# Positional Cloning and Characterization of the Mouse *pudgy* Locus

by
Kenro Kusumi
A.B. Biochemical Sciences, Harvard College

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
February 1997



Signature of Author_____

Department of Biology
February 5, 1997

Certified by____

Eric S. Lander, D.Phil.
Professor, Department of Biology
Thesis Supervisor

Accepted by____

Richard A. Young, Ph.D.
Chairman, Biology Graduate Committee

## TABLE OF CONTENTS

**Positional Cloning and Characterization of the Mouse *pudgy* Locus**

by

KENRO KUSUMI

Submitted to the Department of Biology

on February 5, 1997

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biology

## ABSTRACT

Over 2,000 mutations have been cataloged in the laboratory mouse, but only a few clearly affect early somitic patterning during development. The *pudgy* mutation is one of these few, with *pudgy* mutants exhibiting striking malformations in the vertebrae and ribs due to anomalous somitic segmentation. To understand the genetic basis of this defect, we have undertaken to positionally clone the *pudgy* gene.

As a first step, we characterized the mutant phenotype. We conducted histological and *in situ* hybridization analyses using mesodermal markers and identified irregular somitic segmentation in *pudgy* mutants. In addition, we observed a significant early embryonic lethality in two genetic backgrounds, and lethality prior to embryonic day 9.5.

For positional cloning, we narrowed the region, where the *pudgy* gene must be located, through genetic and physical mapping. We produced over 2,000 meioses in three F2 intercrosses and refined the genetic mapping of the *pudgy* locus to within a 1.8 cM region on mouse chromosome 7. Towards physical mapping of the *pudgy* region, we constructed the first publicly distributed large-insert mouse yeast artificial chromosome (YAC) library. Using this library, we completed YAC-based walks across the *pudgy* interval and have subsequently covered the region in P1 and bacterial artificial chromosome (BAC) clones. Based on the physical maps, we estimate that the *pudgy* gene must be located within a ~600kb region.

To identify genes within the *pudgy* interval, we used direct cDNA selection and identified 113 cDNA fragments. These fragments were localized on the *pudgy* region physical map. For finer resolution, we sequenced ~250kb of the *pudgy* genomic interval and integrated sequences from direct cDNA selection and cDNA library screens with genomic sequences to create a base-pair resolution map.

We evaluated transcripts as candidates for *pudgy* using *in situ* and Northern hybridization, and reverse-transcriptase based sequencing. Through these screens, we identified a ~11kb, novel serine-threonine kinase transcript, expressed in ventral somitic tissues, as a prime candidate for the *pudgy* gene. To test this candidate, we constructed several transgenic lines with BACs and P1s from the region, and are crossing the transgene into a *pudgy* genetic background. Currently, we are cloning and sequencing the complete serine-threonine kinase transcript for evidence of the *pudgy* mutation.

Thesis Supervisor: Eric S. Lander, D.Phil.

Title: Professor of Biology

## PREFACE

### An Overview of Positional Cloning at the *Fin-de-Siècle*

Understanding the function of all 100,000 genes in the mammalian genome remains a holy grail for the community of biologists at the end of the twentieth century. Within a hundred years since the rediscovery of Mendelian genetics, and within 50 years of the identification of DNA as the genetic encoding material, the research community has undertaken the effort to correlate genes with biological function. Gradually, the list of known genes has been expanded up to 10% of the total, with homology-based cloning from other model organisms producing much of this number. Through study of gene expression and through production of targeted mutations and transgenic alleles, the correlation of known genes with their functions has been expanded.

However, sometimes it has been necessary to find the gene for a particular mutation, either because of efforts to solve human diseases or to answer key biological questions. In those instances, alternative methods have been required. One of the key tools toward these efforts has been positional cloning, which requires only a clear mutant phenotype in order to identify the gene. Although the numerical tally of genes cloned by positional cloning is unimpressive, the scientific importance of these genes, which often are novel transcripts unrelated to currently known genes, is great. For example, the following biologically key genes have been identified through positional cloning efforts.

- In the human: CFTR - a chloride transporter involved in the common genetic disease, cystic fibrosis, BRCA1 - a zinc finger protein defective in hereditary breast cancer syndromes, APC - the gene associated with defects in colon cancer syndromes
- In the mouse: *Sry*- the testis determining factor, *brachyury* - a novel transcription factor important in mesoderm induction, *reelin* - an extracellular protein involved in cerebral cortical lamination, and *leptin* - a secreted hormone regulating satiety homeostatis

The starting point in positional cloning efforts is the mutant phenotype. Although the word phenotype is often used synonymously with clinical concepts such as syndrome or disease which encompass a constellation of traits, the investigator should try to limit its definition in as focused a manner

as possible. A mutation may give rise to several phenotypes, with differing modes of inheritance (dominant, recessive). For example, in the mouse *weaver* mutation, there is a seizure phenotype which acts in a dominant fashion, but a cerebellar defect which is inherited in a recessive manner. Regarding the reliability of the physical characteristic selected, the investigator must be sure at what probability the phenotype will appear given the expected genotype (penetrance) and what occurence the mutant phenotype will appear without the mutant genotype (phenocopy). Both of these factors will greatly affect the reliability of any animal phenotype, and uncertain cases will require another generation of matings for confirmation. Sometimes, phenotypes will be observed in a number of different organ systems (pleiotropism) which may aid in the determination of mutant phenotype. For example, in the mouse *nude* mutation, there are phenotypes of hairlessness and athymia. Other factors affecting phenotype are modifier loci (suppressors and enhancers) which may be segregating in the strains used in the genetic crosses, and study of these loci themselves may be highly useful figuring out the genetic pathways involved in the biological function.

Even prior to positional cloning efforts, characterization of the mutant phenotype may yield many clues as to the type of gene involved. Biological issues such as the time at which the mutant phenotype is first observed and the tissues involved will provide essential information for the evaluation of gene candidates. Chimera or transplantation experiments may resolve whether the mutation acts cell-autonomously. Non-cell autonomous phenotypes might suggest extracellular secreted products or membrane associated genes. The more phenotypic information available early in positional cloning efforts, the easier it will be to focus future experimental efforts to the correct tissue and time.

The *modus operandi* of positional cloning is to continually narrow the region being evaluated, first by genetic means and then using the tools of genomics. Starting with a novel trait, the mutation may be completely unmapped, making any of the 100,000 or so genes in the 3 billion base-pair mammalian genome possible candidates. Ideally, we would like to narrow the list of candidates down to one or two genes. Genetic approaches to narrowing the region make use of DNA recombination, whether the result of normal meiosis or through genomic changes (natural or experimentally induced). For meiotic recombination mapping, the key insight is that of linkage, i.e. two loci

that are physically proximate will be less likely to be separated by meiotic recombination than two more distant loci. Similarly for genomic rearrangments, two close loci are more likely to be both affected than two more distant ones, e.g. a deletion which results in loss of several loci.

Once a genetic interval has been achieved, the region must be physically represented in order to isolate genes in the region. This process is evolving rapidly, and the current approaches will be described. With reasonable genetic mapping efforts, it should be possible to narrow the genetic interval down to a 0.1 cM or 200 kb region. In practice, recombination breakpoints are seldom spaced randomly, raising the likely physical interval to 0.5 to 1 Mbp. In order to cover these large regions, DNA clones with inserts of comparable size are required. The only clones currently with such large inserts are Yeast Artificial Chromosomes (YACs). While adequate for covering these large regions, YACs are known for their instability of inserts, with deletions and rearrangements common and a high rate of insertional chimerism. In conjunction with YAC cloning efforts, linked markers can be separated using resources such as radiation hybrids (cell lines of one organism with radiation induced fragments of another organism carried within) or Fluorescence In Situ Hybridization (FISH). Metaphase FISH can localize markers to a chromosomal region, and interphase FISH can be used to physically order markers and determine approximate physical distances.

