Mapping Genes to Human Chromosome 19

Sarah Connolly
University of Illinois at Urbana-Champaign
Lawrence Livermore National Laboratory
Livermore, CA 94550
12/15/95

Prepared in partial fulfillment of the requirements of the Science and Engineering Research Semester under the direction of Dr. Greg Lennon, in the Lawrence Livermore National Laboratory

**This research was supported by an appointment to the U.S. Department of Energy Science and Engineering Research Semester program administered by LLNL under Contract W-7405-Eng-48 with Lawrence Livermore National Laboratory*

Abstract:

For this project, 23 Expressed Sequence Tags (ESTs) were fine mapped to regions of human chromosome 19. An EST is a short DNA sequence that occurs once in the genome and corresponds to a single expressed gene. ^{32}P-radiolabeled probes were made by polymerase chain reaction for each EST and hybridized to filters containing a chromosome 19-specific cosmid library. The location of the ESTs on the chromosome was determined by the location of the ordered cosmid to which the EST hybridized.

Of the 23 ESTs that were sublocalized, 6 correspond to known genes, and 17 correspond to anonymous genes. Five of the anonymous genes have shown strong homologies to other known genes. These localized ESTs may serve as potential candidates for disease genes, as well as detectable markers for future physical mapping.

Introduction:

Goals of the Human Genome Project:

The Human Genome Project is a large coordinated effort, aimed at constructing a detailed map of all three billion base pairs of human DNA. Among the goals of the Human Genome Project are to determine the exact sequence of the genome, to create DNA markers spaced throughout the chromosomes, and to identify the location the genes encoded by this sequence. Since only 10% of the entire genome is estimated to codes for genes, and approximately 100,000 genes reside in the genome, a fast and simple method of locating genes is critical. Identifying the location of genes will provide detectable markers for future physical mapping projects. The localized genes may also serve as candidate genes for various diseases when the approximate location of a diseases gene is known.

Lawrence Livermore National Laboratory focuses its genome effort on chromosome 19. Chromosome 19 is a small, gene-rich chromosome, estimated to contain about 2,000 genes. A high resolution physical map of chromosome 19 has been completed, covering approximately 95% of the chromosome with cosmids contigs. To date, approximately 250 genes have been assigned to these cosmids and mapped to chromosome 19. The mapping effort was continued in this study, with 23 genes more assigned to cosmids of chromosome 19.

Overview of Methods: cDNA mapping

Complementary DNA (cDNA) mapping is a quick method of gene identification and location. A cDNA is a copy of the protein encoding regions of a gene. cDNAs are constructed from mRNA templates in vitro using reverse transcriptase. mRNAs are valuable templates because each mRNA is a continuous stretch of protein-encoding nucleotides that corresponding to a single gene. Constructing a cDNA from an mRNA insures that the cDNA contains only coding information and that it corresponds to a single gene that is expressed in the specific tissue.

Once the cDNA has been made, it is placed in a vector and cloned. Primers are constructed that are capable of selectively amplifying a region of the cDNA by polymerase chain reaction (PCR). PCR is a method used to selectively amplify short regions of DNA. The cDNA primers are 20 base-pair sequences that complement specific regions near the ends of the cDNA sequence. These primers border a target region of the cDNA and allow polymerase to replicate the bordered region repeatedly. The replicated region is called an expressed sequence tag (EST). An EST is a short sequence of DNA that occurs only once in the genome and that corresponds to a single expressed gene. Because an EST sequence is unique in the genome, once it has been mapped to a chromosome it can serve as a physical marker on the chromosome.

Through PCR, the EST is amplified and radiolabeled with ^{32}P to make a probe. A probe is a short stretch of single-stranded DNA whose sequence is complementary to a unique portion of a targeted DNA sequence. The probe is hybridized against filters that contain a chromosome 19-specific cosmid library. Cosmids are 35,000 base pair

segments of chromosome 19 that have been ordered by EcoRI mapping and fluorescence in situ hybridization (FISH). Most of chromosome 19 can be represented by seven cosmid filters.

When a probe is exposed to these cosmid filters, it hybridizes to complementary sequences. The complementary sequence of a probe made from an EST exists in the gene that the EST originated from. The probe will bind to the cosmid that contains the gene, indicating where the gene is located. The hybridized cosmid filters are analyzed by PhosphoImager, and the location of the hybridized probe can be seen on an autoradiogram. The gridded array of the cosmids on the filters allows positive identification of which cosmid the probe hybridized to. The hybridization of the probe to a cosmid indicates that the target gene must be located within that cosmid. To verify the assignment of the gene to the cosmid, the cosmid DNA is used as a template in PCR amplification.

Characteristics of the identified positive cosmids can be found in LLNL databases. If the positive cosmid has been mapped to the chromosome, the location of the gene is known.

