National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Canada

Molecular analysis of the telomeric half of human chromosome 2q

by

Jing Liu

Department of Medicine
Division of the Experimental Medicine
McGill University
Montréal, Québec
Canada

A thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of the requirements
of the degree of Doctor of Philosophy

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-612-12419-3

Canada

ABSTRACT

The first part of my thesis dealt with the physical mapping of human chromosome 2 employing the yeast artificial chromosome (YAC) cloning system. To generate a chromosome 2 YAC sublibrary, over 1,000 interspersed repetitive sequence (IRS)-PCR probes were generated and used to screen the CEPH midi YAC library and approximately 2,000 chromosome 2-specific midi YACs were identified. These YACs were divided into 223 YAC groups, i.e., sets of unordered overlapping YACs, and using publicly available contig analysis software, a tentative order of YACs within each YAC group could be established. To order YAC groups, the chromosome 2 YAC sublibrary was screened with 87 genetically mapped microsatellites and cytogenetically mapped expressed sequence tags (ESTs), and 44 YAC groups were localized along the genetic map of chromosome 2q. In addition, 16 known genes were physically linked with microsatellites within YAC groups, thus providing integration points for genetic, cytogenetic and YAC-based physical maps of chromosome 2q. In a subsequent step of the analysis, the chromosome 2 YAC mapping data created by the Whitehead Institute (WI)/MIT Genome Center were integrated into our dataset. The integrated dataset consisted of 240 YAC groups, of which 14 large groups containing both our and WI/MIT Genome Center YAC groups were located on chromosome 2q. These 14 groups consisted of 1,195 YACs, which will form the backbone for the construction of a complete YAC contig for human chromosome 2q.

The second part of my thesis dealt with the identification of genetic markers within or in the vicinity of *NRAMP1*, a candidate tuberculosis susceptibility locus. The human *NRAMP1* gene was mapped to chromosome 2q35 by PCR analysis in a monochromosomal hybrid panel and by YAC contig analysis. Nine sequence variants and polymorphisms were identified within the *NRAMP1* gene by single strand conformation analysis (SSCA), DNA sequencing and Southern analyses. Furthermore, two highly informative microsatellites, *D2S104* and *D2S173* were shown to be linked to *NRAMP1* within a 1.5 Mbp YAC contig. Together, these markers provide molecular tools for further genetic analysis of inherited susceptibility to tuberculosis and related diseases of the macrophage.

RÉSUMÉ

La première partie de cette thèse décrit la cartographie du chromosome 2 humain par la technique de clonage en chromosomes artificiels de levure (YAC). Plus de mille sondes de séquences répétitives (PCR inter-*Alu*) ont été générées et ensuite utilisées pour le criblage de la librairie de YAC du centre d'étude du polymorphisme humain (CEPH) dans le but de créer une librairie spécifique du chromosome 2. Environ deux mille YAC ont ainsi été identifiés. Ces YAC ont été assemblés en 223 groupes, chaque groupe représentant des clones de YAC se chevauchant mais non ordonnés. L'utilisation de programmmes d'analyse de segments contigus nous a permis d'établir un ordre provisoire de YAC à l'intérieur de chacun de ces groupes. Afin d'établir une collection ordonnée des différents groupes de YAC, la librairie des YAC du chromosome 2 a été criblée avec 87 microsatellites et EST (expressed sequence tags) préalablement localisés par analyse de liaison ou par cytogénétique sur le chromosome 2q. Quarante-quatre groupes de YAC ont ainsi été localisés sur la carte génétique du chromosome 2q. L'utilisation de 16 ADNc de gènes connus a permis l'intégration de la carte physique dans les cartes génétique et cytogénétique existantes. Dans un deuxième temps, la carte physique créée par le centre du génome du Whitehead Institute (WI) et du Massachusetts Institute of Technology (MIT) a été intégrée à nos données de cartographie physique du chromosome 2. La carte physique ainsi générée est représentée par 240 groupes de YAC dont 14 sont localisés sur le chromosome 2q. Ces 14 groupes sont composés de 1,195 YAC et serviront de squelette pour la construction d'un carte physique couvrant la totalité du chromosome humain 2q.

La deuxième partie de ma thèse décrit l'identification de marqueurs d'ADN situés à proximité ou à l'intérieur même du gène *NRAMP1*, un locus candidat pour la susceptibilité à la tuberculose chez l'humain. La localisation de *NRAMP1* dans la région q35 du chromosome 2 s'appuie sur les analyses d'hybrides d'irradiation et de collections ordonnées de YAC contigus. Un total de neuf marqueurs polymorphes a été identifié dans le gène *NRAMP1* par des analyses de séquence directe, de polymorphisme de longueur des fragments de restriction (RFLP) ou de conformation de simple brin (SSCA). Ces

marqueurs polymorphes, de même que les microsatellites *D2S104* et *D2S173*, deux marqueurs étroitement liés au gène *NRAMP1*, servivont d'outils moléculaires pour l'analyse génétique de la susceptibilité innée à la tuberculose et aux maladies impliquant le macrophage chez l'humain.

PREFACE

My Ph. D. thesis project originated from genetic studies of susceptibility to tuberculosis and leprosy. Specifically, my project involved the generation of genetic markers located on human chromosome 2, region q35, for analysis of linkage with susceptibility to tuberculosis and leprosy. The generation of markers from chromosome region 2q35 required the establishment of a physical map of this region. Thus, during the course of my Ph. D. study, my project was expanded to include a large scale physical mapping of human chromosome 2, particularly the telomeric half of the q arm. This thesis therefore consists of two parts: the isolation of yeast artificial chromosome (YAC) clones carrying inserts derived from human chromosome 2 (Chapters 2 and 3) and the generation of genetic markers on human chromosome 2, region q35, specifically from the candidate tuberculosis susceptibility gene *NRAMP1* (Chapter 4).

The work presented in Chapters 2 and 4 of the thesis has been published, and Chapter 3 is to be submitted.

CHAPTER 2: Liu, J., Stanton, V.P., Jr., Fujiwara, T.M., Wang, J-X., Rezonzew, R., Crumley, J., Morgan, K., Gros, P., Housman, D. and Schurr, E. (1995). Genomics 26:178-191.

CHAPTER 3: Liu, J., Wang, J.W., Simkin, L., Stanton, V.P.,Jr., Fujiwara, T.M., Morgan, K., Gros, P., Housman, D., Schurr, E. (to be submitted).

CHAPTER 4: Liu, J., Fujiwara, T.M., Buu, N.T., Sanchez, F.O., Cellier, M., Paradis, A.J., Frappier, D., Skamene, E., Gros, P., Morgan, K. and Schurr, E. (1995). Am. J. Hum. Genet. 56:845-856.

I have made the most significant contributions for the work described in Chapters 2, 3, and 4.

Chapters 2 and 3 describe the large scale cloning of human chromosome 2 YACs. For this part of the work, my contributions included: copying and transferring the CEPH Mark I YAC library from Dr. D. Housman's laboratory at MIT, Cambridge to our laboratory at McGill, generation of all the *Alu*-PCR filters and most of the *Alu*-PCR probes for screening of the YAC libraries, and selection of STSs for screening the chromosome 2-specific YAC sublibrary. I was also responsible for the data analysis, and the integration of our data with the data generated by the Whitehead Institute (WI)/MIT Genome Center. Drs. Vincent Stanton and David Housman (MIT, Cambridge) shared with us the CEPH YAC libraries and the DNA of YAC pools for the generation of the IRS-PCR filters. Jian-Xue Wang provided technical assistance in screening the YAC libraries. Mary Fujiwara and Dr. Kenneth Morgan helped to initiate the development of the YAC mapping database. Dr. Philippe Gros is the co-investigator of the project. In Chapter 2, Rebeca Rezonzew contributed technical assistance in screening the YAC libraries, Joyce Crumley helped to develop the mapping database. In Chapter 3, Leah Simkin wrote the computer programs for the data analysis.

Chapter 4 describes the identification of polymorphisms and sequence variants in the human *NRAMP1* gene. For this part of the work, I have mapped *NRAMP1*, a tuberculosis susceptibility candidate gene, to chromosome 2, region q35, by PCR analysis of the monochromosomal hybrid panel and by YAC contig analysis. I identified 2 and verified 6 polymorphisms and sequence variants by sequence analysis of human *NRAMP1*. In addition, I positioned two highly informative microsatellites to the vicinity of *NRAMP1* by YAC contig analysis. In this chapter, Natalie Buu did the SSCP analysis. Fabio Sanchez did the Southern analysis. Ann Josee Paradis and Danielle Frappier (from Dr. Kenneth Morgan's group) genotyped the families with the markers. Mary Fujiwara supervised the genotyping, and Dr. Kenneth Morgan performed the statistical analysis of *NRAMP1* association with tuberculosis disease in the parent generation of the tuberculosis

families. Drs. Mathieu Cellier and Philippe Gros provided the primer information for the intron-exon boundaries, and located one microsatellite in the immediate 5' region of the *NRAMP1* gene. Dr. Emil Skamene arranged the collection of DNA samples from the tuberculosis families.

Candidates have the option of including, as part of the thesis, the text of a paper(s) submitted or to be submitted for publication, or the clearly-duplicated text of a published paper(s). These texts must be bound as an integral part of the thesis.

If this option is chosen, connecting texts that provide logical bridges between the different papers are mandatory. The thesis must be written in such a way that it is more than a mere collection of manuscripts; in other words, results of a series of papers must be integrated.

The thesis must still conform to all other requirements of the "Guidelines for Thesis Preparation". The thesis must include: A Table of Contents, an abstract in English and French, an introduction which clearly states the rationale and objectives of the study, a comprehensive review of the literature, a final conclusion and summary, and a thorough bibliography or reference list.

Additional material must be provided where appropriate (e.g. in appendices) and in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported in the thesis.

In the case of manuscripts co-authored by the candidate and others, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. Supervisors must attest to the accuracy of such statements at the doctoral oral defense. Since the task of the examiners is made more difficult in these cases, it is in the candidate's interest to make perfectly clear the responsibilities of all the

authors of the co-authored papers. Under no circumstances can a co-author of any component of such a thesis serve as an examiner for that thesis.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 3

LIST OF TABLES

LIST OF ABBREVIATIONS

ARS	autonomously replicating sequence (in yeast)
BAC	bacterial artificial chromosome
cDNA	complementary deoxyribonucleic acid
CEN	centromere
CEPH	Centre d'Etude du Polymorphisme Humain
cM	centiMorgans
DAPI	4'6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
EST	expressed sequence tag
FISH	fluorescence in situ hybridization
GDB	Genome Data Base
ICRF	Imperial Cancer Research Fund
IFGT	irradiation and fusion gene transfer
IRS	interspersed repetitive sequence
IRS PCR	interspersed repetitive sequence PCR
kb	kilobase
LINE	long interspersed repeated sequence
Mbp	megabase pair
NIGMS	National Institute of General Medical Sciences
NRAMP1	natural resistance-associated macrophage protein 1
PAC	P1-derived artificial chromosomes
PCR	polymerase chain reaction
PFGE	pulse-field gel electrophoresis
PIC	polymorphism information content
Q-banding	quinacrine banding
R-banding	reverse banding
RFLP	restriction fragment length polymorphism
RH	radiation hybrid

SCH	somatic cell hybrid
SINE	short interspersed repeated sequence
SSCP	single strand conformation polymorphism
STRP	short tandem repeat polymorphism
STS	sequence-tagged site
TEL	telomere
VNTR	variable number of tandem repeats
WG-RH	whole genome-radiation hybrid
WI/MIT	Whitehead Institute/ Massachusetts Institute of Technology
YAC	yeast artificial chromosome

ACKNOWLEDGMENTS

First of all, I would like to express my deep gratitude to my thesis supervisor Dr. Erwin Schurr for his constant help and guidance throughout my Ph.D. studies. I am particularly grateful for Erwin's patience, encouragement and tolerance while I was in the lows of my project.

I would like to thank Dr. Philippe Gros for his help in establishing the initial collaboration between Dr. David Housman (MIT, Cambridge) and our lab on the chromosome 2 YAC mapping project. The collaboration provided me with the opportunity to work in Dr. Housman's lab to acquire new techniques and to produce experimental material. I appreciate the kindness and generosity of the members of Dr. Housman's lab, which always made my stay in their lab fruitful and enjoyable. In particular, I would like to express my gratitude to Dr. Vincent Stanton for helpful discussions and valuable advise, and to Dr. David Munroe for unpublished protocols and the *Alu* primers. I would also like to thank Dr. Penny Jeggo (MRC Cell Mutation Unit, Brighton, UK) for providing the chromosome 2 reduced hybrid lines.

I could not pass this opportunity to thank Mary Fujiwara and Joyce Crumley from Dr. Kenneth Morgan's group for their help in the initial stages of data management and analysis, and Li-Hong Liu for technical assistance. In particular, I give my special thanks to Jian-Xue Wang for his technical assistance in screening the YAC libraries and to Leah Simkin for writing the computer programs to analyze the data and helping the preparation of the manuscript. Without their effort, the chromosome 2 YAC mapping project would not have advanced to the point it stands today. I also would like to thank Dr. D. Malo and Natalie Buu for their help in translating the abstract of my thesis into French.

I would also like to thank the members of my Ph. D. thesis committee at McGill: Drs. K. Morgan, D. Rosenblatt, G. Rouleau and R. Blostein for their helpful discussion in the direction of the project and the selection of courses.

I would like to extend my deep appreciation to past and present members of the lab. In particular, I would like to thank Helmuth van Es, Hoe Kim, Louis Garcia, Garry Adams, Ulrike Delling, Zbynek Bozdech, Natalie Buu and Fabio Sanchez, whose understanding, moral support and friendship have made my stay in the lab an enjoyable and unforgettable life experience.

I would also like to take this opportunity to thank Drs. Emil Skamene and Dr. Caroline Pietrangelli for their support and encouragement, and the A. and M. Pietrangelli Travel Award for attending the chromosome 2 workshop in Denmark. I would like to acknowledge the financial support provided by the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR), and the faculty of Medicine Internal Scholarship Award. Support of the work described in this thesis was also provided by operating grants from the Medical Research Council of Canada (MRC), the Canadian Genetic Disease Network (Federal NCE Program), and the Canadian Genome Analysis and Technology Program.

Last but not least, I reserve my special thanks to my family and Pierre for their love, understanding and support throughout the highs and lows of my Ph.D. study.

PUBLICATIONS

Liu, J., Stanton, V.P.,Jr., Fujiwara, T.M., Wang, J.W., Rezonzew, R., Crumley, M.J., Morgan, K., Gros, P., Housman, D., Schurr, E. (1995). Large-scale cloning of human chromosome 2-specific yeast artificial chromosome (YACs) using an interspersed repetitive sequences (IRS)-PCR approach. Genomics 26, 178-191.

Liu, J., Fujiwara, T.M., Buu, N.T., Sánchez, F.O., Cellier, M., Paradis, A.J., Frappier, D., Skamene, E., Gros, P., Morgan, K., Schurr, E. (1995). Identification of polymorphisms and sequence variants in the human homologue of the mouse natural resistance-associated macrophage protein gene. Am. J. Hum. Genet. 56, 845-853.

Blunt, T., Taccioli, G.E., Priestley, A., Hafezparast, M., McMillan, T., **Liu, J.**, Cole, C.C., White, J., Alt, F.W., Jackson, S.P., Schurr, E., Lehmann, A.R., Jeggo, P.A. (1995). A YAC contig encompassing the XRCC5 (Ku80) DNA repair gene and complementation of defective cells by YAC protoplast fusion. Genomics. (In press)

Cellier, M., Govoni, G., Vidal, S., Kwan, T., Groulx, N., **Liu, J.**, Sanchez, F., Skamene, E., Schurr, E., Gros, P. (1994). Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression. J. Exp. Med. 180, 1741-1752.

Levee, G., **Liu, J.**, Gicquel, B., Chanteau, S., Schurr, E. (1994). Genetic control of susceptibility to leprosy in French Polynesia; no evidence for linkage with markers on telomeric human chromosome 2. Int. J. Leprosy 62, 499-511.

Spurr, N.K., Barton, H., Bashir, R., Bryson, G.M., Bushby, K., Cox, S., Gringrich, J.C., Hentati, A., Hildebrandt, F., Kao, F-T., Kruse, T., Lai, E., **Liu, J.**, Menke, M., Naylor, S., Nicklin, M., Read, A., Reiner, O., Rocchi, M., Summer, M. (199/). Report of the Third International Workshop on Human Chromosome 2 Mapping. Cytogenet. Cell Genet. 67, 215-244.

Liu, J., Cellier, M., Schurr, E., Skamene, E. (1993). Comparative genome analysis - a novel strategy to study the genetics of host-parasite interaction. J. Parasitol. 79, 463-469.

van Es, H.H.G., Bout, A., **Liu, J.**, Anderson, L., Duncan, A.M.V., Bosma, P., Oude Elferink, R., Jansen, P.L.M., Roy Chowdhury, J., Schurr, E. (1993). Assignment of the human UDP glucuronosyltransferase gene (*UGT1A1*) to chromosome region 2q37. Cytogenet. Cell Genet. 63, 114-116.

ORIGINAL CONTRIBUTIONS TO KNOWLEDGE

1. Generated a large collection of chromosome 2-specific *Alu*-PCR probes, which, in addition to YAC library screening, also can be used to screen other genomic libraries such as P1, BAC or cosmid libraries.

2. Generated a human chromosome 2-specific YAC sublibrary which consists of approximately 2,000 distinct YACs using the IRS-PCR method. These YACs were divided into 223 YAC groups, i.e. sets of unordered overlapping YACs, which provided the backbone for a YAC contig map of human chromosome 2.

3. STS content mapped a set of genetically mapped microsatellites and cytogenetically mapped ESTs in the chromosome 2-specific YAC sublibrary. These STSs were used to localize YAC groups along chromosome 2. Also, they were used as anchor points to integrate genetic and cytogenetic maps with the YAC contig map.

4. Integrated our chromosome 2 YAC mapping data with the data produced by the WI/MIT Genome Center, resulting in 14 large groups consisting of 1,195 YACs on the telomeric half of chromosome 2q.

5. Linked the human *NRAMP1*, a candidate gene for tuberculosis susceptibility, to *VIL* on a 1.5 Mbp YAC contig and thus, inferred the chromosomal localization of *NRAMP1* to be in chromosome region 2q35.

6. Identified 2 and verified 6 polymorphisms and sequence variants in the *NRAMP1* gene, and linked two microsatellites to *NRAMP1* on a 1.5-Mb YAC contig.

CHAPTER 1:

INTRODUCTION

Large scale mapping of mammalian genomes and genetic studies of innate susceptibility to tuberculosis and leprosy

The human genome has $3x10^9$ base pairs of DNA containing an estimated 50,000 to 100,000 genes. To date, we have knowledge about only a very small portion of these genes. It appears likely that the complete sequence of the human genome will provide pivotal information to understand the genetic basis of human biology and diseases. Indeed, the ultimate characterization of the human genome by sequence analysis is the final goal of the Human Genome Initiative which started at the end of the 1980s.

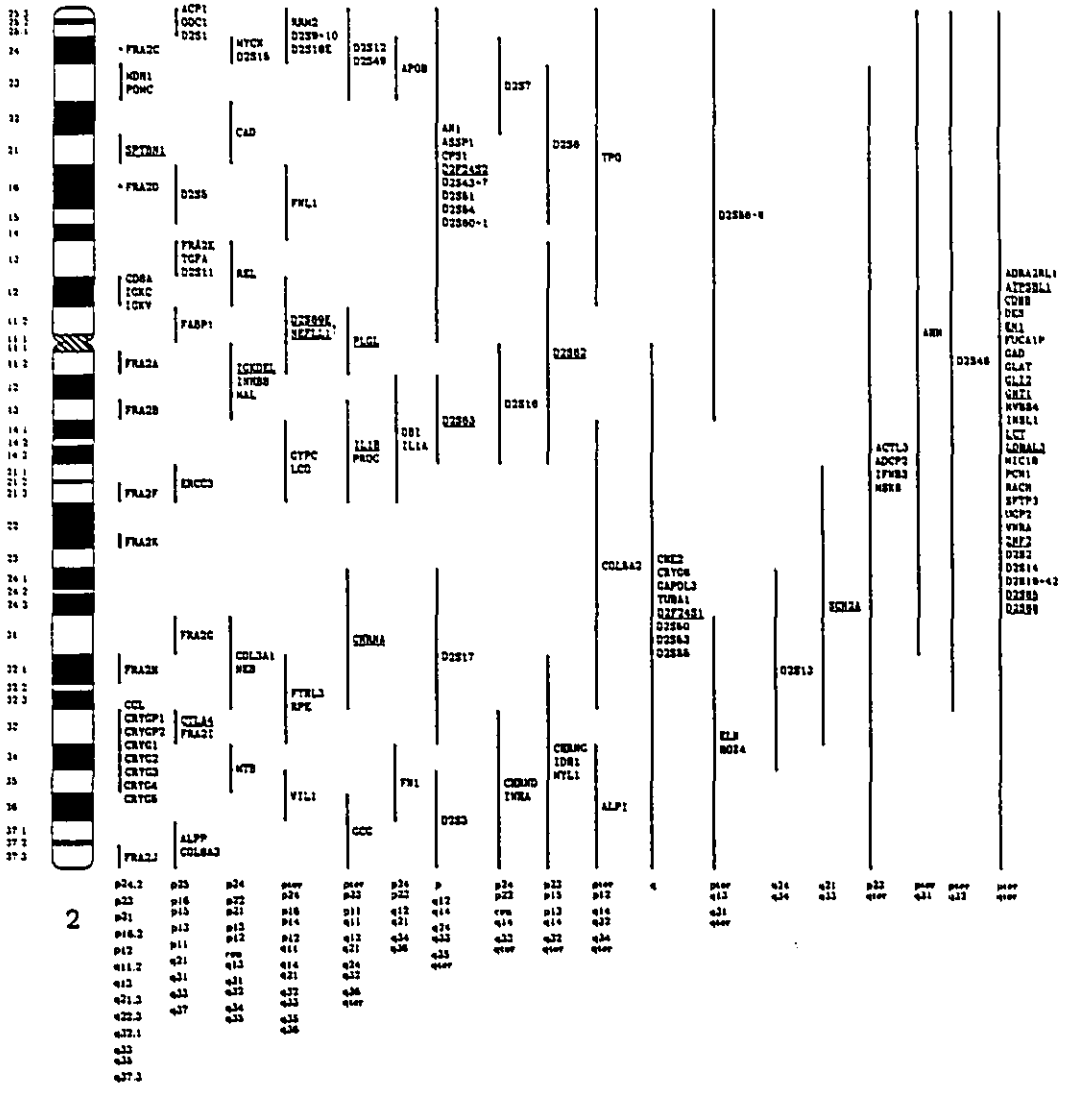
The first step towards sequencing the entire human genome is the construction of physical maps which consist of ordered overlapping cloned fragments of genomic DNA covering each chromosome. Given the large size of mammalian genomes, it is only realistic to clone the entire human genome if there is a vector system which allows the generation of large DNA insert recombinant libraries which are amenable to high throughput library screening. Fortunately, the development of a number of new vector systems including yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), P1-derived artificial chromosomes (PACs), and the advent of novel library screening techniques such as sequence-tagged site (STS) content mapping, interspersed repetitive sequence (IRS)-PCR screening, and high density colony hybridization have made physical mapping of the human genome possible.

Our interest for human chromosome 2 originated from the genetic studies of susceptibility to tuberculosis and leprosy. By comparative genome analysis, the human *Bcg* homologue was suggested to be contained on the telomeric half of human chromosome 2q and closely linked to *VIL*. My original project was the generation of genetic markers localized to the vicinity of the *VIL* gene. During the course of this study, I became involved in the physical mapping and YAC cloning of the chromosomal region surrounding the *VIL* gene. At that time very little chromosome 2-specific physical mapping data were available, eventhough chromosome 2 corresponds to approximately 8% of the human genome (Fig. 1). Thus, it was decided to expand my Ph.D. project to include YAC cloning of chromosome 2 with specific focus on the telomeric half of chromosome 2q.

The aim of our studies was to create a chromosome 2-specific YAC-based physical map. Based on the assumption that data combined from many sources would create the

Figure 1

The compiled human chromosome 2 gene map as of Human Gene Mapping 10 (taken from Cohen et al., in *Genetic Maps*, edited by Stephen J. O'Brien, 1990). Bars are used to indicate a region within which a locus or several loci have been mapped.


2

most accurate YAC-based physical map of chromosome 2, we intended to incorporate our results with data generated through whole genome mapping initiatives. Such an integrated chromosome 2 physical map would not only be useful for gene mapping and identification, but also provide a backbone for establishing a "sequence-ready" physical map for this chromosome.

This chapter therefore consists of two parts: an introductory overview of current techniques and approaches for large scale mapping of mammalian genomes and a short review of genetic studies of innate resistance/susceptibility to tuberculosis and leprosy. In the first part, which is the major focus of the chapter, I will review the development of the major cloning systems and mapping techniques that make it now possible to construct and analyze large-insert recombinant clones and to assemble physical maps encompassing entire mammalian genomes. In the second part of this chapter, I will briefly describe experiments that led to the isolation of the *Nramp1* gene (*Bcg*), which controls innate resistance/susceptibility to mycobacterial infection in the mouse, and the use of a comparative mapping approach to study the genetic component of resistance/susceptibility to tuberculosis and leprosy.

Part I. Large scale mapping of mammalian genomes

1.1. Cloning systems

1.1.1 Cosmids

Cosmids are plasmids that contain a selectable marker, an origin of replication, a cloning site into which foreign DNA can be inserted, and a *cos* site from phage λ. A *cos* site is a DNA segment where a 12-bp cohesive overhang is generated by lambda *ter*. For the construction of recombinant cosmids, the plasmid is cut with a restriction endonuclease, and then ligated with foreign DNA which displays the compatible strand

ends into concatemeric molecules by the enzymatic action of DNA ligase. The ligated DNA is then mixed with a packaging extract containing lambda head proteins, tail proteins, and packaging proteins. Whenever two *cos* sites are present on a concatemer and separated by 40 to 50 kb, these *cos* sites will be cut into two cohesive ends (cos ends), and then packaged into phage heads. The cosmid cloning system thus has the capability to clone foreign DNA as large as 50 kb. Upon injection into the cell, the cosmid DNA circularizes via its complementary cos ends. The ends are covalently joined by the host's ligase, and the resulting circular molecule replicates independently of genomic DNA in the host cell.

Cosmids have been successfully used to construct contiguous overlapping clones (or contigs) and high-resolution physical maps for eukaryotic organisms with small genomes such as *Caenorhabditis elegans*, and fission yeast *Saccharomyces pombe* (Coulson et al., 1986; Hoheisel et al., 1993; Mizukami et al., 1993). More recently, cosmids were also used to construct contigs for specific human genome regions such as the Huntington disease region on human chromosome 4 (Zuo et al., 1993; Baxendale et al., 1993) and entire chromosome 21 (Nizetic et al., 1994; Soeda et al., 1995). However, the relative small insert size (maximum of 40 to 50 kb) has prevented the more widespread use of cosmids as the primary vehicle to obtain long-range contigs for mammalian chromosomes. Nevertheless, a number of features suggest that the cosmid system is useful for high resolution physical mapping of delimited genome regions and for gene identification. First, pure cosmid DNA can be isolated from bacteria by a simple standard minipreparation protocol. This DNA can be used directly to construct a high-resolution restriction enzyme map. Second, cosmid inserts can easily be subcloned into smaller plasmids, which can be used directly as sequencing templates. DNA sequences also can be determined by sequencing directly from cosmid templates using primers complementary to the T3 and T7 promoters present in the vector (Smith et al., 1993). Cosmid contigs are thus considered sequence-ready maps. Third, the use of cosmids can simplify transcript analysis since techniques for isolation of expressed sequences including the 3'-terminal exons by exon amplification from cosmids are well established (Buckler et al., 1991; Krizman and Berget, 1993), and the use of cosmids for direct selection and amplification

of cDNAs (Lovett et al., 1991; Korn et al., 1992) gives a lower background of non-specific hybridization as compared to experiments using yeast artificial chromosomes (YACs) as hybridization probes (Tagle et al., 1993).

1.1.2 Yeast artificial chromosomes (YACs)

1.1.2.1 The YAC cloning system

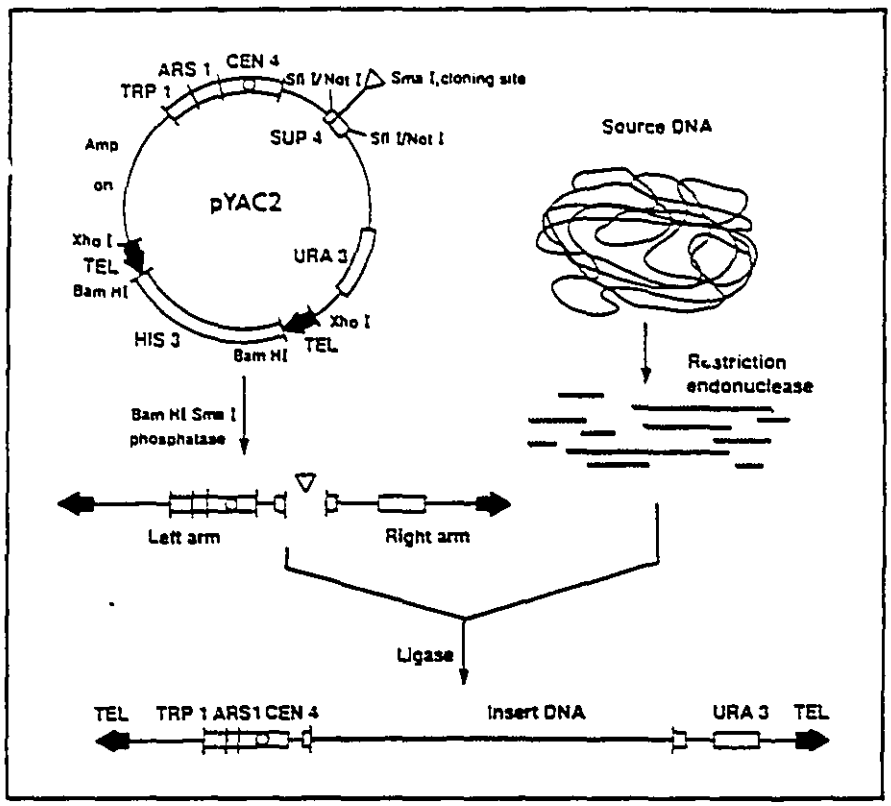
The YAC cloning system is based on the minimal structural requirements for chromosome maintenance in yeast which include a functional centromere, an origin of replication, and two segments containing telomere sequences (Murray and Szostak, 1983). In 1987, Burke, Carle and Olson (1987) developed a vector which incorporated these functions required for the maintenance of a chromosome in yeast into a single plasmid that could be replicated in *Escherichia coli*. This plasmid, termed as 'pYAC2' vector, contained a cloning site within the *SUP4* gene, whose interruption resulted in a phenotypic change from a white color for non-recombinants to a red color for recombinants, an autonomous replication sequence (*ARS1*), a centromere (CEN4), two telomeric sequences (TEL), and selectable markers on both sides of the centromere (*TRP1* and *URA3*). Digestion of the YAC vector with *BamH*I and *Sma*I results in three DNA segments, the left and the right arm, and a "stuffer" region that separates the two TEL sequences in the circular plasmid. The two arms are treated with alkaline phosphatase and then ligated with high molecular weight DNA of up to 1 megabase pairs (Mbp) in length. The ligation products are then transformed into yeast spheroplasts resulting in telocentric YACs. The YAC cloning system thus provides a means for cloning DNA segments one order of magnitude larger than any other cloning systems (Burke et al., 1987; Cohen et al., 1993; Chumakov et al., 1995). The outline of the YAC vector system is shown in Fig. 2.

1.1.2.2 Human YAC libraries

Since the establishment of the YAC technology (Burke et al., 1987), a number of human YAC libraries have been constructed. These YAC libraries can be classified into different categories based on the source of human DNA used. First, whole genome YAC

Figure 2:

The yeast artificial chromosome (YAC) cloning system pYAC2 (taken from Burke et al., 1987). Thin lines represent pBR322-derived sequences. *SUP4*, *TRP1*, *HIS3*, and *URA3* are yeast genes: *SUP4* is an ochre-suppressing allele of a tyrosine tRNA gene that is interrupted when foreign DNA is cloned into the vector; *TRP1* and *URA3* are selectable markers which allow selection for molecules that have acquired both chromosome arms from the vector; *HIS3* is discarded during the cloning process. *ARS*, an autonomous-replication sequence, is the yeast equivalent of an origin of replication. *CEN4* and *TEL* provide centromere and telomere function, respectively.


ARS 1
CEN 4
TRP 1
Sfi I/Not I
Sma I, cloning site
SUP 4
Sfi I/Not I
Amp
on
pYAC2
Xho I
TEL
Bam HI
URA 3
HIS 3
Bam HI
Xho I
TEL
Source DNA
Restriction endonuclease
Bam HI Sma I phosphatase
Left arm
Right arm
Ligase
TEL
TRP 1 ARS1 CEN 4
Insert DNA
URA 3
TEL

libraries which were constructed from DNA of XY males (Brownstein et al., 1989; Traver et al., 1989; Anand et al., 1990; Imai and Olson, 1990; Albertsen et al., 1990; Chumakov et al., 1995). Second, libraries that were targeted to specific chromosomal regions. For example, YAC libraries specific for chromosomal region Xq24-Xq28 were constructed from the DNA of human-hamster hybrid cell lines containing Xq24-Xq28 as their only human material (Abidi et al., 1990). Third, tumor cell-specific YAC libraries, which were constructed from DNA derived from tumor tissues or cells (Feingold et al., 1990), and fourth, chromosome-specific libraries such as the chromosome 21- and chromosome 11-specific YAC libraries (McCormick et al., 1989; Qin et al., 1993). Chromosome-specific YAC libraries were either constructed from monochromosomal human-rodent somatic cell hybrids and then screened with human-specific repeat probes (Wada et al., 1990; Abidi et al., 1990; Green et al., 1995) or from flow sorted human chromosomes (McCormick et al., 1993). An alternative approach to generate a chromosome-specific YAC library is to derive a chromosome-specific YAC collection (or sublibrary) by screening of whole genome YAC libraries with chromosome-specific probes (Chumakov et al., 1992b; Ross et al., 1992; Green et al., 1995).

The average YAC size for clones belonging to the first generation of YAC libraries ranged from 100-300 kb (Guzman and Ecker, 1988; Brownstein et al., 1989; Garza et al., 1989). Using improved techniques such as pulse field gel electrophoresis instead of sucrose gradient fractionation for insert DNA preparation, the size of YACs could be significantly increased (Anand et al., 1989; Imai and Olson, 1990), and several libraries with an average clone size of over one Mbp have been reported (Larin et al., 1991; Albertsen et al., 1990; Cohen et al., 1993; Chumakov et al., 1995). To date, the most commonly used human YAC library is the so-called mega YAC library generated by CEPH/Genethon which is largely made up of clones displaying an average insert size of approximately 1,000 kb (Cohen et al., 1993; Chumakov et al., 1995).

1.1.2.3 YAC chimerism

The YAC cloning system has revolutionized genome mapping and YACs are valuable tools for many purposes, including positional cloning of genes, and construction

of physical maps of human chromosomes (Chumakov et al., 1995). However, the high frequency of chimeric, rearranged, and deleted clones found in YAC libraries has created serious problems for genomic studies, since such clones do not represent a genome region faithfully. It has been estimated that the proportion of chimeric clones - i.e. clones that contain DNA segments which are not adjacent on the chromosome - in mammalian YAC libraries including the widely used CEPH/Genethon mega YAC library, is close to 50% (Bronson et al., 1991; Bates et al., 1992; Dausset et al., 1992; Foote et al., 1992; Cohen et al., 1993; Chumakov et al., 1995). In addition to chimeric clones, the fraction of clones with internal deletions or rearrangements, may also be considerable.

Two mechanisms may be responsible for YAC chimerisms: coligation of multiple genomic fragments into a single YAC molecule, or homologous recombination of repetitive sequences between two or more YACs transformed into same host cell. It is thought that the latter is the major cause of YAC chimerism. An example for the latter mechanism was provided by Green et al. (1991b), who demonstrated that the chimeric junction in at least one chimeric YAC was within an *Alu* sequence and not at the restriction site used for cloning as would be expected for coligation. Recombination between repetitive elements in chimeric YACs is also consistent with the fact that the frequency of chimerism among YACs containing human DNA derived from human-hamster somatic cell hybrid cell lines appears to be significantly lower than that of YACs constructed from total human genomic DNA (Wada et al., 1990; Schlessinger et al., 1991). Presumably, homologous recombination is less likely between human and hamster DNA sequence.

Modifications of YAC library generation protocols aimed at decreasing the rate of chimeric clones include: 1) use of excess vector in ligation reaction to maximize ligation of vectors to all insert ends, and hence decrease coligation of inserts (McCormick et al., 1993); 2) generation of chromosome-specific YAC libraries from rodent-human hybrids, a laborious effort which one would need to repeat for each of the 24 human chromosomes to cover the entire genome; 3) the use of recombination-deficient yeast hosts. Recently, Haldi et al. (1994) constructed megabase-sized human YACs using the recombination-deficient yeast mutant strain *rad52*. Examination of 48 YACs by FISH analysis showed a

rate of chimerism of approximately 8%, in contrast to 50% chimeric clones in the libraries using wild type host AB1380. This result shows that it is possible to significantly decrease the rate of YAC chimerism through the use of appropriate yeast host strains. However, it appears unlikely that any of the alternative techniques mentioned above will have a strong effect on the YAC-mapping of the human genome since mapping efforts in the CEPH YAC libraries have advanced to the point that all but excludes mapping in additional YAC libraries.

1.1.3 Bacteriophage P1 and P1-derived artificial chromosomes (PACs)

1.1.3.1 The bacteriophage P1 cloning system

The bacteriophage P1 cloning system which allows cloning of DNA fragments up to 100 kb was developed in 1990 (Sternberg, 1990; Sternberg et al., 1990). The first generation of P1 vectors such as pAd10 contains a P1 packaging site (*pac*) to initiate the packaging of vector and cloned DNA into phage particles, and two P1 *loxP* recombination sites present in the vector arms flanking the cloned insert. The two *loxP* sites are necessary for circulation of the packaged DNA by the *Cre*-mediated site-specific recombination mechanism following phage infection of *E. coli* cells. The P1 vector also contains a *kan*^{r} gene to select bacterial clones containing the cyclized DNA, a P1 plasmid replicon to maintain P1 DNA in *E. coli* at one copy per cell, and a *lac* promoter-regulated P1 lytic replicon to amplify P1 DNA to 20-30 copies per cell.

A major problem of the first generation P1 libraries was that 10-20% of clones in the libraries contained insertless P1 vector DNA, and these clones outgrew the recombinant clones (Sternberg, 1990; Pierce et al., 1992). To overcome this problem, a second generation of P1 vectors (e.g. pAd10SacBII) incorporating the *SacB* gene, which encodes an enzyme that catalyzes the hydrolysis of sucrose to products lethal to *E. coli*, was constructed (Pierce et al., 1992). In this vector, the cloning site is between *SacB* and its promoter. Clones containing inserts separate *SacB* from its promoter, which inhibits the expression of the *SacB* gene and hence permits the growth of *E. coli* on medium containing sucrose. Thus, recombinant clones are kanamycin and sucrose resistant. In

addition, the cloning site is flanked by rare-cutting restriction enzymes, which can be used for releasing the insert. The T7 and Sp6 promoters in the vector also permit the generation of RNA probes from the vector ends (Pierce et al., 1992).

1.1.3.2 P1-derived artificial chromosomes (PACs)

It was shown that P1 vectors can only accommodate fragments in the 70 kb to 100 kb range, presumably due to the headful capacity of P1 (Sternberg, 1990). The P1 system is also restricted by the requirement for an elaborate *in vitro* packaging system (Sternberg, 1990; Pierce et al., 1992). Recently, an improved cloning vector termed P1-derived artificial chromosomes (PACs) was constructed by deletion of the adenovirus stuffer fragment and by insertion of a modified pUC19 plasmid into the *BamH*I cloning site of the pAD10sacBII vector (Ioannou et al., 1994). In addition, PACs are introduced into *E. coli* via electroporation, hence, avoiding elaborate packaging of DNA and *in vivo* site-specific recombination as required by previous P1 vectors. The PAC vector is able to accommodate inserts up to 300 kb, more than double the maximum insert size described for the P1 vector (Ioannou et al., 1994). The P1 and PAC cloning systems have an intermediate cloning capacity between that of the YAC and cosmid systems.

1.1.4 Bacterial artificial chromosomes (BACs)

The BAC cloning system which is capable of carrying foreign DNA fragments in excess of 300 kb is based on the *E. coli* fertility F factor plasmid (Shizuya et al., 1992). The F factor, whose replication in *E. coli* is strictly controlled, codes for genes that are essential to regulate its own replication. The BAC vector (pBAC) developed by Shizuya et al. (1992) incorporates all of the F factor regulatory genes including *oriS*, *repE*, *parA*, and *parB*. The *oriS* and *repE* genes mediate the unidirectional replication of the F factor, while *parA* and *parB* control copy number to one or two plasmids per *E. coli* cell, thus, decreasing the potential for recombination between DNA fragments carried by the plasmid. In addition, the pBAC vector contains a cloning region which includes two cloning sites (*Hind*III and *BamH*I), flanked by T7 and Sp6 promoters, which can be used

to generate RNA probes from the vector-insert junction, and several rare cutter restriction enzyme sites which can be used to excise the insert.

Similar to the P1 or PAC system, the BAC system also has an intermediate cloning capacity between that of the YAC and the cosmid. Recently, several whole human genome BAC libraries and a human chromosome 2-specific library (Rouquier et al., 1994; Ashworthe et al., 1995; Wang et al., 1994) were constructed. The characterization of these libraries indicated that the BAC cloning system has high transformation efficiency (10-100 times higher than that of YACs) and a low frequency of chimeric clones as a result of low vector copy numbers. However, a critical appraisal of the BAC cloning system in genome mapping still awaits further investigation.

1.2 Mapping techniques

1.2.1 Sequence-tagged site (STS) content mapping

Sequence-tagged sites (STSs) are short unique genomic fragments amplified by PCR with two defined primers. STS content mapping is based on the logic that if an STS is unique to a particular genomic region, then all the YACs in a library containing this STS must carry inserts which are derived from the same genomic region and therefore must overlap. STS content mapping can thus be used to determine which STSs are contained on which YACs, and to construct sets of contiguous overlapping clones (or YAC contigs) out of those YACs containing the same STSs.

STS content mapping has been and still is widely used to identify YACs overlapping candidate regions of disease genes (Green and Olson, 1990a; Huntington's Disease Collaborative Group, 1993). Since most of YAC libraries contain 5-6 x 10^4 clones, the identification of YACs corresponding to specific loci by STS content mapping is generally accomplished by reiterative analysis of a hierarchy of clone pools. The original protocol for STS content mapping of pooled DNA was developed by Green and Olson, and was based on a three-step process (1990b). In this protocol, the first step was a PCR screening of primary DNA pools. A primary pool was made up of 5 sub-pools, each containing 384 clones. In the case of a positive signal detected in a primary DNA

pool, those corresponding 5 sub-pools were subjected to PCR screening. The last step for single-clone identification was accomplished by hybridization of probes to a filter containing the appropriate 384 clones. This step was subsequently replaced by PCR analysis with DNA pools collected from row and column of the 4 plates which corresponed to the positive sub-pool (Chinault and Sternberg, 1994). Although the complexity of pools and the number of successive rounds of PCR screening may differ from one library to the other, the general hierarchical screening procedure for PCR screening of YAC libraries remains the same and has been widely used in genome mapping.

Since STSs are defined by pairs of oligonucleotides and PCR conditions which can be stored in publicly accessible databases, STSs are considered a "common language" in genome mapping (Olson et al., 1989), and STS content mapping is accepted as one of the major physical mapping methods. Indeed, a primary goal for the physical mapping of the human genome is the development of an STS map with an average spacing of 100 kb between consecutive STSs (Collins and Galas, 1993). STS content mapping has initially been used to assemble YAC contigs for the human Y chromosome (Foote et al., 1992) and human chromosome 21 (Chumakov et al., 1992a). In the later case, Chumakov et al. (1992) used 198 human chromosome 21 specific STSs to screen a whole genome YAC library and isolated 810 YACs, which could be grouped into a contiguous array of overlapping clones covering the entire human chromosome 21q. Recently, STS content mapping has been used either alone or in combination with other mapping methods to construct physical maps for a number of human chromosomes including chromosome 3, 4, 7, 10, 11, 12, 13, 16, 19, 21, 22, the X and Y chromosomes, and the whole human genome as reported during the 1995 Cold Spring Harbor Genome Mapping and Sequencing Meeting.

1.2.1.1 Generation of STSs

The application of STS content mapping to mammalian chromosomes or entire genomes requires the generation of large numbers of STSs. For example, about 2,000 STSs are required to construct a map with an average interval of 100 kb between STS

markers for a chromosome of 200 Mbp and a map of the whole genome with the same resolution will require the development and mapping of at least 30,000 STSs. STSs can be obtained from many sources. They can be polymorphic genetic markers such as microsatellites, tri-, or tetra-nucleotide repeats, or genomic fragments randomly derived from either the whole genome or specific chromosomal regions. STSs can also be sequences derived from well characterized cDNA probes or expressed sequence tags (ESTs) stored in public databases such as GDB or dbEST. Like other aspects of genome analysis, the large scale generation of STSs is characterized by a highly integrated data generation and analysis laboratory flow management system. Typically, the first step in such an integrated system is to obtain specific DNA sequences from various sources. This is followed by analysis of sequences to eliminate repetitive elements and sequence duplication, and design and test of PCR primers in standard PCR assays, which includes controls to verify the genomic origin of each STS. Such data flow systems have been used to generate STSs for chromosome 4 (Goold et al., 1993), chromosome 7 (Green et al., 1991a), chromosome 11 (Smith et al., 1993), chromosome 21 (Chumakov et al., 1992a), and chromosome 22 (Bell et al., 1995).

1.2.2 Interspersed repetitive sequence (IRS)

1.2.2.1 IRS in the human genome

In the human genome, SINES (short interspersed repeated sequences) and LINES (long interspersed repeated sequences) are two classes of the most prevalent interspersed repetitive sequences (IRS), which together account for at least 20% of the total human genomic DNA (Jelinek and Schmid, 1982; Singer and Skowronski, 1985; Weiner et al., 1986; Scott et al., 1987; Moyzis et al., 1989). The SINES class is dominated by the *Alu* DNA sequence family, which has a consensus sequence of about 300 bp. It is estimated that 500,000-1,000,000 copies of *Alu* sequence are dispersed in the human genome (Deininger and Schmid, 1976). Thus, on average, one *Alu*-element occurs every 3 to 6 kb in the human genome (Rinehart et al., 1981; Hwu et al, 1986). The human LINES class is dominated by L1 DNA elements which have a consensus sequence of about 6.4 kb. In

contrast to *Alu* sequences which have a similar length, many LINES sequences have serial deletions of the 5' end or internal deletions and rearrangements resulting in considerable length variations among individual L1 repeats. Nevertheless, most of L1-elements share their 3' end sequences. There are about 4,000-20,000 copies for the 5' end and 50,000-100,000 copies of the 3' end of L1 elements in a haploid human genome (Grimaldi et al., 1984; Hwu et al., 1986). Thus, on average, an L1 repeat occurs in the genome once per 30-60 kb.

1.2.2.2 *Alu* family members have a consensus sequence

Alu-elements are ancestrally derived from the 7SL RNA gene (Ullu et al., 1984). *Alu*-elements consist of approximately 300 bp and share a high degree of homology, which allowed the derivation of a consensus sequence corresponding to the entire length of *Alu* elements (Schmid and Jelinek, 1982; Fig. 3a). The human *Alu*-element is an imprecise dimer of two directly repeated, approximately 130 bp monomer sequences with a 31 bp insertion in the second monomer. The murine equivalent sequence (termed as B1 repeat) is 130 bp long, corresponding to a monomer of the human *Alu*-element while the hamster sequence is 134 bp long (Schmid and Jelinek, 1982). All rodent *Alu*-like sequences show considerable homology with one half of the human *Alu* sequence. Presumably *Alu* and B1 elements are descendants of a common ancestral sequence that has been well preserved during recent evolution.