With a large-insert DNA scaffold in place, a second physical map must be made of more stable clones, using either bacterial host vectors systems such as P1s or Bacterial Artificial Chromosomes (BACs), which have smaller insert sizes ranging from 50 kb to 200 kb. In addition to sequence fidelity, the smaller clones are more reasonably sized (containing on average 1-2 genes) for efforts such as cDNA selection and transgenesis. Increasingly, a finer level of physical map is available for large sections of the human genome--genomic sequence. Genomic sequence can serve as the template for gene identification directly through alignment-based methods (e.g. BLAST) with the expressed sequence databases, or through methods which search for gene motifs (e.g. GRAIL).

Even with the physical template in hand, the process of identifying genes within the region is not trivial. The root cause of this problem is the challenge of defining what traits characterize a mammalian gene. Methods which use stereotypical motifs could search for the following: CpG rich sequences in gene

regulatory regions, promoter sequences, open reading frames of translated genes, splice acceptor and splice donor sites, polyA signal sites, and homology to known encoded sequences in DNA databases. The difficulty with analyzing the mammalian genes has to do with the extent to which transcripts are spread out amongst the genome. Indeed, if average cDNA size is 2 kb and average region per gene is 30 kb, we can see that less than 10% of genomic sequence codes for genes, and exons can be spread out over a large "junk" DNA region. Combined with the weak consensus sequences for the above motifs, it is difficult to find genes with certainty without resorting to some homology searches. Experimental techniques which look for the same motifs, such as exon trapping, are subject to the same difficulties. Many experimental approaches have relied on obtaining cDNA clones and identifying those which are located in a region of interest, through methods such as direct cDNA selection. This information will soon become available genome-wide, when EST sequencing information is integrated with genomic sequences.

One of the greatest challenges in positional cloning lies in the ability to evaluate gene candidates within a positionally cloned region. Hypothetically, given vast labor and resources, every candidate could be carefully cloned and characterized in its own right. In large regions with up to 50 genes, this type of analysis is prohibitive to almost all efforts. Clues *have* to be sought from numerous sources. First of all, expression-based clues can be helpful in mutations where the nature of the defect leads to clear ideas of the biological processes affected. Expression can be assayed by *in situ* hybridization for spatial patterns and by reverse-transcriptase polymerase chain reaction for tissue distribution. Northern blot analysis can give clues as to whether the expression pattern of candidate genes is disrupted.

If one candidate is of particular interest, it may be worthwhile to complement the phenotype or to induce a mutation in the candidate gene. As in *Drosophila* where balancer chromosome can be used to complement the phenotype of loss-of-function alleles, large-insert DNA clones can be transgenically introduced and crossed into mutant backgrounds. If gene dosage has an effect on the phenotype, such transgenic lines can also be used to study gene function. Targeted disruption of candidate genes can also be attempted using homologous recombination; however, all of these manipulation require a fair amount of time for embryo manipulation and mouse breeding.

Ultimately, identification of sequence polymorphism associated with the mutation is required. Towards that end, methods which are able to examine DNA changes such as Southern blot for large-scale changes or sequencing for localized polymorphisms can be used. However, careful sequencing will be involved at the end of any positional cloning effort.

What constitutes proof that the gene corresponding to the mutation has been identified? The gold-standard, of course, is identification of a sequence mutation that appears when comparing both the mutant chromosome and the chromosome on which it arose. With many mouse mutations, the mutagenic event has occurred within an inbred strain. Given the extreme homogeneity of mouse inbred strains, high confidence can be taken that the parental chromosome is in hand. Lacking this gold-standard, a suite of secondary evidence is essential in building the argument that the gene is associated with a mutation, including:

- the ability of a genomic clone containing the candidate gene, or a cDNA construct of the candidate gene, to complement the phenotype by transgenesis,
- the identification of sequence defects from many mutant alleles, with no sequence polymorphisms found among the general population or inbred strains,
- the expression of the candidate gene in tissues and time-points that are required to explain the mutant phenotype, with disrupted pattern and/or level of expression or altered transcript size,
- the targeted disruption of the gene produces defects in organ systems involved in the mutant phenotype. Gene knockout alleles may often have more severe phenotypes than spontaneous generated alleles.

Our ability to correlate genes and their function will be greatly enhanced in the future, owing to the imminent completion of the sequencing goals of the Human Genome Project. Within the decade, complete genomic and numerous cDNA sequences will be available to researchers world-wide. What will be possible when this data is available to anyone with a scientific workstation and internet access? We can speculate that this information will expedite many of the steps required to find candidate genes within a region. However, the process of evaluating candidate genes will not be as affected, and this step may become the bottleneck towards reaching the holy grail of understanding the function of all genes.

# KENRO KUSUMI

BIOGRAPHICAL INFORMATION

Date of Birth: 14 February 1967
Place of Birth: Mito, Ibaraki-Prefecture, Japan
Citizenship: U.S.A., naturalized 1985

## Education

Ph.D. candidate, 1990-1997
Department of Biology
Massachusetts Institute of Technology
Cambridge, MA 02139

M.D. candidate, 1988-1997
Division of Health Sciences & Technology
Harvard Medical School/Massachusetts Institute of Technology
Boston, MA 02115

A.B., 1988 *magna cum laude* in Biochemical Sciences
Harvard College
Cambridge, MA 02138

## Training and Experience

*Doctoral Research, Department of Biology, M.I.T. 1990-present*
Advisor: Eric S. Lander, D.Phil., Whitehead Institute for Biomedical Research, Cambridge, MA 02139
Thesis topic: Positional Cloning of the Mouse *pudgy* Locus

*Research Assistant, Division of Health Sciences & Technology, Harvard Medical School/M.I.T., 1988-1990*
Advisor: James I. Mullins†, PhD, Department of Cancer Biology, Harvard School of Public Health, Boston, MA 02115
† present address: Department of Microbiology, University of Washington, Seattle, WA
Research topic: Molecular cloning and characterization of HIV-1 directly from patient tissue by polymerase chain reaction (PCR).

*Undergraduate Honors Research, 1986-1988*
Advisors: Dan T. Stinchcomb, PhD* and Dan Kiehart, PhD†
Department of Cellular & Developmental Biology, Harvard University
*present address: Synergen, Inc. Boulder, CO
† present address: Duke University Medical Center, Durham, NC
Topic: Towards biochemical isolation of a cytoplasmic myosin in *Caenorhabditis elegans* associated with P-granules

## Publications

1. Girard, M., Kieny, M.-P., Pinter, A., Barre-Sinoussi, F., Nara, P., Kolbe, H., Kusumi, K., Chaput, A., Reinhart, T., Muchmore, E., Ronco, J., Kaczorek, M., Gomard, E., Gluckman, J.-C., and Fultz, P.N. (1991) Immunization of chimpanzees confers protection against challenge with human immunodeficiency virus. *Proc. Natl. Acad. Sci. USA.* **88**: 542-546.

2. Jacob, H.J., Lindpaintner, K., Lincoln, S.E., Kusumi, K., Bunker, R.K., Mao, Y.-P., Ganten, D., Dzau, V.J., and Lander, E.S. (1991) Genetic mapping of a major gene causing hypertension in the stroke prone spontaneously hypertensive rat. *Cell.* **67**: 213-224.

3. Kusumi, K., Conway, B., Cunningham, S., Bersen, A., Iversen, A.K.N., Colvin, D., Gallo, M.V., Coutre, S., Shpaer, E.G., Faulkner, D.V., de Ronde, A., Volkman, S., Williams, C., Hirsch, M.S., and Mullins, J.I. (1992) HIV-1 envelope gene structure and diversity *in vivo* and following short-term co-cultivation *in vitro*. *J. Virol.* **66:** 875-885.

