Lecture 8 - Genome Mapping Lab Practicums (AMG text pp. 96-111)
September 13, 2001

Positional Mapping of Human Disease Genes

Research Objective
An autosomal recessive mutation has been mapped by linkage analysis to the human chromosomal region 7q31.3. The team of genomic researchers working on this project plan to identify the candidate disease gene by systematically comparing the sequence of genes in normal individuals (homozygous wild-type), carriers (heterozygous) and patients with the disease (homozygous mutant). The assumption is that homozygous individuals afflicted with the disease will have nucleotide alterations in both copies of the gene, whereas, carriers and normal individuals will have one or two copies of the wild-type gene, respectively. Biochemical evidence suggests that the disease results from a complete loss of a key enzyme required for glucose metabolism.

Available Information and Reagents

1. Materials are available to construct a YAC library covering chromosome 7 using PFGE and YAC vectors.

2. A physical map of human chromosome 7 will be constructed using YACs that are linked together into contigs covering 98% of the chromosome.

3. RFLP analysis was performed using three generations of individuals from four different families which have a high incidence rate of the disease. A low resolution map of 7q31.3 indicated that the disease gene was likely contained within a ~1mb region.

4. Fibroblast cell lines were established from some of the individuals in the study and these can be used as a source of freshly prepared RNA. All of the relevant STS primer pairs, EST clones and YAC-containing yeast strains can be obtained commercially.

Basic Strategy
DNA samples enriched for chromosome 7 will be subjected to restriction enzyme digested, separated by PFGE and ligated into the pYAC4 vector. The resulting YAC library will be analyzed by sequencing the insert junctions using vector primers. Based on the scaffold assembly of the Human Genome, the YACs will be ordered into contigs (the researchers are at a company and wanted to have proprietary rights to their discovery so did not want to use available YAC clones of chromosome 7). Once they have contigs, they will use PCR analysis with known STS primer pairs to further delineate the region most likely to contain the disease gene.

Based on the results of this linkage analysis, EST clones from the region will be used to recover full-length cDNAs from a human cDNA library. The DNA sequence of these cDNAs, and of the genomic regions encoding the transcribed sequences, will be determined. PCR primers flanking each of the predicted exons of the candidate genes in...
the region will be synthesized and used to obtain sequence information from normal and diseased individuals. Northern blots will be performed using cDNA probes and freshly prepared RNA from the fibroblasts cell lines to identify differences in the expression pattern of these genes in normal and diseased individuals.

**Principle of pulse-field gel electrophoresis (PFGE)**

The principle of PFGE is that large DNA fragments require more time to reverse direction in an electric field than do small DNA fragments. By alternating the direction of the current during gel electrophoresis, it is possible to resolve DNA fragments of 100 -1,000 kb.

Very large DNA molecules "squirm" through agarose like snakes and thus migrate parallel to the electric field, when the current is reversed, it takes more time (T) for large molecules to re-orient than smaller molecules.

Contoured clamped homogeneous electric field (CHEF) systems use a hexagonal gel box that alters the angle of the fields relative to the agarose gel. PFGE gels are loaded with DNA samples that are imbedded within agar blocks to minimize random breakage of large molecules.

**Yeast Artificial Chromosome (YAC) vectors**

YAC vectors use DNA sequences that are required for chromosome maintenance in S. cerevisiae. Plasmids containing these yeast sequences are digested with appropriate restriction enzymes and ligated to PFGE size-fractionated genomic DNA.

**Physical maps provide a framework for positional cloning strategies.**

Hypothetical physical map in the 7q31.3 region where the disease gene has been mapped by RFLP linkage analysis using the RFLP7.1 and RFLP7.2 markers. The relative position of STS primer sets (S1-S8), ESTs (E1-E4) and YAC clones (Y123 and Y124) in this 850 kb region are indicated.

Results of PCR reactions using primer pairs for STS4 and STS5 that detect a polymorphic short tandem repeat. The inferred genotype of normal individuals (open symbol), carriers (half-filled symbols) and patients with the disease (filled symbol) is shown above each lane.

