DOE/ER/61232--2

DE93 003066

TECHNICAL PROGRESS DURING THE CURRENT AND PREVIOUS BUDGET PERIODS

A. Summary

During this project year, we have tested and implemented several protocols to increase productivity for mapping EST sequences to human chromosomes. These protocols include (1) adopting PRIMER (8), which permits utilization of batch files, as the standard software for PCR primer design; (2) adding a human 21-only cell line (S. Antonarakis, personal communication) to the NIGMS panel #1 to improve discrimination in discordancy analyses involving chromosome 21, (3) adding a monochromosomal hybrid panel (NIGMS Panel 2) to facilitate chromosome assignment of sequences that are amplified from more than 1 chromosome; (4) combining the products of multiple PCR reactions for electrophoretic analysis (pseudoplexing); (5) routinely multiplexing PCR reactions; and (6) automating data entry and analysis as much as possible.

We have applied these protocols to assign an overall total of 132 human brain cDNA sequences to individual human chromosomes. PCR primers were designed from expressed sequence tags (ESTs) and tested for specific amplification from human genomic DNA. DNA was then amplified using DNA from somatic cell hybrid mapping panels as templates. The amplification products were identified using an automated fluorescence detection system. Chromosomal assignments were made by discordancy analysis. The localized cDNAs include 2 for known human genes, 2 that map to 2 different human chromosomes, and 25 for cDNAs matching existing database records.

B. Introduction

Genes expressed in human tissues are rapidly being identified by partially sequencing cDNA clones (1-4). While some of the genes identified in this way exhibit nucleic acid or protein sequence similarity to known genes in humans or other organisms, most are previously unreported.

Several laboratories are analyzing the cDNA sequence data and using PCR and somatic cell hybrid mapping panels to localize the sequences to individual human chromosomes (5-7). The mapping efforts continue to lag approximately an order of magnitude behind the sequencing projects. The focus of our approach is to maximize sensitivity and throughput by applying the ABI 373A sequencing machine configured with GeneScanner software to analyze fluorescently-tagged PCR products (6-7). The instrumentation permits the combination of high resolution of multiple products separated by denaturing polyacrylamide gel electrophoresis with the four color discrimination of the fluorescence detection system. Because primers can be labeled with different fluorescent dyes, products can be identified by both size and dye label. We have therefore combined multiple pairs of PCR primers in single PCR reactions (multiplexing) and

combined independently amplified products to a single electrophoretic lane (pseudoplexing) to increase throughput.

C. Methods

Figure 1 is a flow chart for our EST mapping strategy. Sequences known to have a low probability of containing coding sequences were provided electronically from the Venter laboratory. Another set showing relatedness to sequence records but not matching human sequences exactly was downloaded from dbEST. Primers were designed from these sequence records using PRIMER (8) with the parameters of 50% GC; primer size of 18-22 nucleotides; primer Tm between 55°-59°C; PCR product size between 80-150 bp, and sequence range within the first 300 bp of each EST sequence. To expedite processing, all source sequence files were analyzed without determining the correspondence between the predicted PCR product and any potential coding sequence. Thus, for the set of ESTs that were selected for mapping because of similarities to previously reported sequences, the amplified product was not restricted to lie either within or without coding regions.

One of each primer pair was synthesized with 5′ Aminolink II and labeled with a fluorescent dye using procedures provided by ABI (9). Template DNA (50 ng) was amplified in the presence of 40 ng of each primer in a total volume of 15 μl. Other components were as described in the Perkin Elmer Cetus GeneAmp Kit protocol (P/N N801-0043). The amplification conditions were: 95°C, 5 min; followed by 25 cycles of 94°C, 1.4 min., 55°C, 2 min., 72°C, 2 min., and a final incubation at 72°C, 10 min. One μl of the PCR product was pooled with 1 μl of fluorescently-labelled internal lane size standard [pGEM3Z sequencing ladder with only ddTTP chain termination products, or GS-2500 (ABI)] and 2 μl deionized formamide. The sample was denatured at 95°C for 2 minutes before loading onto a 6% polyacrylamide denaturing gel (8 M urea). Data from the electrophoretic analyses were collected on an ABI 373A automated sequencer configured with ABI 672 GeneScanner software. Sizes of amplification products were determined relative to the internal lane standard (10) by the GeneScanner software.

If a primer pair failed to give an amplification product, or failed to give a product that could be used to make a chromosome assignment, another oligonucleotide was synthesized that could be used with the the dye-labeled primer. In approximately 20% of the cases, no such oligonucleotide could be identified, and a complete new pair was identified from the EST sequence. If no assignment could be made after 2 sets of primers had been tested, analysis of that EST was terminated.