4. Kusumi, K., Smith, J.S., Segre, J.A., Koos, D.S., and Lander, E.S. (1993) Construction of a yeast artificial chromosome (YAC) library of the mouse genome. *Mammalian Genome.* **4**: 391-392.

5. Lisitsyn, N.A., Segrè, 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 Genetics.* **6**: 57-63.

6. Hästbacka, J., de la Chapelle, A., Mahtani, M.M., Clines, G., Reeve-Daly, M.P., Daly, M., Hamilton, B.A., Kusumi, K., Trivedi, B., Weaver, A., Coloma, A., Lovett, M., Buckler, A., Kaitila, I., and Lander, E.S. (1994) The Diastrophic Dysplasia gene encodes a novel sulfate transporter: positional cloning by fine structure linkage disequilibrium mapping. *Cell.* **78**: 1073-1087.

7. Hamilton, B.A., Frankel, W.F., Kerrebrock, A.W., Hawkins, T.L., FitzHugh, W., Kusumi, K., Russell, L.B., Mueller, K.L., van Berkel, V., Birren, B.W., Kruglyak, L., & Lander, E.S. (1996) Disruption of the nuclear hormone receptor RORα in *staggerer* mice. *Nature.* **379**: 736-739.

## Teaching Assistant Experience

*7.16, Mammalian Project Lab, Department of Biology, M.I.T. 1994*
Instructors: David E. Housman and Richard Mulligan

*7.69, Animal Virology (Graduate Level), Department of Biology, M.I.T. 1991*
Instructors: Phillip A. Sharp, Nancy H. Hopkins, Robert A. Weinberg

**Abbreviations Used in Text**

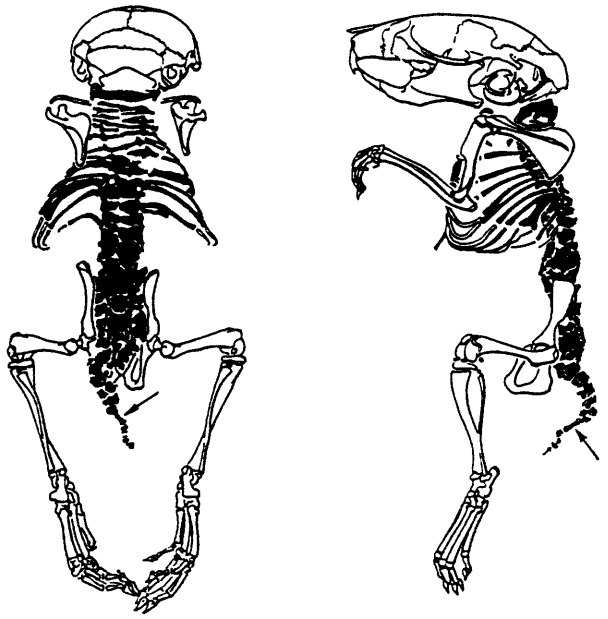
| *Abbreviation* | *Description* |
|---|---|
| BAC | Bacterial Artificial Chromosome |
| BLAST | Basic Local Alignment Search Tool |
| | -X nucleotide query sequence vs. protein sequence database |
| | -N nucleotide query sequence vs. DNA sequence database |
| dbEST | National Center for Biotechnology Information |
| | Non-redundant Database of GenBank+ |
| | European Molecular Biology Laboratory (EMBL)+ |
| | (DNA DataBank of Japan) DDBJ EST Divisions |
| E# | embryonic day # |
| EST | Expressed Sequence Tag, a sequence fragment from |
| | an expressed sequence |
| GenBank | NIH genetic sequence database |
| P1 | P1 phage-based large-insert cloning system |
| *pu* | the *pudgy* allele |
| RT | reverse transcriptase |
| SSCP | Single Stranded Conformational Polymorphism |
| SSLP | Simple Sequence Length Polymorphism |
| STS | Sequence Tagged Site, a sequence-based marker |
| YAC | Yeast Artificial Chromosome |

## Chapter One

# The Genetic Analysis of Somitogenesis and Approaches to the Positional Cloning of the *pudgy* Mutation

"Indeed, it is astonishing that an animal can live in such a frame at all. Considering the pain and discomfort which a single cervical rib, by pressure on the brachial plexus, can cause in man, I personally have no doubt that *pudgy* mice are enduring a martyrdom of pain through their lives. But as mice do not show suffering in any obvious way, this cannot at present be proved to the satisfaction of a behaviourist, let alone a philosopher."

from *Genetic Studies on the Skeleton of the mouse: XXIX pudgy*, Hans Grüneberg, 1961



Text-fig. 1. Pudgy ♀, 23 days old. Drawings based on photographs. A single virtually normal caudal vertebra is indicated by arrows.

The *pudgy* mutation was reported by Hans Grüneberg in 1961 and is one of the most severe skeletal defects described in the mouse (Grüneberg, 1961). In this introduction, we will discuss the following experimental background and approaches that underlie our efforts to positionally clone the *pudgy* gene.

I. Overview of Somite Formation

II. Genetic Analysis of Somitogenesis and Skeletogenesis

III. Approaches to the Positional Cloning of Mouse Mutations

IV. Overview: Positional Cloning of the *pudgy* Locus

## *I. Overview of Somite Formation*

### *A. Somite Development*

One of the most striking features of a vertebrate embryo is the regularly patterned array of somites that appear and disappear during development. Somites are transient collections of mesodermally derived tissue that form bilateral structures around the axis of the neural tube and notochord, as shown in Figure 1.

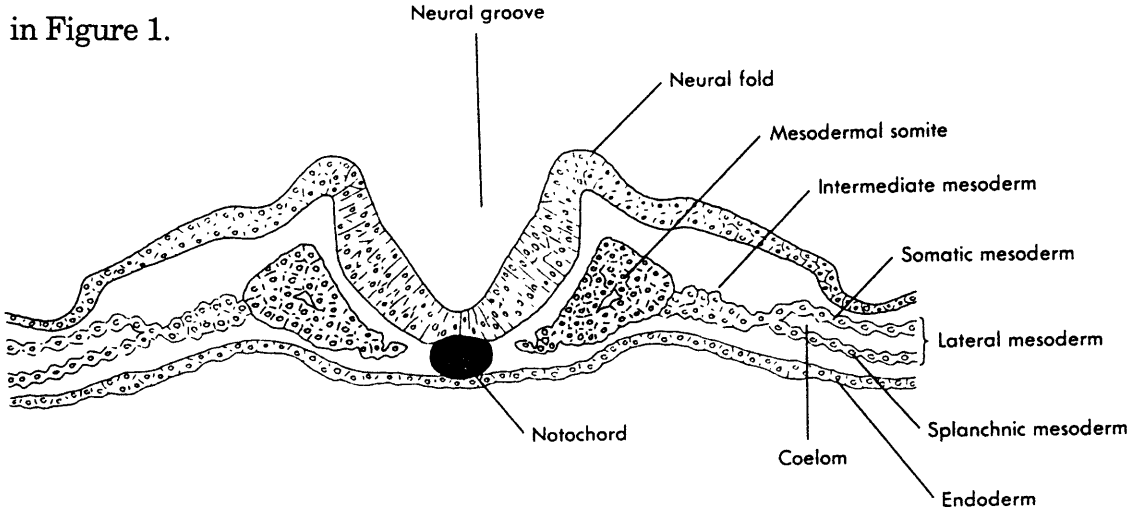



***Figure 1*** *Representation of a transverse vertebrate embryo section. Reproduced from Fig. 5.10, Kent, 1987.*

In the mouse, the most anterior somites coalesce beginning embryonic day 8 and the most posterior somites finally coalesce ending embryonic day 13 (Tam and Tan, 1992). The segmentation is associated with a drastic cellular cytoskeletal rearrangement from mesenchymal to columnar epithelioid cells. When first formed, a somite consists of a single-layer epithelial sphere, bound together through tight junctions at the basal surface, with a lumen filled with mesenchymal mesoderm (Keynes and Stern, 1988). The somite, like the neighboring neural tube, secretes a basal lamina containing extracellular matrix glycoproteins such as fibronectin, laminin, and collagen, and glycosaminoglycans. In the apical zone, there are increased amounts of α-

actinin and actin microfilaments (Ostrovsky et al., 1983). The morphological changes in somite segmentation are clearly associated with cytoskeletal changes in the somitic cells.

