I. Technical Progress During Current Budget Period

DOE/ER/61232--1

DE92 005523

A. Summary

J. Craig Venter, National Institute of Neurological Disorders and Stroke, has begun to identify genes expressed in the human brain by partially sequencing cDNA clones (Adams *et al.*, Science 252:1651-1656, 1991). We are collaborating with the Venter group and using their sequence data to develop methods for rapid localization of newly identified cDNAs to human chromosomes. We are applying the ABI automated DNA sequencer to the analysis of fluorescently-tagged PCR products for assigning sequences to individual human chromosomes. The steps in our mapping protocol are (1) to design PCR primers from the Venter laboratory-generated sequence data, (2) to test the primers for specific amplification from human genomic DNA, (3) to use the primers for PCR amplification from a somatic cell hybrid cell mapping panel, (4) to determine the presence or absence of the specific amplification products from each cell line DNA by electrophoretic analysis using the ABI sequencer, and (5) to analyze the pattern of amplification results from the hybrid panel to identify the chromosomal origin of the cDNA sequence. We have demonstrated the principle by mapping 12 sequences or "Expressed Sequence Tags" (ESTs), providing primer sequence data for subsequent subchromosomal localizations. We will now concentrate on developing methodology to allow multiplexing the amplification reactions and analysis of the reaction products, to achieve a high throughput with a minimum allocation of resources. This project will generate a data set from which to evaluate strategies to identify functional primer sequences from cDNA sequence data.

B. Methods

1. *Primer design.* We have been using the PRIMERS software (Lowe *et al.*, 1990) to identify potential primers from the sequence files supplied by the Venter laboratory. A batch of cDNA sequence files is analyzed under narrowly defined parameters (50% GC, Tm of amplified fragment 78-81°C, amplified fragment size 100-150 bp), and the oligonucleotides identified are used to test for amplification from human, hamster, or mouse genomic DNA.

2. *Amplification and determination of fragment size.* So far, one of each oligonucleotide primer pair was synthesized with 5' Aminolink II and labeled with a fluorescent dye using a procedure provided by Applied Biosystems, Inc. Template DNA (50 ng) was amplified in the presence of 80 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 temperature profile consisted of a 95°C denaturation step for 5 min; 25 cycles of 94°C for 1.4 min, 55°C for 2 min, 72°C for 2 min, and a final 10 min incubation at 72°C to complete product extension. One μl of the PCR product was pooled with a fluorescently-labelled internal lane size standard (pGEM3Z sequencing ladder with only ddTTP chain termination products). This mixture was ethanol-precipitated and re-suspended in 5 μl of de-ionized formamide. Sample was denatured at 95°C for 2 minutes before loading onto a 6% polyacrylamide denaturing gel (8 M urea). Data from the electrophoretic run were collected on an ABI 370A automated sequencer.

During this period, hardware and software were obtained to upgrade the instrumentation to a 373A. Data collection is currently done in parallel because of the lack of software for the 373A to determine fragment sizes. The 373A software (GeneScan) should be installed in November, 1991.

3. *Chromosome assignment.* Primer pairs were tested for amplification of a product from human genomic DNA of the size predicted from the cDNA sequence and distinguishable from mouse or Chinese hamster. Primers meeting this criterion were then tested on DNA from human/rodent somatic cell hybrids obtained from the NIGMS Human Genetic Mutant Cell Repository (NIGMS panel #1) and/or from the BIOS Corporation (PCRable DNA).

Sequences were assigned to a chromosome when the discordancy value was less than 9%. Chromosomes reported as being present in fewer than 10% of metaphases were not included in the discordancy analysis.

C. Results

1. *Representative throughput.* The following table summarizes the mapping process for a set of files received from the Venter laboratory. Of 36 sequences analyzed, 13 could be used to predict primers according to the criteria described in Methods. So far 10 of these have been tested, with 6 resulting in unambiguous chromosome assignment.

