Positional cloning of *nude*, a fork head transcription factor: Genetic, Physical and Transcription Maps of the Region and Mutations in the Mouse and Rat

by

Julia Angela Segrè

Submitted to the Department of Biology in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY in Biology at the

Massachusetts Institute of Technology

December 1995

© 1995 by Julia A. Segrè. All rights reserved.

Signature of Author

Department of Biology
December 18, 1995

Certified by

Eric S. Lander
Professor, Department of Biology
Thesis Supervisor

Accepted by

Frank Solomon
Chairman, Biology Graduate Committee
ACKNOWLEDGEMENTS

Graduate school is a long path, that I started on years ago. For so many reasons, my gratitude goes to my family: Mom, Dad, Katia, and Michele. Mom as a role model - tenacity and joie-de-vivre; Dad, the quintessential scholar with style, and charisma. Katia, always the big sister, for your advice and experience. Michele for your humanity, showing me that my way is not better. Grandma and Grandpa for the intellectual atmosphere they foster and the love and good food they have shared with me. I wish that I had known Nonni better, but I feel they left their indelible mark on us all. Over the years, a few very special people have been added to the family - Bettina, Frank, Daniel, Alexandra - my love and appreciation to all of you.

My second family has been the Lander Lab - an incredible group of individuals. Eric, my appreciation is unbounded for your academic and personal advice. Working with you, has been one of the most rewarding experiences of my life. Thank you for helping me pick primers, motivating me with your enthusiasm, and for being there when I really needed you most. To the old Lander gang (Bill, Howard, Alix, Hillary, Nic, Armand) for setting the tone of the lab - making it always fun to come into the lab. Jennifer - I can’t imagine with whom I would rather work - professional and personal synergism. Armand, who could have imagined that we would have survived so many happy years together (without any serious damage). Kenro, the most energetic and generous of spirits. Johanna for keeping me happy and always being a friend and colleague. Howard for spending Saturdays in the lab with me and administering baby therapy when necessary. Susan for all the great advice - you are the best! Thanks also to the rest of the Lander Lab family: Jenny Sue, Bruce, Arend, Eileen, Mary Pat, Mark, Tau Mu, Leonid, Anne, Shau Neen, Karen, Monique, Tina, Melanie, David, Vicki, Aaron, Renee, Lori, Jessica, Daniel and David. Thanks to my teachers who came before graduate school: Robert Stauffer, Norton Starr, Patricia O'Hara, David Cox.

The third leg of the stool has always been my friends and soulmates: Noga - No time ever passes between seeing you. I’m glad we always find each other again. Anna - Sometimes first impressions are correct. Peggy - You tethered me to this earth when I thought I might float away. Mark for all the acronyms we’ve shared. Kevin for the years of friendship - past, present and future. The MBL gang for sharing so much of yourselves in such a short period of time. Tira los dados de la vida. The friends in California, who made all the difference - for advice, hot tubing, and home cooked meals. And finally, all the wonderful friends I’ve made while wandering the floors of the Whitehead.
TABLE OF CONTENTS

Preface p. 6

Abstract p. 7

Chapter One: Introduction to the nude mouse: Etiology of the hairless and athymic phenotypes. p. 8

Chapter Two: Genetic map of the nude locus: Mapping with simple sequence length polymorphisms and direct isolation of polymorphic markers linked to nude by genetically directed representational difference analysis. p. 26

Chapter Three: Physical map of the nude region: Chromosomal walks in yeast artificial chromosome and bacteriophage P1 clones. p. 46

Chapter Four: Transcription map of the nude region: Direct cDNA selection and exon trapping to isolate expressed sequences in a region dense in transcribed units. p. 64

Chapter Five: Mutation detection and expression studies of the nude gene, a novel fork head transcription factor. p. 80

Materials and Methods p. 107

Figures

Tables
Appendix I: Gene fragments from the nude region:
Sequence, primers, physical map location and
similarity to Genbank entries.

Appendix II: Lisitsyn, N.A., Segre, J.A., Kusumi, K., Lisitsyn, N.M.
Direct isolation of polymorphic markers linked to a trait
by genetically directed representational difference analysis.

Appendix III: Segre, J.A., Nemhauser, J.L., Taylor, B.A., Nadeau,
Genetic, Physical and Transcription Maps of the Region and
Mutations in the Mouse and Rat. *Genomics* 28: 549-559.
Preface

This thesis is organized into five chapters, each describing a specific aspect of the project. As many of the Figures, Tables and Materials and Methods are referred to in more than one chapter, they are included at the end of the thesis. The first appendix is too long to include in the text of the fourth chapter. Some of the work presented in this thesis has already been published and those journal articles appear as Appendix II and III of this thesis.

The nude mapping crosses were set up in collaboration with Joseph Nadeau and Benjamin Taylor at The Jackson Laboratories. The experiments with GD-RDA, described in Chapter Two, were done in collaboration with Nikolai Lisitsyn in Michael Wigler's laboratory at Cold Spring Harbor Laboratories. The in situ hybridization experiments were initiated when I visited Gail Martin's laboratory at UCSF. The cosmid rescue of the nude phenotype was done in conjunction with Hisanori Kurooka in Tasuku Honjo's laboratory at Kyoto University.
Mutations in the *nude* locus in mice and rats produce the pleiotropic phenotype of hairlessness and athymia, resulting in a severely compromised immune system. To identify the underlying causative gene, we utilized modern tools and techniques of positional cloning. Spanning the region in which the *nude* locus resides, we constructed a genetic map of polymorphic markers. To specifically target closely-linked genetic markers to the *nude* region, we developed a method based on the principles of transmission genetics and a recent subtractive cloning procedure, representational difference analysis. We constructed chromosomal walks in yeast artificial chromosomes and bacteriophage P1 clones that span the region between the closest genetic markers that flank the *nude* locus. We assembled a transcription map of the *nude* region from gene fragments, obtained by direct cDNA selection and exon trapping. We identified 7 novel mouse genes with strong similarity to previously identified genes from *Drosophila*, *C. elegans*, rat, rabbit, and human; and 3 previously identified mouse genes. Based on our transcription mapping results, we present a novel approach to estimate the number of genes in a region and estimate that the *nude* locus resides in a region approximately three-fold enriched for genes. Disruptions in a novel fork head domain transcription factor from the *nude* region, *Hfh11*, were detected in all four *nude* alleles. Integration of a cosmid clone containing the wild-type *Hfh11* genomic locus into fertilized *nu/nu* eggs, corrected the hairless phenotype, but not the thymic defect. These results demonstrate that *Hfh11* is the *nude* gene and suggest that the gene is subject to complex regulation. Finally, the expression pattern of *Hfh11* is consistent with the observed *nude* phenotype. *Hfh11* RNA is expressed in the adult thymus, initiating in the developing embryo as thymic organogenesis occurs. High levels of *Hfh11* expression are detected during the active growth phase of the hair follicle in the keratinized shaft.
Chapter One

Introduction to the *nude* mouse:
Etiology of the hairless and athymic phenotypes.

A brief history of the discovery of the *nude* phenotype

The first reference to the *nude* mouse was a brief report in 1962 that a new, transmitting hairless mutant had been picked up at the Virus Laboratory in Glasgow, Scotland (Issacson, et al., 1962). In no way did this one page report anticipate the vast literature that would accumulate about this world-famous mutant.

A breeding colony of *nude* mice was established by Dr. S. Flanagan, who began an intensive genetic analysis of the *nude* locus (Flanagan, 1966). He had an extremely difficult time maintaining his *nude* mouse colony and suspected that a Toxoplasma gondi infection was the cause. Flanagan sent some *nude* mice to a well-known expert on toxoplasmosis, but no sign of infection could be found in the mice. However, Dr. E.M. Pantelouris, who ran the lab next door, discovered the secret to the mutant’s high rate of infection: *Nude* mice are athymic (Pantelouris, 1968). Reinvestigation showed that the mutation in the *nude* locus produced the remarkable pleiotropic phenotype of hairlessness and athymia. An example of *nude* mutant mice is shown in Figure 1.

The discovery of the athymic *nude* mouse came concurrent with a growing understanding of the body’s immune system. The Bursa of Fabricius in the chicken, responsible for producing what we now call the B lymphocytic system, was first described by Glick et al. in 1956. The profound loss of cell-mediated immunity (now known as the T cell system) as a result of a neonatal thymectomy in a mouse was first described by Miller in 1961. In fact, in 1953, workers at the Rowett Institute in Scotland knew that their hairless rats were athymic, but the biological significance of a congenitally athymic rodent was not appreciated and no description of the mutant was published. Only when hairless athymic rats reappeared twenty years later, was a report published and a breeding colony established (Festing, et al., 1978; Festing, 1981).
Jorgen Rygaard states "The history of the nude mouse - in the mind of the scientific world - starts on January 27, 1968, with Dr. E.M. Pantelouris' report in Nature on the "Absence of Thymus in a Mouse Mutant" (Rygaard, 1991). Most of the work done with nude mice uses them as immune-compromised hosts for transplantations of tumors and tissues and very few groups have looked at the etiology of the nude phenotype (Bach-Mortensen, et al., 1976; Bastert, et al., 1977; Houchens, et al., 1978). This thesis presents the work to clone the nude gene based on its chromosomal location and experiments to understand the nude phenotype based on the identification of the gene.

HAIR

embryonic and adult organogenesis

The mouse hair cycle is a fascinating example of organogenesis, tissue development and differentiation. Organogenesis is recapitulated post-natally every 21 days for each of the 500,000 hair follicles that populates the adult mouse skin (Potten, 1985). This section will provide the background on the stages of follicle development, the parts of the hair, the types of hair, and a detailed description of the nude mouse's hairless phenotype.

A hair bulb develops from an invagination of the epidermis (ectodermal in origin), the base of which surrounds a condensation of mesodermal cells known as the dermal papilla (DP). Epidermal cells, known as the germanititive hair matrix cells, rapidly proliferate and differentiate into all parts of the hair fiber and the surrounding sheath. Melanocytes (neural crest in origin) migrate into the skin and secrete pigments, which are linked to proteins, incorporated into granules, and then taken up by the growing hair shaft. Hair bulbs develop embryonically at 14 days post coitum (d.p.c.) and follicles erupt from the skin at 6 days post-natal (Rugh, 1990).

The stages of hair follicle development are tightly controlled with signals coming from both the epidermis and the dermis. Although the molecular nature of this signaling pathway is not known, tissue recombination experiments and histological examination point to the origin of the signals. To perform these experiments, epidermis and dermis of skin from different body regions of embryonic mice, chicks and lizards are recombined as explants. The first message comes from the dermis, instructing the epidermis to initiate a thickening (placode) and then a
downgrowth (plug). The mesenchymal condensates of the DP anlage appear just prior to the very early signs of epidermal placode formation. Even epidermis from a hair-free region such as the footpad forms placodes when combined with ventral dermal mesenchyme. In fact, mouse dermis can also initiate feather buds when combined with chick foot epidermis, or scale placodes when recombined with lizard epidermis. The second message comes from the hair matrix cells, instructing the underlying mesenchymal cells to form an organized dermal papilla. This message probably depends on selective cell adhesion because DP formation occurs only if epidermis and dermis from the same species are recombined. The third message is transmitted from the DP to the adjacent cells of the hair plug, stimulating them to divide rapidly (Sengel, 1986; Sengel, 1975).

Hair growth is cyclic, proceeding from the active growing phase (anagen) to a degenerative phase (catagen) when the matrix cells cease proliferating, leading to a drastic shortening of follicular length. As the hair follicle regresses, the dermal papilla moves up to lie below the hair matrix stem cells, which are positioned below the sebaceous gland (Jones, et al., 1995; Lavker, et al., 1995; Wilson, et al., 1994). The follicle then enters a period of relative inactivity (telogen) until the cycle begins anew with another proliferative phase (Figure 1). The hair stem cells differentiate and extend into the dermis to form a plug, surround the dermal papillae and renew hair follicle differentiation. The new hair grows into the mouth of the old one, which is shed. The recapitulation of hair follicle initiation and growth seems to rely upon the same sequence of signaling molecules that established the first follicles (Hardy, 1992; Messenger, 1993). In mice and rats, the cycle of pelage hair growth is approximately 21 days in a rostral to caudal direction. In the wild, the cycles vary, depending on daylight length and endocrine activity. Humans have a more mosaic pattern of hair cycling, with each follicle type following its own internal clock (Ebling, et al., 1983). None of the signaling molecules, controlling the entry into a new phase of the hair follicle cycle, has yet been identified. However, intriguing candidates exist since many signaling molecules are shown to be expressed in the hair and the targeted disruption of some even giving interesting hair phenotypes (Dolle, et al., 1990; Hebert, et al., 1994; Hirai, et al., 1989; Jones, et al., 1991; Luetteke, et al., 1994; Luetteke, et al., 1993; Mann, et al., 1993; Miettinen, et al., 1995; Ruberte, et al., 1990; Sibilia, et al., 1995; Threadgill, et al., 1995).
The mature hair follicle consists of several cylindrical, concentric cell layers, schematically represented in Figure 2 and as an electron micrograph in Figure 3. The hair matrix daughter cells differentiate into a specific hair cell type, depending upon their position relative to the longitudinal axis of the follicle. The hair shaft is comprised of three distinct layers: the cuticle is a series of thin, overlapping scales that give hair its serrated appearance; the cortex, the bulk of the hair shaft, is a hollowed cylinder of hardened cornified material, made up of longitudinally oriented filaments; the medulla (absent for some portions of the hair fiber) is the central cells, separated by air-filled spaces, arranged in regular rows. The hair matrix cells also give rise to the inner root sheath, which encases and grows with the developing hair fiber, but does not erupt from the skin. The outer root sheath layer of cells is continuous with the epidermis, and attaches to the inner root sheath. These cell layers can be distinguished by their gross morphology and the specific proteins produced in every cell type.

The major structural proteins in the hair follicle are keratin filaments. Keratins belong to a family of structural proteins called intermediate filaments (IFs) because their diameter of 8 to 10 nm falls intermediate between that of actin filaments and microtubules. Keratins are divided into two classes: type I keratins are small (40-56 kDa) and acidic (pKᵢ 4.5-5.5), whereas type II keratins are larger (53-67 kDa) and more basic (pKᵢ 5.5-7.5). Keratin IF have as a basic structural component 400 to 500 amino acid residues, arranged in sequences of heptad repeats that form an α-helical domain. Keratin IFs are obligate heteropolymers of one chain from each of the two superfamilies which interact in the α-helical domain to produce an initial heterodimer of a coiled-coil. The heterodimers go on to form tetramers and then higher order structures resulting in an extensive network of cross-linked keratin filaments (Coulombe, 1993; Stewart, 1993). All three parts of the hair shaft are keratinized. As determined by two dimensional gel electrophoresis, eight keratin filaments are expressed in the hair, 4 from each class (Heid, et al., 1986). The N- and C-terminal domains of hair specific keratins are distinct from all other IFs because of the relatively high content of cysteine residues (Gillespie, 1990).

The different kinds of hair that cover the body are: Zigzag (70%) - fine, varied in length with more than one sharp bend; Awl hairs (28%) - 0.5 cm long and straight; Guard hairs (2% of hairs) - coarse, 1 cm long, and straight;
and Auchene hairs (<1%) - fine hairs with a single bend. Besides the hairs that form the coat there are seven specific types of hair: vibrissae (whiskers), cilia (eyelashes), tail hairs, ear hairs, hairs around the feet, hairs around the nipples, hairs in the perianal/genital regions (Dry, 1926).

There are at least three possible causes of genetic hairlessness in the mouse: suppression of follicle initiation, which is seen in the *crinkled* or *ragged* mutants; abnormal keratinization of the hair shaft or increased resistance of the epidermis to the erupting hair follicle, which is seen in the *naked* mutant; or disruption of the hair cycle, which is seen in the *hairless* mutant (Sundberg, et al., 1991). These classes can be distinguished by examining the hair follicle cycle in the skin of the mice.

**nude hairless phenotype**

*Nude* mice fall into the second class and are not, in fact, hairless. Histological examination of *nude* skin reveals a normal number of hair bulbs, but the hair shafts are bent and crippled, rarely penetrating the epidermis. Whereas in normal skin the follicles are sloped and parallel to one another and the sebaceous glands are closely adjacent to the hair shaft; the follicles of *nude* skin are grossly distorted and widened and the sebaceous glands are abnormally located either at the base of the hair canal or embedded in the dermis. At the end of the growth phase the skin of the *nude* mice decreases in thickness and the follicles shorten as in normal skin. The total number of follicles in *nude* mice decreases with age (Flanagan, 1966; Kopf-Maier, et al., 1990).

The hair defect of *nude* mice could be due to imperfect keratinization of the hair shaft or to increased resistance of the epidermis to the erupting hair tip. To determine if the hair follicle is properly keratinized, Flanagan made use of the fact that the the non-α-helical domains of hair-specific keratin IF, contain a large number of cysteine residues with free sulphydryl groups. During keratinization of the hair follicle, the cysteine sulphydryl groups are oxidized to form disulfide bonds which presumably contributes to filament assembly and increases the strength and rigidity of the hair fiber. Staining sections of the skin of 6-day-old normal and *nude* mice with a reagent that reacts with the sulphydryl groups of the keratin precursors, 1-(4-chloromercuriphenylazo)-naphthol-2, Flanagan observed that in normal skin the mid-follicle region reacted intensely and ended sharply at the distal end of
the mid-follicle region. In the follicles of nude mice, there was a marked
decrease in the sulphydryl reaction in the mid-follicle region, suggesting that
the hair follicles of nude mice are deficient in keratin filament (Flanagan,
1966). All types of pelage hairs (guard, awl, auchene, and zigzag) are abnormal
in nude mice (Sundberg and Schultz, 1991). The vibrissae and eyelashes are
delayed in development, but present in adult mice (Flanagan, 1966 and
personal observation). This suggests that the nude gene affects the
expression of structural protein(s) necessary in hair follicles.

The hair follicle is an ideal mammalian system for studying pattern
formation and organogenesis because of the spatially and temporally well
defined developmental and differentiation programs. Experimentally, the
system is easily manipulated: There exist many cell-type specific promoters to
target gene expression and these genetic manipulations usually give a viable,
interpretable phenotype. As well, follicles form and grow completely
normally from organ cultured embryonic skin, indicating that hair
organogenesis does not require ongoing exposure to circulating factors
(Hardy, 1969). To understand the development of the hair follicle and later
the pathogenesis of diseases of the hair, it is necessary to determine how these
structures are formed, what molecules are involved, and how they are
regulated. The block in the hair cycle of nude mice provides an entry point
into the study of hair morphogenesis.

THYMUS

Embryonic development

The thymus is a bilobed organ located in the upper part of the anterior
mediastinum; i.e. lying in the thoracic cavity above the heart. The thymus is
the major site for generation of immunocompetent T-cell lymphocytes: A
repertoire of mature T-cell receptor positive cells specific for self major
histocompatibility complex molecules is selectively exported out of the
thymus. The stromal cells provide the appropriate microenvironments and
signals necessary for differentiation and maturation of young T cells. Mouse
mutants have shown that T-cell maturation and thymic organogenesis are co-
dependent events (Bosma, et al., 1983; Mombaerts, et al., 1992; Pantelouris,
1973). Transplantation studies show that the thymic dysgenesis in nude mice
is stromal in origin: (i) Bone marrow from a nude mouse can repopulate the
thymus of an irradiated mouse. (ii) Injection of fetal liver cells fails to extend
the survival of nude recipients or to increase lymphocyte numbers (Pantelouris, 1973; Wortis, et al., 1971). This section describes thymic development in both normal and nude mice.

The pharyngeal arches and pouches are transient embryonic structures that give rise to many of the organs in the head and neck region of vertebrates. By the ninth day, four endodermal pouches (pharyngeal) have appeared successively on either side of the primitive foregut. The first pouch is the most cranial and also the largest. Each successive pouch is smaller and more caudal than the preceding one. For each of these endodermal evaginations, there is a corresponding ectodermal invagination or branchial cleft (Figure 4) (Gilbert, 1995; Kaufman, 1992).

To present the etiology of the nude athymic phenotype it is essential to start with normal thymic development. A meticulous three-dimensional reconstruction study from serial sections of the fetal thymus by Cordier and Haumont has given the most detailed description of thymic development for both normal and nude mice: For each day of gestation from 9-17 d.p.c., 10-15 nude and normal embryos were fixed, embedded and serially sectioned into 5μm sections. The sections were stained to distinguish the endoderm and the ectoderm and then photographed. The ectodermal and endodermal components were outlined in each photograph and reconstructed to form a three-dimensional image. This work precisely traced the ectodermal and endodermal contributions to the thymus (Cordier, et al., 1980).

At day 9, two thymic anlage (derivative stromal structures of the thymus) begin to form from three distinct embryonic origins: endoderm from the third pharyngeal pouch, ectoderm from the third branchial cleft, and neural crest-derived mesoderm (LeLievre, et al., 1975). The distal end of the third pharyngeal pouch stretches distally to contact the third ectodermal cleft. At the point of contact, the two epithelia are back to back and clearly distinguishable from each other. At 9.5 d.p.c., the ventral lip of the third branchial cleft (ectoderm) has thickened and protruded into the endoderm of the third pouch whereupon the two tissues continue to develop together. On the eleventh day, the ectoderm begins a rapid proliferation, covering the distal edge and a small part of the cranial and caudal walls of the endoderm. By the twelfth day, the ectoderm has surrounded the ventral and external surfaces of the endoderm, leaving only the extreme caudal region uncovered. By 12.5 d.p.c. the ectoderm has covered all surfaces of the endoderm. This
structure is maintained somewhat even in the adult thymus, where the embryonic ectodermal contributes more significantly to the outer region (cortex) and the endodermal cells to the inner region (medulla).

At day 11.5, the thymic lobes undergo a double rotation and migration in a caudal and medial direction toward the heart. By day 14, migration of the thymic anlage is complete and the two thymic lobes are juxtaposed in the anterior mediastinum. The parathyroids develop at about 11.5 d.p.c. from a small region of the dorsal part of the third endodermal pouch. The parathyroids rotate with the thymic anlage, moving ventrally to the level of the thyroid. The parathyroids only separate from the thymic primordium at 13 d.p.c. and within the next two days become partly incorporated into the thyroid (Cordier and Haumont, 1980).

**nude athymic phenotype**

Cordier and Haumont clearly demonstrated that the thymic defect in nude mice is not due to an absence of thymic precursor cells, but rather to a failure of the ectodermal cells to proliferate and cover the endodermal component. They detected no deviation from normal embryonic development of the pharyngeal region in nude mice between the ninth and eleventh days. However at 11.5 d.p.c., even though the endoderm had developed normally, the ectoderm of the third cleft did not proliferate and as such, covered only the distal extremity and a small part of the cranial and caudal walls of the third pouch. By the twelfth day in the nude mouse, the endodermal portion of the primitive thymus had the same shape, dimensions and orientation as in the normal mouse. However, the ectoderm still only covered the mid-portion of the ventral endoderm, leaving the caudal region bare. By 12.5 d.p.c. the ectoderm had receded while the endoderm continued to proliferate, creating a primitive thymus that is essentially completely endodermal. The parathyroid primordium appeared at the same time in nude and normal mice (between 11 and 11.5 d.p.c.). The rotation of the "thymic anlage" and the parathyroids toward the heart was normal in nude mice, but it occurred slightly later. The thymic rudiment was not invaded either by blood vessels or by lymphoblasts. At the proper stage, however, the parathyroid anlage was engorged by blood vessels.

Morphometric analysis of the volume of the thymus in normal and nude mice showed a striking difference. In normal mice, the volume of the
thymus increases tenfold between the twelfth and the fourteenth day, partially due to the infiltration of the lymphoblasts from the fetal liver. Until the end of embryonic development, the thymus continues to grow exponentially. In the nude mouse, the volume of the thymus merely doubled between the twelfth and the fourteenth day, when it stopped increasing in size completely (Cordier and Haumont, 1980).

To understand the role of the various stromal components in normal thymic development and to assay the stromal/lymphoid etiology of the nude phenotype, Van Ewijk’s group has developed a panel of monoclonal antibodies, directed to various types of stromal cells of the mouse thymus: TR4+ recognizes cortical epithelial cells by day 13; TR5+ recognizes medullary epithelial cells by day 13. Their experiments showed that the majority of stromal cells of the nude thymus are negative for all the antibodies tested. In addition, they showed that the organization of the stroma of thymic lobes remains intact for at least 11 days in organ culture, even when depleted of lymphoid cells (Van Vilet, et al., 1985). This strengthens the argument that the lack of organization of the nude thymus is not simply due to the absence of lymphoid cells, but rather to a true stromal defect.

The ectodermal stromal origin for the thymic dysgenesis is supported by experiments in the developing chick embryo. If the ectoderm of the third and fourth clefts is removed before it fuses with the endoderm of the third and fourth pouches, then the endoderm produces only a rudimentary thymus, while the parathyroids develop normally (Hammond, 1954).

The nude mouse has sometimes been compared to the congenital human athymia condition, known as Di George’s syndrome (Di George, 1965; McCusick, 1990). However, patients with Di George's syndrome also lack the parathyroids, have reduced thyroid tissue, bear craniofacial abnormalities and often are afflicted with cardiac and arterial defects. This human disorder more closely resembles mice with targeted disruptions in the Hoxa-3 gene (Chisaka, et al., 1991; Manley, et al., 1995).

POSITIONAL CLONING

A biochemical analysis of the hairless or the athymic phenotype is not sufficient to allow a direct deduction of the block in the developmental pathway in nude mice. However, since both defects are the result of an inherited disorder, the power of genetics can be used to identify the
underlying defect. Positional cloning is the general strategy to isolate a gene based on their chromosomal location without prior knowledge of its biochemical function. This method bypasses the initial need to understand the primary defect, but allows one to return and address those questions once the gene is cloned.

Positional cloning requires localizing the gene to a specific chromosomal region, and analyzing the genes in the region as candidates. Briefly, genetic markers that co-segregate with the locus in a cross give a gross chromosomal localization. A contiguous set of cloned DNA inserts that span the region is used to further refine the smallest genetic and physical region of the gene. Multiple alleles with chromosomal rearrangements, deletions or translocations, can further narrow the chromosomal location. Expressed sequences are identified from the region based on hallmarks of open reading frames or specific hybridization to the physical DNA clones. Candidate genes from the region are evaluated based on expression patterns and differences observed between wild type and affected individuals. Each stage of this process requires emerging tools and techniques, including a dense sets of genetic markers, libraries of large insert clones, and new methods for identifying transcription units in a physical region.

As the prospects for positional cloning improve, the difficulty of projects that can be tackled increases. When this project began in 1992, T (Brachyury) was the only gene that had been positionally cloned in the mouse -- and then only with heroic scientific efforts and an extraordinary set of reagents, including a dozen alleles, many with chromosomal deletions and rearrangements (Herrmann, et al., 1990). In 1992, cystic fibrosis was the only gene discovered by a positional cloning approach in humans that did not possess alleles with gross cytogenetic rearrangements (Collins, 1992). Chromosomal abnormalities were necessary to delineate an extremely small region in which the gene must lie. With the availability of new tools and techniques, seven additional monogenic traits have been positionally cloned in the mouse from genetically defined regions of several hundred kilobases of DNA (Bultman, et al., 1992; Cordes, et al., 1994; D'Arcangelo, et al., 1995; Hirotsume, et al., 1995; Kingsley, et al., 1992; Miller, et al., 1993; Patl, et al., 1995; Segre, et al., 1995; Vidal, et al., 1993; Zhang, et al., 1994). As the regions that can be scanned for mutations in candidate genes grows, research is focusing on attaining similar goals with complex traits. Finally, to understand the
action of the cloned gene in the development of both normal and affected mice is different for every mutant. Rapid progress is being made in many fields of biology, but it remains a challenge to connect the pathway between the defective gene and the mutant phenotype or between the wild type gene and normal development.
REFERENCES


Chapter Two

Genetic map of the *nude* locus:
Mapping with simple sequence length polymorphisms and direct isolation of polymorphic markers linked to *nude* by genetically directed representational difference analysis.

ABSTRACT

Early genetic studies demonstrated that *nude* segregates as a single autosomal locus on mouse Chromosome 11. We genetically mapped the *nude* locus in 2000 meioses in three separate F$_2$ intercrosses. Initial mapping studies with simple sequence length polymorphisms localized *nude* to a 1.4 cM interval between *D11Mit7* and *D11Mit34*. To specifically target genetic markers to the *nude* region, we developed a method based on the principles of transmission genetics and a recent subtractive cloning procedure, termed representational difference analysis. This method successfully targeted three genetic markers to an interval of less than 0.5 cM around the *nude* locus. Further genetic markers were developed from yeast artificial chromosome and bacteriophage P1 clones that span the *nude* region. The resolution of 2000 meioses and the extremely dense set of genetic markers localized *nude* to a 370 kb interval.
INTRODUCTION

A deep resource of mouse variants with developmental, physiological or behavioral defects exists in the thousands of mutants with single gene alterations and the tens of inbred strains that have been collected over a century of breeding (Festing, 1979; Green, 1989). Although a description of the pathology of many of these mutants exists, to understand the disruption in the underlying developmental pathway requires identifying the causative gene. Positional cloning is the isolation of a gene based on its chromosomal location without prior knowledge of its biochemical function (Collins, 1992). Although this method has been broadly successful in the fruit fly and the nematode, the tools necessary to positionally clone a gene in the mouse are just being developed. Positional cloning depends upon first determining the chromosomal location of the locus, defined by tightly-linked genetic markers that segregate with the phenotype in a cross. Although linkage groups, consisting of phenotypic and protein variants, have existed for all of the mouse chromosomes for two decades, mapping a new locus was cumbersome (Davisson, et al., 1989).

Genetic mapping in the mouse changed dramatically with the increased resolving power of interspecific crosses and DNA-based genetic markers. Interspecific crosses exploit the inherent genetic diversity between a laboratory strain and a distantly related species of Mus; e.g. Mus spretus, Mus musculus castaneus or Mus musculus molossinus (Avner, et al., 1988). The recognition that minor variants in DNA sequence could be followed segregating in a cross, provided a virtually inexhaustible supply of genetic markers (Botstein, et al., 1980). As well, DNA-based polymorphisms provide an immediate entry into the genome, necessary for physical mapping. The first set of DNA based genetic markers were probes that detected restriction fragment length polymorphisms (RFLPs) between two DNA sources (Gusella, et al., 1983). RFLP mapping is extremely useful to determine the chromosomal location of a novel cloned gene. However, to map a phenotype segregating in a cross, RFLPs are not ideal as genetic markers because they require a lot of DNA to genotype each sample, are time consuming to process, and have a low polymorphism rate among inbred laboratory strains. An alternative source of DNA polymorphism is based on variation in the length of simple sequence repeats (SSRs) (e.g. (CA)_n or
(GA)\textsubscript{n}) that densely populate the mouse genome (Love, et al., 1990; Stallings, et al., 1991). These simple sequence length polymorphisms (SSLPs) are easily typed by PCR with primers flanking the SSR. Genetic maps, based on RFLPs and SSLPs independently, and integrated together, have been constructed for the mouse genome (Copeland, et al., 1991; Copeland, et al., 1993; Dietrich, et al., 1992). These genetic maps, consisting solely of DNA-based polymorphisms, have improved prospects for positional cloning in the mouse.

To directly target tightly-linked genetic markers to a locus is essential for organisms with nonexistent genetic maps and ideal for organisms with rudimentary genetic maps. Although there are now 6,000 SSLP that have been genetically mapped in the mouse, when we initiated this project, the map of the mouse consisted of 317 SSLP markers with an average spacing of 4.3 cM between markers. Our method to target markers to a specific region builds on a recently described subtractive technique called representational difference analysis (RDA) for identifying differences between two DNA samples, referred to as Tester and Driver (Lisitsyn, et al., 1993). Specifically, RDA is designed to clone restriction fragments that can be amplified by PCR from Tester but not Driver -- either because the corresponding sequence is completely absent from the Driver due to a homozygous deletion or because it is contained in a small restriction fragment in the Tester but a large and, therefore, poorly amplifiable restriction fragment in the Driver. Thus, RDA can produce clones that detect RFLPs between Tester and Driver. To generate genetic markers linked to a trait, one needs Tester and Driver samples with the property that the Driver contains all of the alleles present in the Tester except in the region surrounding the target gene. As we describe below, such samples can be constructed by using classical transmission genetics.

Here we describe the mapping of the \textit{nude} locus in 2000 meioses in three separate F\textsubscript{2} intercrosses. Initial mapping studies with SSLPs localized \textit{nude} to a 1.7 cM interval between \textit{D11Mit7} and \textit{D11Mit34}. Two implementations of RDA were carried out, using congenic strains and F\textsubscript{2} intercross progeny. Three clones were produced and all mapped within 0.5 cM of \textit{nude}, which comprised less than 1/4000 of the mouse genome. Finally, a dense set of genetic markers was developed from yeast artificial chromosome and bacteriophage P1 clones that span the \textit{nude} region. The
resolution of 2000 meioses and the extremely dense set of genetic markers localized *nude* to a 370 kb interval.

**Alleles of nude rodents**

Over the years, at least four *nude* alleles have arisen independently among rodents. Although many hairless mutants possessing thymuses have been found, all of these mutants have been mapped to genomic regions away from the *nude* locus. This suggests that the hairlessness and athymia are a result of mutations in the same gene. Two alleles of *nude* exist in the mouse, *nude* (*nu*) and *nude-streaker* (*nu*<sup>str</sup>); Two alleles of *nude* exist in the rat: *rnu* and *rnu*<sup>N</sup>. As well, there exists an athymic hairless guinea pig (Reed, et al., 1979). This section gives a description of where each allele was identified and the experiments that showed that *nude* is inherited as a single autosomal recessive phenotype with mutations in homologous genes in the mouse and rat.

*nude* (*nu*): A hairless mutant was discovered in 1962 in a closed but not deliberately inbred albino stock at the Virus Laboratory in Glasgow. Six years later it was reported that these mice are athymic, showing strict correlation between the hairlessness and the athymia (Flanagan, 1966; Issacson, et al., 1962; Pantelouris, 1973).

*nude-streaker* (*nu*<sup>str</sup>): In 1974, one male and two female hairless mice were found in a litter of AKR/J mice in the Animal Resources colony at the Jackson Laboratory. Autopsy of an affected animal revealed absence of the thymus. Linkage tests showed that the new mutation was allelic with *nude* (Eicher, 1976).

Rowett nude (*rnu*): A hairless, athymic rat was first noted in a colony of hooded rats maintained at the Rowett Research Institute, Aberdeen, Scotland, in 1953. The colony was maintained with difficulty due to shortened life span and poor breeding for a number of years, but eventually it died out. More than twenty years later, the colony was reconstituted when a pair of rats at the same Institute produced two female *nude* rats (Festing, et al., 1978).

*rat nude*, New Zealand (*rnu*<sup>N</sup>): In May 1976, hairless, athymic rats were discovered in a colony of outbred albino rats maintained at Victoria University, Wellington, New Zealand. These rats were also athymic and were shown to be allelic with the original *rnu* mutation. As there was no
known prior association between the two stocks, the New Zealand *nude* is considered a second, independent mutation at the same locus (Berridge, et al., 1979).

**Initial genetic mapping studies**

To determine that *nude* segregates as a single autosomal locus with a recessive phenotype, Flanagan set up a *+/nu x +/nu* intercross, which yielded 1349 (25.7%) *nude* and 3890 (74.3%) wild type mice. Flanagan then mapped the *nude* locus to chromosome linkage group VII between two dominant mutant phenotypes, *Rex (Re)* and *Trembler (Tr)*. Crossing heterozygous *nude* mice with each of the seven multiple linkage stocks available at that time, Flanagan discovered that *nude* was not segregating independently of *Re*. Crossing *Re +/+ nu* and *++/+ nu* mice gave a recombination frequency of 19.4 ± 4.6% (n=1413). Finally, three point crosses were made to determine the order of *nu* and *Re* with respect to a third locus, *Trembler (Tr)*, in this linkage group. *Re+Tr/+nu+* females were backcrossed to *+nu+/+nu+* males so that all three phenotypes could be scored in the first generation. The rarest type phenotypes, *Re nu* and *nu Tr*, represent the classes of progeny with double cross-overs, identifying *nu* as central among the three loci. This data gave approximate linkage data of *Re-13 cM-nu-9 cM-Tr* (n=54). Linkage group VII has now been placed on Chromosome 11. Finally, in mice and rats, the *nude* mutations are likely to be in homologous genes because in both species they are tightly linked to the inducible nitric oxide synthase gene(Jenkins, et al., 1994; Zha, et al., 1995).

Since there are many known instances of multiple alleles at the same locus with differences in phenotypic expression, comparative testing was done between the alleles of a rodent species. As a caveat, there is a large variance reported in characteristics even among rodents with the same allele of *nude* because of the variability in genetic background, housing conditions, and methodology performing the measurements. The differences in phenotypic expression between *nu* and *nu<sup>str</sup> mice are not greater than the reported variation among different experiments with *nu* mice(Solomon, et al., 1977). However, this points out a tremendous benefit of working with the *nu<sup>str</sup>* allele: *nude-streaker* mice, arising on the AKR/J line at The Jackson Laboratories, are in effect an inbred line.
RESULTS

Crosses segregating nude

Positional cloning requires defining the chromosomal location of the gene by identifying genetic markers co-segregating with the locus in a cross. We genetically mapped the nude locus in over 2000 meioses in three separate F2 intercrosses. To ensure the greatest rates of polymorphisms between the two strains of the cross, we mated AKR/J-nus tr animals with the inbred subspecies Mus musculus castaneus (CAST/Ei) and Mus musculus molossinus (MOLF/Ei). One concern with intersubspecific crosses is the possibility of recombinational suppression due to structural heterogeneity of the chromosomes (Copeland, et al., 1993; Hammer, et al., 1989). To address this concern we also established an intraspecific cross between AKR/J-nu s tr and C57BL/6J. To minimize background effects, all of our crosses were performed with the nu s tr allele because it arose and has been maintained on the inbred strain AKR/J. Specifically we generated 182 (AKR-nu s tr X C57BL/6J) F2 animals, 650 (AKR-nu s tr X MOLF/Ei) F2 animals, and 226 (AKR-nu s tr X CAST/Ei) F2 animals. When animals are referred to in the text of this thesis, an (AKR-nu s tr X C57BL/6J) F2 animals is abbreviated as B#; an (AKR-nu s tr X CAST/Ei) F2 animals is abbreviated as C#; and an (AKR-nu s tr X MOLF/Ei) F2 animals is abbreviated as M#.

Genetic mapping of nude with simple sequence length polymorphisms

Thirty years ago, Flanagan mapped nude to mouse Chromosome 11 by demonstrating that it was linked to two visible phenotypes, Rex and Trembler (Re-11 cM-nu-7 cM-Tr). As we began the genetic mapping of nude, a map linked to 99% of the mouse genome, consisting of 317 SSLPs with an average spacing of 4.3 cM, was constructed in the Lander Lab by William Dietrich (Dietrich, et al., 1992). This genetic map has been expanded to include 6000 SSLPs (Dietrich, et al., 1994; Dietrich, et al., 1995).

To construct an initial genetic map of the region, progeny were phenotyped at post-natal day 11 for hair growth and genotyped with the original SSLP markers that mapped to the central portion of chromosome 11, where Rex and Trembler have been mapped. These markers gave the map order: (D11Mit4, D11Mit5) - 2.5 cM- (D11Nds1, D11Mit65) -1 cM- D11Mit7--0.7 cM--nu --0.7 cM -- (D11Mit32, D11Mit34) - 0.3 cM-D11Mit8 - 1.1 cM- D11Mit36
(Dietrich, et al., 1992; Dietrich, et al., 1994). An example of the genotyping of animals M#553 to M#593 with D11Mit32, D11Mit34, D11Mit36, and D11Mit65 is shown in Figure 6. The three separate F2 crosses gave a genetic distance of 2.7 ± 0.3 cM between D11Nds1 and D11Mit8, indicating that there is no gross recombinational suppression in the region of nude between the Mus musculus subspecies analyzed. Unfortunately, the intrasubspecific nude cross (AKR/J-nustr x C57BL/6J) F2 did not yield finer structure genetic mapping information because none of the 9 genetic markers in the smallest region around nude were polymorphic between the two strains. As genetic markers, mapping between D11Mit7 and D11Mit8, were developed by the MIT Genome Center, they were mapped with fine resolution on the nude crosses. These markers gave the map order of D11Mit7 -- 0.7 cM -- nu -- 0.05 cM-- D11Mit117 -- 0.05 cM -- D11Mit144 -- 0.3 cM -- (D11Mit118, D11Mit96) -- 0.1 cM -- (D11Mit91, D11Mit94) -- 0.2 cM -- D11Mit34. Table I reports the genotype of the intercross progeny with recombination breakpoints between D11Nds1 and D11Mit8 with the SSLP markers in the interval.

To obtain fine structure mapping information from the intersubspecific crosses, we focused on those progeny that were recombinant in the interval between D11Mit7 and D11Mit34. For all such recombinant animals, the presence or absence of a thymus was checked. (In all cases, the phenotypes of athymia and hairlessness coincided.) We progeny tested unaffected F2 progeny, carrying a recombinant chromosome together with a wild type non-recombinant chromosome (13 animals), to determine which nude allele was carried on the recombinant chromosome. In this manner, each F2 progeny yielded two informative meioses.

Generating additional genetic markers by a subtractive cloning method

To specifically target genetic markers to the nude region, we combined the power of transmission genetics with the recently developed method of representational difference analysis (RDA) (Lisitsyn, et al., 1993). RDA is a subtractive technique to clone differences between PCR amplicons of two DNA samples, referred to as Tester and Driver - either because the corresponding sequence is completely absent from the Driver due to a homozygous deletion or because it is contained in a small restriction fragment in the Tester but a large and, therefore, poorly amplifiable restriction fragment in the Driver. Briefly, the first step is to prepare
amplicons from the Tester and Driver by digesting each sample with a restriction enzyme, ligating the restriction fragments with a compatible adaptor, performing PCR using a primer complementary to the adaptor, and finally removing the adaptor by digestion with the original restriction enzyme. An amplicon contains only a portion of the genome, as it includes only small restriction fragments that are preferentially amplified. The Tester amplicons are then relinkered and subjected to multiple rounds of hybridization-extension-amplification in the presence of excess Driver amplicon, under conditions favoring amplification of fragments present in the Tester amplicon that lack corresponding fragments in the Driver amplicons. Consequently, this procedure should yield small amplifiable restriction fragments which are present in Tester amplicons but absent in Driver amplicons. Thus, RDA can produce clones that detect RFLPs between Tester and Driver.

To isolate genetic markers in the nude region, we used the power of transmission genetics to create Tester and Driver DNA samples that differed only in the nude region. One ideal substrate for genetically-directed representational difference analysis (GD-RDA) is a pair of congenic strains in which nude has been transferred onto an inbred background by successive generations of backcrossing and selection. The congenic strains should be genetically identical to the original inbred line except in a relatively small region surrounding nude - the size of the region determined by the number of backcross generations. In this particular case, the Driver was a C57BL/6J-nu female that was produced by 12 generations of backcrossing nude to C57BL/6J mice. The Tester was a C57BL/6J female. Although the size of the target region differing between the congenic strains is not known precisely, it is estimated to be less than 15 cM based on the breeding schemes used to construct the congenic strains. This congenic region around nude was expected to be large enough to contain polymorphisms detectable by RDA. Also, we expected that the Driver maintained few residual regions from the original non-inbred strain that were not linked to nude. Note that the congenic nude line could be used either as the Tester or the Driver. Since C57BL/6J-nu DNA is used here as Driver, we should clone amplicons from the C57BL/6J genome that are not present in the congenic region. For this experiment, the genomic DNA was digested with the restriction enzyme BglII and three cycles of hybridization-extension-amplification were performed.
The resulting difference-products were separated by agarose gel electrophoresis, revealing one strong band (450 bp), one medium band (500 bp) and a weak background smear (Lane c in Figure 7). We cloned the difference product, selected six clones at random, and identified three clones with distinct insert sizes. To identify which clones showed the desired property of detecting a fragment in the Tester but not the Driver amplicons, the clones were hybridized to Southern blots of DNA from the original Tester and Driver amplicons. This test rapidly eliminated all of the clones: All three clones detected fragments in both the Tester and Driver.

GD-RDA was concurrently applied to congenic strains for Lurcher (Lc), severe combined immunodeficiency (scid), pudgy (pu), tottering (tg) and stargazer (stg) under the same conditions as nude. From these experiments, two clones were isolated that mapped to within the congenic region and detected RFLPs between Tester and Driver genomic DNA. Accordingly, GD-RDA successfully generated polymorphic probes in a region of less than 1% of the mouse genome around the target locus. Although the number of probes produced is limited, amplicons digested with other restriction enzymes would give new and different probes.

Congenic strains are an obvious choice for GD-RDA, but they suffer from a major drawback. Producing congenic strains requires many generations, spanning years, even when selective breeding is used and the congenic regions are often still quite large. Besides congenic lines, it is not enough simply to apply RDA to samples from a single affected and a single unaffected animal to generate genetic markers linked to a trait. The abundant genetic variation between inbred lines will mean that polymorphisms will likely be found throughout the genome. Conversely, unless the nude mutation itself specifically creates an RFLP between AKR/J and AKR-\textsuperscript{nu\textit{str}}, there should be no variation detectable by RDA between these two lines. One requires a new source of Tester and Driver samples with the property that the Driver contains all of the alleles present in the Tester except in the region surrounding nude, resulting in polymorphisms specifically targeted to the vicinity of the gene of interest. As a more practical and rapid approach, we devised a second implementation of GD-RDA that requires a simple two-generation cross.

Transmission genetics is used to produce a collection of siblings with the property that their pooled DNA is homozygous in the region of the nude
gene but heterozygous elsewhere in the genome. The strategy will be discussed in general, first, and then applied to cloning genetic markers in the region of *nude*. Let A and B denote two inbred strains differing at a target locus L of interest. Suppose that A carries a mutant allele causing a recessive phenotype and B carries a wild-type allele causing a dominant phenotype. For a Tester sample, one can use strain B itself. To create a Driver sample, one performs an F₂ intercross between the strains, selects a collection of k progeny showing the recessive phenotype and mixes their DNA together. The principles of mendelian genetics predict that the Driver should contain: (i) no B alleles in the immediate vicinity of L, because progeny were selected for the recessive phenotype; (ii) a deficit of B alleles in a somewhat larger region around L, owing to linkage to L; and (iii) roughly equal proportions of A and B alleles elsewhere in the genome, because a collection of F₂ progeny should have genotypes AA, AB and BB in the ratio 1:2:1 at unselected loci (see Figure 8a,b). If RDA is performed with this Tester and Driver, then one would expect that B alleles should be subtracted everywhere in the genome except in a region around L. GD-RDA should thus yield polymorphic alleles from the wild-type chromosome at loci linked to L.

The targeting of the method can be somewhat improved in the event that the locus L has already been mapped to lie between two flanking genetic markers, X and Y. For the Driver, one can select k/2 progeny in which a crossover has occurred between X and L and k/2 progeny in which a crossover has occurred between L and Y. This would guarantee that the proportion of B alleles is 25% at X and Y, ensuring that the region over which the proportion of B alleles is very low is restricted to the interval X-Y (Figure 8c). As shown below, with the *nude* cross, this refinement can allow targeting of very small intervals.

An important issue in the design of this experiment is the number of progeny that should be pooled. While the proportion of B alleles at unlinked loci in the Driver will have a mean value of 50%, the actual value will fluctuate across the genome. Theoretically, 10 pooled F₂ intercross progeny should suffice to ensure that the proportion of B alleles remains above 10-15% (the proportion necessary for an efficient subtraction) in regions of the genome unlinked to L.

To test this approach, we used *nude* (AKR-\textit{nu}^{str} X MOLF/Ei) F₂ intercross progeny as Driver and the MOLF/Ei parent as the Tester. At this
point, we had generated 416 (AKR-nustr X MOLF/Ei) F2 intercross progeny and genotyped them for genetic markers flanking the nude locus. For the subtraction, we selected the 12 nude progeny with crossovers between nude and closely linked markers (Figure 9). All of the crossovers occurred within a 7 cM interval defined by D11Mit5 and D11Mit36, and 4 of the 12 occurred within a 1.4 cM interval defined by D11Mit7 and D11Mit34. A Driver sample was prepared by pooling equal amounts of DNA from these 12 progeny; the corresponding Tester sample was DNA from the MOLF/Ei parental strain. In principle, GD-RDA should produce MOLF/Ei alleles that detect polymorphisms in the interval between D11Mit5 and D11Mit36. Moreover, if the proportion of MOLF/Ei alleles surrounding the nude locus sufficed to allow efficient subtraction, the polymorphisms might be preferentially targeted to the small interval between D11Mit7 and D11Mit34, or even targeted to the smallest interval between the breakpoints in the recombinant animals.

Using this Tester and Driver combination, we performed three rounds of extension-hybridization-amplification on the PCR amplicons, created by amplifying genomic DNA digested with BglII. The subtraction resulted in two clear bands (700 bp and 450 bp), visible by ethidium bromide staining, which were named RDA-6.1 and RDA-6.2 (Lane f in Figure 7). As above, the probes were initially characterized by hybridization to Southern blots of Tester and Driver amplicons. RDA-6.1 detected a large number of bands in both amplicons and was eliminated. RDA-6.2 showed the desired pattern of hybridizing to the Tester but not Driver amplicons. The probe was then hybridized to Southern blots of mouse DNAs digested with BglII. RDA-6.2 detected an RFLP with a 450 bp allele in MOLF/Ei and a 4 kb allele in AKR/J-nustr. To obtain approximate localization, we genotyped 20 (AKR-nustr X MOLF/Ei) F2 progeny that showed no recombination between genetic markers flanking nude and found that RDA6.2 showed an inheritance pattern completely concordant with that of the nude locus itself. To obtain finer localization, we then genotyped the 12 nude F2 progeny used to create the Driver and found that the RFLP again showed complete concordance with nude -- i.e., the progeny were all homozygous for the AKR allele of the RFLP (Figure 10). This proves that RDA-6.2 maps within the 1.4 cM interval bounded by D11Mit7 and D11Mit34. Subsequent analysis of additional F2 progeny showed that RDA-6.2 recombined with nude six times in 1752
meioses, corresponding to a genetic distance of only 0.3 cM. Thus, GD-RDA successfully targeted a probe to a region less than 1/4000 of the mouse genome.

To generate additional clones we repeated the experiment with the same Tester and Driver, this time digesting the genomic DNA with the restriction enzyme BamHI before amplifying the amplicons. This subtraction produced three clones, two of which showed the expected pattern of hybridizing to the Tester but not the Driver amplicons, RDA-10.2 and RDA-10.4. Both probes detected RFLPs between MOLF/Ei and AKR/J-nu

str genomic DNA digested with BamHI (with allele sizes 600 bp vs. 4-5 kb for RDA-10.2 and 500 bp vs. 3 kb for RDA-10.4). Genetic mapping subsequently showed that both probes mapped extremely close to nude. The 12 nude F2 progeny used to create the Driver were all homozygous for the AKR allele for both RFLPs, indicating that both loci mapped in the 1.3 cM interval between D11Mit7 and D11Mit34. Subsequent analysis of additional F2 progeny showed that RDA-10.4 recombined proximally with nude only once (animal C4 in Table I) and that RDA-10.2 recombined distally with nude only once (animal M952 in Table I). In summary, GD-RDA (subtracting the DNA from nude F2 progeny from the parental wild-type DNA) produced three genetic markers mapping within 0.4 cM of the nude locus.

An analogous experiment was performed with 13 staggerer (C57BL/6J- sg x DBA/2J) F2 intercross progeny with crossovers in a 10 cM interval around staggerer as the Driver and DBA/2J as the Tester. In contrast to the nude experiments, this experiment yielded a probe from a region near sg for which the Driver contained the Tester allele at a proportion of 11.5% (i.e., 3/26), mapping approximately 4.5 cM distal to sg.

These two implementations of GD-RDA, involving congenic strains and two-generation crosses, successfully produced probes mapping near various target genes. Indeed, every clone (6/6) that passed a rapid initial characterization (i.e., detecting a unique fragment in Tester but not Driver amplicon and a unique locus in genomic DNA) mapped to the desired location. The yield of probes was relatively low (6 probes from 9 experiments), which is perhaps not surprising in view of the multiple rounds of competition among PCR products during RDA. The number of probes might be increased through the use of additional restriction enzymes for amplicon preparation, as demonstrated by the successful use of BamHI in the
case of the *nude* experiment. Some restriction enzymes, such as *Taq*I, may produce a higher yield of polymorphisms. It may also be possible to generate new clones with a single restriction enzyme by blocking the amplification of already-identified clones by adding them back to the Driver. Finally, it may be possible to detect less drastic changes in the length of restriction fragments by initially fractionating Tester and Driver by gel electrophoresis and performing subtraction on specific size fractions.

*Generating additional genetic markers from the clones spanning the *nude* locus*

The SSLP and RDA genetic markers were used to screen mouse genomic libraries of yeast artificial chromosomes (YACs) and bacteriophage P1s to identify clones from the *nude* region. The specifics of screening the libraries and isolating the terminal fragments from the clones will be discussed in the next chapter. Here, we focus on identifying and characterizing genetic markers subcloned from the YACs and P1s.

We generated a dense set of genetic markers, based either on size or conformational polymorphisms, across the critical *nude* region. It has been demonstrated that single-stranded DNA molecules take on specific sequence-based secondary structures when electrophoresed under non-denaturing conditions. PCR-generated fragments, differing by as little as a single base substitution, may form structures that migrate differently, identifying a (Orita, et al., 1989). The terminal YAC or P1 fragments were analyzed to determine if they identified either an SSLP or a single-strand conformational polymorphism (SSCP) that segregated in one of the intercrosses. To generate additional genetic markers from the region, we subcloned the YAC that spanned the *nude* interval into pBluescript(pBS) and identified 5 mouse genomic DNA inserts with SSRs. We also sequenced and generated PCR amplicons from the region flanking the -21M13 site for 7 additional pBS clones. We used one standard set of conditions to determine if the PCR amplicon contained an SSCP -- room temperature in 24 cM gels of 0.7 X Hydrolink-MDE with 10% glycerol. To scan for polymorphisms on both strands, we independently end-labeled each primer with $^{32}$P prior to the PCR amplification. We found this method gave a clear SSCP between AKR/J-*nu*str and MOLF-Ei or CAST-Ei for 50% of the YAC ends and almost 70% for a pBS clones. The rate of polymorphism for YAC ends is probably lower.
because the size of the PCR amplicon was limited by the availability of sequence. We never found an SSCP between the strains, AKR/J and C57BL/6J, both of which belong to the same subspecies. This is consistent with the 90% RFLP rate observed between C57BL/6J and Mus. spretus and the low level of RFLPs detected between inbred strains of Mus. musculus (Avner, et al., 1988). Examples of the SSCP typings for BS11.8 and the left end of YAC 32 are shown in Figure 11. A comparative study by Sheffield et al. (1993) indicated that the optimal size fragment for sensitive base substitution detection by SSCP is approximately 150 bp. The average YAC end PCR amplicon was 176 bp with a range of 62 to 600 bp. A comparative analysis of the sensitivity of alternative SSCP methods with 19 known polymorphisms, published by Vidal-Puig and Moller (1994), found that 63% of the sequence changes were detected under our standard conditions. They found that the remaining sequence alterations could be detected when the glycerol was omitted from the Hydrolink-MDE gel.

Five SSCP (BS11.4, BS11.5, BS11.7, BS11.8, BS12.42) and 5 SSLP markers (BS7.10, CA4, CA8, CA69, CA128) were identified from the nude region. The primer sequences and characterization of these genetic markers are given in Table II. The YAC and P1 terminal fragments that yielded polymorphism are underlined in Figures 12, 13 and 14. The primer sequences to amplify the non-chimeric YAC and P1 ends are given in Table III. The type of polymorphism identified by the terminal fragment PCR amplicons is listed in Table IV. Table V gives the genotype of the intercross progeny with recombination breakpoints between RDA6.2 and D11Mit118 with the additional SSCP and SSLP markers.

To identify the boundaries of the nude region, we mapped all of the genetic markers on the animals with the closest flanking crossovers. Based on the first 1000 animals (i.e. 2000 meioses), the non-recombinant region extended from BS12.42 to Y30L (the left end of YAC 30) (Table V). In the last 50 animals, we were fortunate to obtain an animal, (M952), with a proximal breakpoint that cut the region in half. Based on these progeny, nude was determined to lie minimally between P7N (NotI end of P1 #7) to CA8 (230 kb) and maximally between BS12.42 and P15S (SalI end of P1 #15) (370 kb) (Table V and Figure 14).

CONCLUSIONS
General applications of GDRDA

The application of GD-RDA to congenic strains is straightforward. However, the real power of GD-RDA lies in its application to crosses, because the breeding or pedigree collection required is within the realm of practicality for a wide range of organisms. The technique can be applied to any trait whose presence implies homozygosity for a particular allele at a trait-causing locus, so that these homozygotes may be pooled to create a Driver.

An interesting feature of the application to crosses is that the targeting of GD-RDA can be improved by successive iterations. Given a large cross, one could first generate flanking markers that are linked, but perhaps not very closely, to the target locus. Using such flanking markers to identify recombinant progeny, one could perform subsequent subtractions with these progeny to target successively smaller intervals. As shown in the case of the nude and staggerer crosses, the use of recombinant progeny can effectively target quite small intervals. The ultimate resolution of this approach should be limited only by the actual density of polymorphisms detectable by GD-RDA; we estimate this density to be 1-2 per megabase for an enzyme such as BglII. This distance is small enough to permit chromosomal walking from the RDA clone to the target gene.

This project focused on the application of GD-RDA to F2 intercrosses between inbred strains, but the technique is more broadly applicable. It can be applied to backcrosses between inbred strains, two-generation families in an outbred population (for organisms for which inbred lines are not available); and half-sib mating schemes (common in livestock breeding). The application of GD-RDA to random-breeding populations should include the analysis of human families. One might, for example, use an individual affected with a dominant disease as Tester and a collection of unaffected close relatives as Driver. In some families, there may be too few relatives to ensure subtraction of all unlinked regions. In such cases, GD-RDA should at least enrich for linked probes which could then be subsequently screened for linkage. Notwithstanding continuing advances in genomic analysis, construction and application of dense genetic linkage maps remains a daunting task. GD-RDA offers the prospect of obviating the need for such maps, at least for certain purposes. In particular, GD-RDA should open the prospect of genetic mapping and positional cloning of monogenic traits in...
agriculturally important animals, plants and fungi. GD-RDA is unique among molecular genetic techniques in that it provides a way to target DNA probes to the vicinity of a gene without prior knowledge of either the gene’s function or position. By applying classical transmission genetics, one can prepare DNA samples from mixtures of progeny that differ only near the gene of interest and then use the powerful subtraction technique of RDA to clone these differences. The technique opens the prospect of genetic analysis and positional cloning even in organisms without pre-existing genetic maps.

Genetic resolution

We used F2 intercrosses segregating the nude phenotype in order to obtain two informative meioses for every progeny. Initial mapping studies with 200 animals localized nude to a 2.7 cM interval between D11Nds1 and D11Mit34. To map nude with higher resolution, the crosses were expanded to 2000 meioses. To determine fine structure mapping information, we focused on the approximately 3% of the progeny that were recombinant between the closest SSLP markers flanking nude. To obtain the full meiotic power, we progeny-tested those animals carrying recombinant chromosomes over wild-type chromosomes in the nude region to determine which nude allele was carried on the recombinant chromosome.

The mouse genetic map, consisting of 317 SSLPs, had an average spacing of 4.3 cM. The closest markers flanking the nude locus were D11Mit7 and D11Mit8, mapping 0.7 cM proximal and 0.9 cM distal to nude, respectively. An additional 4000 SSLPs were added to the map, yielding an average spacing of 0.8 cM between markers. Eight additional SSLP markers mapped to the interval between D11Mit7 and D11Mit8, including D11Mit117 and D11Mit144 which mapped 0.05 and 0.1 cM distal to nude. D11Mit7 remained the closest proximal flanking SSLP marker. However, D11Mit7 and D11Mit117 map only 0.8 cM from each other, the average spacing between the 4,000 SSLPs. Obviously, the map grew denser, and we were simply fortunate in the beginning.