The formation of the somites follows an anterior to posterior progression in embryonic development, synchronous with the closure of the neural tube. Although somitogenesis occurs in conjunction with neurogenesis, somitogenesis can occur independently of the formation of the neural tube, as demonstrated by transplantation experiments. When sections of segmental plate mesoderm are reversed in anterior-posterior direction, the process of somitogenesis proceeds in a manner ascribed to its former orientation, not that of its new orientation (Keynes and Stern, 1988). Indeed, somitogenesis can sometimes even occur in the absence of notochord, neuroectoderm, or endoderm, suggesting that there is prepatterning.

The unsegmented mesoderm immediately posterior to the wave of coalescent somites is referred to as either the presomitic mesoderm or the segmental plate. Loosely circular discs of mesodermal cells, termed somitomeres, have been identified in the segmental plate. The size and location of these somitomeres roughly correspond to that of future somites (Jacobson, 1988). Somitomeres have been described in numerous vertebrates, ranging from mammals, birds, reptiles, amphibians, and teleosts. Whether there is a one-to-one correspondence between somitomere and somite has not yet been determined, but lineage studies indicate that cell mixing may occur between somitomere and somite stages (Hogan et al., 1994).

During somite maturation, the ventral sections of the somites dissociate, producing mesenchymal cells which migrate to form the sclerotome, as shown in Figure 2. This disaggregation is associated with a loss in N-cadherin expression (Duband et al., 1987). The sclerotome forms the ossified bones of

the vertebrae and the ribs, but not of the pelvic girdles, limbs, or cranium. The remaining epithelial somitic cells develop into the dermomytome, and later differentiate into dermotome, which forms the dermis of the back, and the myotome, which forms the axial skeletal musculature. Myotomal descendants also include muscles in the tongue, extrinsic muscles of the eyes, and possibly muscles in the limbs.

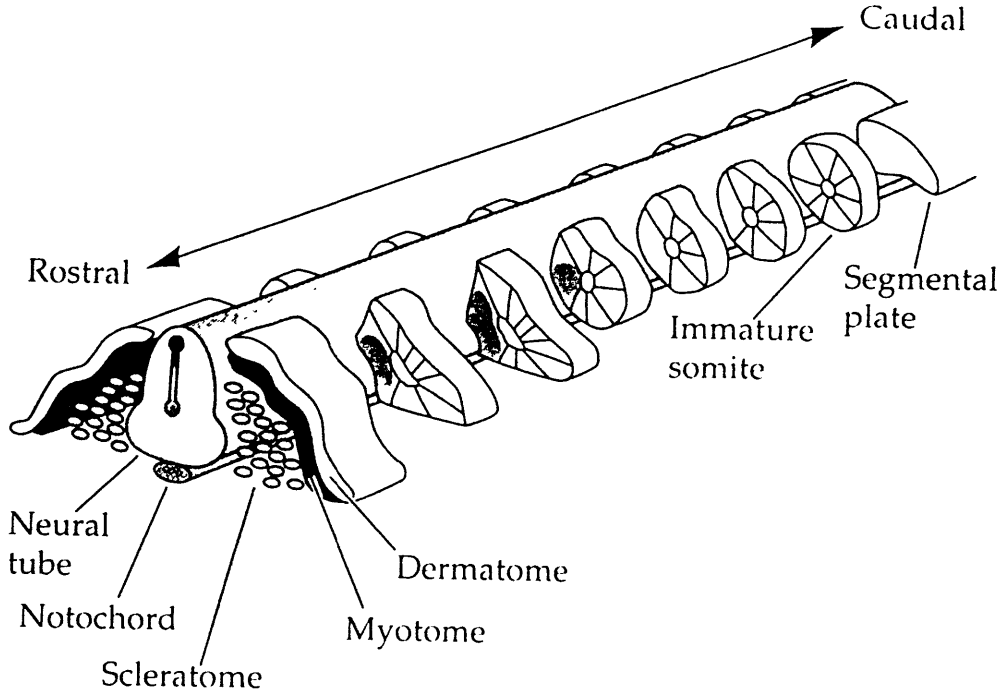



***Figure 2*** *Schematic representation of somite maturation, progressing in a rostral-caudal direction during development. Reproduced from* (Gilbert, 1994).

The process of somitogenesis does not occur isolated from the rest of the organismal development. Tightly associated with the formation of somites is the process of spinal nerve outgrowth from the neural tube, and later the sclerotomal cells will migrate and surround the notochord and neural tube to form the spinal column. On the basis of embryological experiments, somites appear to have anterior-posterior differences (Hogan et al., 1994). Spinal motor neurons migrate specifically through the anterior portion of the somite, and

transplantation experiments reversing the normal polarity of the somite results in neurons migrating through the formerly anterior regions (Serbedzija et al., 1990). Similar results are seen with the migration of neural crest cells through the anterior somite to form the autonomic sympathetic ganglia.

### *B. Adult Vertebral Anatomy*

Lineage studies have also shown that the adult vertebral and rib structures consist of the anterior section of one somite and the adjacent posterior segment of its neighbor (Hogan et al., 1994). Why is there this unusual arrangement? Embryologists have hypothesized that the myotome and sclerotome are offset by a half somite such that axial muscle groups can form with insertion points spanning two adjacent vertebrae, as would be required to produce columnar flexion, as shown in Figure 3.

The vertebral column consists of four major components: the spinal cord and nerves (derived from ectoderm), the ossified vertebral and costal bones (derived from sclerotomal mesoderm), the spinal muscles and ligaments (derived from myotomal mesoderm), and the intervertebral discs (derived from notochordal mesoderm) (Kent, 1987).