Northern blot results using the ES2 cDNA as a probe and RNA isolated from patient-derived fibroblast cell lines. The actin cDNA probe is used as an internal control to monitor RNA sample loading.

**Comments**

A direct correlation between the presence of the disease (phenotype), and a specific nucleotide(s) alteration in a gene (genotype), would be taken as strong evidence that the metabolic disease is due to a functional defect in the encoded gene product. Confirmation requires DNA analysis of samples from other diseased individuals not represented in the initial set of 45 related family members. Due to the nature of this particular disease, it may even be possible to obtain biochemical evidence that the
encoded protein is required for the a key step in the effected metabolic pathway. The fibroblast cell lines would be expected to have a defect in glucose metabolism since the Northern blot showed that diseased individuals lacked the expression of the putative ES2 disease gene. Biochemical experiments could be performed using extracts from the fibroblast cell lines to determine if addition of the ES2 encoded protein restores functional glucose metabolism.

**Prospective**
The identification of a disease gene provides the opportunity to develop molecular genetic tests for diagnostic purposes. The required information comes directly from the DNA sequence of various genetic alleles identified during the initial gene characterization. In some cases, the elucidation of a disease gene coding sequence provides new insights into possible biochemical mechanisms that would explain the physiological defect. For example, if the ES2 gene product encodes a putative kinase, based on sequence homology comparisons to other kinases in the GenBank database, then it might suggest that altered phosphorylation signaling causes a defect in glucose metabolism. Another research direction could be to develop a mouse model of the human metabolic disease as a means to begin testing potential therapeutic agents. One way to develop such an animal model would be to clone the homologous mouse ES2 gene and then use it as a reagent to create a gene knockout mouse.

What is the definition of an STS and briefly describe the rationale behind using these molecular markers as landmarks to identify genomic regions that may contain disease genes.

Based on the data shown in part B of the figure, what is the approximate length of the S4 and S5 PCR products corresponding to the diseased allele?

Can you determine from these PCR results if the S4 or S5 polymorphism is responsible for the disease phenotype? Explain.

If this couple were to have another child, what is the probability that the child would be afflicted with the disease?

Why would it be easier to map a disease gene that displays a dominant phenotype in heterozygous individuals, than to map a disease gene with a recessive phenotype that requires a homozygous genotype?

Based on the data presented in part c of the figure, what is the molecular basis for the disease?

How would these data explain the observation that heterozygous individuals are clinically normal?

Assuming that ES2 were the only gene to exist between S4 and S5, what would have been the next step in this disease gene hunt if the steady-state level of ES2 RNA were the SAME in all patient samples?
Mapping Gene Regulatory Sequences

Research Objective
A researcher is interested in characterizing the DNA sequence control elements of a gene called Capstone (Cap) thought to be involved in kidney cell function. She has already cloned and characterized the cDNA and mapped the 5’ end of the transcript. Her goal now is to find sequences within the regulatory region of the Cap gene that control its expression uniquely in kidney cells. She has evidence that there are kidney-specific transcription factors that bind to region at least 2-4kb upstream of the putative Cap promoter. Her goal is to identify the precise sequences required for kidney-specific Cap transcription and then use them as affinity reagent to purify sequence-specific kidney transcription factors. Her long term objective is to find a way to re-activate the Cap gene in people who have a rare kidney disease in due to lack of Cap gene expression.

Basic Strategy
The plan is to use in vivo and in vitro mapping strategies to locate that transcription factor binding sites and then use gel shift assays to purify the protein for amino acid sequence determination. She will use four methods in the following order:

1. DNase I hypersensitivity to map chromatin alterations in kidney cells compared to liver cells.


3. In vitro DNA binding using DNaseI footprinting.

4. Relative binding affinities using the electrophoretic mobility gel shift assay.

DNaseI hypersensitivity of chromatin in isolated nuclei is a technique that can be used to map gene regulatory regions. (a) DNaseI hypersensitive site mapping strategy. Enzyme digests and probes are designed to localize sites of DNaseI cleavage in isolated nuclei, based on a genomic DNA restriction map in the vicinity of the putative in vivo binding site.