With N meioses, the distance to the closest flanking cross-over on either side of the locus will be exponentially distributed with an expected distance of 1/N Morgans. Thus, the recombinationally inseparable interval containing the gene will have the expected size 2/N Morgans. Mapping nude with a total of 2000 meioses should thus yield an interval of 1/10 cM = 214 kb
(3,000 Mb/1400 cM=physical size of the mouse genome/genetic size of the mouse genome). In fact, as we discuss in the next chapter, we narrowed the *nudef* locus to a minimum region of 230 kb and a maximum of 370 kb.
REFERENCES


Chapter Three

Physical map of the *nude* region: Chromosomal walks in yeast artificial chromosome and bacteriophage P1 clones.

ABSTRACT

We constructed a chromosomal walk, spanning the *nude* region, in large insert yeast artificial chromosome (YAC) clones and smaller insert Bacteriophage P1 clones. The YAC walk was initiated from genetic markers flanking the *nude* locus (*D11Mit7* and *D11Mit34*) and covered approximately 3 Mb. To construct the YAC contig, we (1) identified YACs that contain the genetic markers flanking the *nude* locus; (2) determined which YACs were chimeric; (3) genetically oriented the chromosomal walks; (4) rescreened the mouse YAC libraries for new clones that contain markers closer to the *nude* locus. This process was continued until genetic markers that recombine proximally and distally were contained in the same chromosomal walk. The walk in smaller insert P1 clones was initiated from a dense set of markers in the *nude* region, covers approximately 500 kb, and defines the smallest genetic and physical region in which *nude* must lie. The two chromosomal walks give independent verification that the cloned fragments are an accurate representation of the genomic region. The *nude* YAC and P1 chromosomal walks have a consistent long-range restriction map and both contain over 150 markers, subcloned equally from the two sources.
INTRODUCTION

To refine the chromosomal location of a gene, i.e. to define the minimum genetic and physical regions, requires increased resolving power. Alleles of the gene with chromosomal rearrangements can narrow the region significantly, but this resource is not commonly available. More generally, cloned DNA fragments, spanning the chromosomal region between genetic markers flanking the locus, are necessary to generate closer genetic markers, determine the size and structure of the physical region, and later to identify expressed sequences in the region.

Contiguous cloned fragments (called contigs) have been assembled on a genome-wide basis for the nematode, *Caenorhabditis elegans*, and the human. A total of 700 contigs, covering 80% of the *C. elegans* genome (100 Mb total) were assembled from 17,000 cosmids based on their restriction digestion patterns. Systematic hybridization between yeast artificial chromosome (YAC) clones and selected cosmids revealed previously undetected overlaps in the cosmid contigs and produced a physical map of YACs with only seven gaps (Coulson, et al., 1988). The current human physical map was assembled based on sequence-tagged-site (STS)-content mapping, radiation hybrid mapping, and genetic mapping to yield approximately 600 YAC contigs covering 94% of the genome. Gaps between the contigs were tentatively closed for about 50% of adjacent contigs based on fingerprint analysis and Alu-PCR hybridization (Hudson, et al., 1995). Although the reagents are being developed, contigs covering large regions of the genome do not currently exist for the mouse.

To positionally clone a Mendelian trait in the mouse, it is necessary to create a contig spanning the chromosomal region containing the locus. Contigs are initiated from genetic markers flanking the locus and extended until genetic markers that recombine with the locus both proximally and distally are contained in the same contig. Chromosome walking is based on the concept that it is possible to identify a series of clones from recombinant DNA libraries with overlapping inserts by identifying those that share specific loci (Bender, et al., 1983). The contig is extended by identifying STSs from the existing clones, directed toward the locus of interest, that are contained in novel clones in the library. The number of genetic markers and meiotic
power available in the mouse usually narrows the chromosomal region to several centimorgans, corresponding to a physical region of several megabases of DNA. Cloning physical regions of this size requires hosts that can stably maintain inserts of several hundred kilobases. Burke and Olson first demonstrated that large mammalian inserts can be stably maintained as artificial chromosomes in yeast (Burke, et al., 1987). YAC libraries, clones each containing a unique insert, have been constructed with Drosophila, C. elegans, Arabidopsis, mouse, and human insert DNA (Albertsen, et al., 1990; Chumakov, et al., 1992; Coulson, et al., 1988; Foote, et al., 1992; Garza, et al., 1989; Hwang, et al., 1991; Kusumi, et al., 1993; Rossi, et al., 1994). The technology has recently been developed to maintain artificial chromosomes with inserts of several hundred kilobases in bacteria, as well (Shizuya, et al., 1992). Although the principles of chromosome walking are the same for inserts cloned in all host cells, there are significant practical differences arising from the characteristics of the yeast clones.

YAC libraries contain a high proportion of clones that are not co-linear with the genomic source DNA, including deleted, rearranged, or chimeric inserts. For chromosome walking, chimeric fragments are the most troublesome because an STS cloned from the non-contiguous region could "extend" the walk into a new region of the genome. The proportion of chimeric clones has been estimated at a non-trivial 40-60% for human YAC libraries (Bates, et al., 1992; Haldi, et al., 1994; Selleri, et al., 1992). Deletions and rearrangements are more problematic when the YACs are used to represent the genomic region, either to search for expressed sequences in the region or to complement a mutant phenotype.

To use YAC clones efficiently for chromosome walking, it is critical to determine rapidly whether a YAC contains noncontiguous DNA segments. There are three basic strategies for detecting a chimeric clone: fluorescence in situ hybridization (FISH) of the YAC clone; Alu-PCR dot blot hybridization to a mapping panel; and cloning and mapping either Alu-PCR products or the terminal fragments of the YAC. Alu-PCR specifically amplifies fragments of the mammalian insert DNA from the total YAC DNA with consensus primers from conserved Alu repetitive sequences. FISH mapping has proven very useful to determine rates of chimerism in human YAC libraries and it rapidly identifies noncontiguous segments (even when arising from different regions of the same chromosome) (Haldi, et al., 1994; Selleri, et al., 1992).
However, FISH mapping does not generate new STSs to extend the chromosomal walk. As well, this procedure is technically more difficult in the mouse than in the human because the banding pattern of mouse chromosomes is not as well characterized and the chromosomes are acrocentric, showing a continuous gradation in size. Alu-PCR dot blot hybridization to a mapping panel is an easily streamlined procedure that has been used successfully to characterize human YACs, but there are sensitivity limits to Alu-PCR because of the inherent competition of PCR (Banfi, et al., 1992). Furthermore, somatic cell hybrid panels that carry single chromosomes do not exist in the mouse. In contrast, the mouse has the advantage that it is usually possible to find a variant detected by a clone that can be genetically mapped within an intersubspecific cross. Alu-PCR and cloning the ends of the YACs are useful complementary methods to generate probes from the YAC. Cloning and mapping the ends of each YAC insert gives a detailed characterization of each clone and produces a terminal locus from which to extend the contig.

To determine if a YAC is chimeric, the terminal fragment must be mapped physically and/or genetically. To physically map the YAC end, two distinct sets of DNA that define the region are assayed to see if they contain the terminal sequence. The YAC end is considered to be contiguous if it is contained in either a consistent set of YACs from the region or a hybrid cell line of this specific region of the genome. These results can not be uniquely determined if the terminal sequence is a mouse repetitive element.

To genetically map the YAC end to the region, the terminal fragment must detect a polymorphism in a mapping cross. Genetic mapping of the ends of the YACs also allows one to monitor the progress of the chromosomal walk and eventually to determine when the YACs span the target locus region. Generating polymorphisms is essential to orient the first step of the YAC chromosomal walk. A priori, there is no way to know which end of the contig is oriented toward the locus; i.e. toward the centromere or the telomere of the chromosome. Within a contig of YACs, if a clone is contained in only a subset of the YACs and if it genetically recombines with the original STS, then this orients the walk. All secondary steps of the contig extension can be oriented physically. Chromosome walking continues by rescreening the YAC library with the most distal end of the proximal contig and the most proximal end of the distal contig, until genetic markers that
recombine with the locus proximally and distally are contained in the same chromosomal walk. Briefly, the strategy we employed to construct contigs was (1) to identify YACs that contain the original genetic markers flanking the \textit{nude} locus; (2) to determine which YACs are chimeric (3) to orient the chromosomal walks (4) to rescreen the YAC libraries for new clones that contain the STSs oriented toward the \textit{nude} locus. This process was continued until markers that recombine with \textit{nude} both proximally and distally were contained in the same chromosomal walk.

Although we eventually targeted markers tightly linked to \textit{nude} by genetically-directed representational difference analysis (GD-RDA) the project was initiated with the MIT genetic markers. A chromosomal walk of large insert YAC clones, covering at least 3 Mb, was initiated from \textit{D11Mit7} and \textit{D11Mit34}, which map 0.7 cM proximal and 0.7 cM distal to \textit{nude}, respectively. The walk in smaller insert bacteriophage P1 clones was initiated from a dense set of markers in the \textit{nude} region, covers approximately 500 kb, and defines the smallest genetic and physical region in which \textit{nude} must lie.

RESULTS

Screening YAC libraries

When we initiated the physical mapping of \textit{nude}, Tilghman's lab at Princeton University had developed a publicly available, readily screenable mouse YAC library with 2.2 haploid genome equivalents and an average insert size of 265 kb (Rossi, et al., 1994). The mouse genome is approximately 16M and 3,000 Mb, so that on average 1 cM corresponds to 2 Mb. The distance between the proximal and distal flanking markers, \textit{D11Mit7} and \textit{D11Mit34}, is 1.4 cM which corresponds roughly to 2.8 Mb. A first-order calculation indicates that it would take 20 end-to-end steps to cover this region with YACs from the Princeton library. At that time a mouse YAC library of 20,000 clones with an average insert size of 650 kb was constructed in the Lander Lab, giving 4.3 fold coverage of the genome. We prepared the YAC clones for simple two-step PCR screening based on top-level pools and subpools to give the address of the clones that contain any given STS (Green, et al., 1990; Kusumi, et al., 1993). We switched over to screening the MIT YAC library exclusively, anticipating that we could cover the region in 8 steps.
Initiating the YAC chromosomal walk: Basic characterization of the clones

*D11Mit7* is contained in YACs 5, 7, and 8 from the MIT library and YACs 4 and 6 from the Princeton library. *D11Mit34* is contained in YACs 14 and 15 from the MIT library and YACs 25, 26, and 27 from the Princeton library. The address in the YAC library and the original locus screened to identify the YACs used for this project are given in Table IV. A more complete description of all of the markers contained within the YACs is presented schematically in Figures 12 and 13. The size of the YACs were determined by pulsed-field gel (PFG) analysis (Table IV). At least two independent isolates of each clone were sized, allowing us to work exclusively with the clone that had maintained the largest insert. Of the YAC clones analyzed 18% (5 of 28) had independent isolates of two different sizes as judged by PFG analysis. The largest variance observed was YAC 31: Isolates of this YAC clone ranged in size from 300 kb to 1 Mb.

Detecting chimerism; Orienting and Extending the YAC chromosomal walk

To use YAC clones efficiently for chromosome walking, it is critical to rapidly determine whether a YAC is chimeric; i.e. contains noncontiguous DNA segments. Cloning and mapping (genetically and/or physically) the ends of each YAC insert gives a detailed characterization of the clone and produces a terminal STS from which to extend the contig.

To identify the non-chimeric ends of the contig, we cloned the terminal fragments of the YACs by inverse PCR (Joslyn, et al., 1991; Triglia, et al., 1988) (Note: Bubble-anchored PCR has shown to be an equally successful methods to isolate end fragments from YAC clones (Foote et al., 1992, Riley et al., 1990). ) Briefly, this protocol entails digesting the YAC DNA with a restriction enzyme that cleaves within the vector and the genomic insert to generate relatively small fragments, which are then ligated under dilute conditions to favor monomeric circularization. The products are amplified using two rounds of PCR with nested primers of vector sequence that are in inverse orientation. Originally, this strategy for cloning YAC ends was not popular because of its limited success in generating specific products. We reduced the background problems greatly by redesigning the primers to not reside wholly in the SUP4 sequence of the pYAC4 vector. As well, sequencing the pYAC4 vector, we discovered that there was a 2 bp omission.
in the published sequence that was contained in some of the original inverse PCR vector primers. A key factor in enhancing the likelihood of success using this approach is to use two or three frequent cutting enzymes for each end. Sequencing primer tails (-21M13 and M13REV) were added to the nested set of PCR primers to allow direct sequencing of the amplified inverse PCR products on the ABI 370 automated sequencer. The primer sequences used to rescue YAC ends by inverse PCR are given in Table VI.

The terminal fragments were analyzed to determine whether they mapped to the region contiguous with the nude locus. As well, we determine their orientation within the chromosomosal walk relative to the nude locus. The overlap pattern of the YACs was ascertained at the same time and used to monitor the consistency of the chromosomal walk. To rapidly analyze the clones, PCR primers were selected to amplify the sequence of each end. To physically map the YAC end to the region, a series of DNAs that define the region were assayed. We had two sets of DNAs specific to the nude region - the YAC contig and a hybrid cell line containing mouse Chromosome 11 on a rat host background. A YAC end was considered to physically map to the region if it was contained in YACs other than itself from the contig and if it mapped to a consistent location in the YAC chromosomal walk. Alternatively, a YAC end was considered to physically map to the nude region if it was contained in a mouse Chromosome 11, rat hybrid cell line, and absent from the control rat cell line (Killary, et al., 1984). To eliminate mouse repeats, we also checked that the end was not specifically contained in a hybrid cell line containing mouse Chromosome 1 on a CHO host background (Hunter, et al., 1991). The results of these screens are presented in Table IV. Of the 23 clones identified from the MIT YAC library, 29 ends were cloned and analyzed: 76% (22/29) mapped to the nude region; 17% (5/29) mapped to another region of the genome; and 7% (2/29) were repetitive. The primer sequences to amplify the non-chimeric YAC ends are given in Table III.

To genetically map the YAC end to the region, the terminal fragment must detect a polymorphism in a mapping cross. As described in the previous chapter, we determined if the PCR amplicon of the YAC end detected a size or single-strand-conformational polymorphism (SSCP) that segregated in any of the three nude F2 intercrosses. 54% (15/28) of the non-chimeric ends detected a polymorphism between AKR/J-\textit{nustr} and CAST/Ei
or MOLF/Ei that segregated in the *nude* F2 intercrosses. Genetically mapping the ends of the YACs allowed us to monitor the progress of the chromosomal walk and eventually to determine when the chromosomal walk spanned the *nude* region.

Genetically mapping the ends of the YACs was essential to orient the first step of the YAC chromosomal walk. (The nomenclature used in the text and figures refers to the terminal fragment of mouse sequence adjacent to the centromeric arm of the pYAC4 vector as the left end; such that the left end of YAC 7 is abbreviated as Y7L. Similarly, the terminal mouse sequence adjacent to the noncentromeric arm is referred to as the right end; such that the right end of YAC 8 is abbreviated as Y8R. As well, F2 progeny from the C57BL/6J, CAST/Ei and MOLF/Ei intercrosses will be referred to as B, C, and M, respectively.) The chromosomal walk, originating from the proximal flanking marker, *D11Mit7*, was oriented by the left end of YAC 8. Y8L contains an SSR ((CA)12) that detects a polymorphism between AKR/J-*nustr* and C57BL/6J, CAST/Ei or MOLF/Ei. The primer sequences and the sizes of the PCR products in these inbred strains are given in Table II. Y8L is contained in YACs 7 and 8, and recombines with *D11Mit7*, but not with *nude* in animal C11, orienting Y8L as distal to *D11Mit7* relative to *nude*. The genotypes are shown in Table I and the contig of YACs is shown in Figure 12. Unfortunately, Y8L is not contained in any new YACs in the MIT library, only YACs 7 and 8. Furthermore, we were unable to clone Y7R and Y7L was chimeric. The YAC chromosomal walk from the proximal side stalled as we generated markers from YAC 7 that extended beyond Y8L, but still mapped to the *nude* region.

The chromosomal walk, originating from the distal flanking marker, *D11Mit34*, was oriented by the left end of YAC 14. Y14L is contained in YACs 14,15 and the primers for this end detect a size polymorphism of 180 bp in AKR/J-*nustr* and 130 bp in MOLF-Ei. Y14L recombines with *D11Mit34* in animal M143, orienting Y14L as distal to *D11Mit34* relative to *nude* (Table V). Even though, we were unable to demonstrate that Y14R genetically mapped to the *nude* region, it physically mapped to the *nude* region; i.e. it was specifically contained in the Chromosome 11, rat hybrid cell line. The chromosomal walk toward *nude* was extended from Y14R, which was contained in 3 new MIT YAC clones; YACs 28, 29, and 30 (Figure 13). Cloning the ends of the YACs allowed us to orient the walk, to determine which YACs
were chimeric, and to extend the chromosomal walk. The genetic and physical maps co-evolved: The ends of the YACs were cloned to extend the chromosomal walk, to generate genetic markers closely linked to *nude* and to define the smallest region containing the *nude* locus.

**GD-RDA markers extend and close the YAC chromosomal walks**

GD-RDA, a subtractive cloning method described in the previous chapter, targeted three clones, detecting polymorphisms, to a tightly linked region surrounding the *nude* locus. These markers, RDA6.2, RDA10.2 and RDA10.4, provided new foci of initiation for a chromosomal walk to *nude*.

RDA6.2 mapped distal to Y8L, but proximal to *nude* as shown by the genotype of these markers in animals M100, M210, M433, M486, M564 and animals C4, C100, C190, C216, C223, M152, M465, respectively (Table I). RDA6.2 was contained in MIT YACs 19, 20, 21, and 22, which initiated a new chromosomal walk from a proximal flanking marker toward *nude*. As before, we cloned the ends of these YACs to orient and extend the chromosomal walk. This walk was oriented genetically and physically by the right end of YAC 19. Y19R is contained in YAC 7, which also contains the more proximal markers Y8L and *D11Mit7*. As well, Y19R detected an SSCP that mapped proximal to RDA6.2 relative to *nude* in animals M100, M210, M433, M486, M564 (Table V). Y21L was the most distal marker from this proximal *nude* chromosomal walk and was contained in 4 new MIT YACs.

RDA10.2 and RDA10.4, closed the chromosomal walk of the *nude* region in one step. RDA-10.4 mapped 0.05 cM proximal to *nude*, detecting a cross-over in animal C4 (Table I). RDA-10.2 mapped 0.05 cM distal to *nude*, detecting a cross-over in animal M952 (Table I). RDA-10.2 is contained in YACs 29, 31, 35 and RDA10.4 is contained in YACs 31, 32, 33, and 34 (Figure 13). Since both proximal and distal flanking markers are contained in YAC 31, the *nude* locus must be contained within the genomic region corresponding to this YAC. (Recall from above that Y14R is also contained in YAC 29. Therefore, this contig, originating with RDA10.2 and RDA10.4, joined the chromosomal walk that was initiated with the distal flanking marker, *D11Mit34*. (Figure 13))

**Defining the smallest interval surrounding *nude***
YAC 31 appeared to be colinear with genomic DNA since its ends and all markers cloned from it either mapped genetically and/or physically to the *nude* region by the criteria described above for YAC ends (Table III). The size of YAC 31 was determined to be at least 1 Mb by PFG analysis. However, this YAC was highly unstable: 15 unique isolates of it ranged in size from 150 kb to 1 Mb. Even when individual isolates from a 1 Mb clone of YAC 31 were re-grown, the sizes ranged from 250 kb to 1 Mb.

To increase the genetic resolution and thereby narrow the interval in which to search for the *nude* gene, we cloned an additional 5 SSLPs (BS7.10, CA4, CA8, CA69, CA128) and 5 SSCPsls (BS11.4, BS11.5, BS11.7, BS11.8, BS12.42) markers from YAC 31. Based on which YACs contained the clone, we were able to assign each genetic marker to a unique location in the physical map (Figure 13. BS11.4, BS11.5, BS11.7, and BS11.8 map to YACs 28, 29, and 30, but are not shown for lack of space on the Figure). BS12.42, an SSCP marker contained only in YAC 31, was the closest proximal marker to *nude* and RDA10.2 remained the closest distal marker to *nude* (Figure 13 and Table V).

To determine the physical distance between the closest markers flanking *nude*, we constructed a PFG restriction map of YAC 31. For our analysis, we used a 1 Mb isolate of YAC 31. Total YAC DNA was digested independently with several rare-cutting restriction enzymes: *MluI*, *NotI*, *RsrII*, *SacI* and *SfiI*. Southern blots of these digests were hybridized with a dense set of markers from the region: the ends of the YAC, the clones obtained from direct cDNA selection and exon trapping (described in the next chapter), and the genetic markers subcloned from the YAC. A unique restriction map of the YACs was revealed by the hybridization pattern of this dense set of markers. The PFG map of the region of the YAC that contains the *nude* gene is shown in Figure 14. The physical distance from BS12.42 to RDA10.2 was estimated to be 400 kb.

**Fine scale physical map**

To restrict our analysis to only the smallest region surrounding the *nude* locus, we sought to obtain clones with inserts smaller than YACs that span the *nude* region. We initially subcloned YAC 31 into cosmids in order to construct a cosmid contig around the *nude* locus. We attempted to reassemble the cosmids into a contig based on the results of STS content mapping. However, the instability of YAC 31 thwarted our efforts to create a
contig across the region. Although we selected a clone of YAC 31 that appeared to be a full 1 Mb in size, the DNA apparently contained various internal deletions. In fact, some regions of the DNA appear to be particularly prone to internal deletions because the same genomically non-contiguous regions were co-joined in cosmids constructed independently from YAC 29, which covers part of the nude region. This problem could perhaps have been ameliorated if a YAC DNA band of 1 Mb had been isolated from a PFG before subcloning. A small degree of cosmid chimerism was also attributable to the subcloning procedure. In any case, constructing a contig from cosmids subcloned from a YAC is inherently undesirable, because it does not provide an independent verification of the genomic region.

To obtain an independent representation of the region in smaller insert clones, we constructed a chromosomal walk of the smallest nude region with clones from a genomic bacteriophage P1 library with an average insert size of 85 kb (Genome Systems, Inc., St. Louis, MO). The physical map in P1 clones was initiated at 6 well-spaced loci: BS12.42, ET-28, CA69, Y29L, ET-6, and RDA10.2. (Note: ET-6 and ET-28 are exon trapped clones that will be described in the next chapter.)

When these loci did not recognize an overlapping set of P1 clones, the ends of the P1s were cloned to reveal undetected overlap and to create new STSs from which to extend the contig. The same inverse PCR protocol was used to clone the ends of the P1s as the YACs, except we changed the restriction enzymes and the primers, detailed in the Materials and Methods section and Table VI. The same basic nomenclature is used for naming the ends of the P1s as for the YACs, such that P7N is the end of clone P1 #7 adjacent to the NotI site of the vector and P15S is the end of P1 clone #15 adjacent to the SalI site of the P1 vector. We determined that all of the P1 ends physically mapped to the nude region; i.e. they were contained in YAC 31 and the mouse Chromosome 11 rat hybrid cell line.

To identify the boundaries of the nude region we genetically mapped, by SSCP, P1 ends that physically mapped to the region with an undetermined haplotype. As well, genetic mapping of the P1 ends oriented the chromosomal walks; e.g. P7N is contained in P1s 7 and 9, and recombines with BS12.42, but not with nude in animal C4, orienting P7N as distal to BS12.42 relative to nude. The genotypes of these markers on the recombinant animals are given in Table V and the contig of P1s is shown in Figure 14. The
size of the region covered by the P1s is consistent with the size of the region cloned in YAC 31, as judged by PFG analysis. Finally, *nude* was determined to lie minimally between P7N and 8CA (230 kb) and maximally between BS12.42 and P15S. This chromosomal walk in P1s and the genetic markers contained within represent a complete genetic and physical map of the *nude* region.

**Representation of the genomic structure**

We physically mapped over 150 STSs to this region, including reagents whose development will be discussed in the next chapters. All of the markers showed a unique, consistent physical location in the YAC and P1 contigs. Of the 150 STSs, 44% were subcloned from YAC 31 and 56% were subcloned from the P1 clones. As well, the PFG maps of YACs 29 and 31 and the P1 clones were consistent. Since the YACs and the P1s are completely independent representations of the mouse genome, the chromosomal walks appear to be an accurate reflection of the mouse genomic region.

**CONCLUSIONS**

Chromosomal walks to cover regions of several megabases require large inserts, like YACs. However, YAC libraries are problematic for genomic analysis because they contain inserts that are not colinear with the genomic source DNA; e.g. with deleted, rearranged or chimeric inserts. The analysis of these 23 MIT YACs from the *nude* region, hints at the MIT YAC library's representation of genomic structure. First, we found no unclonable regions: All of the STSs used to screen the MIT YAC library were contained in at least one clone. Second, 76% (22/29) of the cloned YAC ends mapped to the contiguous region; 17% were chimeric, mapping elsewhere in the genome; and 7% (2/29) were repetitive sequence and could not be mapped. Third, the YAC studied in the most detail (YAC 31) was prone to internal deletions, as judged by PFG analysis of individual isolates and the cosmids constructed from the YAC. YAC 31 not only was the most unstable YAC in the region, but it also contained a singly-covered region in the MIT YAC library. A thorough analysis of randomly selected YACs will give a more general characterization of the library since all of these YACs were selected from a specific genomic region. To determine the rate of chimerism of the library, one can take
advantage of the fact that it is usually possible to find a variant within a sequence that can be genetically mapped in an intersubspecific cross. We found that slightly more than half of our YAC ends detected a polymorphism under one standard set of conditions for SSCP.

We were able to compare the correlation between the genetic and physical maps on a fine, but not on a gross level. Mapping *nude* with 2000 meioses should yield an interval of 1/10 cM = 214 kb (3000Mb/1400cM = physical size of the mouse genome/genetic size of the mouse genome). In fact, we narrowed the *nude* locus to a minimum region of 230 kb and a maximum region of 370 kb. We initiated chromosomal walks to span the *nude* region from flanking genetic markers that map 1.4 cM from each other (*D11Mit7* and *D11Mit34*). We spanned the *nude* region before the two contigs joined. Thus, the size of the genomic region covered could only be roughly estimated as larger than 3 Mb. The proximal most marker from the distal walk (Y32L) was not contained in any of the same YACs as the distal most marker from the proximal walk (21L), indicating that the gap between the two contigs was probably at least 1 Mb. However, with the gap uncloned, it is impossible to determine the physical distance between the flanking SSLP markers.

GD-RDA successfully targeted probes that mapped extremely close to *nude*, both genetically and physically. RDA-10.4 mapped 0.05 cM proximal to *nude*, and RDA-10.2 mapped 0.05 cM distal to *nude*. Since both markers were contained in YAC 31, the *nude* locus must be contained within the genomic region corresponding to this YAC. RDA10.2 and RDA10.4 closed the YAC chromosomal walk of the *nude* region in one step and map within 1 Mb of the *nude* gene.

With 6,000 SSLP markers presently available in the mouse, the average spacing between markers is 500 kb and the average non-recombinant interval around a target locus is 1 Mb, distances easily spanned in large insert YACs. In our case, although no MIT SSLP marker is contained in YAC 31, *D11Mit144* and *D11Mit117* are contained in the adjacent YAC 29.

The P1 contig provides an independent verification of the structure of the genomic region. We physically mapped over 150 STSs to this region (44% subcloned from YAC 31 and 56% subcloned from the P1 clones) and they all showed a unique, consistent physical location in the YAC and P1 contigs. As well, the PFG maps of YAC 29, 31 and the P1 clones were consistent. Since the
YACs and the P1s are completely independent representations of the mouse genome in different hosts, the chromosomal walks appear to be an accurate reflection of the genomic region. The P1s are a good starting material for identifying transcription units because the insert DNA is easily purified away from the host DNA.
REFERENCES


Chapter Four

Transcription map of the nude region: Direct cDNA selection and exon trapping to isolate expressed sequences in a region dense in transcribed units.

ABSTRACT

Genetic and physical mapping localized the nude locus to a 370 kb region, minimally spanned by a 1 Mb YAC or a contig of 7 P1 clones. To find expressed sequences in this region, we employed two complementary strategies: direct cDNA selection and exon trapping which yielded 148 and 24 unique gene fragments, respectively. One-quarter of the gene fragments showed strong similarity to Genbank entries, identifying 10 likely transcription units in the region: 7 novel transcripts with similarity to genes from Drosophila, Caenorhabditis elegans, rat, and human; and 3 previously identified mouse genes. To map specific gene fragments into transcription units: (i) we checked the DNA sequences for overlap; (ii) we defined the physical map location of each clone; (iii) we determined their expression pattern by RT-PCR; (iv) we screened cDNA libraries with potentially overlapping gene fragments. Based on our transcription mapping results, we present a novel approach to estimate the number of genes in a region. Specifically, we estimate that the nude locus appears to reside in a region approximately three-fold enriched for genes, with 20-25% of the nucleotides being transcribed.
INTRODUCTION

Construction of dense genetic and physical maps delineate the minimal region that contains a gene of interest. The dense set of genetic markers available in the mouse and the meiotic power of large crosses should narrow the smallest gene-containing region to several hundred kilobase pairs. Rapid isolation of the transcribed sequences from this chromosomal region is critical to identify a causative gene.

Systematic and reliable identification of coding regions within extensive genomic regions is difficult because genes have no common structure: Genes are irregularly dispersed along the chromosome and vary in genomic size and number of exons. As well, many genes are selectively expressed either temporally or spatially. Direct cDNA selection and exon trapping are recently developed methods which alternatively exploit these features of gene structure and expression. Direct cDNA selection is based on recovering cDNA fragments that specifically hybridize to the physical DNA templates (Lovett, et al., 1991). This method is constrained by tissue expression of the gene but not by the genomic structure of the gene. Exon trapping is a strictly genomic approach, that relies upon the fact that most mammalian genes contain multiple internal exons and thus can be spliced into a synthetic vector (Buckler, et al., 1991).

Building a complete transcription map of a region is confounded by the issue of determining if all of the genes have been identified. There are many rough estimates of the number of genes in the mammalian genome, ranging over an order of magnitude from 30,000 to 300,000. The estimates are extrapolated from RNA reassociation studies, the number of CpG islands, or genomic sequencing (Antequera, et al., 1993; Antequera, et al., 1994; Gilbert, 1992; Lewin, 1995; Wagner, et al., 1993). Each of these methods to estimate the total gene number has its own inherent bias. However, it has already become clear that there is a large range of gene density between regions of the genome (Martin-Gallardo, et al., 1992; McCombie, et al., 1992). The gene-rich fraction of the genome probably has at least twice the density of the gene-poor fraction (Fields, et al., 1994). Therefore, methods for calculating gene density for a specific region need to be considered.

This chapter describes our work to identify and characterize expressed sequences in the nude region. To identify genes in the region, we performed
both direct cDNA selection and exon trapping. Based on the results, we were able to compare the ability of the two methods to identify transcription units. In our hands, direct cDNA selection was extremely fruitful, yielding a large number of distinct clones with no redundancy, of which 93% mapped back to the correct physical region. By contrast, exon trapping yielded a smaller set of clones with considerable redundancy.

Based on strong similarity of analyzed gene fragments to Genbank entries, we identified 10 likely transcription units in the region: 7 novel transcripts with similarity to genes from *Drosophila, Caenorhabditis elegans,* rat, and human; and 3 previously identified mouse genes. Based on our transcription mapping results, we present a novel approach to estimate the number of genes in a region. Specifically, we estimate that the *nude* locus appears to reside in a region approximately three-fold enriched for genes, with 20-25% of the nucleotides being transcribed.

RESULTS

Identifying Transcription Units

To find transcription units in the *nude* region, we employed two complementary strategies: direct cDNA selection and exon trapping. Direct cDNA selection recovers cDNA fragments that specifically hybridize to physical DNA templates (Lovett, 1994; Lovett, et al., 1991). This method is constrained by the tissue expression but not by the genomic structure of the gene. Briefly, purified cosmid or P1 DNA, covering the entire 370 kb of the genomic *nude* region, was digested with four-base cutters and ligated to biotin-containing linkers. Primary cDNA was independently prepared by random-priming poly(A)+ selected mRNA from neonatal skin, adult skin, adult testes, and adult thymus. The cDNA and biotinylated genomic DNA were hybridized at high stringency after suppression of repetitive sequences. The biotinylated genomic-cDNA complexes were captured on pre-blocked streptavidin-coated paramagnetic beads and the unbound, nonspecific cDNAs were washed off. The captured cDNAs were eluted and recycled through a second enrichment to increase the selectivity. The double-selected cDNAs were cloned, transformed into bacteria, grown, mini-prepped and sequenced. Each pool of selected cDNAs was analyzed to determine the proportion of
clones that physically mapped back to the *nude* region; i.e. not ribosomal or *E. coli* in origin or non-specifically hybridized clones.

Exon trapping is a strictly genomic approach, relying upon the fact that most mammalian genes contain multiple internal exons which can be spliced into a synthetic vector (Buckler, et al., 1991; Church, et al., 1994). Briefly, pools of cosmids were digested with the restriction enzymes *BamHI*, *BglII* and cloned into the intron of an HIV gene, driven off an SV40 promoter. COS-7 cells are transfected with plasmids containing these constructs, which then process the RNA transcripts in vivo. The resulting COS-7 mRNA is converted to cDNA, amplified with specific primers from the vector flanking sequence and cloned to isolate individual colonies. If the cloned genomic DNA contains an exon, then the transcript will splice together this exon with the flanking 5' and 3' splice sites of the HIV gene. Each pool of clones was analyzed to determine the proportion that contained novel trapped exons; i.e. not splicing from the 5' to the 3' splice site of the HIV gene or to cryptic splice sites in the vector.

Direct cDNA selection provided an extremely deep resource of transcribed sequences. We sequenced and analyzed a total of 180 unique clones with an average insert size of 250 bp without encountering the exact same clone twice. A low background rate of clones was obtained: 2% of the clones were from the *E. coli* genome; 3% were P1 vector sequence; 2% were mouse repetitive sequence. Background problems were reduced by using primary cDNA, rather than a cDNA library grown in bacteria. The sole exception was P1 #21, which yielded 80% bacterial clones; this clone appears on PFG analysis to have deleted all or most of its mouse DNA insert. We found that more than 93% of the clones mapped back to the region by STS content mapping of the P1 clones, yielding 148 unique clones from direct cDNA selection. We attribute this great specificity to the fact that we used two rounds of hybridization and stringent wash conditions. Direct cDNA selection proved to be an extremely effective method to clone transcription units in a physical region from a given tissue source.

With exon trapping, we analyzed a total of 120 clones having an average insert size of 212 bp. In contrast to direct selection, many of the clones occurred multiple times: The 120 clones yielded only 24 distinct sequences. We found at most three unique clones per pool of two cosmids. Even when 8 cosmids were grouped together, a maximum of three unique clones were
identified. An additional problem with the exon trapping procedure was a high degree of background caused by splicing of cryptic splice sites in the vector.

Based on these results, we were able to compare the ability of direct cDNA selection and exon trapping to identify transcription units. In our hands, direct cDNA selection was extremely fruitful, yielding a large number of distinct clones with no redundancy, of which 93% mapped back to the correct physical region. The only caveat is that the method demands knowledge of the expression pattern of the gene to use as a source of cDNA for the selection. By contrast, exon trapping did not give the depth of resources, but it did identify many of the same transcription units as direct selection.

**Sequence Analysis of Transcription Units**

The sequences of the gene fragments were analyzed for nucleotide or amino acid similarities to genes in Genbank, using the programs BLASTN and BLASTX (Altschul, et al., 1990; Altschul, et al., 1990). These search programs were optimized to search for local alignments, allowing for detection of similarities between diverged sequences (Altschul, et al., 1994). Because of the average size of the clones and the number of entries in Genbank, we set the criteria for a significant match to be a Poisson probability score $P(N)$ of less than $10^{-10}$ for nucleotide comparisons or less than $10^{-5}$ for protein comparisons. A number of gene fragments could be grouped as likely to belong to the same transcription units because they shared strong sequence similarity to the same gene in Genbank. All such clones, sharing a strong similarity to a specific gene, mapped to the same physical region in the P1 contig. Appendix I gives the complete details for each unique clone that mapped to the *nude* region: (i) clone name, with the prefix ET indicating that the gene fragment was isolated by exon trapping or DS indicating that the gene fragment was isolated by direct cDNA selection; (ii) primer sequences; (iii) insert sequence; (iv) overlap with other clones; (v) significant nucleotide and protein BLAST matches; (vi) physical map location.

A total of 37 gene fragments showed strong similarity to 10 unique genes in Genbank: 28 of these clones were obtained from direct cDNA selection; 9 of these clones were obtained by exon trapping. In the region that was both exon trapped and direct cDNA selected, the same genes were
identified. The BLAST scores are given in Appendix I and some examples of strong sequence similarities are given in Figure 15. Since one cannot predict the nature of the *nude* gene, many of the transcription units identified were interesting candidates for *nude*. The ten Genbank entries with strong amino acid similarity to gene fragments in the *nude* region are:

1. human HTLF, a winged helix or fork head transcription factor: (5 from direct cDNA selection and 1 from exon trapping) (Li, et al., 1992). A conserved 100 amino acid domain defines fork head transcription factors which have been identified in yeast, *Drosophila*, *C. elegans*, *Xenopus*, mouse, and human. Mouse fork head genes are developmentally regulated during embryogenesis and control cell-specific gene expression in adults.

2. mouse vitronectin gene (2 from direct selection and 1 from exon trapping) (Seiffert, et al., 1993). Perfect nucleotide identity was found to mouse vitronectin, a circulating factor, produced in the liver, that regulates the link between cell adhesion, humoral defense mechanisms, and cell invasion. Although the mouse gene has not been mapped, the human vitronectin gene maps to 17q11, the region that is syntenically conserved in the human with the *nude* region in the mouse.

3. human tumor necrosis factor, alpha induced protein 1 (TNFAIP1). (2 from direct selection and 3 from exon trapping) (Wolf, et al., 1992). The mouse sequence is 95% identical to the human sequence over 1085 bp. Two of the six clones are from the 3’ UTR. TNFAIP1 had been mapped previously to this region of mouse chromosome 11. TNFAIP1 is induced rapidly in endothelial cells in response to tumor necrosis factor-α.

4. *Drosophila* nemo gene (3 from direct selection and 1 from exon trapping) (Choi, et al., 1994). nemo, a serine/threonine protein kinase, is required to initiate the second step of rotation of ommatidia. Rotation is also a common phenomenon in vertebrate embryonic development.

5. *C. elegans* gene emb-5 (4 from direct selection) (Nishiwaki, et al., 1993). emb-5 is required for the correct timing of gut precursor cell division during gastrulation. emb-5 is structurally similar to the *S. cerevisiae* nuclear protein SPT6, which inhibits transcription of various genes, possibly by regulating chromatin assembly.

6. rat Na+/sulfate cotransporter gene (2 from direct selection and 1 from exon trapping) (Markovich, et al., 1993). This transporter is involved in sulfate reabsorption in the kidney, intestine and colon.
(7) *O. cuniculus* Ad-Rab G (3 from direct selection and 1 from exon trapping) (Boll, et al., 1993). Strong nucleotide as well as amino acid similarity was found to this rabbit transcript, cloned in a subtractive hybridization of genes expressed in the intestine of adult but not baby rabbits. No functional characterization of this gene was reported.

(8) mouse fructose aldolase C gene (4 from direct selection) (Paolella, et al., 1986). Perfect nucleotide identity was found to this glycolytic enzyme.

(9) mouse rah GTP-binding protein (2 direct selected clones) (Morimoto, et al., 1991). Strong nucleotide similarity was found to this transcript, whose protein product may function in vesicular trafficking and neurotransmitter secretion.

(10) human MAC30 mRNA, 3' end sequence (1 from direct selection and 1 from exon trapping) (Murphy, et al., 1993). Strong nucleotide similarity was found to this transcript that is down-regulated in meningiomas and in tumors associated with neurofibromatosis 2.

**Estimating the number of genes in the region**

If we assume that the genes in this region are of similar size and that fragments of these genes are recovered at similar frequencies by direct cDNA selection and exon trapping, then the number of genes can be estimated by four independent approaches:

1. **Number of times that fragments from specific genes were recovered.**

   Of the 172 gene fragments examined, 37/172 (=22%) showed strong sequence similarity to previously identified genes and could be grouped into 10 transcription units. The number of gene fragments corresponding to each of the 10 transcription units was 6, 5, 4, 4, 4, 4, 3, 3, 2, 2 with a mean of 3.7 ± 1.2. The actual mean number of hits to these 10 genes is probably somewhat higher, inasmuch as similarities to the untranslated regions would not be expected to have been recognized for the more distant similarities (e.g., *C. elegans* emb-5). Since four of the similarities are only detected on the amino acid level and since untranslated regions are typically about 40% as large as coding regions (based on a random sampling of genes from Genbank), the actual mean might be 20% larger — i.e., about 4.5. Assuming that the hit rate for these 10 genes is a good estimate of the hit rate across the region, we would estimate that the 172 gene fragments represent between 38 (= 172/4.5) and 46 (= 172/3.7) genes.
(2) Number of overlaps among gene fragments. Of the 172 gene fragments with an average insert size of 250 bp, a total of 86 showed significant overlap (>40 bp) with another fragment. For a random collection of fragments, the expected number n of overlaps per clone is given by the formula: 

\[ n = c(1-\theta) \]

where c is the degree of coverage of the region and \( \theta \) is the minimum detectable proportion of overlap (Lander, et al., 1988). In the current case, \( n = \frac{86}{172} \) and \( \theta = \frac{40}{250} \). The estimated coverage of the transcribed portion of the region would thus be \( c = 0.60 \)-fold. Since the clones contain a total of 43 kb of sequence (172 clones x 250 bp/clone), this would suggest that the transcribed portion of the region is about 72 kb (= 43 kb / 0.60). The proportion of the 370 kb region that is transcribed would thus be estimated to be about 20%. Taking the typical size of a mature transcript to be 2 kb (J.S., unpublished observation based on a random sampling of genes from Genbank), this would correspond to about 36 genes.

(3) Degree of coverage of known genes. Of the 10 defined transcription units corresponding to previously known genes, three were known mouse genes (vitronectin, aldolase, rah GTP-binding protein) and three others were mammalian genes showing strong sequence similarity (various winged-helix (fork head) genes, rat sodium-sulfate co-transporter, and rabbit Ad-Rab G). All gene fragments arising from the first group should have been recognized (due to sequence identity) as should most of those arising from coding regions in the second group (due to apparently strong sequence similarity across the coding region). For these six genes, we could thus directly measure the degree of coverage—that is, the average number of times that a given nucleotide is hit. Since the six genes contain a total of 13,775 nucleotides and the recovered gene fragments a total of 6,275 nucleotides, the coverage is 0.46-fold. Assuming that this coverage is representative for the region, the total length of transcribed sequence in the region is estimated to be 93 kb (= 43 kb in gene fragments / 0.46-fold coverage). The proportion of the 340 kb region that is transcribed would thus be estimated to be about 25%. Again taking the typical size of a mature transcript to be 2 kb, this would correspond to about 46 genes.

(4) Proportion of genes similar to known genes. Finally, about 30% of newly-sequenced mammalian genes show strong sequence similarity to previously identified genes in Genbank (Adams, et al., 1993) at present. Since 10 such genes were identified in the region, this would suggest a total of about 33 (=10 / 0.3) genes.
To estimate the total number of genes in the *nude* region, we employed four methods based on the redundancy of gene fragments; the overlap among the total set of gene fragments; the coverage of known genes in the region; and the proportion of gene fragments showing similarity to genes in Genbank. These four independent approaches suggest that about 20-25% of the nucleotides in the *nude* region are transcribed and that this region contains in the range of 33-46 genes. The assumptions that the genes in the region are recovered at similar frequencies and are of similar size are unlikely to be exactly true, but they are probably reasonable approximations. Differential rates of recovery would tend to lead to underestimates of the number of genes, while the presence of a few exceptionally large genes would lead to overestimates.

*Mapping gene fragments into transcription units*

To map the gene fragments into transcription units: (i) we checked for DNA sequence overlap among the gene fragments; (ii) we defined the exact physical map location of each clone within the nude region; (iii) we determined the expression pattern of the gene fragments by RT-PCR; (iv) we screened cDNA libraries with potentially overlapping gene fragments to ascertain if they were contained in the same unique clone. Two clones are considered to come from the same transcription unit if either the DNA sequence of the two clones overlap or they are both contained in the same unique cDNA library clone. It is suggestive that two clones may come from the same transcription unit if the clones map to the same physical region and share the same expression pattern.

Each sequence was assayed for overlap of at least 40 bp with all the other gene fragments. A computer program broke down each clone into 10 bp bits of contiguous sequence and scanned the other clones for 4 or more identical non-overlapping bits. Of the 172 gene fragments with an average insert size of 250 bp, a total of 86 showed significant overlap (>40 bp) with another fragment. The overlap between each pair of clones is shown in Appendix I.

The complete set of P1s delineate 30 non-overlapping physical intervals (BINS) of average size 13 kb. The gene fragments are randomly distributed with the notable exception that very few clones mapped to the region of P1 #21. 80% of the direct cDNA selected clones from P1 #21 were
derived from bacteria, rendering this region depauperized in mouse cDNA clones. Those gene fragments presumed to belong to a common transcription unit (by virtue of strong similarity to a gene in Genbank or overlap in sequence) always mapped to the same or adjoining BIN. The physical map location of each clone is given in Appendix I. The physical map location of each of the transcription units is shown in Figure 16.

To determine the expression pattern of gene fragments, we performed reverse-transcribed PCR (RT-PCR) on a panel of adult mouse tissues (including skin, thymus, liver, testes, brain, heart, kidney) and 10.5 dpc and 11.5 dpc embryos. Since genomic DNA contamination is a source of false positive results in RT-PCR experiments, the first strand cDNA samples were treated with DNase prior to performing PCR. Residual genomic DNA contamination was monitored by performing RT-PCR on each sample with primers, flanking a small intron of the ubiquitously expressed mouse profilin gene. We minimized the impact of lingering genomic DNA contamination in the first strand cDNA template by spanning intron-exon boundaries whenever possible. To determine the reproducibility of the RT-PCR results, we performed two separate trials of 16 gene fragments' expression pattern in 15 independently prepared first strand cDNA samples. 84% of the tissues gave the same result for both trials; 10% of the tissues were positive for only the second trial; and 6% of the tissues were positive for only the first trial. We were unable to define conditions that avoided the stochastic nature of PCR. With this as a caveat, those gene fragments presumed to belong to a common transcription unit (by virtue of strong similarity to an entry in Genbank or overlap in sequence) showed a consistent pattern of expression. The expression pattern for transcription units characterized by RT-PCR are given in Table VII. Examples of the RT-PCR primary results, including those for the control profilin gene, are shown in Figure 17.

Gene fragments, showing no sequence similarity to each other or previously identified genes, were grouped according to their physical map position and expression pattern. We checked whether several pairs of non-overlapping pairs of clones that both mapped to the same BIN and were expressed in the same tissues were from the same gene, by virtue of both being contained in the same unique clone from a cDNA library. We found several examples of clones that were contained in the same cDNA clone. Since genes are alternatively spliced and cDNA clones are often not full-
length, two gene fragments may be from the same transcription unit, but not contained in the same cDNA library clone.

CONCLUSIONS

Finding Transcription Units

In our hands, direct cDNA selection proved to be a powerful method for identifying a deep pool of unique sequences, corresponding to transcription units in a genomic region. Direct cDNA selection identified 148 unique transcripts of average insert size 250 kb from the 370 kb nude region. The only caveat is that the method demands knowledge of the expression pattern of the gene to use as a source of cDNA for the selection. Exon trapping did not give the depth of resources, but it did identify many of the same transcription units as direct cDNA selection. When a mammalian genome has been sequenced, the challenge will be to find the open reading frames in the sequence and to determine which belong to the same gene.

Gene density

Given an estimate of 100,000 total genes in the mammalian genome, there should on average be 12 genes in a 370 kb region, like nude. However, some regions appear to be gene rich (high-GC content, Alu rich or Giemsa light bands), having a much higher gene density than average. In fact, the middle region of mouse Chromosome 11, to which nude maps, has light Giemsa band staining (Buchberg, et al., 1993).

To estimate the total number of genes in the nude region, we employed four methods based on the redundancy of gene fragments; the overlap among the total set of gene fragments; the coverage of known genes in the region; and the proportion of gene fragments showing similarity to genes in Genbank. These four methods estimated that 20-25% of the nucleotides in the nude region are transcribed and that the expected number of genes in the smallest nude region is in the range of 33 to 46; this would seem to indicate that nude lies in a gene-rich region of the genome. Interestingly, the 300 kb human genomic region containing the NF1 gene contains only 4 transcription units (Xu, et al., 1990). Recall that in the mouse, Nf-1 physically maps within 1 Mb of the nude locus. Although these results were gathered in different species, it reminds us that the density of genes may
be very locally determined. With identifying all of the genes as one of the official goals of the U.S. Human Genome Program, some day we will know the answer.

*Building a transcription map of a chromosomal region*

To assemble a complete transcription map from a chromosomal region remains a challenge in positional cloning projects. Building a complete transcription map of a region is confounded by the variety of gene structures and the ability to determine if all of the genes have been identified. Since direct cDNA selection gives a deep pool of non-redundant gene fragments, we could continue cloning new fragments and assemble transcription units based on the sequence overlap. Tissue sources which express a high proportion of transcribed sequences, like brain or testes, could be used as a source of primary mRNA to increase the representation of transcription units.

To improve positional cloning of human loci, genomic resources of the sequence of large numbers of cDNAs and large regions of the genome are currently underway (Adams, et al., 1995). The technology now exists to sequence minimal chromosomal regions containing a locus and to identify open reading frames from the sequence with computer programs, like GRAIL (Lopez, et al., 1994; Roberts, 1991; Xu, et al., 1994). Transcription units still need to be assembled from the putative exons. In conjunction, large numbers of expressed sequences are now being identified in the human and mapped onto the human physical map. Therefore, a computer search can scan a genomic region for expressed sequences that have already been mapped to the region (Hudson, et al., 1995).
REFERENCES


Chapter Five

Mutation detection and expression studies
of the *nude* gene, a novel fork head transcription factor.

A novel fork head transcription factor, *Hfh11*, identified by both direct cDNA selection and exon trapping of the smallest *nude* region, results in the *nude* phenotype when mutated. Disruptions in *Hfh11* have been detected in all four of the *nude* rodent alleles: a single-base-pair deletion in the *nu* allele, a marked decrease in expression levels of the *Hfh11* transcript in the skin of *nu* *str* mice, a nonsense mutation in the *rnu* allele, and a large genomic deletion removing several exons of the *rnu* *N* allele. *Hfh11* produces a 4 kb transcript, which encodes a protein of 648 amino acids. To obtain direct biological proof that *Hfh11* is responsible for the *nude* phenotype, we microinjected a cosmid clone containing the wild-type genomic locus into fertilized *nude* eggs. Two independent founder lines of transgenic mice were generated that correct the hairless phenotype, but not the thymic defect. These complementation results suggest that the *nude* locus is subject to complex regulation. *Hfh11* is expressed in the adult thymus, initiating in the developing embryo as thymic organogenesis occurs. High levels of *Hfh11* expression are detected during the active growth phase of the hair follicle in the keratinized region of the shaft. We detected normal levels of expression in *nude* hair shafts of the cloned keratin genes whose expression patterns would have been consistent with being downstream targets of the *nude* protein. The mutations detected in four alleles of the gene, the expression studies, and the rescue of the hairless phenotype demonstrate that *Hfh11* is the *nude* gene.
INTRODUCTION

Over the past century, hundreds of transmitting mutant mice have been identified and characterized (Green, 1989). Some mutants arose spontaneously, while others have been induced by chemicals, radiation, viral integration, transgene insertion, or targeted disruption of genes in embryonic stem cells. Mutagenic agents create hallmark types of mutations; e.g. The chemical ethynitrosourea (ENU) generally create single base pair changes, whereas X-ray radiation generally creates large genomic rearrangements (Kingsley, et al., 1990; Rinchik, et al., 1990; Rinchik, et al., 1986; Russell, et al., 1989; Russell, 1951; Russell, et al., 1979; Woychik, et al., 1990). To understand the molecular nature of the block in the inductive pathway for a given mutant, the disrupted gene must be identified. When a targeted disruption creates a mutant phenotype, the gene is already known. In all other cases, the gene must be cloned based either on some property of the mutagen or on its chromosomal location. The underlying gene is marked by foreign DNA when the mutation is induced by a viral or transgenic insertion. For the vast class of spontaneously-arising or chemical/radiation induced mutants, the underlying causative gene must be identified based on its chromosomal location. Sometimes candidate genes are genetically map to the same region as the phenotype. Otherwise, the transcription units in the chromosomal region must be identified. In either case, disruptions must be demonstrated in mutant allele(s) of the gene to prove that the causative gene has been identified.

Reviewing the mouse genes that have been positionally cloned, no single method for mutation detection has succeeded to identify the underlying causative gene. The first genes positionally cloned in the mouse, T (Brachyury), short-ears (Bmp-5) and agouti, have multiple alleles, many with chromosomal rearrangements. Disruptions of four spontaneously arising and two radiation induced alleles of T were gross enough to be detected on Southern blots (Herrmann, et al., 1990). No rearrangements or deletions of Bmp-5 sequences were observed in the mice carrying the spontaneously arising mutation. However, Southern blots detected deletions for 4 alleles and altered restriction fragments for 2 alleles of the hundreds isolated in a radiation/chemical mutagenesis experiment (Kingsley, et al., 1992). The agouti gene was first cloned because a radiation-induced inversion
joined the agouti locus and another cloned gene (*limb deformity*). The gene found at the novel end of the breakpoint detected structural alterations in 4 of the 18 induced or spontaneous agouti alleles (Bultman, et al., 1992; Miller, et al., 1993). Multiple alleles with chromosomal rearrangements are a great assistance when attempting to identify a gene based on its chromosomal location.

However, most mutants have only a small number of alleles, many spontaneously arising, perhaps with only single base pair changes in the gene; e.g., *Bcg, kr, ob*. Resistance to Mycobacterium bovis infection (*Bcg*) was genetically mapped to a 400 kb region that contains a putative transporter protein, expressed exclusively in macrophage populations, with a nonconservative Gly to Asp substitution within a predicted transmembrane domain that is associated with the susceptibility to infection in 13 strains. The chromosomal location of the mouse *kreisler* (*kr*) mutation was defined based on an X-ray induced inversion of at least 1 Mb. To confirm that a candidate gene, located at the inversion breakpoint, was *kreisler*, a second *kr* allele was generated by ENU mutagenesis in which a Ser is substituted for an Asp residue in the functional domain of the protein. Mutations in the *obese* (*ob*) gene result in profound obesity and type II diabetes. Both alleles of *ob* showed detectable differences on Northern blots of adipose tissue RNA - one allele lacked the transcript completely and the other allele showed a 20-fold increase. The mutations were an insertion near the promoter region and a nonsense mutation in the coding sequence. The subtle nature of many of the spontaneously-arising mutations and the complexity of transcription units prevent any singular method's ability to identify the underlying causative gene.

We estimated, based on sampling and redundancy, that the 370 kb *nude* region contains in the range of 33-46 genes. We analyzed the gene fragments, obtained by direct cDNA selection and exon trapping, from the region to detect alterations in the DNA sequence or expression levels between wild-type and mutant animals. The temporal and spatial expression patterns hinted at ways to prioritize the candidate genes. As well, the similarities to Genbank entries suggested which transcripts might be important in mouse development.

A novel fork head transcription factor was identified in the smallest *nude* region that showed specific expression in skin and thymus. *Fork head*
(fkh) was identified in the original Drosophila mutant screen as a gene that promotes terminal instead of segmental development (Jurgens, et al., 1988; Nusslein-Volhard, et al., 1985). The fork head gene was cloned and shown to encode a novel DNA-binding protein (Weigel, et al., 1989). In the last six years, however, fork heads have emerged as a major class of transcription factors, characterized by a 100 amino-acid DNA binding domain in the structure of a winged-helix (Weigel, et al., 1990). The class caught the interest of developmental biologists because the first members, fkh, and the rat hepatocyte nuclear factor 3 (HNF-3) proteins (HNF-3α, HNF-3β, HNF-3γ), were involved in the development of gut or gut-derived organs (Jurgens and Weigel, 1988; Lai, et al., 1990; Lai, et al., 1991; Weigel, et al., 1989). Multiple fork heads have since been identified in yeast, flies, nematodes, frogs, chickens, zebrafish, humans and shown to have diverse and important roles in development and differentiation, as demonstrated by expression patterns and genetic mutations (Dirksen, et al., 1992; Galili, et al., 1993; Hermann-Le Denmat, et al., 1994; Hope, 1994; Knochel, et al., 1992; Li, et al., 1993; Miller, et al., 1993; Murphy, et al., 1994; Pierrou, et al., 1994; Ruiz i Altaba, et al., 1992; Strahle, et al., 1993). Some notable examples of FHs in development include: XFKH1, also known as pintallavis or XFD1, is expressed in the blastopore lip of Xenopus, is rapidly induced by activin treatment of animal caps in the presence of cyclohexamide, and is suggested to play a role in the initiation of axis formation (Dirksen and Jamrich, 1992; Knochel, et al., 1992; Ruiz i Altaba and Jessell, 1992); Axial, a fork head cloned in the zebrafish, is expressed in the fish equivalent of the amphibian organizer, and seems to play a crucial role in specification of both the axial mesendoderm and the ventral central nervous system (Strahle, et al., 1993); lin-31, a Caenorhabditis elegans FH, regulates vulval cell fates progenitor cells' fate decisions (Miller, et al., 1993); Finally, the identification that the chicken sarcoma virus qin is a FH, as well as the discovery that the fusion of a FH domain to PAX3 is the cause of a solid tumour alveolar rhabdomyosarcoma, suggest that some FH proteins may also regulate cell proliferation (Galili, et al., 1993; Li and Vogt, 1993). A dozen fork head proteins have been identified in the mouse and shown to be spatially and temporally regulated during development (Ang, et al., 1993; Avraham, et al., 1995; Clevelence, et al., 1994; Kaestner, et al., 1993; Monaghan, et al., 1993; Sasaki, et al., 1993). Expression studies suggest wide-ranging roles for members of this family both early in mammalian...
embryogenesis and later in organogenesis (Ang, et al., 1993; Kaestner, et al., 1993). The best characterized mammalian FHs are the HNF-3 proteins. Though the HNF-3 proteins were cloned based on their expression in adult liver, further experiments suggested that these proteins play a significant role in early embryonic development and later in initiation and maintenance of the endodermal lineage. The expression pattern of the HNF-3s suggest that these proteins define regions of the developing gut: HNF-3β is expressed earliest at the anterior end of the primitive streak in all three germ layers; HNF-3α is transcribed in the invaginating foregut; and HNF-3γ appears upon hindgut differentiation. HNF-3β is expressed earliest in the node, notochord, and floor plate - all populations of cells undergoing commitment to different developmental fates (Ang, et al., 1993; Monaghan, et al., 1993).

A wealth of information has emerged about the structure of the DNA-binding domain. The highly conserved FH domain has been shown to be necessary and sufficient for DNA binding. The functional importance of the FH DNA binding region was delineated by binding assays with deletion mutants of the rat HNF-3α protein to its target site in the transthyretin (TTR) promoter (Lai, et al., 1990). Deletion of 19 amino acids from the FH domain abolishes the DNA binding activity of HNF-3α. As well, one of the Drosophila fkh mutants is an in-frame deletion of only 6 amino acids within the FH domain. (Weigel, et al., 1989) It has recently been shown that the nuclear localization signal overlaps with the FH domain (Qian, et al., 1995).

Despite the strong sequence similarity in the DNA-binding domain, there is a surprising range of DNA sequences recognized by family members. At least 50% of the amino acids are conserved in the binding domain of all family members. Some members of the family from different species, like HNF-3α and fkh, share 90% amino acid identity in the FH domain (Lai, et al., 1991). The three HNF-3 proteins were able to bind with high affinity to a diverse and distinct set of oligonucleotides. In contrast, another member of the FH family, brain-factor-1 (BF-1), binds a selective set of oligos, defining a consensus binding site. Intriguingly, the HNF-3 proteins appear to act as transcriptional activators of TTR in hepatic cells and as inhibitors on the glucagon gene in pancreas cells (Lai, et al., 1993).