DNA from several human/rodent hybrid mapping panels was used (NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2, Coriell Institute for Medical Research, Camden, NJ; PCRable DNA from BIOS Corporation, New Haven, CT). The NIGMS panel 1 was

supplemented with a human 21-only cell line (S. Antonarakis, personal communication).

Sequences were assigned to a chromosome when the discordancy value was less than or equal to 9%. Cell lines reported as containing a particular chromosome in fewer than 12% of metaphases were not included in the discordancy analysis.

D. Results and discussion

1. *Chromosome assignments*. Chromosomal assignments, accession numbers of matching database records, and primer sequences for 132 ESTs are presented in Table 1. Except for 2 cases (EST00884 and EST00858) in which the primers direct amplification across an intron, the products from human genomic DNA were within 5 base pairs of the size predicted by the cDNA sequence.

Figure 2 compares the distribution of chromosome assignments of these ESTS to the percentage of the genome represented by each chromosome. While 132 sequences is too small a set on which to base significant conclusions, it does appear that there is a lack of correlation between chromosome size and localization of sequences expressed in the brain. This observation is consistent with that of McKusick for monochromosomal disease-causing genes (11).

2. *Current status*. Table 2 summarizes our current progress for the set of sequences being tested. Of 626 ESTs analyzed, 198 could not be used to design primers that met the criteria described in the Methods section. For the remaining 428 sequences, analysis of 154 is considered complete, with 132 assigned to chromosomes and 22 not assigned after testing of 2 sets of primer pairs. According to our criteria, no additional attempt will be made to map these sequences.

Testing of 217 primer pairs derived from 168 ESTs has been completed. One hundred six sequences were mapped from the first set of 107 primers designed (one EST was mapped from 2 sets of primers), and 62 required redesign. Of the 48 secondary pairs that have been tested, 26 were useful in making chromosome assignments. The 22 sequences that could not be assigned to chromosomes fell into the categories of inconclusive discordancy analyses (1/22), no amplification of expected human product (17/22), and indistinguishable human, mouse, and hamster products (4/22).

3. *Multiplexing*. Several approaches to multiplexing PCR reactions have been reported. Usually, conditions and primer combinations are empirically optimized to permit successful simultaneous amplification of the same set of multiple primer pairs for repeated use in single reactions (12-13). For our mapping application, taking the time to optimize PCR conditions for selected combinations of primer pairs would decrease, not

enhance, throughput. In our protocol, a primer pair designed from a cDNA sequence is typically used in only two sets of PCR reactions. The first set of proving reactions, used to determine success in amplifying a product from human DNA distinct from potential rodent background, has as template either total human, total mouse, or total hamster genomic DNA. The second set of mapping reactions, used to determine the pattern of amplification based on human chromosome content, has as template DNA samples of either human/mouse or human/hamster hybrid cell lines of known human karyotype. Optimizing multiplex PCR conditions for each primer pair for only two reactions would thus require more effort than mapping the cDNA sequences using only one pair per reaction. Instead, we design primers using narrowly defined ranges of primer and product parameters and determine success using standard reaction conditions (6) in primer proving reactions. We then combine primer pairs for multiplex mapping reactions considering only color of the fluorescent dye label and amplified product size. Of the 132 mapped sequences in this report, 69 were mapped as a result of multiplexed mapping PCR reactions and 85 were analyzed in pseudoplexed gels.

Seventy-one of the primer pairs used to map the sequences described in this report were tested against mapping panels in duplex PCR reactions (4 primers designed to detect 2 different genomic sequences in the same tube). The primers were selected for grouping only on the basis of the dye used for labeling, the lack of background amplification in with the rodent genome in the hybrid cell panel being tested, and the product size. Primer sequences were not tested for interpair sequence complementarity. Of these, only 2 did not permit an assignment to chromosomes based on the multiplexing results and were mapped as a result of single reactions. Of the sequences that could not be mapped, 5 were tested in multiplexed reactions and only one of those included a multiplexing failure. This level of success suggests that higher orders of multiplexing could continue to increase throughput. Experiments are underway to test this.

4. *Sequence database matches.* Several of the ESTs we mapped had previously been reported to match sequence database records (1,2). No new matches were revealed after checking dbEST (September, 1992). We are still in the process of examining these ESTs in more detail to determine whether the assignment of the EST to a human chromosome also permits a new assignment of a human gene product. For a discussion of this analysis, see the draft of (7) presented in the Appendix.

5. *Data management.* The complex data analysis demanded by this project has been simplified by the development of a PC-based database and application software that provide the following functions: automated entry to the database of EST sequences to be analyzed; automated generation of a batch file from the database to serve as input for the execution of PRIMER;

automated ordering of primers to be tested; automated discordancy analysis; data-entry screens that can be easily used by laboratory personnel to record information on labeling reactions, PCR reactions, and electrophoretic analyses; and a suite of automated reports, including electronic files for transferring data to external databases.