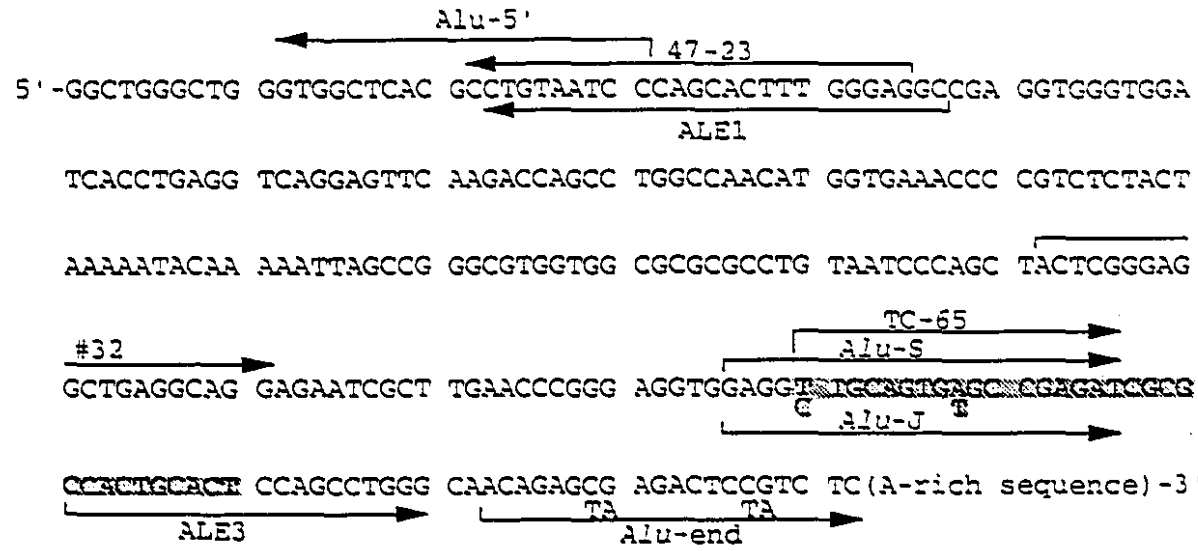
1.2.2.3 The distribution of *Alu* sequences

The distribution of interspersed repetitive DNA sequence elements in the human genome has been investigated by several groups. By *in situ* hybridization, Korenberg and Rykowski (1988) found that the *Alu* family dominated in reverse R bands, and the L1 family dominated in Giemsa/Quinacrine positive bands. These authors concluded that with the exception of some telomeric regions, *Alu*- and L1-elements were precisely inversely distributed along the chromosome. Baldini and Ward (1991) also observed that *Alu* elements were not evenly distributed along the human chromosomes, but rather concentrated in R-bands. In contrast, Moyzis et al. (1989) observed that repetitive

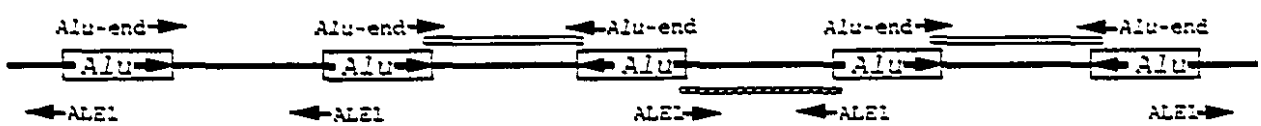
Figure 3:

Consensus *Alu* sequence and scheme for *Alu*-element mediated amplification of IRS. (A) Consensus *Alu* sequence and location of selected *Alu* primers. The orientation of the primers is indicated by arrows. The 31-bp primate-specific region is highlighted in gray. (B) Diagram showing *Alu* PCR amplification with primers *Alu*-end and ALE1. (This figure is adapted and modified from Nelson et al., 1989).

A


Alu-5'
47-23
5'-GGCTGGGCTG GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGCCGA GGTGGGTGGA
ALE1
TCACCTGAGG TCAGGAGTTC AAGACCAGCC TGGCCAACAT GGTGAAACCC CGTCTCTACT
AAAAATACAA AAATTAGCCG GGCGTGGTGG CGCGCGCCTG TAATCCCAGC TACTCGGGAG
TC-65
#32
Alu-S
GCTGAGGCAG GAGAATCGCT TGAACCCGGG AGGTGGAGGT
Alu-J
CCAGCCTGGG CAACAGAGCG AGACTCCGTC TC(A-rich sequence)-3'
ALE3
Alu-end

B


Alu-end
Alu
ALE1

sequences are generally randomly spaced, with an average distance of 3 kb in the human genome. However, local regions of 'preference' or 'exclusion' for the integration of specific classes of interspersed repeat families, such as *Alu* sequences, were also suggested by their data. Another study was conducted by examining the content of *Alu* and L1 sequences in a group of 435 YACs all arising from the Xq24-q28 chromosomal region (Porta et al., 1993). These authors observed a random overall distribution of the *Alu* and L1 repeat sequences, and no apparent enrichment of *Alu* or L1 in DNA of G bands compared to that found in R bands. However, they also found clusters of repetitive elements in delimited regions of up to 100 kb. These authors suggested that the local concentration of repetitive elements may account for the Korenberg and Kykowski finding of differential hybridization of *Alu* and L1 probes to segments of metaphase chromosomes. By sequencing of randomly selected DNA clones from R- and G- bands on chromosome 11, Yoshiura et al. (1993) also did not observe an uneven distribution of *Alu* sequences between R- and G- bands on this chromosome. In summary, localization studies so far have failed to provide a clear and consistent picture with regard to the distribution of repetitive sequences in the human genome.

1.2.2.4 IRS PCR: application in human genome mapping

IRS PCR is a simple technique to amplify genomic fragments between two suitably spaced and oriented IRS (e.g. *Alu* elements; Fig. 3b). IRS PCR was first demonstrated by Nelson et al. (1989) for the *Alu* element (*Alu* PCR or *Alu*-element mediated PCR) to be an effective means of producing human-specific DNA fragments from human/rodent somatic cell hybrid DNA. The principle of this method is based on the fact that although the repetitive sequence in rodent genomes have a high degree of homology to the human *Alu* repeat (up to 70-80%), there is significant sequence divergence through mammalian evolution. Primers can be designed for the primate-specific regions of *Alu* elements which allow specific amplification of human sequences from hybrid cells on mouse or hamster backgrounds. Thus, *Alu* PCR allows the generation of human specific-sequences from somatic cell hybrids without prior knowledge of the human DNA sequence. The simple *Alu* PCR technique was soon widely adapted for rapid generation of DNA probes from

somatic cell hybrids carrying specific chromosomal regions (Ledbetter et al., 1990; Brooks-Wilson et al., 1990; Cole et al., 1991). *Alu* PCR can also be applied to a variety of other DNA sources including YACs, P1s, BACs, and cosmids.

Recently, a new technique termed IRS-bubble PCR was developed (Munroe et al., 1994). Unlike IRS PCR, which requires a pair of IRS in proper orientation and distance with respect to each other, amplification by IRS-bubble PCR requires only a single IRS. Genomic DNA from somatic cell hybrids, YACs, cosmids, or lambda phages is digested with a 4-bp blunt-cutting restriction enzyme (e.g. *Rsa*I, *Alu*I, or *Hae*III) and then ligated to a "bubble" anchor (a pair of preannealed primers with a non-complementary nucleotide bubble in the middle). If a bubble ligated genomic fragment contains a single IRS, an IRS-specific primer is used to direct the first-strand synthesis of the region between the IRS and the annealed "bubble" oligos. The bubble primer and the IRS-specific primer then direct the second-strand synthesis in subsequent rounds of amplification. The resulting amplification product is at least 10-fold more complex than that produced by standard IRS PCR. Thus, this technique is useful for the generation of complex IRS PCR products from YACs, which can be used as probes for FISH analysis or for screening of other genomic libraries such as cosmid or PAC libraries.

Alu PCR was first applied to genome mapping by Nelson et al. (1991), who used the *Alu* PCR method to generate probes from 110 X chromosome specific YACs. These probes were then hybridized to somatic cell hybrid mapping panels for the X chromosome, which allowed the regional assignment of the 110 YACs. More recently, the *Alu* PCR approach was used as one of three methods in the construction of a YAC contig map of the human genome (Cohen et al., 1993; Chumakov et al., 1995). In this study, inter-*Alu* PCR products for 25,000 YACs and a monochromosomal somatic cell hybrid mapping panel were derived by PCR, and spotted at high density onto membranes. The membranes were then hybridized with inter-*Alu* PCR probes derived by amplification of individual YAC DNA, which allowed the identification of overlapping YACs. The information for the overlapping YACs was then used to assembly YAC contigs. Chumakov et al. (1992b) also used the *Alu* PCR technique to generate a chromosome 21-specific YAC sublibrary. These authors performed *Alu* PCR using DNA from two mouse-human somatic cell hybrid

lines containing chromosome 21 as the only human chromosome DNA to generate *Alu* PCR probes. These probes were then hybridized to membranes dotted with *Alu* PCR amplification products derived from individual YACs of a whole genome YAC library. Using this inter *Alu* element hybridization approach, they identified 63 chromosome 21-specific YACs which constituted a chromosome 21 YAC sublibrary.

1.2.2.5 *Alu* PCR primers

Since the introduction of the *Alu* PCR technique, a number of *Alu* primers have been described and used. These primers include: TC-65, 278, 32, 33, 34, 515, 517 (Nelson et al., 1989), PDJ66 and PDJ67 (Aslanidis and de Jong, 1991); *Alu* 559 (Ledbetter et al., 1990), Al (Brooks-Wilson et al., 1990); ALE34, ALE3, ALE1, and ALI (Cole et al., 1991), *Alu*-5' and *Alu*-3' (Tagle and Collins, 1992); A33 (Chumakov et al., 1992b), *Alu*-S, *Alu*-J, 47-23, and *Alu*-end (Munroe et al., 1994). The location of the targeted sequences within the *Alu* consensus for selected primers is indicated in Fig. 3a. Some primers (for example, TC-65, 517, 515, *Alu*-S and *Alu*-J) are targeted to the 31 bp primate-specific sequence in the second repeat of the human *Alu* consensus sequence, and allow specific amplification of primate sequences in human-rodent somatic cell hybrids. However, the 31 bp primate-specific sequence is located 74 bp upstream from the 3' end of the *Alu* repeat element. The *Alu* PCR amplification products using primers derived from the 31 bp region thus contain approximately 74 bp *Alu* sequences at both ends (the exact length of the repeat sequences depends on the targeted position of the *Alu* primer within the 31 bp region). The use of such *Alu* PCR amplification products as hybridization probes requires the suppression of the *Alu* repeat sequences contained in the probes by preannealing to an excess of sheared denatured human DNA. Some *Alu* primers are derived from the very 5' (for example, *Alu*-5', and ALE1) or 3' (for example, *Alu*-3', *Alu*-end, and A1) ends of the *Alu* repeat elements. The amplified inter *Alu* PCR products obtained by *Alu*-PCR using these primers are free of *Alu* repeat sequence. Primers which are derived from the portion of the *Alu* repeat that is conserved among mammalian *Alu*-equivalents (for example, *Alu* primers 278, 32, and 34) can not be used to obtain human specific amplification products from somatic cell hybrid DNA. These

primers are only useful for amplifying inter *Alu* sequences from cloned DNA contained in YACs, P1s, BACs, or cosmids.

1.2.3 Clone DNA fingerprinting

A common type of clone DNA fingerprint is a pattern of DNA fragments resolved on agarose gel which have been generated by cleavage of cloned DNA with restriction endonuclease(s). In the mid-1980s, three pioneering projects were initiated to establish the order of overlapping clones for all or part of the genomes of *C. elegans* (Coulson et al., 1986), *S. cerevisiae* (Olson et al., 1986), and *E. coli* (Kohara et al., 1987) based on DNA fingerprinting of randomly chosen clones. Overlap between clones was inferred when these clones shared similar fingerprints. For example, Olson et al. (1986) fingerprinted 5,000 lambda clones containing inserts with an average size of 15 kb of yeast DNA by a double restriction enzyme digestion with *EcoR*I and *Hind*III. A database was constructed that contained the size of restriction fragments from all clones which was used for pairwise comparison of DNA fingerprints of all clones. Based on the results of the pairwise analysis, all clones were grouped into subsets, in which each member was related to at least one other member by a significant overlap of the fingerprints. A computer program was then used to generate restriction maps for all the clones in each subset. The result of this study showed that the DNA fingerprinting approach could be used to produce regional maps, which could serve as a framework for the construction of a continuous global map of the yeast genome.

Until the early 1990s, attempts of large scale physical mapping by fingerprinting had only been applied to organisms with small genome size such as *C. elegans*, *S. cerevisiae* and *E. coli*. With the development of the YAC cloning systems (Burke et al., 1987) and the repetitive-element hybridization fingerprinting technique (Cangiano et al., 1990; Stallings et al., 1990; Bellanne-Chantelot et al., 1991), it became possible to map the entire human genome by DNA fingerprinting. The principle of the repetitive-element hybridization fingerprinting is the following: a Southern blot of restriction enzyme digested YAC DNAs is hybridized with a selected class of repetitive sequences such as *Alu*, L1, or simple sequence repeats such as (GT)n, or (AC)n. Since *Alu* and L1 repeat

probes are specific to mammalian genomes, and simple sequence repeats only hybridize with a few common sequences in yeast genomic DNA, the use of such repeat-element probes allows the distinguishing of human inserts from the background of yeast DNA. Thus, the combination of restriction enzyme digestion of YAC DNA and hybridization with repeat-element probes produce fingerprints of individual YAC clones, which allowed the detection of overlapping regions of individual YACs by comparing the pattern of bands shared among YACs. Indeed, in 1992, a YAC-based physical map of entire human genome which relied on repetitive-element hybridization fingerprinting was reported (Bellanne-Chantelot et al., 1992). In this study, individual patterns of restriction fragments for 22,000 YACs with average size of 810 kb were obtained by restriction enzyme digestion for all clones. Restriction fragments were size separated, transferred to membranes and hybridized with an L1-probe. The resulting pattern of L1-containing restriction fragments was compared among all clones and more than 1,000 contigs which covered an estimated 15% to 20% of the human genome were assembled.

Despite these successes, DNA fingerprinting is not the first choice for the construction of long contiguous sequence maps because it entails several labor-intensive steps. However, fingerprinting has been and still is widely used to order and analyze clones within sets of overlapping clones in order to obtain high-resolution physical maps (Baxendale et al., 1993; Zuo et al., 1993; Patil et al., 1994). In the past few years, *Alu* PCR fingerprinting which analyzed the pattern of *Alu* PCR amplification products produced by individual clones was developed (Nelson et al., 1991). The *Alu* PCR fingerprinting technique has been used to determine the extent of overlap between large insert clones such as YACs, and to determine the orientation of a set of overlapping YACs relative to each other (Nelson et al., 1991; Sidhu et al., 1992).

1.2.4 Fluorescence in situ hybridization

Fluorescence in situ hybridization (FISH) is a technique that allows the localization of genes or DNA sequences along a chromosome, typically to a specific chromosomal band. Chromosome bands are reproducible landmarks along the length of metaphase chromosomes, which are used to distinguish individual chromosomes within a genome,

and to identify a specific subregion of individual chromosomes. Chromosome bands can be produced using a number of banding techniques such as G-banding (Schweizer, 1976; Tucker et al., 1988), or R-banding (Cherif et al., 1990). For example, the G-banding pattern, which is a series of light and dark bands along the chromosomes, can be produced by pretreatment of metaphase chromosomes with trypsin, followed by Giemsa staining. G-banding can be viewed and photographed with a bright-field microscope. FISH analysis is based on the deposition of fluorescent molecules in chromatin at the sites of specific DNA sequences. Generally, specific DNA or RNA probes are labeled with reporter molecules such as biotin. The biotinylated probe is hybridized to denatured metaphase chromosomes or interphase nuclei. Following incubation with fluorescein-avidin, a discrete fluorescent signal is visible at the site of probe hybridization. The probe hybridized chromatid spreads are then stained with DAPI, which permits chromosome identification by the G-band like pattern, and the hybridized probe can be viewed simultaneously with flurorescein. The chromosomal band location of a probe can then be determined by comparing the position of the fluorescence hybridization signal relative to the DAPI-banded chromosome image.

As mentioned, the targets for FISH analysis can be metaphase chromosomes or interphase nuclei, resulting in different levels of mapping resolution. FISH to metaphase chromosomes allows DNA probes to be ordered with a 2-3 Mbp resolution (Lichter et al., 1990; Lawrence et al., 1990). FISH to interphase nuclei was first introduced by Lawrence et al. (1988), who hybridized Epstein-Barr virus infected nuclei with probes corresponding to two viral DNA regions separated by 130 kb. Using interphase chromatin, the relative order of probes separated by as little as 50 kb can be determined (Lawrence et al., 1988; Trask et al., 1989; Lawrence et al., 1990; Trask et al., 1991). Probe ordering can be derived either from interphase distances measured between pairs of probes (Trask et al., 1989; Trask et al., 1991; Chang et al., 1993) or from multicolor probe labeling (Trask et al., 1991).

Since its development, the FISH technique has been used for a wide variety of experiments including the investigation of the organization of interspersed repetitive sequences in the human genome (Korenberg and Rykowski, 1988), the chromosomal

mapping of genes (Trask et al., 1989; 1991; Lichter et al., 1990; Lawrence et al., 1990; Malo et al., 1993b), the detection of chimeric YAC clones (Chang et al., 1993; Moir et al., 1994), and the chromosome regional assignment of YACs or cosmids (Chang et al., 1993; Cohen et al., 1993; Moir et al., 1994; Lu-Kuo et al., 1993; Ariyama et al., 1995). These studies show that FISH has played an important role in genome research particularly for the localization of DNA sequences to specific chromosome regions.

1.2.5 Colony hybridization

Colony hybridization allows the identification of filter grown and lysed recombinant bacterial or yeast clones by hybridization with radio-labeled probes. The use of colony hybridization in large scale physical mapping is based on robotic systems that produce high density filter grids of cosmids or YACs (Nizetic et al., 1991; Ross et al., 1992). For example, Ross et al., (1992) used a robotic device to spot 20,736 YACs onto a 22 x 22 cm nylon membrane. After two days of growth, filters carrying yeast colonies were subjected to a yeast colony lysis procedure and DNA was fixed to the filters. The filters were then ready to be used for hybridization experiments (Ross et al., 1992).

One advantage of colony hybridization is that the hybridization of high-density gridded clone filters allows the use of complex probes, i.e. a probe composed of many different DNA fragments, to identify clones with inserts from large regions of a chromosome in a single hybridization experiment. As demonstrated by Ross et al. (1992), 388 chromosome 21-specific YACs were identified in a single experiment by hybridization of a complex probe derived from chromosome 21 to a high-density gridded array of YACs on a nylon membrane from a whole genomic human YAC library. Another advantage of the technique is that it does not require sequencing or oligonucleotide synthesis for marker development. Indeed, probes used for colony hybridization can be cDNAs, RFLP probes, or *Alu*-PCR products derived from a variety of DNA templates. The limitations of this method are that hybridization signals are weak for YAC colony hybridization, since the YAC DNA only represents approximately 5% of the total yeast DNA, and that membranes can be reused only a limited number of times.

1.3 Maps

A map is a graphic representation of a milieu usually on a plane surface. Maps are distinguished from other forms of graphical presentation by rigorous application of objective rules for milieu description. Indeed, one can say, that the pivotal step in map creation (mapping) is measurement. Measurement defines the spatial relationship of objects which are usually chosen by convenience, logic and/or necessity to best represent certain features of the milieu. The choice of objects frequently dictates the unit of measurement, and in crude maps, objects themselves sometimes can represent the units of measurement. A map usually represents only one aspect of a milieu which for subjective reasons was defined tc be satisfying or useful. The dichotomy of objective measurements and subjectively chosen objects is a basic feature of any map. A map also has a resolution which is defined by the minimal distance between two objects that can be described accurately in the map. Finally, the change of most milieus with time, the subjectivity of object choice and improvements on the units of measurements are the main reasons why there is never a finite map of a milieu.

The commonly used maps in genome research include physical-, genetic-, cytogenetic-, radiation hybrid-, transcription- and comparative genome- maps. The milieu for these maps generally is the chromosome. Depending on the type of map, the units of measurements are centiMorgan (cM), basepair (bp), centiRay (cR), chromosome bands, cosmids and YACs, while the objects can be loci, markers, STSs or IRS PCR probes. Since each map represents only one aspect of a chromosome, different maps have their own usefulness in gene mapping and identification.

1.3.1 Clone-based physical maps (contig maps)

A contig map is one type of physical map that linearly displays a set of ordered clones along a chromosome. The unit of measurement for a contig map is a clone and the object is a marker such as an STS or a hybridization probe which detects overlapping clones. A contig map that consists of YACs is called a YAC contig map while a map consisting of cosmids is called a cosmid contig map. The resolution of contig maps is

defined by the insert size of clones: the smaller the insert size of clones, the higher the resolution of a contig map.

1.3.1.1 YAC contig maps

YAC contigs are sets of ordered overlapping clones which are constructed manually or through computer programs by lining up YACs that contain common markers. This process is often complicated by inconsistent mapping data such as chimeric YACs, and false positive or false negative identification of YACs. The contig generation is made easier if the order of markers that detect YACs is known. To date, YAC contig maps have been constructed for a large number of chromosomal regions throughout the human genome, including the Huntington disease region (Huntington disease collaborative group, 1993), the q24-q28 region of the X chromosome (Schlessinger et al., 1991; Little et al., 1992), chromosome 21q (Chumakov et al., 1992a), and the Y chromosome (Foote et al., 1992). For example, the chromosome 21q contig map consists of 810 YACs detected by 198 STSs while the Y chromosome contig map displays 196 YACs detected by 160 STSs. Recently, two genome-wide YAC contig maps have been constructed by the CEPH/Genethon group (Chumakov et al., 1995) and the Whitehead Institute/MIT Genome Center (http://www-genome.wi.mit.edu). Both maps consist of overlapping YACs anchored to a comprehensive set of genetic markers and are publicly available on the internet. These two maps provide useful YAC mapping information for further analysis of the genome such as positional cloning of disease genes. However, there are a number of limitations with these two maps including inconsistencies of the data produced by both groups, and the incompleteness of the maps, i.e. the existence of a large number of contig gaps across the genome. Thus, the improved YAC contig maps of the human genome require a careful examination and refinement of the data, and integration of data produced by chromosome-specific mapping efforts.

1.3.1.2 Cosmid maps

Two approaches have been used to construct high resolution cosmid or bacteriophage contig maps. The 'bottom up' approach, uses information of clone DNA

fingerprints to construct overlapping contigs (see section 1.2.3). This approach requires very precise measurements of restriction enzyme digestion fragments and many pairwise comparisons of fingerprints between clones to assemble accurate contig maps. The 'top down' approach, by contrast, utilizes previously determined mapping information, i.e. YAC contigs from regions of interest, to derive a subset of cosmid clones and then to assemble these cosmids into contig maps. This approach was used to construct cosmid contigs and a high-resolution restriction map for the Huntington disease region on human chromosome 4 (Zuo et al., 1993; Baxendale et al., 1993), the Down syndrome region on human chromosome 21 (Patil et al., 1994), and the entire chromosome 21 (Nizetic et al., 1994). The 'top down' approach significantly reduces the number of pairwise comparisons among different clones and the amount of redundancy and overlap required to obtain a contig map.

1.3.2 Genetic maps

Genetic maps display the order of different genetic markers and the interlocus distances along the chromosome. They are based on a biological process termed 'crossing over', which occurs during gamete formation and results in an exchange of homologous markers between pairs of chromosomes. The objects of genetic maps are genetic markers, which have to be polymorphic among the population in order to trace the 'crossing over' event. In other words, each genetic marker need to have two or more alleles, and the rarer allele has to have a frequency of at least 0.01 in the population. The degree of informativeness or polymorphism of a genetic marker can be measured by the polymorphic information content (PIC) value, which depends on the number of alleles at the marker locus and their relative frequencies (Ott, 1991). Markers that are closer together on the same parental chromosome are more likely to segregate together, while markers that are far apart from each other are more likely to separate during meiosis. Genetic maps are constructed by determining the frequency of markers segregating together. The observed recombination frequencies can be transformed into genetic distances and the likely order of loci can be established. The unit of the measurement for a genetic map is centiMorgan

(cM). A cM corresponds to a genetic length over which one observes recombination 1 percent of the time.

There are different classes of genetic maps. Two common classes are comprehensive and framework maps (Keats et al., 1991). Comprehensive maps consist of all the markers which can be uniquely placed at their most likely location regardless of local statistical support. Significant statistical support is generally defined as the likelihood ratio of 1,000:1 for the most likely locus position relative to neighboring loci, as compared to any other location. The comprehensive maps provide unique locations for every locus for which genotyping is available in a data set. Thus, the comprehensive maps are of high-resolution (i.e. smaller average interlocus distance). However, the lack of local statistical support sometimes make the local order of the markers and the interlocus distance unreliable. Therefore, comprehensive maps must be used with caution. In contrast, framework maps only represent loci that can be ordered with odds of 1,000:1 or greater. Framework maps are thus conservative but more reliable for ordering of loci as compared to the comprehensive maps. Markers on the low-resolution framework maps can be used for the placement of other markers, which create a higher-resolution of the comprehensive maps.

1.3.2.1 Development of genetic markers for the construction of genetic maps

The first genetic markers used to construct genetic maps were a limited number of phenotypic markers, including blood group antigens, serum protein markers, and erythrocyte isozymes (Mohr, 1954; Giblett, 1969). Such markers generally displayed low PIC values which limited the development of high resolution genetic maps. In 1980, Botstein et al. observed that there were numerous detectable DNA sequence variations in the human genome, which could be used as markers for genetic mapping (Botstein et al., 1980). The first type of DNA-based marker used in genetic mapping were markers that detected restriction fragment length polymorphisms (RFLPs), which varied in the length of DNA fragments generated by restriction endonuclease digestion of genomic DNA among the population. Markers which detect RFLPs can be cDNA probes or genomic clones free

of repeat-sequences. Due to the relatively low informativeness of most RFLPs, the RFLP probes were soon supplemented by the more informative minisatellites (Jeffreys et al., 1985), or variable number of tandem repeat (VNTR) sequences (Nakamura et al., 1987). Minisatellites or VNTR sequences consist of tandem repeats of a DNA unit and the number of repeats varies among the population. The subsequent discovery of microsatellites, i.e. short tandem repeat polymorphims (STRPs), has provided another source of DNA markers (Weber and May, 1989). The vast majority of microsatellites reported to date display dinucleotide repeats, or (CA)n, although increased numbers of tri- and tetra-nucleotide repeats are also being discovered. Microsatellites are widely dispersed throughout eukaryotic genomes (Hamada et al., 1982; Stallings et al., 1991). It was estimated that there are at least 35,000 (CA)n simple tandem repeats with n >12 uninterrupted dinucleotides per human haploid genome (Weber, 1990). These markers are highly polymorphic (Weber and May, 1989; Weber, 1990), and there are about 12,000 and 7,000 (CA)n repeat blocks with PIC values of 0.50 and 0.70, respectively, in the human genome (Weber, 1990). These markers are also easily typed by PCR since they are defined by a pair of primers flanking the repeat sequences. The above characteristics make microsatellites ideal markers for the construction of high-resolution genetic maps of the human genome. Indeed, the four most recent genetic maps of the whole genome were constructed using a large number of microsatellites including dinucleotide, tri- and tetra-nucleotide repeats (Gyapay et al., 1994; Buetow et al., 1994; Matise et al., 1994; Murray et al., 1994).

1.3.3 Cytogenetic maps

The cytogenetic maps display chromosomal band positions of DNA markers. Metaphase chromosomes can be divided into characteristic bands or subbands by a variety of cytogenetic banding techniques (see section 1.2.4). It is possible to discriminate at least a thousand bands per haploid human genome (Yunis, 1976). The unit of measurement for the cytogenetic maps is chromosome bands or subbands and the objects are markers which can be cDNA probes or genomic clones such as cosmids or YACs. DNA markers can be assigned to a particular chromosome band or subband by in situ hybridization or FISH

analysis (Rudkin, 1977; Trask, 1993). For example, a cDNA probe for *VIL* was assigned to chromosome 2, region q35 (Rousseau-Merck et al., 1988), which indicates that the *VIL* gene is located on the long arm of human chromosome 2 at band q35. To date, a large number of genes, ESTs and genetic markers have been assigned to a particular chromosome band or subband, which provides the basis for the cytogenetic map of the human genome. In addition, a large number of cosmids and YACs which are part of contig maps have also been cytogenetically mapped by FISH (Chang et al., 1993; Trask et al., 1993; Cohen et al., 1993; Moir et al., 1994), which allows the integration of cytogenetic maps with contig maps (Cohen et al., 1993; Chumakov et al., 1995).

1.3.4 Radiation hybrid (RH) maps

RH maps represent markers along chromosomes. The RH mapping technique was first described by Goss and Harris (1975), and subsequently modified and improved by Cox et al. (1990). The basis for RH mapping is the 'irradiation and fusion gene transfer' (IFGT). Briefly, a rodent-human somatic cell hybrid containing a single human chromosome is lethally irradiated with X-rays, which results in fragmentation of the human chromosome. Fragment-bearing hybrid cells which are not viable are then fused with a rodent cell line in a selective medium. The resulting hybrid clones contain unique sets of fragments from the original human chromosome and are named radiation hybrids.

The basic principle for constructing RH maps is that two loci which are close on a chromosome are less likely broken apart by radiation, and thus display correlated retention patterns in a hybrid panel, while markers far apart are independently retained (Cox et al., 1990). A panel of radiation hybrids can be typed for presence or absence of human DNA markers and a RH map is constructed by determining the frequency of marker retention, which can be transformed into order and distance between them. The objects of RH maps are markers which do not need to be polymorphic. Indeed, each unique marker is informative for RH mapping. The unit of measurement in RH maps is the centiRay (cR), which depends on the radiation dose used to generate radiation hybrids (Cox et al., 1990). A distance of 1cR (N rad) between two markers corresponds to 1% frequency of breakage between these two markers after exposure to an irradiation dosage of N rad of X-rays.

The resolution of RH maps can be adjusted by altering the X-ray dose used to fragment a chromosome. The cR distances can also be scaled to physical distances between loci by using the information as determined by other mapping methods such as PFGE or YAC cloning experiments. Such correlation can be used to compare the mapping resolution of different hybrid panels produced under different radiation condition.

1.3.4.1 Whole genome radiation hybrid (WG-RH) mapping

RH maps have been constructed using chromosome-specific hybrid panels for a large number of chromosomal regions of the human genome, including the distal short arm of chromosome 11, containing the Beckwith-Wiedemann and associated embryonic tumor disease loci (Richard et al., 1993) and the distal long arm of chromosome 4, containing the facioscapulohumeral muscular dystrophy (FSHD) gene (Winokur et al., 1993). RH maps were also constructed for individual human chromosomes such as chromosome 11 (James et al., 1994). Recently, the RH mapping technique was further applied to whole genome mapping (Walter et al., 1994). Instead of using a rodent-human somatic cell hybrid containing a known chromosome as a donor cell, WG-RHs were constructed by fusing irradiated diploid human fibroblasts with hamster cells. Each WG-RH contained chromosomal fragments randomly derived from the whole human genome and thus can be used to type markers for each chromosome. The advantage of WG-RH mapping panels is that as few as 100 WG-RH cell lines with marker retention frequencies between 20-50% can be used to systematically produce a high resolution map of the entire human genome (Walter et al., 1994). In contrast, to construct a RH map of the whole genome using chromosome-specific hybrids instead of WG-RHs would require over 4,000 hybrids, since between 100 and 200 chromosome-specific hybrids are needed to map each chromosome (Barrett., 1992). Thus, the WG-RHs provide a useful resource for constructing RH maps of the whole genome.

1.3.5 Integrated maps

As described in the previous sections, each type of map used in human genome research represents unique aspects of human chromosomes. Integrated maps integrate

different aspects of human chromosomes into single maps, and thus provide the most comprehensive picture of human chromosomes. The basis for the integration of different types of map is that they share common objects, i.e. markers such as STSs. An example of such an integrated map is shown by Bell et al. (1995). These authors used 238 chromosome 22-specific STSs which were derived from cytogenetically mapped genes and ESTs or genetically mapped markers. These STSs were assigned to specific regions of the chromosome using a chromosome 22 somatic hybrid mapping panel which was defined by cytogenetic breakpoints, and then used to screen the YAC libraries and identified 587 YACs. These YACs were assembled into contigs based upon their STS content and ordered along the chromosome based upon the cytogenetic breakpoint and genetic map information of the STSs. Thus, the final map provided an integrated view of physical, cytogenetic and genetic maps of chromosome 22. Similarly, several chromosome-specific efforts have created integrated maps for human chromosome 7 (Green et al., 1994), chromosome 11 (Fantes et al., 1995), chromosome 12 (Lebianc-Straceski et al., 1994), and chromosome 13 (Hawthorn et al., 1995).

Part II. Genetic studies of innate susceptibility to tuberculosis and leprosy

1.4 Tuberculosis and leprosy: two major mycobacterial diseases in humans

Tuberculosis, which is caused by *Mycobacterium tuberculosis*, is a highly prevalent infectious disease. It is estimated that about one-third of the world's population is infected with *M. tuberculosis*, and 30 million people are expected to lose their lives in the next decade from this disease (World Health Organization, 1994). Leprosy, which is caused by *Mycobacterium leprae*, is a chronic infectious disease and afflicts an estimated 5.5 million people in the world (Noordeen et al., 1992). Both tuberculosis and leprosy are characterized by a wide spectrum of disease phenotypes, and by the fact that most individuals infected by the causative mycobacteria do not develop clinically evident forms of the disease (Godal and Negassi, 1973; Lenzini et al., 1977). While environmental exposure to mycobacteria is necessary for the development of tuberculosis and leprosy,

several lines of evidence suggest that genetic factors play an important role in both the susceptibility to mycobacterial infection *per se* and the development of the varying clinical manifestations of these two diseases (Buschman et al., 1990). These evidences include: racial differences in disease susceptibility, familial clustering of the disease, survival of certain ethnic groups during large epidemics, and higher concordance of disease among monozygotic twins as compared to dizygotic twins (for review, see Fine, 1981; Stead, 1992). More recently, complex segregation analyses in leprosy families have established the importance of a genetic component for leprosy susceptibility (Abel and Demenais, 1988; Abel et al., 1995). Based on the accumulated experimental data, it is hypothesized that differential susceptibility to mycobacterial infections is influenced by two sets of genes: human leukocyte antigen (HLA)-linked genes that are associated with the subtypes of leprosy or tuberculosis, while non-HLA-linked genes are believed to control susceptibility to the establishment of clinically significant infection (or disease *per se*, i.e. regardless of the disease subtype; for review, see Schurr et al., 1990a).

Since both environmental and genetic factors are involved in the manifestation of tuberculosis and leprosy, the genetic study of these diseases has remained a difficult task. Thus, to dissect the contribution of diverse genetic and environmental factors to disease susceptibility, several mouse models for the study of mycobacterial infection were developed (for review, see Schurr and Skamene, 1995).

1.5 Innate resistance/susceptibility to mycobacterial infection in mice

In the mouse, host response to infection with a small dose (2×10^4 CFUs) of i.v. injected *Mycobacterium bovis* (Bacille Calmette-Guerin, or BCG) is biphasic: an early non-immune phase (0-3 wk), and a late immune phase (3-6 wk) which is controlled by host genes from the major histocompatibility complex (Gros et al., 1981; Lagrange, 1983). The murine innate resistance/susceptibility to *M. bovis* infection during the early phase was found to be under the control of a single, autosomal gene which was designated *Bcg* (Skamene et al., 1982). The *Bcg* gene was also shown to control natural resistance to other mycobacteria such as *M. lepraemurium* (Skamene et al., 1984), *M. intracellulare* (Goto et al., 1989), *M. avium* (Appelberg, 1990), other atypical mycobacteria (Denis,

1986), and antigenically unrelated intracellular parasites such as *L. donovani* and *S. typhimurium*, respectively (Bradley, 1977; Plant and Glynn, 1976).

The cellular basis for the *Bcg* gene mediated phenotype was subsequently identified by *in vivo* cell depletion and cell transfer experiments as the mature tissue macrophage (Gros et al., 1983; Stach et al., 1984). Several *in vivo* and *in vitro* studies have shown that various phenotypic markers of activated macrophages are expressed at higher levels in macrophages derived from Bcg^r mice, and these Bcg^r macrophages have a superior bactericidal ability (for review, see Buschman et al., 1989). The "pleiotropic effects" of the *Bcg* gene suggest that the gene may play a role in controlling the priming/activation of macrophages for antimicrobial activity (for review, see Buschman et al., 1989; Blackwell et al., 1991; 1994).

The similarities of host responses to mycobacterial infections in humans and mice, e.g. the biphasic response in mice and the two sets of genes in humans, suggest that the mouse is a useful animal model for the genetic study of the resistance/susceptibility to tuberculosis and leprosy. A human homologue of the mouse *Bcg* gene would thus be a good candidate for the genetic component of disease susceptibility detected in twin and segregation studies (Comstock, 1978; Abel and Demenais, 1988; for reviews, see Schurr et al., 1991a, b).

1.6 Comparative mapping

Comparison of the genetic maps of different species, in particular mouse and human, indicates a large number of chromosome regions where both gene content and genetic order of loci are conserved (O'Brien, et al., 1988; Nadeau and Taylor, 1984; Nadeau, 1989; Liu et al., 1993). These conserved chromosome regions allow the prediction of locations of homologous genes in other species. For example, if a mouse locus is localized on a mouse-human conserved linkage group, a likely location of the corresponding human homologue can be predicted by comparative mapping information. Such an approach has been used successfully to identify or to establish linkage for a number of disease genes including the genes causing Charcot-Marie-Tooth disease type 1

(CMT1) (Vance et al., 1991; Welcher et al., 1991), and the neural crest mutations of Waardenburg syndrome type I (WS1) (Epstein et al., 1991; Baldwin et al., 1992).

Genetic linkage studies have identified a segment of approximately 32 cM on proximal mouse chromosome 1, flanked by *Col3a1* and *Col6a3* at the proximal and distal sides of *Bcg*, respectively, that is precisely conserved on the telomeric half of human chromosome 2q (Schurr et al., 1990b). Subsequent genetic studies also indicated that the gene order appears to be conserved between mouse and human (Shaw et al., 1993; Spurr et al.,1994). Applying the comparative gene mapping approach to the study of susceptibility to tuberculosis and leprosy, the chromosome 2 region q33-q37 became a primary target for the search of a tuberculosis and leprosy susceptibility gene (for reviews, see Schurr et al., 1991a, b).

1.7 Generation of genetic markers for genetic studies of tuberculosis and leprosy susceptibility

Prior to the identification of a human *Bcg* homologue, chromosome 2q loci syntenic on proximal mouse chromosome 1 were tested for co-segregation with resistance/susceptibility to tuberculosis and leprosy in multiplex families. The vast majority of genetic markers used for these genetic studies were RFLP markers from chromosome 2, region q33-q37 (Jazwinska and Seijeantson, 1988; Shaw et al., 1993; Levee et al., 1994; Boothroyd, 1994). Most of these RFLPs were generated from the following genes: *CRYG1*, *CRYGP1*, *MAP2*, *FN1*, *VIL1*, *DES*, *COL6A3*, *TNP1*, *INHA*, *UGT1A1*. Using the RFLP markers derived from the known genes on the 2q33-q37 region and two anonymous RFLP markers *D2S55* and *D2S3*, Shaw et al. (1993) constructed an RFLP map and proposed the following gene order and loci intervals (cM): *CRYG1*-(4.65)-*MAP2*-(3.45)-*FN1*-(5.95)-*TNP*-(3.41)-*VIL1*-(3.01)-*DES*-(20.14)-*COL6A3*-(10.91)-*D2S55*-(3.67)-*D2S3*. This map represents the most detailed genetic map of expressed genes to date for the telomeric region of human chromosome 2q.

Recently, through the global efforts of the Human Genome Project, a significant number of anonymous markers were generated for the whole human genome, including chromosome 2q33-q37, which can be used for linkage studies (Spurr et al., 1994; Gyapay

et al., 1994; Buetow et al., 1994; Matise et al., 1994; Murray et al., 1994). However, most of these markers are anonymous microsatellites and the order of these markers is not established with respect to known genes. The maximum usefulness of these highly polymorphic microsatellites in linkage studies of disease genes thus awaits the integration of these markers with markers generated from known genes.

1.8 Genetic and physical mapping of *Bcg* on the proximal region of mouse chromosome 1

By segregation analysis of *Bcg* in recombinant inbred mice, *Bcg* had been initially mapped to the proximal region of mouse chromosome 1 between *Idh-1* and *Pep-3* (Skamene et al., 1982). Subsequently, a number of additional RFLP markers, microdissected DNA probes, chromosomal walking probes, and SSRP markers were generated and mapped to proximal mouse chromosome 1. These additional markers greatly increased the resolution of the genetic map overlapping the *Bcg* gene region (Schurr et al., 1989; Malo et al., 1991; Vidal et al., 1992; Malo et al., 1993a; Epstein et al., 1994) and the following locus order and genetic map distances (cM) were established: centromere-*Col3a1*-(8.8)-*Cryg*-(2.6)-*λMm1C163*-(1.6)-*Fn*-1-(2.0)-*Tp-1*-(1.0)-*D1Mcg105*-(0.1)-*λMm1C165/Vil/Bcg*-(0.2)-*λMm1C136*-(0.3)-*Des/D1Mit7*-(0.1)-*Inha*-(2.8)- *λMm1C153*-(2.4)-*λMm1C156*-(1.2)-*Pax-3*-(5.6)-*Akp-3*-(0.8)-*Acrg*-(2.0)-*Sag*-(0.5)-*Col6a3* (Malo et al., 1993 a, b). The two most closely linked markers to *Bcg* were *λMm1C165* and *Vil*, which did not combine with *Bcg* in 1424 meioses tested (Malo et al., 1993a). The minimal genetic *Bcg* interval region provided the framework for the construction of a high resolution physical map of the *Bcg* region.

The construction of a physical map overlapping the *Bcg* gene was undertaken using pulse field gel electrophoresis (PFGE) and fluorescence *in situ* hybridization (FISH) techniques (Malo et al., 1993b). The combined results of these analyses indicated the following locus order and maximal interlocus distances for the 3.9-Mbp segment of proximal mouse chromosome 1 overlapping the *Bcg* gene: *Tp1*-(1000 kb)-*D1Mcg105*-(160 kb)-*λMm1C165*-(180 kb)-*Vil*-(800 kb)-*λMm1C136*-(290 kb)-*Des*-(130 kb)-*Inha* (Malo et al., 1993b). The two markers (*D1Mcg105* and *Vil*) tightly linked to *Bcg*, were used as entry probes to assemble a 400-kb cosmid, phage and YAC contig, which

provided the basis for the molecular cloning of the *Bcg* gene (Vidal et al., 1993; Malo et al., 1993a, b).

1.9 Cloning the *Nramp1* gene - a candidate for *Bcg*

Since there was no visible cytogenetic rearrangements detected in the 400 kb *Bcg* candidate region, the exon amplification technique (Buckle et al., 1991) was used to search for a *Bcg* candidate gene (Vidal et al., 1993). Systematic screening of transcript units for the 400 kb *Bcg* candidate region led to the identification of seven genes, *Vil* and six other unknown genes. Among the six new genes, one was found to be expressed exclusively in spleen and liver, and in mature tissue macrophages, the tissues and cell type known to phenotypically express *Bcg* (Gros et al., 1983). This gene was designated "*Nramp1*" (originally "*Nramp*") for natural resistance-associated macrophage protein. The original *Nramp1* cDNA sequence was subsequently modified by the additional NH2-terminal sequence of *Nramp1*, which encodes a proline/serine-rich putative Src homology 3-binding domain (Barton et al., 1994).

Several lines of evidence suggest that *Nramp1* and *Bcg* are allelic: 1) *Nramp1* lies in the *Bcg* candidate region between *D1Mcg105* and *Vil* (Malo et al., 1993a); 2) *Nramp1* is expressed in the tissues and cell types that display the *Bcg*-mediated phenotype; 3) haplotype analysis of the *Nramp1* region and nucleotide sequence analyses of the *Nramp1* cDNA for 27 inbred mouse strains of either Bcg^r and Bcg^s phenotype showed that the susceptibility trait was associated with a nonconservative glycine to aspartic acid substitution within the predicted transmembrane domain 2 (TM2) of the protein (Malo et al., 1994). More recently, another line of evidence, which supports that *Nramp1* and *Bcg* are allelic was provided by the creation and phenotypic analysis of a null allele at the *Nramp1* locus by homologous recombination in embryonic stem cells of Bcg^r background (Vidal et al., 1995b). The *Nramp1* homozygous null mice have the same susceptibility phenotype to BCG infection as inbred mice carrying the Bcg^s-allele.

1.10 The human *NRAMP1* gene

The human *NRAMP1* gene was subsequently cloned by screening a human adult spleen cDNA library with a murine *Nramp1* cDNA as a hybridization probe (Cellier et al., 1994). The human gene is composed of at least 15 exons and the predicted polypeptide is a 550-amino acid residue membrane protein with 10-12 putative transmembrane domains, two N-linked glycosylation sites, and an evolutionary conserved energy coupling consensus transport motif. Northern blot analyses indicate that the highest sites of expression for *NRAMP1* mRNA are peripheral blood leukocytes, lung, spleen and cultured macrophage cells. Sequence comparison of the predicted human and mouse NRAMP1 protein revealed a remarkably high degree of amino acid conservation between the two species (88% identity, 93% similarity). The human *NRAMP1* gene was also independently isolated and characterized by two other groups (Kishi et al., 1994; Blackwell et al., 1995).

The extensive homology between mouse and human *Nramp1* genes, and the tissue specific expression of the gene suggest that its function could be highly conserved. To date, three hypotheses for the *Nramp1* function have been proposed. First, *Nramp1* codes for a nitrite transporter, which is based on its similarities with an eukaryotic nitrate transporter through a conserved consensus transport motif known as the "binding protein-dependent transport system inner membrane component signature" (Cellier et al., 1994). This hypothesis is attractive since nitric oxide (NO) plays an important role in antimicrobial or tumoricidal activity (Nathan and Hibbs, 1991; Nathan, 1992; 1995). *Nramp1* may be involved in transporting nitrite to the phagolysosome of the infected macrophages, where toxic NO could be regenerated through a dismutation reaction catalyzed by the acid environment. However, this hypothesis is not in agreement with the studies which demonstrated that the TNF-α-dependent inducible nitric oxide synthase (iNOS) gene was responsible for the generation of large amounts of toxic NO (Roach et al., 1994), and the pleiotropic effects of the *Bcg* gene also are difficult to explain by this model. Second, Nramp1 protein may be involved in signal transduction. This hypothesis was based on the presence of a domain rich in proline, serine, and basic amino acids in the N-terminal sequence of *Nramp1*, which shares sequence identity with the SH3 binding domain of proteins involved in signal transduction (Barton et al., 1994). However, this

domain is located in a region that showed the lowest degree of homology among different species including human and mouse NRAMP1 proteins (Cellier et al., 1994; 1995). An alternative hypothesis is that *Nramp1* is associated with the regulation of L-arginine transport in activated macrophages. This hypothesis originated from the study in rodents showing that NO required for the antimicrobial or tumoricidal activity seemed to derive from the pathway of L-arginine through NOS, rather than by its regeneration from nitrate (Blackwell et al., 1994). Preliminary experiments showed a significant enhancement of L-arginine transport in the macrophage cell line derived from Bcg^s mice which was transfected with *Nramp1* isolated from activated Bcg^r macrophage cDNA library as compared to the untransfected controls (Blackwell et al., 1994). However, in man, NOS does not appears to be conserved as antimicrobial system of macrophages (Schneemann et al., 1993). Thus, the biological function of *Nramp1* still remains to be determined.

