(80)

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

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 i 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

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.
Nadeau, J.H., Frankel, W.N., Wigler, M.H., Lander, E.S. (1994).
Direct isolation of polymorphic markers linked to a trait
by genetically directed representational difference analysis.
Nature Genetics 6: 57-63.

Appendix III: Segre, J.A., Nemhauser, J.L., Taylor, B.A., Nadeau,
J.H., Lander, E.S. (1995). Positional cloning of the nude locus:
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 Apendix 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.

ABSTRACT

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 germanitive 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), wheras 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 assemby 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; Hirotsune, 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

Bach-Mortensen, N., Romert, P. and Ballegaard (1976). Transplantation of human adipose tissue to "nude" mice.*Microbiol. Scand.*,84, 283-289.

Bastert, G., Althoff, P., Usadel, K. H., Fortmeyer, H. P. and Schmidt, M. (1977). Heterotransplantation of human fetal pituitaries in "nude" mice.*Endocrinology.*,101, 365-368.

Bosma, G. C., Custer, R. P. and Bosma, M. J. (1983). A severe combined immunodeficiency mutation in the mouse.*Nature.*,301, 527-30.

Bultman, S. J., Michaud, E. J. and Woychik, R. P. (1992). Molecular characterization of the mouse agouti locus.*Cell.*,71, 1195-204.

Chisaka, O. and Capecchi, M. R. (1991). Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene hox-1.5.*Nature.*,350, 473-9.

Collins, F. S. (1992). Positional cloning: Let's not call it reverse anymore.*Nature Genetics.*,1, 3-6.

Cordes, S. P. and Barsh, G. S. (1994). The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor.*Cell.*,79, 1025-34.

Cordier, A. C. and Haumont, S. M. (1980). Development of thymus, parathyroids, and ultimo-branchial bodies in NMRI and nude mice.*Am. H. Anat.*,157, 227-263.

Coulombe, P. A. (1993). The cellular and molecular biology of keratins: beginning a new era.*Curr Opin Cell Biol.*,5, 17-29.

D'Arcangelo, G., Miao, G. G., Chen, S. C., Soares, H. D., Morgan, J. I. and Curran, T. (1995). A protein related to extracellular matrix proteins deleted in the mouse mutant reeler.*Nature.*,374, 719-23.

Di George, A. M. (1965). A new concept of the cellular basis of immunity.*J. Pediatr.*,67, 907-8.

Dolle, P., Ruberte, E., Leroy, P., Morriss-Kay, G. and Chambon, P. (1990). Retinoic acid receptors and cellular retinoid binding proteins. I. A systematic study of their differential pattern of transcription during mouse organogenesis.*Development.*,110, 1133-51.

Dry, F. W. (1926). The coat of the mouse (*Mus musculus*).*J. Genet.*,16, 16-25.

Ebling, F. J. and Hale, P. A. (1983) Biochemistry and Physiology of the Skin. In (ed. L. A. Goldsmith). pp. 522-562. Oxford: Oxford University Press.

Flanagan, S. (1966). "Nude", a new hairless gene with pleiotropic effects in the mouse.*Genetic Research.*,8, 295-309.

Gilbert, S. F. (1995).*Developmental Biology.*. Sunderland, MA: Sinauer Associates, Inc.

Gillespie, J. M. (1990) The proteins of hair and other hard α-keratins. In *Cellular and Molecular Biology of Intermediate Filaments.* (ed. R. D. Golman and P. M. Steinert). pp. 95. New York: Plenum Press.

Hammond, E. M. (1954). Origin of thymus in the chick embryo.*J. Morphol.*,95, 501-521.

Hardy, M. H. (1969).*Advances in Biology of Skin.*. New York: Pergamon Press.

Hardy, M. H. (1992). The secret life of the hair follicle.*Trends Genet.*,8, 55-61.

Hebert, J. M., Rosenquist, T., Gotz, J. and Martin, G. R. (1994). FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations.*Cell.*,78, 1017-25.

Heid, H. W., Werner, E. and Franke, W. W. (1986). The complement of native alpha-keratin polypeptides of hair-forming cells: a subset of eight polypeptides that differ from epithelial cytokeratins.*Differentiation.*,32, 101-19.

Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse.*Nature.*,343, 617-22.

Hirai, Y., Nose, A., Kobayashi, S. and Takeichi, M. (1989). Expression and role of E- and P-cadherin adhesion molecules in embryonic histogenesis. II. Skin morphogenesis.*Development.*,105, 271-7.

Hirotsune, S., Takahara, T., Sasaki, N., Hirose, K., Yoshiki, A., Ohashi, T., Kusakabe, M., Murakami, Y., Muramatsu, M. and Watanabe, S. (1995). The reeler gene encodes a protein with an EGF-like motif expressed by pioneer neurons.*Nature Genetics.*,10, 77-83.

Houchens, D. P. and Ovejera, A. A. (1978).*Proceedings of the symposium on the use of athymic (nude) mice in cancer research.*. New York: Gustav Fischer.

Issacson, J. H. and Cattanach, B. M. (1962). Report.*Mouse News Lett.*,27, 31.

Jones, C. M., Lyons, K. M. and Hogan, B. L. (1991). Involvement of Bone Morphogenetic Protein-4 (BMP-4) and Vgr-1 in morphogenesis and neurogenesis in the mouse.*Development.*,111, 531-42.

Jones, P. H., Harper, S. and Watt, F. M. (1995). Stem cell patterning and fate in human epidermis.*Cell.*,80, 83-93.

Kaufman, M. H. (1992).*The Atlas of Mouse Development*.. London: Academic Press Limited.

Kingsley, D. M., Bland, A. E., Grubber, J. M., Marker, P. C., Russell, L. B., Copeland, N. G. and Jenkins, N. A. (1992). The mouse short ear skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGF beta superfamily.*Cell.*,71, 399-410.

Kopf-Maier, P., Mboneko, V. F. and Merker, H. J. (1990). Nude mice are not hairless. A Morphological Study.*Act Anat.*,139, 178-190.

Lavker, R. M. and Sun, T. T. (1995). Hair follicle stem cells: present concepts.*J Invest Dermatol.*,104, 38S-39S.

LeLievre, C. S. and LeDouarin, N. M. (1975). Mesenchymal derivatives of the neural crest: analysis of chimaeric quail and chick embryos.*J. Embryol. exp. Morph.*,34, 125-154.

Luetteke, N. C., Phillips, H. K., Qiu, T. H., Copeland, N. G., Earp, H. S., Jenkins, N. A. and Lee, D. C. (1994). The mouse waved-2 phenotype results from a point mutation in the EGF receptor tyrosine kinase.*Genes Dev.*,8, 399-413.

Luetteke, N. C., Qiu, T. H., Peiffer, R. L., Oliver, P., Smithies, O. and Lee, D. C. (1993). TGF alpha deficiency results in hair follicle and eye abnormalities in targeted and waved-1 mice.*Cell.*,73, 263-78.

Manley, N. R. and Capecchi, M. R. (1995). The role of Hoxa-3 in mouse thymus and thyroid development.*Development.*,121, 1989-2003.

Mann, G. B., Fowler, K. J., Gabriel, A., Nice, E. C., Williams, R. L. and Dunn, A. R. (1993). Mice with a null mutation of the TGF alpha gene have abnormal skin architecture, wavy hair, and curly whiskers and often develop corneal inflammation.*Cell.*,73, 249-61.

McCusick, V. A. (1990).*Mendelian inheritance in man: Catalog of autosomal dominant, autosomal recessive, and X-linked phenotypes.*. Baltimore: Johns Hopkins Univ Press.

Messenger, A. G. (1993). The control of hair growth: an overview.*J Invest Dermatol.*,101, 4S-9S.

Miettinen, P. J., Berger, J. E., Meneses, J., Phung, Y., Pedersen, R. A., Werb, Z. and Derynck, R. (1995). Epithelial immaturity and multiorgan failure in mice lacking epidermal growth factor receptor.*Nature.*,376, 337-41.

Miller, J. F. A. P. (1961). Immunological function of the thymus.*Lancet.*,2, 748.

Miller, M. W., Duhl, D. M., Vrieling, H., Cordes, S. P., Ollmann, M. M., Winkes, B. M. and Barsh, G. S. (1993). Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation.*Genes Dev.*,7, 454-67.

Mombaerts, P., Iacomini, J., Johnson, R. S., Herrup, K., Tonegawa, S. and Papaioannou, V. E. (1992). RAG-1-deficient mice have no mature B and T lymphocytes.*Cell.*,68, 869-77.

Pantelouris, E. (1968). Absence of Thymus in a Mouse Mutant.*Nature.*,217, 370-371.

Pantelouris, E. M. (1973). Athymic development in the mouse.*Differentiation.*,1, 437-50.

Pantelouris, E. M. (1973). Athymic development in the mouse.*Differentiation.*,1, 437-450.

Patl, N., Cox, D. R., Bhat, D., Faham, M., Myers, R. M. and Peterson, A. S. (1995). A potassium channel mutation in weaver mice implicates membrane excitability in granule cell differentiation.*Nat Genet.*,11, 126-129.

Potten, C. S. (1985).*Radiation and skin*.. London: Taylor and Francis.

Ruberte, E., Dolle, P., Krust, A., Zelent, A., Morriss-Kay, G. and Chambon, P. (1990). Specific spatial and temporal distribution of retinoic acid receptor gamma transcripts during mouse embryogenesis.*Development.*,108, 213-22.

Rugh, R. (1990).*The Mouse: Its Reproduction and Development*.. Oxford: Oxford University Press.

Rygaard, J. (1991) History and Pathology. In *The Nude Mouse in Oncology Research.* (ed. E. Boven and B. Winograd). pp. Boca Raton, FL: CRC Press, Inc.

Segre, J. A., Nemhauser, J. L., Taylor, B. A., Nadeau, J. H. and Lander, E. S. (1995). Positional Cloning of the nude Locus: Genetic, Physical and Transcription Maps of the Region and Mutations in the Mouse and Rat.*Genomics.*,28, 549-559.

Sengel, P. (1986) Biology of the Integument. In *Vertebrates.* (ed. J. Breiter-Hahn, A. G. Matoltsy and K. S. Richards). pp. 374-408. Springer-Verlag.

Sengel, P. (1975).*Morphogenesis of Skin.*. Cambridge: Cambridge University Press.

Sibilia, M. and Wagner, E. F. (1995). Strain-dependent epithelial defects in mice lacking the EGF receptor.*Science.*,269, 234-8.

Stewart, M. (1993). Intermediate filament structure and assembly.*Curr Opin Cell Biol.*,5, 3-11.

Sundberg, J. P. and Schultz, L. D. (1991). Inherited mouse mutations: models for the study of alopecia.*J Invest Dermatol.*,96, 95S-96S.

Threadgill, D. W., Dlugosz, A. A., Hansen, L. A., Tennenbaum, T., Lichti, U., Yee, D., La Mantia, C., Mourton, T., Herrup, K., Harris, R. C. and et, a. (1995). Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype.*Science.*,269, 230-4.

Van Vilet, E., Jenkinson, E. J., Kingston, R., Owen, J. J. and Van Ewijk, W. (1985). Stromal cell types in the developing thymus of the normal and nude mouse embryo.*Eur J Immunol.*,15, 675-81.

Vidal, S. M., Malo, D., Vogan, K., Skamene, E. and Gros, P. (1993). Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg.*Cell.*,73, 469-85.

Wilson, C., Cotsarelis, G., Wei, Z. G., Fryer, E., Margolis-Fryer, J., Ostead, M., Tokarek, R., Sun, T. T. and Lavker, R. M. (1994). Cells within the bulge region of mouse hair follicle transiently proliferate during early anagen: heterogeneity and functional differences of various hair cycles.*Differentiation.*,55, 127-36.

Wortis, H. H., Nehlsen, S. and Owen, J. J. (1971). Abnormal development of the thymus in nude mice.*J. Exp. Med.*,134, 681.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L. and Friedman, J. M. (1994). Positional cloning of the mouse obese gene and its human homologue.*Nature.*,372, 425-32.

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

$(GA)_n$) 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 *nude* locus in 2000 meioses in three separate F_2 intercrosses. Initial mapping studies with SSLPs localized *nude* to a 1.7 cM interval between *D11Mit7* and *D11Mit34*. Two implementations of RDA were carried out, using congenic strains and F_2 intercross progeny. Three clones were produced and all mapped within 0.5 cM of *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 *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^{str}); Two alleles of *nude* exist in the rat: *rnu* and rnu^{N}. 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^{str}): 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^{N}): 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^{str} 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^{str} 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 F_2 intercrosses. To ensure the greatest rates of polymorphisms between the two strains of the cross, we mated AKR/J-*nu*^{str} 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*^{str} and C57BL/6J. To minimize background effects, all of our crosses were performed with the *nu*^{str} allele because it arose and has been maintained on the inbred strain AKR/J. Specifically we generated 182 (AKR-*nu*^{str} X C57BL/6J) F_2 animals, 650 (AKR-*nu*^{str} X MOLF/Ei) F_2 animals, and 226 (AKR-*nu*^{str} X CAST/Ei) F_2 animals. When animals are referred to in the text of this thesis, an (AKR-*nu*^{str} X C57BL/6J) F_2 animals is abbreviated as B#; an (AKR-*nu*^{str} X CAST/Ei) F_2 animals is abbreviated as C#; and an (AKR-*nu*^{str} X MOLF/Ei) F_2 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 F_2 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-nu^{str} x C57BL/6J) F_2 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 F_2 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 F_2 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-*nu*^{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 F_2 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 F_2 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 F_2 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-nu^{str} X MOLF/Ei) F_2 intercross progeny as Driver and the MOLF/Ei parent as the Tester. At this

point, we had generated 416 (AKR-*nu*^{str} X MOLF/Ei) F_2 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 *Bgl*II. RDA-6.2 detected an RFLP with a 450 bp allele in MOLF/Ei and a 4 kb allele in AKR/J-*nu*^{str}. To obtain approximate localization, we genotyped 20 (AKR-*nu*^{str} X MOLF/Ei) F_2 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* F_2 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 F_2 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 *Bam*HI 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 *Bam*HI (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* F_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 cM interval between *D11Mit7* and *D11Mit34*. Subsequent analysis of additional F_2 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* F_2 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) F_2 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 *Bam*HI 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 nondenaturing 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 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 F_2 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 F_2 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 *nude* locus to a minimum region of 230 kb and a maximum of 370 kb.

REFERENCES

Avner, P., Amar, L., Dandolo, L. and Guenet, J. L. (1988). Genetic analysis of the mouse using interspecific crosses.*trend Genet.*,4, 18-23.

Berridge, M., O'Kech, N., McNeilage, L., Heslop, B. and Moore, R. (1979). Rat mutant showing "nude" characteristics.*Transplantation.*,27, 419.

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism.*Am. J. Hum. Genet.*,32, 314-321.

Collins, F. S. (1992). Positional cloning: Let's not call it reverse anymore.*Nature Genetics.*,1, 3-6.

Copeland, N. G. and Jenkins, N. A. (1991). Development and applications of a molecular genetic linkage map of the mouse genome.*Trends Genet.*,7, 113-8.

Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., Steen, R. G., Stein, L. D., Nadeau, J. H. and Lander, E. S. (1993). A genetic linkage map of the mouse: current applications and future prospects.*Science.*,262, 57-66.

Davisson, M. T., Roderick, T. H. and Doolittle, D. P. (1989) Recombination percentages and chromosomal assignments. In *Genetic Variants and Strains of the Laboratory Mouse, Ed. 2.* (ed. M. F. Lyon and A. Searle). pp. 432-505. New York: Oxford University Press.

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H. S., Friedman, J., Dracopoli, N. C. and Lander, E. S. (1992). A genetic map of the mouse suitable for typing intraspecific crosses.*Genetics.*,131, 423-47.

Dietrich, W. F., Miller, J. C., Steen, R. G., Merchant, M., Damron, D., Nahf, R., Gross, A., Joyce, D. C., Wessel, M., Dredge, R. D., Marquis, A., Stein, L. D., Goodman, N., Page, D. C. and Lander, E. S. (1994). A genetic map of the mouse with 4,006 simple sequence length polymorphisms.*Nat Genet.*,7, 220-245.

Dietrich, W. F., Miller, J. C., Steen, R. G., Merchant, M. A., Damron-Boles, D., Husain, Z., Dredge, R. D., Daly, M. J., Ingalls, K. A., O'Connor, T. J., Evans, C. A., DeAngelis, M. M., Levinson, D. M., Kruglyak, L., Goodman, N., Copeland, N. G., Jenkins, N. A., Hawkins, T. L., Stein, L. D., Page, D. C. and Lander, E. S. (1995). A Comprehensive Genetic Map of the Mouse Genome. *Nature.*, submitted.

Eicher, E. (1976). Remutations.*Mouse News Letter.*,54, 90.

Festing, M., May, D., Connors, T., Lovell, D. and Sparrow, S. (1978). An athymic nude mutation in the rat.*Nature.*,274, 365.

Festing, M. F. W. (1979).*Inbred Strains in Biomedical Research.*. New York: Oxford University Press.

Flanagan, S. (1966). "Nude", a new hairless gene with pleiotropic effects in the mouse.*Genetic Research.*,8, 295-309.

Green, M. C. (1989) Catalog of mutant genes and polymorphic loci. In *Genetic Variants and Strains of the Laboratory Mouse, 2 ed.* (ed. M. F. Lyon and A. G. Searle). pp. 12-403. New York: Oxford University Press.

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y. and et, a. (1983). A polymorphic DNA marker genetically linked to Huntington's disease.*Nature.*,306, 234-8.

Hammer, M. F., Schimenti, J. and Silver, L. M. (1989). Evolution of mouse chromosome 17 and the origin of inversions associated with t haplotypes.*Proc Natl Acad Sci U S A.*,86, 3261-5.

Issacson, J. H. and Cattanach, B. M. (1962). Report.*Mouse News Lett.*,27, 31.

Jenkins, N. A., Rothe, H., Gilbert, D. J., Copeland, N. G. and Kolb, H. (1994). Mapping of the gene for inducible nitric oxide (NO) synthase of mouse

macrophages to chromosome 11, close to Evi-2, nu, and Idd-4.*Genomics.*,19, 402-4.

Lisitsyn, N., Lisitsyn, N. and Wigler, M. (1993). Cloning the differences between two complex genomes.*Science.*,259, 946-51.

Love, J. M., Knight, A. M., McAleer, M. A. and Todd, J. A. (1990). Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites.*Nucleic Acids Res.*,18, 4123-30.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms.*Proc Natl Acad Sci U S A.*,86, 2766-70.

Pantelouris, E. M. (1973). Athymic development in the mouse.*Differentiation.*,1, 437-50.

Reed, C. and O'Donoghue, J. (1979). A new guinea pig mutant with abnormal hair production and immunodeficiency.*Lab. Animal Sci.*,29, 744.

Solomon, J.-C. and Lynch, N. (1977). Discrepancies in nude mice.*Biomedicine.*,26, 77.

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E. and Moyzis, R. K. (1991). Evolution and distribution of (GT)n repetitive sequences in mammalian genomes.*Genomics.*,10, 807-15.

Vidal-Puig, A. and Moller, D. E. (1994). Comparative sensitivity of alternative single-strand conformation polymorphism (SSCP) methods.*Biotechniques.*,17, 490-2.

Zha, H., Remmers, E. F., Du, Y., Cash, J. M., Goldmuntz, E. A., Crofford, L. J. and Wilder, R. L. (1995). The rat athymic nude (*rnu*) locus is closely linked to the inducible nitric oxide synthase gene (*Nos2*).*Mammalian Genome.*,6, 137-138.

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 *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 *nude* locus. This process was continued until markers that recombine with *nude* both proximally and distally were contained in the same chromosomal walk.

Although we eventually targeted markers tightly linked to *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 *D11Mit7* and *D11Mit34*, which map 0.7 cM proximal and 0.7 cM distal to *nude*, respectively. The walk in smaller insert bacteriophage 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.

RESULTS

Screening YAC libraries

When we initiated the physical mapping of *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, *D11Mit7* and *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 chromosomal 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* F_2 intercrosses. 54% (15/28) of the non-chimeric ends detected a polymorphism between AKR/J-nu^{str} and CAST/Ei

or MOLF/Ei that segregated in the *nude* F_2 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, F_2 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-nu^{str} 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-nu^{str} 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 inititiated 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 SSCPs (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: *Mlu*I, *Not*I, *Rsr*II, *Sac*I and *Sfi*I. 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 *Not*I site of the vector and P15S is the end of P1 clone #15 adjacent to the *Sal*I 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 a 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

Albertsen, H. M., Abderrahim, H., Cann, H. M., Dausset, J., Le Paslier, D. and Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents.*Proc Natl Acad Sci U S A.*,87, 4256-60.

Banfi, S., Ledbetter, S. A., Chinault, A. C. and Zoghbi, H. Y. (1992). An easy and rapid method for the detection of chimeric yeast artificial chromosome clones.*Nucleic Acids Res.*,20,

Bates, G. P., Valdes, J., Hummerich, H., Baxendale, S., Le Paslier, D. L., Monaco, A. P., Tagle, D., MacDonald, M. E., Altherr, M., Ross, M. and et, a. (1992). Characterization of a yeast artificial chromosome contig spanning the Huntington's disease gene candidate region.*Nat Genet.*,1, 180-7.

Bender, W., Spierer, P. and Hogness, D. S. (1983). Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in Drosophila melanogaster.*J Mol Biol.*,168, 17-33.

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

Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billaut, A., Guasconi, G., Gervy, P., Le Gall, I., Soularue, P., Grinas, L. and et, a. (1992). Continuum of overlapping clones spanning the entire human chromosome 21q [see comments].*Nature.*,359, 380-7.

Coulson, A., Waterston, R., Kiff, J., Sulston, J. and Kohara, Y. (1988). Genome linking with yeast artificial chromosomes.*Nature.*,335, 184-6.

Foote, S., Vollrath, D., Hilton, A. and Page, D. C. (1992). The human Y chromosome: overlapping DNA clones spanning the euchromatic region.*Science.*,258, 60-6.

Garza, D., Ajioka, J. W., Burke, D. T. and Hartl, D. L. (1989). Mapping the Drosophila genome with yeast artificial chromosomes.*Science.*,246, 641-6.

Green, E. D. and Olson, M. V. (1990). Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction.*Proc Natl Acad Sci U S A.*,87, 1213-7.

Haldi, M., Perrot, V., Saumier, M., Desai, T., Cohen, D., Cherif, D., Ward, D. and Lander, E. S. (1994). Large human YACs constructed in a rad52 strain show a reduced rate of chimerism.*Genomics.*,24, 478-84.

Hudson, T. J., Stein, L. D., Gerety, S. S., Ma, J., Castle, A. B., Silva, J., Slonim, D. K., Baptista, R., Kruglyak, L., Xu, S.-H. and others, a. (1995). An STS-Based Map of the Human Genome.*Science.*,in press.

Hunter, K., Housman, D. and Hopkins, N. (1991). Isolation and characterization of irradiation fusion hybrids from mouse chromosome 1 for mapping Rmc-1, a gene encoding a cellular receptor for MCF class murine retroviruses.*Somat Cell Mol Genet.*,17, 169-83.

Hwang, I., Kohchi, T., Hauge, B. M., Goodman, H. M., Schmidt, R., Cnops, G., Dean, C., Gibson, S., Iba, K., Lemieux, B. and et, a. (1991). Identification and map position of YAC clones comprising one-third of the Arabidopsis genome.*Plant J.*,1, 367-74.

Joslyn, G., Carlson, M., Thliveris, A., Albertsen, H., Gelbert, L., Samowitz, W., Groden, J., Stevens, J., Spirio, L., Robertson, M. and et, a. (1991). Identification of deletion mutations and three new genes at the familial polyposis locus.*Cell.*,66, 601-13.

Killary, A. M. and Fournier, R. E. (1984). A genetic analysis of extinction: trans-dominant loci regulate expression of liver-specific traits in hepatoma hybrid cells.*Cell.*,38, 523-34.

Kusumi, K., Smith, J. S., Segre, J. A., Koos, D. S. and Lander, E. S. (1993). Construction of a large-insert yeast artificial chromosome library of the mouse genome.*Mamm Genome.*,4, 391-2.

Rossi, J. M., Chen, H. and Tilghman, S. M. (1994). Genetic map of the fused locus on mouse chromosome 17.*Genomics.*,23, 178-84.

Selleri, L., Eubanks, J. H., Giovannini, M., Hermanson, G. G., Romo, A., Djabali, M., Maurer, S., McElligott, D. L., Smith, M. W. and Evans, G. A. (1992). Detection and characterization of "chimeric" yeast artificial chromosome clones by fluorescent in situ suppression hybridization.*Genomics.*,14, 536-41.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector.*Proc Natl Acad Sci U S A.*,89, 8794-7.

Triglia, T., Peterson, M. G. and Kemp, D. J. (1988). A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences.*Nucleic Acids Res.*,16, 8186.

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 ommatatidia. 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 θ is the minimum detectable proportion of overlap(Lander, et al., 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 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

Adams, M. D. and others, a. (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence.*Nature.*,377, 3-174.

Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C. and Venter, J. C. (1993). Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library.*Nat Genet.*,4, 373-80.

Altschul, S. F., Boguski, M. S., Gish, W. and Wootton, J. C. (1994). Issues in searching molecular sequence databases.*Nat Genet.*,6, 119-29.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool.*J Mol Biol.*,215, 403-10.

Altschul, S. F. and Lipman, D. J. (1990). Protein database searches for multiple alignments.*Proc Natl Acad Sci U S A.*,87, 5509-13.

Antequera, F. and Bird, A. (1993). Number of CpG islands and genes in human and mouse.*Proc Natl Acad Sci U S A.*,90, 11995-9.

Antequera, F. and Bird, A. (1994). Predicting the total number of human genes [letter].*Nat Genet.*,8, 114.

Boll, W., Schmid, C. T., Semenza, G. and Mantei, N. (1993). Messenger RNAs expressed in intestine of adult but not baby rabbits. Isolation of cognate cDNAs and characterization of a novel brush border protein with esterase and phospholipase activity.*J Biol Chem.*,268, 12901-11.

Buchberg, A. M. and Camper, S. A. (1993). Encyclopedia of the mouse genome III. October 1993. Mouse chromosome 11.*Mamm Genome.*,S164-75.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A. and Housman, D. E. (1991). Exon amplification: a strategy to isolate

mammalian genes based on RNA splicing.*Proc Natl Acad Sci U S A.*,88, 4005-9.

Choi, K. W. and Benzer, S. (1994). Rotation of photoreceptor clusters in the developing Drosophila eye requires the nemo gene.*Cell.*,78, 125-36.

Church, D. M., Stotler, C. J., Rutter, J. L., Murrell, J. R., Trofatter, J. A. and Buckler, A. J. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification.*Nat Genet.*,6, 98-105.

Fields, C., Adams, M. D., White, O. and Venter, J. C. (1994). How many genes in the human genome? [news].*Nat Genet.*,7, 345-6.

Gilbert, W. (1992) The Code of Codes. In (ed. D. J. Kevles and L. Hood). pp. 83-97. Harvard University Press.

Hudson, T. J., Stein, L. D., Gerety, S. S., Ma, J., Castle, A. B., Silva, J., Slonim, D. K., Baptista, R., Kruglyak, L., Xu, S.-H. and others, a. (1995). An STS-Based Map of the Human Genome.*Science.*,in press.

Lander, E. S. and Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones: a mathematical analysis.*Genomics.*,2, 231-9.

Lewin, B. (1995).*Genes*.. Oxford: Oxford University Press.

Li, C., Lusis, A. J., Sparkes, R., Tran, S. M. and Gaynor, R. (1992). Characterization and chromosomal mapping of the gene encoding the cellular DNA binding protein HTLF.*Genomics.*,13, 658-64.

Lopez, R., Larsen, F. and Prydz, H. (1994). Evaluation of the exon predictions of the GRAIL software.*Genomics.*,24, 133-6.

Lovett, M. (1994) Direct Selection of cDNAs Using Genomic Contigs. In *Current Protocols in Human Genetics.* (ed. N. C. Dracopoli, J. L. Haines, B. R. Korf, D. T. Moir, C. C. Morton, C. E. Seidman, J. G. Seidman and D. R. Smith). pp. New York: Current Protocols.

Lovett, M., Kere, J. and Hinton, L. M. (1991). Direct selection: a method for the isolation of cDNAs encoded by large genomic regions.*Proc Natl Acad Sci U S A.*,88, 9628-32.

Markovich, D., Forgo, J., Stange, G., Biber, J. and Murer, H. (1993). Expression cloning of rat renal Na+/SO4(2-) cotransport.*Proc Natl Acad Sci U S A.*,90, 8073-7.

Martin-Gallardo, A., McCombie, W. R., Gocayne, J. D., Fitz Gerald, M. G., Wallace, S., Lee, B. M., Lamerdin, J., Trapp, S., Kelley, J. M., Liu, L. I. and et, a. (1992). Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3.*Nat Genet.*,1, 34-9.

McCombie, W. R., Martin-Gallardo, A., Gocayne, J. D., Fitz Gerald, M., Dubnick, M., Kelley, J. M., Castilla, L., Liu, L. I., Wallace, S., Trapp, S. and et, a. (1992). Expressed genes, Alu repeats and polymorphisms in cosmids sequenced from chromosome 4p16.3.*Nat Genet.*,1, 348-53.

Morimoto, B. H., Chuang, C. C. and Koshland, D. J. (1991). Molecular cloning of a member of a new class of low-molecular-weight GTP-binding proteins.*Genes Dev.*,5, 2386-91.

Murphy, M., Pykett, M. J., Harnish, P., Zang, K. D. and George, D. L. (1993). Identification and characterization of genes differentially expressed in meningiomas.*Cell Growth Differ.*,4, 715-22.

Nishiwaki, K., Sano, T. and Miwa, J. (1993). emb-5, a gene required for the correct timing of gut precursor cell division during gastrulation in Caenorhabditis elegans, encodes a protein similar to the yeast nuclear protein SPT6.*Mol Gen Genet.*,239, 313-22.

Paolella, G., Buono, P., Mancini, F. P., Izzo, P. and Salvatore, F. (1986). Structure and expression of mouse aldolase genes. Brain-specific aldolase C amino acid sequence is closely related to aldolase A.*Eur J Biochem.*,156, 229-35.

Roberts, L. (1991). GRAIL seeks out genes buried in DNA sequence [news].*Science.*,254, 805.

Seiffert, D., Poenninger, J. and Binder, B. R. (1993). Organization of the gene encoding mouse vitronectin.*Gene.*,134, 303-4.

Wagner, R., Maguire, M. and Stallings, R. (1993).*Chromosomes: A Synthesis.*. New York: Wiley.

Wolf, F. W., Marks, R. M., Sarma, V., Byers, M. G., Katz, R. W., Shows, T. B. and Dixit, V. M. (1992). Characterization of a novel tumor necrosis factor-alpha-induced endothelial primary response gene.*J Biol Chem.*,267, 1317-26.

Xu, G. F., O'Connell, P., Viskochil, D., Cawthon, R. M., Robertson, M., Culver, M., Dunn, D., Stevens, J., Gesteland, R., White, R. and Weiss, R. (1990). The Neurofibromatosis Type 1 Gene Encodes a Protein Related to GAP.*Cell.*,62, 599-608.

Xu, Y., Mural, R., Shah, M. and Uberbacher, E. (1994). Recognizing exons in genomic sequence using GRAIL II.*Genet Eng.*,16, 241-53.

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 ethylnitrosourea (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 alveloar 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; Clevidence, 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, (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 *nude* genomic region an intriguiging candidate as the causative gene for the *nude* phenotype. Mutations were identified in this novel FH gene in all four *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*Hfh11*, is the *nude* gene, we microinjected a cosmid clone containing the wild-type *nude* genomic locus into fertilized *nude* eggs, correcting the hairless, but not the athymic phenotype. 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. The mutations detected in four alleles of the gene and the rescue of the hairless phenotype demonstrate that *Hfh11* is the *nude* gene.

The expression pattern of *Hfh11* is consistent with the phenotype observed when the gene is disrupted. *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 *Taq*I and *Hind*III 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 rnu^N allele. These authors originally named the gene *whn*, for *w*inged *h*elix in *nude*. To be consistent with mouse nomenclature, the gene has been renamed *Hfh11* for *H*NF-3/ *f*ork *h*ead 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^{str} 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^{str} homozygotes, as compared to AKR/J (Figure 20). Thus, the nu^{str} 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 rnu^N allele has two variant transcripts of about 3 kb and 1.5 kb in adult skin. Analysis of the genomic rnu^N 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 rnu^N 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 rnu^N 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 rnu^N 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 nu^{str} skin, a nonsense mutation in the *rnu* allele, and a large genomic deletion removing several exons of the rnu^N 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. *Hind*III 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 germanitive 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 nu^{str} skin, a nonsense mutation in the *rnu* allele, and a large genomic deletion removing several exons of the rnu^{N} allele. The subtle nature of the genomic changes in all of the *nude* alleles except rnu^{N} 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 *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 *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 *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 *Hfh11* cDNA to resolve this question. Alternatively, the transgenic mice may be expressing *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.

Using a positional cloning approach, isolation of putative genes causing genetic defects have been recently reported in mouse, as well as in human (Bultman, et al., 1992; Cordes, et al., 1994; Herrmann, et al., 1990; Kingsley, et al., 1992; Miller, et al., 1993; Nehls, et al., 1994; Patl, et al., 1995; Segre, et al., 1995; Vidal, et al., 1993; Zhang, et al., 1994). Association of

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 germanitive 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 *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

Ang, S. L. and Rossant, J. (1994). HNF-3 beta is essential for node and notochord formation in mouse development.*Cell.*,78, 561-74.

Ang, S. L., Wierda, A., Wong, D., Stevens, K. A., Cascio, S., Rossant, J. and Zaret, K. S. (1993). The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins.*Development.*,119, 1301-15.

Avraham, K. B., Fletcher, C., Overdier, D. G., Clevidence, D. E., Lai, E., Costa, R. H., Jenkins, N. A. and Copeland, N. G. (1995). Murine chromosomal location of eight members of the hepatocyte nuclear factor 3/fork head winged helix family of transcription factors.*Genomics.*,25, 388-93.

Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff [comment].*Cell.*,74, 773-6.

Bultman, S. J., Michaud, E. J. and Woychik, R. P. (1992). Molecular characterization of the mouse agouti locus.*Cell.*,71, 1195-204.

Campfield, L. A., Smith, F. J., Guisez, Y., Devos, R. and Burn, P. (1995). Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks [see comments].*Science.*,269, 546-9.

Clark, K. L., Halay, E. D., Lai, E. and Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5.*Nature.*,364, 412-20.

Clevidence, D. E., Overdier, D. G., Peterson, R. S., Porcella, A., Ye, H., Paulson, K. E. and Costa, R. H. (1994). Members of the HNF-3/forkhead family of transcription factors exhibit distinct cellular expression patterns in lung and regulate the surfactant protein B promoter.*Dev Biol.*,166, 195-209.

Cordes, S. P. and Barsh, G. S. (1994). The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor.*Cell.*,79, 1025-34.

Coulombe, P. A. (1993). The cellular and molecular biology of keratins: beginning a new era.*Curr Opin Cell Biol.*,5, 17-29.

Dirksen, M. L. and Jamrich, M. (1992). A novel, activin-inducible, blastopore lip-specific gene of Xenopus laevis contains a fork head DNA-binding domain.*Genes Dev.*,6, 599-608.

Flanagan, S. (1966). "Nude", a new hairless gene with pleiotropic effects in the mouse.*Genetic Research.*,8, 295-309.

Galili, N., Davis, R. J., Fredericks, W. J., Mukhopadhyay, S., Rauscher, F. 3., Emanuel, B. S., Rovera, G. and Barr, F. G. (1993). Fusion of a fork head domain gene to PAX3 in the solid tumour alveolar rhabdomyosarcoma [published erratum appears in Nat Genet 1994 Feb; 6(2):214].*Nat Genet.*,5, 230-5.

Green, M. C. (1989) Catalog of mutant genes and polymorphic loci. In *Genetic Variants and Strains of the Laboratory Mouse, 2 ed.* (ed. M. F. Lyon and A. G. Searle). pp. 12-403. New York: Oxford University Press.

Halaas, J. L., Gajiwala, K. S., Maffei, M., Cohen, S. L., Chait, B. T., Rabinowitz, D., Lallone, R. L., Burley, S. K. and Friedman, J. M. (1995). Weight-reducing effects of the plasma protein encoded by the obese gene [see comments].*Science.*,269, 543-6.

Hammer, R. E., Krumlauf, R., Camper, S. A., Brinster, R. L. and Tilghman, S. M. (1987). Diversity of alpha-fetoprotein gene expression in mice is generated by a combination of separate enhancer elements.*Science.*,235, 53-8.

Hammer, R. E., Palmiter, R. D. and Brinster, R. L. (1984). Partial correction of murine hereditary growth disorder by germ-line incorporation of a new gene.*Nature.*,311, 65-7.

Hebert, J. M., Rosenquist, T., Gotz, J. and Martin, G. R. (1994). FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations.*Cell.*,78, 1017-25.

Heid, H. W., Moll, I. and Franke, W. W. (1988). Patterns of expression of trichocytic and epithelial cytokeratins in mammalian tissues. II. Concomitant and mutually exclusive synthesis of trichocytic and epithelial cytokeratins in diverse human and bovine tissues (hair follicle, nail bed and matrix, ligual papill, thymic reticulum).*Differentiation.*,37, 215-230.

Heid, H. W., Werner, E. and Franke, W. W. (1986). The complement of native alpha-keratin polypeptides of hair-forming cells: a subset of eight polypeptides that differ from epithelial cytokeratins.*Differentiation.*,32, 101-19.

Hermann-Le Denmat, S., Werner, M., Sentenac, A. and Thuriaux, P. (1994). Suppression of yeast RNA polymerase III mutations by FHL1, a gene coding for a fork head protein involved in rRNA processing.*Mol Cell Biol.*,14, 2905-13.

Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse.*Nature.*,343, 617-22.

Hope, I. A. (1994). PES-1 is expressed during early embryogenesis in Caenorhabditis elegans and has homology to the fork head family of transcription factors.*Development.*,120, 505-14.

Hubank, M. and Schatz, D. G. (1994). Identifying differences in mRNA expression by representational difference analysis of cDNA.*Nucleic Acids Res.*,22, 5640-8.

Jones, B. and Jesson, M. I. (1995). On the nature of the mutation in the *nude* rat.*Trends in Genetics.*,11, 257-258.

Jones, S. N., Grompe, M., Munir, M. I., Veres, G., Craigen, W. J. and Caskey, C. T. (1990). Ectopic correction of ornithine transcarbamylase deficiency in sparse fur mice.*J Biol Chem.*,265, 14684-90.

Jurgens, G. and Weigel, D. (1988). Terminal versus segmental development in the Drosophila embryo: the role of the homeotic gene fork head.*Roux's Arch. Dev. Biol.*,197, 345-354.

Kaestner, K. H., Lee, K. H., Schlondorff, J., Hiemisch, H., Monaghan, A. P. and Schutz, G. (1993). Six members of the mouse forkhead gene family are developmentally regulated.*Proc Natl Acad Sci U S A.*,90, 7628-31.

Kaytes, P. S., McNab, A. R., Rea, T. J., Groppi, V., Kawabe, T. T., Buhl, A. E., Bertolino, A. P., Hatzenbuhler, N. T. and Vogeli, G. (1991). Hair-specific keratins: characterization and expression of a mouse type I keratin gene.*J Invest Dermatol.*,97, 835-42.

Kingsley, D. M., Bland, A. E., Grubber, J. M., Marker, P. C., Russell, L. B., Copeland, N. G. and Jenkins, N. A. (1992). The mouse short ear skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGF beta superfamily.*Cell.*,71, 399-410.

Kingsley, D. M., Rinchik, E. M., Russell, L. B., Ottiger, H. P., Sutcliffe, J. G., Copeland, N. G. and Jenkins, N. A. (1990). Genetic ablation of a mouse gene expressed specifically in brain.*Embo J.*,9, 395-9.

Knochel, S., Lef, J., Clement, J., Klocke, B., Hille, S., Koster, M. and Knochel, W. (1992). Activin A induced expression of a fork head related gene in posterior chordamesoderm (notochord) of Xenopus laevis embryos.*Mech Dev.*,38, 157-65.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry [see comments].*Nature.*,351, 117-21.

Krumlauf, R., Hammer, R. E., Tilghman, S. M. and Brinster, R. L. (1985). Developmental regulation of alpha-fetoprotein genes in transgenic mice.*Mol Cell Biol.*,5, 1639-48.

Lacy, E., Roberts, S., Evans, E. P., Burtenshaw, M. D. and Costantini, F. D. (1983). A foreign beta-globin gene in transgenic mice: integration at abnormal chromosomal positions and expression in inappropriate tissues.*Cell.*,34, 343-58.

Lai, E., Clark, K. L., Burley, S. K. and Darnell, J., Jr. (1993). Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: a family of transcription factors of diverse biologic function.*Proc Natl Acad Sci U S A.*,90, 10421-3.

Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, R. H. and Darnell, J., Jr. (1990). HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally.*Genes Dev.*,4, 1427-36.

Lai, E., Prezioso, V. R., Tao, W. F., Chen, W. S. and Darnell, J., Jr. (1991). Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the Drosophila homeotic gene fork head.*Genes Dev.*,5, 416-27.

Li, J. and Vogt, P. K. (1993). The retroviral oncogene qin belongs to the transcription factor family that includes the homeotic gene fork head.*Proc Natl Acad Sci U S A.*,90, 4490-4.

Luetteke, N. C., Phillips, H. K., Qiu, T. H., Copeland, N. G., Earp, H. S., Jenkins, N. A. and Lee, D. C. (1994). The mouse waved-2 phenotype results from a point mutation in the EGF receptor tyrosine kinase.*Genes Dev.*,8, 399-413.

Luetteke, N. C., Qiu, T. H., Peiffer, R. L., Oliver, P., Smithies, O. and Lee, D. C. (1993). TGF alpha deficiency results in hair follicle and eye abnormalities in targeted and waved-1 mice.*Cell.*,73, 263-78.

Mann, G. B., Fowler, K. J., Gabriel, A., Nice, E. C., Williams, R. L. and Dunn, A. R. (1993). Mice with a null mutation of the TGF alpha gene have abnormal

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

Miller, L. M., Gallegos, M. E., Morisseau, B. A. and Kim, S. K. (1993). lin-31, a Caenorhabditis elegans HNF-3/fork head transcription factor homolog, specifies three alternative cell fates in vulval development.*Genes Dev.*,7, 933-47.

Miller, M. W., Duhl, D. M., Vrieling, H., Cordes, S. P., Ollmann, M. M., Winkes, B. M. and Barsh, G. S. (1993). Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation.*Genes Dev.*,7, 454-67.

Monaghan, A. P., Kaestner, K. H., Grau, E. and Schutz, G. (1993). Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm.*Development.*,119, 567-78.

Murphy, D. B., Wiese, S., Burfeind, P., Schmundt, D., Mattei, M. G., Schulz-Schaeffer, W. and Thies, U. (1994). Human brain factor 1, a new member of the fork head gene family.*Genomics.*,21, 551-7.

Nehls, M., Pfeifer, D., Schorpp, M., Hedrich, H. and Boehm, T. (1994). New member of the winged-helix protein family disrupted in mouse and rat nude mutations.*Nature.*,372, 103-7.

Nusslein-Volhard, C., Kluding, H. and Jurgens, G. (1985). Genes affecting the segmental subdivision of the Drosophila embryo.*Cold Spring Harb Symp Quant Biol.*,50, 145-54.

Patl, N., Cox, D. R., Bhat, D., Faham, M., Myers, R. M. and Peterson, A. S. (1995). A potassium channel mutation in weaver mice implicates membrane excitability in granule cell differentiation.*Nat Genet.*,11, 126-129.

Pelleymounter, M. A., Cullen, M. J., Baker, M. B., Hecht, R., Winters, D., Boone, T. and Collins, F. (1995). Effects of the obese gene product on body weight regulation in ob/ob mice [see comments].*Science.*,269, 540-3.

Pierrou, S., Hellqvist, M., Samuelsson, L., Enerback, S. and Carlsson, P. (1994). Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending.*Embo J.*,13, 5002-12.

Qian, X. and Costa, R. H. (1995). Analysis of hepatocyte nuclear factor-3 beta protein domains required for transcriptional activation and nuclear targeting.*Nucleic Acids Res.*,23, 1184-91.

Readhead, C., Popko, B., Takahashi, N., Shine, H. D., Saavedra, R. A., Sidman, R. L. and Hood, L. (1987). Expression of a myelin basic protein gene in transgenic shiverer mice: correction of the dysmyelinating phenotype.*Cell.*,48, 703-12.

Rinchik, E. M., Bangham, J. W., Hunsicker, P. R., Cacheiro, N. L., Kwon, B. S., Jackson, I. J. and Russell, L. B. (1990). Genetic and molecular analysis of chlorambucil-induced germ-line mutations in the mouse.*Proc Natl Acad Sci U S A.*,87, 1416-20.

Rinchik, E. M., Russell, L. B., Copeland, N. G. and Jenkins, N. A. (1986). Molecular genetic analysis of the dilute-short ear (d-se) region of the mouse.*Genetics.*,112, 321-42.

Ruiz i Altaba, A. and Jessell, T. M. (1992). Pintallavis, a gene expressed in the organizer and midline cells of frog embryos: involvement in the development of the neural axis.*Development.*,116, 81-93.

Russell, L. B., Hunsicker, P. R., Cacheiro, N. L., Bangham, J. W., Russell, W. L. and Shelby, M. D. (1989). Chlorambucil effectively induces deletion mutations in mouse germ cells.*Proc Natl Acad Sci U S A.*,86, 3704-8.

Russell, W. L. (1951). X-ray-induced mutations in mice.*Cold Spring Harbor Symp. Quant. Biol.*,16, 327-336.

Russell, W. L., Kelly, E. M., Hunsicker, P. R., Bangham, J. W., Maddux, S. C. and Phipps, E. L. (1979). Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse.*Proc. Natl. Acad. Sci.*,76, 5818-5819.

Sasaki, H. and Hogan, B. L. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo.*Development.*,118, 47-59.

Segre, J. A., Nemhauser, J. L., Taylor, B. A., Nadeau, J. H. and Lander, E. S. (1995). Positional Cloning of the nude Locus: Genetic, Physical and Transcription Maps of the Region and Mutations in the Mouse and Rat.*Genomics.*,28, 549-559.

Strahle, U., Blader, P., Henrique, D. and Ingham, P. W. (1993). Axial, a zebrafish gene expressed along the developing body axis, shows altered expression in cyclops mutant embryos.*Genes Dev.*,7, 1436-46.

Vidal, S. M., Malo, D., Vogan, K., Skamene, E. and Gros, P. (1993). Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg.*Cell.*,73, 469-85.

Vrieling, H., Duhl, D. M., Millar, S. E., Miller, K. A. and Barsh, G. S. (1994). Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene.*Proc Natl Acad Sci U S A.*,91, 5667-71.

Weigel, D. and Jackle, H. (1990). The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? [letter].*Cell.*,63, 455-6.

Weigel, D., Jurgens, G., Kuttner, F., Seifert, E. and Jackle, H. (1989). The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo.*Cell.*,57, 645-58.

Weinstein, D. C., Ruiz i Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M. and Darnell, J., Jr. (1994). The winged-helix transcription factor

HNF-3 beta is required for notochord development in the mouse embryo.*Cell.*,78, 575-88.

Winter, H., Siry, P., Tobiasch, E. and Schweizer, J. (1994). Sequence and expression of murine type I hair keratins mHa2 and mHa3.*Exp Cell Res.*,212, 190-200.

Woychik, R. P., Generoso, W. M., Russell, L. B., Cain, K. T., Cacheiro, N. L., Bultman, S. J., Selby, P. B., Dickinson, M. E., Hogan, B. L. and Rutledge, J. C. (1990). Molecular and genetic characterization of a radiation-induced structural rearrangement in mouse chromosome 2 causing mutations at the limb deformity and agouti loci.*Proc Natl Acad Sci U S A.*,87, 2588-92.

Wu, J., Zhou, T., Zhang, J., He, J., Gause, W. C. and Mountz, J. D. (1994). Correction of accelerated autoimmune disease by early replacement of the mutated lpr gene with the normal Fas apoptosis gene in the T cells of transgenic MRL-lpr/lpr mice.*Proc Natl Acad Sci U S A.*,91, 2344-8.

Xuan, S., Baptista, C. A., Balas, G., Tao, W., Soares, V. C. and Lai, E. (1995). Winged helix transcription factor BF-1 is essential for the development of the cerebral hemispheres.*Neuron.*,14, 1141-52.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L. and Friedman, J. M. (1994). Positional cloning of the mouse obese gene and its human homologue.*Nature.*,372, 425-32.

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 nu^{str} 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 *rnu* 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 nu^{str} /+ animals. Informative progeny (i.e., those determined, based on genotype at flanking markers, to carry the recombinant chromosome over a nu^{str}-bearing chromosome) were examined for phenotype to determine whether the recombinant chromosome carried the nu^{str} 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 *Bgl*II or *Bam*HI. 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 *Bam*HI-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'-CGACGTTGTAAAACGACGGCCAGT-3') and Rev25 primer (5'-CACACAGGAAACAGCTATGACCATG-3'), 67 mM Tris-HCl (pH 8.8 at 25°C), 4 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$, 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 restiction-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 picolitter.

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, RsaI 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 RsaI; 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 HpaI or MseI. 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 bair 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'-CTGAGCGGAATTCGTGAGACC-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, transfered 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, transfered 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-*nu/nu*, AKR/J, AKR/J- nu^{str}/nu^{str},rat, and rat- *rnu/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).

Production of Transgenic Mice

Microinjection of oocytes was performed as described by Hogan et al (1986). Homozygous CD1-nu/nu males were mated with homozygous CD1-nu/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.

RNA in situ Hybridization Analysis and Histological Evaluation

RNA *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 ctt gta atc tcc caa taa -3'; mHa2 5'-gga gct taa caa gca ggt ggc -3'; mHa3 aca agc cca ttg gac cct gtg -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.

DNA Analysisof 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 ^{32}P 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'-TCGCATACGGTATAAGAAGATG -3'; T7 side of cosmid vector: 5'-TGATAAGCGGTCAAACATGA -3' (bp 90-71); T7 side of insert DNA: 5'-ATCTCCTCCCCTAACCTGGG -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'-TTCTCACCTGCTCCTAGGGA- 3' from the T7 side of the insert of 1193; 5'-CGCCGACCTGCTTCAC - 3' from exon 1 of *Hfh11*; 5'-CACTTCCAGGCTCCACCC- 3' from exon 2 of *Hfh11*; 5'-ACTGTTCTTCTCAGCCTGCC- 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

Chu, G., Vollrath, D. and Davis, R. W. (1986). Separation of large DNA molecules by countour-clamped homogeneous electric fields.*Science.*,234, 1582-1585.

Church, D. M., Stotler, C. J., Rutter, J. L., Murrell, J. R., Trofatter, J. A. and Buckler, A. J. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification.*Nat Genet.*,6, 98-105.

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H. S., Friedman, J., Dracopoli, N. C. and Lander, E. S. (1992). A genetic map of the mouse suitable for typing intraspecific crosses.*Genetics.*,131, 423-47.

Eicher, E. (1976). Remutations.*Mouse News Letter.*,54, 90.

Feinberg, A. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity.*Anal. Biochem.*,132, 6-12.

Green, E. D. and Olson, M. V. (1990). Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction.*Proc Natl Acad Sci U S A.*,87, 1213-7.

Hebert, J. M., Rosenquist, T., Gotz, J. and Martin, G. R. (1994). FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations.*Cell.*,78, 1017-25.

Hogan, B., Constantini, F. and Lacy, E. (1986).*Manipulating the Mouse Embryo:a Laboratory Manual.*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hunter, K., Housman, D. and Hopkins, N. (1991). Isolation and characterization of irradiation fusion hybrids from mouse chromosome 1 for mapping Rmc-1, a gene encoding a cellular receptor for MCF class murine retroviruses.*Somat Cell Mol Genet.*,17, 169-83.

Kaytes, P. S., McNab, A. R., Rea, T. J., Groppi, V., Kawabe, T. T., Buhl, A. E., Bertolino, A. P., Hatzenbuhler, N. T. and Vogeli, G. (1991). Hair-specific keratins: characterization and expression of a mouse type I keratin gene.*J Invest Dermatol.*,97, 835-42.

Killary, A. M. and Fournier, R. E. (1984). A genetic analysis of extinction: trans-dominant loci regulate expression of liver-specific traits in hepatoma hybrid cells.*Cell.*,38, 523-34.

Kusumi, K., Smith, J. S., Segre, J. A., Koos, D. S. and Lander, E. S. (1993). Construction of a large-insert yeast artificial chromosome library of the mouse genome.*Mamm Genome.*,4, 391-2.

Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M. A., Jaenisch, R. and Berns, A. (1991). Simplified mammalian DNA isolation procedure.*Nucleic Acids Res.*,19, 4293.

Lisitsyn, N., Lisitsyn, N. and Wigler, M. (1993). Cloning the differences between two complex genomes.*Science.*,259, 946-51.

Lovett, M. (1994) Direct Selection of cDNAs Using Genomic Contigs. In *Current Protocols in Human Genetics.* (ed. N. C. Dracopoli, J. L. Haines, B. R. Korf, D. T. Moir, C. C. Morton, C. E. Seidman, J. G. Seidman and D. R. Smith). pp. New York: Current Protocols.

Rossi, J. M., Chen, H. and Tilghman, S. M. (1994). Genetic map of the fused locus on mouse chromosome 17.*Genomics.*,23, 178-84.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989).*Molecular Cloning: A Laboratory Manual.*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Winter, H., Siry, P., Tobiasch, E. and Schweizer, J. (1994). Sequence and expression of murine type I hair keratins mHa2 and mHa3.*Exp Cell Res.*,212, 190-200.

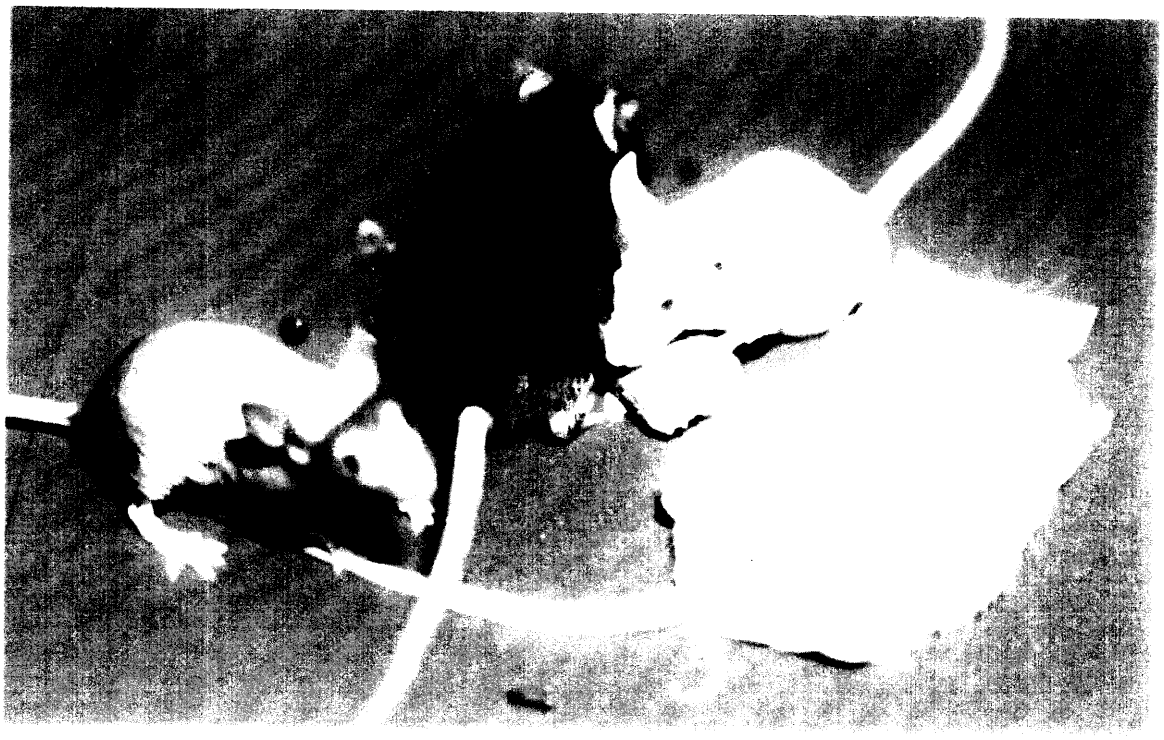


Figure 1. nude mice with wild-type littermates.
From the left: C57BL/6J-nu, C57BL/6J,
AKR-nustr, AKR/J

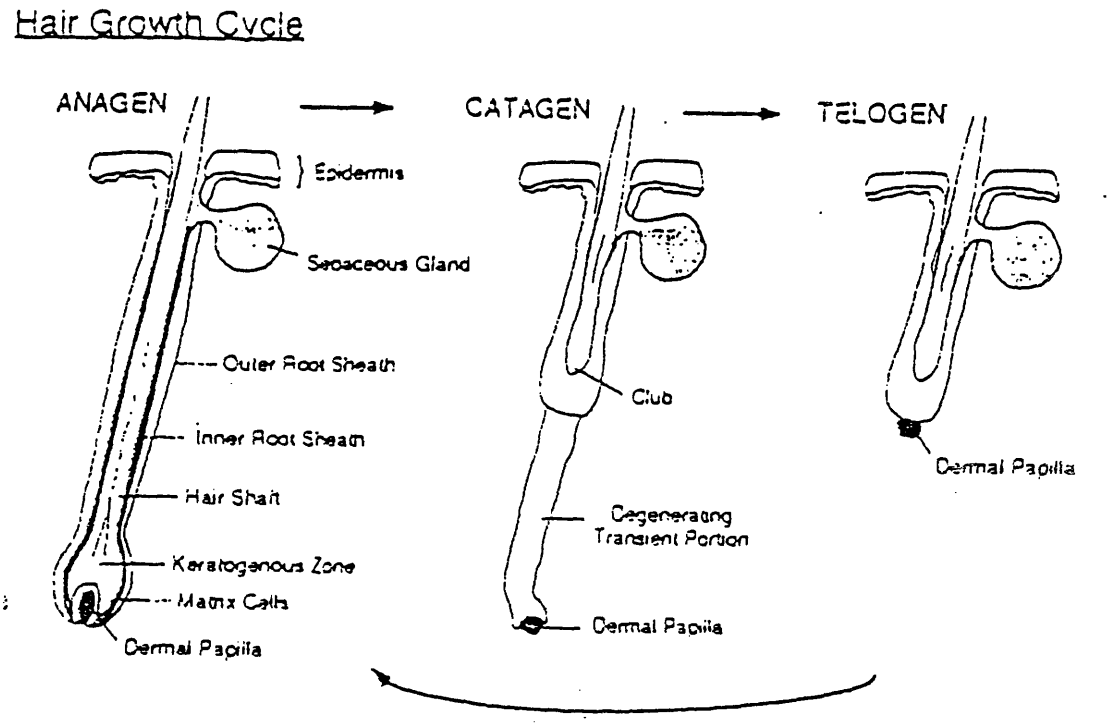


Figure 2. A schematic diagram illustrating the morphology of the hair follicle at the three stages of hair growth cycle: anagen: the stage during which the follicle is regenerated and the hair matrix cells produce a new hair; catagen, the stage during which matrix cell proliferation and hair elongation ceases and many of the structures of the anagen follicle are eliminated; and telogen, the resting phase. Diagram from Herbert et al., 1994.

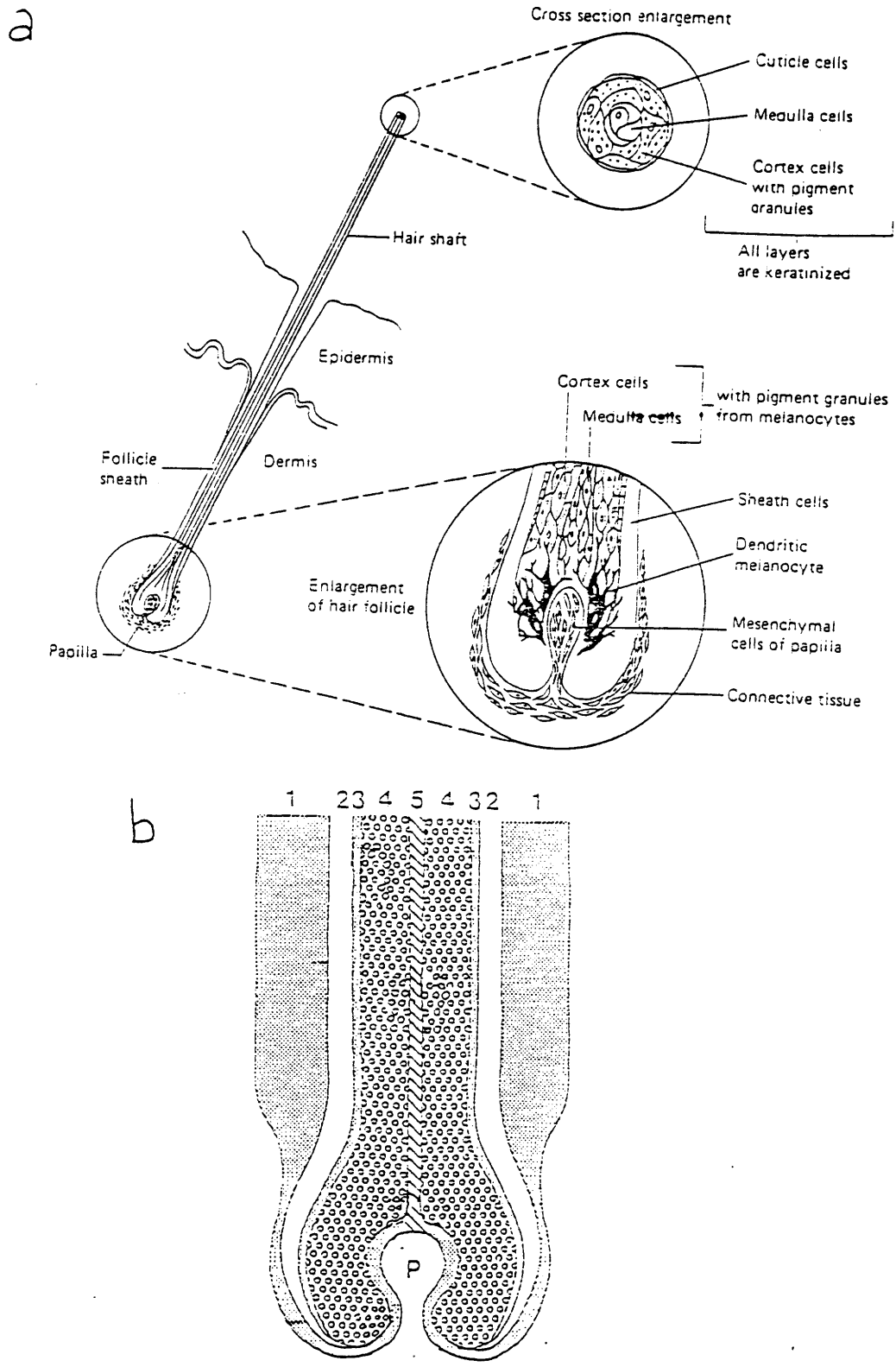


Figure 3. Schematic representation of a hair follicle. (a) The relative arrangement of cells derived from the epidermis (sheath, cortex, medulla), the dermis (papilla and connective tissue) and the neural crest (melanocyte). Diagram from Hogan et al., 1986. (b) Major tissue layers and cell types of the 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 sheat 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

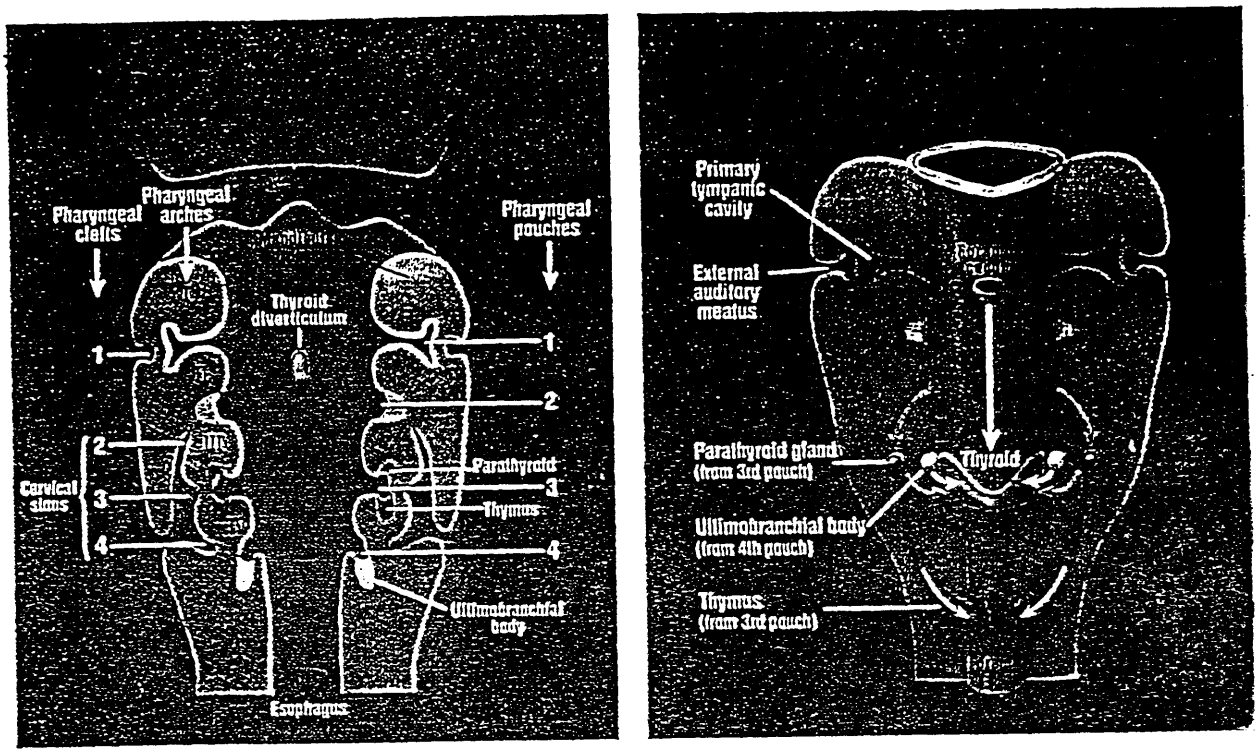


Figure 5. Thymic development from 3rd Pharyngeal arch and pouch. Left side is a representation of a coronal section at embryonic 11 d.p.c. The thymus is beginning to form from the posterior half of the third pharyngeal pouch. The right side is a ventral view of the pharyngeal organ derivatives. The two thymic primordia migrate medially and posteriorly, where the two anlages fuse to become a single gland. Diagram from Manley and Capecchi, 1995.