Materials and Methods:

(1)Construction of cDNA Primers:

Twenty-three cDNA oligonucleotide primer pairs were either designed during the course of this research or already available through the Genome Data Base. Reference numbers are given in Table 1. The primers were constructed to amplify specific EST sequences.

(2)Amplification of Gene Fragment by Polymerase Chain Reaction:

Using these primers, EST fragments were amplified from human placental DNA by polymerase chain reaction. A Perkin Elmer 9600 machine was used for all PCR reactions. For optimization of the PCR, 50µl reactions were performed using 100ng human placental template DNA, 40pm of each primer pair, 200mM dNTPs, 500mM KCl, 100mM Tris pH 8.3, 0.5 units of Taq polymerase, and varying concentrations of $MgCl_2$. The reactions were run through forty cycles of 94°C for 30 seconds, 54-65°C for 30 seconds, and 94°C for 30 seconds. The specific annealing temperatures (Tm) and $MgCl_2$ concentrations for each EST are given in Table 1.

PCR products were separated by gel electrophoresis on a 1.5% agarose gel with 50µg/µl ethidium bromide. The DNA fragments were visualized by ultra-violet fluorescence, and the sizes of the fragments were determined. The desired fragment was cut from the gel and placed in an eppendorf tube with 10µl of distilled water. The tubes were refrigerated for later use.

(3) Construction of Radiolabeled Probe:

A 25µl PCR radiolabeling reaction was performed using α^{32}P-labeled dCTP and the gel slice as the template. The PCR conditions were consistent with the original PCR conditions. The PCR products were purified using Worthington G-50 minispin columns.

(4) Hybridization of Probe to Cosmid Filters:

High density cosmid filters containing cosmid clones from chromosome 19 were prehybridized for at least two hours at 65°C in a hybridization solution that contained 4M NaCl, 0.5M EDTA, 1M Tris, 5% NA pyrophosphate, 50% dextran sulfate, and 20% SDS. 500µl salmon sperm DNA and 50µg Human Cot-1 DNA were added to block repetitive elements.

The radioactive probes were denatured to single-strands by boiling them for five minutes. They were added to the hybridization mixture and hybridized at 65°C overnight. The filters were then washed twice with a high stringency solution (0.1xSSC-0.1% SDS) for thirty minutes, and once with a low stringency solution (2xSSC-1% SDS) for thirty minutes.

The filters were wrapped and exposed to Molecular Dynamics PhosphoImager cassettes for at least one night. The cassettes were scanned and analyzed for identification of the cosmids positive for hybridizations.

(5) Verification by Polymerase Chain Reaction:

The hybridization of a probe to a cosmid was verified by PCR, using the cDNA primers and the cosmid DNA as template. A 50µl PCR was set up and the products were run through a 1.5% agarose gel. Amplification of a correctly sized fragment from the cosmid DNA served as proof that the EST mapped to that cosmid.

(6) Database Analysis:

The location of cosmids on the chromosome and other attributes of cosmids were available through the LLNL Browser program. This database includes links to information such as similarly EcoRI mapped contigs and cytogenetic band information as determined by FISH. If the positive cosmid was contained in a contig that had been FISH mapped, the location of the EST that hybridize to that cosmid was known.

A sequence homology search was run through the NCBI BLAST database for the anonymous genes to determine if the anonymous sequence matched the sequence of a known gene. Anonymous genes that lack a sequence accession numbers (see Table 2) were not run through BLAST because the sequences were not readily available. For the BLAST analysis, a score of over 200 was considered significant. Only the highest scoring homologous sequence was noted.

Results:

Six known genes and seventeen anonymous genes were mapped to the chromosome. Ten genes mapped to the q arm of chromosome 19 and seven mapped to the p arm, indicating no distinct distribution pattern. The six remaining genes were mapped to cosmids that have not yet been placed on the chromosome. Once these cosmids have been placed on the chromosome, the location of the six genes will be known. Figure 1 is an ideogram of human chromosome 19, showing the approximate locations of the ESTs. Table 2 lists the names of the genes and the cosmids to which each was assigned.

The BLAST search for sequence homologies revealed that five of the anonymous genes had strong matches to characterized genes. The names of the homologous genes and the extent of the homology are listed in Table 3. The nine other anonymous cDNA sequences that were sent through BLAST did not show strong homologies to known genes. These sequences can be considered segments of new, unique genes on chromosome 19.