Based on the mouse model and the results of genetic analyses of tuberculosis and leprosy susceptibility, it has been predicted that the human *Nramp1* homologue may be a susceptibility gene for mycobacterial disease of humans. The cloning of human *NRAMP1* provided an essential tool to test its predicted role in resistance/susceptibility to mycobacterial infection by linkage or association analyses. Towards this end, the immediate task was to generate informative markers within or in the immediate vicinity of the *NRAMP1* gene. Recently, a number of polymorphisms have been identified in the human *NRAMP1* gene. These polymorphisms include 8 RFLP-PCR variants, (Chapter 4), a rare 9-nucleotide repeat polymorphism in exon 2 (White et al., 1994), a dinucleotide repeat polymorphism in the promoter region (Chapter 4; Blackwell et al., 1995), and a highly polymorphic 4 bp insertion/deletion in the 3' untranslated region (UTR; Buu et al., 1995).

1.11 *Nramp1* belongs to a small gene family

While cloning the human *NRAMP1* gene, two sets of hybridization fragments (strong and weak ones) were identified in Southern blot analysis of total human genomic DNA with a mouse *Nramp1* cDNA probe under low stringency conditions. This observation led to the identification of two related *NRAMP* genes in humans (Cellier et al.,

1994; Vidal et al., 1995a). The set of strong hybridization fragments corresponds to the human homologue of mouse *Nramp1* (Vidal et al., 1993). The set of weak hybridization fragments corresponds to the second *NRAMP* gene, which was designated as *NRAMP2* (Vidal et al., 1995a). Similarly, by cross-hybridization experiment with mouse *Nramp1* cDNA or *Nramp1*-derived probe, two other mouse Nramp genes were identified. One was designated *Nramp2* (Gruenheid et al., 1995), while the other one was called *Nramp-rs* (Dosik et al., 1994).

Sequence comparison analysis showed that these *Nramp* sequences (except *Nramp-rs*, which sequence awaits to be determined) shared a high degree of similarity. For example, the mouse Nramp1 and Nramp2 proteins share 63% identical residues and an overall homology of 78% (Gruenheid et al., 1995), while the two human NRAMP proteins share 66% identical residues and 80% overall homology (Vidal et al., 1995). Nramp proteins also share other predicted features such as hydropathy profiles, membrane organization, a cluster of predicted N-linked glycosylation sites, and a consensus transport motif (Gruenheid et al., 1995; Vidal et al., 1995a).

In contrast to the high sequence homology, the RNA expression profiles of different members of the *Nramp* family are very different. The mouse *Nramp1* and human *NRAMP1* have a restricted tissue expression pattern, while the mouse *Nramp2* and human *NRAMP2* mRNAs were found to be expressed at low levels in all tissues tested (Gruenheid et al., 1995; Vidal et al., 1995a). Interestingly, *Nramp* genes are not clustered in the genome. Human *NRAMP1* and *NRAMP2* were mapped to chromosome regions 2q35 and 12q13, respectively (Cellier et al., 1994; Chapter 4; Vidal et al., 1995a), while the mouse *Nramp1*, *Nramp2* and *Nramp-rs* genes were mapped to the proximal region of chromosome 1, the distal part of chromosome 15, and the proximal region of chromosome 17, respectively (Vidal et al., 1993; Gruenheid et al., 1995; Dosik et al., 1994).

Recently, *Nramp* homologues from additional species including cattle, fly, plant, and yeast were identified (Cellier et al., 1995). Comparison of amino acid sequences of these proteins with their murine counterpart revealed a remarkable sequence identity and similarity, and the sharing of a common predicted transmembrane topology (Cellier et al.,

1995). Based on the characterization and comparison of these proteins, it was proposed that this group of Nramp proteins forms a novel family of membrane transporters, which share a common mechanism of action on yet to be identified substrates (Cellier et al., 1995).

1.12 Conclusion

The identification of mouse *Nramp1* has opened the door to investigate its role in susceptibility to infection with intracellular pathogens and a better understanding of cytocidal or cytostatic functions of tissue macrophages. Likewise, the cloning of a human *Nramp1* homologue (*NRAMP1*) allows the formal testing of its predicted role in susceptibility to tuberculosis and leprosy. The insight gained from further genetic and functional studies of *Nramp1*/*NRAMP1* will provide crucial information necessary for the development of an immunotherapeutic strategy for treatment of those who are for genetic reasons at increased risk of acquiring mycobacterial disease.

CHAPTER 2:

Large-scale cloning of human chromosome 2-specific yeast artificial chromosomes (YACs) using an interspersed repetitive sequences (IRS)-PCR approach

GENOMICS 26, 178–191 (1995)

Large-Scale Cloning of Human Chromosome 2-Specific Yeast Artificial Chromosomes (YACs) Using an Interspersed Repetitive Sequences (IRS)-PCR Approach

JING LIU,*·† VINCENT P. STANTON, JR.,‡ T. MARY FUJIWARA,†·§·¶ JIAN-XUE WANG,*
REBECA REZONZEW,*·† M. JOYCE CRUMLEY,† KENNETH MORGAN,*·†·§·‖
PHILIPPE GROS,*·** DAVID HOUSMAN,‡ AND ERWIN SCHURR*·†·[1]

*McGill Centre for the Study of Host Resistance, Montreal, Quebec, Canada; Departments of †Medicine, §Human Genetics, ¶Pediatrics, ‖Epidemiology and Biostatistics, and **Biochemistry, McGill University, Montreal, Quebec, Canada; and ‡Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

Received September 1, 1994; revised December 13, 1994

We report here an efficient approach to the establishment of extended YAC contigs on human chromosome 2 by using an interspersed repetitive sequences (IRS)-PCR-based screening strategy for YAC DNA pools. Genomic DNA was extracted from 1152 YAC pools comprised of 55,296 YACs mostly derived from the CEPH Mark I library. *Alu*-element-mediated PCR was performed for each pool, and amplification products were spotted on hybridization membranes (IRS filters). IRS probes for the screening of the IRS filters were obtained by *Alu*-element-mediated PCR. Of 708 distinct probes obtained from chromosome 2-specific somatic cell hybrids, 85% were successfully used for library screening. Similarly, 80% of 80 YAC walking probes were successfully used for library screening. Each probe detected an average of 6.6 YACs, which is in good agreement with the 7- to 7.5-fold genome coverage provided by the library. In a preliminary analysis, we have identified 188 YAC groups that are the basis for building contigs for chromosome 2. The coverage of the telomeric half of chromosome 2q was considered to be good since 31 of 34 microsatellites and 22 of 23 expressed sequence tags that were chosen from chromosome region 2q13–q37 were contained in a chromosome 2 YAC sublibrary generated by our experiments. We have identified a minimum of 1610 distinct chromosome 2-specific YACs, which will be a valuable asset for the physical mapping of the second largest human chromosome.

INTRODUCTION

Chromosome 2, the second largest human chromosome, corresponds to approximately 8% of the human genome. Three genetic maps of chromosome 2 vary in the sex-averaged length (354 cM, NIH/CEPH Collaborative Mapping Group, 1992; 277 cM, Weissenbach *et al.*, 1992; 350 cM, Spurr *et al.*, 1992). To date, almost 200 expressed loci and over 1200 anonymous DNA markers have been mapped to chromosome 2 (Genome Data Base, Johns Hopkins University, Baltimore, MD). Among the genes mapped are several major disease and modifier loci, including the genes responsible for Waardenburg syndrome (Baldwin *et al.*, 1992), Ehlers–Danlos syndrome (Richards *et al.*, 1992), juvenile amyotrophic lateral sclerosis (Spurr *et al.*, 1994), holoprosencephaly (Grundy *et al.*, 1989; Hecht *et al.*, 1991), alveolar rhabdomyosarcoma (Barr *et al.*, 1993), protein C deficiency (Patracchini *et al.*, 1989), cleft lip with or without cleft palate (Ardinger *et al.*, 1989; Chenevix-Trench *et al.*, 1992), and, as suggested by homology mapping, an insulin-dependent diabetes susceptibility modifier gene (Cornall *et al.*, 1991) and a gene controlling resistance to intracellular parasites (Schurr *et al.*, 1990; Vidal *et al.*, 1993). Additional loci of medical interest include a mutator gene (*COCA1, MSH2*) implicated in the development of colon cancer on the short arm (Fishel *et al.*, 1993; Peltomaki *et al.*, 1993), the *PAX8* homeobox gene, possibly involved in developmental defects (Poleev *et al.*, 1992; Stapleton *et al.*, 1993), and the *XRCC5* DNA repair gene (Jeggo *et al.*, 1992; Hafezparast *et al.*, 1993), which are located on the long arm of chromosome 2.

We have initiated the cloning of large segments of human chromosome 2 using the yeast artificial chromosome (YAC) cloning system (Burke *et al.*, 1987; Burke, 1991). Among several strategies for the efficient identification of YAC clones containing defined loci from genomic YAC libraries that are being pursued by different laboratories, hierarchical PCR screening of YAC DNA pools with sequence-tagged site (STS) primers is

[1] To whom correspondence should be addressed at Montreal General Hospital Research Institute, Rm. L11-521, 1650 Cedar Avenue, Montreal, Quebec, H3G 1A4, Canada. Telephone: (514) 937-6011, ext. 4513 or 4554. Fax: (514) 933-7146. E-mail: cxjz@musica.mcgill.ca.

0888-7543/95 $6.00

the most widely used approach (Green and Olson, 1990; Green *et al.*, 1991; Chumakov *et al.*, 1992a; Foote *et al.*, 1992). In addition to being very sensitive, this approach has the advantage that microsatellites that have been widely used as markers for the construction of genetic maps can be used for STS screening, thus allowing the integration of genetic and physical maps. This approach was used to define genetic intervals on a physical map of the human genome (Cohen *et al.*, 1993). Similarly, the generation of physical maps for chromosome 21 and the Y chromosome strongly relied on STS content mapping (Chumakov *et al.*, 1992a; Foote *et al.*, 1992). The main limitations of STS content mapping are the cost of primer synthesis and other PCR reagents and the time required to analyze the PCR products serially.

We have primarily used the interspersed repetitive sequences (IRS)-PCR approach for YAC library screening. Short repeat sequences, interspersed throughout the entire genome, represent approximately 25% of human genomic DNA. The observation that unique genomic sequence between two repeat elements can be amplified by using repeat element-specific primers is the basis for the IRS-PCR protocol. Applied to YAC library screening, individual YACs are used as templates for IRS PCR, and the resulting amplification products representing repeat-free genomic segments are spotted on membranes and hybridized with IRS probes. Probes are generated by IRS-element mediated PCR amplification using, for example, somatic cell hybrids carrying inserts of particular human chromosome regions, individual YAC clones, or cosmids as templates. The hybridization of PCR amplification products results in a very high signal-to-noise ratio. IRS-PCR techniques were used for the development of cosmid contigs (Zucman *et al.*, 1992), and inter-*Alu* PCR products from a chromosome 21 somatic cell hybrid have been successfully used to identify chromosome 21 YACs (Chumakov *et al.*, 1992b). The present investigation is an extension of these studies by using a pooling scheme that was recently developed (Stanton *et al.*, unpublished) for the CEPH Mark I YAC library (Albertsen *et al.*, 1990; Dausset *et al.*, 1992). In this strategy, cultures of YAC clones from the library are first pooled, and then DNA is extracted from the pools and subjected to PCR amplification using IRS element-specific primers. The major advantage of a pooling scheme is that the entire library can be represented in a compact form on small hybridization membranes.

The first step of a physical DNA mapping project is the collection of clones that represent the targeted genome region. This is followed by defining a contig, that is, an ordered set of overlapping clones that contain the DNA of the corresponding genomic segment in precise order. The detailed analysis of the individual clones in the contig then leads to the establishment of a physical map. Here, we describe an approach to the molecular cloning of close to 2000 human chromosome 2-specific YACs that constitute a resource for the generation of physical maps of the chromosome.

MATERIALS AND METHODS

YAC library. The YAC library was contained in 576 96-well microtiter plates. Of these 576 plates, 552 plates correspond to the first-generation CEPH Mark 1 YAC library (Albertsen *et al.*, 1990; Dausset *et al.*, 1992) and 24 plates were YAC clones from the original Washington University YAC library (Burke *et al.*, 1987; Brownstein *et al.*, 1989). The combined libraries constitute 8 blocks of 72 plates. Each block is made up of 12 floors; each floor has 6 plates and is divided into 2 half-floors. Each YAC clone is assigned an address based on its location in a block, plate, row, and column. The average insert size of a YAC clone has been estimated to be 430 kb for the CEPH YAC clones (Albertsen *et al.*, 1990) and 250 kb for the Washington University YAC clones (Brownstein *et al.*, 1989).

Pooling scheme for the library and YAC DNA preparation. Within each block, aliquots of individual YAC cultures grown to confluency were combined from 288 YACs derived from the rows, columns, or half-floors. Thus, from all 8 blocks, there were 192 row pools, 192 column pools, and 192 half-floor pools. In addition, YAC cultures grown to confluency from each plate were pooled to make 576 plate pools. DNA of high purity was prepared from each YAC pool in agarose plugs using a lyticase/LiCl protocol (Smith, 1990).

PCR conditions for amplification of pooled YAC DNA. Two sets of *Alu* primers were employed for IRS amplification from all templates. The first set, Alu S (5'-GAGGTTGCAGTGAGCCGAGAT-3') and Alu J (5'-GAGGCTGCAGTGAGCCGTGAT-3'), was derived from the primate-specific region of *Alu* elements and corresponds to the most common *Alu* subfamily (Jurka and Smith, 1988). These two primers differed at 2 of 21 nucleotides. *Alu*-PCR reactions were performed with a 4:1 ratio of Alu S:Alu J primer, which approximately corresponds to their relative abundance in the genome. The second primer set corresponds to the 3' end of human *Alu* elements. It consisted of Alu 263–283.1 (5'-ACAGAGCGAGACTC[C/T(3:1)][G/A(2:1)]TCTC-3') and Alu 263–283.2 (5'-ACAGAG[C/T(3:1)][G/A(3:1)]AGACTCCGTCTC-3'). These primers were used in a 1:1 ratio for *Alu* PCR. *Alu*-PCR amplification of YAC DNA pools was performed in a 100-μl reaction volume containing 10 m*M* Tris (pH 8.3), 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 0.001% gelatin, 200 μM each dNTP, ~50 ng DNA, and 2.5 units of *Taq* DNA polymerase. Parameters for thermocycling were as follows: incubation for 4 min at 94°C, followed by 35 cycles of 1 min at 94°C, 30 s at 59°C, 1 min at 58°C, and 2 min at 72°C. A final extension step of 10 min at 72°C was added to facilitate complete replication of amplified products.

Preparation of hybridization membranes. *Alu*-PCR reactions for YAC DNA pools were set up with a Biomek 1000 workstation (Beckman, Mississauga, Ontario). A Perkin Elmer 9600 thermal cycler was used to amplify *Alu*-PCR products under oil-free conditions. The success of PCR amplification for each pool was tested by visual inspection of amplification products on 1.5% TAE agarose gels stained with ethidium bromide. Poorly amplified pools were identified and subjected to a second round of *Alu*-PCR amplification and inspection. Amplification products derived from 576 row, column, or half-floor pools, and from 576 plate pools, were each mixed with a small amount of India ink and stamped on a 2.5 × 9 inch GeneScreen Plus hybridization membrane (NEN Research Products, Boston, MA) using a 96-well replicating device from Washington University. All membranes were baked for 1 h at 80°C; filter-bound DNA was denatured in 0.4 *N* NaOH, 0.6 *M* NaCl for 15 min followed by neutralization in 0.5 *M* Tris–HCl (pH 7.4), 0.5 *M* NaCl for 5 min. Filters were prewashed in 0.1× SSC, 0.5% SDS at 65°C for 1 h according to the manufacturer's recommendation (NEN).

Somatic cell hybrids. All somatic cell hybrids used in our experiments were kindly provided by Dr. P. Jeggo (MRC Cell Mutation Unit, Sussex, UK) except hybrid GM10826B, which was obtained from the NIGMS Human Genetic Mutant Cell Repository (Camden,

NJ) (Table 1). Somatic cell hybrid GM10826B is described as containing all of human chromosome 2. Hybrids b3-1, D6-2, D7-2, and D4-2 (Jeggo *et al.*, 1993; Hafezparast *et al.*, unpublished data) were obtained by fusing the irradiated H22-oua^R hybrid to the XR-V15B Chinese hamster ovary cell line, which is deficient in double strand-break rejoining. They contain small segments of human chromosome 2 from region 2q34–q36. Hybrid 014-1 contains approximately the telomeric half of chromosome 2q but has lost sequences near the telomere.

Generation of IRS-PCR probes. Inter-*Alu* PCR products were directly excised from agarose gels and labeled except for amplification products from somatic cell hybrids GM10826B and 014-1, which were cloned into a T-modified vector (Marchuk *et al.*, 1991). Individual colonies carrying the recombinant plasmid were picked as candidate probes and resuspended in 60 μl of 1× PCR buffer. Inserts were amplified using T3 and T7 primers in 96-well polycarbonate plates (Fisher Scientific, Montreal, Canada) in a BioOven III (BioTherm, Fairfax, VA) in 20 μl of reaction volume containing 50 m*M* Tris (pH 8.3), 0.05% Tween, 0.05% NP40, 187.5 μ*M* each dNTP, 1.8 m*M* $MgCl_2$, 0.2 to 0.4 μ*M* each primer, 2 μg BSA, and 1.5 units of *Taq* polymerase. PCR parameters were as described above. Following amplification, all inserts were sized by electrophoresis on 1.3% agarose gels stained with ethidium bromide and grouped according to their fragment size. The PCR product was digested with a combination of *Rsa*I, *Hae*III, and *Hin*fI restriction enzymes, and the digested products were size separated on 1.8% agarose gels. The resulting pattern of restriction fragments was used as a fingerprint for each probe.

Probe labeling. Primers Alu S and Alu J are located 78 bp upstream from the 3′ end of *Alu* elements. Thus, *Alu*-PCR products amplified using this set of primers carry a 78-bp *Alu* sequence at both ends. To avoid nonspecific hybridization by this 78-bp *Alu* sequence, probes were labeled using Alu-3′ (5′-C[T/C][A/G]TCTCAAA-3′), a 10-mer oligonucleotide directed to the very 3′ end of *Alu* elements, thus avoiding radiolabeling of repeat sequences. Furthermore, all probes generated using Alu S and Alu J primers were preannealed with an excess of sheared denatured total human placental DNA for 4 h at 65°C. Alu 263–283.1 and Alu 263–283.2 are located close to the 3′ end of *Alu* elements; therefore, no preannealing was required. Labeling of probes was performed using 0.75 μl of [α^{32}P]dATP (sp. act. >3000 Ci/mmol; Amersham) and the random hexanucleotide labeling technique (Feinberg and Vogelstein, 1983).

YAC library screening. Labeled probes were heat denatured and hybridized to 2 row/column/half-floor and 2 plate pool filters. Prehybridization was performed for 4 to 16 h at 42°C in 8 ml of solution containing 50% deionized formamide, 1% SDS, and 3× SSC (1× SSC is 0.15 *M* NaCl, 0.015 *M* sodium citrate). Hybridization was performed using the same solution at 42°C overnight in the presence of radiolabeled probe. Filters were washed for 20 min at room temperature in a solution containing 2× SSC, 0.5% SDS, followed by a wash with 0.5× SSC, 0.5% SDS for 15 min, and a final wash for 30 to 60 min at 65°C in 0.1× SSC, 0.5% SDS. Filters were exposed for 16 to 72 h at −70°C with intensifying screens to Hyperfilm-MP (Amersham, Dakville, Ontario). Filters were deprobed by gently washing for 1 h at 65°C in the same prehybridization solution. Filters have been reused for up to 50 consecutive hybridizations.

Data processing. Data are maintained in the Microsoft FoxPro Relational Database Management System for Windows on a Hewlett–Packard Vectra 25N microcomputer with 16 MB of RAM. The main table of the database contains information obtained from screening the pooled YAC filters, including the name of the probe, YAC address, codes to qualify the address and to differentiate unambiguous from ambiguous addresses, and a variable to link the record to laboratory notebooks. The address for each YAC clone that hybridized to a probe is derived from the location of the hybridization signals on the filters. Ideally, an unambiguous address consisting of block, plate, row, and column for each YAC clone can be assigned from the location of hybridization signals on the filters. However, a set of ambiguous addresses arises, for example, when there are 2 row pool signals and 2 column pool signals that are consistent with a plate pool. The YAC address is stored in the database as a single variable, which is a character string of genome, floor, plate, row, and column, and can be linked to the corresponding location in the original CEPH YAC plates. The probe–address combinations generated by our experiments are available via anonymous ftp. To obtain the data, *ftp* to ftp.mcgill.ca and *cd* to the directory pub/McGill-Contrib/Chromosome2.

As an aid to building contigs, we have developed a program to perform the initial step of finding all of the YACs that are part of a contig. We call the set of YACs in a contig a YAC group. The input to the program is all of the probe–YAC address combinations derived from the location of hybridization signals. As a starting point a YAC group is defined by the addresses associated with an arbitrary probe. We find all of the other probes that share any of the YAC addresses with the starting probe. Then for each probe found by this search, we add the set of all associated addresses that have not already been included in the YAC group. This process is repeated until there are no new probes and no new addresses to be added to the YAC group. Only unambiguous YAC addresses or ambiguous YAC addresses that have been confirmed by single clone hybridization experiments were included. A preliminary analysis was based on 3470 probe–address combinations. Finally, the program CONTIGMAKER (version 0.9), available from the Whitehead Institute ftp site, was used for building contigs of YAC groups using ambiguous and unambiguous probe–address combinations. The default postscript output files from CONTIGMAKER were edited to improve the graphical presentation of contigs.

STS screening. For PCR screening of chromosome 2 YACs with expressed sequence tag (EST) and microsatellite primers, DNA was isolated from chromosome 2-specific YAC clones identified by IRS-PCR screening following the protocol of Chumakov *et al.* (1992a) with minor modifications. YAC clones grown in 150 μl of AHC medium in 96-well plates were pelleted at 2000 rpm, washed once in 20 m*M* Tris–HCl and 1 m*M* EDTA, resuspended in 20 μl of 20 m*M* Tris–HCl (pH 8.0), 10 m*M* DTT containing 800 μg/ml Zymolase 100T (ICN, Costa Mesa, CA), and incubated for 45 min at 30°C. Next, 60 μl of 140 m*M* NaOH was added and YAC clones were incubated for 7 min at room temperature prior to the addition of 60 μl of 1 *M* Tris–HCl (pH 8.0). The resulting DNA preparations were stored in 96-well microtiter plates at −20°C. Aliquots of DNA samples from half of a plate (48 YACs) were pooled, diluted 1:2 with H_2O; 10 μl of diluted pooled DNA was used as the template for STS PCR in a 30-μl total reaction volume. For PCR amplification of individual YAC clones, a small aliquot of DNA solution was inoculated into 20 μl of PCR buffer using the 96-well replicating device from Washington University. The conditions for STS screening of YAC DNA pools with EST primers were the same as those described for plasmid insert amplification using T3 and T7 primers (see above). For STS screening with microsatellites, PCR parameters were as described for EST primers using the buffer described for *Alu* PCR.

Primers for microsatellite markers *D2S102, D2S104, D2S115, D2S116, D2S117, D2S118, D2S120, D2S124, D2S125, D2S126, D2S128, D2S130, D2S137, D2S138, D2S140, D2S143, D2S148, D2S152, D2S153, D2S154, D2S155, D2S157, D2S159, D2S163, D2S164, D2S172, D2S173, D2S206,* and *D2S211* and primers for the *CRYG1B* locus were obtained from Research Genetics (Huntsville, AL). Primer sequences for the loci *D2S295, D2S301, D2S317, D2S334,* and *D2S371* were obtained from the Genome Data Base (Johns Hopkins University); oligonucleotides were synthesized at the Institut de Recherche en Biotechnologie (Montreal, Canada) on an ABI DNA synthesizer. STS primers for the loci *AGXT* (Purdue *et al.*, 1991), *ALPI* (Griffin *et al.*, 1987), *ALPP* (Martin *et al.*, 1987), *COL3A1* and *COL5A2* (Huerre-Jeanpierre *et al.*, 1985), *FN1* (Henry *et al.*, 1985), *GCG* (Schroeder *et al.*, 1984; Tricoli *et al.*, 1984; Spurr *et al.*, 1994), *GYPC* (Mattei *et al.*, 1986), *IL1B* (Webb *et al.*, 1986), *INHA* (Barton *et al.*, 1989), *PROC* (Kato *et al.*, 1988), and *TNP1* (Luerssen *et al.*, 1990) were synthesized according to the sequences listed by Spurr *et al.* (1994). Sequences for *CD28* (Lafage-Pochitaloff *et al.*, 1990), *DES* (Viegas-Péquignot *et al.*, 1989), and *MYL1* (Cohen-

Haguenauer *et al.*, 1988) were obtained from Jeggo *et al.* (1993). Primer sequences for *CPS* (Hoshide *et al.*, 1993) were obtained from Dr. W. Cramers (University of Amsterdam, The Netherlands); primers for *ERCC3* were kindly provided by Dr. G. Weeda (Erasmus University, Rotterdam, The Netherlands); and primers for *UGT1A1* (Van Es *et al.*, 1993) were kindly provided by Dr. P. J. Bosma (Academic Medical Centre, Amsterdam, The Netherlands). STS primers were designed for *IL8R* (Morris *et al.*, 1992), *NRAMP* (Cellier *et al.*, 1994), *PAX3* (Foy *et al.*, 1990), and *VIL* (Rousseau-Merck *et al.*, 1988; Pringault *et al.*, 1991); these primer sequences are available on request from the authors.

RESULTS

YAC Library Pooling Scheme

The long-term aim of our study is to generate a set of overlapping YAC clones covering human chromosome 2. The results presented in this communication are a first step toward this goal. We have successfully adapted IRS PCR to high-throughput YAC library screening by using a novel pooling scheme for 55,296 YAC clones that correspond to plates 1 to 552 of the CEPH Mark I YAC libraries supplemented with copies of 24 plates from the original Washington University YAC library. These 55,296 YAC clones contain human genomic DNA inserts that are expected to constitute approximately a 7- to 7.5-fold genome coverage. The pooling scheme allowed representation of the 55,296 clones on one small 2.5 × 9 inch membrane and thus greatly facilitated the screening of hundreds of probes. YAC DNA from row, column, half-floor, and plate pools was used as templates for IRS PCR. The amplification products of individual pools were dot blotted onto membranes to generate two types of filters: a row/column/half-floor filter containing 192 row, 192 column, and 192 half-floor pools (that is, a total of 576 pools) and a plate pool filter that contained the 576 plate pools. Each plate pool represents the IRS-PCR products of 96 YAC clones, whereas each of the row, column, and half-floor pools represents 288 YAC clones. Thus, each YAC should be represented four times, once on the plate pool filter and three times on the row/column/half-floor filter. A schematic outline of the IRS-PCR strategy is given in Fig. 1. An example of an IRS probe hybridization screen of the YAC IRS-PCR pool filters is given in Fig. 2. The probe designated 70 was derived from a chromosome 2 somatic cell hybrid, GM10826B, and hybridized to 5 plate pools and 15 row, column, or half-floor pools. Thus, 5 distinct YAC clones were identified by probe 70. This example illustrates that screening of IRS-PCR products from pools of YACs with IRS probes results in hybridization signals that clearly distinguish positive clones from background.

YAC Clone Identification

To identify the specific YAC clones in the pools that hybridized with the IRS probes in screening the IRS YAC pool filters, the YAC address was derived from

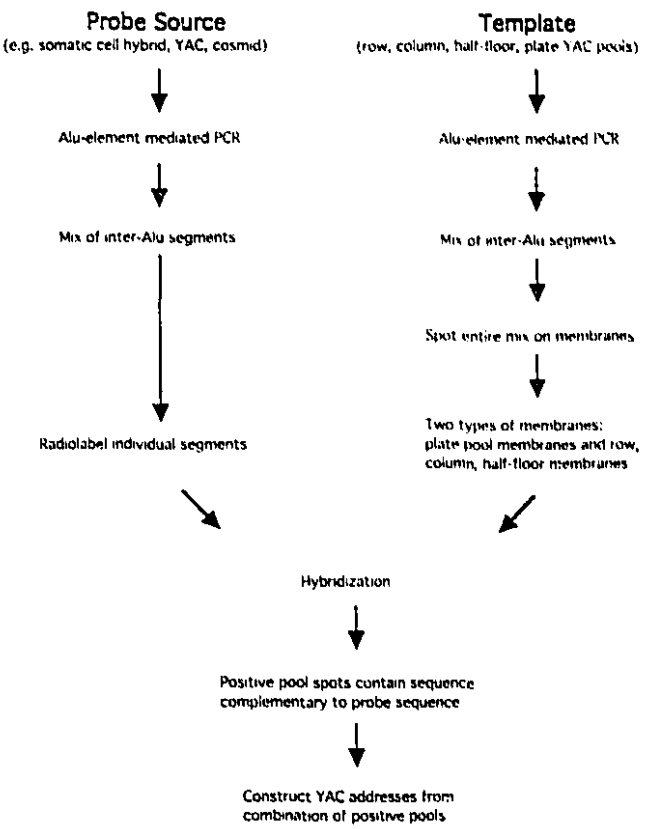


FIG. 1. Flow diagram of IRS-PCR hybridization screening of YAC pools employed in our experiments (for details see text).

the plate, row, and column coordinates, which are represented by the IRS-PCR products spotted on the filters. The half-floor pools introduce a redundancy that serves as an internal control for the assignment of the YAC address. Initially, the autoradiograms were visually inspected, and hybridization signals that were detected on duplicate filters were recorded and used for the assignment of addresses. We have recently standardized the assignment of YAC addresses and reduced the error rate associated with pool scoring and address construction by digitizing the autoradiograms, assigning coordinates to positive pools, and implementing an algorithm for the assignment of corresponding YAC addresses. As in all complex pooling systems, unambiguous addresses could not be assigned to all YAC clones detected by specific probes. For example, if a probe was found to hybridize to two different YAC clones that are contained in the same plate, an unequivocal assignment of row and column could not be made. We have successfully employed two approaches to resolve such ambiguities. First, in regions of high density of IRS probes degenerate addresses could be resolved since adjacent probes in the same contig hybridized to only one of the ambiguous sets of addresses. Second, IRS-PCR products generated from individual YAC clones were dot blotted and rescreened with the probe that

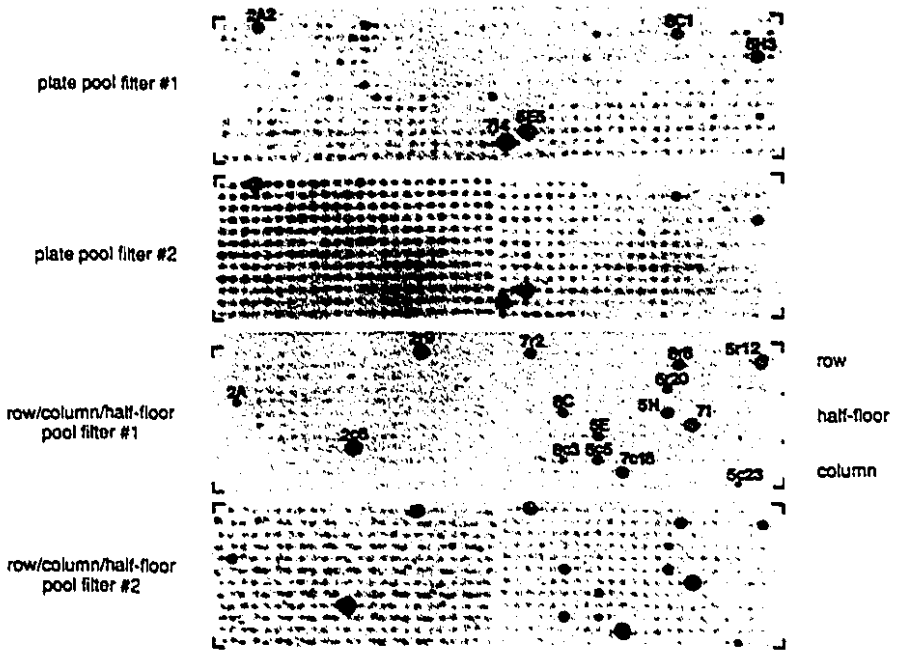


FIG. 2. Autoradiogram of YAC library screening with *Alu*-PCR probe 70. Duplicate filters are indicated as 1 and 2. The locations for the IRS-PCR products from row, half-floor, and column pools are indicated to the right of row/column/half-floor pool filter 1. Signals that appeared on duplicate filters were considered positive, and the corresponding pools are indicated above each spot on one of the duplicate filters. Probe 70 identified plate pools 2A2, 5E5, 5H3, 7I4, and 8C1; row pools 2r9, 5r12, 5r20, 7r2, and 8r8; column pools 2c6, 5c23, 5c5, 7c18, and 8c3; and half-floor pools 2A, 5E, 5H, 7I, and 8C. In the pooling scheme, row pools 1 to 8, 9 to 16, and 17 to 24 correspond to rows A to H of plates 1 and 4, 2 and 5, and 3 and 6, respectively. Column pools 1 to 12 and 13 to 24 correspond to columns 1 to 12 of plates 1 to 3 and 4 to 6, respectively. Thus, the combination of these address components identified 5 distinct YAC clones: 2A2A06 from pools 2A2, 2r9, 2c6, and 2A; 5E5D11 from pools 5E5, 5r12, 5c23, and 5E; 5H3D05 from pools 5H3, 5r20, 5c5, and 5H; 7I4B06 from pools 7I4, 7r2, 7c18, and 7I; 8C1H03 from pools 8C1, 8r8, 8c3, and 8C.

detected the ambiguous address. To improve the efficiency of rescreening individual YAC clones, groups of 10 to 12 probes that had detected ambiguous YAC addresses were pooled in such a way that none of the probes in a pool shared any potential YAC addresses.

Probe Generation

To generate IRS probes for screening of the YAC library, we used the same Alu S/Alu J and Alu 263–283.1/2 primers that had been used for IRS PCR of the YAC pool DNA. Both primer pairs are specific for human *Alu* repeats and can therefore be used for the selective amplification of human DNA from human–rodent somatic cell hybrids and from human DNA inserts in YAC or cosmid clones in the presence of bacterial or yeast genomic DNA. The human chromosome 2 rodent–somatic cell hybrids used in our experiments are summarized in Table 1. The predominant choice of somatic cell hybrids carrying inserts of DNA derived from chromosome region 2q32–q37 reflects our short-term goal of obtaining continuous YAC coverage of the telomeric region of chromosome 2q. In addition to somatic cell hybrids, a number of YAC and cosmid clones carrying loci previously assigned to chromosome 2q33–q37 were used as templates for probe generation. YAC library clones that hybridized to the IRS-PCR probes derived from these chromosomally assigned YAC and cosmid clones were used as physical anchor points. Finally, several YAC clones were used for chromosome walking experiments.

The number of bands and the size of the IRS-PCR amplification products obtained from different templates were found to vary widely when analyzed on 1.3% agarose gels (data not shown). Individual cosmid templates yielded 0 to 3 bands, while YAC clones yielded 3 to 5 bands. Somatic cell hybrids that con-

TABLE 1

Somatic Cell Hybrids Used in the Generation of *Alu*-PCR Probes

Somatic cell hybrid line	Estimated chromosome 2 content	Reference
GM10826B	Entire chromosome	Drwinga *et al.* (1993)
014-1	2q24–q37	Jeggo *et al.* (unpublished data)
b3-1	Vicinity of *CRYG1*	Hafezparast *et al.* (unpublished data)
D6-2	Vicinity of *DES* and *MYL1*	Hafezparast *et al.* (1993)
D4-2	Vicinity of *FN1*	Hafezparast *et al.* (1993)
D7-2	Vicinity of *FN1*	Hafezparast *et al.* (1993)

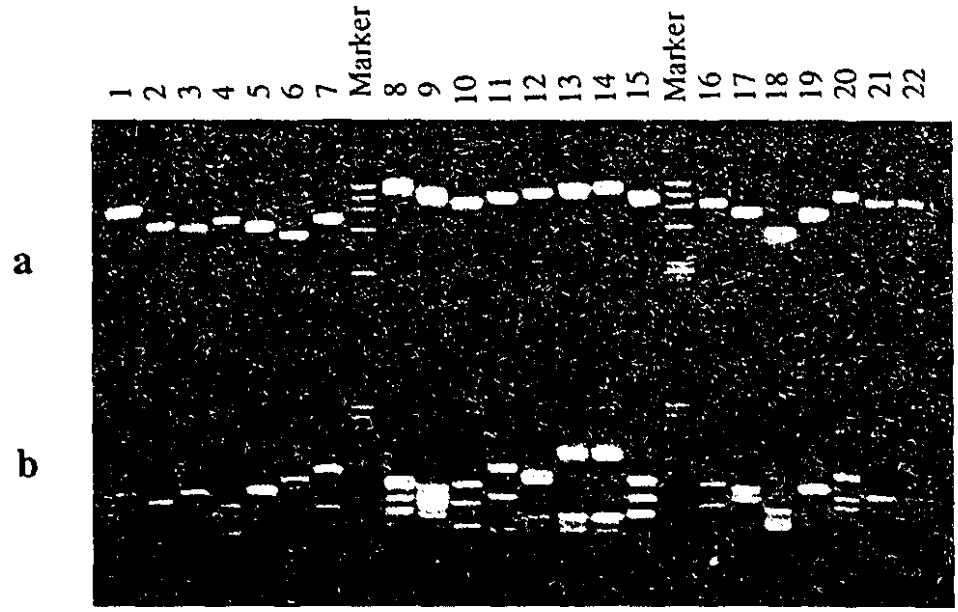


FIG. 3. Fingerprint of 22 arbitrarily chosen inter-*Alu* fragment probes. *Alu*-element-mediated PCR amplification products from somatic cell hybrids GM10826B and 014-1 were cloned into T-modified vectors. Amplified inserts were characterized according to their molecular weight (a) and the restriction fragment pattern obtained after digestion with *Rsa*I, *Hin*fI, and *Hae*III (b) on a 1.8% agarose gel stained with ethidium bromide. For example, IRS probes 3 and 5 are likely to be identical. The size marker is ΦX174 DNA digested with *Hae*III.

tained only relatively small amounts of chromosome 2 DNA yielded 1 to 6 bands. A total of 100 individual bands of *Alu*-PCR products derived from cosmids, YACs, or highly reduced somatic cell hybrids (b3-1, D6-2, D4-2, D7-2) that were clearly visible on ethidium bromide-stained agarose gels were directly excised from low-melting agarose gels and used as probes. However, somatic cell hybrids GM10826B and 014-1 gave a smear of *Alu*-PCR amplified products, which were cloned into T-modified vectors (Marchuk *et al.*, 1991), and several thousand recombinant clones were picked for further analysis. Prior to their use for library screening, inter-*Alu* sequence amplification products from somatic cell hybrids GM10826B and 014-1 were further characterized to avoid testing identical probes. A fingerprint pattern defined by the number and size of bands was obtained by digesting the amplified inserts of the recombinant T-vector clones with three frequent-cutting restriction enzymes, *Rsa*I, *Hin*fI, and *Hae*III (Fig. 3). Simply by sorting by the number and approximate molecular weight of the bands, we were able to compare newly generated probes with a database of previously generated probes and thus identify unique probes. Using this approach, we were able to select 696 unique probes from 2700 candidate probes derived from somatic cell hybrids GM10826B and 014-1.

Screening of YAC IRS-DNA Pools with Inter-Alu Probes

We used 796 unique chromosome 2-specific inter-*Alu* probes to screen filters of pooled DNA representing the 55,296 YAC clones. The success rate of detecting at least 1 library YAC clone ranged from 80 to 92% depending on the source of the probes, except for the 8 probes generated from cosmids, which all failed (Table 2). A probe failure is defined by the complete absence of any hybridization signal or the presence of a large number of hybridization signals that are probably due to the presence of unknown repeat elements in the IRS probe (less than 1%). No significant difference was observed in the rate of successful library screenings between probes generated from different somatic cell hybrids or individual YAC clones (χ^2 test, $P > 0.1$). However, probes derived from somatic cell hybrids that contained only a small piece of chromosome 2q or from individual YAC clones usually gave weaker hybridization signals and thus required longer exposure times.

TABLE 2

Analysis of Hybridization Screenings for 796 *Alu*-PCR Probes

Souce of *Alu*-PCR probes	Number of probes	Number successful[a]	% success
GM-10826B	345	300	87
014-1	351	289	82
Reduced somatic cell hybrids (b3-1, D6-2, D7-2 and D4-2)	12	11	92
YACs	80	64	80
Cosmids	8	0	0

[a] A probe was scored as successful if at least one YAC clone, either unambiguously or as part of a set of ambiguous addresses, was identified by that probe.

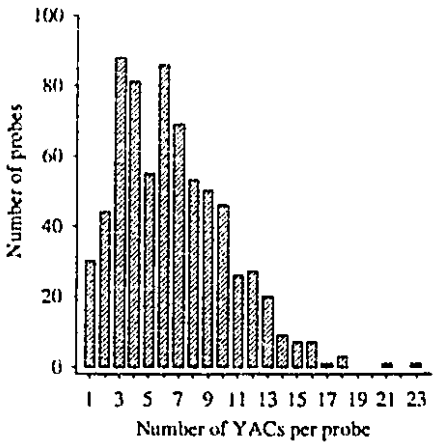


FIG. 4. Distribution of the minimum number of YACs detected per *Alu*-PCR probe. For example, 88 probes detected 3 YAC clones, that is, 3 different unambiguous addresses or sets of ambiguous addresses.

The number of YAC clones detected per probe ranged from 1 to 23, indicating variation in the number of YAC clones that contain inserts of a specific genomic region (Fig. 4). The overall distribution of the number of YACs hybridizing to specific probes was consistent with the suggestion that cross-hybridization is not a major problem in the screening strategy. We counted the number of YAC clones detected by a probe as the number of unambiguous addresses plus the number of ambiguous sets of addresses. Using this scoring system, each successful probe detected an average of 6.6 YACs, which is consistent with the data obtained for YAC cloning of chromosome 3 by other investigators using the same YAC library (Stanton *et al.*, unpublished data). The chimera rate for the CEPH Mark I YAC library was estimated to be 40% (Dausset *et al.*, 1992); therefore, on average, 4 YAC clones detected per probe should not be chimeric. Since the average insert size of YAC clones was determined to be approximately 430 kb (Albertsen *et al.*, 1990; Dausset *et al.*, 1992), and assuming that on average a core YAC at each end is extended 50% by additional YACs detected by the same probe, we estimate that on average each IRS probe detected YAC clones that covered approximately 800 kb of contiguous genomic DNA. We call this estimated 800 kb bin around each probe (the average contig size per probe) a library entry point.

Distribution of IRS Probes

The usefulness of IRS screening of YAC libraries for YAC contig assembly is dependent on the even and dense distribution of *Alu* elements along the chromosome. We looked for clustering of IRS probes by counting the number of probes that share at least one YAC address at each entry point. Based on genetic markers present in the hybrids, we estimate that the total human chromosome 2 genomic DNA contained in somatic cell hybrid GM10826B and 014-1 is 270 and 70 Mb, respectively. Assuming that each successful probe covers an average region of 800 kb, the average number of probes per entry point for hybrids GM10826B and 014-1 is 1 and 4, respectively. In other words, if the probes were evenly distributed one probe derived from somatic cell hybrid GM10826B and four probes derived from 014-1 should be detected per entry point. We further assumed that any number of probes in excess of three times the calculated average probe number per entry point reflected clustering of *Alu* repeats on chromosome 2. Under these assumptions, approximately 40% of the probes were clustered (Fig. 5). This percentage does not increase with higher probe density, which is to be expected if clustered and nonclustered repeats are amplified with the same efficiency. Thus, the results of our analysis show that regions of uneven probe density exist along the chromosome. However, this analysis also showed that the majority of inter-*Alu* probes were derived from regions of low to moderate probe density, and we conclude that *Alu* clustering should not be a major impediment to contig assembly in the chromosome region under investigation.

YAC Group and Contig Analysis

Next, we looked to what extent the 664 inter-*Alu* probes successfully used for library screenings allowed the identification of YAC groups. We defined a YAC group as the set of YAC clones that belong to the same contig; that is, a YAC group is composed of the unordered YAC clones in a contig. We assigned different probes and their associated YAC addresses to a YAC group if the probes shared at least 1 YAC address in common. In an initial analysis of unambiguous addresses, we identified 188 YAC groups. The number of YACs and probes per YAC group is summarized in Fig. 6. Next, we analyzed contig depth within the YAC groups and identified within the 6 largest YAC groups several points where the contig depth was 1. To avoid false connections caused by experimental error or chimeric YACs, YAC groups were tentatively subdivided at such positions. To test our assumption that the experimentally defined YAC groups can be used to construct contigs and to test the quality of the resulting contigs, we used the program CONTIGMAKER to order YACs within a YAC group. This analysis showed that YAC groups provide an excellent source of YAC clones for contig construction. An example of one of the resulting contigs is shown in Fig. 7. Further experiments aimed to connect different YAC groups are in progress.

STS Content Mapping

We determined the extent to which the telomeric region of chromosome 2q, which by design of our experiments had better probe coverage than the rest of the chromosome, was included in the YAC clones. For this purpose, all YAC clones identified in our previous ex-

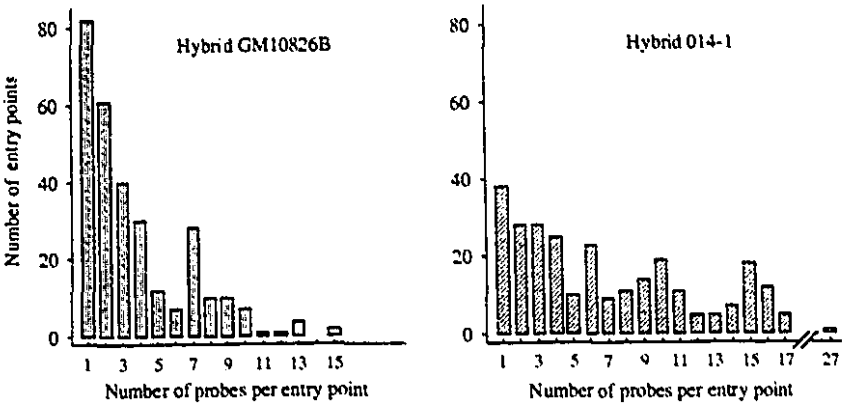


FIG. 5. Distribution of the number of inter-*Alu* probes per entry point. Each probe derived from somatic cell hybrid GM10826B or 014-1 defines an entry point (see text). The number of probes per entry point is the total number of probes that hybridize to at least one of the YAC clones detected by the entry probe. A high number of probes per 800-kb entry point bin indicates inter-*Alu* probe clustering in the region corresponding to the entry probe.

periments were screened for the presence of 23 EST and 34 microsatellite markers predicted to be in chromosome region 2q13–q37. As shown in Table 3, 22 of 23 (96%) of the ESTs and 31 of 34 (91%) of the microsatellites identified at least 1 YAC clone. Assuming that each half-plate pool of the chromosome 2 sublibrary that displayed a PCR product of expected size during STS screening contains at least 1 positive YAC, the average number of YACs per EST and microsatellite is 4.4 and 3.8, respectively. The expected number of YACs per STS in a chromosomal sublibrary is a function of the probe density applied to the initial library screen. Assuming a chimeric rate of 40% for the CEPH Mark I library and an average of 6.6 YACs per IRS probe, approximately 4 YACs per randomly chosen STS probe are expected to be detected by sublibrary screening. As the probe density in the IRS probe screen increases, so will the number of YACs per probe in the sublibrary STS screen. Further analysis revealed that only 4 ESTs and 2 of the microsatellites hybridized to regions of high *Alu*-element density, which suggests that ESTs and microsatellites used in our experiments were generally distributed independently of *Alu* elements along the chromosome. The 53 successful STSs detected YAC clones that belonged to 29 different YAC groups. Since the microsatellites were previously mapped genetically and the majority of expressed genes tested had previously been mapped cytogenetically, 20 of these YAC groups containing 605 distinct YACs could be ordered along the long arm of chromosome 2 (Fig. 8).