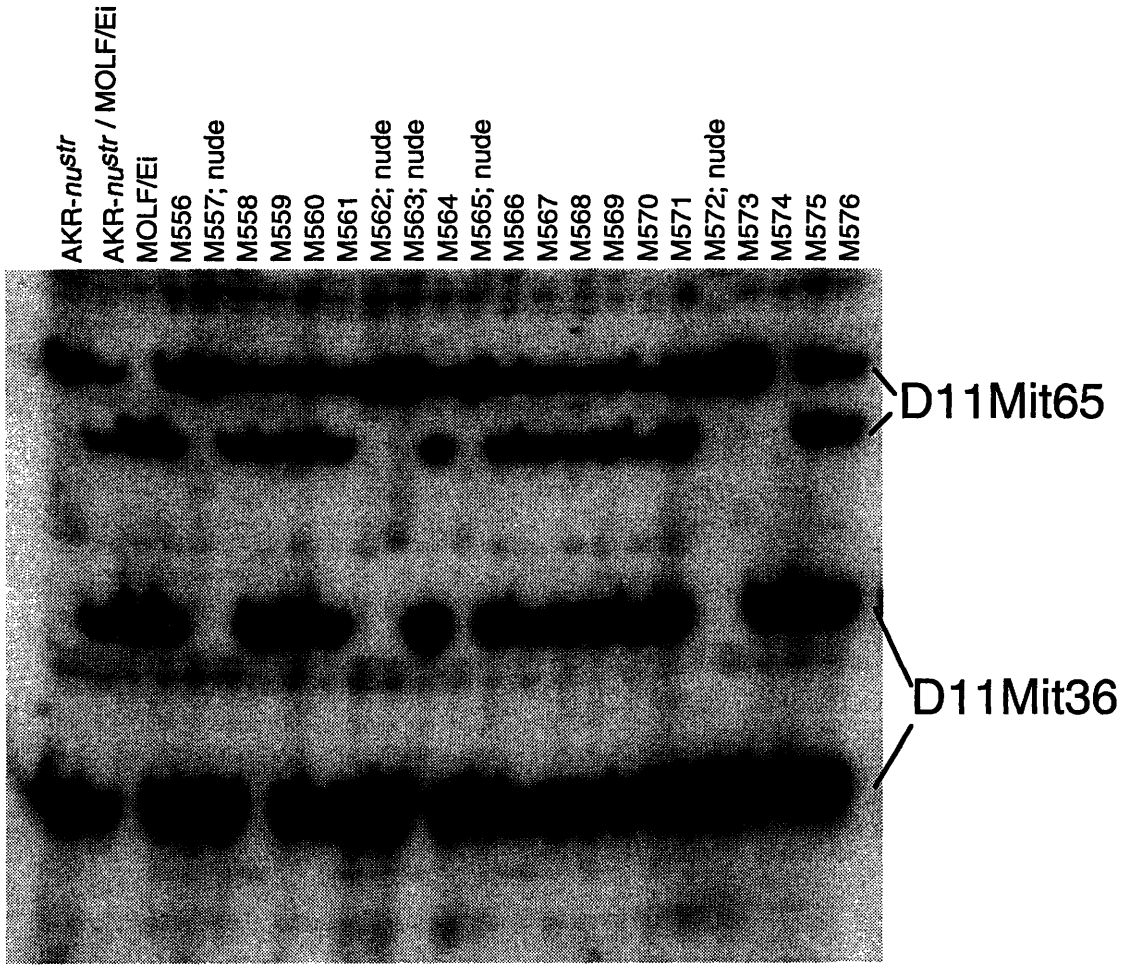


Figure 6. The genotype of (AKR-*nu*^{str} x MOLF/Ei)F_2 progeny M566 to M586 with simple sequence length polymorphism markers D11Mit65 and D11Mit36. The AKR-*nu*^{str} and MOLF/Ei sizes are 382 and370 for D11Mit65 and 234 and274 for D11Mit36, respectively. The *nude* animals are indicated. Note that there are cross-overs between D11Mit65 and D11Mit36 in animals M558 and M564.

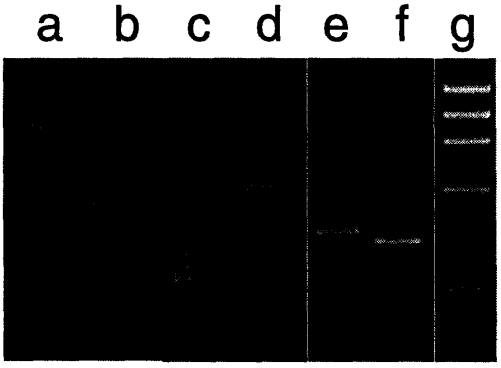


Figure7. Agarose gel electrophoresis of difference-products obtained after the first (lanes a,b), second (lanes c,d) and third (lanes e,f) hybridization-extension-amplification steps as pairs for the nude congenic and the nude cross (first iteration). Lane g is the size standard of ϕX174 RF DNA digested with *HaeIII*.

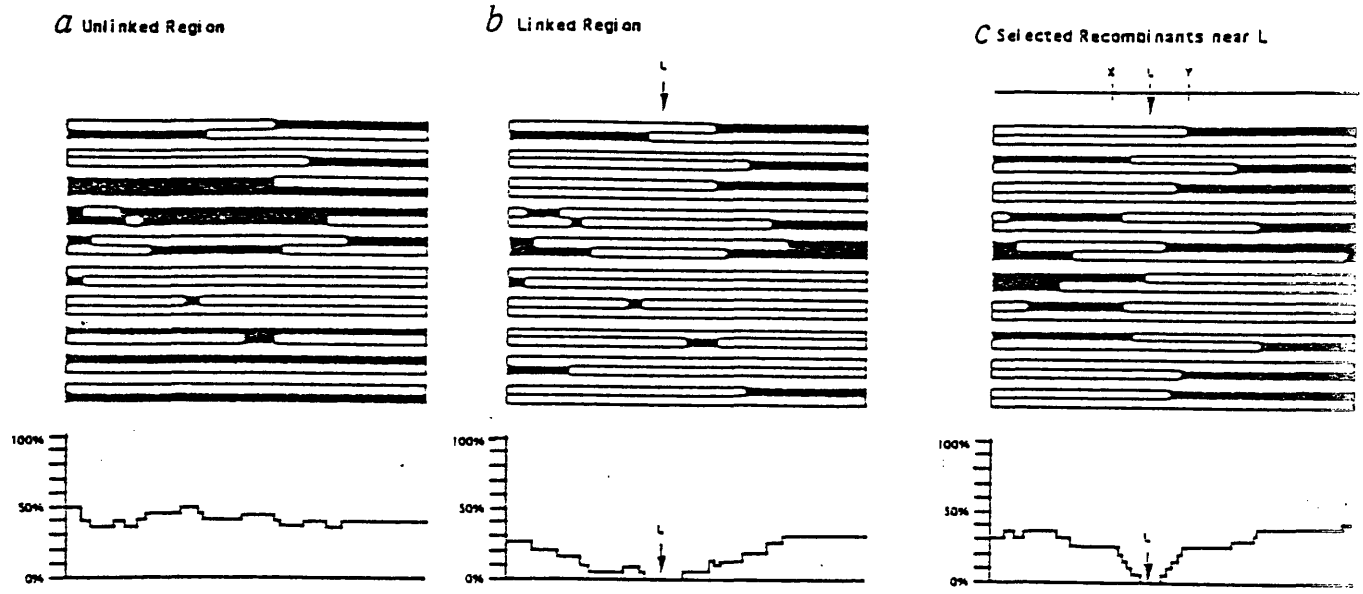


Fig. 8 Schematic diagram representing the principle underlying GDRDA with progeny from an F_2 intercross. Each panel shows hypothetical chromosomal genotypes from 10 progeny to be pooled to create a Driver; each chromosome is arbitrarily drawn to be 100 cM. Strain A carries a recessively-acting allele 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).

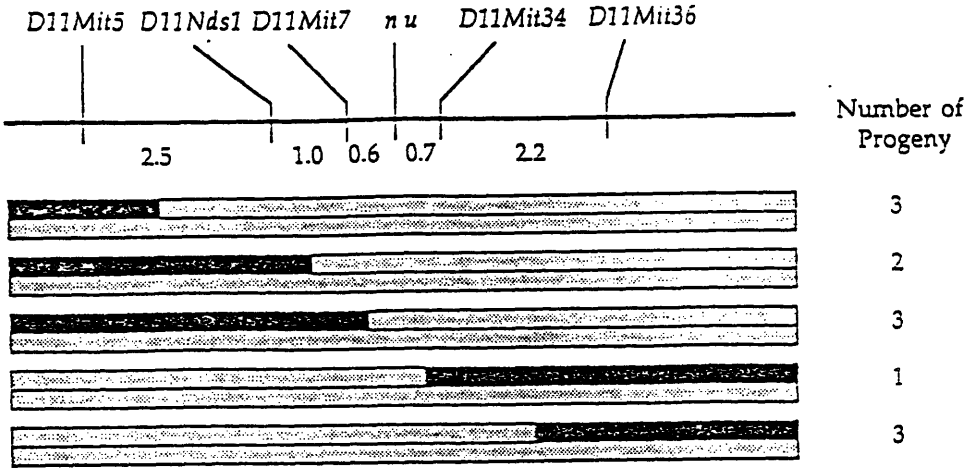


Figure 9. Schematic diagram indicating chromosomal genotypes of the 12 (MOLF/Ei x AKR/J-*nu*^{str}) F_2 progeny pooled to create the Driver for GD-RDA, relative to a genetic map of polymorphic markers near the *nude* locus. Black indicates regions derived from MOLF/Ei. Shading indicates regions derived from AKR. The number of progeny of each type is indicated at the right.

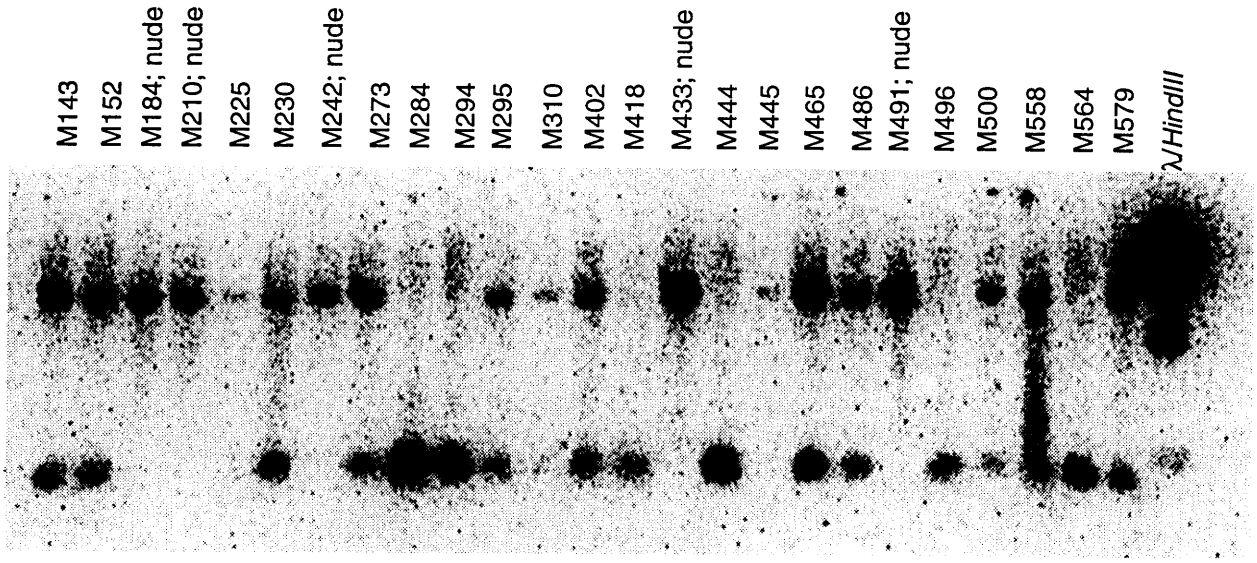


Figure 10a. Southern blot of DNAfrom F2 intercross progeny, digested with *BglII* and hybridized with GD-RDA 6.2. The animals are listed above the lane. All the nude animals, indicated above, are homozygous for the AKR-nu^{str} allele of 4 kb. The size standard is λ digested with *HindIII*.

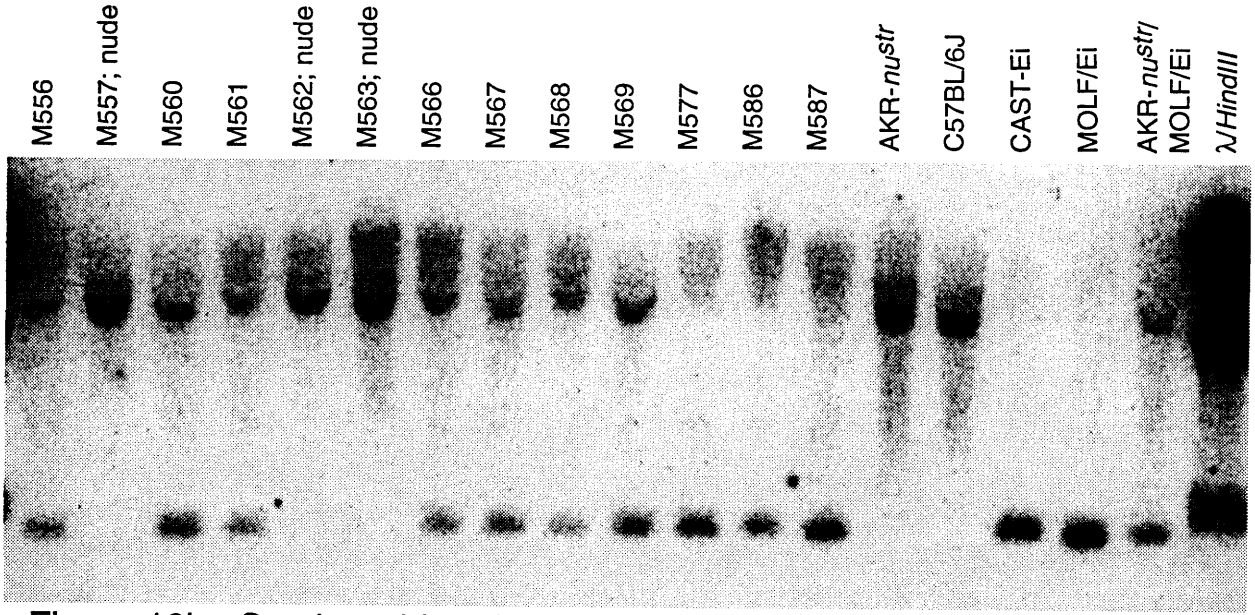


Figure 10b. Southern blot of DNAfrom F2 intercross progeny, digested with *BamHI* and hybridized with GD-RDA 10.4. The animals are listed above the lane. All the nude animals, indicated above, are homozygous for the AKR-nu^{str} allele of 3 kb. The size standard is λ digested with *HindIII*.

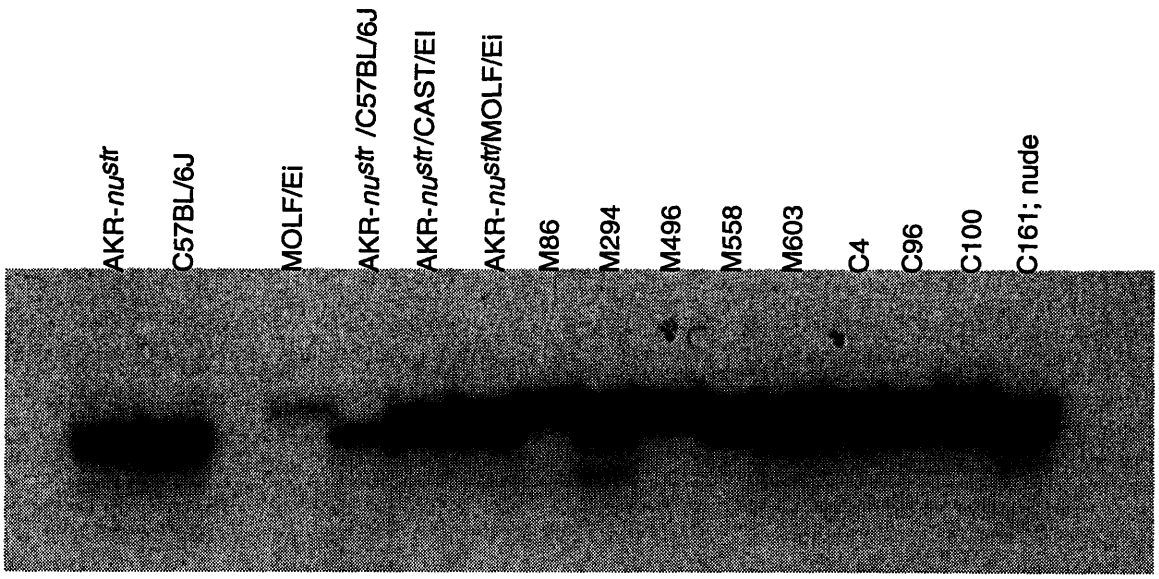


Figure 11a. The genotype of nude intercross progeny with the single strand conformational polymorphism BS11.8. The nude animal is indicated.

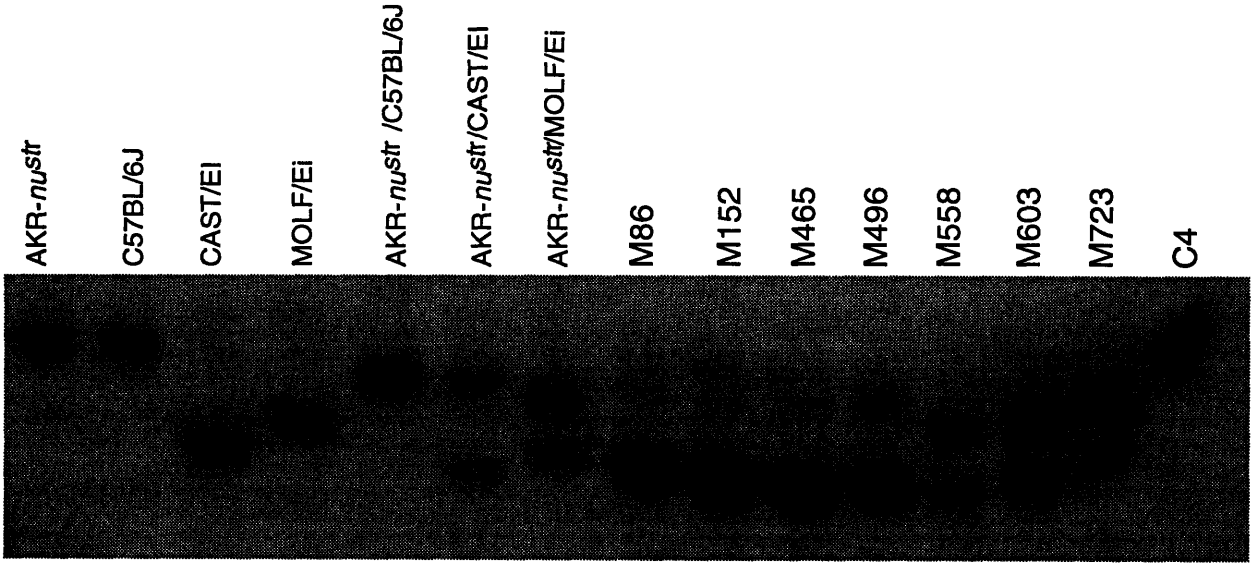


Figure 11b. The genotype of nude intercross progeny with the single strand conformational polymorphism Y32L. All of the animals are unaffected.

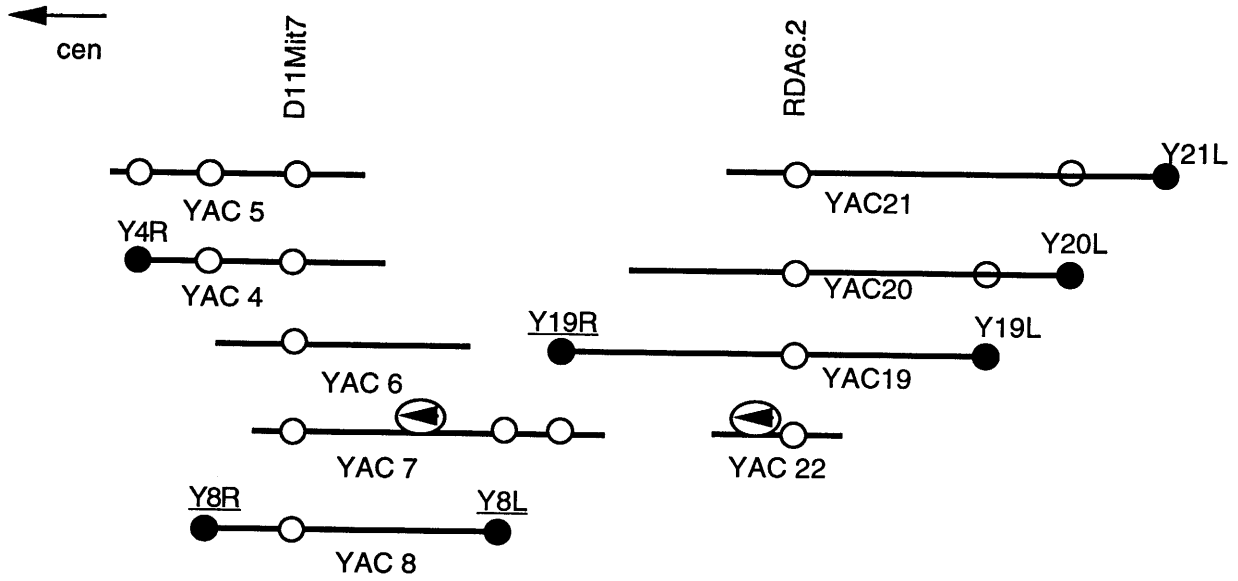


Figure 12. YAC chromosomal walk proximal to nude. Contig is centered around D11Mit7 and RDA6.2. The direction to the centromere is indicated. The orientation of the contig is suggested by the genetic markers. Underlined names indicate terminal fragments that contain a sequence polymorphism, assayed as 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.

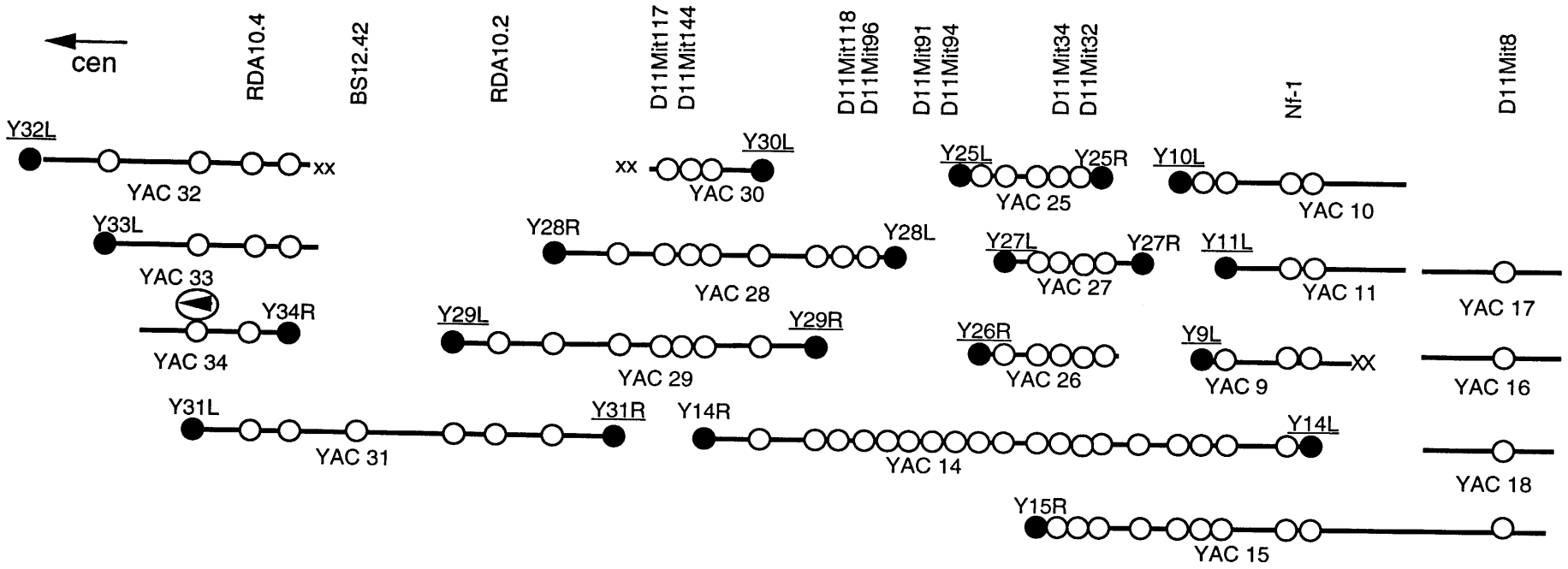


Figure 13. YAC chromosomal walk surrounding nude. Contig is centered around genetic markers: both MIT SSLPs and GD-RDA clones. The orientation of the contig is suggested by the genetic markers. The direction to the centromere is indicated. An underlined name indicates a terminal fragment that contains 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.

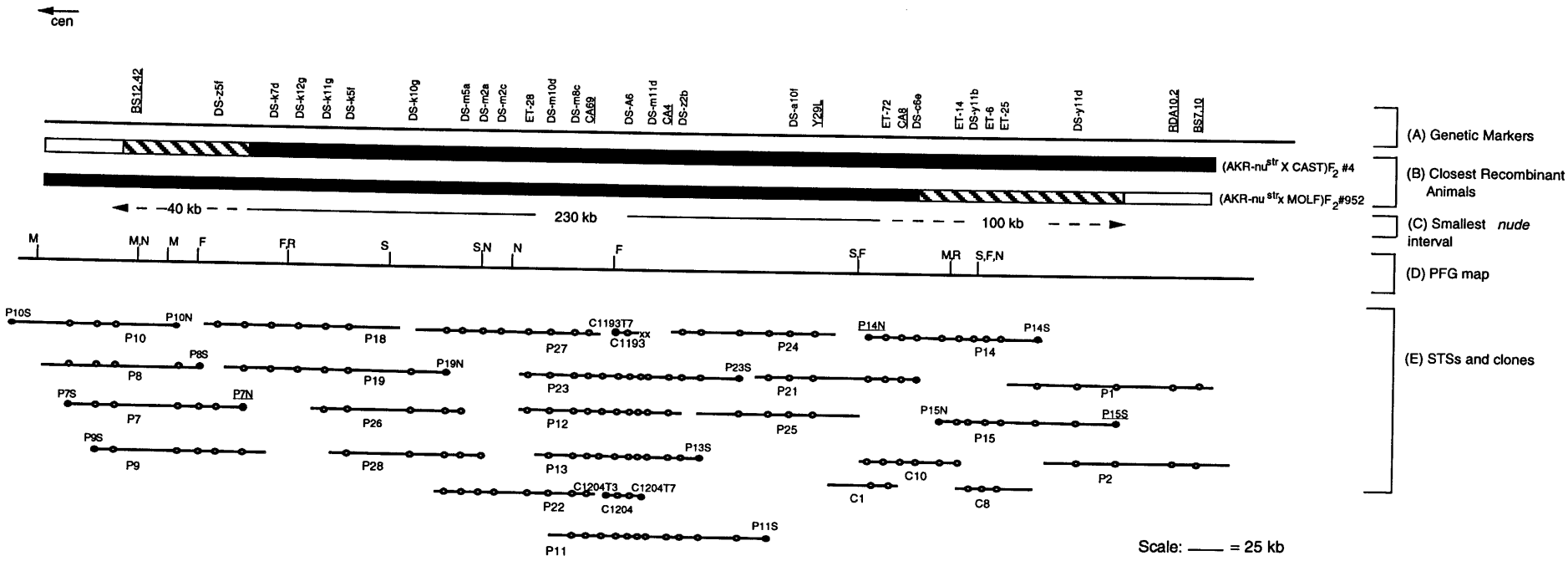


Figure 14. Fine structure physical mapping of the nude region. The direction to the centromere is indicated. The orientation of the contig is suggested by the genetic markers. (A) Genetic markers and anchoring loci. An underlined name indicates a locus that contain a sequence polymorphism, assayed as 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 haplotye. (C) The smallest region that does not recombine with the *nude* locus. (D) Pulsed-field gel-based restriction map of the region. S, *Sac*II; F, *Sfi* I; M, *Mlu*I; N, *Not*I; R, *Rsr*II. (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 GGGGCATTGAGGGCCCCATCGATGCTGCCTTCACTCGCATCAACTGTCAGGGGNAGACCT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Vtn:     596 GGGGCATTGAGGGCCCCATCGATGCTGCCTTCACTCGCATCAACTGTCAGGGGAAGACCT 655

cY8A :    61 ACTTGTTCAAGGGTAGTCAGTACTGGCGCTTTGAGGATGGGGTCCTGGACCCTGGTTATC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Vtn:     656 ACTTGTTCAAGGGTAGTCAGTACTGGCGCTTTGAGGATGGGGTCCTGGACCCTGGTTATC 715

cY8A :   121 CCCGAAACATCTCCGAAGGCTTCAGTGGCATACCAGACAATGTTGATGCAGCGTTCG 177
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Vtn:     716 CCCGAAACATCTCCGAAGGCTTCAGTGGCATACCAGACAATGTTGATGCAGCGTTCG 772

B.  DS-K12G vs. emb-5

cK12G:     4 EAVLEGARYMVALQIARVPLVRQVLRQTFQERAKLNITPTKKGRKDVDEAHYAYSFKYLK 183
             + VL GA++M+A +I+R P VR  +RQ F++ A   I PTKKGR  +D+ H  Y  +Y+K
emb-5:   552 DMVLNGAKFMLAKEISRQPQVRHSVRQEFRQSAHFWIKPTKKGRDTIDQTHPLYDKRYIK 611

cK12G:   184 NKPVKELRDDQFLKIGLAEDEGLLTI 261
             +KPV+ L  ++FL    A+++GL+ +
emb-5:   612 SKPVRSLTAEEFLFYHKAKEDGLVDV 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.

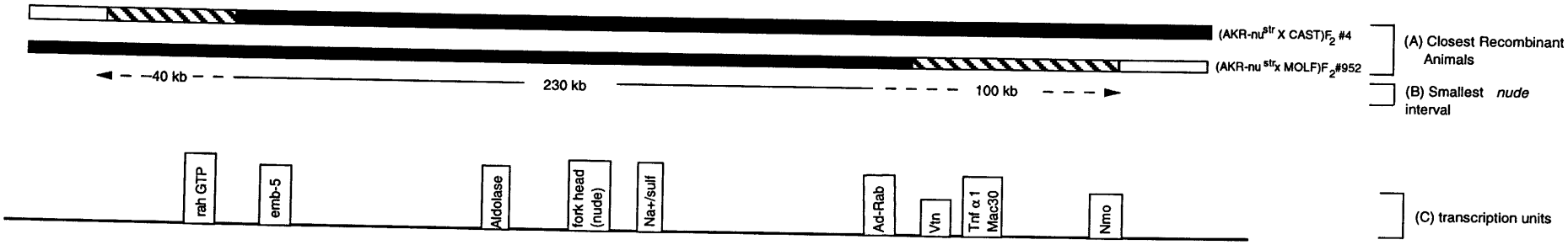


Figure 16. Transcription map of the nude region. (A) Recombinational breakpoints for the closest recombinant animals in *nude* crosses. The empty bar denotes the chromosomal region that recombines with the *nu* gene. The black bar denotes the chromosomal region that does not recombine with the *nu* gene. The hatched bar denotes the chromosomal region with an undetermined haplotye. (B) The smallest region that does not recombine with the *nude* locus. (C) The location of transcription units with strong sequence similarity to genes in Genbank.

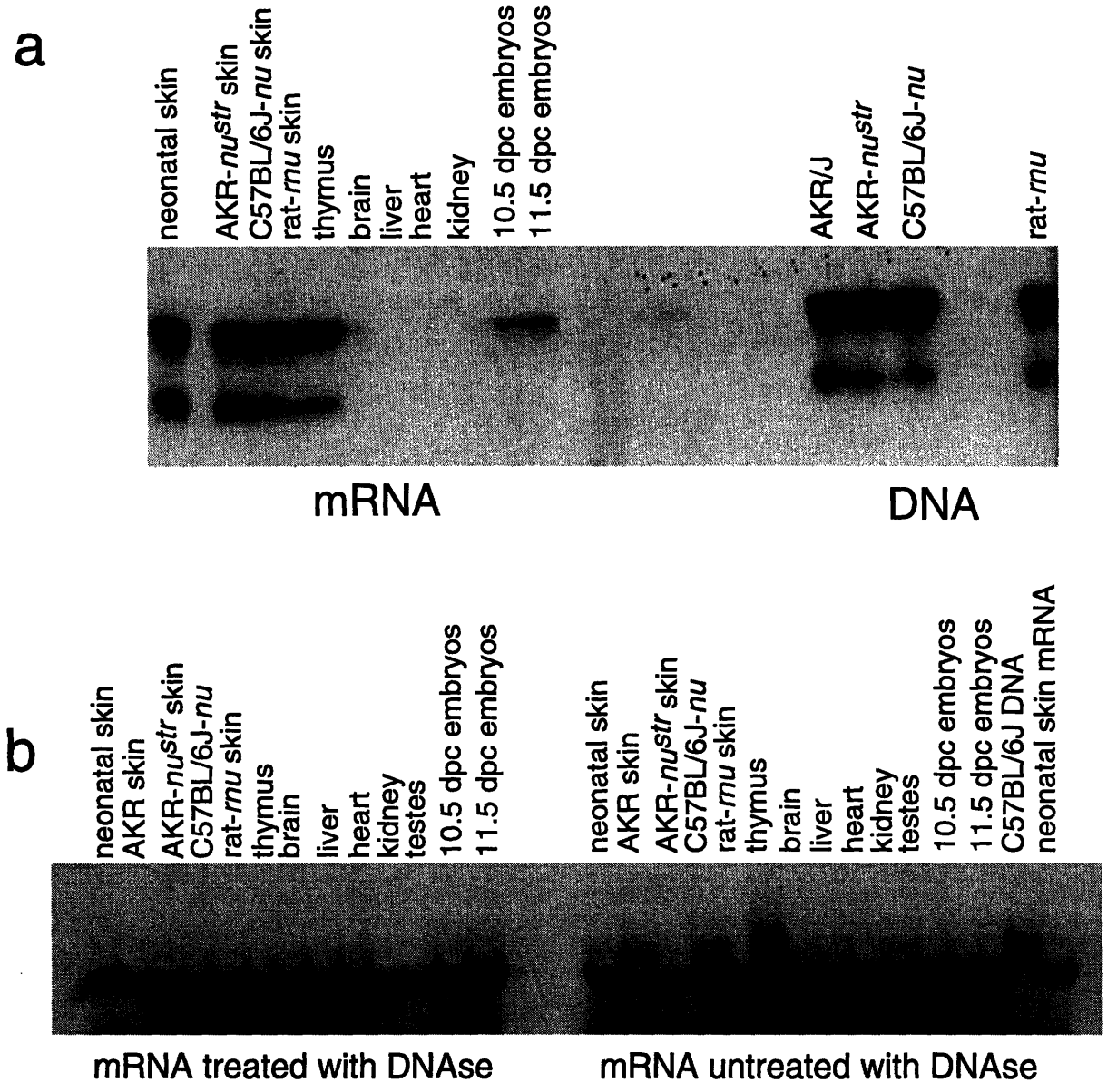


Figure 17. Expression studies by Reverse-transcribed PCR. Screening for genomic deletions in mutant mice by PCR . a. ET-90 expression. b. Profilin expression with primers that flank an intron on samples treated with DNAse and untreated.

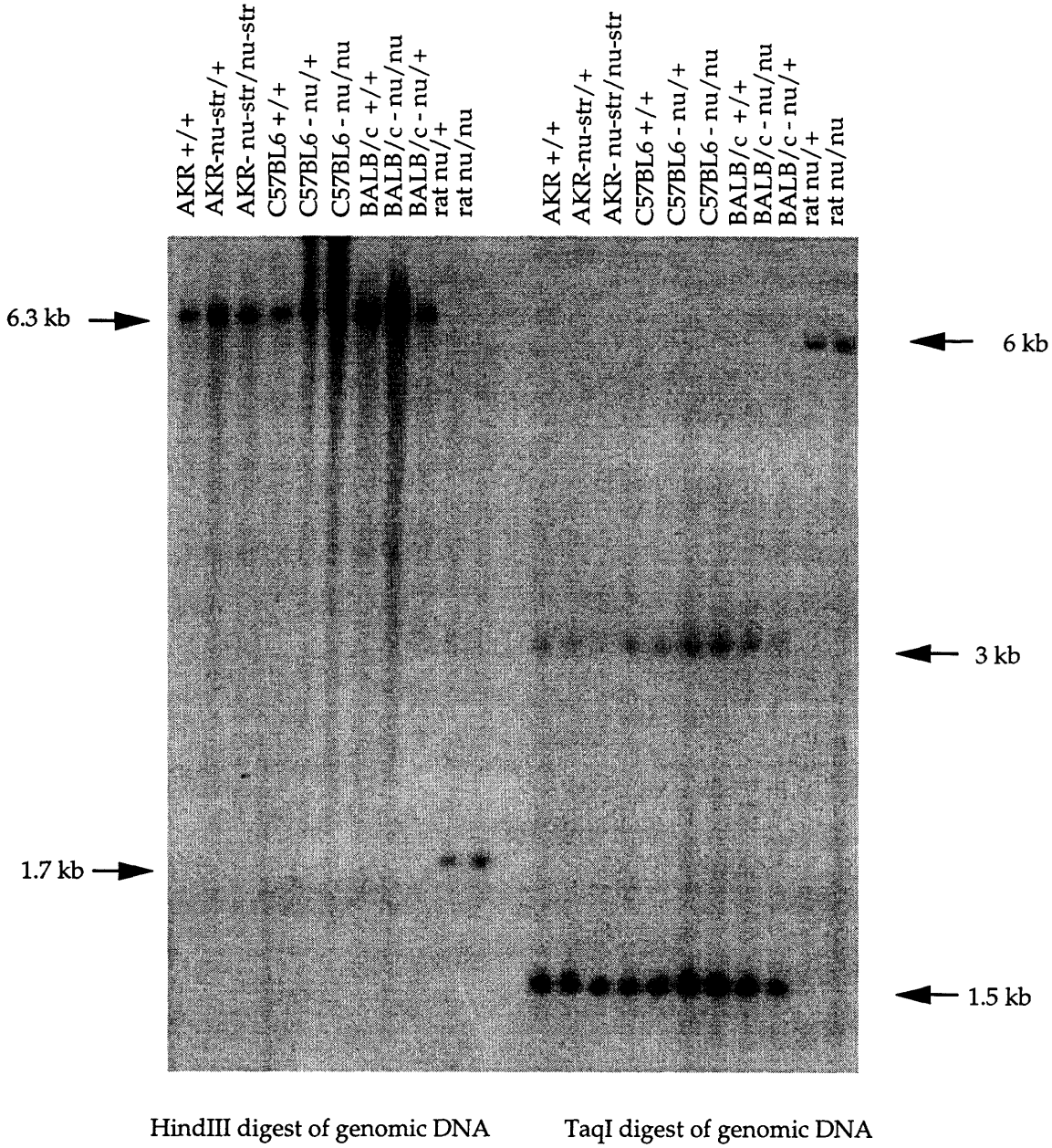


Figure 18. Southern blot of restriction digested genomic DNA hybridized with DS-a10g.

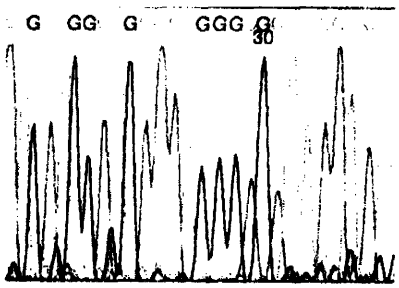


Figure 19a. mouse wild-type nude sequence (-21M13 primer)

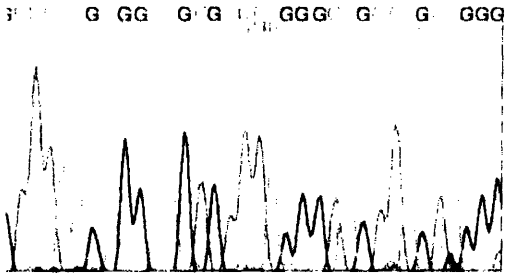


Figure 19b. mouse wild-type nude sequence (M13Rev primer)

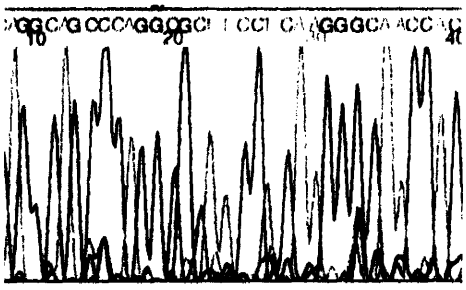


Figure 19c.C57BL/6J-nu sequence (-21M13 primer). Deletion of single G at bp19.

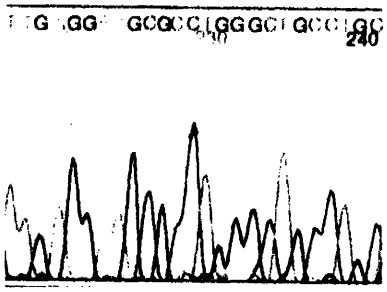


Figure 19d.C57BL/6J-nu sequence (M13Rev primer). Deletion of single C at bp 228.

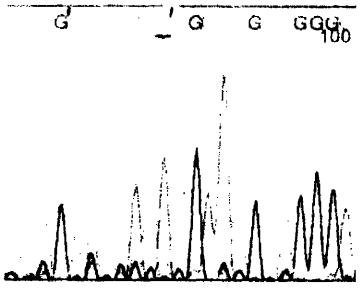


Figure 19e. rat wild-type nude sequence (-21M13 primer)

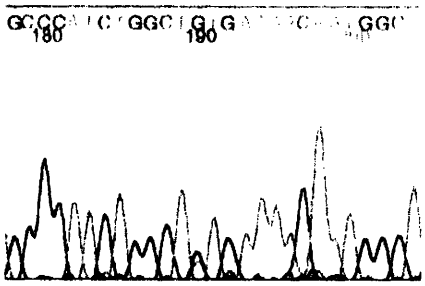


Figure 19f. rat rnu/+ sequence (M13Rev primer)

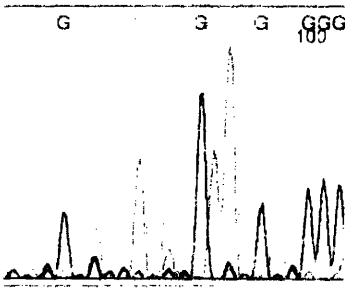


Figure 19g. rat rnu sequence (-21M13 primer). C to T change at bp 91.

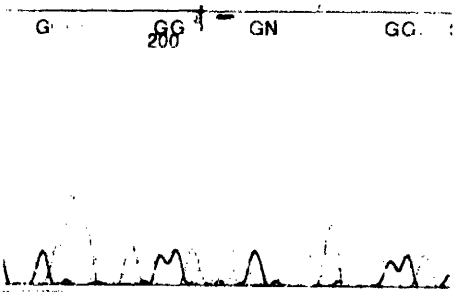


Figure 19h. rat rnu sequence (M13Rev primer). G to A change at bp 204.

Figure 19. Sequencing wild-type and mutant alleles to detect mutations.

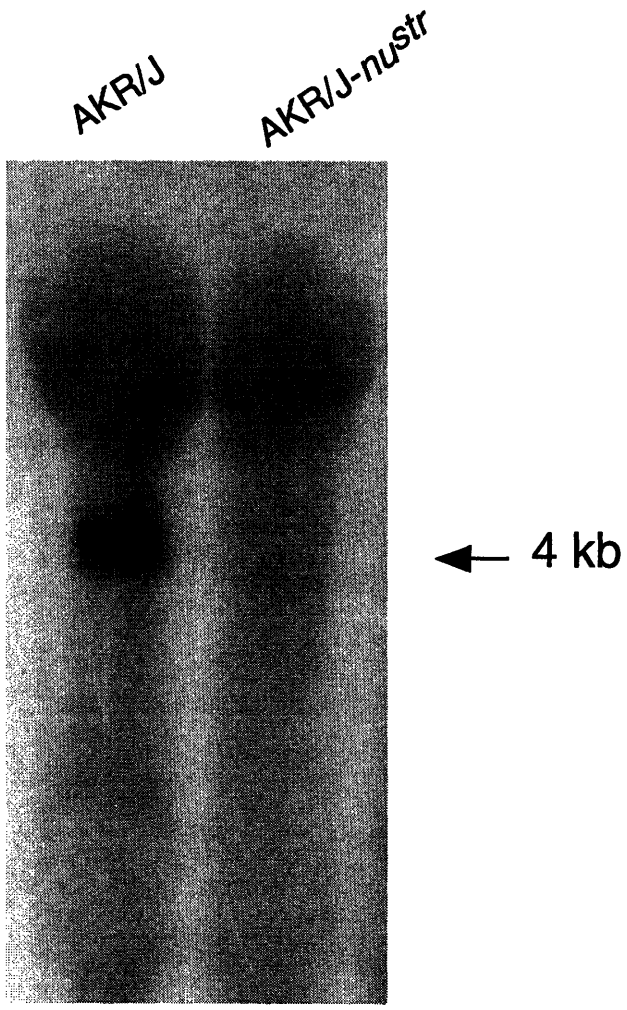


Figure 20. Northern blot analysis of nude gene expression in AKR/J and AKR/J-*nu*^{str} skin.

```
          M  V  S  L  L  P  P  H  S  D  V  T  L  P  G  S  T
GACTGGGTGATGGTGTCGCTACTCCCTCCGCACTCTGACGTCACACTTCCAGGCTCCACC      60
 R  L  E  G  E  P  Q  G  D  L  M  Q  A  P  G  L  P  G  S  P
CGACTGGAGGGCGAACCCCAAGGGGACCTCATGCAGGCTCCGGGCCTCCCAGGCTCCCCT     120
 A  P  Q  N  K  H  A  N  F  S  C  S  S  F  V  P  D  G  P  P
GCCCCACAGAACAAGCACGCCAACTTCAGCTGCTCATCATTTGTGCCTGATGGCCCTCCA     180
 E  R  A  P  S  L  P  P  H  S  P  S  I  A  S  P  G  P  E  Q
GAGAGGGCCCCCTCGCTGCCCCCCCACAGCCCAAGCATCGCATCTCCAGGCCCAGAGCAG     240
 I  Q  S  H  C  T  A  G  P  G  P  G  S  F  R  L  S  P  S  D
ATCCAGAGCCACTGCACAGCTGGACCTGGCCCAGGCTCCTTCCGCCTTTCTCCCTCAGAC     300
 K  Y  P  G  F  G  F  E  E  G  P  A  G  S  P  G  R  F  L  R
AAGTATCCTGGCTTTGGCTTTGAGGAGGGCCCAGCAGGCAGCCCAGGGCGCTTCCTCAGG     360
 G  N  H  M  P  F  H  P  Y  K  G  H  F  H  E  D  I  F  S  E
GGCAACCACATGCCTTTCCACCCTTACAAGGGGCACTTCCATGAAGACATCTTCTCTGAG     420
 A  Q  T  A  M  A  L  D  G  H  S  F  K  T  Q  G  A  L  E  A
GCCCAGACGGCCATGGCGCTTGATGGACACTCCTTTAAGACTCAGGGGGCACTGGAAGCC     480
 F  E  E  I  P  V  D  V  G  D  A  E  A  F  L  P  S  F  P  A
TTTGAGGAGATCCCTGTGGATGTGGGGGATGCTGAGGCCTTTCTGCCTAGCTTCCCAGCA     540
 E  A  W  C  N  G  L  P  Y  P  S  Q  E  H  N  Q  I  L  Q  G
GAGGCTTGGTGCAATGGACTCCCTTACCCCAGCCAGGAACACAACCAAATTCTGCAGGGG     600
 S  E  V  K  V  K  P  Q  A  L  D  N  G  P  G  M  Y  C  Y  Q
TCAGAGGTCAAGGTCAAGCCCCAAGCTTTGGACAATGGTCCTGGGATGTACTGCTACCAG     660
 P  P  L  Q  H  V  Y  C  S  S  Q  P  T  F  H  Q  Y  S  P  G
CCTCCCTTGCAGCATGTGTACTGTTCTTCTCAGCCCACCTTTCATCAGTACTCCCCGGGT     720
 G  G  S  Y  P  V  P  Y  L  G  S  T  H  Y  P  Y  Q  R  I  A
GGAGGCAGCTACCCTGTGCCCTACCTGGGCTCCACTCACTACCCCTATCAGAGGATTGCG     780
 P  Q  A  N  A  D  G  H  Q  P  L  F  P  K  P  I  Y  S  Y  S
CCCCAGGCCAACGCAGACGGTCACCAGCCACTCTTCCCAAAGCCCATCTACTCCTACAGC     840
 I  L  I  F  M  A  L  K  N  X  K  T  G  S  L  P  V  S  E  I
ATCCTCATCTTCATGGCCCTTAAGAACAGCAAGACCGGAAGCCTTCCAGTCAGTGAAATC     900
 Y  N  F  M  T  E  H  F  P  Y  F  K  T  A  P  D  G  W  K  N
TACAATTTCATGACGGAGCACTTCCCTTATTTCAAGACTGCGCCTGATGGCTGGAAGAAT     960
 S  V  R  H  N  L  S  L  N  K  C  F  E  K  V  E  N  K  S  G
TCTGTTCGCCATAACCTGTCTCTCAACAAGTGCTTTGAGAAGGTAGAGAATAAATCCGGA    1020
 S  S  S  R  K  G  C  L  W  A  L  N  P  S  K  I  D  K  M  Q
AGTTCCTCCCGAAAGGGCTGTCTGTGGGCCCTCAATCCTTCCAAAATCGACAAGATGCAG    1080
 E  E  L  Q  K  W  K  R  K  D  P  I  A  V  R  K  S  M  A  K
GAAGAGCTGCAGAAGTGGAAGAGGAAAGACCCCATTGCTGTGCGCAAAAGCATGGCCAAA    1140
 P  E  E  L  D  S  L  I  G  D  K  R  E  K  L  G  S  P  L  L
CCAGAAGAGCTGGACAGCCTCATCGGAGACAAAAGGGAGAAACTGGGTTCTCCTCTGCTA    1200
 G  C  P  P  P  G  L  A  G  P  G  P  I  R  P  L  A  P  S  A
GGATGTCCACCCCCTGGGCTGGCAGGCCCAGGTCCCATCCGGCCCCTGGCACCTTCAGCT    1260
 G  L  T  Q  P  L  H  P  M  H  P  A  P  G  P  M  P  G  K  N
GGTCTTACCCAGCCTCTACACCCAATGCATCCAGCTCCAGGTCCCATGCCTGGCAAGAAC    1320
 P  L  Q  D  L  L  G  G  H  A  P  S  C  Y  G  Q  T  Y  P  H
CCCCTGCAGGACCTACTGGGTGGCCATGCACCCTCCTGCTACGGGCAGACCTACCCACAC    1380
 L  S  P  S  L  A  P  S  G  H  Q  Q  P  L  F  S  Q  P  D  G
CTTTCTCCCAGCCTGGCCCCTTCTGGACACCAGCAGCCATTGTTTTCACAGCCAGATGGG    1440
 H  L  D  L  Q  A  Q  P  G  T  P  Q  D  S  P  L  P  A  H  T
CATCTTGATCTGCAGGCCCAGCCAGGCACCCCCCAGGACTCACCTCTACCTGCCCACACA    1500
 P  P  S  H  G  A  K  L  L  A  E  P  S  S  A  R  T  M  H  D
CCACCCAGCCACGGTGCCAAGCTGCTGGCTGAGCCTTCCTCAGCCAGGACCATGCACGAT    1560
 T  L  L  P  D  G  D  L  G  T  D  L  D  A  I  N  P  S  L  T
ACTCTGCTACCAGACGGAGACCTTGGCACTGACCTGGACGCCATCAACCCCTCTCTCACT    1620
 D  F  D  F  Q  G  N  L  W  E  Q  L  K  D  D  S  L  A  L  D
GACTTCGACTTCCAGGGAAATCTGTGGGAGCAGCTGAAGGATGACAGCTTGGCCCTGGAC    1680
 P  L  V  L  V  T  S  S  P  T  S  S  S  M  L  P  P  P  P  A
CCCCTCGTACTGGTGACCTCATCCCCAACATCATCCTCCATGTTGCCACCCCCACCAGCA    1740
```

```
 A  H  C  F  P  P  G  P  C  L  A  E  T  G  N  E  A  G  E  L
GCCCATTGCTTCCCCCCAGGGCCTTGTCTGGCAGAAACAGGCAATGAGGCAGGTGAACTG      1800
 A  P  P  G  S  G  G  S  G  A  L  G  D  M  H  L  S  T  L  Y
GCACCTCCAGGCAGCGGCGGCTCCGGTGCTCTGGGAGACATGCACCTCAGCACTCTCTAC      1860
 S  A  F  V  E  L  E  S  T  P  S  S  A  A  A  G  P  A  V  Y
TCCGCCTTTGTGGAACTGGAGTCCACGCCCTCCTCAGCAGCTGCCGGCCCTGCCGTGTAC      1920
 L  S  P  G  S  K  P  L  A  L  A.
CTCAGTCCCGGCTCAAAGCCATTGGCTCTGGCTTGAGCTATGCCCAACATCAGCTGCGGC      1980
CTTGCCTAGCTGGCTGCCCATATAGGGCTCACCTTAAAGGTCAAAGAAGGAAAACACTAC      2040
TTGTCTCCTATGTCACTCAGCCAACTTATTTGTTAGCCAGAAGCTAGGGGATCCACCTAG      2100
GATGCTACCGGGTGACGCGCTCCACACGCGGtGCCCCCAGCAAGGAAAGTGTCGGAAAAG      2160
AAGCAACAGCGCGCCCCCTAGCGCCAGCTCACTCGAACTTCAGCTCTCGAGCGTGAACTC      2220
AAGCTCTACACACCTGCTCATCATGTGCCTTACACTCAGAGGAAGCCTTGGGTCACAGAG      2280
TCTGATTTGATTTCTGGGGCAGCTGAGAGCTAAAAGCTTTAGTTAGCAAAGCTCAGGGCC      2340
AGTGTGACAGGTCAAAGATCACCCCTCCAAACCTCATCCCAATCCCCATGTGCTCTACAG      2400
ACTTGGCCCTCTCCC                                                   2415
```

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.

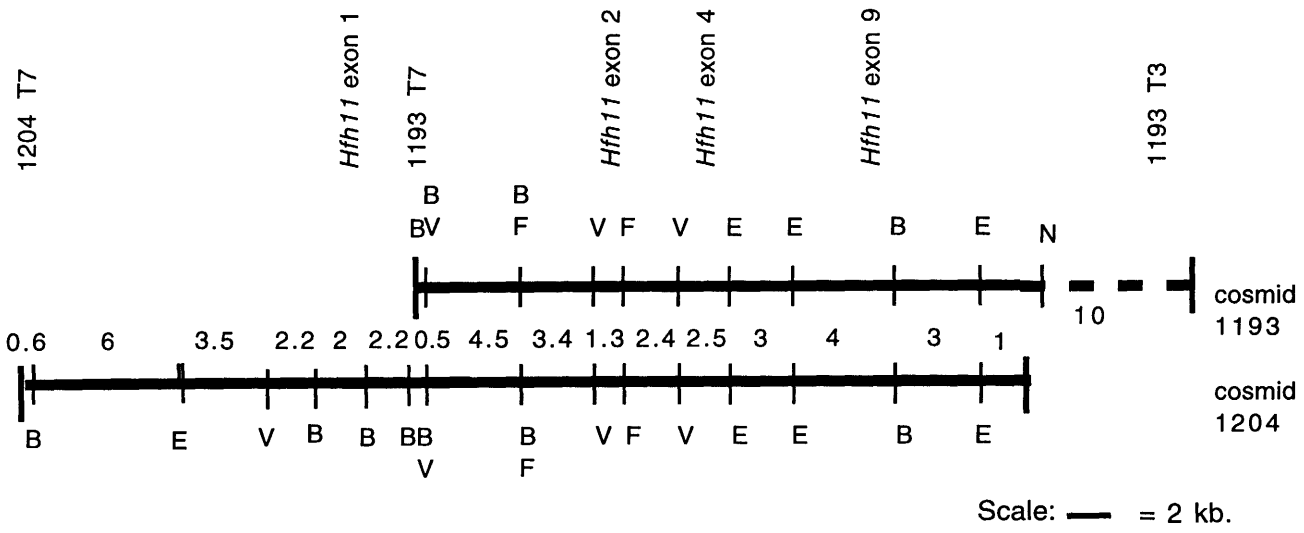


Figure 22. Restriction Map of cosmids 1193 and 1204. B, BamHI; E, EcoRI; F, SfiI; N, NotI; V, EcoRV. Total insert size of 1193 is 36 kb; 1204 is 42 kb. Cosmid 1193 has a chimeric insert: ___ is mouse DNA. _ _ is yeast DNA.

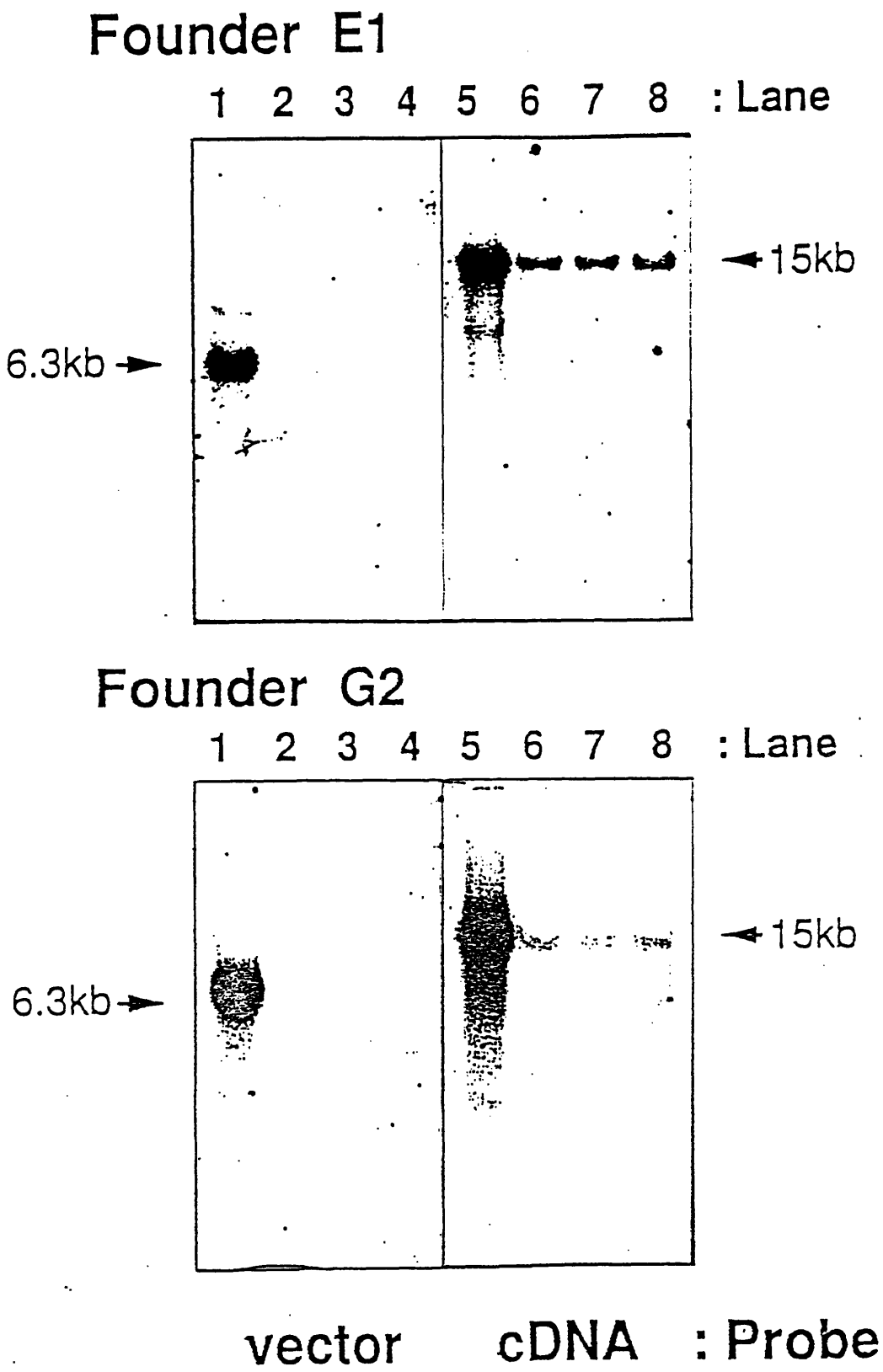


Figure 23. Southern blot analysis of DNA from *Hfh11* transgenic mice. Tail DNA samples from *Hfh11* transgenic mice (lane 1,5) and their non-transgenic littermates (lane 2-4,6-8) were digested with HindIII. Membranes were hybridized with a cosmid vector probe (lane 1-4) and a probe for the carboxy terminal of a *Hfh11* cDNA (lane 5-8). Expected transgene bands are indicated by arrows.