The three-dimensional structure of the FH DNA-binding domain bound to its target revealed a new DNA-binding structure, termed a winged-helix. Burley and co-workers determined the 2.5A resolution X-ray structure
of the DNA-binding domain (residues 107-223) of HNF-3γ complexed to the 13 bp TTR promoter target site, \((\text{GACTAAGTCAACC})\). The DNA binding domain takes on an α-helix/β-sheet structure that is composed of three N-terminal helices and a three-stranded antiparallel β sheet. In total, 14 residues, scattered throughout the primary sequence, make direct or water-mediated contacts with the phosphate backbone of one or both strands. The amount of surface area buried in the HNF-3γ-TTR complex ensures high affinity binding by the monomeric winged-helix motif (Clark, et al., 1993). HNF-3γ has a domain spanning two helices that is structurally related to the helix-turn-helix (HTH) motif of E. coli catabolite gene activator protein (CAP), the eukaryotic homeodomain of the engrailed protein, and the nucleosome binding organizer, histone H5 (Brennan, 1993). The structure of the binding domain has been termed a winged-helix motif, since the protein when complexed to its binding site somewhat resembles “a butterfly perched on a straight rod” (Clark, et al., 1993).

The role of FH proteins in developmental decisions made the novel fork head transcription factor identified in the \(\text{nude}\) genomic region an intriguing candidate as the causative gene for the \(\text{nude}\) phenotype. Mutations were identified in this novel FH gene in all four \(\text{nude}\) alleles by T. Boehm’s laboratory and our group (Nehls, et al., 1994; Segre, et al., 1995). To obtain direct biological proof this gene, named \(\text{Hfh11}\), is the \(\text{nude}\) gene, we microinjected a cosmid clone containing the wild-type \(\text{nude}\) genomic locus into fertilized \(\text{nude}\) eggs, correcting the hairless, but not the athymic phenotype. This partial rescue of the \(\text{nude}\) phenotype in two independent transgenic lines demonstrates that the \(\text{Hfh11}\) gene is indeed the \(\text{nude}\) gene—or at least the gene responsible for the hairless defect in \(\text{nude}\) mice. Although it is formally possible that a second nearby gene is responsible for the thymic defect, we interpret the results to suggest that the \(\text{nude}\) locus is subject to complex differential regulation in skin and thymus, such that the genomic DNA introduced contains regulatory signals sufficient for expression in skin but not thymus. The mutations detected in four alleles of the gene and the rescue of the hairless phenotype demonstrate that \(\text{Hfh11}\) is the \(\text{nude}\) gene.

The expression pattern of \(\text{Hfh11}\) is consistent with the phenotype observed when the gene is disrupted. \(\text{Hfh11}\) is expressed in the adult thymus and turns on in the developing embryo during thymic organogenesis. Since hairs cycle synchronously and the cellular parts of the hair are visibly
different, the exact spatial and temporal expression pattern of *nude* in the hair follicle can be determined. High levels of *Hfh11* expression are detected during the active growth phase of the hair follicle in the keratinized region of the shaft.

The hair follicle is an ideal system to attempt to identify downstream targets of the *nude* protein because it is straight-forward to test for altered expression of candidate genes in the skin of *nude* mice. Flanagan’s initial characterization of the *nude* phenotype suggested that the extensive network of cross-linked keratin filaments is not formed in the hair shafts of *nude* mice (Flanagan, 1966). Keratins are intermediate filaments (IFs) with a characteristic α-helical domain of 400 to 500 amino acids, arranged in sequences of heptad repeats. Keratins are divided into two superfamilies: type I keratins are smaller and acidic, whereas type II keratins are larger and more basic. Keratin IFs are obligate heteropolymers of one chain from each of the two superfamilies (Coulombe, 1993). As determined by two-dimensional gel electrophoresis of keratin extracts of hair follicles, 8 major keratins are produced in the hair, 4 from each superfamily. The type I keratins were designated Ha1-4 and the type II keratins were designated Hb1-4 (Heid, et al., 1986). A panel of antibodies generated against the hair keratins showed that some of these “hair-specific” keratins were produced in cells forming nails, the tongue and, surprisingly, the thymus (Heid, et al., 1988). Although these proteins may be important in thymic development, the cells expressing these proteins may also represent the broad spectrum of cell-differentiation-related antigens that are involved in the self-tolerance restriction of maturing lymphocytes. The connection of keratin expression in the two organs affected in *nude* mice is intriguing, however. We investigated the expression in *nude* and wild-type skin of cloned murine hair keratins as possible downstream targets of *Hfh11*. The expression levels of Ha1, Ha2, Ha3 were normal in *nude* skin.

RESULTS

*Mutation detection of the gene fragments*

To detect gross alterations in the genomic DNA or the expression levels between wild-type and mutant animals, we analyzed the gene fragments on Southern blots and by reverse-transcribed PCR (RT-PCR).
Clones were characterized by hybridization to a Southern blot of restriction-digested genomic DNA from wild type and nude mice. No deletions or alterations were found. An example of a gene fragment hybridized to TaqI and HindIII digests of nude and wild-type digested genomic DNA is given in Figure 18. We scanned for loss of expression of the gene fragments in nude skin by reverse-transcribed PCR (RT-PCR). Although RT-PCR, as performed, was not sufficiently quantitative to detect subtle alterations in expression levels, we hoped to identify any allele that resulted in a complete lack of RNA. As well, we amplified each gene fragment by PCR from nude and wild-type animals, searching for a genomic deletion. As an example, the amplification of ET-X90 from first-strand cDNA of mutant and wild type skins and from nude genomic DNA is given in Figure 17. No gross alterations in expression level or genomic deletions were detected with the gene fragments.

To detect more subtle alterations in the DNA sequence or the expression levels of the transcription units, we began (i) sequencing the coding region of transcription units from nude and wild-type cDNA; and (ii) analyzing the size and expression of transcription units by Northern blot analysis. To identify larger transcription units, gene fragments were hybridized to skin, thymus, embryonic 10.5 or embryonic 11.5 cDNA libraries with average insert sizes of 2 to 2.5 kb. Individual clones were picked and the insert was sequenced with primers from the vector sequence. Specific primers were selected from the insert sequence and used to determine the sequence of the wild-type and mutant cDNAs. Northern blots were prepared with polyA+-selected mRNA from nude and wild type skins, thymus, heart, liver, kidney. As these experiments are time consuming and require some precious reagents, we needed to focus first on our best candidate genes from the region.

The temporal and spatial expression patterns, as determined by RT-PCR, hint at ways to prioritize the candidate genes from the region. As well, the similarities to Genbank entries suggest which transcripts might be important in mouse development. In particular, five clones (ET-X90, DS-A6 DS-a10g, DS-a8d, DS-z11h) show strong similarity to fork head transcription factors, a class of proteins defined by a conserved 100 amino acid DNA binding domain with the structure of a winged helix (Appendix I) (Brennan, 1993; Clark, et al., 1993; Weigel and Jackle, 1990). Other mouse fork head
transcription factors are essential to proper embryonic development, with spatially and temporally restricted patterns of expression (Ang, et al., 1994; Weinstein, et al., 1994; Xuan, et al., 1995). These gene fragments were all contained in P11, P12, P13, P22, P23 and sequence overlap was detected between DS-a8d, DS-z11h, DS-a10g, and DS-A6 (Appendix I). Furthermore, this transcription unit showed specific expression in the adult skin and thymus and in the developing embryo during thymic organogenesis (Shown as ET-90 in Figure 17).

Mutations at the nude locus and characterization of the nude gene

While we were completing the analysis of the candidate genes from the nude region, Nehls et al. (1994) reported that the fork head/winged helix homologue above has mutations in the mouse nu allele and the rat rnuN allele. These authors originally named the gene whn, for winged helix in nude. To be consistent with mouse nomenclature, the gene has been renamed Hfh11 for HNF-3/ fork head homologue 11. Hfh11 is a 4 kb transcript in adult skin, with 8 coding exons, producing a protein of 648 amino acids. The nude allele, nu, has a single base-pair (G) deletion in exon 3. The sequence of cDNA from C57BL/6J and C57BL/6J-nu from this region is shown in Figure 19. No mutations were found in the coding region in the nu s t r allele. However, Northern blot analysis indicates that expression of the Hfh11 transcript is reduced at least ten-fold in adult skin from AKR/J-nu s t r homozygotes, as compared to AKR/J (Figure 20). Thus, the nu s t r mutation appears to be a more subtle change in the untranslated region of the gene and the exact sequence change remains to be identified. The rat rnuN allele has two variant transcripts of about 3 kb and 1.5 kb in adult skin. Analysis of the genomic rnuN DNA suggested that there was an intergenic deletion encompassing exons 5 and 6, which encode the N-terminal half of the DNA-binding domain. Sequence analysis of the two variant transcripts from the rnuN allele indicated that the mutant transcripts diverge from the wild-type sequence after exon 4, before the presumptive DNA-binding domain. The aberrant transcripts have translational stops 24 and 12 amino acids after they diverge from the wild-type sequence. Based on sequence similarity between an endogenous rat retrovirus and the aberrant rnuN transcripts, Jones and Jesson suggest that the disruption in the rat nude locus arose as a result of an integration of an endogenous retrovirus into the fourth intron that is then
spliced to the *nude* transcript (Jones, et al., 1995). As the *rnuN* allele of *nude* rats is not available in the United States, we were unable to verify these results. To analyze the rat *rnu* allele, we cloned and determined the entire coding sequence of the wildtype and mutant rat *Hfh11* genes (shown in Figure 21). The *rnu* allele is a nonsense mutation at basepair 1429. The change from a C to a T in the cDNA sequence of wild-type versus *rnu* cDNA is shown in Figure 19. Disruptions in *Hfh11* have been detected for all four of the *nude* rodent alleles: a single-base-pair deletion in the *nu* allele, a marked decrease in expression levels of the *Hfh11* transcript in *nustr* skin, a nonsense mutation in the *rnu* allele, and a large genomic deletion removing several exons of the *rnuN* allele.

Our results in conjunction with those from Boehm's group determined the genomic structure of the *nude* gene in the mouse. By examining clones from a skin cDNA library, Nehls et al. assembled a partial sequence of this transcription unit consisting of 9 exons comprising 2.5kb in length. Since a polyadenylation site was not located and the cDNA sequence is 1.5 kb smaller than the transcript size on a Northern, probably neither the complete 5' nor the 3' untranslated sequences have been determined. The first reported exon is 64 bp and contains only 5' untranslated sequence, including an in-frame stop. The translation starts in exon 2 and ends in the middle of exon 9. The lengths of the reported exons are 64, 151, 463, 114, 131, 97, 208, 492, and 783 bp. We determined the lengths of the introns within the coding region by long-range PCR on cosmids and P1 clones from the *Hfh11* region to be 0.5, 2, 1.8, 1, 5.8, 0.3 and 1.6 kb. The first intron was too large to amplify by PCR, but restriction mapping of cos1204 (position in physical map shown in Figure 14) indicated that the intron is approximately 10.5 kb. These lengths are consistent with the restriction maps of the cosmids 1193 and 1204, shown in Figure 22. Thus, the genomic distance from exon 1 to exon 9 is approximately 26 kb.

**Cosmid rescue of the hairless, but not the athymic phenotype of nude mice**

To obtain direct biological proof that the *Hfh11* gene is the *nude* gene, we sought to correct the phenotype by inserting the wild-type genomic *Hfh11* locus into a *nude* background. The hairless, but not the athymic phenotype was corrected when fertilized *nu/nu* eggs were microinjected with cosmid 1193, that includes the coding exons of the *Hfh11* gene. Cosmid 1193
contains 8.5 kb of sequence 5' to exon 2 and 4 kb of sequence 3' of exon 9, but
does not contain exon 1. Recall that the first reported exon is non-coding
and may not be present in all forms of the nude transcript. Cosmid 1193 has
a chimeric insert, consisting of 26 kb from the Hfh11 locus and 10 kb of yeast
DNA (from the YAC host) 3' to exon 9. Cosmid 1204 contains exons 1
through 9 of Hfh11, as well as 12 kb of sequence 5' to exon 1 and 6 kb of
sequence 3' to exon 9; the total insert size is 42 kb. Unfortunately, we were
unable to obtain transgenic mice with cosmid 1204. (Figure 22).

To produce the transgenic mice, male pronuclei of fertilized nu/nu
eggs were injected with approximately 100 copies of the circular cosmid clone
1193. From 800 microinjected eggs, 213 survived until the two-cell stage and
were introduced into host mothers. Twenty-two newborns were obtained,
including one female and one male transgenic mice (named E1 and G2,
respectively), which both had white hair over their entire body. Integration
of the cosmid clone into the genome was confirmed by both Southern blot
analysis and PCR. HindIII digested DNA from the two founder mice,
hybridized with the cosmid vector showed the expected strong band of 6.3 kb,
which was absent in non-transgenic littermates (Figure 23). E1 and G2
contained roughly 7 and 30 copies of the transgene, respectively (Figure 23).
To confirm that cosmid 1193 had integrated intact into the mouse DNA, we
performed PCR with one primer from the cosmid vector and one primer
from the cosmid insert. Reactions for both the T3 and the T7 ends of the
cosmid clones yielded specific PCR products in DNA from the uninjected
cosmid and from the transgenic 1193 mouse line.

Even though the two lines of transgenic mice both had white hair over
their entire body, the correction of the hairless phenotype was not complete
and the levels differed between the two founder mice (Figure 24). Compared
to wild type, the density of hair was much less in the E1 but only slightly less
in G2 animal. Founder E1 was backcrossed to a homozygous nude male, and
half of its offspring showed a phenotype essentially identical phenotype to E1,
demonstrating that the partial correction of the hairless phenotype seen in
founder E1 was not due to mosaicism for the transgene.

The hair length and structure of both founders appeared normal. To
investigate the phenotype on a finer level, we examined the histology of skin
sections from the transgenic mice. Whereas hair shafts from nude mice are
bent and typically fail to break the skin surface, the shafts from the transgenic
mice were straight and protruded from the skin. No obvious histological differences were observed between the wild-type and transgenic mice skin or hairs (Figure 25).

Neither a thymus nor a thymus-like organ could be found in either founder mouse, nor in any of the transgenic offspring of E1 (Figure 26). Splenocytes and peripheral blood cells were also analyzed in Hfh11 transgenic, control wild-type, and control nude mice. Two-color flow cytometric analysis was performed on these cells with combinations of antibodies against CD4/CD8, or Thy-1/B220. Spleen and peripheral blood cells from the Hfh11 transgenic mice were negative for all T cell surface antigens tested (Figure 27).

This partial rescue of the nude phenotype in two independent transgenic lines demonstrates that the Hfh11 gene is indeed the nude gene—or at least the gene responsible for the hairless defect in nude mice. Although it is formally possible that a second nearby gene is responsible for the thymic defect, we interpret the results to suggest that the nude locus is subject to complex differential regulation in skin and thymus, such that the genomic DNA introduced contains regulatory signals sufficient for expression in skin but not thymus.

Expression studies with the nude transcript in the hair

To explore the inductive pathway involving the nude gene in the hair follicle, we first sought to define the exact temporal and spatial expression pattern of this gene in wild-type animals. We carried out an in situ hybridization analysis of Hfh11 RNA in skin taken from the dorsal midline of wild-type mice at birth and at frequent intervals through postnatal day 24 (P0 - P24) when the follicles undergo the first round of the hair cycle. During the period examined, the follicles progress synchronously through the three stages of the hair cycle, anagen (follicle generation and hair production), catagen (follicle regression), and telogen (resting phase).

Hfh11 is expressed during the hair shaft production phase of anagen, consistent with the phenotype observed when the gene is disrupted. Hfh11 is first detected at P2, is more strongly expressed at P3 and persists through P14 (Figure 28). In catagen (P18), no Hfh11 RNA is detected. Strong Hfh11 expression returns at P24 in newly forming hair follicles. With a 21 day hair cycle, this is equivalent to P3 for the new hairs. Hfh11 mRNA is expressed in
the cortical cells, beginning several cell layers above the apex of the dermal papillae. The *Hfh11* transcript continues to be detected as the hair shaft extends from the germinative hair bulb. No *Hfh11* mRNA was detected above background levels in other components of the dermis or epidermis (Figure 28).

We tested for the altered expression of cloned keratin genes in *nude* skin, as possible downstream targets of the *nude* protein. Two keratin genes, mHa1 and mHa3, have been cloned and shown by *in situ* hybridization to be expressed in the cortical cells of the hair shaft (Kaytes, et al., 1991; Winter, et al., 1994). To determine the temporal and to confirm the spatial expression patterns of these transcripts, we carried out independent *in situ* hybridization analyses with mHa1 and mHa3 RNA on the P0-P24 staged wild-type mouse skin samples. The expression of both mHa1 and mHa3 begin in cortical cells at P2 with stronger expression at P3 – completely consistent with the *Hfh11* RNA expression. We detected completely normal levels and pattern of expression of mHa1 and mHa3 in *nude* skin (Figure 29). mHa2 was also expressed normally in the cuticle of the wild-type and *nude* hair shaft, starting at P2 (Figure 29) (Winter, et al., 1994).

CONCLUSIONS

*Mutations in the nude gene*

Positional cloning identified a novel fork head transcription factor, that when mutated results in the *nude* phenotype. Disruptions in *Hfh11* have been detected for all four of the *nude* rodent alleles: a single-base-pair deletion in the *nu* allele, a marked decrease in expression levels of the *Hfh11* transcript in *nustr* skin, a nonsense mutation in the *rnu* allele, and a large genomic deletion removing several exons of the *rnuN* allele. The subtle nature of the genomic changes in all of the *nude* alleles except *rnuN* reinforces the value of having several distinct alleles of a gene in a positional cloning project.

*Partial rescue of the nude phenotype with the genomic Hfh11 locus*

Transgenic insertion of cosmid clone 1193, containing the genomic *Hfh11* coding region, into fertilized *nu/nu* eggs corrected the hairless, but not the athymic phenotype of the *nude* mice. The cosmid contains the entire
coding region of the mouse \textit{Hfh11} gene, together with 8.5 kb of 5' and 4 kb of 3' flanking regions, but does not contain the first reported, non-coding exon. Because the genetic evidence strongly suggests that a single gene is responsible for both the hairless and athymic phenotypes, the partial rescue seems likely to be due to differential regulation of the \textit{Hfh11} gene in the skin and thymus.

Two interesting questions are raised by these results: (i) why cosmid 1193 rescues the hair defect, despite lacking a reported initial exon; and (ii) why cosmid 1193 fails to rescue the thymus defect. These results indicate that the reported exon 1 of \textit{Hfh11} is not required for rescue of the hair phenotype. Although this short (63 bp) untranslated exon was found in a clone from a skin cDNA library, it is nonetheless possible that it represents a rare or aberrant alternative splice product and that the major skin promoter lies between exons 1 and 2 and is contained in cosmid 1193. It will be necessary to obtain the structure of the complete 4 kb \textit{Hfh11} cDNA to resolve this question. Alternatively, the transgenic mice may be expressing \textit{nude} from a promoter adjacent to the insertion site of the cosmid. This argument seems less likely since the same phenotype was observed in two different transgenic lines.

The failure of cosmid 1193 to rescue the thymus phenotype suggests that its insert lacks a critical transcriptional regulatory sequences, such as a promoter or enhancer elements, required for thymus expression. It is possible, for example, that the thymus promoter lies 5' to exon 1 and that exon 1 is usually found in the thymus specific transcripts. Alternatively, the integration sites might affect the transgene’s specific expression in the thymus. The influence of chromosomal position has been observed and used to account for the lack of and variation in the expression of several other genes introduced into mice (Hammer, et al., 1984; Krumlauf, et al., 1985; Lacy, et al., 1983). A third possibility is that a sequence critical for thymic expression might have been inadvertently damaged in the original YAC or in the cosmid.

expression patterns with mutant phenotypes and detecting mutations in the gene provide circumstantial evidence, but they are not sufficient to prove that a gene is responsible for the genetic defect. In rare cases, the wild-type protein can be administered to the mice to correct the mutant phenotype (Campfield, et al., 1995; Halaas, et al., 1995; Pelleymounter, et al., 1995). Otherwise, transgenic and knock-out experiments are essential for proving that a candidate gene is correct (Jones, et al., 1990; Koopman, et al., 1991; Readhead, et al., 1987; Wu, et al., 1994). Furthermore, introduction of genomic regions into the germline of mutant mice provides the opportunity to identify cis-acting DNA sequences and unravel the biochemical basis of their tissue specific expression (Hammer, et al., 1987; Krumlauf, et al., 1985). Our Hfh11 transgenic mice provide an excellent experimental system to begin to address the complicated but interesting problem of how the expression of the Hfh11 gene is differentially regulated in two distinct tissues, the skin and thymus.

Expression studies with the Hfh11; Searching for downstream targets

The expression pattern of Hfh11 is consistent with the phenotype observed when the gene is disrupted. Hfh11 is expressed in the adult thymus and in the developing embryo during thymic organogenesis. Hfh11 is expressed during the anagen phase of hair cycle in the cortex of the hair shaft, beginning several cell layers above the apex of the dermal papillae and continuing as the hair shaft extends from the germinative hair bulb. Since Flanagan postulated that the defect in the hair shafts was improper keratinization, we assayed the expression of cloned murine hair-specific keratins (mHa1, mHa3) in nude and wild type skins. The transcription of Hfh11 and the keratins commence at post-natal day 2, but we found no difference in the expression levels between normal and nude skin. As the other mouse hair keratin genes are cloned, this analysis will be continued. However, since all of the hair keratin proteins are already associated with spots on two-dimensional gel electrophoresis, a biochemical analysis could also identify if the protein levels of any of the hair keratins is affected in nude mice.

The hair follicle is a wonderful system to identify and characterize the genes involved in an inductive pathways because (i) the samples are abundant; (ii) temporal expression can be specifically determined since hairs cycle synchronously; (iii) spatial expression can be specifically determined
since the cellular parts of the hair can be distinguished histologically. Although the molecular nature of the signaling molecules involved in hair follicle initiation have not been identified, many intriguing genes have been shown to be expressed specifically in the hair bulb, including TGF-β, Bmp-2, Bmp-4, sonic hedgehog, E-cadherin, P-cadherin (personal observation and Dolle, et al., 1990; Hirai, et al., 1989; Jones, et al., 1991; Ruberte, et al., 1990). Furthermore, the induction of follicle initiation and growth may involve conserved pathways with less genetic redundancy than observed in other organ systems, as demonstrated by the hair-specific phenotypes of point mutations in the EGF-R and targeted disruptions in TGF-α and FGF-5, resulting in the waved-2, waved-1 and angora phenotypes, respectively (Hebert, et al., 1994; Luetteke, et al., 1994; Luetteke, et al., 1993; Mann, et al., 1993). Subtractive cloning between the cDNA of nude and wild type skin is a possible directed approach to identify downstream targets of the nude protein (Hubank, et al., 1994). Since the nu allele does not affect transcript levels, the skin samples for the subtraction could be selected just as Hfh11 turns on in both the nude and wild type animals.

Function of fork head transcription factors in development

FH transcription factors have diverse and important roles in development and differentiation, as demonstrated by expression patterns and genetic mutations in a variety of organisms. The targeted disruptions of HNF-3β and BF-1, as well as the spontaneously-arising nude mice correlate a strict phenotype with the loss of a FH transcription factor in mice. These three mutants also underscore the diversity of FH transcription factors’ function in development. The targeted disruption of HNF-3β leads to embryonic lethality by 11.5 dpc but mutant embryos are already morphologically distinguishable at gastrulation, 6.5 dpc (Ang and Rossant, 1994; Weinstein, et al., 1994). Mutant embryos lack a discernible node, notochord, and head process. BF-1-/- embryos die at birth and have a reduction in the proliferative rate in the telencephalic neuroepithelium, resulting to a smaller cerebral hemispheres with alterations in neuronal differentiation (Xuan, et al., 1995). nude mice are the first mouse mutants with disruptions in a FH gene that is consistent with viability. The specific hairless, athymic phenotype points to possible directed roles of FHs in development. To understand if the specificity of the FH transcription factors
comes from spatial and temporal regulation or from variant amino acids within the conserved DNA binding domain awaits further experimentation.

**Future directions**

Disruptions in the fork head transcription factor, \(Hfh11\), result in the hairless, athymic phenotypes of \(nude\) mice and rats. Much experimental work remains, however, to connect this cloned gene with the development of these organs in normal rodents. To understand the inductive pathways in thymic stroma and hair shaft development, the genes upstream and downstream of \(Hfh11\) must be identified. The hair follicle is a wonderful system to study genetic regulation in morphogenesis. The accessibility of the system suggests the potential to identify and characterize downstream genes of \(Hfh11\): mRNA samples are abundant to apply subtractive cloning techniques and the spatial and temporal expression patterns of candidate genes can be rapidly analyzed in both normal and \(nude\) skin by \textit{in situ} hybridization. To determine direct induction of transcription, the upstream regions of these candidate genes should be searched for conserved FH binding sites.

To understand the induction of the \(Hfh11\) gene, the upstream regulatory regions must be identified. The rescue of the hairless, but not the athymic phenotype by the genomic \(nude\) coding region contained in cosmid 1193, suggests that the \(nude\) gene is subject to complex regulation. To define the boundaries of the regulatory regions, it will be important to demonstrate a complete rescue of the thymic and hairless phenotype with a clone that contains the entire \(nude\) genomic region, perhaps cosmid 1204 or P11, P12, P13, P23. Recall that cosmid 1193 does not contain the published first noncoding exon of \(Hfh11\). In a similar, but probably more extreme case, Barsh and colleagues showed that the alternative isoforms of agouti mRNA contain different noncoding first exons located 100 kb apart, with independent regulatory elements that are ventral and hair cycle specific (Vrieling, et al., 1994). A combination of cDNA cloning and RNA expression studies should be used to determine if there are distinct regulatory regions for thymic and hair specific expression. The upstream regulatory regions of \(nude\) are probably nearby since the region is so gene-rich: In fact, fructose aldolase is less than 50 kb from the 5' end of the \(nude\) gene.
There are also some very interesting genomic questions raised by this project: (i) how many genes are in the 370 kb nude region or how many genes would be recognized by sequencing the region? (ii) what percent of the region is transcribed? (iii) if we sequenced this region both in the mouse and the human, would we find syntenic conservation of the gene level? (iv) could the comparison of the mouse and human sequence allow us to identify coding exons? regulatory regions?

Ultimately, we would like to understand how the thymic stroma develops and how T-cells home to this micro-environment. Likewise, hair and skin have the potential to serve as a non-redundant model of genetic regulation in morphogenesis.
REFERENCES


skin architecture, wavy hair, and curly whiskers and often develop corneal inflammation. *Cell.*, 73, 249-61.


MATERIALS AND METHODS

When the method was specifically designed for this project a complete
description is offered. When the method utilized was from other’s published
work only a brief description is offered.

Animals and cell lines
Congenic C57BL/6-nu mice were developed by repeatedly backcrossing the
original nude (nu) allele onto C57BL/6 mice; For the mRNA studies, the
animals were purchased from Taconic Farms (Germantown, NY) For the GD-
RDA experiments the DNA was purchased from The Jackson Laboratory (Bar
Harbor, Maine). The nustr mutation arose at The Jackson Laboratory on the
AKR/J inbred line (Eicher, 1976) and has been maintained on this
background; these animals were purchased from The Jackson Laboratory (Bar
Harbor, Maine). The rat nnu mutation arose on an outbred strain and has
been maintained by randomly mating nude males with heterozygous
females; these animals were purchased from Harlan Sprague Dawley
(Indianapolis, IN). C57BL/6J, CAST/Ei and MOLF/Ei animals, used in
mapping crosses, were a gift from the breeding facility of The Jackson
Laboratory. The mouse Chromosome 11; rat hybrid cell line and control rat
cell line were a gift from Dr. Christine Kozak (Killary, et al., 1984). The mouse
Chromosome 1; CHO hybrid cell line and CHO DNA were a gift from Dr.
Kent Hunter (Hunter, et al., 1991). For the RNA in situ hybridization, Swiss-
Webster mice were purchased from Simenson Laboratories, Gilroy, CA. To
obtain postnatal nude skin samples, a pregnant Swiss-Webster nu/+ female,
mated with a Swiss-Webster nu/nu male, was also purchased.

Cross progeny were phenotyped at post-natal day 11 for hair growth.
Presence or absence of a thymus was checked for all key recombinant animals.
Unaffected animals carrying chromosomes with key crossovers in the nude
region were progeny tested by mating them to nustr /+ animals. Informative
progeny (i.e., those determined, based on genotype at flanking markers, to
carry the recombinant chromosome over a nustr-bearing chromosome) were
examined for phenotype to determine whether the recombinant
chromosome carried the nustr allele.

Genotype Analysis
DNA from cross progeny was prepared from tail biopsies as described by Laird
et al. (1991). Simple sequence length polymorphism (SSLP) markers were
genotyped as described by Dietrich et al.(1992). Single strand conformational
polymorphism (SSCP) markers were amplified exactly as for SSLP markers,
diluted nine-fold with 95% formamide, 10 mM NaOH containing
bromphenol blue and xylene cyanol, denatured for 5 mins on a 100° heating
block, allowed to cool to room temperature, and electrophoresed on MDEE
gels (0.5X for PCR products >500bp; 0.7X for PCR products <500bp) (AT
Biochem, Inc, Malvern, PA) for 16 hrs at 10V/cm.
GD-RDA procedure

Representational Difference Analysis was performed essentially as published (Lisitsyn, et al., 1993). In this work, all amplicons were prepared by digesting 2 µg each of Tester and Driver DNAs with either BglII or BamHI. The iterative hybridization-amplification step was repeated three times. The resulting material was digested with the same restriction enzyme as used to prepare the amplion, ligated to BamHI-digested and dephosphorylated pBluescript II S-, and transformed into E.coli XL-Blue competent cells according to the supplier's recommendations. To maximize the success of RDA, it may be helpful to employ the following controls: (1) Ligation of PCR products with new adaptors on each round of RDA should be monitored by gel electrophoresis, which should show a detectable increase in fragment size distribution; (2) Concentration of Tester and Driver DNA at each step should be determined by gel electrophoresis, using Sau3A digested human DNA as a control; (3) Experiment 1 from Lisitsyn et al., 1993 can be performed in parallel with the main experiment, as a positive control.

For each experiment, six white colonies were picked at random and the inserts were immediately analyzed by PCR. The colonies were resuspended in 100 µl LB medium containing ampicillin (for subsequent growth and plasmid isolation) and a 5 µl aliquot was immediately transferred to 100 µl of a PCR reaction containing 1 µM each of Seq24 primer (5'-CGACGTTGTAAACGACGGCCAGT-3') and Rev25 primer (5'-CACACAGGAAACAGCTATGACCATG-3'), 67 mM Tris-HCl (pH 8.8 at 25°C), 4 mM MgCl2, 16 mM (NH4)2SO4, 10mM β-mercaptoethanol, 170 µg/ml bovine serum albumin, and 200 µM (each) of dATP, dGTP, dCTP and dTTP. The mixtures were incubated at 95°C for 5 min and cooled to 72°C, after which 5 U of AmpliTaq polymerase (Perkin Elmer Cetus) was added and the mixture was thermocycled for 30 cycles (95°C for 1 min, 72°C for 3 min) followed by a final incubation at 72°C for 10 min. The amplified plasmid inserts were analyzed by agarose gel electrophoresis to identify those having distinct sizes. These were purified on QiaGen-tip20 columns (Qiagen Inc), according to supplier's recommendations. To determine whether the clones represented sequences which were selectively present in the Tester but not Driver amplicons, selected inserts were radioactively labelled using Megaprime DNA labelling system (Amersham) according to supplier's recommendations, and hybridized to Southern blots containing DNA from Tester and Driver amplicons, which had been electrophoresed in a 2% agarose gel and transferred using a vacuum blotting apparatus to GeneScreen Plus membranes. Finally, clones were tested to determine whether they detected a unique genomic locus by hybridizing them to Southern blots of restriction-digested genomic DNA, with washing at moderate stringency (two 30 min washes in 0.5X SSC, 0.1% SDS at 65°C). Clones detecting a fragment present in Tester but not Driver amplicons were hybridized to Southern blots containing...
restiction-digested mouse genomic DNA to test whether they detected a RFLP between Tester and Driver. Clones detecting RFLPs were subsequently genetically mapped in the mouse genome, by hybridizing them to Southern blots containing restriction-digested DNA from progeny of various two-generation mouse crosses. The inheritance pattern of the RFLPs was compared to that of various simple sequence length polymorphisms (SSLPs) that mapped to the regions of interest.

Isolation of YACs, P1s and cosmids
MIT YACs were obtained by PCR-based screening of the the library (Kusumi, et al., 1993), which had been prepared with a two-level pooling scheme based on screening of "super-pools" and "sub-pools", yielding a unique address (Green, et al., 1990). Princeton YACs were obtained by PCR-screening of the YAC DNA pools provided by the Princeton University Mouse YAC library resource (Rossi, et al., 1994). A stab of the frozen YAC clone was plated onto AHC (-ura, -trp) selective plates and individual colonies were picked and analyzed by PCR to confirm that they contained the desired STS. Bacteriophage P1 clones were obtained from commercially available libraries made from either mouse cell line RIII (P1-P2 and P7-P23) or mouse ES cell line from strain 129 (P24,P25) (Genome Systems Inc., St. Louis, MO). A stab of the frozen P1 clone was plated onto LB-KAN (50 μg/ml) selective plates and individual colonies were picked and analyzed by PCR to confirm that they contained the desired STS. Cosmids were constructed from high molecular weight YAC DNA (prepared in agarose), partially digested with MboI using restriction-minus packaging extracts and hosts (Stratagene, La Jolla, CA). The library was plated at 5000 colonies/plate, lifted onto nylon membranes and hybridized with cloned mouse repetitive elements (SINE and LINE) to identify cosmids containing mouse genomic inserts. Total yeast DNA for PCR was prepared from YAC clones as described by Treco (1991). High molecular weight YAC DNA for PFG analysis was prepared as described by Gemmill (1994). P1 and cosmid DNAs were prepared according to the supplier's recommendation as single copy plasmids (Genome Systems Inc, St. Louis, MO; Stratagene, La Jolla, CA). To prepare the cosmid DNA for microinjection, the DNA was purified with the gene-clean kit (BIO 101) according to supplier's recommendation, and resuspended in 10mM Tris and 0.1mM EDTA (pH 7.5) to a concentration of 50 copies per picoliter.

Cloning of YAC, P1, and cosmid ends
We improved upon the published protocols to clone YAC ends by inverse PCR. To isolate the insert DNA adjacent to the centromeric arm ("left end"), 500 ng of YAC DNA was digested independently with 10 units of HaeIII, Sau3a I, and TaqI. To isolate the insert DNA adjacent to the noncentromeric arm ("right end"), 500 ng of YAC DNA was digested independently with 10 units of HaeIII, AluI, or HhaI. 50 ng of the digestion products were then ligated under dilute conditions to favor monomeric circularization. PCR reactions were performed with 5 ng of ligation product using InvL-2 and Left
1 as primers for the left end and Right 2 and Right 7 as primers for the right end. Thirty cycles of 30 s at 94°C, 1 min at 55°C, 1 min at 72°C, and finally 6 mins at 72°C were performed. Two microliters of the amplified DNA was reamplified with chimeric primers to introduce M13 sites. Primers for left end amplifications were M13ForL1 and M13RevL2. Primers for the right end, digested with HaeIII or HhaI, were M13ForR7 and M13RevR2. Primers for the right end digested with AluI were M13ForAluR and M13RevR2. All of the primer sequences used for end rescue are given in Table VI. The products from this amplification were resolved on a low-melting-point agarose gel. Excised bands were melted at 65°C, digested with agarase (purchased from New England Biolabs, Beverly, MA) at 37°C for 1 hr, and sequenced following the Applied Biosystems -21M13/M13REV dye-primer cycle-sequencing protocol using an ABI373 machine (Applied Biosystems, Foster City, CA). In all cases, sequence was obtained from the vector/insert junction to ensure that the fragment represented the end of the insert.

To clone the ends of the P1s and cosmids we selected new enzymes and primers from the vector sequence. To clone the insert DNA adjacent to the SalI site in the P1 vector, 50 ng P1 insert DNA was digested with HinfI, Rsal or AluI. To clone the insert DNA adjacent to the NotI site in the P1 vector, digestion was performed with HhaI or AluI. Digestion products were religated to promote circularization. PCR was performed on 0.5 ng of ligation product using PN_-30For and PN_-40Rev for the NotI end and PS_64For and PS_18Rev for the SalI end. The resulting PCR products were then reamplified to introduce M13 sequencing primer sites. The chimeric primers M13FOR-20 and M13REV-50 were used for the NotI end; M13FOR79 and M13REV18 to reamplify the SalI end digested with HinfI and Rsal; and M13FOR312 and M13REV18 to reamplify the SalI end digested with AluI.

To clone the insert DNA adjacent to the T3 site in the cosmid vector, 50 ng cosmid insert DNA was digested with DdeII or HhaI. To clone the insert DNA adjacent to the T7 site in the cosmid vector, digestion was performed with Hpal or Msel. Digestion products were religated to promote circularization. PCR was performed on 0.5 ng of ligation product using CT3_7931For and CT3_7921Rev for the T3 end and CT7_147For and CT7_94Rev for the T7 end. The resulting PCR products were then reamplified to introduce M13 sequencing primer sites. The chimeric primers CT3_M13For3 and CT3_M13Rev7901 were used for the T3 end; CT7_M13For170 and CT7_M13Rev77 were used for the T7 end. The name of the primer indicates the first base pair of the primer in the published vector sequence. All of the primer sequences used for end rescue are given in Table VI. PCR primer pairs were then picked to amplify the genomic sequence at the end of the insert, using the PRIMER computer program (S. Lincoln, M. Daly and E.S. Lander).

Characterization of YACs, P1s, and cosmids
The sizes of the P1s and YACs were determined by pulse-field gel (PFG) electrophoresis (Chu, et al., 1986), followed by transfer to nylon membrane
and hybridization of the blot to pBR322 DNA labeled with $^{32}$P by the random priming method (Feinberg, et al., 1983). To determine the restriction map of the region, P1 DNA or high molecular weight YAC DNA (in the agarose plugs) was digested according to manufacturer's recommendations (New England Biolabs, Beverly, MA) and fractionated by PFG electrophoresis. After Southern transfer the blots were hybridized with gene fragments labeled with $^{32}$P specifically primed with an oligonucleotide from the vector sequence flanking the insert (5'-CTGAGCGAATTCGTGAGACC-3' for direct cDNA selected clones and 5'-CTCGAGGTCGACCCAGCA-3' for exon trapped clones). To determine the STSs content of each cosmid, 100 ng of cosmid DNA was independently denatured with NaOH, transferred to nylon membrane and hybridized with direct cDNA selected or exon trapped clones, labeled with $^{32}$P as above.

**Direct cDNA selection and Exon Trapping**

Direct cDNA selection was performed on P1 and cosmid clones according to the protocol of (Lovett, 1994) with several modifications. Biotin was incorporated into the digested genomic DNA by ligating biotin-containing linkers (ligation of the two oligonucleotides: BIO-Blunt-1 (5'-BIOTIN-GCGGTGACCCGGGAGATCTGAATTC-3') and Blunt-2 (5'-GAATTCAGATC-3')) These biotinylated primers were also used to amplify the genomic DNA. Primary cDNA for the selection was independently prepared by random-priming poly A+ selected mRNA, from BALB/cJ post-natal day 0.5 skin, adult C57BL/6J skin, adult C57BL/6J testes, and adult C57BL/6J thymus. Streptavidin coated magnetic beads were pre-blocked with 0.1% BSA and 0.2 µg/mL mouse COT-1 DNA (GIBCO BRL, Gaithersburg, MD). After two rounds of hybridization, selected cDNA fragments were cloned by using the uracil DNA glycosylase cloning system (GIBCO BRL, Gaithersburg, MD). Exon trapping was performed with the SPL3 plasmid according to manufacturer's conditions on pools of 8 cosmids, digested with BamHI and BglII (GIBCO BRL, Gaithersburg, MD). RT-PCR was performed on first strand cDNA made from DNAsed total mRNA, according to manufacturer's instructions (GIBCO BRL, Gaithersburg, MD).

**Northern analysis**

mRNA was prepared from shaved or nude skin sections with TRizol according to manufacturer's recommendations (Gibco BRL Life Technologies, Gaithersburg, MD). Poly A+ selected mRNA was isolated on Oligo (dT) cellulose columns (Gibco BRL Life Technologies, Gaithersburg, MD). The RNA was denatured with glyoxal and dimethyl sulfoxide, transferred to nylon membrane, hybridized, and washed under stringent conditions as described in Section 7.40 - 7.50 of Sambrook et al. (1989).

**Sequencing of the mouse and rat nude genes**

Primary cDNA was independently prepared by specific priming with an oligonucleotide from the 3' untranslated region of the mouse nude gene (5'-
GGGAGAGGGCCAAGTCTGT-3') poly A+ selected mRNA from C57BL/6J, C57BL6J-\textit{nu}/\textit{nu}, AKR/J, AKR/J-\textit{nu}stress/\textit{nu}stress, rat, and rat-\textit{rnu}/\textit{rnu} adult skin. Twelve overlapping fragments were amplified by PCR from the primary cDNA and both strands were sequenced according to manufacturer’s instructions (Applied Biosystems, Foster City, CA).

\textit{Production of Transgenic Mice}

Microinjection of oocytes was performed as described by Hogan et al (1986). Homozygous CD1-\textit{nu}/\textit{nu} males were mated with homozygous CD1-\textit{nu}/\textit{nu} females (both purchased from Charles River Laboratories). Fertilized eggs were isolated from oviducts, freed from follicle cells by hyaluronidase treatment. About 2 pl of the DNA solution (5ng/ul) was microinjected into the male pronucleus. After injection, eggs were incubated overnight until they reached the two-cell stage. Two-cell embryos were then transferred to the distal oviducts of pseudopregnant ICR mice.

\textit{RNA in situ Hybridization Analysis and Histological Evaluation}

RNA \textit{in situ} hybridization analyses were performed using a digoxigenin-labeled antisense RNA probe, essentially as described by Herbert et al. (1994). Briefly, dorsal skin samples were removed from the mice and fixed in 4% paraformaldehyde and frozen in OCT compound (Miles Corporation, Elkhart, IN). Sections (8 µM) were collected on SuperFrost Plus slides (Fisher Scientific, Pittsburgh) and then treated with 10 µg/ml of proteinase K for 10 mins at room temperature. The sections were then hybridized at 65° to an RNA probe prepared with digoxigenin-labeled UTP(purchased from Boehringer Mannheim, Indianapolis, IN) according to manufacturer’s conditions (Ambion Inc., Austin, TX). The slides were treated with an anti-digoxigenin antibody coupled to alkaline phosphatase and were stained for alkaline phosphatase activity. To detect the nude transcript, the PCR product of the DS-a8d was used as a template for in vitro transcription. To detect the keratin transcripts, partial cDNAs, amplified from a mouse skin library (Mouse Skin cDNA Library in the Uni-ZAP XR vector fromStratagene, La Jolla, CA) with a specific primer picked from the published sequence and the -21M13 primer, were used as a template. The keratin specific primers were: mHa1 5'-ccc tcc tct gta atc tcc caa taa -3'; mHa2 5'-gga gct taa caa gca ggt ggc -3'; mHa3 aca agc cca ttg gag cct gtt -3' (Kaytes, et al., 1991; Winter, et al., 1994). Skins were removed and fixed in 10% formaldehyde and prepared for routine light microscopy according to standard techniques. Histological evaluation was carried out on sections stained with hematoxylin and eosin.

\textit{DNA Analysis of the transgenic mice}

Genomic DNA from the transgenic mice was prepared from approximately 1.5cm of tail biopsy samples at 2 weeks of age. The presence of the transgene was examined by Southern blot analysis using 10ug of genomic DNA digested with HindIII. The digests were separated on 0.75% agarose gels and transferred
onto nylon membrane (Hybond-N;Amersham). The probe template DNAs were the neomycin resistance gene specific for the cosmid DNA and the 3' half of the Hfh11 cDNA. They were amplified by PCR and radiolabeled with 32P by random primer extension. Hybridization was performed as described (18) and autoradiograms were taken on an imaging plate for analysis with a Bio-Image Analyzer (BAS2000;Fuji Film). Primers to confirm correct integration of cosmid 1193 are as follows: T3 side of cosmid vector: 5'-ATAGGCGTATCACGAGGCC -3' (bp 7908-7925); T3 side of insert DNA: 5'-TCGCATACGGTGATAAGAGATG -3'; T7 side of cosmid vector: 5'-TGATAAGCGGTCAAACATGA -3' (bp 90-71); T7 side of insert DNA: 5'-ATCTCTCCCCCTAACCCCTGGG -3. To construct a restriction map of the cosmids relative to the Hfh11 gene, the cosmid genomic DNA was digested to completion singly and in pairs with BamHI, EcoRI, EcoRV, SfiI and NotI. The digests were separated on 0.7% agarose gels and transferred onto nylon membrane and hybridized with end-labeled oligonucleotides: 5'-GATCACAACCATCTGTAATGGG- 3' from the T7 side of the insert of 1204; 5'-TTCTCACCTGCTCTAGGGA- 3' from the T7 side of the insert of 1193; 5'-CGCCGACCTGCTCTC - 3' from exon 1 of Hfh11; 5'-CACTTCCAGGCTCCCACCC- 3' from exon 2 of Hfh11; 5'-ACTGTTCTTCTCAGGCCCTGCC- 3' from exon 4 of Hfh11; and 5'-GCTCGAGAGCTGAAGTTCG- 3' from exon 9 of Hfh11. Long-range PCR to determine the size of the introns between coding exons was carried out according to manufacturer’s instructions (Boehringer Mannheim).

Flow cytometric analysis

Single-cell suspensions from spleen and peripheral blood were prepared free from red blood cells following standard procedures. Samples of 5x10^5 cells were treated with 10ul of normal mouse serum for blocking the nonspecific binding of antibodies. After incubation, a pair of phycoerythrin (PE)-conjugated anti-CD4 / fluorescein isothiocyanate (FITC)-conjugated anti-CD8, or of biotin-conjugated anti-Thy1.2 / FITC-conjugated anti-B220 antibodies, diluted in PBS to the appropriate concentration, were added directly. As a secondary antibody against biotin-conjugated anti-Thy1.2, PE-conjugated streptavidin was used. Cells were suspended with PBS containing 2% FCS and 0.05% sodium azide (staining buffer), and analyzed by flow cytometry on a FACScan (Becton Dickinson) equipped with logarithmic scales, and data were processed in a LYSISII soft ware.
REFERENCES


Figure 1. nude mice with wild-type littermates. From the left: C57BL/6J-nu, C57BL/6J, AKR-nustr, AKR/J
Hair Growth Cycle

Figure 2. A schematic diagram illustrating the morphology of the hair follicle at the three stages of hair growth cycle: anagen: the stage during which the follicle is regenerated and the hair matrix cells produce a new hair; catagen, the stage during which matrix cell proliferation and hair elongation ceases and many of the structures of the anagen follicle are eliminated; and telogen, the resting phase. Diagram from Herbert et al., 1994.
Figure 3. Schematic representation of a hair follicle. (a) The relative arrangement of cells derived from the epidermis (sheath, cortex, medulla), the dermis (papilla and connective tissue) and the neural crest (melanocyte). Diagram from Hogan et al., 1986. (b) Major tissue layers and cell types of the hair follicle are indicated by numbers at the top margin: 1, outer root sheath; 2, inner root sheath; 3, cuticle; 4, cortex; 5, medulla; P, dermal papilla. Diagram from Heid et al., 1988.
Figure 4. Electron micrograph of a longitudinal section through the main layers of a nonpigmented mouse hair follicle. At the upper left is the medulla, then the cortex and cuticle of the hair shaft, then the three layers of the inner root sheath and at the bottom is part of the outer root sheath. (Magnification x 6000) from Rogers and Powell, “Hair Follicle Keratins” in Handbook of Mouse Mutants with Skin and Hair Abnormalities, ed. Sunerberg, J.P. p. 106
Figure 5. Thymic development from 3rd Pharyngeal arch and pouch. Left side is a representation of a coronal section at embryonic 11 d.p.c. The thymus is beginning to form from the posterior half of the third pharyngeal pouch. The right side is a ventral view of the pharyngeal organ derivatives. The two thymic primordia migrate medially and posteriorly, where the two anlages fuse to become a single gland. Diagram from Manley and Capecchi, 1995.
Figure 6. The genotype of (AKR-nu<sup>str</sup> x MOLF/Ei)<sub>F<sub>2</sub></sub> progeny M566 to M586 with simple sequence length polymorphism markers D11Mit65 and D11Mit36. The AKR-nu<sup>str</sup> and MOLF/Ei sizes are 382 and 370 for D11Mit65 and 234 and 274 for D11Mit36, respectively. The nude animals are indicated. Note that there are cross-overs between D11Mit65 and D11Mit36 in animals M558 and M564.
Figure 7. Agarose gel electrophoresis of difference-products obtained after the first (lanes a,b), second (lanes c,d) and third (lanes e,f) hybridization-extension-amplification steps as pairs for the nude congenic and the nude cross (first iteration). Lane g is the size standard of φX174 RF DNA digested with HaeIII.
Fig. 8 Schematic diagram representing the principle underlying GORDOA with progeny from an F$_1$ intercross. Each panel shows hypothetical chromosomal genotypes from 10 progeny to be pooled to create a Driver; each chromosome is arbitrarily drawn to be 100 cM. Strain A carries a recessively-acting allele at locus L and is shown in white; strain B is shown in black. Graphs show percentage of B alleles present in Driver at each location along the chromosome. a, A chromosome unlinked to L (the percentage of B alleles remains close to 50%). b, The chromosome containing L, with progeny having the recessive phenotype selected at random (the percentage of B alleles drops slowly to 0% at L). c, The chromosome containing L, with progeny having the recessive phenotype selected to be recombinant between L and one of two flanking genetic markers, X or Y (the percentage of B alleles drops sharply to 0% in the X-Y interval).
Figure 9. Schematic diagram indicating chromosomal genotypes of the 12 (MOLF/Ei x AKR/J-nu<sup>str</sup>) F<sub>2</sub> progeny pooled to create the Driver for GD-RDA, relative to a genetic map of polymorphic markers near the nude locus. Black indicates regions derived from MOLF/Ei. Shading indicates regions derived from AKR. The number of progeny of each type is indicated at the right.
Figure 10a. Southern blot of DNA from F2 intercross progeny, digested with BglII and hybridized with GD-RDA 6.2. The animals are listed above the lane. All the nude animals, indicated above, are homozygous for the AKR-\textit{nu}^{str} allele of 4 kb. The size standard is \lambda digested with HindIII.

Figure 10b. Southern blot of DNA from F2 intercross progeny, digested with BamHI and hybridized with GD-RDA 10.4. The animals are listed above the lane. All the nude animals, indicated above, are homozygous for the AKR-\textit{nu}^{str} allele of 3 kb. The size standard is \lambda digested with HindIII.
Figure 11a. The genotype of nude intercross progeny with the single strand conformational polymorphism BS11.8. The nude animal is indicated.

Figure 11b. The genotype of nude intercross progeny with the single strand conformational polymorphism Y32L. All of the animals are unaffected.
Figure 12. YAC chromosomal walk proximal to nude. Contig is centered around D11Mit7 and RDA6.2. The direction to the centromere is indicated. The orientation of the contig is suggested by the genetic markers. Underlined names indicate terminal fragments that contain a sequence polymorphism, assayed as a size or conformational difference. Closed circles indicate the original source of the STS. Open circles indicate the presence of the marker in other clones. ☐ indicates that the orientation of the YAC was not determined.
Figure 13. YAC chromosomal walk surrounding nude. Contig is centered around genetic markers: both MIT SSLPs and GD-RDA clones. The orientation of the contig is suggested by the genetic markers. The direction to the centromere is indicated. An underlined name indicates a terminal fragment that contains a sequence polymorphism, assayed as a size or conformational difference. Closed circles indicate the original source of the STS. Open circles indicate the presence of the marker in other clones. xx indicates that the cloned end is chimeric. indicates that the orientation of the YAC was not determined.
Figure 14. Fine structure physical mapping of the nude region. The direction to the centromere is indicated. The orientation of the contig is suggested by the genetic markers. (A) Genetic markers and anchoring loci. An underlined name indicates a locus that contain a sequence polymorphism, assayed as a size or conformational difference. Prefixes for the loci: BS is an STS cloned from YAC 31; CA is a random SSLP cloned from YAC 31; DS is a direct cDNA selected clone; ET is an exon trapped clone. (B) Recombinational breakpoints for the closest recombinant animals in nude crosses. The empty bar denotes the chromosomal region that recombines with the nu gene. The black bar denotes the chromosomal region that does not recombine with the nu gene. The hatched bar denotes the chromosomal region with an undetermined haplotype. (C) The smallest region that does not recombine with the nude locus. (D) Pulsed-field gel-based restriction map of the region. S, Sad1; F, Sfi1; M, Mlu1; N, Not1; R, Rsrl. (E) STS-content mapping of P1 clones (denoted P) and cosmids (denoted C). Closed circles indicate the original source of the STS. Open circles indicate the presence of the marker in other clones.
A. DS-Y8A vs. vit

cY8A : 1 GGGGCATTGAGGGCCCCATCGATGCTGCCTTCACTCAACTGTCAGGGGNAGACCT 60

Vtn: 596 GGGGCATTGAGGGCCCCATCGATGCTGCCTTCACTCAACTGTCAGGGGAAGACCT 655

cY8A : 61 ACTTGTTCAAGGGTAGTCAGTACTGGCGCTTTGAGGATGGGGTCGGACCCTGGTTATC 120

Vtn: 656 ACTTGTTCAAGGGTAGTCAGTACTGGCGCTTTGAGGATGGGGTCGGACCCTGGTTATC 715

cY8A : 121 CCCGAAACATCTCGGAAGGCTTCAGTGGCATACCAGACAATGTTGATGCAGCGTTCG 177

Vtn: 716 CCCGAAACATCTCGGAAGGCTTCAGTGGCATACCAGACAATGTTGATGCAGCGTTCG 772

B. DS-K12G vs. emb-5

cK12G: 4 EAVLEGARYMVALQIARVPLVRQVLRQTFQERAKLNITPTKKGRDVDEAHYAYSPKYLK 183

+ VL GA++M+A +I+R P VR +SQ F++ A I PTKKGR +D+ H Y +Y+K

emb-5: 552 DMVLNGAKMLAEISQPQVRHSVRFQSAHFWIINKPDTIDQTHPLYDKRYLK 611

cK12G: 184 NKPVKELRDDQFLKIGLAEDEGLLITI 261

+KFV+ L ++FL A+++GL+ +

emb-5: 612 SKPVRSILTAEFFLYHKAKEDGLVW 637

Figure 15. Examples of strong sequence similarities. (A) Nucleotide sequence alignment of a direct cDNA selected clone, DS-Y8A, with mouse vitronectin mRNA, produced by the BLASTN program. The sequences have perfect nucleotide identity over 177 nucleotides with the exception of a single undetermined nucleotide in DS-Y8A. Numbering corresponds to nucleotide position for both the mouse vitronectin mRNA (Vtn) and DS-Y8A. (B) Amino acid alignment of direct cDNA selected clone, DS-K12G, with the C. elegans protein emb-5 produced by the BLASTX program. The two clones share 41% amino acid identity over 86 residues. Amino acids listed between the lines are identical between the two sequences; + denotes conservative amino acid substitution; blank denotes non-conservative substitution. Numbering is in amino acid residues for emb-5 and in nucleotides for DS-K12G.
Figure 16. Transcription map of the nude region. (A) Recombinational breakpoints for the closest recombinant animals in *nu* crosses. The empty bar denotes the chromosomal region that recombines with the *nu* gene. The black bar denotes the chromosomal region that does not recombine with the *nu* gene. The hatched bar denotes the chromosomal region with an undetermined haplotype. (B) The smallest region that does not recombine with the *nu* locus. (C) The location of transcription units with strong sequence similarity to genes in Genbank.
Figure 17. Expression studies by Reverse-transcribed PCR. Screening for genomic deletions in mutant mice by PCR. a. ET-90 expression. b. Profilin expression with primers that flank an intron on samples treated with DNAse and untreated.
Figure 18. Southern blot of restriction digested genomic DNA hybridized with DS-a10g.
Figure 19a. mouse wild-type nude sequence (-21M13 primer)
Figure 19b. mouse wild-type nude sequence (M13Rev primer)
Figure 19c. C57BL/6J-nu sequence (-21M13 primer). Deletion of single G at bp 19.
Figure 19d. C57BL/6J-nu sequence (M13Rev primer). Deletion of single C at bp 228.
Figure 19e. rat wild-type nude sequence (-21M13 primer)
Figure 19f. rat mu/+ sequence (M13Rev primer)
Figure 19g. rat mu sequence (-21M13 primer). C to T change at bp 91.
Figure 19h. rat mu sequence (M13Rev primer). G to A change at bp 204.