DISCUSSION

We applied IRS PCR to the screening of pooled YAC DNA to identify YAC clones with chromosome 2 DNA inserts. In total, 664 chromosome 2-specific IRS probes were generated, which on average detected 6.6 YAC clones. These YACs were distributed along the entire chromosome, and it is estimated that they cover up to 90% of chromosome region 2q13–q37, assuming that expressed genes and microsatellites were distributed randomly along this region of the chromosome. Al-

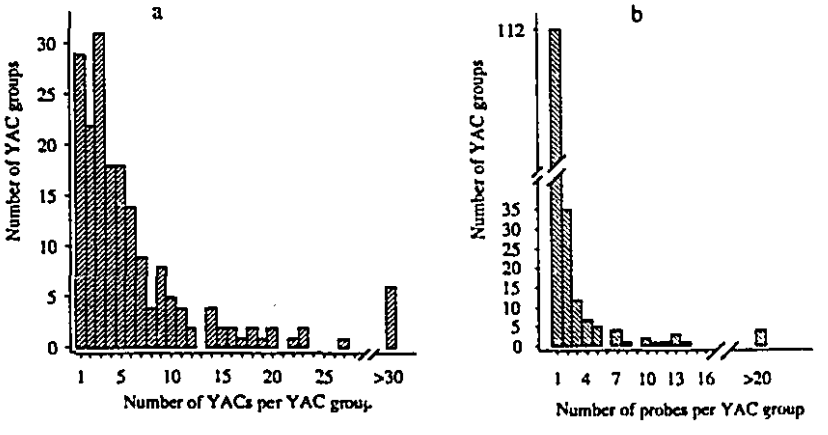


FIG. 6. Distribution of YAC groups by number of YAC addresses (a) and by number of probes (b). In a 6 YAC groups contain more than 30 YACs (33, 53, 59, 67, 193, and 199 YACs). In b 4 groups contain more than 20 probes (21, 30, 103, and 139 probes).

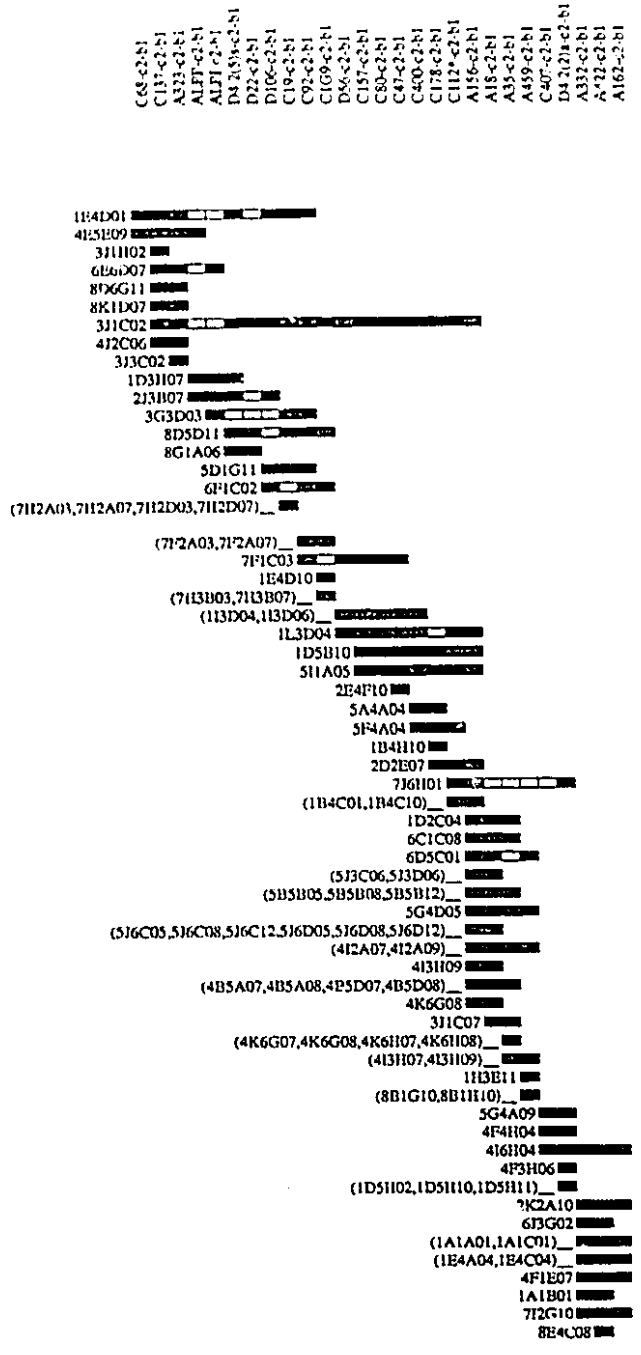


FIG. 7. Contig of the core region of YAC group 1 around the *ALPP* and *ALPI* loci in human chromosome region 2q37. The names of IRS probes are given at the top of the figure. YAC clones are schematically represented as small rectangular boxes. A solid area on a YAC clone underneath a probe indicates that the YAC was identified unambiguously by that probe, an open area indicates that the probe did not hybridize with the YAC, and a shaded area indicates that the YAC was identified as part of an ambiguous set of addresses by the probe. YAC addresses or sets of ambiguous addresses (in parentheses) are given to the left of each schematically represented YAC clone. The estimated length of the contig is 2 Mb.

though we have not yet defined one contiguous set of YAC clones spanning the entire telomeric region of the long arm of chromosome 2, we assume that contig gaps are uncommon if 90% coverage of this region has been achieved. Furthermore, the availability of chromosome region-specific YAC clones should advance the molecular analysis of human chromosome 2q even before the establishment of a complete contig. The high-resolution genetic maps of chromosome 2 established by the NIH/CEPH consortium and by Généthon use different sets of microsatellites, and only very few expressed loci are included in these maps (NIH/CEPH Collaborative Mapping Group, 1992; Spurr *et al.*, 1992; Weissenbach *et al.*, 1992). Sequential STS screening of the close to 2000 distinct chromosome 2 YAC clones should help to integrate these different genetic maps with physical maps of chromosome 2. Such integrated maps will prove particularly useful in positional cloning of medically important loci on chromosome 2q, such as *XRCC5* (Jeggo *et al.*, 1992), the insulin-dependent diabetes susceptibility modifier gene (Cornall *et al.*, 1991), and the gene causing familial juvenile nephronophtisis (Antignac *et al.*, 1993).

We have chosen the *Alu* repeat as the IRS element for IRS-PCR-based chromosome 2 YAC cloning experiments (Schmid and Jelinek, 1982). *Alu* elements occur in approximately 500,000 to 1,000,000 copies in the haploid human genome and represent the most common IRS element in human DNA (Schmid and Deininger, 1975; Singer, 1982). Thus, if *Alu* elements were evenly distributed, one *Alu* element should be found every 3 to 5 kb in human DNA. The utility of an *Alu*-element-based strategy for the cloning of extended genome regions depends on the even distribution of *Alu* repeats in the human genome. However, experiments using fluorescence *in situ* hybridization of labeled *Alu* element probes to metaphase chromosomes suggested that *Alu* elements are not evenly distributed along human chromosomes but concentrated in R-bands (Baldini and Ward, 1991). These findings are not in agreement with a recent report analyzing the nucleotide composition of individual microdissected chromosome bands. Sequencing of randomly selected DNA clones from R- and G-bands did not detect any uneven distribution of *Alu* elements between R- and G-bands on chromosome 11 (Yoshiura *et al.*, 1994). A detailed analysis of IRS localization over 2 Mb of cloned chromosome DNA showed a significantly uneven distribution of both *Alu* and L1 elements (Arveiler and Porteous, 1992). Likewise, regions of preference or exclusion for the integration of *Alu* sequences in the human genome were reported by other investigators (Moyzis *et al.*, 1989). Our data suggest that clusters of *Alu* elements occur along human chromosome 2. However, our data also suggest that the *Alu* density on chromosome 2 is sufficiently high that *Alu*-poor regions should not be a major obstacle for an *Alu*-PCR-based contig assembly strategy for chromosome 2.

TABLE 3

Analysis of Chromosome 2-Specific YACs for the Presence of Expressed Loci and Microsatellite Markers[a]

Expressed loci				D segments			
Success	Localization	Failure	Localization	Success	Localization	Failure	Localization
IL1B	2q13–q21	*AGXT*	2q37.3	D2S102	2q33–q37	D2S128	2q33–q37
PROC	2q13–q21			D2S104	2q33–q37	D2S140	2q
GYPC	2q14–q21			D2S115	2q21–q33	D2S152	2q21–q33
ERCC3	2q21			D2S116	2q		
GCG	2q23			D2S117	2q		
COL3A1	2q31–q32.3			D2S118	2q		
COL5A2	2q31–q32.3			D2S120	2q33–q37		
CRYGA	2q33–q35			D2S124	2q21–q33		
CPS1	2q33–q36			D2S125	2q		
MYL1	2q33–q34			D2S126	2q33–q37		
CD28	2q33			D2S130	2q		
FN1	2q34			D2S137	2q33–q37		
TNP1	2q34			D2S138	2q21–q33		
NRAMP	2q35			D2S143	2q33–q37		
IL8R	2q35			D2S148	2q21–q33		
VIL1	2q35			D2S153	2q		
INHA	2q33–q34			D2S154	2q		
DES	2q35			D2S155	2q33–q35		
PAX3	2q36			D2S157	2q		
ALPP	2q37.1			D2S159	2q33–q37		
ALPI	2q37.1			D2S163	2q33–q37		
UGT1A1	2q37			D2S164	2q33–q37		
				D2S172	2q33–q37		
				D2S173	2q33–q37		
				D2S206	2q33–q37		
				D2S211	2q34–q37		
				D2S295	2q		
				D2S301	2q		
				D2S317	2q		
				D2S334	2q		
				D2S371	2q		
Success rate: 22/23 (96%)				Success rate: 31/34 (91%)			

[a] The cytogenetic localizations were taken from the Genome Data Base (Johns Hopkins University, Baltimore, MD).

Human *Alu* elements are approximately 300 bp in length and are composed of two highly homologous direct repeats of an ancestral sequence (Schmid and Jelinek, 1982; Britten *et al.*, 1988; Jurka and Milosavljevic, 1991). Although the direct repeat structure is specific to primates, *Alu*-homologous sequences are found in many other species, including hamster and mouse, which provide the most common genomic background for human–rodent somatic cell hybrids. The primers used in our experiments are complementary to primate-specific sequences in the *Alu* element and thus allowed the selective amplification of human sequences from a panel of human–rodent somatic cell hybrids. Similar primers have been generated by other investigators for the same purpose (Nelson *et al.*, 1989; Bernard *et al.*, 1991; Nelson, 1991; Tagle and Collins, 1992). Thus, by using appropriately selected human–rodent somatic cell hybrids, IRS probe generation can be used in well-defined genome regions. This notion is supported by our results, which showed that 85% of IRS probes derived from somatic cell hybrids of different chromosome coverage can be used for successful screening of YAC DNA pools. Somatic cell genetics and IRS-PCR-based YAC contig construction therefore emerge as complementary approaches for the generation of physical maps of human chromosomes. In addition, the high frequency of identified YAC clones by library screenings with inter-*Alu* probes derived from individual YAC clones demonstrates that IRS PCR is very suitable for contig extension using YAC clones from the ends of a contig as templates for IRS probe generation. We conclude from the overall high success rate of YAC library screenings using IRS probes derived from a wide variety of templates that the IRS-PCR approach is a fast, efficient, and economical approach for cloning of a targeted genome region.

The IRS strategy for YAC cloning assumes that the IRS probes used for screening of the YAC library will hybridize only to unique DNA fragments in the human genome. However, non-*Alu* repeat elements (Schmid and Deininger, 1975; Schmid and Jelinek, 1982; Singer, 1982; Rubin *et al.*, 1993; Singer and Skowron-

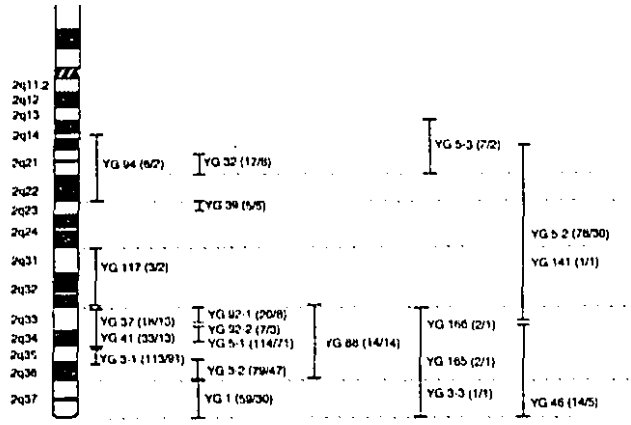


FIG. 8. Distribution of YAC groups along human chromosome 2q. A total of 20 YAC groups comprising 605 distinct YACs contained either microsatellites or ESTs for which the regional localization on chromosome 2 was known. It was therefore possible to order these YAC groups along the chromosome. An idiogram of chromosome 2q with corresponding chromosomal band designations is given at the left. Bars indicate the cytogenetic localization of YAC groups according to previously mapped markers that are included in the YAC groups. In cases where YAC groups contained several markers with known subchromosomal localization, the most narrow cytogenetic assignment was used to localize the YAC group. Following the name of the YAC group (YG), the number of YACs and probes contained in the YAC group is given in parentheses.

ski, 1985) may occur in inter-*Alu* probes generated by *Alu* PCR. These probes would probably hybridize to YAC clones that are not physically linked to the chromosome region of interest. If undetected, probes containing dispersed repetitive sequences would limit the validity of contig construction. For example, repeat-element mediated cross-hybridization has been observed for the *D5F40S1* locus on chromosome 5 (Bernard and Wood, 1993). In this case a low-copy repetitive element was located 3′ to the locus but also was present on chromosomes 2 and 17. In our experiments such low-copy repeats are unlikely to be an impediment to contig assembly. Like ambiguous addresses or chimeric YAC clones, errors in ordering YAC clones in a contig will be identified by having sufficient YAC overlap as a result of high probe density and by STS content mapping as we illustrated for the telomeric region of chromosome 2q. The presence of high-copy repeat elements in IRS probes used for library screening will give rise to many positive hybridization signals on the pooled DNA filters, which does not allow the identification of individual YAC addresses. It is noteworthy that such sweeping hybridization of inter-*Alu* probes to YAC DNA pools was observed for less than 1% of probes tested in our experiments.

A cloning strategy that is based on the random selection of clones present the problem of contig closure and thus requires experimental strategies that allow the preferential generation of probes to close gaps between contigs. The IRS strategy allows the identification of probes that fall into such contig gaps in a relatively straightforward manner. We have spotted approximately 4000 additional probes from somatic cell hybrid 014-1 on hybridization membranes, which will then be probed with pools of labeled *Alu*-PCR products obtained from each of the 188 YAC groups. Probes that do not hybridize with YAC contig-derived *Alu*-PCR products fall into contig gaps and can be efficiently used for extending the contig. In addition, by focusing on probe generation from YAC clones located at contig ends, it should be possible to extend and close gaps between contigs efficiently. We consider that this direct IRS-tag approach is less work-intensive and less error-prone than analyzing *Alu* fingerprints and their overlaps (Bellanné-Chantelot *et al.*, 1992). At the present stage of our experiments, possible sources of contig errors could come from intrachromosomal YAC chimeras if contig extension is achieved by only single YACs. Using the approach that we have described, we plan to concentrate on the establishment of a complete contig of chromosome region 2q24–q37 and the integration of our data with those obtained from different YAC libraries before extending our efforts to the entire chromosome.

Recently, a first-generation physical map of the human genome was published by Généthon (Cohen *et al.*, 1993). The challenge provided to other groups is to confirm and refine these data. The main source of errors in this first-generation physical map is the occurrence of false connections within contigs caused by syntenic chimeric YACs or the use of multilocus probes that recognize highly homologous sequences that are present on different parts of the same chromosome (Bellanné-Chantelot *et al.*, 1992). In addition, the nonclonability of certain genomic sequences in YACs will contribute to contig gaps in the YAC map. The first source of error is directly addressed by our experiments. While the published Généthon map relied predominantly on the later generation YAC library with average insert size of 900 kb (mega-YAC library), we used the CEPH Mark I library, which has a smaller insert size of approximately 430 kb. It is highly unlikely that the same syntenic chimeras will be found in both YAC libraries. Thus, our data provide additional contig depth, which will help to resolve false connections caused by chimeric YAC clones, and the integration of results achieved with different libraries will significantly increase the likelihood of a valid contig. In addition, we have initiated the screening of the CEPH Mark I and mega-YAC libraries with the same IRS probes. Preliminary results of these experiments suggest that the genomic sequences contained in both libraries are not identical: six probes did not hybridize with mega-YAC library clones but detected a set of YAC clones in the CEPH Mark I library. Similarly, one probe detected two mega-YAC clones but none in the CEPH Mark I library. The complementarity of our results and already existing physical maps is best illustrated by the different strategies applied by us and the Généthon group. The Géné-

thon map, which is based predominantly on STS probes, provides an excellent integration of genetic maps with the YAC map. On the other hand, most cosmid and BAC libraries are set up for hybridization screening, and the probes generated in our experiments will be useful for the integration of physical maps generated from smaller insert genomic libraries with the YAC map. Likewise, preliminary results from our laboratory suggest that IRS probes may be useful for the establishment of PFGE maps of selected genome regions.

In addition to the human and mouse genome projects, numerous efforts to generate genetic and physical maps for economically important livestock species, such as cattle (Libert *et al.*, 1993), are under way. While STS-based strategies will play an important role in the generation of physical maps for these species, it is reasonable to predict that the high cost of STS screening and the logistical requirements needed for the efficient application of hierarchical STS-PCR screening will be a major impediment to STS-based physical mapping projects. IRS-PCR-based screening strategies offer a cost-efficient complement to STS-based strategies. *Alu* elements are primate-specific; however, IRS elements are widely found in the genomes of mammalian species, and we suggest that IRS strategies can be applied to the genome analysis of nonprimate species. Recent results indicate that *Alu*-like elements are evenly distributed throughout the bovine and ovine genomes (Rajcan-Separovic and Sabour, 1993). Of course, species-specific *Alu*-like sequence segments need to be defined prior to the application of IRS PCR to other mammalian species. For several livestock species, panels of somatic cell hybrids have already been established and could be used as templates for the generation of regional chromosomal probes to be used in targeted cloning of defined genome regions. The cost-efficiency and the simplicity of the experimental design strongly argue in favor of a repeat-element PCR-based cloning strategy as a major component for the physical analysis of genomes of other mammalian species in addition to human and mouse.

ACKNOWLEDGMENTS

We thank Leah Simkin for advice on database management and for the program to define YAC groups. We also thank Dr. Penny Jeggo (MRC Cell Mutation Unit, Brighton, UK) for the generous gift of somatic cell hybrids and for sharing unpublished data. This work was supported by grants from the Canadian Genome Analysis and Technology Program and the Canadian Genetic Diseases Network (Federal NCE program). Work in the laboratory of D.H. is supported by a grant from the Human Genome Office (000297). J.L. is supported by a predoctoral studentship from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche, P.G. is supported by an E. W. R. Steacie Memorial Fellowship from the Natural Sciences and Engineering Research Council of Canada and is an International Research Scholar of the Howard Hughes Medical Institute, and E.S. is a scholar of the Medical Research Council of Canada.

REFERENCES

Albertsen, H. M., Abderrahim, H., Cann, H. M., Dausset, J., Le Paslier, D., and Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. USA* **87:** 4256–4260.

Antignac, C., Arduy, C. H., Beckmann, J. S., Benessy, F., Gros, F., Medhioub, M., Hildebrandt, F., Dufier, J.-L., Kleinknecht, C., Broyer, M., Weissenbach, J., Habib, R., and Cohen, D. (1993). A gene for familial juvenile nephronophthisis (recessive medullary cystic kidney disease) maps to chromosome 2p. *Nature Genet.* **3:** 342–345.

Ardinger, H. H., Buetow, K. H., Bell, G. I., Bardach, J., Van Demark, D. R., and Murray, J. C. (1989). Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. *Am. J. Hum. Genet.* **45:** 348–353.

Arveiler, B., and Porteous, D. J. (1992). Distribution of Alu and L1 repeats in human YAC recombinants. *Mamm. Genome* **3:** 661–668.

Baldini, A., and Ward, D. C. (1991). In situ hybridization banding of human chromosomes with Alu-PCR products: A simultaneous karyotype for gene mapping studies. *Genomics* **9:** 770–774.

Baldwin, C. T., Hoth, C. F., Amos, J. A., da Silva, E. O., and Milunsky, A. (1992). An exonic mutation in the HuP2 paired domain gene causes Waardenburg's syndrome. *Nature* **355:** 637–638.

Barr, F. G., Galili, N., Holick, J., Biegel, J. A., Rovera, G., and Emanuel, B. S. (1993). Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. *Nature Genet.* **3:** 113–117.

Barton, D. E., Yang-Feng, T. L., Mason, A. J., Seeburg, P. H., and Francke, U. (1989). Mapping of genes for inhibin subunits α, β_A, and β_B on human and mouse chromosomes and studies of *jsd* mice. *Genomics* **5:** 91–99.

Bellanné-Chantelot, C., Lacroix, B., Ougen, P., Billault, A., Beaufils, S., Bertrand, S., Georges, I., Glibert, F., Gros, I., Lucotte, G., Susini, L., Codani, J.-J., Gesnouin, P., Pook, S., Vaysseix, G., Lu-Kuo, J., Ried, T., Ward, D., Chumakov, I., Le Paslier, D., Barillot, E., and Cohen, D. (1992). Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* **70:** 1059–1068.

Bernard, L. E., and Wood, S. (1993). Human chromosome 5 sequence primer amplifies Alu polymorphisms on chromosomes 2 and 17. *Genome* **36:** 302–309.

Bernard, L. E., Brooks-Wilson, A. R., and Wood, S. (1991). Isolation of DNA fragments from a human chromosomal subregion by *Alu* PCR differential hybridization. *Genomics* **9:** 241–246.

Britten, R. J., Baron, W. F., Stout, D. B., and Davidson, E. H. (1988). Sources and evolution of human Alu repeated sequences. *Proc. Natl. Acad. Sci. USA* **85:** 4770–4774.

Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J., Schlessinger, D., and Olson, M. V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244:** 1348–1351.

Burke, D. T. (1991). The role of yeast artificial chromosome clones in generating genome maps. *Curr. Opin. Genet. Dev.* **1:** 69–74.

Burke, D. T., Carle, G. F., and Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236:** 806–812.

Cellier, M., Govoni, G., Vidal, S., Kwan, T., Groulx, N., Liu, J., Sanchez, F., Sklamene, E., Schurr, E., and Gros, P. (1994). Human natural resistance associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization and tissue specific expression. *J. Exp. Med.* **180:** 1741–1752.

Chenevix-Trench, G., Jones, K., Green, A. C., Duffy, D. L., and Martin, N. G. (1992). Cleft lip with or without cleft palate: Associa-

tions with transforming growth factor alpha and retinoic acid receptor loci. *Am. J. Hum. Genet.* **51:** 1377–1385.

Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billaut, A., Guasconi, G., Gervy, P., LeGall, I., Soularue, P., Grinas, L., Bougueleret, L., Bellanné-Chantelot, C., Lacroix, B., Barillot, E., Gesnouin, P., Pook, S., Vaysseix, G., Frelat, G., Schmitz, A., Sambucy, J.-L., Bosch, A., Estivill, X., Weissenbach, J., Vignal, A., Riethman, H., Cox, D., Patterson, D., Gardiner, K., Hattori, M., Sakaki, Y., Ichikawa, H., Ohki, M., Le Paslier, D., Heilig, R., Antonarakis, S., and Cohen, D. (1992a). Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359:** 380–387.

Chumakov, I. M., Le Gall, I., Billault, A., Ougen, P., Soularue, P., Guillou, S., Rigault, P., Bui, H., de Tand, M.-F., Barillot, E., Abderrahim, H., Cherif, D., Berger, R., Le Paslier, D., and Cohen, D. (1992b). Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nature Genet.* **1:** 222–225.

Cohen, D., Chumakov, I., and Weissenbach, J. (1993). A first-generation physical map of the human genome. *Nature* **366:** 698–701.

Cohen-Haguenauer, O., Barton, P. J. R., Van Cong, N., Serero, S., Gross, M.-S., Jegou-Foubert, C., de Tand, M.-F., Robert, B., Buckingham, M., and Frezal, J. (1988). Assignment of the human fast skeletal muscle myosin alkali light chains gene (MLC1F/MLC3F) to 2q32.1–2qter. *Hum. Genet.* **78:** 65–70.

Cornall, R. J., Prins, J.-B., Todd, J. A., Pressey, A., DeLarato, N. H., Wicker, L. S., and Peterson, L. B. (1991). Type 1 diabetes in mice is linked to the interleukin-1 receptor and Lsh/Ity/Bcg genes on chromosome 1. *Nature* **353:** 262–266.

Dausset, J., Ougen, P., Abderrahim, H., Billault, A., Asmbucy, J.-L., Cohen, D., and Le Paslier, D. (1992). The CEPH YAC library. *Behring Inst. Mitt.* **91:** 13–20.

Drwinga, H. L., Toji, L. H., Kim, C. H., Greene, A. E., and Mulivor, R. A. (1993). NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. *Genomics* **16:** 311–314.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–13.

Fishel, R., Lescoe, M. K., Rao, M. R. S., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. *Cell* **75:** 1027–1038.

Foote, S., Vollrath, D., Hilton, A., and Page, D. C. (1992). The human Y chromosome: Overlapping DNA clones spanning the euchromatic region. *Science* **258:** 60–66.

Foy, C., Newton, V., Wellesley, D., Harris, R., and Read, A. P. (1990). Assignment of the locus for Waardenburg syndrome type I to human chromosome 2q37 and possible homology to the splotch mouse. *Am. J. Hum. Genet.* **46:** 1017–1023.

Green, E. D., and Olson, M. V. (1990). Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **87:** 1213–1217.

Green, E. D., Mohr, R. M., Idol, J. R., Jones, M., Buckingham, J. M., Deaven, L. L., Moyzis, R. K., and Olson, M. V. (1991). Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: Application to the mapping of human chromosome 7 using yeast artificial chromosomes. *Genomics* **11:** 548–564.

Griffin, C. A., Smith, M., Henthorn, P. S., Harris, H., Weiss, M. J., Raducha, M., and Emanuel, B. S. (1987). Human placental and intestinal alkaline phosphatase genes map to 2q34–q37. *Am. J. Hum. Genet.* **41:** 1025–1034.

Grundy, H. O., Niemeyer, P., Rupani, M. K., Ward, V. F., and Wassman, E. R. (1989). Prenatal detection of cyclopia associated with interstitial deletion of 2p. *Am. J. Med. Genet.* **34:** 268–270.

Hafezparast, M., Kaur, G. P., Zdzienicka, M., Athwal, R. S., Lehmann, A. R., and Jeggo, P. A. (1993). Subchromosomal localization of a gene (XRCC5) involved in double strand break repair to the region 2q34–36. *Somatic Cell and Mol. Genet.* **19:** 413–421.

Hecht, B. K.-M., Hecht, F., and Munke, M. (1991). Forebrain cleavage gene causing holoprosencephaly: Deletion mapping to chromosome band 2p21. *Am. J. Med. Genet.* **40:** 130.

Henry, I., Jeanpierre, M., Weil, D., Grzeschik, K. H., Ramirez, F., and Junien, C. (1985). The structural gene for fibronectin (FN) maps to 2q32.3–qter. *Cytogenet. Cell Genet.* **40:** 650.

Hoshide, R., Matsuura, T., Haraguchi, Y., Endo, F., Yoshinaga, M., and Matsuda, I. (1993). Carbamyl phosphate synthetase I deficiency. *J. Clin. Invest.* **91:** 1884–1887.

Huerre-Jeanpierre, C., Henry, I., Mattei, M. G., Weil, D., Grzeschik, K. H., Chu, M. L., Bernard, M., Ramirez, F., and Junien, C. (1985). The gene for human alpha-1 type III collagen (COL3A1) is physically linked to alpha-2 type V collagen COL5A2 on chromosome 2 (2q31–2q32.3). *Cytogenet. Cell Genet.* **40:** 657.

Jeggo, P. A., Hafezparast, M., Thompson, A. F., Broughton, B. C., Kaur, G. P., Zdzienicka, M. Z., and Athwal, R. S. (1992). Localization of a DNA repair gene (XRCC5) involved in double-stranded-break rejoining to human chromosome 2. *Proc. Natl. Acad. Sci. USA* **89:** 6423–6427.

Jeggo, P. A., Hafezparast, M., Thompson, A. F., Kaur, G. P., Sandhu, A. K., and Athwal, R. S. (1993). A hamster–human subchromosomal hybrid cell panel for chromosome 2. *Somatic Cell Mol. Genet.* **19:** 39–49.

Jurka, J., and Smith, T. (1988). A fundamental division in the Alu family of repeated sequences. *Proc. Natl. Acad. Sci. USA* **85:** 4775–4778.

Jurka, J., and Milosavljevic, A. (1991). Reconstruction and analysis of human Alu genes. *J. Mol. Evol.* **32:** 105–121.

Kato, A., Miura, O., Sumi, Y., and Aoki, N. (1988). Assignment of the human protein C gene (PROC) to chromosome region 2q14–q21 by in situ hybridization. *Cytogenet. Cell Genet.* **47:** 46–47.

Lafage-Pochitaloff, M., Costello, R., Couez, D., Simonetti, J., Monnoni, P., Mawas, C., and Olive, D. (1990). Human CD28 and CTLA-4 Ig superfamily genes are located on chromosome 2 at bands q33–q34. *Immunogenetics* **31:** 198–201.

Libert, F., Lefort, A., Okimoto, R., Womack, J., and Georges, M. (1993). Construction of a bovine genomic library of large yeast artificial chromosome clones. *Genomics* **18:** 270–276.

Luerssen, H., Mattei, M.-G., Schroter, M., Grzeschik, K.-H., Adham, I. M., and Engel, W. (1990). Nucleotide sequence of the gene for human transition protein 1 and its chromosomal localization on chromosome 2. *Genomics* **8:** 324–330.

Marchuk, D., Drumm, M., Saulino, A., and Collins, F. S. (1991). Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19:** 1154.

Martin, D., Tucker, D. F., Gorman, P., Sheer, D., Spurr, N. K., and Trowsdale, J. (1987). The human placental alkaline phosphatase gene and related sequences map to chromosome 2 band q37. *Ann. Hum. Genet.* **51:** 145–152.

Mattei, M. G., Colin, Y., Le Van Kim, C., Mattei, J. F., and Cartron, J. P. (1986). Localization of the gene for human erythrocyte glycophorin C to chromosome 2,q14–q21. *Hum. Genet.* **74:** 420–422.

Morris, S. W., Nelson, N., Valentine, M. B., Shapiro, D. N., Look, A. T., Kozlosky, C. J., Beckmann, M. P., and Cerretti, D. P. (1992). Assignment of the genes encoding human interleukin-8 receptor types 1 and 2 and an interleukin-8 receptor pseudogene to chromosome 2q35. *Genomics* **14:** 685–691.

Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J.-R., Burks, C., Sirotkin, K. M., and Goad, W. B. (1989). The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* **4:** 273–289.

Nelson, D. L. (1991). Interspersed repetitive sequence polymerase chain reaction (IRS PCR) for generation of human DNA fragments from complex sources. *Methods: a Comp. Methods Enzymol.* **2:** 60–74.

Nelson, D. L., Ledbetter, S. A., Corbo, L., Victoria, M. F., Ramirez-

Solis, R., Webster, T. D., Ledbetter, D. H., and Caskey, C. T. (1989). Alu polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA* **86:** 6686–6690.

NIH/CEPH Collaborative Mapping Group (1992). A comprehensive genetic linkage map of the human genome. *Science* **258:** 67–86.

Patracchini, P., Aiello, V., Palazzi, P., Calzolari, E., and Bernardi, F. (1989). Sublocalization of the human protein C gene on chromosome 2q13–q14. *Hum. Genet.* **81:** 191–192.

Peltomaki, P., Aaltonen, L. A., Sistonen, P., Pylkkanen, L., Mecklin, J.-P., Jarvinen, H., Green, J. S., Jass, J. R., Weber, J. L., Leach, F. S., Petersen, G. M., Hamilton, S. R., de la Chapelle, A., and Vogelstein, B. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260:** 810–812.

Poleev, A., Fickenscher, H., Mundlos, S., Winterpacht, A., Zabel, B., Fidler, A., Gruss, P., and Plachov, D. (1992). PAX8, a human paired box gene: Isolation and expression in developing thyroid, kidney and Wilms' tumors. *Development* **116:** 611–623.

Pringault, E., Robine, S., and Louvard, D. (1991). Structure of the human villin gene. *Proc. Natl. Acad. Sci. USA* **88:** 10811–10815.

Purdue, P. E., Lumb, M. J., Fox, M., Griffo, G., Hamon-Benais, C., Povey, S., and Danpure, C. J. (1991). Characterization and chromosomal mapping of a genomic clone encoding human alanine:glyoxylate aminotransferase. *Genomics* **10:** 34–42.

Rajcan-Separovic, E., and Sabour, M. P. (1993). Fluorescence in situ hybridization of bovine Alu-like sequences to bovine and ovine chromosomes. *Genome* **36:** 984–985.

Richards, A. J., Lloyd, J. C., Narcisi, P., Ward, P. N., Nicholls, A. C., De Paepe, A., and Pope, F. M. (1992). A 27-bp deletion from one allele of the type III collagen gene (COL3A1) in a large family with Ehlers–Danlos syndrome type IV. *Hum. Genet.* **88:** 325–330.

Rousseau-Merck, M. F., Simon-Chazottes, D., Arpin, M., Pringault, E., Louvard, D., Guenet, J. L., and Berger, R. (1988). Localization of the villin gene on human chromosome 2q35–q36 and on mouse chromosome 1. *Hum. Genet.* **78:** 130–133.

Rubin, C. M., Leeflang, E. P., Rinehart, F. P., and Schmid, C. W. (1993). Paucity of novel short interspersed repetitive element (SINE) families in human DNA and isolation of a novel MER repeat. *Genomics* **18:** 322–328.

Schmid, C. W., and Deininger, P. L. (1975). Sequence organization of the human genome. *Cell* **6:** 345–358.

Schmid, C. W., and Jelinek, W. R. (1982). The Alu family of dispersed repetitive sequences. *Science* **216:** 1065–1070.

Schroeder, W. T., Lopez, L. C., Harper, M. E., and Saunders, G. F. (1984). Localization of the human glucagon gene (GCG) to chromosome segment 2q36–37. *Cytogenet. Cell Genet.* **38:** 76–79.

Schurr, E., Skamene, E., Morgan, K., Chu, M. L., and Gros, P. (1990). Mapping of Col3a1 and Col6a3 to proximal murine chromosome 1 identifies conserved linkage of structural protein genes between murine chromosome 1 and human chromosome 2q. *Genomics* **8:** 477–486.

Singer, M. F. (1982). SINEs and LINEs: Highly repeated short and long interspersed sequences in mammalian genomes. *Cell* **28:** 433–434.

Singer, M. F., and Skowronski, J. (1985). Making sense of LINEs: Long interspersed repeat sequences in mammalian genomes. *Trends Biochem. Sci.* 119–122.

Smith, D. R. (1990). Genomic long-range restriction mapping. *Methods: Comp. Methods Enzymol.* **1:** 195–203.

Spurr, N. K., Cox, S., Bryant, S. P., Attwood, J., Robson, E. B., Shields, D. C., Steinbrueck, T., Jenkins, T., Murray, J. C., Kidd, K. K., Summar, M. L., Tsipouras, P., Retief, A. E., Kruse, T. A., Bale, A. E., Vergnaud, G., Weber, J. L., McBride, O. W., Donis-Keller, H., and White, R. L. (1992). The CEPH consortium linkage map of human chromosome 2. *Genomics* **14:** 1055–1063.

Spurr, N. K., Kruse, T., and Naylor, S. (1994). Report of the Third International Workshop on Human Chromosome 2 Mapping. *Cytogenet. Cell Genet.* **67:** 215–244.

Stapleton, P., Weith, A., Urbánek, P., Kozmik, Z., and Busslinger, M. (1993). Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9. *Nature Genet.* **3:** 292–298.

Tagle, D. A., and Collins, F. S. (1992). An optimized Alu-PCR primer pair for human-specific amplification of YACs and somatic cell hybrids. *Hum. Mol. Genet.* **1:** 121–122.

Tricoli, J. V., Bell, G. I., and Shows, T. B. (1984). The human glucagon gene is located on chromosome 2. *Diabetes* **33:** 200–202.

Van Es, H. H. G., Bout, A., Liu, J., Anderson, L., Duncan, A. M. V., Bosma, P., Oude Elferink, R., Jansen, P. L. M., Roy Chowdhury, J., and Schurr, E. (1993). Assignment of the human UDP glucuronosyltransferase gene (UGT1A1) to chromosome region 2q37. *Cytogenet. Cell Genet.* **63:** 114–116.

Vidal, S. M., Malo, D., Vogan, K., Skamene, E., and Gros, P. (1993). Natural resistance to infection with intracellular parasites: Isolation of a candidate for Bcg. *Cell* **73:** 469–485.

Viegas-Péquignot, E., Lin, L. Z., Dutrillaux, B., Apiou, F., and Paulin, D. (1989). Assignment of human desmin gene to band 2q35 by nonradioactive in situ hybridization. *Hum. Genet.* **83:** 33–36.

Webb, A. C., Collins, K. L., Auron, P. E., Eddy, R. L., Nakai, H., Byers, M. G., Haley, L. L., Henry, W. M., and Shows, T. B. (1986). Interleukin-1 gene (IL1) assigned to long arm of human chromosome 2. *Lymphokine Res.* **5:** 77–85.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., and Lathrop, M. (1992). A second-generation linkage map of the human genome. *Nature* **359:** 794–801.

Yoshiura, K. I., Kubota, T., Soejima, H., Tamura, T., Izumikawa, Y., Niikawa, N., and Jinno, Y. (1994). A comparison of GC content and the proportion of *AluI/KpnI*-repetitive sequences in a single dark- and light-band region from a human chromosome. *Genomics* **20:** 243–248.

Zucman, J., Delattre, O., Desmaze, C., Azambuja, C., Rouleau, G., De Jong, P., Aurias, A., and Thomas, G. (1992). Rapid isolation of cosmids from defined subregions by differential *Alu*-PCR hybridization on chromosome 22-specific library. *Genomics* **13:** 395–401.

The aim of the YAC mapping of human chromosome 2 in our laboratory is to construct a YAC contig overlapping the entire chromosome and as part of this project, my focus has been on the telomeric half of the long arm. To this end, we have taken a multi-step mapping approach including the generation of a collection of chromosome 2-specific YACs, YAC group analysis and ordering of YAC groups along the chromosome.

In Chapter 2, we described the use of the IRS-PCR approach to screen the CEPH Mark I YAC library. In total, close to 800 IRS-PCR probes were used to screen the YAC library and close to 2,000 chromosome 2-specific YACs were identified. This collection of chromosome 2-specific YACs was divided into 188 YAC groups, which formed the basis for the construction of a YAC contig map for human chromosome 2. The IRS-PCR approach was demonstrated by these experiments to be practical and efficient. Since most of the IRS-PCR probes were randomly derived from chromosome 2, a limitation of this approach was the difficulty of ordering YACs along the chromosome. Thus, in the subsequent step, the chromosomal localization of the YACs detected by the IRS-PCR probes needed to be determined by other means. Furthermore, we primarily used the CEPH Mark I library for the mapping of chromosome 2 while many other laboratories used the CEPH mega YAC library. The integration of mapping data which were derived from different libraries with our own results would provide a better map for chromosome 2.

In Chapter 3, we used the STS content mapping approach with a set of genetically or cytogenetically mapped STS markers to screen our collection of chromosome 2-specific YACs (chromosome 2 YAC sublibrary). These markers enabled us to order a subset of YACs along the chromosome as well as to integrate the cytogenetic, genetic maps with the YAC contig map. Furthermore, since most of the genetically mapped STS markers we used to screen the CEPH YAC Mark I library were also used by the WI/MIT Genome Center to screen the CEPH mega YAC library, these common markers were used as anchors to integrate the chromosome 2 YAC mapping data for both CEPH YAC libraries. Together, the experiments described in Chapter 2 and 3 allowed us to create an integrated genetic, cytogenetic and physical map of the telomeric region of human chromosome 2q.

CHAPTER 3:

Towards an integrated genetic, cytogenetic and physical map of human chromosome 2

Abstract

We previously generated a human chromosome 2 yeast artificial chromosome (YAC) sublibrary from the CEPH Mark I YAC library by the interspersed repetitive sequence (IRS)-PCR method. In the present study, we extended our mapping effort by screening the CEPH Mark I and mega YAC libraries with additional 78 and 112 IRS probes, respectively, and by sequence-tagged site (STS) screening of the chromosome 2 YAC sublibrary using 63 genetically mapped microsatellites and 24 cytogenetically mapped expressed sequence tags (ESTs). The resulting data was organized into 223 YAC groups and the STS content mapping allowed us to order 39 of the YAC groups along the long arm of chromosome 2. We found that 17 microsatellites were physically linked with 16 known genes within YAC groups. In addition, we integrated our data with the chromosome 2 YAC mapping data generated by the Whitehead Institute/MIT (WI/MIT) Genome Center on the basis of common microsatellites and mega YACs. The integrated dataset contained 240 YAC groups; 22 of these integrated groups contained both our YAC groups and WI/MIT Genome Center contig groups. Fourteen of the 22 integrated groups are located in chromosome region 2q and consist of 1,195 YAC clones corresponding to approximately one third of all chromosome 2 YAC clones analyzed. Thus, we have integrated information from genetic, cytogenetic and YAC-based physical maps and have built the backbone for a complete YAC contig for the long arm of chromosome 2.

Introduction

Human chromosome 2 is the second largest chromosome and corresponds to approximately 8% of the human genome. Four recent genetic maps for this chromosome have an average marker spacing of 2 to 5 cM, making it possible to map any polymorphic genetic marker to a 5 cM interval, except for a few regions such as the telomeric region of chromosome 2q (Buetow et al. 1994; Gyapay et al. 1994; Matise et al. 1994; Murray et al. 1994). The development of clone-based physical maps for chromosome 2 has lagged behind the advances in genetic map construction (Spurr et al. 1994). Recently, chromosome 2-specific bacterial artificial chromosome (BAC), cosmid, and bacteriophage P1-derived artificial chromosome (PAC) libraries were constructed (Wang et al. 1994; Gingrich et al. 1995). These resources will be useful for high-resolution physical mapping of chromosome 2 once an ordered array of overlapping YAC clones spanning chromosome 2 has been achieved.

We initiated YAC contig assembly for human chromosome 2 based on a two-step mapping process. We used the IRS-PCR approach which relies on the detection of YAC clones by hybridization of individual IRS-PCR products to IRS-PCR products of pooled YAC DNA (Liu et al. 1995a). The IRS-PCR approach is a cost-effective, rapid and efficient method for YAC contig assembly, since large numbers of IRS probes can be hybridized to IRS-PCR products of an entire YAC library (spotted on 7 x 11 cm filters), allowing simultaneous screening of multiple probes and libraries (Chumakov et al. 1992; Hunter et al. 1994; Liu et al. 1995a; Stanton et al. unpublished). We previously used close to 796 IRS probes to identify a minimum of 1,610 chromosome 2-specific YAC clones, which constitute our chromosome 2-specific YAC sublibrary. These chromosome 2-specific YAC clones were arranged into 188 groups, of which only 20 could be assigned to specific cytogenetic regions of chromosome 2 (Liu et al. 1995a).

In the present study, we determined the chromosomal localization of YAC groups along chromosome 2. We conducted STS content mapping of the chromosome 2-specific YAC sublibrary, rather then the entire YAC library, using a set of genetically mapped microsatellites and cytogenetically mapped ESTs, thereby reducing the work load.

Moreover, relatively few STSs were needed to order YAC groups since detection of a few YAC clones in each group was sufficient to localize the group on the chromosome. Finally, STS content mapping allowed us to integrate our mapping data with those available on the Whitehead Institute/MIT Genome Center www-sites, and with other genetic and cytogenetic maps into a YAC map of chromosome 2.

Materials and methods

Total genomic YAC libraries

Two total genomic YAC libraries were used in this study. The first one is the extended CEPH Mark I YAC library consisting of plates numbered 1 to 552. (Albertsen et al. 1990; Dausset et al. 1992), plus 24 plates from the original Washington University YAC library (Burke et al. 1987; Brownstein et al. 1989). The second library is the mega YAC library consisting of plates numbered 709 to 972, excluding plates 861 to 868, and encompasses 24,576 YAC clones with an average insert size of 1 Mb (Dausset et al. 1992).

Determination of YAC sizes

Yeast cultures and agarose blocks for YAC size determination by pulse field gel electrophoresis (PFGE) were prepared following a previously described protocol (Chiu et al. 1994). Yeast chromosomes were separated on 1% agarose gels using a Biorad CHEF Mapper apparatus. Electrophoresis was carried out in 0.5x TBE buffer (1x TBE is: 0.1 M Tris, 0.1 M boric acid, 0.2 mM Na_2EDTA) at 14°C with a linear switch ramp from 60 sec to 100 sec for 24 hr. DNA was then transferred to Hybond-N nylon membranes (Amersham). The probe used for hybridization was a gel purified 2.67 kb fragment of *Pvu*II and *BamH*I digested pBR322, which corresponds to the left arm of the pYAC2 vector (Burke et al. 1987). Labeling of probes was done using α-^{32}P-dATP (>3000 Ci/mmole) and the random hexanucleotide labeling technique (Feinberg and Vogelstein 1983). Hybridizations were carried out according to the manufacturer's instructions (Amersham). YAC size was estimated by comparison of the mobility of the band of YACs with those of yeast chromosomes of known size from *Saccharomyces cerevisiae* strain AB1380 and with a 50 kb Lambda cocatemer ladder (New England Biolabs, Missisauga, Ontario).

IRS-PCR hybridization membranes

The generation of IRS-PCR hybridization membranes for the CEPH Mark I YAC library was previously described by Liu et al. (1995a). A similar procedure was used to generate IRS-PCR hybridization membranes for the mega YAC library except that a different pooling scheme was used. The mega YAC library, which consists of 256 plates, was assembled in 32 vertical stacks of 8 plates. From each stack of 8 plates, 28 pools were collected: 8 plate pools (1-8), 8 row pools (A-H), and 12 column pools (1-12). Thus a total of 28 x 32 = 896 pools were collected. DNA was extracted from each pool at the WI/MIT Genome Center and the *Alu* PCR products were produced as described by Liu et al. (1995a). The set of 896 pools of *Alu* PCR products was spotted in duplicate on a 7 x 11 cm membrane. Each YAC address is composed of three components corresponding to the plate, row and column pool coordinates.

IRS-PCR probes for screening YAC libraries

A set of 78 new IRS-PCR probes were generated from somatic cell hybrids containing only portions of human chromosome 2 and used to screen the CEPH Mark I YAC library. In addition, a subset of 112 IRS-PCR probes from our collection of 874 probes which had been used to screen the Mark I YAC library was also used to screen the CEPH mega YAC library. The protocols for the generation of IRS-PCR probes from somatic cell hybrids, IRS-PCR probe labeling, and for IRS-PCR probe hybridization to membranes were described previously (Liu et al. 1995a).

Chromosome 2 YAC sublibrary

Previously, we identified 4,572 YAC clones from the CEPH Mark I YAC library using the IRS-PCR screening method with chromosome 2-specific probes (Liu et al. 1995a). We individually picked these 4,572 YAC clones and stored them in 48 ninety-six well plates as a chromosome 2 enriched YAC sublibrary. Of these 4,572 YAC clones, a minimum of 1610 were chromosome 2-specific; and a subset of them were part of ambiguous groups and likely were not chromosome 2-specific. In addition, we obtained 18 YAC clones from the ICRF YAC library which were detected by hybridization

screening with cDNA probes for *FN1, CRYG, TNP1, COL3A1, COL5A2*, and *UGT1A1*. We included these YAC clones in our chromosome 2 YAC sublibrary, which we then used as the template for further screening with STSs.