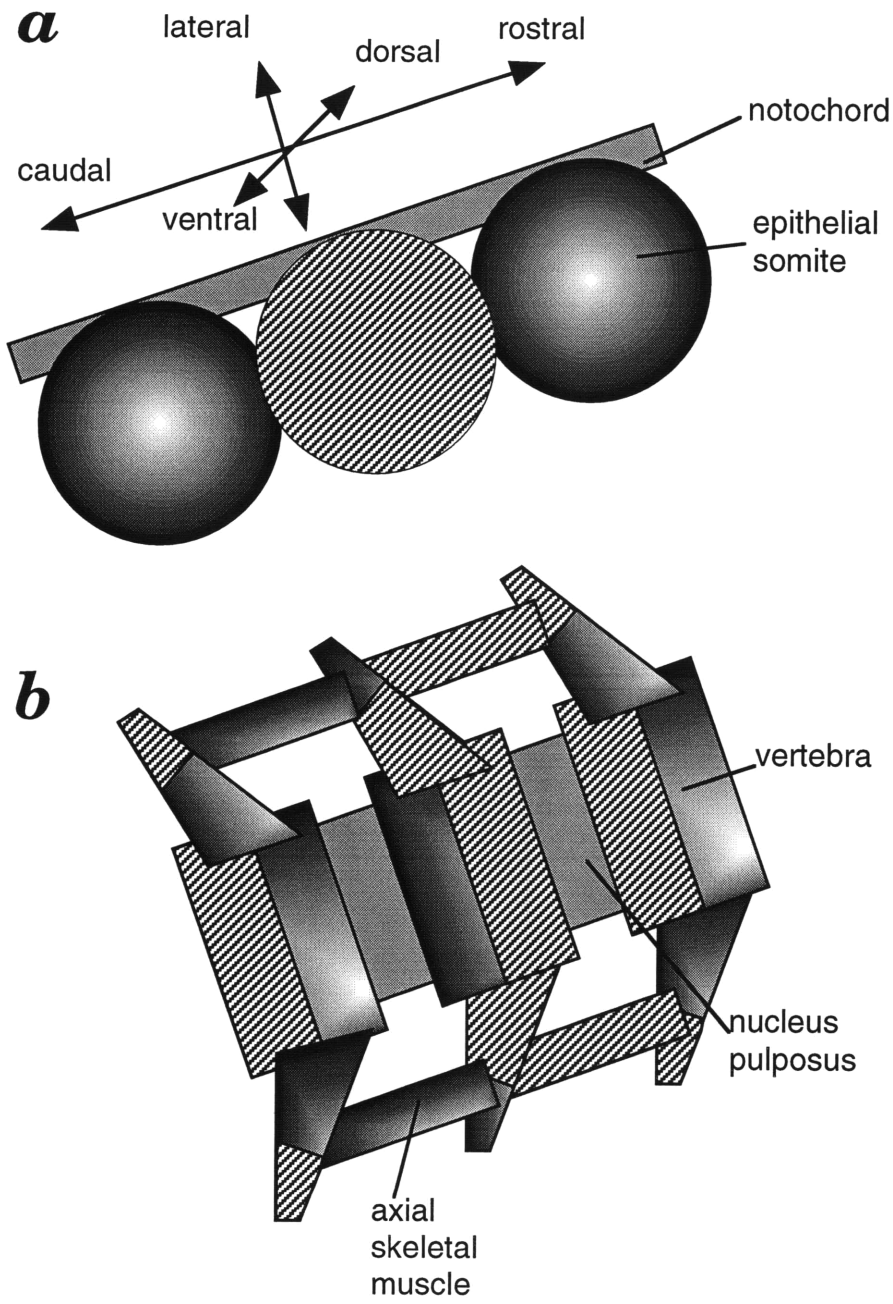



**Figure 3** Somitic cell lineages in the formation of the vertebral column.
*a* A representation showing epithelial somites and notochord.
*b* A representation of the adult vertebral column, with nucleus pulposus derived from the notochord, vertebral bodies and spinous processes derived from sclerotome, and axial musculature derived from myotome. Hatch marks represent cell lineages and the distribution of their descendants in the adult column.

The process of somitogenesis and skeletal formation is not invariant among vertebrates. The number of vertebrae in tetrapods varies widely from 9 in frogs to over 400 in some snakes (Kent, 1987). Furthermore, ribs, which develop from the same precursors as vertebrae, also range in number in from 9 to 24 pairs. Associated with these skeletal variations is regulation of the number of somites.

## *II. Genetic Analysis of Somitogenesis and Skeletogenesis*

### *A. Overview*

What are the cellular processes that underlie somitogenesis? We are at the beginning of understanding this complex process. Clues from aggregation studies have shown that separated somitic cells are more self-adherent than cells in the presomitic mesoderm, pointing to changes in cell adhesions as important in somitogenesis. Morphologically, mesodermal cells undergo numerous obvious changes from mesenchymal to epithelial cell. During the formation of sclerotome, those precursor cells transform yet again to a mesenchymal state. Furthermore, many changes in gene expression occur within mesodermal cells during somitogenesis that we can observe using techniques such as *in situ* hybridization.

For the last 50 years, the process of somitogenesis has been examined in a number of vertebrate model systems--chick, *Xenopus*, mouse, and recently zebrafish--using the approaches of both embryology and genetics. The process of somitogenesis involves patterning, migration, and differentiation across several axes. Somitic patterning in the mouse progresses along the embryonic anterior-posterior axis from E8 to E12, with divisions developing in a proximal-lateral axis from lateral mesoderm as well. Within each somite,

heterogeneity develops between the anterior and posterior sections of each somite, and dorsal-ventral asymmetries arise due to the interaction with other structures such as the notochord and neural tube.

Where does the hunt for genes involved in these developmental processes begin? The majority of currently known genes involved in somitogenesis originate from screens for vertebrate homologs of *Drosophila* patterning genes. After mouse homologs have been cloned, targeted mutations have been generated, and some of these genes are allelic with classical mouse mutations. Due to their origin from homolog screens, these developmentally regulatory genes can be classified into families, which include:

- the fibroblast growth factor (FGF) family
- the FGF receptor/tyrosine kinase family
- the *hedgehog* family
- the *Wnt* family
- the TGFβ/bone morphogenetic protein (BMP) family
- the homeobox gene family
- the *paired* box (*Pax*) family
- the bHLH (basic helix-loop-helix) gene family
- the EGF repeat/*Notch* family
- the zinc finger gene family

A much smaller number of genes involved in somitogenesis have been identified by positional cloning efforts. Many mouse mutations have been identified which affect the normal patterning and development of tail vertebrae. Almost half of the >65 pairs of vertebrae are found in the tail, where defects can be easily noted and are not necessarily life threatening. Positionally cloned mutations with vertebral and somitic defects have included

the classical mutants *Brachury* and *fused.* These genes can be added to the above list:

- the *Brachyury* (T-box) family
- the *fused*/RGS box family

Why attempt to find vertebrate genes involved in somitogenesis when *Drosophila* homologs remained unexamined? There are clearly benefits in both approaches. In this initial period of examining gene function during vertebrate development, the availability of *Drosophila* homologs has completely transformed our understanding of the process. Entire frameworks of key gene interactions appear to be conserved between *Drosophila* and vertebrates, e.g., the conservation of *patched* and *hedgehog* interactions (Carroll, 1995; Goodrich et al., 1996; Johnson et al., 1996; Marigo et al., 1996). Complicating matters is the presence of several vertebrate homologs for every *Drosophila* gene. During the period from which arthropods and chordates shared a common ancestor, the vertebrate genome has undergone what appears to be two sets of genomic duplications (reviewed in Sidow, 1996). Consequently, there are on average four vertebrate homologs of each *Drosophila*-related patterning gene, greatly complicating the investigation of gene function. Although genetic frameworks can be conserved, they need not necessarily be. Somitic mesoderm forms most of the vertebral column, a process unique to vertebrates and only analogous to *Drosophila*, where ectodermal mineralization produces the insect exoskeleton.

Positional cloning efforts have yielded genes that are not *Drosophila* homologs, and hence provide us with genes that may be specific to vertebrate development. For example, the *Brachyury* gene was positionally cloned from the mouse in 1990, and only subsequently have homologs been identified in a number of other vertebrate systems (Schulte-Merker et al., 1994). Due to the immense effort required to positionally clone genes until recently, only a

handful of the 94 mutations affecting the tail or limbs have been cloned (Doolittle et al., 1996).

What roles do the currently known genes play during somitogenesis? Below, summaries are presented of key experiments for each gene family, and a graphical summary is shown in Figure 4. Since genes can play a number of roles throughout development, it is nearly impossible to neatly categorize them in any fashion. Below, I have divided the families into three groups: genes expressed in presomitic mesoderm, genes expressed beginning in early somite formation, and genes expressed later in somitogenesis.