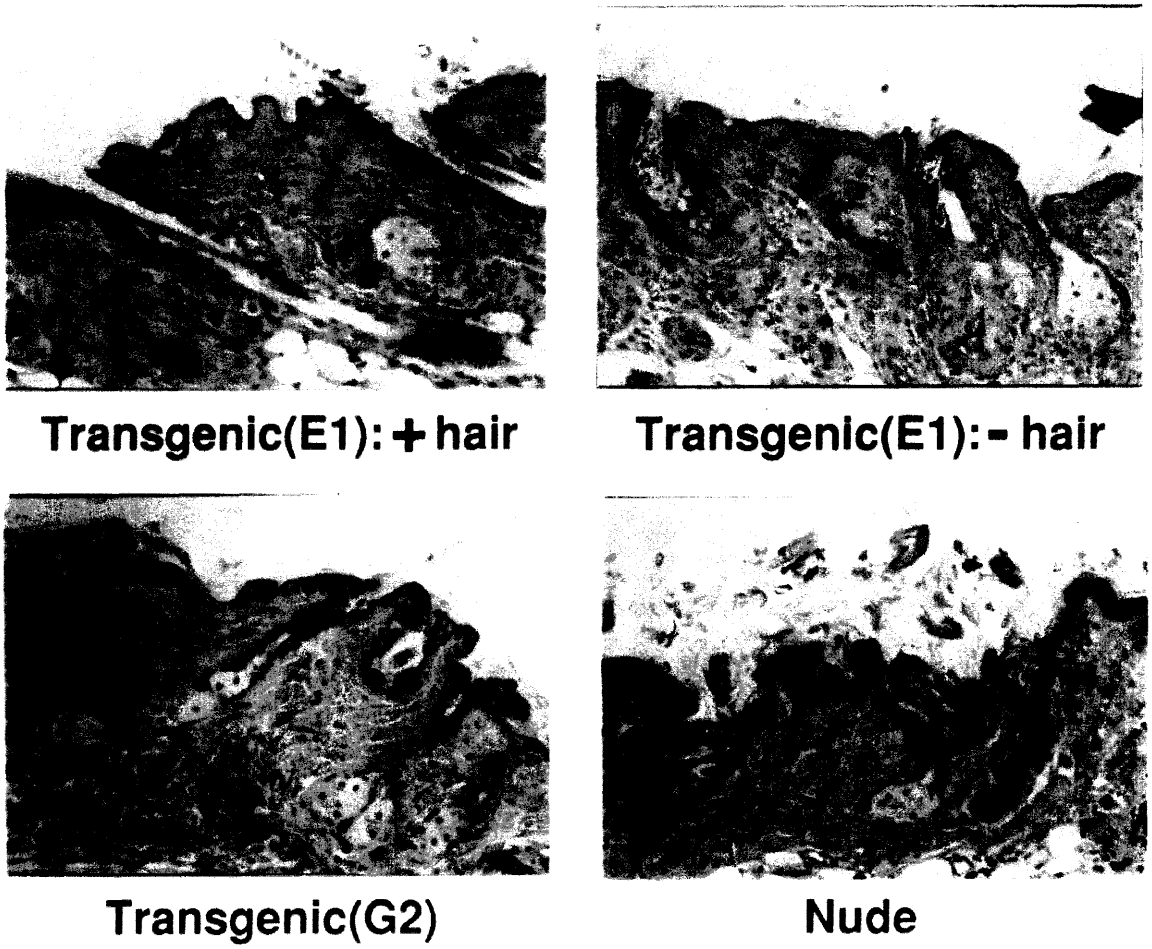
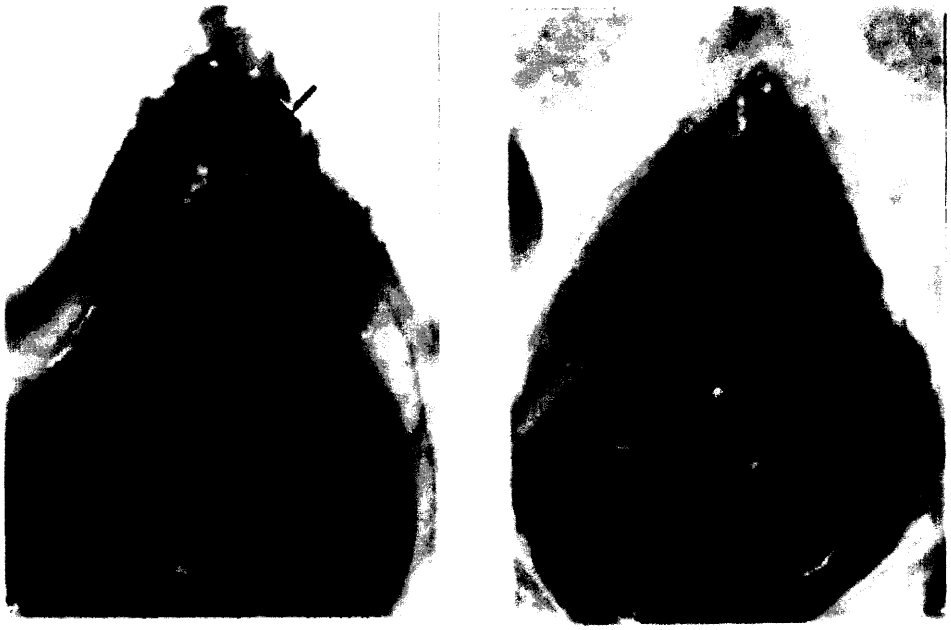


Figure 25. Histological sections of the skin of transgenic mice, E1 and G2. The tissue sections were stained with hematoxylin and eosin. For transgenic E1, both haired and non-haired regions are shown.



Normal Transgenic

Figure 26. Phenotype of control wild-type and Hfh11 transgenic mice. Thymus, indicated by the arrow in the the wild-type, is absent in transgenic mice.

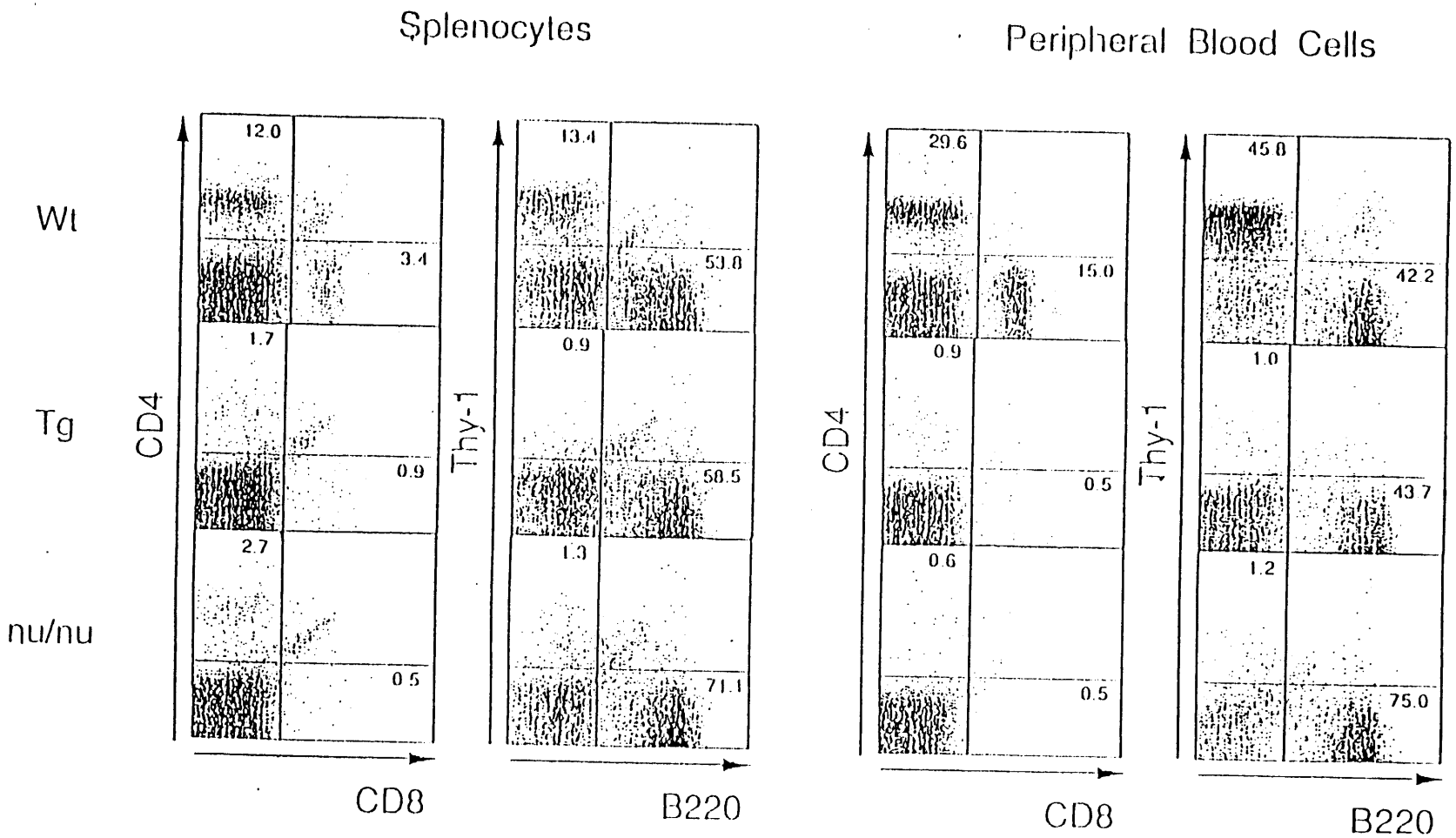


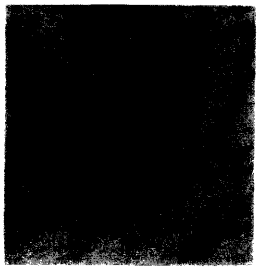
Figure 27. Analysis of the T cell populations in *Hfh11*transgenic mice. Two-color flow cytometry of splenocytes and peripheral blood cells from 1.5-month-old-control wild type (wt), *Hfh11* transgenic (E1), and control nude (nu/nu) mice. Single cell suspensions were stained with antibodies specific to CD4, CD8, Thy-1, B220 and analyzed with FACScan. The number in each quadrant indicates the calculated percentage of cells.

a

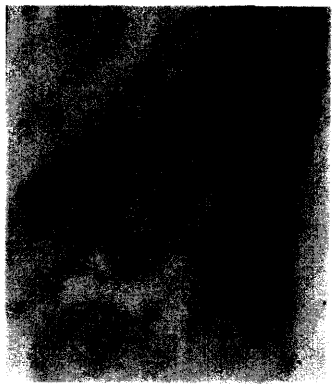
b

Figure 28. Hfh11 expression in the cortex of the hair follicle at post natal day 4. (a) x 100. (b) x200

a



b



c

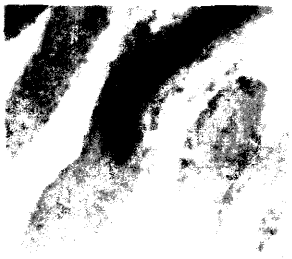


Figure 29. Expression of keratin genes in skin of *nude* mouse at post-natal day 4. Magnification is x200. (a) Ha1 (b) Ha2 (c) Ha3.

ANIMAL	D11Nds1	D11Mit7	Y8L	RDA6.2	RDA10.4	CA69	CA128	Phenotype	Genotype	CA4	CA8	RDA10.2	BS7.10	D11Mit117	D11Mit144	D11Mit118	D11Mit96	D11Mit91	D11Mit94	D11Mit34	D11Mit32	D11Mit8
B11	A	-	A	-	-	-	-	wt	H	-	-	-	-	-	H	H	-	-	-	H	-	H
B13	B	-	H	-	-	-	-	wt	H	-	-	-	-	-	H	H	-	-	-	H	-	H
B25	B	-	H	-	-	-	-	wt	H	-	-	-	-	-	H	H	-	-	-	H	-	H
B63	H	-	H	-	-	-	-	wt	?	-	-	-	-	-	B	B	-	-	-	B	-	B
B86	B	-	H	-	-	-	-	wt	H	-	-	-	-	-	H	H	-	-	-	H	-	H
B119	A	-	A	-	-	-	-	nude	A	-	-	-	-	-	H	H	-	-	-	H	-	H
B140	H	-	H	-	-	-	-	nude	A	-	-	-	-	-	A	A	-	-	-	A	-	A
B147	H	-	H	-	-	-	-	nude	A	-	-	-	-	-	A	A	-	-	-	A	-	A
B171	A	-	A	-	-	-	-	nude	A	-	-	-	-	-	A	A	-	-	-	H	-	H
B173	A	-	A	-	-	-	-	nude	A	-	-	-	-	-	?	H	-	-	-	H	-	H
B180	A	-	A	-	-	-	-	nude	A	-	-	-	-	-	A	A	-	-	-	A	-	H
C4	A	A	A	A	A	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
C11	H	H	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A
C12	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	A
C56	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	C	C	C	C	C	C	C
C76	A	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
C100	A	A	A	A	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
C161	A	A	A	A	A	A	A	nude	A	A	A	A	A	H	H	H	H	H	H	H	H	H
C163	C	C	C	C	C	C	C	wt	C	C	C	C	C	C	C	C	C	?	?	?	H	H
C176	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	?	?	?	A	A
C190	H	H	H	H	C	C	C	wt	C	C	C	C	C	C	C	C	C	C	C	C	C	C
C216	A	A	A	A	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
C223	A	A	A	A	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H

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) F_2 animals is abbreviated as B#; an (AKR-*nu*^{str} X CAST/Ei) F_2 animals is abbreviated as C#; and an (AKR-*nu*^{str} X MOLF/Ei) F_2 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.

ANIMAL	D11Nds1	D11Mit7	Y8L	RDA6.2	RDA10.4	CA69.4	CA128.1	Phenotype	Genotype	CA4	CA8	RDA10.2	BS7.10	D11Mit117	D11Mit144	D11Mit118	D11Mit96	D11Mit91	D11Mit94	D11Mit34	D11Mit32	D11Mit8
M86	M	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	H	H	H	H	H	H	H
M100	H	H	H	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M143	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	M
M152	H	H	H	H	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M
M184	A	A	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	H	H	H
M210	H	H	H	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M225	M	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
M230	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	M	M	M
M242	A	A	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	H
M273	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	A	A	A
M284	H	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M
M294	M	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	H	H	H	H	H
M295	M	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
M310	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	A	A	A
M402	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	M
M418	H	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M
M433	H	H	H	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M445	H	A	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M465	H	H	H	H	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M
M486	A	A	A	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
M491	H	A	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M496	M	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	H	H	H	H	H	H	H
M558	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	A	A	A	A	A	A	A
M564	H	H	H	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M
M579	M	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
M603	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	A	A	A	A	A	A	A
M627	H	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M
M722	M	M	M	?	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
M723	A	A	A	?	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H
M952	H	H	H	H	H	H	H	wt	H	H	H	A	A	A	A	A	A	A	A	A	A	A

Assay name	Forward Primer 5'-3'	Reverse Primer 5'-3'	Amplicon Size (bp)			
SSLP markers:						
			AKR/J	C57BL6/J	CAST-Ei	MOLF-Ei
Y8L	ATGCGTGCTGTAGCGTGTAG	GGCCAGTGAGCTAAAGAATATC	140	130	132	144
BS7.10	CTTGAGAGATGGTTAAGGGTGG	ACTGNAAATTATGAACTACCTACCCA	300	300	0	304
CA4	AACTCCCATCCAGGGAGTCT	TGGAGAGAATAATGGGGCAG	210	210	196	214
CA8	CACAGCAACAAGCCAAGAAA	TCACTGTGCATAGCAATCAGC	180	180	190	186
CA69	CTCCACGGACATTCTGCC	TCAACTGTCTTAAGCATCCTGA	160	160	140	130
CA128	GATCCCAGGAGATTGGCC	TCTGAAGAGAAACCAGAGTATGACA	125	125	127	135
SSCP markers:						
BS11.4	TCACAAGACAGAAGCAAAAGG	TGGCTCTCCCCTTTTCTCC	161			
BS11.5	GAGACAATCTCATTCCGTTACG	CCGTAAGTGGTGCCATCC	102			
BS11.7	AGCAGGCACTTCCACACC	CCTCACTGGATAAGAGTGTGG	176			
BS11.8	CTTGCTGGAGACCAGCCC	CAACTTGAGCTACACAATGAGC	223			
BS12.42	TTACCCGAGGCCTTCATG	TCATGAGCTGCTTGCCTAAG	222			

Table II. Primers to amplify polymorphisms in the *nude* region. The allele sizes are given for the SSLP markers in the relevant strains.

Assay name	Forward Primer 5'-3'	Reverse Primer 5'-3'	Amplicon Size (bp)
Y4R	tcgtaggccctcatgcttgca	agacggtatcacgtcctcatgc	100
Y8L	atgcgtgctgtagcgtgtag	ggccagtgagctaaagaatatc	140
Y8R	ttctgggacgtggaactctc	cctgggcaaatcctagtaacc	600
Y9L	ttaggctgcagaaagtcttca	aagccagtcatgtgatcca	120
Y10L	cctttgttgtgaatttcactttc	atcaagcaaggcaatactgatg	122
Y11L	aaggtggtcaacctaatgattg	tccattggacaaaatactcaga	258
Y14L	agtaaggcacgttctttttaaagaaacaa	gctctctctcatggatggct	180
Y14R	gctttgccaaaaatcgtcc	tgtttacgatgtaatcacttaggg	104
Y15R	gagtgttagattcccatggagc	ccaagatggcctaactcaca	161
Y19L	tgtatttggctaccatgaaatg	tgaaagactgtggcgaactc	156
Y19R	ccactgagctaaatccccaa	tggcagagtgtctccttgc	293
Y20L	cagatgttaatatcccacttacaactg	agaaagcagtactgtaggtaggtc	183
Y21L	gtggaaggtgaaggatggac	ctgcatgctaggtgtgcag	124
Y25L	aattctgcttcattgctgagtc	gcccagccnttctccctag	62
Y25R	gacgcatcgtggatcatttt	tatcccaccagccacagc	112
Y26R	tgctcggacgagcttcttc	ggtggtttctaggtgaatcca	199
Y27L	ctttctgttaccacagcccc	gtccacggaccttctgcact	275
Y27R	gggaactgaacctagactcgt	ggtgttgaaattaacatcaagaatt	126
Y28L	tagttccctcagatgccaga	aggacccaggtttgattacc	99
Y28R	atcaagccatggaaaagcac	atgacaacgtgtttcaggcc	130
Y29L	cctcgtgtcttctctgtaac	ggcctttgtcttgtggtgtc	185
Y29R	gcccagaacttaagaagagca	cgtgactgtggaggccag	144
Y30L	tccctcagctagatagtaagttcc	aaaccctaactaacctaggctctg	222
Y31L	gtncatgcgtatggcatg	cactactgccaaaacttggaa	136
Y31R	ttgaaagttaggatggaggtgg	tctgtgcatgtgcagacaga	165
Y32L	gggttccaggcagctctt	caaagcttatcctagaccctga	149
Y33L	tagggagaatcaggggcga	agattttgcaccaaagtgtgg	175
Y34R	gaaaattcccacccttaagtcc	catcagcctttgcctctagg	213
P7N	tcacctgggtactgagaatgg	gctttcatggaagtccaagg	270
P7S	ccctttttgaatacctcaccc	gcaccgacagcaaggtga	181
P8S	ggctgtaccggtcagagatg	tccaatgagactgaagcttcc	177
P9S	ggaactctccgtagatgaggc	ctggcatcccagcatctg	163
P10N	ctcaggcatcgctcacattt	gcactgaggcacaagtcct	196
P10S	taactggacttcctggttaggc	tcttcccagtcccctacctc	97
P11S	tatgcacatgtacacacaatgc	ctggctactgttttctggct	183
P13S	tcaccaatgcgtgtgacct	tcttgactcctttgaggtagcc	174
P14N	aggggtcttgaacctgcca	aaaccacgtgcgaatgct	170
P14S	tctagctacctataggctcgca	tgcgttaactcagctaactcg	149
P15N	ctctaccctcccgcgtct	gattccggtaaaggtgctgg	123
P15S	ccactcactaagaggccagc	tgcaaaagaaggggtaaga	121
P19N	gagaaatggcaagaccctcg	aacactttgtttagtgctggc	156
P23S	ttgtatcagccaactgtcttgg	tggctggagtcatggctaag	166
C1193T7	atctcctcccctaacctggg	ttctcacctgctcctaggga	237
C1204T3	tgaaactctgggcaggagg	cagctcccaattagccactg	191
C1204T7	gatcacaaccatctgtaatggg	taaagatgtatttattttatgtatg	114

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 refered 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 *Not*I site of the vector and P#S is the end of the P1 clone adjacent to the *Sal*I site of the vector.

YAC #	Address	Size (kb)	Locus Screened	Left end	Right end
4	PR C100 G10	380	D11Mit7	repetitive	contig
5	MIT fbw G4	375	D11Mit7	N.D.	N.D.
6	PR D21 E7	250	D11Mit7	N.D.	N.D.
7	MIT fcd A10	500	D11Mit7	chimeric	N.D.
8	MIT fap F10	150	D11Mit7	contig;SSLP	contig;SSCP
9	MIT ffk E6	500	Nf1,exon1	contig;SSCP	chimeric
10	MIT fck C3	600	Nf1,exon1	contig;SSCP	N.D.
11	MIT fex C6	625	Nf1,exon1	contig;SSCP	N.D.
14	MIT fcl F6	700	D11Mit34	contig;size	contig
15	MIT fal B7	700	D11Mit34	N.D.	contig
16	MIT adg D4	450	D11Mit8	N.D.	N.D.
17	MIT feh C7	575	D11Mit8	N.D.	N.D.
18	MIT ffh E9	650	D11Mit8	N.D.	N.D.
19	MIT fao E1	700	RDA6.2	contig	contig;SSCP
20	MIT fbr A8	700	RDA6.2	contig	N.D.
21	MIT fbn E6	1000	RDA6.2	contig	repetitive
22	MIT fee F8	550	RDA6.2	N.D.	chimeric
25	PR A15 B3	225	D11Mit34	contig;SSCP	contig
26	PR C189 G5	200	D11Mit34	N.D.	contig;SSCP
27	PR C190 D10	300	D11Mit34	contig;SSCP	contig
28	MIT aan G1	500	Y14R	contig	contig
29	MIT fel D9	625	Y14R	contig;SSCP	contig;SSCP
30	MIT ffq G1	675	Y14R	contig;SSCP	chimeric
31	MIT fal G2	1000	RDA10.4	contig	contig;SSCP
32	MIT feq B12	550	RDA10.4	contig;SSCP	chimeric
33	MIT fes F2	625	RDA10.4	contig	N.D.
34	MIT fex D10	675	RDA10.4	repetitive	contig
35	MIT far F12	400	RDA10.2	N.D.	N.D.

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 refered 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, RDA10.2, RDA10.4 are markers obtained by GD-RDA. Nf-1 is a previously identified locus from this region (Andre Bernards, 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 amplicon from the terminal sequence is a genetic marker, the type of polymorphism is described.

	D11Nds1	D11Mit7	Y8L	Y19R	RDA6.2	Y32L	RDA10.4	BS12.42	P7N	CA69	CA128	Phenotype	Genotype	CA4	Y29L	P14N	CA8	P15S	RDA10.2	BS7.10	Y31R	D11Mit117	BS11.8	BS11.7	BS11.5	BS11.4	D11Mit144	Y30L	Y29R	D11MIT118	D11Mit96	D11Mit91	D11Mit94	D11Mit34	D11Mit32	Y14L	D11Mit8
C11	H	H	A	A	A	A	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M100	H	H	H	H	A	A	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M210	H	H	H	H	A	A	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M433	H	H	H	H	A	A	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
M486	A	A	A	A	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
M564	H	H	H	H	M	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
M152	H	H	H	H	H	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
M465	H	H	H	H	H	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
C100	A	A	A	A	A	?	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
C190	H	H	H	H	H	?	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
C216	A	A	A	A	A	?	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
C223	A	A	A	A	A	?	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
C4	A	A	A	A	A	A	A	A	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
M952	H	H	H	H	H	H	H	H	H	H	H	wt	H	H	H	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
C161	A	A	A	A	A	A	A	A	A	A	A	nude	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H	H	H	H	H	H	H
M86	M	M	M	M	M	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	H	H	H	H	H	H	H	H
M496	M	M	M	M	M	M	M	M	M	M	M	wt	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	H	H	H	H	H	H	H	H
M558	H	H	H	H	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	A	A	A	A	A	A	A	A
M603	H	H	H	H	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	A	A	A	A	A	A	A	A
M143	H	H	H	H	H	H	H	H	H	H	H	wt	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	M	M

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-*nu*^{str} X CAST/Ei) F_2 animals is abbreviated as C#; and an (AKR-*nu*^{str} X MOLF/Ei) F_2 animals is abbreviated as M#. For the genotypes: A is an AKR-*nu*^{str} homozygote; C is a CAST/Ei homozygote; H is a heterozygote; M is a MOLF/Ei homozygote; and ? is an undetermined genotype.

YACs	Primer name	Primer Sequence 5'-3'
Left side	LEFT1	TTT CTT CAA CAA TTA AAT ACT CTC GG
	INVL-2	AAG AAT TGA TCC ACA GGA CGG GTG T
	M13FORL1	TGT AAA ACG ACG GCC AGT GTA GCC AAG TTG GTT T
	M13REVL2	CAG GAA ACA GCT ATG ACC TGG TCG CCA GAT GGT T
Right side		
	RIGHT2	CAA AAA AAT CTC CCG GGG G
	RIGHT7	TTC AAG AAT TGA TCC TCT ACG C
	M13REVR2	CAG GAA ACA GCT ATG ACC CGA GTC GAA CGC CCG A
	M13FORR7	TGT AAA ACG ACG GCC AGT GCC GGA CGC ATC GTG
	M13FORALUR	TGT AAA ACG ACG GCC AGT CAT AAG GGA GAG CGT C
COSMIDS	Primer name	Primer Sequence 5'-3'
T3 side	CT3_7931For	CGT CTT CAA GAA TTC GCG G
	CT3_M13For3	TGT AAA ACG ACG GCC AGT TTC GCG GCC GCA ATT AAC
	CT3_7921For	CGT GAT ACG CCT ATT TTT ATA GG
	CT3_M13Rev7901	CAG GAA ACA GCT ATG ACC AGG TTA ATG TCA TGA TAA TAA TGG
T7 side		
	CT7_147For	GAA AGG AAA CGA CAG AGG CC
	CT7_94Rev	TCG ATG ATA AGC GGT CAA AC
	CT7_M13For170	TGT AAA ACG ACG GCC AGT AAG CTC GCT TTC AGC ACC T
	CT7_M13Rev77	CAG GAA ACA GCT ATG ACC AAC ATG AGA ATT CGC GGC
P1	Primer name	Primer Sequence 5'-3'
NotI side	PN_-30For	GCC GCT AAT ACG ACT CAC TAT AGG
	PN_-40Rev	GAG CAA TAT AGT CCT ACA ATG TCA AGC
	PN_M13For-20	TGT AAA ACG ACG GCC AGT CAC TAT AGG GAG AGG ATC
	PN_M13Rev-50	CAG GAA ACA GCT ATG ACC AGT CCT ACA ATG TCA AGC
SfiI side	PS_64For	GAG GAT CGA AAC GGC AGA
	PS_18Rev	CCG TCG ACA TTT AGG TGA CAC
	PS_M13For79	TGT AAA ACG ACG GCC AGT AGA TCG CAA AAA ACA GTA
	PS_M13For312	TGT AAA ACG ACG GCC AGT GCC TCA ATT CAA TCA ACG
	PS_M13Rev18	CAG GAA ACA GCT ATG ACC TGA CAC TAT AGA AGG ATC

Table VI. Primers to clone terminal fragments of YACs, P1s, and cosmids by inverse PCR.

	size	neonatal skin	AKR skin	nustr skin	B6 skin	B6-nu skin	rat-nu skin	thymus	brain	liver	heart	testes	kidney	10.5 dpc	11.5 dpc	10 cDNA	11 cDNA	SK cDNA	THY cDNA	TES cDNA	BR cDNA	AKR DNA	nustr DNA	B6 DNA	nu DNA	nufl DNA	rat DNA	rat-nu DNA
Hfh11; ET90	100	*	*	*	*	*		*						*	*	*	*	*	*	*	*	*	*	*	*	*	*	-?
Na/sulf;ET4.1	283	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*
VIT; ET14	150	*	*	*	*	*		*	*	*	*	*		*	*	*	*	*			*	*	*	*	*	*		
nmo; ET11.10	110	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*		*			*	*		
Mac30; ET6	130	*	*	*	*	*		*	*	*		*	*	*	*	*	*	*	*	*	*							
Ad-Rab; ET7.3	110	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*		*	*		*		*	*	*
TNFa1;ET11.1	114		*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*
ET2.8	95	*	*	*	*	*			*	*		*	*	*	*	*	*	*	*	*		*	*	*	*	*		
ET3.8	110	*	*	*	*		*	*	*			*	*	*	*	*	*	*	*		*	*			*	*		
ET101	140	*	*		*	*	*	*	*	*			*	*	*	*	*					*	*	*	*	*	*	*
ET98	140	*	*	*	*	*	*		*		*	*	*	*	*	*	*	*		*		*	*	*	*	*	*	*
ET69	110	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		
ET3.2	153	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*
ET3.4	105	*	*	*	*	*	*	*	*		*	*	*	*	*			*	*			*	*	*	*	*	*	*
ET101	140	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*		*	*	*
ET99	110	*	*	*	*	*			*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*		
ET98	140	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ET25	80	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*		*			*	*		
ET6.10	100	*	*	*	*	*	*	*		*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

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 alphabatize the entries and some other minor modifications.

SEQUENCE FILENAME: A1
FORWARD PRIMER: CATGTCCCTGCTTAGTTCTGG
REVERSE PRIMER: TTGGGGATCCAGGCGTTGTA
LENGTH OF INSERT: 274

```
AGCTCCCATGCAGAACCATGTCCCTGCTTAGTTCTGGACATCCAAGACAGGGGAAGAACATTTCCCCAT
                >>>>>>>>>>>>>>>>>>>>
TCGAGGGTACGTCTTGGTACAGGGACGAATCAAGACCTGTAGGTTCTGTCCCCTTCTTGTAAAGGGGTA
------------------------------------------------------------------------
CTCCTGACCTGAGAGCAGCCAGCTTGAGTAAGGTGAGGTTGCCAGCCTCCCATAAGGGTCAATACAACG
                                                                  <<<<<<<
GAGGACTGGACTCTCGTCGGTCGAACTCATTCCACTCCAACGGTCGGAGGGTATTCCCAGTTATGTTGC
------------------------------------------------------------------------
CCTGGATCCCCAAACCTGTGCTCAGGAGGAAGCTGGGGCTGAGAATNCCCCAAGGCTATCTACTCACAT
<<<<<<<<<<<<
GGACCTAGGGGTTTGGACACGAGTCCTCCTTCGACCCCGACTCTTANGGGGTTCCGATAGATGAGTGTA
------------------------------------------------------------------------
TTGTGCCTGGACCCGGGCTCACCTGGNTTNCTCCTGCCTNATGNGCTTNGCTCTGTAACTANCTGGG

AACACGGACCTGGGCCCGAGTGGACCNAANGAGGACGGANTACNCGAANCGAGACATTGATNGACCC
------------------------------------------------------------------------
```

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: ACTTCTGCAGTTCTTCCTGCA
REVERSE PRIMER: CCTGTCCCTCAACAAGTGCTTT
LENGTH OF INSERT: 147

```
TCCACTTCTGCAGTTCTTCCTGCATCTTGTCGATTTTGGAAGGATTGAGGGCCCACAGACAGCCCTTTC
   >>>>>>>>>>>>>>>>>>>>
AGGTGAAGACGTCAAGAAGGACGTAGAACAGCTAAAACCTTCCTAACTCCCGGGTGTCTGTCGGGAAAG
------------------------------------------------------------------------
GAGAGGAACTTCCGGATTTATTCTCCACCTTCTCAAAGCACTTGTTGAGGGACAGGTTATGGCGAACAG
                                  <<<<<<<<<<<<<<<<<<<<<<
CTCTCCTTGAAGGCCTAAATAAGAGGTGGAAGAGTTTCGTGAACAACTCCCTGTCCAATACCGCTTGTC
------------------------------------------------------------------------
```

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: GTGNAAANAAGGGGTGCTT
LENGTH OF INSERT: 268
--------------------------------------------------------------------------
 TGGGGCTTCCCTGAGCAGGTGGGTGACAAGCTCTGTGTGGGTGAGGAGCTGGGGAAGTCAGATGGAGCA
                              >>>>>>>>>>>>>>>>>>>>
 ACCCCGAAGGGACTCGTCCACCCACTGTTCGAGACACACCCACTCCTCGACCCCTTCAGTCTACCTCGT
--------------------------------------------------------------------------
 GGCTAGGTAAGAGCCTTAGGACGNACAAAGGTTTGGGCTAGAGTGAGCTCTTATGGTCTCCTTGCATAA

 CCGATCCATTCTCGGAATCCTGCNTGTTTCCAAACCCGATCTCACTCGAGAATACCAGAGGAACGTATT
--------------------------------------------------------------------------
 GCAAAGCACCCCTTNTTTNCACCAGGAGGAGCAACAAGGNCAGTGCANGAGGTCCAGTTNAGCCACAAC
    <<<<<<<<<<<<<<<<<<<
 CGTTTCGTGGGGAANAAANGTGGTCCTCCTCGTTGTTCCNGTCACGTNCTCCAGGTCAANTCGGTGTTG
--------------------------------------------------------------------------
 CTNACTCCCAATATTGCAAAAGGAGGCTTGCCTAAAGGNANGTNTTAGAATGGGGTACCAG

 GANTGAGGGTTATAACGTTTTCCTCCGAACGGATTTCCNTNCANAATCTTACCCCATGGTC
--------------------------------------------------------------------------
 KEY:

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: ACCTCCTGCCGTGTGGTG
REVERSE PRIMER: CCCACATTGAGCTGCACGTAC
LENGTH OF INSERT: 291
--------------------------------------------------------------------------
 ATAAGGGCCACTCACTCGTGATCACTTCCACCACGGCAGAACCTTCTAAGCCTACCTCCTGCCGTGTGG
                                                  >>>>>>>>>>>>>>>>
 TATTCCCGGTGAGTGAGCACTAGTGAAGGTGGTGCCGTCTTGGAAGATTCGGATGGAGGACGGCACACC
--------------------------------------------------------------------------
 TGATCTACCTGCAGCGGGAGATGTCAGGGACACCTGTCTGTGTCCAGCCTCGGGGCCAAGCCCAAGATC
 >>
 ACTAGATGGACGTCGCCCTCTACAGTCCCTGTGGACAGACACAGGTCGGAGCCCCGGTTCGGGTTCTAG
--------------------------------------------------------------------------
 AGTGGCTTCAAGGGAGGCGGCTGGGCAACAAGTACGTGCAGCTCAATGTGGGAGGCTCCCTGTACTACA
                                    <<<<<<<<<<<<<<<<<<<<
 TCACCGAAGTTCCCTCCGCCGACCCGTTGTTCATGCACGTCGAGTTACACCCTCCGAGGGACATGATGT
--------------------------------------------------------------------------
 CCACAGTGCGGGCCCTCACCCGGCACGACACCATGCTCAAGGCCATGTTCAGTGGGCGAATGGAGGTGC

 GGTGTCACGCCCGGGAGTGGGCCGTGCTGTGGTACGAGTTCCGGTACAAGTCACCCGCTTACCTCCACG
--------------------------------------------------------------------------
 TGACTGACAAAAAGA

 ACTGACTGTTTTTCT
--------------------------------------------------------------------------

DUPLICATE REPORT FOR ABX11_1: <no duplicates>
NUCLEOTIDE BLAST HITS:
    ['M80783','Human B12 protein mRNA, compl...',3.0e-53]
PROTEIN BLAST HITS:
    ['A41784','tumor necrosis factor-alpha-induced ...',3.0e-33]
    ['M80783','B12 protein [Homo sapiens]',3.4e-25]
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
```

```
SEQUENCE FILENAME: ABX11_10
FORWARD PRIMER: TGAGCCAGTCACAAACCCC
REVERSE PRIMER: TCGTGTTGATGAGAGAGGCA
LENGTH OF INSERT: 177
-----------------------------------------------------------------
 ACCAAGTGACTTTGAGCCAGTCACAAACCCCCAAATTTGATGACACCTTTGAGAAGAACCTAAGCTCTG
             >>>>>>>>>>>>>>>>>>>
 TGGTTCACTGAAACTCGGTCAGTGTTTGGGGGTTTAAACTACTGTGGAAACTCTTCTTGGATTCGAGAC
-----------------------------------------------------------------
 TCCGACAGGTTAAAGAAATTATCACAGTTCATTCTGGAACAGCAGAAAGGAAACCAGAGTGCCTCTCTC
                                                          <<<<<<<<<<
 AGGCTGTCCAATTTCTTTAATAGTGTCAAGTAAGACCTTGTCGTCTTTCCTTTGGTCTCACGGAGAGAG
-----------------------------------------------------------------
 ATCAACACGAAGTCCGCTGCCTTTAAGAGCTTTATTAGA
 <<<<<<<<<<
 TAGTTGTGCTTCAGGCGACGGAAATTCTCGAAATAATCT
-----------------------------------------------------------------

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
-----------------------------------------------------------------
 ACTCTATGAGACCCCTCGAGTCCCTGACAATTCCTTACTGGAGGCCACAAGCCAAGCCGCAGCCAGGCT

 TGAGATACTCTGGGGAGCTCAGGGACTGTTAAGGAATGACCTCCGGTGTTCGGTTCGGCGTCGGTCCGA
-----------------------------------------------------------------
 TCCCCCAGTGAAGATGAAGACACCTTTGAACTACGGGACCGTGTCCGTGCATTCATTAAGCGCTATAGA
                                          >>>>>>>>>>>>>>>>>>>>
 AGGGGGTCACTTCTACTTCTGTGGAAACTTGATGCCCTGGCACAGGCACGTAAGTAATTCGCGATATCT
-----------------------------------------------------------------
 CTTACGATGACCGGCAACTTGGCCATCAGTCCACTCATCGTGACTGATGAGACCCTTAAGAAATCGGGG

 GAATGCTACTGGCCGTTGAACCGGTAGTCAGGTGAGTAGCACTGACTACTCTGGGAATTCTTTAGCCCC
-----------------------------------------------------------------
 CACAAGAAGTCCTGCTGCCACCCTGTGGAACGTACCCATTGGCCACCCTATGCTGCTCCTGCTAGCTGG
                          <<<<<<<<<<<<<<<<<<<
 GTGTTCTTCAGGACGACGGTGGGACACCTTGCATGGGTAACCGGTGGGATACGACGAGGACGATCGACC
-----------------------------------------------------------------
 AGCCTCTCTAA

 TCGGAGAGATT
-----------------------------------------------------------------

DUPLICATE REPORT FOR ABX12_1: <no duplicates>
NUCLEOTIDE BLAST HITS:
    ['M80783','Human B12 protein mRNA, compl...',1.2e-31]
PROTEIN BLAST HITS:
    ['M80783','B12 protein [Homo sapiens]',9.6e-28]
    ['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

 CGAATGCTACTGGCCGTAGAACCGGTAGTCAGGTGAGTAGCACTGACTACTCTGGGAATTCTTTAGCCC
--------------------------------------------------------------------
 GCACAAGAAGTCCTGTCTGCCACCCTGTGGAACGTACCCATTGGCCACCCTATGCTGCTCCTGCTAGCT
                                                  >>>>>>>>>>>>>>>>>>>>
 CGTGTTCTTCAGGACAGACGGTGGGACACCTTGCATGGGTAACCGGTGGGATACGACGAGGACGATCGA
--------------------------------------------------------------------
 GGGAGCCTGCTCTAGTATGTACAGGTGACAGCCCTGCTTGGGAGGTTGGGGAGCTGACGAGGGAAGGCC

 CCCTCGGACGAGATCATACATGTCCACTGTCGGGACGAACCCTCCAACCCCTCGACTGCTCCCTTCCGG
--------------------------------------------------------------------
 ACATTGTCTTGGCTGGCCCCCCTCCCCCCAGCACTTCCAGATGCCACCCCACTTCCTGAATAAGGAGCT
                            <<<<<<<<<<<<<<<<<<<<
 TGTAACAGAACCGACCGGGGGGAGGGGGGTCGTGAAGGTCTACGGTGGGGTGAAGGACTTATTCCTCGA
--------------------------------------------------------------------
 ACTTCTCTCTAGGA

 TGAAGAGAGATCCT
--------------------------------------------------------------------
DUPLICATE REPORT FOR ABX12_3: ABX12_1 (22)
     vs ABX12_1 (score = 143)
      ABX12_1 ACTCTATGAGACCCCTCGAGTCCCTGACAATTCCTTACTGGAGGCCACAAGCCAAGCCGCAGCCA

      ABX12_1 GGCTTCCCCCAGTGAAGATGAAGACACCTTTGAACTACGGGACCGTGTCCGTGCATTCATTAAGC

                    GCTTACGATGACCGGCATCTTGGCCATCAGTCCACTCATCGTGACTGATGAGACCCTT
                     |||||||||||||||| ||||||||||||||||||||||||||||||||||||||
      ABX12_1 GCTATAGACTTACGATGACCGGCAACTTGGCCATCAGTCCACTCATCGTGACTGATGAGACCCTT

              AAGAAATCGGGGCACAAGAAGTCCTGTCTGCCACCCTGTGGAACGTACCCATTGGCCACCCTATG
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
      ABX12_1 AAGAAATCGGGGCACAAGAAGTCCTG-CTGCCACCCTGTGGAACGTACCCATTGGCCACCCTATG

              CTGCTCCTGCTAGCTGGGAGCCTGCTCTAGTATGTACAGGTGACAGCCCTGCTTGGGAGGTTGGG
              ||||||||||||||| ||||||| |||||
      ABX12_1 CTGCTCCTGCTAGCT-GGAGCCT-CTCTAA

              GAGCTGACGAGGGAAGGCCACATTGTCTTGGCTGGCCCCCCTCCCCCCAGCACTTCCAGATGCCA
      ABX12_1

              CCCCACTTCCTGAATAAGGAGCTACTTCTCTCTAGGA
      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

```
TTAACCTTAACTAAAGGAACAAAACTGGTACGGGCCCCCCCCTTCAGGTTCAAGGTAACCATAACTTTG

AATTGGAATTGATTTCCTTGTTTTGACCATGCCCGGGGGGGGAAGTCCAAGTTCCATTGGTATTGAAAC
```

```
ATCTCACAAGTTGCACGCTTCTAAATTCGACCCCAAGCAAAAGATGAAAATCAAGTGCCGTTTCCAAGA

TAGAGTGTTCAACGTGCGAAGATTTAAGCTGGGGTTCGTTTTCTACTTTTAGTTCACGGCAAAGGTTCT
```

```
AAGCCTACCGGAAGCTTTGAGCACGAGGAGGAGGCGCTGTCCCTGGCAGTGTGCACGGACATGATGAGC
     >>>>>>>>>>>>>>>>>>>
TTCGGATGGCCTTCGAAACTCGTGCTCCTCCTCCGCGACAGGGACCGTCACACGTGCCTGTACTACTCG
```

```
TACATTGGGCCTGAGAGGACGGCAGTTGTCCGGGGGCTGATCACAGGGAAGCCTTCAACATCATCGGCC

ATGTAACCCGGACTCTCCTGCCGTCAACAGGCCCCCGACTAGTGTCCCTTCGGAAGTTGTAGTAGCCGG
```

```
GTCGCATAGTCCAGGTCGCCCAGGCCATGTCCTTGACGGAGGATGTGCTTGCTGCTGCCTTGGCCGACC
   <<<<<<<<<<<<<<<<<<
CAGCGTATCAGGTCCAGCGGGTCCGGTACAGGAACTGCCTCCTACACGAACGACGACGGAACCGGCTGG
```

```
ACCTGCCGAGGACAAATAAGCTCTGACAAGAGCGGCCGCTCAAGTCCAGCTTAG

TGGACGGCTCCTGTTTATTCGAGACTGTTCTCGCCGGCGAGTTCAGGTCGAATC
```

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

```
AGCACCTGTTGCGCCTGTCCGCTTCTGAAAGGCGACCATTCTCCGTTCTGCATGCCAAGCCTCCCCCAT
               >>>>>>>>>>>>>>>>>>
TCGTGGACAACGCGGACAGGCGAAGACTTTCCGCTGGTAAGAGGCAAGACGTACGGTTCGGAGGGGGTA
```

```
AGAATCTGGGTCGTAGACCCAAGACCAGGTATTCCAGAGAGATGGAAGTTGCCAGAGTGGGGGGGGGGG
                                                                  <<<<
TCTTAGACCCAGCATCTGGGTTCTGGTCCATAAGGTCTCTCTACCTTCAACGGTCTCACCCCCCCCCCC
```

```
TACCCAACTTTCCACTTAGCCAGCCACTCAGTCCTCGGCAGGGAATCTGCCATCTGTCTATCTTTAACT
<<<<<<<<<<<<<<<<
ATGGGTTGAAAGGTGAATCGGTCGGTGAGTCAGGAGCCGTCCCTTAGACGGTAGACAGATAGAAATTGA
```

```
CTAGCTGTCCATCTGGAGTATAATTGCTCCCACCCGAGA

GATCGACAGGTAGACCTCATATTAACGAGGGTGGGCTCT
```

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
           >>>>>>>>>>>>>>>>>>>>
 TTGGGGAGCCCCGTTCCCCTTCGACTATATATTAATAACCGAGTGGGTCTCGACGGAGGGGAGTGGACT
--------------------------------------------------------------------------
 TGTCAGCCCAGACCCAGGACAATCAGGCTGACGGACTACAGACGGTCGGAGCTCCTGGCCCCCCAGACC
                                                                  <<<<<<
 ACAGTCGGGTCTGGGTCCTGTTAGTCCGACTGCCTGATGTCTGCCAGCCTCGAGGACCGGGGGGTCTGG
--------------------------------------------------------------------------
 CGGCCCAACGCCGACCTGCTTCACTGAGCAGA
 <<<<<<<<<<<<<
 GCCGGGTTGCGGCTGGACGAAGTGACTCGTCT
--------------------------------------------------------------------------
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
--------------------------------------------------------------------------
 TAATGGACCCATTTATCAGAAGTTTCGCAACCAGTTCTTAGCATTTTCCTTTTCAGAA
      <<<<<<<<<<<<<<<<<<<
 ATTACCTGGGTAAATAGTCTTCAAAGCGTTGGTCAAGAATCGTAAAAGGAAAAGTCTT
--------------------------------------------------------------------------
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: GCAAATTGCCCCCAAATG
LENGTH OF INSERT: 360
------------------------------------------------------------------------
 TAAGCCTTGATCTCAACAAACTGACGCTCCTAGAAGTCGACCCAGCAATTAAAGGTCAGCTGTATCTGT
                    >>>>>>>>>>>>>>>>>>>>>
 ATTCGGAACTAGAGTTGTTTGACTGCGAGGATCTTCAGCTGGGTCGTTAATTTCCAGTCGACATAGACA
------------------------------------------------------------------------
 TTCCTTCGGAGCAAAGAGCTAGAAGACCAAAAGAGCATCACCCGCTTTCACCAGGGCCTGAAGTCTCTG

 AAGGAAGCCTCGTTTCTCGATCTTCTGGTTTTCTCGTAGTGGGCGAAAGTGGTCCCGGACTTCAGAGAC
------------------------------------------------------------------------
 CATCTGCTACTCAGCCAGCATTTGGGGGCAATTTGCCCTGACGGGTACACACAAACCTGGTGCTCAAGG
                    <<<<<<<<<<<<<<<<<<
 GTAGACGATGAGTCGGTCGTAAACCCCCGTTAAACGGGACTGCCCATGTGTGTTTGGACCACGAGTTCC
------------------------------------------------------------------------
 CCAGGTCAACTCGATCTTCCCTGAAAATAGTAACGTGGTGAACTTTGCTTCATGGTTTGGTTTTGCCTT

 GGTCCAGTTGAGCTAGAAGGGACTTTTATCATTGCACCACTTGAAACGAAGTACCAAACCAAAACGGAA
------------------------------------------------------------------------
 CCCCACCATGGTGATCTTGCTGCTACTGGCTTGGCTATAGCTACAGGTCCTCTTCCTGGGTGTCAAGTA

 GGGGTGGTACCACTAGAACGACGATGACCGAACCGATATCGATGTCCAGGAGAAGGACCCACAGTTCAT
------------------------------------------------------------------------
 AGTGGTGCCTTAGAC

 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: CTTCGCAACCTCTTGCCTAG
REVERSE PRIMER: CCTACTTCATTCATCATGGAAG
LENGTH OF INSERT: 275
--------------------------------------------------------------------------
 AGTGTGACCCTGAACAGTTCCCTGTAAAAGACTAGGGAGATGGACAGACTCAAGATTCCCATCCAGCAA

 TCACACTGGGACTTGTCAAGGGACATTTTCTGATCCCTCTACCTGTCTGAGTTCTAAGGGTAGGTCGTT
--------------------------------------------------------------------------
 GGTACTGCCCACATGAAGGTCCACACCTCCAGGACTTCGCAACCTCTTGCCTAGCTGCAGTGGACTCAA
                                      >>>>>>>>>>>>>>>>>>>>
 CCATGACGGGTGTACTTCCAGGTGTGGAGGTCCTGAAGCGTTGGAGAACGGATCGACGTCACCTGAGTT
--------------------------------------------------------------------------
 ACGTGTGGGGAGAAGGCACCGGCGGGAAGGAGAGGGTCCCAGCAAGGGGCAGTTAGTGGACCCACGCAG

 TGCACACCCCTCTTCCGTGGCCGCCCTTCCTCTCCCAGGGTCGTTCCCCGTCAATCACCTGGGTGCGTC
--------------------------------------------------------------------------
 AAGACTCACCTGGGAAGATAAACTGCTTCCATGATGAATGAAGTAGGCAGCTTTAAACAGAAGAGGAG
                              <<<<<<<<<<<<<<<<<<<<<<
 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
--------------------------------------------------------------------------
 ACACCTATGGCCTGGTCAGCTCGGTCTGGACCGCTCCCTGCTGCACCGCGTGTCAGAGCAGCAGCTCCT
          >>>>>>>>>>>>>>>>>>
 TGTGGATACCGGACCAGTCGAGCCAGACCTGGCGAGGGACGACGTGGCGCACAGTCTCGTCGTCGAGGA
--------------------------------------------------------------------------
 GGAGGACTGTGCATCCGCCTGGGAGTGCACCGCACGCGCATCCTCTCTGCAGCCAGAGGAGTCCTCCGC
                                                  <<<<<<<<<<<<<<<<
 CCTCCTGACACGTAGGCGGACCCTCACGTGGCGTGCGCGTAGGAGAGACGTCGGTCTCCTCAGGAGGCG
--------------------------------------------------------------------------
 TGCGATGGGTACCAGAACAGGAGTAAGGAGA
 <<<<
 ACGCTACCCATGGTCTTGTCCTCATTCCTCT
--------------------------------------------------------------------------

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:GGGTGTGGATTAACCTTTTCC
REVERSE PRIMER: CGGATTGCTCTTCCTGTCAT
LENGTH OF INSERT: 185

GTACGGATTGCTCTTCCTGTCATTGGTCACGACGCCTGTCATCCGGGCCAAGCTTTCCAAGNTGGTGAGC

AACTCGNAGNAGGGGTACTGGAAAAGGTTAATCCACACCCGGANTTCNAGCAGTTTCCTCAGAGCCCCTG

ATCTGCCCGAACCCATGAGGAGCAATGGACCCACACTGACGATA

DUPLICATE REPORT FOR ABX7_3: <no duplicates>
NUCLEOTIDE BLAST HITS:
emb|Z34286|OCRADI2 O.cuniculus (AdRab-G) mRNA expressed i... 285 3.6e-18 2
PROTEIN BLAST HITS: <none>
gi|498983|gp|Z34286|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: AATTTCCCTGATTCTCCCAG
REVERSE PRIMER: TGCAGCTGGGACTCCAGG
LENGTH OF INSERT: 276

GTTTACTGGGCCCCTAATAACACTTGGTCAAAATCCATTTTCTTCTGGCCCCAGAGCCCAGGTGTTACC

CAAATGACCCGGGGATTATTGTGAACCAGTTTTAGGTAAAAGAAGACCGGGGTCTCGGGTCCACAATGG

AATTTCCCTGATTCTCCCAGAGATGTCCTTCTGGTTTTGCTCCTTTTCTCATAACTCCATCCTCCTAGG
>>>>>>>>>>>>>>>>>>>>
TTAAAGGGACTAAGAGGGTCTCTACAGGAAGACCAAAACGAGGAAAAGAGTATTGAGGTAGGAGGATCC

CCCCGAGCAGTGGCTCTCTCATCCTAAGGCCCAGGATCTGTAGCAATATACCTGGAGTCCCAGCTGCAC
<<<<<<<<<<<<<<<<<<
GGGGCTCGTCACCGAGAGAGTAGGATTCCGGGTCCTAGACATCGTTATATGGACCTCAGGGTCGACGTG

CCCTAAGCCTGCAGAGCCCCTCTACTGCTCAGGTGTCGGACAAATAGACCTTGTTACCTTAAGAGGAGA

GGGATTCGGACGTCTCGGGGAGATGACGAGTCCACAGCCTGTTTATCTGGAACAATGGAATTCTCCTCT

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: TGGAGGTAGGGGAGGATGTC
REVERSE PRIMER: CAGAGACTGCCCACATATAT
LENGTH OF INSERT: 209
--------------------------------------------------------------------
 CAGACAGGAATCTTTTAAAGAGCTCGTGTTGAATGGGACCTGATCCGTGGAGGTAGGGGAGGATGTCTC
                                                  >>>>>>>>>>>>>>>>>>>>
 GTCTGTCCTTAGAAAATTTCTCGAGCACAACTTACCCTGGACTAGGCACCTCCATCCCCTCCTACAGAG
--------------------------------------------------------------------
 ATGGCTGACATAGTCCAACCCTGGGATAGCATAGAGCCCCGGCTGTCATTATTGGATAGCCAAGATGGC

 TACCGACTGTATCAGGTTGGGACCCTATCGTATCTCGGGGCCGACAGTAATAACCTATCGGTTCTACCG
--------------------------------------------------------------------
 ACAGTTCTGTTCTGAATTCTAGATATATGTGGGCAGTCTCTGGCTCAAATCAGCACTTTGGTAGTAAGT
                         <<<<<<<<<<<<<<<<<<<
 TGTCAAGACAAGACTTAAGATCTATATACACCCGTCAGAGACCGAGTTTAGTCGTGAAACCATCATTCA
--------------------------------------------------------------------
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: TTAGGACATGATACCTTTGA
LENGTH OF INSERT: 227
--------------------------------------------------------------------
 AAGGCTAGGGTAGCTCAGTATGCAAAGCCCTGGGTTTTATCTCCACTGCTACGAAACCCAAGAAACAGT
                                     >>>>>>>>>>>>>>>>>>>>>>
 TTCCGATCCCATCGAGTCATACGTTTCGGGACCCAAAATAGAGGTGACGATGCTTTGGGTTCTTTGTCA
--------------------------------------------------------------------
 GACTCCCCTTCTTCATGGCTATAGCTAGTGCAAAGTTTTGATGGCCTCTTTGTAGTTGTCAGGCCCTTC

 CTGAGGGGAAGAAGTACCGATATCGATCACGTTTCAAAACTACCGGAGAAACATCAACAGTCCGGGAAG
--------------------------------------------------------------------
 CGTGGCAGCAGGCTGACAGCACTCAAAGGTATCATGTCCTAATCCATGTCCAATAGGCCACAGGTTTAC
                         <<<<<<<<<<<<<<<<<<<
 GCACCGTCGTCCGACTGTCGTGAGTTTCCATAGTACAGGATTAGGTACAGGTTATCCGGTGTCCAAATG
--------------------------------------------------------------------
 AGAAGGATTTTCAAAGGGGGGGGTCTCACG
                        ..........
 TCTTCCTAAAAGTTTCCCCCCCCAGAGTGC
--------------------------------------------------------------------

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
------------------------------------------------------------------------
 ATCACGTGTTCATCGACCTGCAGGCGGTGCTGCCGCCGAACTATACCCGAGTGTTCAGCAACCTGTTGC
       >>>>>>>>>>>>>>>>>>>
 TAGTGCACAAGTAGCTGGACGTCCGCCACGACGGCGGCTTGATATGGGCTCACAAGTCGTTGGACAACG
------------------------------------------------------------------------
 GTGGTACTCAAAGGAGTTCAAAGACCCTCTGATGCAGGAGCCCCCAGTGTGGTTCAAGTCCTTCCTGCT
                                                  <<<<<<<<<<<<<<<<<<<<
 CACCATGAGTTTCCTCAAGTTTCTGGGAGACTACGTCCTCGGGGGTCACACCAAGTTCAGGAAGGACGA
------------------------------------------------------------------------
 CTGTGAGTTTGTTCCAGCTCCGGTTCTCACGAATA
                        ..............
 GACACTCAAACAAGGTCGAGGCCAAGAGTGCTTAT
------------------------------------------------------------------------

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
------------------------------------------------------------------------
 TCAGCTTCCTAGAGAGAGCGAGCAAAGCTTCAAGCAGTGACTTTCTGATACAGAGCCTGGGGCTTCTGA
    >>>>>>>>>>>>>>>>>>>>                                          <<<
 AGTCGAAGGATCTCTCTCGCTCGTTTCGAAGTTCGTCACTGAAAGACTATGTCTCGGACCCCGAAGACT
------------------------------------------------------------------------
 GCCCTAGGGCAGCGGTTCCCAGGCTGTGAGTCACACATTAGATATCCTGCATATCAGATATTTACCTTA
 <<<<<<<<<<<<<<<<
 CGGGATCCCGTCGCCAAGGGTCCGACACTCAGTGTGTAATCTATAGGACGTATAGTCTATAAATGGAAT
------------------------------------------------------------------------
 TAATTCATAACAGTAGGAAAATTAGTCGTGAGGTAGTGACAAAAATAATTTTATGGTTGGGGGTCTCAC
                                                              .........
 ATTAAGTATTGTCATCCTTTTAATCAGCACTCCATCACTGTTTTTATTAAAATACCAACCCCCAGAGTG
------------------------------------------------------------------------

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: CCCATGGTGCAGGTTCTG
REVERSE PRIMER: GCGTGCTTAAGGCTCTTACC
LENGTH OF INSERT: 182
--------------------------------------------------------------------------
 GCTCGCCTTTTTTCCCATGGTGCAGGTTCTGCTCGGGCTACTGGTGGTGTTGGGGCTTCCCTGAGCAGGT
             >>>>>>>>>>>>>>>>>>
 CGAGCGGAAAAAGGGTACCACGTCCAAGACGAGCCCGATGACCACCACAACCCCGAAGGGACTCGTCCA
--------------------------------------------------------------------------
 GGGTGACAAGCTCTGTGTGGGTGAGGAGCTGGGGAAGTCAGATGGAGCAGGCTAGGTAAGAGCCTTAAG
                                                  <<<<<<<<<<<<<<<
 CCCACTGTTCGAGACACACCCACTCCTCGACCCCTTCAGTCTACCTCGTCCGATCCATTCTCGGAATTC
--------------------------------------------------------------------------
 CACGCACAAAGGTTTGGGCTAGAGTAGCTCTTATGGTCTCCTTG
 <<<<<
 GTGCGTGTTTCCAAACCCGATCTCATCGAGAATACCAGAGGAAC
--------------------------------------------------------------------------
 KEY:
DUPLICATE REPORT FOR C16: A8 (21)
     vs A8 (score = 126)
           GCTCGCCTTTTTTCCCATGGTGCAGGTTCTGCTCGGGCTACTGGTGGTGTTGGGGCTTCCCTGAGC
                                                                ||||||||||||||||
         A8                                                     TGGGGCTTCCCTGAGC

           AGGTGGGTGACAAGCTCTGTGTGGGTGAGGAGCTGGGGAAGTCAGATGGAGCAGGCTAGGTAAGA
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         A8 AGGTGGGTGACAAGCTCTGTGTGGGTGAGGAGCTGGGGAAGTCAGATGGAGCAGGCTAGGTAAGA

           GCCTTAAGCACGCACAAAGGTTTGGGCTAGAGT-AGCTCTTATGGTCTCCTTG
           ||||| || ||| |||||||||||||||||||| ||||||||||||||||||
         A8 GCCTT-AGGACGNACAAAGGTTTGGGCTAGAGTGAGCTCTTATGGTCTCCTTGCATAAGCAAAGC

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: TGCATGACCCCAGTACCCCA
LENGTH OF INSERT: 191
------------------------------------------------------------------------
 AGGGTCTCTAACTGCCCTGAAGCTCACTATGCAGACCAGGCTGGCCTCAAACTCATAGAGGTCCAGCTG
       >>>>>>>>>>>>>>>>>>>>
 TCCCAGAGATTGACGGGACTTCGAGTGATACGTCTGGTCCGACCGGAGTTTGAGTATCTCCAGGTCGAC
------------------------------------------------------------------------
 CCTCTGCTTCCTGAGTGCTTTGATTCTGTCTCTTCCAGTTCAAGCAGAGTTATGGGGTACTGGGGTCAT
                                                  <<<<<<<<<<<<<<<<<<
 GGAGACGAAGGACTCACGAAACTAAGACAGAGAAGGTCAAGTTCGTCTCAATACCCCATGACCCCAGTA
------------------------------------------------------------------------
 GCATGATGCTCATGATTTCCGTAATCTTTCAGATCAGCCCATGGGGCCACTTCCGGTTCACGAAGTCCG
 <<<                                                  .................
 CGTACTACGAGTACTAAAGGCATTAGAAAGTCTAGTCGGGTACCCCGGTGAAGGCCAAGTGCTTCAGGC
------------------------------------------------------------------------
 CTGCTGG
 .......
 GACGACC
------------------------------------------------------------------------
DUPLICATE REPORT FOR C2: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
------------------------------------------------------------------------
SEQUENCE FILENAME: C3
FORWARD PRIMER: AGAAGGTGGGGACAGGAGAC
REVERSE PRIMER: CCTGGATGGATGCCATGCACAA
LENGTH OF INSERT: 200
------------------------------------------------------------------------
 TGAACTCACAGAAAAGCTTAGAAGGTGGGGACAGGAGACCCACTAAGCATCACTACCCGGGACTCATGC
                    >>>>>>>>>>>>>>>>>>>>
 ACTTGAGTGTCTTTTCGAATCTTCCACCCCTGTCCTCTGGGTGATTCGTAGTGATGGGCCCTGAGTACG
------------------------------------------------------------------------
 TCCTCAAGACCCTGAAGTCACTCTCCACACCTACATATATGTAACACATGTGCACACTTGTGCATGGCA
                                                      <<<<<<<<<<<<
 AGGAGTTCTGGGACTTCAGTGAGAGGTGTGGATGTATATACATTGTGTACACGTGTGAACACGTACCGT
------------------------------------------------------------------------
 TCCATCCAGGGCCATTATTTACTGAGCAAGCTTTGAAGAACTCAGTTTAACCCTCTAAGCTTCTGTCTC
 <<<<<<<<<<                                                  .......
 AGGTAGGTCCCGGTAATAAATGACTCGTTCGAAACTTCTTGAGTCAAATTGGGAGATTCGAAGACAGAG
------------------------------------------------------------------------
 ACGATTCCCGGCTGGTTGGAAGTAAGTATAGTCGGCCCCGTGAAATTCCCGGCCGTACCTGAAGGGGGT
 ........................................................................
 TGCTAAGGGCCGACCAACCTTCATTCATATCAGCCGGGGCACTTTAAGGGCCGGCATGGACTTCCCCCA
------------------------------------------------------------------------
 CCATTTTCCCTTTAGAGGGTCGTTTAGGGTTTTTGGGGAATTAGGGATACTTGGCTAATGGGGGGGGTG
 ........................................................................
 GGTAAAAGGGAAATCTCCCAGCAAATCCCAAAAACCCCTTAATCCCTATGAACCGATTACCCCCCCCAC
------------------------------------------------------------------------
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