Selection of STSs in chromosome region 2q

We selected microsatellites from a recently published genetic map (Gyapay et al. 1994) to use them as anchors to order YAC groups and to integrate our data with the data from the WI/MIT Genome Center. We selected two proximal markers, *D2S356* and *D2S124* and then starting from the map position where *D2S300* is located and extending to the telomeric end of chromosome region 2q, we selected 65 microsatellites, choosing at least one microsatellite for each position resolved on the genetic map (see Fig. 1). We also selected two additional microsatellites, *D2S102* and *D2S211* (Spurr et al. 1994; Barber et al. 1993), which were used in a genetic analysis of susceptibility to tuberculosis (Boothroyd 1994). Thirty-four of these 67 microsatellites were described previously (Liu et al. 1995a); the other 33 were *D2S72, D2S133, D2S161, D2S280, D2S307, D2S309, D2S313, D2S315, D2S322, D2S325, D2S336, D2S338, D2S342, D2S344, D2S345, D2S346, D2S348, D2S350, D2S351, D2S353, D2S355, D2S356, D2S360, D2S360, D2S362, D2S364, D2S369, D2S377, D2S384, D2S389, D2S395, D2S396*, and *D2S401*. Primers for all 67 microsatellites were purchased from Research Genetics (Huntsville, AL).

In addition, we selected primer pairs for 24 genes (*IL1B, PROC, GYPC, ERCC3, GCG, COL3A1, COL5A2, CRYGA, CPS1, MYL1, CD28, FN1, XRCC5, TNP1, NRAMP1, IL8R, VIL1, INHA, DES, PAX3, ALPP, ALPI, UGT1A1*, and *AGXT*) located in chromosome region 2q13-q37. The source and the sequence of these primers, except for *XRCC5*, were described previously (Liu et al. 1995a). Primers for an STS derived from the 3' UTR of the *XRCC5* gene (Taccioli et al. 1994) were kindly provided by Dr. P. Jeggo (MRC Cell Mutation Unit, Brighton, UK).

STS content mapping of the chromosome 2 YAC sublibrary

The procedure for isolation of YAC DNA and the PCR conditions for STS screening of the chromosome 2 YAC sublibrary were described previously (Liu et al. 1995a). Yeast genomic DNA isolated from *Saccharomyces cerevisiae* strain AB1380 and human genomic DNA were used as negative and positive controls, respectively, in amplification reactions. Positive clones were identified by examining ethidium bromide stained agarose gels for the presence of amplification products of expected size.

Montreal General Hospital (MGH) chromosome 2 YAC database

Previously, we developed a database for storing the results of IRS-PCR screening of the CEPH Mark I YAC library and STS screening of the chromosome 2 YAC sublibrary (Liu et al. 1995a). We expanded the database by adding new mapping information from IRS-PCR screening of both the Mark I and mega YAC libraries, and from additional STS content mapping of the chromosome 2 YAC sublibrary.

The WI/MIT Genome Center chromosome 2 YAC mapping data

Human chromosome 2 YAC mapping data from the WI/MIT Genome Center was downloaded from the World Wide Web (http://www-genome.wi.mit.edu) (Release 6, March 1995). We retrieved all the chromosome 2 contigs designated as WC from this dataset. In addition, we also retrieved YAC data corresponding to those microsatellites which detected Mark I YAC clones in our chromosome 2 YAC sublibrary, but were not contained in any WC-group on chromosome 2.

YAC group analysis

We previously defined a set of overlapping YAC clones as a YAC group (YG) and developed a program in FoxPro (Microsoft) to assemble these YAC groups (Liu et al. 1995a). For the previous YAC group analysis, we used a dataset consisting of unambiguously detected YAC clones and others which were ambiguously detected initially but later confirmed by individual YAC hybridization or PCR experiments. By matching ambiguous with unambiguous addresses, we were also able to resolve a subset of

ambiguously detected Mark I and mega YAC clones and included them in the YAC group analysis presented here. The principle for this disambiguation process was the following: if a YAC clone X was detected as part of an ambiguous group by probe A, and clone X was the only member of this ambiguous group to be unambiguously detected by another probe B, then if probes A and B belonged to the same or adjacent previously defined YAC groups, the probe-address combination A-X was considered disambiguated and included in the YAC group analysis of the present study. Finally, we eliminated from the YAC group analysis probes that detected more than 15 YAC clones, under the assumption that such probes most likely contained repeat sequences. The resulting MGH dataset was subjected to YAC group analysis.

Since our program for the YAC group analysis uses marker/YAC pairs as input, it was possible to combine the WI/MIT Genome Center data with our data for YAC group analysis. Among the chromosome 2 WC-groups, we eliminated those YAC clones which appeared in more than two WC-groups, under the assumption that these clones are intrachromosomal chimeric YAC clones. The integrated MGH-WI/MIT Genome Center dataset was also subjected to YAC group analysis (see Fig. 2).

The information for our YAC groups and the integrated MGH-WI/MIT YAC groups is available via anonymous ftp. To obtain the data, *ftp* to ftp.mcgill.ca and *cd* to the directory pub/McGill-Contrib/Chromosome2.

Results

Scheme for generating and ordering groups of overlapping YAC clones and for integrating our data with the WI/MIT Genome Center data

The scheme we used to generate and order groups of overlapping YAC clones, and to integrate our data with the WI/MIT Genome Center data, is summarized in Fig. 2. Previously, we used the IRS-PCR screening approach to generate a Mark I YAC sublibrary enriched for chromosome region 2q (Liu et al. 1995a). In the present study, we continued our YAC mapping effort by screening the CEPH Mark I YAC library with 78 addition randomly derived IRS probes and by STS content mapping of the chromosome 2 Mark I YAC sublibrary with 30 additional STSs. In total, we used 67 genetically mapped microsatellites and 24 cytogenetically mapped ESTs to screen the chromosome 2 YAC sublibrary. These STSs were used as internal controls to verify the validity of STS-anchored YAC groups and to order them along the chromosome. Although it is possible that in some cases unrelated YAC groups might be falsely connected through an intrachromosomal chimeric YAC clone, such false connections could be detected easily by means of internal controls such as STSs with known chromosomal localization. The final STS-anchored YAC groups which resulted from analysis of our data were therefore consistent with the map positions of the STSs and could be ordered along the chromosome 2 (Fig. 2).

The YAC library which we used to derive the chromosome 2 sublibrary was the CEPH Mark I YAC library, while the WI/MIT Genome Center data are based on the CEPH mega YAC library. Thus, integrating the chromosome 2 mapping data from both libraries allowed us to create a more complete chromosome 2 YAC contig map. The resulting groups are called integrated groups (IG) (Fig. 2).

STS content mapping of the chromosome 2 YAC sublibrary

Of the 67 microsatellites which were selected, 4 did not amplify control genomic DNA and were not be used to screen the sublibrary. Among the 63 markers which were used, 53 identified at least one YAC in the sublibrary. In addition, all of the 24 ESTs

selected from chromosome region 2q13-q37, except for *AGXT*, identified YAC clones in the sublibrary. Thus, of the 87 STSs (63 microsatellites and 24 ESTs) used for screening the chromosome 2 YAC sublibrary, 87% (53 microsatellites and 23 ESTs) identified at least one YAC in the sublibrary. Table 1 lists the YAC clones identified by each STS. The number of YAC clones detected per STS for the 76 STSs ranged from 1 to 14 with an average of 4, which is consistent with our previous report (Liu et al. 1995a). In total, 245 STS-positive YAC clones were identified. The size for 130 of these YAC clones is also shown in Table 1.

Approximately half of the IRS-PCR probes used for the generation of the chromosome 2 YAC sublibrary were derived from somatic cell hybrid 014-1 which contained the telomeric half of chromosome 2q excluding the tip of the telomere (see Fig. 1) while the majority of the other half were derived from hybrid GM-10826B containing entire chromosome 2 (Liu et al. 1995a). The higher the density of IRS-PCR probes for a given genomic region, the higher the coverage of YAC clones one would expect for this region in the sublibrary. Therefore, we investigated whether success of the STS screening was associated with the genomic region contained in the hybrid 014-1. We determined the boundaries of this hybrid by PCR analysis for the presence or absence of amplification products of specific primers from chromosome region 2q to be at the map region of *D2S72* and *D2S307*, and in the region between *D2S338* and *D2S395* (see Fig. 1). Of the 10 microsatellites which failed to detect YAC clones in the sublibrary, 7 of them were not contained in hybrid 014-1 (Fig. 1). *D2S351* is inside the map boundaries of hybrid 014-1 but was not detected in hybrid 014-1. The likely explanation is that there is a small internal deletion around *D2S351* in hybrid 014-1, since microsatellites adjacent to *D2S351* were found in this hybrid. In addition, there are other rearrangements in hybrid 014-1 since we observed that genomic sequences from the region around *D2S138*, which is proximal to the expected map boundaries of hybrid 014-1, were contained in this hybrid.

YAC group analysis

The updated dataset from our chromosome 2 YAC mapping database, that includes new data from IRS-PCR screening of both Mark I and mega YAC libraries as

well as from new STS screening of the chromosome 2 YAC sublibrary, consisted of 1,733 Mark I and 301 chromosome 2-specific mega YAC clones which were detected by 752 IRS-PCR probes and 76 STSs. After the initial analysis we identified several YAC groups which contained false connections based on the map positions of the STSs contained in these groups. We re-analyzed these YAC groups via structured database queries and identified single YAC clones which were causing the false connection between unrelated YAC groups. After excluding these chimeric YAC clones, we did another YAC group analysis and identified 220 YAC groups which are now consistent with the map positions of their included STSs. Three additional YAC groups consisting only of chimeric YAC clones detected by *D2S153*, *D2S161* and *D2S172*, which were not included in the YAC group analysis, were named as YG221', YG222' and YG223'. Thus, in total, our dataset consisted of 223 YAC groups.

Chromosomal localization of YAC groups and physical linkage of microsatellites with ESTs

Among the 223 YAC groups which were identified in the present study, 44 were anchored by at least one STS. We denoted these 44 YAC groups as STS-anchored YAC groups. Thus, the 76 STSs were associated with 44 STS-anchored YAC groups comprised of 837 distinct YAC clones. Table 2 lists the 44 STS-anchored YAC groups, their associated STSs, and the number of YAC clones and the number of markers (IRS-PCR probes and STSs) contained in each group. Since all the microsatellites except *D2S102* and *D2S211* were selected from the Gyapay map, the YAC groups could be ordered along chromosome region 2q based on the genetic map position of the markers (see Fig. 3). Of the 44 STS-anchored YAC groups, 39 were anchored by at least one microsatellite from the Gyapay map and are shown in the center of Fig. 3. *D2S211* and *D2S102* were physically linked to groups that contained microsatellites from the Gyapay map.

Among the 44 STS-anchored YAC groups, 16 contained two or more STS markers (first 16 groups in Table 2) and can be further subdivided into three categories based on the type of STS contained in the group: 9 YAC groups containing both ESTs

and microsatellites; 5 YAC groups containing only microsatellites; and 2 YAC groups containing only ESTs. The 9 STSs-anchored groups that contain both ESTs and microsatellites include 17 microsatellites that can be ordered along the genetic map of chromosome region 2q and 16 ESTs that map to distinct cytogenetic bands. These 9 STSs-anchored groups represent the integration of genetic, cytogenetic and physical map information in chromosome region 2q, as shown in Fig. 3.

Integration of our data with the WI/MIT Genome Center data

In order to gain more information for chromosome 2 YAC mapping, we integrated our data with the WI/MIT Genome Center chromosome 2 mapping data. The integration was based on common microsatellites and/or common mega YAC clones. Microsatellites that were used for screening the chromosome 2 Mark I YAC sublibrary in our laboratory and for screening the mega YAC library by the WI/MIT Genome Center allowed us to integrate our STS-anchored YAC groups with the WI/MIT Genome Center data. In addition, we used a subset of IRS-PCR probes to screen the mega YAC library and detected a set of mega YAC clones, some of which were also detected by the WI/MIT Genome Center. Therefore, these common mega YAC clones provided a second way of integrating our data with the WI/MIT Genome Center data.

In the WI/MIT Genome Center data, there were 67 chromosome 2 WC-groups which consisted of 1,574 YAC clones and 495 STSs (including genetically mapped microsatellites and other STSs). Of the 53 microsatellites used in our study, 48 were also used by the WI/MIT Genome Center for screening the mega YAC library. The exceptions were *D2S72*, *D2S104*, *D2S211*, *D2S322* and *D2S371*. Among the 48 common microsatellites, 38 were found in the 67 chromosome 2 WC-groups. We therefore retrieved the WI/MIT Genome Center data corresponding to the 10 microsatellites not in WC-groups. Thus, the chromosome 2 WI/MIT Genome Center data which were integrated with our data consisted of 67 WC-groups plus 10 microsatellites and their associated mega YAC clones. The integrated dataset consisted of 3,329 YAC clones detected by 752 IRS-PCR probes and 533 STSs.

After the initial analysis of the combined MGH and WI/MIT data, we identified and eliminated chimeric YAC clones. A final analysis produced 240 groups which we called integrated groups (IG). Twenty-two integrated groups included both our YAC groups and WC-groups; 14 of these 22 groups were located in chromosome region 2q, which was the major focus of our YAC mapping (see Fig. 4). Of the 14 integrated groups, the minimum YAC coverage was determined for 6 integrated groups based on the boundary markers from the Gyapay map: 4 cM (IG12), 4 cM (IG20), 13 cM (IG5), 8 cM (IG42), 5 cM (IG3), and 2 cM (IG25). Integrating the information from the WI/MIT Genome Center has allowed us to increase the continuous YAC coverage, for example, IG5 contains 11 of our YAC groups and 3 WI/MIT Genome Center WC-groups. IG5 consists of 307 YAC clones and 130 markers, and overlaps at least 13 cM of the genomic region from *D2S155* to *D2S295*. Together, the 6 integrated groups cover a minimum of 36 cM, a genomic region which corresponds to approximately one fourth of the entire long arm of chromosome 2. The actual YAC coverage for chromosome region 2q is significantly higher since there were another 8 integrated groups (IG148, IG14, IG151, IG7, IG184, IG31, IG11 and IG38) for which minimal genomic coverage could not be accurately determined because there was only one genetic marker from the Gyapay map contained in each group. Several of these groups fall within the gaps between the 6 other integrated groups. In total, the 14 integrated groups consist of 1,195 YAC clones (approximately 1/3 of the YAC clones in the integrated dataset) which were detected by 529 markers.

Discussion

In the present study, we continued our effort in the assembly of a complete YAC contig for human chromosome 2. We subjected the YAC clones contained in the sublibrary to STS content mapping using primer pairs derived from 24 expressed genes and 63 microsatellites which localize to the long arm of chromosome 2. We were able to order 44 YAC groups along the chromosome and to integrate 16 expressed genes into the genetic map of chromosome 2 on the basis of our YAC group analysis. The integration of expressed genes with microsatellite-based genetic maps will be useful for linkage analysis and for cloning disease genes. Indeed, it has been suggested that the positional candidate approach which is based on the integration of transcript and genetic maps will eventually replace the positional cloning approach for cloning of disease genes (Collins 1995). In addition, there are a number of other applications for such integrated maps, for example, once a candidate disease gene is identified in an animal model, the human homologue could be tested for linkage analysis with the human disease phenotype. Frequently, no polymorphic variants with high degree of heterozygosity are found in the human gene and it would be desirable to identify highly informative microsatellites located in the vicinity of the disease gene which can be used for linkage studies. This problem can be illustrated by the genetic analysis of tuberculosis susceptibility. The human *NRAMP1* gene has been suggested by studies in inbred strains of mice as a candidate tuberculosis susceptibility gene (Vidal et al. 1993; Cellier et al. 1994; Liu et al. 1995b). To date, a number of polymorphisms have been identified in the *NRAMP1* gene (White et al. 1994; Liu et al. 1995b; Blackwell et al. 1995; Buu et al. 1995); however, most of these polymorphisms are di-allelic and have low degree of heterozygosity, and genetic studies are thus restricted by the lack of highly informative genetic markers. The establishment of physical linkage within a small YAC contig between two microsatellites, *D2S104* and *D2S173*, and *NRAMP1* provided additional highly informative markers for further genetic analysis of susceptibility to tuberculosis (Liu et al. 1995b).

The integration of our data with the WI/MIT Genome Center data greatly improves the YAC contig information available for human chromosome 2. Whereas our

data was created mainly using the CEPH Mark I YAC library, the WI/MIT Genome Center data was generated from the CEPH mega YAC library. The integration of the two datasets thus increases the depth of YAC coverage of this chromosome. The two datasets are complementary to each other to some extent. For example, among the 63 microsatellites tested in our study, 10 failed to detect YAC clones in the sublibrary. But, 5 of these 10 STSs detected mega YAC clones in the WI/MIT Genome Center experiments. On the other hand, 5 markers (*D2S72*, *D2S104*, *D2S211*, *D2S322* and *D2S371*) were not detected in the WI/MIT Genome Center screen but were detected in Mark I YAC clones in our sublibrary. Furthermore, in the WI/MIT Genome Center dataset, *D2S173* and *D2S163* were not part of any of the 67 WC-groups. In contrast, these two markers were part of two extended YAC groups in our dataset (YG55 and YG92), with 52 YAC clones and 39 YAC clones, respectively. Finally, the integration between 2 datasets results in larger YAC groups. For example, in chromosome region 2q (Fig. 4), in 8 instances (IG14, IG12, IG20, IG7, IG5, IG42, IG31) our adjacent YAC groups were connected via the WI/MIT Genome Center data, while in 3 instances (IG20, IG5 and IG3), the WI/MIT Genome Center WC-groups were connected via our data. Thus, the integration between these two datasets has created 14 large YAC groups in human chromosome region 2q.

The localization of 44 STS-anchored YAC groups and the creation of 14 integrated YAC groups comprising 1,195 YAC clones and 529 probes on the telomeric half of the long arm of chromosome 2 represents significant progress toward the construction of a YAC-based physical map of chromosome 2. To complete the physical map for this region of the genome, it is necessary to localize YAC groups which are not already part of the 14 integrated YAC groups and to initiate chromosome walking from these groups. To date, we have identified 55 YAC groups which contain probes derived from hybrid 014-1 and are not part of the integrated YAC groups. We intend to select a minimal set of YAC clones from these groups and derive IRS probes to screen both the Mark I and mega YAC libraries. This effort will likely extend the present YAC groups and result in additional contig closures. Currently, we are also exploring the possibility of using "inner product mapping" to localize the YAC clones of the chromosome 2 sublibrary (Perlin and Chakravarti 1993; Perlin et al. 1995). This information would aid us

in selecting IRS probes that are located in the contig gaps. Together, these efforts will help us to obtain a complete set of overlapping clones for the long arm of human chromosome 2.

Acknowledgments

We thank Li-Hong Liu for laboratory assistance and Joyce Crumley for the initial development of the database. This work was supported by grants from the Canadian Genome Analysis and Technology Program and the Canadian Genetic Diseases Network (Federal NCE program). Work in the laboratory of D.H. was supported by a grant from the Human Genome Office (HG00299). J.L. is supported by a predoctoral studentship from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche and the McGill Faculty of Medicine Internal Scholarship Award. P.G. is supported by an E.W.R. Steacie Memorial Fellowship from the Natural Sciences and Engineering Research Council of Canada and is an International Research Scholar of the Howard Hughes Medical Institute, and E.S. is a scholar of the Medical Research Council of Canada.

References

Albertsen, H.M., Abderrahim, H., Cann, H.M., Dausset, J., Le Paslier, D., Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. Proc. Natl. Acad. Sci. USA 87, 4256-4260.

Barber, T.D., Morell, R., Johnson, D.H., Asher, J.H.,Jr., Friedman, T.B. (1993). A highly informative dinucleotide repeat polymorphism at the D2S211 locus linked to ALPP, FN1 and TNP1. Hum. Mol. Genet. 2, 88.

Blackwell, J.M., Barton, C.H., White, J.K., Searle, S., Baker, A-M., Williams, H., Shaw, M-A. (1995). Genomic organization and sequence of the human NRAMP gene: identification and mapping of a promoter region polymorphism. Molecular Medicine 1, 194-205.

Boothroyd, L.J. (1994). Genetic susceptibility to tuberculosis. MS thesis, McGill University, Montreal.

Brownstein, B.H., Silverman, G.A., Little, R.D., Burke, D.T., Korsmeyer, S.J., Schlessinger, D., Olson, M.V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. Science 244, 1348-1351.

Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Duyk, G.M., Sheffield, V.C., Wang, Z., Murray, J.C. (1994). Integrated human genome-wide maps constructed using the CEPH referemce panel. Nature Genet. 6, 391-393.

Burke, D.T., Carle, G.F., Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236, 806-812.

Buu, N.T., Cellier, M., Gros, P., Schurr, E. (1995). Identification of a highly polymorphic length variant in the 3'UTR of NRAMP1. Immunogenetics (in press).

Cellier, M., Govoni, G., Vidal, S., Kwan, T., Groulx, N., Liu, J., Sanchez, F., Skamene, E., Schurr, E., Gros, P. (1994). Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression. J. Exp. Med. 180, 1741-1752.

Chiu, I.M., Gilmore, E.C., Liu, Y., Payson, R.A. (1994). Construction of a yeast artificial chromosome contig encompassing the human acidic fibroblast growth factor (FGF1) gene: toward the cloning of the ANLL/MDS tumor-suppressor gene. Genomics 19, 552-560.

Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billaut, A., Guasconi, G., Gervy, P., LeGall, I., Soularue, P., Grinas, L., Bougueleret, L., Bellanné-Chantelot, C., Lacroix, B., Barillot, E., Gesnouin, P., Pook, S., Vaysseix, G., Frelat, G., Schmitz, A., Sambucy, J.-L., Bosch, A., Estivill, X., Weissenbach, J., Vignal, A., Riethman, H., Cox, D., Patterson, D., Gardiner, K., Hattori, M., Sakaki, Y., Ichikawa, H., Ohki, M., Le Paslier, D., Heilig, R., Antonarakis, S., Cohen, D. (1992). Continuum of overlapping clones spanning the entire human chromosome 21q. Nature 359, 380-387.

Collins, F.S. (1995). Positional cloning moves from perditional to traditional. Nature Genet. 9, 347-350.

Dausset, J., Ougen, P., Abderrahim, H., Billault, A., Asmbucy, J.-L., Cohen, D., Le Paslier, D. (1992). The CEPH YAC library. Behring Inst. Mitt. 91, 13-20.

Feinberg, A.P., Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132(1), 6-13.

Gingrich, J., Garnes, J., Boehrer, D., Johnson, W., Wong, B., Bergmann, A., Eveleth, J., Langlois, R., Lamerdin, J., Lee, D., Carrano, A. (1995). Generation and chracterization of chromosome 2-specific cosmid, fosmid and PAC clone libraries and new chromosome 2 STSs derived from micro-dissection clones. Genome mapping and sequencing, Cold Spring Harbor laboratory, Cold Spring Harbor, N.Y.

Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., Weissenbach, J. (1994). The 1993-1994 Généthon human genetic linkage map . Nature Genet. 7 (Special), 246-339.

Hunter, K.W., Ontiveros, S.D., Watson, M.L., Stanton, V.P., Jr., Gutierrez, P., Bhat, D., Rochelle, J., Graw, S., Ton, C., Schalling, M., Aburatani, H., Brown, S.D.M., Seldin, M.F., Housman, D.E. (1994). Rapid and efficient construction of yeast artificial chromosome contigs in the mouse genome with interspersed repetitive sequence PCR (IRS-PCR): generation of a 5-cM, > 5 megabase contig on mouse chromosome 1. Mamm. Genome 5, 597-607.

Liu, J., Stanton, V.P.,Jr., Fujiwara, T.M., Wang, J.W., Rezonzew, R., Crumley, M.J., Morgan, K., Gros, P., Housman, D., Schurr, E. (1995a). Large-scale cloning of human chromosome 2-specific yeast artificial chromosome (YACs) using an interspersed repetitive sequences (IRS)-PCR approach. Genomics 26, 178-191.

Liu, J., Fujiwara, T.M., Buu, N.T., Sanchez, F.O., Cellier, M., Paradis, A.J., Frappier, D., Skamene, E., Gros, P., Morgan, K., Schurr, E. (1995b). Identification of polymorphisms and sequence variants in the human homologue of the mouse natural resistance-associated macrophage protein gene. Am. J. Hum. Genet. 56, 845-853.

Matise, T.C., Perlin, M., Chakravarti, A. (1994). Automated construction of genetic linkage maps using an expect system (MultiMap): a human genome linkage map. Nature Genet. 6, 384-390.

Murray, J.C., Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Manion, F., Quillen, J., Sheffield, V.C., Sunden, S., Duyk, G.M., Weissenbach, J., Gyapay, G., Dib, C., Morrissette, J., Lathrop, G.M., Vignal, A., White, R., Matsunami, N., Gerken, S., Melis, R., Albertsen, H., Plaetke, R., Odelberg, S., Ward, D., Dausset, J., Cohen, D., Cann, H. (1994). A comprehensive human linkage map with centimorgan density. Science 265, 2049-2054.

Perlin, M., Chakravarti, A. (1993). Efficient construction of high-resolution physical maps from yeast artificial chromosomes using radiation hybrids: inner product mapping. Genomics 18, 283-289.

Perlin, M.W., Duggan, D.J., Davis, K., Farr, J.E., Findler, R.B., Higgins, M.J., Nowak, N.J., Evans, G.A., Qin, S., Zhang, J., Shows, T.B., James, M.R., Richard, C.W.,III. (1995). Rapid construction of integrated maps using inner product mapping: YAC coverage of human chromosome 11. Genomics 28, 315-327.

Spurr, N.K., Barton, H., Bashir, R., Bryson, G.M., Bushby, K., Cox, S., Gringrich, J.C., Hentati, A., Hildebrandt, F., Kao, F-T., Kruse, T., Lai, E., Liu, J., Menke, M., Naylor, S., Nicklin, M., Read, A., Reiner, O., Rocchi, M., Summer, M. (1994). Report of the Third International Workshop on Human Chromosome 2 Mapping. Cytogenet. Cell Genet. 67, 215-244.

Taccioli, G.E., Gottlieb, T.M., Blunt, T., Priestley, A., Demengeot, J., Mizuta, R., Lehmann, A.R., Alt, F.W., Jackson, S.P., Jeggo, P.A. (1994). Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. Science 265, 1442-1445.

Vidal, S.M., Malo, D., Vogan, K., Skamene, E., Gros, P. (1993). Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. Cell 73, 469-485.

Wang, M., Chen, X-N., Shouse, S., Manson, J., Wu, Q., Li, R., Wrestler, J., Noya, D., Sun, Z-G., Korenberg, J., Lai, E. (1994). Construction and characterization of a human chromosome 2-specific BAC library. Genomics 24, 527-534.

White, J.K., Shaw, M-A., Barton, C.H., Cerretti, D.P., Williams, H., Mock, B.A., Carter, N.P., Blackwell, J.M. (1994). Genetic and physical mapping of 2q35 in the region of NRAMP and IL8R genes: identification of a polymorphic repeat in exon 2 of NRAMP . Genomics 24, 295-302.

FIGURES AND TABLES

Figure 1:

Microsatellites in human chromosome region 2q. The genetic map was adapted from the map produced by Gyapay et al. (1994). We selected all the markers except those which are underlined. In total, 65 markers were selected. The asterisk (*) indicates four microsatellites that failed to amplify control human genomic DNA and were not used for STS screening. One of the two primary sources for IRS probes was hybrid 014-1. The chromosomal region contained in hybrid 014-1 is estimated to be 70 Mbp and is indicated by the vertical line at the far left of the figure. Ten microsatellites failed to identify any YAC clones in the sublibrary: three enclosed in [ ] were in the region contained in hybrid 014-1; the seven others are enclosed in < >.

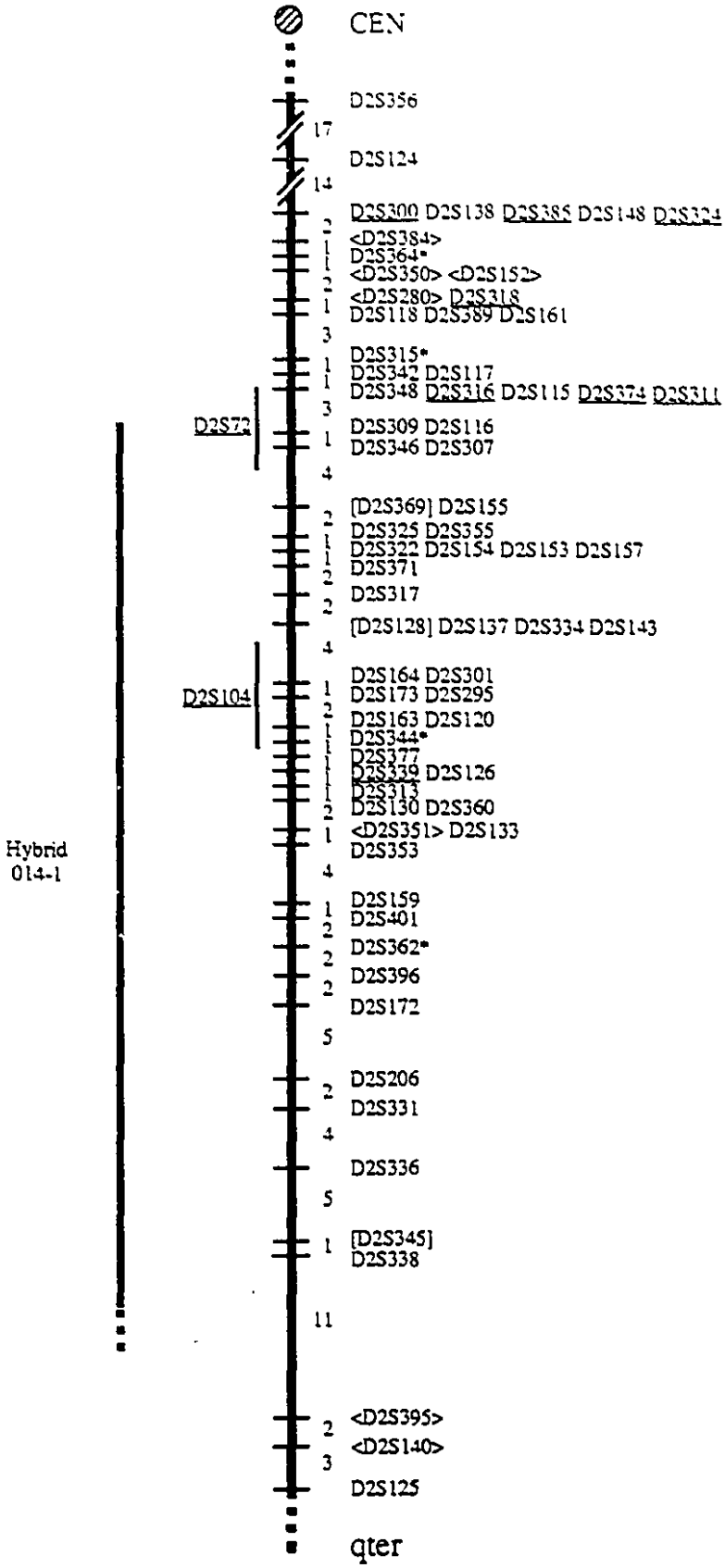

CEN
D2S356
17
D2S124
14
D2S300 D2S138 D2S385 D2S148 D2S324
2
1
1
2
1
3
1
1
3
1
4
2
1
1
2
2
4
1
2
1
1
1
1
2
1
4
1
2
2
2
5
2
4
5
1
11
2
3
<D2S384>
D2S364*
<D2S350> <D2S152>
<D2S280> D2S318
D2S118 D2S389 D2S161
D2S315*
D2S342 D2S117
D2S348 D2S316 D2S115 D2S374 D2S311
D2S72
D2S309 D2S116
D2S346 D2S307
[D2S369] D2S155
D2S325 D2S355
D2S322 D2S154 D2S153 D2S157
D2S371
D2S317
[D2S128] D2S137 D2S334 D2S143
D2S164 D2S301
D2S104
D2S173 D2S295
D2S163 D2S120
D2S344*
D2S377
D2S339 D2S126
D2S313
D2S130 D2S360
<D2S351> D2S133
D2S353
D2S159
D2S401
D2S362*
D2S396
D2S172
D2S206
D2S331
D2S336
[D2S345]
D2S338
<D2S395>
<D2S140>
D2S125
qter
Hybrid
014-1

Figure 2:

Flow diagram for generating and ordering groups of overlapping YAC clones and for the integration of our data with the WI/MIT Genome Center data (for details see text).

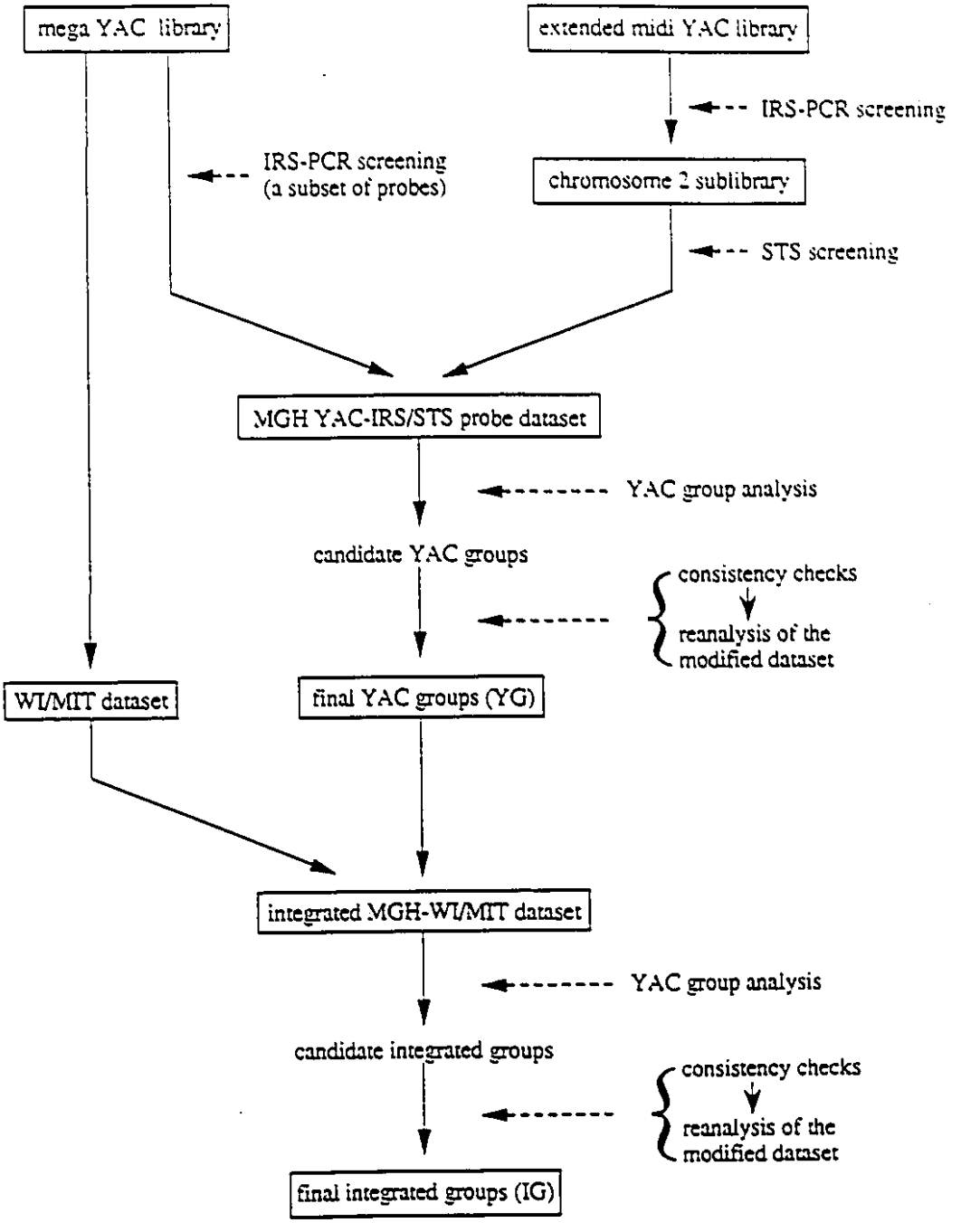

mega YAC library
extended midi YAC library
IRS-PCR screening
IRS-PCR screening
(a subset of probes)
chromosome 2 sublibrary
STS screening
MGH YAC-IRS/STS probe dataset
YAC group analysis
candidate YAC groups
consistency checks
reanalysis of the
modified dataset
WI/MIT dataset
final YAC groups (YG)
integrated MGH-WI/MIT dataset
YAC group analysis
candidate integrated groups
consistency checks
reanalysis of the
modified dataset
final integrated groups (IG)

Figure 3:

Chromosome localization of YAC groups and physical linkage of microsatellite markers and ESTs. An ideogram of chromosome 2 with chromosomal band designations is given at the left. Vertical bars to the right of the ideogram indicate the cytogenetic localization (obtained from GDB) of ESTs which were contained in the STS-anchored YAC groups. Where STS-anchored YAC groups contain more than one EST, the narrowest cytogenetic assignment was used. The 5 ESTs which are not physically linked with genetically mapped microsatellites are labeled next to the corresponding vertical bars. On the far right, horizontal tick marks indicate the location of the STSs; the number between two tick marks indicates the genetic distance (cM). Physically linked STS markers within an STS-anchored YAC group are separated by slashes, while markers within different STS-anchored YAC groups at the same genetic map position are separated by semi-colons. In the center of the figure, 39 of the microsatellite-anchored YAC groups are ordered on the basis of the genetic order of the microsatellites. The numbers in parentheses after each STS-anchored YAC group indicate the number of YAC clones and the number of markers (IRS-PCR probes and STSs) in the group. Thin lines show the connections between the cytogenetic and genetic maps and the STS-anchored YAC groups. Braces are used when the markers at a map position do not fit on one line.

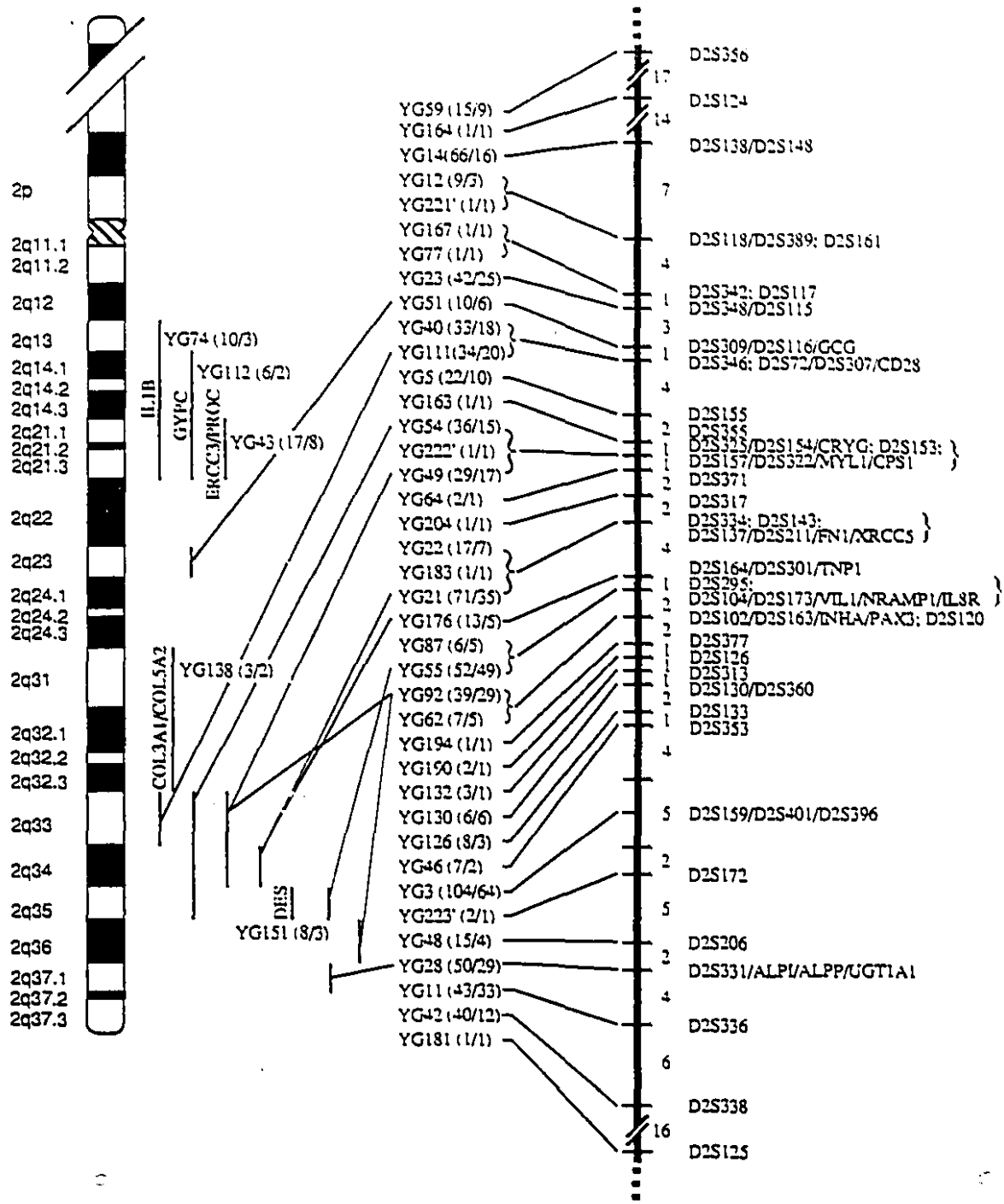

2p
2q11.1
2q11.2
2q12
2q13
2q14.1
2q14.2
2q14.3
2q21.1
2q21.2
2q21.3
2q22
2q23
2q24.1
2q24.2
2q24.3
2q31
2q32.1
2q32.2
2q32.3
2q33
2q34
2q35
2q36
2q37.1
2q37.2
2q37.3
IL1B
GYPC
ERCC3/PROC
COL3A1/COL5A2
DES
YG74 (10/3)
YG112 (6/2)
YG43 (17/8)
YG138 (3/2)
YG151 (8/3)
YG59 (15/9)
YG164 (1/1)
YG14(66/16)
YG12 (9/3)
YG221' (1/1)
YG167 (1/1)
YG77 (1/1)
YG23 (42/25)
YG51 (10/6)
YG40 (33/18)
YG111(34/20)
YG5 (22/10)
YG163 (1/1)
YG54 (36/15)
YG222' (1/1)
YG49 (29/17)
YG64 (2/1)
YG204 (1/1)
YG22 (17/7)
YG183 (1/1)
YG21 (71/35)
YG176 (13/5)
YG87 (6/5)
YG55 (52/49)
YG92 (39/29)
YG62 (7/5)
YG194 (1/1)
YG190 (2/1)
YG132 (3/1)
YG130 (6/6)
YG126 (8/3)
YG46 (7/2)
YG3 (104/64)
YG223' (2/1)
YG48 (15/4)
YG28 (50/29)
YG11 (43/33)
YG42 (40/12)
YG181 (1/1)
D2S356
17
D2S124
14
D2S138/D2S148
7
D2S118/D2S389; D2S161
4
1
D2S342; D2S117
D2S348/D2S115
3
1
D2S309/D2S116/GCG
D2S346; D2S72/D2S307/CD28
4
D2S155
2
D2S355
1
D2S323/D2S154/CRYG; D2S153;
1
D2S157/D2S322/MYL1/CPS1
2
D2S371
2
D2S317
2
D2S334; D2S143;
D2S137/D2S211/FN1/XRCC5
4
D2S164/D2S301/TNP1
1
D2S295;
2
D2S104/D2S173/VIL1/NRAMP1/IL3R
2
D2S102/D2S163/INHA/PAX3; D2S120
2
D2S377
1
D2S126
1
D2S313
1
D2S130/D2S360
2
D2S133
1
D2S353
4
5
D2S159/D2S401/D2S396
2
D2S172
5
2
D2S206
2
D2S331/ALPI/ALPP/UGT1A1
4
D2S336
6
D2S338
16
D2S125

Figure 4:

An integrated map for human chromosome region 2q. On the left is a genetic map, modified from the map of Gyapay et al. (1994). Horizontal tick marks indicate the location of microsatellites; the number between two tick marks indicates the genetic distance (cM). In the center, vertical bars indicate the location of our YAC groups (YG) and the WI/MIT Genome Center contigs (WC-groups) that are contained in the corresponding integrated groups (IG) shown on the right. A solid vertical bar indicates the minimum YAC coverage contained in an IG based on the positions of the boundary markers from the Gyapay map. A striped vertical bar is shown at the map position for a marker when an IG contains only one marker from the Gyapay map and the minimum YAC coverage could not be determined. Following the name of each IG, the number of YAC clones and the number of markers (IRS-PCR probes and STSs) contained in the group are given in parentheses. Note that *D2S346* and *D2S307* are at the same position in the Gyapay map, however, *D2S346* is part of IG20 and *D2S307* is part of IG7. Also, IG184 consists of YG204 and WC-54.

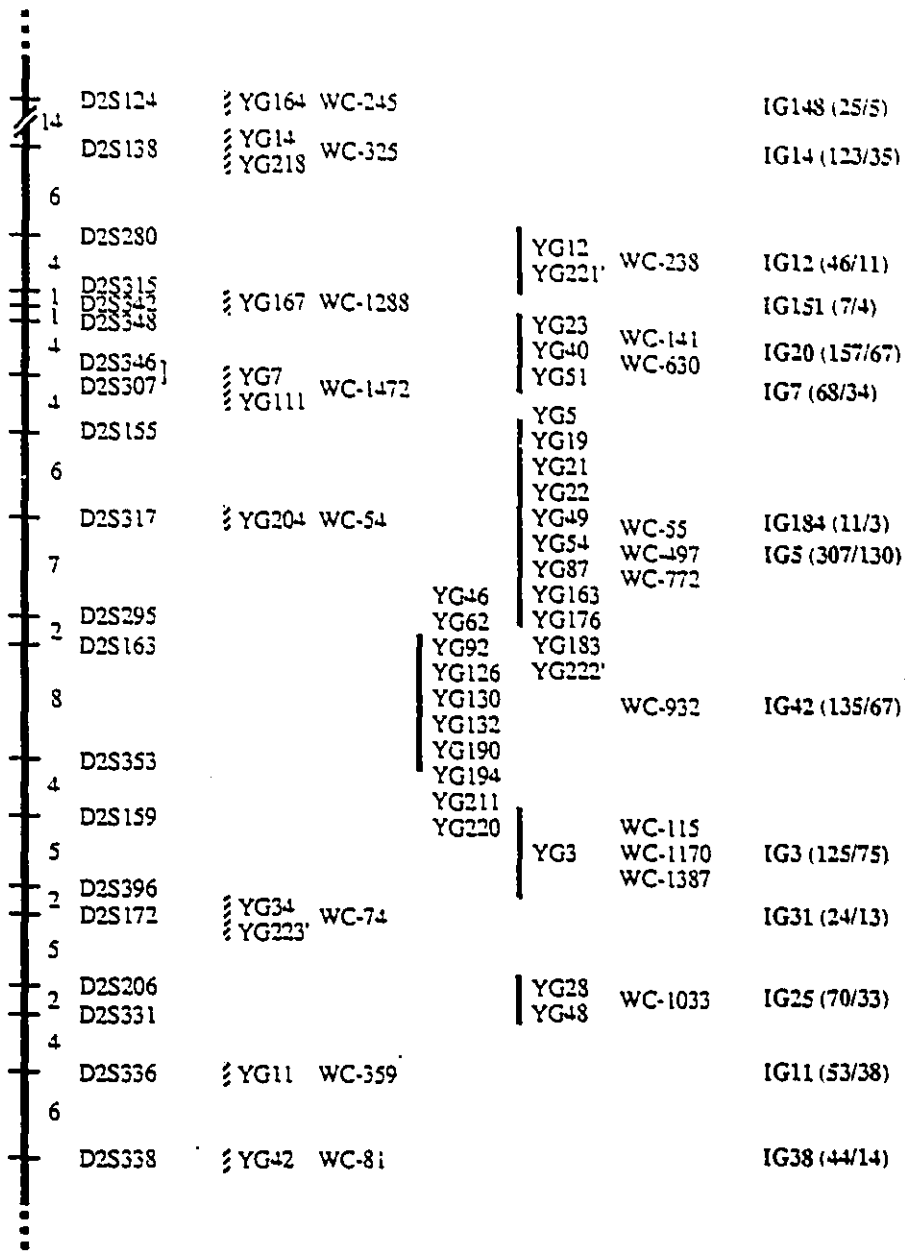

D2S124
YG164
WC-245
IG148 (25/5)
14
D2S138
YG14
YG218
WC-325
IG14 (123/35)
6
D2S280
YG12
YG221'
WC-238
IG12 (46/11)
4
D2S315
1
D2S343
1
D2S348
YG167
WC-1288
IG151 (7/4)
YG23
YG40
YG51
WC-141
WC-630
IG20 (157/67)
4
D2S346
D2S307
YG7
YG111
WC-1472
IG7 (68/34)
4
YG5
YG19
YG21
YG22
YG49
YG54
YG87
YG163
YG176
YG183
YG222'
D2S155
6
D2S317
YG204
WC-54
WC-55
WC-497
WC-772
IG184 (11/3)
IG5 (307/130)
7
YG46
YG62
YG92
YG126
YG130
YG132
YG190
YG194
YG211
YG220
D2S295
2
D2S163
8
WC-932
IG42 (135/67)
D2S353
4
D2S159
YG3
WC-115
WC-1170
WC-1387
IG3 (125/75)
5
D2S396
2
D2S172
YG34
YG223'
WC-74
IG31 (24/13)
5
D2S206
2
D2S331
YG28
YG48
WC-1033
IG25 (70/33)
4
D2S336
YG11
WC-359
IG11 (53/38)
6
D2S338
YG42
WC-81
IG38 (44/14)

Table 1. Chromosome 2 specific STSs and corresponding YACs.

