2 gene
CAAT box binding transcription factor (CTF)
Calpain (calcium activated neutral protease)
Cathepsin B
Cathepsin L (2)
CCAAT/enhancer binding protein delta
Cystein-rich sequence binding protein
DNA binding protein(neurodap 1)
DNA binding protein, KET
DNA binding protein, TAXREB107
eEF-1 alpha (20)
eEF-1 beta (2)
eEF-1 gamma related protein
Enhancer of split m9/m10
FKBP (FK506-binding protein)

Appendix 1 (*continued*).

G17 gene
Helicase (218kD Mi-2 protein) (2)
Helix-loop-helix-leucine zipper (SREBP-1)
Heterogenous ribonucleoprotein homolog (3)
Hlark mRNA
hnRNP core protein A1
hnRNP D (2)
hnRNP-C
hnRNP-C2
hnRNP-E1 (2)
hnRNP-H
hnRNP-K
MAD-3 encoding IkB-like acitivity
Myocyte specific enhancer factor (2)
Nuclear p68 protein (3)
Plasminogen activator inhibitor-1
Polyadenylation specificity factor
Prolyl 4 hydroxylase, alpha subunit (4)
Proteasome component C9
Proteasome related gene (LMP 2)
Proteasome subunit HC8 (2)
Proteasome subunit p55
Proteasome-like subunit
putative RNA binding protein
Ribosomal protein L1
Ribosomal protein L10 (3)
Ribosomal protein L17
Ribosomal protein L21 (2)
Ribosomal protein L23 (5)
Ribosomal protein L24
Ribosomal protein L26 (4)
Ribosomal protein L27 (2)
Ribosomal protein L28 (2)
Ribosomal protein L3 (2)
Ribosomal protein L30 (7)
Ribosomal protein L32
Ribosomal protein L35 (4)
Ribosomal protein L36
Ribosomal protein L37 (3)
Ribosomal protein L37a
Ribosomal protein L38
Ribosomal protein L4
Ribosomal protein L41, yeast homologue
Ribosomal protein L44, yeast homologue
Ribosomal protein L5 (3)
Ribosomal protein L6
Ribosomal protein L7 (2)
Ribosomal protein L7a (3)
Ribosomal protein L9 (4)
Ribosomal protein S10 (3)
Ribosomal protein S11 (6)
Ribosomal protein S14 (3)
Ribosomal protein S16
Ribosomal protein S17 (2)
Ribosomal protein S18 (4)
Ribosomal protein S20 (4)
Ribosomal protein S24 (2)
RIbosomal protein S25 (2)
Ribosomal protein S26 (2)
Ribosomal protein S27 (2)
Ribosomal protein S28, yeast homolog
Ribosomal protein S29
Ribosomal protein S2a
Ribosomal protein S3a (6)
Ribosomal protein S4 (2)
Ribosomal protein S5
Ribosomal protein S6
Ribosomal protein S7
Ribosomal protein S8 (2)
Ribosomal protein S9
Ribosomal protein YL30
RNA binding protein, HU-K4
RNA helicase #46, ATP-dependent
RNA polymerase I
Seb4
Sec61(protein transport protein)
SOX-4 (Sry-like HMG box protein)
Splicing factor mRNA
Splicing factor SRp75
Splicing factor, SF1-HL1 isoform
SW1/SNF complex 155 kDa subunit (BAF155)
TCF-1 (T-cell specific transcription factor)
TFIID 55 kDa subunit
TFIIB mRNA
Transcription factor (Staf 50)
Transcription factor AF-1p
Transcription factor, ASF
Transglutaminase type I
Translation factor Sui1 homolog (3)
Translation initiation factor p40 subunit
Translation initiation factor 3
Translation initiation factor, Eif4g2
Translation repressor NAT1 mRNA (7)
U21.1 hnRNP
Ubiquitin (2)
Ubiquitin conjugating enzyme (ubc4)
Ubiquitin conjugating enzyme E2 (2)
USF-2(upstream stimulatory factor-2)
Wilm's tumor-related protein (6)
Zinc finger protein RING (RZF)
Zinc finger protein, HPF1 (3)
Zinc finger protein, ZNF