REFERENCES

1. Adams, M.D., Kelly, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, O., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991) *Science* **252**: 1651-1656.
2. Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelly, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992) *Nature* **355**: 632-634.
3. Wilcox, A.S., Khan, A.S., Hopkins, J.A. and Sikela, J.M. (1991) *Nucleic Acids Res.* **19**: 1837-1843.
4. Gieser, L. and Swaroop, A. (1992) *Genomics* **13**: 873-876.
5. Polymeropoulos, M.H., Xiao, H., Glodek, A., Gorski, M., Adams, M.D., Moreno, R.F., Fitzgerald, M.G., Venter, J.C., Merril, C.R. (1992) *Genomics* **12**: 492-496.
6. Durkin, A.S., Maglott, D.R., Nierman, W.C. (1992) *Genomics*, In Press.
7. Durkin, A.S., Nierman, W.C., Maglott, D.R. (1992) in preparation.
8. Lincoln, S.E., Daly, M.J., Lander, E.S. (1991) PRIMER: A computer program for automatically selecting PCR primers. Vers 0.5. @1991, the Whitehead Institute for Biomedical Research.
9. Applied Biosystems User Bulletin, Synthesis of fluorescent dye-labeled oligonucleotides for use as primers in fluorescent-based DNA sequencing, Issue **11**, (1989).
10. Carrano, A.V., Lamerdin, J., Ashworth, L.K., Watkins, B., Branscomb, E., Slezak, T., Raff, M., De Jong, P.J., Keith, D., McBride, L., Meister, S., Kronick, M. (1989) *Genomics* **4**:129-136.
11. McKusick, V.A. (1988). The Morbid Anatomy of the Human Genome, Howard Hughes Medical Institute.
12. Chamberlain, J. S., Gibbs, R. A., Rainer, J. E., Nguyen, P.-N., and Caskey, C. T. (1988). *Nucleic Acids Res.* **16**: 11141-11156.
13. Gibbs, R.A., Nguyen, P., Edwards, A., Civitello A.B., Caskey, C.T. (1990). *Genomics* **7**: 235-244.

DISCLAIMER

FIGURE 1.