Locus	Address[a] (MIT)	Address[b] (CEPH or ICRF)	Size[c] (kb)	Locus	Address[a] (MIT)	Address[b] (CEPH or ICRF)	Size[c] (kb)
ALPI	1D3H07	12_h_7	150	IL1B	1G3C12	21_e_12	350
	2J3B07	142_b_7	200		5B4E06	401_e_6	380
	3G3D03	213_d_3			7F5G07	309_g_7	
	4D5E03	320_e_3		IL8R	2A4D07	149_d_7	
	4E5E09	323_e_9			2A4F07	149_f_7	345
	6E6D07	515_d_7			5E3D02	366_d_2	430, 365
ALPP	1D3H07	12_h_7	150		5F3D09	369_d_9	
	2J3B07	142_b_7	200	INHA	2B2E05	98_e_5	
	3G3D03	213_d_3			2B5E05	157_e_5	175
	4D5E03	320_e_3			3C1B10	199_b_10	435
	4E5E09	323_e_9			3C1C11	199_c_11	400
	6E6D07	515_d_7		MYL1	1D3G08	12_g_8	445
CD28	2E5A04	168_a_4	595		4D3D02	279_d_2	430
	2I3H04	138_h_4	650		7I5D07	381_d_7	270
	2L3A11	148_a_11	425		7I5D09	381_d_9	270
	2L3A12	148_a_12	435	NRAMP1	2A3A03	96_a_3	470
	6B3D12	454_d_12	270		3G1H06	211_h_6	500
	7D3F12	135_f_12			4G6F03	330_f_3	460
	7D5D04	154_d_4			5E3D02	366_d_2	430, 365
	7K3G10	101_g_10	460	PAX3	2D3G08	120_g_8	510
	8E2B09	493_b_9	400		3A5E01	230_e_1	
COL3A1	2L3B04	148_b_4	455		3D5E07	239_e_7	595
	9A1A09	900H0925	555, 360		3F4B03	244_b_3	420
	9A1A11	900C06131	810		4E4A01	322_a_1	330
COL5A2	2L3B04	148_b_4	455		4E5G04	323_g_4	420
	9A1A09	900H0925	555, 360		6A2A09	437_a_9	595
	9A1A11	900C06131	810		7J5G03	384_g_3	270
CPS1	3G6G03	250_g_3	405, 315		7J5G04	384_g_4	200
	5G4C10	416_c_10	350		8F1E11	496_e_11	280
CRYG	1D5G04	66_g_4		PROC	1C6C06	64_c_6	180
	2C5H09	160_h_9	475		2C4F03	159_f_3	280
	6H4A08	522_a_8	320		2E4A12	167_a_12	535
	6J2B11	477_b_11	260		4F4A08	325_a_8	380
	7E5A11	272_a_11			4F4A10	325_a_10	320
	9A1A06	900E0563	445		4G4D03	328_d_3	520
	9A1A08	900G04148	690, 360		5L3C11	397_c_11	
	9A1A10	900F05123	610	TNP1	4G4E02	328_e_2	370, 220
	9A1A12	900H0758	920		4J1H09	302_h_9	280
DES	6C4B03	491_b_3	200		5C4F06	404_f_6	430
	6F1E09	464_e_9	300		5D4H10	407_h_10	310
	8F2E05	498_e_5	480		9A1A01	900CO2127	500
ERCC3	1C6C06	64_c_6	180		9A1A02	900D0120	500
	5L3C11	397_c_11			9A2G08	Am11	
	5L6G04	436_g_4			9A2G09	Am5	685
FN1	1C5B04	63_b_4	520		9A2G10	Am8	595
	1L4G12	89_g_12	710, 365	UGT1A1	1J6A10	85_a_10	
	2A4G09	149_g_9	180		7B2G02	61_g_2	950
	3E2F11	206_f_11	490		7C1G07	92_g_7	500
	4J1A04	302_a_4			9A1B01	900C0742	350
	4J5G06	340_g_6			9A1B12		935, 160
	5C4E10	404_e_10	480	VIL1	2A3A03	96_a_3	470
	5I4F08	424_f_8	565		3G1H06	211_h_6	500
	5J1D03	385_d_3	320		4A2F03	268_f_3	150
	6J5B03	541_b_3	375		4G6F03	330_f_3	460
	9A1B03	900F0374		XRCC5	1B1B11	4_b_11	370
	9A1B04	900C028			1D2C09	11_c_9	405
	9A1B05	900D044			3K3B05	225_b_5	
GCG	1D4H12	65_h_12	1000		5G3D05	372_d_5	
	2C6A01	161_a_1			7K1C09	55_c_9	880
	5H6C11	422_c_11			7K1D10	55_d_10	575
	5J1B05	385_b_5					

Table 1. *continued*

Locus	Address[a] (MIT)	Address[b] (CEPH or ICRF)	Size[c] (kb)	Locus	Address[a] (MIT)	Address[b] (CEPH or ICRF)	Size[c] (kb)
D2S102	2D3G08	120_g_8	510	cont'd	4D3D02	279_d_2	430
	2F5A09	172_a_9	630		7I5D07	381_d_7	270
D2S104	4L3E04	312_e_4	690		7I5D09	381_d_9	270
D2S115	1D6B09	67_b_9	500	D2S159	2J3E09	142_e_9	
	3D6D04	240_d_4	500		3B6G05	234_g_5	
D2S116	1D4H12	65_h_12	1000		3E4F01	241_f_1	
	1F6B05	73_b_5	605		4K4A01	343_a_1	
D2S117	1G4C12	74_c_12	540		6E4G02	513_g_2	
D2S118	1A3D09	3_d_9	455		6F5G04	517_g_4	
	1A3D10	3_d_10	470		6F5G05	517_g_5	
	1A3D11	3_d_11	365		8J3G02	513_g_2	
	1F6G10	73_g_10	850		8J3H01	513_h_1	
	2I2G08	137_g_8	460	D2S161	2B3D04	99_d_4	
D2S120	3A4F04	229_f_4		D2S163	3C1B10	199_b_10	435
	3D1A08	202_a_8			3C1C11	199_c_11	400
D2S124	3F1D05	208_d_5			4A6F04	290_f_4	
D2S125	4D4A08	319_a_8			6F5G09	517_g_9	365
D2S126	4K6D03	345_d_3	455		8C2H08	447_h_8	395
	7G6D06	342_d_6	430	D2S164	3L1G11	226_g_11	480
D2S130	2I3E11	138_e_11	410		4J1H09	302_h_9	280
	3K1D02	223_d_2	560		9A1A01	900C02127	500
	6H6H12	524_h_12	355		9A2G09	Am5	685
D2S133	4K6D05	345_d_5	190		9A2G10	Am8	595
	4L2F07	310_f_7		D2S172	1F1H08	16_h_8	375, 270
	4L4B04	346_b_4			3J4E09	257_e_9	530
D2S137	1D2C09	11_c_9	405	D2S173	4L3E04	312_e_4	690
	3A4H07	229_h_7	250	D2S206	4K6H07	345_h_7	460
	3I4G04	254_g_4	480		7C4F06	112_f_6	
	3L6D11	265_d_11	430, 365	D2S211	3I4G04	254_g_4	480
	4B3F01	273_f_1	535		4B3F01	273_f_1	535
	4B3F10	273_f_10	555		4B3F10	273_f_10	555
D2S138	1B6C02	58_c_2			4D6A02	321_a_2	400
	2J6A03	185_a_3			4D6A07	321_a_7	
	3B3B02	198_b_2	605		4D6G09	321_g_9	
D2S143	4F2E08	287_e_8	550		7E4F05	156_f_5	370
D2S148	1A4B03	50_b_3	560		7K1C09	55_c_9	880
	1C4G09	62_g_9	760		7K1D10	55_d_10	575
	2A3G01	96_g_1	425	D2S295	1I5H01	81_h_1	390, 265
	3A4D05	229_d_5	460		7B1A02	60_a_2	580
	3B3B02	198_b_2	605	D2S301	3L1G11	226_g_11	480
	3D4E10	238_e_10			4J1H09	302_h_9	280
	3E3H11	207_h_11			9A1A01	900C02127	500
	3F1F05	208_f_5	560		9A1A02	900D0120	500
	3H1A07	214_a_7	705	D2S307	2E5A04	168_a_4	595
	6A2D05	437_d_5		D2S309	1D4H12	65_h_12	1000
	6G3A10	469_a_10			1D5E08	66_e_8	
	7F2B11	170_b_11			1F6B05	73_b_5	605
D2S153	4F1B05	286_b_5	450, 380		6K6G02	549_g_2	
D2S154	1D5G04	66_g_4		D2S313	2I2G08	141_g_8	
	2D6A06	164_a_6			2L1D11	146_d_11	
	5D1E08	358_e_8			5I1A04	378_a_4	
	5D1E09	358_e_9		D2S317	6D3B04	460_b_4	355
	6G6D03	521_d_3		D2S322	1L3B02	49_b_2	415
	9A1A06	900E0563			4F5A02	326_a_2	
	9A1A10	900F05123			4G3B07	294_b_7	
D2S155	1A1D09	1_d_9	380		5I5A08	425_a_8	350
	1F1C02	16_c_2	540		7G4E03	333_e_3	350
	1F5C02	72_c_2	540	D2S325	1L6G02	91_g_2	
	4F4D03	325_d_3			3D4B03	238_b_3	
	5I2D10	379_d_10	320		5L4B01	434_b_1	
D2S157	1D3G08	12_g_8	445	D2S331	1D5B10	66_b_10	

Table 1. *continued*

Locus	Address[a] (MIT)	Address[b] (CEPH or ICRF)	Size[c] (kb)
cont'd	1I3D06	27_d_6	
	1L3D04	49_d_4	
	3J1C02	220_c_2	615
D2S334	3D2D08	203_d_8	390
	4A2E11	268_e_11	380
	4A2G11	268_g_11	415
	4K5B09	344_b_9	300, 270
	5B1H03	352_h_3	320
	5D2G02	360_g_2	320
D2S336	3C1C08	199_c_8	595
	6H4H12	522_h_12	
	7F1C01	166_c_1	
	7F2B11	170_b_11	
D2S338	1C5H02	63_h_2	
	1J6E03	85_e_3	
	2F2E02	126_e_2	
	2J6E08	185_e_8	
	4B5G05	314_g_5	
	6A3F04	438_f_4	
	8B3B04	445_b_4	
D2S342	3F6H04	246_h_4	
D2S346	3L2E02	227_e_2	
	4C5G07	317_g_7	
	4F4D08	325_d_8	
D2S348	3D6D04	240_d_4	500
	3G4E10	247_e_10	
D2S353	4B4D09	313_d_9	
	8E2A02	493_a_2	
	8F2F02	498_f_2	
D2S355	3E5A08	242_a_8	
D2S356	1E6D02	70_d_2	
	2B6A06	158_a_6	
	2D5A02	163_a_2	
	2D5C02	163_c_2	
	4E5G04	323_g_4	420
	4G1B05	289_b_5	
	6K5A10	545_a_10	
D2S360	6H6H12	524_h_12	355
D2S371	1E5E12	69_e_12	445
	1K4D04	86_d_4	540
D2S377	5D2H08	360_h_8	
D2S389	1A3D11	3_d_11	365
	1F6G10	73_g_10	850
	2I2G08	137_g_8	460
	4E1H06	280_h_6	
D2S396	4F2E01	287_e_1	
	4F4F05	325_f_5	
	5A5G03	399_g_3	
D2S401	4E1A06	280_a_6	
	4E1A08	280_a_8	
	4E2E11	281_e_11	
	4E2H11	281_h_11	
	4E3D06	284_d_6	
	6F4A12	516_a_12	
	6F5H01	517_h_1	
D2S72	2L3A11	148_a_11	425
	2L3A12	148_a_12	435
	7D3F12	135_f_12	

[a] Corresponds to MIT's YAC library pooling scheme (Liu et al. 1995a). YAC clones with plate address 9A1 or 9A2 are from the ICRF YAC library.

[b] Corresponds to the CEPH YAC designation where applicable except those underlined which correspond to the ICRF YAC library.

[c] Size provided where known.

Table 2. Forty-four STS-anchored YAC groups (YG)

YG no.	Linked STS(s)[a]	YACs[b]	Markers[c]	YG no.	Linked STS(s)[a]	YACs[b]	Markers[c]
21	D2S137/D2S211/FN/XRCC5	71	35	48	D2S206	15	4
28	D2S331/ALPP/ALPI/UGT1A1	50	29	59	D2S356	15	9
49	D2S322/D2S157/MYL1/CPS1	29	17	62	D2S120	7	5
51	D2S309/D2S116/GCG	10	6	64	D2S371	2	1
54	D2S325/D2S154/CRYG	36	15	74	IL1B	10	3
55	D2S104/D2S173/VIL1/NRAMP1/IL8R	52	49	77	D2S117	1	1
92	D2S102/D2S163/PAX3/TNHA	39	29	87	D2S295	6	5
111	D2S72/D2S307/CD28	34	20	112	GYPC	6	2
176	D2S164/D2S301/TNP1	13	5	126	D2S133	8	3
3	D2S159/D2S401/D2S396	104	64	132	D2S313	3	1
12	D2S118/D2S389	9	3	151	DES	8	3
14	D2S138/D2S148	66	16	163	D2S355	1	1
23	D2S348/D2S115	42	25	164	D2S124	1	1
130	D2S130/D2S360	6	6	167	D2S342	1	1
43	ERCC3/PROC	17	8	181	D2S125	1	1
138	COL3A1/COL5A2	3	2	183	D2S143	1	1
5	D2S155	22	10	190	D2S126	2	1
11	D2S336	43	33	194	D2S377	1	1
22	D2S334	17	7	204	D2S317	1	1
40	D2S346	33	18	221'	D2S161	1	1
42	D2S338	40	12	222'	D2S153	1	1
46	D2S353	7	2	223'	D2S172	2	1

[a] The STS(s) contained in the group.

[b] The number of YACs contained in the group.

[c] The number of markers (IRS-PCR probe or STS).

As indicated in the preface, my thesis project originated from the generation of genetic markers for the study of susceptibility to tuberculosis and leprosy (Chapter 4). This project was subsequently expanded to include large scale YAC mapping of human chromosome 2 (Chapters 2 and 3). A major impetus for the human genome project was provided by the suggestion that investigation of genetic diseases would greatly benefit from a detailed knowledge of the genomic organization of chromosomal regions carrying disease genes. In our experiments, this proposal has been proven correct, and the study of susceptibility to tuberculosis and leprosy has definitely benefited from the mapping information obtained in Chapters 2 and 3. In Chapter 3, we physically linked 17 microsatellites with 16 ESTs in 9 YAC groups. Among them, two highly informative microsatellites, *D2S173* and *D2S104*, and an EST which was derived from exon 2 of *NRAMP1* were found to belong to the same YAC group. Subsequent YAC contig analysis indicated that *D2S173* and *D2S104* were physically linked with *NRAMP1* on a 1.5-Mb YAC contig. Together with 9 other genetic markers identified within the *NRAMP1* gene as described in Chapter 4, these molecular markers can be included in an extended *NRAMP1* haplotype and will be useful to assess the role of *NRAMP1* in susceptibility to tuberculosis and leprosy.

CHAPTER 4:

Identification of polymorphisms and sequence variants in the human homologue of the mouse natural resistance-associated macrophage protein gene

Am. J. Hum. Genet. 56:845–853, 1995

Identification of Polymorphisms and Sequence Variants in the Human Homologue of the Mouse Natural Resistance–Associated Macrophage Protein Gene

Jing Liu,[1,2] T. Mary Fujiwara,[2,3,4] Natalie T. Buu,[1,2] Fabio O. Sánchez,[1,2] Mathieu Cellier,[1,5] Ann Josée Paradis,[6] Danielle Frappier,[2] Emil Skamene,[1,2] Philippe Gros,[1,5] Kenneth Morgan,[1,2,3,6] and Erwin Schurr[1,2]

[1]McGill Centre for the Study of Host Resistance and Departments of [2]Medicine, [3]Human Genetics, [4]Pediatrics, [5]Biochemistry, and [6]Epidemiology and Biostatistics, McGill University, Montreal

Summary

The most common mycobacterial disease in humans is tuberculosis, and there is evidence for genetic factors in susceptibility to tuberculosis. In the mouse, the *Bcg* gene controls macrophage priming for activation and is a major gene for susceptibility to infection with mycobacteria. A candidate gene for *Bcg* was identified by positional cloning and was designated "natural resistance-associated macrophage protein gene" (*Nramp1*), and the human homologue (*NRAMP1*) has recently been cloned. Here we report on (1) the physical mapping of *NRAMP1* close to *VIL* in chromosome region 2q35 by PCR analysis of somatic cell hybrids and YAC cloning and (2) the identification of nine sequence variants in *NRAMP1*. Of the four variants in the coding region, there were two missense mutations and two silent substitutions. The missense mutations were a conservative alanine-to-valine substitution at codon 318 in exon 9 and an aspartic acid–to-asparagine substitution at codon 543 in the predicted cytoplasmic tail of the NRAMP1 protein. A microsatellite was located in the immediate 5′ region of the gene, three variants were in introns, and one variant was located in the 3′ UTR. The allele frequencies of each of the nine variants were determined in DNA samples of 60 Caucasians and 20 Asians. In addition, we have physically linked two highly polymorphic microsatellite markers, *D2S104* and *D2S173*, to *NRAMP1* on a 1.5-Mb YAC contig. These molecular markers will be useful to assess the role of *NRAMP1* in susceptibility to tuberculosis and other macrophage-mediated diseases.

Received July 12, 1994; accepted for publication January 9, 1995.

Address for correspondence and reprints: Dr. Erwin Schurr, Montreal General Hospital Research Institute, Room L11-521, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada.

0002-9297/95/5604-0005$02.00

Introduction

Tuberculosis is a major public health concern worldwide. About one-third of the world's population is infected with *Mycobacterium tuberculosis*, and 30 million people are expected to die in the next decade from tuberculosis (World Health Organization 1994). While environmental exposure to mycobacteria is necessary for the development of tuberculosis, intrinsic host factors are also important in determining the outcome of an infection with *M. tuberculosis*. For example, human twin studies show significantly greater concordance of disease among MZ twins compared with DZ twins of the same sex (Kallmann and Reisner 1943; Vogel and Motulsky 1986, pp. 213–214).

Genetic studies in the mouse have demonstrated that innate susceptibility to *M. bovis* (BCG), *M. lepraemurium*, *M. intracellulare*, and *M. avium*, as well as two nonmycobacterial species, *Salmonella typhimurium* and *Leishmania donovani*, is under the control of a single gene located in the proximal region of mouse chromosome 1 (Schurr et al. 1990*a*). This gene has been alternatively designated "*Lsh*," "*Ity*," or "*Bcg*" (Blackwell 1989). Susceptibility to infection is recessive to resistance. Resistant mice restrict the growth of the above-mentioned infectious agents in their reticuloendothelial organs, while rapid uncontrolled proliferation of injected pathogens occurs in susceptible mice. Analysis of the resistant and susceptible mice has implicated the mature tissue macrophage in resistance to infection and has led to the proposal that the *Bcg/Lsh/Ity* gene regulates macrophage priming for activation (Buschman et al. 1989; Blackwell et al. 1991).

A candidate gene for *Bcg*, originally designated "*Nramp*" (which we refer to as "*Nramp1*"), was isolated by positional cloning and was shown to be expressed in the macrophage (Vidal et al. 1993). Sequencing of the *Nramp1* gene from 27 inbred strains of mice showed that susceptibility to mycobacterial infection was associated with a nonconservative substitution of glycine to aspartic acid in the second putative transmem-

brane domain (Malo et al. 1994). A human homologue of the mouse gene has recently been cloned, and the exon-intron organization has been determined for 16 exons (Cellier et al. 1994). In the present communication we report on the localization of the human *NRAMP1* gene (originally designated "*NRAMP*") and on the identification of nine sequence variants. These variants should be useful for the genetic analysis of susceptibility to tuberculosis and other nonparasitic diseases of the macrophage in humans.

Material and Methods

Somatic Cell Hybrids

The National Institute of General Medical Sciences (NIGMS) mapping panel 2 (Human Genetic Mutant Cell Repository, Coriell Institute for Medical Research, Camden, NJ), which is made up of 24 hybrid cell lines that contain individual chromosomes as their only human complement (except for hybrid cell lines 07299 and 10478, which contain human chromosomes 1 and X, and 4 and 20, respectively), was used for the chromosomal assignment of *NRAMP1*. The somatic cell hybrids were grown according to the conditions specified by the supplier. After one passage of cells, DNA was extracted according to standard protocols using proteinase K treatment and phenol/chloroform extraction (Sambrook et al. 1989). Primers 2F (5′-CCCTCCCTTAATGAAGGATC-3′) and 4R (5′-CCACCACTCCCCTATGAGGTG-3′), which hybridize to the 5′ and 3′ exon-intron boundaries of exons 5 and 6 of *NRAMP1*, were used to amplify DNA from the hybrid panel.

Construction of a YAC Contig in the Vicinity of NRAMP1

We previously identified close to 2,000 chromosome 2–specific YAC clones from the CEPH Mark I YAC library by the interspersed repetitive sequence (IRS)–PCR approach (Liu et al., in press). These YACs were subsequently used for sequence-tagged site (STS) screening with a set of expressed sequence tags and microsatellite markers including primers specific for *NRAMP1* and *VIL* (Rousseau-Merck et al. 1988; Lu-Kuo et al. 1993) and *D2S104* and *D2S173* (Spurr et al. 1994). The primer sequences for *VIL* are Vil-3 (5′-TACAGTGAGGACCCATGTGC-3′) and Vil-4 (5′-TCTGTTAAGGGTTTCGGCTC-3′). YAC clones positive for *NRAMP1* and *VIL* were used for chromosome walking in order to construct a YAC contig. The protocols for preparing YAC DNA, amplifying YAC DNA with IRS-PCR primers, obtaining walking probes, and screening of the CEPH Mark I YAC library are described elsewhere (Liu et al., in press). The Baylor College YAC library was screened by Dr. G. Chinault with the Vil-3/4 primers. The mega YACs, 829_c_12 and 872_h_8, were isolated from a copy of the CEPH YAC libraries available in our laboratory.

Human Genomic DNA Samples

DNA samples were obtained from Epstein-Barr virus (EBV)–transformed cell lines of nine Native Canadians who are members of an extended multiplex tuberculosis family (Mah and Fanning 1991), from EBV-transformed cell lines of 27 individuals from 12 tuberculosis families from Hong Kong, and from white blood cells of 60 unrelated Caucasians and 20 unrelated Asians. High-molecular-weight genomic DNA was extracted from EBV-transformed cells and white blood cells, according to standard protocols (Sambrook et al. 1989), was quantified by standard UV spectroscopy, and was stored at 4°C in Tris-EDTA (pH 8.0).

Molecular Cloning

Subcloning of restriction fragments containing the 5′ end of the *NRAMP1* gene was done according to standard procedures by digestion of cosmid DNA with restriction enzymes and size separation of restriction products on agarose gels. A 1.6-kb *PstI* fragment overlapping the 5′ region of the *NRAMP1* gene was identified by Southern analysis using a partial cDNA clone as a probe. The 1.6-kb fragment was extracted from agarose by using the Qiagen DNA purification kit and was subcloned in pBluescript KS+; fragment ends were sequenced by the dideoxy-chain termination method (Sanger et al. 1977) using the Sequenase kit (U.S. Biochemical) and T3- or T7-specific sequencing primers. Sequence analysis revealed the presence of a microsatellite ~150 bp into the 5′ region of the fragment. A 5′ PCR primer (5′-GACATGAAGACTCGCATTAG-3′) was designed and, together with the 3′ primer (5′-TCAAGTCTCCACCAGCCTAGT-3′), was used to test for microsatellite length variation. To allow visualization of the length polymorphism, amplification was done with a ^{33}P-labeled 5′ primer. The amplification products were digested with *NcoI*, and end-labeled fragments were sized on standard sequencing gels. The ligation and transformation of recombinant vectors into DH5α, as well as the selection and growth of recombinant bacteria, were done according to standard protocols (Sambrook et al. 1989). Plasmid DNA was isolated using the QIA prep-spin plasmid kit (Qiagen) and was used for dideoxy sequencing of the inserts. T-vector was prepared according to the protocol described by Marchuk et al. (1991), with minor modifications.

Single-Strand Conformation Analysis (SSCA)

The PCR reaction mix used for all SSCA PCR reactions contained 200 ng of genomic DNA, 0.2 mM dGTP, 0.2 mM dCTP, 0.2 mM dTTP, 0.07 mM dATP, 16 pmol of each primer, 1 mM $MgCl_2$, 10 mM Tris-

HCl (pH 8.0), 0.05% Tween-20, 0.05% NP40, 0.2 units of *Taq* DNA polymerase (Perkin Elmer Cetus), and 8 μCi of ^{35}S-dATP in a total reaction volume of 15 μl. The conditions for amplification were a single denaturation step of 3 min at 94°C, followed by 30 cycles of denaturation (1 min at 94°C), annealing (1 min at 55°C–63°C primer-specific annealing temperature), and extension (1 min at 72°C), and a final extension step of 7 min at 72°C. The primers used for amplification of *NRAMP1* gene segments for which sequence variants were identified are listed in table 1. All SSCA variants were detected on 6% polyacrylamide/5% glycerol sequencing gels run at room temperature at a constant current of 28 mA for 5 h.

DNA Sequencing

PCR amplifications were carried out in 20-μl reaction volumes containing 100 ng of genomic DNA, 50 mM Tris-HCl (pH 8.3), 1.8 mM $MgCl_2$, 0.05% Tween-20, 0.05% NP40, 187.5 μM of each dNTP (Promega), 0.4 μM of each primer, and 1 unit of *Taq* polymerase (Perkin Elmer Cetus). Parameters for thermocycling were as follows: incubation for 4 min at 94°C, followed by 33 cycles of 1 min at 94°C, 30 s at 59°C, 1 min at 58°C, 1 min 40 s at 72°C, and a final extension step of 10 min at 72°C. For direct sequencing, the genomic PCR-amplification products were diluted 1:15 in H_2O without further purification and were subjected to thermal cycle sequencing using the Hot Tub DNA Sequencing System (Amersham). For clone sequencing, 1–2 μg of purified plasmid DNA were used for dideoxy sequencing using the Sequenase kit (U.S. Biochemical) with T3- or T7-specific sequencing primers. For each ligation, at least four clones were sequenced in both sense and antisense orientation.

Southern Analysis

Five micrograms of genomic DNA or 1 μg of cosmid DNA was digested with restriction enzymes, with conditions and quantities recommended by the supplier (New England BioLabs). The digested DNA samples were size separated on agarose gels, were transferred onto Hybond nylon membranes (Amersham), and hybridized with probes labeled to high specific activity (5×10^8 cpm/μg DNA) with α-^{32}P-dATP (specific activity >3,000 Ci/mmol; Amersham [Feinberg and Vogelstein 1983, pp. 213–214]) under high-stringency conditions as described elsewhere (Schurr et al. 1989). All *NRAMP1*-derived probes were preannealed with 400 μg of human placental DNA in 1 ml of hybridization medium for 1 h at 65°C.

Restriction-Site Analysis

Analysis of the restriction-endonuclease recognition sites in DNA sequences was carried out with the MacVector software package, version 4.1 (International Biotechnologies). Restriction-endonuclease digestions were done by using 5–10 U of enzyme per PCR reaction, under conditions recommended by the supplier (New England BioLabs). Restriction-enzyme digestion products were resolved by electrophoresis on 12% polyacrylamide gels stained with ethidium bromide; a 100-bp ladder was used as a size marker (Pharmacia).

Results

Physical Mapping of NRAMP1

The chromosomal localization of *NRAMP1* was determined by PCR amplification of genomic DNA extracted from the monochromosomal somatic cell hybrids of NIGMS panel 2 by using primers 2F and 4R. As shown in figure 1, the predicted PCR product of ~700 bp was amplified in human genomic DNA and in DNA obtained from hybrid 10826B, which contains human chromosome 2 as its only human chromosome. The fact that only one clearly identifiable band was amplified in human genomic DNA and in hybrid 10826B demonstrated the *NRAMP1* specificity of the primers and showed that *NRAMP1* is located on chromosome 2.

In the mouse, *Nramp1* is closely linked to *Vil* (Malo et al. 1993). To estimate the physical distance between the human homologues of these genes, we carried out STS screening of a previously generated chromosome 2 CEPH Mark I YAC sublibrary, with a set of STS markers including primers specific for *NRAMP1*, *VIL*, and 50 microsatellite markers. Three YAC clones, 2A3A03 (96_a_3), 3G1H06 (211_h_6), and 4G6F03 (330_f_3), were identified that carried both *NRAMP1* and *VIL* (the CEPH YAC designations are given in parentheses). STS screening of the Baylor College YAC library, with primers specific for *VIL*, identified a YAC clone, B2, with an insert size of 200 kb. PCR analysis with primers specific for *NRAMP1* revealed that B2 also contained the *NRAMP1* sequence, suggesting a maximum distance of 200 kb separating these two loci. Thus, *NRAMP1* can be assigned to chromosome region 2q35, since *VIL* had previously been assigned to this chromosomal region (Rousseau-Merck et al. 1988; Lu-Kuo et al. 1993). A YAC contig of the region encompassing *NRAMP1* and *VIL* was constructed by chromosome walking in the CEPH Mark I YAC library by using IRS-PCR probes derived from YAC clones that included *NRAMP1* and *VIL* (fig. 2). Among the YAC clones identified by chromosome walking, clone 4L3E04 (312_e_4) tested positive for microsatellite markers *D2S104* and *D2S173*. We then searched Généthon's GenomeView database via the Internet (URL is http://www.genethon.fr) for *D2S173* and *D2S104*, and we identified a 210-kb YAC clone (829_e_12) that was positive for both markers. In

Table 1

***NRAMP1* Polymorphisms and Sequence Variants**

Name	Nucleotide/Amino Acid Change	Primers, 5′ to 3′ (PRC Product Size)	Polymorphic Enzyme	Allele[a]	Allele Frequency: Caucasian (n = 120)	Allele Frequency: Asian (n = 40)
$(GT)_n$	Microsatellite, 5′ to exon 1	GAC ATG AAG ACT CGC ATT AG TCA AGT CTC CAC CAG CCT AGT (~800 bp)	*Nco*I	Allele 1 = 286 bp Allele 2 = 288 bp Allele 3 = 290 bp	.73 .25 .02	.85 .10 .05
274C/T	C or T at nucleotide 274; TTC or TTT (Phe) at codon 66 in exon 3	TGC CAC CAT CCC TAT ACC CAG TCT CGA AAG TGT CCC ACT CAG (216 bp)	*Mnl*I	Allele 1 (T) = 167, 37, and 12 bp Allele 2 (C) = 102, 65, 37, and 12 bp	.27 .73	.12 .88
469+14G/C	G or C at nucleotide + 14 of intron 4	TCT CTG GCT GAA GGC TCT CC TGT GCT ATC AGT TGA GCC TC (624 bp)	*Apa*I	Allele 1 (G) = 624 bp Allele 2 (C) = 455 and 169 bp	.73 .27	.92 .08
577−18G/A	G or A at nucleotide −18 of intron 5	CTG GAC CAG GCT GGG CTG AC CCA CCA CTC CCC TAT GAG GTG (146 bp)	*Msp*I	Allele 1 (A) = 146 bp Allele 2 (G) = 125 and 21 bp	0 1.0	.02 .98
823C/T	C or T at nucleotide 823; GGC or GGT (Gly) at codon 249 in exon 8	CTT GTC CTG ACC AGG CTC CT CAT GGC TCC GAC TGA GTG AG (234 bp)	*Nar*I	Allele 1 (T) = 234 bp Allele 2 (C) = 135 and 99 bp	.02 .98	.15 .85
A318V	C or T at nucleotide 1029; GCG (Ala) or GTG (Val) at codon 318 in exon 9	TCC TTG ATC TTC GTA GTC TC GGC TTA CAG GAC ATG AGT AC (232 bp)	*Bso*FI	Allele 1 (Val) = 232 bp Allele 2 (Ala) = 171 and 61 bp	0 1.0	0 1.0
1465-85G/A	G or A at nucleotide −85 of intron 13	GCA AGT TGA GGA GCC AAG AC ACC TGC ATC AAC TCC TCT TC (241 bp)	*Bsr*I	Allele 1 (A) = 142, 75, and 24 bp Allele 2 (G) = 102, 75, 40, and 24 bp	.38 .62	.35 .65
D543N	G or A at nucleotide 1703; GAC (Asp) or AAC (Asn) at codon 543 in exon 15	GCA TCT CCC CAA TTC ATG GT AAC TGT CCC ACT CTA TCC TG (240 or 244 bp)	*Ava*II	Allele 1 (Asn) = 201 and 39 bp Allele 2 (Asp) = 126, 79, and 39 bp	.01 .99	.18 .82
1729+55del4	Deletion of TGTG in the 3′ UTR (55 nt 3′ to the last codon in exon 15)	Same as for D543N	*Fok*I	Allele 1 (−TGTG) = 240 bp Allele 2 (+TGTG) = 211 and 33 bp	.01 .99	.18 .82

[a] All alleles show Mendelian segregation (data not shown).

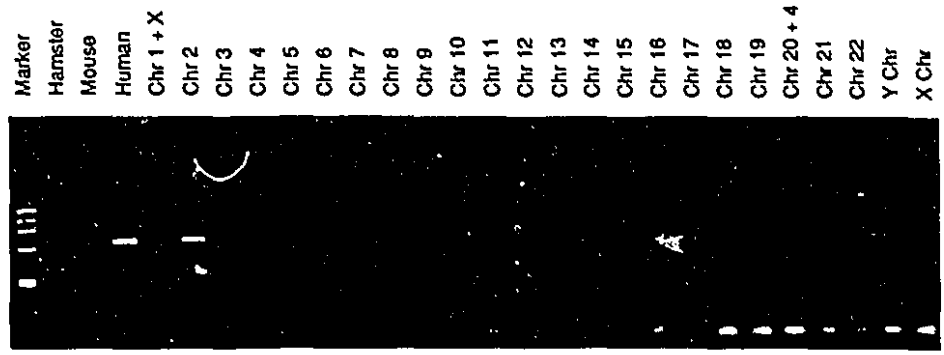


Figure 1 PCR analysis of monochromosomal somatic cell hybrids. Human, hamster, and mouse genomic DNAs were used as controls. *Hae*III-digested ΦX174 DNA was used as the size marker.

addition, we identified a mega YAC clone (872_h_8), from the Quickmap database (Cohen et al. 1993), that was positive for *D2S173*. DNA from this YAC clone was tested using primers specific for *VIL* and *NRAMP1*, and positive amplification was observed for both markers. Taken together, the results of these experiments demonstrated that *VIL*, *NRAMP1*, *D2S173*, and *D2S104* are physically linked on a chromosomal segment with an estimated size of 1.5 Mbp (fig. 2).

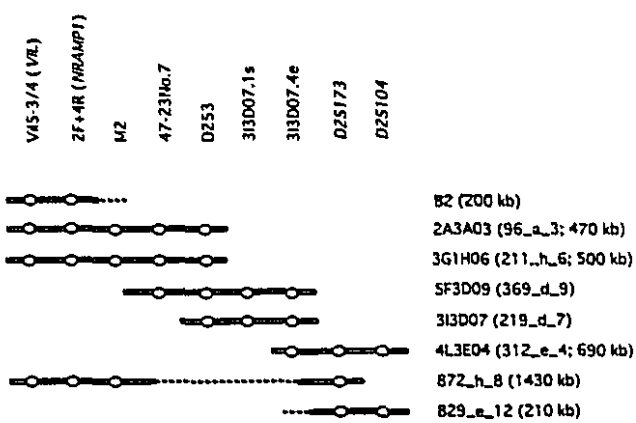


Figure 2 Genomic cloning of the region that includes *NRAMP1* and *VIL*. A likely order of markers and loci is given at the top of the figure. Only the YAC clones belonging to a minimum tiling path and five of the chromosome-walking probes (M2, 47-23No.7, D253, 313D07.1s, and 313D07.4e) are shown; however, the tiling path and the order of probes were derived from all the data. The order of probes has not been independently established, and other orders are likely. YACs, which are not drawn to scale, are represented as shaded bars; the names of YAC clones are given to the right of the bars. The CEPH YAC designations for CEPH Mark I YAC clones and the known sizes of YACs are indicated in parentheses. An open ellipse on a bar indicates that the YAC clone was identified by the corresponding probe. Dotted lines on three of the YACs (B2, 872_h_8, and 829_e_12) indicate that these YACs were not tested for the presence of the chromosome-walking probes. Probe M2 is an IRS walking probe derived from clone 872_h_8 and was used to screen the CEPH Mark I YAC library. The estimated size of the contig is 1.5 Mb.

Identification of Polymorphisms and Sequence Variants in the NRAMP1 *Gene*

We used several methods to search for mutations in the *NRAMP1* gene in genomic DNA of a panel of individuals that included 9 members of a Native Canadian kindred and 24 individuals from 11 Hong Kong families. Primers for 15 of the 16 exons were derived in the process of identifying the exon-intron organization of *NRAMP1* (Cellier et al. 1994). The majority of these primers were located within intronic sequences, to allow amplification of the exon-intron splice consensus sequences. The DNA samples were screened by SSCA of all *NRAMP1* exons except for exon 1. Direct sequencing revealed four SSCA variants that were due to mutations in the coding region and one that was in intronic sequences. Two variants were predicted to cause amino acid substitutions: A318V, an alanine-to-valine substitution at codon 318 in exon 9; and D543N, an aspartic acid–to-asparagine substitution at codon 543 in exon 15 (table 1). Two variants were silent nucleotide substitutions: 274C/T in codon 66 (phenylalanine) in exon 3 and 823C/T in codon 249 (glycine) in exon 8. The intronic variant, 577−18G/A, was a G-to-A nucleotide substitution near the 3′ end of intron 5. The mutations have been designated according to the nomenclature suggested by Beaudet and Tsui (1993); the nucleotides and codons were numbered according to the sequence of GenBank accession number L32185.

In the process of DNA sequencing to determine the sequence changes of the SSCA variants, we found two additional mutations: 1465−85G/A, a G-to-A nucleotide substitution near the 3′ end of intron 13; and 1729+55del4, a 4-bp TGTG deletion located 55 nt downstream of the last codon in exon 15. This 4-bp deletion creates a heteroduplex of the PCR products of individuals who are heterozygous for the insertion/deletion polymorphism. By Southern analysis, an *Apa*I RFLP was detected (allele 1 = 5 kb; and allele 2 = 4 kb + 1 kb) by using an *NRAMP1* cDNA probe. Sequential

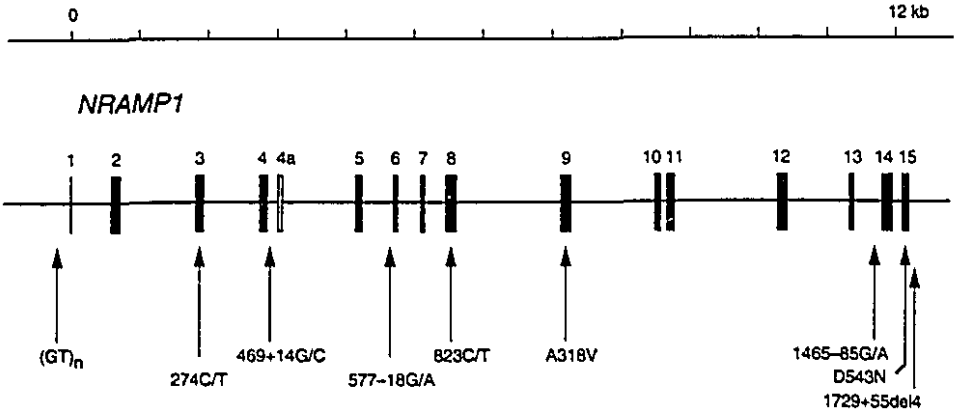


Figure 3 Location of polymorphisms and sequence variants in *NRAMP1*. The schematic diagram of the exon-intron organization was adapted from Cellier et al. (1994).

hybridization using *NRAMP1* exon probes to DNA from individuals heterozygous for the RFLP, along with direct sequencing, identified the polymorphism, designated "469+14G/C," as a G-to-C nucleotide substitution in intron 4. Finally, a microsatellite was identified on a 1.6-kb genomic *Pst*I fragment overlapping exons 1 and 2 of *NRAMP1* and was located ~800 bp 5′ to the *NRAMP1* start codon. The structure of the repeat was determined to be $(GT)_7AC(GT)_5AC(GT)_9$. We did not find any variants in exon 4a, which is an alternatively spliced exon (fig. 3) and codes for an inverted Alu Sx element (Jurka and Milosavljevic 1991), either by direct sequencing of 6 individuals or by SSCA of 33 individuals composing our screening panel and of 24 unrelated Caucasians.

Analysis of the sequence variants revealed that the nucleotide changes either created or destroyed a recognition site for restriction endonucleases (except for the microsatellite), which allowed us to develop a PCR assay for each of the variants (table 1). YAC clone 2A3A03 was used as a positive control for specificity of PCR assays of *NRAMP1* sequences. To determine if the sequence variants were polymorphic, we tested DNA samples from 60 Caucasians and 20 Asians. The D543N missense mutation in exon 15 and the 4-bp deletion, 1729+55del4, in the 3′ UTR appear to be in absolute linkage disequilibrium. The allele frequencies differed between the two ethnic groups, for four polymorphisms: 469+14G/C, 823C/T, D543N, and 1729+55del4 (Fisher exact test; $P < .05$). The 1465−85G/A polymorphism in intron 13 had the highest expected heterozygosity, in both the Caucasian and Asian samples (.47 and .46, respectively).

The observed allele counts for the nine variants are shown in table 2, for 22 unrelated members of 12 Hong Kong families and for 8 unrelated members of the Canadian kindred. We detected no significant association between tuberculosis disease status and the *NRAMP1* alleles in the small sample from either of these groups (Fisher exact test; $P > .05$).

Discussion

The evolutionary conservation of chromosomal segments between mouse and human is well documented, and comparative mapping has been employed successfully for the identification of a number of genes or disease loci (for reviews, see Nadeau et al. 1992; Liu et al. 1993; Searle et al. 1994). We have shown previously that a 35-cM chromosomal segment that includes *Bcg* on proximal mouse chromosome 1 is conserved in human chromosome region 2q33-qter (Schurr et al. 1990*b*). Among the conserved loci, *Vil* was the locus most closely linked to *Bcg*, and the physical distance between *Vil* and *Nramp1*, the candidate gene for *Bcg*, was determined to be ~50 kb (Malo et al. 1993). In the present study, we have physically mapped the human *NRAMP1* gene to within a maximum distance of 200 kb of *VIL*, in chromosome region 2q35, by PCR analysis of somatic cell hybrids and by YAC cloning. Both this chromosomal location of *NRAMP1* and the observed high sequence similarity with mouse *Nramp1* (Cellier et al. 1994) strongly argue against the possibility that human homologues of either the mouse chromosome 17 *Nramp*-related sequence (Dosik et al. 1994) or the mouse chromosome 15 *Nramp2* gene (Gruenheid et al., in press) were amplified in our experiments. A possible human homologue of *Nramp*-related sequence is expected to map to chromosome region 6q27 (Dosik et al. 1994), and human *NRAMP2* has been mapped, by in situ hybridization, to 12q13 (S. Vidal and P. Gros, personal communication). Furthermore, the estimated maximum distance between *VIL* and *NRAMP1* on chromosome 2q35 was similar to the observed physical distance between *Nramp1* and *Vil* on mouse chromosome 1. Our results, therefore, support the validity of mouse-human

Table 2

Distribution of Alleles of *NRAMP1* Polymorphisms and Sequence Variants in Unrelated Individuals from Tuberculosis Families

		Sequence Variant																			
		$(GT)_n$			274C/T		469+14G/C		577−18G/A		823C/T		A318V		1465-85G/A		D543N		1729+55del4		
Disease Status[a]	No. of Individuals	1	2	3	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	
Hong Kong families:																					
TB(+)	9	18	0	0	0	18	18	0	0	18	2	16	1	17	2	16	3	15	3	15	
TB(−)	13	22	2	2	2	24	24	2	2	24	1	25	0	26	8	18	3	23	3	23	
Canadian kindred:																					
TB(+)	3	4	2	0	2	4	4	2	0	6	1	5	0	6	3	3	1	5	1	5	
TB(−)	5	8	2	0	1	9	9	1	0	10	2	8	3	10	3	7	3	7	3	7	

[a] TB(+) = affected with tuberculosis; and TB(−) = not affected with tuberculosis.

comparative genome analysis on the level of physical maps.

We have identified nine sequence variants in *NRAMP1*. Seven of these variants are silent substitutions or are located in introns or UTRs of the gene. Only two of the sequence variants are predicted to cause amino acid substitutions: an alanine-to-valine substitution at codon 318 (A318V), which changes an amino acid located between the predicted fifth and sixth transmembrane domain, and an aspartic acid–to-asparagine substitution at codon 543 (D543N) in the predicted cytoplasmic carboxyl-terminal end of the protein. The alanine-to-valine substitution is a conservative substitution and is not expected to influence the function of the NRAMP1 protein. However, the substitution of a negatively charged aspartic acid by an uncharged asparagine residue could affect function of the protein. The homologous codon at the corresponding position in the mouse Nramp1 protein codes for a glutamic acid (Vidal et al. 1993), another amino acid with a negatively charged side chain. This provides support for the functional relevance of a negatively charged residue at codon 543. Additional knowledge of the function of NRAMP1 protein is required for inference of the effects of these missense mutations.

Heterozygosities for the nine polymorphisms detected in the *NRAMP1* gene were low or moderate, with the highest expected value, .47, being seen for the 1465 −85G/A polymorphism in intron 13. Two microsatellite markers, *D2S104* and *D2S173*, were found to be closely linked to the *NRAMP1* region by physical mapping. The expected heterozygosities of *D2S104* and *D2S173* were .72 and .70, respectively (Spurr et al. 1994). The combination of these two highly informative microsatellite markers in the vicinity of the *NRAMP1* gene and the *NRAMP1* sequence variants should greatly facilitate the study of the role of the *NRAMP1* gene in disease susceptibility.