Chromosome 19

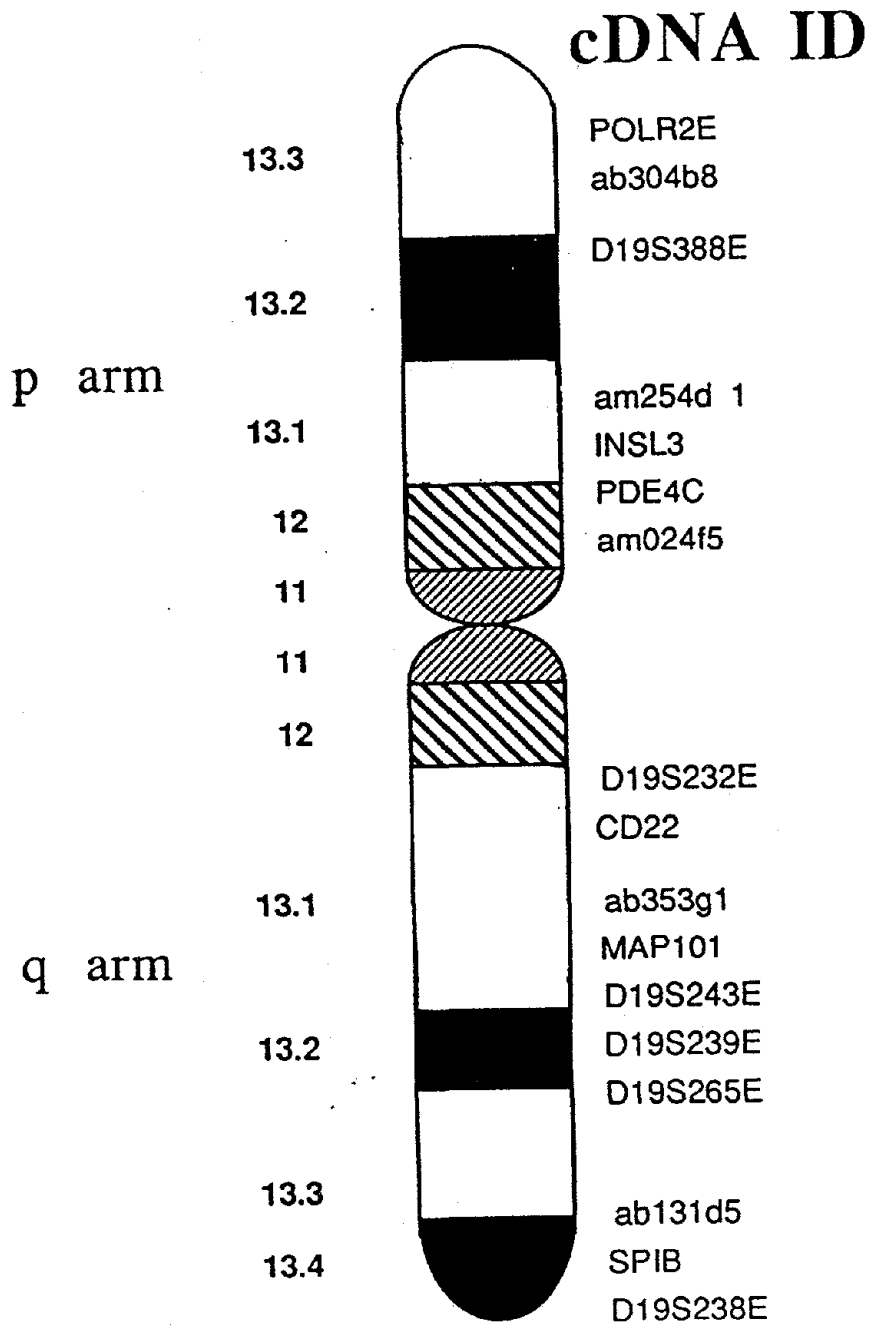

Figure 1: Ideogram of Chromosome 19

Locus	Sequence Accession Number	Genome Database Locus ID	Genome Database PCR probe ID	cDNA ID	Gene Name	Source of PCR Primers and GDB Reference numbers		PCR Optimization Conditions	
								Tm	[MgCl2]
D19S238E	M78256,M79273	G00-130-359	G00-189-000		Anonymous	Polymeropolous	G00-101-107	55	2.0mM
D19S239E	M78106	G00-130-220	G00-188-995		Anonymous	Polymeropolous	G00-101-107	55	2.0mM
D19S265E	M78736	G00-130-834	G00-191-600		Anonymous	Durkin	G00-042-878	60	2.0mM
D19S388E	M85728	G00-139-508	G00-198-958		Anonymous	Maglott	G00-043-830	60	2.0mM
D19S390E	M85570	G00-139-582	G00-199-032		Anonymous	Maglott	G00-043-830	60	2.0mM
D19S819E		G00-130-937	G00-581-496		Anonymous	Polymeropolous	G00-101-107	54	2.0mM
D19S243E		G00-133-968	G00-188-887		Anonymous	Polymeropolous	G00-101-107	55	2.0mM
ab353g1	H11320			47518	Anonymous-LLNL	LLNL		55	2.0mM
ab178c8	R40345			28807	Anonymous-LLNL	LLNL		60	2.0mM
am024f05	T54094, T54185				Anonymous-LLNL	LLNL		60	2.0mM
ab304b8	R59631,R59691			42233	Anonymous-LLNL	LLNL		60	2.0mM
ab389b11	H19436			51303	Anonymous-LLNL	LLNL		60	2.0mM
am254d1	R75610, R75700				Anonymous-LLNL	LLNL		60	2.0mM
ab321b12	H06151,H06152			44051	Anonymous-LLNL	LLNL		60	2.0mM
ab131d5	T77499,R38383			23804	Anonymous-LLNL	LLNL		60	2.0mM
D19S232E		G00-133-858	G00-188-874		Anonymous	Polymeropolous	G00-101-107	60	1.0 mM
D19S234E	M62234	G00-122-882	G00-188-890		Anonymous	Polymeropolous	G00-101-107	55	1.5 mM
CD22-		G00-127-545			CD22 antigen	LLNL		60	2.0mM
INSL3		G00-230-307			Insulin-like 3 (Leydig cell)	LLNL		60	2.0mM
PDE4C	Z46632	G00-132-539			Phosphodiesterase 4C, cAMP specific	LLNL		60	2.0mM
POLR2E		G00-250-743			Polymerase (RNA) II polypeptide E	LLNL		60	2.0mM
SPIB		G00-374-517			Spi-B transcription factor	LLNL		60	2.0mM
MAP101					Map 101	Ross, David		60	1.0 mM

Table 1: References and PCR Conditions

Locus/EST	Location on Chromosome	Positive Cosmid IDs
D19S238E	19q13.4	24602, 25446, 16251
D19S239E	19q13.2	16651, 24533, 25361
D19S265E	19q13.2	26165, 26630, 26834, 27577, 28943, 30006, 30012, 30563, 31696, 32336
D19S388E	19p13.2-13.3	20259
D19S390E		25056
D19S819E		17025
D19S243E	19q13.2	18930, 23864, 24762, 25796
ab353g1	19q13.1	31418, 31710
ab178c8		21915
am024f05	19p12-13.1	29957, 30465, 30777, 31164, 31165, 33114, 34330