Nuclei from mouse kidney cells were isolated and treated for various amounts of time (0, 1, 2 or 4 minutes) with DNaseI under conditions that permit limited digestion of chromatin. The reaction was stopped and genomic DNA was extracted and digested with restriction enzymes (XhoI or NotI+ClaI) that release DNA fragments encompassing the region to be mapped. The appearance of DNA fragments shorter than the expected restriction enzyme fragment using Southern blots, was used to map a kidney-specific DNaseI hypersensitive site.

Reporter genes can be used for in vivo mapping of gene regulatory sequences. Genomic segments containing putative gene regulatory sequences can be inserted into reporter
gene plasmids encoding enzymes that are readily measured by biochemical assays such as the firefly luciferase protein.

In vitro DNaseI footprinting is used to map high affinity protein binding sites on naked DNA. An end-labeled double-stranded DNA fragment is incubated in binding buffer with and without a purified DNA binding protein. Limited digestion with DNaseI (~1 cleavage per molecule) produces a population of DNA fragments of various sizes that can be resolved by denaturing acrylamide gel electrophoresis.

The electrophoretic mobility gel shift assay, or EMSA, is another type of in vitro DNA binding assay that is used to map transcription factor binding sites in gene regulatory regions. EMSA is based on the reduced electrophoretic mobility of a DNA-protein complex, compared to unbound DNA.

An autoradiogram of an EMSA assay performed with increasing amounts of kidney cell extracts (left side). The autoradiogram on the right shows that specific Cap binding activity can be identified in kidney cell extracts based on control binding reactions. An antibody against a known transcription factor (CBP) suggest that the kidney-specific binding activity is part of a bigger multi-subunit protein complex that includes the histone acetyltransferase proteins CBP/p300.

**Comments**
Understanding transcriptional control mechanisms is an important part of functional genomics. Developmental (temporal) and cell-specific (spatial) gene expression are hallmarks of multicellular organisms. A combination of physical mapping and functional mapping strategies, using both in vivo and in vitro assays, are required to accurately indentify cis (DNA sequence) and trans (proteins) determinants. In this case, in vivo DNaseI hypersensitivity was used to find the general location of "chromatin perturbations" that reflected kidney-specific expression. The functional significance of this finding was confirmed using reporter gene assays in transiently transfected kidney cells. Finally, evidence was obtained for a protein(s) displaying specific in vitro DNA binding activity using DNaseI footprinting and gel shift assays.

**Prospective**
The identification of a specific DNA binding activity would be the start for biochemical purification of the cognate protein(s). This would most likely be done using the EMSA assay to purify protein fractions that contain the activity until a highly purified protein could be isolated for protein sequencing. An alternative approach would be to using the identified binding site in a yeast one-hybrid assay to find cDNA sequences that encode a Cap gene specific binding protein. Once the cDNA was obtained, either by protein sequencing leads (query the GenBank database for cDNA clones that potentially encode the protein) or by the yeast one hybrid function screening method, it could be used for protein structure ad function studies and antibody production for in vivo immunocytochemistry.

*Based on these data, which segment contains the basal promoter activity required for transcription?*
Which segment of the Cap gene contains an "enhancer" function?

What controls would need to be done in each of these methods to show that the Cap gene binding activity was kidney-specific?

What is thought to be the mechanism of enhancer-mediated transcriptional activation, i.e., how does this upstream DNA sequence function to stimulate transcription?

How could the EMSA assay be used to purify the putative enhancer binding protein identified using the luciferase reporter gene?

What could be done to decrease the interference of non-specific DNA binding proteins in the EMSA assay?

How could the EMSA assay (or in vitro DNaseI footprinting assay) be used to measure DNA binding affinities of a purified protein?

What might explain why the yeast one hybrid strategy would work when biochemical purification by EMSA failed?

Why might EMSA work but the yeast one hybrid strategy fail?