"The problem of the extent to which genetic factors enter into susceptibility to tuberculosis is one of the oldest in human genetics" (Neel and Shull 1954, p. 292). Infection by *Mycobacterium tuberculosis* results in a wide spectrum of phenotypic manifestations, ranging from skin-test sensitivity to purified protein derivative in individuals without detectable tubercle bacilli to fully developed pulmonary disease characterized by multibacillary granuloma that may cause death if the infection is left untreated (Lenzini et al. 1977). The clinical phenotype is almost certainly modulated by acquired immune responses. Among immunologically competent individuals, the quality and quantity of the specific antitubercle immune response depend on genetic factors, prior exposure to *M. tuberculosis* or other related environmental mycobacteria, and/or BCG vaccination (Dubos and Dubos 1987; Sifford and Bates 1991). Thus it appears from the complex etiology of tuberculosis that inadequate classification of tuberculosis patients with respect to clinical, vaccination, and exposure history will result in both etiologic heterogeneity and loss of power either to detect a major susceptibility gene or to confirm a candidate gene such as *NRAMP1*. We failed to detect a statistically significant association between *NRAMP1* variants and tuberculosis disease status in two samples of unrelated individuals. This could indicate that there is no predominant predisposing mutation in these families. However, our sample sizes were small, and both cases and controls were selected on the basis of being parents or grandparents of multiplex tuberculosis families. In addition, incomplete penetrance and sporadic cases of tuberculosis would further reduce the power to detect an association. A recessive susceptibility model with liability classes based on clinical status, age, exposure, and BCG vaccination (Boothroyd 1994) will be used in linkage analysis of *NRAMP1* and tuberculosis susceptibility. The *NRAMP1* variants described in the present report will also be useful for the genetic analysis of human

susceptibility to leprosy, typhoid fever, and Kala Azar, in analogy to susceptibility to intracellular parasites under *Bcg/Lsh/Ity* control in the mouse (for review, see Blackwell 1989; Blackwell et al. 1991; Shaw et al. 1993).

Numerous studies have established mouse *Nramp1* as an important regulatory element in the pathways of macrophage differentiation (Buschman et al. 1989; Blackwell et al. 1991). This suggests human *NRAMP1* as a candidate disease gene for a large number of genetic defects involving the macrophage, most notably autoimmune diseases (Ivanyi 1994). For example, genetic studies in NOD mice have mapped a diabetes susceptibility gene to the vicinity of *Bcg* (*Nramp1*) on proximal mouse chromosome 1 (Cornall et al. 1991), suggesting that human *NRAMP1* may act as a diabetes susceptibility gene. In addition, *NRAMP1* may play a role in the efficacy of BCG immunotherapy of patients with bladder cancer. Of the patients who suffered from superficial bladder cancer and were eligible for BCG immunotherapy, only 30%–50% responded to BCG immunotherapy (Morales et al. 1992). Failure of BCG treatment has been linked to an inefficient inflammatory response to instilled BCG (Torrence et al. 1988), which implies that *NRAMP1* may control the response to BCG immunotherapy. Thus, the study of *NRAMP1* expression and mutation analysis can be applied to a wide range of human diseases in addition to infectious diseases.

Acknowledgments

The blood samples from tuberculosis families were obtained as part of an ongoing collaboration with Drs. Anne Fanning, David Higgins, and Mark Miller. We thank Isabel Anacleto, Lucy Boothroyd, Leah Simkin, and Jian-Xue Wang for technical assistance. This work was supported by grants from the Medical Research Council of Canada (MRC), the Canadian Genetic Diseases Network (Federal NCE program), and the National Institutes of Health (grant RO1 AI35237-02). J.L. is supported by a studentship from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche, F.O.S. was supported by a fellowship from WHO/TDR/UNDP, N.T.B. is the recipient of an MRC studentship, P.G. is supported by an E.W.R. Steacie Memorial Fellowship from the Natural Sciences and Engineering Research Council of Canada and is an International Research Scholar of the Howard Hughes Medical Institute, and E.S. is an MRC scholar.

References

Beaudet AL, Tsui L-C (1993) A suggested nomenclature for designating mutations. Hum Mutat 2:245–248

Blackwell J (1989) The macrophage resistance gene *Lsh/Ity/Bcg*. Res Immunol 140:767–828

Blackwell JM, Roach TI, Atkinson SE, Ajioka JW, Barton CH, Shaw M-A (1991) Genetic regulation of macrophage priming/activation: the *Lsh* gene story. Immunol Lett 30(2): 241–248

Boothroyd LJ (1994) Genetic susceptibility to tuberculosis. MS thesis, McGill University, Montreal

Buschman E, Taniyama T, Nakamura R, Skamene E (1989) Functional expression of the *Bcg* gene in macrophages. Res Immunol 140:793–797

Cellier M, Govoni G, Vidal S, Kwan T, Groulx N, Liu J, Sanchez F, et al (1994) Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization and tissue-specific expression. J Exp Med 180:1741–1752

Cohen D, Chumakov I, Weissenbach J (1993) A first-generation physical map of the human genome. Nature 366:698–701

Cornall RJ, Prins JB, Todd JA, Pressey A, Delarato NH, Wicker LS, Peterson LB (1991) Type 1 diabetes in mice is linked to the interleukin-1 receptor and *Lsh/Ity/Bcg* genes on chromosome 1. Nature 353:262–266

Dosik JK, Barton CH, Holiday DL, Krall MM, Blackwell JM, Mock BA (1994) An *Nramp*-related sequence maps to mouse chromosome 17. Mamm Genome 5:458–460

Dubos R, Dubos J (1987) The white plague. Rutgers University Press, New Brunswick and London

Feinberg AP, Vogelstein B (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Gruenheid S, Cellier M, Vidal S, Gros P. Identification and characterization of a second mouse *Nramp* gene. Genomics (in press)

Ivanyi J (1994) Molecular biology of natural resistance-associated macrophage protein. Parasitol Today 10:416–477

Jurka J, Milosavljevic A (1991) Reconstruction and analysis of human Alu genes. J Mol Evol 32:105–121

Kallmann FJ, Reisner D (1943) Twin studies on the genetic significance of genetic factors in tuberculosis. Am Rev Tuberc 47:549–574

Lenzini L, Rottoli P, Rottoli L (1977) The spectrum of human tuberculosis. Clin Exp Immunol 27:230–237

Liu J, Cellier M, Schurr E, Skamene E (1993) Comparative genome analysis—a novel strategy to study the genetics of host-parasite interaction. J Parasitol 79(4): 463–469

Liu J, Stanton VP, Fujiwara TM, Wang JX, Rezonzew R, Crumley J, Morgan K, et al. Large scale cloning of human chromosome 2 specific yeast artificial chromosomes (YACs) using an interspersed-repetitive-sequences (IRS) PCR approach. Genomics (in press)

Lu-Kuo J, Ward DC, Spritz RA (1993) Fluorescence *in situ* hybridization mapping of 25 markers on distal human chromosome 2q surrounding the human Waardenburg syndrome, type 1(WS1) locus (*PAX3* gene). Genomics 16:173–179

Mah MW, Fanning EA (1991) An epidemic of primary tuberculosis in a Canadian community. Can J Infect Dis 2:133–141

Malo D, Vidal S, Leiman JH, Ward DC, Gros P (1993) Physical delineation of the minimal chromosomal segment encompassing the murine host resistance locus *Bcg*. Genomics 17:667–675

Malo D, Vogan K, Vidal S, Hu J, Cellier M, Schurr E, Fuks A, et al (1994) Haplotype mapping and sequence analysis

of the *Nramp* gene predict susceptibility to infection with intracellular parasites. Genomics 23:51–61

Marchuk D, Drumm M, Saulino A, Collins FS (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res 19:1154

Morales A, Nickel JC, Wilson JWL (1992) Dose response of Bacillus Calmette-Guérin in the treatment of superficial bladder cancer. J Urol 147:1256–1258

Nadeau JH, Davisson MT, Doolittle DP, Grant P, Hillyard AL, Kosowsky M, Roderick TH (1992) Comparative map for mice and humans. Mamm Genome 3:480–536

Neel JV, Shull WJ (1954) Human heredity. University of Chicago Press, Chicago

Rousseau-Merck MF, Simon-Chazottes D, Arpin M, Pringault E, Louvard D, Guenet JL, Berger R (1988) Localization of the villin gene on human chromosome 2q35-q36 and on mouse chromosome 1. Hum Genet 78:130–133

Sambrook J, Fritsch F, Maniatis T (1989) Molecular cloning: a laboratory manual, 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schurr E, Buschman E, Malo D, Gros P, Skamene E (1990*a*) Immunogenetics of mycobacterial infections: mouse-human homologies. J Infect Dis 161:634–641

Schurr E, Skamene E, Forget A, Gros P (1989) Linkage analysis of the *Bcg* gene on mouse chromosome 1: identification of a tightly linked marker. J Immunol 141:4507–4513

Schurr E, Skamene E, Morgan K, Chu ML, Gros P (1990*b*) Mapping of *Col3a1* and *Col6a3* to proximal murine chromosome 1 identifies conserved linkage of structural protein genes between murine chromosome 1 and human chromosome 2q. Genomics 8:477–486

Searle AG, Edwards JH, Hall JG (1994) Mouse homologues of human hereditary disease. J Med Genet 31:1–19

Shaw M-A, Atkinson S, Dockrell H, Hussain R, Lins-Lainson Z, Shaw J, Ramos F, et al (1993) An RFLP map for 2q33-q37 from multicase mycobacterial and leishmanial disease families: no evidence for an *Lsh/Ity/Bcg* gene homologue influencing susceptibility to leprosy. Ann Hum Genet 57:251–271

Sifford M, Bates JH (1991) Host determinants of susceptibility to Mycobacterium tuberculosis. Semin Respir Infect 6:44–50

Spurr NK, Barton H, Bashir R, Bryson GM, Bushby K, Cox S, Gringrich JC, et al (1994) Report of the third international workshop on human chromosome 2 mapping 1994. Cytogenet. Cell Genet 67:216–244

Torrence RJ, Kavoussi LR, Catalona WJ, Ratliff TL (1988) Prognostic factors in patients treated with intravesical bacillus Calmette-Guérin for superficial bladder cancer. J Urol 139(5): 941–944

Vidal SM, Malo D, Vogan K, Skamene E, Gros P (1993) Natural resistance to infection with intracellular parasites: isolation of a candidate for *Bcg*. Cell 73:469–485

Vogel F, Motulsky AG (1986) Human genetics: problems and approaches. Springer, Berlin

World Health Organization (1994) TB: a global emergency. WHO report on the TB epidemic. WHO/TB/94.177

CHAPTER 5:
GENERAL DISCUSSION

I. Large scale YAC mapping of human chromosome 2

5.1 Strategy for the YAC mapping of chromosome 2

The selection of a suitable mapping strategy is essential for the success of a mapping project. Two major factors that one generally takes into account for the selection of a strategy are cost and efficiency of map generation. We chose IRS-PCR screening as the primary mapping method and STS content mapping as additional approach to complement the IRS-PCR method based on the following considerations.

First, the IRS-PCR approach allows the rapid and cost-effective generation of large numbers of markers. Assuming that the physical length of chromosome 2 is 250 Mbp, it would require the generation of at least 2,500 markers to cover the entire chromosome to create a map with an average marker density of 100 kb/marker. The generation of this number of STSs would be a laborious, time- and material- consuming effort since many steps are involved in STS marker generation, including DNA sequencing and sequence analysis, primer selection and synthesis, PCR optimization and data analysis. Nevertheless, since STS content mapping is a PCR-based technique which makes the screening process more amenable to automation, and STS markers can be derived for any unique sequence in the genome, this method is the most commonly used approach for YAC mapping of human chromosomes. In addition, for several human chromosomes, a large collection of STSs has been generated independent of the YAC mapping projects which can be used for YAC cloning. Unfortunately, at the initiation of our experiments only a small number of chromosome 2-specific STS markers were available. Thus, taking an STS content mapping strategy as the primary approach for YAC mapping of chromosome 2 would have meant that at least at the early phase of the project a significant amount of effort and time would have been required for STS marker generation.

In contrast to the tedious process of STS marker generation, the generation of a large number of IRS-PCR probes only requires cloning of IRS-PCR products into T-modified vectors, selection of recombinant clones and fingerprinting of inserts. In our hands, this process is straightforward, fast and cost-effective (Chapter 2). Thus, with respect to marker generation, the IRS-PCR approach is preferable over the STS content

mapping approach. In addition, the hybridization-based IRS-PCR approach can be adapted to high probe throughput screening of pooled YAC libraries. In our laboratory, up to 96 loci were screened in parallel by one person in one week, and, in theory, the screening of a YAC library with 2,500 markers could be done by one person within one year. Thus, even without the highly automated or robotized procedures used by the majority of large genome centers, the IRS-PCR approach enables small laboratories to be competitive in the fast moving field of genome research.

The IRS-PCR approach allowed us to quickly generate a chromosome 2 YAC sublibrary, which formed the basis for YAC contig assembly of chromosome 2 (Chapter 2). However, the major limitation of the IRS-PCR approach is that by this method alone, it is difficult to order YACs (Chapter 3). To order YACs along the chromosome, we chose STS content mapping of the chromosome 2 sublibrary as additional approach. The use of STS content mapping at this stage of the experiments had the benefit that the STS screening could be done in the chromosome 2 sublibrary instead of the total genomic YAC library, greatly reducing the number of PCR reactions and amplification products which needed to be analyzed. Finally, STS content mapping with genetically mapped microsatellites and cytogenetically mapped ESTs allowed the integration of genetic and cytogenetic maps into YAC contig maps, a result which is of particular importance for a better understanding of the genomic organization of chromosome 2.

5.2 Estimation of YAC coverage of the telomeric half of chromosome 2q

The estimation of YAC coverage at different stages of the mapping process provides an important indication for the progress towards assembly of a single YAC contig covering chromosome 2. Since most of my mapping efforts were focused on the telomeric half of chromosome 2q, I will attempt to estimate the YAC coverage for this region of the chromosome.

To estimate the extent to which the telomeric half of chromosome 2q is contained in the chromosome 2 YAC sublibrary, the results of the STS content mapping of the sublibrary can be used. Of the 87 STSs used to screen the sublibrary, 67 (87%) detected at least one YAC in the sublibrary. Assuming that these STSs were distributed randomly

along the telomeric half of chromosome 2q, an estimated 87% of the chromosomal segment would have been cloned in the sublibrary. However, this assumption is likely incorrect since the STSs used are not randomly distributed (Chapter 3; Gyapay et al., 1994; Bell et al., 1995), and we observed that the distributions of IRS-PCR probes and STS loci are to some extend complementary. While clustering of STS markers on sublibrary YACs could result in an overestimate of the cloned chromosome 2 region, an inverse distribution of IRS and STS markers would suggest that the coverage of the telomeric half of chromosome 2q segments in the sublibrary could be higher than 87% since the sublibrary was generated based on IRS-PCR screening. Indeed, this conclusion is consistent with the following estimate. In total, there are 223 YAC groups in our dataset, of which 65 contain probes are derived from hybrid 014-1. This hybrid is estimated to carry an approximate 70 Mbp segment of the telomeric half of the long arm. Assuming the remaining 158 YAC groups are evenly distributed along the chromosome, approximately 40 of them should overlap with the chromosomal segment contained in hybrid 014-1. Based on the average YAC size of 450 kb and the contig depth of 6.7 for the chromosome 2 sublibrary, we estimate the average genomic coverage for a one probe contig (the smallest contig) to be approximately 700 kb. Assuming each YAC group to be represented by one probe only, the genomic coverage of a 70 Mbp segment by 105 YAC groups would be close to 100%.

Next, I estimate the chromosomal coverage by the YAC groups that can be ordered along the telomeric half of chromosome 2q. In our dataset, 44 YAC groups are anchored by genetically or cytogenetically mapped STS(s) and thus can be ordered. The minimum genomic coverage for these groups would be 31 Mbp (assuming a minimum coverage of 0.7 Mbp/contig; see above), which corresponds to approximately one-fourth of chromosome 2q. After integration of our data with the WI/MIT Genome Center data, there are 14 integrated groups which contain 39 YAC groups and 19 WI/MIT WC-groups on chromosome 2q. These 14 integrated groups are all anchored by genetic markers and can be ordered. Unfortunately, there is no simple and accurate method to estimate the chromosome coverage by these 14 integrated groups. However, it is possible to estimate the minimal coverage for 6 of them based on the genetic markers from the Gyapay map

which were contained in each group. Using this approach, the total minimal coverage for these 6 groups that can be estimated as 36 cM. For the remaining 8 integrated groups, the minimal genomic coverage could not be determined because only one genetic marker from the Gyapay map was contained in each group.

5.3 YAC contigs

A YAC contig is defined as a set of ordered YAC clones carrying overlapping inserts. One example of such a contig is shown in Chapter 2 (Fig. 7). However, unless a dataset which is subjected to contig analysis is free of errors, the contigs generated are unlikely to represent the true order of YACs. Generally, most errors in YAC mapping data are caused by false positive or false negative clones, markers that contain repeat sequences, and chimeric YACs. The error rate of primary mapping data appears to be similar for STS content mapping or IRS-PCR screening of YAC libraries (Chumakov et al., 1992; Foote et al., 1992; Bell et al., 1995; Green et al., 1995). Thus, additional methodologies are needed to establish true orders of YACs. Currently, two approaches which are either clone-oriented or marker-oriented have been used to overcome the difficulty caused by incorrect data and to establish true YAC contigs. When the clone-oriented approach is used, a fingerprint is created for each YAC within a YAC group and an order of overlaps among YACs is then established by comparing the fingerprints (Bellanne-Chantelot et al., 1992). As described in Chapter 1, fingerprint analysis of overlapping clones can be laborious and a high percentage of YAC chimerism can further complicate the analysis. Alternatively, in the marker-oriented approach, an order of markers is established first and each clone is then positioned along these ordered markers. Using the marker-oriented approach to establish contigs for YAC groups likely requires less effort than fingerprinting of YACs since STS markers can be ordered efficiently by using the whole genome-radiation hybrid (WG-RH) mapping panel (Walter et al., 1994). Moreover, ordering of markers avoids the problem caused by chimeric YACs. Most importantly, however, an order of markers is independent of clone libraries and the same order of markers can be used to position clones from YAC, BAC or PAC libraries.

5.4 Ambiguity

Currently, most YAC libraries, including the CEPH YAC libraries, are arranged in 96-well plates, i.e. individual YAC clones are stored in wells and identified by their "addresses" which are composed of plate, row and column coordinates. Most YAC libraries contain large numbers of clones and screening of individual clones either by PCR or by hybridization would be a tremendous task. Thus, to increase the efficiency of library screening, different multi-dimensional pooling schemes of library YACs have been developed. For example, we used a pooling scheme for the CEPH Mark I YAC library that enabled us to identify YACs by testing only 1,152 pools instead of 55,296 individual clones. The efficiency of YAC library screening is thus increased 48 fold (Chapter 2). However, several problems are associated with YAC pooling (Bell et al., 1995; Green et al., 1995). First, cross well contamination during preparation of the pools can create false positives during YAC library screening. Second, absence of a positive signal in only one of the dimensions can result in false negatives. Third, the presence of more than one positive YAC in a pool can result in ambiguous YAC addresses (Chapter 2). Such ambiguities are a major limitation for the use of pooled YACs in genome analysis (Cohen et al., 1993; Bell et al., 1995; Green et al., 1995), and ambiguous addresses are by far the most common problem in our study. Currently, the chromosome 2 YAC sublibrary consists of close to 4,500 YACs, of which approximately 2,000 were directly shown to contain chromosome 2-specific inserts. A large proportion of the remaining YACs are non-chromosome 2 specific YACs identified as part of ambiguous sets of addresses. Such non-chromosome 2 YACs could potentially create connections between YAC groups which are not adjacent on the chromosome. To avoid this problem, we only included YACs in the YAC group analysis which had been demonstrated to be chromosome 2-specific. The resulting YAC groups thus are based on incomplete mapping information, i.e. these YAC groups miss the true chromosome 2 YACs contained in sets of ambiguous YACs which can not be resolved. Such incomplete data can contribute to the incapability of extending true contigs. In our study, a number of techniques have been used to resolve the ambiguity of YAC addresses, including screening of individual YAC clones with pooled IRS-PCR probes detecting non-overlapping sets of ambiguous addresses (Chapter

2), and STS screening of individual clones in the chromosome 2 YAC sublibrary (Chapter 3). The screening of individual YACs is the most effective approach for resolving ambiguities, but unfortunately also requires the most effort. In addition, we also disambiguated a subset of ambiguously detected YACs by matching ambiguous with unambiguous addresses (Chapter 3). Together, the above techniques for disambiguating ambiguous YAC addresses have significantly reduced the proportion of ambiguous YAC clones in the chromosome 2 sublibrary.

5.5 IRS-PCR probes

During the course of our study, we have generated approximately 2,000 distinct chromosome 2-specific IRS-PCR probes, and over 1,000 probes were used to screen YAC libraries. The probes used can be classified into three categories based on the results of the screening of YAC libraries. The first category includes probes that identified an average of 6.7 YACs in the library and thus hybridized to one or few specific sequence(s) on chromosome 2. As described in Chapter 2, close to 90% of IRS-PCR probes belong in this category. The second category includes probes that hybridized to a large number of YAC pools. These probes likely contain a high-copy number of repeat sequences. Since all probes correspond to inter-*Alu* sequences, it would be interesting to compare the sequence of these probes with DNA database sequences. Perhaps such an analysis would reveal novel repeats in the human genome. The third category includes probes that failed to produce any positive signals. One explanation for the failure of the IRS-PCR screening is that a fraction of inter-*Alu* fragments contained in YAC pool DNA are not amplified. Due to the complexity of YAC DNA pools (most pools contain 288 YACs) and the competition for amplification of different IRS segments during the PCR process, not all IRS segments present in YAC pool DNA are represented in the IRS-PCR products spotted on the hybridization membranes, resulting in the lack of hybridization signals for some IRS-PCR probes during library screening. An alternative explanation is that the failure of IRS-PCR screening could be caused by genomic *Alu* polymorphisms, i.e. differences in the number and location of *Alu* elements in the DNA used to create the YAC library and the DNA used for the generation of probes (Zietkiewicz et al., 1992).

For YAC group analysis and subsequent YAC contig analysis, probes with a low-copy number repeat are a problem. Without detailed analysis of the IRS-PCR probes, it is difficult to distinguish single copy probes which represent a unique landmark in the genome from probes with a low-copy number of repeats which represent multiple locations in the genome. For example, probe C77(L) detected 7 Mark I YACs and 12 mega YACs. Further analysis revealed that probe C77(L) detected two groups of YACs overlapping the well separated *VIL1* and the *D2S159* regions. However, low copy repeat probes can be detected by including the IRS-PCR products of the National Institute of General Medical Sciences (NIGMS) monochromosomal hybrid panel 2 and the IRS-PCR products of a reduced chromosome 2 hybrid panel as controls in the hybridization experiments. If a probe only detects the monochromosomal hybrid containing chromosome 2 and a set of hybrids that carry a common region of chromosome 2, it is unlikely that this probe contains repeat sequences.

In conclusion, the combined IRS-PCR and STS content mapping strategy used in this study has proven to be practical and efficient for large scale physical mapping. The resources created by our experiments, such as the chromosome 2-specific YAC library, the large number of IRS-PCR probes and the YAC groups anchored by genetically or cytogenetically mapped markers, are useful reagents for studying the genomic structure and organization of chromosome 2, and will likely prove very useful for the mapping and isolation of disease genes located on this chromosome.

II. Genetic studies of innate susceptibility to tuberculosis and leprosy

5.6 Comparative mapping and genome-wide mapping

The murine *Bcg* gene has long been considered a pivotal element in the control of innate susceptibility to mycobacterial infections. In the late 1980s when the mapping of a human tuberculosis and leprosy susceptibility gene was initiated, it was already known that a 30 cM chromosomal segment of proximal mouse chromosome 1, including murine *Bcg*, was syntenic with the telomeric end of human chromosome 2, region q32-qter (Schurr et

al., 1990b). Therefore it was reasonable to believe that a human homologue of *Bcg* would also be found on the evolutionary conserved telomeric half of chromosome 2q. Thus, this region provided a target for the search of a human tuberculosis and leprosy susceptibility gene. A candidate gene approach which focused on chromosome 2, region q32-qter, was a practical choice since the density of genetic markers across this region of the genome was low at that time and a genome-wide search for the tuberculosis and leprosy susceptibility gene therefore would have been very tedious.

While the comparative candidate gene mapping strategy is a time saving approach when compared with whole genome screening strategies, the extent to which experimental species, such as the mouse, are similar to humans is still an open question. Especially since susceptibility to tuberculosis and leprosy is a complex trait which is likely characterized by incomplete penetrance of susceptibility alleles and by the occurrence of phenocopies. In contrast, the mouse *Bcg* gene model is free of these complications. Human *NRAMP1* -the human *Bcg* homologue gene - was recently mapped on chromosome 2, region q35, as predicted by mouse-human comparative maps. However, if *NRAMP1* is a major tuberculosis and leprosy susceptibility gene is still under investigation. Data from our group indicate that the human *NRAMP1* gene does play a role in susceptibility to tuberculosis and leprosy diseases for some ethnic groups but not for others (Boothroyd, 1994; Levee et al., 1994; Sanchez et al., unpublished). In contrast, results obtained by other groups have excluded *NRAMP1* as a susceptibility gene for leprosy for some genetic models (White et al., 1994). These results suggest that several loci may control susceptibility to tuberculosis and leprosy, and that the relative importance of these loci for expression of the susceptibility phenotype may vary among different ethnic groups. Where do these additional tuberculosis and leprosy susceptibility genes map? One approach which possibly will lead to the answer of the above question is a genome-wide search. A large set of microsatellites which can be rapidly typed by PCR is now available for the whole genome, which provides the basis for a genome-wide search for the tuberculosis and leprosy susceptibility loci. In fact, an increasing number of disease genes including a collection of diabetes susceptibility genes have been mapped by the genome-wide mapping approach (Field et al., 1994; Davies et al., 1994). It is reasonable

to expect that the same genome-wide search for the tuberculosis and leprosy susceptibility loci will tell us more about the genetic basis for susceptibility to tuberculosis and leprosy disease.

5.7 Genetic markers for the study of innate susceptibility to tuberculosis and leprosy

One essential tool to test for linkage of a given phenotypic trait with specific regions of the human genome is a set of highly polymorphic genetic markers. Until recently, the creation of such markers has been a very tedious process. In the case of the genetic study of susceptibility to tuberculosis and leprosy, an extensive effort has been made to generate genetic markers on chromosome 2, region q33-q37 (Shaw et al., 1993; Levee et al., 1994). The first set of markers we generated were RFLP markers derived from a number of expressed loci located on chromosome 2, region q33-q37 (Levee et al., 1994). Most of these markers are di-allelelic and of low heterozygosity, and they are derived from genes with known cytogenetic locations. The second set of markers we generated are RFLP-PCR markers directly derived from the human *NRAMP1* gene (Chapter 4). Again, most of these markers are di-allelelic and have a relative low heterozygosity. However, if these *NRAMP1* markers are combined, the resulting haplotypes are much more informative and provide a useful tool to test the role of the *NRAMP1* gene in susceptibility to tuberculosis and leprosy. In addition, we have identified 17 highly informative microsatellites which were physically linked with 16 expressed genes on the telomeric half of chromosome 2q (Chapter 3). Among them, two microsatellites, *D2S104* and *D2S173*, were found to be linked to *NRAMP1* within a 1.5 Mbp YAC contig (Chapter 4). The establishment of physical linkage of a set of microsatellites with expressed genes on the telomeric half of chromosome 2q achieved in this study not only provides additional highly informative markers for further genetic analysis of susceptibility to tuberculosis and leprosy, but also for linkage analysis of other disease traits located in this region of the genome including a type 1 diabetes susceptibility gene (*IDDM7*) (Copeman et al., 1995) and the amyotrophic lateral sclerosis type 2 (ALS2) locus (Hentati et al., 1994). More importantly, they allow the integration of cytogenetic, genetic mapping information with a YAC-based physical map, a result that is

of significance for a better understanding of the genomic structure and organization of chromosome 2q.

REFERENCES

Abel, L., Demenais, F. (1988). Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. Am. J. Hum. Genet. 42, 256-266.

Abel, L., Lap, V.D., Oberti, J., Thuc, N.V., Cua, V.V., Guilloud-Bataille, M., Schurr, E., Lagrange, P.H. (1995). Complex segregation analysis of leprosy in Southern Vietnam. Genetic Epidemiology 12, 63-82.

Abidi, F.E., Wada, M., Little, R.D., Schlessinger, D. (1990). Yeast artificial chromosomes containing human Xq24-Xq28 DNA: library construction and representation of probe sequences. Genomics 7, 363-376.

Albertsen, H.M., Abderrahim, H., Cann, H.M., Dausset, J., Le Paslier, D., Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. Proc. Natl. Acad. Sci. USA 87, 4256-4260.

Anand, R., Riley, J.H., Butler, R., Smith, J.C., Markham, A.F. (1990). A 3.5 genome equivalent multi access YAC library: construction, characterization, screening and storage. Nucleic Acids Res. 18, 1951-1956.

Antignac, C., Arduy, C.H., Beckmann, J.S., Benessy, F., Gros, F., Medhioub, M., Hildebrandt, F., Dufier, J.-L., Kleinknecht, C., Broyer, M., Weissenbach, J., Habib, R., Cohen, D. (1993). A gene for familial juvenile nephronophthisis (recessive medullary cystic kidney disease) maps to chromosome 2p. Nature Genetics 3, 342-345.

Appelberg, R., Sarmento, A.M. (1990). The role of macrophage activation and of Bcg-encoded macrophage function(s) in the control of Mycobacterium avium infection in mice. Clin. Exp. Immunol. 80, 324-331.

Ardinger, H.H, Buetow, K.H., Bell, G.I., Bardach, J., VanDenmark, D.R., Murray, J.C. (1989). Association of genetic variant of the transforming growth factor-alpha gene with cleft lip palate. Am. J. Hum. Genet. 45, 348-353.

Ariyama, T., Inazawa, J., Ezaki, T., Nakamura, Y., Horri, A., Abe, T. (1995). High-resolution cytogenetic mapping of the short arm of chromosome 1 with newly isolated 411 cosmid markers by fluorescence *in situ* hybridization: the precise order of 18 markers on 1p36.1 on prophase chromosomes and "stretched" DNAs. Genomics 25, 114-123.

Arveiler, B., Porteous, D.J. (1992). Distribution of Alu and L1 repeats in human YAC recombinants. Mamm. Genome 3, 661-668.

Ashworth, L.K., Alegriahartman, M., Burgin, M., Devlin, L., Carrano, A.V., Batzer, M.A. (1995). Assembly of high-resolution bacterial artificial chromosome, P1-derived artificial chromosome, and cosmid contigs. Anal. Biochem. 224, 564-571.

Aslanidis, C., de Jong, P.J. (1991). Coincidence cloning of Alu PCR products. Proc. Natl. Acad. Sci. USA 88, 6765-6769.

Baldini, A., Ward, D.C. (1991). In situ hybridization banding of human chromosomes with Alu-PCR products: a simultaneous karyotype for gene mapping studies. Genomics 9, 770-774.

Baldwin, C.T., Hoth, C.F., Amos, J.A., da-Silva, E.O., Milunsky, A. (1992). An exonic mutation in the HuP2 paired domain gene causes Waardenburg's syndrome. Nature 355, 637-638.

Barber, T.D., Morell, R., Johnson, D.H., Asher, J.H.,Jr., Friedman, T.B. (1993). A highly informative dinucleotide repeat polymorphism at the D2S211 locus linked to ALPP, FN1 and TNP1. Hum. Mol. Genet. 2, 88.

Barr, F.G., Galili, N., Holick, J., Biegel, J.A., Rovera, G., Emanuel, B.S. (1993). Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. Nature Genetics 3, 113-117.

Barrett, J.H. (1992). Genetic mapping based on radiation hybrid data. Genomics 13, 95-103.

Barton, C.H., White, J.K., Roach, T.I.A., Blackwell, J.M. (1994). NH_2-terminal sequence of macrophage-expressed natural resistance-associated macrophage protein (*Nramp*) encodes a proline/serine-rich putative *Src* homology 3-binding domain. J. Exp. Med. 179, 1683-1687.

Barton, D.E., Yang-Feng, T.L., Mason, A.J., Seeburg, P.H., Francke, U. (1989). Mapping of genes for inhibin subunits α, β_a and β_b on human and mouse chromosomes and studies of *jsd* mice. Genomics 5, 91-99.

Bates, G.P., Valdes, J., Hummerich, H., Baxendale, S., Le Paslier, D.L., Monaco, A.P., Tagle, D., MacDonald, M.E., Altherr, M., Ross, M., Brownstein, B.H., Bentley, D., Wasmuth, J.J., Gusella, J.F., Cohen, D., Collins, F., Lehrach, H. (1992). Characterization of a yeast artificial chromosome contig spanning the Huntington's disease gene candidate region. Nature Genet. 1, 180-187.

Baxendale, S., MacDonald, M.E., Mott, R., Francis, F., Lin, C., Kirby, S.F., James, M., Zehetner, G., Hummerich, H., Valdes, J., Collins, F.S., Deaven, L.J., Gusella, J.F., Lehrach, H., Bates, G.P. (1993). A cosmid contig and high resolution restriction map of the 2 megabase region containing the Huntington's disease gene. Nature Genet. 4, 181-186.

Beaudet, A.L., Tsui, L-C. (1993). A suggested nomenclature for designating mutations. Hum. Mutat. 2, 245-248.

Bell, C.J., Budarf, M.L., Nieuwenhuijsen, B.W., Barnoski, B.L., Buetow, K.H., Campbell, K., Colbert, A.M.E., Collins, J., Daly, M., Desjardins, P.R., DeZwaan, T., Eckman, B., Foote, S., Hart, K., Hiester, K., Van Het Hood, M.J., Hopper, E., Kaufman, A., McDermid, H.E., Overton, G.C., Reeve, M.P., Searls, D.B., Stein, L., Valmiki, V.H., Watson, E., Williams, S., Winston, R., Nussbaum, R., Lander, E.S., Fischbeck, K.H., Emanuel, B.S., Hudson, T.J. (1995). Integration of physical, breakpoint and genetic maps of chromosome 22. Localization of 587 yeast artificial chromosomes with 238 mapped markers. Hum. Mol. Genet. 4, 59-69.

Bellanné-Chantelot, C., Barillot, E., Lacroix, B., Le Paslier, D., Cohen, D. (1991). A test case for physical mapping of human genome by repetitive sequence fingerprints: construction of a physical map of a 420 kb YAC subcloned into cosmids. Nucleic Acids Res. 19, 505-510.

Bellanné-Chantelot, C., Lacroix, B., Ougen, P., Billault, A., Beaufils, S., Bertrand, S., Georges, I., Glibert, F., Gros, I., Lucotte, G., Susini, L., Codani, J.-J., Gesnouin, P., Pook, S., Vaysseix, G., Lu-Kuo, J., Ried, T., Ward, D., Chumakov, I., Le Paslier, D., Barillot, E., Cohen, D. (1992). Mapping the whole human genome by fingerprinting yeast artificial chromosomes. Cell 70, 1059-1068.

Bernard, L.E., Brooks-Wilson, A.R., Wood, S. (1991). Isolation of DNA fragments from a human chromosomal subregion by Alu PCR differential hybridization. Genomics 9, 241-246.

Bernard, L.E., Wood, S. (1993). Human chromosome 5 sequence primer amplifies Alu polymorphisms on chromosomes 2 and 17. Genome 36, 302-309.

Blackwell, J. (1989). The macrophage resistance gene *Lsh/Ity/Bcg*. Res. Immunol. 140, 767-828.

Blackwell, J.M., Barton, C.H., White, J.K., Roach, T.I.A., Shaw, M-A., Whitehead, S.H., Mock, B.A., Searle, S., Williams, H., Baker, A-M. (1994). Genetic regulation of leishmanial and mycobacterial infections: the *Lsh/Ity/Bcg* gene story continues. Immunology Letters 43, 99-107.

Blackwell, J.M., Barton, C.H., White, J.K., Searle, S., Baker, A-M., Williams, H., Shaw, M-A. (1995). Genomic organization and sequence of the human NRAMP gene: identification and mapping of a promoter region polymorphism. Molecular Medicine 1, 194-205.

Blackwell, J.M., Roach, T.I.A., Atkinson, S.E., Ajioka, J.W., Barton, C.H., Shaw, M-A. (1991). Genetic regulation of macrophage priming/activation: the *Lsh* gene story. Immunol. Lett. 30, 241-248.

Blunt, T., Taccioli, G.E., Priestley, A., Hafezparast, M., McMillan, T., Liu, J., Cole, C.C., White, J., Alt, F.W., Jackson, S.P., Schurr, E., Lehmann, A.R., Jeggo, P.A. (1995). A YAC contig encompassing the XRCC5 (Ku80) DNA repair gene and complementation of defective cells by YAC protoplast fusion. Genomics. in press.

Boehnke, M., Lange, K., Cox, D.R. (1991). Statistical methods for multipoint radiation hybrid mapping. Am. J. Hum. Genet. 49, 1174-1188.

Boothroyd, L.J. (1994). Genetic susceptibility to tuberculosis. MS thesis, McGill University, Montreal.

Botstein, D., White, R.L., Skolnick, M.H., Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction length polymorphisms. Am. J. Hum. Genet. 32, 314-331.

Bradley, D.J. (1977). Genetic control of *Leishmania* populations within the host. II. Genetic control of acute susceptibility of mice to *L. donovani* infections. Clin. Exp. Immunol. 30, 130-140.

Britten, R.J., Baron, W.F., Stout, D.B., Davidson, E.H. (1988). Sources and evolution of human Alu repeated sequences. Proc. Natl. Acad. Sci. USA 85, 4770-4774.

Bronson, S.K., Pei, J., Taillon-Miller, P., Chorney, M.J., Geraghty, D.E., Chaplin, D.D. (1991). Isolation and characterization of yeast artificial chromosome clones linking the HLA-B and HLA-C loci. Proc. Natl. Acad. Sci. USA 88, 1676-1680.

Brooks-Wilson, A.R., Goodfellow, P.N., Povey, S., Nevanlinna, H.A., de Jong, P.J., Goodfellow, P.J. (1990). Rapid cloning and characterization of new chromosome 10 DNA markers by *Alu* element-mediated PCR. Genomics 7, 614-620.

Brownstein, B.H., Silverman, G.A., Little, R.D., Burke, D.T., Korsmeyer, S.J., Schlessinger, D., Olson, M.V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. Science 244, 1348-1351.

Buckler, A.J., Chang, D.D., Graw, C.S., Brook, D., Haber, D.A., Sharp, P.A., Housman, D.E. (1991). Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. Proc. Natl. Acad. Sci. USA 88, 4005-4009.

Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Duyk, G.M., Sheffield, V.C., Wang, Z., Murray, J.C. (1994). Integrated human genome-wide maps constructed using the CEPH reference panel. Nature Genet. 6, 391-393.

Burke, D.T. (1991). The role of yeast artificial chromosome clones in generating genome maps. Curr. Opin. Gen. Develop. 1, 69-74.

Burke, D.T., Carle, G.F., Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236, 806-812.

Buschman, E., Schurr, E., Gros, P. and Skamene, E. *Microbial determinants of virulence and host response*, ASM Publication, 1990. pp. 93-111.

Buschman, E., Taniyama, T., Nakamura, R., Skamene, E. (1989). Functional expression of the *Bcg* gene in macrophages. Res. Immunol. 140, 793-797.

Buu, N.T., Cellier, M., Gros, P., Schurr, E. (1995). Identification of a highly polymorphic length variant in the 3'UTR of *NRAMP1*. Immunogenetics. in press.

Cangiano, G., Ameer, H., Waterston, R., La Volpe, A. (1990). Use of repetitive DNA probes as physical mapping strategy in *Caenorhabditis elegans*. Nucleic Acids Res. 18, 5077-5081.

Cellier, M., Govoni, G., Vidal, S., Kwan, T., Groulx, N., Liu, J., Sanchez, F., Skamene, E., Schurr, E., Gros, P. (1994). Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression. J. Exp. Med. 180, 1741-1752.

Cellier, M., Privé, G., Belouchi, A., Kwan, T., Rodrigues, V., Chia, W., Gros, P. (1995). The natural resistance associated macrophage protein (NRAMP) defines a new family of membrane proteins conserved through evolution. Proc. Natl. Acad. Sci. USA. in press.

Chang, E., Welch, S., Luna, J., Giacalone, J., Francke, U. (1993). Generation of a human chromosome 18-specific YAC clone collection and mapping of 55 unique YACs by FISH and fingerprinting. Genomics 17, 393-402.

Chenevix-Trench, G., Jones, K., Green, A.C., Duffy, D.L., Martin, N.G. (1992). Cleft lip with or without cleft palate: associations with transforming growth factor alpha and retinoic acid receptor. Am. J. Hum. Genet. 51, 1377-1385.

Cherif, D., Julier, C., Delattre, O., Derre, J., Lathrop, G.M., Berger, R. (1990). Simultaneous localization of cosmids and chromosome R-banding by fluorescence microscopy: Application to regional mapping of human chromosome 11. Proc. Natl. Acad. Sci. USA 87, 6639-6643.

Chiu, I.M., Gilmore, E.C., Liu, Y., Payson, R.A. (1994). Construction of a yeast artificial chromosome contig encompassing the human acidic fibroblast growth factor (FGF1) gene: toward the cloning of the ANLL/MDS tumor-suppressor gene. Genomics 19, 552-560.

Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billaut, A., Guasconi, G., Gervy, P., LeGall, I., Soularue, P., Grinas, L., Bougueleret, L., Bellanné-Chantelot, C., Lacroix, B., Barillot, E., Gesnouin, P., Pook, S., Vaysseix, G., Frelat, G., Schmitz, A., Sambucy, J.-L., Bosch, A., Estivill, X., Weissenbach, J., Vignal, A., Riethman, H., Cox, D., Patterson, D., Gardiner, K., Hattori, M., Sakaki, Y., Ichikawa, H., Ohki, M., Le Paslier, D., Heilig, R., Antonarakis, S., Cohen, D. (1992a). Continuum of overlapping clones spanning the entire human chromosome 21q. Nature 359, 380-387.

Chumakov, I.M., Le Gall, I., Billault, A., Ougen, P., Soularue, P., Guillou, S., Rigault, P., Bui, H., de Tand, M.-F., Barillot, E., Abderrahim, H., Cherif, D., Berger, R., Le Paslier, D., Cohen, D. (1992b). Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. Nature Genetics 1, 222-225.

Chumakov, I.M., Rigault, P., Le Gall, I., BellannÄ-Chantelot, C., Billault, A., Guillou, S., Soularue, P., Guasconi, G., Poullier, E., Gros, I., Belova, M., Sambucy, J-L., Susini, L., Gervy, P., Glibert, F., Beaufils, S., Bui, H., Catherine, M., De Tand, M-F., Dukasz, F., Lecoulant, S., Ougen, P., Perrot, V., Saumier, M., Doravito, C., Bahouayila, R., Cohen-Akenine, A., Barillot, E., Bertrand, S., Codani, J-J., Caterina, D., Georges, I., Lacrox, B., Lucotte, G., Sahbatou, M., Schmit, C., Sangouard, M., Tubacher, E., Dib, C., FaurÄ, S., Fizames, C., Gyapay, G., Millasseau, P., Nguyen, S., Muselet, D., Vignal, A., Morissette, J., Menninger, J., Lieman, J., Desai, T., Banks, A., Bray-Ward, P., Ward, D., Hudson, T., Berety, S., Foote, S., Stein, L., Page, D.C., Lander, E.S., Weissenbach, J., Le Paslier, D., Cohen, D. (1995). A YAC contig map of the human genome. Nature 377, 175-.

Chinault, A.C., Sternberg, N.L. (1994). Sreening of large-insert libraries by PCR. In *Current protocols in human genetics*, N.C. Dracopoli, J.L. Haines, B.R. Korf, D.T. Moir, C.C. Morton, C.E. Seidman, J.G. Seidman and D.R. Smith, eds. (A joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc), pp5.5.1

Cohen, D., Chumakov, I., Weissenbach, J. (1993). A first-generation physical map of the human genome. Nature 366, 698-701.

Cohen-Haguenauer, O., Barton, P.J.R., Van Cong, N., Serero, S., Gross, M.-S., Jegou-Foubert, C., de Tand, M.-F., Robert, B., Buckingham, M., Frezal, J. (1988). Assignment of the human fast skeletal muscle myosin alkali light chains gene (MLC1F/MLC3F) to 2q32.1-2qter. Hum. Genet. 78, 65-70.

Cohen I.H., Chan, H.S., Track, R.K., Kidd, K.K. (1990). The human gene map (*Homo Sapiens*) (2N = 46) as of HGM10. In *Genetic Maps*, S.J.O'Brien editor, Cold Spring Harbor Laboratory Press, pp5.6

Cole, C.G., Goodfellow, P.N., Borrow, M., Bentley, D.R. (1991). Generation of novel sequence tagged sites (STSs) from discrete chromosomal regions using *Alu*-PCR. Genomics 10, 816-826.

Collins, F., Galas, D. (1993). A new five-year plan for the U.S. Human Genome Project. Science 262, 43-46.

Collins, F.S. (1995). Positional cloning moves from perditional to traditional. Nature Genet. 9, 347-350.

Comstock, G.W. (1978). Tuberculosis in twins: a re-analysis of the Prophit survey. American Review of Respiratory Disease 117, 621-624.

Copeman, J.B., Cucca, F., Hearne, C.M., Cornall, R.J., Reed, P.W., Ronningen, K.S., Undlien, D.E., Nistico, L., Buzzetti, R., Tosi, R., Fociot, F., Nerup, J., Cornelis, F., Barnett, A.H., Bain, S.C., Todd, J.A. (1995). Linkage disequilibrium mapping of a type 1 diabetes susceptibility gene (IDDM7) to chromosome 2q31-q33. Nature Genet. 9, 80-85.

Cornall, R.J., Prins, J-N., Todd, J.A., Pressey, A., DeLarato, L.S., Wicker, L.S., Peterson, L.B. (1991). Type 1 diabetes in mice is linked to the interleukin-1 receptor and Lsh/Ity/Bcg genes on chromosome 1. Nature 353, 262-266.

Coulson, A., Sulston, J., Brenner, S., Karn, J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 83, 7821-7825.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S., Myers, R.M. (1990). Radiation hybrid mapping: a somatic cell genetic method for construction high-resolution maps of mammalian chromosomes. Science 250, 245-250.

Cox, D.R., Green, E.D., Lander, E.S., Cohen, D., Myers, R.M. (1994). Assessing mapping progress in the Human Genome Project. Science 265, 2031-2032.

Dausset, J., Ougen, P., Abderrahim, H., Billault, A., Asmbucy, J.-L., Cohen, D., Le Paslier, D. (1992). The CEPH YAC library. Behring Inst. Mitt. 91, 13-20.

Davies, J.L., Kawaguchi, Y., Bennett, S.T., Copeman, J.B., Cordell, H.J., Pritchard, L.E., Reed, P.W., Gough, S.C.L., Jenkins, S.C., Palmer, S.M., Balfour, K.M., Rowe, B.R., Farrall, M., Barnett, A.H., Bain, S.C., Todd, J.A. (1994). A genome-wide search for human type 1 diabetes susceptibility genes. Nature 371, 130-136.

Deininger, P.L., Schmid, C.W. (1976). An electron microscope study of the DNA sequence organization of the human genome. J. Mol. Biol. 106, 773-790.

Denis, M., Forget, A., Pelletier, M., Turcotte, R., Skamene, E. (1986). Control of the Bcg gene of early resistance in mice to infections with BCG substrains and atypical mycobacteria. Clin. Exp. Immunol. 63, 517-525.

Dosik, J.K., Barton, C.H., Holiday, D.L., Krull, M.M., Blackwell, J.M., Mock, B.A. (1994). An Nramp-related sequence maps to mouse Chromosome 17. Mamm. Genome 5, 458-460.

Drwinga, H.L., Toji, L.H, Kim, C.H., Greene, A.E. (1993). NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. Genomics 16, 311-314.

Dubos, R., Dubos, J. (1987). The white plague. Rutgers University Press, New Brunswick and London.

Epstein, D.J., Bardeesy, N., Vidal, S., Malo, D., Weith, A., Vekemans, M., Gros, P. (1994). Characterization of a region-specific library of microclones in the vicinity of the Bcg and splotch loci on mouse chromosome 1. Genomics 19, 163-166.

Epstein, D.J., Vekemans, M., Gros, P. (1991). Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. Cell 67, 767-774.