Metabolism (10.4%)
2′ oxidoglutarate dehydrogenase
Adenosin Deaminase (ADA)
Adenosin monophosphate deaminase1
Adenosine triphosphatase
Aldehyde dehydrogenase
Alpha enolase gene (2)
Apolipoprotein A-IV gene
ATP synthase, beta subunit (2)
ATP synthase, gamma subunit (2)
ATP synthase, subunit a
ATP synthase, subunit c (2)
ATP synthase, subunit d
ATP synthase, subunit f
ATP synthase, subunit g
ATPase, H+ transport, vacuolar (2)
ATPase, NA+ / K+ transport, alpha subunit
ATPase, proton subunit D, vacuolar (3)
Co-beta glucosidase(proactivator)
CYP3B6 gene cryptic exon (cytochrom p450)
Cytochrome C oxidase polypeptide I (COX I)
Cytochrome C oxidase subunit IV (COX 4)
Cytochrome c oxidase subunit VIa
Cytochrome C oxidase subunit Vib
Cytochrome C oxidase subunit VIIb
Deoxycytidine kinase (3)
Enyol CoA dehydrogenase
Glycealdehyde-3-phosphate dehydrogenase
GTP cyclohydrolase I
Iduronate sulphate sulphatase (IDS)
Importin beta subunit
Inosine monophosphate dehydrogenase type II
Lactate dehydrogenase A
Lactate dehydrogenase B
Liver phosphatase 2A (2)
Lysophosphatidic acid acyltransferease-alpha (2)
Malate dehydrogenase
mGLT(glutamate transporter)-1 mRNA
Monocarboxylate transporter 1 (2)
Muscle glycogen phosphorylase
N-acetylglucosaminyltransferase III DNA,
NAD(P)+ dependent malic enzyme mRNA (2)
NADH:ubiquitin oxidoreductase
Nm23 protein (nucleoside diphosphate kinase)
Oligoadenylate synthetase
Phospholipase A2
Phosphotyrosyl phosphatase activator
Purin 5′-nucleotidase
r-aminobutylaldehyde dehydrogenase
Ribonuclease 6 precursor
Ribonucleoside diphosphate reductase M1
Ribonucleotide reductase M2
Secretory carrier membrane protein
sterol carrier protein (3)
Ubiquinal-cytochrome C reductase

Unclassified (17.9%)
23KD highly basic protein(Ig gamma)
AICL(activation induced c-type lectin)
Ataxia telagiectusa (ATM) gene
Autoantigen DFS70
BBC1(breast basic conserved) protein
Beta amyloid protein(APP) gene
Brain 0–44 mRNA
BRCA2(breast cancer susceptibility)
CG1 protein
CLE7 (*G.gallus*) (3)
COX7RP
CpG DNA (5)
Dead box, x isoform(DBX) alternative transcript 2
Down syndrome region
DT1P1A10, CAG repeat
DT1P1A2, CAG repeat
Duffy blood group antigen Fya-b+

Appendix 1 (*concluded*).

E25 mRNA (2)	Hypothetical protein, KIAA0161 gene	Nuclear Factor IV
EB virus small RNA associated protein (EAP)	Hypothetical protein, KIAA0217 gene	Nucleolar phosphoprotein B23(nucleophomin)
EBI1-ligand chemokine (ELC)	Hypothetical protein, KIAA0220 gene	Nucleolar protein P120(ribonucleoprotein)
GP91-PHOX gene promoter region	Hypothetical protein, KIAA0252 gene	Ocular melanoma associated antigen
GRSF-1(G-rich sequence factor-1)	Hypothetical protein, KIAA0261 gene	ORF mRNA
GT233 mRNA	Hypothetical protein, KIAA0323 gene	p38–2G4
HE5 (CDw52-like epididymal protein) (5)	IEF7422	Proliferation-associated protein pag
Heat shock protein hsp70	Interferon inducible gene (3)	Prostatein C3 subunit gene
Hepatocyte nuclear factor 1 promoter	JKA10 mRNA induced upon T-cell activation	putative nuclear pore complex protein(Npap60)
HepG2 3' region cDNA	Leucine rich protein	PxF protein (3)
HLA class III containing NOTCH4 gene (2)	LLReps3	Retinoblastoma produt binding protein (RBQ-1)
Housekeeping	LTG9/MLLT3, C-terminal	Retroposon SINE-R11 (2)
Human farmilial Alzheimer's disease gene	Lymphoid restricted membrane proein, Jaw1	S153 clone mRNA
Human mRNA for 3–7 gene product	Lysosomal-associated membrane glycoprotein1	Serine-rich neutrophil protein
Huntington's disease region	Lysosome associated membrane protein2	Split hand/split foot 1 (DSS1) (3)
Hypothetical protein, KIAA0034 gene (3)	Melanoma associated antigen	T-cell surface glycoprotein E2
Hypothetical protein, KIAA0037 gene	Metallopanstimulin 1	Tera mRNA
Hypothetical protein, KIAA0040 gene	MGC-24 (PNA-binding protein)	Thymocyte antigen CD1b
Hypothetical protein, KIAA0040 gene	MHC class I HLA-C 1 gene	Transformation related protein (2)
Hypothetical protein, KIAA0045 gene	MHC class I HLA-C 1 locus C heavy chain	Transmembrane protein (4)
Hypothetical protein, KIAA0083 gene	Na+/H+ exchang regulatory co-factor	Unknown antigen
Hypothetical protein, KIAA0101 gene	NAC, alpha (4)	Unknown protein within p53 intron 1
Hypothetical protein, KIAA0107 gene	NifU-like protein (hNifU) mRNA	X11gene
Hypothetical protein, KIAA0128 gene	Ninein (centromal protein)	XG blood group
Hypothetical protein, KIAA0159 gene	NK-tumor recognition molecule-related protein	ZFM 1 protein alternatively spliced product

Note: Human infant ESTs matching to 416 distinct known genes were divided into 7 groups according to the putative function (Adams et al. 1995). Numbers in parentheses indicate the frequency of the ESTs. In the case of non-human matched ESTs, the organism is indicated in parentheses.