ab304b8	19p13.3	21834, 22077
ab389b11		22115
am254d1	19p13.1	16351
ab321b12		15616
ab131d5	19q13.3-13.4	16450
D19S232E	19q13.11	20520, 21363
D19S234E		22986, 24563
CD22-	19q13.1	16661, 18534, 18536, 24202
INSL3	19p13.2-p12	19847, 25186, 25196
PDE4C	19p12-13.1	19896, 22037
POLR2E	19p13.3	15978, 23821
SPIB	19q13.3-q13.4	20372, 24917
MAP101	19q13.1	16584, 19754, 24590

Table 2: EST Locations

Locus	Homologous Gene Name	GDB ID	SCORE
ab321b12	Human Natural Killer Cell Enhancing Factor	L19185	1632
D19S239E	H. sapiens HZF4 mRNA for Zinc Finger	X78927	1433
D19S388E	Human Interleukin 11	M81890	810
ab131d5	Human Tristetraproline mRNA	M63625	703
D19S238E	Human Retinoblastoma Susceptability	L11910	264

Table 3: BLAST search results

Conclusions:

The twenty-three ESTs that were localized to chromosome 19 in this study may now serve as detectable landmarks on the physical map of the chromosome. The mapping of these ESTs will aid in further physical mapping projects. ESTs can be used to find pairs of overlapping clones in the construction of contig maps of chromosomes. In addition, the anonymous ESTs may serve as candidate genes for various diseases when the approximate location of a disease gene has been determined by genetic linkage techniques. Further study is needed to determine the location of the six genes that mapped to cosmids that were not currently mapped to the chromosome. The homologies found for five of the anonymous genes may deserve further inquiry also. Eventually, a metric ordering of these ESTs will be determined, based on the locations of these cosmids on the chromosome.

References:

Cooper, Necia ed. (1992) *Los Alamos Science: The Human Genome Project* . Number 20: 108-140.

Mohrenweiser, H.W., S.Tsuijimoto, K. Tynan, J. Lamerdin, and A.V. Carrano (1995) Unique sequence STSs for 21 cytogenetically mapped loci on human chromosome 19. *Cytogenet Cell Genet* **71:** 58-61.

Mohenrenweiser HW, Swartz A, Johnson K: *Report of the Committe on the Genetic Constitution of Chromosome 19 in Cuicchia AJ, Pearson PL (eds):Human Gene Mapping 1993*, pp.575-597 (Johns Hopkins University Press, Baltimore 1994).

Pappas, George J., Michael H. Polymeropolous, John M. Boyle, and Jeffery M. Trent (1995) Regional Assignment by Hybrid Mapping of 36 Expressed Sequence Tags (ESTs) on Human Chromosome 6. *Genomics* **25:** 124-128.