EST MAPPING STRATEGY

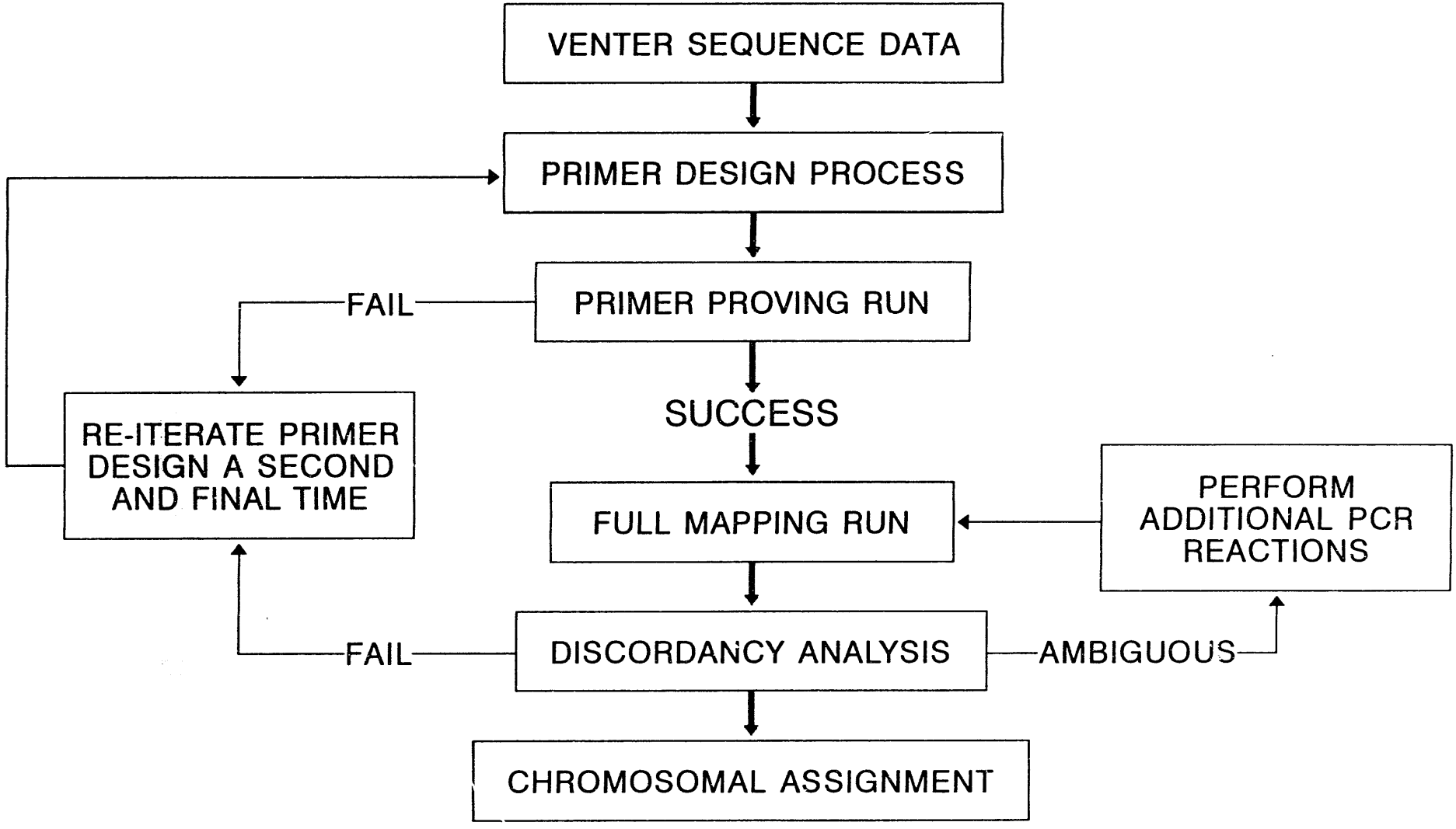