Figure 19. Sequencing wild-type and mutant alleles to detect mutations.
Figure 20. Northern blot analysis of nude gene expression in AKR/J and AKR/J-nuStr skin.
Figure 21. cDNA sequence of the rat *nude* gene. Amino acids differing from those in the mouse transcript appear in bold. The site of the single base-pair change from C to T at nucleotide 1429, resulting in the nonsense mutation of the *rnu* allele is enlarged and underlined.
Figure 22. Restriction Map of cosmids 1193 and 1204. B, BamHI; E, EcoRI; F, Sfil; N, NotI; V, EcoRV. Total insert size of 1193 is 36 kb; 1204 is 42 kb. Cosmid 1193 has a chimeric insert: ___ is mouse DNA. ___ is yeast DNA.
Figure 23. Southern blot analysis of DNA from Hfh11 transgenic mice. Tail DNA samples from Hfh11 transgenic mice (lane 1,5) and their non-transgenic littermates (lane 2-4,6-8) were digested with HindIII. Membranes were hybridized with a cosmid vector probe (lane 1-4) and a probe for the carboxy terminal of a Hfh11 cDNA (lane 5-8). Expected transgene bands are indicated by arrows.
Figure 25. Histological sections of the skin of transgenic mice, E1 and G2. The tissue sections were stained with hematoxylin and eosin. For transgenic E1, both haired and non-haired regions are shown.
Figure 26. Phenotype of control wild-type and Hfh11 transgenic mice. Thymus, indicated by the arrow in the wild-type, is absent in transgenic mice.
Figure 27. Analysis of the T cell populations in Hfhi11 transgenic mice. Two-color flow cytometry of splenocytes and peripheral blood cells from 1.5-month-old-control wild type (wt), Hfhi11 transgenic (E1), and control nude (nu/nu) mice. Single cell suspensions were stained with antibodies specific to CD4, CD8, Thy-1, B220 and analyzed with FACScan. The number in each quadrant indicates the calculated percentage of cells.
Figure 28. Hfh11 expression in the cortex of the hair follicle at post natal day 4. (a) x 100. (b) x200
Figure 29. Expression of keratin genes in skin of *nude* mouse at postnatal day 4. Magnification is x200. (a) Ha1 (b) Ha2 (c) Ha3.
| ANIMAL | D11Nds1 | D11Mit7 | Y8L | RDA6.2 | RDA10.4 | CA128 | Phenotype | Genotype | CA4 | CA8 | RDA10.2 | BS7.10 | D11Mit117 | D11Mit118 | D11Mit96 | D11Mit91 | D11Mit94 | D11Mit34 | D11Mit32 | D11Mit8 |
|-------|---------|---------|-----|--------|---------|-------|-----------|----------|-----|-----|---------|--------|------------|----------|----------|------------|----------|----------|----------|
| B11   | A       | A       |     |        |         |       | wt        | H        |     |     |         |        |            |          |          |             |          |          |          |
| B13   | B       | H       |     |        |         |       | wt        | H        |     |     |         |        |            |          |          |             |          |          |          |
| B25   | B       | H       |     |        |         |       | wt        | H        |     |     |         |        |            |          |          |             |          |          |          |
| B63   | H       | H       |     |        |         |       | ?         | B        |     |     |         |        |            |          |          |             |          |          |          |
| B86   | B       | H       |     |        |         |       | wt        | H        |     |     |         |        |            |          |          |             |          |          |          |
| B119  | A       | A       |     |        |         |       | nude      | A        |     |     |         |        |            |          |          |             |          |          |          |
| B140  | H       | H       |     |        |         |       |           | A        |     |     |         |        |            |          |          |             |          |          |          |
| B147  | H       | H       |     |        |         |       | nude      | A        |     |     |         |        |            |          |          |             |          |          |          |
| B171  | A       | A       |     |        |         |       | nude      | A        |     |     |         |        |            |          |          |             |          |          |          |
| B173  | A       | A       |     |        |         |       |           | ?        |     |     |         |        |            |          |          |             |          |          |          |
| B180  | A       | A       |     |        |         |       | nude      | A        |     |     |         |        |            |          |          |             |          |          |          |

Table I. The phenotype and genotype with the SSLP and the RDA genetic markers of the animals with recombination breakpoints in the *nude* region. An (AKR-*nu*str X C57BL/6J) F2 animals is abbreviated as B#; an (AKR-*nu*str X CAST/Ei) F2 animals is abbreviated as C#; and an (AKR-*nu*str X MOLF/Ei) F2 animals is abbreviated as M#. For the genotypes: A is an AKR-*nu*str homozygote; B is a C57BL/6J homozygote; C is a CAST/Ei homozygote; H is a heterozygote; M is a MOLF/Ei homozygote; and ? is an undetermined genotype.
<table>
<thead>
<tr>
<th>ANIMAL</th>
<th>D11Nds1</th>
<th>D11Mit7</th>
<th>Y8L</th>
<th>RDA6.2</th>
<th>RDA10.4</th>
<th>CA69.4</th>
<th>CA128.4</th>
<th>Phenotype</th>
<th>Genotype</th>
<th>CA4</th>
<th>CA8</th>
<th>RDA10.2</th>
<th>BS7.10</th>
<th>D11Mit117</th>
<th>D11Mit144</th>
<th>D11Mit98</th>
<th>D11Mit94</th>
<th>D11Mit34</th>
<th>D11Mit32</th>
<th>D11Mit8</th>
</tr>
</thead>
<tbody>
<tr>
<td>M86</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
<td>MM</td>
</tr>
<tr>
<td>M100</td>
<td>H</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>M</td>
<td>A</td>
<td>AA</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M143</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M152</td>
<td>H</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M184</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M210</td>
<td>H</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M225</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M230</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M242</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M273</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M284</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M294</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M295</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M310</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M402</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M418</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M433</td>
<td>H</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M455</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M465</td>
<td>H</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M486</td>
<td>A</td>
<td>A</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M491</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M496</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M558</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M564</td>
<td>H</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M579</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M603</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M627</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M722</td>
<td>M</td>
<td>M</td>
<td>?</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M723</td>
<td>A</td>
<td>A</td>
<td>?</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M952</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Assay name</td>
<td>Forward Primer 5'-3'</td>
<td>Reverse Primer 5'-3'</td>
<td>Amplicon Size (bp)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SSLP markers:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y8L</td>
<td>ATGCCGTGCTGTAGCGTGTAG</td>
<td>GCCCAGTGAGCTAAAGAATATC</td>
<td>AKR/J 140 132 144</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS9.10</td>
<td>CTTGAGAGATGGTAAAGGTG</td>
<td>ACTGNAATTAATGAACCTACCTACCCA</td>
<td>C57BL6/J 300 304</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA4</td>
<td>AACTCCCCATCCAGGGAGTCT</td>
<td>TGGAGAGATAATGGGCGAG</td>
<td>CAST-Ei 210 196 214</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA8</td>
<td>CAGCAGCAAAACGGGAAGA</td>
<td>TCACTTGCTAGCAATCAGC</td>
<td>MOLF-Ei 180 190 186</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA69</td>
<td>CTCCACGGACATTCGCCC</td>
<td>TCAACTGCTTTAAGCATCATCTGA</td>
<td>160 140 130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA128</td>
<td>GATCCCAGGAGATTGGCC</td>
<td>TCTGAGAGAAACCAGATATGACA</td>
<td>125 127 135</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SSCP markers:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS11.4</td>
<td>TCACCAAGACAAGCAAGCAAAGG</td>
<td>TGGCTTCCCTTTTCTCC</td>
<td>161</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS11.5</td>
<td>GAGACAATCTCATCCAGATGCCAG</td>
<td>CGTAGTTGACTCCATCC</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS11.7</td>
<td>AGGGAGGAGCTCATCCAC</td>
<td>CCTAAGGATGAGTGAGTG</td>
<td>176</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS11.8</td>
<td>CTGCTGGAGAACAGGCC</td>
<td>CAACATGAGCTACACAATGACC</td>
<td>223</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS12.42</td>
<td>TTACCCGAGGCTTCATG</td>
<td>TCAGAGGTGTGGCTAAG</td>
<td>222</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table II. Primers to amplify polymorphisms in the nu/nu region. The allele sizes are given for the SSLP markers in the relevant strains.
<table>
<thead>
<tr>
<th>Assay name</th>
<th>Forward Primer 5'-3'</th>
<th>Reverse Primer 5'-3'</th>
<th>Amplicon Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y4R</td>
<td>tcgtaagccccctcatgttgc</td>
<td>agacggtatcacgtcctcatgc</td>
<td>100</td>
</tr>
<tr>
<td>Y8L</td>
<td>atgcgctctgtgtagcttga</td>
<td>gcacgtgagcttaagacatct</td>
<td>140</td>
</tr>
<tr>
<td>Y8R</td>
<td>ttcgagagactggaaacttc</td>
<td>cctggccaaatctctatgaacc</td>
<td>600</td>
</tr>
<tr>
<td>Y9L</td>
<td>ttgaagctctggagcttta</td>
<td>aagccagctctgtgtcatca</td>
<td>120</td>
</tr>
<tr>
<td>Y10L</td>
<td>ccttttggtgaaatttcaaccttcc</td>
<td>atcaagaagccttaactgtagt</td>
<td>122</td>
</tr>
<tr>
<td>Y11L</td>
<td>aagtggtcaacctgaacttca</td>
<td>tccattggacaaataacctgaga</td>
<td>258</td>
</tr>
<tr>
<td>Y14L</td>
<td>agtaagggagctttttaaaagaaacc</td>
<td>gccctctctctaggtggttc</td>
<td>180</td>
</tr>
<tr>
<td>Y14R</td>
<td>gctttgcccacatctgcc</td>
<td>ttgattacacgttagttacctaggg</td>
<td>104</td>
</tr>
<tr>
<td>Y15R</td>
<td>gatgtttagattcccatggagc</td>
<td>caaagatttggccctactaacc</td>
<td>161</td>
</tr>
<tr>
<td>Y19L</td>
<td>tgtattttgccacattgagaatg</td>
<td>tttaaattggaagctgcca</td>
<td>156</td>
</tr>
<tr>
<td>Y19R</td>
<td>ccacattgtcataaaatccca</td>
<td>ttgacagctgtctttgctgagc</td>
<td>293</td>
</tr>
<tr>
<td>Y20L</td>
<td>cagatggtataatcccaacctcaactg</td>
<td>agaagacgctactgtagttaggtc</td>
<td>183</td>
</tr>
<tr>
<td>Y21L</td>
<td>tgtggaagctctaaatgagcag</td>
<td>caggcttcctctctcctcttag</td>
<td>124</td>
</tr>
<tr>
<td>Y25L</td>
<td>aatctctgtcctattgtgtgagtc</td>
<td>gccgagctntctctctctcag</td>
<td>62</td>
</tr>
<tr>
<td>Y25R</td>
<td>gacgcatctggtagatctttt</td>
<td>tatccacccagcaccagc</td>
<td>112</td>
</tr>
<tr>
<td>Y26R</td>
<td>tgtgctgagagctttcttcc</td>
<td>ggtggtttctcctaggttaacct</td>
<td>199</td>
</tr>
<tr>
<td>Y27L</td>
<td>cttcttgtgacactcaccgcc</td>
<td>gccctcggagcctcttcgct</td>
<td>275</td>
</tr>
<tr>
<td>Y27R</td>
<td>ggaactgtaacctagtactgt</td>
<td>ggtggtggaatttacaaatcataaagtcttagcttgct</td>
<td>126</td>
</tr>
<tr>
<td>Y28L</td>
<td>tagttcctccagatgccgaga</td>
<td>agcagacctgctctctctctataggg</td>
<td>99</td>
</tr>
<tr>
<td>Y28R</td>
<td>atcaacgcttagaagacac</td>
<td>atgacaagctgcttccagcccc</td>
<td>130</td>
</tr>
<tr>
<td>Y29L</td>
<td>ccctgtgctctctctctgtaac</td>
<td>gccctcctgctctttggtgtgct</td>
<td>185</td>
</tr>
<tr>
<td>Y29R</td>
<td>ggcacagaacttaaagagac</td>
<td>ctgactgtggagggcag</td>
<td>144</td>
</tr>
<tr>
<td>Y30L</td>
<td>tccctcagctagatagtagtctc</td>
<td>aaaaaacccttaaactaacttagcttgcttgct</td>
<td>222</td>
</tr>
<tr>
<td>Y31L</td>
<td>gtcatcgtgcatggccag</td>
<td>cactactgcaccaaaaactggaa</td>
<td>136</td>
</tr>
<tr>
<td>Y32L</td>
<td>tgtgaagtttaggtgagggaggtg</td>
<td>tctggtcagttgctagcaga</td>
<td>165</td>
</tr>
<tr>
<td>Y33L</td>
<td>gggttcaccagcgaccttct</td>
<td>caaagcttttaatccgacccctta</td>
<td>149</td>
</tr>
<tr>
<td>Y34L</td>
<td>taggagacattacggggccga</td>
<td>agaattttgcacaaataagtgtagtgag</td>
<td>175</td>
</tr>
<tr>
<td>Y35L</td>
<td>gaaatctcctccctaatagcctcc</td>
<td>catcagcctttgctcctagg</td>
<td>213</td>
</tr>
<tr>
<td>P7N</td>
<td>tcaacgctgtactagtaagatgg</td>
<td>gctttcatgagattcctaggcc</td>
<td>270</td>
</tr>
<tr>
<td>P7S</td>
<td>cccctttggaataacctcacc</td>
<td>gcacggacagcaaggtga</td>
<td>181</td>
</tr>
<tr>
<td>P8S</td>
<td>ggcggaccgctggcaagacag</td>
<td>tccaatgagactgagctctcc</td>
<td>177</td>
</tr>
<tr>
<td>P9S</td>
<td>ggattctcctctttgttagaacgc</td>
<td>cttgcatccacctgacattct</td>
<td>163</td>
</tr>
<tr>
<td>P10N</td>
<td>ccctttgggccagctcttttc</td>
<td>gcacttgaggccacaagttcct</td>
<td>196</td>
</tr>
<tr>
<td>P10S</td>
<td>taaactgtgactctctcttgagggc</td>
<td>tttccggctaacctctcctcctct</td>
<td>97</td>
</tr>
<tr>
<td>P11S</td>
<td>tacactcgtcttacacacagctgct</td>
<td>tctgctacgttcttttctctgct</td>
<td>183</td>
</tr>
<tr>
<td>P13S</td>
<td>tacacacatgcctgtgtaacct</td>
<td>tctgctacgttcttttctctgct</td>
<td>174</td>
</tr>
<tr>
<td>P14N</td>
<td>agggttggtttggaccactctggtg</td>
<td>aaaaaccgtggctgagttctggtg</td>
<td>170</td>
</tr>
<tr>
<td>P14S</td>
<td>tctactcactcctattagctgctca</td>
<td>tgctgtacattcagctattgc</td>
<td>149</td>
</tr>
<tr>
<td>P15N</td>
<td>cttcacctcctttgcctgctc</td>
<td>gattccgtaaatgggtgcttgct</td>
<td>123</td>
</tr>
<tr>
<td>P15S</td>
<td>ccacgtactcagaaagggccag</td>
<td>tgcacaaagacgaggggtagtaaga</td>
<td>121</td>
</tr>
<tr>
<td>P19N</td>
<td>gagaatttggaacacctctctcctcctc</td>
<td>aacactttgattggctgctgcc</td>
<td>156</td>
</tr>
<tr>
<td>P23S</td>
<td>ttgctactctcacttgtgttgg</td>
<td>tggcttgagctctgtctgaag</td>
<td>166</td>
</tr>
<tr>
<td>C1193T7</td>
<td>atctctctctctatcctggg</td>
<td>ttctaccgtctctctctgg</td>
<td>237</td>
</tr>
<tr>
<td>C1204T3</td>
<td>tgaactctggtggcgagatg</td>
<td>ccagctcctcctattagcagct</td>
<td>191</td>
</tr>
<tr>
<td>C1204T7</td>
<td>gacacacaacctctgtaatggg</td>
<td>taaagatgactattttattttttatagttagtc</td>
<td>114</td>
</tr>
</tbody>
</table>

Table III. Primers to amplify terminal fragments of YACs, Bacteriophage P1s, and cosmids. The mouse sequence adjacent to the centromeric arm of the pYAC4 vector is referred to as the left end, or Y#L. Similarly, the terminal mouse sequence adjacent to the noncentromeric arm is referred to as the right end, or Y#R. P#N is the end of the P1 clone adjacent to the NotI site of the vector and P#S is the end of the P1 clone adjacent to the SalI site of the vector.
<table>
<thead>
<tr>
<th>YAC #</th>
<th>Address</th>
<th>Size (kb)</th>
<th>Locus Screened</th>
<th>Left end</th>
<th>Right end</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>PR C100 G10</td>
<td>380</td>
<td>D11Mit7</td>
<td>repetitive</td>
<td>contig</td>
</tr>
<tr>
<td>5</td>
<td>MIT ftw G4</td>
<td>375</td>
<td>D11Mit7</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>6</td>
<td>PR D21 E7</td>
<td>250</td>
<td>D11Mit7</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>7</td>
<td>MIT fcd A10</td>
<td>500</td>
<td>D11Mit7</td>
<td>chimeric</td>
<td>N.D.</td>
</tr>
<tr>
<td>8</td>
<td>MIT fap F10</td>
<td>150</td>
<td>D11Mit7</td>
<td>contig; SSLP</td>
<td>contig; SSCP</td>
</tr>
<tr>
<td>9</td>
<td>MIT ftk E6</td>
<td>500</td>
<td>Nf1, exon1</td>
<td>contig; SSCP</td>
<td>chimeric</td>
</tr>
<tr>
<td>10</td>
<td>MIT fck C3</td>
<td>600</td>
<td>Nf1, exon1</td>
<td>contig; SSCP</td>
<td>N.D.</td>
</tr>
<tr>
<td>11</td>
<td>MIT fex C6</td>
<td>625</td>
<td>Nf1, exon1</td>
<td>contig; SSCP</td>
<td>N.D.</td>
</tr>
<tr>
<td>14</td>
<td>MIT fcl F6</td>
<td>700</td>
<td>D11Mit34</td>
<td>contig; size</td>
<td>contig</td>
</tr>
<tr>
<td>15</td>
<td>MIT fai B7</td>
<td>700</td>
<td>D11Mit34</td>
<td>N.D.</td>
<td>contig</td>
</tr>
<tr>
<td>16</td>
<td>MIT adg D4</td>
<td>450</td>
<td>D11Mit8</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>17</td>
<td>MIT fex C7</td>
<td>575</td>
<td>D11Mit8</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>18</td>
<td>MIT fex E9</td>
<td>650</td>
<td>D11Mit8</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>19</td>
<td>MIT fao E1</td>
<td>700</td>
<td>RDA6.2</td>
<td>contig</td>
<td>contig; SSCP</td>
</tr>
<tr>
<td>20</td>
<td>MIT fbx A8</td>
<td>700</td>
<td>RDA6.2</td>
<td>contig</td>
<td>N.D.</td>
</tr>
<tr>
<td>21</td>
<td>MIT fbn E6</td>
<td>1000</td>
<td>RDA6.2</td>
<td>contig</td>
<td>repetitive</td>
</tr>
<tr>
<td>22</td>
<td>MIT fee F8</td>
<td>550</td>
<td>RDA6.2</td>
<td>N.D.</td>
<td>chimeric</td>
</tr>
<tr>
<td>25</td>
<td>PR A15 B3</td>
<td>225</td>
<td>D11Mit34</td>
<td>contig</td>
<td>contig</td>
</tr>
<tr>
<td>26</td>
<td>PR C189 G5</td>
<td>200</td>
<td>D11Mit34</td>
<td>N.D.</td>
<td>contig; SSCP</td>
</tr>
<tr>
<td>27</td>
<td>PR C190 D10</td>
<td>300</td>
<td>D11Mit34</td>
<td>contig; SSCP</td>
<td>contig</td>
</tr>
<tr>
<td>28</td>
<td>MIT aan G1</td>
<td>500</td>
<td>Y14R</td>
<td>contig</td>
<td>contig</td>
</tr>
<tr>
<td>29</td>
<td>MIT fel D9</td>
<td>625</td>
<td>Y14R</td>
<td>contig</td>
<td>contig; SSCP</td>
</tr>
<tr>
<td>30</td>
<td>MIT ftfq G1</td>
<td>675</td>
<td>Y14R</td>
<td>contig</td>
<td>chimeric</td>
</tr>
<tr>
<td>31</td>
<td>MIT fa1 G2</td>
<td>1000</td>
<td>RDA10.4</td>
<td>contig</td>
<td>contig; SSCP</td>
</tr>
<tr>
<td>32</td>
<td>MIT feq B12</td>
<td>550</td>
<td>RDA10.4</td>
<td>contig</td>
<td>chimeric</td>
</tr>
<tr>
<td>33</td>
<td>MIT fes P2</td>
<td>625</td>
<td>RDA10.4</td>
<td>N.D.</td>
<td>contig</td>
</tr>
<tr>
<td>34</td>
<td>MIT fex D10</td>
<td>675</td>
<td>RDA10.4</td>
<td>repetitive</td>
<td>contig</td>
</tr>
<tr>
<td>35</td>
<td>MIT far F12</td>
<td>400</td>
<td>RDA10.2</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
</tbody>
</table>

Table IV. Characterization of the ends of the YACs from the nude region. MIT is a YAC from the MIT Library; PR is a YAC from the Princeton Library. The mouse sequence adjacent to the centromeric arm of the pYAC4 vector is referred to as the left end. Similarly, the terminal mouse sequence adjacent to the noncentromeric arm is referred to as the right end. D11Mit7, D11Mit8 and D11Mit34 are SSLP markers from the MIT genome center. RDA6.2, RDA10.4 are markers obtained by GD-RDA. Nf-1 is a previously identified locus from this region (Andre Bernard, unpublished sequence). Y14R is the right end of YAC14. N.D. is an end that was not determined; repetitive is an end whose terminal sequence is repetitive and therefore could not be mapped; contig is an end that maps to the region; chimeric is an end that maps elsewhere in the genome. As well, if the ampiclon from the terminal sequence is a genetic marker, the type of polymorphism is described.
<table>
<thead>
<tr>
<th></th>
<th>D11Nds1</th>
<th>D11Mit7</th>
<th>Y8L</th>
<th>Y19R</th>
<th>RDA62</th>
<th>Y22L</th>
<th>RDA10.4</th>
<th>P7N</th>
<th>CA69</th>
<th>CA128</th>
<th>Phenotype</th>
<th>Genotype</th>
<th>C4</th>
<th>CA4</th>
<th>Y22L</th>
<th>P14N</th>
<th>P75S</th>
<th>PDA10.2</th>
<th>RDA11.17</th>
<th>Y31R</th>
<th>BS11.4</th>
<th>BS11.5</th>
<th>BS11.7</th>
<th>BS11.8</th>
<th>BS11.9</th>
<th>D11Mit117</th>
<th>D11Mit144</th>
<th>Y30R</th>
<th>D11Mit118</th>
<th>D11Mit96</th>
<th>D11Mit94</th>
<th>D11Mit92</th>
<th>D11Mit8</th>
</tr>
</thead>
<tbody>
<tr>
<td>C11</td>
<td>H</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>nude</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M100</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>nude</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M210</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>nude</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M433</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>nude</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M486</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M564</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M152</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M465</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>C100</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>?</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>C190</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>?</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>C216</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>?</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>C223</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>?</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>C4</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M952</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>C161</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>nude</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>M06</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M496</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>wt</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>M558</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M603</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>M143</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>wt</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
</tbody>
</table>

Table V. The phenotype and genotype with the SSCP, SSLP and the RDA genetic markers of the animals with recombination breakpoints in the smallest nude region. An (AKR-nusstr X CAST/Ei) F₂ animals is abbreviated as C#: and an (AKR-nusstr X MOLF/Ei) F₂ animals is abbreviated as M#. For the genotypes: A is an AKR-nusstr homozygote; C is a CAST/Ei homozygote; H is a heterozygote; M is a MOLF/Ei homozygote; and ? is an undetermined genotype.
<table>
<thead>
<tr>
<th>YACs</th>
<th>Primer name</th>
<th>Primer Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left side</td>
<td>LEFT1</td>
<td>TTT CTT CAA CAA TTA AAT ACT CTC GG</td>
</tr>
<tr>
<td></td>
<td>INVL-2</td>
<td>AAG AAT TGA TCC ACA GGA CCG GTG T</td>
</tr>
<tr>
<td></td>
<td>M13FORL1</td>
<td>TGT AAA ACG ACG GCC AGT GTA GCC AAG TTA GTT T</td>
</tr>
<tr>
<td></td>
<td>M13REVl2</td>
<td>CAG GAA ACA GCT ATG ACC TGG TCG CCA GAT GGT T</td>
</tr>
<tr>
<td>Right side</td>
<td>RIGHT2</td>
<td>CAA AAA AAT CTC CCC CCC C</td>
</tr>
<tr>
<td></td>
<td>RIGHT7</td>
<td>TTC AAG AAT TGA TCC TCT ACG C</td>
</tr>
<tr>
<td></td>
<td>M13REVR2</td>
<td>CAG GAA ACA GCT ATG ACC CGA GTC GAA CGC CCG A</td>
</tr>
<tr>
<td></td>
<td>M13FORR7</td>
<td>TGT AAA ACG ACG GCC AGT GCC GGA CGC ATC GTG</td>
</tr>
<tr>
<td></td>
<td>M13FORALUR</td>
<td>TGT AAA ACG ACG GCC AGT CAT AAG GGA GAG CGT C</td>
</tr>
<tr>
<td>COSMIDS</td>
<td>Primer name</td>
<td>Primer Sequence 5'-3'</td>
</tr>
<tr>
<td>T3 side</td>
<td>CT3_7931For</td>
<td>CGT CTT CAA GAA TTC GCG G</td>
</tr>
<tr>
<td></td>
<td>CT3_M13For3</td>
<td>TGT AAA ACG ACG GCC AGT TTC GCG GCC GCA ATT AAC</td>
</tr>
<tr>
<td></td>
<td>CT3_7921For</td>
<td>CGT GAT ACG CCT ATT TTT ATA GG</td>
</tr>
<tr>
<td></td>
<td>CT_M13Rev7901</td>
<td>CAG GAA ACA GCT ATG ACC AGG TTA ATG TCA TGA TAA TAA TGG</td>
</tr>
<tr>
<td>T7 side</td>
<td>CT7_147For</td>
<td>GAA AGG AAA CGA CAG AGG CC</td>
</tr>
<tr>
<td></td>
<td>CT7_94Rev</td>
<td>TCG ATG ATA AGC GGT CAA AC</td>
</tr>
<tr>
<td></td>
<td>CT7_M13For170</td>
<td>TGT AAA ACG ACG GCC AGT AAG CTC GCT TTC AGC ACC T</td>
</tr>
<tr>
<td></td>
<td>CT7_M13Rev77</td>
<td>CAG GAA ACA GCT ATG ACC AAC ATG AGA ATT CGC GGC</td>
</tr>
<tr>
<td>P1</td>
<td>Primer name</td>
<td>Primer Sequence 5'-3'</td>
</tr>
<tr>
<td>NotI side</td>
<td>FN_-30For</td>
<td>GCC GCT AAT ACG ACT CAC TAT AGG</td>
</tr>
<tr>
<td></td>
<td>FN_40Rev</td>
<td>GAG CAA TAT AGT CCT ACA ATG TCA AGC</td>
</tr>
<tr>
<td></td>
<td>FN_M13For-20</td>
<td>TGT AAA ACG ACG GCC AGT CAC TAT AGG GAG AGG ATC</td>
</tr>
<tr>
<td></td>
<td>FN_M13Rev-50</td>
<td>CAG GAA ACA GCT ATG ACC AGT CCT ACA ATG TCA AGC</td>
</tr>
<tr>
<td>SfII side</td>
<td>PS_64For</td>
<td>GAG GAT CGA AAC GCC AGA</td>
</tr>
<tr>
<td></td>
<td>PS_18Rev</td>
<td>CGG TCG ACA TTT AGG TGA CAC</td>
</tr>
<tr>
<td></td>
<td>PS_M13For79</td>
<td>TGT AAA ACG ACG GCC AGT AGA TCG CAA AAA ACA GTA</td>
</tr>
<tr>
<td></td>
<td>PS_M13For312</td>
<td>TGT AAA ACG ACG GCC AGT GCC TCA ATT CAA TCA ACG</td>
</tr>
<tr>
<td></td>
<td>PS_M13Rev18</td>
<td>CAG GAA ACA GCT ATG ACC TGA CAC TAT AGA AGG ATC</td>
</tr>
</tbody>
</table>

Table VI. Primers to clone terminal fragments of YACs, P1s, and cosmids by inverse PCR.
<table>
<thead>
<tr>
<th></th>
<th>size</th>
<th>neonatal skin</th>
<th>AKR skin</th>
<th>nustr skin</th>
<th>B6 skin</th>
<th>B6-nu skin</th>
<th>rat-nu skin</th>
<th>thymus</th>
<th>brain</th>
<th>liver</th>
<th>heart</th>
<th>testes</th>
<th>kidney</th>
<th>10.5 dec</th>
<th>11.5 dec</th>
<th>10 cDNA</th>
<th>11 cDNA</th>
<th>SK cDNA</th>
<th>THY cDNA</th>
<th>TES cDNA</th>
<th>BR cDNA</th>
<th>AKR DNA</th>
<th>nustr DNA</th>
<th>B6 DNA</th>
<th>nu DNA</th>
<th>rat DNA</th>
<th>rat-nu DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hlh11; ET90</td>
<td>100</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Na/sulf; ET4.1</td>
<td>283</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>VIT; ET14</td>
<td>150</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>nmo; ET11.10</td>
<td>110</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Mac30; ET6</td>
<td>130</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Ad-Rab; ET7.3</td>
<td>110</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>TNFa1;ET11.1</td>
<td>114</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET2.8</td>
<td>95</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET3.8</td>
<td>110</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET101</td>
<td>140</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET98</td>
<td>140</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET69</td>
<td>110</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET3.4</td>
<td>153</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET101</td>
<td>140</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET99</td>
<td>110</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET98</td>
<td>140</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET25</td>
<td>80</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ET6.10</td>
<td>100</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Table VII. Expression patterns of transcription units from the *nude* region. PCR amplification of reverse-transcribed cDNA from a variety of tissues; cDNA libraries and genomic DNA.
APPENDIX I: Gene fragments identified in the nude region by direct cDNA selection and exon trapping. For each entry the following information is provided: (i) Fragment name; (ii) Primers to amplify the locus; (iii) size of the insert; (iv) duplicate sequences, showing the overlap; (v) nucleotide blast matches; (vi) protein blast matches; and (vii) physical map location on the P1 contig. Please note that this appendix will be revised for the final copy to alphabetize the entries and some other minor modifications.

SEQUENCE FILENAME: A1
FORWARD PRIMER: CATTTCTCCTGCTTAGTTCTGG
REVERSE PRIMER: TGGGGAATCCAGGCTTGTGA
LENGTH OF INSERT: 274

AGCTCCCATGCAGAACCATGTCCCTGCTTAGTTCTGGACATCCAAGACAGGGGAAGAACATTTCCCCAT
TCGAGGGATACCTTGGTGACAGGGGAAGAACATTTCCCCAT
CTCCTGACCTGAGACCGACGACGCTTTGTAAGGGAGTTGCGCCAGCCTCCCATAGGGTCAATACAGC
GAGGACTGGACTCTCGTCGGTCCTATTCCACTCCAAAAACGGTCGGAGGGTATTCCCAGTTATGGTC
CCTGGATCCCCAAACCTGTGCTCAGGAGGAAGGGGGCAGTCCCCAAGGCTATCTACTCACAT
GGACCTAGGGGTTTGGACAGAATCTCTCTGTGACCCCGACTCTTANGGGGTTCCGATAGATGAGTGTA
TTGTGCCTGGACCCGGGCTCACCTGGNCTNCTCCTGCCTNATGGCTNGCTGTAACTANCTGGG
AACACGGACCTGGCCGGGAGTGACCNGAAGAGGCAGGANTACNGAAANCAGACATTGATNGACC

KEY:
DUPLICATE REPORT FOR A1: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,21,24,25
END OF RECORD.

SEQUENCE FILENAME: A6
FORWARD PRIMER: ACTTCTCAGATTCTTCCTGCA
REVERSE PRIMER: CCTGTCCTCCACAAGTCTGT
LENGTH OF INSERT: 147

TCCACTCTCAGATTCTTCCTGCAATTTGGGAAGATTGAGGGCCACAGACACGGCCCTTC
AGGTGAAGACGTCAAGAAGGACGTAGAACAGCTAAAACCTTCCTAACTCCGGTTCTGGGAAAGG
GAGAGGAACTTCCGGATTTATTCTCCACCTTCTCAAAGCACTTGTTGAGGGAAGGTTATGGCGAACAG
CTCTCCTGACCCGGATTTATCTCCACCTTCTCAGAAAGCATTGTTAGGGACAGGGTTATTGCGGACAG

DUPLICATE REPORT FOR A6: <no duplicates>
NUCLEOTIDE BLAST HITS:
['M87634', 'Rat BF-1 mRNA, complete cds.', 0.00010]
['X74143', 'H.sapiens HBF-2 mRNA for transcription...', 0.00023]
['X74142', 'H.sapiens HBF-1 mRNA for transcription...', 0.00023]
['X74144', 'H.sapiens HBF-3 mRNA', 0.0013]
PROTEIN BLAST HITS:
['A42826', 'HTLF=---T-cell leukemia virus enhancer ...', 1.7e-10]
['P32314', 'HUMAN T-CELL LEUKEMIA VIRUS ENHAN...', 1.7e-10]
['A49395', 'glutamine (Q)-rich factor 1, QRF-1 ...', 5.5e-05]
['L13201', 'HNF-3/forkhead homolog-1...', 5.7e-05]
PHYSICAL MAP LOCATION: 11,12,13,22,23
END OF RECORD.
SEQUENCE FILENAME: A8
FORWARD PRIMER: ACAAGCTCTGTGTGGGTGAG
REVERSE PRIMER: GTGAAAAAANAGGGTGCTT
LENGTH OF INSERT: 268

TGGGGCTTCCCTGAGCGAGGTTGGTGCAAGCTCTGTGTGGGTGAGGAGCTGGGGAAGTCAGATGGAGCA

ACCCCGAAGGACTCGTCCACCCACTGTTCGAGACACACCCACTCCTCGAGCCCTCTCTCTAGTCTACCTCGT

GTTTGCTGGGGAAANAAANGGTGTCCTACTGCTGGTCTCNGTCAGTNCTACTGCAANTCCTCTGTGTG

CTNACTCCCAAATGGAAAGGAGCTGGCTAANAGGNTAGAATGGGGTACCAG

DUPLICATE REPORT FOR A8: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,21,24,25
END OF RECORD.

SEQUENCE FILENAME: ABX11_1
FORWARD PRIMER: ACCTCCTGCCTGTGGTG
REVERSE PRIMER: CCCACATTGAGCTGCACGTAC
LENGTH OF INSERT: 291

ATAAGGGCCACTCAGTGATCCTCTCACCCACTCACCAGGCGAGAACCTCTTACCTCTCCCTCGGCTGTGGG

TTATCCCGGGGATTGAGCACACTAGTGAAAGGGTCGCTCTGGAAAGATTCCGATGAGGAGCGAGCACCC

TGATCTACCCATGGGAGAGGAGTCAGGGCAAGCCTCTCTGCTGAGAGGGCGGCCAAGCCCAAGATCA

ACTAGATGGACGTGCTCTCGAGTCGCAGGCCAGAGGAGGTAGTGGGTCAGGGTCTTAG

AGTGGCCCTTCAAGGGGCGCTGGGCTGGGCAACAAAGTACCTGGCAGCTCAAMTGAGGGGCTCCCTCTAGA

TCACCCGAGGCTCTCCCGCCCGACCCGCTGCTGACTGCGAGTCTCAACCTCAGGACAGAATGGGAGTCC

CCACAGTGCCCGGCCCTCACCAGCCACACCTCTCAAGGCACTGTTCAAGGGGAAATGGGAGTGC

DUPLICATE REPORT FOR ABX11_1: <no duplicates>
NUCLEOTIDE BLAST HITS:
- ['MB0783', 'Human B12 protein mRNA, compl...', 3.0e-53]

PROTEIN BLAST HITS:
- ['A41784', 'tumor necrosis factor-alpha-induced ...', 3.0e-33]
- ['MB0783', 'B12 protein [Homo sapiens]', 3.4e-25]

PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: ABX11_10
FORWARD PRIMER: TGACCGACTCACAACACCC
REVERSE PRIMER: TGCTGTTGAATGAGAGGCA
LENGTH OF INSERT: 177

ACCAAGTGACCTTTGGACAGCAGTCAACACACCCCAATACTTGATGAACACCTTTTGAGAAGAACCTAAGCTCTG

TCGTTCTGACATCTGACAGTGGAGGTGGTAAAACCTACTGTGGAACATCTTTTGACTGATCGGAC

TCGCAGACGTAAAGAAGAATTACAGATCAGACATTCTGGAACAGCAGAAAGGAAACCAGAGTGGCCTTCTTC

AGGCTGTCACATCTTTAAATAGTGACATGACGGACCTTGCTGGGCTCAGGAGAG

ATCAACAGAATGCGCCATCCTTAAGAGCAGTCTTATAGA

TAGTGTGCTTCAGGCGAGAAATCTCTGAAATAATCT

DUPLICATE REPORT FOR ABX11_10: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
["U12010", "nemo gene product [Dros...", 1.4e-07]
["U12009", "nemo gene product [Dros...", 1.5e-07]
PHYSICAL MAP LOCATION: 1,2,15
END OF RECORD.

SEQUENCE FILENAME: ABX12_1
FORWARD PRIMER: GTGCATTCATTAAGCGCTAT
REVERSE PRIMER: GTGGCCAATGGGTACGTTCC
LENGTH OF INSERT: 287

ACTCTATGGAACCCCTCGAGTCCCTGACAATTCCTTACTGGAGGCCACAAGCCAAGGCCAGCCAGGCT

TGAGATCTCTGGGAGATCAGGACTGTGAATGAAGACTCAGGCGACTGTTAAGGAATGACCTCCGGTGTTCGGTTCGGCGTCGGTCCGA

TCCCCCAGTGGAGATGAGAAGACACCTTTTPGAACCTGGACGGACGTGTCCTGCCATCATTAAAGGCTATATAGA

AGGGGGTCATTCTCTACTCTCTCTCAGGAAAATCAGGTGGCAAGGAGCAAGGTAAGTAATTCGCGATATCT

CTTACAGTGAACCCGCAACTTGGGCATCACTGCCATCTGTATCTGATAGGACCATACATAGAATTTAAGAATGGGG

GAATGCATCTGGGCTTTGGAACGGGTATCTACGGTGAATGACCTCAGTACCTCTGGAGAAATTTTATGGC

CAGAAGAGAGTTCTGACATCTGACAGGACCATTTGGGCGACCTTATGGGATACCGACGCTGATTCTGGAG

GTGGTCTTCTGCAGACGCTTGGGACACTTGCAATCAGGATGAACTGAGGATAGGACGACTGACGAG

AGCCCTCTCTAA

TGGGAGGATT

DUPLICATE REPORT FOR ABX12_1: <no duplicates>
NUCLEOTIDE BLAST HITS:
["M80783", 'Human B12 protein mRNA, compl...', 1.2e-31]
PROTEIN BLAST HITS:
["A41784", 'tumor necrosis factor-alpha-induced ...', 1.2e-27]
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: ABX12_3
FORWARD PRIMER: CCCTATGCTGCTCCTGCTAG
REVERSE PRIMER: GGTGGCATCTGGAAGTGCTGG
LENGTH OF INSERT: 290

GCTTACGATGACCGGCATCTTGGCCATCAGTCCACTCATCGTGACTGATGAGACCCTTAAGAAATCGGG
CGAATGCTACTGGCCCTAAGACCGTATGCTGAGTGATGACGACTCTACCTCTGGAAATCTTTAGGCC
GACAAGAGACTCTGGTCCACCCCTGGAGACTGACTGCTGGGTACCTGTTGGGCTGCTTTGATGACGTCGT
GGGGAGCCTGCTCTAGTATGTACAGGTGACAGCCCTGCTTGGGAGGTTGGGGAGCTGACGAGGGAAGGCC
CCCTCGGACGAGATCATACATGTTCATCTGCCTGGAAGAACAACCCTCAAACCTCCTGACCTCTCCCTCCG
ACATGTCTGCTGCTGCCCCCTCCCCCAACCCGACTCCCAAGCGACCACCTCTTGCCTGATATAAAGAGCT

TGTAACAGAACCGGACGAGGAGGGGTCTGGAAGTGCTACCGTTGGGATAGGACTTTACTTCTCGA

DUPLICATE REPORT FOR ABX12_3: ABX12_1 (22)
vs ABX12_1 (score = 143)
  ABX12_1  ACTCTATGAGACCCCTCGAGTCCCTGACAATTCCTTACTGGAGGCCACAAGCCAAGCCGCAGC
  ABX12_1  GGCTTCCCCCAGTAAGATGAAGACACCTTTGAACTACGGGACCGTGTCCGTGCATTCATTAAGC
    GCTTACGATGACCGGCATCTTGGCCATCAGTCCACTCATCGTGACTGATGAGACCCTTAAGAAATCGGG
    CGAATGCTACTGGCCCTAAGACCGTATGCTGAGTGATGACGACTCTACCTCTGGAAATCTTTAGGCC
    GACAAGAGACTCTGGTCCACCCCTGGAGACTGACTGCTGGGTACCTGTTGGGCTGCTTTGATGACGTCGT
    GGGGGAGCCTGCTCTAGTATGTACAGGTGACAGCCCTGCTTGGGAGGTTGGGGAGCTGACGAGGGAAGGCC
    CCCTCGGACGAGATCATACATGTTCATCTGCCTGGAAGAACAACCCTCAAACCTCCTGACCTCTCCCTCCG
    ACATGTCTGCTGCTGCCCCCTCCCCCAACCCGACTCCCAAGCGACCACCTCTTGCCTGATATAAAGAGCT

ABX12_1

NUCLEOTIDE BLAST HITS:
['M80783', 'Human B12 protein mRNA, compl...', 6.8e-18]

PROTEIN BLAST HITS:
['P10495', 'GLYCINE-RICH CELL WALL STRUCTURAL...', 1.3e-05]

PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: ABX2_8
FORWARD PRIMER: TACCGGAAGCTTTGAGCACG
REVERSE PRIMER: GGGCGACCTGGACTATGC
LENGTH OF INSERT: 399

TTAACCTTAACTAAAGGAACAAAACTGGTACGGGCCCCCCCCTTCAGGTTCAAGGTAACCATAACTTTTG
AATTTGGAATTGTTTCTCTTTGAGCAATGCGGGGAGGGAAATGTCCAAGTTCTCATGTAATGAAAC
ATCTTCAAGTTCGACCTTTCTAAATTCGACCCCAAGCAGCAAGATGAAACTCAAGTGCCTGGTTTCAAGAA
TAGAGTGTTTCAACGGTCGAAGATTTAAGCTGGGCTGTTTCTACTTPTTAGTTCACGGCAAGAGTCTC
AAGCTCTACCGGAAGCTTTGAGCAAGGAGGAGGCGCTGTCCCTGGCAGTGCTGACGGACATGATGAGC
TTCGGATGGCCTTCGAAACTCGTGCTCCTCCGCGACAGGGACCGTCACACGTGCCTGTACTACTCG
TAGATTGGGCTTGGAGGACCGAGCATGTGCTCAGGCTTTGCTGCTGCTGGCCAG

DUPLICATE REPORT FOR ABX2_8: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 12,22,23,27
END OF RECORD.

SEQUENCE FILENAME: ABX3_2
FORWARD PRIMER: GTCCGCTTCTGAAAGGCG
REVERSE PRIMER: AGTGGAAAGTTGGGTACCCC
LENGTH OF INSERT: 246

AGCACGCTGGGCTGGCCTGGCTTCTGAAAGGCGACCATTCTCCGTTCTGCATGCCAAGCCTCCCCCAT
TCGTGGACAACGCGGACAGGCAC

DUPLICATE REPORT FOR ABX3_2: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: ABX3_4
FORWARD PRIMER: GCAAGGGGAAGCTGATATAT
REVERSE PRIMER: TCGGCGTTGGGCCGGGTCTG
LENGTH OF INSERT: 170

AACCCCTCGGGGCAAGGGGAAGCTGATATATAATTATTGGCTCACCCAGAGCTGCCTCCCCTCACCTGA
<<<<<<<<<<<<<<<<<<<<<<<
TTGGGAGGCCCTTTCCCTTTCCACTATAATTATAATACCGGAGTGGGTCTGACAGCGAGGGGAGTGGGCT
<<<<<<<<<<<<<<<<<<<<<<
TGTCAGGCCAAGCCAGCAAACTCAGGCTGACCCGACATACAGGACTGCGGAGCTCTTGCGCCCCCAGACCC

DUPLICATE REPORT FOR ABX3_4: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.

SEQUENCE FILENAME: ABX3_8
FORWARD PRIMER: AAGGCTGGTGGGTGTCTCA
REVERSE PRIMER: AAACTTCTGATAAATGGGTC
LENGTH OF INSERT: 127

ATTAAAGGCTGGTGGGTGTCTCAACCACTACGTCTCCACATTCCTGTCCGGAGTGATGCTGACCTGGCC
<<<<<<<<<<<<<<<<<<<<<<
TAATTTCCGACCACCCACAGAGTTGGTGATGCAGAGGTGTAAGGACAGGCCTCACTACGACTGGACCGG
TAATGGACCCATTTATCAGAAGTTTCGCAACCAGTTCTTAGCATTTTCCTTTCAGAA

DUPLICATE REPORT FOR ABX3_8: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: ABX4_1
FORWARD PRIMER: AACTGACGCTCCTAGAAGTCGA
REVERSE PRIMER: GCAAATTGCCCCAAATG
LENGTH OF INSERT: 360

TAAGCCTTGACACAAACTGACGCTCCTAGAAGTCGAACCCAGCAAATTAAAGGTGACGCTGTAATCTGT

ATCCGAACTAGAGTTGGTACGGGAGGCATTCCAAGTCGACATAGACA

TTCCCTGAGCAAAGACTAGAAGAACAAAAAGACATCACCAGGGTCTAGTCTCTG

AAGGAGCTGCTGTCTCTGCTGATTGTTTGACTGCGAGGATCTTCAGCTGGGTCGTTAATTTCCAGTCGACATAGACA

CATCTGCTACTCAGCCGAGGATGTTGGCGCTGACGCGTACACACAAACCTGTGGCTCAAGG

<<<<<<<<<<<

GTAGACGATGAGTCGGTCGTAAACCCCCGTTAAACGGGATGCCCATGTGTGTTTGGACCACGAGTTCC

GTCCTGTTGAGCTAGAAGGGACATTTATACATTGACACCCTTGAAAGGATAGTAACACAAAAACCCAGAA

CCCCACCATGATGATCCTGCTCTACTGCTGCTGATGCTTAGTACACAGGTGCTCTCTCTGCTGGTGTCAGTA

GGGTTGATCACTAGAAGCGAGATGACCAGGAACAGATCGATGTCCAGGAGAAGGACCACAGCTCAT

AAGTGCTGCCCTAGAC
TCACCACGGAATCTG

DUPLICATE REPORT FOR ABX4_1: <no duplicates>

NUCLEOTIDE BLAST HITS: <none>

PROTEIN BLAST HITS:
['A47714', 'Na+/sulfate cotransporter, renal - r...', 2.1e-13],
['Z30974', 'K08E5.2, similar to Yeas...', 4.3e-07],
['A45484', 'delta-like dlk homeotic protein - mo...', 1.1e-05],
['D16847', 'Stromal Cell derived Pro...', 1.1e-05],
['Z34533', 'B0285.6, similar to C. e...', 7.3e-05]

PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: ABX6_10
FORWARD PRIMER: CTTCGCAACCCTCTTGCTTAG
REVERSE PRIMER: CCTACTTCATCATCAAGGAAG
LENGTH OF INSERT: 275

AGTGTGACCCCTGACAGTTCTCCTGGAAAAAGACTAGGGGAGATGGACAGACTCAAGATTCCCATCCAGCAA

TCACACTGGGACTTCAAGGGAAGTTTCTGATCCCTACTGCTCTGAGTCTCTAAGGGTATGGTCTGTT

GGTACCTGCCACATGAAGTCCACATCCACCGAGCTCTGCAACCTCTTCTGCTAGCTGCAATGGACTCAAA

CCATGAGGGGAGGGTTGCTGAGGTCCAGGACGGTGGAGTGGAGAAGGGATGACGCTCAGCTGAGTT

ACGTTGCGGAGAAGGCCACCACCGGGAGGAGGAGGATCTGAGGACACGCGAGAGGAGAGGAGAAGGAGG

TGCAACCCCTCTCCGGGCACTCCTCTCCAGGCTCTGGTCTCCCTGACCGACCAATCATCCTGGGCCTGC

AAGACTCACCTGGGAGATAAATAACTGCTTTCCATGATACTGAAATGATAGGACGCTTTAAAAACAGAGAGGAG

<<<<<<<<<<<<<<<<<<<<<<

TTCTGAGTGGACCCTTCTATTTGACGAAGGTACTACTTACTTCATCCGTCGAAATTTGTCTTCTCCTC

DUPLICATE REPORT FOR ABX6_10: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14, 21, 24, 25
END OF RECORD.

SEQUENCE FILENAME: ABX7_2
FORWARD PRIMER: CTGGTCAGCTCGGTCTGG
REVERSE PRIMER: CGCAGCGGAGGACTCCTCTG
LENGTH OF INSERT: 169

ACACTGGTGGCCGGCGCCGGGGACCGCTCCCTGCTGCACCGCGTGTCAGAGCAGCAGCTCCT

>>>>>>>>>>>>>>>>>>

TGTGGATACCGGACCAGTCGAGCCAGACCTGGCGAGGGACGACGTGGCGCACAGTCTCGTCGTCGAGGA

GGAGGACTGTGACATCCCGCTTGGAGTGCACCGACCCGCTACTTCTCTGACGCCAGAGGAGTCTCCGCC

<<<<<<<<<<<

CCCTCTGACAGCCGACGCGCTACTGCTGCTGGTCTCGTAGAGAACATCTGGGCTCTCTGACGGAGCGG

<<<<<<<<<<<

TGGGATGGTTACCAGAACAGGAGTGTAAGGAGA

<<<

ACGCTACCATGGCTCTGCTCCATCCTCCT

DUPLICATE REPORT FOR ABX7_2: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14, 21, 24, 25
END OF RECORD.
SEQUENCE FILENAME: ABX7_3
FORWARD PRIMER: GGGTGTCGATTAACCTTCC
REVERSE PRIMER: CGGATTGCTCTTCTCTGTCAT
LENGTH OF INSERT: 185

GTACGGATTGCTCTTCTGCTATTGGTCAGACGCTCTGCTCATCCGGCAAGCTTTCAAGNTGGGTGAC
AACTCGNAGGCTAATGGAAAAGGTAAATCCACACCGGANGTCTGAGTTTCTTCAGAGCCCTG
ATCTGCCCAGGCCATGAGGAATGGAGCACCACACTGACGATA

DUPLICATE REPORT FOR ABX7_3: <no duplicates>
NUCLEOTIDE BLAST HITS: emb|234286|OCRADI2 O.cuniculus (AdRab-G) mRNA expressed i... 285 3.6e-18 2
PROTEIN BLAST HITS: <none>
gi|498893|gp|234286|OCRADI2_1 O.cuniculus (AdRab-G) mR... +1 100 1.1e-07 1
PHYSICAL MAP LOCATION: 14,21,24,25
END OF RECORD.

SEQUENCE FILENAME: ABX8_1
FORWARD PRIMER: AATTTCCTGATTCCAG
REVERSE PRIMER: TGCCATGGGACTCCAG
LENGTH OF INSERT: 276

GTTTACTGGGCCCTTAATAACACTTGGTCAAAAATTCATTTTCTTGCCCAGAGGGTGTTACC

DUPLICATE REPORT FOR ABX8_1: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 1,2
END OF RECORD.
SEQUENCE FILENAME: C11
FORWARD PRIMER: TGGAGGTAGGGGAGGATGTCT
REVERSE PRIMER: CAGAGACTGCCCCCATATAT
LENGTH OF INSERT: 209

CAGACAGGAATCTI'TAAAGAGCTCGTGTTGAATGGGACCTGATCCGTGGAGGTAGGGGAGGATGTCTC

GTCTGTCCTTAGAAAATTTCTCGAGCACAACTTACCCTGGACTAGGCACCTCCATCCCCTCTCAGAG

ATGGCTGACATAGTCCAAACCCTGAGGATAGCATAGGCGCTGTCATTATTGGATAGCCAAGATGGCTACCGACTGTATCAGGTTGGGACCCTATCGTATCTCGGGGCCGACAGTAATAACCTATCGGTTCTACCGACAGTTCTGTTCTGAA

TTCTAGATATAGTGGGCAGTCTCTGGCTCAAATCAGCACTTTGGTAGTAAGT

TGTCAAGACAAGACTTAAGATCTATATACACCGTCAGAGCCGAGTTTAGTCGTGAAACCATCATTCA
DUPLICATE REPORT FOR C11: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: C12
FORWARD PRIMER: GGTTTTATCTCCACTGCTACGA
REVERSE PRIMER: TTAGGACATGATACCTITTGA
LENGTH OF INSERT: 227

AAGGCTAGGGTAGCTCAGTATGCAAAGCCCTGGGTATCTCCACTGCTACGAAACCCAAGAAACAGT

TTCCGATCCCATCGAGTCATACGTTTCGGGACCCAAAATAGAGGTGACGATGCTTTGGGTTCTTTGTCA

GACTCTCCCTTCTCTATGCTATAGCTAGTGAAAGTTTTGATGGCCTCTTTGTAGTTGTCAGGCCCTTC

CTGAGGGGAAGAGTACCCGNTATGCAACGGGTTCATCAAAACTACCCGGAGAACATCAACAGTCCGGGAAG

CGTGGCAGCAGCTGACAGCACTCCAAAGGATCATAGGCTCTAAATCCAGTGGAAATTCPGCAGGTTCAC

GCCCGTCGCCGACAGTCTGCGAGTTGTTCATAGTACAGGATTAGTACAGGGTGTTATCCTCGGTTCCTCA

AGAAAGATTTTCAAAAAGGGGGGCTCTCAG

TCTTCTCTAAAAGTTTCTCCCCCCCCAGAGTGC

DUPLICATE REPORT FOR C12: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: C14
FORWARD PRIMER: TGTTCATCGACCTGCAGGC
REVERSE PRIMER: GCAGGAAGGACTTGAACCACA
LENGTH OF INSERT: 158

ATCACTGTTTCTATGCACTGCAGGCGGTCTGCCGCCGAACTATACCCGAGTGTTCAGCAACCTGTTGC
<<<<<<<<<<<>
TAGTGCAACAGTAGTCGAGGTTCCGGCAGAGCGGGTCTGATATGGGCTCACAAGTCGTTGGCAACG
CTGGTGATTTGTGTCCAGTCCGGGTTCAGCTCAAGAAAGGACCA
<<<<<<<<<<<
CACCATGAGTTTCCTCAAGTTTCTGGAGACTACGTCTCGGGGCTTGACACCAAGTCCAGGAAGGAGGA
CTGTCAGGTTCCTTGGTGCTGGTGAGCCAGCCAGCTCCGGTTCTCACGAATA
<<<<<<<<<<<
GACACTCAAAACAAGGGTGAGCGGCCAAGGTYGGTATAT

DUPLICATE REPORT FOR C14: <no duplicates>
NUCLEOTIDE BLAST HITS:
["L19183","Human MAC30 mRNA, 3 end.",6.9e-15]
PROTEIN BLAST HITS:
["L19183","MAC30 gene product [Ho...",2.8e-13]
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.

SEQUENCE FILENAME: C15
FORWARD PRIMER: GCTTCCTAGAGAGAGCGAGCA
REVERSE PRIMER: ACCGCTGCCCTAGGGCTCA
LENGTH OF INSERT: 198

TCAGCTTCTAGAGAGAGCGAGCCTTTCTCAAGCTTGACTTTCTGATACAGAGCCTGGGGCTTCTGA
<<<<<<<<<<<>
AGTCGAAGGATCTCTCTCTCGCCTCTCAGTTCGTCACTGAAAGACTATGTCTCGGACCCCGAAGACT
GCCCTAGGGCAGGGTCTCGCAGAGCTTCTGACATATACGATTATCTGCATACCAAGATATTCTCTTA
<<<<<<<<<<<
GGGATGCCCTGCGCAAGGGTGCCCGACACTCAGTGCTGTAATATTTAATTAGGACTATAGCTATAAAATGGAAT
TAAATCTAAACAGTGGAAAATTAGCTGAGGATGACAAAAATAATTTATGGGGGGGTCTCAC
ATTATATGATCTTCCTTTTTTTACTAGACCTTCATCATCAGTTTTTTTATTAAAAATACCAACCCCGAGAGG

DUPLICATE REPORT FOR C15: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
SEQUENCE FILENAME: C16
FORWARD PRIMER: CCCATGGTGCAAGTCTG
REVERSE PRIMER: GCGTGCTTAAGGCTCTTACC
LENGTH OF INSERT: 182

GCTGCCCTTTTCCCATGGTGCAAGTCTGCTGCTGGGCTACTGTTGGGTTGGGCTTCCCTGAGCAGGTA
CGAGGCCAAGAGGTACCAGCTTCAAGACGAGCCCGATGACCACCACAACCCCGAAGGGACTCGTCCA
GGTGCACAGCTCTCTGTTGAGGGGAGTCAGATGGAGCAGGCTAGGTAAGAGCCTTAAGCCCACTGTTCGAGACACACCCACTCCTCGACCCCTTCAGTCTACCTCGTCCGATCCATTCGGAATTC

KEY:
DUPLICATE REPORT FOR C16: A8 (21)
vs A8 (score = 126)

GCTGCCCTTTTCCCATGGTGCAAGTCTGCTGCTGGGCTACTGTTGGGTTGGGCTTCCCTGAGC

A8

AGGTGGGTCACAGCTCTCTGTTGAGGGGAGTCAGATGGAGCAGGCTAGGTAAGAGCCTTAAGCCCACTGTTCGAGACACACCCACTCCTCGACCCCTTCAGTCTACCTCGTCCGATCCATTCGGAATTC

A8

GCTCTAAGCAGCGACAAAAAGGTTTGGGCTAGAGTAGCTCTTATGGTCTCCTTG

A8 GCCTTAAGCAGCGACAAAAAGGTTTGGGCTAGAGTAGCTCTTATGGTCTCCTTG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,21,24,25

END OF RECORD.
SEQUENCE FILENAME: C2
FORWARD PRIMER: TCTAACTGCCCTGAAGCTCAC
REVERSE PRIMER: TCGTGCAACCCCGATACCCCA
LENGTH OF INSERT: 191

AGGGTCTCTAACTGCCCTGAAGCTCAGTATGCAGACCAGGCTGGCCTCAAACTCATAGAGGTCCAGCTG
BTCCTCGCTCCCTGATGGCTTTGATTCTGTCTCTTCCAGTTCAAGCAGAGTTATCGGGTACTGGGGTCAT
GGAGACGAAGGACTCAAGAAACTAAGACAGAGAAGGTCAAGTTCGTCTCAATACCCCATGACCCCAGTA
GCATGATGATTTCCGTATATGACTTTCAGATCAGCCCATGGGGCCACTTCCGGTTCACGAAGTCCGC
GTACTACGAGTACTAAAGGCATTAGAAAGTCTAGTCGGGTACCCCGGTGAAGGCCAAGTGCTTCAGGC
CTGCTGG

DUPLICATE REPORT FOR C2: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: C3
FORWARD PRIMER: AGAAGGTGGGGACAGGAGAC
REVERSE PRIMER: CCTGGATGGATGCCATGCACAA
LENGTH OF INSERT: 200

TGAACTCACAGAAAAGC2TAGAAGGTGGGGACAGGAGACCCACTAAGCATCACTACCCGGGACTCATGC
ACTITGAGTGTCTTTTCGAATCTTCCACCCCTGTCCTCTGGGTGATrTGAGTAGTGAGGGCCCTGAGTACG
-- - - - - - - - - - - - - - - - - - - - - - - - - - - ------------
TCCTCAAGACCCTGAAGTCACTCTCCACACCTACATATATGTAACACATGTGCACACTTGTGCATGGCA
AGGAGTTCTGGGACTTCAGTGAGAGGTGTGTGGATGTATATACATTGTGTACACGTGTGAACACGTACCGT
-------------------------------------------------------
TCCATCCATTATTTACTGCAAAAGCTTTGAAGAACTCAGTTTAACCCTCTAAGCTTCTGTCTC
<<<<<<<<
AGGTAGGTCCCGGTAATAAATGACTCGTTCAAATTTTTTCTGATCAAATTGGGAGACGAAGACAGAG
-------------------------------------------------------
ACGATTCCCGGCTGGTTGGAAGTAAGTATAGTCGGCCCCGTGAAATTCCCGGCCGTACCTGAAGGGGGT
o,°
ooo° ..
........................................................
TGCTAAGGGCCGACCAACCTTATATCACC
CCATTTTCCCTTTAGAGGGTCGTTTAGGGTTTTGGGGAATTAGGGATACTGCTAATGGGGGGGGTG
......... ,,, .. .......................................................
GGTAAAAGGGAAATCTCCCAGCAAATCCCAAAAACCCCTTAATCCCTATATCGAGACTTTCCCCCA
-----
----------------------------------------------------------
DGTTAAAAGGGAAATCTCCCAGCAAATCCCAAAAACCCCTTAATCCCTATATCGAGACTTTCCCCCA

DUPLICATE REPORT FOR C3: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,21,24,25
END OF RECORD.
SEQUENCE FILENAME: C4
FORWARD PRIMER: CCAGGCAGTGATGATAAAGGC
REVERSE PRIMER: GGGTTTAACCTGTAAAACATT
LENGTH OF INSERT: 361

AAGCTTTTAGCAGGGGAGTGATGTGATCCAAGGACAACTTGATGCTGTGTGGGATGAAAAGTACATGAA

TCCTTTAACCTTTTAGCAGGGGAGTGATGTGATCCAAGGACAACTTGATGCTGTGTGGGATGAAAAGTACATGAA

GAGAAATGGCAAAATCCAGAGGAGGCAATGCTCGGATGATAGACACACAGAGCATGC

<<<<<<<<<<<<<<<<<<<<

TTGACCGGTTAAAACAGGTGTTTCGGAACGTGAAGAAGAACTCGAGGTGTTCAGAACAACCCCACTGGA

<<<<<<<<<<<<<<<<<<<<

GACCAGGTTGTGAGCCCTCGAACCAAGTCAAGTAGTTTTACTTCGGGAACGACGGTTTCTACAGGAAAC

---------- ---------- ---- - - -- - - - -- - - -- - - - -- - ------"

DUPLICATE REPORT FOR C4: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.