TTCGAAAATCGTCCCCTCACTACACTAGGTTCCTGTTGAACTACGACACACCCTACTTTTCATGTACTT
----------------------------------------------------------------------
AGGAAATTGGAAAATCAGAGGAAGGACTGTCCAGGCAGTGATGATAAAGGCTTACACCACAGAGCATGC
                                >>>>>>>>>>>>>>>>>>>>
TCCTTTAACCTTTTAGTCTCCTTCCTGACAGGTCCGTCACTACTATTTCCGAATGTGGTGTCTCGTACG
----------------------------------------------------------------------
TTAAACAGAAACACAACAACCACAACGCCCAAAGGGACCAACCTCTTTTAATACCGGAACACTCAAGTC

AATTTGTCTTTGTGTTGTTGGTGTTGCGGGTTTCCCTGGTTGGAGAAAATTATGGCCTTGTGAGTTCAG
----------------------------------------------------------------------
GTTTCTTTATTCCTCAATGTTTTACAGGTTAAACCCACTAACAACTGGTCCCTCAAACCAAGGATTTAG
                <<<<<<<<<<<<<<<<<<<<
CAAAGAAATAAGGAGTTACAAAATGTCCAATTTGGGTGATTGTTGACCAGGGAGTTTGGTTCCTAAATC
----------------------------------------------------------------------
GTATACCATCAGCTTTGCTAAAACTTAAAACTGACCCTTGAAAGAGGATTTTCAGTTATGGCCCTTTGG

CATATGGTAGTCGAAACGATTTTGAATTTTGACTGGGAACTTTCTCCTAAAAGTCAATACCGGGAAACC
----------------------------------------------------------------------
GAATTGGGTCCGGGGA

CTTAACCCAGGCCCCT
----------------------------------------------------------------------
```

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: ACTGGCCAATTTTGTCCACA
REVERSE PRIMER: AAGGACATCTTTGGCAGCAA
LENGTH OF INSERT: 292

```
GGCCTCAGCTGCTGCTTGGGCTTCTCTCTGCTTCGCTATGGACTTCAGCAAGTTCCGAGCACCAATGGC

CCGGAGTCGACGACGAACCCGAAGAGAGACGAAGCGATACCTGAAGTCGTTCAAGGCTCGTGGTTACCG
----------------------------------------------------------------------
CTTCATCTGCTGCTCTCTTCTTTGGCAAGCTGATCAACAAGCTCAATTAGACCACCAACAATTTTCTGA

GAAGTAGACGACGAGAGAAGAAACCGTTCGACTAGTTGTTCGAGTTAATCTGGTGGTTGTTAAAAGACT
----------------------------------------------------------------------
AACTGGCCAATTTTGTCCACAAAGCCTTGCACTTCTTCTTGAGCTCCACAAGTCTTGTTGGGGTGACCT
  >>>>>>>>>>>>>>>>>>>>
TTGACCGGTTAAAACAGGTGTTTCGGAACGTGAAGAAGAACTCGAGGTGTTCAGAACAACCCCACTGGA
----------------------------------------------------------------------
CTGGTCCAACACTCGGGAGCTTGGTTCAGTTCATCAAAATGAAGCCCTTGCTGCCAAAGATGTCCTTTG
                                                  <<<<<<<<<<<<<<<<<<<<
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: GCAGCCTGATGGTGACTCTG
REVERSE PRIMER: GGTTTCCATGACCTTAATAAA
LENGTH OF INSERT: 281
------------------------------------------------------------------
 ATTAGGAGGCTGAGAGCCACTGCTGTAGAGGATGGTAGGTACAAAAGGAGAAGAGGGAGGGGTAGGAGG

 TAATCCTCCGACTCTCGGTGACGACATCTCCTACCATCCATGTTTTCCTCTTCTCCCTCCCCATCCTCC
------------------------------------------------------------------
 AGGGGAGAGGAGGAGGAAGAGGAAGCAGCCTGATGGTGACTCTGCAGCCCTAGTTTATCCCTAAATCCA
                          >>>>>>>>>>>>>>>>>>>>
 TCCCCTCTCCTCCTCCTTCTCCTTCGTCGGACTACCACTGAGACGTCGGGATCAAATAGGGATTTAGGT
------------------------------------------------------------------
 AAGTTCTTTCCCATTTAGCCTGAGTCCCAGACATTTAGAGGGCGTCCTTATGGTTTAAGCTTCCCTTTT
                                                                    <<
 TTCAAGAAAGGGTAAATCGGACTCAGGGTCTGTAAATCTCCCGCAGGAATACCAAATTCGAAGGGAAAA
------------------------------------------------------------------
 TATTAAGGTCATGGAAACCCTTTTGAGCCATTTTTTTCTCAAAGGGGACTTGGGGCCCTGTTTCCTAAG
 <<<<<<<<<<<<<<<<<<
 ATAATTCCAGTACCTTTGGGAAAACTCGGTAAAAAAAGAGTTTCCCCTGAACCCCGGGACAAAGGATTC

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
  >>>>>>>>>>>>>>>>>>>>
 TGTTACTACTTCTCCGGCAGTCTACCTGTCTCCTTCGTCTCCGTCTTTGACGCTAGTTCTCCTTCCCCG
------------------------------------------------------------------
 TTGGCACGTCCAACAGAAGACTATTGGAAGGAGTGGTGATGGGCGCCTGTACATCCCAGCTCTTGGTGC
                                          <<<<<<<<<<<<<<<<<<<<
 AACCGTGCAGGTTGTCTTCTGATAACCTTCCTCACCACTACCCGCGGACATGTAGGGTCGAGAACCACG
------------------------------------------------------------------
 AGACAGGAGATT

 TCTGTCCTCTAA
------------------------------------------------------------------

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: TCTTTGTCCCCATCCCCA
REVERSE PRIMER: GGCCTTTCCTTCTTNGTTCT
LENGTH OF INSERT: 159
--------------------------------------------------------------------------
 AAACAGAAGCCCCGTGCCNACCGGTTCTTTGTCCCCATCCCCATTNAAGTAGGGCCCAGCAAACTTNCA
                              >>>>>>>>>>>>>>>>>>
 TTTGTCTTCGGGGCACGGNTGGCCAAGAAACAGGGGTAGGGGTAANTTCATCCCGGGTCGTTTGAANGT
--------------------------------------------------------------------------
 AAGCCAACTCTGAAGTCANACTTCCCACCTTGCCACCCAGGTCCTTAAGAACNAAGAAGGAAAGGCCTG
                                                  <<<<<<<<<<<<<<<<<<<<
 TTCGGTTGAGACTTCAGTNTGAAGGGTGGAACGGTGGGTCCAGGAATTCTTGNTTCTTCCTTTCCGGAC
--------------------------------------------------------------------------
 TGAAGACAGCTCAGCACAATG

 ACTTCTGTCGAGTCGTGTTAC
--------------------------------------------------------------------------
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
               >>>>>>>>>>>>>>>>>>>>>
 TCGTCATGACACTCATGCTTAAATTCGTCGTTGGGTCGGTCCTCCTCACGCTTCCGTCGAGAGACAGTC
--------------------------------------------------------------------------
 CCGTGTTTAGCACTTTGCCTTGCTTCAGCGGGACAGCTGGGAGAACATTTTCAACTCCTCTTCTGGGGC
                                        <<<<<<<<<<<<<<<<<<<
 GGCACAAATCGTGAAACGGAACGAAGTCGCCCTGTCGACCCTCTTGTAAAAGTTGAGGAGAAGACCCCG
--------------------------------------------------------------------------
 AGATCCTCCG

 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]
    ['M55442','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]
    ['M77123','vitronectin [Mus musculus]',3.7e-16]
    ['JT0662','Vitronectin - Mouse>gi 441466 gp X72...',3.8e-16]
    ['X03168','Human mRNA for S-protein....',8.2e-13]
    ['P04004','VITRONECTIN PRECURSOR (SERUM SPRE...',8.6e-13]
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
```

```
SEQUENCE FILENAME: EX25
PRIMERS NOT PICKED
LENGTH OF INSERT: 81
--------------------------------------------------------------------
 AACCGTGCAGAGGGCAAAGTGCTGGAGACAGTTGGTGTGTTTGAGGTGCCAAAACAAAATGGAAAATAT

 TTGGCACGTCTCCCGTTTCACGACCTCTGTCAACCACACAAACTCCACGGTTTTGTTTTACCTTTTATA
--------------------------------------------------------------------
 GAGACTGGGCAG

 CTCTGACCCGTC
--------------------------------------------------------------------

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
--------------------------------------------------------------------
 TTCAGCAACCTGTTGCGGNTGGTACTCAAAGGAGTTCAAAGACCCTCTGATGCAGGAGCCCCCAGTGTG
                                      >>>>>>>>>>>>>>>>>>>>
 AAGTCGTTGGACAACGCCNACCATGAGTTTCCTCAAGTTTCTGGGAGACTACGTCCTCGGGGGTCACAC
--------------------------------------------------------------------
 GTTCAAGTCCTTCCTGCTCTGTGAGCTTGTGTTCCAGCTGCCTTTCTTTCCCATTGCGGCATATGCCTT
                                    <<<<<<<<<<<<<<<<<<<
 CAAGTTCAGGAAGGACGAGACACTCGAACACAAGGTCGACGGAAAGAAAGGGTAACGCCGTATACGGAA
--------------------------------------------------------------------
 CTTCAA

 GAAGTT
--------------------------------------------------------------------

DUPLICATE REPORT FOR EX6: C14 (16)
     vs C14 (score = 100)
                                                                TTCAGCAACCTG
                                                                ||||||||||||
          C14 ATCACGTGTTCATCGACCTGCAGGCGGTGCTGCCGCCGAACTATACCCGAGTGTTCAGCAACCTG

              TTGCGGNTGGTACTCAAAGGAGTTCAAAGACCCTCTGATGCAGGAGCCCCCAGTGTGGTTCAAGT
              |||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          C14 TTGC-G-TGGTACTCAAAGGAGTTCAAAGACCCTCTGATGCAGGAGCCCCCAGTGTGGTTCAAGT

              CCTTCCTGCTCTGTGAGCTTGTGTTCCAGCTGCCTTTCTTTCCCATTGCGGCATATGCCTTCTTC
              |||||||||||||||| || ||||||||||
          C14 CCTTCCTGCTCTGTGAG-TT-TGTTCCAGCTC

              AA
          C14

NUCLEOTIDE BLAST HITS:
    ['L19183','Human MAC30 mRNA, 3 end.',4.0e-31]
PROTEIN BLAST HITS:
    ['L19183','MAC30 gene product [Ho...',2.1e-24]
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
```

```
SEQUENCE FILENAME: EX69
FORWARD PRIMER: ACTCAACCCAGACCCCAAGT
REVERSE PRIMER: AGCGGGGGTTCACGCCGAG
LENGTH OF INSERT: 246
--------------------------------------------------------------------
 AATCACTTCAGTTTACTCAACCCAGACCCCAAGTCCCACGATGTCCACTGGGACATTGAGGACCTGTCC
              >>>>>>>>>>>>>>>>>>>
 TTAGTGAAGTCAAATGAGTTGGGTCTGGGGTTCAGGGTGCTACAGGTGACCCTGTAACTCCTGGACAGG
--------------------------------------------------------------------
 AGCGCTTTGTGCAGCCTTTCCTGAATCGCCTCAGTGTTGCGGGCAATTTCTCTGTGGGCTCTCAGATCC

 TCGCGAAACACGTCGGAAAGGACTTAGCGGAGTCACAACGCCCGTTAAAGAGACACCCGAGAGTCTAGG
--------------------------------------------------------------------
 TATACTATGCCATGCTCGGCGTGAACCCCCGCTTTGACCCAGCCTCATCACTACTCCTTGGCTATGCAC
               <<<<<<<<<<<<<<<<<<<
 ATATGATACGGTACGAGCCGCACTTGGGGGCGAAACTGGGTCGGAGTAGTGATGAGGAACCGATACGTG
--------------------------------------------------------------------
 AGCCTGCCCCACGTCATCAACCCAGTGAGTCCCGACTCG

 TCGGACGGGGTGCAGTAGTTGGGTCACTCAGGGCTGAGC
--------------------------------------------------------------------

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
                    >>>>>>>>>>>>>>>>>>>
 ACATGTAAGGGCGACGGGACTGACCTCCGTTCAGTCACCCCTGTGGGGTCTACAGAAATAGTCAATGGC
--------------------------------------------------------------------
 GAGGAACTCAGGGTCCCAGCTGGCCAGCCTCCTGAAGTTGCACCTGCAGCTTCACGGCTTCAGCGTCTT

 CTCCTTGAGTCCCAGGGTCGACCGGTCGGAGGACTTCAACGTGGACGTCGAAGTGCCGAAGTCGCAGAA
--------------------------------------------------------------------
 CATCGACGTGGAGAAGCTGGAACCGGCAAATTCGAGGACAAGCTTATCCAAAGCGTCATAGGCTCGCAA
      <<<<<<<<<<<<<<<<<<<<
 GTAGCTGCACCTCTTCGACCTTGGCCGTTTAAGCTCCTGTTCGAATAGGTTTCGCAGTATCCGAGCGTT
--------------------------------------------------------------------
 TTTTGTCCTGGTGCTGTCTGCTGGGGCGCTGGATAAGTGCATGAGGCCATGACTGCAAGGACTGGGTGC

 AAAACAGGACCACGACAGACGACCCCGCGACCTATTCACGTACTCCGGTACTGACGTTCCTGACCCACG
--------------------------------------------------------------------
 ACAAGACCTGGGA

 TGTTCTGGACCCT
--------------------------------------------------------------------

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: ACATCCTCATCTTCATGGCC
LENGTH OF INSERT: 91
--------------------------------------------------------------------
 CCTCATCTTCATGGCCCTTAAGAACAGTAAGACCGGAAGCCTTCCAGTCAGTGAAATCTACAATTTCAT

 GGAGTAGAAGTACCGGGAATTCTTGTCATTCTGGCCTTCGGAAGGTCAGTCACTTTAGATGTTAAAGTA
--------------------------------------------------------------------
 GACGGAGTGCCATTACTTCAAG

 CTGCCTCACGGTAATGAAGTTC
--------------------------------------------------------------------
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
--------------------------------------------------------------------
 AGAACTGGGGAGACTCTAGGAAACTCTTCAAATAATCCCAGGTTTACACGGCACTATCCACACTGAGAG
      >>>>>>>>>>>>>>>>>>>>
 TCTTGACCCCTCTGAGATCCTTTGAGAAGTTTATTAGGGTCCAAATGTGCCGTGATAGGTGTGACTCTC
--------------------------------------------------------------------
 AATTAGGACTTCAGCAGGAGAGCTTCCTCAGGCTGCCATTTTGAATGAGGGTCCATCTTCCTGAGCACC
                                                  <<<<<<<<<<<<<<
 TTAATCCTGAAGTCGTCCTCTCGAAGGAGTCCGACGGTAAAACTTACTCCCAGGTAGAAGGACTCGTGG
--------------------------------------------------------------------
 TGGCATGGGTCAGCCTGTCTCCTCTGATCTG
 <<<<
 ACCGTACCCAGTCGGACAGAGGAGACTAGAC
--------------------------------------------------------------------

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: TGACTGCTTTAAGCTGTGGCA
REVERSE PRIMER: GTGGCCTCTGGTACTCATGG
LENGTH OF INSERT: 245
--------------------------------------------------------------------
 GACCCAGCAGAGATTGTGACTGCTTTAAGCTGTGGCAAGAACATTGTGCCCATCATTGATGGCTTTGAG
                  >>>>>>>>>>>>>>>>>>>>
 CTGGGTCGTCTCTAACACTGACGAAATTCGACACCGTTCTTGTAACACGGGTAGTAACTACCGAAACTC
--------------------------------------------------------------------
 TGGCCTGAAGCCTAGCGCTGCCTGAGGATATGCAGGCTGTACTCACCTTCAACGGCATCAAATGGTCCC
                                                                    <<
 ACCGGACTTCGGATCGCGACGGACTCCTATACGTCCGACATGAGTGGAAGTTGCCGTAGTTTACCAGGG
--------------------------------------------------------------------
 ATGAGTACCAGAGGCCACCATCGAGAAGATCATCCGCTTCCTACAGGGCCGCCCCTCTCAGGACTCCCT
 <<<<<<<<<<<<<<<<<<
 TACTCATGGTCTCCGGTGGTAGCTCTTCTAGTAGGCGAAGGATGTCCCGGCGGGGAGAGTCCTGAGGGA
--------------------------------------------------------------------
 GCCGGATCGGATACCAGTTTGGAGGGACTACGCAATGG

 CGGCCTAGCCTATGGTCAAACCTCCCTGATGCGTTACC
--------------------------------------------------------------------

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
--------------------------------------------------------------------
 GCCTACTGTGCTTACTCCATCATCCTCATGGCGCTGCTGTGTGTACAGAGGCCCTGCCCTTGGCTGTAG
           >>>>>>>>>>>>>>>>>>>>
 CGGATGACACGAATGAGGTAGTAGGAGTACCGCGACGACACACATGTCTCCGGGACGGGAACCGACATC
--------------------------------------------------------------------
 CGCCCTCTTCCCCATCATCCTCTTCCCTTTGATGGGTATCATGGAAGCCTCCAAG
       <<<<<<<<<<<<<<<<<<<
 GCGGGAGAAGGGGTAGTAGGAGAAGGGAAACTACCCATAGTACCTTCGGAGGTTC
--------------------------------------------------------------------

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
-------------------------------------------------------------------
 TCGTGAGACCCAGGGTGTCCCACTGTAGGAATAGTCTGCTTTGTTATGCATCACCAGCCGGTCATCCAC
 .........                          >>>>>>>>>>>>>>>>>>>>
 AGCACTCTGGGTCCCACAGGGTGACATCCTTATCAGACGAAACAATACGTAGTGGTCGGCCAGTAGGTG
-------------------------------------------------------------------
 GAAGTAGAAGCTGTCAGACTGGGTCTCATACGGGTGACGAATCATCTCACTGATTAGCTCCTCGGAGAG

 CTTCATCTTCGACAGTCTGACCCAGAGTATGCCCACTGCTTAGTAGAGTGACTAATCGAGGAGCCTCTC
-------------------------------------------------------------------
 AGGAGGGAAGTCATAGATGTGCTCACAGGTGTTCTTGCTTGCTGGGGATGCAGGTCTCACGAATACNAG
             <<<<<<<<<<<<<<<<<<                  ..................
 TCCTCCCTTCAGTATCTACACGAGTGTCCACAAGAACGAACGACCCCTACGTCCAGAGTGCTTATGNTC
-------------------------------------------------------------------
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: AGGAGTCCTCCTCTTGCAGG
REVERSE PRIMER: GGGTGGCTCTGGACCTTGAAG
LENGTH OF INSERT: 487
-------------------------------------------------------------------
 ATTCGTGAGACCGACCATTTCCTGTCCCTAACCTGATCTAGTCACGGTCAAGAAGGGGGCACTACTCTG
 ..........
 TAAGCACTCTGGCTGGTAAAGGACAGGGATTGGACTAGATCAGTGCCAGTTCTTCCCCCGTGATGAGAC
-------------------------------------------------------------------
 AAACTCAGGCCCAGTGGACCTAGCGAGGCATCCAGGAGTCCTCCTCTTGCAGGAGGCCTTCCTGAACAT
                                    >>>>>>>>>>>>>>>>>>>>
 TTTGAGTCCGGGTCACCTGGATCGCTCCGTAGGTCCTCAGGAGGAGAACGTCCTCCGGAAGGACTTGTA
-------------------------------------------------------------------
 GGGGACAGCCAGCCTCCCCCTGCTGCCCCTCCCACTAGCTCCAGAGGTTTGCCTAGGGATGCTGGACAA

 CCCCTGTCGGTCGGAGGGGGACGACGGGGAGGGTGATCGAGGTCTCCAAACGGATCCCTACGACCTGTT
-------------------------------------------------------------------
 CTTTCCTCTCGAGACTGACTTCAAGGTCCAGAGCCACCCGGAATCTTTTGTCTGTTGAACCAACAACTT
                 <<<<<<<<<<<<<<<<<<<<
 GAAAGGAGAGCTCTGACTGAAGTTCCAGGTCTCGGTGGGCCTTAGAAAACAGACAACTTGGTTGTTGAA
-------------------------------------------------------------------
 ATAAATTAGCCTGGAAGGTTGGTGGGGCAAGGGGACAAGGTCTCAACGGNTTTACAGGNTNATGGNGTG

 TATTTAATCGGACCTTCCAACCACCCCGTTCCCCTGTTCCAGAGTTGCCNAAATGTCCNANTACCNCAC
-------------------------------------------------------------------
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: TGNCCTGTGTGAATAGCTTG
LENGTH OF INSERT: 290
--------------------------------------------------------------------
 GGAATTCGTGAGACCGAGACAGGAAACTTCTAGAAAAATCAGATTCCATGATGAGCAGGCAAGGTCAGA
 .............
 CCTTAAGCACTCTGGCTCTGTCCTTTGAAGATCTTTTTAGTCTAAGGTACTACTCGTCCGTTCCAGTCT
--------------------------------------------------------------------
 GCCCTCCATTACCCAATCAGAATGGCTGGAATGAAGAGGAGGCCAGCTCCCAGGAGGAAGGGGAAGCCC
            >>>>>>>>>>>>>>>>>>>>
 CGGGAGGTAATGGGTTAGTCTTACCGACCTTACTTCTCCTCCGGTCGAGGGTCCTCCTTCCCCTTCGGG
--------------------------------------------------------------------
 TTCATGAAGTTCAAGAGTGGCCGTGTAGATGGNGTTGAAGATGCCGGGAGCATTCAACATGGNCAAGCT
                                                              <<<<<<
 AAGTACTTCAAGTTCTCACCGGCACATCTACCNCAACTTCTACGGCCCTCGTAAGTTGTACCNGTTCGA
--------------------------------------------------------------------
 ATTCACACAGGNCACAGCAGAGGAACAGAGCACCNTGNTCCGACTCGCTCACCANCTTTNGAAAGGTTG
 <<<<<<<<<<<<<<
 TAAGTGTGTCCNGTGTCGTCTCCTTGTCTCGTGGNACNAGGCTGAGCGAGTGGTNGAAANCTTTCCAAC
--------------------------------------------------------------------

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) mR...',4.1e-20]
PHYSICAL MAP LOCATION: 14,21
END OF RECORD.
--------------------------------------------------------------------
SEQUENCE FILENAME: a10e
FORWARD PRIMER: TTTTGGACCTGCGTCATTTG
REVERSE PRIMER: GCATTTGGGAGGCACCTG
LENGTH OF INSERT: 248
--------------------------------------------------------------------
 AATTCGTGAGACCAAATGGGGAAGCCTTGGTCAAGGCGGGTTTTGGACCTGCGTCATTTGCTGCTCATC
 ............                          >>>>>>>>>>>>>>>>>>>>
 TTAAGCACTCTGGTTTACCCCTTCGGAACCAGTTCCGCCCAAAACCTGGACGCAGTAAACGACGAGTAG
--------------------------------------------------------------------
 TCTCTGCTCCCTCACTCCCCACCCCCACTCAGTAAAACTGAGAACACCTGTCTGCAGCGACCATGGTCT

 AGAGACGAGGGAGTGAGGGGTGGGGGTGAGTCATTTTGACTCTTGTGGACAGACGTCGCTGGTACCAGA
--------------------------------------------------------------------
 TTCTGTGCCCCCCACAGGCTGCTCTTTGGAATTGCTCAGCCCCAGGTGCCTCCCAAATGCCTGCTCTCA
                                        <<<<<<<<<<<<<<<<<<
 AAGACACGGGGGGTGTCCGACGAGAAACCTTAACGAGTCGGGGTCCACGGAGGGTTTACGGACGAGAGT
--------------------------------------------------------------------
 GGACCAAAGAGTGAAGGACTCATGACCTGGGGAGCTGGACCGATTTCTCAAGGGGGTCTCACGANTTCN
                                                          ..............
 CCTGGTTTCTCACTTCCTGAGTACTGGACCCCTCGACCTGGCTAAAGAGTTCCCCCAGAGTGCTNAAGN
--------------------------------------------------------------------

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: CCACTCCCCTCTAACCCCG
REVERSE PRIMER: CAGCTGACAGACCTTGGCCTCT
LENGTH OF INSERT: 231

```
GCGGAATTCGTGAGACCCACCACTCCCCTCTAACCCCGCCCATACAGAGCTTTTACAAGTGGGATCAGC
.................  >>>>>>>>>>>>>>>>>>>
CGCCTTAAGCACTCTGGGTGGTGAGGGGAGATTGGGGCGGGTATGTCTCGAAAATGTTCACCCTAGTCG
------------------------------------------------------------------------
TCAGGGAGGGAGACAGCTAAAACCTTGTAAATACTTCCTAGAGAGGAGCTGAGGTCAGGAGAGGCCAAG
                                                          <<<<<<<<<<
AGTCCCTCCCTCTGTCGATTTTGGAACATTTATGAAGGATCTCTCCTCGACTCCAGTCCTCTCCGGTTC
------------------------------------------------------------------------
GTCTGTCAGCTGAGAGCTAGTCAGGGTGTGACTCAGTCTGTNGTGTCTGATAACTAGGGGCCTCTTCTN
<<<<<<<<<<<<
CAGACAGTCGACTCTCGATCAGTCCCACACTGAGTCAGACANCACAGACTATTGATCCCCGGAGAAGAN
------------------------------------------------------------------------
GACTCCTTGATAACTTCACTTTTCCTGGAGGATGTGAACCCGGTCTCACGAATNCNAGGTTGGGGNTGN
                                          ............................
CTGAGGAACTATTGAAGTGAAAAGGACCTCCTACACTTGGGCCAGAGTGCTTANGNTCCAACCCCNACN
------------------------------------------------------------------------
```

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: GCTACCCTGTGCCCTACCTG
REVERSE PRIMER: GTTCTTAAGGNCCATGAAGA
LENGTH OF INSERT: 304
--------------------------------------------------------------------------
 AATTCGTGAGACCTATGTACTGTTCTTCTCAGCCTGCCTTCCATCAGTACTCCCCGGGTGGAGGCAGCT
 .............                                                       >>>
 TTAAGCACTCTGGATACATGACAAGAAGAGTCGGACGGAAGGTAGTCATGAGGGGCCCACCTCCGTCGA
--------------------------------------------------------------------------
 ACCCTGTGCCCTACCTGGGCTCACCTCACTATCCCTATCAGAGGATTGCACCCNAGGNCAACGNCGAAG
 >>>>>>>>>>>>>>>>>
 TGGGACACGGGATGGACCCGAGTGGAGTGATAGGGATAGTCTCCTAACGTGGGNTCCNGTTGCNGCTTC
--------------------------------------------------------------------------
 TCACCAGCCACTCTTNCCAAAGNCCATCTACTCTTACAGCATCCTCATCTTCATGGNCCTTAAGAACAG
                                                    <<<<<<<<<<<<<<<<<<<<
 AGTGGTCGGTGAGAANGGTTTCNGGTAGATGAGAATGTCGTAGGAGTAGAAGTACCNGGAATTCTTGTC
--------------------------------------------------------------------------
 TAAGACCGGAAGGCCTTCCAGTCAAGTGAAATCTTACANTTTTCATGACGGGAGCACTTNCCCTTACTT

 ATTCTGGCCTTCCGGAAGGTCAGTTCACTTTAGAATGTNAAAAGTACTGCCCTCGTGAANGGGAATGAA
--------------------------------------------------------------------------
 CAAGGTGAGTNTAAAACCCACCCCTTGNCCCCAAGGCCCAGGGTCTTCACGGANTTCCGGGTTTGTGNG
                                                 ..........................
 GTTCCACTCANATTTTGGGTGGGGAACNGGGGTTCCGGGTCCCAGAAGTGCCTNAAGGCCCAAACACNC
--------------------------------------------------------------------------
DUPLICATE REPORT FOR a10g: EX90 (5)
     vs EX90 (score = 74)
               NCGTNGCATGCACGCGTNACGTAAAGCTTGNGATCCTCTAGAGCGGCGNNTACTACTACTACTGA
          EX90

               GCGGAATTCGTGAGACCTATGTACTGTTCTTCTCAGCCTGCCTTCCATCAGTACTCCCCGGGTGG
          EX90

               AGGCAGCTACCCTGTGCCCTACCTGGGCTCACCTCACTATCCCTATCAGAGGATTGCACCCNAGG
          EX90

               NCAACGNCGAAGTCACCAGCCACTCTTNCCAAAGNCCATCTACTCTTACAGCATCCTCATCTTCA
                                                                  ||||||||||
          EX90                                                    CCTCATCTTCA

               TGGNCCTTAAGAACAGTAAGACCGGAAGGCCTTCCAGTCAAGTGAAATCTTACANTTTTCATGAC
               ||| |||||||||||||||||||||| |||||||||| ||||||||| ||||  |||||||||
          EX90 TGGCCCTTAAGAACAGTAAGACCGGAA-GCCTTCCAGTC-AGTGAAATC-TACA-ATTTCATGAC

               GGGAGCACTTNCCCTTACTTCAAGGTGAGTNTAAAACCCACCCCTTGNCCCCAAGGCCCAGGGTC
                ||||    | || ||||||||||
          EX90 -GGAG----TGCCATTACTTCAAG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD:
```

SEQUENCE FILENAME: a11g
FORWARD PRIMER: CTTCTTTTTCAAGGGGACCA
REVERSE PRIMER: ACCTTGCAATGAGGCACACT
LENGTH OF INSERT: 388

```
ACGTNGCATGCACGCGTGACGTNAAGCTTGGGATCCTCTAGAGCGGCCGCCTACTACTACTACTACTGA
.....................................................................
TGCANCGTACGTGCGCACTGCANTTCGAACCCTAGGAGATCTCGCCGGCGGATGATGATGATGATGACT
---------------------------------------------------------------------
GCGGAATTCGTGAGACCCCAAGCCTGAAATCTCTGNGCTCTTTCCCTTNGTCTCTTGATGCGNCCCCTC
.................
CGCCTTAAGCACTCTGGGGTTCGGACTTTAGAGACNCGAGAAAGGGAANCAGAGAACTACGCNGGGGAG
---------------------------------------------------------------------
CTCTCCTTGAAGTGGCTTCTTTTCACCCCACTCCTTCTTTTTCAAGGGGACCACCCCTTTTCCCATGGT
                                    >>>>>>>>>>>>>>>>>>>
GAGAGGAACTTCACCGAAGAAAAGTGGGGTGAGGAAGAAAAAGTTCCCCTGGTGGGGAAAAGGGTACCA
---------------------------------------------------------------------
GGAGGTNGGGCAGCAGCTGCATCTNTAACCCTTTNCCCAGTGGAGTCTTCGGAACCCTTTGAGGGTGCT

CCTCCANCCCGTCGTCGACGTAGANATTGGGAAANGGGTCACCTCAGAAGCCTTGGGAAACTCCCACGA
---------------------------------------------------------------------
CGNACAGAGGAGTGTGCCTCATTGCAAGGTTGTCTTCAAGTGGCCACATNAGGCTAGGTGGNGGGCAAA
          <<<<<<<<<<<<<<<<<<<
GCNTGTCTCCTCACACGGAGTAACGTTCCAACAGAAGTTCACCGGTGTANTCCGATCCACCNCCCGTTT
---------------------------------------------------------------------
GNGAAAACCTTTNAAGNATCCCTGANCCNGNCCGNTTATAGGCCCTNAATTCGTGGTTTNCCCGAGGAA

CNCTTTTGGAAANTTCNTAGGGACTNGGNCNGGCNAATATCCGGGANTTAAGCACCAAANGGGCTCCTT
---------------------------------------------------------------------
AGNNTTAGGGCCATTAATTTTNGCCCANGGNACAATTTGNTNNATAACGGGGGNNCTTTT

TCNNAATCCCGGTAATTAAAANCGGGTNCCNTGTTAAACNANNTATTGCCCCCNNGAAAA
---------------------------------------------------------------------
```

DUPLICATE REPORT FOR a11g: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.

```
SEQUENCE FILENAME: a12b
FORWARD PRIMER: TGGATAAGTGCATGCAGGAC
REVERSE PRIMER: CCCAGGCAGCGCCTTGAGCTC
LENGTH OF INSERT: 384
--------------------------------------------------------------------
 AATTCGTGAGACCAGAAGCTGGAAGCCGGCAAATTCGAGGACAAGCTTATCCAAAGCGTCATAGCGGCT
 ............
 TTAAGCACTCTGGTCTTCGACCTTCGGCCGTTTAAGCTCCTGTTCGAATAGGTTTCGCAGTATCGCCGA
--------------------------------------------------------------------
 CGCAATTTTGTCCTGGTGCTGTCTGCTGGGGCGCTGGATAAGTGCATGCAGGACCATGACTGCAAGGAC
                                  >>>>>>>>>>>>>>>>>>>>
 GCGTTAAAACAGGACCACGACAGACGACCCCGCGACCTATTCACGTACGTCCTGGTACTGACGTTCCTG
--------------------------------------------------------------------
 TGGGTGCACAAGGAGATTGTGACTGCTTTAAGCTGTGNCAAGAACATTGTGCCCATCATTGATGGCTTT

 ACCCACGTGTTCCTCTAACACTGACGAAATTCGACACNGTTCTTGTAACACGGGTAGTAACTACCGAAA
--------------------------------------------------------------------
 GAGTGGCCTNGAGCTCAAGGCGCTGCCTGGGGATATGCAGGCTGTACTCACCTTCAACGGGNATCAAAA
        <<<<<<<<<<<<<<<<<<<
 CTCACCGGANCTCGAGTTCCGCGACGGACCCCTATACGTCCGACATGAGTGGAAGTTGCCCNTAGTTTT
--------------------------------------------------------------------
DUPLICATE REPORT FOR a12b: EX76 (20)  EX98 (14)
     vs EX76 (score = 133)

              G-AGCGGAATTCGT-GA-GACCAGAAGCTGGAAGCCGGCAAATTCGAGGACAAGCTTATCCAAAG
                |||  |  |||  | ||  |   |||||||||||  |||||||||||||||||||||||||||||
          EX76 TCAGC-GTCTTCATCGACGTGGAGAAGCTGGAA-CCGGCAAATTCGAGGACAAGCTTATCCAAAG

              CGTCATAGCGGCTCGCAATTTTGTCCTGGTGCTGTCTGCTGGGGCGCTGGATAAGTGCATGCAGG
              |||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
          EX76 CGTCATA--GGCTCGCAATTTTGTCCTGGTGCTGTCTGCTGGGGCGCTGGATAAGTGCATG-AGG

              ACCATGACTGCAAGGACTGGGTGCACAAGGAGATTGTGACTGCTTTAAGCTGTGNCAAGAACATT
               |||||||||||||||||||||||||||||
          EX76 -CCATGACTGCAAGGACTGGGTGCACAAGACCTGGGA

       vs EX98 (score = 166)
               GACTGCAAGGACTGGGTGCACAAGGAGATTGTGACTGCTTTAAGCTGTGNCAAGAACATTGTGCC
                                          |||||||||||||||||||||||||| ||||||||||||||||
          EX98                    GACCCAGCAGAGATTGTGACTGCTTTAAGCTGTGGCAAGAACATTGTGCC

              CATCATTGATGGCTTTGAGTGGCCTNGA-GCTCAAGGCGCTGCCTGGGGATATGCAGGCTGTACT
              |||||||||||||||||||||||||| || || | | |||||||||| |||||||||||||||||
          EX98 CATCATTGATGGCTTTGAGTGGCCT-GAAGC-CTA-GCGCTGCCTGAGGATATGCAGGCTGTACT

              CACCTTCAACGGGNATCAAAATTGGTCCCATGNGTACCAGGAGGGCCACCATTNGAGAAGATAAT
              |||||||||| || ||| ||| |||||||||| |||||| || |||||||| | |||||||| ||
          EX98 CACCTTCAAC-GGCATC-AAA-TGGTCCCATGAGTACCA-GA-GGCCACCA-TCGAGAAGATCAT

              CCGNTTTCTACAAGGGCCTGCCNTTNTAANGACTTCTNTNNNCGGNTCGGGTTACCANTTTNGGG
              |||  ||  ||||  ||||||  |||   | | |  |||  ||   |   |||  ||  ||  |||||  |||   ||
          EX98 CCGCTTCCTAC-AGGGCC-GCCCCTCTCAGGAC-TCCCT-GCCGGATC-GGATACCAGTTT--GG

              GGGGGCTACGNC
               |||  |||||
          EX98 AGGGACTACGCAATGG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,21,24,25
END OF RECORD.
```

```
SEQUENCE FILENAME: a12c
FORWARD PRIMER: AGAGGTTTGCCTAGGGATGC
REVERSE PRIMER: TGCCCCTGAGGCTCTGGCC
LENGTH OF INSERT: 228
--------------------------------------------------------------------
 CGGAATTCGTGAGACCGACTCCCCCTGCTGCCCCTCCCACTAGCTCCAGAGGTTTGCCTAGGGATGCTG
 ..............                                  >>>>>>>>>>>>>>>>>>>>
 GCCTTAAGCACTCTGGCTGAGGGGGACGACGGGGAGGGTGATCGAGGTCTCCAAACGGATCCCTACGAC
--------------------------------------------------------------------
 GACACCTTTCCTCTCGAGACTGACTCAGGGTCCAGAGCCACCGAATCCATCCCCAAGTGGAGATGATGT

 CTGTGGAAAGGAGAGCTCTGACTGAGTCCCAGGTCTCGGTGGCTTAGGTAGGGGTTCACCTCTACTACA
--------------------------------------------------------------------
 CGGAGAGGTTACAGCACAGGGTCTAGGCCAGAGCCTCAGGGGCAGCATTAGGGGTCAGGGTGTCCCACT
                          <<<<<<<<<<<<<<<<<<
 GCCTCTCCAATGTCGTGTCCCAGATCCGGTCTCGGAGTCCCCGTCGTAATCCCCAGTCCCACAGGGTGA
--------------------------------------------------------------------
DUPLICATE REPORT FOR a12c: a10a (6)  a10c (13)
     vs a10a (score = 138)
              GTCAGGGTGTCCCACTGTAGGAATAGTCTGCTTTNNTTATGCATCACCAGCCCGGTCATNCACGA
                ||||||||||||||||||||||||||||||||  |||||||||||||| ||||||||  |||||
          a10a  CAGGGTGTCCCACTGTAGGAATAGTCTGCTTT-GTTATGCATCACCAG-CCGGTCATCCACGA

              A-TAGAAAGCTNTCAGACTTGGGTCTAATACGGGTGACGGAATCATCTCAATTGATTAGGTTCTC
              | ||| |||||| |||||| |||||||| |||||||||| |||||||||| | ||||||| | |||
          a10a AGTAG-AAGCTGTCAGAC-TGGGTCTCATACGGGTGAC-GAATCATCTC-ACTGATTAGCTCCTC

              GGAGAGANGGAGGGGNAAGTNATANGATGTTNTTCACAGGGTNGTTCTTTGCNTGCTGGGGGGAT
              ||||||| ||| ||| |||| ||| |||| |  ||||| ||| |||| |||| ||||  ||||||
          a10a GGAGAGA-GGA-GGG-AAGTCATA-GATG-TGCTCACA-GGT-GTTC-TTGCTTGCT--GGGGAT


     vs a10c (score = 102)
                                                                  ACGTNCGCATGTCAC
                                                                        |||
          a10c GACCATTTCCTGTCCCTAACCTGATCTAGTCACGGTCAAGAAGGGGGCACTACTCTGAAACTCAG

              GC---GTG-ACGTAA--AG-CTT---GGNA-TCCTC-TA--G-AGC-GGCCGTNTACT-ACTACT
              ||   ||| || ||   || | |   || | ||||| |   | ||  |||| | | || |  ||
          a10c GCCCAGTGGACCTAGCGAGGCATCCAGG-AGTCCTCCTCTTGCAGGAGGCC-T-TCCTGA--AC-

              ACTGANCGGAATTCGTGAGACC-GACTCCCCCTGCTGCCCCTCCCACTAGCTCCAGAGGTTTGCC
              | ||   || | |   || || | ||||||||||||||||||||||||||||||||||||||||
          a10c A-TG---GGGA--C---AG-CCAGCCTCCCCCTGCTGCCCCTCCCACTAGCTCCAGAGGTTTGCC

              TAGGGATGCTGGACACCTTTCCTCTCGAGACTGACT-CAGGGTCCAGAGCCACC-G-AATCCATC
              |||||||||||||| |||||||||||||||||| || |||||||||||||| | ||||  |
          a10c TAGGGATGCTGGACAACTTTCCTCTCGAGACTGACTTCAAGGTCCAGAGCCACCCGGAATCTTTT

              CCCAAGTGGAGATGATGTCGGAGAGGTTACAGCACAGGGTCTAGGCCAG-A-GCCTCAGG-GGCA
                | || || |  |   |  | | ||| | | |  | || ||| | | |  |  || ||
          a10c GTC-TGTTGA-ACCA--AC--A-A-CTTATA--A-A---T-TA-GCCTGGAAG-GT-TGGTGG-G

              GCATTAGGGGTCAG
              |||  ||||| ||
          a10c GCA--AGGGGACAA

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
--------------------------------------------------------------------
 GCTCCCAGAAACACTTCAACCACAGCTCCTTCTTCAGAGATCTTGGAACTCCAGCATGAAAATGGGGAA

 CGAGGGTCTTTGTGAAGTTGGTGTCGAGGAAGAAGTCTCTAGAACCTTGAGGTCGTACTTTTACCCCTT
--------------------------------------------------------------------
 GCCTTGGTCAAGGCGGGTTTTGGACCTGCGTCATTTGCTGCTCATCTCTCTGCTCCCTCACTCCCCACC
                                                  >>>>>>>>>>>>>>>>>>>>
 CGGAACCAGTTCCGCCCAAAACCTGGACGCAGTAAACGACGAGTAGAGAGACGAGGGAGTGAGGGGTGG
--------------------------------------------------------------------
 CCCACTCAGTAAAACTGAGAACACCTNGTCTGCAGCGACCTTGGTCTTCCTGTGCCCCCCACAGGCTTG

 GGGTGAGTCATTTTGACTCTTGTGGANCAGACGTCGCTGGAACCAGAAGGACACGGGGGGTGTCCGAAC
--------------------------------------------------------------------
 CTCTTTGGGGATTGCTTCAGCCCCAAGGTGGCTCCCAAAATGNCTGCTCTTCAGGGACCCAAAGAGTTG
                                  <<<<<<<<<<<<<<<<<<<
 GAGAAACCCCTAACGAAGTCGGGGTTCCACCGAGGGTTTTACNGACGAGAAGTCCCTGGGTTTCTCAAC
--------------------------------------------------------------------
 AANGGGACTNCATTGACAATNGGGAGCTTGGGCCCGATTNGCCCCTGGGGGNTNNGNTTCCGTTGGAGG

 TTNCCCTGANGTAACTGTTANCCCTCGAACCCGGGCTAANCGGGGACCCCCNANNCNAAGGCAACCTCC
--------------------------------------------------------------------
 AAACCTTTNGCGGGTTNTTAACGGAATTT

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

              AAGTTGGGGCTCCCAGAAACACTTCAACCACAGCTCCTTCTTCAGAGATCTTGGAACTCCAGCAT
         a10e

              GAAAATGGGGAAGCCTTGGTCAAGGCGGGTTTTGGACCTGCGTCATTTGCTGCTCATCTCTCTGC
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         a10e   AAATGGGGAAGCCTTGGTCAAGGCGGGTTTTGGACCTGCGTCATTTGCTGCTCATCTCTCTGC

              TCCCTCACTCCCCACCCCCACTCAGTAAAACTGAGAACACCTNGTCTGCAGCGACCTTGGTCTTC
              |||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||||||
         a10e TCCCTCACTCCCCACCCCCACTCAGTAAAACTGAGAACACCT-GTCTGCAGCGACCATGGTCTTT

              CTGTGCCCCCCCACAGGCTTGCTCTTTGGGGATTGCTTCAGCCCCAAGGTGGCTCCCAAAATGNCT
              |||||||||||||||||| |||||||| || ||||| |||||||| ||||| ||||| |||||| ||
         a10e CTGTGCCCCCCCACAGGC-TGCTCTTT-GGAATTGC-TCAGCCCC-AGGTGCCTCCC-AAATGCCT

              GCTCTTCAGGGACCCAAAGAGTTGAANGGGACTNCATTGACAATNGGGAGCTTGGGCCCGATTNG
              ||||  ||| ||| |||||||| ||||  ||||| || ||||  | |||||| | || ||||||
         a10e GCTC-TCA-GGA-CCAAAGAG-TGAA--GGACT-CA-TGAC-CTGGGGAGC-T-GGACCGATTTC

              CCCCTGGGGGNTNNGNTTCCGTTGGAGGAAACCTTTNGCGGGTTNTTAACGGAATTT
         a10e TCAAGGG

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

```
SEQUENCE FILENAME: a7b
FORWARD PRIMER: CTCGGGTAGTTTCTTGGCC
REVERSE PRIMER: GGACCATTTCCTGTCCCTTGCC
LENGTH OF INSERT: 153
--------------------------------------------------------------------
 TCCTCTAGAGCGGCCGAGTACTACTACTACTGAGCGGAATTCGTGAGACCCCTCGGGTAGTTTCTTGGC
 ..................................................  >>>>>>>>>>>>>>>>>>
 AGGAGATCTCGCCGGCTCATGATGATGATGACTCGCCTTAAGCACTCTGGGGAGCCCATCAAAGAACCG
--------------------------------------------------------------------
 CCCTGATGGAGGAGGCCAGGGCTCAGAGCAGGGTACCCACTGTGGATCTTTGTCTGTNAACCGACAACT
 >
 GGGACTACCTCCTCCGGTCCCGAGTCTCGTCCCATGGGTGACACCTAGAAACAGACANTTGGCTGTTGA
--------------------------------------------------------------------
 ATAAATAGCCTGGAGGTNGTNGGCAAGGGACAGGAAATGGTCCCTAACCTGATCTAGTCACGGTCGGTC
                        <<<<<<<<<<<<<<<<<<<<<<                    ....
 TATTTATCGGACCTCCANCANCCGTTCCCTGTCCTTTACCAGGGATTGGACTAGATCAGTGCCAGCCAG
--------------------------------------------------------------------
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: CGGTTAAGACCATGAGCGCCCC
LENGTH OF INSERT: 271
--------------------------------------------------------------------
 TGAGCGGAATTCGTGAGACCCGCCATATTCAGGAATAAAAGTGTTTTCTGGTCTTTAGAGAAAAGAACA
 ..................                                  >>>>>>>>>>
 ACTCGCCTTAAGCACTCTGGGCGGTATAAGTCCTTATTTTCACAAAAGACCAGAAATCTCTTTTCTTGT
--------------------------------------------------------------------
 GAGGCCAAGGCAGAAAGGGTGTGGTGGACACAGCAGGTACAGCAGGGTAAAGAGGAGGCAAGAAGAGTC
 >>>>>>>>>>>
 CTCCGGTTCCGTCTTTCCCACACCACCTGTGTCGTCCATGTCGTCCCATTTCTCCTCCGTTCTTCTCAG
--------------------------------------------------------------------
 CTGGGCTCAGAAGGAAACCAAGAGAGCCAAGCATTGCCCAGGGGCGCTCATGGTCTTAACCGTTACCAC
                                        <<<<<<<<<<<<<<<<<<<<<<
 GACCCGAGTCTTCCTTTGGTTCTCTCGGTTCGTAACGGGTCCCCGCGAGTACCAGAATTGGCAATGGTG
--------------------------------------------------------------------
 TCTTCTGTTGCTGTTTTTAAGTCAGGGTCCATGTAGCCTANGGCTAACCTCAAATTTCACTATGTNGCA

 AGAAGACAACGACAAAAATTCAGTCCCAGGTACATCGGATNCCGATTGGAGTTTAAAGTGATACANCGT
--------------------------------------------------------------------
DUPLICATE REPORT FOR a7c: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
```

```
SEQUENCE FILENAME: a7e#2
FORWARD PRIMER: TTTCTTTGAGCAAATGGGCT
REVERSE PRIMER: GTTAAGGGGTCACCTGACAC
LENGTH OF INSERT: 193
--------------------------------------------------------------------
 GCGGCCGNNTACTACTACTACTGAGCGGAATTCGTGAGACCCCCCTCTCTGCATCCTTTCTTTGAGCAA
 ........................................                >>>>>>>>>>>>>
 CGCCGGCNNATGATGATGATGACTCGCCTTAAGCACTCTGGGGGGAGAGACGTAGGAAAGAAACTCGTT
--------------------------------------------------------------------
 ATGGGCTCCGTGGAACTGGCATGATAGTTCCCCTTTGACAAGGTGTGTGTGCTTGAAGGTACTGCTAAT
 >>>>>>>
 TACCCGAGGCACCTTGACCGTACTATCAAGGGGAAACTGTTCCACACACACGAACTTCCATGACGATTA
--------------------------------------------------------------------
 CATCAGAATACACTCCTAGTTAGCCCAGGCACCTCCTTAGCATCTTCAACATAGTGTCAGGTGACCCCT
                                                 <<<<<<<<<<<<<<<<
 GTAGTCTTATGTGAGGATCAATCGGGTCCGTGGAGGAATCGTAGAAGTTGTATCACAGTCCACTGGGGA
--------------------------------------------------------------------
 TAACTCAGGGGTTAGCACAGGATAAATGGTCTCACGNATTCCGGNGGANGGNGNNNNNNNANNGNNNTT
 <<<<                          ..........................................
 ATTGAGTCCCCAATCGTGTCCTATTTACCAGAGTGCNTAAGGCCNCCTNCCNCNNNNNNNTNNCNNNAA
--------------------------------------------------------------------
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: GGCACTGGCTTGTGTTGGAGT
LENGTH OF INSERT: 394
--------------------------------------------------------------------
 GGCCGNNTACTACTACTACTGAGCGGAATTCGTGAGACCCTCTTTTGGAGGCCTCAAAGTGTCTGATAT
 ..................................                    >>>>>>>>
 CCGGCNNATGATGATGATGACTCGCCTTAAGCACTCTGGGAGAAAACCTCCGGAGTTTCACAGACTATA
--------------------------------------------------------------------
 GGCCCGTGCAGGATTCCTGCTCAATATCATCGGAGTGCTGACTATCACATTATCCATAAACAGCTGGAG
 >>>>>>>>>>
 CCGGGCACGTCCTAAGGACGAGTTATAGTAGCCTCACGACTGATAGTGTAATAGGTATTTGTCGACCTC
--------------------------------------------------------------------
 TATCCCTATCTTCAAGCTGGACACATTTCCCACCTGGGCCTACTCCAACACAAGCCAGTGCCTTCTCAA
                                          <<<<<<<<<<<<<<<<<<<<
 ATAGGGATAGAAGTTCGACCTGTGTAAAGGGTGGACCCGGATGAGGTTGTGTTCGGTCACGGAAGAGTT
--------------------------------------------------------------------
 TCCGCCTAATTCCACTGTACCAGGCCACTAGACTATGGTACAACCTAGCTCTGCTAGGGGAAAAGCCCG

 AGGCGGATTAAGGTGACATGGTCCGGTGATCTGATACCATGTTGGATCGAGACGATCCCCTTTTCGGGC
--------------------------------------------------------------------
 CTGGTGCCACTCCANGGGAGCCAGGAGAGGGGGCGNTCCAAACTTCCAAAGTNCCCGGGGACAGAGGGT

 GACCACGGTGAGGTNCCCTCGGTCCTCTCCCCCGCNAGGTTTGAAGGTTTCANGGGCCCCTGTCTCCCA
--------------------------------------------------------------------
 NCAGAATTCGGGNTNGGGGGCATGGGAAATCATGGNATTTCGGNGTTGTTTTNGCCTAACAATNCTGNT

 NGTCTTAAGCCCNANCCCCCGTACCCTTTAGTACCNTAAAGCCNCAACAAAANCGGATTGTTANGACNA
--------------------------------------------------------------------
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: AGCAGGGTGCTCTGTTCTCT
REVERSE PRIMER: TGGAATGAAGAGGAGGCC
LENGTH OF INSERT: 306
------------------------------------------------------------------------
 GCGGAATTCGTGAGACCAGCTTTCCAAGCTGTGAAGCGAGTCGGAGCAGGGTGCTCTGTTCTCTGCTGT
 .................                                >>>>>>>>>>>>>>>>>>>>
 CGCCTTAAGCACTCTGGTCGAAAGGTTCGACACTTCGCTCAGCCTCGTCCCACGAGACAAGAGACGACA
------------------------------------------------------------------------
 GGCCTGTGTGAATAGCTTGGCCATGTTGATGGCTCCGGATCTTCAACTCCATCTACCCGGCCACTCTGA

 CCGGACACACTTATCGAACCGGTACAACTACCGAGGCCTAGAAGTTGAGGTAGATGGGCCGGTGAGACT
------------------------------------------------------------------------
 ACTTCATGAAGGGCTTCCCCTTCCTNCTGGGAGCTGGCCTCCTCTTCATTCCAGCCATTCTGATTGGGG
                                  <<<<<<<<<<<<<<<<<<
 TGAAGTACTTCCCGAAGGGGAAGGANGACCCTCGACCGGAGGAGAAGTAAGGTCGGTAAGACTAACCCC
------------------------------------------------------------------------
 TACTGGAAAAGGTTAATCCACACCCGGAGTTCCAGCAGTTTCCTNAGAGCCCCTGATCTGCCCGAAGCC

 ATGACCTTTTCCAATTAGGTGTGGGCCTCAAGGTCGTCAAAGGANTCTCGGGGACTAGACGGGCTTCGG
------------------------------------------------------------------------
 ATGAGGAGCAATGGACCCACATTGACGATGTATTGCCCAAAGCCNAGGGTNTCACGAATTCCGNGGAAA
                                                  ....................
 TACTCCTCGTTACCTGGGTGTAACTGCTACATAACGGGTTTCGGNTCCCANAGTGCTTAAGGCNCCTTT
------------------------------------------------------------------------

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

              CA--T--TCCAGCCATTCTGATTGGGGTACTGGAAAAGGTTAATCCACACCCGGAGTTCCAGCAG
              ||  |  |  ||| | || ||  |||||||||||||||||||||||||||||| || ||||||
       ABX7_3 CAAGTGGT-GAGCAACTC-GAGAGGGGTACTGGAAAAGGTTAATCCACACCCGGA-TT-CAGCAG

              TTTCCTNAGAGCCCCTGATCTGCCCGAAGCCATGAGGAGCAATGGACCCACATTGACGATGTATT
              |||||| |||||||||||||||||||||| |||||||||||||||||||||||| |||||||
       ABX7_3 TTTCCTCAGAGCCCCTGATCTGCCCGAACCCATGAGGAGCAATGGACCCACACTGACGATA

              GCCCAAAGCCNAGGGTNTCACGAATTCCGNGGAAANNANNTTNGANTNNNCCNGGGAATTNTGGA

     vs a10d (score = 167)
              CTGAGCGGAATTCGTGAGACCAGC-TTTCCAA-GCTG-TGAAGCGAGTCGGAGCAGGGTGCTCTG
                                  ||| | |||| || | || || |||||||||| || |||||||||
         a10d NGCCTTGTCAANNCNGGGCCCAACCTTTCNAAAGNTGGTG-AGCGAGTCGGANCANGGTGCTCTG

              TTC-TCTGCTGTGGCCTGTGTGAATAGCTTGGCCATGTTGA-TGGCTCC-GG-ATCTTCAACTCC
              ||| ||||||||| |||||||||||||||| ||||||||| | ||||| || ||||||||| ||
         a10d TTCCTCTGCTGTGNCCTGTGTGAATAGCTTGNCCATGTTGAAT-GCTCCCGGCATCTTCAACNCC

              ATCTACCCGGCCACTCT-GAACTTCATGAAGGGCTTCCCCTTCCTNCTGGGAGCTGGCCTCCTCT
              |||||| |||||||||| |||||||||||||||||||||||||| ||||||||||||||||||
         a10d ATCTACACGGCCACTCTTGAACTTCATGAAGGGCTTCCCCTTCCTCCTGGGAGCTGGCCTCCTCT

              TCATTCCAGCCATTCTGATTGGGGTACTGGAAAAGG-T-TAA--T--CCA-CACC-C--GGAGTT
              |||||||||||||||||||| ||||| |||  | || | | |  |  ||  | |  |  |||  |
         a10d TCATTCCAGCCATTCTGATT-GGGTAATGG--AGGGCTCTGACCTTGCCTGCTCATCATGGA-AT
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: GCTCCCACCACATTCTAAAG
REVERSE PRIMER: CCTTTAGCCAAAAACCTTATCA
LENGTH OF INSERT: 368
--------------------------------------------------------------------
 ATAAGGCCTCAACCAGAGAGGGCCGGGTCCTGATTCTCAAGATGGAGATAGTGAGAACTCTATTGATGC

 TATTCCGGAGTTGGTCTCTCCCGGCCCAGGACTAAGAGTTCTACCTCTATCACTCTTGAGATAACTACG
--------------------------------------------------------------------
 AACCCTGGAAGGAAATCTCCTGGCTGAGTGACAAGGAGCTACAACGAGTGGACAATGGGGGACAACTGC

 TTGGGACCTTCCTTTAGAGGACCGACTCACTGTTCCTCGATGTTGCTCACCTGTTACCCCCTGTTGACG
--------------------------------------------------------------------
 TGGCTTTCTGGGTGCCAAGCTGTGGCTCCCACCACATTCTAAAGGCTTGAGAGAAAGCAGTCATGGCTC
                          >>>>>>>>>>>>>>>>>>>>
 ACCGAAAGACCCACGGTTCGACACCGAGGGTGGTGTAAGATTTCCGAACTCTCTTTCGTCAGTACCGAG
--------------------------------------------------------------------
 CAGAAACAGCAAGAAGGTCCTGTCTTTGGGCCACAACTGACCCACCTNGGGTGTTCAACTTGTTTTTCC

 GTCTTTGTCGTTCTTCCAGGACAGAAACCCGGTGTTGACTGGGTGGANCCCACAAGTTGAACAAAAAGG
--------------------------------------------------------------------
 CCCCTTCTTNCCTTCCCTTGATAAGGTTTTTGGCTAAAGGACTAGGAATTGGGCCTATTTTNGGAAGTA
                     <<<<<<<<<<<<<<<<<<<<<
 GGGGAAGAANGGAAGGGAACTATTCCAAAAACCGATTTCCTGATCCTTAACCCGGATAAAANCCTTCAT
--------------------------------------------------------------------
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: ATTGCCCGGCTATATGTGG
REVERSE PRIMER: GTCCCAGCNAGAGGGGTGCG
LENGTH OF INSERT: 287
--------------------------------------------------------------------
 CGGAATTCGTGAGACCAGCCAAAGTCCACCAGGCTTTTNACGCTCCGCCATCATCGATCCATTGCCCGG
 ..............                                          >>>>>>>>>
 GCCTTAAGCACTCTGGTCGGTTTCAGGTGGTCCGAAAANTGCGAGGCGGTAGTAGCTAGGTAACGGGCC
--------------------------------------------------------------------
 CTATATGTGGTTCCAGCCCCAGAAAAGTCCAGGATGCATCTAGCCCTGTACTCTTTGGCCATATTCGTG
 >>>>>>>>>>
 GATATACACCAAGGTCGGGGTCTTTTCAGGTCCTACGTAGATCGGGACATGAGAAACCGGTATAAGCAC
--------------------------------------------------------------------
 GTGGTTACTGTGCATTTTGGGGTCAAGATATCCTGACACTCTATGAGCTGAGCGCACCCCTCTNGCTGG
                                                  <<<<<<<<<<<<<<<<<
 CACCAATGACACGTAAAACCCCAGTTCTATAGGACTGTGAGATACTCGACTCGCGTGGGGAGANCGACC
--------------------------------------------------------------------
 GACTCCAAGCTGATTGGCTGTGGCTCAGCAGCCCAGCACCTACCCTANCTCACCAGCCTTCTGGGCCTA
 <<<
 CTGAGGTTCGACTAACCGACACCGAGTCGTCGGGTCGTGGATGGGATNGAGTGGTCGGAAGACCCGGAT
--------------------------------------------------------------------
 AGGNTTCTGCAATTCTGCCTGGCCGACGGTCTCACGNATNCCGGNTTTTGNGNNGTAGNNTGGNNTTTG
                              ........................................
 TCCNAAGACGTTAAGACGGACCGGCTGCCAGAGTGCNTANGGCCNAAAACNCNNCATCNNACCNNAAAC
--------------------------------------------------------------------
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: CTGATTTTGGACAATGTTTGGC
REVERSE PRIMER: CTGCCTTGGCCCTGAACGAT
LENGTH OF INSERT: 249
--------------------------------------------------------------------
 AACAATGTTACCTCCCTCAAGGTTATTGTGAAAAATAAAGCCAGTCTTCTGATTTTGGACAATGTTTGG
                                                >>>>>>>>>>>>>>>>>>>>>
 TTGTTACAATGGAGGGAGTTCCAATAACACTTTTTATTTCGGTCAGAAGACTAAAACCTGTTACAAACC
--------------------------------------------------------------------
 CATGCTGTGCACGCTGTCAACAGGGTTCCTTTATCAACAGTGTTGGAAGCTACATAGGACTGTTCAGAG
 >
 GTACGACACGTGCGACAGTTGTCCCAAGGAAATAGTTGTCACAACCTTCGATGTATCCTGACAAGTCTC
--------------------------------------------------------------------
 TGGCTGCCTGCAGCAGNATAAGTGATCGTTCAGGGCCAAGGCAGGCACCGCTACAAGACAGTGCACAAG
                           <<<<<<<<<<<<<<<<<<<
 ACCGACGGACGTCGTCNTATTCACTAGCAAGTCCCGGTTCCGTCCGTGGCGATGTTCTGTCACGTGTTC
--------------------------------------------------------------------
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: ATTCACCTCTCTGGGCTTGT
REVERSE PRIMER: TTCCCAAAATCATTTGCGTT
LENGTH OF INSERT: 280
--------------------------------------------------------------------
 TNGAGCGGAATTCGTGAGACCGCGGGGCTTAGAAGGGTAGAGGGTGGGGCTTGGCTGCCTGTAGGATGG
 ......................
 ANCTCGCCTTAAGCACTCTGGCGCCCCGAATCTTCCCATCTCCCACCCCGAACCGACGGACATCCTACC
--------------------------------------------------------------------
 AGCGGGCAAGNTGGAGGAGAATGTAAAGATTCACCTCTCTGGGCTTGTACCNTCAAATCTCTCTAGGGA
                                >>>>>>>>>>>>>>>>>>>>
 TCGCCCGTTCNACCTCCTCTTACATTTCTAAGTGGAGAGACCCGAACATGGNAGTTTAGAGAGATCCCT
--------------------------------------------------------------------
 CATGCTTAGCTCTGAGGTCTTCCATTCCATGGAAAGTCTGTCTCCTGAAGTGAAAGCTAAATTCCTGAT

 GTACGAATCGAGACTCCAGAAGGTAAGGTACCTTTCAGACAGAGGACTTCACTTTCGATTTAAGGACTA
--------------------------------------------------------------------
 AAAACGCAAATGATTTTGGGAACGTTTAAGCCTCCCCAGTCCCTCACAGTCAAAGTACANCAATGGAAC
   <<<<<<<<<<<<<<<<<<<
 TTTTGCGTTTACTAAAACCCTTGCAAATTCGGAGGGGTCAGGGAGTGTCAGTTTCATGTNGTTACCTTG
--------------------------------------------------------------------

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: CTCCGTGGAACTGNCATGAT
LENGTH OF INSERT: 194
--------------------------------------------------------------------
 GGAATTCGTGAGACCATTTATCCTGTGCTAACCCCTGAGTTAAGGGGTCACCTGACACTATGTTGAAGA
 ..............      >>>>>>>>>>>>>>>>>>>>
 CCTTAAGCACTCTGGTAAATAGGACACGATTGGGGACTCAATTCCCCAGTGGACTGTGATACAACTTCT
--------------------------------------------------------------------
 TGCTAAGGAGGTGCCTGGGCTAACTAGGAGTGTATTCTGATGATTAGCAGTACCTTCAAGCACACACAC

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

              GAGCGGAATTCGTGAGACCATTTATCCTGTGCTAACCCCTGAGTTAAGGGGTCACCTGACACTAT
                                  ||||||||||||||||||||||||||||||||||||||||||||
        a7e#2                     ATTTATCCTGTGCTAACCCCTGAGTTAAGGGGTCACCTGACACTAT

              GTTGAAGATGCTAAGGAGGTGCCTGGGCTAACTAGGAGTGTATTCTGATGATTAGCAGTACCTTC
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        a7e#2 GTTGAAGATGCTAAGGAGGTGCCTGGGCTAACTAGGAGTGTATTCTGATGATTAGCAGTACCTTC

              AAGCACACACACCTTGTCAAAGGGGAACTATCATGNCAGTTCCACGGAGCCCATTTGCTCAAAGA
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
        a7e#2 AAGCACACACACCTTGTCAAAGGGGAACTATCATGCCAGTTCCACGGAGCCCATTTGCTCAAAGA

              AAGGATGCAAGNGAGGGG
              |||||||| || ||||||
        a7e#2 AAGGATGC-AGAGAGGGG

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