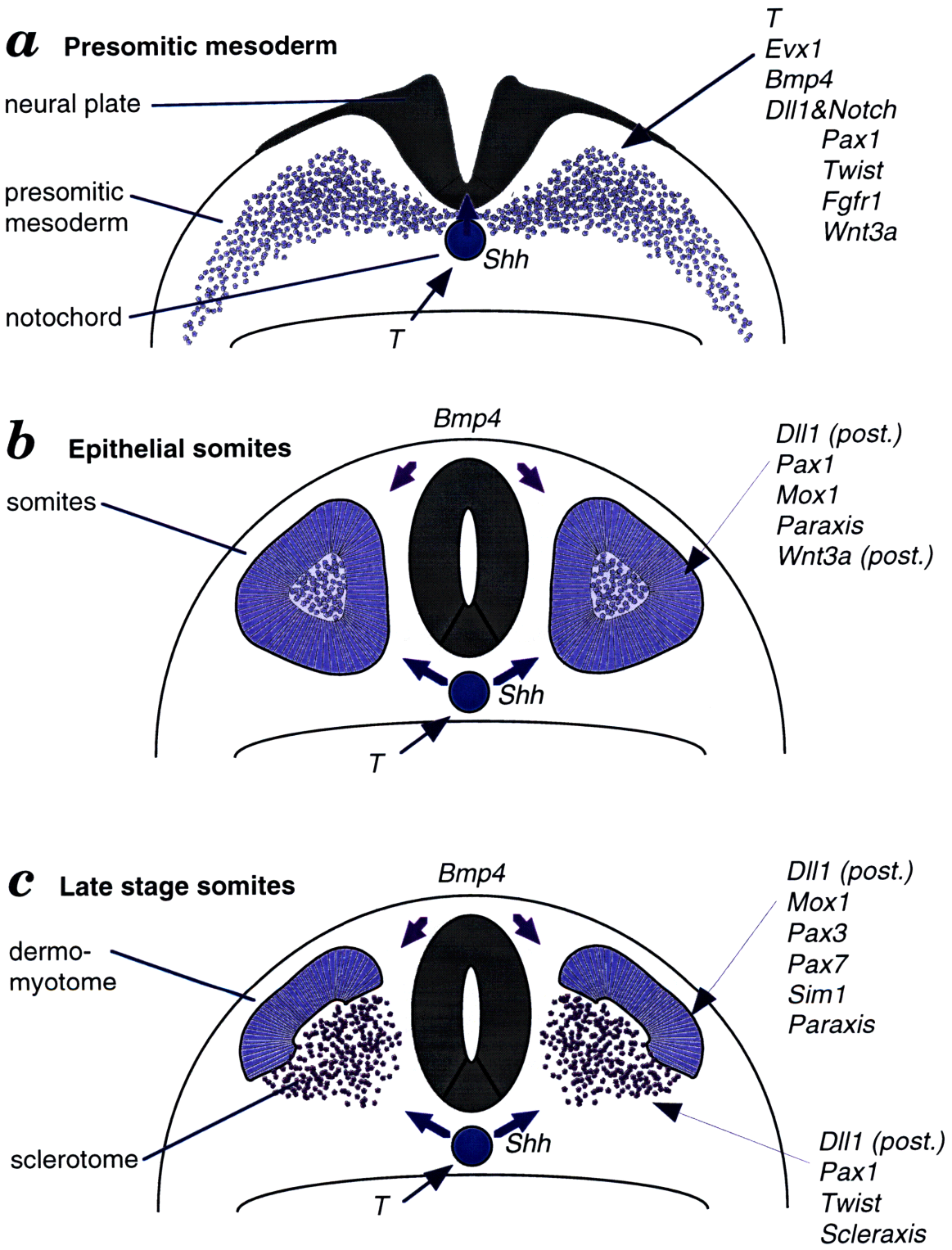



**Figure 4** Gene Expression during Somitogenesis
Three stages are diagrammed, *a* presomitic mesoderm, *b* epithelial sphere stage somites, and *c* somites differentiating into sclerotome and dermomyotome. General regions where genes are expressed are indicated.

### *B. Genes Expressed in Presomitic Mesoderm*

One gene expressed in presomitic mesoderm has quite a long history of investigation. From the time the *Brachyury* (*T*) mutant was identified in 1927 until the gene was positionally cloned in 1990, the locus and its many alleles affecting notochord development have been an intense focus of study. When cloned, the *Brachyury* (*T*) gene was novel and predicted to be a DNA binding transcription factor (Herrmann and Kispert, 1994; Herrmann et al., 1990). The gene is expressed throughout mesoderm during gastrulation and gradually becomes restricted in expression to the notochord and presomitic mesoderm. The function of *T* as a transcription factor has been demonstrated (Kispert et al., 1995) but the downstream targets of the gene have yet to be identified. Numerous alleles of *T* exist, with the more severe alleles, such as $T^{2J}$ resulting in large gaps in the development of notochord, leading to an absence of somites.

*Brachyury* homologs have been identified in other vertebrates (the frog *Xenopus laevis* and the zebrafish *Brachydanio rerio*), with the *no tail* mutant in zebrafish closely resembling the original mouse *Brachyury* mutant. A homolog has even been identified in the non-vertebrate chordate ascidian *Halocynthia roretzi*, where *T* is not expressed in the mesenchyme, and restricted to the notochord. This suggests either that the mesodermal expression is lost in ascidians or that has been acquired since the divergence of ascidian and vertebrate ancestors. Thus, the *Brachyury* mutations emphasize the important inductive role of notochord in somitogenesis.

What are the factors released from the notochord that affects somitogenesis? Candidates include the vertebrate genes homologous to *Drosophila hedgehog*. The *hedgehog* product is a secreted factor that plays an important role in patterning. There exists several vertebrate homologs of *hedgehog*, including *sonic hedgehog* (*shh*/*vhh1*), *cephalic hedgehog*, and *banded*

*hedgehog*. The *sonic hedgehog* gene has been the focus of many experiments examining the interaction between somites and notochord, where *shh* is expressed (Fan and Tessier-Lavigne, 1994; Johnson et al., 1994). Ectopic expression of *shh* is able to induce sclerotomal markers in somites.

Many members of the mammalian fibroblast growth factor (FGF) family are expressed quite early, prior to and during gastrulation, as well as during the somitic patterning. The family currently includes 9 members and 4 tyrosine kinase receptors (Yamaguchi and Rossant, 1995). *Fgf3* is expressed in mesoderm exiting the primitive streak, and targeted gene mutation results in viable homozygotes with only disorganized somites and tail abnormalities. One of the most severe defects can be seen in targeted mutations of the FGF receptor, *Fgfr1*. The result is a complete lack of somites, although the remaining anterior mesodermal tissue is thickened and expresses *T* and the somitic marker *Mox1* (Yamaguchi et al., 1994). The FGF family of genes appears to be expressed in a variety of mesodermal derivatives and expression of family members is required during somitogenesis.

Another family of genes expressed early in mesodermal development are the *Wnt* genes. The *Wnt* gene family is large, with over 15 vertebrate members involved in various aspects of early embryonic patterning (Takada et al., 1994). One member of this family, *Wnt3a* is expressed beginning in the primitive streak and tailbud during gastrulation. The mouse *vestigial tail* (*vt*) mutation was found to be allelic to the *Wnt3a* locus (Greco et al., 1996), and mutant mice are characterized by a loss of pre-sacral vertebrae (Doolittle et al., 1996). Gene targeting of the *Wnt3a* gene results in homozygotes with a loss of somites, notochord, and tailbud beyond the level of the forelimb, at somites 8-12 (Takada et al., 1994). *In situ* hybridization with *T* probe in these mutant

embryos shows scattered cells expressing these mesodermal markers, pointing to a role for *Wnt3a* in organizing somitic cells.