SEQUENCE FILENAME: C5
FORWARD PRIMER: ACTGGCCAATTITTTGTCACCA
REVERSE PRIMER: AAGGACATCTTTGGCAGCAA
LENGTH OF INSERT: 292

GGCCCTAGTGCTGCTTGCTGGGTCTTCTCGTCGCTGTATGAGCTACGGCGGCCACTGAAGCTCTGAGCTGTTACCG

CTTCATCTGCTGCTCTCTCTTTGCGAACCTGATCAACACCTCTAGTATACCGACCAACACATTTCTGTA

GAAGTACAGGACAGAAGAAGAAGGCCCTTCCGACTAGTTTGTCAGATATACTGGGTGTTTTAAAGAAGCT

---------- ---------- ---- - - -- - - - -- - - -- - - - -- - ------"

TTGACCGGTTAAAACAGGTGTTTCGGAACGTGAAGAAGAACTCGAGGTGTTCAGAACAACCCCACTGGA

---------- ---------- ---- - - -- - - - -- - - -- - - - -- - ------"

GACCAGGTTGTGAGCCCTCGAACCAAGTCAAGTAGTTTTACTTCGGGAACGACGGTTTCTACAGGAAAC

---------- ---------- ---- - - -- - - - -- - - -- - - - -- - ------"

DUPLICATE REPORT FOR C5: <no duplicates>
NUCLEOTIDE BLAST HITS: ['T16883','NIB1963-5R Homo sapiens cDNA 5...',8.0e-10]
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: C7
FORWARD PRIMER: GCAGCCTGATGTTGACTCTG
REVERSE PRIMER: GGTTTCCATGACCTTAATAAA
LENGTH OF INSERT: 281

ATTAGGAGGCTGAGAGCCACTGCTGTAGAGGATGTAGGTACAAAAGGAGAAGAGGGAGGGGTAGGAGG
TAATCCTCCGACTCTCGGTGACGACATCTCCTA ATAT TTCCATCC TCC
AGGGGAGAGGAGGAGGAAGAGGAAGCAGCCTGATGGTGACTCTGCAOCCCTAGTTTATCCCTAAATCCA

<<<<<<<<<<<
TCCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCGGACTACCACTGAGACGTCGGGATCAAATAGGGATTTAGGT
AAGTTCTTTTCCCAATTAGCCTAGCCCAAGACATTTAGAGGGGCTCTATGTCTAAGTTTTAUAGCTTCTCC

<<<<<<<<<<<
ATAATTTCCAGTACCTCTGGGAAAACTCGGTAAAAAAAGAGTTTCCCCTGAACCCCGGGACAAAGGATTC

DUPLICATE REPORT FOR C7: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: EX101
FORWARD PRIMER: CAATGATGAAGAGGCCGTCA
REVERSE PRIMER: AGAGCTGGGATGTACAGGCG
LENGTH OF INSERT: 150

ACAATGATGAAGAGGCCGTCAGATGGACAGAGGAAGCAGAGGCAGAAACTGCGATCAAGAGGAAGGGGC

TTGTTACTACTTCTCCGGCAGTCTACCTGTCTCCITCGTCTCCGTCTTTGACGCTAGTTCTCCTTCCCCG
TTGGCACGTCCAACAGAAGACTATTGGAAGGAGTGGTGATGGGCGCCTGTACATCCCAGCTCTlGGTGC
AACCGTGCAGGTTGTCTTCTGATAACCTTCCTCACCACTACCCGCGGACATGTAGGGTCGAGAACCACG

AGACAGGATT

DUPLICATE REPORT FOR EX101: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: EX12
FORWARD PRIMER: TCTTTGTCCTCCCCATCCCCAC
REVERSE PRIMER: GGCCTTTCCTTTNGTTCT
LENGTH OF INSERT: 159

AAACAGAACCCTGGCTGCACGCGTTCCTTTGTCCCCCTATCCCAAGTTAGGCCACGAAAACCTTTCA

TTGGTTCTCCGGGCAAGGTGTCGCAAGAAAGCGGTTGGGTAAATTGTACTCCCCGCTGTTTGAANCT

AAGGCAACTCTGAAGTGCGACCCCTCCACCTCCGGCCACCCAGGTTCTTAACNGTNAAGAAGGAGCCCTG

TTCGTTGAGACTCTCAGTNTGAAAGGGTGGAACGGTGAGTCCAGGAAATTCTCTGNTCTTTCTTTCCGAC

TGAAGACAGCTCAAGCAACTG

ACTTCTCTGAGCTGCTGTTTAC

DUPLICATE REPORT FOR EX12: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: EX14
FORWARD PRIMER: TACGAATTTAAGCAGCAACCCA
REVERSE PRIMER: GTTGAAAATGTTCTCCCAGC
LENGTH OF INSERT: 148

AGCAGTACTGTGAGTACGAATTTAAGCAGCAACCCAGCCAGGAGGAGTGCGAAGGCAGCTCTCTGTCAG

TCGTCATGACACTCATGCTTAAATTCGTCGTTGGGTCGGTCCTCCTCACGCTTCCGTCGCGGTCAGAGACAGTC

CCGCAAAATCTGGAAAGGGAAGCAAGCTGCTGACCCTCTTGTAAAAGTTGAGGAGAACGACCCCGAGATCCTCCG

TCTAGGAGGC

DUPLICATE REPORT FOR EX14: <no duplicates>
NUCLEOTIDE BLAST HITS:

['M77123', 'Mouse vitronectin mRNA, compl...', 1.9e-25]
['X63003', 'M.musculus mRNA for vitronectin', 2.0e-25]
['X72091', 'M.musculus gene for vitronectin', 7.1e-25]
['M59442', 'O.cuniculus domesticus vitron...', 5.0e-06]
['X03168', 'Human mRNA for S-protein', 3.1e-05]

PROTEIN BLAST HITS:

['P29788', 'VITRONECTIN PRECURSOR (SERUM SPRE...', 3.7e-16]
['M71223', 'Vitronectin [Mus musculus]', 3.7e-16]
['J700662', 'Vitronectin - Mouse-gi 44166 gp X72...', 3.8e-16]
['X03168', 'Human mRNA for S-protein....', 8.2e-13]

PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: EX25
PRIMERS NOT PICKED
LENGTH OF INSERT: 81
--------------------------------------------------
AACCGTGCAGAGGGCAAAGTGCTGGAGACAGTTGGTGTGTTTGAGGTGCCAAAACAAAATGGAAAATAT
TTGCAAGCTTCCTCGTCTACGACCTCAGACGAACTCTCAGACGAGTTTTGTTTTACCTTTTATATA
GAGACTGGGACAG
CTCTGGCCGTC
--------------------------------------------------
DUPLICATE REPORT FOR EX25: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
--------------------------------------------------
SEQUENCE FILENAME: EX6
FORWARD PRIMER: AAGACCCTCTGATGCAGGAG
REVERSE PRIMER: ATGGGAAAGAAAGGCAGCT
LENGTH OF INSERT: 144
--------------------------------------------------
TTCCCGACACCTGCTGGGAGTGGACTCTCAAAAGGAGCTTCAAAAGGACCGCTCTGATGCAGGAGCCCTCAAAGTGTGG
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>> >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>iores
SEQUENCE FILENAME: EX69
FORWARD PRIMER: ACTCAACCCAGACCCCAAGT
REVERSE PRIMER: AGCGGGGTTCAGCGGAG
LENGTH OF INSERT: 246

AATCACTTCAGTTACTCAACCCAGAACCCCAAGTCCACGATGTCCACTGGGACATTGAGGACCTGTCC
TTAGTGAAGTCAAATGAGT± T G TTCAGGGTGCTACAGGTGACCCTGTAACTCCTGGACAGG
AGCGCTrTGTGCAGCCTTTCCTGAATCGCCTCAGTGTTGCGGGCAATTTCTCTGTGGGCTCTCAGATCC
TCGCGAAACACGTCGGAAAGGACTAGCGGAGTCACAACGCCCGTTAAAGGAGACACCCGAGAGTCTAGG
TATACTATGCCACTGTCGGCGTGAACCCCCGCTTTGACCCAGCCTCATCACTACTCCTGGCTATGCAC
ATATGATACGGTACGAGCCGCACTTGGGGGCGAAACTGGGTCGGAGTAGTGATGAGGAACCGATACGTG
AGCCTGCCCCACGTCATCAACCCAGTGAGTCCCGACTCG
TCGGACGGGGTGCACTTAGTTGGGTCACCTCAGGGCTGAGC

DUPLICATE REPORT FOR EX69: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,14,22,23,27
END OF RECORD.

SEQUENCE FILENAME: EX76
FORWARD PRIMER: GACTGGAGGCAAGTCAGTGG
REVERSE PRIMER: GCCGGTTCCAGCTTCTCCACGT
LENGTH OF INSERT: 289

TGTACATTCCCGCTGCCCTGACTGGAGGCAAGTCAGTGGGGACACCCCAGATGTCTTTATCAGTTACCG
ACATCCTAGGGGGAAGGCAACGGTGACTCCTGGGCTCTGCGTACTGTCCGTGGGGGGGCCCTGAGAATGCA
GAGAACTCAGGGTCCCAGCTGGCCAGCCTCCTGAAGTTGCACCTGCAGCTTCACGGCTTCAGCGTCTT
CTCCTTGAGTCCCAGGGTCGAGCCGGTCGGAGGACTTCAACGTGGACGTCGAAGTGCCGAAGTCGC
AAAACAGGACCACACGACAGACCACCCCGAGACCCTACCTCAGCTCGACTGAGTGACGTACGCAAGC
TTGTTCTGGACCT

DUPLICATE REPORT FOR EX76: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,21,24,25
END OF RECORD.
SEQUENCE FILENAME: EX90
FORWARD PRIMER: CTTGAAGTAAGGGAAGTGC
REVERSE PRIMER: ACATCCTCATCTCATGGCC
LENGTH OF INSERT: 91

CCTCATCTCAGGCCCTTAAAGACAGTAAGACCGGAAGGCTTCCAGTCAGTGAAATCTACAATTTCAT
GGAGTAGAAGTACCGGGAATTCTTGTCATTGCCITCGGAAGGTCGTCACTTTAGATGTTAAAGTA
GACGGAGTGCATTACTCCAAG

CTGCCCTCAOGGTAATGAAAGTTCC

DUPLICATE REPORT FOR EX90: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
['A42826', 'HTLF-T-cell leukemia virus enhanc...',0.039]
['P32314', 'HUMAN T-CELL LEUKEMIA VIRUS ENHAN...',0.039]
['P14734', 'FORK HEAD PROTEIN->pir A32380 A32...',0.041]

PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.

SEQUENCE FILENAME: EX92
FORWARD PRIMER: TGGGGAGACTCTAGGAAACTCT
REVERSE PRIMER: GCCAGGTGCTCAGGAAGA
LENGTH OF INSERT: 169

AGAACTGCGGGAGACTCTGAAGAAACTCTTCATCAAATATCCCGAGCTTTACAGGCTACTACAGGAGAC
>>><<<<<<<<<<
TCTTGACCTCTGAGATCTGAGATCCT
AGTATTAGGGTCCAAATGTGCCGTGATAGGTGTGACTCTC
AATTAGGACTTCAGGAGAGCTCTTCCATCAGGCGCTTGGGACTGAGACG
<<<<<<<<<<<<<<
TTAATCCCTAGGCTCCCTGCGGAGGCAGCTGGATACACAATCTTCTCCAGTGGAGACGACTCTC
TGCCATGGGTACGCGTACTCCTGCTGATCTG
<<<<
ACGTTACCAGTCGGACAGGAGACGGACTAGAC

DUPLICATE REPORT FOR EX92: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: EX98
FORWARD PRIMER: TGACTGCTTTTAAGCTTGCC
REVERSE PRIMER: GTGGCCTCTTGCTAATG
LENGTH OF INSERT: 245

GACCGCGCAGAGATTGTGACTGCT'TAAGCTGTGGCA
>>>>>>>>>>>>>>>>>>>>>

CTGGCTGTCGCTACACGACCAAGCTTCACTTTCTTTACTGCAG

TGGCTAAGGCTGCTCTCAGGGGTATGCGAGGACGCTGTAATCGCTTCAACGAGCTCATC

ACCGGACTTCCGATCGGACGCGACTTCTTCCATGTGGCACATGAGTGGAAAGTTGCGCTAAGG

ATGAGTACCATGAGCGGACCATGAGCATCTCGCGCTTTGTAGGGGAGGCGCGCGCTCGAGCTCTCT

TACCTAAGCTCCGCTGCTGTAGCTCTAATCTAAGCCGAGGAGTCCCGGCGGAGAGCTCCCT

CGGGATCGGATACCACTTGGAGGAGCCTACGCAATGG

DUPLICATE REPORT FOR EX98: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14, 21, 24, 25
END OF RECORD.

SEQUENCE FILENAME: EX99
FORWARD PRIMER: TACTCCATCATCCTCATGGCG
REVERSE PRIMER: GGAAGAGGATGATGGGGAAG
LENGTH OF INSERT: 124

GCCTACTGCTTACCCATCATCCCTCATCGGCGCTCGCTGTAGTACAGAGCGGCGGGCTGCTCTG

CGGATGACACGAATGAGGTAGTAGGAGTACCGCGACGACACACATGTCTCCGGGACGGGAACCGACATC

CGCCCTCTCCACTCCATCATCCTCCTCCTCCTTTGTAGTACAGAGCGGCGGGCTGCTCTCAAG

CGGGGAAGGGGTAGTAGGAGAAGGGAAACTACCCATAGTACCTTCGGAGGTTC

DUPLICATE REPORT FOR EX99: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11, 12, 13, 23
END OF RECORD.
SEQUENCE FILENAME: a10a
FORWARD PRIMER: TGCTTTGTTATGCATCACCAG
REVERSE PRIMER: ACACCTGTGAGCACATCTA
LENGTH OF INSERT: 180

TCGTGAGACCCAGGGTGTCCTCAGTGAAGATCATCCTTGGTATGCAATCCGCGGCTATCCAC

AGACTCTCTGGTCACAGGGTGACATCTCTTTCAGAAGCAGAAACATAGTGATGTTCGCGCAGTAGGATG

GAAGTAGAAAGCTGTGACCTGGTTCTCACATAGGCGCTGAGAATCATCCTGGAGAGAG

CTCTCTTCGTACATACAGAGTGTCGCTGACCTGGTAAAGGACAGGGATTGGACTAGATCAGTGCCAGTTCTTCCCGTGATGAGAC

DUPLICATE REPORT FOR a10a: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.

SEQUENCE FILENAME: a10c
FORWARD PRIMER: AGGAGTGCTCTCCTCTGAGG
REVERSE PRIMER: GGCTGCCTGGAGGCTTAAGG
LENGTH OF INSERT: 487

ATTCGTGAGACCGACATTTCCTGTCCTCCTAACCCTGATGCTACGTCAGGCGTTAGAAAGGGCACTACCTTG

TAAGCAGCTCTGGCTGTTTAAAGGACAGGGATTGGACTAGATGTCCTGGACACTCTCTCCTCCGCTGAGAC

AAACTCAAGGCCCAGTGACCTGGAGGAGGCTCCAGGACCTCCTGTCCTGACAGGATCTCCGTCGAAACAT

TTTGAGTCCGGGTACTCGTCACTGGCTGGCTGGAGAGGAGAACGTCCTCCGGAAGGACTTGTA

GGGACAGCCAGCTCTCCCTGCTGCCCTCCACTGCTCCAGAGGTTGCTAGGATGCTGACAA

CCCTGTCGGCTGGAGGAGGAGGACAGGGAGTGGATGCGTCTCACAAGGATCCAGGCACGACCTGGTTTCTG

CTTCTCTGACAGCTCTCAAGGTCAGGGCCAGGCCCCGAAATCTTTGCTGTGTAACCAAACACTTT

GAAAAGAGACTCTGACTAGTCTCCAGGTTGGGCTTGAAGAAAACAGCAACTTGGTTGTGAAA

ATAAATCCTGGAAGTTGCTGAGGCAAGGCAAGCTCTCAACGGNTTCACAGGNTNATGGNTG

TATTTAATCGACCCCTCCACACCCCGTCTCCCTGTTCCAGAGCTGCAANAAAGTCCNANTACCCNAC

DUPLICATE REPORT FOR a10c: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: a10d
FORWARD PRIMER: CAATCAGAATGGCTGGAATG
REVERSE PRIMER: TGNCTGGTGGATAGTCTTG
LENGTH OF INSERT: 290

GGAATTCGTGAGACCGAGACAGGAAACTTCTAGAAAAATCAGATCCATGAGCAGGCAAGGTCAGACCTTAAGCACTCTGGCTCTGTCCTTTGAAGATCTTTTTAGTCTAAGGTACTACTCGTCCGTCTCAGGGAAGGAGGCC

GCGGAGTAAAGGTTATGCATTACGGACCTCTTCTTCTCTCAGGTGCTGCTGCTCCCGG

TTATGAAGTTCAAGAGTGGCCGTGTAGATGGNGTTGAAGATGCCGGGAGCATTCAACATGGNCAAGCTAAGTACTTCAAGTTCTCACCGGCACATCTACCNCAACTTCTACGGCCCTCGTAAGTTGTACCNGTTCGATATTCACACAGGNCACAGCAGAGGAACAGAGCACCNTGNTCCGACTCGCTCACCNCNTTTNGAAAGGTG

CGGGAGGTAATGGGTTAGTCTTACCGACCTTACTTCTCCTCCGGTCGAGGGTCCTCCITCCCCTTCGGG

TAAGTGCTCCNGTCTGCTCTTGGNACNAGGCTGAGCGAGTGGTNGAAANCTTTCCAAC

DUPLICATE REPORT FOR a10d: <no duplicates>
NUCLEOTIDE BLAST HITS: 
['Z34286', 'O.cuniculus (AdRab-G) mRNA expressed i...', 8.9e-26]
PROTEIN BLAST HITS: 
['Z34286', 'O.cuniculus (AdRab-G) mRNA...', 4.1e-20]
PHYSICAL MAP LOCATION: 14,21
END OF RECORD.

SEQUENCE FILENAME: a10e
FORWARD PRIMER: TTTTGGACCTGCGTCATTTG
REVERSE PRIMER: GCATTTGGGAGGCACCTG
LENGTH OF INSERT: 248

AATTCGTGAGACCAAATGGGGAAGCCTTGGTCAAGGCGGGTTIGGACCTGCGTCATTTGCTGCTCATC

TTAAGCACTCTGGTTTACCCCTTCGGAACCAGTTCCGCCCAAAACCTGGACGCAGTAAACGACGAGTAGTCTCTGCTCCCTCACTCCCCACCCCCACTCAGTAAAAACTGAGAAACCTCCTGTCGACGAGCACCATTGTTCT

AGACACGAGGAGGAGGGGTGGGTTGAGGCAGTCATTTTGACTCTCGTGACAGCTCGCTGTTACCCA

TTCTGTCGGGCCCCACAGGGTGCTCCTTTTGGAAATGTTGCTCGCCCAGGGCACCTCCGCTCCAAATGCTGCTCTCTCA

AAGACACGAGGAGGAGGGGTGGGTTGAGGCAGTCATTTTGACTCTCGTGACAGCTCGCTGTTACCCA

GCACAAAGATGAGGAGCTCATCAGTACCTGGGAGGACTGACCCAGATTCAGTCTCATCCAGTACAGTTCN

CCTGCGTTCTCCTAGCTGAGTACTGACCCTGAGCCTCTGCTGCTGCTAAAGAGT TTCCCCAAGAGTGCNTNAAGN

DUPLICATE REPORT FOR a10e: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: a10f
FORWARD PRIMER: CCACTCCCCCTTACCCCG
REVERSE PRIMER: CAGCTGACAGACCTTGCCCTCT
LENGTH OF INSERT: 231

GCGGAATTCGTGAGACCCACCACTCCCCCTTACCCCGGACGCTTTTACAAGTGGGATCAGC

TCAGGGAGGGAGAAGCTAAAAACCGTTGAAAATCTCTAGAGAGAGCTACGAGGAGTTGGGCAAG

AGTCCCTCCCTGTCGATTGGACATTTATGAAGGATCTCTCCTCGACTCCAGTCCTCTCCTGTTTC

GACTCTGATACCTCTCTACTGACTCCACACTCAGTCAAGCAGCAGACTGACCTTGGGCCAGAGTG

DUPLICATE REPORT FOR a10f: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 21,24,25
END OF RECORD.
SEQUENCE FILENAME: a10g
FORWARD PRIMER: GCTACCCTCTGCCCATTGCT
REVERSE PRIMER: GTCTTTAACGCCATGAAAG
LENGTH OF INSERT: 304

AATTCGTGAGACCTATGTACTGTTCTTCTCAGCCTGCCTTCCATCAGTACTCCCCGGGTGGAGGCAGCT

TTAAGCACTCTGGAGACGTGGAAGAAGGTGTGAGACGTGGAACGTGGAAGAAGGT

ACCCCTGCTGCCCTACCTGGGCTCACCTCACTATCCCTATCAGAGGATTGCACCCNNAGGGNCACGNNCGAAG

TCACCAGCCACCTCTTNCCAAAGNCATCTACTTCTACAGCATCCTCTTCAAGCCTTTAAGACAG

AGTGCTGGTGGAAGANGTGGTACAGTGAATGCTCTAGTGAAGTGAAGTACNNGGAATTCCTGTC

TAAGGCGGAAAGCCTTCCAGCTCAAGTGAATTTTACTAATTTTCTAGCAGCGGACACTTNCCCTTACTT

ACCCCTGCTGCCCTACCTGGGCTCACCTCACTATCCCTATCAGAGGATTGCACCCNNAGGGNCACGNNCGAAG

CAAGGCTGGTGGAAGANGTGGTACAGTGAATGCTCTAGTGAAGTGAAGTACNNGGAATTCCTGTC

GTTCACCTCAANATTGTTGGGTGGAACNGGGGTTCCTAGCTCCACAGAGTGTCCNATAGGGCGGCAACACNC

DUPLICATE REPORT FOR a10g: EX90 (5)

vs EX90 (score = 74)

NCCTGCATGCAAGCCTCNACGTGTAAGCTTGNQAATCCTCTAGAGCGGCCNNTACCTACTACTACTGA

GGGAAATCCTGGAGAAGCCTATGACTGTTCTTCTCAGCCTCTTCCACTCAGTACTCCCCGGGTGGG

AGGCAGCTACCCTCTGCCCATTGCT

NCAACGNCGAAGCTACCCAGCGCCTACTTNCCAAAGNCATCTACTTCTAACGATCCCTACTCTTGCA

TCCTGTCTCA

TGGNCTTTAGGAACAGTAAGACGCCGTCTCCAGTCAAGTGAAATCCTTTACANTTTTTCATGAC

EX90 TGGGCCCTTAGAAGACAGTACCGCCCTACCCATAGCTGAGATGTTCTGGGCAGTCTCTCAGCTGAACT

GGGAGCACTNCTTACTGCCAGGTGAGTNTAAAACCCACCCCTTGNCCCCAAGGCCCAGGGTC

NUCLEOTIDE BLAST HITS: <none>

PROTEIN BLAST HITS: <none>

PHYSICAL MAP LOCATION: 11,12,13,23

END OF RECORD:
SEQUENCE FILENAME: allg
FORWARD PRIMER: CTTCTITTCAAGGGGACCA
REVERSE PRIMER: ACCTTGCAATGAGGCACACT
LENGTH OF INSERT: 388

ACGTNGCATGCACGGGCTGACGTTAAGCTTCTGGAGATCCCTTAGAGCGGCCGCCTACTACTACTACTA
TGCACTCGAGACCAGCTGACTCTCTGACATGCGGAGAAGTCACGCCCTGGGAAGAGGTTACCA
GGGACCACCCCTTTCATGGT

C-NTGTCTCACACGGAGTAACGTTCCAACAGAAGTTCACCGGTGTAATCCGATCCACCCCGTTT
GNGAAAACCTTTNAGSMARTCCCTGANCGCGGCMCAGMTATACCGCCCTNAAACGTTGTTTNCQGAGGA

DUPLICATE REPORT FOR allg: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: al2b
FORWARD PRIMER: TGGATAAGTGCATGCAGGAC
REVERSE PRIMER: CCCAGGCAGCGCTTGAAGCTC
LENGTH OF INSERT: 384

AATTCGTCGAGACCAGACCGAAGCTTCGAGGACAAGCTTATCCAAAG
TGTCATAGCGGCTCGCAATTTTGTCCTGGTGCTGTCTGCTGGGGCGCTGGATAAGTGCATGCAGGAC

--------------

EX76 (score = 133)

G-AGCGGAATTCGT-GA-GACCAGAACTGGGAAGCCGAAATTCGAGGACAAGCTTATCCAAAG

EX76 TCAGC-GTCTCTAGCTGAGCTGGAGAAGCTGGAA-CCGGCAAATTCGAGGACAAGCTTATCCAAAG

EX76 GCTGCA--GGCTGCAATTTTGTCCTGGTGCTGTCTGCTGGGGCGCTGGATAAGTGCATGCAGGAC

ACCATGACTGCAAAGGACTGGAAGATGACTGCTTTAAGCTGT冈CAAGAACATTGTGCCCATCATTGAGGCTTT

EX76 -CCATGACTGCAAAGGACTGGAAGATGACTGCTTTAAGCTGT冈CAAGAACATTGTGCC

EX76 CATCATTGATGGCTITGAGTGGCCT-GAAGC-CTA-GCGCTGCCTGAGGATATGCAGGCTGTACT

EX98 CATCATTGATGGCTITGAGTGGCCT-GAAGC-CTA-GCGCTGCCTGAGGATATGCAGGCTGTACT

EX98 CCCTTCACCCACTGAGGACAATGG

GAGGACTACGCAATGG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,21,24,25
END OF RECORD.
SEQUENCE FILENAME: al2c
FORWARD PRIMER: AGAGGTTTGGCTTAGGGATGTC
REVERSE PRIMER: TGCCCCCTAGGCCCTCGGCC
LENGTH OF INSERT: 228

CGGAATTCGTGAGACCGACTCCCCTGCTGCCCCTCCCACTAGCTCCAGAGGTTTGCCTAGGGATGCTG

CTGTTGAAAGGAGAGCTCTGACTGAGCTCCAGGGGACGGGCTGATAGGTTACCTCCTCTACA

CGAGAGTTACAGCACAGGCTCAGGCCCAGCTCAGGAACTAGGTCAGGTGCCCTCCTC

DUPLICATE REPORT FOR al2c: al0a (6) al0c (13)
vs al0a (score = 138)

GTCAGGGGTTTACCTAGGATCTGACGACTCCCCCTGCTGCCCTCCCACTAGCTCCAGAGGTTTGCCTAGGGATGCTG

GATGGATGACGACTCCCCCTGCTGCCCCTCCCACTAGCTCCAGAGGTTTGCCTAGGGATGCTG

vs al0c (score = 102)

ACGTGCCGATGTCAC

GACCATTTTCCGTCCCTAACCTGATCTAGCTGACGACTCCCCCTGCTGCCCTCCCACTAGCTCCAGAGGTTTGCCTAGGGATGCTG

GCA--AGGGGA--CATG--TTT--ATTGACGACTCCCCCTGCTGCCCTCCCACTAGCTCCAGAGGTTTGCCTAGGGATGCTG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: a12d
FORWARD PRIMER: TCTCTCTGCTCCCTCACTCCC
REVERSE PRIMER: AAGAGCAGNCATTTTGGGAG
LENGTH OF INSERT: 400

GCTCCCAAGAAACACTTTCAACCCACAGCTCCTTCTCAGAGATCTTGGAACCTCAGACATGAAAATGGGAA
CGAGGGTCTTTGTGAAGTTGGTGTCGAGGAAGAAGTCTCTAGAACCTGAGGTCGTACTTTTACCCCTT
GCCTTGGTCAAGCCGGGTTTTCGCCACCTGCTACATTTCCTGCTGTCTCTGTCCCTCTATCCCCACC
CCACACTGATTEDAATCTGAGACCTTGCTGACGACCTCTGCTGTGCCCCACCACAGGTTG
GCTGGAATTTTGATCTTTCTTGGGGAAATTGCCTTAAA

DUPLICATE REPORT FOR a12d: a10e (23)
vs a10e (score = 206)

AGAGCGGCCACTTACTACTACTACTACTGAGCGGAATTCGTGAGACCCAGAGAGGGGACTAAACT

PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: a7b
FORWARD PRIMER: CTCGGGTAGTTTCTTGCC
REVERSE PRIMER: GGACCATTTCCTGTCCTGGCC
LENGTH OF INSERT: 153

TCCTCTAGAGGGCCGCTAGTACTACTATCTGAACGGAAATTCGGAGACCCCTGCGGTAGGTCTTGCC
AGGAGATCTCGGGCAGCTAGTACTACTATCTGAACGGAAATTCGGAGACCCCTGCGGTAGGTCTTGCC
CCCTGATGGAGGCCGCCAGGGCTCAGAGCAGGCTCAGACTGTGGACTTTGTCTGTNAAACCGCAACT
GGACTACCTCTCCGCGTCTCGCCTCCTCCCATGGGATCGACCTAGAAGACTACTAAGACGAGTGCTTGGA
ATAAATAGCTGAGGTNGTNGCAGAGGACGGAATTTGCTTCCCCTAATACGTGTACGTGCAGGTGTC
TTTTATCGGACCTCCACANCGCTCTCCCCCTCTTTTACCCAGGATTTGAGATACGTGCAGCCAG

DUPLICATE REPORT FOR a7b: <none>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.

SEQUENCE FILENAME: a7c
FORWARD PRIMER: GAAAAGAACAGAGGCCAAGGC
REVERSE PRIMER: CGGGTTAAGACCATGAGCGCCCC
LENGTH OF INSERT: 271

TGAGCGGAATTCGGAGACCCGGCATATTCAGGAATAAAGTGTCTTCTGCTCTTTTAGAAGAAAGAACA
ACTCCGCTTAAGCGACTGGGGGCTAAATAGGTCTTATTTCACAAAAGAAGCCAAATTCCTTTTCTGT
GAGCCCAAGCCAAAGGGGTGGTGGGACACAGCAPCAGAGGTGAAAAGAGACACGAAAGAGAGCT
CTCGGCTCTGCTTTTCCACACCCACCTGTGCTGTCCCTCTGCTCTTCTCCAG
CTGGGCTCAAGAGAACAAAGAAGACACCAAGACCGTCAGGGCTCAGGTCTTTAACGOCTTACAC
GACCGAGCTCTCTGCTTGTCTTCTGCTGTCTATCAGCCGGTCCCCGCAGTACGAAATTTCCAGTGT
TCTGCTCTGCTTTTATAGCGGAGTCCTAGCTGCTGACCTCAACCCATTCCTGATTATAGTGTGCA
AGAAGCGACGCAAAAAATTCAGTTCAGGAGTACATCGGATNCCGATTGATTAAAGGATAGACANCGT

DUPLICATE REPORT FOR a7c: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: a7e#2
FORWARD PRIMER: TTTCCTTGAAGCAGGAATGGGCT
REVERSE PRIMER: GTTAAGGGTCACCTGACAC
LENGTH OF INSERT: 193

GCGGCCGNNACTACTACTACTGACGGAATTCGTGAGACCCTCTCTGATCCTTTCTTGGAGC

CGGCCGNNNATGATGAGTATGATGACTCTCCTTAAAGCAGCTCTGGGAGAGAGCTAGAAAG

ATGGGCTCCCTGGAATGCTGACATGATTTCCTCTTTGGAAGGTTGTTGTTGTTGAAGGAT

TCGGCAGGACCTTGAACCTCGTCAACTCAAGGAAACTTCTCCACACCAAGACTTTCAAGAT

CATCAGAATACTCCTAGTTAGCCACAGCCTCTTTTCACTCTTCACATTGTCGAGACCCTCCT

GTAGCTTCTATGGAGAGTATGGGCTCCGTGGAACTGGCATGATAGTTCCCCGACAAGGTGTGT

TACCCGAGGCACCTTGACCGTACTA

CATCAGAATACTCCTAGTTAGCCACAGCCTCTTTTCACTCTTCACATTGTCGAGACCCTCCT

DUPLICATE REPORT FOR a7e#2: <none>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.

SEQUENCE FILENAME: a7f#2
FORWARD PRIMER: TCTGATATGGCCCGTGCA
REVERSE PRIMER: GGCACTGGCTTGTTGGAGTG
LENGTH OF INSERT: 394

GGCCGNNTACTACTACTACTGAGCGGAATTCGTGAGACCCTCTCTGATCCTTTCTTGGAGC

CGGCCGNNNATGATGAGTATGATGACTCTCCTTAAAGCAGCTCTGGGAGAGAGCTAGAAAG

ATGGGCTCCCTGGAATGCTGACATGATTTCCTCTTTGGAAGGTTGTTGTTGTTGAAGGAT

TCGGCAGGACCTTGAACCTCGTCAACTCAAGGAAACTTCTCCACACCAAGACTTTCAAGAT

CATCAGAATACTCCTAGTTAGCCACAGCCTCTTTTCACTCTTCACATTGTCGAGACCCTCCT

ATTGAGTCCCCAATCGTGTCCTATTTACCAGAGTGCNTAAGGCCNCCTNCCNCNNNN TNNCNNNAA

DUPLICATE REPORT FOR a7f#2: <none>
NUCLEOTIDE BLAST HITS: ['X59677','Rattus sp. cDNA for M2 gene (clone M2-...',8.4e-35]
PROTEIN BLAST HITS: ['S36784','mucin - rat',2.7e-21]
['A47714', 'Na+/sulfate cotransporter, renal - r...',6.0e-09]
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: a7g#2
FORWARD PRIMER: AGCAGGTTGCTCTGTTCTCT
REVERSE PRIMER: TGGAATGAAGGAGGCC
LENGTH OF INSERT: 306

GCGGAATTCGTGAGACCAGCTTTCCAGGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTG AGTACTGAGGGGTTCCATGGATGAGTGGCTTGCTCCGGATCTTCAACTCCATCTACCGCCACTCTTT

DUPLICATE REPORT FOR a7g#2: ABX7_3 (12) a10d (-16)
vs ABX7_3 (score = 97)

vs a10d (score = 167)

NUCLEOTIDE BLAST HITS:
['Z34286', 'O.cuniculus (AdRab-G) mRNA expressed i... ', 1.7e-41]

PROTEIN BLAST HITS:
['Z34286', 'O.cuniculus (AdRab-G) mR... ', 1.0e-25]

PHYSICAL MAP LOCATION: 14,21
END OF RECORD.
SEQUENCE FILENAME: a7h
FORWARD PRIMER: GCTCCCACCACCATCAAG
REVERSE PRIMER: CTTTACACAAAACCTCTTATCA
LENGTH OF INSERT: 368

ATAGGCCCCTAACGAGGGAGGGCGGCTGCTGATTCTCAAGATGGATGAGACTCTAATTTGATGC
TATTCCGAGATTGTGCTCTCCGAGCCAGAATAGCTCTCTACTTCATCGACTCTTGAGATATACAG
ACCCCTGGAAGAAAATCCTGCGCTGATCCAAGAGGAGCTACAGTACGTGAGG
TTGGGACCTTTCTTGAGGACCAGACTCTGTTCTCCTGATGTGCTTACCTGTTACCCCTGCTGAG
TGCTTTCCTGTGCTGAGCCAGACACGAGGTTGGAACCATTTGGAAGGCTTTTTTCTGACTAG
ACCCCAAAGGCTTCATCTGGAGTTGATGGAAGATGCGCAAGGAGCCTCAATTTTNGAAAGTA
<<<<<<<<<<<<<<<<<<
GGGGGAAGGGGGAAGGGAACTATTCAAAAAACCGATTTCCTGATCCTTAACCCGGATAAAANCCTTCAT

DUPLICATE REPORT FOR a7h: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.

SEQUENCE FILENAME: a8a
FORWARD PRIMER: ATGCCCCGCTATATGTGG
REVERSE PRIMER: GTCCCAGCNAGAGGGGTGCG
LENGTH OF INSERT: 287

CGGAATTCGTGAGACCAGCCAAAGTCCACCCAGGCTTTTCNAACTCTCCCGCATTAGATCTTCTCCCG
GCTTAAAGCTCCTCGTTTCGTGTTTCGTGAGGAANCTCAGCTCCGCCATCATCGATCCATTGCCCGG
CTATATGTGGTTCCAGCCCCAGAAAAGTCCAGGATGCATCTAGCCCTGTACTCTGGCCATATTCGTG
GATATACACCAAGGTCGGGGTCTTTTCAGGTCCTACGTAGATCGGGACATGAGAAACCGGTATAAGCAC
GTGGTTACTGTCATTTTGGGTCAGATACTCTGACACTCTAAG6CTGAGGCGCACCCCTGTGNGG
<<<<<<<<<<<<<<<<<
CACCAATGACACGTAAAACCCCAGTTCTATAGGACTGTGAGATACTCGACTCGCOTGGGGAGANCGACC
GACTCCAGCTGATTTGCCTGCTACGACGGCCACCACACCCTACCTACCTACACCAGCCTCTTGGGCTA
CTGGAGGGTCACTACTGTGAGGAGTGCATCCTCGATGGATGGGATNGAGTGGTCGGAAGACCCGGAT
AGGNTTCTGCAATTCTGCCTGGCCGACGGTCTCACGNATNCCGGNTTTTGNGNNGTAGG
TCCNAAAGACGTTAAGCGACCGGGCTGCCAGACGGCAGGCTGCTGNTANGGCCNAAAACNCAACACGN

DUPLICATE REPORT FOR a8a: <no duplicates>
NUCLEOTIDE BLAST HITS:
['Z34286', 'O.cuniculus (AdRab-G) mRNA expressed i...', 4.2e-42]
PROTEIN BLAST HITS:
['Z34286', 'O.cuniculus (AdRab-G) mR...', 1.3e-36]
PHYSICAL MAP LOCATION: 14,21
END OF RECORD.
SEQUENCE FILENAME: a8b
FORWARD PRIMER: CTGATTTTTGACAATGTGTG
REVERSE PRIMER: CTGCTTTGCCCCGAGAT
LENGTH OF INSERT: 249

ACAAATGTTACTCCCTCCAAGGTTATTGGAAGGGTTTGGGACAGTGGTTTGG

TTGTTACAATGGAGGGAGTTCCAATAACACTTATTTCGGTCAGAAGACTAAAACCTGTTACAAACC

CAGACAGTGCACAGTCAGAAGACTTTGGAACATCAGAGAGGGGAGAATGTAAAGATTTCTCTCTAGGA

TCGCCCGTTCNACCTCCTCTTACATTTCTAAGTGGAGAGACCCGAACATGGTTAGAGAGATCGCTCATGCTCAGCTCTGAGGTCTTCATTCCATGGAAAGTCTGTCTCCTGAAGTGAAAGCTAAATTCCTGAT

DUPLICATE REPORT FOR a8b: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.

SEQUENCE FILENAME: a8c
FORWARD PRIMER: ATTACAACCTCTCTGCGTTG
REVERSE PRIMER: TTCCCAAAATCATTTGCGTT
LENGTH OF INSERT: 280

TNGAGCGGAATTCGTGAGACCGGGGCTTAGGGTAGAGGGGGCTTGGCTGTAGGATGGANCTCGCCTTAAGCACTCTGGCGCCCCGAATCTCCCATCTCCCACCCCGAACCGACGGACATCCTACCAGCGGCAGNGGGGAGAATGTAAAGATTCACCTCTCTGGGCTTACNTCAAATCTCTCTAGGA

AAAACGCAAATGATGAAAGCTCCCCGTAAGTCCCTCACAGTCAAAGTACAATGGAAC

TMGCGTTTACTAAAACCCTTGCAAATTCGGAGGGGTCAGGGAGTCAGTTTCATGTGTAAAACCTTGCAGACCGAGTTTTAGAGGACTTCACTTTCGATTTAAGGACTA

DUPLICATE REPORT FOR a8c: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: a8h
FORWARD PRIMER: CCTGTGCTAACCCCTGAGTT
REVERSE PRIMER: CTCCGTGGAAACTGNCATGAT
LENGTH OF INSERT: 194

GGAATTCGTGAGAACCATTTATCCTGTGCTAACCCCTGAGTTAAGGGGTCACCTGACACTATGTTGAAGA

CCTTAAGCACTCTGTAAGATAGGACACGATTGGGGACTCAATTCCCCAGTGGAGGTACAACTTCTGCTAAGGAGGTGCCTGGGCTAACTAGGAGTGTATTCTGATGATTAGCAGTACCTTC

ACGATTCCTCCAGGACCCCGATTTGATCCTCACAATAAGACTAATAATCGTCAGTGAAGTTGTTGCTGATGAGTTG

CTTGTCAAGGGAAACTATATCATGNCAGTTAAGGTAAGGTAAGGGTCACCTGACACTAT

GAACAGTATCCCTGATGATAGGACACGATTGGGGACTCAATTCCCCAGTGGAGGTACAACTTCTGCTAAGGAGGTGCCTGGGCTAACTAGGAGTGTATTCTGATGATTAGCAGTACCTTC

DUPLICATE REPORT FOR a8h: a7e#2 (-33) vs a7e#2 (score = 188)

GAGCGGAATTCGTGAGACCATTTATCCTGTGCTAACCCCTGAGTTAAGGGGTCACCTGACACTATGTTGAAGA

GTTGAAAGATCTGCTGAAACGGGCTGTTAACGACTAGTTAAGGGGTCACCTGACACTATGTTGAAGA

AAGCACACACACTTTGTCACAGGGAACATCATGAGCAGTCCAGGAGCCCATTTGCTCAAAGA

AAGGATC-AGAGAGGG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: c5a
FORWARD PRIMER: CAATCTAGAAGCTGTGGTCC
REVERSE PRIMER: TGTCATGCCTCTGGCTCTCG
LENGTH OF INSERT: 406

ACTACTACTACTGAGCGGAATTCGTGAGACCTGCTCTTTCTCTAGAAACCCTCTGTGGT
........................................ >>>>>>>>>>>>>>>>>>
TGATGATGATGACTCGCCTTAAGCACTCTGGACGAGAAAGAGGGGGTGGTTAGATCTTCGACACCA
GCGACCACCCCCCGAAGTAGACAAGAAGAGTCAACAAAACAGTCAGAGGTCCTGCGGACTCCCTCTT
CTCTTTTTTCACAGGCAATAAACAGAAGCCAGACAGACAGGACAGACAGACAGGGAAG
GAGAAAAGTGTCCTCAGCACAGATGCAAGATGTCTGTCTCCAGCTCAAGTTCTACAACAAGGGNACT
ANACTCGANGAGCTGTGCTNAGATCTACAGAAGAGTCAAGGTTCAGTTCTCCCTCTTTA
TGAAGGCAAGNCCACACCACAGGAGGACTNTACATCATTACTAGGCTCTCCNNTGA
ACAGGTTCGTGTCCNGGTGGGTCCCTCCNACTACTTCAGAAGTTCCCGAGGANAGATNGGT
NGNANATTTTTCMTTNAACAGTCTNAAGTNCNANGTTTCAANGGAACTAACAANTGTNACCCAGGTCC
AAAAGTTTTCCTNAGACAGAAGT
TTTCCAGGANTTCTGTCCA

DUPLICATE REPORT FOR c5a: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 24,25
END OF RECORD.
SEQUENCE FILENAME: c5c
FORWARD PRIMER: AGGACATCTTTGGGCGAGG
REVERSE PRIMER: CAACNCCNCTTGAGCTCCA
LENGTH OF INSERT: 419

GGAATTCGTGAGACCCAGGACGCTNCTTGGTTTCNCTGCAGCTATNGCCAAGGACATCTTGGGCGAGG
GATCTTTNATTTGAACTGAACGTCGATANCGGTTCCTGTAGAAACCCGCTCC
GGAATTCGTGAGACCCAGGACGCTNCTTGGTTTCNCTGCAGCTATNGCCAAGGACATCTTGGGCGAGG

DUPLICATE REPORT FOR c5c: C5 (-4) vs C5 (score = 96)

GAGCGGAAATTGGAGACCCAGGACGCTNCTTGGTTTCNCTGCAGCTATNGCCAAGGACATCT
C5
GAAACCAGCTATTTGACCCAAAGGACATCT

TTGGGCGAGGCA--GGGCTGCAATTTTNATTGGAACCTCAAC-AAGCTCC-GAGTTNNTNGACCCAGAG
C5
TT-GGC-A-CGAAGGGCTTCTA-TTTTA-GTAACTGGAACCAAGGCTCAGTTGCTTGAG-CAAGAG

GTCAACC-AGGA-GACT-GTGGAGCTCAAGGGNQNGTTCGGAAGACTTTGTGGACAG--T-GAG
C5
GTCAACCAGGAAAACAGGCTTGGCAAGGACAGTGAAGGCTCAGTGCTTGGAG-CAAGAG-C-TTTTGAGACAAAATG-G

NACCA-TCTNGAGACACAGTTTTGTCTCTNCT--TNGAGAGA--GTNTNCTNACCCNACCTNN
C5
--CCAGT--TTCAGA-A-AA-TTGTGGTTGC-TTAATT-GAGCTTGTTTGAGCGCTTGGCAGAA

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15

END OF RECORD.
SEQUENCE FILENAME: c5d
FORWARD PRIMER: CACTAAAAAGGAAGGCTCA
REVERSE PRIMER: CCACACTTCCTCTAGG
LENGTH OF INSERT: 279

CACATCATCACTGAGACTGACAGTAAACAGGGAAGGCTCCAACAAATGATCACCCTGGGGTGAAAGAAGTA

GTTAGTACCTCTCATGTGTTTTTTTCCACTCCAGGTGGTTTACTAGTGGGACCCACTCTTTCAT

TAGCTCTCTCGTTTGGCTTTTTTGCTATTCCATAGGAGGTTGCTCTGGGAGTTGCTATTCTCCTTC

ATTCAGGAAACCCGAGGCTACCCAGAGGAGATATAGTCTGCTGCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

ATTACGCTGAGACCAATGACCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

ACTGCAGCCCTCTGCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCTCTCTCGTTTGGCTTTTTTGCTATTCCATAGGAGGTTGCTCTGGGAGTTGCTATTCTCCTTC

ATTCAGGAAACCCGAGGCTACCCAGAGGAGATATAGTCTGCTGCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

ATTACGCTGAGACCAATGACCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

ACTGCAGCCCTCTGCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

--------------------------------------------------------------------------------------------------
DUPLICATE REPORT FOR c5d: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.

--------------------------------------------------------------------------------------------------
SEQUENCE FILENAME: c5g
FORWARD PRIMER: TTCAACACTTGCGAGAACAGGACAACCTGAGGCTTTTATTCGTCTCTAGGAAAAAGGAAACAGGGCACCAAGGGATG
REVERSE PRIMER: TTACACCCAGGTTGGGCAA
LENGTH OF INSERT: 381

AATTCGTGAGACCAATGACCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TTACACACTTGCTGAGAACAGGACAACCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

GTTGCTAACACCATTTCTTTCTTTGAAAACCTCTTTTCTTTCTTTGTGACGTTTCCCTAC

GTTGCTAACACCATTTCTTTCTTTGAAAACCTCTTTTCTTTCTTTGTGACGTTTCCCTAC

CAGATTGCTGCTGAGAACAGGACAACCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

ACAGCAAAAGCAGACCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TGCTCAGCAGGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

TAGCAATTGGACAATGGATTTGGGATTTTGAGCTTGCCTGACTTGCTCTGCTGCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG

--------------------------------------------------------------------------------------------------
DUPLICATE REPORT FOR c5g: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: c6g
FORWARD PRIMER: CGATACCTTTCTAGCTGCGT
REVERSE PRIMER: AGAAGCAGAGAACGAGAAGA
LENGTH OF INSERT: 311

DUPLICATE REPORT FOR c6g: C5 (17) vs CS (score = 141)

NUCLEOTIDE BLAST HITS:
[{'T16882', 'NIB1963 Homo sapiens cDNA 3end.', 1.3e-19],
{'T09998', 'seq692 Homo sapiens cDNA clone...', 6.5e-101],
{'T16883', 'NIB1963-5R Homo sapiens cDNA 5...', 1.4e-08]

PROTEIN BLAST HITS: <none>

PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: yllb
FORWARD PRIMER: AGCCCATGACTAGACAGCAC
REVERSE PRIMER: CCCAAGAAGCCTTTAAC
LENGTH OF INSERT: 254

GGAATTCGTGAGACCCCAACCCGGGACTCTCTGCTATGTTTCTCAGCCCATGACTAGACAGCACT

CTTTAAGCCTCTGCTGAGAGTCTCCCTGATCCCTGGAATAGACCCCAACCCGGGACTCTCTGCTATG

AGCCAGAAGAATGCTAAGCTAAAGGGGTTTGGCCCTGAGGACGATACAAAGGAGTCGGGTACTGATCT

GAAACCCCCAGGGGTCTTGGAATCTTGAGGTTAAAGCGCTCTTCTGAGAGCAACTCTTTTGTTTC

TTTGGCGTCCCCAGAACTCTTACAACTCTTTTGTTCGAACTCTTTTGTTCGAACTCTTTTGTTCG

ATTGGTGGCCTCATGCTGTCACATTCCAAAGTCTNTTNGGAAAGNCTAAAGCCAAAGCCTCGGTCTC

TAACCACGGAGTACAGACATGTAAGTTTTCCAGANAANCCTTTCNCTAATTTCGTTTCCAGGCAGAGT

DUPLICATE REPORT FOR yllb: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: ylld
FORWARD PRIMER: ATGGTGTTTGTGTTTTCTCCACAG
REVERSE PRIMER: CCAAAGTGGTGAGACGTCATACATA
LENGTH OF INSERT: 422

GGAATTCGTGAGACCGACGCTGGTTCTCCACAGAAATTATCCACCAGTTCTTCTGGAACAG
CAGAAAGGAAAC-AGAGTGCCTCTCTCAGATCAACCCGACGCTGCCCTTTTAAAGCTTTATCTAGTAACGTCTNGT

DUPLICATE REPORT FOR ylld: ABX11_10 (8) vs ABX11_10 (score = 80)

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
['U12010', 'nemo gene product [Dros...]', 0.00017]
['U12009', 'nemo gene product [Dros...]', 0.00017]

PHYSICAL MAP LOCATION: 1,2,15
END OF RECORD.
SEQUENCE FILENAME: ylle
FORWARD PRIMER: CACCCCAACCTAGATAAAAATCT
REVERSE PRIMER: GGAACACTGGAAANATCTCTC
LENGTH OF INSERT: 287

AGCGGAAATTCGAGACCACAAAGAGCAGCTGCCTTAGCCAGGCGGTTGATGCCGCTCTATCATCTCCCTGTAC
TCGGCTTAAGCCACTCTGAGTCCTGACCAGATGGTCCGCTCCACTACGGCGAGTATGGAGGACGAGT
CTGGGGGACGTACAGAGTCAAGTGGATGCAAGGTTCAAAGCCACGCCACCTAGTAAATCTCTCTCAGCA
GAGGGTGGAGGAGCAGAGGATGCTCAGTGGGATTAGGACAGAGGATGGGCTCAGTGTTAAAGGCACTTNGCATCAATCCTGACAACTTTGAGTT
-------------------------------------------------------
TGGCCCTAAGGACAGAACTAGTAGAAGCACCAYATGAGAGATATNTTCCAAGTTTCCCTTACTCA
ACCCGAGGTCTCTGGTACTGTCATCTTGAGTTAAAATCTCTTTANAGGTCTCAAAAGACGAAAGAGGAT
CAAGGCCATTNGGAAGACACANCCACCGATACGCTACAGNNTCCGTGNGGAGTTNNTNTNTNTNTNGN
GTTCCGGTANCCTCTCTGTTNGGGGTGCTCAGAGTGCNNAAAGCANCNNCCNAAANANANANANACN

DUPLICATE REPORT FOR ylle: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.

SEQUENCE FILENAME: yllf
FORWARD PRIMER: CATAACGACGCTGACCTTCC
REVERSE PRIMER: GCTTCAACGGGACAAGCTTGG
LENGTH OF INSERT: 261

ATTCGTGAGACCCATGGGATAGAGAAGACAGATGGGAGCAGTACACGCTCATTTCTC
TTGCACTCTGGCACCCTTATTCTTGCTCANCNTACCTACGCCTGAG(TGCACGGYGAAGG
TGCCACACATGGCAGTTCGGCTGAGATTTGCGACTTCGCTCGACTCAGGAGATCCTGCCAGAA
ACCTCTGACTGCTAGGGAGCAGACTTTAAACCCCAAGAGACCGAGGTAGCTCTTCGAGGCTGT
GAGAGTTCGAAAATATCTCCCAACTTGTCCCCGTTGAAGCAAGCAGTTCAAACACGGNTGAC
-------------------------------------------------------
CTTGGGGGACGTACAGAGTCAAGTGGATGCAAGGTTCAAAGCCACGCCACCTAGTAAATCTCTCTCAGCA
GAGGGTGGAGGAGCAGAGGATGCTCAGTGGGATTAGGACAGAGGATGGGCTCAGTGTTAAAGGCACTTNGCATCAATCCTGACAACTTTGAGTT
-------------------------------------------------------

DUPLICATE REPORT FOR yllf: EX14 (-7)
vs EX14 (score = 112)

ACTTTCCCTGGCCACACATGGCAGTACACGCTCATTTCTCGGCCATCACAGAG

EX14

TCCTGGCCAGAGAGTCTGAGGGATAGAGGAGGACAGATGGGAGCAGTACACGCTCATTTCTCGGCCATCACAGAG

SEQUENCE FILENAME: y4a
FORWARD PRIMER: CTCGATTCAGTATAGCAGCC
REVERSE PRIMER: CTTTTATTGCCTTCAGGAG
LENGTH OF INSERT: 262

```
GCCTGATTCAGTATAGCAGCCCATGTCTCTCTGCAAGCTTCCAGTNGGCACATGTGGGGCACATTTCG
CGGACGTAAGGTCATATCGTCGGTACAGAGAGACGTTCGAAGGTCANCCGTGTACACCCCGTGTAAAGC
TTITGAGAGACCGGATGGCTCCCACTTTGATGTCCGGATCCCACCATTCTCCTTGGAAAGCAATAAAGA
AAAACTCTCTGGCTACCGAGGGTGAAACTACAGGCCTAGGGTGGTAAGAGGAACCTTTCGTTATTTCT
TTGAGAGACCGGATGGCTCCCACTTTGATGTCCGGATCCCACCATTCTCCTTGGAAAGCAATAAAGA
```

DUPLICATE REPORT FOR y4a: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.

SEQUENCE FILENAME: y4b
FORWARD PRIMER: CACTGTATGCTATCTCAGCCGC
REVERSE PRIMER: TGCTCTACACCCTCTCCAG
LENGTH OF INSERT: 215

```
GCGGCCCGTGAACTACTACTACTGAGCGGAATTCGTGAGACCCCCCGTTG GAAAAGCAGCACTTGCACATACACGTATGGTATCTCAGCCGCCCTTCATCTAGGTAGGNGTNGGCTAAGGCACARTGCGGATGTGTATGTGCATACCATAGAGTCGGCGGGAAGTAGATCCATCCNANCCGATTCCGTAGGAATCGCCTA
```

DUPLICATE REPORT FOR y4b: x9c (17)
vs x9c (score = 114)

```
AGCACTTGCACATACTAGTATCTCAGCAGCAGCCCTTCTTCTAGGATGNGGCTAAGGCACATACACGTATGGTATCTCAGCCGCCCTTCATCTAGGTAGGNGTNGGCTAAGGCACARTGCGGATGTGTATGTGCATACCATAGAGTCGGCGGGAAGTAGATCCATCCNANCCGATTCCGTAGGAATCGCCTA
```

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
['U12010', 'nemo gene product [Dros...]', 3.6e-24]
['U12009', 'nemo gene product [Dros...]', 3.7e-24]
PHYSICAL MAP LOCATION: 1,2
END OF RECORD.
SEQUENCE FILENAME: y4c
FORWARD PRIMER: TGGAGACTTGGTCTCTCAAGCT
REVERSE PRIMER: GTACATTTAGAGATTTCCGT
LENGTH OF INSERT: 182

GGAATTCGTGAGACGGGGCTCCATCATCATGGGAGCTTGTTCTCTTCAAGCTTCTCTCAAGGACAGTTCTG
............
CCTTTAACGCTCTGCGGCACAGTATGATACCTCTCAAGGACAGTTCTGCTGACAGCCTCGAGCCACCCN

ANAGTGACCTGATTCCAGAAAAGTGGTCACCTCTCAGCAGGACAGGAAGTTCTGACGCTTTAGG

TTNCTACTGAGACTGACACCTAGAAAATTAGGTGACGTGGCATCTCCTTCTCAGCAGCCGAAACCTGGAAGAGATCCTGAGCTTCTG

GACACTGACCGGATCTCCTAATCTGACCTCCTGACGTGACGCTGGGACGGTTACCTTCCTCAGCAGGAGGTTCTGACCA

<<<<<<<<<<<<<<<<<<<
CTGTGACTGAGCTTAGAGATTTGAGACAGGACTTGGAGTGCAGCGATCAGGACACTGAGTTCG

DUPLICATE REPORT FOR y4c: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,24
END OF RECORD.

-----------------------------------------------------------------------

SEQUENCE FILENAME: y4e
FORWARD PRIMER: GGACATCTTGACCTGGACCTGG
REVERSE PRIMER: GCCTAGTGAAGCCAGAAGT
LENGTH OF INSERT: 241

CGGAATTCGTGAGACCAGAGGCTGGACATCTTGGGCACTGCACTCAAGTACAGGACACTGAGTAACTG
............
GCTTTAACGCTCTGCGGCACAGTATGATACCTCTCAAGGACAGTTCTGCTGACAGCCTCGAGCCACCCN

CCCGGTCTCTTCTACCGNCGCGAGCTTGTAGATGCTGAGGACATCTGACCTCAGTACTGAC

GCCTTTAGGAGATGGGGGCTCCAGGAGGAGATGGGCTCTGACGACCGCCAGAGAGAATGGGCTCTGAGATTTGAGCAAAGCTT

<<<<<<<<<<<
TGCTTACTGAGTTGAGACGGAAAGTGATCGAGGTCACTGACCTGATTTTTGTTGACTG

CTGGAACCTGAGCTTAGAGATTTGAGACAGGACTTGGAGTGCAGCGATCAGGACACTGAGTTCG

GACCCCTTGCACCCACAGAAGACTTCTCAGGATCTCTCAGGAGGTGCTGACGATCGGGGT

DUPLICATE REPORT FOR y4e: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
SEQUENCE FILENAME: y7a
FORWARD PRIMER: CCTTGGCTACTCTTATGATGGC
REVERSE PRIMER: GCACCTTAGTGACGACGCAAC
LENGTH OF INSERT: 348

CAGCTGTATCCCTGCTACTCTTATGATGGCCTCATGCTGTCACATTCCAAAGTCTGTGGGAAAGACT
GTCGACATAGGGAACCGATGAGAATACTACCGGAGTACGACAGTGTAAGGTTTCAGACACCCTTTCTG
GAAAACCAAAAGCCTCTATTTGCGACAAAGATATGTCACCTTTGCTTCCCAGGACGATCTGGAGACAA
CTTTGCTTCTCAGAGTATAAAACCTGNTTTTCCACCTAAGTAGAATCAAGGTGCCTGACCCAAGCATCTT
GCAAAGCCTCTATTTGGNCANAAAGTTGGATTGTCATCTTAGTTCCAGCGAGATCTGGACGAA
CTTTCGGTTrCGGAGATAAACCNGTNTTTTCAACCTAAGTAGAATCAAGGTCGCTCTAGACCTGCTT
<<<<<<<<<<<
ATTACACAATACGCTCTGATAATAGTACGCTACTATCTCTTTCTCCCTGCTTCAGCTGATCCTT
GATTATCAACGAGCAGTGATTCCACGACTGGTTCTTGAATACTCGATAAAAAATGTTAAGTGCTTACAC
CTAATAGTTGTGCACTAAGGTCGCTTGGAGACCCACAGTTTTAGCTTGTATCGTATCTCCAGGAGAA
<<<<<<<<<<<
GGAGAGGGGTCCTGACCCCGATGATCTTGACCCTAATTTAGTACACACCACCAGGCCCTGGCTAACACAA
CTTTCAAAACCTAACTCAGGAGAGCTGAGAAAAGACTN

DUPLICATE REPORT FOR y7a: <no duplicates>
NUCLEOTIDE BLAST HITS:

[ 'M80783', 'Human B12 protein mRNA, compl...', 6.8e-271]
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.