```
SEQUENCE FILENAME: c5a
FORWARD PRIMER: CAATCTAGAAGCTGTGGTGCC
REVERSE PRIMER: TGTCTGTCTGTCTGGGCTTCTG
LENGTH OF INSERT: 406
--------------------------------------------------------------------
 ACTACTACTACTGAGCGGAATTCGTGAGACCTGCTCTTTCTCTCTCCCCACCAATCTAGAAGCTGTGGT
 ..............................                >>>>>>>>>>>>>>>>>>
 TGATGATGATGACTCGCCTTAAGCACTCTGGACGAGAAAGAGAGAGGGGTGGTTAGATCTTCGACACCA
--------------------------------------------------------------------
 GCCTGGTGGGGGGCTTCATCTGTTTGCTTCAGTGTTGTGTTCAGATCTCCACGGCACCAAGTAAGGGCA
 >>>
 CGGACCACCCCCCGAAGTAGACAAACGAAGTCACAACACAAGTCTAGAGGTGCCGTGGTTCATTCCCGT
--------------------------------------------------------------------
 CTCCTTTTTCACAGGCAATAAAACAGAAGCCCAGACAGACAGACAGGACAGACAAGACAGACAGGGAAG
                          <<<<<<<<<<<<<<<<<<<<<<
 GAGGAAAAAGTGTCCGTTATTTTGTCTTCGGGTCTGTCTGTCTGTCCTGTCTGTTCTGTCTGTCCCTTC
--------------------------------------------------------------------
 TNTNAGCTNGCTCAGCAACAGNTGCAAGATGTCTGTTCTCCAGCTCAAGTTCCTACAACAAGGGGNACT

 ANANTCGANCGAGTCGTTGTCNACGTTCTACAGACAAGAGGTCGAGTTCAAGGATGTTGTTCCCCNTGA
--------------------------------------------------------------------
 TGAAGAGCACAGGNCCACACCCAGGGGGCNTTNCAGTCTTCAGGGCTCCTNTNTGNNCTCTAATTNCCA

 ACTTCTCGTGTCCNGGTGTGGGTCCCCCGNAANGTCAGAAGTCCCGAGGANANACNNGAGATTAANGGT
--------------------------------------------------------------------
 NCNTNTAAAAGNGAANTTGCAGANTTCAAGNTNCCAAGTNCCTTGGATTGTTNACANTAGGTTCCAGGG

 NGNANATTTTCNCTTNAACGTCTNAAGTTCNANGGTTCANGGAACCTAACAANTGTNATCCAAGGTCCC
--------------------------------------------------------------------
 AAAAGGTTTCCTNAGGACAAGGT

 TTTTCCAAAGGANTCCTGTTCCA
--------------------------------------------------------------------

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: CAACNCCNCCTTGAGCTCCA
LENGTH OF INSERT: 419
--------------------------------------------------------------------
 GGAATTCGTGAGACCCAGGACGCTNCTTGGTTTCNCTGCAGCTATNGCCAAGGACATCTTTGGGCGAGG
 ..............                                  >>>>>>>>>>>>>>>>>>>
 CCTTAAGCACTCTGGGTCCTGCGANGAACCAAAGNGACGTCGATANCGGTTCCTGTAGAAACCCGCTCC
--------------------------------------------------------------------
 CAGGGCTGCATTTTTNATTGAACTGAACAAGCTCCGAGTNTTNGACCCAGAGGTCACCCAGCAGACTGT
                                                                     <
 GTCCCGACGTAAAAANTAACTTGACTTGTTCGAGGCTCANAANCTGGGTCTCCAGTGGGTCGTCTGACA
--------------------------------------------------------------------
 GGAGCTCAAGGNGGNGTTGCAAGGACTTTTGTGGACAGTGAGNACCATCTNGAGACACAAGTTTTTGTT
 <<<<<<<<<<<<<<<<<<<
 CCTCGAGTTCCNCCNCAACGTTCCTGAAAACACCTGTCACTCNTGGTAGANCTCTGTGTTCAAAAACAA
--------------------------------------------------------------------

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

            GAGCGGAATTCGTGAGACCCAGGACGCTNCTTGGTTTCNCTGCAGCTATNGCC-AA-GGACATCT
                                                                 |||| ||| || ||||||||
         C5                                                GAAACCAGCTATTGCCCAAAGGACATCT

            TTGGGCGAGGCA-GGGCTGCATTTTTNATTGAACTGAAC-AAGCTCC-GAGTNTTNGACCCAGAG
            || ||| | ||| ||||| || |||| | |||||||||| ||||||| |||| || || ||||||
         C5 TT-GGC-A-GCAAGGGCTTCA-TTTTGA-TGAACTGAACCAAGCTCCCGAGTGTTGGA-CCAGAG

            GTCACCC-AGCA-GACT-GTGGAGCTCAAGGNGGNGTTGCAAGGACTTTTGTGGACAG--T-GAG
            ||||||| | || |||| |||||||||||  |  | ||||||| | ||||||||||   | | |
         C5 GTCACCCCAACAAGACTTGTGGAGCTCAAGAAGAAG-TGCAAGG-C-TTTGTGGACAAAATTG-G

            NACCA-TCTNGAGACACAAGTTTTTGTTCTTNCT--TTNGAGCA-GTTTNCTNACCTNACCTTNN
              ||| | |  ||| | || || ||| |  | ||  || ||||  |||
         C5 --CCAGT-TTCAGA-A-AA-TTGTTGGTGGT-CTAATT-GAGCTTGTTGATCAGCTTGCCAAAGA


NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
```

```
SEQUENCE FILENAME: c5d
FORWARD PRIMER: CAGTAAAAAAGGGAAGGCTCCA
REVERSE PRIMER: CCCNAGTCNCCCTCTAGG
LENGTH OF INSERT: 279
--------------------------------------------------------------------
 CACATCATCACTGGAGTACAGTAAAAAAGGGAAGGCTCCACAAATGATCACCCTGGGGTGAAAGAAGTA
                    >>>>>>>>>>>>>>>>>>>>>>
 GTGTAGTAGTGACCTCATGTCATTTTTTCCCTTCCGAGGTGTTTACTAGTGGGACCCCACTTTCTTCAT
--------------------------------------------------------------------
 TAAGGTCCTTGTGGTCTTGGGGTTCAGGTAACAGGTGGGCTCTTTTGTCTTCAAGGAGGTGCCCGGAGC

 ATTCCAGGAACACCAGAACCCCAAGTCCATTGTCCACCCGAGAAAACAGAAGTTCCTCCACGGGCCTCG
--------------------------------------------------------------------
 GGTAATCGCATGGGCCCTAGAGGGNGACTNGGGTGAAGCTTGGCTTGGAGGTTGCGTTTGAGCAAGAAA
                 <<<<<<<<<<<<<<<<<<
 CCATTAGCGTACCCGGGATCTCCCNCTGANCCCACTTCGAACCGAACCTCCAACGCAAACTCGTTCTTT
--------------------------------------------------------------------
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: TCCCAACATTCACTCAGAGGA
REVERSE PRIMER: TTACACCCAGGTTGGGCAAA
LENGTH OF INSERT: 381
--------------------------------------------------------------------
 AATTCGTGAGACCAATGACCTGAGGCTTTTATTCGTCTCTAGGAAAAGGAAACAGGGCACCAAGGGATG
 ............
 TTAAGCACTCTGGTTACTGGACTCCGAAAATAAGCAGAGATCCTTTTCCTTTGTCCCGTGGTTCCCTAC
--------------------------------------------------------------------
 GGTGCTAAGCAGACTTTCTTTGAAAACCTCTTTCCCAACATTCACTCAGAGGAACAAGTGCCAGACACA
                                    >>>>>>>>>>>>>>>>>>>>>
 CCACGATTCGTCTGAAAGAAACTTTTGGAGAAAGGGTTGTAAGTGAGTCTCCTTGTTCACGGTCTGTGT
--------------------------------------------------------------------
 ACATCAGGCCCTCTGCCTGCTTGATTTTATTCTCTTGGTGACCGTGAAGTACTATTGTGACCCTTAATT

 TGTAGTCCGGGAGACGGACGAACTAAAATAAGAGAACCACTGGCACTTCATGATAACACTGGGAATTAA
--------------------------------------------------------------------
 TAGCAATAAGAAACATTGGGTTCCATAATTTGCCCAACCTGGGTGTAAAAGCACACACCCTCGATCCCA
                               <<<<<<<<<<<<<<<<<<<
 ATCGTTATTCTTTGTAACCCAAGGTATTAAACGGGTTGGACCCACATTTTCGTGTGTGGGAGCTAGGGT
--------------------------------------------------------------------
 GTACTACAAAGGCAAGAGGGATTAAAGGCGTGTGCCACCANCGGCCTGCTTGGTGATATGCTTTGCAAT

 CATGATGTTTCCGTTCTCCCTAATTTCCGCACACGGTGGTNGCCGGACGAACCACTATACGAAACGTTA
--------------------------------------------------------------------
 ATTTCACTTAGGNGNGTCACCCGTNGGGAGACTTNAGGGTNTAATTNTT

 TAAAGTGAATCCNCNCAGTGGGCANCCCTCTGAANTCCCANATTAANAA
--------------------------------------------------------------------
DUPLICATE REPORT FOR c5g: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
```

```
SEQUENCE FILENAME: c6d
FORWARD PRIMER: TTCTTGCCGGCTTTGAGG
REVERSE PRIMER: GGGCCNAATGGAGGTGCTTG
LENGTH OF INSERT: 334
--------------------------------------------------------------------------
 TCTTGTCCTGCAGGGCAGTCTGGCACGTGCTCACCAGCCCTTGAATGAGGTAATATTTTGCTTCAGCCA

 AGAACAGGACGTCCCGTCAGACCGTGCACGAGTGGTCGGGAACTTACTCCATTATAAAACGAAGTCGGT
--------------------------------------------------------------------------
 TCAACTCCTGGATTTCTTGCCGGCTTTGAGGGAGGGTGATGGTGTCATCTCGGAGGTAATTCANGATGG
             >>>>>>>>>>>>>>>>>>
 AGTTGAGGACCTAAAGAACGGCCGAAACTCCCTCCCACTACCACAGTAGAGCCTCCATTAAGTNCTACC
--------------------------------------------------------------------------
 TGCCAAAGTGCTTTCCACATCGGGTCTATTGAGAATCCCAGCCTTCTTTGGTCAAGTCAAGCACCTCCA
                                                          <<<<<<<<<<<<
 ACGGTTTCACGAAAGGTGTAGCCCAGATAACTCTTAGGGTCGGAAGAAACCAGTTCAGTTCGTGGAGGT
--------------------------------------------------------------------------
 TTNGGCCCACTTGAACAATGGCCTTTGAGCATGGGTGTTCGGTGCCCGGGGTGAGGGGCCCNGNAACTG
 <<<<<<<<
 AANCCGGGTGAACTTGTTACCGGAAACTCGTACCCACAAGCCACGGGCCCCACTCCCCGGGNCNTTGAC
--------------------------------------------------------------------------
DUPLICATE REPORT FOR c6d: <no duplicates>
NUCLEOTIDE BLAST HITS:
    ['M80783','Human B12 protein mRNA, compl...',4.5e-62]
PROTEIN BLAST HITS:
    ['M80783','B12 protein [Homo sapiens]',6.4e-39]
    ['A41784','tumor necrosis factor-alpha-induced ...',8.0e-39]
    ['Z35639','D2045.8, similar to TNF-...',7.3e-11]
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
--------------------------------------------------------------------------
SEQUENCE FILENAME: c6e
FORWARD PRIMER: GCGCAGAGCCACTGATTATA
REVERSE PRIMER: CCAGAGTCACATAGCCCCAGGT
LENGTH OF INSERT: 392
--------------------------------------------------------------------------
 AGGGAACTTTCCTTGCCCTGCCCTCCTCTTACTGCAGCGCAGAGCCACTGATTATAACTGTCTGCTTTA
                                          >>>>>>>>>>>>>>>>>>>>
 TCCCTTGAAAGGAACGGGACGGGAGGAGAATGACGTCGCGTCTCGGTGACTAATATTGACAGACGAAAT
--------------------------------------------------------------------------
 AGGATTTGTTTTACTTTAGTTATATGTGTGGATCTGTATCTGGGTACCATGCACTTGAGTNCAGGGACC

 TCCTAAACAAAATGAAATCAATATACACACCTAGACATAGACCCATGGTACGTGAACTCANGTCCCTGG
--------------------------------------------------------------------------
 CAAGGAAGTCAGAGGCATTGGGTTACCTGGGGCTATGTGACTCTGGAGTCAGTTCTGTCCTCACCTTGT
                          <<<<<<<<<<<<<<<<<<<<<<
 GTTCCTTCAGTCTCCGTAACCCAATGGACCCCGATACACTGAGACCTCAGTCAAGACAGGAGTGGAACA
--------------------------------------------------------------------------
 ACTTCTAGGGNTTNCTCAGCAAAGTGCCTTTGACCAGTTGGAGCCAACATCGGCCTTGNTTTCTNAACT

 TGAAGATCCCNAANGAGTCGTTTCACGGAAACTGGTCAACCTCGGTTGTAGCCGGAACNAAAGANTTGA
--------------------------------------------------------------------------
 TGTGTTTTTAAAGGAGTTATCACTTNCAANTNGGAACCTACCAACTTTTGAAGGGGGNTCCCTANTTGC

 ACACAAAAATTTCCTCAATAGTGAANGTTNANCCTTGGATGGTTGAAAACTTCCCCCNAGGGATNAACG
--------------------------------------------------------------------------
DUPLICATE REPORT FOR c6e: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15,21
END OF RECORD.
```

```
SEQUENCE FILENAME: c6g
FORWARD PRIMER: CGATACCTTTCTAGCTGCGT
REVERSE PRIMER: AGAAGCAGAGAACGAGAAGA
LENGTH OF INSERT: 311
--------------------------------------------------------------------
 GAATTCGTGAGACCATAGCAGGAAGGCAAAAAAACTGAAATTTTCAGTTCATTTCTGAAAAATAAATTG
 ..............
 CTTAAGCACTCTGGTATCGTCCTTCCGTTTTTTTGACTTTAAAAGTCAAGTAAAGACTTTTTATTTAAC
--------------------------------------------------------------------
 GTCAATAAATTCATTTTGTTCTGCTTCTACTTTACACAAAGCTTCATATTCAACCCGATACCTTTCTAG
                                                      >>>>>>>>>>>>>>
 CAGTTATTTAAGTAAAACAAGACGAAGATGAAATGTGTTTCGAAGTATAAGTTGGGCTATGGAAAGATC
--------------------------------------------------------------------
 CTGCGTCTTCTTTTCTGCTATCAGGGCCTGCAGCTGCTGCTNTGGGCTTCTCTCTGCTTCGCTATGGAC
 >>>>>>
 GACGCAGAAGAAAAGACGATAGTCCCGGACGTCGACGACGANACCCGAAGAGAGACGAAGCGATACCTG
--------------------------------------------------------------------
 TTCAGCGAGTTCCGAGCACCAATGGCCTTCATCTTCTCGTTCTCTGCTTCTTTGGCAAGCTGATCAACA
                                  <<<<<<<<<<<<<<<<<<<<
 AAGTCGCTCAAGGCTCGTGGTTACCGGAAGTAGAAGAGCAAGAGACGAAGAAACCGTTCGACTAGTTGT
--------------------------------------------------------------------
 AGCTCAATTNAGGACCACCAACAATTTTCTGAAACTGGGCCAATTNTNGGGTCTCACGGAATTCCGTNG
                                                      ....................
 TCGAGTTAANTCCTGGTGGTTGTTAAAAGACTTTGACCCGGTTAANANCCCAGAGTGCCTTAAGGCANC
--------------------------------------------------------------------
DUPLICATE REPORT FOR c6g: C5 (17)
     vs C5 (score = 141)
            ACGTCGNCATGCACGCGTGACGTGAAGCTTGGATCCTCTAGAGCGGCCGCNTACTACTACTACTG
         C5

            AGCGGAATTCGTGAGACCATAGCAGGAAGGCAAAAAAACTGAAATTTTCAGTTCATTTCTGAAAA
         C5

            ATAAATTGGTCAATAAATTCATTTTGTTCTGCTTCTACTTTACACAAAGCTTCATATTCAACCCG
         C5

            ATACCTTTCTAGCTGCGTCTTCTTTTCTGCTATCAGGGCCTGCAGCTGCTGCTNTGGGCTTCTCT
                                                  ||||| |||||||||||| ||||||||||||
         C5                                       GGCCT-CAGCTGCTGCT-TGGGCTTCTCT

            CTGCTTCGCTATGGACTTCAGCGAGTTCCGAGCACCAATGGCCTTCATCTTCTCGTTCTCTGCTT
            |||||||||||||||||||||| |||||||||||||||||||||||||| || |  |||| |||
         C5 CTGCTTCGCTATGGACTTCAGCAAGTTCCGAGCACCAATGGCCTTCATCTGCT-G--CTCT-CTT

            CTTTGGCAAGCTGATCAACAAGCTCAATTNAGGACCACCAACAATTTTCTGAAACTGGGCCAATT
            |||||||||||||||||||||||||||||| | |||||||||||||||||||||||| ||||||||
         C5 CTTTGGCAAGCTGATCAACAAGCTCAATT-A-GACCACCAACAATTTTCTGAAACT-GGCCAATT


NUCLEOTIDE BLAST HITS:
    ['T16882','NIB1963 Homo sapiens cDNA 3end.',1.3e-19]
    ['T09998','seq692 Homo sapiens cDNA clone...',6.5e-10]
    ['T16883','NIB1963-5R Homo sapiens cDNA 5...',1.4e-08]
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
```

SEQUENCE FILENAME: y11b
FORWARD PRIMER: AGCCCATGACTAGACAGCAC
REVERSE PRIMER: CCCAAGAAGCCTTTAACACT
LENGTH OF INSERT: 254

```
GGAATTCGTGAGACCCCAACCCGGGACTCCTGCTATGTTTCCTCAGCCCATGACTAGACAGCACTNAAC
..............                                  >>>>>>>>>>>>>>>>>>>>
CCTTAAGCACTCTGGGGTTGGGCCCTGAGGACGATACAAAGGAGTCGGGTACTGATCTGTCGTGANTTG

AGCAGAGGAGATGGATGTCAAAGCGCCTCATGGAGGCTAAAGGGGGTGTGGCGGTAAAGAGACAAACTG

TCGTCTCCTCTACCTACAGTTTCGCGGAGTACCTCCGATTTCCCCCACACCGCCATTTCTCTGTTTGAC

GAAACACCGCAGGGGTCTTGAGAATCTTGAGTGTTAAAGGCTTCTTGGGGGACAGCTTGAGAAAACAAG
                                <<<<<<<<<<<<<<<<<<<
CTTTGTGGCGTCCCCAGAACTCTTAGAACTCACAATTTCCGAAGAACCCCCTGTCGAACTCTTTTGTTC

ATTGGTGGCCTCATGCTGTCACATTCCAAAGTCTNTTNGGAAAGNCTTAAAGCCAAAGCCTCGGTCTCA
                                                                .......
TAACCACCGGAGTACGACAGTGTAAGGTTTCAGANAANCCTTTCNGAATTTCGGTTTCGGAGCCAGAGT
```

DUPLICATE REPORT FOR y11b: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.

```
SEQUENCE FILENAME: y11d
FORWARD PRIMER: ATGGTGTTTGTTTTCTCCACAG
REVERSE PRIMER: CCAAAGTGGTGAGAGCTACATA
LENGTH OF INSERT: 422
--------------------------------------------------------------------
 GGAATTCGTGAGACCGACGCTGGCTGTGTTAGCTGGGAATCTACACTGCCAGCCTGCGCGTTGGCAGTG
  ..............
 CCTTAAGCACTCTGGCTGCGACCGACACAATCGACCCTTAGATGTGACGGTCGGACGCGCAACCGTCAC
--------------------------------------------------------------------
 GAGTGGATGGTGTTTGTTTTCTCCACAGAAATTATCCACCAGTTCATTCTGGAACAGCAGAAAGGAAAC
       >>>>>>>>>>>>>>>>>>>>
 CTCACCTACCACAAACAAAAGAGGTGTCTTTAATAGGTGGTCAAGTAAGACCTTGTCGTCTTTCCTTTG
--------------------------------------------------------------------
 AGAGTGCCTCTCTGCATCAACCCGCAGTCCCGCTGCCTTTAAGAGCTTTATTAGGTAACTACGTCTNGT
                                                                     <
 TCTCACGGAGAGACGTAGTTGGGCGTCAGGGCGACGGAAATTCTCGAAATAATCCATTGATGCAGANCA
--------------------------------------------------------------------
 ATGTAGCTCTCACCACTTTGGAAGTAAAGAAAAGGCTGGGTNTGATTTTTGATATTCAGCATATTTCTG
 <<<<<<<<<<<<<<<<<<<<
 TACATCGAGAGTGGTGAAACCTTCATTTCTTTTCCGACCCANACTAAAAACTATAAGTCGTATAAAGAC
--------------------------------------------------------------------

DUPLICATE REPORT FOR y11d: ABX11_10 (8)
     vs ABX11_10 (score = 80)

               TGGCAGTGGAGTGGATGGTGTTTGTTTTCTCCACAGAAATTATCCACCAGTTCATTCTGGAACAG
                                                    ||||| || ||||||||||||||||||
     ABX11_10  TTGAGAAGAACCTAAGCTCTGTCCGACAGGTTAAAGAAATTAT-CA-CAGTTCATTCTGGAACAG

               CAGAAAGGAAAC-AGAGTGCCTCTCTGCATCAACCCGCAGTCCCGCTGCCTTTAAGAGCTTTATT
               |||||||||||| |||||||||||| ||||||| || ||| ||||||||||||||||||||||
     ABX11_10  CAGAAAGGAAACCAGAGTGCCTCTCT-CATCAACACGAAGT-CCGCTGCCTTTAAGAGCTTTATT

               AGGTAACTACGTCTNGTATGTAGCTCTCACCACTTTGGAAGTAAAGAAAAGGCTGGGTNTGATTT
               ||
     ABX11_10  AGA


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: y11e
FORWARD PRIMER: CAGCCCACCTAGTAAAATCCT
REVERSE PRIMER: GGAAACTTGGAANATTCTCTC
LENGTH OF INSERT: 287
--------------------------------------------------------------------------
 AGCGGAATTCGTGAGACCAAAGCAGCTGCTTAGCCAGGCAGGGTGATGCCCGTCTATCATCCCTGCTAC
 ..................
 TCGCCTTAAGCACTCTGGTTTCGTCGACGAATCGGTCCGTCCCACTACGGGCAGATAGTAGGGACGATG
--------------------------------------------------------------------------
 CTGGGGGAGCTGAGGTCAGGAGGATGGNGNGTTCAAAGCCAGCCCACCTAGTAAAATCCTGTCTCAGCA
                                          >>>>>>>>>>>>>>>>>>>>>>>
 GACCCCCTCGACTCCAGTCCTCCTACCNCNCAAGTTTCGGTCGGGTGGATCATTTTAGGACAGAGTCGT
--------------------------------------------------------------------------
 GAGGGGTGGAGGAGCAGAGGATGGCTCAGTGGGTAAAGGCACTTNGCATCAATCCTGACAACTTGAGTT

 CTCCCCACCTCCTCGTCTCCTACCGAGTCACCCATTTCCGTGAANCGTAGTTAGGACTGTTGAACTCAA
--------------------------------------------------------------------------
 TGGCCCTAAGGACAGAACTAGTAGAACCCAAATTGGAGAGAATNTTCCAAGTTTCCTGTCTTTCACTCA
                                     <<<<<<<<<<<<<<<<<<<<<<
 ACCGGGATTCCTGTCTTGATCATCTTGGGTTTAACCTCTCTTANAAGGTTCAAAGGACAGAAAGTGAGT
--------------------------------------------------------------------------
 CAAGGCCATNGGAAGACACANCCACCAACGGTCTCACGNNTTCGGTNGGNGGGNTTTNNTNTTTNTNGN
                                    ......................................
 GTTCCGGTANCCTTCTGTGTNGGTGGTTGCCAGAGTGCNNAAGCCANCCNCCCNAAANNANAAANANCN
--------------------------------------------------------------------------

DUPLICATE REPORT FOR y11e: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.
--------------------------------------------------------------------------
SEQUENCE FILENAME: y11f
FORWARD PRIMER: CATAGCAGCGTCCACTTTCC
REVERSE PRIMER: GCTTCAACGGGACAAGCTTGG
LENGTH OF INSERT: 261
--------------------------------------------------------------------------
 ATTCGTGAGACCGTGGGATAAGGAGCCAGTNACGTAGATGCGGCCGGCCATAGCAGCGTCCACTTTCCC
 ............                                        >>>>>>>>>>>>>>>>>>>>
 TAAGCACTCTGGCACCCTATTCCTCGGTCANTGCATCTACGCCGGCCGGTATCGTCGCAGGTGAAAGGG
--------------------------------------------------------------------------
 TGGCACACCATGCCAGTTCCGGCTGATGAATTGGGGTTCTCTGGCTCCATCAGAGGATCTGCCCCAGAA

 ACCGTGTGGTACGGTCAAGGCCGACTACTTAACCCCAAGAGACCGAGGTAGTCTCCTAGACGGGGTCTT
--------------------------------------------------------------------------
 GAGGAGTTCGAAAATATTCTCCCAAGCTTGTCCCGTTGAAGCAAGGCAAAGTTCTCAAACACGGNTGAC
                        <<<<<<<<<<<<<<<<<<<<
 CTCCTCAAGCTTTTATAAGAGGGTTCGAACAGGGCAACTTCGTTCCGTTTCAAGAGTTTGTGCCNACTG
--------------------------------------------------------------------------
 AGAGAGCTGCCTTCNGACTCCTCCTTGGTTGGNTTGTNGTNGAAATTCGTACTCCCAATACTTCTTGGN
                                                                      ...
 TCTCTCGACGGAAGNCTGAGGAGGAACCAACCNAACANCANCTTTAAGCATGAGGGTTATGAAGAACCN
--------------------------------------------------------------------------
DUPLICATE REPORT FOR y11f: EX14 (-7)
     vs EX14 (score = 112)
             ACTTTCCCTGGCACACCATGCCAGTTCCGGCTGATGAATTGGGGTTCTCTGGCTCCATCAGAGGA
                                                                          |||||
         EX14                                                            CGGAGGA

             TCTGCCCCAGAAGAGGAGTTCGAAAATATTCTCCCAAGCTTGTCCCGTTGAAGCAAGGCAAAGTT
             |||||||||||||||||||| |||||| ||||||| ||| ||||||| ||||||||||||||||
         EX14 TCTGCCCCAGAAGAGGAGTT-GAAAATGTTCTCCC-AGC-TGTCCCGCTGAAGCAAGGCAAAGTG
```

```
          CTCAAACACGGNTGACAGAGAGCTGCCTTCNG-ACTCCTCCTTGGTTGGNTTG-TNG-TNGAAAT
          || |||||||| |||||||||||||||||| | |||||||| ||| ||| ||| | | |  ||||
    EX14  CT-AAACACGGCTGACAGAGAGCTGCCTTC-GCACTCCTCC-TGGCTGGGTTGCT-GCT-TAAAT

          TCGTACTCCCAATACTTCT
          |||||||| || |||| ||
    EX14  TCGTACTCACAGTACTGCT
```

NUCLEOTIDE BLAST HITS:
['M77123','Mouse vitronectin mRNA, compl...',1.4e-53]
['X63003','M.musculus mRNA for vitronectin',1.4e-53]
['X72091','M.musculus gene for vitronectin',7.8e-37]
['X03168','Human mRNA for S-protein',8.2e-33]
['M55442','O.cuniculus domesticus vitron...',1.8e-32]
PROTEIN BLAST HITS:
['P29788','VITRONECTIN PRECURSOR (SERUM SPRE...',3.1e-33]
['M77123','vitronectin [Mus musculus]',3.1e-33]
['JT0662','Vitronectin - Mouse>gi 441466 gp X72...',3.1e-33]
['P22458','VITRONECTIN PRECURSOR (SERUM SPRE...',1.7e-25]
['X03168','Human mRNA for S-protein....',4.5e-25]
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.

SEQUENCE FILENAME: y1h
FORWARD PRIMER: TGAGTGTATGCACAAGGCCT
REVERSE PRIMER: TCAGATTCAGCTAGAGCA
LENGTH OF INSERT: 340

CCNTTGGTGGAATCGTTTTATTTTAGTGAAAATAAGCCAGANAAAGCACGGATACANCTGAGTGTATGC
                                                          >>>>>>>>>>>
GGNAACCACCTTAGCAAAATAAAATCACTTTTATTCGGTCTNTTTCGTGCCTATGTNGACTCACATACG

ACAAGGCCTGGACCCCAGNACTAGGGAGNCTAAGTTAGAATCATGAGTTTGAGGCAACATNAACATACA
>>>>>>>>>
TGTTCCGGACCTGGGGTCNTGATCCCTCNGATTCAATCTTAGTACTCAAACTCCGTTGTANTTGTATGT

TAGTGAGGCTTGTCTCGAAAAAAGGAAGNAANNTCCCTTGCTCTAGCTGAATCTGAATGCCCACTGGAC
                                    <<<<<<<<<<<<<<<<<<
ATCACTCCGAACAGAGCTTTTTTCCTTCNTTNNAGGGAACGAGATCGACTTAGACTTACGGGTGACCTG

TGGGCCGNTTAAGCNCTANGAAAGGAGGTTCTGGGAGTCAAGAGCCAAGANTTGANACTTNCCTTNCGT

ACCCGGCNAATTCGNGATNCTTTCCTCCAAGACCCTCAGTTCTCGGTTCTNAACTNTGAANGGAANGCA

GNTCCTCAANNACTACCTNTAACTTCAAACAACACATCTTGCACCAAGGNAGNAACATANGATT

CNAGGAGTTNNTGATGGANATTGAAGTTTGTTGTGTAGAACGTGGTTCCNTCNTTGTATNCTAA

DUPLICATE REPORT FOR y1h: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 14,15
END OF RECORD.

```
SEQUENCE FILENAME: y4a
FORWARD PRIMER: CTGCATTCCAGTATAGCAGCC
REVERSE PRIMER: CTTTATTGCTTTCCAAGGAG
LENGTH OF INSERT: 262
--------------------------------------------------------------------
 GCCTGCATTCCAGTATAGCAGCCATGTCTCTCTGCAAGCTTCCAGTNGGCACATGTGGGGCACATTTCG
   >>>>>>>>>>>>>>>>>>>>
 CGGACGTAAGGTCATATCGTCGGTACAGAGAGACGTTCGAAGGTCANCCGTGTACACCCCGTGTAAAGC
--------------------------------------------------------------------
 TTTTGAGAGACCGGATGGCTCCCACTTTGATGTCCGGATCCCACCATTCTCCTTGGAAAGCAATAAAGA
                                                <<<<<<<<<<<<<<<<<<<<
 AAAACTCTCTGGCCTACCGAGGGTGAAACTACAGGCCTAGGGTGGTAAGAGGAACCTTTCGTTATTTCT
--------------------------------------------------------------------
 TGAGAAGACCCCACCCTCAAGGCCTTCACTGGTAGGCAAGCTTGAGGCCCAGGTTGGCCCTAAGAACAA

 ACTCTTCTGGGGTGGGAGTTCCGGAAGTGACCATCCGTTCGAACTCCGGGTCCAACCGGGATTCTTGTT
--------------------------------------------------------------------
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: CACGTATGGTATCTCAGCCGC
REVERSE PRIMER: TGCTCTACACCCTCTCCAG
LENGTH OF INSERT: 215
--------------------------------------------------------------------
 GCGGCCGCGTACTACTACTACTGAGCGGAATTCGTGAGACCCCCCGTTGAGGTGGAAAAGCAGCACTTG
 ....................................
 CGCCGGCGCATGATGATGATGACTCGCCTTAAGCACTCTGGGGGGCAACTCCACCTTTTCGTCGTGAAC
--------------------------------------------------------------------
 CACATACACGTATGGTATCTCAGCCGCCCTTCATCTAGGTAGGNGTNGGCTAAGGCATCCTTAGCGGAT
       >>>>>>>>>>>>>>>>>>>>
 GTGTATGTGCATACCATAGAGTCGGCGGGAAGTAGATCCATCCNCANCCGATTCCGTAGGAATCGCCTA
--------------------------------------------------------------------
 ATTCTTTTTGATGGATCAAAGACCAACATCCTGCAAAGAGGTGAACAGCTTCATGTGTGGCCTGGCTGG
                                                          <<<<
 TAAGAAAAACTACCTAGTTTCTGGTTGTAGGACGTTTCTCCACTTGTCGAAGTACACACCGGACCGACC
--------------------------------------------------------------------
 AGAGGGTGTAGAGCACAGGAAGGGATGGCTTTAAGAGAGCAGAAAATTCGGTCTCACGAATGCGGGCTT
 <<<<<<<<<<<<<<<                                  ....................
 TCTCCCACATCTCGTGTCCTTCCCTACCGAAATTCTCTCGTCTTTTAAGCCAGAGTGCTTACGCCCGAA
--------------------------------------------------------------------
DUPLICATE REPORT FOR y4b: x9c (17)
     vs x9c (score = 114)
              AGCACTTGCACATACACGTATGGTATCTCAGCCGCCCTTCATCTAGGTAGGNGTNGGCTAAGGCA
                                                                  ||||||||||
          x9c                                                     GGCTAAGGCA

              TCCTTAGCGGATATTCTTTTTGATGGATCAAAGACCAACATCCTGCAAAGAGGTG-AACAGCTTC
              |||||||||||||||||||||||||||||||||||||||||||||||||||||  ||| |||||||||
          x9c TCCTTAGCGGATATTCTTTTTGATGGATCAAAGACCAACATCCTGCAAAGCNGTGGAACAGCTTC

              ATGTGTGGCCTGGCTGGAGAGGGTGTAGAGCACAGGAAGGGATGGCT-TT-A--AGAGAGCAGAA
              |||||||||||||||||||||||| |||||||||||||| ||||||| || |  || ||
          x9c ATGTGTGGCCTGGCTGGAGAGGGTCTAGAGCACAGGAAGAGATGGCTGTTTATNAG-GACCCCTG
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: GTACAGTTAGAGATTCGCGT
LENGTH OF INSERT: 182
--------------------------------------------------------------------
 GGAATTCGTGAGACCGGGGTCCATCATCATGGAGACTTGGTCTCTCAAGCTTCTCAGGCCAGTCTGGGN
 ..............              >>>>>>>>>>>>>>>>>>>>>>
 CCTTAAGCACTCTGGCCCCAGGTAGTAGTACCTCTGAACCAGAGAGTTCGAAGAGTCCGGTCAGACCCN
--------------------------------------------------------------------
 ANAGTGACCTGATTCGGAAAAGTTGTCATCTCTCAGCAAGCCCTGGACAGGAAGGATCTGAGCTTTAGG

 TNTCACTGGACTAAGCCTTTTCAACAGTAGAGAGTCGTTCGGGACCTGTCCTTCCTAGACTCGAAATCC
--------------------------------------------------------------------
 GACACTGACGCGAATCTCTAACTGTACCTTCCTTCTGAACTGCGAACTGCCTTAGCCAGGGTCTCACGA
       <<<<<<<<<<<<<<<<<<<                          ..........
 CTGTGACTGCGCTTAGAGATTGACATGGAAGGAAGACTTGACGCTTGACGGAATCGGTCCCAGAGTGCT
--------------------------------------------------------------------

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: GGACATCTTGGGGCACTG
REVERSE PRIMER: GCCTAGTGAAGCCAGAAGT
LENGTH OF INSERT: 241
--------------------------------------------------------------------
 CGGAATTCGTGAGACCAGAGGCTGGACATCTTGGGGCACTGGTAACAGNANGCAGCACTGGACTAAGTT
 ..............       >>>>>>>>>>>>>>>>>
 GCCTTAAGCACTCTGGTCTCCGACCTGTAGAACCCCGTGACCATTGTCNTNCGTCGTGACCTGATTCAA
--------------------------------------------------------------------
 CCGGGTCTCTCTACCNGGNGNCCAGTNTAGAGGAGGCCAGTGTAGAGTTGGACTGACCTGACTATAGAC

 GGCCCAGAGAGATGGNCCNCNGGTCANATCTCCTCCGGTCACATCTCAACCTGACTGGACTGATATCTG
--------------------------------------------------------------------
 ACGGATTACCCACCAGCACTTCTGGCTTCACTAGGCCTGGAGAAGAGTGAGGCAAAACCAGTCAGGTTT
                <<<<<<<<<<<<<<<<<<
 TGCCTAATGGGTGGTCGTGAAGACCGAAGTGATCCGGACCTCTTCTCACTCCGTTTTGGTCAGTCCAAA
--------------------------------------------------------------------
 CTGGGAACTGTGGGTTCTGAGATTTGAGCAAAGCTNTCAAGCATCCAGACGGTCTCACGGATGCGGGTT
                                                  ..................
 GACCCTTGACACCCAAGACTCTAAACTCGTTTCGANAGTTCGTAGGTCTGCCAGAGTGCCTACGCCCAA
--------------------------------------------------------------------
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: GCACCTTAGTGACGAGCAAC
LENGTH OF INSERT: 348

```
CAGCTGTATCCCTTGGCTACTCTTATGATGGCCTCATGCTGTCACATTCCAAAGTCTGTGGGAAAGACT
          >>>>>>>>>>>>>>>>>>>>
GTCGACATAGGGAACCGATGAGAATACTACCGGAGTACGACAGTGTAAGGTTTCAGACACCCTTTCTGA
```

```
GAAAGCCAAAGCCTCTATTTGGNCANAAAAGTTGGATTGTCATCTTAGTTCCAGCGAGATCTGGACGAA

CTTTCGGTTTCGGAGATAAACCNGTNTTTTCAACCTAACAGTAGAATCAAGGTCGCTCTAGACCTGCTT
```

```
CTAATAGTTGCTCGTCACTAAGGTGCCTGACCCAAGCATCTTAGAGCTATTTATACATTCAAGATCGTG
      <<<<<<<<<<<<<<<<<<<<
GATTATCAACGAGCAGTGATTCCACGGACTGGGTTCGTAGAATCTCGATAAATATGTAAGTTCTAGCAC
```

```
ATTAACATAACATGCCTGATTATTAAGTCACATTTCTCATTTCCCCTTGCCCTGTCAGGCTAGATCCTT

TAATTGTATTGTACGGACTAATAATTCAGTGTAAAGAGTAAAGGGGAACGGGACAGTCCGATCTAGGAA
```

```
GTGAGAGCCCAACCCTNGGTTTGGCATCTTGCCATAGGTCTAAGAGGTGAGCTTTGTTGTGCCATGTNG

CACTCTCGGGTTGGGANCCAAACCGTAGAACGGTATCCAGATTCTCCACTCGAAACAACACGGTACANC
```

DUPLICATE REPORT FOR y7a: <no duplicates>
NUCLEOTIDE BLAST HITS:
['M80783','Human B12 protein mRNA, compl...',6.8e-27]
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

```
GTGGGAATGAGCCAGAGCTACACGGTGGGACCCAGCGGAAAGGAAAAAGAAAGTTGCCTTAGACAATGG
                  >>>>>>>>>>>>>>>>>
CACCCTTACTCGGTCTCGATGTGCCACCCTGGGTCGCCTTTCCTTTTTCTTTCAACGGAATCTGTTACC
```

```
GGCTCTTGTTTGGCAGTCAGAGGCCTCATCTGCCTCCCCAGGACTGGGGCTACTAGAACTGGGATTAAA
                                                                    <<<
CCGAGAACAAACCGTCAGTCTCCGGAGTAGACGGAGGGGTCCTGACCCCGATGATCTTGACCCTAATTT
```

```
AGTACACACCACCAGGCCCTGGCTAACACAACTTTCAAAACCTAACTCAGGAGAGCTGAGAAAAGACTN
<<<<<<<<<<<<<<<<<<
TCATGTGTGGTGGTCCGGGACCGATTGTGTTGAAAGTTTTGGATTGAGTCCTCTCGACTCTTTTCTGAN
```

```
GTGGTGTCATTCATAATTAGCCCCTTGCAGAGCCCGTTGCTCTGCCCATCTTGCCCACATAACACTGG

CACCACAGTAAGTATTAATCGGGGAACGTCTCGGGCAACGAGACGGGTAGAACGGGTGTATTGTGACC
```

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
--------------------------------------------------------------------
 GGGGCATTGAGGGCCCCATCGATGCTGCCTTCACTCGCATCAACTGTCAGGGGNAGACCTACTTGTTCA
                        >>>>>>>>>>>>>>>>>>
 CCCCGTAACTCCCGGGGTAGCTACGACGGAAGTGAGCGTAGTTGACAGTCCCCNTCTGGATGAACAAGT
--------------------------------------------------------------------
 AGGGTAGTCAGTACTGGCGCTTTGAGGATGGGGTCCTGGACCCTGGTTATCCCCGAAACATCTCCGAAG
                                                                   <<<
 TCCCATCAGTCATGACCGCGAAACTCCTACCCCAGGACCTGGGACCAATAGGGGCTTTGTAGAGGCTTC
--------------------------------------------------------------------
 CTTCAGTGGCATACCAGACAATGTTTGATGCAGCGTTCGNCCTTCCCGGCCACTGTAACGGTGGGCAGG
 <<<<<<<<<<<<<<<<<
 GAAGTCACCGTATGGTCTGTTACAAACTACGTCGCAAGCNGGAAGGGCCGGTGACATTGCCACCCGTCC
--------------------------------------------------------------------

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]
    ['X05006','Human S-protein gene, complete cds.',1.3e-37]
PROTEIN BLAST HITS:
    ['M77123','vitronectin [Mus musculus]',8.1e-33]
    ['JT0662','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: AGTGCTGCTTTTCCACCTCA
REVERSE PRIMER: TGAACTGGTGGATAATTTCTT
LENGTH OF INSERT: 249
--------------------------------------------------------------------------
 AAGTGCTGCTTTTCCACCTCAACGGGGAGAGTCTACACCAGTNACTTTGAGCCTGTCACCAAGGGNAAA
  >>>>>>>>>>>>>>>>>>>>
 TTCACGACGAAAAGGTGGAGTTGCCCCTCTCAGATGTGGTCANTGAAACTCGGACAGTGGTTCCCNTTT
--------------------------------------------------------------------------
 TTTGATGACACCTTTGAGAAGAACCTCAGCTCTGTCCGACAGGTTAAAGAAATTATCCACCAGTTCATT
                                                <<<<<<<<<<<<<<<<<<<<<<
 AAACTACTGTGGAAACTCTTCTTGGAGTCGAGACAGGCTGTCCAATTTCTTTAATAGGTGGTCAAGTAA
--------------------------------------------------------------------------
 CTGGAACAGCAGAAAGGAAACAGAGTGCCTCTCTGCATCAACCCGCAGTCCCNTGCCTTTAAGAGCTTT

 GACCTTGTCGTCTTTCCTTTGTCTCACGGAGAGACGTAGTTGGGCGTCAGGGNACGGAAATTCTCGAAA
----------------------------------------------------------------------------
 ATTAGTTCCACCGTCGGTCAAGCCGTCTTGAGATTGCNGCCG

 TAATCAAGGTGGCAGCCAGTTCGGCAGAACTCTAACGNCGGC
----------------------------------------------------------------------------

DUPLICATE REPORT FOR y8c: ABX11_10 (16)  y11d (15)
     vs ABX11_10 (score = 144)
               AAGTGCTGCTTTTCCACCTCAACGGGGAGAGTCTACACCA-GTNACTTTGAGCCTGTCACCAAGG
                                                       |||| || |||||||||| ||||| ||
     ABX11_10                                          ACCAAGTGACTTTGAGCCAGTCACAAACC

               GN-AAATTTGATGACACCTTTGAGAAGAACCTCAGCTCTGTCCGACAGGTTAAAGAAATTATCCA
                  |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||| ||
     ABX11_10 CCCAAATTTGATGACACCTTTGAGAAGAACCTAAGCTCTGTCCGACAGGTTAAAGAAATTAT-CA

               CCAGTTCATTCTGGAACAGCAGAAAGGAAAC-AGAGTGCCTCTCTGCATCAACCCGCAGTCC-CN
                ||||||||||||||||||||||||||||||| |||||||||||| ||||||| || ||||| |
     ABX11_10 -CAGTTCATTCTGGAACAGCAGAAAGGAAACCAGAGTGCCTCTCT-CATCAACACGAAGTCCGC-

               TGCCTTTAAGAGCTTTATTAGTTCCACCGTCGGTCAAGCCGTCTTGAGATTGCNGCCG
               |||||||||||||||||||
     ABX11_10 TGCCTTTAAGAGCTTTATTAGA

     vs y11d (score = 93)

               NAAATTTGATGACACCTTTGAGAAGAACCTCAGCTCTGTCCGACAGGTTAAAGAAATTATCCACC
                                                                    ||||||||||
          y11d ACTGCCAGCCTGCGCGTTGGCAGTGGAGTGGATGGTGTTTGTTTTCTCCACAGAAATTATCCACC

               AGTTCATTCTGGAACAGCAGAAAGGAAACAGAGTGCCTCTCTGCATCAACCCGCAGTCCCN-TGC
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||
          y11d AGTTCATTCTGGAACAGCAGAAAGGAAACAGAGTGCCTCTCTGCATCAACCCGCAGTCCCGCTGC

               CTTTAAGAGCTTTATTAGTTCCAC--CGTCG-GTCA---AGC-CGTCTTGAGATTGCNGCCG
               |||||||||||||||||| |  ||  ||||  || |   ||| | ||
          y11d CTTTAAGAGCTTTATTAGGT-AACTACGTCTNGT-ATGTAGCTC-TCACCACTTTGGAAGTAAAG

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: z11e
FORWARD PRIMER: GCTGATTAGGAACTTCCTGGC
REVERSE PRIMER: GCTGTAGCTCNTGTTGTTGAG
LENGTH OF INSERT: 296
--------------------------------------------------------------------
 GCGGAATTCGTGAGACCGGGTCTCAGGTGTAGCTTGGCTTTAAGTTCACTGTGTAGCTGAGGCTGATTA
 .................                                          >>>>>>>>
 CGCCTTAAGCACTCTGGCCCAGAGTCCACATCGAACCGAAATTCAAGTGACACATCGACTCCGACTAAT
--------------------------------------------------------------------
 GGAACTTCCTGGCTCTTCTCTCTCNGNNNGGACCAGGGTTGTAGGTTTGCCTGACACATCTGATTTCTT
 >>>>>>>>>>>>>
 CCTTGAAGGACCGAGAAGAGAGAGNCNNNCCTGGTCCCAACATCCAAACGGACTGTGTAGACTAAAGAA
--------------------------------------------------------------------
 TGGTGTTATGGATGGATCCAAGGGCTTCCTGAAAGCTAGAGAAGCACTCAACAACANGAGCTACAGCCT
                                                  <<<<<<<<<<<<<<<<<<<<<
 ACCACAATACCTACCTAGGTTCCCGAAGGACTTTCGATCTCTTCGTGAGTTGTTGTNCTCGATGTCGGA
--------------------------------------------------------------------
 AAGGCCTTTACTTGTGGTCCTCCTGCATCTAGGGACACTTAAGAATGTTTTAGNTAAGGCCTAGTAGCG

 TTCCGGAAATGAACACCAGGAGGACGTAGATCCCTGTGAATTCTTACAAAATCNATTCCGGATCATCGC
--------------------------------------------------------------------
 NAAGCTTGTAATGTAAACATGAGTACGTTGATTGNAGGGTCTCACGGATTNCGGTNGGGNANAGNNGNN
                                          ..............................
 NTTCGAACATTACATTTGTACTCATGCAACTAACNTCCCAGAGTGCCTAANGCCANCCCNTNTCNNCNN
--------------------------------------------------------------------
DUPLICATE REPORT FOR z11e: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,23
END OF RECORD.
```

```
SEQUENCE FILENAME: z11g
FORWARD PRIMER: AGAGCATCAAGCTAAGCCGC
REVERSE PRIMER: GCCTGAATCAGACTCCGG
LENGTH OF INSERT: 198
------------------------------------------------------------------------
 GCGGAATTCGTGAGACCAACAGACTAGAGAGCATCAAGCTAAGCCGCTCCTCGCTCCTCACCTGAGCCT
 .................          >>>>>>>>>>>>>>>>>>>>
 CGCCTTAAGCACTCTGGTTGTCTGATCTCTCGTAGTTCGATTCGGCGAGGAGCGAGGAGTGGACTCGGA
------------------------------------------------------------------------
 CAGGTNNGNATTAGTGGGAAAGCACTGATGATGTCACCCTCAGAAAGGGGCTGGATTAAATGTGATAAG

 GTCCANNCNTAATCACCCTTTCGTGACTACTACAGTGGGAGTCTTTCCCCGACCTAATTTACACTATTC
------------------------------------------------------------------------
 CCTGTAAAGTGCAGAAGCCGGAGTCTGATTCAGGCCTTTCGGGTTCCAAGTCAANAACGNTTTCTCCTG
                   <<<<<<<<<<<<<<<<<<
 GGACATTTCACGTCTTCGGCCTCAGACTAAGTCCGGAAAGCCCAAGGTTCAGTTNTTGCNAAAGAGGAC
------------------------------------------------------------------------

DUPLICATE REPORT FOR z11g: a7d (16)
     vs a7d (score = 112)
               CGTTCGNATGTCACGGCGTNACGTAAAGCTTGNGATCCTCTAGAGCGGCCGTGTACTACTACTAC
                                                                  |||| || | | | |  ||
          a7d                                                   ACAGGCC-TG-AAT-C-A-GAC

               TTGAGCGGAATTCGTGAGACC-AACAGACTAGAGAGCATCAAGCTAAGCCGCTCCTCGCTCCTCA
                |   |||  ||| ||  ||  |||||||||||||||||| |||||||||||||| | | ||||
          a7d -T--CCGG-CTTC-TG-CACTTAACAGACTAGAGAGCATC-AGCTAAGCCGCTCCTNGGTNCTCA

               CCTGAGCCTCAGGTNNGNATTAGTGGGAAAGCACTGATGATGTCACCCTCAGAAAGGGGCTGGAT
               ||||||||||||||    | | ||||||||||||||||||||||||||||||||||||||||
          a7d CCTGAGCCTCAGGTTTCCACTGGTGGGAAAGCACTGATGATGTCACCCTCAGAAAGGGGCTGGAT

               TAAATGTGATAAGCCTGTAAAGTGCAGAAGCCGGAGTCTGATTCAGGCCTTTCGGGTTCCAAGTC
               ||||||||||||||||||   || |  ||| | ||
          a7d TAAATGTGATAAGCCTGTTGGGTTC-CAAGTCAGAAATGTTTTCTCC

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,22,23,27
END OF RECORD.
```

```
SEQUENCE FILENAME: z11h
FORWARD PRIMER: GTCAAGCCCCAAGCTCTG
REVERSE PRIMER: CAGGTAGGGCACAGGGTAGC
LENGTH OF INSERT: 203
--------------------------------------------------------------------
 ATTCGTGAGACCNGAGGNAAGGTCAAGCCCCAAGCTCTGGACAGTGGTCCTGGGATGTACTGCTACCAG
 ............          >>>>>>>>>>>>>>>>>>
 TAAGCACTCTGGNCTCCNTTCCAGTTCGGGGTTCGAGACCTGTCACCAGGACCCTACATGACGATGGTC
--------------------------------------------------------------------
 GNNCNTTGCGACATATGTACTGTTCTTCTCAGCCTGCCTTCCATCAGTACTCCCCGGGTGGAGGCAGCT
                                                                    <<<
 CNNGNAACGCTGTATACATGACAAGAAGAGTCGGACGGAAGGTAGTCATGAGGGGCCCACCTCCGTCGA
--------------------------------------------------------------------
 ACCCTGTGCCCTACCTGGGTTCACCTCACTATCCCTATCAGAGGATTGCACCCCAGGGCCAACGCCGAA
 <<<<<<<<<<<<<<<<<
 TGGGACACGGGATGGACCCAAGTGGAGTGATAGGGATAGTCTCCTAACGTGGGGTCCCGGTTGCGGCTT
--------------------------------------------------------------------

DUPLICATE REPORT FOR z11h: a10g (19)  a8d (-24)
     vs a10g (score = 123)
               GCTACCAGGNNCNTTGCGACATATGTACTGTTCTTCTCAGCCTGCCTTCCATCAGTACTCCCCGG
                                    ||||||||||||||||||||||||||||||||||||||||||
          a10g                      TATGTACTGTTCTTCTCAGCCTGCCTTCCATCAGTACTCCCCGG

               GTGGAGGCAGCTACCCTGTGCCCTACCTGGGTTCACCTCACTATCCCTATCAGAGGATTGCACCC
               |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
          a10g GTGGAGGCAGCTACCCTGTGCCCTACCTGGGCTCACCTCACTATCCCTATCAGAGGATTGCACCC

               CAGGGCCAACGCCGAAGTCACCAGGGTCTCACGAATTCNGGTGGGGANNNNNNGNTNNNNTGGGT
                | || ||||| ||||||||||||
          a10g NA-GGNCAACGNCGAAGTCACCAGCCACTCTTNCCAAAGNCCATCTACTCTTACAGCATCCTCAT

     vs a8d (score = 149)

               GTNACGTAA-AG-CTTGNGATCCTCTAGAGCGGCCGT-ACT--A-CTACTA-CTGAGCGGAATTC
                 || | || || |||| | | |  || | |  || | ||   | | |  | |  | |  |||||
           a8d CCNA-GCAAGAGGCTTG-G-TGCAATA-AACTTCCCTTACCCCAGCCAGGAACACAACCAAATTC

               GTG-AGAC--CNGAGGN-AAGGTCAAGCCCCAAGCTCTGGACAGTGGTCCTGGGATGTACTGCTA
                || ||    | ||||  ||||||||||||||||||||||||||||||||||||||||||||||
           a8d -TGCAGGGGTCAGAGGTCAAGGTCAAGCCCCAAGCTCTGGACAGTGGTCCTGGGATGTACTGCTA

               CCAGGNNCN-TTGCGACATATGTACTGTTCTTCTCAGCCTGCCTTCCATCAGTACTCCCCGGGTG
               ||||    |  |||| ||||||||||||||||||||||||||||||||||||||||||||||||
           a8d CCAGCCTCCCTTGCAACATATGTACTGTTCTTCTCAGCCTGCCTTCCATCAGTACTCCCCGGGTG

               GAGGCAGCTACCCTGTGCCCTACCTGGGTTCACCTCACTATCCCTATCAGAGGATTGCACCCCAG
               |||||||||||||||||||||||||||| |||||||||||||||
           a8d GAGGCAGCTACCCTGTGCCCTACCTGGGCTCACCTCACTATCCC