The BMP (bone morphogenetic protein) family of TGFβ superfamily-related cytokines and their serine-threonine kinase family receptors are expressed at a number of different times during development. In *Drosophila*, the TGFβ cytokine *decapentaplegic* plays a key role in the induction of the dorsal-ventral axis. *Bmp4* appears to lateralize somitic tissue in chicks when expressed ectopically (Pourquie et al., 1996). This is contradictory to the pattern of *Bmp4* expression which is expressed in the neural tube and dorsal ectoderm. Mice deficient in *Bmp4* expression by targeted mutagenesis never produce mesoderm during gastrulation, indicating a crucial role played during mesoderm induction (Winnier et al., 1995).

The connection of BMPs with skeletal patterning was first made clear by a spontaneous mutation in the *Bmp5* gene, resulting in the mouse *short ear* phenotype (Kingsley et al., 1992). Numerous skeletal defects are seen in *short ear* mice, due to the absence of *Bmp5* induced condensation sites throughout the developing fetus (Kingsley, 1994). Unlike *Bmp4*, *Bmp5* appears to act much later and does not have effects on somitogenesis.

Homologs of the *Drosophila* cell-determination genes *Notch* and receptors *serrate* and *Delta* have been cloned in vertebrates. All genes in this family contain epidermal growth factor (EGF) repeats and are also transmembrane proteins. A mouse *Delta* homolog, *Dll1*, which is expressed in numerous tissues including the presomitic mesoderm and caudal half of the posterior somites (Bettenhausen et al., 1995). Mouse *Notch1* itself is also expressed in presomitic mesoderm in E8.5 embryos, with RNA also in cephalic mesenchyme, and endothelium of heart and aorta (del Amo et al., 1992). Gene targeting of the *Notch1* gene resulted in embryonic defects first noticeable at

E9, with reduced size, distended pericardia, and discontinuous notochord (Conlon et al., 1995; Swiatek et al., 1994). Detailed examination of somite formation in homozygous mutant embryos showed subtle defects in the timing, regularity and organization of epithelial somites. These initial defects did not affect the development of the sclerotome. Given the importance of *Notch* in *Drosophila* differentiation events, the finding that *Notch* related genes affected somitic organization is not surprising.

### *C. Genes Expressed in Early Somites*

Like their *Drosophila* counterparts, some mammalian homeobox genes are organized in gene clusters, with the gene order corresponding to the temporal and spatial order of expression along the anterior-posterior axis. Expression of the *Hox* genes in vertebrates is required for the differentiation of skeletal and hindbrain segment identity (Krumlauf, 1993). For example, disruption of the *Hoxc8* gene results in the transformation of the L1 vertebra towards the T7 vertebrae and an additional pair of ribs. However, the transformations are not always so simple, as demonstrated by disruptions of *Hoxb4* which only produces a partial transformation of cervical vertebra C2 to C1. In general, the clustered *Hox* genes appear to play a downstream role in the differentiation of somites into adult structures.

Nonclustered genes include the *caudal* homeobox genes, *Cdx1* and *Cdx2*. *Cdx1* transcript is expressed during gastrulation, and its protein is first detectable at E7.5 in the primitive streak (Meyer and Gruss, 1993). Expression can be found in numerous tissues including all mesodermal tissues. However, by E10.5, expression of *Cdx1* diminishes and remains only in the forelimbs. Targeted disruption of the *Cdx1* gene results in the anterior vertebral homeotic transformations, with the homozygous mice still viable and

fertile (Subramanian et al., 1995). Other nonclustered vertebrate homeobox genes are the *Mox* (mesoderm homeobox) genes. Two members of the *Mox* family have been identified, and both are expressed in developing somites (Candia et al., 1992). *Mox1* is expressed in throughout somitogenesis in presomitic mesoderm, lateral plate mesoderm, and mesenchyme in the heart and vasculature. *Mox2* is expressed starting with the epithelial somite stage, and is restricted to the sclerotome during somitogenesis. To date, no targeted mutations have been reported for these genes.

The *paired box* (*Pax*) genes, vertebrate homologs of the *Drosophila paired* gene. To date, nine *Pax* genes have been isolated, with defective alleles of some corresponding to classical mouse mutations (Noll, 1993). *Pax1* is first detected beginning embryonic day 8.5, in the ventromedial part of the formed somites, thus representing one the earliest markers for somitic differentiation (Wallin et al., 1994). During somitogenesis, *Pax1* is expressed in the sclerotome, reflecting the periodic patterning of the somites, and later goes on to be limited to the perichordal region and then the intervertebral discs. Loss of Pax-1 expression results in vertebral malformations along the entire vertebral column. Intervertebral discs are slightly enlarged while vertebral centers are reduced in size. Mutants can first be detected at embryonic day 10.5. In the adult, there is loss of vertebral bodies, intervertebral discs, and proximal sections of the ribs.

An amino acid substitution was found in the paired-box of the *Pax1* gene in the *undulated* mutation which diminished specific DNA binding (Balling et al., 1988; Chalepakis et al., 1991) The *undulated* mouse mutant is characterized by a loss of the posterior portion of vertebrae, and a distortion of vertebrae leading to the "kinky tail" phenotype. In addition, two other mutants were found to have *Pax1* defects including *extensive* ($Un^{ex}$) and *short-tail* ($Un^{s}$),

with the $Un^s$ allele representing a null allele. Examination of the $Un^s$ mutant revealed mice with absent proximal costae, vertebral bodies, and intervertebral discs (descended from notochord), with the neural arch portion of vertebrae intact, i.e., ventral structures defective in this mouse. In the E12 mutant embryos, the notochord is pronouncedly enlarged, and fails to properly form the nucleus pulposus of the intervertebral disk. These findings indicate that there are continued notochord-sclerotome interactions after the effects of signals such as *shh*.

### *D. Genes Expressed Late in Somitogenesis*

Several basic helix-loop-helix (bHLH) genes play a role later in somitogenesis. The mouse homolog of the *Drosophila twist* gene, a bHLH protein, was found to expressed beginning at E8 in the somites and lateral mesoderm (Wolf et al., 1991). The paraxis and scleraxis genes were isolated by a search for additional bHLH genes, motivated by the identification of *myoD* and *myf5* bHLH genes as key players in the differentiation of the myotome lineage (Burgess et al., 1995; Cserjesi et al., 1995). Scleraxis is first detected in somitic sclerotome and mesenchymal cells in the limb and body wall at embryonic day 9.5. It continues to be expressed in bone and chondrocyte precursor cells during skeletogenesis. Disruption of the paraxis gene results in an absence of epithelial somite formation, but does not prevent the differentiation of somitic lineage cells (Rawls et al., 1996). Like their myogenic counterparts, the bHLH genes appear to play a role in the differentiation steps during somitic segmentation.

Zinc finger proteins are an abundant class of genes, with an estimated 300-400 zinc finger genes of the *Krüppel* class alone in vertebrate genomes (Bellefroid et al., 1989). However, not many zinc finger genes have been shown

to participate in the process of somitic patterning. Mutations in the *Gli3* gene have been identified in the mouse mutants *extra-toes* (*Xt*) and *Polydactyly Nagoya* (*Pdn*), producing a number of limb deformities (Ruppert et al., 1990; Ruppert et al., 1991; Schimmang et al., 1992; Schimmang et al., 1994). In the vertebrae of (*Pdn*/*Xt*) mice, the cervical vertebrae showed broadening of the neural arches, but expression of *Gli3* in somites has not been documented. Given recent findings that *Gli3* is downstream in the *hh* and *ptc* pathway in vertebrates as well as *Drosophila*, more zinc finger genes may be involved in these transduction pathways.