SEQUENCE FILENAME: y7h
FORWARD PRIMER: TACACGGTGGGACCCAGC
REVERSE PRIMER: GGCCTGGTGGTGTGTACTTTT
LENGTH OF INSERT: 275

GTGGGAATGAGCCAGAGCTACACGGTGGGACCCAGCGGAAGGAAAAAGAAAGTTGCTTAGACAATGG
CACCCTTACTCGTCTCGATGTGCCACCCTGGGTCGCCTTTCCTTTTTCTTTCAACGGAATCTGTTACC
GGCTCTTGTTTGGCAGTCAGAGGCCTCATCTGCCTCCCCAGGACTGGGGCTACTAGAACTGGGATTAAA
CCGAGAACAAACCGTAGTCTCCGAGTGAGGAGGGGTCCTGACCCCGATGATCTTGACCCTAATTT
AGTACAACCCACCCAGCGCCCTTGCGTAACAACACTTTTCAAAAACCTAATCAGGAGACTGAGAAAGACTN
<<<<<<<<<<<
TCATGTGGTGGTCTCGGAGCGATTTGTTGAAAGTTTGGATGAGTTCGCTCTCGACTTTTCTG
GTGGTGTACATATGCACCCCCCTGGGACCGACCGCCCTGGCTTCGCTGCTGGCCACATAACACTTG
CACCACGTAAGTATTTATACCGGGGAACCTCGGGGCAACAGACGGGTTAGAAACGGGGTGATTTGACC
<<<<<<<<<<<

DUPLICATE REPORT FOR y7h: <none>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 1,2,15
END OF RECORD.
SEQUENCE FILENAME: y8a
FORWARD PRIMER: ATGCTGCCTTCACTCGCA
REVERSE PRIMER: CTGGTATGCCACTGAAGCTT
LENGTH OF INSERT: 207

GGGGCATGAGGGCCCCATCGATGCTGCCTTCACTCGCATCAACTGTCAGGGGNAGACCTACTTGTTCA

CCCCGGTTAATCTCCCCGGGTAGCTACGACGGAATGGACGGTACAGGACGTCCCGNTCTGGAATAGAAGAT

AGGGTATGCTACTGCGCTTTTGAGGATGAGGCGCTCTGACCTGGTTTATACCCGAAACATCTCCGAAG

TCCCCATCAATGCACTGACGCAAAACTCCCTACCCCGGACCTGGGACAAATAGGGCTTTGTAGAGGCTTC

CTTCACTGCACTGGCATACGAGGGATGACCTTCTGGACCCGACTGTAACGGTGGCAGG

GAAGTCACGTTATGGTGTTACAAACTACGTCGCAAGGGAAGGGCCGGTACAITGCCACCCGTCC

DUPLICATE REPORT FOR y8a: <no duplicates>

NUCLEOTIDE BLAST HITS:
['M77123', 'Mouse vitronectin mRNA, compl...', 4.4e-57]
['X63003', 'M.musculus mRNA for vitronectin', 2.6e-49]
['X72091', 'M.musculus gene for vitronectin', 6.5e-46]
['X03168', 'Human mRNA for S-protein', 8.1e-46]
['X50006', 'Human S-protein gene, complete cds.', 1.3e-37]

PROTEIN BLAST HITS:
['M77123', 'Vitronectin [Mus musculus]', 8.1e-33]
['J0T662', 'Vitronectin - Mouse>gi 441466 gp X72...', 8.1e-33]
['P22458', 'VITRONECTIN PRECURSOR (SERUM SPRE...)', 5.8e-32]
['P04004', 'VITRONECTIN PRECURSOR (SERUM SPRE...)', 1.1e-30]
['X03168', 'Human mRNA for S-protein....', 2.9e-30]

PHYSICAL MAP LOCATION: 14,15

END OF RECORD.
SEQUENCE FILENAME: y8c
FORWARD PRIMER: AGTGCTGCTTTCCACCTCA
REVERSE PRIMER: TGAACGTGGTGAATTTCTT
LENGTH OF INSERT: 249

AAGTGCTGCTTTCCACCTCAACGGGAGAGTCTACACCAGTNACTTTGAGCCTGTCACCAAGGG

TAATTCAAGGCCACAGGCAAGGAAATTTGATGACACCTTTGAGAAGAACCTCAGCTCTGTCCGACAGGTTAAAGAAATTATCAACCTACCTCATTT

ABX11_10 (16) ylld (15) vs ABX11_~t (score = 144)

ABX11_10
AAGTGCTGCTTTCCACCTCAACGGGAGAGTCTACACCAGTNACTTTGAGCCTGTCACCAAGGG

ABX11_10
CCCAAAATTTGATGACACCTTTGAGAAGAACCTCAGCTCTGTCCGACAGGTTAAAGAAATTATCA

yld (score = 93)

yld
AAGTGCTGCTTTCCACCTCAACGGGAGAGTCTACACCAGTNACTTTGAGCCTGTCACCAAGGG

yld
AAGTGCTGCTTTCCACCTCAACGGGAGAGTCTACACCAGTNACTTTGAGCCTGTCACCAAGGG

yld
AAGTGCTGCTTTCCACCTCAACGGGAGAGTCTACACCAGTNACTTTGAGCCTGTCACCAAGGG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
['U12010', 'nemo gene product [Dros...]', 2.2e-22]
['U12009', 'nemo gene product [Dros...]', 1.6e-20]
PHYSICAL MAP LOCATION: 1,2,15
END OF RECORD.
SEQUENCE FILENAME: z1le
FORWARD PRIMER: GCTGATTAGGAACTTCCCTGC
REVERSE PRIMER: GCTGTAGCTCNTGTTGTTGAG
LENGTH OF INSERT: 296

GCGGAATTCTGAGACCGGGGTCTCAGTGAGTTCTTGCTTAAGTTTCAGCTCAGCGGCGTGATTA

GAACTTCTCTGCTCCTCTCTCTCGNNNGACCAGGGTTGAGGTTGGGCTGACACATCGATTTCTT

CTTGAGAACCAGAGAAGAGAGAGNNNCCCTGGTCCCACATCCCAACCGACTGATAGACTAAAGAA

TGCGTTATGGATGGATCCAAAGGCTTTCTCCCTGAAGCTAGACCTCACCAAACANGAGCTACAGCT

ACCACAAATACCTACCTAGGATTCCCGAAGACTTTGATTCAGACCTCGAGTTGTGNTCTGATATGCGA

AAAGCCCTTTACTTTGCTCTTTGATCCACCTAAGATGTGCTAAGCTAAGCGTTAGTACGGG

TTCCGGAAATGAACACCACTGAGGAGATCTCCTGGAATTTACAAAATCNATTCGGATCATCGC

NAAGCTGTCTAATGAAACAGAGACAGTCTGGTTGAGCTGGCACGGTTNCGGTNNGHHNNNGN

NTCCGAAACCTACATTGTTGACTGACATCGAACATCGAGCTTAAAGGGACCTCCTNCCCNNTCN

DUPPLICATE REPORT FOR z1le: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: zllg
FORWARD PRIMER: AGAGCATCAAGCTAAGCCGC
REVERSE PRIMER: GCCTGAATCAGACTCCGG
LENGTH OF INSERT: 198

GCGGAATTCGTGAGACCAACAGACTAGAGAGCATCAAGCTAAGCCGCTCCTCGCTCCTCAGCTGAGCCTCCTTAAGCACTCTGGTTGTCTGATCTCTGTAGTTCGATTCGGCGAGGAGCGAGGAGTGGACTCGGACAGGTNNGNATTAGG CACTGATGATGTCACCTCAGAAAGGGGCTGGATTAAATGTGATAAG

GCGGAATTCGTGAGACCAACAGACTAGAGAGCATCAAGCTAAGCCGCTCCTCGCTCCTCAGCTGAGCCTCCTTAAGCACTCTGGTTGTCTGATCTCTGTAGTTCGATTCGGCGAGGAGCGAGGAGTGGACTCGGA

DUPLICATE REPORT FOR zllg: a7d (16)

vs a7d (score = 112)

CGTTCGNNATGTCACGGCGATNACGTAAGCTTGGATCTCTAGGACTGTACGACTACTACGGC

ACAGGCC-TG-AAT-C-A-GAC

TTGAGCGGAATTTCGTGAGACC-AACAGACTAGAGAGCATCAGCTAAGCGGCTCGCTCCTCGCTCCTCA

a7d -T--CCGG-CTTC-TG-CACTTAACAGACTAGAGAGCATC-AGCTAAGCGGCTCGCTCCTC

CCTGAACTCAGATGNNATTAATGGGAAGCAGACTAGATGTGTCACCCCTGAGAAGGGGCTGGAT

a7d CCTGAACTCAGATGNNATTAATGGGAAGCAGACTAGATGTGTCACCCCTGAGAAGGGGCTGGAT

TAAATGGTAAGCCTGTAAGCTCAGAGAGCGGAGTCTGATTCAGGCTTGGGGTTCCAGTCAAGT

a7d TAAATGGTAAGCCTGTAAGCTCAGAGAGCGGAGTCTGATTCAGGCTTGGGGTTCCAGTCAAGT

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,22,23,27
END OF RECORD.
SEQUENCE FILENAME: zllh
FORWARD PRIMER: GTCAAGCCCCAAGCTCTG
REVERSE PRIMER: CAGGTAGGGCACAGGGTAGC
LENGTH OF INSERT: 203

ATTCGTGAGACCNGAGGNAAGGTCAAGCCCCAAGCTCTGGACAGTGGTCCTGGGATGTACTGCTACCAG

DUPLICATE REPORT FOR zllh: a10g (19) a8d (-24)
vs a10g (score = 123)

GCTACCAAGGNNCNNTGGCACATATCGTACTCTCTCAGCTCGCTCTCCATCATCGTACTCCCGTGG

vs a8d (score = 149)

GTAAGGAAGCCGTACCCCTGTCGCTGCTACTCGTGGGTTCACTCATCTTCTCATCACAGAGATGCAAAC

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: zle#2
FORWARD PRIMER: GACTGAGCCAGAAGTGTCTG
REVERSE PRIMER: CCGGTGTGTGACCTTCTTA
LENGTH OF INSERT: 267

GGAATTCGTGAGACCGGTAAAGACCCCTGTGTAAAGTCATGTGTGTTGTTACAGAATGCTGAGGACTTTGGA

GAGCCAGAAGTGTCTGTACCAGTGGCCTGGCTCCTGTAGTACAGAAAGTGCAGTGAAGGATTTCTGGGG

CTCGGTCTTCACAGACATGGTCACCGGACCGAGGACATCATGTCTTTCACGTCACTTCCTAAAGACCCC

AGGATCATAACCACAAAGTCTCATATGAGGTGTTCATGATGAAAGGTACACACAGCGTGTC

TCCTAGTATGTTTCTACGGAGTATAAGCTTAAAGACTCAACAAAGTACTACCTTTCCAGGTGTGC

TCCTAGTATGTTTCTACGGAGTATAAGCTTAAAGACTCAACAAAGTACTACCTTTCCAGGTGTGC

ACGTCCCTACTAGTTGGAAGGGAAGTTCCCTAAAGACTCAGAAGAAACTACAGAAGAAACTC

DUPLICATE REPORT FOR zle#2: <none>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7,9,18
END OF RECORD.

--------------------------------------------------
SEQUENCE FILENAME: zlg#2
FORWARD PRIMER: CACCATGTCCACCTCAACC
REVERSE PRIMER: AGGACAAGGACGGAGCTCCA
LENGTH OF INSERT: 229

GCCNCCTACTACTACTTTACTGAGCGGAATTCGTGAGACCNCCAGAATACCTATCATCACCCGCACCAT

GTCCACCTCAACCCTCACTCACTCAAAGCATTATAGATCTTCGGATCAACANTGTATACCATAAT

CAGTGGGATGTTGAGGTACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTG

NCCACCCCGGCCAGGACCTGTGAAGGACCTATTGTGCGGCCTGCGCTTGTGCTCTTGAAATGTACAG

NGGTGGGGTCGCCCTGGACACCTTCCGTAACCACCGGATCCCTGGAAGCCAGATCCCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTG

CAGGAGGAGGCAAGCTCACAGGCAAGAAAGTGGGATGTTGAGGTACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTGCCCAAAAGTGTTGACCGGACCTTG

GCCCCCTACCGGCTCGATCGTATCTCGCTCTCACCTCGATGCCTCCCTCCGCTCCAGAG

DUPLICATE REPORT FOR zlg#2: <none>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,22,23,27
END OF RECORD.
SEQUENCE FILENAME: z2b#2
FORWARD PRIMER: GTCAGGCCCCCAGTAAACTG
REVERSE PRIMER: AGGCAGCTTGGGGTGCTTC
LENGTH OF INSERT: 312

AATTCGTGAGACCGGCAAATAAAATAAAAGCCTTTTGGATCTCTCACGCCAGTCAAGGCCCCCAGTAAACTGTA

TTAACGGACTCTGCCCCTTTATTTTTATCAGAAAACCTAGAAGTCGTCGCAGTCCGGGGGTCATTTGACATT

ACAGACGTGAACTTGTGAGTGAGAAGGAATTTGGGCAGCTTCTCTCTCTCTCTCTCTTGAGAAAGCAGTCTTTATCC

TGTCATGCCTTGACACTACCCCTCTCAACCGTGAAGGAGGAAGAGGACTCTTTCTGAAGATAGGG

CTGCTGGCCAGATGCCCTGTACACCACAATGGGCGAAGCAACCCCAAGGAGGCTCTGCTGACTGGCCACTTTGCAAGAC

GACACGGTCTAACCAGGACATGGGTAACGCCATTGCGGTTGACGGATCGGAGTCCAGGATTTCTCTG

TTGAAATGTCATCCCTAAGGGGACTAGAGCTGTCGCTCCCTTAAACCAGTTGTCAGGAATTCTGGAAGGGGA

AACTTACAGANGATTCCTCTGAATCTCTCAAGGGGAAATTGCTCAACAAGTCTTAACTGAAACTCTCCCCT

CTGCTAGGCGTTCCTCTTTAAAGCCTTTAACCCTNCGCCCGAGGTTTCAAGGCTCTGACGATTCTCTGTCG

GAGCTCCGCCAAGAAGAATTCGAATTGGACAGGGGACCTGCGAAAGTCCCAGACACCCCAAGCCTCGCAGCTACGT

DUPLICATE REPORT FOR z2b#2: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11, 13, 23, 24
END OF RECORD.
SEQUENCE FILENAME: z2e#2
FORWARD PRIMER: GGAACTTTCCTGAAGGGGAG
REVERSE PRIMER: GCCTGGCTTCTGATAGCAGAA
LENGTH OF INSERT: 238

GTTCGNATGCAGGCGGTACGTAAAAGCTTGGGATCTCTTAGAGGCGCGCCNGTACTACTACTAGCGG

CAAGCGTACGGGCTCACTTTCTGAGACCTCTGGGAGTGAACCTCTCGGGAGCACTGAGATGAGCTC

GAATTCGTGAGACCCTGAAACTGAGGTTTGAATGTGACTACTG

CTTAAGCTTGGCCAACCCAC

CCCAGAATATCCTCAGCTCTTAGGAGCAGCCAGGACTGTAAGACACACTTCGGCT

GGGGCTTTTGAAGAGGAGAAACATAGATGTCGTCCTGCGTGAGCAGTTCTGAGACAGCAG

CAGTCAGGATGAGCTTTAAGGTTTCTGATGACCAACCGCCGCTCAGGGAGGTTGAC

GTACAGACACTTCTGCGTACGGCGAAGCATAGACACTTTGAGG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7, 9, 18

END OF RECORD.
SEQUENCE FILENAME: z2f#2
FORWARD PRIMER: ATATTCCCAATCTATGGAGCA
REVERSE PRIMER: CGACCAGGATACCTCAAACCT
LENGTH OF INSERT: 216

ATTCGTGAGACCGACCCGACGATTTCCCAATCTATGGAGCAATGGGAGATGGCAGACTTGGAT
............
TAAGCCTCTGTGGCTGCTATGAAGGGTTTGAGATACCTGTATACCTTACCTGTATGAAACCTGGAACCTA
................
AGAGATTCACACTCTAGAATCTCAAGTGGACGTTCGTTCTGCTTCCTC

TCCTAGGTTGGAGACTTGGGAGAATTCCTTGGACAAAGAGAGGATGAGAACGGAGAGG
............
AGGTGGAGGTACCTGGCTGCTATGGGAAAAAGCCACCTCTGCACTTTCGAGGCAG
<<<<<<<<<<<
TCCAAACTTCCATTGGGACACGAGTCATTATCCACTCTACCTTTCGGTGGGAGACGTAAGGTGAGTCGTC

DUPLICATE REPORT FOR z2f#2: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.

SEQUENCE FILENAME: z2h#2
FORWARD PRIMER: CTGTTGAGCAAAGTCGGCC
REVERSE PRIMER: TTCTCTTGCCCCTGGCCCT
LENGTH OF INSERT: 326

AATTCGTGAGACCCTCTAAGGGTTCCCTCTACCCTCTAGCTAAGCTTCCTCCCATTGGGAGATCAGCAGAGG
TT.AAGCACTCTGGGAGATTCCCAAGGGAGATAAACGATTCTGAAGGAGGGTAACCCTCTAGTCGTCTCC
ATCTGTTGAGCAAAGTCGGCCCCTTCTGTCCCAGTAGCCTGGGCTTCCTGGCTGAGTTAGCAGAGTCC
>>>>>>>>>>>>>>>>>>
TAGACCACTCTGTTTCAGCCGGGGAAGACACGGTCATCGGACCCGAAGGACCGAACTCAATCGTCTCAGG
CATAAGCTGCTTAAAGAGACTGCTGCTAGTAAGTTGCTTACGGTGGGGGAGGGCCAGGGGCAAGAG
<<<<<<<<<<<
GTATTCGACGAATTCTCTCGACCCGATCGTATTCAACGAATGCCACCCCCCTCCCCGGTCCCCGTTCTC

AAAGGGGAAGGGTTACCACCATGTCGCCCTCATACGTTNCATGCAAAGAGAAGTNCTCTNAAAGCCAGCCT
<<
TGGCCCTCCCCCATGGTGGACCGGAGGCTAGTCTCAGTCTCTCANGAGANTTCCGTCGGG

GNCATCTGGAGACCCCAACTGTCAGAGGGACTTTCTACCTTNAAGGGGGAGGCAGGAGCTCAGGT'TTA

CNGATGAACTCTGCGGCCGTNGACTCTCCCTGAAAAGTGAGANWTCCCCNCGTCCCTGAGGCCAATT

DUPLICATE REPORT FOR z2h#2: <none>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: z3e
FORWARD PRIMER: TGTCCATTGCTGGCGATG
REVERSE PRIMER: AAAAAGAAAGAGGGGACGG
LENGTH OF INSERT: 317

ATTCGTGAGACCGCGGCTNACCCCACGACCCTCTCCAGCCCCATGATCGAGAGCACCCCCATGTCCAT

TAAGCACTCTGGCCGCGACNYGGGGTGGCTGGGAGAAGGGGGTTACTAGCTCTCCGTGGGGAAGGATTAA

TGCTGGCGATGCGACCCCACTCCTCGACGAGATGGATCGGTAGGGGGCCTGTTCCTCGGACTCTGGTCA

ACGACCGCTACGGCGGCTGAGGACCTGCTCTACCTAGCCATCCCCCGAAGAGGAGCTGAGACGTGCTCTGAAGCTCCCAAGAGGCCTGACTTATGCTCACTCCCCTGCCCCCTCTTTTTTTGGTCCAATAAA

GGAGACTTGGGGTCTCCGGACCGTGAAATACCGAGTGAGGGCAGGGAAGAAGAAAACAGGGTTTT

G7GCCG7GAAATGACGTTCCTTTTTTAAAGTGCTCAAGCCCTGCTGGATGTGGCCTTGGGTTGGTGTGAC

CACCAGCTATTTACGCAAGAAAAATCCACCACTTGCGACCCGAACCCACACCCCAACCACACGA

ACCGAGCACAGGTGCTCCACCTCCGCGGTTTTGTGCCTAACANCTGTCTTCTTGGGGGTCTCAACGAATTTC

TGCCCTCGTGCAACAAGGTGANGCCAAACAGATTTTGGACAAGAAACCCCCCATGAGTTGCTTAAAGC

DUPLICATE REPORT FOR z3e: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 9,18,19
END OF RECORD.
SEQUENCE FILENAME: z4e#2
FORWARD PRIMER: AGAAGCCGGAGTCTGATTCA
REVERSE PRIMER: TCACCACCTGCGGACAGCCT
LENGTH OF INSERT: 298

-------------------------------------
CGGAATTCGTGAGACCGCGGCTrAGCTGATGTCTCTAGTCTGTTAAGGGT

TCAGGGCTCTGGGTGGTCCAGTGGAGAAGGTTCTCTCTCTCTCAACATCTTNTCTGAAAGGGTAGAG

GGGAACCTTTAGCCACAAGGTTAGAAGCATCCTAAGGCCATTGTGCTCTACTAGGCTG

-------------------------------------
CTCTGAAATCCTGTCTCCACTCTTAACTGCTGCCGAGAGATCGACAATTCACGTCTTCGGCCTCAGACTA

ACGCCATAAGGAAAGCTGAAGGAAGGGGAAGTGGCTCAGAAAGTNGAGAGGACAAGGCTGCTACAATTCAA

CTCAAGTCTCCCTCTATCAATCATCTCAAGTGGTTCTCTCCACGGAGTGACTCTCTGCTCCGGATGTTAAGTT

GAGTTGATGAGGGGATTAAAGTNGAGAGTTCACGCAAAACCCAGAGTGCCTANNGGCCANANCCNNNNNN

DUPLICATE REPORT FOR z4e#2: zllg (7)
vs zllg (score = 65)

GTACTACTACTACTGAGCGGAATTCGTGAGACCGCGGCTrAGCTGATGTCTCTAGTCTGTTAAGGGT

zllg GTGGGAAAGCACTGATGATGTCACCCTCAGAAAGGGGCTGGATTAAATGTGATAAGCCTGTAAAG
TGCGGTATTCCTTCGACTCTCTCTCCACCAGAGTCTCTCTGTCCGGATGTTAAGTT

zllg CACACTCTCTCTAGGAAGGGGATTAAAGTNGAGAGTTCACGCAAAACCCAGAGTGCCTANNGGCCANANCCNNNNNN

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,22,23,27
END OF RECORD.
SEQUENCE FILENAME: z4f#2
FORWARD PRIMER: CACTACCGAGCAGGGACTG
REVERSE PRIMER: GGCTCGGGTGCCAGGAGCTCG
LENGTH OF INSERT: 206

CGGAATTCGTGAGACCGGGGCCCACTACCGAGCAGGGACTGGGGGTGGTCCAGTGGCAAGCCAGAACAG

GCCCTAAGCAGCTGGGGCTCGCTCCCTGACCCCCACCAGGTCACCGTTCGGTCTTGTCCTGATGCAAACGGTGGNNNANCTGATNGGGATTTCCAGCAGGCCTGCTTCCGGAAACA CCTGGC

GAACTACGTTTGCCACCNNTNGACTCANCCCTAAAGGTCTG/TCCGGACGAAGGCTCTTTGCTCGGACC

CATGTGAGCAAAGCCACCGAGCTCCTGGCACCCGAGCCCCTGATCCCACAAGATAGTCGAAGTCTTCA

GTACAACTCGTTCTGCTCGAGGACCCTGGGCTCGGGACTAGGGTGTTCTATCAGCTTCAGAAAT

DUPLICATE REPORT FOR z4f#2: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7,8,9,10
END OF RECORD.
SEQUENCE FILENAME: z5a
FORWARD PRIMER: TTGAGAGGAGCTGGAGAAGGG
REVERSE PRIMER: GCCCCATAGATAGCACCCTT
LENGTH OF INSERT: 272

GAATTCTGAGACCTAGCTTGCCTCTGGTACATACACGCAATTGAGAGGAGTCGAGTTGGG
CTTAACTATTCTCTCACTCACACTCTCTCTGAGGCTTTAGCCCTGACAGGCTTCGGAGCTCAGCTACTTACCTCCCAATTGAGAGGAGTGAGAAGGG

GAAGTTCGAGACCTAGCTGCTTGCTCTGGGATACACAGCAATTGAGAGGAGCTGGAGAAGGG

CAGAAAAGCGCGAGCTGTGTTCCCAAGATATGCACCCGCGCTCTTAAAGCTGTCGACTGACTATG

TAAGGGTGACCTGCATTGCGTCTCCAAGATNAAHNTCNATNACACTCTCTAGTGAAGNAGATTT

ATTCCCACTGNGACCGAAGCTTCTCTCANTTNAGANTGCANGTTGAGAGCTACCCNTCTCAATT

DUPLICATE REPORT FOR z5a: zld#2 (20) vs zld#2 (score = 163)

CGTNNCAATGCAGCGTTAANAANCTTGGGAGCTACTTCAGAAGCGGCAATTCGTGAGACCTAGCTGCTGCTCTGGGATACACAGCAATTGAGAGGAGTA

AGCGGAATTCGTAGACCTAGCTTGGTATAGCCCTGAGCCTCTAGAGCGGCCGTGTA

GTCTGAGACCTAGCTTGGTATAGCCCTGAGCCTCTAGAGCGGCCGTGTA

CAGTNAAATGGAGCTGTGTTCCCAAGATATGCACCCGCGCTCTTAAAGCTGTCGACTGACTATG

TAAGGGTGACCTGCATTGCGTCTCCAAGATNAAHNTCNATNACACTCTCTAGTGAAGNAGATTT

ATTCCCACTGNGACCGAAGCTTCTCTCANTTNAGANTGCANGTTGAGAGCTACCCNTCTCAATT

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: z5e
FORWARD PRIMER: CTCCTCTGGAGAGACGGATAGG
REVERSE PRIMER: GTGCGGCCAGGACAGGAAG
LENGTH OF INSERT: 171

TGAGCGGAATTCGTGAGACCGGAACAGGCTCCTCGAGGGATAGGTTGATCGCCAG

ACTCGCCCTAAGCCTGCTGCTGAGAGGCCATGTCCTGAGGCCATGTCCTGAGG

ACCTGTCGGAACCCAGGACCGCTCCNCCCTGGAGAGGTGATCGTACAGGAGAAT

TGAGACAGGACAGGACCGCTCCNCCCTGGAGAGGTGATCGTACAGGAGAAT

<<<<<<<<<<<<<<<<<

GTTGAGAGACTCGAAGGACAGGACCCCGGTGTTCCTGGACGAACTCCCGTCAACCAGA

DUPLICATE REPORT FOR z5e: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7,8,9
END OF RECORD.

SEQUENCE FILENAME: z5f
FORWARD PRIMER: GCCCTTGTCATCCCATTTTGC
REVERSE PRIMER: GCCAAATCCCTACTTGAGC
LENGTH OF INSERT: 354

GAATTCGTGAGACCGATAGCCTTGTCATCCCATTTTGCACACCCTATTATCTGGGTTCTGCTCTCCTTT

CTTAAGCACTCTGGCTATCGGAACAGTAGGGTAAAACGTGTGGGATAATAGACCCAAGACGAGAGGAAA

CTTCCAGTTGGGACAGCGCCTGGAATGGTTATCTACGTCCTCCTAGTATCCCTTCCCATGAC

<<<<

GAAGGTCACCCCTTGCGGACCAGCCTTCTCCCAATCCGAGTTGTCAGTCAAGGATGGAGGAGGAG

AAGGTAGAGGATTGCCCTGGAATGCCAGAAGCTCTTAGGCGCAGTTCGATCTGGAGAGG

<<<<

TTCCATCTAGCTACGGACGTTGGCTATCGTAGTGTTCTCCCAATCCGAGTTGTCAGTCAAGGATGGAGG

<<<<

TGAGGGAAGGGAAAAAGCTACGTTGGCTATCGTAGTGTTCTCCCAATCCGAGTTGTCAGTCAAGGATGGAGG

ACTACCTGTCCCCCGGTCGTTCTGCTATCGTAGTGTTCTCCCAATCCGAGTTGTCAGTCAAGGATGGAGG

CTCTTCAACCCCTATCCGTTGGCTATCGTAGTGTTCTCCCAATCCGAGTTGTCAGTCAAGGATGGAGG

GAGAAATGGCTGATACCTACGCTAGGGGNCTACGGCAGTTGGCTATCGTAGTGTTCTCCCAATCCGAGTTG

DUPLICATE REPORT FOR z5f: <no duplicates>
NUCLEOTIDE BLAST HITS:
['S72304', 'rah=ras-related homolog [mice,...', 1.1e-16]
['X13905', 'Rat cDNA for ras-related rab1B protein', 1.2e-06]
['X63278', 'Z.mays yptm1 cDNA', 4.3e-05]
['X75593', 'H.sapiens mRNA for rab13', 5.3e-05]
['D16064', 'Rice cDNA, partial sequence...', 0.0091]

PROTEIN BLAST HITS:
['A41636', 'GTP-binding protein rah - mouse [fra...3.6e-08]
['X15747', 'Mouse ypt1 gene for ras-r...', 3.1e-06]
['L17070', 'GTPase-activating protein ...', 4.0e-06]
['S30273', 'GTP-binding protein ypt2 - rice>gi 2...', 5.6e-06]
['P35286', 'RAS-RELATED PROTEIN RAB-13 [FRA...6.6e-06]

PHYSICAL MAP LOCATION: 7,8,9,10
END OF RECORD.
SEQUENCE FILENAME: z7d#2
FORWARD PRIMER: TCACATTCCCCCAAGGCG
REVERSE PRIMER: CCNAAGTGGCAATTAGTTGTCT
LENGTH OF INSERT: 258

AATTCGTGAGACCGCTAAGGAACAGTGAGAATAAACCCCAGACCTTGACATCAGCTCACATTCCCCC

TTAACACCTTGGCAGTTCTTTGATCTTCTTATTGTGGGCTGAAAGTCAACTGGTGCAAGAAGGAGGGG

AAGGCCAGAAAGC2NGGCTTGAATCATTATTTCTCAGGCGCAACTGCAACGAGACAGAC

>>>

TCGCTCCTTTGCCNCCGAATCATGATAATTAGGAATCTGGGTTGAACACGTGGCTCTGTTCTGTCCAC

CAAGAATTTAACAAACAGCTGAAGGTAAGGTGAAATTCTAGACACATATTGCCACT

<<<<<<<<<<<<<<<<<

GTCTATTTAGTTGTCTGCAAGTGCACACTCTCCTTCACACACTTAAGAATCTGTGATTAACGGTA

<<<<<<<<<<<<<<<<<

TNGGAATGGACAACATTTGGTCTCGAGACATGGAAAGNTGCTTNGAATGGGNAAAGGGGCTCT

<<<<......

ANCCCTAACTCAGTIAAAACAGACCCCTTGACACCTCCTCAGAGAANCTTACCNTTTCCCCCAGA

DUPLICATE REPORT FOR z7d#2: z3b (32)
vs z3b (score = 226)

CGTCGCAATGCACGGCTNACGTAAGGTGGGCTTCTGAAGGCGGCGTNTACTACTACTACTGA

z3b

GCCGAATTCGGAACAGCCCTAGGAAAATCTTGGGATCCTCTAGAGCGGCCGTNTACTACACTACTGA

z3b

AAGGCACAGTTGCAATATCCAGAATACCCACAGCTTGATATCTGCAAGAAGGAGGGG

z3b

ATCCCCCAGGGAAGAAAGCNGGCTTAGTAACATTACTATTCTTCTAGACCCCAACTGCAACA

z3b

ATCCCCCAGGGAAGAAAGCNGGCTTAGTAACATTACTATTCTTCTAGACCCCAACTGCAACA

z3b

GAAGACAAAGTCCAGAATTTTACAACAAACAGCCTAAGGTGAAATTCTTA

z3b

GAAGACAAAGTCCAGAATTTTACAACAAACAGCCTAAGGTGAAATTCTTA

z3b

GACAACATATTGGCCACTTNGGGAATAGGAAAGCAGTTGGAAGNT-GCTCT

z3b

GACAACATATTGGCCACTTNGGGAATAGGAAAGCAGTTGGAAGNT-GCTCT

z3b

TNGAAATGGGN-A-----AAGGGGGGTCTCA--CGGAATTCCTCCGTGGAGNAGNNGNNTTTNNTNNTATT

z3b

T-CAATCCCTCTTGACACTTGCAAGGCG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7, 8, 9, 10
END OF RECORD.
SEQUENCE FILENAME: z7e#2
FORWARD PRIMER: TTCTCCTCGGACACTGCTCC
REVERSE PRIMER: CCCTGGAATGGATGCTGA
LENGTH OF INSERT: 288

GGAATTCTGAGACCGTTCCTGAGCAGCTCCATACAGGCTCAAAGGACAAGCTTTACAAGAA

TTCTCCTCGTACAGGCTGGTGCNAAACCANATACGCAAGGTGTTTTGAGACAGGGGTTATTTAGA

AAAGGAGGAGCATCCCAACNTTTCGTTATGCTCCTCCTCCACAAAACTTCTGTCTCCCTCCTACCTTCT

ATTGTCAAGCAGCATTTCAGGCGATGATGCTGGCTGGCAGCTTTTNGACAGTTCTTNTCCTCCTACCAT

TTAGNCATGCCACACCGTTGTTGGCTCTGCCTCAACTCTTGGCCTTCCCAAGCACAGGACTCTGAGGT

AATCNGTACGGGCTTCACAGAGATGGGAACGGTTCCTGTCCCTGAAGTTCTCTCA

DUPLICATE REPORT FOR z7e#2: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7,9,18
END OF RECORD.

SEQUENCE FILENAME: z7h#2
PRIMERS NOT PICKED
LENGTH OF INSERT: 435

ATTCGTGAGACCAAAAAAAACTTGTCCTCAGAAGCTGACNNNTCTCTTAGTCCTCAGGCGAGAAA

TAAGCAGCTTGTTTTTTTTGACAGAGATCCTTGAGTGGNNAAAAGAATACGAGTGGCTGTCCCTTT

GCCGCCTTCAGGGTGCAAGCTCACTCTTTCGTCAAATGAGGTATGTATGACTCAAGGCCAGCTCTA

***************

CGGGGGGAGTCTCAACGGGTACAGATGGAAAGATTTACTGCCAGATTACATCAAGTCGCGGCGGAT

TTGATGCGGTGTAATGTGCTATAGCACTCTTGGATGGGTTGTCACACCGTACAGGATGATGT

***************

AACCTAGCGCAGGGACAGAATTCTGTAGGACTCTAAGCTCCCTCAAAAAACTTGCGGAGATCCCTACPACA

GTTGTTGTTGTTGTCTTNTGTTTGAANACCCCTTGCTCCGACANAGATNNTATAGANAACCTTT

CACACACACACACACACACANACATNTGGGCAAGCCGCTNTCTANNNTATCTNTTGAAA

DUPLICATE REPORT FOR z7h#2: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: z8b#2
FORWARD PRIMER: AGTGCCATCCGTAGAGAGCC
REVERSE PRIMER: GAAGATAGACTGATCCGGGC
LENGTH OF INSERT: 287

CGGAATTCGTGAGACCGCATGGAGGAGCAAGCCAGTGGACTAGGAGGGGAAGGGAGCCCATCGTGGAGT
GCTTAAGCACTCTCGGCTACCTCTGCTGTCGTCACCTGATCCTCCCTCTCTCCTGGGTAGCACTCTCA
GCCATCCGTAGAGACCGGCACGGGCACNANAGTGAAGACATGTCATCACCAGCAGACCTTGCCTTGGTCTCT
CGGTAGACCTCTCTCGGTCGCCCTNNTCACTTCTGTACAGGTTGTCGTCTGGAACGGAACCAGAGA
TGCCCAACACCCCTCCCAGCTGGAAGAGGCTCCTCTGCAAATGGCCTCAGTAT
ACCGGGTGTGGGAGGGGCCTGCGACCGATCTCTCCTCCCGAGGGAGACTGTACGAACGGGCCTAGTCAGAT
TCCTCGTCAAGGCTTTTCTCAGACTTNNNGAGTTCTCGGAAAGTTACCTAACCCAGGGACAGAGGATTOG
AGAGAGGGATGACCTGGAAGAGGCTCTGACATGGTCCCGGATCAGTCTA
ACCGGGGTTGTGGGAGGGGCCTGCGACCGATCTCTCCTCCCGAGGGAGACTGTACGAACGGGCCTAGTCAGAT

DUPPLICATE REPORT FOR z8b#2: z10f (24) vs z10f (score = 221)

TGAOCGGAAATTAGTAGGACGCATGGAGGACGAAACGCAGTGAGACTAGGAGGGGAAGGGAGCCCAT
AGGGCCATCGGTAGGACG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,22,23,27
END OF RECORD.
SEQUENCEFILENAME: z9a
FORWARD PRIMER: CTCAGCAGCAAGGCAAGG
REVERSE PRIMER: GGGGCTCCCTTTCAGGC
LENGTH OF INSERT: 355

AATTCGTGAGACCATCCTGTGGCTGCATCACAGTGGCAGACCATCCTGCTGGACACACATCTCACCTGCATCGGGACCTCAA
TTAAGCACTCTGTGAGACCATCCTGTGGCTGCATCACAGTGGCAGACCATCCTGCTGGACACACATCTCACCTGCATCGGGACCTCAA
GACACAGAACATCCCTTGGAACACACGACATGGGCTAGGATCGTGAACAGTTACAGCTCAGTTTTGCAATCCTTAAGAT
CTTGCTCTTGTAGGAAGAAGACTGTGTGTGGGTCTGACCCACTGATTCTAGGCACATGCACTCTGCTGCTG

DUPLICATE REPORT FOR z9a: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7,9
END OF RECORD.

SEQUENCEFILENAME: k10d.Seq
FORWARD PRIMER: CCTGTGTGGGGAGCTGTG
REVERSE PRIMER: TAGTCTAACTATGGACAGAG
LENGTH OF INSERT: 179

GTTCCTTCCTGTGGGGAGTGTGCTCAGCACATGGCAGGAAGTGAGACCCAACATCACAAACAAG
TTCTTGGTGGGGAGCTGTGCTCAGCACATGGCAGGAAGTGAGACCCAACATCACAAACAAG

DUPLICATE REPORT FOR k10d.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: K10F.Seq
FORWARD PRIMER: CACCTCAGCTGGCCTITCA
REVERSE PRIMER: GGGCCTTTG GCCACA CACTTGG
LENGTH OF INSERT: 265

CAGGGAGTCTGTGGGATTCTATGGGAAGCCACGGTGTTGAATCTTACTTCAGGTCGTCACCTCAGCTGG

GTCCCTCAGACACCATACCGATACCCGTGGTCCCAACAATTAATGAGTCGACGAGCTGGAGTGGAC

CCTTCATCAAATTCAGCTTTACAGGCGCTGCTGCACCTGTGGGAGGCACTGATATGGCGG

>>>>>>>

GGAAGTAGTGTTTGTAAGGCTCAAGATGTCGCGCGTGAGCTCAGACCGAGTGGAGTGGAC

<<<<<<<<<<<<<<

AAAGGCCACCTTCTAGCAGGTGATGACCTGAGATGCCAACGGCTGGCCAGAGTTGCACAT

TTTCGCGTCAATTTTTCTCAGTTACTATCGGTTCGATACAGGCTGTTCCCGGAGTCTCACAGGTCTA

GTGACCAGTGACAGCTGAACAGATGCTGTATCTCTNAGGNMTTTAGATTTGG

CAGCGCTCACTGCTGAGCCTGTCCTCAGTACATAGANATCCGAAATTTCCAAACCC

DUPLICATE REPORT FOR K10F.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: K11F.Seq
FORWARD PRIMER: TAAAGTGCGGTAATGAAGACGT
REVERSE PRIMER: TCCCAGCTGCCCCAGAAGC
LENGTH OF INSERT: 271

CCTGTNCCTCGGACTCTGTATTTCCCTGAGCTCCCAAGAGGCCTGGCCACTTATGGCCTCACTCCCTG
GGACANGGAGCTGAGCAATGGAGACTTCGAGGGTTCTCCGGACCGGTGAATACCGGAGTGAGGGAC
CCCCTCCTCTTTTTTGGCAATAAAGTGGGCTGAAAGATGACGTCTTTTTATGTCAGCCTGGATCTGG
GCTGGCCTGTG6GCTGTTGGCAACCGGACAGCGCTTCCACTCGGCTGGCTGGATCTACACCTGTGTT
CCACGGGAACACTGACCCACACTGCTGCTGCGAAAGGTAGGGAACACGGCAGATGCTGACAAAA
CTGGGGCCAGGCTGGGAAAGGCAGACATCTG7AATACNGCAG
<<<<<
GACCACGGCTCGGACCTCTCGCATAACGGCCCTCTCCCTCCTGTTCCCCCCTCTCCTGTTACAGATTTA

DUPLICATE REPORT FOR K11F.Seq: MR2084 (18) vs z3e (score = 182)

\[ z3e \] GGGGCTNACCCCACGACCCTCTCCCAGCCCCATGATCGAGAGCACCCCCATGTCCATTGCTGGCG
\[ z3e \] CCTGTNCCTCGGACTCTGTATTTCCCTGAGCTCCCAAGAGGCCTGGCCACTTATGGCCTCACTCCCTG
\[ z3e \] TGAAGCTCCAAGAGGCTGGCCACTTTATGCTAGGCTCCACTCCCTGCCCCCTCTCTTTTTTTGTCAT
\[ z3e \] TGAAGCTCCAAGAGGCTGGCCACTTTATGCTAGGCTCCACTCCCTGCCCCCTCTCTTTTTTTGTCAT
\[ z3e \] AAAGTGGGCTGAAATGACGTCTTTTTATGTCAGCCTGGATCTGGCCCTCTCCCTGCTGCTGCT
\[ z3e \] GGTGTGCAACCCAGGACAGCTGCTGCTGCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG
\[ z3e \] GGGGCTNACCCCACGACCCTCTCCCAGCCCCATGATCGAGAGCACCCCCATGTCCATTGCTGGCG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 9,18,19
END OF RECORD.
SEQUENCE FILENAME: kld.Seq
PRIMERS NOT PICKED
LENGTH OF INSERT: 101

ACTAGGCAAGGTTGAGGGGAGNGAGAACATGAAATTCAAGAGCCACACTAGCTCACAAAAGCCTGTG:
TGATCGCTTTTCACATCCCGGTCTGTTACCTTCTCGGTGGACTCGGAGTGTTTTCGGACAC:
CTCATTAATATGTCAGCAGAGTAGTTTCAGGCAGTAATTAATACAGTCGTCTCATCAAAGTCC

DUPLICATE REPORT FOR kld.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: K1F.Seq
FORWARD PRIMER: TGACTGGGAGTTGGACCC
REVERSE PRIMER: CCAGTGTCATGCAGGTCAG
LENGTH OF INSERT: 239

CTCGGGGGTCTATTGATGAGGGACTCTATTTGAGATGAGGCCACTGGAGGAGCCACCTTT
GAGCCGCCAGCTATCTGATGAGGATATGACTTCACCTCCCTGGAAN
TTTCTGTTGAAAAACTGCTAGTTTTGTTGTTCAGAGGGGACTGAGGACCTACCTGTAGGAAN
AAAATGTAACCTTGGACACGCAAAGAATCTCCCTGGACCTACAGTTTCTCTGGACAGATT
GATATGACGACTTCCTCTCCTAGCCCTATGAGCAGCTAAGTTGCTAGGCTAGTACCA
CATTACCGAGCTGAAAGACTCGGAGGAGCATTGACCTGGACCTACAGTG
CGGGGAATCTCATGATCATCACAGTGTCAG
GCCCCTTAGGAGTACTAGTAGTTCAGAGCCACATGGTAGGGGAAATCCTCATGATCATCACAGTGTCAG

DUPLICATE REPORT FOR K1F.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 19,22,26,27,28
END OF RECORD.
SEQUENCE FILENAME: K1OF.Seq
FORWARD PRIMER: CACCTCAGCTGGCCTTCA
REVERSE PRIMER: GGGCCTTGTGCCACACTTG
LENGTH OF INSERT: 265

CAGGGAGTCTGTGCGGATTCTATGGGAAGCCACGGTGTTGAATCTTACTTCAGGTCGTCACCTCAGCTGG
GTCCCTCAGACACCCCTAAGATCCCTTCGGTGCCACAACTTAGAATGAAGTCCAGCAGTGGAGTCGACC
CCTTCATCAAACATTTCCAGTCTTTACAGGGCCCTGCGGCTGCCACTCTGCTGCGGAGGCACTGATATGACGG
GGAAGTATGTGTAAAGGTCAGAATGTCCCGGGACCGACGGTGAGACACCCTCCGTGACTATACTGCCC
AAAGCCACCTTAGAAAAAGCAGATGATATCTCTCAATGCAAGCAGCCCAGACGTTGCACAT

TTTGGTGGAAATCTTTTCTCTGTACTACAGACGTACGGTTCTACACCCGTGCCGGGTCTCAGCTGTA

GTGACCAGTACAGACTAGAAGATGTGTATCTNTNAGGGNNTTTAAGATTGG

CACTGGGATGTCATGCAGATCTTGCTCACTACATAGANACTCCNAAAATCTAAACCC

DUPLICATE REPORT FOR K1OF.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18, 19
END OF RECORD.
SEQUENCE FILENAME: K11F. Seq
FORWARD PRIMER: TAAAGTGCGTGAGATTGACGT
REVERSE PRIMER: TCCCCGCTGCCCCAGAAC
LENGTH OF INSERT: 271

CTGTNCCTCGGACTCTGGTTACCTCTGAAGCTCCCAAGAGGCCTGGCACTTATGGCCTCACTCCCTG
GGACANGGAGCCTGAGACCAATGGAGACTTCGAGGGTTCTCCGGACCGGTAATACCGGAGTGAGGGAC
CCCCTC

DUPLICATE REPORT FOR K11F. Seq: MR2084 (18)
vs z3e (score = 182)
z3e  GCGGCTNACCCCACGACCCTCTCCCAGCCCCATGATCGAGAGCACCCCCATGTCCATTGCTGGCG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 9,18,19
END OF RECORD.
SEQUENCE FILENAME: kl2d.Seq
FORWARD PRIMER: TGGATTAGCCCTCTTGGGG
REVERSE PRIMER: GTATACCAATTTTTTCGATAAC
LENGTH OF INSERT: 289

ATHER PRIMER: open
REVERSE PRIMER: close
LENGTH OF INSERT: 289

DUPLICATE REPORT FOR kl2d.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19,26,28
END OF RECORD.

SEQUENCE FILENAME: K12F.Seq
FORWARD PRIMER: GTGTTGCTAGGAGCATGTTGT
REVERSE PRIMER: CTGGATCAAGGCCTGNCTGT
LENGTH OF INSERT: 318

DUPLICATE REPORT FOR K12F.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: 
['gp', 'Z35719', 'T04A8.14, emb-5 protein ...', 1.1e-05]
['sp', 'P34703', 'EMB-5 PROTEIN.>pir S35241 S35241 ...', 1.5e-05]
['gp', 'L21156', 'STE12 gene product [Kl...', 0.046]
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: k2d.Seq
FORWARD PRIMER: GCAAGGGGGAACCATTTAN
REVERSE PRIMER: AGGCTAGTGGTCTGTTGAA
LENGTH OF INSERT: 188

AAACACAGCAAGGGGAACCATTTANAGTCAATCTAGAAACAAGGGCTGGCAATTACGTAGCA
TTTTGTGCTCCCCTTTGGAATATGGTATGATCTCTGTGTCGACGGGACGTGTAACGATCTGTTCA
AAACACAGTCACACGCGACCACATAGCCAAANGGGAAGGGACAGNGNGANCATGAAATTCAGAGGAGC
TTTTGTCCAGTGCCTGGTGATCCGTTKNCNCTNGTACTTTTTAAGCTCTGCGTTGA
AGCTTCACAAAAGCCTGCTCTACTACTATGCAAGCTGTAGTTTCAGG
TGGAGTGTTTTTCCGACAGGATAAGTACATCGTCACATCAAGGTC

DUPLICATE REPORT FOR k2d.Seq: kld(8)
vs kid (score = 85)

AAACACAGCAAGGGGAACCATTTANAGTCAATCTAGAAACAAGGGCTGGCAATTACGTAGCA
kld

AGATAAAAACGATCACACGGGACCACATAGCCAAANGGGAAGGGACAGNGNGANCATGAAATTCAGAGGAGC

kld

GCCACACTAGGCTCACAAGACCGAAGGCTGGCAATTACGTAGCA

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: k3d.Seq
FORWARD PRIMER: GGATTTTTGTTGTTGAGGCA
REVERSE PRIMER: GCNTGGGTAGAGCAGGA
LENGTH OF INSERT: 207

CCACCACGTGCTAATTGGGATATTGTTTGGTGGAGGCAAGATGCTACGCTGTTAAGGCTGGCTGCTTGA
GATGGGTGTCACGATTAAACCACATCTCGTCAAGAAGCACACATGCAGAAGCAGGACGACGAA
AAATGAGGTGTTGGCTCTTCAAGATCAGGACTCTCTCTCTCCACGAAATATTAGAT
TTTNGAGATTTATCCTCGAACAGAAGATGCTGTGAGGAGGATTTTATAATCTCTNATC
AGGCCTGGTCTCCTACGACAGGCCCCAGCTAATCCTTATTTTATTTGATGACACGACGACGAGAAAC
TACGGAACACAGGATGTTGGCGAGATTAAAAATCAATCTGGTCTCCTT

DUPLICATE REPORT FOR k3d.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
SEQUENCE FILENAME: K3F.Seq
FORWARD PRIMER: ACTAATGGCGGCGGCTTG
REVERSE PRIMER: CTCTAAACACTGAAGTAATTA
LENGTH OF INSERT: 285

GAGCACTTTTAGGTGCTCTGATTACCCGTCCGCACCACCGGATCCAGACATTAACATCGTGAGTTTTCCAACTTCGTCCTAACAT
ACTTCAAGATCACCTAGGGCTCTATATTAGGAGGGGCTGTCAAGTTTCTGACGGGTGTTG TCAAGTTTCCGACCATTA
TACTTCAGTTAGTATTTGCGGCAATACGGGAAGCTTCGCCCTACCACACTGGGGGGAATTA
<<<<<<<<<<<
ATGAAGTCACAATCTCATAAACGGCTCGTATGCCCTTTCGAAACGGGGGGTGTGANCCCCCTATTT
AGCCCTGAA

TCGCGGACTT

FASTN REPORT FOR K3F.Seq: <no fastn matches found>
DUPLICATE REPORT FOR K3F.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19,26,28
END OF RECORD.

SEQUENCE FILENAME: k4d.Seq
FORWARD PRIMER: CTTAAACCACCTGAAGAGTCCA
REVERSE PRIMER: GAATATGTAACAGTGACTCC
LENGTH OF INSERT: 330

GGCTGGNGTAGCATTNANAGAAGCAAGGGGTCCGTNTTCTGTAACTACTACTGGNAGTAATGCCTGGNAC
CCGACCNCATCGTTAANNTTCTCTCGTTCCCCAGGCGANAAAGACATTGAGTATGACCNTCATAGGGACNTG
AGNATCTTAAATGGTAACCTTCTAACCACCTGAGAGTCTCCATCACAGTGGGAAAATCTGGCGCCGCTTA
TCNTAGGATTATCATTAGGAATTTGGTGGACTTCTCAGGTAAGTGTCAACCCTTTAGACCGGGGCCAT
CCGGAATCCTTCAGGGGAGCTGGTACATACATCTCTCTGGTACTACCTCGGGTATCCNAGGAG
<<<<<<<<<<<
GGCCTTAGGAAGTCCTCAGTGACAATGTATAAGATAGGACCCAAATGGGGCCCCGACCATAGGNTCCTC
GAACTTNCCGGGCNGCTCCTGGCNNAGCTGCTNCTNNGT
CTTGAANGGCCCGNCGAGGAACNNCCGNNTCGACNTCATNCCTGACNTCCACCCGAANAGGANGNCA
--------- ----------------------------------------------------
DUPLICATE REPORT FOR k4d.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: ['gp','Z35663','T04A8.15, Short region ...',0.0043]
['sp','P34703', 'EMB-5 PROTEIN.>pir S35241 S35241 ...',0.0047]
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: K5F.Seq
FORWARD PRIMER: TAACCAATCAGAGTGACGGA
REVERSE PRIMER: GGGNATGGTTTGTGCATC
LENGTH OF INSERT: 331

CTCTGCTCACCCTGTTCTCTGATGGCCACACCTGCTGGTNCTCCTCATATCTCTTTTCTCTCTCTCCCCA
GAGACGAGGCTGGACACAAAGATACCTCCGAGGAGCCANGGAGOTATAGAAGAAAAGAAGAAGAGGGGT
ACCCCTTCTCACCCTCTCTGAGGGACCCCATGAGCTGGAGGGGAATGCTGCTCCTCCCTCCTCTCTCTGC
TGSSGAGAAGGTAGAGGACACCCCTGGGTTACTGAACCCTGCCCTCGGAGGAGGGAGAGGAGAGACGS
TCAGTCATTGGCTGAATCAGCTCTTTATTAACCAATCAGAGGTGAGCAGGAATACATATTTTTGCATAAAC
AGTCACTAACGAACTTTAGTGAGAAATAATGGTTATCTCCACTGCTATTGCTTTAAAAAGCTATTTTG
TGAAAAGTTGAAAGCTTTAACCCTGAGCACCAGTTAAAGAAAGACATAGAAATTAGGCATTTTTGAGGCAT
ACTTTGACCTCAACGAAATTTTGAGGCTGCAATTTCTCTCTCGATCTTTAAATCACTGTA
CTGAGAGAAGTTGATGACAAACCCATCCCCCAAGAGGCTAATCCAAGTTCTTGA

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19,26,28
END OF RECORD.
SEQUENCE FILENAME: k6d.Seq
PRIMERS NOT PICKED
LENGTH OF INSERT: 178

CAGNNACCATCCCAANNACGATGACCCCTCGTCGGNNANCTCTCTCAAGNTNCTNTGTTGAGGCGCC

GTGGCTGATTGGGTATGCTGAGCCNACTGGNACNNTGAGAGTGCNCNANNNACTGCGCC

AGNNNTCTCTCTCATNTNTCTTNCNNNTCGGGNCTTTTGTATAAAAANNNNNTAAANTNNTNNGCCN

TCNCAGAAAGTNCNANANAGAGTNGACGNACCNNGAAGTACATATTTTTNNNNTANNANANNNTGNN

TACANNNANCTNTNCAANTAAAGACNNTPTAAACACACAGAC

AGNNTNNTGGAANANAGTNTATTCTGGNAAATATCCGCTTG

DUPLICATE REPORT FOR k6d.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS:
['S72304', 'rah=ras-related homolog [mice,...', 7.4e-29]
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7,9,18
END OF RECORD.

SEQUENCE FILENAME: K6F.Seq
FORWARD PRIMER: CTTTPTTTCTTTTTTTTAGGGGA
REVERSE PRIMER: GGCACTGTCTATCTGTCACAT
LENGTH OF INSERT: 257

TCCTGAAGACCAAGAGATTGTGTAAGGAAGAGAAAGAGATTTTCTTTCTTTTTTTTTTTTTTTTT

AGACCTCCGTCTCTCATTTAACACACTCCCTTCTCTTTTCTTAAAAAGAAAGAAAAAAAGAAA

TTTGTGGGGAACGGTGGTGCAAGGGCCTCTTAATCCAACATCTACGAACCGAAATACCA

GCTCAAGGATTCTCTAAGAGATGATGCTCTAGACGACGAATACCTTTAATCCAGACTTCA

CGAATCTCTCAAGAGCTACTCCACAGGCAGTACACGCTAACTGTTACGAGATGTTAGGCTCTGGAAGT

GCGAGNCCAGAGGACACAAAAGATTTAAAAAGATTCTACGCCCAAAACNTTG

CGGTCAAGGTCTCTCTCTCTCTCTCAGTCTTCGTCGGTGNNAC

DUPLICATE REPORT FOR K6F.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 19,22,26,27,28
END OF RECORD.
SEQUENCE FILENAME: k7d.Seq
FORWARD PRIMER: CTTAGTTAACTTGCCCTTGGCT
REVERSE PRIMER: CCCATAGNAAAGTGTCAGGAAA
LENGTH OF INSERT: 216

CGGTCCTCTTTACGGATGGCTCAGATCAGCAGGGGAAGACTCCGACTTCACTGGGGTTTTCTTCTTTAG

GCCCAAGAGATGCCCTTACCGAGTACTAGTGCTCCCTTCTGAGGGCTGAAGTGACCCCAAAAGAGAATC

TTAATTGCTCCTTGGCTTTTTTCTTTCAAGTTATGCTTTATCCACCATCTCTTCTGAGTTGCTCTT

AATTGAAACGGGAAACGGAAAAAGAATTCATACCGAAAGTGAATAGGTTGTAGAGGAGGAGTAGAAGAA

TCCTCTTCATTGTACATTTGCGAAGGGAGAGCTTTTGCCGACACTTTINCTAGGGACCTTTCAAGACC

AGAGGANAGAGANTATAGGAACTAGCAAGCTCAGGAAAGGACTGTGAAANGATACCCTGGAAAGATACAGGCGG

AGATAGTGA

TCTATACACT

DUPLICATE REPORT FOR k7d.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 9,18,19
END OF RECORD.
SEQUENCE FILENAME: K7F.Seq
FORWARD PRIMER: TAGCATGGGTCATCCC
REVERSE PRIMER: ATGACACTTCAGATCTCC
LENGTH OF INSERT: 267

CTGGTATTTGCCAACAATATCCAGGGTTCTGAGAATCTCTTTTAGGAAGGGGGTCTCAGGAGAG
GACCATAACGCTTGTTTATAGTGCTCAAGAAGCTATGTAAGGAATCCCTTCCCCCAAGTGAAGTAAT
GGCATACCTAGCATGGGATCCTCAATGACAGGATCGAGGAAGCTAATCGATGACCCAGGCAAGAG
CCGTATGGATCGTACCCATGGGGAGTTTTTAGTCTCCTAGTCTCCTTCCGATTGACTACTGGGCA
TGCCAAGACATCGACATGCCAGATGTCAATGGTGCGGCCCCCATCTGGAGAGCTCAAGTGAGA
AGAGTCTGGTAGTGTTACGTACCCACACCACCCCCGGGGGTAGAACTGCTGAGTCACCCCT
GTCGGCATACCTAGCATGGGTCATCCC

DUPLICATE REPORT FOR K7F.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: k8d.Seq
FORWARD PRIMER: CCACAAGTTGCATCTCC
REVERSE PRIMER: AGCATCCCGTACTGAGC
LENGTH OF INSERT: 382

GGCTTACCCCAGACTTGAAGCAAGCTTCTAGCTTCAGTCTTTAAATATTCAATCATGCACTGCCACTCC
CCGAA.TGGGGTCTGAACTTCGTTCGAAGATCGAAGTCAGAAATTATAAGTTAGTACGTGACGGTGAGG
AACCAAGTTGCATTTCCTTAAGATACCTTAGCTGTGCTTGAGCTGAAGCTAGGAACTATTGTATTACTT

TTTGCTCCTACCTTAATGGAATCGACACGAACTCGACTTCGATCCTTGATAACATAATGAA
TGTTGTCCTACTTGATGTTGCCTCCTGGGACCTAGACACCTTACAGCAGAGCTCGCTCAGTAACGGGA
-------------- ------------------------------------------------------
TTTTCTTCATTTGAAATACGGGGCGGGGACATCCCCCAAGATGCAGAATGCTGCCCAAGGNTTAGCCGGAA
AAAGAATGAAATTATGCCCCCGCCCTGTAGGGGGTTCTACGTCTTACGACGGGTTCCNAATCGGCCTT
----------------------------------------------------------------
GGAAGCTNGAAACGCATNAAGGGNGGATGGNCGNTGC
CCTTCGANCTTTGCGTANTTCCCNCCTACCNGCNACG
---------------------------------------------------------------------
AAAGAATGAAATTATGCCCCTGGGCTTCTAGCTTACGACGGGTTCCNAATCGGCCTT

DUPLICATE REPORT FOR k8d.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: K8F. Seq
FORWARD PRIMER: GCCTTTCTGCTCCCTTCTAG
REVERSE PRIMER: CACACATGGAAGGCAACGA
LENGTH OF INSERT: 333

AAAAAAAAAAAACAAACAAAAAAAACAAAGAGATGCCTTGAAGACTAGAACAGGCCTTCTGCTCCCTTTTTCAGCAG

TTTTTGTGTTGGTTTTTTTTGTCTCTACGGAGATCTGATCTGCTGCTGCCTTTCCTTTCTAGAGGCGGAAAGGTGCTGTC

CCCTAGGACACTGAGTGAGCCATAGGCAAGGTTGCTCTGCTCTTGCTCTTTCTCTTTCTCTTTAAGCTAG

GGGATCCGTGACGTCACTGCCGTATCCGTTCCCACAGGACGACGAGAACGGAAAGACGAGGGAAGATCT

ATTTTCCAGACGCTCTCAGACATGATGCTCACAGATGCCTTCTACTAG

TAGAAAGGTCTGAGAAGCTCTAATGTCGAGTGGCTACGGTTCTCTGATACCAACAGCGATGACTC

GATGGGGTGAAGATGTTGTAACACTCTTCTCTTGTGCTCTTTCCAATGTGACTGAGGCGCTTTT

GTACCACACTTACCTACCAAGATATGGGACAAGACGAAACGAAGTACTGACCTCCTCCGACGAAAAAAA

GAGGCATACTTTTTCTTTAGTTAACCCCTTTTGACTGTTGAAGCTTGCAAAA

CTTCTTAGAAGAAAAAGCTGGGTTTACCCCAATTGGGACCGAGGAGGTTTT

DUPLICATE REPORT FOR K8F. Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.