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

```
SEQUENCE FILENAME: z1e#2
FORWARD PRIMER: GACTGAGCCAGAAGTGTCTGT
REVERSE PRIMER: CCGGCTGTGTGACCTTTCTA
LENGTH OF INSERT: 267
--------------------------------------------------------------------
 GGAATTCGTGAGACCGGTAAGACCCTGTTGTTAATGATGTCATATGTGTTGGTTACAGGACTTGGGACT
 ..............                                                >>>>
 CCTTAAGCACTCTGGCCATTCTGGGACAACAATTACTACAGTATACACAACCAATGTCCTGAACCCTGA
--------------------------------------------------------------------
 GAGCCAGAAGTGTCTGTACCAGTGGCCTGGCTCCTGTAGTACAGAAAGTGCAGTGAAGGATTTCTGGGG
 >>>>>>>>>>>>>>>>
 CTCGGTCTTCACAGACATGGTCACCGGACCGAGGACATCATGTCTTTCACGTCACTTCCTAAAGACCCC
--------------------------------------------------------------------
 AGGATCATAACCCAAAGTCCTATTCAGAATTCTGAGTGTTTCATGATAGAAAGGTCACACAGCCGGTTC
                                                 <<<<<<<<<<<<<<<<<<<
 TCCTAGTATTGGGTTTCAGGATAAGTCTTAAGACTCACAAAGTACTATCTTTCCAGTGTGTCGGCCAAG
--------------------------------------------------------------------
 TGCAGGGAGATCAAACATTCTCCCCTTCAAGGGATTTCCTGAAGCCTTTGAGTTGTTTACTTCCTTGAG

 ACGTCCCTCTAGTTTGTAAGAGGGGAAGTTCCCTAAAGGACTTCGGAAACTCAACAAATGAAGGAACTC
--------------------------------------------------------------------
DUPLICATE REPORT FOR z1e#2: <none>
NUCLEOTIDE BLAST HITS:  <none>
PROTEIN BLAST HITS:  <none>
PHYSICAL MAP LOCATION: 7,9,18
END OF RECORD.
--------------------------------------------------------------------
SEQUENCE FILENAME: z1g#2
FORWARD PRIMER: CACCATGTCCACCTCAACC
REVERSE PRIMER: AGGACAAGGACGGAGCTCCA
LENGTH OF INSERT: 229
--------------------------------------------------------------------
 GCCNCCTACTACTACTATTACTGAGCGGAATTCGTGAGACCNCCAGAATACCTCCATCACCCGCACCAT
 ....................................                          >>>>>>
 CGGNGGATGATGATGATAATGACTCGCCTTAAGCACTCTGGNGGTCTTATGGAGGTAGTGGGCGTGGTA
--------------------------------------------------------------------
 GTCCACCTCAACCCTCACTGGCAGCTCAGAAGCATTATAGATCTTCGGATCAACANTGTATACCATAAT
 >>>>>>>>>>>>>
 CAGGTGGAGTTGGGAGTGACCGTCGAGTCTTCGTAATATCTAGAAGCCTAGTTGTNACATATGGTATTA
--------------------------------------------------------------------
 NCCACCCCAGCGGGACCTGTGGAAGGCATTGGTGGCCACTGGAGCTCCGTCCTTGTCCTGAATGTACAG
                                          <<<<<<<<<<<<<<<<<<<<
 NGGTGGGGTCGCCCTGGACACCTTCCGTAACCACCGGTGACCTCGAGGCAGGAACAGGACTTACATGTC
--------------------------------------------------------------------
 CGGGGAGTGGGCAAGCTCAGGCACATAGAGCAGAAAGTGGAGTACAGGGTACAAGGGAGGCAGGGTCTC
                                                                .......
 GCCCCTCACCCGTTCGAGTCCGTGTATCTCGTCTTTCACCTCATGTCCCATGTTCCCTCCGTCCCAGAG
--------------------------------------------------------------------
DUPLICATE REPORT FOR z1g#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
------------------------------------------------------------------------
 AATTCGTGAGACCGGCAAATAAATAAAGCCTTTTGATCTTCAGCCAGTCAGGCCCCCAGTAAACTGTAA
 .............                                >>>>>>>>>>>>>>>>>>>>
 TTAAGCACTCTGGCCGTTTATTTATTTCGGAAAACTAGAAGTCGGTCAGTCCGGGGGTCATTTGACATT
------------------------------------------------------------------------
 ACAGACGTGAACTTGTGAGTGGGAGAGTTGGCAGCTTCCTCCTTCTCTCCTGAGAAAGCATCTTATTCC

 TGTCTGCACTTGAACACTCACCCTCTCAACCGTCGAAGGAGGAAGAGAGGACTCTTTCGTAGAATAAGG
------------------------------------------------------------------------
 CTGTGCCAGATGGGCCTGTACACCCATGGCCGAAGCACCCCAAGCTGCCTCGCTAGCCACTTGCAAGAC
                                  <<<<<<<<<<<<<<<<<<<
 GACACGGTCTACCCGGACATGTGGGTACCGGCTTCGTGGGGTTCGACGGAGCGATCGGTGAACGTTCTG
------------------------------------------------------------------------
 TTGAATGTCTNCCTAAGGGACTAAGAGTGTCCCCTTAACCAGTTGTCAGGAATTGACTTGAGGAGGGGA

 AACTTACAGANGGATTCCCTGATTCTCACAGGGGAATTGGTCAACAGTCCTTAACTGAACTCCTCCCCT
------------------------------------------------------------------------
 CTTGCAGGCGGTTCTTCTTAAAGCTTTAACCCTNGCCCGGACGGTTCAGGGTCTCACGATTTCCGTGGG
                                                  ..................
 GAACGTCCGCCAAGAAGAATTTCGAAATTGGGANCGGGCCTGCCAAGTCCCAGAGTGCTAAAGGCACCC
------------------------------------------------------------------------
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: GCCTGGCTCCTGTAGTACAGAA
LENGTH OF INSERT: 238
------------------------------------------------------------------------
 GTTCGNATGNCACGCGTNACGTAAAGCTTGGGATCCTCTAGAGCGGCCGNGTACTACTACTACTGAGCG
 ........................................................................
 CAAGCNTACNGTGCGCANTGCATTTCGAACCCTAGGAGATCTCGCCGGCNCATGATGATGATGACTCGC
------------------------------------------------------------------------
 GAATTCGTGAGACCCTGAAACTGAGCTTAAGGAACTTTCCTGAAGGGGAGAATGTTTGATCTCCCTGCA
 ..............              >>>>>>>>>>>>>>>>>>>>
 CTTAAGCACTCTGGGACTTTGACTCGAATTCCTTGAAAGGACTTCCCCTCTTACAAACTAGAGGGACGT
------------------------------------------------------------------------
 GAACCGGCTGTGTGACCTTTCTATCATGAAACACTCAGAATTCTGAATAGGACTTTGGGTTATGATCCT

 CTTGGCCGACACACTGGAAAGATAGTACTTTGTGAGTCTTAAGACTTATCCTGAAACCCAATACTAGGA
------------------------------------------------------------------------
 CCCCAGAAATCCTTCACTTGCACTTTCTGTACTACAGGAGCCAGGCCACTGGTACAGACACTTCTGGCT
                         <<<<<<<<<<<<<<<<<<<<<<
 GGGGTCTTTAGGAAGTGAACGTGAAAGACATGATGTCCTCGGTCCGGTGACCATGTCTGTGAAGACCGA
------------------------------------------------------------------------
 CAGTCCCAAGTCCTNTAACCAACACATATGACATCATTAACAACCGGTCTCACGGATTCGGTTGGGNNN
                                                 ......................
 GTCAGGGTTCAGGANATTGGTTGTGTATACTGTAGTAATTGTTGGCCAGAGTGCCTAAGCCAACCCNNN
------------------------------------------------------------------------
DUPLICATE REPORT FOR z2e#2: z1e#2 (-37)
     vs z1e#2 (score = 222)
              GTTCGNATGNCACGCGTNACGTAAAGCTTGGGATCCTCTAGAGCGGCCGNGTACTACTACTACTG
        z1e#2                                                          CTTAAGCT

              AGCGGAATTCGTGAGACC-CTGAAACTGAGCTTAAGGAACTTTCCT-GAAGGGGAGAATGTTTGA
                 ||||   || | ||  || |||  | |||| |||||  | ||| ||||||||||||||||||
        z1e#2 CAAGGAA---GT-AAACAACTCAAA--G-GCTTCAGGAA-ATCCCTTGAAGGGGAGAATGTTTGA

              TCTCCCTGCAGAACCGGCTGTGTGACCTTTCTATCATGAAACACTCAGAATTCTGAATAGGACTT
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        z1e#2 TCTCCCTGCAGAACCGGCTGTGTGACCTTTCTATCATGAAACACTCAGAATTCTGAATAGGACTT

              TGGGTTATGATCCTCCCCAGAAATCCTTCACTTGCACTTTCTGTACTACAGGAGCCAGGCCACTG
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
        z1e#2 TGGGTTATGATCCTCCCCAGAAATCCTTCAC-TGCACTTTCTGTACTACAGGAGCCAGGCCACTG

              GTACAGACACTTCTGGCTCAGTCCCAAGTCCTNTAACCAACACATATGACATCATTAACAACCGG
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
        z1e#2 GTACAGACACTTCTGGCTCAGTCCCAAGTCCTGTAACCAACACATATGACATCATTAACAACAGG

              -TCTCACGGATTCGGTT
               ||| ||
        z1e#2 GTCTTACC
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7,9,18
END OF RECORD.
```

```
SEQUENCE FILENAME: z2f#2
FORWARD PRIMER: ATATTCCCAATGCTATGGAGCA
REVERSE PRIMER: CGACCAGGGTACCTCAAACCT
LENGTH OF INSERT: 216
------------------------------------------------------------------------
 ATTCGTGAGACCGACCCGACGATATTCCCAATGCTATGGAGCAATGGAATGGACATTTTGGACTTGGAT
 .............          >>>>>>>>>>>>>>>>>>>>>>
 TAAGCACTCTGGCTGGGCTGCTATAAGGGTTACGATACCTCGTTACCTTACCTGTAAAACCTGAACCTA
------------------------------------------------------------------------
 AGAGATCCACACTCTAGAACTCAGACTCTCAAAGGTATCGAAGCTGTTTTCCTCCAGCTCTTGCCTCTC

 TCTCTAGGTGTGAGATCTTGAGTCTGAGAGTTTCCATAGCTTCGACAAAAGGAGGTCGAGAACGGAGAG
------------------------------------------------------------------------
 AGGTTTGAGGTACCCTGGTCGTCATGGTAATAGGGTGGTGAAAAGCCACCCTCTGCACTTTCCAGGCAG
 <<<<<<<<<<<<<<<<<<<<
 TCCAAACTCCATGGGACCAGCAGTACCATTATCCCACCACTTTTCGGTGGGAGACGTGAAAGGTCCGTC
------------------------------------------------------------------------

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: TTCTCTTGCCCCTGGCCCCT
LENGTH OF INSERT: 326
------------------------------------------------------------------------
 AATTCGTGAGACCCTCTAAGGGTTCCCTCTATTTGCTAAGACTTCCTCCCATTGGGAGATCAGCAGAGG
 ............
 TTAAGCACTCTGGGAGATTCCCAAGGGAGATAAACGATTCTGAAGGAGGGTAACCCTCTAGTCGTCTCC
------------------------------------------------------------------------
 ATCTGTTGAGCAAAGTCGGCCCCTTCTGTCCCAGTAGCCTGGGCTTCCTGGCTTGAGTTAGCAGAGTCC
   >>>>>>>>>>>>>>>>>>>
 TAGACAACTCGTTTCAGCCGGGGAAGACAGGGTCATCGGACCCGAAGGACCGAACTCAATCGTCTCAGG
------------------------------------------------------------------------
 CATAAGCTGCTTAAGAGAGCTGGGCTAGCATAAGTTGCTTACGGTGGGGGGAGGGGCCAGGGGCAAGAG
                                                  <<<<<<<<<<<<<<<<<<
 GTATTCGACGAATTCTCTCGACCCGATCGTATTCAACGAATGCCACCCCCCTCCCCGGTCCCCGTTCTC
------------------------------------------------------------------------
 AACAGGGAAGGGGTACCAACCTGCCCCCATNCAGTNCATGCAAAGAGAAGTNCTCTNAAAGCCAGCCCT
 <<
 TTGTCCCTTCCCCATGGTTGGACGGGGGTATGTCANGTACGTTTCTCTTCANGAGANTTTCGGTCGGGA
------------------------------------------------------------------------
 GNCTACTTGGAAGCCCCCAGNCTGAGAGGGACTTTCACTCTNNAAGGGGNGCAGGGGAGCTCGGGTTTA
                                                                 ......
 CNGATGAACCTTCGGGGGTCNGACTCTCCCTGAAAGTGAGANNTTCCCCNCGTCCCCTCGAGCCCAAAT
------------------------------------------------------------------------
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: AAAAAGAAAGAGGGGCAGGG
LENGTH OF INSERT: 317
------------------------------------------------------------------------
 ATTCGTGAGACCGCGGCTNACCCCACGACCCTCTCCCAGCCCCATGATCGAGAGCACCCCCATGTCCAT
 ...........                                                  >>>>>>>
 TAAGCACTCTGGCGCCGANTGGGGTGCTGGGAGAGGGTCGGGGTACTAGCTCTCGTGGGGGTACAGGTA
------------------------------------------------------------------------
 TGCTGGCGATGCGACCCCACTCCTGGACGAGATGGATCGGTAGGGGGCCTGTTCCTCGGACTCTGGTCA
 >>>>>>>>>>>
 ACGACCGCTACGCTGGGGTGAGGACCTGCTCTACCTAGCCATCCCCCCGGACAAGGAGCCTGAGACCAGT
------------------------------------------------------------------------
 CCTCTGAAGCTCCCAAGAGGCCTGGCACTTATGGCTCACTCCCTGCCCCTCTTTCTTTTTGTCCATAAA
                                          <<<<<<<<<<<<<<<<<<<<
 GGAGACTTCGAGGGTTCTCCGGACCGTGAATACCGAGTGAGGGACGGGGAGAAAGAAAAACAGGTATTT
------------------------------------------------------------------------
 GTGGCGTGAAATGACGTTCTTTTTAAGTGGTCAAGCCTGGCTGGATGGTGGCCTTGTGGGTTGGTGTGC

 CACCGCACTTTACTGCAAGAAAAATTCACCAGTTCGGACCGACCTACCACCGGAACACCCAACCACACG
------------------------------------------------------------------------
 ACCGAGCAGCAGTTGTTCCACTNCGGGTTTGTGCTAAACANCTGTTCTTTGGGGGTCTCAACGAATTCG
                                                          ..............
 TGGCTCGTCGTCAACAAGGTGANGCCCAAACACGATTTGTNGACAAGAAACCCCCAGAGTTGCTTAAGC
------------------------------------------------------------------------
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: TCACCCACTGGGGACAGCCT
LENGTH OF INSERT: 298
--------------------------------------------------------------------
 GCGGAATTCGTGAGACCGCGGCTTAGCTGATGCTCTCTAGTCTGTTAAGTGCAGAAGCCGGAGTCTGAT
 .................                                  >>>>>>>>>>>>>>>>>
 CGCCTTAAGCACTCTGGCGCCGAATCGACTACGAGAGATCAGACAATTCACGTCTTCGGCCTCAGACTA
--------------------------------------------------------------------
 TCAGGCCTGTTGGGTTCCAAGTCAGAAATGTTTTCTCCTGTTCCACACTACCTNTCTGAAAGGGTAGAG
 >>>
 AGTCCGGACAACCCAAGGTTCAGTCTTTACAAAAGAGGACAAGGTGTGATGGANAGACTTTCCCATCTC
--------------------------------------------------------------------
 GGGAACCTTAGCCACAAGGTAGAAGATCCTAAGCCATTGCCTCCTAGACCAGGCTGTCCCCAGTGGGTG
                                                    <<<<<<<<<<<<<<<<<<<
 CCCTTGGAATCGGTGTTCCATCTTCTAGGATTCGGTAACGGAGGATCTGGTCCGACAGGGGTCACCCAC
--------------------------------------------------------------------
 ACGCCATAAGGAAGCTGAGGAAGGAGGGAGGTGGTCTCAGAAGTNGAGAGGACAAGGGCTACAATTCAA
 <
 TGCGGTATTCCTTCGACTCCTTCCTCCCTCCACCAGAGTCTTCANCTCTCCTGTTCCCGATGTTAAGTT
--------------------------------------------------------------------
 CTCAAGTACTCCCCTAAATTCANCTCTCAAGTGCGTTTTGGTCTCACGGATTNCCGGTNTNGGNNNNNN
                                          ..........................
 GAGTTCATGAGGGGATTTAAGTNGAGAGTTCACGCAAAACCAGAGTGCCTAANGGCCANANCCNNNNNN
--------------------------------------------------------------------

DUPLICATE REPORT FOR z4e#2:  z11g (7)
     vs z11g (score = 65)

             GTACTACTACTACTGAGCGGAATTCGTGAGACCGCGGCTTAGCTGATGCTCTCTAGTCTGTTAAG
                                                                 || |||| |||
         z11g GTGGGAAAGCACTGATGATGTCACCCTCAGAAAGGGGCTGGATTAAATGTGATAAGCCTGTAAAG

             TGCAGAAGCCGGAGTCTGATTCAGGCCTGTT-GGGTTCCAAGTCAGA-AATGTTTTCTCCTGTTC
             ||||||||||||||||||||||||||||| || |||||||||||| | || | ||||||||||||
         z11g TGCAGAAGCCGGAGTCTGATTCAGGCCT-TTCGGGTTCCAAGTCA-ANAACGNTTTCTCCTGTTC

             CACACTACCTNTCTGAAAGGGTAGAGGGGAACCTTAGCCACAAGGTAGAAGATCCTAAGCCATTG
             |||||
         z11g CACAC

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
 .............       >>>>>>>>>>>>>>>>>>>
 GCCTTAAGCACTCTGGCCCCGGGTGATGGCTCGTCCCTGACCCCCACCAGGTCACCGTTCGGTCTTGTC
--------------------------------------------------------------------
 CTTGATGCAAACGGTGGNNNANCTGAGTNGGGATTTCCAGCAGGCCTGCTTCCGAGAACAGAGCCTGGC

 GAACTACGTTTGCCACCNNNTNGACTCANCCCTAAAGGTCGTCCGGACGAAGGCTCTTGTCTCGGACCG
--------------------------------------------------------------------
 CATGTTGAGCAAAGCCACCGAGCTCCTGGCACCCGAGCCCCTGATCCCACAAGATAGTCGAAGTCTTCA
                  <<<<<<<<<<<<<<<<<<<
 GTACAACTCGTTTCGGTGGCTCGAGGACCGTGGGCTCGGGGACTAGGGTGTTCTATCAGCTTCAGAAGT
--------------------------------------------------------------------
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
--------------------------------------------------------------------------
 GAATTCGTGAGACCTAGCTGCTTGCTCTGGGATACACAGCAATTGAGAGGAGCTGGAGAAGGGGTTGGG
 ..............                              >>>>>>>>>>>>>>>>>>>>>
 CTTAAGCACTCTGGATCGACGAACGAGACCCTATGTGTCGTTAACTCTCCTCGACCTCTTCCCCAACCC
--------------------------------------------------------------------------
 GGGAGTCCTGACAGGCTTCGGAGCTCAGCTACTTACCTCCCAATTGAGAGAGGAAGGCAAGTNATGGAA

 CCCTCAGGACTGTCCGAAGCCTCGAGTCGATGAATGGAGGGTTAACTCTCTCCTTCCGTTCANTACCTT
--------------------------------------------------------------------------
 GTCTTTTCCGGCCTCAGACAAGGGTGCTATCTATGGGGCTGGAGGAATTCCAAGTCAAGAGCTGATCAC
                   <<<<<<<<<<<<<<<<<<<
 CAGAAAAGGCCGGAGTCTGTTCCCACGATAGATACCCCGACCTCCTTAAGGTTCAGTTCTCGACTAGTG
--------------------------------------------------------------------------
 TAAGGGTGACCNTCGATTGCCTTCCAAGAGTNAAANTCTNCATTNCACCTCTATGAGGNAAGATGTTAA

 ATTCCCACTGGNAGCTAACGGAAGGTTCTCANTTTNAGANGTAANGTGGAGATACTCCNTTCTACAATT
--------------------------------------------------------------------------
DUPLICATE REPORT FOR z5a: z1d#2 (20)
     vs z1d#2 (score = 163)
              CGTNNGCATGCACGCGTNACGTAAANCTTGGGATCCTCTAGAGCGGCCGTGTACTACTACTACTG
        z1d#2

              AGCGGAATTCGTGAGACCTAGCTGCTTGCTCTGGGATACACAGCAATTGAGAGGAGCTGGAGAAG
        z1d#2

              GGGTTGGGGGGAGTCCTGACAGGCTTCGGAGCTCAGCTACTTACCTCCCAATTGAGAGAGGAAGG
                || ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
        z1d#2   GTCGGGGGGAGTCCTGACAGGCTTCGGAGCTCAGCTACTTACCTCCCAATTGAGGGAGGAAGG

              CAAGTNATGGAAGTCTTTTCCGGCCTCAGACAAGGGTGCTATCTATGGGGCTGGAGGAATTCCAA
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
        z1d#2 CAAGTNATGGAAGTCTTTTCCGGCCTCAGAC-AGGGTGCTATCTATGGGGCTGGAGGAATTCCAA

              GTCAAGAGCTGATCACTAAGGGTGACCNTCGATT-GCCTTCCAAGAGTNAAANTCTN-CATTNCA
              ||| |||||||||||||| ||||||  |||||| ||||| | ||||||| | | |  || ||
        z1d#2 GTC-AGAGCTGATCACTAA-GGTGACGATCGATTTGCCTTNC-AGAGTNAGA-T-TAACA-TNTG

              CCTCTATGAGG
              ||||||
        z1d#2 CCTCTAG


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

```
SEQUENCE FILENAME: z5e
FORWARD PRIMER: CTCCTCGTCCAGGGATAGG
REVERSE PRIMER: GTGGCCCCAGGACAGGAAG
LENGTH OF INSERT: 171
-----------------------------------------------------------------------
 TGAGCGGAATTCGTGAGACCGGAACAGGCTCCTCGTCCAGGGATAGGTGGTCCAAGCCTAGCTTGTTGA
 .....................        >>>>>>>>>>>>>>>>>>>
 ACTCGCCTTAAGCACTCTGGCCTTGTCCGAGGAGCAGGTCCCTATCCACCAGGTTCGGATCGAACAACT
-----------------------------------------------------------------------
 ACCTGTTACTTCCACAGGCCAGGACCCTNCCTGGAGAGGTGAGGATCATGGAAGAGTCAATACCACAGA

 TGGACAATGAAGGTGTCCGGTCCTGGGANGGACCTCTCCACTCCTAGTACCTTCTCAGTTATGGTGTCT
-----------------------------------------------------------------------
 CAACTCTCTGAGCTTCCTGTCCTGGGGCCACAAGGACCTGCTTGAGGGCAGTTGGTCTCACGAATTCNG
           <<<<<<<<<<<<<<<<<<<                          ................
 GTTGAGAGACTCGAAGGACAGGACCCCGGTGTTCCTGGACGAACTCCCGTCAACCAGAGTGCTTAAGNC
-----------------------------------------------------------------------
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: GCCTTGTCATCCCATTTTGC
REVERSE PRIMER: GCCAAATCCCTACCTTGAGC
LENGTH OF INSERT: 354
-----------------------------------------------------------------------
 GAATTCGTGAGACCGATAGCCTTGTCATCCCATTTTGCACACCCTATTATCTGGGTTCTGCTCTCCTTT
 ..............    >>>>>>>>>>>>>>>>>>>>
 CTTAAGCACTCTGGCTATCGGAACAGTAGGGTAAAACGTGTGGGATAATAGACCCAAGACGAGAGGAAA
-----------------------------------------------------------------------
 CTTCCAGTTGGGACACGGCTGGTCAGGAAAGGTTCAAGTGCATTGCTTCCACCTGCTACCGTGGAGCTC
                                                                  <<<<
 GAAGGTCAACCCTGTGCCGACCAGTCCTTTCCAAGTTCACGTAACGAAGGTGGACGATGGCACCTCGAG
-----------------------------------------------------------------------
 AAGGTAGGGATTTGGCTCTGGAATGCAGGAAGTCTAAGAGCACATCAAGTCAAGTGTTCTGAGGAGAGG
 <<<<<<<<<<<<<<<<
 TTCCATCCCTAAACCGAGACCTTACGTCCTTCAGATTCTCGTGTAGTTCAGTTCACAAGACTCCTCTCC
-----------------------------------------------------------------------
 TGATGGACAAGGGAAAGGCAAAGACATCCATCCATTCATCTTCCTGTCAANCTGCAGCTGTTTCTGTCA

 ACTACCTGTTCCCTTTCCGTTTCTGTAGGTAGGTAAGTAGAAGGACAGTTNGACGTCGACAAAGACAGT
-----------------------------------------------------------------------
 CTCTTTCAAGCCATCATCATTGNCTTCAACCTNGAATGACCGTGGATCCCTTGGNACACACCAAGTATT

 GAGAAAGTTCGGTAGTAGTAACNGAAGTTGGANCTTACTGGCACCTAGGGAACCNTGTGTGGTTCATAA
-----------------------------------------------------------------------
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 yptm2 cDNA',4.3e-05]
    ['X75593','H.sapiens mRNA for rab 13',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 rgp2 - rice>gi 2...',5.6e-06]
    ['P35286','RAS-RELATED PROTEIN RAB-13 (FRAGMEN...',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
------------------------------------------------------------------------
 AATTCGTGAGACCGCTAAGGAACAGTGAGAATAAACCCCAGACCTTCAGTTATCAGCTCACATTCCCCC
 .............                                              >>>>>>>>>>>>
 TTAAGCACTCTGGCGATTCCTTGTCACTCTTATTTGGGGTCTGGAAGTCAATAGTCGAGTGTAAGGGGG
------------------------------------------------------------------------
 AAGGCGAGGAAAGGCNNGGGCTTAGTACTATTACTTTCAGAGCCCAACTGTGCAACAGAAGACAAGGTG
 >>>>>>
 TTCCGCTCCTTTCCGNNCCCGAATCATGATAATGAAAGTCTCGGGTTGACACGTTGTCTTCTGTTCCAC
------------------------------------------------------------------------
 CAAGAATTTACAACAAAACAGTCTAACGTTACAGAGAAGTGTTGAATTTCTAGACAACTAATTGCCACT
                                                 <<<<<<<<<<<<<<<<<<
 GTTCTTAAATGTTGTTTTGTCAGATTGCAATGTCTCTTCACAACTTAAAGATCTGTTGATTAACGGTGA
------------------------------------------------------------------------
 TNGGGAATTGAGACACATTTGGTCCTGGAACAGTTGGAAGNTGCTCTTNGAATGGGNAAAGGGGGGTCT
 <<<<                                                          .....
 ANCCCTTAACTCTGTGTAAACCAGGACCTTGTCAACCTTCNACGAGAANCTTACCCNTTTCCCCCCAGA
------------------------------------------------------------------------
DUPLICATE REPORT FOR z7d#2: z3b (32)
     vs z3b (score = 226)
               CGTCGCATGCACGCGTNACGTAAAGCTTGGGATCCTCTAGAGCGGCCGTNTACTACTACTACTGA
          z3b

               GCGGAATTCGTGAGACCGCTAAGGAACAGTGAGAATAAACCCCAGACCTTCAGTTATCAGCTCAC
                                 ||||||||||||||||||||||||||||||||||||||||||||||||
          z3b                    AAGGAACAGTGAGAATAAACCCCAGACCTTCAGTTATCAGCTCAC

               ATTCCCCCAAGGCGAGGAAAGGCNNGGGCTTAGTACTATTACTTTCAGAGCCCAACTGTGCAACA
               |||||||||||||||||||||||  |||||||||||||||||||||| ||||||||||||||||||
          z3b ATTCCCCCAAGGCGAGGAAAGGCAAGGGCTTAGTACTATTACTTTCAAAGCCCAACTGTGCAACA

               GAAGACAAGGTGCAAGAATTTACAACAAAACAGTCTAACGTTACA-GAGAAGTGTTGAATTTCTA
               |||||| |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
          z3b GAAGAC-AGGTGCAAGAATTTACAACAAAACAGTCTAACGTTACAAGAGAAGTGTTGAATTTCTA

               GACAACTAATTGCCACTTNGGGAATTGAGACACATTTGGTCCTGGAACAGTTGGAAGNT-GCTCT
               |||||||||||||||||| ||||| |||||||||||||||||||||||||  |||| | | |||
          z3b GACAACTAATTGCCACTT-GGGAA-TGAGACACATTTGGTCCTGGAACAGTANGAAGGTTG-TCT

               TNGAATGGGN-A----AAGGGGGGTCTCA--CGGAATTCCCGTTGGGNAGNGNNNTTTNNNNATT
               | |||||||  |    ||| |||| | ||  |  ||||
          z3b T-GAATGGGTGAGGGNAAGAGGGG-C-CATTCCCAATTGTAAGCCCTCTGCAGCG
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7,8,9,10
END OF RECORD.
```

```
SEQUENCE FILENAME: z7e#2
FORWARD PRIMER: TTCTCCTGGGACACTGCTCC
REVERSE PRIMER: CCCTGAAATGGATGCTGA
LENGTH OF INSERT: 288
--------------------------------------------------------------------------
 GGAATTCGTGAGACCGGTTCTCCTGGGACACTGCTCCATCACAGGCTCAAAGGACAAGCTTTACAAGAA
 .............. >>>>>>>>>>>>>>>>>>>>
 CCTTAAGCACTCTGGCCAAGAGGACCCTGTGACGAGGTAGTGTCCGAGTTTCCTGTTCGAAATGTTCTT
--------------------------------------------------------------------------
 TTCCTTCCCTGCTACAGGCTGGTGNGAAACCAATACCAGGGTTTGGAAGAGCAGAGGGGTTATTTTAGA

 AAGGAAGGGACGATGTCCGACCACNCTTTGGTTATGGTCCCAAACCTTCTCGTCTCCCCAATAAAATCT
--------------------------------------------------------------------------
 ATTGTCAGCATCCATTTCAGGGCAGTATGGGTTGGCAGGCTTTTNGACAGTGTCTTNGTCCTCACCCAT
      <<<<<<<<<<<<<<<<<<
 TAACAGTCGTAGGTAAAGTCCCGTCATACCCAACCGTCCGAAAANCTGTCACAGAANCAGGAGTGGGTA
--------------------------------------------------------------------------
 TTAGNCATGCCCAACCGGTTTGGTCTGCTTCCCTCAACCCTTGCCCAAGGACAGGGACTTCAAGAGAGT

 AATCNGTACGGGTTGGCCAAACCAGACGAAGGGAGTTGGGAACGGGTTCCTGTCCCTGAAGTTCTCTCA
--------------------------------------------------------------------------
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
--------------------------------------------------------------------------
 ATTCGTGAGACCAAAAAAAACTTGTCCTCAGAAACTTGAACCNNNTCTCTTAGTCTCAGCCACAGGAAA
 ............
 TAAGCACTCTGGTTTTTTTTTGAACAGGAGTCTTTGAACTTGGNNNAGAGAATCAGAGTCGGTGTCCTTT
--------------------------------------------------------------------------
 GGCCCCCTTCAGGTTGCCAAGTCTCACTCTTTCTCAATGAGAGGTTCATATAGTTCAAGCCCCAGCCTA
                                                 ***************
 CCGGGGGAAGTCCAACGGTTCAGAGTGAGAAAGAGTTACTCTCCAAGTATATCAAGTTCGGGGTCGGAT
--------------------------------------------------------------------------
 TTGGATGCGGTCAACTGTCTTAAGCATCCTGAGTAACTAGGGGTTTGCAACCCGCTACAGGGATGATGT
 ***************
 AACCTACGCCAGTTGACAGAATTCGTAGGACTCATTGATCCCCAAACGTTGGGCGATGTCCCTACTACA
--------------------------------------------------------------------------
 GTGTGTGTGTGTGTGTGTGNNTNTGTGTAANACCCCGTTCCCGCANANAGATNNNTATAGANAACCTTT

 CACACACACACACACACACNNANACACATTNTGGGGCAAGGGCGTNTNTCTANNNATATCTNTTGGAAA
--------------------------------------------------------------------------
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
 ..............                                                  >>>
 GCCTTAAGCACTCTGGCGTACCTCCTCGTTCGGTCACCTGATCCTCCCCTTCCCTCGGGTAGCACCTCA
--------------------------------------------------------------------
 GCCATCCGTAGAGAGCCCACGGGCANNAGTGAAGACATGTCATCACCAGCAGACCTTGCCTTGGTCTCT
 >>>>>>>>>>>>>>>>
 CGGTAGGCATCTCTCGGGTGCCCGTNNTCACTTCTGTACAGTAGTGGTCGTCTGGAACGGAACCAGAGA
--------------------------------------------------------------------
 TGGCCCAACACCCTCCCAGCGTGGCTAGAGAGGAGGGCTCCCTCTGACATGCTTGCCCGGATCAGTCTA
                                                 <<<<<<<<<<<<<<<
 ACCGGGTTGTGGGAGGGTCGCACCGATCTCTCCTCCCGAGGGAGACTGTACGAACGGGCCTAGTCAGAT
--------------------------------------------------------------------
 TCTTCTCAAGGCTTTTTCCAGGACTTNNCGAGTTTCCTGAAAGATCTTAACCAGGGACAGGAGGATTGG
 <<<<<
 AGAAGAGTTCCGAAAAAGGTCCTGAANNGCTCAAAGGACTTTCTAGAATTGGTCCCTGTCCTCCTAACC
--------------------------------------------------------------------
DUPLICATE REPORT FOR z8b#2: z10f (24)
     vs z10f (score = 221)

            TGAGCGGAATTCGTGAGACCGCATGGAGGAGCAAGCCAGTGGACTAGGAGGGGAAGGGAGCCCAT
                                          |||||||||| ||||||||||||||||||||||||
        z10f                            NAGCAAGCCAGTNGACTAGGAGGGGAAGGGAGCCCAT

            CGTGGAGTGCCATCCGTAGAGAGCCCACGG-GCANNAGTGAAGACATGTCATCACCAGCAGACCT
            |||||||||||||||||||||||||||||| ||   ||| ||||||||||||||||||||||||
        z10f CGTGGAGTGCCATCCGTAGAGAGCCCACGGCGC-GTAGTNAAGACATGTCATCACCAGCAGACCT

            TGCCTTGGTCTCTTGGCCCAACACCCTCCCAGCGTGGCTAGAGAGGAGGGCTCCCTCTGACATGC
            |||||||||||||| |||||||||||||| ||||||| |||||||||||||||||||||| ||||||
        z10f TGCCTTGGTCTCTNGGCCCAACACCCTNCCAGCGTNGCTAGAGAGGAGGGCTCCCTCTAACATGC

            TTGCCCGGATCAGTCTATCTTCTCAAGGCTTTTTCCAGGA-CTTNNCGAGTTTCCTGAAAGATCT
             ||||| |||||||||||||  ||| || ||    ||||  ||| |||| ||||||||||||||||
        z10f -TGCCCGAATCAGTCTATCTNNTCANGGNTTNNNNCAGGGNCTT-NCGANTTTCCTGAAAGATCT

            TAACCAGGGACA-GGAGGATTGGGACAGCATGG--CAAGGAA-GAGGGGAATGG-GNCTCACG-A
            |||||||||||| ||||| | || |||||||||  || |||| |||||||||   || | | | |
        z10f TAACCAGGGACAAGGAGGGTNGGCACAGCATGGNGCAGGGAAAGAGGGGAATTTAGN-T-ANGCA
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 11,12,13,22,23,27
END OF RECORD.
```

```
SEQUENCE FILENAME: z9a
FORWARD PRIMER: CTCAGCAGCAAGAGCAAGG
REVERSE PRIMER: GGGGCTCCCTCTTCAGGC
LENGTH OF INSERT: 355
--------------------------------------------------------------------
 AATTCGTGAGACCATCCTGCTGGCTCTGCATCACGTGCACACACATCTCATCCTGCATCGGGACCTCAA
 ............
 TTAAGCACTCTGGTAGGACGACCGAGACGTAGTGCACGTGTGTGTAGAGTAGGACGTAGCCCTGGAGTT
--------------------------------------------------------------------
 GACACAGAACATCCTTCTTGACAAACACCGCATGGTGGTCAAGATCGGTGACTTTGGCATCTCTAAGAT

 CTGTGTCTTGTAGGAAGAACTGTTTGTGGCGTACCACCAGTTCTAGCCACTGAAACCGTAGAGATTCTA
--------------------------------------------------------------------
 CCTCAGCAGCAAGAGCAAGGCCTACACGGTGCCTAAGACGTGCGGGGGACACTGGGGGTCCTGAGAGCT
  >>>>>>>>>>>>>>>>>>>
 GGAGTCGTCGTTCTCGTTCCGGATGTGCCACGGATTCTGCACGCCCCCTGTGACCCCCAGGACTCTCGA
--------------------------------------------------------------------
 GAGGTTGCTCCACAGTCCTGCCCCTTCCCCTTCCCTCAAGTCCCTGTCTAGATGGCCTGAAGAGGGAGC
                                                    <<<<<<<<<<<<<<<
 CTCCAACGAGGTGTCAGGACGGGGAAGGGGAAGGGAGTTCAGGGACAGATCTACCGGACTTCTCCCTCG
--------------------------------------------------------------------
 CCCAAGGCCTGCTTTGAGAAGNTATCCCCAAATGAGTCTTTGGCCAAGAGCTTTTGTCCCTNTTCTCCC
 <<<
 GGGTTCCGGACGAAACTCTTCNATAGGGGTTTACTCAGAAACCGGTTCTCGAAAACAGGGANAAGAGGG
--------------------------------------------------------------------

DUPLICATE REPORT FOR z9a: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 7,9
END OF RECORD.
--------------------------------------------------------------------
SEQUENCE FILENAME: k10d.Seq
FORWARD PRIMER: CCTGTGTGGGGAGCTGTG
REVERSE PRIMER: TAGTCTAACTATGGACAGAG
LENGTH OF INSERT: 179
--------------------------------------------------------------------
 GTTCCTTCCTGTGTGGGGAGCTGTGCTCTCAGCACATGGCAGGAAGTGAGACCCAACATCACAAACAAG
        >>>>>>>>>>>>>>>>>>
 CAAGGAAGGACACACCCCTCGACACGAGAGTCGTGTACCGTCCTTCACTCTGGGTTGTAGTGTTTGTTC
--------------------------------------------------------------------
 AAAGCTCTGCCCGTTCTCTGTGTTNGGCAGTGTTNCTCAGGAGATTTCAAAGGCTATGTGGCACAGCTC
                                                              <<<
 TTTCGAGACGGGCAAGAGACACAANCCGTCACAANGAGTCCTCTAAAGTTTCCGATACACCGTGTCGAG
--------------------------------------------------------------------
 TGTCCATAGTTAGACTACCTCTGAGCTTCACAGCGGTGCAG
 <<<<<<<<<<<<<<<<<
 ACAGGTATCAATCTGATGGAGACTCGAAGTGTCGCCACGTC
--------------------------------------------------------------------
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: CACCTCAGCTGGCCTTCA
REVERSE PRIMER: GGGCCTTGTGCCACACTTGG
LENGTH OF INSERT: 265

```
CAGGGAGTCTGTGGGATTCTATGGGAAGCCACGGTGTTGAATCTTACTTCAGGTCGTCACCTCAGCTGG
                                                          >>>>>>>>>>>>
GTCCCTCAGACACCCTAAGATACCCTTCGGTGCCACAACTTAGAATGAAGTCCAGCAGTGGAGTCGACC

CCTTCATCAAACATTTCCAGTCTTACAGGGCCCTGGCTGCCACTCTGTGGGAGGCACTGATATGACGGG
>>>>>>
GGAAGTAGTTTGTAAAGGTCAGAATGTCCCGGGACCGACGGTGAGACACCCTCCGTGACTATACTGCCC

AAAGCCACCTTAGAAAAGAGCATGATAGTCTCGATGCCAAGTGTGGCACAAGGCCCCAGAGTTGCACAT
                                        <<<<<<<<<<<<<<<<<<<<
TTTCGGTGGAATCTTTTCTCGTACTATCAGAGCTACGGTTCACACCGTGTTCCGGGGTCTCAACGTGTA

GTGACCAGTACAGAGCTAGAACACAGATGTGTATCTNTNAGGGNNNTTTAAGATTTGGG

CACTGGTCATGTCTCGATCTTGTGTCTACACATAGANANTCCCNNAAATTCTAAACCC
```

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: TAAAGTGGCGTGAAATGACGT
REVERSE PRIMER: TCCCAGCCTGGCCCCAGAAC
LENGTH OF INSERT: 271
------------------------------------------------------------------------
 CCTGTNCCTCGGACTCTGGTTACCTCTGAAGCTCCCAAGAGGCCTGGCCACTTATGGCCTCACTCCCTG

 GGACANGGAGCCTGAGACCAATGGAGACTTCGAGGGTTCTCCGGACCGGTGAATACCGGAGTGAGGGAC
------------------------------------------------------------------------
 CCCCTCTTCTTTTTTTGTCCATAAAGTGGCGTGAAATGACGTTCTTTTTAGTGGTCAGCCTGGCTGGAT
                       >>>>>>>>>>>>>>>>>>>>>
 GGGGAGAAGAAAAAAACAGGTATTTCACCGCACTTTACTGCAAGAAAAATCACCAGTCGGACCGACCTA
------------------------------------------------------------------------
 GGTGGCCTTGTGAGCTGGTGTGCACCGAGCAGCAGCTGTTCCACTCCGGGCTTGTGCTAACCACCTGTT
                                                                    <<<
 CCACCGGAACACTCGACCACACGTGGCTCGTCGTCGACAAGGTGAGGCCCGAACACGATTGGTGGACAA
------------------------------------------------------------------------
 CTGGGGCCAGGCTGGGAGAGCAGTATGGCAGGGAGGAGAGCAAAGGTGGGAGCAATCTGTAAAT
 <<<<<<<<<<<<<<<<<
 GACCCCGGTCCGACCCTCTCGTCATACCGTCCCTCCTCTCGTTTCCACCCTCGTTAGACATTTA
------------------------------------------------------------------------

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

                                                    CCTGTNCCTCGGACTCTGGTTACCTC
                                                    ||||| ||||||||||||| |||||
       z3e    ATGCGACCCCACTCCTGGACGAGATGGATCGGTAGGGGGCCTGTTCCTCGGACTCTGGTCACCTC

              TGAAGCTCCCAAGAGGCCTGGCCACTTATGGCCTCACTCCCTGCCCCTCTTCTTTTTTTGTCCAT
              |||||||||||||||||||| ||||||||| |||||||||||||||||| | ||||||||||
       z3e    TGAAGCTCCCAAGAGGCCTGG-CACTTATGG-CTCACTCCCTGCCCCTCTT-TCTTTTTGTCCAT

              AAAGTGGCGTGAAATGACGTTCTTTTTA-GTGGTCA-GCCTGGCTGGATGGTGGCCTTGTGAGCT
              |||||||||||||||||||||||||||| ||||||| ||||||||||||||||||||||| | |
       z3e    AAAGTGGCGTGAAATGACGTTCTTTTTAAGTGGTCAAGCCTGGCTGGATGGTGGCCTTGTGGGTT

              GGTGTGCACCGAGCAGCAGCTGTTCCACTCCGGGCTTGTGCTAACCACCTGTTCT--GGGGCCAG
              |||||||||||||||||| ||||||||| |||| ||||||||| || ||||||| |||
       z3e    GGTGTGCACCGAGCAGCAGTTGTTCCACTNCGGGTTTGTGCTAAACANCTGTTCTTTGGG

              GCTGGGAGAGCAGTATGGCAGGGAGGAGAGCAAAGGTGGGAGCAATCTGTAAAT

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 9,18,19
END OF RECORD.
```

```
SEQUENCE FILENAME: k1d.Seq
PRIMERS NOT PICKED
LENGTH OF INSERT: 101
--------------------------------------------------------------------------
 ACTAGGCAAAGGTGAGGGGCAGNGAGAACATGNAATTCAAGAGCCACACTAGCCTCACAAAAGCCTGTG

 TGATCCGTTTCCACTCCCCGTCNCTCTTGTACNTTAAGTTCTCGGTGTGATCGGAGTGTTTTCGGACAC
--------------------------------------------------------------------------
 CTCATTAATTATGTCAGCAGAGTAGTTTCAGG

 GAGTAATTAATACAGTCGTCTCATCAAAGTCC
--------------------------------------------------------------------------
DUPLICATE REPORT FOR k1d.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: TGCACTGGGAGTGGAGCC
REVERSE PRIMER: CCAGTGTCATGCAGAGGTCAG
LENGTH OF INSERT: 239
--------------------------------------------------------------------------
 CTCGGGCGGGTCATTGTAGGACCCCTATAATACAACTNAGAACATGCACTGGGAGTGGAGCCACCTTTN
                                                  >>>>>>>>>>>>>>>>>>>
 GAGCCCGCCCAGTAACATCCTGGGGATATTATGTTGANTCTTGTACGTGACCCTCACCTCGGTGGAAAN
--------------------------------------------------------------------------
 TTTTCATTGGAAACCTGGTCAGTTTGTGTTCAGAGGGGACCTGGATGTCAACGTAAGAACAGCTTCTAA

 AAAAGTAACCTTTGGACCAGTCAAACACAAGTCTCCCCTGGACCTACAGTTGCATTCTTGTCGAAGATT
--------------------------------------------------------------------------
 GTATGACGACTGCTGCTTCCTTCACCTGCCTGACCTCTGCATGACACTGGACCTACAGTGGCTGTACCA
                                  <<<<<<<<<<<<<<<<<<<<
 CATACTGCTGACGACGAAGGAAGTGGACGGACTGGAGACGTACTGTGACCTGGATGTCACCGACATGGT
--------------------------------------------------------------------------
 CGGGGAATCCTCATGATCATCACAGTGTCACG

 GCCCCTTAGGAGTACTAGTAGTGTCACAGTGC
--------------------------------------------------------------------------
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: K10F.Seq
FORWARD PRIMER: CACCTCAGCTGGCCTTCA
REVERSE PRIMER: GGGCCTTGTGCCACACTTGG
LENGTH OF INSERT: 265
--------------------------------------------------------------------------
 CAGGGAGTCTGTGGGATTCTATGGGAAGCCACGGTGTTGAATCTTACTTCAGGTCGTCACCTCAGCTGG
                                                          >>>>>>>>>>>>
 GTCCCTCAGACACCCTAAGATACCCTTCGGTGCCACAACTTAGAATGAAGTCCAGCAGTGGAGTCGACC
--------------------------------------------------------------------------
 CCTTCATCAAACATTTCCAGTCTTACAGGGCCCTGGCTGCCACTCTGTGGGAGGCACTGATATGACGGG
 >>>>>>
 GGAAGTAGTTTGTAAAGGTCAGAATGTCCCGGGACCGACGGTGAGACACCCTCCGTGACTATACTGCCC
--------------------------------------------------------------------------
 AAAGCCACCTTAGAAAAGAGCATGATAGTCTCGATGCCAAGTGTGGCACAAGGCCCCAGAGTTGCACAT
                                          <<<<<<<<<<<<<<<<<<<<
 TTTCGGTGGAATCTTTTCTCGTACTATCAGAGCTACGGTTCACACCGTGTTCCGGGGTCTCAACGTGTA
--------------------------------------------------------------------------
 GTGACCAGTACAGAGCTAGAACACAGATGTGTATCTNTNAGGGNNNTTTAAGATTTGGG

 CACTGGTCATGTCTCGATCTTGTGTCTACACATAGANANTCCCNNAAATTCTAAACCC
--------------------------------------------------------------------------
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: TAAAGTGGCGTGAAATGACGT
REVERSE PRIMER: TCCCAGCCTGGCCCCAGAAC
LENGTH OF INSERT: 271
--------------------------------------------------------------------------
 CCTGTNCCTCGGACTCTGGTTACCTCTGAAGCTCCCAAGAGGCCTGGCCACTTATGGCCTCACTCCCTG

 GGACANGGAGCCTGAGACCAATGGAGACTTCGAGGGTTCTCCGGACCGGTGAATACCGGAGTGAGGGAC
--------------------------------------------------------------------------
 CCCCTCTTCTTTTTTTGTCCATAAAGTGGCGTGAAATGACGTTCTTTTTAGTGGTCAGCCTGGCTGGAT
                    >>>>>>>>>>>>>>>>>>>>>
 GGGGAGAAGAAAAAAACAGGTATTTCACCGCACTTTACTGCAAGAAAAATCACCAGTCGGACCGACCTA
--------------------------------------------------------------------------
 GGTGGCCTTGTGAGCTGGTGTGCACCGAGCAGCAGCTGTTCCACTCCGGGCTTGTGCTAACCACCTGTT
                                                                      <<<
 CCACCGGAACACTCGACCACACGTGGCTCGTCGTCGACAAGGTGAGGCCCGAACACGATTGGTGGACAA
--------------------------------------------------------------------------
 CTGGGGCCAGGCTGGGAGAGCAGTATGGCAGGGAGGAGAGCAAAGGTGGGAGCAATCTGTAAAT
 <<<<<<<<<<<<<<<<<
 GACCCCGGTCCGACCCTCTCGTCATACCGTCCCTCCTCTCGTTTCCACCCTCGTTAGACATTTA
--------------------------------------------------------------------------

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

                                                        CCTGTNCCTCGGACTCTGGTTACCTC
                                                        ||||| ||||||||||||| |||||
       z3e    ATGCGACCCCACTCCTGGACGAGATGGATCGGTAGGGGGCCTGTTCCTCGGACTCTGGTCACCTC

              TGAAGCTCCCAAGAGGCCTGGCCACTTATGGCCTCACTCCCTGCCCCTCTTCTTTTTTTGTCCAT
              |||||||||||||||||||| ||||||||| |||||||||||||||||| | ||||||||||
       z3e    TGAAGCTCCCAAGAGGCCTGG-CACTTATGG-CTCACTCCCTGCCCCTCTT-TCTTTTTGTCCAT

              AAAGTGGCGTGAAATGACGTTCTTTTTA-GTGGTCA-GCCTGGCTGGATGGTGGCCTTGTGAGCT
              |||||||||||||||||||||||||||| ||||||| |||||||||||||||||||| | |
       z3e    AAAGTGGCGTGAAATGACGTTCTTTTTAAGTGGTCAAGCCTGGCTGGATGGTGGCCTTGTGGGTT

              GGTGTGCACCGAGCAGCAGCTGTTCCACTCCGGGCTTGTGCTAACCACCTGTTCT--GGGGCCAG
              |||||||||||||||||| ||||||||| |||| ||||||||| || ||||||| |||
       z3e    GGTGTGCACCGAGCAGCAGTTGTTCCACTNCGGGTTTGTGCTAAACANCTGTTCTTTGGG

              GCTGGGAGAGCAGTATGGCAGGGAGGAGAGCAAAGGTGGGAGCAATCTGTAAAT

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 9,18,19
END OF RECORD.
```

```
SEQUENCE FILENAME: k12d.Seq
FORWARD PRIMER: TGGATTAGCCCTCTTGGGG
REVERSE PRIMER: GGATAACAATTTTTGCATAAC
LENGTH OF INSERT: 289
------------------------------------------------------------------
 AAAAATCAAAATCATTTAAACGAAGAACTTGGATTAGCCCTCTTGGGGGATGGGTTTGTGCATCATCTA
                                >>>>>>>>>>>>>>>>>>>
 TTTTTAGTTTTAGTAAATTTGCTTCTTGAACCTAATCGGGAGAACCCCCTACCCAAACACGTAGTAGAT
------------------------------------------------------------------
 TTCAGATGCTAATTTCTATGCTTCTCTTTAACTNCAGTCCAGTAGTTGGCCAGGTTAAGCACTTCCAGT

 AAGTCTACGATTAAAGATACGAAGAGAAATTGANGTCAGGTCATCAACCGGTCCAATTCGTGAAGGTCA
------------------------------------------------------------------
 TTCAGTGTTATGCAAAAATTGTTATCCGTCACCTCTGATTGGTTAATAAAGAGCTGATTCAGCCAATGA
       <<<<<<<<<<<<<<<<<<<<
 AAGTCACAATACGTTTTTAACAATAGGCAGTGGAGACTAACCAATTATTTCTCGACTAAGTCGGTTACT
------------------------------------------------------------------
 CTGAGCAGAGGAGAGAGGGAAGGCAGGACTTCCCGTCCCAGTCATGGGGTCCCACAGAGGAGTNAGAAG

 GACTCGTCTCCTCTCTCCCTTCCGTCCTGAAGGGCAGGGTCAGTACCCCAGGGTGTCTCCTCANTCTTC
------------------------------------------------------------------
 GGGTTGGGGGAGG

 CCCAACCCCCTCC
------------------------------------------------------------------
DUPLICATE REPORT FOR k12d.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
------------------------------------------------------------------
 CAGATGAAGACATCCTGTGTCTCANGTTTCATCCCTTGCAGGTGAGTAGAATTGGCAAGTAGGCCAGAG

 GTCTACTTCTGTAGGACACAGAGTNCAAAGTAGGGAACGTCCACTCATCTTAACCGTTCATCCGGTCTC
------------------------------------------------------------------
 GGTAAGAAGGCACGGTAACCTTACTGCTGCTGTGTTGCTAGGAGCATGTTGTGAAAGAGGAGCTGCTTA
                                >>>>>>>>>>>>>>>>>>>>
 CCATTCTTCCGTGCCATTGGAATGACGACGACACAACGATCCTCGTACAACACTTTCTCCTCGACGAAT
------------------------------------------------------------------
 ACGCCTTGTACTGTGAATTCATCAACCGGGTCAATGAGGTCGGGGTGGATGTCAACCGTGCCATTTGCC

 TGCGGAACATGACACTTAAGTAGTTGGCCCAGTTACTCCAGCCCCACCTACAGTTGGCACGGTAAACGG
------------------------------------------------------------------
 CATCCTTACAGNCAGGCCTTGATCCAGTATGTCTGTGGGCTTGGGACCCCCGAAAAGGGGACCCACCTT
       <<<<<<<<<<<<<<<<<<<
 GTAGGAATGTCNGTCCGGAACTAGGTCATACAGACACCCGAACCCTGGGGGCTTTTCCCCTGGGTGGAA
------------------------------------------------------------------

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: GCAAGGGGAAACCATTTAN
REVERSE PRIMER: AGGCTAGTGTGGCTCTTGAA
LENGTH OF INSERT: 188
--------------------------------------------------------------------
 AAACAAGCAAGGGGGAAACCATTTANAGTCAATCTAGAGAACAAGGGCTGGCAATTACGTAGCAAAGAT
       >>>>>>>>>>>>>>>>>>>>
 TTTGTTCGTTCCCCCTTTGGTAAATNTCAGTTAGATCTCTTGTTCCCGACCGTTAATGCATCGTTTCTA
--------------------------------------------------------------------
 AAAACAGGTCACAGCGACCACTAGGCAANGGTGAGGGACAGNGNGANCATGAAATTCAAGAGCCACACT
                                                  <<<<<<<<<<<<<<<
 TTTTGTCCAGTGTCGCTGGTGATCCGTTNCCACTCCCTGTCNCNCTNGTACTTTAAGTTCTCGGTGTGA
--------------------------------------------------------------------
 AGCCTCACAAAAGCCTGTGCTCATTACTATGTCAGCAGTGTAGTTTCAGG
 <<<<<
 TCGGAGTGTTTTCGGACACGAGTAATGATACAGTCGTCACATCAAAGTCC
--------------------------------------------------------------------
DUPLICATE REPORT FOR k2d.Seq: k1d(8)
     vs k1d (score = 85)
            AAACAAGCAAGGGGGAAACCATTTANAGTCAATCTAGAGAACAAGGGCTGGCAATTACGTAGCAA
       k1d

            AGATAAAACAGGTCACAGCGACCACTAGGCAANGGTGAGGGACAGNGNGANCATGAAATTCAAGA
                                  ||||||||| |||||||| ||||| || |||| |||||||||
       k1d                        ACTAGGCAAAGGTGAGGGGCAGNGAGAACATGNAATTCAAGA

            GCCACACTAGCCTCACAAAAGCCTGTGCTCATTACT-ATGTCAGCAGTGTAGTTTCAGG
            |||||||||||||||||||||||||||||||||| | |||||||||| ||||||||||
       k1d  GCCACACTAGCCTCACAAAAGCCTGTGCTCATTAATTATGTCAGCAGAGTAGTTTCAGG

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
--------------------------------------------------------------------
SEQUENCE FILENAME: k3d.Seq
FORWARD PRIMER: GGATTTTTGTTTGTTGAGGCA
REVERSE PRIMER: GCGNTGGTGGTAGAGCAGGA
LENGTH OF INSERT: 207
--------------------------------------------------------------------
 CCACCCGACTGCTAATTTGGATTTTTGTTTGTTGAGGCAATGTCTCACTGTGTAGCACTGGCTGGTCTG
                   >>>>>>>>>>>>>>>>>>>>
 GGTGGGCTGACGATTAAACCTAAAAACAAACAACTCCGTTACAGAGTGACACATCGTGACCGACCAGAC
--------------------------------------------------------------------
 AAATNCTCTAGGTAGAGTGGCTGTCTTCAGAGATCCTGCCTCTCTCTGCCTCCAGAATATTAAGANTAG

 TTTANGAGATCCATCTCACCGACAGAAGTCTCTAGGACGGAGAGAGACGGAGGTCTTATAATTCTNATC
--------------------------------------------------------------------
 ATGCCTCCTGCTCTACCACCANCGCCCAGCCTAACTCCATTTTTATTTAGTAGAGCCAGAGCAGGAAAC
      <<<<<<<<<<<<<<<<<<<
 TACGGAGGACGAGATGGTGGTNGCGGGTCGGATTGAGGTAAAAATAAATCATCTCGGTCTCGTCCTTTG
--------------------------------------------------------------------
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: ACTAATGGGCAGGCGTGG
REVERSE PRIMER: CTCTAACACTGAAGTAATTA
LENGTH OF INSERT: 285
--------------------------------------------------------------------
 GAGCACTTTTAGTGTGTCTCGTATTGTTCTAGATACAGGAAATACAGTAGTGAATCAAACAATAAAACA

 CTCGTGAAAATCACACAGAGCATAACAAGATCTATGTCCTTTATGTCATCACTTAGTTTGTTATTTTGT
--------------------------------------------------------------------
 AGACTAATGGGCAGGCGTGGTGGCCTAGGTCTGTAATTGTAGCACTCAAAAGGTTGAAGCAGGATTGTA
   >>>>>>>>>>>>>>>>>>
 TCTGATTACCCGTCCGCACCACCGGATCCAGACATTAACATCGTGAGTTTTCCAACTTCGTCCTAACAT
--------------------------------------------------------------------
 ACTTCAAGATCAGCCTAGCCTATATAGTGGGGGCTGTCTCAAAACAACAGAAGTTCAAAGGCTGGTAAT
                                                                  <<<<
 TGAAGTTCTAGTCGGATCGGATATATCACCCCCGACAGAGTTTTGTTGTCTTCAAGTTTCCGACCATTA
--------------------------------------------------------------------
 TACTTCAGTGTTAGAGTATTTGCCGAGCATACGGGAAAGCTTCTGCCCCCCACACTNGGGGGAAATAAA
 <<<<<<<<<<<<<<<<
 ATGAAGTCACAATCTCATAAACGGCTCGTATGCCCTTTCGAAGACGGGGGGTGTGANCCCCCTTTATTT
--------------------------------------------------------------------
 AGCCCTGAA

 TCGGGACTT
--------------------------------------------------------------------
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
--------------------------------------------------------------------
 GGCTGGNGTAGCATTNANAGAAGCAGGGGTCCGTNTTCTGTAACTACTACTGGNAGTAATGCCTGGNAC

 CCGACCNCATCGTAANTNTCTTCGTCCCCAGGCANAAGACATTGATGATGACCNTCATTACGGACCNTG
--------------------------------------------------------------------
 AGNATCCTAATAGTAATCCTTAAACCACCTGAAGAGTCCATTCACAGTTGGGAAAATCTGGCCCCGGTA
                    >>>>>>>>>>>>>>>>>>>>>
 TCNTAGGATTATCATTAGGAATTTGGTGGACTTCTCAGGTAAGTGTCAACCCTTTTAGACCGGGGCCAT
--------------------------------------------------------------------
 CCGGAATCCTTCAGGAGTCACTGTTACATATTCTATCCTGGGTTTACCCCGGGGCTGGTATCCNAGGAG
            <<<<<<<<<<<<<<<<<<<
 GGCCTTAGGAAGTCCTCAGTGACAATGTATAAGATAGGACCCAAATGGGGCCCCGACCATAGGNTCCTC
--------------------------------------------------------------------
 GAACTTNCCGGGCNGCTCCTTGNNGGCNNAGCTGNAGTANGGACTGNAGGTGGGCTTNTCCTNCTNNGT

 CTTGAANGGCCCGNCGAGGAACNNCCGNNTCGACNTCATNCCTGACNTCCACCCGAANAGGANGANNCA
--------------------------------------------------------------------
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: TAACCAATCAGAGGTGACGGA
REVERSE PRIMER: GGGNATGGGTTTGTGCATC
LENGTH OF INSERT: 331
--------------------------------------------------------------------
 CTCTGCTCCACCTGTGTTCTCATGGCCACTCCTGGTNCCTCCATATCTTCTTTTCTTCTTCCTCCCCCA

 GAGACGAGGTGGACACAAGAGTACCGGTGAGGACCANGGAGGTATAGAAGAAAAGAAGAAGGAGGGGGT
--------------------------------------------------------------------
 ACCCCTTCTCACTCCTCTGTGGGACCCCATGACTGGGACGGGAAGTCCTGCCTTCCCTCTCTCCTCTGC

 TGGGGAAGAGTGAGGAGACACCCTGGGGTACTGACCCTGCCCTTCAGGACGGAAGGGAGAGAGGAGACG
--------------------------------------------------------------------
 TCAGTCATTGGCTGAATCAGCTCTTTATTAACCAATCAGAGGTGACGGATAACAATTTTTGCATAACAC
                              >>>>>>>>>>>>>>>>>>>>
 AGTCAGTAACCGACTTAGTCGAGAAATAATTGGTTAGTCTCCACTGCCTATTGTTAAAAACGTATTGTG
--------------------------------------------------------------------
 TGAAACTGGAAGTGCTTAACCTGGGCAACTTCTGGACCGCAGTTAAAGAGAAGCATAGAAATTAGGCAT

 ACTTTGACCTTCACGAATTGGACCCGTTGAAGACCTGGCGTCAATTTCTCTTCGTATCTTTAATCCGTA
--------------------------------------------------------------------
 CTGAATAGGTGATGCACAAACCCATNCCCCAAGAGGGCTAATCCAAGTTCTTTGA
          <<<<<<<<<<<<<<<<<<
 GACTTATCCACTACGTGTTTGGGTANGGGGTTCTCCCGATTAGGTTCAAGAAACT
--------------------------------------------------------------------
DUPLICATE REPORT FOR K5F.Seq: k12d   (-40)
     vs k12d   (score = 252)
               CTCTGCTCCACCTGTGTTCTCATGGCCACTCCTGGTNCCTCCATATCTTCTTTTCTTCTTCCTCC
                                                                            |||||
        k12d                                                                CCTCC

               CCCAACCCCTTCTCACTCCTCTGTGGGACCCCATGACTGGGACGGGAAGTCCTGCCTTCCCTCTC
               |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
        k12d   CCCAACCCCTTCTNACTCCTCTGTGGGACCCCATGACTGGGACGGGAAGTCCTGCCTTCCCTCTC

               TCCTCTGCTCAGTCATTGGCTGAATCAGCTCTTTATTAACCAATCAGAGGTGACGGATAACAATT
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        k12d   TCCTCTGCTCAGTCATTGGCTGAATCAGCTCTTTATTAACCAATCAGAGGTGACGGATAACAATT

               TTTGCATAACACTGAAACTGGAAGTGCTTAACCTGGGCAACTTCTGGACCGCAGTTAAAGAGAAG
               |||||||||||||||||||||||||||||||||| ||||| |||||| | ||||||||||||||
        k12d   TTTGCATAACACTGAAACTGGAAGTGCTTAACCTGGCCAACTACTGGACTGNAGTTAAAGAGAAG

               CATAGAAATTAGGCATCTGAATAGGTGATGCACAAACCCATNCCCCAAGAGGGCTAATCCAAGTT
               |||||||||||| |||||||||||| |||||||||||||||| ||||||||||||||||||||||
        k12d   CATAGAAATTA-GCATCTGAATAGATGATGCACAAACCCATCCCCCAAGAGGGCTAATCCAAGTT

               CTTTGA
               |||
        k12d   CTTCGTTTAAATGATTTTGATTTTT

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
--------------------------------------------------------------------
 CAGGNACCATGCCAANNACTGGATGACCCTCCGTNCCNNTGNNACTCTTCAGNGTNCCTNGTGAGCGGC

 GTCCNTGGTACGGTTNNTGACCTACTGGGAGGCANGGNNACNNTGAGAAGTCNCANGGANCACTCGCCG
--------------------------------------------------------------------
 AGNGTTCTTCATGNTNTNTCTCANNCTGCNGGGNCTTTCATGTATAAAANNNNTNAATNNTNNNAGCCN

 TCNCAAGAAGTACNANANAGAGTNNGACGNCCCNGAAAGTACATATTTTNNNNANTTANNANNNTCGGN
--------------------------------------------------------------------
 TACANNNANCCTTNTNCANTAAAGAGCNTTATAAGCAGAC

 ATGTNNNTNGGAANANGTNATTTCTCGNAATATTCGTCTG
--------------------------------------------------------------------
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: CTTTTTTCTTTTTTTTGTGGGGA
REVERSE PRIMER: GGCATGTGCTATCTGTCACAT
LENGTH OF INSERT: 257
--------------------------------------------------------------------
 TCCTGAAGACCAAGAGATTTGTTAAGAGGGAGAGAAGAAAAGGAATTTTCTTTCTTTCTTTTTTCTTTT
                                                      >>>>>>>>>>>>
 AGGACTTCTGGTTCTCTAAACAATTCTCCCTCTCTTCTTTTCCTTAAAAGAAAGAAAGAAAAAAGAAAA
--------------------------------------------------------------------
 TTTGTGGGGAACGGTGGTGCCAGGGATTGAACTCATTAATCCAATACCTGACACTGAACCGAATACCCA
 >>>>>>>>>>
 AAACACCCCTTGCCACCACGGTCCCTAACTTGAGTAATTAGGTTATGGACTGTGACTTGGCTTATGGGT
--------------------------------------------------------------------
 GCTCAGGAGTTCTCAAGGATGATGTCGCTATGTGACAGATAGCACATGCCTGTAAATCCCAGCACTTCA
                                 <<<<<<<<<<<<<<<<<<<<
 CGAGTCCTCAAGAGTTCCTACTACAGCGATACACTGTCTATCGTGTACGGACATTTAGGGTCGTGAAGT
--------------------------------------------------------------------
 GGCAGTNGCAAGAGGACAGAAAAGTAAAAAGGATTCAAGGCCAAACNTTG

 CCGTCANCGTTCTCCTGTCTTTTCATTTTTCCTAAGTTCCGGTTTGNAAC
--------------------------------------------------------------------
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
--------------------------------------------------------------------
 CGGGTCCTCTTACGGATTGGCTCAGATCAGCAGGGGAAGACTCCGACTTCACTGGGGTTTTCTTCTTAG
                                                                >>>>>
 GCCCAGGAGAATGCCTAACCGAGTCTAGTCGTCCCCTTCTGAGGCTGAAGTGACCCCAAAAGAAGAATC
--------------------------------------------------------------------
 TTAACTTGCCCTTGGCTTTTTTCTTCAAGTATGCCTTTTCATCCACCATCTCCTCCTCATCTTGCTCTT
 >>>>>>>>>>>>>>>>>
 AATTGAACGGGAACCGAAAAAAGAAGTTCATACGGAAAAGTAGGTGGTAGAGGAGGAGTAGAACGAGAA
--------------------------------------------------------------------
 TCTCCTNTCTCTNTATATCCTTGATGACGTNGCTTTCCTGACACTTTNCTATGGGACTTTCAGAACGCC
                                    <<<<<<<<<<<<<<<<<<<<<<
 AGAGGANAGAGANATATAGGAACTACTGCANCGAAAGGACTGTGAAANGATACCCTGAAAGTCTTGCGG
--------------------------------------------------------------------
 AGATAGTGA

 TCTATCACT
--------------------------------------------------------------------
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: TAGCATGGGTACCCCTCAAT
REVERSE PRIMER: ATGACAACTTCACATCTCCC
LENGTH OF INSERT: 267
--------------------------------------------------------------------------
 CTGGTATTTGCCAACAATATCCAGGTTCTTGGATACATTCCTTAGGAAGGGGGTCTCAATCCTAGAGAG