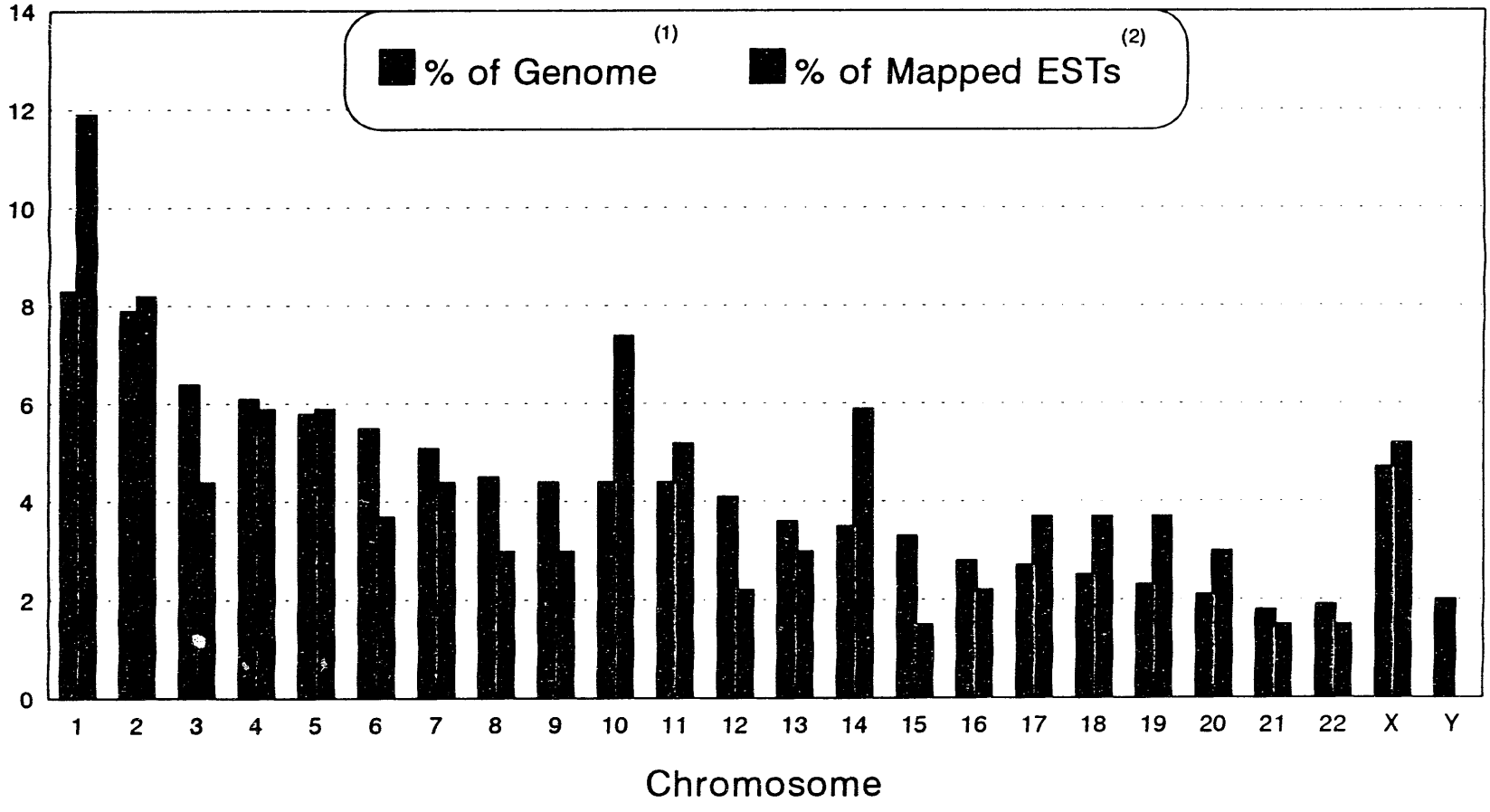