SEQUENCE FILENAME: k9d. Seq
PRIMERS NOT PICKED
LENGTH OF INSERT: 171

CTCTATATTACTTNCACACAGAGGGCGCTNTANAGNNCTANCTANACNACNATNOCGTACCTTTAC

GGATTAATGAAAGTGTGCGTCGCTCGGNNATNTGGNNNATNTGNNNTGNNNATNTGNNTGNTTCTT

ANCACAGGACATCTCTGAGATACGATCAGCTACNCTAATCTCGACTTNTGACTGTCN

TNTGCTCTGTAAGCAGCCNTAATGCTNTACNTACAGNTTNTACNTATGGNNTCAAGCTGAAATT

NATNOCCGCTNOCTNTAGTACCTCAGACGCAAGCAGCAG

NTANGGCGANGCGNATNTACGTAGATGNTNCNTNG

DUPLICATE REPORT FOR K9D. Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
SEQUENCE FILENAME: K9F.Seq
FORWARD PRIMER: CTAAGCAGATCGGCCGTAAGG
REVERSE PRIMER: TTGCTGAGATGTTTTATCA
LENGTH OF INSERT: 254

GTCACGACGCTTGGGCTGGTCAAGTCAAGCAGACAAAGGTTATGACTGATGCACTGAGTGTCATAGGCC

CAGTCGTCNCAACTCCAGCTTCGCTCCAGTTCGTCGTCGTTTCTGTGTGTTCTACTGAGTATCTCAAGCCTTATCCG

CTAAGCAGATCGGCCGTAAGGAGAGCGCAATGGGTTAGGTCAGTACCTTGAGCTTGATGTCATATGATGTCCTCTATC

GATTGCTAGCCGCCACCTTCTCGACTCTCGTGGCGAGTCCCACTCGATCGCCTACCCGCACTTCTTCTTCTTT

GTCTCAAGACGCTGAGCTCAACGATGGCAAGCGAGAGGTTAGTTGACTGATGTCATATGATGTCCTCTATC

CAGATTTTTCTGCTGGACGCTGATGTCTACCTCGTTGTCTGTCCACTATGCTGTTAGTGAGTT

CTGCTCAGTGGGCAAAACTATTTCCAGTCTTCATTTGGCAATNGGTGACTT

GACGAGTCATCCCCGTTTTGTTATACAGTTAACCNGGTANCCACTGAA

DUPLICATE REPORT FOR K9F.Seq: K8F (-13) vs K8F (score = 130)

K8F

GTCACGACGCTTGGGCTGGTCAAGTCAAGCAGACAAAGGTTATGACTGATGCACTGAGTGTCATAGGCC

AGGCTAAGCAGATCGGCCGTAAGGAGAGCGCAATGGGTTAGGTCAGTACCTTGAGCTTGATGTCATATGATGTCCTCTATC

CAGATTTTTCTGCTGGACGCTGATGTCTACCTCGTTGTCTGTCCACTATGCTGTTAGTGAGTT

CTGCTCAGTGGGCAAAACTATTTCCAGTCTTCATTTGGCAATNGGTGACTT

GACGAGTCATCCCCGTTTTGTTATACAGTTAACCNGGTANCCACTGAA

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: ml0a.Seq  
FORWARD PRIMER: AAGGGGACACCATTGTCATC  
REVERSE PRIMER: GCCAAATTNGGGTGCTTGAAC  
LENGTH OF INSERT: 259  
TCCTGGATGGTGCGGACGAAGGGGACACCATTGTCATCTTTCTGGTAGAGTGTCTCANGGATGAAGATG  
>>>>>>>>>>>>>>>>>>>  
AGGACCTACCAAGCTCTTCCCCCTGGTGGAAACAGTGGAAAGAACACTGCTCTAC  
ACCCCCCAATGCCACCTTTTTTACACGGAACATCAACANTTNNATAGGCACTGGTGACAG  
TCCTGGATGGTGCGGACGAAGGGGACACCATTGTCATCTTTCTGGTAGAGTGTCTCANGGATGAAGATG  
>>>>>>>>>>>>>>>>>>>  
AGGACCTACCAAGCTCTTCCCCCTGGTGGAAACAGTGGAAAGAACACTGCTCTAC  
ACCCCCCAATGCCACCTTTTTTACACGGAACATCAACANTTNNATAGGCACTGGTGACAG  
----------

DUPLICATE REPORT FOR ml0a.Seq: <no duplicates>  
NUCLEOTIDE BLAST HITS:  
['X03796', 'Mouse mRNA 5-region for aldolase C (...', 7.4e-48]  
['X06984', 'Rat brain mRNA for aldolase C (EC 4...', 2.2e-47]  
['M63656', 'Rat aldolase C gene, complete...', 8.2e-47]  
['X07292', 'Human aldolase C gene for fructose-1,6...', 1.3e-38]  
['X05196', 'Human aldolase C gene', 1.8e-38]  
PROTEIN BLAST HITS:  
[ 'sp', 'P05063', 'FRUCTOSE-BISPHOSPHATE ALDOLASE (E...', 1.1e-21]  
[ 'pir', 'A25388', 'fructose-bisphosphate aldolase (EC 4...', 1.1e-21]  
[ 'sp', 'P09117', 'FRUCTOSE-BISPHOSPHATE ALDOLASE (EC ...', 4.9e-21]  
[ 'gp', 'X06984', 'Rat brain mRNA for aldol...', 4.9e-21]  
[ 'sp', 'P09972', 'FRUCTOSE-BISPHOSPHATE ALDOLASE (E...', 4.9e-21]  
PHYSICAL MAP LOCATION: 22,27  
END OF RECORD.

SEQUENCE FILENAME: mlld.Seq  
FORWARD PRIMER: GGCACTATGGGCCTATCGC  
REVERSE PRIMER: ACACCACAGCAGCGCCAT  
LENGTH OF INSERT: 275  
TTTCGTA-CTCCTGGAAGCAGCTGCCGGGACAGGCTGGTCCTCACCATGCCACCTGCTGGCAGGCA  
CTATGGGCTATGCCCTACCCATGTTGCTCTGATAGCTGCGCTGCTGGCGCTGCTGCCTGCTGTGGCCCA  
ATTGCCCAACTAAAGGAACCTACTGTGCTCTTACACGACATGCTCGTTACCAGC  
GAGGCCCTGGGCCCTGTCACCCGCGTCTTTCCCATCATGCTTTTGGATGGGATACAT  
GAAGGCCCTACA  

DUPLICATE REPORT FOR mlld.Seq: <no duplicates>  
NUCLEOTIDE BLAST HITS:  
PROTEIN BLAST HITS:  
pir|A47714|A47714 Na+/sulfate cotransporter, renal - r... +3  174  7.8e-19  
PHYSICAL MAP LOCATION: 11,12,13,23  
END OF RECORD.
SEQUENCE FILENAME: m2a.Seq
FORWARD PRIMER: AAGGCATTCATGGTGAGCTA
REVERSE PRIMER: AGGAAAAAAATAACCAT
LENGTH OF INSERT: 227

GTTGTTITAGTTGGTGAAGACTTAAACACATAAGAA/TGTTCAAGGAAAGGATCTCTAGCGTACCTAGAAG
>>>>>>>>>>>>>>>>>>>>

CAACAAAAACTACCACCTCTGAATTTGGTGAATCTCTTACAAGTCCCTTTGGTAAGTACCCTGCTAAT

CGAGACTGCTGTAAGTGTAAGATCGTGAACTGCTACCCCTGGAGGACAATTCCAGAGCTAGGAAATGATCT

CTCTTGACGACATTCCACCTCGAGAGTTGGGACTTCCCTGGTATTAGTCTAGCCACTTCTTTTACTAGA

AAAGTTGTTAAGTACACGCAGAACGTTGGAATGGTAATTTTTTTTTCTTTTGTGCCACAGGCTAAATCAGAG
<<<<<<<<<<<

TTTCAACATTTCATGGCCTGCTACCTACCAATAAAAAAAAAGAAACAGGGGTCGACTTTAGCTTC

ATCAGTTCCCCAGACGAGG

TGTATAGACGCGGTCGCTTCC

DUPLICATE REPORT FOR m2a.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 22,27,28
END OF RECORD.

-- - - - -------------'----" ---- --- -- -- --- -- -- --- -- -- --- -- --

SEQUENCE FILENAME: m4a.Seq
FORWARD PRIMER: ATCCAGACCGTCAAGGC
REVERSE PRIMER: AAAAAAAGAAAATGATTCTT
LENGTH OF INSERT: 237

GGAAAAAGAAAGAAATCGTGAACATATCCTTAAACCGTTTGGCATTTTCTGAGACTGAATCAT
>>>>>>>>>>>>>>>

GGTCTGGCAGGTTCCGTTCTTTACAATTCGTCTCTAGAAGTAAAGACCTCGAGTCTTAACTGGGTCT

GCCTCGCTGCTCTAGGAATACGAGACGATACTTATTTAGCCCAAAGCCAGGGAAGTTTTATAAGAATCAT

CGGGACACACGCTCGTTAGGTCCGCTCCTCCTCAATAGAAATCCTGCTTTCCTCCCTAACATATTCTTACAT
<<<<<<<<<<<

TTTCTTTTTTTTTTTCTGATGGG

<<<<<<<<<<<

AAAGAAAAAACAAAAAGCTCAAGGCC

DUPLICATE REPORT FOR m4a.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
SEQUENCE FILENAME: m5a.Seq
FORWARD PRIMER: TCGGCACTACCCACCTATCT
REVERSE PRIMER: TCCTTTCAGCTCTTCCTCTG
LENGTH OF INSERT: 242

CTCGGCACTACCCACCTATCTCCTCCTCCCTCTGTTTTATTTCAAGTATGTTTTTCTAAAGGCTCACAGAC
GACCCGATGGGGATGATAGAACACAAATGATGTTACAAAAAGATTTTCGGACTCTGCTCTCT
TGTTACTGGTTGCTCTGAGGCTGCTTTTCAATATGATATAATGATGCTACGTCGAGGCGAGGTAGAG
ACCATGACAAACAGAGATCTCATCGTCACAAGTTATCTTCATATTACTTACGTGAGCCCTCCGCTCTCCT
AGAGCTGAAAGGAGGCCCAGCTCTATACCCAGAGACTAGTTAAAAAACAAAAAGCAGTGTAGAT
<<<<<<<<<<<
TCTCAGCTTCCTCGGGCTAGATATGGTGGTCTCTGATCGAATTTTGTTTTTCTGTACATACTA
CTTTCGGCACTACCCACCTATCTTGTCTCCCCCTGTTTTTATTCAAGTAGTTTTTCTAAAGCCTACAGAC
GAAGCCGTGATGGGTGGATAGAACACAAATGATGTTACAAAAAGATTTTCGGACTCTGCTCTCT
<<<<<<<<<<<
AAGAGCTGAAAGGAGGCCCAGCTCTATACCCAGAGACTAGTTAAAAAACAAAAAGCAGTGTAGAT
<<<<<<<<<<<
TCTCAGCTTCCTCGGGCTAGATATGGTGGTCTCTGATCGAATTTTGTTTTTCTGTACATACTA
CTTTCGGCACTACCCACCTATCTTGTCTCCCCCTGTTTTTATTCAAGTAGTTTTTCTAAAGCCTACAGAC
GAAGCCGTGATGGGTGGATAGAACACAAATGATGTTACAAAAAGATTTTCGGACTCTGCTCTCT
<<<<<<<<<<<
AAGAGCTGAAAGGAGGCCCAGCTCTATACCCAGAGACTAGTTAAAAAACAAAAAGCAGTGTAGAT
<<<<<<<<<<<
TCTCAGCTTCCTCGGGCTAGATATGGTGGTCTCTGATCGAATTTTGTTTTTCTGTACATACTA
CTTTCGGCACTACCCACCTATCTTGTCTCCCCCTGTTTTTATTCAAGTAGTTTTTCTAAAGCCTACAGAC
GAAGCCGTGATGGGTGGATAGAACACAAATGATGTTACAAAAAGATTTTCGGACTCTGCTCTCT
<<<<<<<<<<<
AAGAGCTGAAAGGAGGCCCAGCTCTATACCCAGAGACTAGTTAAAAAACAAAAAGCAGTGTAGAT
<<<<<<<<<<<
TCTCAGCTTCCTCGGGCTAGATATGGTGGTCTCTGATCGAATTTTGTTTTTCTGTACATACTA
CTTTCGGCACTACCCACCTATCTTGTCTCCCCCTGTTTTTATTCAAGTAGTTTTTCTAAAGCCTACAGAC
GAAGCCGTGATGGGTGGATAGAACACAAATGATGTTACAAAAAGATTTTCGGACTCTGCTCTCT
<<<<<<<<<<<
AAGAGCTGAAAGGAGGCCCAGCTCTATACCCAGAGACTAGTTAAAAAACAAAAAGCAGTGTAGAT
<<<<<<<<<<<
TCTCAGCTTCCTCGGGCTAGATATGGTGGTCTCTGATCGAATTTTGTTTTTCTGTACATACTA

DUPLICATE REPORT FOR m5a.Seq: <none>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 22,26,27,28
END OF RECORD.

SEQUENCE FILENAME: m6a.Seq
FORWARD PRIMER: TCAGTGAGGGCTCGCTCTG
REVERSE PRIMER: GTGNAANNNACCAGGATN
LENGTH OF INSERT: 211

CCTCAGTGAGGGCTCGCTCTGCCCCAAGNAAAGNATCNTCTGGGAGTNTAAAGCCAAATGCGGNTGNANT
GAGTCAACTCCCGAGACCGGCTTCTCTCTCCTANGACCCCTTCTCANATCCTCTTTACGNNACNNAT
GCCCTGAAGAAGNAGAGGAGAAGCCNMCNTNGNGCCAGATGCACAAGCTNAGGGNCTGNTATCCTGNNNT
<<<<<<<<<<<
CGGGACCTCCCTCCTCCTCCTCCTCGGGGCCACCCCMGGGTCTACGTGTGAGTCCNGCACTAGAACANNNNA
TNCACGCCTCTCGGGGAGGCTCTCCTNGCTCTCGTTGAGGCTAGGGGTATAATTCTCCACACTGG
<<<<<<<<<<<
ANGTCGAGACACCCCTCCGGGAGACGCCAGACCCGACTCCCGATCGCCCACAATAAGGTTGTGATCG
ACCC

DUPLICATE REPORT FOR m6a.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,22,23
END OF RECORD.
SEQUENCE FILENAME: m7a.Seq
FORWARD PRIMER: GGTCTCCAAGGAGACAAGAA
REVERSE PRIMER: GAACCAAGGAAGCATCCNC
LENGTH OF INSERT: 140

AGGGTCTCCAAGGAGACAAGAAAGTAAGGTGATTACACCCACAAATAGCCAGGTTGCCCACTAGGNCTCT
TTCCAGAGGTTCCCTCCCTTCCATCTCCATAAAATTGGGTGGTTATCCTCCAACCGGTGATCCNGGA
NCANTCCTNACCCTACGGGCTCAAGGGNACAGNAGGATGCTTCCTTGGTTCCAGCTGGGTACCAACTT
NGGTNAGGANTAGGGTGACCAGGCTGTCNTCCCTACGAAGGAACCAAGGTCGACCCATGGTTGAA

DUPLICATE REPORT FOR m7a.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: m8a.Seq
FORWARD PRIMER: AAACGCCACTGAGAGGTGAAT
REVERSE PRIMER: TAGACCAGGCTGTCCCCAGT
LENGTH OF INSERT: 237

GTCCTTGCAGCCTGTCAGCACAGCATACAAACGCACTGAGAGGTGAATTAGGGGAGTACTTGAGTTGA
CTCCAGAGGTTCCCTCCCTTCCATCTCCATAAAATTGGGTGGTTATCCTCCAACCGGTGATCCNGGA
NCANTCCTNACCCTACGGGCTCAAGGGNACAGNAGGATGCTTCCTTGGTTCCAGCTGGGTACCAACTT
NGGTNAGGANTAGGGTGACCAGGCTGTCNTCCCTACGAAGGAACCAAGGTCGACCCATGGTTGAA

DUPLICATE REPORT FOR m8a.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11, 12,13,22,23,27
END OF RECORD.
SEQUENCE FILENAME: m9a.Seq
PRIMERS NOT PICKED
LENGTH OF INSERT: 142

GGGACCATGGCGCCGCAGAGCTGCGGCACGNACCTAGCTCCGGNTCATGGTGCCAGTCATAGTCGTGT

CCCCTGTCACCCGCGGGTCTCGAGCCGGCTCACTGCTCTTTCTACCCCGTGGTATGCGAGAGAT

AATGAGCCCTGAGACACGGGGACCTGCTNNCTCTTCGATGGAAAGTGGCAACCGTATACGCTCTCTA

DUPLICATE REPORT FOR m9a.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS:
['X05196', 'Human aldolase C gene', 7.5e-20]
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 22, 27
END OF RECORD.

SEQUENCE FILENAME: kl0g.Seq
FORWARD PRIMER: ACTGCCATCTGGTGGAATTT
REVERSE PRIMER: CCACGCCTGGGTTAGCTGC
LENGTH OF INSERT: 234

CTAGACTTTGCTTCTACATCTTATAGCTGCTGAGGAACTGACTGCCATCTGGTGCAATTT

<<<<<<<<<<<<<<<<<<<<<<<<<

GATCTGAAACGAGGATGTAAGAATCCGACCCGTCACCACTCGTCTTGACTGACGGTAGACCACCTTAAA

CAGTGACTGCAGACGGTTTTTTATGCTTTTATGTGTAATAATGTTTGCTGGATGTATGTGTATCAC

GTCACTGACCTGCAAAAATGCGAGAAATACACATTTATTTCACAAAACGAACTACATAACAGAGATG

ATGGGCAAAAGGGCCAGAGACATCTAGCCGCTGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT

TACAGCTTCTGCTGTTCTCTGCTAGATCGCGGATGGGACCCGGGACCTCAAATGCCAGAACACT

GCAGCTGAGCCTGGAAATATAACTCGGAC

CGACACTGCACTTTTTAAATTGGGACCTCG

DUPLICATE REPORT FOR kl0g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 19, 26, 28
END OF RECORD.
SEQUENCE FILENAME: kllg.Seq
FORWARD PRIMER: TGGGCACCTGTGTTCTTTA
REVERSE PRIMER: CCCAGACTACCGGGGGGA
LENGTH OF INSERT: 279

GGGAAGGAGAGCCTTGTCTCATCGTGGGCACCTTGTTGTCTTTAGCATCTTTTGATGACCGCCTGGAAG

TACTACTCAAATGGAGTAAACCTCTCTCAAACCCCAAAATTCAGTTTTGCCTCTCTATTACCAGAAGAAT

GGNAAAAAGGGGGGCCATAGACCCCCCACTGGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

DDUPPLICATE REPORT FOR kllg.Seq: <zo duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19,26
END OF RECORD.

SEQUENCE FILENAME: kl2g.Seq
FORWARD PRIMER: CTTGTCAGACAGGTGCTGAG
REVERSE PRIMER: GCTCTTTAACAGGCTTGTTC
LENGTH OF INSERT: 33%

CCAGAAGCTGTGCTTGAAGGTGCCCGATACATGGTTGCACTACAGATTGCCAGGGTGCCCCTTGTCAGA

GTTCTCTGCAAGAACACTCCACGGGCTTCTGTACACACGACGACGGCACTGCTCTCTCCTCTCCCTCTCT

DDUPPLICATE REPORT FOR kl2g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: ['gp','Z35719', 'T04A8.14, emb-5 protein ...
['sp','P34703','EMB-5 PROTEIN.>pir S35241 S35241 ...
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: klg.Seq
FORWARD PRIMER: CCCTTTCAACACACTGTGGC
REVERSE PRIMER: TCAATACACGCCACAATGTC
LENGTH OF INSERT: 250

CCCCTTTCAACACACTGTGGC
>GCTCTCCACCAGATTTAGTGAGGGATGAGGGTTTCCTTGCTGTTCTGCTGCTGCAGC
<<<<<<
GCCCCAAAGCTCATGCTCATATGCTTCACATCCTTCCAGCATCTGGCATCCCTAGAAGGATGAGC
>>>>>>>>>>>>>>>

DUPLICATE REPORT FOR klg.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: ['T24569', 'EST144 Homo sapiens cDNA clone...', 2.1e-66]
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

-------------------------------

SEQUENCE FILENAME: k2g.Seq
FORWARD PRIMER: CCTCAGTCAGCATCTGGGC
REVERSE PRIMER: TCTGGAGAGAGGTGGTTCTTTG
LENGTH OF INSERT: 321

GGCCAACAGCTCAAGCACTATCCGCTTCACATCCACGACTGCTCGCATCTGGAACAGATGACG
>GCTCTCCACCAGATTTAGTGAGGGATGAGGGTTTCCTTGCTGTTCTGCTGCTGCAGC
<<<<<<
CCTGCTCGAGACTTACCGCAATAAGATAGAAGAGGAACCTGGGGTACCGGCTCAAGCGTACATCG
<<<<<<<<<<<<<<

DUPLICATE REPORT FOR k2g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: k3g.Seq
FORWARD PRIMER: CCTCAGTCACATCTGGGC
REVERSE PRIMER: GCTCAGAGTGTTTCAGCTACAA
LENGTH OF INSERT: 251

CAAGGCCAACAGCTCATGCATCTGGGCATCCCTGTTNCTCC
TTCCGGTTTGTCGAGTACGTGATAGGCGAAGTGTAGGAGTCAGTCGTAGACCCGTAGGGACNAGAGG
TGCGATAGTGACCCATGAGGCTTCTTATTCACAAGAAACTTCTTAGGTTATGACCTGAGGAC
TTAATTCTTTTTACATCGACTTGGAGACTCGTACCTTCAAAAGACCCATGCCACATACAGAAA
TTAACACCCATGGGAGGGGACAGATCCTACCGGCGTGNCNCC

DUPLICATE REPORT FOR k3g.Seq: k2g (11)
vs k2g (score = 64)

k2g

CAAGGCCAACAGCTCATGCATCTGGGCATCCCTGTTNCTCC

k2g

GCTCAGAGTGTTTCAGCTACAA

k2g

TTAATTCTTTTTACATCGACTTGGAGACTCGTACCTTCAAAAGACCCATGCCACATACAGAAA

k2g

CAAGGGTGTCAACATCACCCCCTGACACAAAGCCTCTGCTCCAGACTTGTACCNGATTAGCAAT

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: k4g.Seq
FORWARD PRIMER: TGCTCAATGTTGGAAAGTCGT
REVERSE PRIMER: CCATGATCCTCCTGGCTCAGCT
LENGTH OF INSERT: 293

ACGGTTAACCTGCCCATAAAACTCTGAGGAAGGAGGAGCTGCTCCCCTCTCCCAGTACATTGAGATGCT
TGCCCAATGGCGGTTATTTGGAATCCTTCTTCTTGGACAGGGGAGGCTGTTACTTCTACATACT
AACAAGTGGAAGCTCTGAAATTTAATATTTTGGGAGAAAAGTGTATTTCTATTCTGCAGAACACT
TGTTACACACCTTACGCAACCTAAAAATTTAATTTAAAAACCCCTCTTTTTAAATCAAGATAACAGCTTGTGA
CACAGTTTGGAGCTCTCAGTTTTTAGTTACAAAACAATATTAGCACAATGGAGAAGCTGAGCCAGG
GTTGCAATTGACGGGTAMvTGGACATCCTTCCCTCCTCGGACCAGGGGGAGGGGTCTGAACTTCATACGA
ACAATGG

DUPLICATE REPORT FOR k4g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18, 19, 26, 28
END OF RECORD.

SEQUENCE FILENAME: k5g.Seq
FORWARD PRIMER: GCCAATGAATTGGATCAATGG
REVERSE PRIMER: CTTTACTATACAGTCCTAAC
LENGTH OF INSERT: 237

TCAGGGTCAACAGCGGGCAAGTCGTTACACAATTTGATTAGCATATAAAAAAGAGCCATAGGATGAATATGGTT
AGTCCCAGTGGTTCGGGCACTATGTGTTTAGATTCGTATTTTCTCGGTTACTTAACCTAGTTACCAAAAGGTATTTGGAGTTATACTTGAGAAATTCATGAAGATTCACAGTAGGAGACAGTCAACACGCCCC
TTCCATTAAACCTCAATAGCAACTCAAGTTAAGACTTCTAAGTGTCATCCTCTGTCAGTTGTGGGGAAGTTGTCTTTTGGCAAGTTTTACTCTAGTrAGGACTGTATAGTAAAGCCTTGTCTCAAAAAGCCACAAA
TTCAACAGAAAAAATGAGATCAATCCTGACATATCCGACAACATTTCGGAACAGAGTTTTTCGGTGTTAAAATATTTTTAGATAATACAAAITTGTA
ATTTTATAAAAATCTATTATGTTTAAACAT

DUPLICATE REPORT FOR k5g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
SEQUENCE FILENAME: k6g.Seq
FORWARD PRIMER: CTTTGAGGTGAACCTTCAGG
REVERSE PRIMER: GCTGCAGTTGCTAGATGCCA
LENGTH OF INSERT: 165

TGACA/CTTTTGAGGTGAACCTTCAGG
ACTGTGAAACTCCACTTTGAAGTGACTTTGAGCTTCTCTCTCATCTGCAGA
CAATTTTGAGGTGAACCTTCAGG
GTTAAAAGGCTCGACCGTTATCGAAGGTAAGGCGTGGCCGGTGAAACCGTAGATCGTTGAC

DUPLICATE REPORT FOR k6g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
SEQUENCE FILENAME: mllc.Seq
FORWARD PRIMER: GTTTCCTAGTTGGGTTGACTGT
REVERSE PRIMER: TGAGAGAGAGAGAGAGANAGAG
LENGTH OF INSERT: 328

AGCAAGGTGGCTCCCTTCAAGGTAGCAGGCATTTCTCATATTCAGAAACCATGACACCTTTTGAGAACG
TCGTTCCACGAAAGGGGATCTCCATATTCAGAAACCATGACACCTTTTGAGAACG
CAGAGGGGAAAAAGCCCGACGTCCACCTTTCCAGGGGAATTTTGAGACCCCGACGTGGGGGACAGAC
GTCTCTGCTTTTTGGGGCTCAGGGTGAAGGCTCCCTTTAAAAACTTCCCCTGGGGCTCTCACCCTCTGTG
TGACAAAGATCTGAGATTGAAGACTTCTACGGTGAATCTTATGACTTCAGGAGGT
ACTGTTTCTTAGACCTCTATCTTGACAAAGGATCAACCCAACTGACACATTAGAATCATGAGTCCTCCA
GGAGAGAGAGAATCACAAATGTAAGGTCAGCCTGGGCTNGCATTTGTGAGACCCCACCTCTCTCTC
***********
CCTCTCTCTCTTTAGTGTGTTTACATTCCAGTGCCAGTGTNGGCTNGCATTTGTGAGACCCCACCTCTCTCT
TCTCTCTCTCTTCTCTCTCTCTCTCTCTTCACACACACACACCGGTCTTTACGG
***************
AAGAGAGAGAGAGAGAGAGAGAGAGAGAGGTGTGTGTGTGTCCCAATGCC

DUPLICATE REPORT FOR mllc.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: ml2c.Seq
FORWARD PRIMER: CAGCACCCAGTCCAATACTC
REVERSE PRIMER: GCTGATCTTTCAGATTGGAA
LENGTH OF INSERT: 177

ATAACCAGCCCCTGGGGCTTGCGCAGGACCCCATCTCAATCTTCTTTNCTTGGACTGACTATTCAGGGAC
>>>>>>>>>>>>>>>>>>>>>
TATGGTCCGGGAAACCCGACAGGCTGGGCTAGGTTACTGAGAANNGAACTGACATGATAAGCTCTTGG
TTCAAGGTAGAAATATCGAGATCCGGTCCCAGTGTCTGTCTCGAGTCCTCTCCTTGGTGATTCCACAAT
<<<<<<<<<
AAGTCATCTTCTTPAGTCTGTAGACAGGCCTACAGACAGACTGAGGGG
<<<<<<<<<
CTGAAAGATCACGCTCTTCTAGAGGATCTGTCACNTGGGG
<<<<<<<<<
GACTTCTAGTGAGATCCTTCTGACAGGTANCCC

DUPLICATE REPORT FOR ml2c.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 19,22,26,27,28
END OF RECORD.
SEQUENCE FILENAME: k8g.Seq
FORWARD PRIMER: AAGTTTGGCCTCACACCTGA
REVERSE PRIMER: AGCACAGCTTCTGGTGTAGG
LENGTH OF INSERT: 392

GGGCTAGATGGCCTGCCAAAAGTTTGGCCTCACACCTGAGCACTTGGGGAAGCACCTCCGTCGACGC

CCGACTACGGCAGACGGTTTTCTCAACCGGAGTGGGATGGGACACCTCATTGGAGGAGCACTTGCG

TACACGCGCACAGACACAGCATCCTCAAGACGGTTTGGGAGCAGTTGGACGAGGACACATTCG

ATGCTCCGCTGTCTCCTGCTCAGAGTTGCTGAAACTCCGACGCTGCTGTTGTAATGCGGGTGTGGCT

CAGTTCCCTACACCGAGACGCTDGTTGAAAGTCCCGGAGTACAGGTTGCACTACGATTTCCAGGGAG

GTTCAAGGGAAGGCTGCTGACCAGACACTCCACGCGGTATGTCAGTCATGACGACGTGGTCCTC

CCCCCTTTCGACAGAGCTGCTGAGCAAACTTCACGAGGACGAGCTAAACATACACACCAACCGAG

GGGACACAGCTCCTGACAGCTCCTGCTGAGCAAACTTCACGAGGACGAGCTAAACATACACACCAACCGAG

AAAGGTACAGGAGATGGTGATGAGGGCCCACTTTGGTACCTACCTTTTAAAGGAGACC

TTTCCATCCTTTCTCACCACCTCCTCCGCTGAGCTCAATGAGAAAAATTCCTATGAAATTCTCTG

AAGCCTNTTAAAGGCTAGAGACGACGCCAGTTCCCTCACAGGNTA

TTCTGAGAAATTTCCGAGCTTCTCTCGGGTCGAAGAGTAGTCTCGNAT

DUPLICATE REPORT FOR k8g.Seq: k12g (23)
  vs k12g  (score = 194)
    k12g
      GGGCTAGATGGCCTGCCAAAAGTTTGGCCTCACACCTGAGCACTTGGGGAAGCACCTCCGTCGACGC
    k12g
      CAGCTACAGCGCCAGACAGAGCAGATCCTCAAGACGGTTTGGGAGCAGTTGGACGAGGACACATTCG
    k12g
      GGGTACGCGTCTGCTGAGCAAACTTCACGAGGACGAGCTAAACATACACACCAACCGAG
    k12g
      AAGGCTTGCAGAGAAGGCTGCTGACCAGACACTCCACGCGGTATGTCAGTCATGACGACGTGGTCCTC
    k12g
      CAGTTCCCTACACCGAGACGCTDGTTGAAAGTCCCGGAGTACAGGTTGCACTACGATTTCCAGGGAG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
  ['gp', 'Z35719', 'T04A8.14, emb-5 protein ... ', 2.5e-19]
  ['sp', 'P34703', 'EMB-5 PROTEIN.>pir S35241 S33241 ... ', 2.6e-18]
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
SEQUENCE FILENAME: k5d.Seq
FORWARD PRIMER: GGGTTCCCTTCTGTTG
REVERSE PRIMER: CACATAGGCTTTGAAATCTC
LENGTH OF INSERT: 368

GGGAGAGCATCTGAATGAGTAGTGGCTTTTGTCTACTGTTGTAAAGCAGTTTCTGCTGGTTGCCAGCCC
CTCTCTGAGAGTGTTAGTGGGCTGTGCTCTCAGCACATGGCAGGAAGTGAGACCCAACA
AAAAAAAAAGAAGCTCTGCTTTCTCTCTGTTGTAGATGTTCGGTACACGGTCCGAGTCAAA

DUPLICATE REPORT FOR k5d.Seq: k10d (27)
vs k10d (score = 171)

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18, 19
END OF RECORD.
SEQUENCE FILENAME: k9g.Seq
FORWARD PRIMER: AGGGGGACTCTTACCATGAA
REVERSE PRIMER: TCCCTNCGGTGAGCATCGG
LENGTH OF INSERT: 351
CTGAGCTAAGGGAGTTATGTCGTCTGAGTTACTCTCGTAAGACTTTCTCTCTCTGACTAATATTTTCTGAGAGAGA
GACTCAGATTCTGCTCCATACGAGGCACAATGACGCGATTCGAAGGAACTGATTTCAAAAACGACTCCTCT
GATAAATTCTGCTTCTTCTAACGGCAGTGCTTCTCCACAAACAATGACAGCAAAAGGCGAGGGAGCTCTTACCA
CTATTACACACAACAGTCAGACGACACAATGACGGCTTTCAGGTCATGTCTTCTCACAAACAGTGCACAAAGGCACAGGG
CTGAGCTAAGCGAGGTATGTGCTGCGTGTTACTGCCGTAAGCTTCCTTGACTAAAGTTTGCTGAGGAGA
ACTCGATTCGCTCCATACACGACGCACAATGACGGCATTCGAAGGAACTGATTTCAAACGACTCCTCT
CTGAGCTAAGCGAGGTATGTGCTGCGTGTTACTGCCGTAAGCTTCCTTGACTAAAGTTTGCTGAGGAGA

DUPLICATE REPORT FOR k9g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.

SEQUENCE FILENAME: ml0c.Seq
FORWARD PRIMER: GTGGCAGACTGCCTTAGGG
REVERSE PRIMER: GTAGTGGGTACAAGTGACAG
LENGTH OF INSERT: 259
GCTGGGCAGAGCAAGGGCACTTAGCGTCGTCCGTCCACCTTTATGCTGTGGTGACTTC-GTCCCTGNTGTGGC
CGACCGTCTCGTCTCGTCGACTTAGGACAGACAGCTAAATAGCACACACTGGAACCACACCCAGAGAGCA
AGACTGCTTTAGGGGCAACAGCGCCCTGAATACGACCAGAAACCTCAGAAAATTTAGGCTTGTGAC
TCTGAGGCAGACTTAGGCTCATTACTTTACTGGGCTGTGAATACGACACCCAGCCAGAGAT
GAGCCCCCTCTCTCAAAGACAGACAGATCGTGTGTTGCTGTCGTCTTTATTTTCTTTTTCCTGAGCACG
ACCCACTACAGGGCAACATCGACACATCAGATTGCAGSGCAGCTCTACCAAGACAGAGAGAGAGAGGA

DUPLICATE REPORT FOR ml0c.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
SEQUENCE FILENAME: mlc.Seq
FORWARD PRIMER: TCTGTCTGACTGCCTGGCTC
REVERSE PRIMER: CCACTGCTCAGGGACTCTGC
LENGTH OF INSERT: 372

GCCAGNGTTTNTGGGAACGCGACTGAAGAGAAGAACGAGAGAAATNACGTTCCTTNAGCCATTTGCC

CGTGCTGAANACCCCCGGTACCTTCTCGCTGGCTCTTTAAATGCAAGAANYTGCTGGTACCCG

ATTCCCATGCAATGCCCAGGCTGACCCCTAATTAAACCTCAGACCTGTGCTGACTGCCCTGG

TAAAGGTTACGGAGAATCAGAAACCAGCGAGAGAGGAAGTAAATAGTATGAATAGACACTACGAGC

CTCTGACCAAGGTGGTGCTTCTGCCTGAAGATTTACCCCGAGCCGGCCGGCCGCTTCCCTCAG

GAGACTGGTCAACACACAGAAGACAGACCATGGGTCGCCGACGGGACGGGAGGTGCC

GNCCTCTTCAGTGGCTAGTTAGTGACAGGACCAGTGACAGTGCCCTGTAGGCTGACCTCAAG

CCCGAAGATGACGAGACGACTCCCAATCAGCTCCCTGTTGCTAAGTCCAGGACTGCTGGGAGTTTC

ACCTCGAGGGGGGTCTCACCTGACGCCTGCTATGGCAGAAGCTTTCTCAACAAAGGCGAGAGTT

duplicate report for mlc.Seq: <no duplicates>

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.

SEQUENCE FILENAME: m2c.Seq
FORWARD PRIMER: ACCATGCCTACTGAGGCACG
REVERSE PRIMER: GTAAGACATGAGATGTGA
LENGTH OF INSERT: 263

GGGCAGAGATGAACGGGCTGCTAGCCCAGGCAGATATGAAGGCAGTGGAGATGGCCGGANNNCAGCACA

CCCGTCTCTACTTGGCCAGACGGGTCTCCCGCTCTTACCTCTCGTCTACCCCTGCTCCTACGCTGGC

GTCCCCACATGCAACAGCAGCTACTGAGCGGACGCGCTGCTTACCGGACCACTCCTGCTGGCTCGAAT

CAGGGGATGAGCGGTTGCTGAGTCCAGTAGCTGACTCGACTGACGACTGAGTAGGGTCCTCGGACCGGATCGTTA

GCACATCCCATTGGCTCATAGCTCAACCGCCCGCCGGGGCGCTCTATAGCTCAGAATATCTTATGC

CGTGAGGGGTTAAACCCGATATCTAGGGTCGCTCTGGCGCTCTATGTTATTCAGAAGAATAGG

AGCGCCAGTCTCATCAGATCTCAGTGTTCAATGCGGACAGCAG

duplicate report for m2c.Seq: <no duplicates>

NUCLEOTIDE BLAST HITS:
["X06984", 'Rat brain mRNA for aldolase C (EC 4.1...", 1.4e-52]
["M83656", 'Rat aldolase C gene, comple...", 1.8e-52]
["X05277", 'Rat mRNA for aldolase C", 1.3e-30]
["X07292", 'Human aldolase C gene for fructose-1,6...", 3.8e-25]
["X05196", 'Human aldolase C gene", 6.4e-25]

PROTEIN BLAST HITS:
['gp', 'X05277', 'Rat mRNA for aldolase C ...', 6.7e-09]
['gp', 'X06984', 'Rat brain mRNA for aldol...', 4.2e-08]
['pir', 'S00326', 'fructose-bisphosphate aldolase (EC 4....', 4.3e-08]
['sp', 'P09117', 'FRUCTOSE-BISPHOSPHATE ALDOLASE (EC ...', 9.3e-07]
['gp', 'M63656', 'aldolase C [Rattus nor...', 9.3e-07]

PHYSICAL MAP LOCATION: 22,27
END OF RECORD.
SEQUENCE FILENAME: m3c.Seq
FORWARD PRIMER: CTGTGTTGGTTTTCCC
REVERSE PRIMER: CTGAGATAATAGAGTATGGC
LENGTH OF INSERT: 226

CCCAGAAGGGCCAAGTCAAGAGCTCTTCCGTTCAATCTCTTGTTTTGGGTGCTCCCGATATGGAACCCATC

GGGGCTTTCCGGGTCCAGTCTCTGAGACAGAGGAACCCAAAAACAGAGGGCTATACCTTGGGTTAG

AGGCTCGGGGCAAGGGTTTTAGAGAAAGAAAAAAGACAGAGGTTGAGTTGGAGTAAGCTCCATTCCCTTCTCGC

TCGGAGTCCCCCTGTTCTCAAAACCTTCTTTTAGCAACATATGCTTAACCTTTAGGGGTTCTCTGTCACAA

TCCCCGCTACATTTGGAACCTCGCTCATATATTAGTTACACAGACAGGGTTCTCGG

CAGATGCCCCGTCAGCCACC

GTCTACGGGACCAGGGTG

DUPLICATE REPORT FOR m3c.Seq: <no duplicates>

NUCLEOTIDE BLAST HITS: <none>

PROTEIN BLAST HITS: <none>

PHYSICAL MAP LOCATION: 11,12,13,23

END OF RECORD.

SEQUENCE FILENAME: m4c.Seq
FORWARD PRIMER: TGCTCTCAGAGTGAAGTAA
REVERSE PRIMER: GTGGAGACTAGGAAATAATT
LENGTH OF INSERT: 333

GAGGAATTCAGTCAGAGCTGATCTCAAGGGCTAAGGGTGAAGAGAAGTGGCTAGATCCTTTCCGGGATACCA

CTCTAAGAGGCTAGCTCTGAGACCCACCTTCCCTCTCTCAGCTACAGTGCTTCTGATCTTATCTCTAGGGT

GCCTAATAGGGCTTCAGGTGGCTCTGGTTGAGATTGAGAGGACACGAGCTCTCTCTCTGGGTAACAGGCC

AGATTCACCGCTCTCTCTATGCTACCTACTTTAATATAGGCACAGTGCTCTCTCTCTGAGCTTTACCTGGC

TTTCCTAACTCGCCCAAGAGGGGG'TGGGACCCGTCGTAGATTAATAAAGGATCAGAGGTCCTrCTCTG

TTCCAGAATTCCTCGAGGCTACAGGTGATTTAAGCTGAGCCCTGCCACCTGGAGCTGGGAGATGTGTCAG

GAAATTTCGGCAGTAAAAGCTGAGGACACTTGAGTCGGTATGAGATAATAGAGTCCATGTCGTGTCCCAGGAC

AGAGCTGATCACTAAGGTGACGATCGATTGCCTTCCAGAGTGAGATTAACATCTGCCTCTAGAG

GAGGAATTCCAAGTCAGAGCTGATCACTAAGGTGACGATCGATT

DUPLICATE REPORT FOR m4c.Seq: MR2058 (9) MR2102 (4)

vs MR2058 (score = 70)

MR2058 GTGCGGGGCAAGTCGACAGGCTCGGAGCTGCTACACCTCTCCAAATGAGGGAGGAAGCC

GAGGAATTCAGAG

MR2058 AGTNATAGGAATCTTTTCCCGCTAGACAGGGTCTGCTATCTGATGGGCGTGAGGAAATTCAGAGTC

AGAGCTGATCATAAGGGTCAGAGTAGAAGGTTGCAAAATTACAGCCTGGCTCTTAGAG

MR2058 AGAGCTGATCATAAGGGTCAGAGTAGAAGGTTGCAAAATTACAGCCTGGCTCTTAGAG
vs MR2102 (score = 58)

GAGGAATTCCAGTCAAGAGCTGATCACTA

MR2102 TCCGGCTCAGACAAAGGCTGCTATCTAGGGGCTGGAATTCCAGTCAAGAGCTGATCACTA

AGG-TGACGATCGATTGCCTTCCAAGAGTGAGAT-T-TAACATCTGC-C-TCTA-GAGGTG-GCAT

MR2102 AGGGTGACCNTGATTGCCTTCCAAGAGTNAAANTC-T-TAACATCTGC-C-TCTA-GAGGTG-GCAT

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
SEQUENCE FILENAME: m5c.Seq
FORWARD PRIMER: AAGTGCCCTTGCTCTGCC
REVERSE PRIMER: GTCTTTTNCAGGTGCCCG
LENGTH OF INSERT: 310

---

TNCTTGCTTTTTAGAAAGGGCCTNGCCTGCAAGNCCTAAACTTTTCTTANAGTTTATGTCTTATTTACAGN

ANAAACAGAAAATCTCTTCGGANGGCAGCAGTTTNGCAGGTGGAAAGANCCTAATACAGATTTAAGTCN

GCNGTGTTGCCCTAAGNCAGTCTGCCACACATCAGGGACAAAGTACACACAGCATAGTGGAGAGC

CGNGCACAAGGGGATCNGCAGCTGCTGATGTCCTCCTGTTGAGGTGAGTTATTTCTCACTAGCTG

NCTAGTGCCCTTGGCCTGCAGCTGTTAAGNCTTCTTACAGTCCTGAGCTTACAGNAGATG

NATTCAGGAACAGAGCGGTGGAAGGCCTACTCNAGGAGAAATGCGAAGATGCAGTTACAGACAT

GCCCACCACAAGGCCCTTTATGAGGGCCAGAGGGACTTNGTCAGACGGTGTAAGTCCCTGGTTAGTGGTGTCGTATTTCCACTAGCCTGC

NTAAGTGCCCTTGCTGCTGCAAGNCCTTATACAGTCAACGTCCTTGATACANGATGGCCGCGCACAANGGCTATGGAGCCCAGAGAGC

GGCGCGGTTINCGAATAACTCTGGGTCTCTCNCTCCTGCCACTGAGTCCCTGNATTTTGCAGAAAAGCNC

GGACTCTNNAAAAGCACACCTTTTGAAGCTGAGCA

<<<<<<<<<

CTGGANANNTTCTGGTAAGAANCGGAGCTGACCTOP

DUPLICATE REPORT FOR m5c.Seq: m10c (-14)
vs m10c (core = 140)
m10c CAAGTTGCAAGGGGATCNGCAGCAGATTAGGGTCCTCACTGAGCTTACAGT

TNCTTGCTTTTTAGAAAGGGCCTNGCCTGCAAGN

---

m10c AGGGCAAGAAGACCATATACATGCATGAGGGGCACTGCTTTTAGAAAGGGCCTGCTGCAGN

NCTAAGATTGTTTTATGGCTTAAATTCAGNCAGTGGTGACTCTTANAGNAGTCTGCCTAGGAGGC

m10c GCCTAAACTTTTTCTGCTGTTTCTGCCTGATTCAGGCAGGCTGCGTGTCCCTAGAGCTGCAC

ATCAGGGGACCAAGTCCACACAGCATGAAAGGGTAGTGATCCAGACGGTAAAGTGCTTGGCTTCCCGACC

m10c ANCAGGGGACCAAGTCCACACAGCATGAAAGGGTAGTGATCCAGACGGTAAAGTGCTT

m10c CTGGCTTTGAAGNCCCTTTTCAAGCTCCTTGGATACANGATGGGCCGCGCACAAGGCTTTATG

m10c GAGCCCAAGAGGGGOTGACTNAGGCCANTAAAGNGCTTTTGCGGACCTNNGAAGAACC

---

m10c AGTTTTAGGCGAGCA

---

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
SEQUENCE FILENAME: m7c.Seq
FORWARD PRIMER: AAAAGTGCAATTGGGGG
REVERSE PRIMER: CCAGGCAACCACAGGAC
LENGTH OF INSERT: 185

GAAAAAGTGCAATTGGGGGTCATCTTCTCTCTCTACGACATCAAGCTACAGGATGATATGCTGCCC

CTTTTCCATCTACACCTACTATAGAGGTACTCTGGAATCTTGCTTTCTAATTACCCAGGG

CTCCGTACCCATACCCAGGATAGGGCTTCTCTAGGCTCATCAAGGTACACAGTCCCTGGTGGTGCCTGGGGGCTAGGTGTGGAAACCAGTNTATATAAGAGGGTGACAAGCCAG

GAAGCAGGCGTGAGGTTGATCTATTCCCCATTAGAGCATTGCCGTAAGTCCATGTGAGTGGCAGGGACCGAGAGGAGC

GAGGGCTTAGGTTGGAAGCCAGTTTATTTAAGAGGGTGACAAGCAG

CCCCGATCCCACACTTTTGCTGACANATAMTCTCCCACTGGTCGC

DUPLICATE REPORT FOR m7c.Seq: m10a (-13) vs m10a (score = 85)

m10a TTGNCCNGANTATTGCTCNACGGATCGTGACCCCCCCGCCAAGGGNATTCTGGCTGCACATGAGTC

m10a CGCAGNCAGCTATGGCCAAAAAGCTGAGCCAAATTNGGGTGTAGAACACTTNGGANAATCGNCGN

m10a CTGTACCGCCAGNTCCTATNNANTGTTGATGTCCGTGTGAAAAAGGTGCATTGGGGGGGTCATCT

m10a TCATCCNTGAGACACTCTACCAGAAAGATGACAATGGTGTCCCCTTCGTCCGCACCATCCAGGAAGGGCATTCTCGTAGGCATCAAGGTACACAGTCCCTGGTGGTGCCTGGGGGCTAGGTGTGGAAACCAGTNTATATAAGAGGGTGACAAGCCAG

m10a CCAOINTATATAAGAGGGTGACAAGCCAG

NUCLEOTIDE BLAST HITS:
['M63656', 'Rat aldolase C gene, comple...', 1.8e-54]
['X03796', 'Mouse mRNA 5-region for aldolase C (...)', 4.9e-42]
['X06984', 'Rat brain mRNA for aldolase C (EC 4.1...)', 1.2e-39]
['X07292', 'Human aldolase C gene for fructose-1,6...', 3.0e-34]
['X05196', 'Human aldolase C gene', 3.0e-34]

PROTEIN BLAST HITS:
['sp', 'P05063', 'FRUCTOSE-BISPHOSPHATE ALDOLASE (E...)', 4.8e-24]
['pir', 'A25380', 'fructose-bisphosphate aldolase (EC 4...)', 4.8e-24]
['sp', 'P09972', 'FRUCTOSE-BISPHOSPHATE ALDOLASE (E...)', 2.0e-23]
['pir', 'A25861', 'fructose-bisphosphate aldolase (EC 4...)', 2.0e-23]
['gp', 'X07292', 'aldolase C [Homo sapiens]', 2.0e-23]

PHYSICAL MAP LOCATION: 22,27
END OF RECORD.
SEQUENCE FILENAME: m8c.Seq
FORWARD PRIMER: CAAGTGCTGGGGAAGTCTAG
REVERSE PRIMER: CCCAGGGCTCCTCCACAG
LENGTH OF INSERT: 230

AGCCCTCAGGGGAGTGGCGCTTGCCGAAATCTAAGTGGGGAAGTCTAGGCTAGCTGCAAGGAG
TCGGATCACTCCCGAGCCACGCCGTACCTGGGCTTACGTACGACGACGCTGCTGCTTACGGAC
CTAGCCCTGGAAGGAGGAGGGAGGC'CTGGGGACCCAGATGCACAAGCTCAGGCCTGTNACCTGGTC
GATCAGGAGAAGCGGAGGAgGACCGCTGCAAGGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG
AGAACCCTGCTCCATNCACAAGAG
TCTGGGACACAGGTGTTTCT

DUPPLICATE REPORT FOR m8c.Seq: m6a (15)
vs m6a  (score = 165)

m6a

AGCCCTCAGGGGAGTGGCGCTTGCCGAAATCTAAGTGGGGAAGTCTAGGCTAGCTGCAAGGAG
TCGGATCACTCCCGAGCCACGCCGTACCTGGGCTTACGTACGACGACGCTGCTGCTTACGGAC
CTAGCCCTGGAAGGAGGAGGGAGGC'CTGGGGACCCAGATGCACAAGCTCAGGCCTGTNACCTGGTC
GATCAGGAGAAGCGGAGGAgGACCGCTGCAAGGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG
AGAACCCTGCTCCATNCACAAGAG
TCTGGGACACAGGTGTTTCT

m6a

AGCCCTCAGGGGAGTGGCGCTTGCCGAAATCTAAGTGGGGAAGTCTAGGCTAGCTGCAAGGAG
TCGGATCACTCCCGAGCCACGCCGTACCTGGGCTTACGTACGACGACGCTGCTGCTTACGGAC
CTAGCCCTGGAAGGAGGAGGGAGGC'CTGGGGACCCAGATGCACAAGCTCAGGCCTGTNACCTGGTC
GATCAGGAGAAGCGGAGGAgGACCGCTGCAAGGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG
AGAACCCTGCTCCATNCACAAGAG
TCTGGGACACAGGTGTTTCT

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,22,23
END OF RECORD.
SEQUENCE FILENAME: m9c.Seq
FORWARD PRIMER: ATCCAGACCGTCCAAGGC
REVERSE PRIMER: AAAAAAGAAATGTTCTTT
LENGTH OF INSERT: 239

GAAAAAAAAAGAAAAAGCTGAAAACCTCGAACATATCTTTAACCCTTGGACATTTCTCTNGACTGAGTACAT

CTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTACTGCTGATATGATTTGCAAAACCTGAAAGCTGAGTAATAG

CCAGACCGTCCAAGGCAGAAATGTAAGCAGAGATCTTCAACTCTGGAGCTCAGAATTGAACCCAGAGGCTCTGGCA

GGAAAAAGAGAAAAAGCTGAAACTCTGGACATATCTTAAACCGTTTGGCATTTCCTNGACTGAGTACAT

TCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTACTGCTGATATGATTTGCAAAACCTGAAAGCTGAGTAATAG

CCCTCTGCTGCTAGGAAGTGACAGCACTTTAGCCCAACGCCGAGAGTTTTATATATCCAGTCGAGGG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis

Nikolai A. Lisitsyn1, Julia A. Segre2, Kenro Kusumi2, Natalia M. Lisitsyn1, Joseph H. Nadeau1, Wayne N. Frankel1, Michael H. Wigler1 & Eric S. Lander2

We describe a technique, genetically directed representational difference analysis (GDRDA), for specifically generating genetic markers linked to a trait of interest. GDRDA is applicable, in principle, to virtually any organism, because it requires neither prior knowledge of the chromosomal location of the gene controlling the trait nor the availability of a pre-existing genetic map. Based on a subtraction technique described recently called representational difference analysis, GDRDA uses the principles of transmission genetics to create appropriate Tester and Driver samples for subtraction. We demonstrate the usefulness of GDRDA by, for example, successfully targeting three polymorphisms to an interval of less than 1 cM of the mouse nude locus of chromosome 11.

Positional cloning, the isolation of genes based on their chromosomal location without prior knowledge of their biochemical function, is a powerful general approach that is applicable, in principle, to any organism1. Its actual use, however, has been much more restricted. Positional cloning depends on the ability to find tightly-linked genetic markers near a locus of interest, and hence the method has been practical only in the handful of organisms for which dense genetic maps have been constructed — principally, the fruit fly, nematode, mouse and human. For most organisms, genetic maps are either nonexistent or too rudimentary to allow routine positional cloning. To make positional cloning broadly applicable, one would ideally want a method for directly generating tightly-linked markers without recourse to a pre-existing genetic map. Here, we describe such a procedure, called genetically directed representational difference analysis (GDRDA).

Our method is based on a recently described subtractive technique called representational difference analysis (RDA) for identifying differences between two DNA samples, referred to as Tester and Driver4. Specifically, RDA is designed to clone restriction fragments that can be amplified by the polymerase chain reaction (PCR) from Tester but not Driver — either because the corresponding sequence is completely absent from the Driver due to a homozygous deletion or because it is contained in a small restriction fragment in the Tester but not Driver. Thus, RDA can produce clones that detect restriction fragment length polymorphisms (RFLPs) between Tester and Driver.

To generate genetic markers linked to a trait, it is not enough simply to apply RDA to samples from a single affected and a single unaffected individual in a population or family. The abundant genetic variation among even close relatives in most populations, will mean that polymorphisms will likely be found throughout the genome. One requires a way to find polymorphisms specifically in the vicinity of the gene of interest. To ensure this, one needs Tester and Driver samples with the property that the Driver contains all of the alleles present in the Tester except in the region surrounding the target gene. As we describe below, such samples can be constructed by using classical transmission genetics. Although the methods are most easily applied to organisms that can be bred, they are applicable to natural populations as well.

Here, we describe two specific implementations of GDRDA. The first involves using congenic strains, while the second involves using progeny from an appropriate cross or pedigree. We tested the methods by using them to produce genetic markers linked to various mouse mutations and found them to be remarkably effective: of the one-third of clones that passed a simple initial screen, all (6/6) mapped to the desired region. Using congenic strains, genetic markers were produced near pudgy on chromosome 7 and tottering on chromosome 8. Using progeny from F1 intercrosses, genetic markers were produced near nude on chromosome 11 and staggerer on chromosome 9. The GDRDA experiment with nude was aimed at finding polymorphisms within an interval of less than 1 cM around the locus. Three clones were produced and all mapped with 0.2 cM of nude, which comprised less than 1/2,000 of the mouse genome.

GDRDA with congenic strains
One ideal substrate for RDA would be a pair of congenic

1Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, New York 11724, USA
2Whitehead Institute for Biomedical Research, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Cambridge, Massachusetts 02142, USA
3The Jackson Laboratory, Bar Harbor, Maine 04609, USA

Correspondence should be addressed to E.S.L.
strains' in which a particular gene has been transferred from one genetic background onto another by successive generations of backcrossing and selection. Congenic strains will be genetically identical except in a relatively small region surrounding the gene of interest. The region will typically be small enough to permit chromosomal walking to the target gene, but large enough to contain polymorphisms detectable by RDA. (RDA can detect only the minority of polymorphisms that cause gross differences in restriction fragments and thus, for example, comparison of isogenic strains that were identical except for a single mutation would likely fail to yield an RDA polymorphism.)

To test this implementation of GDRDA, we turned to the laboratory mouse, for which congenic strains have been developed for many interesting mutations. We selected congenic strains for Lurcher (Lc), severe combined immunodeficiency (scid), pudgy (pu), tottering (tg), stargazer (stg) and nude (nu). The congenic strains were constructed using between 11 and 40 generations of backcrosses (see Methodology for details of the strains).

RDA was performed in each case (see Methodology), using one of the pair of congenic strains as Tester and the other as Driver. Briefly, the first step involves preparing ‘amplicons’ from the Tester and Driver, which entails digesting each sample with a restriction enzyme, ligating the restriction fragments with a compatible adaptor, performing PCR using a primer complementary to the adaptor, and finally removing the adaptor by digestion with the original restriction enzyme. An amplicon contains only a portion of the genome, as it includes only small restriction fragments that are preferentially amplified. The Tester amplicons are then subjected to multiple rounds of hybridization-extension-amplification in the presence of excess Driver amplicon, under conditions favouring amplification of fragments present in the Tester amplicon that lack corresponding fragments in the Driver amplicon. Consequently, this procedure should yield small amplifiable restriction fragments which are present in Tester amplicons but absent or reduced in Driver amplicons. In these experiments, the restriction enzyme BglII was used and three cycles of hybridization-extension-amplification were performed. The resulting difference-products were separated by agarose gel electrophoresis. Several strong bands were visible upon staining with ethidium bromide, as well as a weak background smear (Fig. 1).

For each experiment, we cloned the difference product and selected six clones at random. We initially identified clones with distinct insert sizes (a total of 18 clones from the six experiments) and then characterized the clones by hybridizing them to Southern blots containing the Tester and Driver amplicons, to identify which clones showed the desired property of detecting a fragment in the Tester but not the Driver amplicon (Fig. 2). Of a total of 18 clones, this rapid test eliminated 15. The ‘failures’ could be grouped into three categories: First, seven clones detected a high-copy repeat in both Tester and Driver. Second, seven clones detected fragments in both the Tester and Driver amplicons. Finally, one clone failed to detect a signal in either Tester or Driver amplicon. Interestingly, all clones whose insert sizes did not correspond to one of the clear bands visible in the ethidium-stained difference product (11/18) failed the initial characterization. With a single exception, this was also true for the experiments described in the next section and suggests that this criterion might be useful for eliminating clones directly. Three clones (one each for pudgy, tottering and stargazer) showed the expected behaviour of hybridizing to the Tester but not the Driver amplicon. These three clones were then hybridized to Southern blots of Tester genomic DNA (as opposed to amplicon DNA) digested with BglII to determine whether they detected a unique genomic locus. Two clones (RDA-4.5 for pudgy and RDA 8.2 for tottering) detected a unique locus, whereas
Fig. 2 Autoradiograms obtained after hybridization of probe RDA-4.5 (a) from *pudgy* congenic strains and probe RDA-8.2 (b) from *tottering* congenic strains to Southern blots containing *BglII* amplicons from Tester (lane a) and Driver (lane b). Sizes (bp) are indicated to the right. The faint band above the major DNA fragment is an unidentified PCR byproduct frequently observed on blots of *BglII* amplicons.

one clone (for *stargazer*) detected multiple loci and was eliminated. This rapid initial characterization thus eliminated all but two clones.