Fantes, J.A., Oghene, K., Boyle, S., Danes, S., Fletcher, J.M., Bruford, E.A., Williamson, K., Seawright, A., Schedl, A., Hanson, I., Zehetner, G., Bhogal, R., Lehrach, H., Gregory, S., Williams, J., Little, P.F.R., Sellar, G.C., Hoovers, J., Mannens, M., Weissenbach, J., Junien, C., Heyningen, V.V., Bickmore, W.A. (1995). A high-resolution integrated physical, cytogenetic, and genetic map of human chromosome 11: distal p13 to proximal p15.1. Genomics 25, 447-461.

Feinberg, A.P., Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132(1), 6-13.

Feingold, J.M., Ogden, S.D., Denny, C.T. (1990). Streamlined approach to creating yeast artificial chromosome libraries from specialized sources. Proc. Natl. Acad. Sci. USA 87, 8637-8641.

Field, L.L., Tobias, R., Magnus, T. (1994). A locus on chromosome 15q26 (IDDM3) produces susceptibility to insulin-dependent diabetes mellitus. Nature Genet. 8, 189-194.

Fine, P.E.M. (1981). Immunogenetics of susceptibility to leprosy, tuberculosis, and leishmaniasis. An epidemiological perspertive. Int. J. Leprosy 49, 437-454.

Fishel, R., Lescoe, M.K., Rao, M.R.S., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., Kolodner, R. (1993). The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. Cell 75, 1027-1038.

Foote, S., Vollrath, D., Hilton, A., Page, D.C. (1992). The human Y chromosome: overlapping DNA clones spanning the euchromatic region. Science 258, 60-66.

Foy, C., Newton, V., Wellesley, D., Harris, R., Read, A.P. (1990). Assignment of the locus for Waardenburg syndrome type I to human chromosome 2q37 and possible homology to the splotch mouse. Am. J. Hum. Genet. 46, 1017-1023.

Garza, D., Ajioka, J.W., Burke, D.T., Hartl, D.L. (1989). Mapping the *Drosophila* genome with yeast artificial chromosomes. Science 246, 641-646.

Giblett, E.R. (1969). *"Genetic Markers in Human Blood,"*, Blackwell, Oxford, UK.

Gnirke, A., Barnes, T.S., Patterson, D., Schild, D., Featherstone, T., Olson, M.V. (1991). Cloning and in vivo expression of the human GART gene using yeast artificial chromosomes. EMBO J. 10, 1629-1634.

Godal, R., Negassi, K. (1973). Subclinical infection in leprosy. Brit. Med. J. 13, 557-559.

Goold, R.D., diSibio, G.L., Xu, H., Lang, D.B., Dadgar, J., Magrane, G.G., Dugaiczyk, A., Smith, K.A., Cox, D.R., Masters, S.B., Myers, R.M. (1993). The development of sequence-tagged sites for human chromosome 4. Hum. Mol. Genet. 2, 1271-1288.

Goss, S.J., Harris, H. (1975). New method for mapping genes in human chromosomes. Nature 255, 680-684.

Goto, Y., Buschman, E., Skamene, E. (1989). Regulation of host response to *Mycobacterium intracellulare* in vivo and in vitro by the *Bcg* gene. Immunogenetics 30, 218-221.

Green, E.D., Braden, V.V., Fulton, R.S., Lim, R., Ueltzen, M.S., Peluso, D.C., Mohr-Tidwell, R.M., Idol, J.R., Smith, L.M., Chumakov, I., Le Paslier, D., Cohen, D., Featherstone, T., Green, P. (1995). A human chromosome 7 yeast artificial chromosome (YAC) resource: construction, characterization, and screening. Genomics 25, 170-183.

Green, E.D., Idol, J.R., Mohr-Tidwell, R.M., Braden, V.V., Peluso, D.C., Fulton, R.S., Massa, H.F., Magness, C.L., Wilson, A.M., Kimura, J., Weissenbach, J., Trask, B.J. (1994). Integration of physical, genetic and cytogenetic maps of human chromosome 7: isolation and analysis of yeast artificial chromosome clones for 117 mapped genetic markers. Hum. Mol. Genet. 3, 489-501.

Green, E.D., Mohr, R.M., Idol, J.R., Jones, M., Buckingham, J.M., Deaven, L.L., Moyzis, R.K., Olson, M.V. (1991a). Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: application to the mapping of human chromosome 7 using yeast artificial chromosomes. Genomics 11, 548-564.

Green, E.D., Olson, M.V. (1990a). Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: a model for human genome mapping. Science 250, 94-98.

Green, E.D., Olson, M.V. (1990b). Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci. USA 87, 1213-1217.

Green, E.D., Riethman, H.C., Dutchik, J.E., Olson, M.V. (1991b). Detection and characterization of chimeric yeast artificial-chromosome clones. Genomics 11, 658-669.

Griffin, C.A., Smith, M., Henthorn, P.S., Harris, H., Weiss, M.J., Raducha, M., Emanuel, B.S. (1987). Human placental and intestinal alkaline phosphatase genes map to 2q34-q37. Am. J. Hum. Genet. 41, 1025-1034.

Grimaldi, G., Skowronski, J., Singer, M.F. (1984). Defining the beginning and end of KpnI family segments. EMBO J. 3, 1753-1759.

Gingrich, J., Garnes, J., Boehrer, D., Johnson, W., Wong, B., Bergmann, A., Eveleth, J., Langlois, R., Lamerdin, J., Lee, D., Carrano, A. (1995). Generation and chracterization of chromosome 2-specific cosmid, fosmid and PAC clone libraries and new chromosome 2 STSs derived from micro-dissection clones. Genome mapping and sequencing, Cold Spring Harbor laboratory, Cold Spring Harbor, N.Y.

Gros, P., Skamene, E., Forget, A. (1983). Cellular mechanisms of genetically controlled host resistance to *Mycobacterium bovis* (BCG). J. Immunol. 131, 1966-1972.

Gros, P., Skamene, E., Forget, A.J. (1981). Genetic control of natural resistance to *Mycobacterium bovis* (BCG) in mice. J. Immunol. 127, 2417-2421.

Gruenheid, S., Cellier, M., Vidal, S., Gros, P. (1995). Identification and characterization of a second mouse *Nramp* gene. Genomics 25, 514-525.

Grundy, H.O., Niemeyer, P., Rupani, M.F., Wassman, E.R. (1989). Prenatal detection of cyclopia associated with interstitial deletion of 2p. Am. J. Med. Genet. 34, 268-270.

Guzman, P., Ecker, J.R. (1988). Development of large DNA methods for plants: Molecular cloning of large segments of *Arabidopsis* and carrot DNA into yeast. Nucleic Acids Res. 16, 11091-11105.

Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., Weissenbach, J. (1994). The 1993-1994 Généthon human genetic linkage map . Nature Genet. 7 (Special), 246-339.

Hafezparast, M., Kaur, G.P., Zdzienicka, M., Athwal, R.S., Lehmann, A.R., Jeggo, P.A. (1993). Subchromosomal localization of a gene (XRCC5) involved in double strand break repair to the region 2q34-36. Somatic Cell and Molecular Genetics 19, 413-421.

Haldi, M., Perrot, V., Saumier, M., Desai, T., Cohen, D., Cherif, D., Ward, D., Lander, E.S. (1994). Large human YACs constructed in a *rad52* strain show a reduced rate of chimerism. Genomics 24, 478-484.

Hamada, H., Kakunaga, T. (1982). Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298, 396-398.

Hamada, H., Petrino, M.G., Takunaga, T. (1982). A novel repeated element with Z-DNA formaing potential is widely found in evolutionary diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA 79, 6465-6469.

Hawthorn, L., Cowell, J.K. (1995). Integration of the physical and genetic linkage map for human chromosome 13. Genomics 27, 399-404.

Hecht, B.K.-M., Hecht, F., Munke, M. (1991). Forebrain cleavage gene causing holoprosencephaly: deletion mapping to chromosome band 2p21. Am. J. Med. Genet. 40, 130.

Henry, I., Jeanpierre, M., Weil, D., Grzeschik, K.H., Ramirez, F., Junien, C. (1985). The structural gene for fibronectin (FN) maps to 2q32.3-qter. Cytogenet. Cell Genet. 40, 650.

Hentati, A., Bejaoui, K., Pericak-Vance, M.A., Hentati, F., Speer, M.C., Hung, W-Y., Figlewicz, D.A., Haines, J., Rimmler, J., Hamida, C.B., Hamida, M.B., Brown, R.H.Jr., Siddique, T. (1994). Linkage of recessive familial amyotrophic lateral sclerosis to chromosome 2q33-q35. Nature Genet. 7, 425-428.

Hoheisel, J.D., Maier, E., Mott, R., McCarthy, L., Grigoriev, A.V., Schalkwyk, L.C., Nizetic, D., Francis, F., Lehrach, H. (1993). High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast S. pombe. Cell 73, 109-120.

Hoshide, R.T., Matsuura, T., Haraguchi, Y., Endo, F., Yoshinaga, M., Matsuda, I. (1993). Carbamyl phosphate synthetase I deficiency. J. Clin. Invest. 91, 1884-1887.

Huerre-Jeanpierre, C., Henry, I., Mattei, M.G., Weil, D., Grzeschik, K.H., Chu, M.L., Bernard, M., Ramirez, F., Junien, C. (1985). The gene for human alpha-1 type III collagen (COL3A1) is physically linked to alpha-2 type V collagen COL5A2 on chromosome 2 (2q31-2q32.3). Cytogenet. Cell Genet. 40, 657.

Hunter, K.W., Ontiveros, S.D., Watson, M.L., Stanton, V.P.,Jr., Gutierrez, P., Bhat, D., Rochelle, J., Graw, S., Ton, C., Schalling, M., Aburatani, H., Brown, S.D.M., Seldin, M.F., Housman, D.E. (1994). Rapid and efficient construction of yeast artificial chromosome contigs in the mouse genome with interspersed repetitive sequence PCR (IRS-PCR): generation of a 5-cM, > 5 megabase contig on mouse chromosome 1. Mamm. Genome 5, 597-607.

Huntington's Disease Collaborative Research Group, (1993). A novel gene containing a tri-nucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72, 971-983.

Hwu, H.R., Roberts, J.W, Davidson, E.H., Britten, R.J. (1986). Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc. Natl. Acad. Sci. USA 83, 3875-3879.

Imai, T., Olson, M.V. (1990). Second-generation approach to the construction of yeast artificial-chromosome libraries. Genomics 8, 297-303.

Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A., de Jong, P.J. (1994). A new bacteriophage P1-derived vector for the propagation of large DNA fragments. Nature Genet. 6, 84-89.

Ivanyi, J. (1994). Molecular biology of natural resistance-associated macrophage protein. Parasitol. Today 10, 416-477.

James, M.R., Richard III, C.W., Schott, J-J., Yousry, C., Clark, K., Bell, J., Terwilliger, J.D., Hazan, J., Dubay, C., Vignal, A., Agrapart, M., Imai, T., Nakamura, Y., Polymeropoulos, M., Weissenbach, J., Cox, D.R., Lathrop, G.M. (1994). A radiation hybrid map of 506 STS markers spanning human chromosome 11. Nature Genet. 8, 70-76.

Jazwinska, E.C., Serjeantson, S.W. (1988). HLA-DR.-DQ DNA genotyping and T-cell receptor RFLPs in leprosy. Disease Markers 6, 173-183.

Jeffreys, A.J., Wilson, V., Thein, S.L. (1985). Hypervariable 'minisatellite' regions in human DNA. Nature 314, 67-73.

Jeggo, P.A., Hafezparast, M., Thompson, A.F., Broughton, B.C., Kaur, G.P., Zdzienicka, M.Z., Athwal, R.S. (1992). Localization of a DNA rpair gene (XRCC5) involved in double-stranded-break rejoining to human chromosome 2. Proc. Natl. Acad. Sci. USA 89, 6423-6427.

Jeggo, P.A., Hafezparast, M., Thompson, A.F., Kaur, G.P., Sandhu, A.K., Athwal, R.S. (1993). A Hamster-human subchromosomal hybrid cell panel for chromosome 2. Somatic Cell and Molecular Genetics 19, 39-49.

Jelinek, W.R., Schmid, C.W. (1982). Repetitive sequences in eukaryotic DNA and their expression. Annu. Rev. Biochem. 51, 813-844.

Jurka, J., Milosavljevic, A. (1991). Reconstruction and analysis of human Alu genes. J. Mol. Evol. 32, 105-121.

Jurka, J., Smith, T. (1988). A fundamental division in the Alu family of repeated sequences. Proc. Natl. Acad. Sci. USA 85, 4775-4778.

Kallmann, F.J., Reisner, D. (1943). Twin studies on the genetic significance of genetic factors in tuberculosis. Am. Rev. Tuberc. 47, 549-574.

Kato, A., Miura, O., Sumi, Y., Aoki, N. (1988). Assignment of the human protein C gene (PROC) to chromosome region 2q14-q21 by in situ hybridization. Cytogenet. Cell Genet. 47, 46-47.

Keats, B.J., Sherman, S.L., Morton, N.E., Robson, E.B., Buetow, K.H., Cartwright, P.E., Chakravarti, A., Francke, U., Green, P.P., Ott, J. (1991). Guidelines for human linkage maps: an international system for human linkage maps (ISLM, 1990). Genomics 9, 557-560.

Kim, U-J., Shizuya, H., Birren, B., Slepak, T., de Jong, P., Simon, M.I. (1994). Selection of chromosome 22-specific clones from human genomic BAC library using a chromosome-specific cosmid library pool. Genomics 22, 336-339.

Kishi, F. (1994). Isolation and characterization of human *Nramp* cDNA. Biochem. Biophys. Res. Commun. 204, 1074-1080.

Kohara, Y., Akiyama, K., Isono, K. (1987). The physical map of the whole E. coli chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell 50, 495-508.

Korenberg, J.R., Rykowski, M.C. (1988). Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. Cell 53, 391-400.

Korn, B., Sedlacek, Z., Manca, A., Kioschis, P., Konecki, D., Lehrach, H., Poustka, A. (1992). A strategy for the selection of transcribed sequences in the Xq28 region. Hum. Mol. Genet. 1, 235-242.

Krizman, D.B., Berget, S.M. (1993). Efficient selection of 3'-terminal exons from vertebrate DNA. Nucleic Acids Res. 21, 5198-5202.

Lafage-Pochitaloff, M., Costello, R., Couez, D., Simonetti, J., Monnoni, P., Mawas, C., Olive, D. (1990). Human CD28 and CTLA-4 Ig superfamily genes are located on chromosome 2 at bands q33-q34. Immunogenetics 31, 198-201.

Lagrange, P.H., Hurtrel, B., Brandely, M., Thickstun, P.M. (1983). Immunological mechanisms controlling mycobacterial infections. Bull. Eur. Physiopathol. Respir. 19, 163-172.

Larin, Z., Monaco, A.P., Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc. Natl. Acad. Sci. USA 88, 4123-4127.

Lawrence, J.B., Singer, R.H., McNeil, J.A. (1990). Interphase and metaphase resolution of different distances within the human dystrophin gene. Science 249, 928-932.

Lawrence, J.B., Villnave, C.A., Singer, R.H. (1988). Sensitive, high-resolution chromatin and chromosome mapping in situ: presence and orientation of two closely integrated copies of EBV in a lymphoma line. Cell 52, 51-61.

LeBlanc-Straceski, J.M., Montgomery, K.T., Kissel, H., Murtaugh, L., Tsai, P., Ward, D.C., Krauter, K.S., Kucherlapati, R. (1994). Twenty-one polymorphic markers from human chromosome 12 for integration of genetic and physical maps. Genomics 19, 341-349.

Ledbetter, S.A., Nelson, D.L., Warren, S.T., Ledbetter, D.H. (1990). Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. Genomics 6, 475-481.

Lenzini, L., Rottoli, P., Rottoli, L. (1977). The spectrum of human tuberculosis. Clinical and Experimental Immunology 27, 230-237.

Levee, G., Liu, J., Gicquel, B., Chanteau, S., Schurr, E. (1994). Genetic control of susceptibility to leprosy in French Polynesia; no evidence for linkage with markers on telomeric human chromosome 2. Int. J. Leprosy 62, 499-511.

Libert, F., Lefort, A., Okimoto, R., Womack, J., Georges, M. (1993). Construction of a bovine genomic library of large yeast artificial chromosome clones. Genomics 18, 270-276.

Lichter, P., Tang, C.J., Call, K., Hermanson, G., Evans, G.A., Housman, D., Ward, D.C. (1990). High resolution mapping of human chromosome 11 by *in situ* hybridization with cosmid clones. Science 247, 64-69.

Little, R.D., Porta, G., Carle, G.F., Schlessinger, D., D'Urso, M. (1989). Yeast artificial chromosomes with 200- to 800-kilobase inserts of human DNA containing HLA, V kappa, 5S, and Xq24-Xq28 sequences. Proc. Natl. Acad. Sci. USA 86, 1598-1602.

Liu, J., Cellier, M., Schurr, E., Skamene, E. (1993). Comparative genome analysis - a novel strategy to study the genetics of host-parasite interaction. J. Parasitol. 79, 463-469.

Liu, J., Stanton, V.P.,Jr., Fujiwara, T.M., Wang, J.W., Rezonzew, R., Crumley, M.J., Morgan, K., Gros, P., Housman, D., Schurr, E. (1995a). Large-scale cloning of human chromosome 2-specific yeast artificial chromosome (YACs) using an interspersed repetitive sequences (IRS)-PCR approach. Genomics 26, 178-191.

Liu, J., Fujiwara, T.M., Buu, N.T., Sánchez, F.O., Cellier, M., Paradis, A.J., Frappier, D., Skamene, E., Gros, P., Morgan, K., Schurr, E. (1995b). Identification of polymorphisms and sequence variants in the human homologue of the mouse natural resistance-associated macrophage protein gene. Am. J. Hum. Genet. 56, 845-853.

Lovett, M., Kere, J., Hinton, L.M. (1991). Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. Proc. Natl. Acad. Sci. USA 88, 9628-9632.

Lu-Kuo, J., Ward, D.C., Spritz, R.A. (1993). Fluorescence in situ hybridization mapping of 25 markers on distal human chromosome 2q surrounding the human Waardenburg syndrome, type I (WS1) locus (PAX3 gene). Genomics 16, 173-179.

Luerssen, H., Mattei, M.-G., Schroter, M., Grzeschik, K.-H., Adham, I.M., Engel, W. (1990). Nucleotide sequence of the gene for human transition protein 1 and its chromosomal localization on chromosome 2. Genomics 8, 324-330.

Mah, M.W., Fanning, E.A. (1991). An epidemic of primary tuberculosis in a Canadian community. Can. J. Infect. Dis. 2, 133-141.

Malo, D., Schurr, E., Epstein, D.J., Vekemans, M., Skamene, E., Gros, P. (1991). The host resistance locus *Bcg* is tightly linked to a group of cytoskeleton-associated protein genes that include Villin and Desmin. Genomics 10, 356-364.

Malo, D., Vidal, S., Hu, J., Skamene, E., Gros, P. (1993a). High-resolution linkage map in the vicinity of the host resistance locus *Bcg*. Genomics 16, 655-663.

Malo, D., Vidal, S., Lieman, J.H., Ward, D.C., Gros, P. (1993b). Physical delineation of the minimal chromosomal segment encompassing the murine host resistance locus Bcg. Genomics 17, 667-675.

Malo, D., Vogan, K., Vidal, S., Hu, J., Cellier, M., Schurr, E., Fuks, A., Bumstead, N., Morgan, K., Gros, P. (1994). Haplotype mapping and sequence analysis of the mouse Nramp gene predict susceptibility to infection with intracellular parasites. Genomics 23, 51-61.

Marchuk, D., Drumm, M., Saulino, A., Collins, F.S. (1991). Construction of T-Vectors, a rapid and general system for direct cloning of unmodified PCR products. Proc. Natl. Acad. Sci. USA 19, 1154.

Martin, D., Tucker, D.F., Gorman, P., Sheer, D., Spurr, N.K., Trowsdale, J. (1987). The human placental alkaline phosphatase gene and related sequences map to chromosome 2 band q37. Ann. Hum. Genet. 51, 145-152.

Matise, T.C., Perlin, M., Chakravarti, A. (1994). Automated construction of genetic linkage maps using an expect system (MultiMap): a human genome linkage map. Nature Genet. 6, 384-390.

Mattei, M.G., Colin, Y., Le Van Kim, C., Mattei, J.F., Cartron, J.P. (1986). Localization of the gene for human erythrocyte glycophorin C to chromosome 2,q14-q21. Hum. Genet. 74, 420-422.

McCormick, M.K., Campbell, E., Deaven, L., Moyzis, R. (1993). Low-frequency chimeric yeast artificial chromosome libraries from flow-sorted human chromosomes 16 and 21. Proc. Natl. Acad. Sci. USA 90, 1063-1067.

McCormick, M.K., Shero, J.H., Cheung, M.C., Kan, Y.W., Hieter, P.A.; Antonarakis, S.E. (1989). Construction of human chromosome 21-specific yeast artificial chromosomes. Proc. Natl. Acad. Sci. USA 86, 9991-9995.

Mizukami, T., Chang, W.I., Garkavtsev, I., Kaplan, N., Lombardi, D., Matsumoto, T., Niwa, O., Kounosu, A., Yanagida, M., Marr, T.G., Beach, D. (1993). A 13 kb resolution cosmid map of the 14 Mb fission yeast genome by nonrandom sequence-tagged site mapping. Cell 73, 121-132.

Mohr, J. (1954). "A Study of Linkage in Man," Munksgaard, Copenhagen.

Moir, D.T., Dorman, T.E., Day, J.C., Ma, N.S-F, Wang, M-T., Mao, J-I. (1994). Toward a physical map of human chromosome 10: isolation of 183 YACs representing 80 loci and regional assignment of 94 YACs by fluorescence *in situ* hybridization. Genomics 22, 1-12.

Morales, A., Nickel, J.C., Wilson, J.W.L. (1992). Dose response of Bacillus Calmette-Guerin in the treatment of superficial bladder cancer. J. Urol. 147, 1256-1258.

Morris, S.W., Nelson, N., Valentine, M.B., Shapiro, D.N., Look, A.T., Kozlosky, C.J., Beckmann, M.P., Cerretti, D.P. (1992). Assignment of the genes encoding human interleukin-8 receptor types 1 and 2 and an interleukin-8 receptor pseudogene to chromosome 2q35. Genomics 14, 685-691.

Moyzis, R.K., Torney, D.C., Meyne, J., Buckingham, J.M., Wu, J-R., Burks, C., Sirotkin, K.M., Goad, W.B. (1989). The distribution of interspersed repetitive DNA sequences in the human genome. Genomics 4, 273-289.

Munroe, D.J., Haas, M., Bric, E., Whitton, T., Aburatani, H., Hunter, K., Ward, D., Housman, D.E. (1994). IRE-bubble PCR: a rapid method for efficient and representative amplification of human genomic DNA sequences from complex sources. Genomics 19, 506-514.

Murray, A.W., Szostak, J.W. (1983). Construction of artificial chromosomes in yeast. Nature 305, 189-193.

Murray, J.C., Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Manion, F., Quillen, J., Sheffield, V.C., Sunden, S., Duyk, G.M., Weissenbach, J., Gyapay, G., Dib, C., Morrissette, J., Lathrop, G.M., Vignal, A., White, R., Matsunami, N., Gerken, S., Melis, R., Albertsen, H., Plaetke, R., Odelberg, S., Ward, D., Dausset, J., Cohen, D., Cann, H. (1994). A comprehensive human linkage map with centimorgan density. Science 265, 2049-2054.

Nadeau, J.H. (1989). Maps of linkage and synteny homologies between mouse and man. Trends Genet. 5, 82-86.

Nadeau, J.H., Davisson, M.T., Doolittle, D.P., Grant, P., Hillyard, A.L., Kosowsky, M., Roderick, T.H. (1992). Comparative map for mice and humans. Mamm. Genome 3, 480-536.

Nadeau, J.H., Taylor, B.A. (1984). Lengths of chromosomal segments conserved since divergence of man and mouse. Proc. Natl. Acad. Sci. USA 81, 814-818.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235, 1616-1622.

Nathan, C. (1992). Nitric oxide as a secretory product of mammalian cells. FASEB J. 6, 3051-3061.

Nathan, C. (1995). Natural resistance and nitric oxide. Cell 82, 873-876.

Nathan, C.F., Hibbs, J. (1991). Role of nitric oxide synthesis in macrophage antimicrobial activity. Cuur. Opin. Immunol. 3, 65-70.

Neel, J.V., Shull, W.J. (1954). Human heredity. University of Chicago Press, Chicago.

Nelson, D.L., Ballabio, A., Victoria, M.F., Pieretti, M., Bies, R.D., Gibbs, R.A., Maley, J.A., Chinault, A.C., Webster, T.D. (1991). *Alu*-primed polymerase chain reaction for regional assignment of 110 yeast artificail chromosome clones from the human X chromosome: Identification of clones associated with a disease locus. Proc. Natl. Acad. Sci. USA 88, 6157-6161.

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H., Caskey, C.T. (1989). *Alu* polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. Proc. Natl. Acad. Sci. USA 86, 6686-6690.

NIH/CEPH Collaborative Mapping Group. (1992). A comprehensive genetic linkage map of the human genome. Science 258, 67-86.

Nizetic, D., Gellen, L., Hamvas, R.M., Mott, R., Grigoriev, A., Vatcheva, R., Zehetner, G., Yaspo, M-L., Dutriaux, A., Lopes, C., Delabar, J-M., Broeckhoven, C.V., Potier, M-C., Lehrach, H. (1994). An integrated YAC-overlap and 'cosmid-pocket' map of the human chromosome 21. Hum. Mol. Genet. 3, 759-770.

Nizetic, D., Zehetner, G., Monaco, A.P., Gellen, L., Young, B.D., Lehrach, H. (1991). Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: Their potential use as reference libraries. Proc. Natl. Acad. Sci. USA 88, 3233-3237.

Noordeen, S.K., Lopez Bravo, L., Sundaresan, T.K. (1992). Estimated number of leprosy cases in the world. WHO Bull. 70, 707-710.

O'Brien, S.J., Seuanez, H.N., Womack, J.E. (1988). Mammalian genome organization: an evolutionary view. Annu. Rev. Genet. 22, 323-351.

Olson, M., Hood, L., Cantor, C., Botstein, D. (1989). A common language for physical mapping of the human genome. Science 245, 1434-1435.

Olson, M.V., Dutchik, J.E., Graham, M.Y., Brodeur, G.M., Helms, C., Frank, M., MacCollin, M., Scheinman, R., Frank, T. (1986). Random-clone strategy for genomic restriction mapping in yeast . Proc. Natl. Acad. Sci. USA 83, 7826-7830.

Ott, J. (1991). Analysis of Human Genetic Linkage. Baltimore: The Johns Hopkins University Press.

Patil, N., Peterson, A., Rothman, A., de Jong, P.J., Myers, R.M., Cox, D.R. (1994). A high resolution physical map of 2.5 Mbp of the Down syndrome region on chromosome 21. Hum. Mol. Genet. 3, 1811-1817.

Patracchini, P., Aiello, V., Palazzi, P., Calzolari, E., Bernardi, F. (1989). Sublocalization of the human protein C gene on chromosome 2q13-q14. Hum. Genet. 81, 191-192.

Peltomaki, P., Aaltonen, L.A., Sistonen, P., Pylkkanen, L., Mecklin, J.-P., Jarvinen, H., Green, J.S., Jass, J.R., Weber, J.L., Leach, F.S., Petersen, G.M., Hamilton, S.R., de la Chapelle, A., Vogelstein, B. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. Science 260, 810-812.

Perlin, M., Chakravarti, A. (1993). Efficient construction of high-resolution physical maps from yeast artificial chromosomes using radiation hybrids: inner product mapping. Genomics 18, 283-289.

Perlin, M.W., Duggan, D.J., Davis, K., Farr, J.E., Findler, R.B., Higgins, M.J., Nowak, N.J., Evans, G.A., Qin, S., Zhang, J., Shows, T.B., James, M.R., Richard, C.W.,III. (1995). Rapid construction of integrated maps using inner product mapping: YAC coverage of human chromosome 11. Genomics 28, 315-327.

Pierce, J.C., Sauer, B., Stemberg, N. (1992). A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: Improved cloning efficacy. Proc. Natl. Acad. Sci. USA 89, 2056-2060.

Plant, J.E., Glynn, A. (1976). Genetics of resistance to infection with *Salmonella typhimurium* in mice. J. Infect. Dis. 133, 72-78.

Poleev, A., Fickenscher, H., Mundlos, S., Winterpacht, A., Zabel, B., Fidler, A., Gruss, P., Plachov, D. (1992). PAX8, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilms' tumors. Development. 116, 611-623.

Porta, G., Zucchi, I., Hillier, L., Green, P., Nowotny, V., D'Urso, M., Schlessinger, D. (1993). *Alu* and L1 sequence distributions in Xq24-q28 and their comparative utility in YAC contig assembly and verification. Genomics 16, 417-425.

Pringault, E., Robine, S., Louvard, D. (1991). Structure of the human villin gene. Proc. Natl. Acad. Sci. USA 88, 10811-10815.

Purdue, P.E., Lumb, M.J., Fox, M., Griffo, G., Hamon-Benais, C., Povey, S., Danpure, C.J. (1991). Characterization and chromosomal mapping of a genomic clone encoding human alanine:glyoxylate aminotransferase. Genomics 10, 34-42.

Qin, S., Zhang, J., Isaacs, C.M., Nagafuchi, S., Jani Sait, S.N., Abel, K.J., Higgins, M.J., Shows, T.B. (1993). A chromosome 11 YAC library. Genomics 16, 580-585.

Rajcan-Separovic, E., Sabour, M.P. (1993). Fluorescence in situ hybridization of bovine Alu-like sequences to bovine and ovine chromosomes. Genome 36, 984-986.

Richard, C.W., Boehnke, M., Berg, D.J., Lichy, J.H., Meeker, T.C., Hauser, E., Myers, R.M., Cox, D.R. (1993). A radiation hybrid map of the distal short arm of human chromosome 11, containing the Beckwith-Wiedemann and associated embryonal disease loci. Am. J. Hum. Genet. 52, 915-921.

Richards, A.J., Lloyd, J.C., Narcisi, P., Ward, P.N., Nicholls, A.C., De Paepe, A., Pope, F.M. (1992). A 27-bp deletion from one allele of the type III collagen gene (COL3A1) in a large family with Ehlers-Danlos syndrome type IV. Hum. Genet. 88, 325-330.

Rinehart, F.P., Ritch, T.G., Deininger, P.L., Schmid, C.W. (1981). Renaturation rate studies of a single family of interspersed repeated sequences in human deoxyribonucleic acid. Biochemistry 20, 3003-3010.

Roach, T.I., Chatterjee, D., Blackwell, J.M. (1994). Induction of early-response genes KC and JE by mycobacterial lipoarabinomannans: regulation of KC expression in murine macrophages by *Lsh/Ity/Bcg* (candidate *Nramp*). Infection and Immunity 62, 1176-1184.

Ross, M.T., Nizetic, D., Nguyen, C., Knights, C., Vatcheva, R., Burden, N., Douglas, C., Zehetner, G., Ward, D.C., Baldini, A., Lehrach, H. (1992). Selection of a human chromosome 21 enriched YAC sub-library using a chromosome-specific composite probe. Nature Genetics 1, 284-290.

Rouquier, S., Batzer, M.A., Giorgi, D. (1994). Application of bacterial artificial chromosomes to the generation of contiguous physical maps: a pilot study of human ryanodine receptor gene (RYR1) region. Analytical Biochemistry 217, 205-209.

Rousseau-Merck, M.F., Simon-Chazottes, D., Arpin, M., Pringault, E., Louvard, D., Guenet, J.L., Berger, R. (1988). Localization of the villin gene on human chromosome 2q35-q36 and on mouse chromosome 1. Hum. Genet. 78, 130.

Rubin, C.M., Leeflang, E.P., Rinehart, F.P., Schmid, C.W. (1993). Paucity of novel short interspersed repetitive element (SINE) families in human DNA and isolation of a novel MER repeat. Genomics 18, 322-328.

Rudkin, G.T., Stollar, B.D. (1977). High resolution detection of DNA-RNA hybrids in situ by indirect immunoflurorescence. Nature 265, 472-473.

Sambrook, J., Fritsch, F., Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., Coulson, A.R. (1977). DAN sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Schedl, A., Montoliu, L., Kelsey, G., Schutz, G. (1993). A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic. Nature 362, 258-261.

Scherer, S.W., Tompkins, B.J., Tsui, L.C. (1992). A human chromosome 7-specific genomic DNA library in yeast artificial chromosomes. Mamm. Genome 3, 179-181.

Schlessinger, D., Little, R.D., Freije, D., Abidi, F., Zucchi, I., Porta, G., Pilia, G., Nagaraja, R., Johnson, S.K., Yoon, J-Y., Srivastava, A., Kere, J., Palmieri, G., Ciccodicola, A., Montanaro, V., Romano, G., Casamassimi, A., D'Urso, M. (1991). Yeast artificial chromosome-based genome mapping: some lessons from Xq24-q28. Genomics 11, 783-793.

Schmid, C.W., Deininger, P.L. (1975). Sequence organization of the human genome. Cell 6, 345-358.

Schmid, C.W., Jelinek, W.R. (1982). The Alu family of dispersed repetitive sequences. Science 216, 1065-1070.

Schneemann, M., Schoedon, G., Hofer, S., Blau, N., Guerrero, L., Schaffner, A. (1993). Nitric oxide synthase is not a constituent of the antimicrobial armature of human mononuclear phagocytes. J. Infect. Dis. 167, 1358-1363.

Schroeder, W.T., Lopez, L.C., Harper, M.E., Saunders, G.F. (1984). Localization of the human glucagon gene (GCG) to chromosome segment 2q36-37. Cytogenet. Cell Genet. 38, 76-79.

Schurr, E., Buschman, E., Malo, D., Gros, P., Skamene, E. (1990a). Immunogenetics of mycobacterial infections: mouse-human homologies. The Journal of Infectious Diseases 161, 634-639.

Schurr, E., Malo, D., Radzioch, D., Buschman, E., Morgan, K., Gros, P., Skamene, E. (1991a). Genetic control of innate resistance to mycobacterial infections. Cambridge:Elsevier Science Publishers Ltd. A42-A45.

Schurr, E., Radzioch, D., Malo, D., Gros, P., Skamene, E. (1991b). Molecular genetics of inherited susceptibility to intracellular parasites. Behring Inst. Mitt. 88, 1-12.

Schurr, E., Skamene, E., Forget, A., Gros, P. (1989). Linkage analysis of the *Bcg* gene on mouse chromosome 1: identification of a tightly linked marker. J. Immunol. 142, 4507-4513.

Schurr, E., Skamene, E., Morgan, K., Chu, M.L., Gros, P. (1990b). Mapping of Col3a1 and Col6a3 to proximal murine chromosome 1 identifies conserved linkage of structural protein genes between murine chromosome 1 and human chromosome 2q. Genomics. 8, 477-486.

Schurr, E., Skamene, E. (1995). The role of the *Bcg* gene in mycobacterial infections. In *Tuberculosis*, W. Rom and S. Garay, eds. (Little, Brown and Company Press). in press.

Schweizer, D. (1976). Reverse fluorescent chromosome banding with chromomycin and DAPI. Chromosoma 58, 307-324.

Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, C.T., O'hara, B., Rossiter, J.P., Cooley, T., Heath, P., Smith, K.D., Margolet, L. (1987). Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics 1, 113-125.

Searle, A.G., Edwards, J.H., Hall, J.G. (1994). Mouse homologues of human hereditary disease. J. Med. Genet. 31, 1-19.

Shaw, M-A., Atkinson, S., Dockrell, H., Hussain, R., Lins-Lainson, Z., Shaw, J., Ramos, F., Silveira, F., Mehdi, S.Q., Kautab, F., Khaliq, S., Chiang, T., Blackwell, J. (1993). An RFLP map for 2q33-q37 from multicase mycobacterial and leishmanial disease families: no evidence for an Lsh/Ity/Bcg gene homologue influencing susceptibility to leprosy. Ann. Hum. Genet. 57, 251-271.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc. Natl. Acad. Sci. USA 89, 8794-8797.

Sidhu, M.S., Helen, B.K., Athwal, R.S. (1992). Fingerprinting human chromosomes by polymerase chain reaction-mediated DNA amplification. Genomics 14, 728-732.

Sifford, M., Bates, J.H. (1991). Host determinants of susceptibility to Mycobacterium tuberculosis. Simin. Respir. Infect. 6, 44-50.

Singer, M.F. (1982). SINEs and LINEs: Highly repeated short and long interspersed sequences in mammalian genomes. Cell 28, 433-434.

Singer, M.F., Skowronski, J. (1985). Making sense of LINEs: Long interspersed repeat sequences in mammalian genomes. Trends Biochem. Sci. 119-122.

Skamene, E., Gros, P., Forget, A., Patel, P.J., Nesbitt, M. (1984). Regulation of resistance to leprosy by chromosome 1 locus in the mouse. Immunogenetics 19, 117-120.

Skamene, E., Gros, P., Forget, A.J., Kongshavn, P.A.L., St.-Charles, C., Taylor, B.A. (1982). Genetic regulation of resistance to intracellular pathogens. Nature 297, 505-509.

Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw, H., Deininger, P. (1987). Clustering and subfamily relationships of the Alu family in the human genome. Mol. Biol. Evol. 4, 19-29.

Smith, D.R. (1990). Genomic long-range restriction mapping. Methods: A Companion to Methods in Enzymology 1, 195-203.

Smith, M.W., Clark, S.P., Hutchinson, J.S., Wei, Y.H., Churukian, A.C., Daniels, L.B., Diggle, K.L., Gen, M.W., Romo, A.J., Lin, Y., Selleri, L., McElligott, D.L., Evans, G.A. (1993). A sequence-tagged site map of human chromosome 11. Genomics 17, 699-725.

Soeda, E., Hou, D-X., Osoegawa, K., Atsuchi, Y., Yamagata, T., Shimokawa, T., Kishida, H., Okano, S., Chumakov, I., Cohen, D., Raff, M., Gardiner, K., Graw, S.L., Patterson, D., de Jong, P., Ashworth, L.K., Slezak, T., Carrano, A.V. (1995). Cosmid assembly and anchoring to human chromosome 21. Genomics 25, 73-84.

Spurr, N.K., Cox, S., Bryant, S.P., Attwood, J., Robson, E.B., Shields, D.C., Steinbrueck, T., Murray, J.C., Kidd, K.K., Summar, M.L., Tsipouras, P., Retief, A.E., Kruse, T.A., Bale, A.E., Vergnaud, G., Weber, J.L., McBride, O.W., Donis-Keller, H., White, R.L. (1992). The CEPH consortium linkage map of human chromosome 2. Genomics 14, 1055-1063.

Spurr, N.K., Barton, H., Bashir, R., Bryson, G.M., Bushby, K., Cox, S., Gringrich, J.C., Hentati, A., Hildebrandt, F., Kao, F-T., Kruse, T., Lai, E., Liu, J., Menke, M., Naylor, S., Nicklin, M., Read, A., Reiner, O., Rocchi, M., Summer, M. (1994). Report of the Third International Workshop on Human Chromosome 2 Mapping. Cytogenet. Cell Genet. 67, 215-244.

Stach, J-L., Gros, P., Forget, A., Skamene, E. (1984). Phenotypic expression of genetically controlled natural resistance to *Mycobacterium bovis* (BCG). J. Immunol. 132, 888-892.

Stallings, R.L., Ford, A.F., Nelson, D., Torney, D.C., Hildebrand, C.E., Moyzis, R.K. (1991). Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. Genomics 10, 807-815.

Stallings, R.L., Torney, D.C., Hildebrand, C.E., Longmire, J.L., Deaven, L.L., Jett, J.H., Doggett, N.A., Moyzis, R.K. (1990). Physical mapping of human chromosomes by repetitive sequence fingerprinting. Proc. Natl. Acad. Sci. USA 87, 6218-6222.

Stapleton, P., Weith, A., Urbanek, P., Kozmik, Z., Busslinger, M. (1993). Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9. Nature Genet. 3, 292-298.

Stead, W.W. (1992). Genetics and resistance to tuberculosis. Could resistance be enhanced by genetic engineering? Ann. Intern. Med. 116, 937-941.

Sternberg, N. (1990). Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc. Natl. Acad. Sci. USA 87, 103-107.

Sternberg, N., Ruether, J., de Riel, K. (1990). Generation of a 50,000-member human DNA library with an average DNA insert size of 75-100 kbp in a bacteriophage P1 cloning vector. New Biologist 2, 151-162.

Taccioli, G.E., Gottlieb, T.M., Blunt, T., Priestley, A., Demengeot, J., Mizuta, R., Lehmann, A.R., Alt, F.W., Jackson, S.P., Jeggo, P.A. (1994). Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. Science 265, 1442-1445.

Tagle, D.A., Collins, F.S. (1992). An optimized *Alu*-PCR primer pair for human-specific amplification of YACs and somatic cell hybrids. Hum. Mol. Genet. 1, 121-122.

Tagle, D.A., Swaroop, M., Lovett, M., Collins, F.S. (1993). Magnetic bead capture of expressed sequences encoded within large genomic segments. Nature 361, 751-753.

Torrence, R.J., Kavoussi, L.R., Catalona, W.J., Ratliff, T.L. (1988). Prognostic factors in patients treated with intravesical bacillus Calmette-Guerin for superficial bladder cancer. J. Urol. 139, 941-944.

Trask, B.J., Massa, H.F., Kenwrick, S., Gitschier, J. (1991). Mapping of human chromosome Xq28 by two-color fluorescence *in situ* hybridization of DNA sequences to interphase cell nuclei. Am. J. Hum. Genet. 48, 1-15.

Trask, B.J., Pinkel, D., van den Engh, G. (1989). The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. Genomics 5, 710-717.

Traver, C.N., Klapholz, S., Hyman, R.W., Davis, R.W. (1989). Rapid screening of a human genomic library in yeast artificial chromosomes for single-copy sequences. Proc. Natl. Acad. Sci. USA 86, 5898-5902.

Tricoli, J.V., Bell, G.I., Shows, T.B. (1984). The human glucagon gene is located on chromosome 2. Diabetes 33, 200-202.

Tucker, J.D., Christensen, M.L., Carrano, A.V. (1988). Simultaneous identification and banding of human chromosome material in somatic cell hybrids. Cytogenet. Cell Genet. 48, 103-106.

Ullu, E., Tschudi, C. (1984). Alu sequences are processed 7SL RNA genes. Nature 312, 171-172.

van Es, H.H.G., Bout, A., Liu, J., Anderson, L., Duncan, A.M.V., Bosma, P., Oude Elferink, R., Jansen, P.L.M., Roy Chowdhury, J., Schurr, E. (1993). Assignment of the human UDP glucuronosyltransferase gene (UGT1A1) to chromosome region 2q37. Cytogenet. Cell Genet. 63, 114-116.

Vance, J.M. (1991). Hereditary motor and sensory neuropathies. Journal of Medical Genetics 28, 1-5.

Vidal, S., Belouchi, A.-M., Cellier, M., Beatty, B., Gros, P. (1995a). Cloning and characterization of a second human NRAMP gene on chromosome 12q13. Mamm. Genome 6, 224-230.

Vidal, S., Tremblay, M.L., Govoni, G., Gauthier, S., Sebastiani, G., Malo, D., Skamene, E., Olivier, M., Jothy, S., Gros, P. (1995b). The *Ity/Lsh/Bcg* locus: natural resistance to infection with intracellular parasites is abrogated by disruption of the *Nramp1* gene. J. Exp. Med. 182, 655-666.

Vidal, S.M., Malo, D., Vogan, K., Skamene, E., Gros, P. (1993). Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. Cell 73, 469-485.

Vidal, S.V., Epstein, D.J., Malo, D., Weith, A., Vekemans, M., Gros, P. (1992). Identification and mapping of six microdissected genomic DNA probes to the proximal region of mouse chromosome 1. Genomics 14, 32-37.

Viegas-Pequignot, E., Lin, L.Z., Dutrillaux, B., Apiou, F., Paulin, D. (1989). Assignment of human desmin gene to band 2q35 by nonradioactive in situ hybridization. Hum. Genet. 83, 33-36.

Vogel, F., Motulsky, A.G. (1986). Human genetics: problems and approaches. Springer, Berlin.

Wada, M., Little, R.D., Abid, F., Porta, G., Labella, T., Cooper, T., Della Valle, G., D'Urso, M., Schlessinger, D. (1990). Human Xq24-28: approach to mapping with yeast artificial chromosome. Am. J. Hum. Genet. 46, 95-106.

Walter, M.A., Spillett, D.J., Thomas, P., Weissenbach, J., Goodfellow, P.N. (1994). A method for constructing radiation hybrid maps of whole genome. Nature Genet. 7, 22-28.

Wang, M., Chen, X-N., Shouse, S., Manson, J., Wu, Q., Li, R., Wrestler, J., Noya, D., Sun, Z-G., Korenberg, J., Lai, E. (1994). Construction and characterization of a human chromosome 2-specific BAC library. Genomics 24, 527-534.

Ward, T., Davies, K.E. (1993). The leading role of STSs in genome mapping. Hum. Mol. Genet. 2, 1097-1098.

Webb, A.C., Collins, K.L., Auron, P.E., Eddy, R.L., Nakai, H., Byers, M.G., Haley, L.L., Henry, W.M., Shows, T.B. (1986). Interleukin-1 gene (IL1) assigned to long arm of human chromosome 2. Lymphokine Res. 5, 77-85.

Weber, J.L. (1990). Informativeness of human $(dC\text{-}dA)_n(dG\text{-}dT)_n$ polymorphims. Genomics 7, 524-530.

Weber, J.L., May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44, 388-396.

Weiner, A.M., Deininger, P.L., Efstratiadis, A. (1986). Nonviral retroposons: Genes, pseudogens, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. 55, 631-661.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., Lathrop, M. (1992). A second-generation linkage map of the human genome. Nature 359, 794-801.

Welcher, A.A., Suter, U., De Leon, M., Snipes, G.J., Shooter, E.M. (1991). A myelin protein is encoded by the homologue of a growth arrest-specific gene. Proc. Natl. Acad. Sci. USA 88, 7195-7199.

White, J.K., Shaw, M-A., Barton, C.H., Cerretti, D.P., Williams, H., Mock, B.A., Carter, N.P., Blackwell, J.M. (1994). Genetic and physical mapping of 2q35 in the region of NRAMP and IL8R genes: identification of a polymorphic repeat in exon 2 of NRAMP. Genomics 24, 295-302.

Winokur, S.T., Schutte, B., Weiffenbach, B., Washington, S.T., McElligott, D., Chakravarti, A., Wasmuth, J.H., Altherr, M.R. (1993). A radiation hybrid map of 15 loci on the distal long arm of chromosome 4, the region containing the gene responsible for Facioscapulohumeral Muscular Dystrophy (FSHD). Am. J. Hum. Genet. 53, 874-880.

World Health Organization. (1994). TB: a global emergency. WHO report on the TB epidemic. WHO/TB/94.177.

Yoshiura, K.I., Kubota, T., Soejima, H., Tamura, T., Izum[illegible]a, Y., Nikawa, N., Jinno, Y. (1994). A comparison of GC content and the proportion of *Hu/kpnI*-repetitive sequences in a single dark- and light-band region from a human chromosome. Genomics 20, 243-248.

Yunis, J.J. (1976). High resolution of human chromosomes. Science 191, 1268-1270.

Zietkiewicz, E., Labuda, M., Sinnett, D., Glorieux, F.H., Labuda, D. (1992). Linkage mapping by simultaneous screening of multiple polymorphic loci using Alu oligonucleotide-directed PCR. Proc. Natl. Acad. Sci. USA 89, 8448-8451.

Zucman, J., Delattre, O., Desmaze, C., Azambuja, C., Rou[illegible], G., de Jong, P., Aurias, A., Thomas, G. (1992). Rapid isolation of cosmids from chromosomal subregions by differential Alu-PCR hybridization on chromosome 22-specific libraries. Genomics 13, 395-401.

Zuo, J., Robbins, C., Baharloo, S., Cox, D.R., Myers, R.M. (1993) Construction of cosmid contigs and high-resolution restriction mapping of the Huntington disease region of human chromosome 4. Hum. Mol. Genet. 2, 889-899.