Table 1

Total files in batch:	36
Total files giving at least 1 possible primer pair:	13
Total primer pairs tested:	10
Total primer pairs giving amplification in human distinct from rodent:	7
Total unambiguously mapped:	6

2. *Chromosome assignment.* A subset of primers successful in mapping human sequences is given in Table 2. Additional details about fragments detected in human and rodent cells (MMU = Mus musculus, CHO = hamster) cells are given in Table 3. The results of the discordancy analyses are summarized in Table 4. Three sequences were assigned to a chromosome in spite of a single discordant result. Two discordancies resulted from the inability to detect a fragment when the chromosome was reported to be present, one resulted from the detection of a weak band, possibly of a smaller size, which was not detected in any other reaction.

Table 2

ATCC#	EST	NAME	PRIMERS (5' -> 3')	(bp)	CH
61986	EST104	HHCG36	AL2-CAGATCAATACATCCTCTGGGG CTGTGCAGTGGTGAGTAAAAGG	181	5
61906	EST109	HHCG44	AL2-CTAACCACAACCCACACATTGG CCTCAGCACAAGAGAAGAATGG	144	11
61922	EST113	HHCG70	AL2-TCGGAGAAGTTGCAGTTTCTGG GTTAAAAGCTGTTAGACGGGGC	159	20
37920	EST94	HHCE15	AL2-GCAGGATGTCAGTCTTTTGAGG AGCACACATTATCTACCACGGC	137	3
61830	EST38	HHCA23RZ	AL2-GGAAGTACAGGATTTGGC TTAGAGATGGGATGATGCCG	150	4

Table 3

ATCC#	EST	NAME	E[1]	O[2]	CH	NOTES
61896	EST00104	HHCG36	181	183	5	
61906	EST00109	HHCG44	144	147	11	CHO band approximately 380 bp
61922	EST00113	HHCG70	159	148	20	
37920	EST00094	HHCE15	137	141	3	
61830	EST00038	HHCA23RZ	150	151	4	MMU band approximately 400 bp
		HHCF44F	170	173	1	
		HHCMC10F	133	135	5	MMU band 145 bp strong; MMU band 285 bp weak
		HHCMC37F	122	123	4	
		HHCMC78F	139	143	3	
		HHCMC82F	143	450	X	MMU band 445 bp, >600 bp
		HHCMC70F	147	161	7	
		HHCMC13F	126	125	7	

[1]Expected fragment size in nucleotides
[2]Observed fragment size in nucleotides

LOCALIZATION OF cDNA SEQUENCES USING HUMAN x RODENT CELL HYBRIDS

Sequence name	Percent discordant for each human chromosome																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
HHCE15	44	44	6	44	41	44	31	24	61	27	75	47	27	24	20	59	28	22	47	41	35	47	53	53
HHCF44F	0	22	41	31	47	44	38	47	28	27	44	53	13	47	33	18	61	22	13	35	41	40	20	35
HHCG36	33	44	59	50	6	33	56	53	50	40	38	24	53	41	53	53	39	28	40	41	12	33	60	24
HHCG44	56	44	59	31	41	44	50	41	33	60	0	29	53	47	53	47	50	56	47	29	35	40	33	41
HHCG70	33	33	47	31	29	33	19	29	61	33	44	24	20	18	33	53	28	17	33	6	29	20	60	53
HHCMC10F	50	39	53	50	0	39	50	47	56	40	31	18	40	35	47	59	44	33	47	35	6	20	60	29
HHCMC13F	39	44	29	19	53	39	0	24	67	47	44	35	53	18	20	59	22	39	33	18	64	47	60	53
HHCMC37F	50	39	29	24	29	17	19	24	78	40	50	24	33	0	20	76	11	28	33	33	41	27	73	53
HHCMC70F	61	50	29	25	29	33	19	12	89	33	50	18	47	12	20	82	0	39	40	24	35	40	87	59
HHCMC78F	44	39	0	31	41	33	31	29	61	40	63	41	27	24	20	64	27	33	40	35	47	40	60	35
HHCMC82F	24	12	25	12	72	20	20	16	16	21	17	28	28	32	25	21	20	28	40	40	40	16	0	28

END

DATE FILMED

4/02/92