If GDRDA performed as intended, RDA-4.5 and RDA-8.2 should detect *BglII* polymorphisms mapping near *pudgy* and *tottering*, respectively. RDA-4.5 detected a *BglII* RFLP with a much smaller fragment in Tester than Driver (580 bp and 3.5 kb, respectively). Based on a genetic mapping panel consisting of 22 progeny from a (CAST/Ei × C57BL/6J)-*mnd*)F2 intercross, this fragment mapped to the 9 cM interval between *D7Mit56* and *D7Mit25*, which is consistent with the location of *pudgy*. Based on subsequent genetic mapping in a cross segregating *pu*, we determined that RDA-4.5 maps approximately 3 cM distal to *pu*, within the *pu-p* interval that was retained intact by the breeding scheme used to construct the stock (KK., W.F. and E.S.L., unpublished observations). RDA-8.2 detected a *BglII* RFLP with a much smaller fragment in Tester than Driver (400 bp and >3 kb, respectively). Using the same (CAST/Ei × C57BL/6J)-*mnd*)F2 intercross as above, RDA-8.2 was found to map to the 7 cM interval between *D8Mit51* and *D8Mit9*, which is consistent with the location of *tottering*.

Thus, both GDRDA probes mapped to the desired region. Although the size of the target region differing between the congenic strains is not known precisely, it is estimated to be less than 15 cM based on the breeding schemes used in constructing the congenic strains. Accordingly, GDRDA successfully generated polymorphic probes in a region of less than 1% of the mouse genome around the target locus.

**GDRDA with two-generation crosses**

Congenic strains are an obvious choice for GDRDA, but they suffer from a major drawback. Producing congenic strains requires many generations of breeding, which can span years or decades depending on the organism. To develop a more practical and rapid approach, we devised a second implementation of GDRDA that requires only a simple two-generation cross.

Transmission genetics is used to produce a collection of

---

**Fig. 3** Schematic diagram representing the principle underlying GDRDA with progeny from an F2 intercross. Each panel shows hypothetical chromosomal genotypes from 10 progeny to be pooled to create a Driver; each chromosome is arbitrarily drawn to be 100 cM. Strain A carries a recessively-acting allele at locus L and is shown in white; strain B is shown in black. Graphs show percentage of B alleles present in Driver at each location along the chromosome. a, A chromosome unlinked to L (the percentage of B alleles remains close to 50%). b, The chromosome containing L, with progeny having the recessive phenotype selected at random (the percentage of B alleles dips slowly to 0% at L). c, The chromosome containing L, with progeny having the recessive phenotype selected to be recombinant between L and one of two flanking genetic markers, X or Y (the percentage of B alleles drops sharply to 0% in the X-Y interval).
siblings with the property that their pooled DNA is homozygous in the region of a target gene but heterozygous elsewhere in the genome. Let A and B denote two inbred strains differing at a target locus L of interest. (As discussed below, outbred strains can also be used with only minor modifications in the procedure.) Suppose that A carries a mutant allele m causing a recessive phenotype and B carries a wildtype allele + causing a dominant phenotype. For a Tester sample, one can use strain B itself. To create a Driver sample, one performs an F₁ intercross between the strains, selects a collection of k progeny showing the recessive phenotype, and mixes their DNA together. The principles of mendelian genetics predict that the Driver should contain: (i) no B alleles in the immediate vicinity of L, because progeny were selected for the recessive phenotype; (ii) a deficit of B alleles in a somewhat larger region around L, owing to linkage to L; and (iii) roughly equal proportions of A and B alleles elsewhere in the genome, because a collection of F₁ progeny should have genotypes AA, AB and BB in the ratio 1:2:1 at unselected loci (see Fig. 3a,b). If RDA is performed with this Tester and Driver, then one would expect that B alleles should be subtracted everywhere in the genome except in a region around L. GDRDA should thus yield polymorphic alleles from the wild-type chromosome at loci linked to L.

The targeting of the method can be somewhat improved in the event that the locus L has already been genetically mapped between two flanking genetic markers, X and Y (which might have been taken from a pre-existing genetic map or might have been generated by a previous application of GDRDA). For the Driver, one can select k/2 progeny in which a crossover had occurred between X and L and k/2 progeny in which a crossover had occurred between L and Y. This would guarantee that the proportion of B alleles is 25% at X and Y, ensuring that the region over which the proportion of B alleles is very low is restricted to the interval X-Y (Fig. 3c). As we demonstrate below, this refinement can allow targeting of very small intervals.

An important issue in the design of this experiment is the number of progeny that should be pooled. While the proportion of B alleles at unlinked loci in the Driver will have a mean value of 50%, the actual value will fluctuate across the genome. In the accompanying box, we discuss how many progeny should be pooled to ensure that the proportion of B alleles remains high enough throughout the genome to ensure efficient subtraction. For an F₁ intercross in the mouse, we conclude that 10 progeny should suffice.

To test this approach, we applied it to two mouse crosses involving the nude (nu) locus on chromosome 11 and the staggerer (sg) locus on chromosome 9. In both cases, GDRDA successfully generated probes mapping close to the target loci.

In the course of studies on nude, we had generated 416 (MOLF/Ei × AKR/J-nu⁺⁻) F₂ intercross progeny, genotyped them for various genetic markers on Chromosome 11 and determined the position of nude relative to these markers (J.S., J.N., Benjamin Taylor and E.S.L., unpublished data). Using this information, we selected 12 nude progeny having crossovers between nude and closely linked markers (Fig. 4). All of the crossovers occurred within a 7 cM interval defined by DII1Mit5 and DII1Mit36, and 4 of the 12 occurred within a 1.3 cM interval defined by DII1Mit7 and DII1Mit34. A Driver sample was prepared by pooling equal amounts of DNA from these 12 progeny; the corresponding Tester sample was DNA from the MOLF/Ei parental strain. In principle, GDRDA should produce MOLF/Ei alleles of polymorphisms in the interval DII1Mit5 and DII1Mit36. Moreover, if the proportion of B alleles outside this interval sufficed to allow efficient subtraction, the polymorphisms might be targeted preferentially to the small interval between DII1Mit7 and DII1Mit34.

Using this Tester and Driver combination, we performed RDA with the restriction enzyme BglII. In the resulting difference product, two clear bands (700 bp and 450 bp) were visible by ethidium bromide staining. These were cloned to produce probes RDA-6.1 and RDA-6.2. As above, the probes were initially characterized by hybridization to Southern blots of Tester and Driver amplicons. RDA-6.1 turned out to detect a large number of bands in both amplicons and was eliminated. RDA-6.2 showed the expected pattern of hybridizing to the Tester but not Driver amplicon. The probe was then hybridized to Southern blots of mouse DNAs digested with BglII. It detected an RFLP with a 450 bp allele in MOLF/Ei and a 4 kb allele in AKR/J-nu⁺⁻. Using this RFLP, the locus detected by RDA-6.2 was genetically mapped. To obtain approximate localization, we genotyped 20 (MOLF/Ei × AKR/J-nu⁺⁻) F₂ progeny that showed no recombination between genetic markers flanking nude and found that the RFLP showed an inheritance pattern completely concordant with that of the nude locus itself (Fig. 4). To obtain finer localization, we then genotyped the 12 nude F₂ progeny used to create the Driver and found that the RFLP again showed complete concordance with nude — i.e., the progeny were all homozygous for the AKR allele of the RFLP. This proves that RDA-6.2 maps within the 1.3 cM interval bounded by DII1Mit7 and DII1Mit34. Subsequent analysis of additional F₂ progeny (J.A.S., J.H.N., Benjamin Taylor and E.S.L., unpublished data) has shown that RDA-6.2 recombined with nude only twice in 1290 meioses, corresponding to a genetic distance of only 0.2 cM. Thus, GDRDA successfully targeted a probe to a region less than 1/2,000 of the mouse genome.

We next attempted to generate additional clones by repeating GDRDA using the restriction enzyme BamHI. Two of three clones, RDA-10.2 and RDA-10.4, showed...
strains used were: Lurcher (Lc, chromosome 6). This dominantly-acting mutation arose in the Mi^+^ stock and a congenic strain was produced by 40 generations of backcrosses to Balb/cBy. The Tester was a Lc/+ female from the BALB/cBy congenic strain (N40) and the Driver was a BALB/cBy female. Severe combined immunodeficiency (scid, chromosome 16). This recessively-acting mutation arose on C.B-17 (a BALB/c-like strain) and a congenic strain was produced by 11 generations of backcrosses to C3H/HeJ. The Tester was a C3H/HeJ-scid male (N11) and the Driver was a C3H/HeJ male. Pudgy (pu, chromosome 7). This recessively-acting mutation arose on a non-inbred stock. It was maintained on a homozygous chinchilla (c^e^) stock, in trans to the nearby p mutation (that is, p^e^ x p^+/^ x p^e^) and was subsequently brother-sister mated for 42 generations with selection for heterozygotes in every other generation. The breeding scheme should maintain two alternative forms of a congenic region including the p-p interval, but the animals should be identical outside this region. The Tester was a pu/+ female and the Driver a pu/ pu female from this stock (N42). Töttering (tg, chromosome 8). This recessively-acting mutation arose on a DBA/2J genetic background and a congenic strain was produced by 34 generations of backcrosses to C57Bl/6J. The Tester was C57Bl/6J-tg female (N34) and the Driver was a C57Bl/6J female. Stargazer (stg, chromosome 15). This recessively-acting mutation arose on a A/J background and a congenic strain was produced by 19 generations of backcrosses to (C3H/HeJ x C57Bl/6J) hybrid background. The Tester was a stg/stg female from the congenic strain (N19) and the Driver was a 1:1 mixture of C3H/HeJ and C57Bl/6J female DNA. Nude (chromosome 11). This recessively-acting mutation arose in an non-inbred strain and a congenic strain was produced by 12 generations of backcrosses to C57Bl/6J. The Tester was a C57Bl/6J female and the Driver was a C57Bl/6J-nu female (N12). For further information about the mutations discussed in this paper, see ref. 6.

RDA procedure. RDA was performed essentially as described. A detailed protocol is available directly from the authors. To maximize the success of RDA, it may be helpful to employ the following controls: (i) ligation of PCR products with new adaptors on each round of RDA can be monitored by PCR and subsequent gel electrophoresis before hybridization, which should show a detectable increase in fragment size distribution; (ii) concentration of Tester and Driver DNA at each step should be determined by gel electrophoresis, using StuI/A digested human DNA as a control; (iii) experiment 1 from ref. 2 can be performed in parallel with the main experiment, as a positive control.

In this work, all amplicons were prepared by digesting 2 μg of either of Tester and Driver DNA with either BglII or BamHI. The iterative hybridization-extension-amplification step was repeated three times. The resulting material was digested with the same restriction enzyme and electrophoresed in a 2% agarose gel and transferred using a vacuum blotting apparatus to GeneScreen Plus membranes. Finally, clones were purified on Qiagen.

Genetic mapping of RFLPs. Clones detecting a fragment present in Tester but not Driver amplicons, which had been electrophoresed in a 2% agarose gel and transferred using a vacuum blotting apparatus to GeneScreen Plus membranes. Finally, clones were tested to determine whether they detected a unique genomic locus by hybridizing them to Southern blots of restriction-digested genomic DNA, with washing at moderate stringency (two 30 min. washes in 0.1× SSC, 0.1% SDS at 65°C).

Acknowledgements. This work was supported in part by grants from the National Center for Human Genome Research, the National Science Foundation, and the Mary Kay Foundation to E.S.L., the National Institute of Child Health and Development to J.H.N. and the National Cancer Institute and the American Cancer Society to M.H.W.
the expected pattern of hybridizing to the Tester but not the Driver amplicon. Both probes detected RFLPs between MOLF/IEI and AKR/J-nu"\textsuperscript{"m"} (with allele sizes of 600 bp and 4–5 kb for RDA-10.2 and 500 bp and 3 kb for RDA-10.4). Genetic mapping subsequently showed that both probes mapped close to \textit{mu}d. The 12 F\textsubscript{2} progeny used to create the Driver were all homozygous for the AKR allele for both RFLPs, indicating that both loci mapped in the 1.3 \textit{cM} interval between \textit{D11Mit7} and \textit{D11Mit34}. Subsequent analysis of additional F\textsubscript{2} progeny (J.A.S., J.H.N., Benjamin

In applying GDRDA to an F\textsubscript{2} intercross, how many progeny should be pooled to create the Driver? The method requires that the proportion of B alleles is sufficient to ensure subtraction at all unlinked loci. While the expected proportion will be 50\% by mendelian segregation, the actual proportion will fluctuate across the genome. The more progeny pooled, the smaller the fluctuations.

The critical proportion \( \alpha \) of B alleles needed to ensure subtraction at a locus is not known precisely and can only be determined based on empirical evidence from many RDA experiments. Indeed, it may depend on the nature of the sequence, the hybridization conditions used and the ratio of Tester and Driver at each stage. Nonetheless, the current RDA protocol — which employs a 80-fold excess of Driver on the first round — seems to allow efficient subtraction at all loci present at 10–15\% in the Driver (N.L., unpublished data). Thus, one might set the critical threshold for subtraction at \( \alpha = 0.10–0.15 \).

Given a choice for critical threshold \( \alpha \), how many progeny \( n \) should be pooled? Let \( a_1(a) \) denote the expected length of the region linked to \( L \), for which the proportion \( \pi_1 \) of B alleles is less than \( \alpha \) and \( b_1(a) \) denote the expected length of unlinked regions of the genome for which \( \pi_1 < \alpha \), where \( n \) is the number of recombinant haploid genomes pooled to create the Driver (that is, \( n = 2k \)). The ratio \( c_1(a) = a_1(a)/b_1(a) \) should be such a proportion in the ratio of linked to unlinked clones that should be produced by RDA. (For the calculation of \( c_1(a) \), see Methodology.) If \( c_1(a) > 1 \), then linked clones should constitute a minority of fragments surviving subtraction. If \( c_1(a) < 1 \), linked clones will be a majority which must be identified by subsequent screening. The number of progeny to pool is thus \( n \geq 2k \), where \( n \) is the smallest integer such that \( c_1(a) \geq C \), for a chosen lower bound \( C \).

A graph of \( c_1(a) \) is shown in Fig. 5. Applying this to the mouse genome (genetic length \( = 16 \) Morgans) and choosing a critical threshold \( \alpha = 0.15 \) for subtraction, one has \( c_1(0.15) = 7.7, 13.6, 24.1, \) and 182.6 for \( k = 6, 8, 10 \) and 12 F\textsubscript{2} progeny pooled (with 

\( n = 2k \)). To ensure \( c_1(a) > 1 \), it might thus be prudent to pool at least 10 F\textsubscript{2} progeny.

How close to L will the linked polymorphisms be? Assuming the simple model of a critical threshold \( \alpha \) for subtraction, they would be expected to lie roughly within the fraction \( \alpha = 0.1 \) of \( L \). If \( \alpha = 0.10 \), the target interval thus may be about 20 \textit{cM}. If subtraction is not all-or-none, there should be a bias toward the centre of the region because the proportion of B alleles will be lowest there. If polymorphisms can be selected having recombinations near \( L \), the interval targeted can be made much smaller.

GDRDA can be applied to a backcross between inbred strains with only a minor modification in the analysis. In a \((A \times B)F1 \times A\) backcross, use the \((A \times B)F1\) animals as Tester and use a collection of \( k \) backcross progeny showing the recessive phenotype as the Driver. The Tester and Driver for these \( k \) backcross progeny are identical to those that would be obtained by taking the Tester and Driver for \( k/2 \) F\textsubscript{2} intercross progeny and mixing each 1:1 with strain A; the mixing with strain A should have no effect, since A alleles should be efficiently subtracted. The number of progeny to pool is thus \( n \) is the smallest integer such that \( c_1(a) \geq C \).

GDRDA can also be applied to crosses involving non-inbred matings. Consider a mating in an outbred population between individual C who is heterozygous (\( m/\)) and another individual D. The Tester can be C and the Driver can be a collection of progeny who inherited the allele \( m \) from C. If \( m \) causes a dominant phenotype, these progeny can be readily identified based on their phenotype. If \( m \) causes a recessive phenotype, they could be identified either by progeny testing or by using a parent D who is also heterozygous and selecting homozygous progeny. Subtraction should yield alleles present only on the chromosome carrying the \( + \) allele in C. The situation differs from a backcross with inbred strains only in one respect: one must ensure subtraction of two possibly different alleles at unlinked loci in C. To account for regions in which the proportion of either allele is too low, the function \( b_1(a) \) should be replaced by \( 2b_1(a) \). The minimum number of backcross progeny that should be pooled is thus \( k = n \), where \( n \) is the smallest integer such that \( c_1(a/2) \geq C \). This is only a slight increase over the corresponding backcross.

Finally, the progeny in the outbred matings need not be full sibs. One could use progeny from matings of C to multiple partners, D, D, . . . , D. The potential drawback is that a linked C allele could be subtracted if it is present in any of the Ds, which would decrease the number of detectable polymorphisms as \( n \) increases. The half-sib design may be especially convenient in the case of livestock, for which a single male is often mated to multiple females.
Taylor and E.S.L., unpublished data) showed that RDA-10.4 recombined with nude only once and that RDA-10.2 never recombined with nude in 1290 meioses. In summary, GDRDA produced three distinct polymorphisms mapping within 0.2 cm of the target locus.

Finally, we performed an analogous experiment using the staggerer mutation. In another project, we had genotyped 270 (C57BL/6J-sg × DBA/2J)F2 intercross progeny for various genetic markers on chromosome 9 and determined the position of staggerer relative to these markers (K.K., W.N.F., Muriel Davison, and E.S.L., unpublished data). Using this information, we selected 13 sg/sg homozygous progeny having crossovers in a 10 cm interval containing staggerer defined by D9Mit48 and D9Mit11. A Driver sample was prepared by pooling DNA from these 13 progeny; the corresponding Tester sample was DNA from the DBA/2J parental strain. In principle, GDRDA should produce DBA/2J alleles of polymorphisms in the interval between D9Mit48 and D9Mit11.

RDA was performed with the restriction enzyme BgIII. A single strong band (500 bp) was visible by ethidium bromide staining and was cloned to produce the probe RDA-1.1, which passed the initial characterization tests. The probe detected a BgIII RFLP between Tester and Driver (500 bp allele in DBA/2J and >4 kb in C57BL/6J-sg). The RFLP mapped to the interval between D9Mit48 and D9Mit11 on a (CAST/Ei × C57BL/6J)F2 intercross. When the 12 recombinant progeny that had been used in the Driver were genotyped, we found that 9 progeny were homozygous for the C57BL/6J-sg allele but three progeny were heterozygous. In contrast to the nude experiments in which all three probes derived from a region for which the Driver completely lacked Tester alleles, this experiment yielded a probe from a region near sg for which the Driver contained the Tester allele at a proportion of 11.5% (that is, 3/26). Subsequent genotyping of the 270 (C57BL/6J-sg × DBA/2J)F2 progeny has shown that RDA-1.1 maps approximately 4.5 cm distal to sg(K.K., W.N.F., Muriel Davison and E.S.L., unpublished data). In summary, GDRDA produced closely linked markers in both the nude and staggerer crosses.

Discussion

GDRDA is unique among molecular genetic techniques in that it provides a way to target DNA probes to the vicinity of a gene without prior knowledge of either the gene’s function or position. By applying classical transmission genetics, one can prepare DNA samples from mixtures of progeny that differ only near the gene of interest and then use the powerful subtraction technique of RDA to clone these differences. The technique opens the prospect of genetic analysis and positional cloning even in organisms without pre-existing genetic maps.

We describe two particular implementations of GDRDA, using congenic strains and two-generation crosses. Both approaches successfully produced probes mapping near various target genes. Indeed, every clone (6/6) that passed a rapid initial characterization (detecting a unique fragment in Tester but not Driver amplicon and a unique locus in genomic DNA) mapped to the desired location. In the case of the nude cross, we obtained three different probes that mapped within 0.2 cm of the target locus.

The yield of probes was relatively low (6 probes from 9 experiments), which is perhaps not surprising in view of the multiple rounds of exponential competition among PCR products during RDA. The number of probes might be increased through the use of additional restriction enzymes for amplicon preparation, as demonstrated by the successful use of BamHI in the case of the nude experiment. Some restriction enzymes, such as TogI, may produce a higher yield of polymorphisms. It may also be possible to generate new clones with a single restriction enzyme by blocking the amplification of already-identified clones by adding them back to the Driver. Finally, it may be possible to detect less drastic changes in the length of restriction fragments by initially fractionating Tester and Driver by gel electrophoresis and performing subtraction on specific size fractions.

Application of GDRDA to congenic strains is straightforward. However, the real power of GDRDA lies in its application to crosses, because the breeding or pedigree collection required is within the realm of practicality for a wide range of organisms. The technique can be applied to any trait whose presence implies homozygosity for a particular allele at a trait-causing locus, so that these homozygotes may be pooled to create a Driver.

An interesting feature of the application to crosses is that the targeting of GDRDA can be improved by successive iteration. Given a large cross, one could first generate flanking markers that are linked, but perhaps not very closely, to the target locus. Using such flanking markers to identify recombinant progeny, one could perform subsequent subtractions with these progeny to target successively smaller intervals. As shown in the case of the nude and staggerer crosses, the use of recombinant progeny can effectively target quite small intervals. The ultimate resolution of this approach should be limited only by the actual density of polymorphisms detectable by GDRDA; we estimate this density to be 1–2 per megabase for an enzyme such as BgIII.

We have focussed here on the application of GDRDA to F2 intercrosses between inbred strains, but the technique is more broadly applicable. It can be applied to backcrosses between inbred strains, two-generation families in an outbred population (for organisms for which inbred lines are not available), and half-sib mating schemes (common in livestock breeding). Considerations in designing such experiments are discussed in the accompanying box.

The application of GDRDA to random-breeding populations should include the analysis of human families. One might, for example, use an individual affected with a dominant disease as Tester and a collection of unaffected close relatives as Driver. In some families, there may be too few relatives to ensure subtraction of all unlinked regions. In such cases, GDRDA should at least enrich for linked probes which could then be subsequently screened for linkage. We will discuss this issue in more detail elsewhere.

Notwithstanding continuing advances in genomic analysis, construction and application of dense genetic linkage maps remains a daunting task. GDRDA offers the prospect of obviating the need for such maps, at least for certain purposes. In particular, GDRDA should open the prospect of genetic mapping and positional cloning of monogenic traits in most experimentally and agriculturally important animals, plants and fungi.

Methodology

Mouse strains. All mouse strains used were maintained at The Jackson Laboratory, with the exception of those used for the Lc congenic experiment, which were maintained at the Wadsworth Center, Albany, NY and provided by Anne Messer. The congenic
Mutations in the nude locus in mice and rats produce the pleiotropic phenotype of hairlessness and athymia, resulting in severely compromised immune system. To identify the causative gene, we utilized modern tools and techniques of positional cloning. Specifically, spanning the region in which the nude locus resides, we constructed a genetic map of polymorphic markers, a physical map of yeast artificial chromosomes and bacteriophage P1 clones, and a transcription map of genes obtained by direct cDNA selection and exon trapping. We identified seven novel transcripts with similarity to genes from Drosophila, Caenorhabditis elegans, rat, or human and three previously identified mouse genes. Based on our transcription mapping results, we present a novel approach to estimate the number of genes in a region and estimate that the nude locus resides in a region approximately threefold enriched for genes. We confirm a recently published report that the nude phenotype is caused by mutations in a gene encoding a novel winged helix or fork head domain transcription factor, whn (Nehls et al., Nature 372: 103–107, 1994). We report as well the mutations in the rat rnu allele and the complete coding sequence of the rat whn mRNA.

INTRODUCTION

There are hundreds of biologically important mouse mutants for which the defective gene has not yet been identified (Green, 1989). Positional cloning, identifying genes based on their chromosomal location, is the most general way to identify the genes responsible for these diverse phenotypes. Prospects for positional cloning have improved rapidly in recent years with the avail-ability of new tools and techniques, including dense sets of genetic markers (Dietrich et al., 1994), large insert yeast artificial chromosome (YAC) libraries (Kusumi et al., 1993), and new methods for identifying transcription units in a physical region (Buckler et al., 1991; Lovett et al., 1991).

Mutations in the nude locus produce the remarkable pleiotropic phenotype of hairlessness and athymia (Flanagan, 1966; Pantelouris, 1968). Genetic studies demonstrated that the nude mutation segregates as a single autosomal locus on mouse chromosome 11 (Flanagan, 1966). Because of the athymia, the mice lack T lymphocytes and therefore have a highly impaired immune system. Accordingly, nude mice are extensively used in cancer research for transplants of tumors and tissues from other species.

From a developmental standpoint, the hair defect appears to result from improper keratinization of the hair follicles, resulting in short, bent hairs that only rarely emerge from the epidermis; the number of hair bulbs is normal (Flanagan, 1966; Kopf-Maier et al., 1990). The thymic defect appears to occur because the expected rapid proliferation of ectoderm of the developing thymus at 11.5 days postcoitum (dpc) fails to occur (Cordier and Haumont, 1980). Transplantation studies show that the thymic rudiment of a nude mouse fails to attract normal lymphoid cells, but bone marrow from a nude mouse can repopulate the thymus of an irradiated mouse, indicating that the thymic dysgenesis in nude mice is stromal, and ectodermal, in origin (Pantelouris, 1973; Wortis et al., 1971). There is no obvious explanation for how these two discrete developmental phenotypes are related. To address this question, the causative gene must be identified.

nude is a good locus for positional cloning, because there are five nude alleles among rodents. Two alleles of nude exist in the mouse: nude (nu), which arose in an outbred strain in the Virus Laboratory in Glasgow in 1966, and nude-streaker (nu"), which arose within the inbred AKR/J mouse colony at The Jackson Laboratory in 1974 (Eicher, 1976; Flanagan, 1966). Two alleles
of \textit{nude} exist in the rat: \textit{nru} and \textit{nru}'\textsuperscript{x}, both of which arose on outbred strains (Berridge \textit{et al.}, 1979; Festing \textit{et al.}, 1978). An athymic hairless guinea pig has also been reported (Reed and O'Donoghue, 1979). In mice and rats, the mutations are likely to be in homologous genes because in both species they are tightly linked to the inducible nitric oxide synthase gene (Jenkins \textit{et al.}, 1994; Zha \textit{et al.}, 1995). In the guinea pig, the mutation has not been mapped.

We report here an extensive characterization of the genomic region surrounding the \textit{nude} locus, including genetic distance, physical distance, and gene density. We genetically mapped the \textit{nude} locus and constructed a YAC contig across the region. We then generated a dense set of sequence tagged sites (STSs) to construct a physical map in bacteriophage P1 clones that covers the smallest region in which the \textit{nude} locus must lie.

To identify genes in the region, we performed both direct cDNA selection and exon trapping. Based on the results, we were able to compare the ability of the two methods to identify transcription units. In our hands, direct cDNA selection was extremely fruitful, yielding a large number of distinct clones with no redundancy, of which 93\% mapped back to the correct physical region. By contrast, exon trapping yielded a smaller set of clones with considerable redundancy.

Based on strong similarity of analyzed gene fragments to GenBank entries, we identified 10 likely transcription units in the region: 7 novel transcripts with similarity to genes from \textit{Drosophila}, \textit{Caenorhabditis elegans}, rat, or human, and 3 previously identified mouse genes. Based on our transcription mapping results, we present a novel approach to estimate the number of genes in a region. Specifically, we estimate that the \textit{nude} locus appears to reside in a region approximately threefold enriched for genes, with 20–25\% of the nucleotides being transcribed.

Finally, we confirmed a recently published report that the \textit{nude} phenotype in mice and rats is caused by mutations in a gene encoding a novel winged helix (or fork head) domain transcription factor, \textit{whn} (Nehls \textit{et al.}, 1994). In that report, the mutations in the mouse \textit{nu} and the rat \textit{nru}'\textsuperscript{N} alleles are described. Here, we report the complete coding sequence of the rat \textit{whn} mRNA together with the mutation present in the rat \textit{nru} allele, as well as confirming the mutation in the mouse \textit{nu} allele. We have also analyzed the mouse \textit{nru}'\textsuperscript{r} allele but find no mutation in the coding region.

**MATERIALS AND METHODS**

**Animals.** Congenic C57BL/6-\textit{nru} mice were developed by repeatedly backcrossing outbred \textit{nude} (\textit{nu}) mice onto C57BL/6; they were purchased from Taconic Farms (Germantown, NY). The \textit{nru}'\textsuperscript{r} mutation arose at The Jackson Laboratory on the AKR/J inbred line (Eicher, 1976) and has been maintained on this background; these animals were purchased from The Jackson Laboratory (Bar Harbor, ME). The rat \textit{nru} mutation was discovered on an outbred strain and has been maintained by randomly mating \textit{nude} males with heterozygous females; these animals were purchased from Harlan Sprague Dawley (Indianapolis, IN). C57BL/6J, CAST/Ei, and MOLF/Ei animals, used in mapping crosses, were from The Jackson Laboratory. Pregenic were progeny at postnatal Day 11 for hair growth. Presence or absence of a thymus was checked for all key recombinant animals. Unaffected animals carrying chromosomes with key crossovers in the \textit{nude} region were progeny tested by mating them to \textit{nru}'\textsuperscript{r}/+ animals. Informative progeny (i.e., those determined, based on genotype at flanking markers, to carry the recombinant chromosome over a \textit{nru}'\textsuperscript{r}-bearing chromosome) were examined for phenotype to determine whether the recombinant chromosome carried the \textit{nru}'\textsuperscript{r} allele.

**Genotype analysis.** DNA from cross progeny was prepared from tail biopsies as described by Laird \textit{et al.} (1991). Simple sequence length polymorphism (SSLP) markers were genotyped as described by Dietrich \textit{et al.} (1992). Single-strand conformational polymorphism (SSCP) markers were amplified exactly as for SSLP markers and electrophoresed in 1x MDEE gel for 16 h at 200 V, according to the manufacturer's recommendation (AT Biochem, Inc., Malvern, PA).

**Isolation of YACs, P1s, and cosmids.** YACs were obtained by PCR-based screening of the MIT YAC library (Kusumi \textit{et al.}, 1993), with the markers described in Table 1. Bacteriophage P1 clones were obtained from commercially available libraries made from either mouse cell line IIII (P1-P2 and P1-2S) or mouse ES cells from strain 129 (P24, P25) (Genome Systems Inc., St. Louis, MO). Primers used for PCR-based screening are listed in Table 1. Cosmids were prepared from YAC DNA partially digested with Mbo1 using restriction-minus packaging extracts and hosts (Stratagene, La Jolla, CA). The library was plated at 5000 colonies/plate, lifted onto nylon membranes, and hybridized with cloned mouse repetitive elements B1, B2, and L1 to identify cosmids containing mouse genomic inserts.

**Cloning of YAC and P1 ends.** Total yeast DNA was prepared from YAC clones as described by Treco (1991). P1 clone DNA was prepared according to the supplier's recommendation (Genome Systems Inc., St. Louis, MO). We have previously described an improved protocol used to clone YAC ends by inverse PCR (Haidi \textit{et al.}, 1994). Essentially the same protocol was used to clone ends of P1 clones, with the only changes being in the choice of restriction enzymes and the primer sequences. Specifically, to clone the insert DNA adjacent to the SalI site in the P1 vector, 50 ng P1 insert DNA was digested with HindIII, Rsal, or Alul. To clone the insert DNA adjacent to the NotI site in the P1 vector, digestion was performed with HhaI or Alul. Digestion products were religated to promote circularization. PCR was performed on 0.5 ng of ligation product using P1UP (5'-GCCGCTAATGACCTCATTAGG-3') and P1UR (5'-GAGGAAATAATGCTTCAAATGCAG-3') for the NotI end and P1FL (5'-GAGGATCCAGTTGCGAGC-3') and P1LR (5'-CCGTGACATTAGTGGCAAC-3') for the SalI end. The resulting PCR products were then reamplified to introduce 3' and 5' cloning primer sites. The chimeric primers M13FORUF (5'-TTTAGAGAGGAGGATC-3') and M13REVUR (5'-CAGGAAACACGTATGACCC-3') were used for the NotI end, M13FORFL1 (5'-TGAATACCCGGGCGGCTAC-3') and M13REVFL1 (5'-TGAATACCCGGGCGGCTAC-3') or M13REVFL2 (5'-TGAATACCCGGGCGGCTAC-3') to reamplify the SalI end digested with HindIII and Rsal, and M13FORFL3 (5'-CCGTGACATTAGTGGCAAC-3') and M13REVFL3 to reamplify the SalI end digested with Alul. All sequencing was performed by following the cycle-sequencing protocol for fluorescently labeled M13 sequencing primers (Applied Biosystems, Foster City, CA) and reaction products were electrophoresed on an ABI370A (Applied Biosystems, Foster City, CA). Sequence was obtained from the vector/insert junction to ensure that the fragment represented the end of the insert.

**Characterization of YACs.** The sizes of the YACs were determined by pulsed-field gel electrophoresis (Chu \textit{et al.}, 1986), followed by transfer to nylon membrane and hybridization of the blot to \textit{pBR322} DNA labeled with \textit{32P} by the random priming method (Feinberg and Vogelstein, 1983). To determine the restriction map of the region, YAC DNA was digested in agarose plugs according to the manufacturer's recommendations (New England Biolabs, Beverly, MA) and...
fractionated by pulsed-field gel electrophoresis. After Southern transfer the blots were hybridized with gene fragments labeled with 32P specifically primed with an oligonucleotide from the vector sequence flanking the insert (5'-CTGACGGGAATTCGAGACC-3' for direct cDNA selected clones and 5'-CTCGAGGTCGACCCAGCA-3' for exon trapped clones).

Direct cDNA selection and exon trapping. Direct cDNA selection was performed on P1 and cosmids clones according to the protocol of Lovett (1994) with several modifications. Biotin was incorporated into the digested genomic DNA by ligating biotin-containing linkers (ligation of the two oligonucleotides: Bio-Blunt-1 (5'-BIOTIN-GGAGTGAGTTATAATCC-3') and Blunt-2 (5'-GAAATCGATGATGATGC-3')). Primer cDNA for the selection was independently prepared by random-primer poly(A)⁺ selected mRNA, from BALB/cJ postnatal Day 0.5 skin, adult C57BL/6J skin, adult C57BL/6J testes, and adult C57BL/6J thymus. Streptavidin-coated magnetic beads were preblocked with 0.1% BSA and 0.2 μg/ml mouse COT-1 DNA (GIBCO BRL, Gaithersburg, MD). After two rounds of hybridization, selected cDNA fragments were cloned by using the uracil DNA glycosylase cloning system (GIBCO BRL). Exon trapping was performed with the SPL3 plasmid under the manufacturer's conditions on pools of eight cosmids, digested with BamHI and BglII (GIBCO BRL). RT-PCR was performed on first-strand cDNA made from DNased total mRNA, according to the manufacturer's instructions (GIBCO BRL).

Sequencing of the mouse and rat nude genes. Primary cDNA was independently prepared by specific priming with an oligo from the 3' untranslated region of the mouse nude gene (5'-GGGAGAGGG-CCATTCTTCTG-3') poly(A)⁺ selected mRNA from C57BL/6J, C57BL/6J nu/nu, AKR/J, AKR/J-nu+/nu-, rat, and rat-nu/rnu adult skin. Twelve overlapping fragments were amplified by PCR from the primary cDNA, and both strands were sequenced according to the manufacturer's instructions (Applied Biosystems, Foster City, CA).

RESULTS

Mouse Crosses and Genetic Mapping

We genetically mapped the nude locus in over 2000 meioses in three separate F₂ intercrosses. To ensure the greatest rates of polymorphisms between the two strains of the cross, we mated AKR/J-nu+/nu animals with the inbred subspecies Mus musculus castaneus (CAST/Ei) and Mus musculus molossinus (MOLF/Ei). One concern with intersubspecific crosses is the possibility of recombinational suppression due to structural heterogeneity of the chromosomes (Copeland et al., 1993; Hammer et al., 1989). To address this concern we also made an intraspecific cross with C57BL/6J. Specifically we generated 182 (AKR-nu+/nu × C57BL/6J)F₂ animals, 700 (AKR-nu+/nu × MOLF/Ei)F₂ animals, and 226 (AKR-nu+/nu × CAST/Ei)F₂ animals.

To construct an initial genetic map of the region, all progeny were phenotyped at postnatal Day 11 for hair growth and genotyped with SSLP markers flanking the nude locus. These markers gave the map order D11Mit7-0.7 cM-nu-0.9 cM-D11Mit34 (Dietrich et al., 1992). All of the crosses gave a genetic distance of 1.6 ± 0.2 cM between D11Mit7 and D11Mit34, indicating that there is no gross recombinational suppression in the region of nude between the Mus musculus subspecies analyzed. Unfortunately, the intrasubspecific nude cross did not yield finer structure genetic mapping information because none of the nine genetic markers in the smallest region around nude were polymorphic between AKR/J and C57BL/6J.

To obtain fine structure mapping information from the intersubspecific crosses, we focused on those progeny that were recombinant in the interval between D11Mit7 and D11Mit34. For all such recombinant animals, the presence or absence of a thymus was checked. (In all cases, the phenotypes of athymia and hairlessness coincided.) We progeny-tested selected unaffected F₂ progeny, carrying such a recombinant chromosome together with a wildtype nonrecombinant chromosome (13 animals), to determine which nude allele was carried on the recombinant chromosome. In this manner, each F₂ progeny yielded two informative meioses.

To generate markers in the nude region, we devised the method of genetically directed representational difference analysis (GD-RDA), which we have described elsewhere (Lisitsyn et al., 1994). We obtained three new markers, RDA6.2, RDA10.2, and RDA10.4. RDA6.2 and RDA10.4 mapped proximal to nude, and RDA10.2 did not recombine with nude in any of the crosses, giving the map order D11Mit7-0.55 cM-RDA6.2-0.1 cM-RDA10.4-0.05 cM-(nu, RDA10.2)-0.9 cM-D11Mit34. The location of the genetic markers on the YAC contig from RDA10.4 to D11Mit34 is shown in Fig. 1.

Gross Physical Map

We eventually narrowed the region that contained the nude locus to a single YAC, designated YAC 31 (formal designation WI FALG2) from the library of Kusumi et al. (1993). The proximal marker RDA10.4 and the nonrecombinant marker RDA10.2 were both contained in this YAC. One end of YAC 31, cloned by inverse PCR, was shown by SSCP to map 0.05 cM distal to nude (data not shown). Therefore, the nude locus must be contained within the genomic region corresponding to YAC 31.

YAC 31 does not appear to be chimeric since its ends and all markers cloned from it either map by SSCP analysis to the appropriate region of mouse chromosome 11 or were contained in other clones in the region. The size of YAC 31 was determined to be at least 1 Mb by pulsed-field gel (PFGE) electrophoresis. However, this YAC was highly unstable: 15 unique isolates of it ranged in size from 150 kb to 1 Mb. Even when individual isolates from a 1-Mb clone of YAC 31 were regrown, the sizes ranged from 250 kb to 1 Mb.

To determine the physical distance between the closest markers flanking nude, we constructed a PFGE map of YAC 31. For our analysis, we used a 1-Mb isolate of YAC 31. The YAC DNA was digested independently with several rare-cutting restriction enzymes: MluI, NotI, RsrII, SacI, and SfiI. Southern blots of these digests were hybridized with a dense set of markers from the region: the ends of the YAC, the clones obtained from direct cDNA selection and exon trapping (described below), and random markers subcloned from the YAC. A unique restriction map of the YACs was
revealed by the hybridization pattern of this dense set of markers. The PFG map of the region of the YAC that contains the nude gene is shown in Fig. 2.

**Fine Scale Physical Map**

We next sought to obtain smaller insert clones that span only the smallest region of YAC 31 in which nude must lie. We initially subcloned the YAC into cosmids to construct a cosmid contig around the nude locus. However, the instability of the YAC thwarted our initial efforts to create a contig across the region. Although we selected a clone of YAC 31 that appeared to be a full 1 Mb in size, the DNA apparently contained various internal deletions. In fact, some regions of the DNA appear to be particularly prone to internal deletions because the same genomically noncontiguous regions were cojoined in cosmids constructed independently from a different YAC that covered part of the nude region. This problem could perhaps have been ameliorated if a YAC DNA band of 1 Mb had been isolated from a PFG before subcloning. A small degree of cosmid chimerism was also attributable to the subcloning procedure. In any case, constructing a contig from cosmids subcloned from a YAC is inherently undesirable, because it does not provide independent verification of the genomic region.

To obtain an independent representation of the region in smaller insert clones, we constructed an extensive P1 contig of the nude region with clones from a genomic library. We physically mapped over 150 STSs to this region: All of the markers subcloned from YAC 31 (44% of the total) and the P1 clones (56% of the total) showed a unique, consistent physical map.

**FIG. 1.** Large-scale map of the nude region. (A) Genetic map of polymorphic markers. (B) STS content mapping of YACs with genetic markers. The location of the 370-kb region containing nude is shown. Primer sequences are given in Table 1 or published in Dietrich et al. (1992, 1994).

**FIG. 2.** Fine structure map of the nude region. (A) Genetic markers. (B) Recombinational breakpoints for the closest recombinant animals in nu crosses. The empty bar denotes the chromosomal region that recombines with the nu gene. The black bar denotes the chromosomal region that does not recombine with the nu gene. The hatched bar denotes the chromosomal region with an undetermined haplotype. (C) The smallest region that does not recombine with the nude locus. (D) Pulsed-field gel-based restriction map of the region. S, SacII; F, SfiI; M, MluI; N, NotI; R, RsalII. (E) STS-content mapping of P1 clones (denoted P) and cosmids (denoted C). (F) The location of transcription units with strong sequence similarity to genes in GenBank.
cal location in the YAC and P1 contigs. Of the 150 STSs, 64% are derived from direct cDNA selected clones, 13% are from exon trapped clones, 14% are from the YACs, and 9% are from the P1s. The primer sequences for the framework markers are given in Table 1. The size of the region covered by the P1s is consistent with the size of the region in the YAC, as judged by pulsed-field gel analysis. To identify the boundaries of the **nudef** region, we genetically mapped many ends of the P1 clones by SSCP on the animals with the closest flanking crossovers. Based on the first 1000 animals (i.e., 2000 meioses), the nonrecombinant region had a minimum size of 800 kb. In the last 50 animals, we were fortunate to obtain an animal with a proximal breakpoint that cut the region in half. Based on these progeny, the smallest physical region in which **nudef** was determined to lie is a minimum of 230 kb (from **D11Seg8** to **D11Seg30**) and a maximum of 370 kb (from **D11Seg4** to **D11Seg41**) (Fig. 2).

**Finding Transcription Units**

To find transcription units in the **nudef** region, we employed two complementary strategies: direct cDNA selection and exon trapping. Direct cDNA selection is based on recovering cDNA fragments that specifically hybridize to the physical DNA templates (Lovett, 1991). This method is constrained by tissue expression of the gene but not by the genomic structure of the gene. We performed two rounds of direct cDNA selection of primary cDNA from adult thymus, adult testes, adult skin, and neonatal skin hybridized to groups of cosmids or P1 clones covering the entire 370 kb of the genomic **nudef** region. Exon trapping is a strictly genomic approach, relying upon the fact that most mammalian

### Table 1

<table>
<thead>
<tr>
<th>Locus</th>
<th>Forward sequence</th>
<th>Reverse sequence</th>
<th>Product size</th>
<th>Gene recognized</th>
<th>Clones recognized</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D11Seg41</strong></td>
<td><strong>TAATGACTTCTGGTATGC</strong></td>
<td><strong>GCTTCTGTTAGTCTGCT</strong></td>
<td>97</td>
<td><strong>P10</strong></td>
<td></td>
</tr>
</tbody>
</table>
genes contain multiple internal exons and thus can be spliced into a synthetic vector (Church et al., 1994). We obtained cosmids that covered half of the nude region and used these for exon trapping.

Using both methods to clone transcription units provided an opportunity to compare the results of direct selection and exon trapping. Direct cDNA selection provided an extremely deep resource of transcribed sequences. We sequenced and analyzed a total of 170 unique clones with an average insert size of 250 bp without encountering the exact same clone twice. A low background rate of clones was obtained: 2% of the clones were from the Escherichia coli genome; 3% were P1 vector sequence; 2% were mouse repetitive sequence. Background problems were reduced by using primary cDNA, rather than a cDNA library grown in bacteria. The sole exception was the P1 clone, P21, which yielded 80% bacterial clones; this appears to have occurred because this clone deleted most of its mouse DNA insert. We found that more than 93% of the clones mapped back to the region by STS content mapping of the P1 clones, yielding 148 unique clones from direct cDNA selection. We attribute this great specificity to the fact that we used two rounds of hybridization and stringent washing conditions. Direct cDNA selection proved to be an extremely effective method to clone transcription units in a physical region from a selection proved to be an extremely effective method to clone transcription units in a physical region from a

Sequence Analysis of Transcription Units

The DNA sequences of the gene fragments were analyzed to look for similarities to known genes. We used a computer program to check each sequence for overlap of at least 40 bp with other gene fragments and for similarity to all other sequences in GenBank, using the programs BLASTN and BLASTX to search for nucleotide and amino acid sequence similarity, respectively. These search programs are optimized to search for local alignments, allowing for detection of similarities between diverged sequences. Because of the average size of the clones and the number of entries in GenBank, we set the criterion for a significant match to be a Poisson probability score P(N) of less than 10^{-10} for nucleotide comparisons or less than 10^{-5} for protein comparisons. A number of gene fragments could be grouped as likely to belong to the same transcription units because they shared strong sequence similarity to the same gene in GenBank. All such clones, sharing a strong similarity to a specific gene, mapped to the same physical region in the P1 contig. To confirm that the clones were part of the same transcription unit, we demonstrated in several cases that they were contained within a common clone from a cDNA library.

A total of 37 gene fragments showed strong similarity to 10 unique genes in GenBank: 28 of these clones were obtained from direct cDNA selection; 9 of these clones were obtained by exon trapping (see Fig. 3 for examples). In the subregion that was both exon trapped and direct cDNA selected, the same genes were identified. Since one cannot predict the nature of the nude gene, many of the transcription units identified were interesting candidates for nude. The 10 transcripts with strong amino acid similarities are:

1. human HTLF, a winged helix or fork head transcription factor (five from direct cDNA selection and one from exon trapping) (Li et al., 1992). A conserved 100-amino-acid domain defines fork head transcription factors that have been identified in yeast, Drosophila, C. elegans, Xenopus, mouse, and human. Mouse fork head genes are developmentally regulated during embryogenesis and control cell-specific gene expression in adults.

2. mouse vitronectin gene (two from direct selection and one from exon trapping) (Seiffert et al., 1993). Perfect nucleotide identity was found to mouse vitronectin, a circulating factor, produced in the liver, that regulates the link between cell adhesion, humoral defense mechanisms, and cell invasion. Although the mouse gene has not been mapped, the human vitronectin gene maps to 17q11, the region that is syntenically conserved in the human with the nude region in the mouse.

3. human tumor necrosis factor, alpha induced protein 1 (TNFAIP1) (two from direct selection and three from exon trapping) (Wolf et al., 1992). The mouse sequence is 95% identical to the human sequence over 1085 bp. Two of the six clones are from the 3' UTR. TNFAIP1 had been mapped previously to this region of mouse chromosome 11. TNFAIP1 is induced rapidly in endothelial cells in response to tumor necrosis factor-α.

4. Drosophila nemo gene (three from direct selection and one from exon trapping) (Choi and Benzer, 1994). nemo, a serine/threonine protein kinase, is required to initiate the second step of rotation of ommatidia. Rotation is also a common phenomenon in vertebrate embryonic development.

5. C. elegans gene emb-5 (four from direct selection) (Nishiwaki et al., 1993). emb-5 is required for the correct timing of gut precursor cell division during gastrulation. emb-5 is structurally similar to the Saccharomyces cerevisiae nuclear protein SPF6, which inhibits transcription of various genes, possibly by regulating chromatin assembly.

6. rat Na'/sulfate cotransporter gene (two from direct selection and one from exon trapping) (Markovich et al., 1993). This transporter is involved in sulfate reabsorption in the kidney, intestine, and colon.
FIG. 3. Examples of strong sequence similarities. (A) Nucleotide sequence alignment of a direct cDNA selected clone, cY8A, with mouse vitronectin mRNA, produced by the BLASTN program. The sequences have perfect nucleotide identity over 177 nucleotides with the BLASTX program. The two clones share 41% amino acid identity over 86 residues. Amino acids listed between the lines are identical (Vtn) and cY8A. (B) Amino acid alignment of a direct cDNA selected clone, cK12G, with the exception of a single undetermined nucleotide in cY8A. Numbering corresponds to nucleotide position for both the mouse vitronectin mRNA and cK12G, with the C. elegans protein emb-5 produced by the BLASTX program. The two clones share 41% amino acid identity over 86 residues. Amino acids listed between the lines are identical (emb-5) and cK12G. (7) Oryctolagus cuniculus Ad-Rab G (three from direct selection and one from exon trapping) (Boll et al., 1993). Strong nucleotide as well as amino acid similarity was found to this rabbit transcript, cloned in a subtractive hybridization of genes expressed in the intestine of adult but not baby rabbits. No functional characterization of this gene was reported.

(8) mouse fructose aldolase C gene (four from direct selection) (Paolella et al., 1986). Perfect nucleotide identity was found to this glycolytic enzyme.

(9) mouse rah GTP-binding protein (two direct selected clones) (Morimoto et al., 1991). Strong nucleotide similarity was found to this transcript, whose protein product may function in vesicular trafficking and neurotransmitter secretion.

(10) human MAC30 mRNA, 3' end sequence (one from direct selection and one from exon trapping) (Murphy et al., 1993). Strong nucleotide similarity was found to this transcript that is down-regulated in meningiomas and in tumors associated with neurofibromatosis 2.

Estimating the Number of Genes in the Region

If we assume that the genes in this region are of similar size and that fragments of these genes are recovered at similar frequencies by direct cDNA selection and exon trapping, then the number of genes can be estimated by four independent approaches.

(1) Number of times that fragments from specific genes were recovered. Of the 172 gene fragments examined, 37/172 (22%) showed strong sequence similarity to previously identified genes and could be grouped into 10 transcription units. The numbers of gene fragments corresponding to each of the 10 transcription units were 6, 5, 4, 4, 4, 3, 3, 2, and 2, with a mean of 3.7 ± 1.2. The actual mean number of hits to these 10 genes is probably somewhat higher, inasmuch as similarities to the untranslated regions would not be expected to have been recognized for the more distant similarities (e.g., C. elegans emb-5). Since four of the similarities are detected only on the amino acid level and since untranslated regions are typically about 40% as large as coding regions (J.S., unpublished observation based on a random sampling of genes from GenBank), the actual mean might be 20% larger—i.e., about 4.5. Assuming that the hit rate for these 10 genes is a good estimate of the hit rate across the region, we estimate that the 172 gene fragments represent between 38 (=172/4.5) and 46 (=172/3.7) genes.

(2) Number of overlaps among gene fragments. Of the 172 gene fragments with an average insert size of 250 bp, a total of 86 showed significant overlap (>40 bp) with another fragment. For a random collection of fragments, the expected number n of overlaps per clone is given by the formula $n = c(1 - \theta)$, where c is the degree of coverage of the region and $\theta$ is the minimum detectable proportion of overlap (Lander and Waterman, 1988). In the current case, $n = 86/172$ and $\theta = 40/250$. The estimated coverage of the transcribed portion of the region would thus be $c = 0.60$-fold. Since the clones contain a total of 43 kb of sequence (172 clones × 250 bp/clone), this would suggest that the transcribed portion of the region is about 72 kb (=43 kb/0.60). The proportion of the 370-kb region that is transcribed would thus be estimated to be about 20%. Taking the typical size of a mature transcript to be 2...
boundaries whenever possible. The direct cDNA se-
lected clones were superior to the exon trapped clones
because they tended to be longer and encompass se-
quence from two exons. Those gene fragments pre-
ferred to about 36 genes.

(3) Degree of coverage of known genes. Of the 10
defined transcription units corresponding to previously
known genes, three were known mouse genes (vitronect-
in, aldolase, rah GTP-binding protein) and three oth-
ers were mammalian genes showing strong sequence
similarity (various winged helix (fork head) genes, rat
sodium-sulfate cotransporter, and rabbit Ad-Rab G).
All gene fragments arising from the first group should
have been recognized (due to sequence identity), as
should most of those arising from coding regions in
the second group (due to apparently strong sequence
similarity across the coding region). For these six
genes, we could thus directly measure the degree of
coverage—that is, the average number of times that a
given nucleotide is hit. Since the six genes contain a
total of 13,775 nucleotides and the recovered gene frag-
ments a total of 6275 nucleotides, the coverage is 0.46-
fold. Assuming that this coverage is representative
for the region, the total length of transcribed sequence in
the region is estimated to be 93 kb (=43 kb in gene
fragments/0.46-fold coverage). The proportion of the
370-kb region that is transcribed would thus be esti-
imated to be about 25%. Again taking the typical size
of a mature transcript to be 2 kb, this would correspond
to about 46 genes.

(4) Proportion of genes similar to known genes. Fi-
ally, about 30% of newly sequenced mammalian genes
show strong sequence similarity to previously identi-
ified genes in GenBank (Adams et al., 1993) at present.
Since 10 such genes were identified in the region, this
would suggest a total of about 33 (=10/0.3) genes.

These four independent approaches suggest that
about 20–25% of the nucleotides in the nude region
are transcribed and that this region contains in the
range of 33–46 genes. The assumptions that the genes
in the region are recovered at similar frequencies and
are of similar size are unlikely to be exactly true, but
they are probably reasonable approximations. Differential
rates of recovery would tend to lead to underesti-
mates of the number of genes, while the presence of a
few exceptionally large genes would lead to overesti-
mates.

Expression Analysis of Transcription Units

To determine the expression pattern of each tran-
scription unit, we performed RT-PCR on a panel of
adult mouse tissues (including skin, thymus, liver, tes-
tes, etc.) and whole embryos at different days of devel-
ment (8.5, 9.5, 10.5, and 11.5 dpc). We minimized
the impact of genomic DNA contamination in the first-
strand cDNA template by spanning intron–exon
boundaries whenever possible. The direct cDNA se-
lected clones were superior to the exon trapped clones
because they tended to be longer and encompass se-
quence from two exons. Those gene fragments pre-
sumed to belong to a common transcription unit (by
virtue of strong similarity to a gene in GenBank) al-
ways showed a consistent pattern of expression (data
not shown).

Gene fragments, showing no sequence similarity to
previous genes, were grouped according to their physi-
cal position and expression pattern. We checked
whether several pairs of clones that did not overlap in
sequence, but both mapped to the same P1 clones and
were expressed in the same tissues were from the same
gene, by virtue of both being contained in the same
unique clone from a cDNA library. Clones were charac-
terized by hybridization to a Southern blot of restric-
tion-digested genomic DNA from wildtype and nude
mice. No deletions or alterations were found.

Mutations at the nude Locus

While we were completing the analysis of the candi-
date genes, Nehls et al. (1994) reported that the winged
helix homologue above has mutations in the mouse nu
allele and the rat rnuN allele. These authors originally
named the gene whn, for winged helix in nude. To be
consistent with mouse nomenclature, the gene has
been renamed Hfh11 for HNF-3/fork head homolog 11.

To analyze the rat rnu allele, we cloned and deter-
mined the entire coding sequence of the wildtype and
mutant rat Hfh11 genes (shown in Fig. 4). The rnu
allele is a nonsense mutation at bp 1429. We also con-
firmed the reported mutation in the mouse nu allele
and searched for the mutations in the nuwr allele. Inter-
estingly, no mutations were found in the coding region
in the nuwr allele. Northern blot analysis also indicates
that the Hfh11 transcript is present at approximately
normal levels in adult skin from AKR/J-nuwr/nuwr ho-
mozygotes (data not shown). Thus, the nuwr mutation
appears to be a more subtle change and remains to be
identified.

DISCUSSION

Genetic Resolution

We used F2 intercrosses segregating the nude pheno-
type to obtain two informative meioses for every prog-
eeny. Initial mapping studies with 200 animals localized
nude to a 1.6-cM interval between D11Mit7 and
D11Mit34. To determine fine structure mapping infor-
mation, we focused on the approximately 3% of the
progeny that were recombinant between the closest
SSLP markers flanking nude. To obtain the full meiotic
power, we progeny-tested those animals carrying re-
combinit chromosomes over wildtype chromosomes in
the nude region to determine which nude allele was

carried on the recombinant chromosome. With N
meioses, the distance to the closest flanking crossover
on either side of the locus will be exponentially distrib-
uted with an expected distance of 1/N morgans. Thus,
the recombinationally inseparable interval containing
the gene will have the expected size 2/N morgans. Map-
 POSITIONAL CLONING OF THE nude LOCUS 557

FIG. 4. cDNA sequence of the rat nude gene. Amino acids differing from those in the mouse transcript appear in boldface. The site of the single basepair change from T to A at nucleotide 1429 resulting in the nonsense mutation of the gene. Amino acids differing from those in the mouse transcript appear in boldface. The site of the single basepair change from T to A at nucleotide 1429 resulting in the nonsense mutation of the nude allele is enlarged and underlined.
ping *nude* with a total of 2000 meioses should thus yield an interval of 1/10 cM = 214 kb (3000 Mb/1400 cM = physical size of the mouse genome/genetic size of the mouse genome). In fact, we narrowed the *nude* locus to a minimum region of 230 kb and a maximum of 370 kb.

**Physical Map**

With nearly 6000 SSLP markers currently available in the mouse, the average spacing between markers is 500 kb and the average interval around a target locus is 1 Mb, distances easily spanned in large insert YACs. In our case, although no MIT SSLP marker is contained in YAC 31, D11Mit144 and D11Mit117 are contained in the adjacent YAC, WI FEL D9 (see Fig. 1). To identify the closest markers flanking a locus and to initiate a walk in P1s, a dense set of genetic markers can be subcloned from the YACs. The P1s are a good starting material for identifying transcription units because the insert DNA is easily purified away from the host DNA. The P1 contig also provides an independent verification of the structure of the genomic region.

**Finding Transcription Units**

In our hands, direct cDNA selection proved to be a powerful method for identifying a deep pool of unique sequences, corresponding to transcription units in a genomic region. Direct cDNA selection identified 148 unique transcripts of average insert size 250-bp from the 370-kb *nude* region. The only caveat is that the method demands knowledge of the expression pattern of the gene to use as a source of cDNA for the selection. Exon trapping did not give the depth of resources, but it did identify many of the same transcription units as direct selection.

**Gene Density**

The number of genes in the mammalian genome is estimated to be about 100,000. This would correspond to about 12 genes in the 370 kb shown to contain the *nude* locus. However, some regions appear to be gene rich (high GC content or Alu-rich Giemsa light bands), having a much higher gene density than average. In fact, the middle region of mouse chromosome 11, to which *nude* maps, has light Giemsa band staining (Buchberg and Camper, 1993). To estimate the total number of genes in the *nude* region, we employed four methods based on the redundancy of gene fragments, the overlap among the total set of gene fragments, the coverage of known genes in the region, and the proportion of gene fragments showing similarity to genes in GenBank. These four methods indicated that 20–25% of the nucleotides in the *nude* region are transcribed and that the expected number of genes in the smallest *nude* region is in the range of 33 to 46; this would seem to indicate that *nude* lies in a gene-rich region of the genome. Alternatively, if the *nude* locus resides in a region of typical gene density, then the mouse genome would contain more than 300,000 genes—which seems unlikely.

**Mutations in the *nude* Gene**

This method succeeded in finding the gene, a novel winged helix/fork head protein, that when disrupted results in the *nude* phenotype. Fork head genes are a family of transcription factors that are varied in their expression patterns in embryonic development and adult expression. We report here a novel rat allele, *rnu*, in the *nude* gene, that results in a nonfunctional protein. Further studies are required to address precisely how this novel transcription factor produces the pleiotropic phenotype of hairlessness and athymia. The subtle nature of the genomic changes in all of the *nude* alleles except *rnu* reinforces the value of having several distinct alleles of a gene in a positional cloning project.

Finally, we mention that pronuclear injection of a cosmid containing the *Hfh11* gene into a *nude* blastocyst has successfully rescued the hairless phenotype of the mouse (H. Kurooka, T. Honjo, J.A.S., and E.S.L., unpublished data). The results of these experiments will be reported elsewhere.

**ACKNOWLEDGMENTS**

We thank Aaron Chang and Armand MacMurray for help with direct cDNA selection; Richard Resnick, Lincoln Stein, Mary Pat Reeve-Daly, Mark Daly, and Nathan Goodman for computer analysis of the gene fragments; Ricardo Proenca, Jeffrey Friedman, and Alan Buckler for help with exon trapping; and Don Varnum and Carolyn Dunbar for help with the crosses and progeny testing. This work was supported in part by Grants NIH HG00189 (to E.S.L.), HD25389 (to J.H.N.), and GM18694 (to B.A.T.).

**REFERENCES**