FIGURE 2. DISTRIBUTION OF MAPPED EXPRESSED SEQUENCE TAGS (ESTs) WITHIN THE HUMAN GENOME

(As of 29 October 92)



(1) *Science* **250**:239-244, 1990.

(2) All ESTs were derived from human brain cDNA libraries. (N=133)

TABLE 1. SUMMARY OF MAPPED ESTS

CH	*EST#*	*Locus/ DB match*[1]	*Sequence*	*Sequence*
1	EST00100		23T ACCATACACAACACTCCAGAGC	23B CTTCCATCCCATTTAGCC
1	EST00503		40S CAAAGACTCCTCTGCTAGAACC	40A AAAGAGCTACGAGGCAGATG
1	EST00552		51S CATGTTGTAGCTCATCGCAG	51A AGGATGGCAATACAACGG
1	EST00745		73S GAATAGACACTGGGTATCCTCC	73A CATGAATTCTCTCAGGGGTC
1	EST00751		74S CTGCCAGAACCCAAAACTAC	74A AGAAATGCGTGTAGAGGGAG
1	EST00767		76S TGAAACCTTTGTCCTGGG	76A TCCATTGAGTCTGAGTGTGG
1	EST00883		30S GGATGCCTTTAATCAAGCCTGG	30B TGACAAAAAGCCTTGGGC
1	EST00885		180S CTGGAACCTGTGAGTACTGTTC	180A TACTGGTGCTGCCATAGTCT
1	EST00903		117S AACCACCATTGCTGCTGT	117A GTGGGTTAAGGGCATGTAGA
1	EST00904		118S GTGCATTGTGGCCTGTTACT	118A CGCAAGCAGCAATCATAG
1	EST01041	P:B35308	139S AGAAGCCCTTTCCTGTTG	139A GTTCAAGGTCTTGGTGCTCA
1	EST01676	G:M20866	187S GGGTCATCTCCACAACATTC	187A CATGCTGCCAACTTCTAACC
1	EST01764	P:A36427	159S TAGGCAAAGGAAAGGTGG	159A AAAGCAGAAGCAGATCGG
1	EST01919		98S GGCATCAGTTTCCTCATCTG	98A AAAGTCACAACCTTGGGC
1	EST02447	G:M37744	174S TCCAGGAAAGAGACCTCCTA	174A GTTGTGATATGCCACGTCAG
1	EST02679		10S CAGCAAACCTGGAATGAAAGGG	10A TAAGAAGGGGATGAGTCG
2	EST00587		57S GACCAGCCTAGATTCCATTG	57A CCCTTGGTATAAAGCAGCAG
2	EST00707		70S CTTTCTGCTAGGTTGTTGGG	70A ACAAAGAGGTTGAGGGCTAG
2	EST00707		102S ATGAATGGAAAGGCTGGG	102A TAGTCATGGCCCGATTCA
2	EST00876		32S ATAACTCCACCAGTCTCTCC	32A TACTCTCCCTTATCCCCATACC
2	EST00892		183S TCGTTACGTTGACAGGTAGG	183A GGCAGAACTTGAAGGGTTAC
2	EST00915		126S GGAAAGGGAAAAAGGTGC	126A TAACACAGAGGGAAAGGGAG
2	EST01870		82S CTTTGTGCCAGGGATTTG	82A AAGGCTCCATCTTCTGGAAC
2	EST01880		88S GGTTTGTCTCAGCTGTGTTG	88A CTTTGTGGGGTGTGAGTATG
2	EST01911		96S ACCAGTTCTCAGTTGGGTTC	96A CACAATATCAACCTGAGGGG
2	EST01972		198S GAAAGCAGGATGAGTGATGG	198A TAGAGTCCATGGGTCCTGTT
2,10	EST02530	G:M64279	175S GTGCTTCTTTTGCAGACTGG	175A GCTTTGGTCGAACTTGGT
3	EST00094		7S GCAGGATGTCAGTCTTTTGAGG	7A AGCACACATTATCTACCACGGC
3	EST00476		34S AAAATCAGTGCTGGAGCC	34A AGAGAACTGTCCAAGGGTTG
3	EST00514		42S AAAAGTCTGCTGAGTCTGGG	42B GGAGCAATTCCACATTCG
3	EST00605		62S GTTCCTCTTCCTATCACTGGTG	62B ACTTTATGCCCACCTCTTCC
3	EST00857		17S GCAGGATGTCAGTCTTTTGAGG	17A CCAGCACACATTATCTACCACG
3	EST01894		91S TCACAATCTTCCATGGGG	91A TCTGACCCTCACCTTTATCC
4	EST00038	G:M63325	8S GGAAGTACAGGATTTGGC	8A TTAGAGATGGGATGATGCCG
4	EST00530		44S GTCATGTCTGTTGAACGAGG	44B CCTACACCAGGAAATAGGGT
4	EST00754		75S TTCCATGTAGCGTCTTCCAC	75A CAACCCTGGTCTTGAGAAAC
4	EST00863		24S GGGATAGAGTCTGGTATTCTGC	24A GGATACTGTGCTTAGTTCCTGC
4	EST01650	G:M64930	150S GACGTGTACAGCACTTTCCA	150A TGGGAGCCACTTGATCTT
4	EST01873		83S CCATCTGTAACACTAGCAGAGG	83A TCAGAAACTGCGAACTAGGG
4	EST02370	G:K03021	173S TGGTGGGAAGGTGGTATT	173A GTTAGAAAGGTAGTGGTCCCTG
4	EST02533	P:JU0319	176S CCTTTGTGTTGGTGAGACAG	176A CCGAACACCTGGTTCATT
5	EST00104		3S CAGATCAATACATCCTCTGGGG	3A CTGTGCAGTGGTGAGTAAAAGG
5	EST00478		35S AACAGAAAGCTCTGGGAGAC	35A CACTGCAGCTTGAGAATCAC
5	EST00835		13S TCTCCAACACAGTCATGC	13A CGGATGCCATCATATACC