 GACCATAAACGGTTGTTATAGGTCCAAGAACCTATGTAAGGAATCCTTCCCCCAGAGTTAGGATCTCTC
--------------------------------------------------------------------------
 GGCATACCTAGCATGGGTACCCCTCAATCAGAGGATCAGAGGAAGGCTAACTGATGACCCCGTGGGTGA
         >>>>>>>>>>>>>>>>>>>>
 CCGTATGGATCGTACCCATGGGGAGTTAGTCTCCTAGTCTCCTTCCGATTGACTACTGGGGCACCCACT
--------------------------------------------------------------------------
 TCTCAGACCATCACACATGCCATGATTCATGGTGGTGGGGCCCCCCATCTTGGCAGGACCTCAGTGGGA
                                                                    <<<<
 AGAGTCTGGTAGTGTGTACGGTACTAAGTACCACCACCCCGGGGGGTAGAACCGTCCTGGAGTCACCCT
--------------------------------------------------------------------------
 GATGTGAAGTTGTCATTGGCAGTCACAGCAACAGCTTGCTTGCAGTNCCTCTTTCTCTNT
 <<<<<<<<<<<<<<<<
 CTACACTTCAACAGTAACCGTCAGTGTCGTTGTCGAACGAACGTCANGGAGAAAGAGATA
--------------------------------------------------------------------------
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: CCAACAAGTTGCATTTCCTT
REVERSE PRIMER: AGCATCCCGTTACTGAGC
LENGTH OF INSERT: 382
--------------------------------------------------------------------------
 GGCTTACCCCAGACTTGAAGCAAGCTTCTAGCTTCAGTCTTTAAATATTCAATCATGCACTGCCACTCC
                                                                     >
 CCGAATGGGGTCTGAACTTCGTTCGAAGATCGAAGTCAGAAATTTATAAGTTAGTACGTGACGGTGAGG
--------------------------------------------------------------------------
 CAACAAGTTGCATTTCCTTAAGATACCTTAGCTGTGCTTGAGCTGAAGCTAGGAACTATTGTATTACTT
 >>>>>>>>>>>>>>>>>>>
 GTTGTTCAACGTAAAGGAATTCTATGGAATCGACACGAACTCGACTTCGATCCTTGATAACATAATGAA
--------------------------------------------------------------------------
 TGTTCGTCCTACTTGATGTTGCCTCCTGGGACCTAGACACCTTACAGCAGAGCTCGCTCAGTAACGGGA
                                                        <<<<<<<<<<<<<<
 ACAAGCAGGATGAACTACAACGGAGGACCCTGGATCTGTGGAATGTCGTCTCGAGCGAGTCATTGCCCT
--------------------------------------------------------------------------
 TGCTGCACCTTTCCCCTTCAGGCTCAAAGATGTTCAGTCAATGGATGAACTGAAAGACGTCTACAACCA
 <<<<
 ACGACGTGGAAAGGGGAAGTCCGAGTTTCTACAAGTCAGTTACCTACTTGACTTTCTGCAGATGTTGGT
--------------------------------------------------------------------------
 TTTCTTACTTTAATACGGGGGCGGGGACATCCCCCAAGATGCAGAATGCTGCCCAAGGNTTAGCCGGAA

 AAAGAATGAAATTATGCCCCCGCCCCTGTAGGGGGTTCTACGTCTTACGACGGGTTCCNAATCGGCCTT
--------------------------------------------------------------------------
 GGAAGCTNGAAACGCATNAAGGGNGGATGGNCGNTGC

 CCTTCGANCTTTGCGTANTTCCCNCCTACCNGCNACG
--------------------------------------------------------------------------
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: CACACATGGAAGGCAACAGA
LENGTH OF INSERT: 333

AAAAACAACAACAAAAAAACAAGAGAGTCGCCTAGAACTAGAACAGGCGTCGTCCCCTTTTCCAGCACG

TTTTTGTTGTTGTTTTTTTTGTTCTCTCAGCGGATCTTGATCTTGTCCGCAGCAGGGGAAAAGGTCGTGC

CCCTAGGCACTGCAGTGACGGCATAGGCAAGGGTGTCCTGCTGCTCTTGCCTTTCTGCTCCCTTCTAGA
>>>>>>>>>>>>>>>>>>>>

GGGATCCGTGACGTCACTGCCGTATCCGTTCCCACAGGACGACGAGAACGGAAAGACGAGGGAAGATCT

ATCTTTCCAGAGCTCTCAGCACATGATAGCTCACCGATGCCACAATGGACATATGTTGTGCCTACTGAG

TAGAAAGGTCTCGAGAGTCGTGTACTATCGAGTGGCTACGGTGTTACCTGTATACAACACGGATGACTC

CAGTGGGTGAGATGGTGTTATCACCTGCTTCTGTTGCCTTCCATGTGTGACTGGAAGGCTGCTTTTTTT
<<<<<<<<<<<<<<<<<<<<

GTCACCCACTCTACCACAATAGTGGACGAAGACAACGGAAGGTACACACTGACCTTCCGACGAAAAAAA

GAGGCATACCTTTTTTTCTTGCAACCCCTTTAAGGTTAACCCTGNGGCCTGGGTCCAA

CTCCGTATGGAAAAAAGAACGTTGGGGAAATTCCAATTGGGACNCCGGACCCAGGTT

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

CCTATATTACTTNCACAGCAAGGNCCNTNANAGCNNANTCTANACNNACNATNNCNCGNTNANAAAAGC

GGATATAATGAANGTGTCGTTCCNGGNANTNTCGNNTNAGATNTGNNTGNTANNGNGCNANTNTTTTCG

ANCAGGACNATTNCNTGGGNATTAGCANTAGNATGTCNAANTCCNNGTCNNAGTTGCNACTTTNAGTCN

TNGTCCTGNTAANGNACCCNTAATCGTNATCNTACAGNTTNAGGNNCAGNNTCAACGNTGAAANTCAGN

NATNGCCGCTNGCGCNTTATGACTTCANAGNCA

NTANCGGCGANCGCGNAATACTGAAGTNTCNGT

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: CTAAGCAGATCGGCGTAAGG
REVERSE PRIMER: TGGGTGAGATGGTGTTATCA
LENGTH OF INSERT: 254
------------------------------------------------------------------------
 GTCAGCAGNGTTCAGGTCAGGTCAGGTCAAGACAACAAGTATGACTGATCACTAGAGTGTGCAATAGGC

 CAGTCGTCNCAAGTCCAGTCCAGTCCAGTTCTGTTGTTCATACTGACTAGTGATCTCACACGTTATCCG
------------------------------------------------------------------------
 CTAAGCAGATCGGCGTAAGGAGACTGGAGCCAGGCTCAGGGTGAGCTAAAGGGGTGCAGAAAAAGGAAT
 >>>>>>>>>>>>>>>>>>>>
 GATTCGTCTAGCCGCATTCCTCTGACCTCGGTCCGAGTCCCACTCGATTTCCCCACGTCTTTTTCCTTA
------------------------------------------------------------------------
 GTCTCAAAAAAGCAGCCTGCCAGTCACACATGGAAGGCAACAGAGCAGGTGATAACACCATCTCACCCA
                                                  <<<<<<<<<<<<<<<<<<<<
 CAGAGTTTTTTCGTCGGACGGTCAGTGTGTACCTTCCGTTGTCTCGTCCACTATTGTGGTAGAGTGGGT
------------------------------------------------------------------------
 CTGCTCAGTAGGGCACAACATATGTCCATTGTNGGCATNGGTGACTT

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

               AGGCCTAAGCAGATCGGCGTAAGGAGACTGGAGCCAGGCT-CAGGGTGAGC-T-AAAGGGGT-GC
                                                  ||| ||||||  |||||| | | | |||||||| ||
         K8F                                      TTGGACCCAGGCCNCAGGGTTAACCTTAAAGGGGTTGC

               A-GAAAAA-GGAATGTCTCAAAAAA-GCAGCCTGCCAGTCACACATGGAAGGCAACAGA-GCAGG
               | |||||| || ||| ||||||||| ||||||| |||||||||||||||||||||||| |||||
         K8F   AAGAAAAAAGGTATGCCTCAAAAAAAGCAGCCTTCCAGTCACACATGGAAGGCAACAGAAGCAGG

               TGATAACACCATCTCACCCACTGCTCAGTAGGGCACAACATATGTCCATTGTNGGCATNGGTGA-
               |||||||||||||||||||||||||||| |||||||||||||||||||| ||||| |||||
         K8F   TGATAACACCATCTCACCCACTGCTCAGTA-GGCACAACATATGTCCATTGT-GGCATCGGTGAG

               CTT
               ||
         K8F   CTATCATGTGCTGAGAGCTCTGGAAAGATTCTAGAAGGGAGCAGAAAGGCAAGAGCAGCAGGACA

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
```

```
SEQUENCE FILENAME: m10a.Seq
FORWARD PRIMER: AAGGGGACACCATTGTCATC
REVERSE PRIMER: GCCAAATTNGGGTGTAGAAC
LENGTH OF INSERT: 259
--------------------------------------------------------------------
 TCCTGGATGGTGCGGACGAAGGGGACACCATTGTCATCTTTCTGGTAGAGTGTCTCANGGATGAAGATG
                    >>>>>>>>>>>>>>>>>>>>
 AGGACCTACCACGCCTGCTTCCCCTGTGGTAACAGTAGAAAGACCATCTCACAGAGTNCCTACTTCTAC
--------------------------------------------------------------------
 ACCCCCCCAATGCACCTTTTTCACACGGACATCAACANTNNATAGGANCTGGCGGTACAGNCGNCGATT

 TGGGGGGGTTACGTGGAAAAAGTGTGCCTGTAGTTGTNANNTATCCTNGACCGCCATGTCNGCNGCTAA
--------------------------------------------------------------------
 NTCCNAAGTGTTCTACACCCNAATTTGGCTCAGCTTTTTGGCCATAGCTGNCTGCGGACTCATGTGCAG
         <<<<<<<<<<<<<<<<<<<
 NAGGNTTCACAAGATGTGGGNTTAAACCGAGTCGAAAAACCGGTATCGACNGACGCCTGAGTACACGTC
--------------------------------------------------------------------
 CCAGAATNCCCTTGGCGGGGGGGTCACGATCCGTNGAGCAATANTCNGNCAA

 GGTCTTANGGGAACCGCCCCCCCAGTGCTAGGCANCTCGTTATNAGNCNGTT
--------------------------------------------------------------------
DUPLICATE REPORT FOR m10a.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 4.1...',2.2e-47]
    ['M63656','Rat aldolase C gene, comple...',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: m11d.Seq
FORWARD PRIMER: GGCACTATGGGCCTATCGC
REVERSE PRIMER: ACACCACAGCAGCGCCAT
LENGTH OF INSERT: 275
--------------------------------------------------------------------
TTTCGTAGCTCCTGGAAGCAGCTGCCGGGACAGGCTGGTCCTCACCATGCCACCTGCTGGCAGGCA

CTATGGGCCTATCGCTCCTACCTGATTGTGCTATGTCTGCCCATTTTCCTGTTGCCTCTGCCACTC

ATTGTCCAAACTAAGGAAGCCTACTGTGCTTACTCCATCATCCTCATGGCGCTGCTGTGGTGTACA

GAGGCCCTGCCCTTGGCTGTCACCGNCCTCTTCCCCATCATCCTCTTCCCTTTTGATGGGTATCAT

GGAAGCCTNCAA .

DUPLICATE REPORT FOR m11d.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
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: AGGAAAAAAAAATAACCAT
LENGTH OF INSERT: 227
--------------------------------------------------------------------
 GTTGTTTTAGTGGGTGAGACTTAAACACATAAGAATGTTCAGGGAAAGGCATTCATGGTGAGCTAGAGA
                                                  >>>>>>>>>>>>>>>>>>>>
 CAACAAAATCACCCACTCTGAATTTGTGTATTCTTACAAGTCCCTTTCCGTAAGTACCACTCGATCTCT
--------------------------------------------------------------------
 GAGAACTGCTGTAAGGTAGATCGTGGATGCTCACCCTGGAAGGACAATTCAGAGCTCAGGAAATGATCT

 CTCTTGACGACATTCCATCTAGCACCTACGAGTGGGACCTTCCTGTTAAGTCTCGAGTCCTTTACTAGA
--------------------------------------------------------------------
 AAAGTTGTAAAGTACAGCGGAACAGTGGATGGTTATTTTTTTTTCCTTTGTCCCCAGGCTGAATCAGAG
                               <<<<<<<<<<<<<<<<<<<
 TTTCAACATTTCATGTCGCCTTGTCACCTACCAATAAAAAAAAAGGAAACAGGGGTCCGACTTAGTCTC
--------------------------------------------------------------------
 ATCATTCTGCCCAGCACAGG

 TAGTAAGACGGGTCGTGTCC
--------------------------------------------------------------------
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: ATCCAGACCGTCCAAGGC
REVERSE PRIMER: AAAAAAAGAAAATGATTCTT
LENGTH OF INSERT: 237
--------------------------------------------------------------------
 GGAAAAAGAGAAAAAGCTGAAACTCTGGACATATCTTAAACCGTTTGGCATTTCCTGGACTGAGTACAT
                                                                    >>
 CCTTTTTCTCTTTTTCGACTTTGAGACCTGTATAGAATTTGGCAAACCGTAAAGGACCTGACTCATGTA
--------------------------------------------------------------------
 CCAGACCGTCCAAGGCAAGAAATGTAAGCAGAGATCTTCATTTCTCTGGAGCTCAGAATTGAACCCAGA
 >>>>>>>>>>>>>>>>
 GGTCTGGCAGGTTCCGTTCTTTACATTCGTCTCTAGAAGTAAAGAGACCTCGAGTCTTAACTTGGGTCT
--------------------------------------------------------------------
 GCCTCTGTGTGCTAGGCAAATCCAGAGCAGTATCTTTAGCCCAAAGCAGGGAAGTTTTATAAGAATCAT
                                                          <<<<<<<<<
 CGGAGACACACGATCCGTTTAGGTCTCGTCATAGAAATCGGGTTTCGTCCCTTCAAAATATTCTTAGTA
--------------------------------------------------------------------
 TTTCTTTTTTTTTTTGGTTTTTCGAGTCGGG
 <<<<<<<<<<<
 AAAGAAAAAAAAAAACCAAAAAGCTCAGCCC
--------------------------------------------------------------------
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
--------------------------------------------------------------------
 CTTCGGCACTACCCACCTATCTTGTCTCCCCCTGTTTTTATTCAAGTAGTTTTTCTAAAGCCTACAGAC
   >>>>>>>>>>>>>>>>>>>>
 GAAGCCGTGATGGGTGGATAGAACAGAGGGGGACAAAAATAAGTTCATCAAAAAGATTTCGGATGTCTG
--------------------------------------------------------------------
 TGGTACTGTTTGTCTCTAGAGTAGCAGTGTTCAATAGAAGTATAATGAATGCACTCGGGAGGCAGAGGA
                                                                <<<<<<<
 ACCATGACAAACAGAGATCTCATCGTCACAAGTTATCTTCATATTACTTACGTGAGCCCTCCGTCTCCT
--------------------------------------------------------------------
 AGAGCTGAAAGGAGCCCGATCTATACCACCAGAGACTAGCTTAAAACAAAACAAAAAGACATGTATGAT
 <<<<<<<<<<<<<
 TCTCGACTTTCCTCGGGCTAGATATGGTGGTCTCTGATCGAATTTTGTTTTGTTTTTCTGTACATACTA
--------------------------------------------------------------------
 TTTAATCCTAGTACCCAGGAGGTAGAGGGCCAGTG

 AAATTAGGATCATGGGTCCTCCATCTCCCGGTCAC
--------------------------------------------------------------------
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: TCAGTGGAGGGCTGCCTG
REVERSE PRIMER: GTGNAANNNACCAGGATN
LENGTH OF INSERT: 211
--------------------------------------------------------------------
 CCTCAGTGGAGGGCTGCCTGGCCAAGNAAGNAAGTNCTGGGGAAGTNTAGCCAAGAATGCCGNTGNNTA
   >>>>>>>>>>>>>>>>>>
 GGAGTCACCTCCCGACGGACCGGTTCNTTCNTTCANGACCCCTTCANATCGGTTCTTACGGCNACNNAT
--------------------------------------------------------------------
 GCCCTGGAAGAAGGAGGAAGGCCCNTGGGGNCCCAGATGCACAAGCTNAGGNCTGTNATCCTGGTNNNT
                                                    <<<<<<<<<<<<<
 CGGGACCTTCTTCCTCCTTCCGGGNACCCCNGGGTCTACGTGTTCGANTCCNGACANTAGGACCANNNA
--------------------------------------------------------------------
 TNCACAGCTGTCTGGGGAGCCCTGGGCTCTGNTGCTCCTGAGGGCTAGCGGGTTTATTTCCACACTAGC
 <<<<<
 ANGTGTCGACAGACCCCTCGGGACCCGAGACNACGAGGACTCCCGATCGCCCAAATAAAGGTGTGATCG
--------------------------------------------------------------------
 ACCC

 TGGG
--------------------------------------------------------------------
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: GAACCAAGGAAGCATCCTNC
LENGTH OF INSERT: 140
--------------------------------------------------------------------
 AGGGTCTCCAAGGAGACAAGAAAGTAAGGTGATTACACCCACAAATAGCCAGGTTGCCCACTAGGNCCT
   >>>>>>>>>>>>>>>>>>>>
 TCCCAGAGGTTCCTCTGTTCTTTCATTCCACTAATGTGGGTGTTTATCGGTCCAACGGGTGATCCNGGA
--------------------------------------------------------------------
 NCCANTCCTNATCCCACTGGTCTCAAGGNACAGNAGGATGCTTCCTTGGTTCCAGCTGGGTACCAACTT
                                  <<<<<<<<<<<<<<<<<<<<
 NGGTNAGGANTAGGGTGACCAGAGTTCCNTGTCNTCCTACGAAGGAACCAAGGTCGACCCATGGTTGAA
--------------------------------------------------------------------
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: AAACGCACTGAGAGGTGAAT
REVERSE PRIMER: TAGACCAGGCTGTCCCCAGT
LENGTH OF INSERT: 237
--------------------------------------------------------------------
 GTCCTTGCAGCCTGTCAGCACAGCATACAAACGCACTGAGAGGTGAATTTAGGGGAGTACTTGAGTTGA
                                >>>>>>>>>>>>>>>>>>>>
 CAGGAACGTCGGACAGTCGTGTCGTATGTTTGCGTGACTCTCCACTTAAATCCCCTCATGAACTCAACT
--------------------------------------------------------------------
 ATTGTAGCCCTGTCCTCTCCACTCCTGAGACCACCTCCCTCCTTCCTCAGCTTCTTGACTGGCGTCACC

 TAACATCGGGACAGGAGAGGTGAGGACTCTGGTGGAGGGAGGAAGGAGTCGAAGAACTGACCGCAGTGG
--------------------------------------------------------------------
 CACTGGGGACAGCCTGGTCTAGGAGGCAGTGGCTTAGGATCTTCTACCTTGTGGCTAAGGTTCCCCTCT
  <<<<<<<<<<<<<<<<<<<
 GTGACCCCTGTCGGACCAGATCCTCCGTCACCGAATCCTAGAAGATGGAACACCGATTCCAAGGGGAGA
--------------------------------------------------------------------
 ANCCTTTCAGAGAGGTAGCGTGGAACAGGA

 TNGGAAAGTCTCTCCATCGCACCTTGTCCT
--------------------------------------------------------------------
DUPLICATE REPORT FOR m8a.Seq:  <none>
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
------------------------------------------------------------------------
 GGGACCATGGCGGCCGCAGAGCTGCGGCACGNACCTAGCTCCGGNTCATGGTGCCAGTCATAGTCGTGT

 CCCTGGTACCGCCGGCGTCTCGACGCCGTGCNTGGATCGAGGCCNAGTACCACGGTCAGTATCAGCACA
------------------------------------------------------------------------
 TTACTCGGGACTCTGTGCCCCTGGACGANCAAGAGAAGCTACCTTTCACCGTTGGCATATGCGAGAGAT

 AATGAGCCCTGAGACACGGGGACCTGCTNGTTCTCTTCGATGGAAAGTGGCAACCGTATACGCTCTCTA
------------------------------------------------------------------------
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: k10g.Seq
FORWARD PRIMER: ACTGCCATCTGGTGGAATTT
REVERSE PRIMER: CCAGGCCCAGGGGTTAGCTGC
LENGTH OF INSERT: 234
------------------------------------------------------------------------
 CTAGACTTTGCTCCTACATTCTTAGGCTGGGCAGTGGTGAGCAGAACTGACTGCCATCTGGTGGAATTT
                                                    >>>>>>>>>>>>>>>>>>>>
 GATCTGAAACGAGGATGTAAGAATCCGACCCGTCACCACTCGTCTTGACTGACGGTAGACCACCTTAAA
------------------------------------------------------------------------
 CAGTGACTGCAGGCTTTTTAGTCTTTATGTGTAATAAAGTGTTTGCTTGGATGTATGTCTGTGTATCAC

 GTCACTGACGTCCGAAAAATCAGAAATACACATTATTTCACAAACGAACCTACATACAGACACATAGTG
------------------------------------------------------------------------
 ATGTCCAAAGAGGCCAGAAGGAGACATCTAGGCAGCTAACCCCTGGGCCTGGAGTTACAGGTGCTGTGA
                                 <<<<<<<<<<<<<<<<<<<<
 TACAGGTTTCTCCGGTCTTCCTCTGTAGATCCGTCGATTGGGGACCCGGACCTCAATGTCCACGACACT
------------------------------------------------------------------------
 GCTGTGAGCTGGAAATTAACCTGGGAC

 CGACACTCGACCTTTAATTGGACCCTG
------------------------------------------------------------------------
DUPLICATE REPORT FOR k10g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: 19,26,28
END OF RECORD.
```

```
SEQUENCE FILENAME: k11g.Seq
FORWARD PRIMER: TGGGCACCTTGTTGTCTTTA
REVERSE PRIMER: CCCAGACTACGCGGGGGGA
LENGTH OF INSERT: 279
--------------------------------------------------------------------
 GGGAAGGAGAGCCTTGTCTCATCGTGGGCACCTTGTTGTCTTTAGCATCTTTTGATGACCGCCTGGAAG
                           >>>>>>>>>>>>>>>>>>>>
 CCCTTCCTCTCGGAACAGAGTAGCACCCGTGGAACAACAGAAATCGTAGAAAACTACTGGCGGACCTTC
--------------------------------------------------------------------
 ATGATGATTTTAACCTCATTGAGGAGAATTTGGGTGTTAAAGTCAAACGAGGAGTAAGTGCTTTCTTTA

 TACTACTAAAATTGGAGTAACTCCTCTTAAACCCACAATTTCAGTTTGCTCCTCATTCACGAAAGAAAT
--------------------------------------------------------------------
 CCNTTTTTCCCCCCGCGTAGTCTGGGGTCACCAGGGAGAAGAAGGATGACCAGTAGCATCATGCTTTGG
        <<<<<<<<<<<<<<<<<<
 GGNAAAAAGGGGGGGCGCATCAGACCCCAGTGGTCCCTCTTCTTCCTACTGGTCATCGTAGTACGAAACC
--------------------------------------------------------------------
 GGCCCAGATCTAAAGCGTATGAGAAAAAGATTTGCCAGTTTGCAGGGTGGGCCTTGAGGCTGGATAAAC

 CCGGGTCTAGATTTCGCATACTCTTTTTCTAAACGGTCAAACGTCCCACCCGGAACTCCGACCTATTTG
--------------------------------------------------------------------
DUPLICATE REPORT FOR k11g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:  <none>
PHYSICAL MAP LOCATION: 18,19,26
END OF RECORD.
--------------------------------------------------------------------
SEQUENCE FILENAME: k12g.Seq
FORWARD PRIMER: CTTGTCAGACAGGTGCTGAG
REVERSE PRIMER: GCTCTTTAACAGGCTTGTTC
LENGTH OF INSERT: 332
--------------------------------------------------------------------
 CCAGAAGCTGTGCTTGAAGGTGCCCGATACATGGTTGCACTACAGATTGCCAGGGTGCCCCTTGTCAGA
                                                        >>>>>>>>>
 GGTCTTCGACACGAACTTCCACGGGCTATGTACCAACGTGATGTCTAACGGTCCCACGGGGAACAGTCT
--------------------------------------------------------------------
 CAGGTGCTGAGACAAACCTTTCAGGAGAGAGCCAAGCTAAACATAACCCCAACCAAGAAAGGTAGAAAG
 >>>>>>>>>>>
 GTCCACGACTCTGTTTGGAAAGTCCTCTCTCGGTTCGATTTGTATTGGGGTTGGTTCTTTCCATCTTTC
--------------------------------------------------------------------
 GATGTGGATGAGGCCCACTACGCTTACTCTTTTAAGTACTTAAAGAACAAGCCTGTTAAAGAGCTGAGA
                                          <<<<<<<<<<<<<<<<<<<
 CTACACCTACTCCGGGTGATGCGAATGAGAAAATTCATGAATTTCTTGTTCGGACAATTTCTCGACTCT
--------------------------------------------------------------------
 GACGACCAGTTCCTCAAGATAGGCCTGGCTGAGGATGAAGGGCTGCTCACCATCGATCTNCTCAAAATA

 CTGCTGGTCAAGGAGTTCTATCCGGACCGACTCCTACTTCCCGACGAGTGGTAGCTAGANGAGTTTTAT
--------------------------------------------------------------------
 TGTCTGGGTCATTACCATACCCTTTCCACTNCCTTNATATCTATGCTTGATGTTAA

 ACAGACCCAGTAATGGTATGGGAAAGGTGANGGAANTATAGATACGAACTACAATT
--------------------------------------------------------------------
DUPLICATE REPORT FOR k12g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
    ['gp','Z35719','T04A8.14, emb-5 protein ...',7.8e-20]
    ['sp','P34703','EMB-5 PROTEIN.>pir S35241 S35241 ...',8.1e-20]
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
```

```
SEQUENCE FILENAME: k1g.Seq
FORWARD PRIMER: CCCTTTCACACACTGTGGC
REVERSE PRIMER: TCAATACACGCCACAATGTC
LENGTH OF INSERT: 250
--------------------------------------------------------------------
 CCCCTTTCACACACTGTGGCTGTCTTCCCCCGTATCCTCCAGTAACTGTTGCTGTCATCCACAGAGGTC
  >>>>>>>>>>>>>>>>>>>
 GGGGAAAGTGTGTGACACCGACAGAAGGGGGCATAGGAGGTCATTGACAACGACAGTAGGTGTCTCCAG
--------------------------------------------------------------------
 GCACCTGTCACTGACTGCTGCCCACTACCTGACCCATAGCGCACGTCGTGAGAATGCAGTCGGACATTG
                                                              <<<<<<<
 CGTGGACAGTGACTGACGACGGGTGATGGACTGGGTATCGCGTGCAGCACTCTTACGTCAGCCTGTAAC
--------------------------------------------------------------------
 TGGCGTGTATTGAGTAGCTTCACCACGGAGCCGCAAGTANCGACAGCCATGTTGGACACTCCCACAGCG
 <<<<<<<<<<<<<
 ACCGCACATAACTCATCGAAGTGGTGCCTCGGCGTTCATNGCTGTCGGTACAACCTGTGAGGGTGTCGC
--------------------------------------------------------------------
 CTCCACAAACCACCCAACANCAGAAGCGACAAGACAGCCATCC

 GAGGTGTTTGGTGGGTTGTNGTCTTCGCTGTTCTGTCGGTAGG
--------------------------------------------------------------------
DUPLICATE REPORT FOR k1g.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: TCTGGAGAGGTGGTTCTTTG
LENGTH OF INSERT: 321
--------------------------------------------------------------------
 GGCCAACAGCTCATGCACTATCCGCTTCACATCCTCAGTCAGCATCTGGGCATCCCTAGAAGGATGAGC
                                     >>>>>>>>>>>>>>>>>>>
 CCGGTTGTCGAGTACGTGATAGGCGAAGTGTAGGAGTCAGTCGTAGACCCGTAGGGATCTTCCTACTCG
--------------------------------------------------------------------
 AGTACTGGTTAAAGGATGCCCTTCCTCAGGGATGGAGAGTGTTGCTGATGCGTGAGTNTGGAGAAAAGA

 TCATGACCAATTTCCTACGGGAAGGAGTCCCTACCTCTCACAACGACTACGCACTCANACCTCTTTTCT
--------------------------------------------------------------------
 GCTTTAAAGACAAAGAACCACCTCTCCAGAGAAACCCAAGGGTGTCAACATCACCCCCTGACACAAAGC
        <<<<<<<<<<<<<<<<<<<
 CGAAATTTCTGTTTCTTGGTGGAGAGGTCTCTTTGGGTTCCCACAGTTGTAGTGGGGGACTGTGTTTCG
--------------------------------------------------------------------
 CTCTGCTCCAGACTTGTACCNGATTTAGCAATAAGAGACGAGGGACTGGGTCACAGCCTTGCCCCAGGC

 GAGACGAGGTCTGAACATGGNCTAAATCGTTATTCTCTGCTCCCTGACCCAGTGTCGGAACGGGGTCCG
--------------------------------------------------------------------
 TTTCCCAGAGCCCCGCGCTCCTCCTTACCTGTTCTNTCCTGCGAT

 AAAGGGTCTCGGGGCGCGAGGAGGAATGGACAAGANAGGACGCTA
--------------------------------------------------------------------
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: CCTCAGTCAGCATCTGGGC
REVERSE PRIMER: GCTCAGAGTTGTTCAGCTACAA
LENGTH OF INSERT: 251
--------------------------------------------------------------------
 CAAGGCCAACAGCTCATGCACTATCCGCTTCACATCCTCAGTCAGCATCTGGGCATCCCTGTTNTCTCC
                                      >>>>>>>>>>>>>>>>>>>
 GTTCCGGTTGTCGAGTACGTGATAGGCGAAGTGTAGGAGTCAGTCGTAGACCCGTAGGGACAANAGAGG
--------------------------------------------------------------------
 TGCGATAGTGACCACATGAGGCTTCTTATTCACAAGAAACTTCTTTAGGGTTTCAATATCTTGAGCCTG

 ACGCTATCACTGGTGTACTCCGAAGAATAAGTGTTCTTTGAAGAAATCCCAAAGTTATAGAACTCGGAC
--------------------------------------------------------------------
 AATTAAAAGAAAATTGTAGCTGAACAACTCTGAGCATGGAAAGTTTTCTGGGTACGGTGGTATGTCTTT
            <<<<<<<<<<<<<<<<<<<<<
 TTAATTTTCTTTTAACATCGACTTGTTGAGACTCGTACCTTTCAAAAGACCCATGCCACCATACAGAAA
--------------------------------------------------------------------
 AATTCCAGCATTTNGGAGGGCAGATCCTACCGGCGTGNGNCACC

 TTAAGGTCGTAAANCCTCCCGTCTAGGATGGCCGCACNCNGTGG
--------------------------------------------------------------------
DUPLICATE REPORT FOR k3g.Seq: k2g (11)
     vs k2g  (score = 64)
             CAAGGCCAACAGCTCATGCACTATCCGCTTCACATCCTCAGTCAGCATCTGGGCATCCCT-GTTN
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
       k2g         GGCCAACAGCTCATGCACTATCCGCTTCACATCCTCAGTCAGCATCTGGGCATCCCTAGAAG

             TCTCCTGCGA-TAGTGACCACATGAGGCTTCTTATTCACAAGAAACTTCTTTAGGGT-TT-CAAT
               |   || | || ||   | | || ||  | | | | | || |    |  || || || |  |
       k2g     GAT-GAGC-AGTACTGGTTAAAGGATGC--CCT-TCCTC-AGGGA----TGGAGAGTGTTGC--T

             -AT-CTTGAGCCTGAATTAAAAGAAAATTGTA--G-CT--GAACAAC-TCTGAGCATG-GAAAGT
              || | ||||  || |  ||||| |  || ||  | |   |||| || |||   || | ||||
       k2g     GATGCGTGAGTNTGGA-GAAAAG-AGCTT-TAAAGACAAAGAACCACCTCT--CCA-GAGAAACC

             TTTCTGGGTACGGTGGTATGTCTTTAATTCCAGCATTTNGGAGGGCAGATCCTACCGGCGTGNGN
       k2g     CAAGGGTGTCAACATCACCCCCTGACACAAAGCCTCTGCTCCAGACTTGTACCNGATTTAGCAAT

             CACC
       k2g     AAGAGACGAGGGACTGGGTCACAGCCTTGCCCCAGGCTTTCCCAGAGCCCCGCGCTCCTCCTTAC

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:  <none>
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
```

```
SEQUENCE FILENAME: k4g.Seq
FORWARD PRIMER: TGCTACAATGTGTGGAAGTCGT
REVERSE PRIMER: CCATGATCCTCCTGGCTCAGCT
LENGTH OF INSERT: 293
--------------------------------------------------------------------
 ACCGTTAACTGCCCATAAACCTGTAGGAAGGGAGGAGCCTGGTCCCCCTCCCCAGCATTGAAGTATGCT
                                                                >>>>
 TGGCAATTGACGGGTATTTGGACATCCTTCCCTCCTCGGACCAGGGGGAGGGGTCGTAACTTCATACGA
--------------------------------------------------------------------
 ACAATGTGTGGAAGTCGTTGAATTTAATTAATTTTGGGGAGAAAAAGTTTAGTTCTATTGTCGAACACT
 >>>>>>>>>>>>>>>>>>
 TGTTACACACCTTCAGCAACTTAAATTAATTAAAACCCCTCTTTTTCAAATCAAGATAACAGCTTGTGA
--------------------------------------------------------------------
 CACAGTTTGTGAGTCTTCAGTGTTTTAGTTACAAAACAATTATAGCAACAATGGAGAAGCTGAGCCAGG
                                                 <<<<<<<<<<<<
 GTGTCAAACACTCAGAAGTCACAAAATCAATGTTTTGTTAATATCGTTGTTACCTCTTCGACTCGGTCC
--------------------------------------------------------------------
 AGGATCATGGAATTTACGTTTTGGTTGCATTTGTAGACCCTGTCTCAAAAACAGCTCTCCCCTTCAACC
 <<<<<<<<<<
 TCCTAGTACCTTAAATGCAAAACCAACGTAAACATCTGGGACAGAGTTTTTGTCGAGAGGGGAAGTTGG
--------------------------------------------------------------------
 ACGAGTTTAGGGCCCCC

 TGCTCAAATCCCGGGGG
--------------------------------------------------------------------
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
--------------------------------------------------------------------
 TCAGGGTCACCAAGCCCAGTGATACACAAATCTAAGCATAAAAAGAGCCAATGAATTGGATCAATGGTT
                                          >>>>>>>>>>>>>>>>>>>>>
 AGTCCCAGTGGTTCGGGTCACTATGTGTTTAGATTCGTATTTTTCTCGGTTACTTAACCTAGTTACCAA
--------------------------------------------------------------------
 AAAGGTATTTTGGAGTTATACTTGAGTTCAATTCTGAAGATTCACAGTAGGAGACAGTCAACACGCCCC

 TTTCCATAAAACCTCAATATGAACTCAAGTTAAGACTTCTAAGTGTCATCCTCTGTCAGTTGTGCGGGG
--------------------------------------------------------------------
 AAGTTGTCTTTTGGCAAGTTTTACTCTAGTTAGGACTGTATAGTAAAGCCTTGTCTCAAAAAGCCACAA
                              <<<<<<<<<<<<<<<<<<<
 TTCAACAGAAAACCGTTCAAAATGAGATCAATCCTGACATATCATTTCGGAACAGAGTTTTTCGGTGTT
--------------------------------------------------------------------
 TAAAATATTTTTAGATAATACAAATTTGTA

 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: CTTTGAGGTGAAACTTCATGAT
REVERSE PRIMER: GCTGCAGTTGCTAGATGCCAA
LENGTH OF INSERT: 165
--------------------------------------------------------------------------
 TGACATCTTTGAGGTGAAACTTCATGATAACTACGAACTGATGAAGGNTGAAAGAGGCGTAGAGCTCCT
       >>>>>>>>>>>>>>>>>>>>>
 ACTGTAGAAACTCCACTTTGAAGTACTATTGATGCTTGACTACTTCCNACTTTCTCCGCATCTCGAGGA
--------------------------------------------------------------------------
 CAATTTTCCGAGCTGGCAATAGCTCTCCATGCTGCTTCCGCAGGCAGCCACTTTGGCATCTAGCAACTG
                                                  <<<<<<<<<<<<<<<<<
 GTTAAAAGGCTCGACCGTTATCGAGAGGTACGACGAAGGCGTCCGTCGGTGAAACCGTAGATCGTTGAC
--------------------------------------------------------------------------
 CAGCCTTTTGGCACTCTCCTTACTTTC
 <<<<
 GTCGGAAAACCGTGAGAGGAATGAAAG
--------------------------------------------------------------------------
DUPLICATE REPORT FOR k6g.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely detrmined
END OF RECORD.
```

```
SEQUENCE FILENAME: m11c.Seq
FORWARD PRIMER: GTTTCCTAGTTGGGTTGACTGT
REVERSE PRIMER: TGAGAGAGAGAGAGAGANAGAG
LENGTH OF INSERT: 328
--------------------------------------------------------------------
 AGCAAGGTGGCTTCCCCTTCAAGGTTAGCAGGCATTCTCATATTCACGAACATGACACCTTTGGATAGC

 TCGTTCCACCGAAGGGGAAGTTCCAATCGTCCGTAAGAGTATAAGTGCTTGTACTGTGGAAACCTATCG
--------------------------------------------------------------------
 CAGAGAGGGAAAAGCCCCCAGCTCCACCTTTCCAGGGGAATTTTGAAGCCCCGAGCAGTGGGGACAGAC

 GTCTCTCCCTTTTCGGGGGTCGAGGTGGAAAGGTCCCCTTAAAACTTCGGGGCTCGTCACCCCTGTCTG
--------------------------------------------------------------------
 TGACAAAGAATCTGGAGATAGAACTGTTTCCTAGTTGGGTTGACTGTGTAATCTTAGTACTCAGGAGGT
                        >>>>>>>>>>>>>>>>>>>>>
 ACTGTTTCTTAGACCTCTATCTTGACAAAGGATCAACCCAACTGACACATTAGAATCATGAGTCCTCCA
--------------------------------------------------------------------
 GGAGAGAGGAGAATCACAAATGTAAGGTCAGCCTGGGCTNGCATTTGTGAGACCCCACCTCTCTCTCTC
                                                          **********
 CCTCTCTCCTCTTAGTGTTTACATTCCAGTCGGACCCGANCGTAAACACTCTGGGGTGGAGAGAGAGAG
--------------------------------------------------------------------
 TCTCTCTCTCTNTCTCTCTCTCTCTCACACACACACACAGGGTCTTACGG
 *******<<<<<<<<<<<<<<<<<<<<<<
 AGAGAGAGAGANAGAGAGAGAGAGAGTGTGTGTGTGTGTCCCAGAATGCC
--------------------------------------------------------------------
DUPLICATE REPORT FOR m11c.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
--------------------------------------------------------------------
SEQUENCE FILENAME: m12c.Seq
FORWARD PRIMER: CAGCACCCAGTCCAATACTC
REVERSE PRIMER: GCTGATCTTTCAGATTGTGGAA
LENGTH OF INSERT: 177
--------------------------------------------------------------------
 ATAACCAGCCCCTGGGCTTGTCCAGCACCCAGTCCAATACTCTTTNNCTTGACTGTACTATTCGAGGAC
                      >>>>>>>>>>>>>>>>>>>
 TATTGGTCGGGGACCCGAACAGGTCGTGGGTCAGGTTATGAGAAANNGAACTGACATGATAAGCTCCTG
--------------------------------------------------------------------
 TTCAGGTAGAAATATCAGACATCTGTGCCCAGTGTCTGTCTCGAGTCCTCTCCTTGGTGATTCCACAAT
                                                        <<<<<<<<<
 AAGTCCATCTTTATAGTCTGTAGACACGGGTCACAGACAGAGCTCAGGAGAGGAACCACTAAGGTGTTA
--------------------------------------------------------------------
 CTGAAAGATCAGCTGTCTCTAGGAAGTCGTCACTNGGGG
 <<<<<<<<<<<<<
 GACTTTCTAGTCGACAGAGATCCTTCAGCAGTGANCCCC
--------------------------------------------------------------------
DUPLICATE REPORT FOR m12c.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
--------------------------------------------------------------------
 GGGCTAGATGGCCTGGCCAAAAAGTTTGGCCTCACACCTGAGCAGTTTGGGGAGAACCTCCGTGACAGC
                        >>>>>>>>>>>>>>>>>>>>
 CCCGATCTACCGGACCGGTTTTTCAAACCGGAGTGTGGACTCGTCAAACCCCTCTTGGAGGCACTGTCG
--------------------------------------------------------------------
 TACCAGCGCCACGAGACAGAGCAGTTCCCTACAGAGCCTTTGGAGCTGGCCAAAGATTACGTGTGCAGC

 ATGGTCGCGGTGCTCTGTCTCGTCAAGGGATGTCTCGGAAACCTCGACCGGTTTCTAATGCACACGTCG
--------------------------------------------------------------------
 CAGTTCCCTACACCAGAAGCTGTGCTTGAAGGTGCCCGATACATGGTTGCACTACAGATTGCCAGGGAG
       <<<<<<<<<<<<<<<<<<<
 GTCAAGGGATGTGGTCTTCGACACGAACTTCCACGGGCTATGTACCAACGTGATGTCTAACGGTCCCTC
--------------------------------------------------------------------
 CCCCTTGTCAGACAGGTGCTGAGACAAACCTTTCAGGAGAGAGCCAAGCTAAACATAACCCCAACCAAG

 GGGGAACAGTCTGTCCACGACTCTGTTTGGAAAGTCCTCTCTCGGTTCGATTTGTATTGGGGTTGGTTC
--------------------------------------------------------------------
 AAAGGTAGGAAAGGATGTTGGATNGAGGGCCCANTTACGGTTACTCTTTTTAAGGTACTTTAAAGGACC

 TTTCCATCCTTTCCTACAACCTANCTCCCGGGTNAATGCCAATGAGAAAAATTCCATGAAATTTCCTGG
--------------------------------------------------------------------
 AAGNCTNTTTAAAGNGGCTNAGAGACGGCCCAGTTTCCTCAAGGNTA

 TTCNGANAAATTTCNCCGANTCTCTGCCGGGTCAAAGGAGTTCCNAT
--------------------------------------------------------------------
DUPLICATE REPORT FOR k8g.Seq: k12g   (23)
     vs k12g   (score = 194)
              GGGCTAGATGGCCTGGCCAAAAAGTTTGGCCTCACACCTGAGCAGTTTGGGGAGAACCTCCGTGA
       k12g

              CAGCTACCAGCGCCACGAGACAGAGCAGTTCCCTACAGAGCCTTTGGAGCTGGCCAAAGATTACG
       k12g

              TGTGCAGCCAGTTCCCTACACCAGAAGCTGTGCTTGAAGGTGCCCGATACATGGTTGCACTACAG
                                  ||||||||||||||||||||||||||||||||||||||||||||
       k12g                    CCAGAAGCTGTGCTTGAAGGTGCCCGATACATGGTTGCACTACAG

              ATTGCCAGGGAGCCCCTTGTCAGACAGGTGCTGAGACAAACCTTTCAGGAGAGAGCCAAGCTAAA
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
       k12g   ATTGCCAGGGTGCCCCTTGTCAGACAGGTGCTGAGACAAACCTTTCAGGAGAGAGCCAAGCTAAA

              CATAACCCCAACCAAGAAAGGTAGGAAAGGATGTTGGATNGAGGGCCCANTTACGGTTACTCTTT
              |||||||||||||||||||||||| |||||||| ||||| || ||||||  |||| |||||| ||
       k12g   CATAACCCCAACCAAGAAAGGTA-GAAAGGATG-TGGAT-GA-GGCCCA-CTACGCTTACTC-TT

              TTAAGGTACTTTAAAGGACCAAGNCTNTTTAAAGNGGCTNAGAGACGGCCCAGTTTCCTCAAGGN
              |||| |||| ||||| || |||| ||  |||||| ||| |||||| | |||| |||||||||
       k12g   TTAA-GTAC-TTAAA-GAACAAGCCT-GTTAAAG-AGCTGAGAGAC-GACCAG-TTCCTCAAGAT

NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS:
    ['gp','Z35719','T04A8.14, emb-5 protein ...',2.5e-19]
    ['sp','P34703','EMB-5 PROTEIN.>pir S35241 S35241 ...',2.6e-19]
PHYSICAL MAP LOCATION: 18,19
END OF RECORD.
```

```
SEQUENCE FILENAME: k5d.Seq
FORWARD PRIMER: GGGTTCCTTCCTGTGTGGG
REVERSE PRIMER: CACATAGCCTTTGAAATCTC
LENGTH OF INSERT: 368
--------------------------------------------------------------------------
 GGGAGAGCATCTGAATGAGTAGTGGCTTTTGTCTACTGTTGTAAAGCAGTTTCTGCTGGTTGCCAGCCC

 CCCTCTCGTAGACTTACTCATCACCGAAAACAGATGACAACATTTCGTCAAAGACGACCAACGGTCGGG
--------------------------------------------------------------------------
 TTTGAAAGAGGGTTCCTTCCTGTGTGGGGAGCTGTGCTCTCAGCACATGGCAGGAAGTGAGACCCAACA
          >>>>>>>>>>>>>>>>>>>
 AAACTTTCTCCCAAGGAAGGACACACCCCTCGACACGAGAGTCGTGTACCGTCCTTCACTCTGGGTTGT
--------------------------------------------------------------------------
 TCACAAACAAGAAAGCTCTGCCTGTTCTCTGTGTTGGGCAGTGTTGCTCAGGAGATTTCAAAGGCTATG
                                                  <<<<<<<<<<<<<<<<<<<
 AGTGTTTGTTCTTTCGAGACGGACAAGAGACACAACCCGTCACAACGAGTCCTCTAAAGTTTCCGATAC
--------------------------------------------------------------------------
 TGGCACAGCTCTGTCCATAGATAGACTACCTCTGAGCTTCACAGCGGTGCAGGGCCACACAGANGGAAC
 <<
 ACCGTGTCGAGACAGGTATCTATCTGATGGAGACTCGAAGTGTCGCCACGTCCCGGTGTGTCTNCCTTG
--------------------------------------------------------------------------
 CTNCCTCCGAGTNAGAAGTTGGCAGTAAATCTNCTAACTNCCTTNTGTNCNGTCGTCTCCTCCTGNGGG

 GANGGAGGCTCANTCTTCAACCGTCATTTAGANGATTGANGGAANACANGNCAGCAGAGGAGGACNCCC
--------------------------------------------------------------------------
 CTGCCATGTGCCAGGCTCAGTTT

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

              GCCCTTTGAAAGAGGGTTCCTTCCTGTGTGGGGAGCTGTGCTCTCAGCACATGGCAGGAAGTGAG
                              ||||||||||||||||||||||||||||||||||||||||||||||||||
      k10d                   GTTCCTTCCTGTGTGGGGAGCTGTGCTCTCAGCACATGGCAGGAAGTGAG

              ACCCAACATCACAAACAAGAAAGCTCTGCCTGTTCTCTGTGTTGGGCAGTGTTGCTCAGGAGATT
              |||||||||||||||||||||||||||||| |||||||||||| |||||||||| ||||||||||
       k10d   ACCCAACATCACAAACAAGAAAGCTCTGCCCGTTCTCTGTGTTNGGCAGTGTTNCTCAGGAGATT

              TCAAAGGCTATGTGGCACAGCTCTGTCCATAGATAGACTACCTCTGAGCTTCACAGCGGTGCAGG
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
       k10d   TCAAAGGCTATGTGGCACAGCTCTGTCCATAGTTAGACTACCTCTGAGCTTCACAGCGGTGCAG

              GCCACACAGANGGAACCTNCCTCCGAGTNAGAAGTTGGCAGTAAATCTNCTAACTNCCTTNTGTN
       k10d

              CNGTCGTCTCCTCCTGNGGGCTGCCATGTGCCAGGCTCAGTTT
       k10d

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: TCCCTNCGGTGAGCATGCG
LENGTH OF INSERT: 351
--------------------------------------------------------------------
 CTGAGCTAAGCGAGGTATGTGCTGCGTGTTACTGCCGTAAGCTTCCTTGACTAAAGTTTGCTGAGGAGA

 GACTCGATTCGCTCCATACACGACGCACAATGACGGCATTCGAAGGAACTGATTTCAAACGACTCCTCT
--------------------------------------------------------------------
 GATAAATGGTNGTGTTCAGGTCATGTTCTCACCAAACAAGTGCACAAAGGCACAGGGGGACTCTTACCA
                                                  >>>>>>>>>>>>>>>>
 CTATTTACCANCACAAGTCCAGTACAAGAGTGGTTTGTTCACGTGTTTCCGTGTCCCCCTGAGAATGGT
--------------------------------------------------------------------
 TGAAGTCCTTTCCTGGGTGTCCCCATCTATTCTGCTGCAGTCACTGTAACCTTTAACTCTTGTCCCACT
 >>>>
 ACTTCAGGAAAGGACCCACAGGGGTAGATAAGACGACGTCAGTGACATTGGAAATTGAGAACAGGGTGA
--------------------------------------------------------------------
 TTCCTCCTGTCACTACAGCAAGTTAGGCCATGAGAATCATGCTCACCGCTCAACGCATGCTCACCGNAG
                                                  <<<<<<<<<<<<<<<<
 AAGGAGGACAGTGATGTCGTTCAATCCGGTACTCTTAGTACGAGTGGCGAGTTGCGTACGAGTGGCNTC
--------------------------------------------------------------------
 GGAGGCGCAGTAGGCCTTACTTTACACTGGGCTTGAAAGTACATGTCAGCTTTACGNTTTTTAATATTA
 <<<
 CCTCCGCGTCATCCGGAATGAAATGTGACCCGAACTTTCATGTACAGTCGAAATGCNAAAAATTATAAT
--------------------------------------------------------------------
 TTCCCC

 AAGGGG
--------------------------------------------------------------------
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: m10c.Seq
FORWARD PRIMER: GTGGCAGACTGCCTTAGGG
REVERSE PRIMER: GTAGTGGGTACAAGTGACAG
LENGTH OF INSERT: 259
--------------------------------------------------------------------
 GCTGGGCAGAGCAAGGGCACTTAGCGTCCGATCACCTTTATGCTGTGGTGACTTGGTCCCTGNTGTGGC
                                                        >>>>>
 CGACCCGTCTCGTTCCCGTGAATCGCAGGCTAGTGGAAATACGACACCACTGAACCAGGGACNACACCG
--------------------------------------------------------------------
 AGACTGCCTTAGGGACACACGCGCCCTGAATCAGACCAGAAACCTCAGAAAAGTTTAGGCCTTTGCAGC
 >>>>>>>>>>>>>>
 TCTGACGGAATCCCTGTGTGCGCGGGACTTAGTCTGGTCTTTGGAGTCTTTTCAAATCCGGAAACGTCG
--------------------------------------------------------------------
 GAGGCCCTTTCTAAAAGACAAGGATGCCCCCTCATGTGATATGTCTCTTTGTCCCTCTCTGTCACTTGT
                                                  <<<<<<<<<<<
 CTCCGGGAAAGATTTTCTGTTCCTACGGGGGAGTACACTATACAGAGAAACAGGGAGAGACAGTGAACA
--------------------------------------------------------------------
 ACCCACTACAGGCAACCTGACACATCANCTNCNNCGANCCCATNCACAACTG
 <<<<<<<<<
 TGGGTGATGTCCGTTGGACTGTGTAGTNGANGNNGCTNGGGTANGTGTTGAC
--------------------------------------------------------------------
DUPLICATE REPORT FOR m10c.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS: <none>
PROTEIN BLAST HITS: <none>
PHYSICAL MAP LOCATION: not uniquely determined
END OF RECORD.
```

```
SEQUENCE FILENAME: m1c.Seq
FORWARD PRIMER: TCTGTCTGACTGCCTGGCTC
REVERSE PRIMER: CCACTGCTCAGGGACTCTGC
LENGTH OF INSERT: 372
--------------------------------------------------------------------
 GCCAGNCTTTNTGGGAGGCCACTGAAGCAGAACCCGAGAAAGGAAATNACGTTCCTTNAGCCATTGGCC

 CGGTCNGAAANACCCTCCGGTGACTTCGTCTTGGGCTCTTTCCTTTANTGCAAGGAANTCGGTAACCGG
--------------------------------------------------------------------
 ATTCCCATGCATGTCCTTGATGCTTGGCTGCCTCTTCTCATTAATCTCATCATCTGTCTGACTGCCTGG
                                                    >>>>>>>>>>>>>>>>>
 TAAGGGTACGTACAGGAACTACGAACCGACGGAGAAGAGTAATTAGAGTAGTAGACAGACTGACGGACC
--------------------------------------------------------------------
 CTCTGAGCAGTGTGTGTGTTCTGTCCTGGAACTGTCTCTGTACCCCAGGCTGGCCCCTGCCTCCCACAG
 >>>
 GAGACTCGTCACACACACAAGACAGGACCTTGACAGAGACATGGGGTCCGACCGGGGACGGAGGGTGTC
--------------------------------------------------------------------
 GGNGCCTTCACTGCTGCTGAGGTTAGTGCAGGGACCACAGTGCAGAGTCCCTGAGCAGTGGCACTCAAG
                                          <<<<<<<<<<<<<<<<<<<
 CCNCGGAAGTGACGACGACTCCAATCACGTCCCTGGTGTCACGTCTCAGGGACTCGTCACCGTGAGTTC
--------------------------------------------------------------------
 ACCTCAGGAGGCGGGTCAACTTNATCCTCAGGGGTCCCTTGACAAGTNTGTTCAAAAGGGACAGAGTTA

 TGGAGTCCTCCGCCCAGTTGAANTAGGAGTCCCCAGGGAACTGTTCANACAAGTTTTCCCTGTCTCAAT
--------------------------------------------------------------------
 CCAAAACAAAGGGGTCTTAANGGATTT

 GGTTTTGTTTCCCCAGAATTNCCTAAA
--------------------------------------------------------------------
DUPLICATE REPORT FOR m1c.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: GTAAGNCATGTGAGATGTGA
LENGTH OF INSERT: 263
--------------------------------------------------------------------
 GGGCAGAGATGAACGGGCTTGCAGCCCAGGGCAGATATGAAGGCAGTGGAGATGGCGGANNNCAGCACA

 CCCGTCTCTACTTGCCCGAACGTCGGGTCCCGTCTATACTTCCGTCACCTCTACCGCCTNNNGTCGTGT
--------------------------------------------------------------------
 GTCCCTCTACATCGCCAACCATGCCTACTGAGGCACGCTGACTCTACCCACAGCCCTTGGCCTAGCAAT
                  >>>>>>>>>>>>>>>>>>>>
 CAGGGAGATGTAGCGGTTGGTACGGATGACTCCGTGCGACTGAGATGGGTGTCGGGAACCGGATCGTTA
--------------------------------------------------------------------
 GCACATCCCATTTTGCCTATAGTCACAGCCAGGGCCAAATAGCTATGCAGAACAGAGATATCTTTATCC

 CGTGTAGGGTAAAACGGATATCAGTGTCGGTCCCGGTTTATCGATACGTCTTGTCTCTATAGAAATAGG
--------------------------------------------------------------------
 AGCGCCAAGTCATCTCACATCTCACATGNCTTACATTGCTGTACCCGAGACCACAG
               <<<<<<<<<<<<<<<<<<<
 TCGCGGTTCAGTAGAGTGTAGAGTGTACNGAATGTAACGACATGGGCTCTGGTGTC
--------------------------------------------------------------------
DUPLICATE REPORT FOR m2c.Seq: <no duplicates>
NUCLEOTIDE BLAST HITS:
    ['X06984','Rat brain mRNA for aldolase C (EC 4.1...',1.4e-52]
    ['M63656','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: CTTGGTTTGGTGTTTCCCC
REVERSE PRIMER: CTGAGATAATAGAGTATGGC
LENGTH OF INSERT: 226
--------------------------------------------------------------------
 CCCAGAAGGGCCAAGTCACAGATCTCTGTCCATCCTTGGTTTGGTGTTTCCCCNGATATGGAACCCATC
                                      >>>>>>>>>>>>>>>>>>>
 GGGTCTTCCCGGTTCAGTGTCTAGAGACAGGTAGGAACCAAACCACAAAGGGGNCTATACCTTGGGTAG
--------------------------------------------------------------------
 AGCCTCAGGGGACCCAGGTTTAGAAGAAATATGCTTGTATCAGATTGGAATCCCAAGAGCAGGTTGCCT

 TCGGAGTCCCCTGGGTCCAAATCTTCTTTATACGAACATAGTCTAACCTTAGGGTTCTCGTCCAACGGA
--------------------------------------------------------------------
 TCCCCGTCATTTTGGGACTCCTGTGAACTCAGCCATACTCTATTATCTCAGTACAGCACAGGGTCCTGG
                                     <<<<<<<<<<<<<<<<<<<
 AGGGGCAGTAAAACCCTGAGGACACTTGAGTCGGTATGAGATAATAGAGTCATGTCGTGTCCCAGGACC
--------------------------------------------------------------------
 CAGATGCCCTGGTCCAACC

 GTCTACGGGACCAGGTTGG
--------------------------------------------------------------------
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: TGCCTTCCAGAGTGAGATTAA
REVERSE PRIMER: GTGGAGACTAGGAAATAATT
LENGTH OF INSERT: 333
--------------------------------------------------------------------
 GAGGAATTCCAAGTCAGAGCTGATCACTAAGGTGACGATCGATTGCCTTCCAGAGTGAGATTAACATCT
                                              >>>>>>>>>>>>>>>>>>>>
 CTCCTTAAGGTTCAGTCTCGACTAGTGATTCCACTGCTAGCTAACGGAAGGTCTCACTCTAATTGTAGA
--------------------------------------------------------------------
 GCCTCTAGAGGTGGCATGGCGAGCTTCACTTGTACCCAAAAGTGGGGCTCACAGTCTTGGGGTGTCCCA

 CGGAGATCTCCACCGTACCGCTCGAAGTGAACATGGGTTTTCACCCCGAGTGTCAGAACCCCACAGGGT
--------------------------------------------------------------------
 AGGATTGAGCGGGTTCTCCCCCAAACCCATGGGCAGCATTAATTATTTCCTAGTCTCCACGGAAGAGAC
                                          <<<<<<<<<<<<<<<<<<<
 TCCTAACTCGCCCAAGAGGGGGTTTGGGTACCCGTCGTAATTAATAAAGGATCAGAGGTGCCTTCTCTG
--------------------------------------------------------------------
 TTCAGGCAACTCCCAGCAGTTTCTTCCCCAGTAGNCCTGGGACAGGNGGANTTGAGAGGGACCCCTGGG

 AAGTCCGTTGAGGGTCGTCAAAGAAGGGGTCATCNGGACCCTGTCCNCCTNAACTCTCCCTGGGGACCC
--------------------------------------------------------------------
 GAAACTTGAGGCATGAGATTTATCAATATGTAACAGNTACAGTTTCATGANGGTNGG

 CTTTGAACTCCGTACTCTAAATAGTTATACATTGTCNATGTCAAAGTACTNCCANCC
--------------------------------------------------------------------
DUPLICATE REPORT FOR m4c.Seq: MR2058 (9)  MR2102 (4)
     vs MR2058 (score = 70)
       MR2058 GTCGGGGGGAGTCCTGACAGGCTTCGGAGCTCAGCTACTTACCTCCCAATTGAGGGAGGAAGGCA

                                                                GAGGAATTCCAAGTC
                                                                |||||||||||||||
       MR2058 AGTNATGGAAGTCTTTTCCGGCCTCAGACAGGGTGCTATCTATGGGGCTGGAGGAATTCCAAGTC

              AGAGCTGATCACTAAGGTGACGATCGATT-GCCTTCCAGAGTGAGATTAACATCTGCCTCTAGAG
              ||||||||||||||||||||||||||||| ||||| |||||| |||||||||| |||||||||
       MR2058 AGAGCTGATCACTAAGGTGACGATCGATTTGCCTTNCAGAGTNAGATTAACATNTGCCTCTAG
```

vs MR2102 (score = 58)

```
                                                  GAGGAATTCCAAGTCA-GAGCTGATCACTA
                                                  |||||||||||||||| ||||||||||||
MR2102 TCCGGCCTCAGACAAGGGTGCTATCTATGGGGCTGGAGGAATTCCAAGTCAAGAGCTGATCACTA

       AGG-TGACGATCGATTGCCTTCCA-GAGTGAGA-T-TAACATCTGC-C-TCTA-GAGGTG-GCAT
       ||| ||||  ||||||||||||| |||| | | | |  ||| | | | |||| ||||   | ||
MR2102 AGGGTGACCNTCGATTGCCTTCCAAGAGTNAAANTCT-NCAT-TNCACCTCTATGAGGNAAG-AT
```

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

```
SEQUENCE FILENAME: m5c.Seq
FORWARD PRIMER: AAGTGCCCTTGCTCTGCCC
REVERSE PRIMER: GTTCTTTNCNAGGTCCCNCG
LENGTH OF INSERT: 310
--------------------------------------------------------------------------
 TNCTTGTCTTTTAGAAAGGGCCTNGCTGCAAAGNCCTAAACTTTTCTNAGGTTTATGGTCTAATTCAGN

 ANGAACAGAAAATCTTTCCCGGANCGACGTTTCNGGATTTGAAAAGANTCCAAATACCAGATTAAGTCN
--------------------------------------------------------------------------
 GCNCGTGTGTCCCTAAGNCAGTCTGCCACATCAGGGACCAAGTCACCACAGCATAAAGGTGATCGGACG

 CGNGCACACAGGGATTCNGTCAGACGGTGTAGTCCCTGGTTCAGTGGTGTCGTATTTCCACTAGCCTGC
--------------------------------------------------------------------------
 NTAAGTGCCCTTGCTCTGCCCAGCCCTGCTTTGAAGNCCTCTTACAGTCAACGTCCTTGATACANGATG
   >>>>>>>>>>>>>>>>>>>
 NATTCACGGGAACGAGACGGGTCGGGACGAAACTTCNGGAGAATGTCAGTTGCAGGAACTATGTNCTAC
--------------------------------------------------------------------------
 GCCGCGCCCAANGGCTTTATGGAGCCCAGAGAGNAGGGGTGACTNAGGGCANTAAAGNGTCTTTTCGNG
                                                                      <<<<
 CGGCGCGGGTTNCCGAAATACCTCGGGTCTCTCNTCCCCACTGANTCCCGTNATTTCNCAGAAAAGCNC
--------------------------------------------------------------------------
 GGACCTNGNAAAGAACCACTTTAAGCCTGGAGCA
 <<<<<<<<<<<<<<<<
 CCTGGANCNTTTCTTGGTGAAATTCGGACCTCGT
--------------------------------------------------------------------------
DUPLICATE REPORT FOR m5c.Seq: m10c (-14)
     vs m10c  (score = 140)
       m10c  CAGTTGTGNATGGGNTCGNNGNAGNTGATGTGTCAGGTTGCCTGTAGTGGGTACAAGTGACAGAG

                                                 TNCTTGTCTTTTAGAAAGGGCCTNGCTGCAAAG
                                                 |||||||||||||||||||| |||||||||
       m10c    AGGGACAAAGAGACATATCACATGAGGGGGCATCCTTGTCTTTTAGAAAGGGCCTCGCTGCAAAG

               NCCTAAACTTTTCTNAGGTTTATGGTCTAATTCAGNGCNCGTGTGTCCCTAAGNCAGTCTGCCAC
                |||||||||||| |||||| |||||| |||||| || |||||||||||||| ||||||||||
       m10c    GCCTAAACTTTTCTGAGGTTTCTGGTCTGATTCAGGGCGCGTGTGTCCCTAAGGCAGTCTGCCAC

               ATCAGGGACCAAGTCACCACAGCATAAAGGTGATCGGACGNTAAGTGCCCTTGCTCTGCCCAGCC
               | |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
       m10c    ANCAGGGACCAAGTCACCACAGCATAAAGGTGATCGGACGCTAAGTGCCCTTGCTCTGCCCAGC

