# PHYSICAL MAPPING OF HUMAN CHROMOSOME 16

Sinoula Apostolou B.Sc. (Hons.)

A thesis submitted for the Degree of Doctor of Philosophy to The University of Adelaide

Department of Cytogenetics and Molecular Genetics,
Women's and Children's Hospital, North Adelaide, South Australia

Faculty of Medicine, Department of Paediatrics,
University of Adelaide, South Australia

August, 1997

# CONTENTS

## Physical Mapping of Human Chromosome 16

This thesis involved the construction of a detailed physical map of the distal band of the long arm of human chromosome 16, 16q24. Physical maps are composed of cloned DNA segments which allow chromosomes to be more amenable to detailed analysis. The 16q24 region has been demonstrated to possess a high gene density. These genes include the Fanconi anaemia group A gene (FAA) which has been localised to the 16q24.3 region by classic linkage analysis. Also, loss of heterozygosity (LOH) of the 16q24 region, further refined to 16q24.3-qter, has been demonstrated in sporadic breast tumours. This LOH indicates the presence of a tumour suppressor gene (TSG) in the region. The integration of this physical map with the chromosome 16 genetic map is of great importance as it can be used as a framework map which can benefit the positional cloning of candidate disease genes, including the TSG and FAA, mapped to this specific chromosomal region. During the course of this thesis, the International Fanconi Anaemia/Breast Cancer (FAB) consortium was established to expedite the positional cloning of the FAA and tumour suppressor genes localised to 16q24.3-qter.

The 16q24 region was deficient in markers, thus the first step toward constructing the physical map was the identification of cosmids from this chromosomal region. Cloned DNA segments, 35-45 kb in cosmid vectors, facilitate access to any chromosomal region for further analysis. An Alu PCR strategy was used to isolate DNA, as cosmids, in this region. This technique allows the specific PCR amplification of human DNA of unknown sequence, from complex mixtures of human and rodent DNA. It was applied to the isolation of human specific sequences directly from two human/rodent somatic cell hybrids, CY2 and CY18A, which contain the distal part of 16q as the only human chromosome 16 material. The Alu PCR amplified products from these hybrids were used as hybridisation probes for screening approximately 4000 clones of a gridded chromosome 16 specific cosmid library to enable the

identification of cloned DNAs from this predefined chromosomal region. Los Alamos National Laboratory (LANL) have assembled these cosmid clones into contigs, that is sets of clones which possess overlap, by repetitive sequence fingerprinting. They have been estimated to represent an 84% coverage of the euchromatin of chromosome 16. A total of 32 identified cosmids were confirmed to map to 16q24. An estimation of the amount of 16q24 represented by these cosmids, together with their contigs, is about 2 megabases.

The next step was the use of a selection of the cosmid clones localised to 16q24 to identify transcribed sequences encoded by these cosmids, using the approach of direct cDNA selection. In this study, the direct cDNA selection protocol was modified in an attempt to rapidly isolate longer cDNA fragments and to obtain greater transcribed sequence information. An enriched region specific cDNA library was generated from hybridisation of cDNA inserts to cloned genomic DNA localised to the region of interest. The PCR products from the second round of selection were used as hybridisation probes to screen 40,000 clones of a normalised infant brain cDNA library to identify homologous clones. Five cDNA clones were demonstrated to show specific homology to cosmids from which they were derived. Thus, the newly isolated cosmids were useful in the identification of transcribed sequences. These resources greatly benefit the positional cloning approach for the identification of disease genes assigned to this region of interest.

Subsequently, a collaborative effort led to the identification of additional cosmids mapping to the 16q24.3-qter region. Expressed sequences and microsatellite repeats already mapped to this region, were used as hybridisation probes for screening approximately 14,600 clones of a ten times coverage gridded chromosome 16 cosmid library. Cosmid walking was also performed by probing these filters with cosmid ends to extend the identified singleton cosmids and cosmid contigs in this region. These identified cosmids were assembled into contigs that extended over 650 kb of genomic DNA in the 16q24.3 region.

The cDNA clones localised to 16q24.3 are possible candidates for disease genes localised to this region, including the TSG and FAA. Thus, two novel cDNA clones, yc81e09 and yh09a04, were further characterised. This involved confirmation of their localisation to 16q24.3-qter using a panel of human/rodent somatic cell hybrids. Northern blots and reverse transcription polymerase chain reaction (RT-PCR) were used to determine the full-length sizes of the transcripts and expression patterns in various tissues. Clone yc81e09 demonstrated expression in peripheral blood lymphocytes (PBL), muscle, the frontal lobe and occipital lobe brain sections and a number of cell lines. This clone detected a transcript of approximately 3.7 kb in size. Clone yh09a04 was expressed in PBL, tonsil, muscle, the brain stem and frontal lobe brain sections and numerous cell lines. A transcript of approximately 2.5 kb was detected on a Northern blot. Subsequent Northern analysis from the FAB consortium detected transcripts of several sizes including 2, 3, 4.7, 5.5 and 7.5 kb, the most prominent of which was 4.7 kb in length.

Sequences of the two cDNA clones were obtained and compared to sequences in accessible nucleotide databases to identify overlapping sequences that may extend the sequence of the clones. The sequence of clone yh09a04 was extended with a homologous clone, yf14a03 but the sequence of clone yc81e09 was not extended. The remaining sequence of the 5.5 kb transcript, containing yh09a04, was finally identified by the FAB consortium. Additional sequence for yc81e09 has been obtained by the FAB consortium but more sequence is required to attain the full-length of this transcript. Homologies of two transcribed sequences to any known genes or protein motifs were also investigated to determine function, but no significant homologies were identified.

The identification of a candidate gene that may be responsible for a disease initiates a search for disease causing mutations, an essential step in the positional cloning of disease genes. Single stranded conformation polymorphism (SSCP) analysis of the two transcripts was conducted to determine if their sequence is altered in breast cancer patient samples displaying LOH at 16q24.3-qter, when compared to normal DNA sequence. The FAB consortium

conducted SSCP analysis of the 5.5 kb transcript sequence in Fanconi anaemia (FA) patient samples. Three polymorphisms were identified for transcript yh09a04 and no differences were detected for yc81e09 in the segments that were investigated in the breast tumour samples. Four mutations were detected in the sequence of the 5.5 kb transcript, which included yh09a04/yf14a03, in FA samples. This transcript was therefore determined to be the FAA gene. Thus, the positional cloning strategy involving the physical mapping of cosmids, transcript identification and mutation analysis of candidate genes has been successful in the identification of the FAA gene. Additional work is required for the identification of the TSG.

The approach of direct cDNA selection was also applied to a second project involving the positional cloning of the MEF gene responsible for familial Mediterranean fever (FMF) localised to 16p13.3 by genetic linkage mapping. The international FMF consortium has constructed a YAC/cosmid contig encompassing the FMF candidate region. An enriched cDNA library was generated from hybridisation of cDNA inserts from a foetal brain cDNA library to seven of the cosmids localised to 16p13.3. Analysis of the cloned PCR products from one round of selection identified three transcripts which demonstrated homology to cosmids from which they were derived. These transcribed sequences contribute to the transcript map of the region and the FMF consortium is continuing in its efforts to identify the MEF gene by screening for disease specific mutations in the transcripts.