CH	EST#	Locus/ DB match[1]		Sequence		Sequence
5	EST01471	MTAP5	143S	GGAAAACTACCAAGACCCCT	143A	ATGAGTACCCACCCTCTTCA
5	EST01744	P:DEBOXM	157S	ACAGCACGTCTTGGCAAT	157A	TGAGCTAGTAATTCTGGGCC
5	EST01943		188S	TCAGTTCATAGGAGATGGGG	188A	CTGACAGTTTTCCACAGCCT
5	EST01977		203S	CCAAACACCCTTTCCACTAC	203A	CTCATATTGCCTCTGAACCC
5q	EST00488		37S	AAGAACCAGGCAACTTGG	37A	CTGGTCTCAACCAATTCACC
6	EST00770		77S	CCCCTTAGAATGTCATGGTC	77A	AAGCCTAGCTGTAAAGGTGG
6	EST00833		12S	ACCCAGTTCTCAAAGACC	12B	TTTACCATTCAGAGGCAGCG
6	EST00874		27S	GGTACACAGGGAAAGTACATGG	27A	TGATTCACAGCTACGTGTGC
6	EST01915		97S	AATCTAAGGGTTGGGTCTCC	97A	GAACTGGAAGAACTGTTGGC
6	EST01920		99S	AGCCTAGTAGCTGTATTGGGTG	99A	ATTGGTAAGTAGGGTCCTGC
7	EST00548		50S	ATTACACGTCTCGGGAGTTC	50A	TGTCTATCTGTGAACGTGGC
7	EST00601		60S	AAGCACTCCATAGATCCACC	60A	ACAGAACACCAAGGAGAACC
7	EST00654		66S	ATAGACCGATTGAGGTCAGC	66A	ATGAGACTGCACTGCCTTTC
7	EST00838		20S	GTTCTTTCCCAGGTATGC	20A	TTGTTGGTACTGAGGAAGTGCG
7	EST01888		89S	CCTTTATCCTCAGCTGGTTG	89A	ATCTATAGCATCCTCAGGGG
7	EST02113	G:M58318	169S	CCTTCTGCAGGAACTCAAAC	169A	ACTCCCTACTATCATGTACCCC
8	EST00582		56S	CTGAGTTGCTGTTTGGTCTG	56B	ACCTCTCTGCCTTCTCTTAGTG
8	EST00614		63S	GTAATGCATGTCGAGTGCTC	63A	TGCTTCTGTGGGAAAAGG
8	EST00680		67S	AGATTTCACAGAGGCAGTGG	67B	ATTACCCCTCTGGGCATTAG
8	EST01905		94S	GGCTTATTTGTGGATGCC	94B	AATGCAACAACCCAGCTC
9	EST00510		41S	TGGCCTTTGTCTCAACTCAC	41A	TAGAGAAACTGGGACTTGGG
9	EST00895		111S	TTTCCACAATGACTCCCC	111B	TGGAGTTTGGGTGGTTCT
9	EST01906		95S	ACAACTATCTGGCAGGGATG	95A	TAGTGCACACCCAAATGG
9	EST02055	P:S05054	167S	AGAAGAGAAGCAGCCCAATC	167A	GTGGTGCTCAAGTCCATCAT
10	EST00577		55S	CCCTGGCACAAAAGATTC	55B	TCACCGAGATGAGTTTGC
10	EST00684		68S	TTGTCCCCACCAAAACTC	68A	AAGAGTCTCGACTAACCCAGTC
10	EST00695		69S	GTCCATTTGCTCAACCAGAC	69A	GGTGGGATAAGTGAAGCAAC
10	EST00856		16S	GCCATCCTTTTTGAGAGG	16B	CTGTAAGAGTTGCAGCATTCCG
10	EST01876		85S	CTTCATCAGGGAACTTGAGC	85A	TGGGTCAGAAGTTTAGGAGC
10	EST01945		190S	GGCTTCTGTGAAAGACCAGT	190A	TTCTCATCTCCCTTTGGG
10	EST01958		196S	TCTAAGTGACCTTCTTGCCC	196A	TCACGCTGCTAGAACAGAAG
10	EST01982	P:B33283	164S	CTAGCGGAATTACTGGAGGA	164A	TGCTGTGAGCAAATCCTG
10,14	EST02146	G:M31178	170S	TCCTTGGGTAGAACTGCACT	170A	GCTAATCCGTCCAGACTCAT
10,2	EST02530	G:M64279	175S	GTGCTTCTTTTGCAGACTGG	175A	GCTTTGGTCGAACTTGGT
11	EST00109		4S	CTAACCACAACCCACACATTGG	4A	CCTCAGCACAAGAGAAGAATGG
11	EST00501		39S	TCCATAACCTCTTCCTGGAG	39A	GGTGCTTGGCTTTCATTC
11	EST00532		45S	TACTTTCCTTCCTTCCCTCC	45A	GTATTGCCGGATTACCAGTC
11	EST00547		49S	TCTGGTCGGAGTTAAAGGAG	49B	CTTAAACGCAGTATGGGGAG
11	EST00776		79S	GGTTTTACTTGGTTGGGG	79A	AACTGACTTGCCAATCCC
11	EST00901		115S	CTGAAACAGCTTCCCCTCTA	115A	AGCTCTTCCCTCCTTTCTTC
11	EST00926		136S	AGAGTCCACATGCAACACCT	136A	GCCTGTGGGTGTCAGTATTA
12	EST00867		25S	GAAGATTATTGTGGTGCCCG	25B	CTCATTTCTCTTTTCCTCCCCC
12	EST00899		114S	CCATAGGGAATGTATGGGAG	114A	CCCTGACTGTCCTTGTTTTC
12	EST00908		120S	TCAAGGAAAGGTGGACCTAG	120B	CCTGAGCCATGCTGATTT
13	EST00527		43S	CCAGGGATAGGTTGTTTGAG	43A	GTTCCATAGTAGAGGGAAGGTG
13	EST00537		46S	GGTATATCAGGCCAAAGGTG	46A	TTAGCACAGAACCTGACCTG
13	EST01889		90S	CACAGCAAGGAACTGTAACG	90A	TAATGAAACCCTGGCTGC
13	EST01896		92S	CTGTTGTATCACATGCCAGG	92A	CTCTGGGTTCAGTTTCTCTACC
14	EST00795		86S	TAAGAGCCGCCAGATTTC	86A	AGGTTGTTCTCAAGGTCTGC