               CTGCTTTGAAGNCCTCTTACAGTCAACGTCCTTGATACANGATGGCCGCGCCCAANGGCTTTATG
       m10c

               GAGCCCAGAGAGNAGGGGTGACTNAGGGCANTAAAGNGTCTTTTCGNGGGACCTNGNAAAGAACC
       m10c

                 CTTTAAGCCTGGAGCA
       m10c

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

```
SEQUENCE FILENAME: m7c.Seq
FORWARD PRIMER: AAAAGTGCATTGGGGGGG
REVERSE PRIMER: CCAGGCACCACCAGGGAC
LENGTH OF INSERT: 185
--------------------------------------------------------------------
 GAAAAAGTGCATTGGGGGGGTCATCTTCTTCCATGAGACACTCTACCAGAAAGATGACAATGGTGTCCC
   >>>>>>>>>>>>>>>>>
 CTTTTTCACGTAACCCCCCCAGTAGAAGAAGGTACTCTGTGAGATGGTCTTTCTACTGTTACCACAGGG
--------------------------------------------------------------------
 CTTCGTCCGCACCATCCAGGATAAGGGCATTCTCGTAGGCATCAAGGTACACAGTCCCTGGTGGTGCCT
                                                  <<<<<<<<<<<<<<<
 GAAGCAGGCGTGGTAGGTCCTATTCCCGTAAGAGCATCCGTAGTTCCATGTGTCAGGGACCACCACGGA
--------------------------------------------------------------------
 GGGGGCTAGGTGTGGAAACCAGTNTATATAAGAGGGTGACAAGCCAG
 <<
 CCCCCGATCCACACCTTTGGTCANATATATTCTCCCACTGTTCGGTC
--------------------------------------------------------------------
DUPLICATE REPORT FOR m7c.Seq: m10a    (-13)
     vs m10a    (score = 85)
       m10a    TTGNCNGANTATTGCTCNACGGATCGTGACCCCCCCGCCAAGGGNATTCTGGCTGCACATGAGTC

       m10a    CGCAGNCAGCTATGGCCAAAAAGCTGAGCCAAATTNGGGTGTAGAACACTTNGGANAATCGNCGN

                                                  GAAAAAG-TGCATTGGGGGGGTCATCT
                                                  ||||||| ||||||||||||||||||||
       m10a    CTGTACCGCCAGNTCCTATNNANTGTTGATGTCCGTGTGAAAAAGGTGCATTGGGGGGGTCATCT

               TCTTCCATGAGACACTCTACCAGAAAGATGACAATGGTGTCCCCTTCGTCCGCACCATCCAGGAT
               || ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       m10a    TCATCCNTGAGACACTCTACCAGAAAGATGACAATGGTGTCCCCTTCGTCCGCACCATCCAGGA

               AAGGGCATTCTCGTAGGCATCAAGGTACACAGTCCCTGGTGGTGCCTGGGGGGCTAGGTGTGGAAA
       m10a

               CCAGTNTATATAAGAGGGTGACAAGCCAG
       m10a

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','A25388','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: CCCAGGGCTCCCAGACAG
LENGTH OF INSERT: 230
--------------------------------------------------------------------------
 AGCCCTCAGTGGAGGGCTGCCTGGCCAAGGAAGCAAGTGCTGGGGAAGTCTAGCCAANAATGCCGCTGT
                                      >>>>>>>>>>>>>>>>>>>>
 TCGGGAGTCACCTCCCGACGGACCGGTTCCTTCGTTCACGACCCCTTCAGATCGGTTNTTACGGCGACA
--------------------------------------------------------------------------
 CTAGCCCTGGAAGGAGGAGGGAGGCCCCTGGGGACCCAGATGCACAAGCTCAGGCCTGTNATCCTGGTC

 GATCGGGACCTTCCTCCTCCCTCCGGGGACCCCTGGGTCTACGTGTTCGAGTCCGGACANTAGGACCAG
--------------------------------------------------------------------------
 CNCTCCACAGCTGTCTGGGAGCCCTGGGCTCTTGCTTGCTCCTGAGGCTAGCGGGTTTATTTCCACACT
           <<<<<<<<<<<<<<<<<<
 GNGAGGTGTCGACAGACCCTCGGGACCCGAGAACGAACGAGGACTCCGATCGCCCAAATAAAGGTGTGA
--------------------------------------------------------------------------
 AGAACCCGTGTCCATNCCAAGAG

 TCTTGGGCACAGGTANGGTTCTC
--------------------------------------------------------------------------
DUPLICATE REPORT FOR m8c.Seq: m6a   (15)
     vs m6a   (score = 165)
              AGCCCTCAGTGGAGGGCTGCCTGGCCAAGGAAGCAAGTGCTGGGGAAGTCTAGCCAANAATGCCG
                 ||||||||||||||||||||||||||| ||| |||| |||||||||| ||||||| |||||||
       m6a        CCTCAGTGGAGGGCTGCCTGGCCAAGNAAGNAAGTNCTGGGGAAGTNTAGCCAAGAATGCCG

              CTGTCTAGCCCTGGAAGGAGGAGGGAGGCCCCTGGGGACCCAGATGCACAAGCTCAGGCCTGTNA
                ||  |||||||||||| |||||| |||||| |||| |||||||||||||||| ||| ||||||
       m6a    NTGNNTAGCCCTGGAAGAAGGAGGAAGGCCCNTGGGGNCCCAGATGCACAAGCTNAGGNCTGTNA

              TCCTGGTCCNCTCCACAGCTGTCTGGG-AGCCCTGGGCTCTTGCTTGCTCCTGAGG-CTAGCGGG
              |||||||  | | |||||||||||||| |||||||||||| ||  |||||||||| |||||||||
       m6a    TCCTGGTNNNTTNCACAGCTGTCTGGGGAGCCCTGGGCTC-TG-NTGCTCCTGAGGGCTAGCGGG

              TTTATTTCCACACTAGAACCCGTGTCCATNCCAAGAG
              |||||||||||||||| ||||
       m6a    TTTATTTCCACACTAGCACCC

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: AAAAAAAGAAAATGATTCTT
LENGTH OF INSERT: 239
------------------------------------------------------------------------
 GGAAAAAGAGAAAAAGCTGAAACTCTGGACATATCTTAAACCGTTTGGCATTTCCTNGACTGAGTACAT
                                                                      >>
 CCTTTTTCTCTTTTTCGACTTTGAGACCTGTATAGAATTTGGCAAACCGTAAAGGANCTGACTCATGTA
------------------------------------------------------------------------
 CCAGACCGTCCAAGGCAAGAAATGTAAGCAGAGATCTTCATTTCTCTGGAGCTCAGAATTGAACCCAGA
 >>>>>>>>>>>>>>>>
 GGTCTGGCAGGTTCCGTTCTTTACATTCGTCTCTAGAAGTAAAGAGACCTCGAGTCTTAACTTGGGTCT
------------------------------------------------------------------------
 GCCTCTGTGTGCTAGGCAAATCCAGAGCAGTATCTTTAGCCCAAAGCAGGGAAGTTTTATAAGAATCAT
                                                                <<<<<<<<<
 CGGAGACACACGATCCGTTTAGGTCTCGTCATAGAAATCGGGTTTCGTCCCTTCAAAATATTCTTAGTA
------------------------------------------------------------------------
 TTTCTTTTTTTTTTGGTTTTTCGAGTCGGGGG
 <<<<<<<<<<<
 AAAGAAAAAAAAAAACCAAAAAGCTCAGCCCCC
------------------------------------------------------------------------
DUPLICATE REPORT FOR m9c.Seq: m4a   (44)
     vs m4a    (score = 235)
               GGAAAAAGAGAAAAAGCTGAAACTCTGGACATATCTTAAACCGTTTGGCATTTCCTNGACTGAGT
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
       m4a     GGAAAAAGAGAAAAAGCTGAAACTCTGGACATATCTTAAACCGTTTGGCATTTCCTGGACTGAGT

               ACATCCAGACCGTCCAAGGCAAGAAATGTAAGCAGAGATCTTCATTTCTCTGGAGCTCAGAATTG
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       m4a     ACATCCAGACCGTCCAAGGCAAGAAATGTAAGCAGAGATCTTCATTTCTCTGGAGCTCAGAATTG

               AACCCAGAGCCTCTGTGTGCTAGGCAAATCCAGAGCAGTATCTTTAGCCCAAAGCAGGGAAGTTT
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       m4a     AACCCAGAGCCTCTGTGTGCTAGGCAAATCCAGAGCAGTATCTTTAGCCCAAAGCAGGGAAGTTT

               TATAAGAATCATTTTCTTTTTTTTTTGGTTTTTCGAGTCGGGGG
               |||||||||||||||||||||||||||||||||||||||||
       m4a     TATAAGAATCATTTTCTTTTTTTTTTGGTTTTTCGAGTCGGG

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. Lisitsyn[1], Julia A. Segre[2], Kenro Kusumi[2], Natalia M. Lisitsyn[1], Joseph H. Nadeau[3], Wayne N. Frankel[3], Michael H. Wigler[1] & Eric S. Lander[2]

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 usefullness 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.

[1]*Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, New York 11724, USA*
[2]*Whitehead Institute for Biomedical Research, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA*
3 *The Jackson Laboratory, Bar Harbor, Maine 04609, USA*

Correspondance should be addressed to E.S.L.

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 organism[1]. 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 Driver[2]. 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 a large and, therefore, poorly amplifiable restriction fragment in the 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 F_2 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

strains[3] 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.)

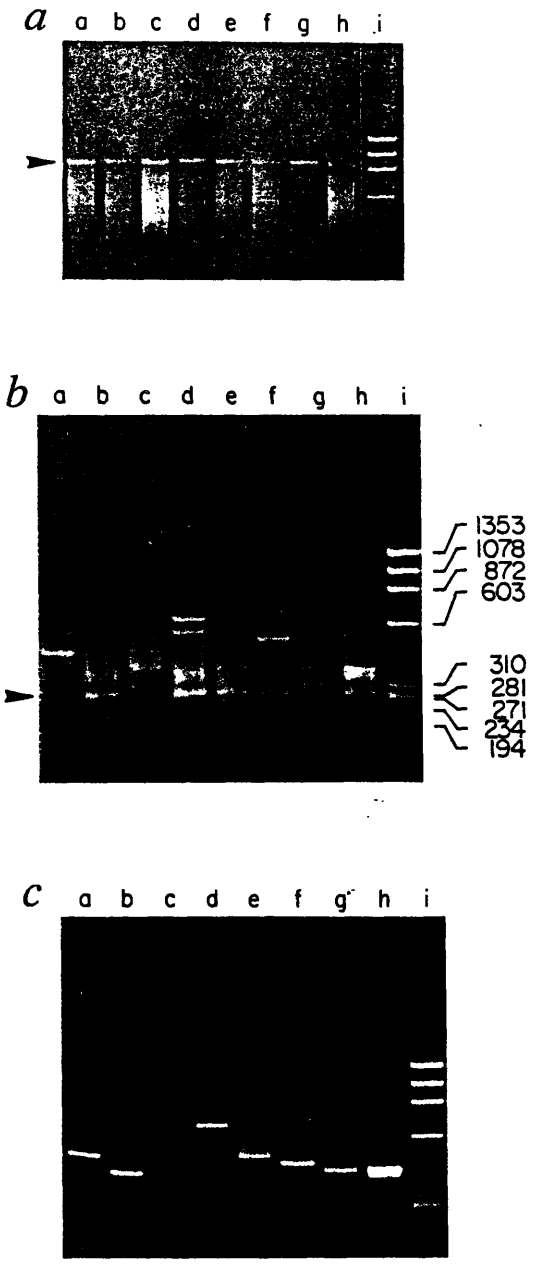


Fig. 1 Agarose gel electrophoresis of difference -products obtained after the first (a), second (b) and third (c) hybridization-extension-amplification steps in various experiments. Lanes are: *a*, *staggerer* cross; *b*, *Lurcher* congenic; *c*, *stargazer* congenic; *d*, *pudgy* congenic; *e*, *nude* congenic; *f*, *nude* cross; *g*, *severe combined immunodeficiency* congenic; *h*, *tottering* congenic; and *i*, *Hae*III digest of ϕX174 RF DNA. Sizes (bp) are indicated to the right. Arrows on the left show abundant mouse repeats removed by subsequent subtractions.

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 *Bgl*II 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 *Bgl*II 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

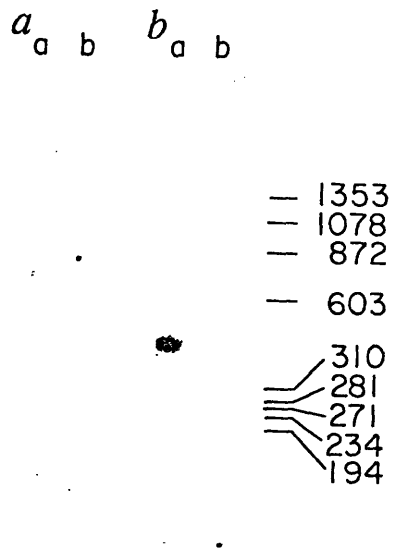


Fig. 2 Autoradiograms obtained after hybridization of probe RDA-4.5 (*a*) from *pudgy* congenic strains and probe RDA-8.2 (*b*) from *tottering* congenic strains to Southern blots containing *Bgl*II amplicons from Tester (lane a) and Driver (lane b). Sizes (bp) are indicated to the right. The faint band above the major DNA fragment is an unidentified PCR byproduct frequently observed on blots of *Bgl*II 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 *Bgl*II polymorphisms mapping near *pudgy* and *tottering*, respectively. RDA-4.5 detected a *Bgl*II 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*[4]. 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 *Bgl*II 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*[4].

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

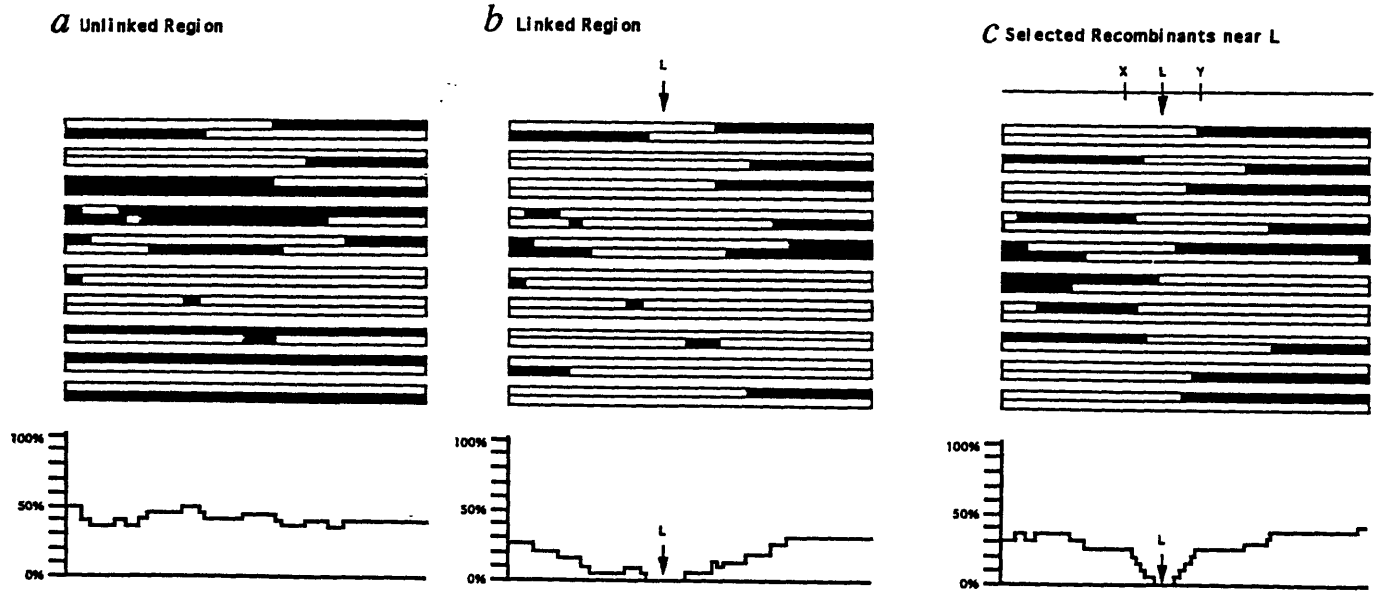


Fig. 3 Schematic diagram representing the principle underlying GDRDA with progeny from an F_2 intercross. Each panel shows hypothetical chromosomal genotypes from 10 progeny to be pooled to create a Driver; each chromosome is arbitrarily drawn to be 100 cM. Strain A carries a recessively-acting allele 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).

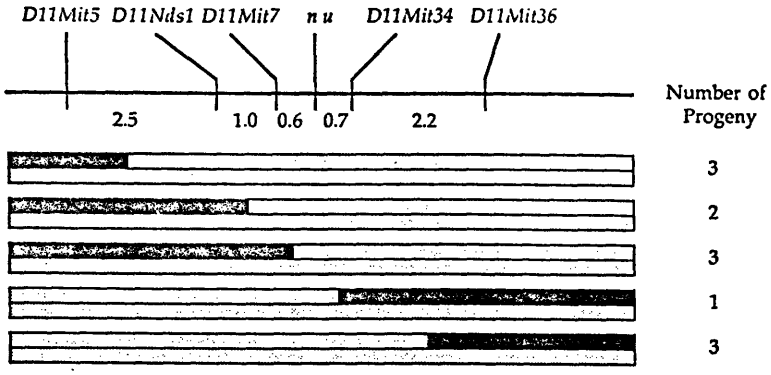


Fig. 4 Schematic diagram indicating chromosomal genotypes of the 12 (MOLF/Ei × AKR/J-nu^{str}) F2 progeny pooled to create the Driver for GDRDA, relative to a genetic map of polymorphic markers near the *nude* locus. Black indicates regions derived from MOLF/Ei. Shading indicates regions derived from AKR. The number of progeny of each type is indicated at the right.

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 F_2 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 F_2 progeny should have genotypes AA, AB and BB in the ratio 1:2:1 at unselected loci (see Fig. 3*a*,*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. 3*c*). 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 F_2 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^{str}) F_2 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 *D11Mit5* and *D11Mit36*, and 4 of the 12 occurred within a 1.3 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, GDRDA should produce MOLF/Ei alleles of polymorphisms in the interval *D11Mit5* and *D11Mit36*. 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 *D11Mit7* and *D11Mit34*.

Using this Tester and Driver combination, we performed RDA with the restriction enzyme *Bgl*II. 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 *Bgl*II. It detected an RFLP with a 450 bp allele in MOLF/Ei and a 4 kb allele in AKR/J-nu^{str}. 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^{str}) F_2 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 F_2 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 *D11Mit7* and *D11Mit34*. Subsequent analysis of additional F_2 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 *Bam*HI. 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^{wh} 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 chincilla (c^{ch}) stock, in *trans* to the nearby *p* mutation (that is, $pu + c^{ch}/+ p\ c^{ch}$) 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 *pu-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). *Tottering* (*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 a (C3H/HeJ × 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 a 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[2]. 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 *Sau*3A 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 each of Tester and Driver DNAs with either *Bgl*II or *Bam*HI. The iterative hybridization-extension-amplification step was repeated three times. The resulting material was digested with the same restriction enzyme as used to prepare the amplicon, ligated to *Bam*HI-digested and dephosphorylated pBluescript II SK(−), and transformed into *E. coli* XL-Blue competent cells according to the supplier's recommendations.

Initial characterization of RDA clones. 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'-CGACGTTGTAAAACGACGGCCAGT-3') and Rev25 primer (5'-CACACAGGAAACAGCTATGACCATG-3'), 67 mM Tris-HCl (pH 8.8 at 25 °C), 4 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$, 10mM β-mercaptoethanol, 170 μg ml^{-1} 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 Ampli*Taq* polymerase (Perkin-Elmer Cetus) are 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 analysed 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 a Megaprime DNA labelling system (Amersham) according to the 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 0.5× SSC, 0.1% SDS at 65 °C).

Genetic mapping of RFLPs. 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. Southern blotting and hybridization was essentially as described[7]. The inheritance pattern of the RFLPs was compared to that of various simple sequence length polymorphisms (SSLPs) that mapped to the regions of interest. The SSLPs and the genotyping protocol were previously described[8,9].

Calculation of $c_n(\alpha)$. As described in Box 1, the proportion of linked to unlinked clones produced by GDRDA depends on the ratio $c_n(\alpha) = a_n(\alpha)/b_n(\alpha)$, where $a_n(\alpha)$ is the expected length of the region linked to L for which the proportion π_B of B alleles is less than α, $b_n(\alpha)$ is the expected length of the unlinked regions of the genome for which $\pi_B < \alpha$, and n is the number of recombinant haploid genomes pooled (that is, $n = 2k$, where k is the number of F2 progeny pooled). The function was calculated as follows:

$$a_n(\alpha) = \int_{-\Delta}^{\Delta} p_n(\alpha,\theta(x))\,dx \quad \text{and} \quad b_n(\alpha) = p_n(\alpha,0.5)\,G,$$

where

$$p_n(\alpha,x) = \sum_{i=0}^{i\le\alpha n} \binom{n}{i} x^i (1-x)^{n-i}$$

denotes the cumulative probability distribution of the binomial distribution for n independent events having probability x, $\theta(x)$ is a chosen map function converting genetic distance to recombination frequency (we used Haldane's map function, $\theta(x) = (1 - e^{-2x})/2$), Δ denotes the maximum distance that should be considered to be 'linked' to L (we used $\Delta = 0.50$ Morgans), G denotes the genetic length of the 'unlinked' genome, and all lengths are measured in Morgans. The first equation follows immediately from the definition of $p_n(\alpha,x)$. The second equation follows by first noting that the proportion of B alleles at distance x from L is binomially distributed with probability equal to the recombination fraction $\theta(x)$ and then integrating over the 'linked' points in an interval $[-\Delta, \Delta]$ centred on L.

Acknowledgements

We thank L. Rodgers, M. Riggs, J. Smith, A. Weaver, D. Varnum, J. Duffy, M. Okler, A. Messer, L. Shultz and M. Davisson for their help and contribution to this work. W.N.F. was a Special Fellow of the Leukemia Society of America. This work was supported in part by grants from the National Center for Human Genome Research, the National Science Foundation, and the Markey 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.

1. Collins, F. Positional cloning: let's not call it reverse anymore. *Nature Genet.* 1, 3–6 (1992).
2. Lisitsyn, N., Lisitsyn, N. & Wigler, M. Cloning the difference between two complex genomes. *Science* 259, 946–951 (1993).
3. Snell, G.D. Histocompatibility genes of the mouse. II. Production and analysis of isogenic resistant lines. *J. natn. Cancer Inst.* 21, 843–877 (1958).
4. Hillyard, A.L., Doolittle, D.P., Davisson, M.T., Maltais, L., & Roderick, T.H. Locus map of the mouse with comparative map points of human on mouse. (GBASE Electronic Database, Jackson Laboratory, Bar Harbor, Maine, March, 1993).
5. Williams, J.G.K., Reiter, R.S., Young, R.M. & Scolnick, P.A. Genetic Mapping of mutations using phenotypic pools and mapped RAPD markers. *Nucl. Acids Res.* 21, 2697–2702 (1993).
6. Lyon, M.F. & Searle, A.G. *Genetic Variants and Strains of the Laboratory Mouse* 2nd edn (Oxford Univeristy Press, Oxford, 1989).
7. Brown, T. in *Current Protocols in Molecular Biology* (eds Ausubel, F. *et al.*) 2.9.1–2.9.15 (Green, New York, 1993).
8. Dietrich, W. *et al.* A Genetic Map of the Mouse Suitable for Typing Intraspecific Crosses. *Genetics* 131, 423–447 (1992).
9. Whitehead Institute/MIT Center for Genome Research, Mouse Genome Public Electronic Database, "genome_database@genome.wi.mit.edu" (July, 1993).

Received 17 August; accepted 18 October 1993.

the expected pattern of hybridizing to the Tester but not the Driver amplicon. Both probes detected RFLPs between MOLF/Ei and AKR/J-*nu*^{nu} (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 *nude*. The 12 F_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 cM interval between *D11Mit7* and *D11Mit34*. Subsequent analysis of additional F_2 progeny (J.A.S., J.H.N., Benjamin

Box1 Experimental design issues

In applying GDRDA to an F_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 α 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 of alleles 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 α, how many progeny k should be pooled? Let $a_n(\alpha)$ denote the expected length of the region linked to L for which the proportion π_B of B alleles is less than α and $b_n(\alpha)$ denote the expected length of *unlinked* regions of the genome for which $\pi_B < \alpha$, where n is the number of recombinant haploid genomes pooled to create the Driver (that is, $n = 2k$). The ratio $c_n(\alpha) = a_n(\alpha)/b_n(\alpha)$ should give a good indication of the ratio of linked to unlinked clones that should be produced by RDA. (For the calculation of $c_n(\alpha)$, see Methodology.) If $c_n(\alpha) >> 1$, then linked clones should consititute the majority of fragments surviving subtraction. If $c_n(\alpha) < 1$, linked clones will be a minority which must be identified by subsequent screening. The number of progeny should thus be $k \geq n/2$, where n is the smallest integer such that $c_n(\alpha) \geq C$, for a chosen lower bound C.

A graph of $c_n(\alpha)$ 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 $c_n(0.15) = 7.7, 13.6, 24.1$, and 182.6 for $k = 6, 8, 10$ and 12 F_2 progeny pooled (with $n = 2k$). To ensure $c_n(\alpha) >> 1$, it might thus be prudent to pool at least 10 F_2 progeny.

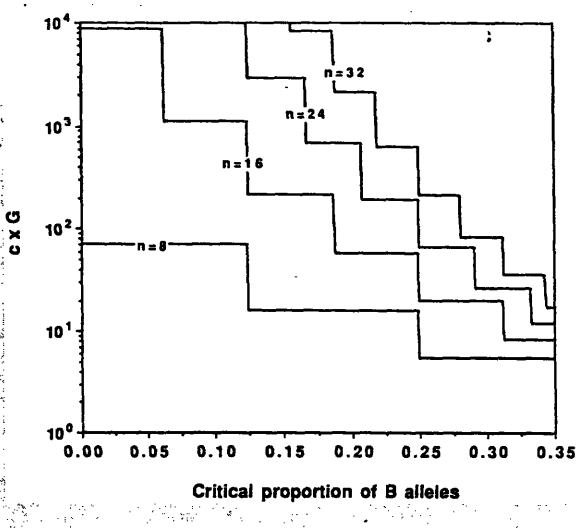


Fig. 5 Mathematical analysis of construction of Driver for an F2 intercross. Curves show the value of $c_n(\alpha) \times G$, where G is the length of the target genome measured in Morgans and $c_n(\alpha)$ is closely related to the ratio of linked to unlinked probes expected to occur in the difference-product, assuming a critical proportion α of B alleles necessary for efficient subtraction and a Driver pool containing n haploid genomes (see text for exact description). Given α, the number n should be chosen to ensure that $c_n(\alpha) >> 1$ or, using the graph, that $c_n(\alpha) \times G >> G$. (Note, the function $c_n(\alpha) \times G$ is given because, unlike $c_n(\alpha)$ itself, the value does not depend on G.) The curves are step functions because, if n haploid genomes are pooled, the proportion of B alleles at any locus must be an integral multiple of $1/n$.

How close to L will the linked polymorphisms be? Assuming the simple model of a critical threshold α for subtraction, they would be expected to lie roughly within recombination fraction $\theta \approx \alpha$ of L. If $\alpha = 0.10$, the target interval thus may be about 20 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 progeny can be selected having recombinations near L (as discussed in the text), 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 x B)F1 x A backcross, use the (A x 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_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 twice as many as that for the corresponding intercross (that is, $k = n$, where n is the smallest integer such that $c_n(\alpha) \geq C$).

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 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 $b_n(\alpha)$ should be replaced by $2b_n(\alpha)$. The minimum number of backcross progeny that should be pooled is thus $k = n$, where n is the smallest integer such that $c_n(\alpha)/2 \geq C$. 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_1, D_2, ..., D_j. The potential drawback is that a linked C allele could be subtracted if it is present in any of the D_j, which would decrease the number of detectable polymorphisms as j 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 Davisson, 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 *D9Mit9*. 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 *D9Mit49* and *D9Mit11*.

RDA was performed with the restriction enzyme *Bgl*II. 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 *Bgl*II 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 *D9Mit49* and *D9Mit11* based 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 Davisson 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 *Bam*HI 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.

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 L. 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 *Bgl*II.

We have focussed here on the application of GDRDA to F_2 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[5], 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

GENOMICS 28, 549–559 (1995)

Positional Cloning of the *nude* Locus: Genetic, Physical, and Transcription Maps of the Region and Mutations in the Mouse and Rat

JULIA A. SEGRE,*,† JENNIFER L. NEMHAUSER,* BENJAMIN A. TAYLOR,‡ JOSEPH H. NADEAU,‡,[1] AND ERIC S. LANDER*,†,[2]

*Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142; †Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; and ‡The Jackson Laboratory, Bar Harbor, Maine 04609

Received March 16, 1995; accepted May 26, 1995

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. © 1995 Academic Press, Inc.

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 availability 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^{str}), which arose within the inbred AKR/J mouse colony at The Jackson Laboratory in 1974 (Eicher, 1976; Flanagan, 1966). Two alleles

[1] Present address: Dept. of Human Genetics, Montreal General Hospital, Montreal, Quebec H3G1A4, Canada.

[2] To whom correspondence should be addressed at the Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142. Telephone: (617) 252-1906. Fax: (617) 252-1933.

0888-7543/95 $12.00

of *nude* exist in the rat: *rnu* and rnu^N, both of which arose on outbred strains (Berridge *et al.*, 1979; Festing *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 *et al.*, 1994; Zha *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 *nude* locus, including genetic distance, physical distance, and gene density. We genetically mapped the *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 *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 *Drosophila, 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 *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 *nude* phenotype in mice and rats is caused by mutations in a gene encoding a novel winged helix (or fork head) domain transcription factor, *whn* (Nehls *et al.*, 1994). In that report, the mutations in the mouse *nu* and the rat rnu^N alleles are described. Here, we report the complete coding sequence of the rat *whn* mRNA together with the mutation present in the rat *rnu* allele, as well as confirming the mutation in the mouse *nu* allele. We have also analyzed the mouse nu^{str} allele but find no mutation in the coding region.

MATERIALS AND METHODS

Animals. Congenic C57BL/6-*nu/nu* mice were developed by repeatedly backcrossing outbred *nude* (*nu*) mice onto C57BL/6; they were purchased from Taconic Farms (Germantown, NY). The nu^{str} 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 *rnu* mutation was discovered 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 from The Jackson Laboratory. Progeny were phenotyped 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 *nude* region were progeny tested by mating them to nu^{str}/+ animals. Informative progeny (i.e., those determined, based on genotype at flanking markers, to carry the recombinant chromosome over a nu^{str}-bearing chromosome) were examined for phenotype to determine whether the recombinant chromosome carried the nu^{str} 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 and electrophoresed in 1× 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 *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 RIII (P1-P2 and P7-P23) or mouse ES cell line 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 *Mbo*I 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 (Haldi *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 *Sal*I site in the P1 vector, 50 ng P1 insert DNA was digested with *Hin*fI, *Rsa*I, or *Alu*I. To clone the insert DNA adjacent to the *Not*I site in the P1 vector, digestion was performed with *Hha*I or *Alu*I. Digestion products were religated to promote circularization. PCR was performed on 0.5 ng of ligation product using P1UF (5′-GCCGCTAATACGACTCACTATAGG-3′) and P1UR (5′-GAGCAATATAGTCCTACAATGTCAAGC-3′) for the *Not*I end and P1LF (5′-GAGGATCGAAACGGCAGA-3′) and P1LR (5′-CCGTCGACATTTAGGTGACAC-3′) for the *Sal*I end. The resulting PCR products were then reamplified to introduce M13 sequencing primer sites. The chimeric primers M13FORUF (5′-TGTAAAACGACGGCCAGTCACTATAGGGAGAGGATC-3′) and M13REVUR (5′-CAGGAAACAGCTATGACCAGTCCTACAATGTCAAGC-3′) were used for the *Not*I end, M13FORLF1 (5′-TGTAAAACGACGGCCAGTAGATCGCAAAAAACAGTA-3′) and M13REVLR (5′-CAGGAAACAGCTATGACCTGACACTATAGAAGGATC-3′) to reamplify the *Sal*I end digested with *Hin*fI and *Rsa*I, and M13FORLF2 (5′-TGTAAAACGACGGCCAGTGCCTCAATTCAATCAACG-3′) and M13REVLR to reamplify the *Sal*I end digested with *Alu*I. 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 *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 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 ^{32}P specifically primed with an oligonucleotide from the vector sequence flanking the insert (5′-CTGAGCGGAATTCGTGAGACC-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 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-GCG-GTGACCCGGGAGATCTGAATTC-3′) and Blunt-2 (5′-GAATTC-AGATC-3′)). Primary cDNA for the selection was independently prepared by random-priming 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 *Bam*HI and *Bgl*II (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-CCAAGTCTGT-3′) poly(A)$^+$ selected mRNA from C57BL/6J, C57BL6J-*nu/nu*, AKR/J, AKR/J-*nu^{str}/nu^{str}*, rat, and rat-*rnu/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 F_2 intercrosses. To ensure the greatest rates of polymorphisms between the two strains of the cross, we mated AKR/J-*nu^{str}* 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^{str}* × C57BL/6J)F_2 animals, 700 (AKR-*nu^{str}* × MOLF/Ei)F_2 animals, and 226 (AKR-*nu^{str}* × CAST/Ei)F_2 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 F_2 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 F_2 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 (PFG) 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 PFG 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: *Mlu*I, *Not*I, *Rsr*II, *Sac*I, and *Sfi*I. 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

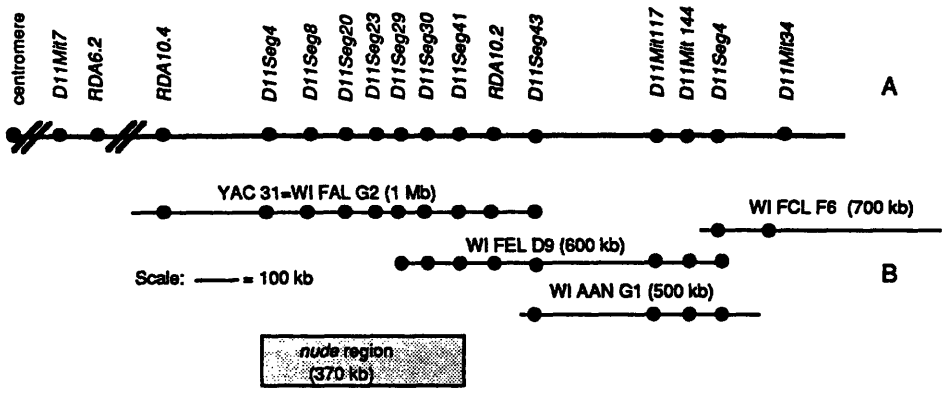


FIG. 1. Large-scale map of the *nude* region. (**A**) Genetic map of polymorphic markers. (**B**) STS content mapping of YACs with genetic markers. The location of the 370-kb region containing *nude* is shown. Primer sequences are given in Table 1 or published in Dietrich *et al.* (1992, 1994).

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 physi-

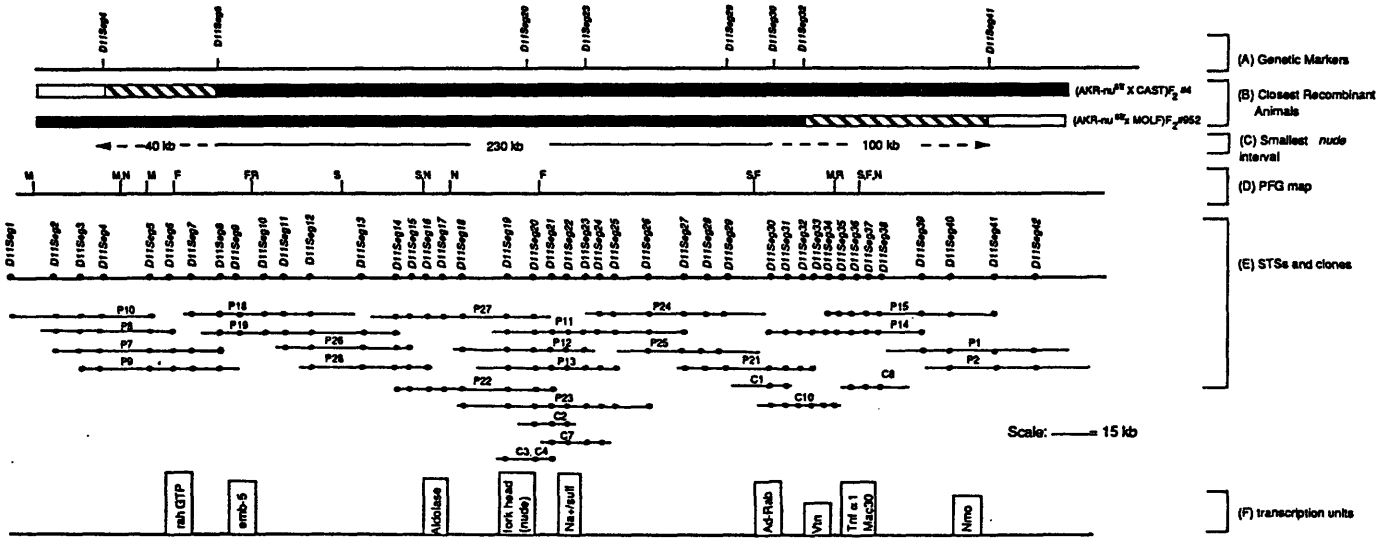


FIG. 2. Fine structure map of the *nude* region. (**A**) Genetic markers. (**B**) Recombinational breakpoints for the closest recombinant animals in *nu* crosses. The empty bar denotes the chromosomal region that recombines with the *nu* gene. The black bar denotes the chromosomal region that does not 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, *Sac*II; F, *Sfi*I; M, *Mlu*I; N, *Not*I; R, *Rsr*II. (**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.

TABLE 1

Primer Sequences for a Subset of Gene Fragments Identified, Together with Product Size, Gene Showing Sequence Similarity, and Clones Containing the Sequence Based on STS Content

Locus	Forward sequence	Reverse sequence	Product size	Gene recognized	Clones recognized
D11Seg1	TAACTGGACTTCCTGGTTAGGC	TCTTCCCAGTCCCCTACCTC	97		P10
D11Seg2	CCCTTTTGAATACCTCACCC	GCACCGACAGCAAGGTGA	181		P10, 8, 7
D11Seg3	GGAACTCTCCGTAGATGAGGC	CTGGCATCCCAGCATCTG	163		P7, 8, 9, 10
D11Seg4	TTACCCGAGGCCTTCATG	TCATGAGCTGCTTGCCTAAG	222		P7, 8, 9, 10
D11Seg5	CTCAGGCATCGCTCACATTT	GCACTGAGGCACAAGTCCT	196		P7, 8, 9, 10
D11Seg6	GGCTGTACCGGTCAGAGATG	TCCAATGAGACTGAAGCTTCC	178		P7, 8, 9
D11Seg7	GCCTTGTCATCCCATTTTGC	GCCAAATCCCTACCTTGAGC	134	rah GTP	P7, 9, 18
D11Seg8	TCACCTGGGTACTGAGAATGG	GCTTTCATGGAAGTCCAAGG	200		P7, 9, 18, 19
D11Seg9	CTTAGTTAACTTGCCCTTGGCT	CCCATAGGAAAGTGTCAGGAAA	129		P18, 19
D11Seg10	CCAACAAGTTGCATTTCCTT	AGCATCCCGTTACTGAGC	143	emb-5	P18, 19
D11Seg11	TGGGCACCTTGTTGTCTTTA	CCCAGACTACGCGGGGGGA	140		P18, 19, 26
D11Seg12	CACCTGTGTTCTCATGGCC	CCAGGTTAAGCACTTCCAGTT	220		P18, 19, 26, 28
D11Seg13	TCCTACATTCTTAGGCTGGGC	CCAGGCCCAGGGGTTAGCTGC	179		P19, 26, 28
D11Seg14	GAGAAATGGCAAGACCCTCG	AACACTTTGTTTAGTGCTGGC	157		P19, 26, 27, 28
D11Seg15	TCGGCACTACCCACCTATCT	TCCTTTCAGCTCTTCCTCT	149		P22, 26, 27, 28
D11Seg16	AAGGCATTCATGGTGAGCTA	CAGAATGATCTCTGATTCAGCC	140		P22, 27, 28
D11Seg17	ACCATGCCTACTGAGGCACG	GTAAGGCATGTGAGATGTGA	155	aldolase	P22, 27
D11Seg18	GGCCGATGATGTTGAAGG	AAGTGCCGTTTCCAAGAAAG	160		P12, 22, 23, 27
D11Seg19	CACTTCTGCAGTTCTTCCTGC	CTGTTCGCCATAACCTGTCC	136	fork head (*Hfh11*)	P11, 12, 13, 22, 23, 27
D11Seg20	CTCCACGGACATTCTGCC	TCAACTGTCTTAAGCATCCTGA	160		P11, 12, 13, 22, 23, 27
D11Seg21	CAAGTGCTGGGGAAGTCTAG	CCCAGGGCTCCCAGACAG	133		P11, 12, 13, 22, 23, C2, C7
D11Seg22	GGCACTATGGGCCTATCGC	ACACCACAGCAGCGCCAT	133	Na^+/sulf	P11, 12, 13, 23, C2, C7
D11Seg23	AACTCCCATCCAGGGAGTCT	TGGAGAGAATAATGGGGCAG	210		P11, 12, 13, 23, C7
D11Seg24	GTCAGGCCCCCAGTAAACTG	AGGCAGCTTGGGGTGCTTC	143		P11, 13, 23, 24, C7
D11Seg25	TCACCAATGCGTGTGACCT	TCTTGACTCCTTTGAGGTAGCC	174		P11, 13, 23, 24
D11Seg26	TTGTATCAGCCAACTGTCTTG	TGGCTGGAGTCATGGCTAG	166		P11, 23, 24, 25
D11Seg27	TATGCACATGTACACACAATGC	CTGGCTACTGTTTTCTGGCT	183		P21, 24, 25
D11Seg28	CCACTCCCCTCTAACCCCG	CAGCTGACAGACCTTGGCCTCT	131		P21, 24, 25
D11Seg29	CTCTGTAACTGCCACCTCACC	CTTGTGGTGTCTATAAATTGTAAAGG	166		P21, 24, 25
D11Seg30	TCTAGCTACCTATAGGCTCGCA	TGCGTTAACTCAGCTAACTCG	148		P14, 21, C1, C10
D11Seg31	CCTTACTCCTGTTCTGGTACCC	CCTGGTCAGCTCGGTCTG	157	Ad-Rab	P14, 21, C1, C10
D11Seg32	CACAGCAACAAGCCAAGAAA	TCACTGTGCATAGCAATCAGC	180		P14, 21, C10
D11Seg33	ACCTTTGGCCTCCAGGGA	CCAGAGTCACATAGCCCCAGGT	197		P14, 21, C10
D11Seg34	CTCTACCCTCCCGCGTCT	GATCCGGTAAAGGTGCTGG	123		P14, 15, C10
D11Seg35	TTTAAGCAGCAACCCAGCC	GAAAATGTTCTCCCAGCTGTC	102	Vtn	P14, 15, C10
D11Seg36	GGACACGGTCCCGTAGTTC	TCTATGAGACCCCTCGAGTCC	114	TNFAIP1	P14, 15, C8
D11Seg37	ACTCAAAGGAGTTCAAAGACCC	AATGGGAAAGAAAGGCAGCT	100	Mac30	P14, 15, C8
D11Seg38	AACCGTGCAGAGGGCAAA	TCTGCCCAGTCTCATATTTTCC	81		P14, 15, C8
D11Seg39	AGGGGTCTTGAACCTGCCA	AGGGGTCTTGAACCTGCCA	170		P1, 14, 15
D11Seg40	GACGCTGGCTGTGTTAGCTG	TTTCCTTTCTGCTGTTCCAG	121	nmo	P1, 2, 15
D11Seg41	CCACTCACTAAGAGGCCAGC	TGCAAAAGAAGGGGGTAAGA	122		P1, 2, 15
D11Seg42	GTCAGGAGAAAAAACAACCAGG	TCAGACTCCTCTACCAGGATGG	584		P1, 2
D11Seg43	TTGAAAGTTAGGATGGAGGTGG	TCTGTGCATGTGCAGACAGA	165		None
D11Seg44	TGACTCTATGGCTGTGACATTG	CTGGGTCCTGGTACTGAACC	109		None

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 *nude* 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 *nude* 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 *nude* 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 *nude* region. Exon trapping is a strictly genomic approach, relying upon the fact that most mammalian

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 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 3 unique clones per pool of two cosmids. Even when eight cosmids were grouped together, a maximum of 3 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.

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 SPT6, 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.

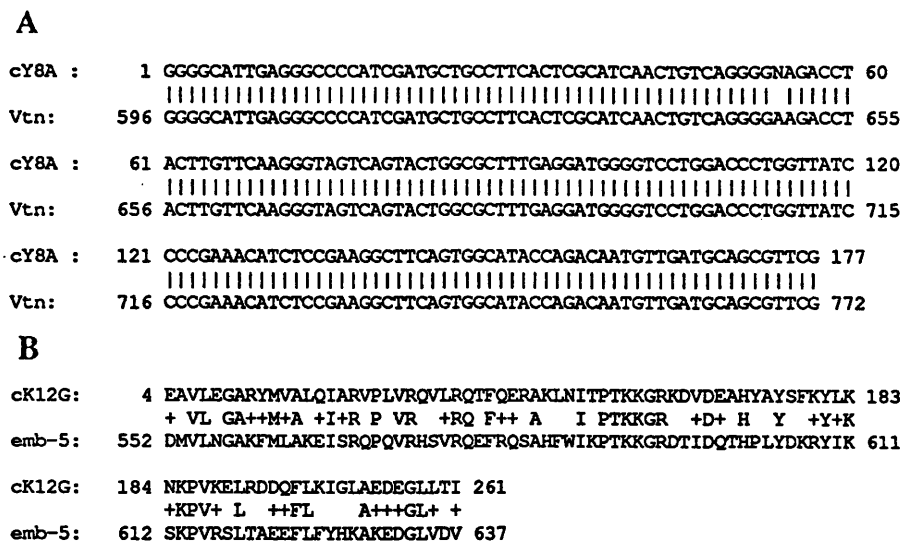


FIG. 3. Examples of strong sequence similarities. (**A**) Nucleotide sequence alignment of a direct cDNA selected clone, cY8A, with mouse vitronectin mRNA, produced by the BLASTN program. The sequences have perfect nucleotide identity over 177 nucleotides with the exception of a single undetermined nucleotide in cY8A. Numbering corresponds to nucleotide position for both the mouse vitronectin mRNA (Vtn) and cY8A. (**B**) Amino acid alignment of a direct cDNA selected clone, 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 between the two sequences; + denotes conservative amino acid substitution; blank denotes nonconservative substitution. Numbering is in amino acid residues for emb-5 and in nucleotides for 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, 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 θ 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

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 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 fragments 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 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.

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.

Expression Analysis of Transcription Units

To determine the expression pattern of each transcription unit, we performed RT-PCR on a panel of adult mouse tissues (including skin, thymus, liver, testes, etc.) and whole embryos at different days of development (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 selected clones were superior to the exon trapped clones because they tended to be longer and encompass sequence from two exons. Those gene fragments presumed to belong to a common transcription unit (by virtue of strong similarity to a gene in GenBank) always showed a consistent pattern of expression (data not shown).

Gene fragments, showing no sequence similarity to previous genes, were grouped according to their physical 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 characterized by hybridization to a Southern blot of restriction-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 candidate genes, Nehls *et al.* (1994) reported that the winged helix homologue above has mutations in the mouse *nu* allele and the rat rnu^N 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 determined 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 confirmed the reported mutation in the mouse *nu* allele and searched for the mutations in the nu^{str} allele. Interestingly, no mutations were found in the coding region in the nu^{str} allele. Northern blot analysis also indicates that the *Hfh11* transcript is present at approximately normal levels in adult skin from AKR/J-nu^{str}/nu^{str} homozygotes (data not shown). Thus, the nu^{str} mutation appears to be a more subtle change and remains to be identified.

DISCUSSION

Genetic Resolution

We used F_2 intercrosses segregating the *nude* phenotype to obtain two informative meioses for every progeny. Initial mapping studies with 200 animals localized *nude* to a 1.6-cM interval between *D11Mit7* and *D11Mit34.* 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 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 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. Map-

```
          M  V  S  L  L  P  P  H  S  D  V  T  L  P  G  S  T
GACTGGGTGATGGTGTCGCTACTCCCTCCGCACTCTGACGTCACACTTCCAGGCTCCACC      60
 R  L  E  G  E  P  Q  G  D  L  M  Q  A  P  G  L  P  G  S  P
CGACTGGAGGGCGAACCCCAAGGGGACCTCATGCAGGCTCCGGGCCTCCCAGGCTCCCCT     120
 A  P  Q  N  K  H  A  N  F  S  C  S  S  F  V  P  D  G  P  P
GCCCCACAGAACAAGCACGCCAACTTCAGCTGCTCATCATTTGTGCCTGATGGCCCTCCA     180
 E  R  A  P  S  L  P  P  H  S  P  S  I  A  S  P  G  P  E  Q
GAGAGGGCCCCCTCGCTGCCCCCCCACAGCCCAAGCATCGCATCTCCAGGCCCAGAGCAG     240
 I  Q  S  H  C  T  A  G  P  G  P  G  S  F  R  L  S  P  S  D
ATCCAGAGCCACTGCACAGCTGGACCTGGCCCAGGCTCCTTCCGCCTTTCTCCCTCAGAC     300
 K  Y  P  G  F  G  F  E  E  G  P  A  G  S  P  G  R  F  L  R
AAGTATCCTGGCTTTGGCTTTGAGGAGGGCCCAGCAGGCAGCCCAGGGCGCTTCCTCAGG     360
 G  N  H  M  P  F  H  P  Y  K  G  H  F  H  E  D  I  F  S  E
GGCAACCACATGCCTTTCCACCCTTACAAGGGGCACTTCCATGAAGACATCTTCTCTGAG     420
 A  Q  T  A  M  A  L  D  G  H  S  F  K  T  Q  G  A  L  E  A
GCCCAGACGGCCATGGCGCTTGATGGACACTCCTTTAAGACTCAGGGGGCACTGGAAGCC     480
 F  E  E  I  P  V  D  V  G  D  A  E  A  F  L  P  S  F  P  A
TTTGAGGAGATCCCTGTGGATGTGGGGGATGCTGAGGCCTTTCTGCCTAGCTTCCCAGCA     540
 E  A  W  C  N  G  L  P  Y  P  S  Q  E  H  N  Q  I  L  Q  G
GAGGCTTGGTGCAATGGACTCCCTTACCCCAGCCAGGAACACAACCAAATTCTGCAGGGG     600
 S  E  V  K  V  K  P  Q  A  L  D  N  G  P  G  M  Y  C  Y  Q
TCAGAGGTCAAGGTCAAGCCCCAAGCTTTGGACAATGGTCCTGGGATGTACTGCTACCAG     660
 P  P  L  Q  H  V  Y  C  S  S  Q  P  T  F  H  Q  Y  S  P  G
CCTCCCTTGCAGCATGTGTACTGTTCTTCTCAGCCCACCTTTCATCAGTACTCCCCGGGT     720
 G  G  S  Y  P  V  P  Y  L  G  S  T  H  Y  P  Y  Q  R  I  A
GGAGGCAGCTACCCTGTGCCCTACCTGGGCTCCACTCACTACCCCTATCAGAGGATTGCG     780
 P  Q  A  N  A  D  G  H  Q  P  L  F  P  K  P  I  Y  S  Y  S
CCCCAGGCCAACGCAGACGGTCACCAGCCACTCTTCCCAAAGCCCATCTACTCCTACAGC     840
 I  L  I  F  M  A  L  K  N  X  K  T  G  S  L  P  V  S  E  I
ATCCTCATCTTCATGGCCCTTAAGAACAGCAAGACCGGAAGCCTTCCAGTCAGTGAAATC     900
 Y  N  F  M  T  E  H  F  P  Y  F  K  T  A  P  D  G  W  K  N
TACAATTTCATGACGGAGCACTTCCCTTATTTCAAGACTGCGCCTGATGGCTGGAAGAAT     960
 S  V  R  H  N  L  S  L  N  K  C  F  E  K  V  E  N  K  S  G
TCTGTTCGCCATAACCTGTCTCTCAACAAGTGCTTTGAGAAGGTAGAGAATAAATCCGGA    1020
 S  S  S  R  K  G  C  L  W  A  L  N  P  S  K  I  D  K  M  Q
AGTTCCTCCCGAAAGGGCTGTCTGTGGGCCCTCAATCCTTCCAAAATCGACAAGATGCAG    1080
 E  E  L  Q  K  W  K  R  K  D  P  I  A  V  R  K  S  M  A  K
GAAGAGCTGCAGAAGTGGAAGAGGAAAGACCCCATTGCTGTGCGCAAAAGCATGGCCAAA    1140
 P  E  E  L  D  S  L  I  G  D  K  R  E  K  L  G  S  P  L  L
CCAGAAGAGCTGGACAGCCTCATCGGAGACAAAAGGGAGAAACTGGGTTCTCCTCTGCTA    1200
 G  C  P  P  P  G  L  A  G  P  G  P  I  R  P  L  A  P  S  A
GGATGTCCACCCCCTGGGCTGGCAGGCCCAGGTCCCATCCGGCCCCTGGCACCTTCAGCT    1260
 G  L  T  Q  P  L  H  P  M  H  P  A  P  G  P  M  P  G  K  N
GGTCTTACCCAGCCTCTACACCCAATGCATCCAGCTCCAGGTCCCATGCCTGGCAAGAAC    1320
 P  L  Q  D  L  L  G  G  H  A  P  S  C  Y  G  Q  T  Y  P  H
CCCCTGCAGGACCTACTGGGTGGCCATGCACCCTCCTGCTACGGGCAGACCTACCCACAC    1380
 L  S  P  S  L  A  P  S  G  H  Q  Q  P  L  F  S  Q  P  D  G
CTTTCTCCCAGCCTGGCCCCTTCTGGACACCAGCAGCCATTGTTTTCACAGCCAGATGGG    1440
 H  L  D  L  Q  A  Q  P  G  T  P  Q  D  S  P  L  P  A  H  T
CATCTTGATCTGCAGGCCCAGCCAGGCACCCCCCAGGACTCACCTCTACCTGCCCACACA    1500
 P  P  S  H  G  A  K  L  L  A  E  P  S  S  A  R  T  M  H  D
CCACCCAGCCACGGTGCCAAGCTGCTGGCTGAGCCTTCCTCAGCCAGGACCATGCACGAT    1560
 T  L  L  P  D  G  D  L  G  T  D  L  D  A  I  N  P  S  L  T
ACTCTGCTACCAGACGGAGACCTTGGCACTGACCTGGACGCCATCAACCCCTCTCTCACT    1620
 D  F  D  F  Q  G  N  L  W  E  Q  L  K  D  D  S  L  A  L  D
GACTTCGACTTCCAGGGAAATCTGTGGGAGCAGCTGAAGGATGACAGCTTGGCCCTGGAC    1680
 P  L  V  L  V  T  S  S  P  T  S  S  S  M  L  P  P  P  P  A
CCCCTCGTACTGGTGACCTCATCCCCAACATCATCCTCCATGTTGCCACCCCCACCAGCA    1740
 A  H  C  F  P  P  G  P  C  L  A  E  T  G  N  E  A  G  E  L
GCCCATTGCTTCCCCCCAGGGCCTTGTCTGGCAGAAACAGGCAATGAGGCAGGTGAACTG    1800
 A  P  P  G  S  G  G  S  G  A  L  G  D  M  H  L  S  T  L  Y
GCACCTCCAGGCAGCGGCGGCTCCGGTGCTCTGGGAGACATGCACCTCAGCACTCTCTAC    1860
 S  A  F  V  E  L  E  S  T  P  S  S  A  A  A  G  P  A  V  Y
TCCGCCTTTGTGGAACTGGAGTCCACGCCCTCCTCAGCAGCTGCCGGCCCTGCCGTGTAC    1920
 L  S  P  G  S  K  P  L  A  L  A.
CTCAGTCCCGGCTCAAAGCCATTGGCTCTGGCTTGAGCTATGCCCAACATCAGCTGCGGC    1980
CTTGCCTAGCTGGCTGCCCATATAGGGCTCACCTTAAAGGTCAAAGAAGGAAAACACTAC    2040
TTGTCTCCTATGTCACTCAGCCAACTTATTTGTTAGCCAGAAGCTAGGGGATCCACCTAG    2100
GATGCTACCGGGTGACGCGCTCCACACGCGGtGCCCCCAGCAAGGAAAGTGTCGGAAAAG    2160
AAGCAACAGCGCGCCCCCTAGCGCCAGCTCACTCGAACTTCAGCTCTCGAGCGTGAACTC    2220
AAGCTCTACACACCTGCTCATCATGTGCCTTACACTCAGAGGAAGCCTTGGGTCACAGAG    2280
TCTGATTTGATTTCTGGGGCAGCTGAGAGCTAAAAGCTTTAGTTAGCAAAGCTCAGGGCC    2340
AGTGTGACAGGTCAAAGATCACCCCTCCAAACCTCATCCCAATCCCCATGTGCTCTACAG    2400
ACTTGGCCCTCTCCC                                                  2415
```

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 C to T at nucleotide 1429 resulting in the nonsense mutation of the *rnu* 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^N 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 GM18684 (to B.A.T.).

REFERENCES

Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C., and Venter, J. C. (1993). Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nature Genet.* **4:** 373–380.

Berridge, M., O'Kech, N., McNeilage, L., Heslop, B., and Moore, R. (1979). Rat mutant showing "nude" characteristics. *Transplantation* **27:** 419.

Boll, W., Schmid, C. T., Semenza, G., and Mantei, N. (1993). Messenger RNAs expressed in intestine of adult but not baby rabbits. Isolation of cognate cDNAs and characterization of a novel brush border protein with esterase and phospholipase activity. *J. Biol. Chem.* **268:** 12901–12911.

Buchberg, A. M., and Camper, S. A. (1993). Encyclopedia of the mouse genome III. October 1993. Mouse chromosome 11. *Mamm. Genome* S164–S175.

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

Choi, K. W., and Benzer, S. (1994). Rotation of photoreceptor clusters in the developing Drosophila eye requires the nemo gene. *Cell* **78:** 125–136.

Chu, G., Vollrath, D., and Davis, R. W. (1986). Separation of large

DNA molecules by contour-clamped homogeneous electric fields. *Science* **234:** 1582–1585.

Church, D. M., Stotler, C. J., Rutter, J. L., Murrell, J. R., Trofatter, J. A., and Buckler, A. J. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6:** 98–105.

Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., Steen, R. G., Stein, L. D., Nadeau, J. H., and Lander, E. S. (1993). A genetic linkage map of the mouse: Current applications and future prospects. *Science* **262:** 57–66.

Cordier, A. C., and Haumont, S. M. (1980). Development of thymus, parathyroids, and ultimo-branchial bodies in NMRI and nude mice. *Am. J. Anat.* **157:** 227–263.

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H. S., Friedman, J., Dracopoli, N. C., and Lander, E. S. (1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131:** 423–447.

Dietrich, W. F., Miller, J. C., Steen, R. G., Merchant, M., Damron, D., Nahf, R., Gross, A., Joyce, D. C., Wessel, M., Dredge, R. D., Marquis, A., Stein, L. D., Goodman, N., Page, D. C., and Lander, E. S. (1994). A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genet.* **7:** 220–245.

Eicher, E. (1976). Remutations. *Mouse News Lett.* **54:** 90.

Feinberg, A., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–12.

Festing, M., May, D., Connors, T., Lovell, D., and Sparrow, S. (1978). An athymic nude mutation in the rat. *Nature* **274:** 365.

Flanagan, S. (1966). "Nude," a new hairless gene with pleiotropic effects in the mouse. *Genet. Res.* **8:** 295–309.

Green, M. C. (1989). Catalog of mutant genes and polymorphic loci. *In* "Genetic Variants and Strains of the Laboratory Mouse," 2nd ed. (M. F. Lyon and A. G. Searle, Eds.), 12–403, Oxford Univ. Press, New York.

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

Hammer, M. F., Schimenti, J., and Silver, L. M. (1989). Evolution of mouse chromosome 17 and the origin of inversions associated with t haplotypes. *Proc. Natl. Acad. Sci. USA* **86:** 3261–3265.

Jenkins, N. A., Rothe, H., Gilbert, D. J., Copeland, N. G., and Kolb, H. (1994). Mapping of the gene for inducible nitric oxide (NO) synthase of mouse macrophages to chromosome 11, close to Evi-2, nu, and Idd-4. *Genomics* **19:** 402–404.

Kopf-Maier, P., Mboneko, V. F., and Merker, H. J. (1990). Nude mice are not hairless: A morphological study. *Acta Anat.* **139:** 178–190.

Kusumi, K., Smith, J. S., Segre, J. A., Koos, D. S., and Lander, E. S. (1993). Construction of a large-insert yeast artificial chromosome library of the mouse genome. *Mamm. Genome* **4:** 391–392.

Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M. A., Jaenisch, R., and Berns, A. (1991). Simplified mammalian DNA isolation procedure. *Nucleic Acids Res.* **19:** 4293.

Lander, E. S., and Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics* **2:** 231–239.

Li, C., Lusis, A. J., Sparkes, R., Tran, S. M., and Gaynor, R. (1992). Characterization and chromosomal mapping of the gene encoding the cellular DNA binding protein HTLF. *Genomics* **13:** 658–664.

Lisitsyn, N. A., Segre, J. A., Kusumi, K., Lisitsyn, N. M., Nadeau, J. H., Frankel, W. N., Wigler, M. H., and Lander, E. S. (1994). Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. *Nature Genet.* **6:** 57–63.

Lovett, M. (1994). Direct selection of cDNAs using genomic contigs. *In* "Current Protocols in Human Genetics" (N. C. Dracopoli, J. L. Haines, B. R. Korf, D. T. Moir, C. C. Morton, C. E. Seidman, J. G. Seidman, and D. R. Smith, Eds.), Current Protocols, New York.

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

Markovich, D., Forgo, J., Stange, G., Biber, J., and Murer, H. (1993). Expression cloning of rat renal Na+/SO4(2−) cotransport. *Proc. Natl. Acad. Sci. USA* **90:** 8073–8077.

Morimoto, B. H., Chuang, C. C., and Koshland, D. J. (1991). Molecular cloning of a member of a new class of low-molecular-weight GTP-binding proteins. *Genes Dev.* **5:** 2386–2391.

Murphy, M., Pykett, M. J., Harnish, P., Zang, K. D., and George, D. L. (1993). Identification and characterization of genes differentially expressed in meningiomas. *Cell Growth Differ.* **4:** 715–722.

Nehls, M., Pfeifer, D., Schorpp, M., Hedrich, H., and Boehm, T. (1994). New member of the winged-helix protein family disrupted in mouse and rat nude mutations. *Nature* **372:** 103–107.

Nishiwaki, K., Sano, T., and Miwa, J. (1993). emb-5, a gene required for the correct timing of gut precursor cell division during gastrulation in *Caenorhabditis elegans,* encodes a protein similar to the yeast nuclear protein SPT6. *Mol. Gen. Genet.* **239:** 313–322.

Pantelouris, E. (1968). Absence of thymus in a mouse mutant. *Nature* **217:** 370–371.

Pantelouris, E. M. (1973). Athymic development in the mouse. *Differentiation* **1:** 437–450.

Paolella, G., Buono, P., Mancini, F. P., Izzo, P., and Salvatore, F. (1986). Structure and expression of mouse aldolase genes: Brain-specific aldolase C amino acid sequence is closely related to aldolase A. *Eur. J. Biochem.* **156:** 229–235.

Reed, C., and O'Donoghue, J. (1979). A new guinea pig mutant with abnormal hair production and immunodeficiency. *Lab. Animal Sci.* **29:** 744.

Seiffert, D., Poenninger, J., and Binder, B. R. (1993). Organization of the gene encoding mouse vitronectin. *Gene* **134:** 303–304.

Treco, D. A. (1991). Preparation of yeast DNA. *In* "Current Protocols in Molecular Biology" (F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, Eds.), Wiley, New York.

Wolf, F. W., Marks, R. M., Sarma, V., Byers, M. G., Katz, R. W., Shows, T. B., and Dixit, V. M. (1992). Characterization of a novel tumor necrosis factor-alpha-induced endothelial primary response gene. *J. Biol. Chem.* **267:** 1317–1326.

Wortis, H. H., Nehlsen, S., and Owen, J. J. (1971). Abnormal development of the thymus in nude mice. *J. Exp. Med.* **134:** 681.

Zha, H., Remmers, E. F., Du, Y., Cash, J. M., Goldmuntz, E. A., Crofford, L. J., and Wilder, R. L. (1995). The rat athymic nude (*rnu*) locus is closely linked to the inducible nitric oxide synthase gene (*Nos2*). *Mamm. Genome* **6:** 137–138.

4796.80