CH	EST#	Locus/ DB match[1]	Sequence	Sequence
14	EST00845		15S AGGAGGAAGCTGAAATCC	15A GGAAGTCCATAAGAGACTCACC
14	EST00902		116S GGCGGCACTGTTGTAGTTTA	116A TATTCCTTGGCTTGCCTC
14	EST01583	P:R5RT18	145S GGAACTTGATCTTGGAGTCG	145A AGATCATGAAGGTGGAGGAG
14	EST01627	P:A24577	148S AGAAGATCCATCGCAGAGTC	148A CTGGCGAAGAATGGTGTT
14	EST01868		81S TATCCCCATGCCCTAAAG	81A TGGCAAGATCAGAAACCC
14	EST01875		84S TTCACACTGCAGCAAAGC	84A ATATTTTAGGCCCCGTGG
14,10	EST02146	G:M31178	170S TCCTTGGGTAGAACTGCACT	170A GCTAATCCGTCCAGACTCAT
15	EST00780		80S TTGAGAAGTGGAAGGCAGAG	80A AAGGGCAGTGTTGGTGTAAG
15	EST01678	P:36479	153S AACGGTGGTTTATGCGAG	153A TGTACGAGGGGAAGACATCT
16	EST00566		53S TTATGGAAAGCCAGCAGC	53B CCCATAAGGGGTTGACTAGA
16	EST00831		11T GCACTTCATTTTCTTCCC	11B TTCATCTGCCAGGAGAATGC
16	EST00889		181S AAAAGCCAGAACACCCTACC	181A AGTAATGACGTGTGTGGCTG
17	EST00483		36S GGTGTAGAATGCCAGATTCC	36A TCTTCCTCATCCTGCTTCTC
17	EST00675	T:RICGOS2G_1	138S GGGGTCCGCTTTTAACTT	138A GCCTATTGCTTGACCTCTGT
17	EST00854		19S TTGCTGTGGAATCCATGAGAGC	19A GGCAAGTGATCTGTTCTTGG
17	EST00869		28S AGTGAGGAGGAAAGAGGAAAGG	28A ATAGCCCCAAGCCCTTATTTGC
17	EST01667	P:JQ0771	186S TTGGCACACCAAGAAGCT	186A GTTGATCTCAGTGCGGTGAT
18	EST00542		48S CATCAAAACCCGAAGCTC	48A GGGCCTTGCAATAGTTCTAC
18	EST00775		78S TAAAGCCATGCAGGAAGG	78A AGCAGTTAGGGACTGGTTTG
18	EST00893		109S CCACGTATCTGTTCCATGTG	109A CCCCCAGAAGGCTTATAGTT
18	EST00906		119S TAGAGAGGAGCTGGATTTCG	119A CGTGGTAGATTCTGGGTGAT
18	EST01784	P:A28209	160S AACTAATTGGGTTGGCCC	160A TGGACATCAACGTCAGACTC
19	EST00708		71S AAACCTTGGTCTGCATGG	71A AAGGGCAAGGAGAGGTTATC
19	EST00875		29S CCTTTTTCCCTCTGAGTTCG	29A AATGAGAAGCAGCTCTGG
19	EST00884	ATP1A3	31S GTAGTAGGTTTCCTTCTCCACC	31A TGTGCCTTCCCCTAAAGTTTCC
19	EST00894		110S CAGGTTGGCACGTGTATAAG	110A GTCTCCATTTCCAGTTCCAG
19	EST01791	P:JN0129	161S GTGGCACATGACACAATCTC	161A CAGGTTAATTGGATGGCG
20	EST00113		5S TCGGAGAAGTTGCAGTTTCTGG	5A GTTAAAAGCTGTTAGACGGGGC
20	EST00890		182S TTAGATGCGTCCCTTTGG	182A TCTGATTTTCCTGCCCTG
20	EST00909		121S TCTTCCCTTTGGGCTCTT	121A GAACCCAATACGAACTCGAC
20	EST00925		135S CAAACCACAGGGGAAATG	135A GATGGCTACGTTTTGGAGAG
21	EST00541		47S CCGTCGGTAACAGAAAACTC	47A ACACAGTCCTGTAATACCCTCC
21	EST00591		58S TCAAAGACTGGGCGAAAG	58A CCAAGATGGTCTGGATTCTC
22	EST00101		2S TATATGGGTCCCTTGTGC	2B CTCCGATCTATACACACC
22	EST01601	P:ROBO	146S GTCATCTCCTTCACCTAAGAGG	146A GAAAGTCTGAGAAGGCCTGA
X	EST00574		54S CATCCTCCAGATGAGGAAAG	54A TTTTGTATCCCCAGCAGC
X	EST00737		72S GACCTTATAACCATACGGCG	72B CCCAAATTCTGGTTCCCT
X	EST00858		18S AAGGCGAGGATTATGTGC	18A TTCTACTGGGTACACTTCGACC
X	EST00887		104S AATGAGAGAGGGTGAGATGG	104A AGATACAAGCAGGAGATCGG
X	EST00896		112S AATGCTTGGCATGACTCC	112A TCCCCAAACTCCTACACCTA
X	EST01879		87S GAGGCATCAGAGAGAAAACC	87A GAACCTGAGAAGACCCTTACTG
X	EST02087	G:M55284	168S TGCTTCTCTGTCTGTACCCA	168A GATTTGCTCCGGGTAATAGG

[1]Locus = human gene mapping workshop locus symbol; DB match = accession number of matched database file, with P = PIR, G = GenBank, and T = GenPept.

TABLE 2. PROCESSING SUMMARY

Total EST files processed:	626	
Total files meeting primer design criteria:	428	
Total files tested with at least 1 primer pair:	168	
Subtotal mapped:	106	
Total files tested with 2 primer pairs:	48	
Subtotal mapped:	26	
	Total mapped:	132

preprint & reprint removed, ok

DATE
FILMED
12/4/92