Finally, positional cloning efforts are beginning to highlight the role of new gene families in somitogenesis. An example is the recent cloning of the *fused* gene. The *fused* mutation displays a semidominant phenotype of shortened and/or kinked tail, with homozygotes displaying a more severe tail malformation including vertebral and rib fusions (Doolittle et al., 1996). These appear due to ectopic duplication of axes during development. Three alleles have been isolated, with the third transgenic insertion allele disrupting a candidate gene, a novel transcription factor in the RGS (regulator of G-protein) box family. The RGS genes play a negative regulatory role in the signal transduction pathway of signals such as *Sst2*, a pheromone signal in *S. cerevisiae* (Goodrich et al., 1996; Johnson et al., 1996; Marigo et al., 1996).

With over 224 mutations in the mouse affecting skeleton, great potential exists for isolating new gene families involved in somitogenesis (Doolittle et al., 1996). Below, I will outline current strategies for the positional cloning of developmental mutations.

## *III. Approaches to Positional Cloning of Mouse Developmental Mutations*

### *A. Introduction*

How are the genes corresponding to a development defect cloned? Given that we have identified less than 10% of all genes within the genome, simply testing available candidate genes is unlikely to yield a result. However, starting with the mutant phenotype, we can use genetic information to positionally clone the gene. In *Drosophila* and *C. elegans*, the availability of numerous genetic tools allows for the rapid cloning of mutant alleles in those systems.

Within vertebrate systems, many resources have been developed for mouse genetics over the past century. The tools available for identifying mouse mutant genes have improved vastly over the last decade, due to the progress of the Mouse Genome Project. As part of the project, key infrastructural resources have been developed and added to the classic repertoire:

- Genetic Resources: The development of inbred mouse strains allows researchers to have common genetic reagents to compare work, especially useful in the studies of modifiers and suppressors. Also, this enables simple biallelic crosses, and the construction of congenic strains, especially useful in dissecting multigenic traits. Combined with a dense genetic map of over 6,000 simple sequence length polymorphism based markers (Dietrich et al., 1996) and over 2,000 mapped genes, fine-scale genetic mapping is possible throughout the genome.
- Physical Resources: Currently, a wealth of DNA and cDNA clone resources are available for the mouse, including YAC, BAC, P1 and cosmid libraries for genomic DNA clones (Haldi et al., 1996; Kusumi et al., 1993; Larin et al., 1991)

and cDNA libraries from numerous adult tissue and embryonic sources. Furthermore several cDNA libraries are being sequenced to generate expressed sequence tags (EST) data to be deposited in GenBank.

• Genetic Manipulation: Mutations can be induced, either through the insertion of DNA clones through transgenesis, or through the disruption of normal sequence through homologous recombination.

• Environmental Control: As with any experimental system, environmental conditions can be tightly controlled, to decrease environmentally caused phenotypic variance.

### *B. Genetic Mapping*

The first experiments towards positional cloning involve the genetic localization of a mutation. Several approaches have been developed for genetic mapping in the mouse, but linkage analysis is one of the most useful and common techniques for modern positional cloning efforts, given the ability to accumulate high-resolution genetic information. Large numbers of meiotic recombination events can be generated by intercrosses or backcrosses, given the relatively large litter size of mice (average of 8-12 per litter) and the short gestation and nursing times (20 days gestation and 20 days until weaning). Combined with concurrent genotype screening of recombinant progeny using either tail or ear clippings, a large-scale cross of over 2,000 meioses can be generated given a limited amount of mouse cage space by saving only the key recombinant animals. With N meioses, one would expect a flanking recombinant marker from a given locus at 1/N Morgans, and a recombinationally inseparable region of 2/N Morgans. Given 2,000 meioses, we would expect a 0.1cM nonrecombinant interval, and with an average of 2 Mbp of mouse genomic DNA/ cM, we would expect a physical region of 200 kb.

Prior to the availability of a dense array of over 6,000 mouse SSLP markers, deletion alleles of a mutation have been extremely useful for cloning efforts. If genetic markers can be identified which appear to be hemizygous in a mutation region, this may be due to a large deletion, especially if the mutation arose during a x-ray or chlorambucil mutagenesis screen. Smaller deletion alleles of a candidate gene can greatly expedite the process of mutation identification. However, not all deletions are simple to analyze. Insertions and deletions associated with transgenic alleles can have a very complex organization with multiple rearrangements. In addition, the frequency of embryonic lethal loci in mammals makes it difficult to generate very large deletions (> 1MB) that can be maintained in a homozygous state. Overall, deletion alleles are extremely useful when available, but unfortunately since large-scale mutagenesis screens have not been carried out in the mouse, such alleles are usually not available.

Linkage disequilibrium is typically more useful for human disease mapping than for mouse mapping, but can be useful for initial low-resolution mapping of a mutation. For mouse mutations that arose on an inbred strain and were subsequently introgressed into another strain (e.g. arose on DBA/J and introgressed onto C57BL/6J background), determining the extent of ancestral chromosome surrounding the mutation locus may be initially helpful, e.g., with older lines that may have been maintained for 50 backcrosses or more. However, these lines would represent the equivalent of 50 or more meiotic recombination events. In the mouse, development of congenic mutant strains is a poor substitute for linkage analysis.

### *C. Physical Mapping*

Given a small genetic interval, the physical region corresponding to this region must be obtained. Despite the high-resolution genetic mapping possible in the mouse (to within 1 cM), this still leaves one with a large physical interval, since 1 cM in the mouse is approximately 2 Mbp. To span such regions, YACs are currently the largest available clones, with average insert sizes of modern YAC libraries averaging approximately 700 kb. Increasingly, large regions of the mouse genome have been assembled into YAC contigs as part of the Mouse Genome Project. Whereas YACs are well suited for spanning regions, their low fidelity makes them problematic for accurate representation of genomic regions. For that reason, it is common practice to isolate sequence tagged site (STS) markers from the YACs to assemble BAC or P1 clones in a secondary high-resolution physical map, complementing the YAC-based map.

### *D. Transcriptional Mapping*

Candidate genes must next be isolated from the physical region in question. Current estimates of gene number (70,000 genes in the mouse) combined with the size of the genome (3 billion base pairs) yield an average of 23 genes in a 1 Mbp region. How can one identify these genes?

Although prohibitive in the past, the ability to sequence in a high-throughput fashion using automated DNA sequencing technology increasingly makes medium scale (100 kb) sequencing approachable. Also, the rapid pace of the Human Genome Project will result in large amounts of human genomic sequence becoming available in the upcoming years. Unfortunately, the mouse genome is currently not included in these efforts. However, due to the relatively close evolutionary relationship between laboratory mouse strains and humans, estimated at 60 million years divergence, large stretches of genes

are conserved in order between the two organisms, referred to as synteny conservation. Given the number of gene-based markers already available as genetic markers, it is possible to identify the syntenically conserved equivalent region in the human, and scan the genomic sequence for genes.

The availability of raw, assembled genomic sequence does not immediately allow the investigator to develop a transcript map. There are both computational and empirical methods of identifying genes in the region, neither of which are yet 100% efficient. Numerous algorithms have been designed to find gene sequences, but basically they fall into two classes: those that look for homology to known gene sequences in the databases (e.g. BLAST) and those that look for gene motifs (e.g. splice donor/acceptor sites with GRAIL) (Fickett, 1996). However, the current computational methods are not adequate to identify all transcripts.

More traditional approaches to transcript mapping are also available. With the human genome, EST transcript fragments are being sequenced and localized on the human physical map (Schuler, 1996). Mouse EST sequencing efforts are underway at Washington University and Kyoto University, and a substantial amount of expressed sequence will be available in the near future for similar efforts in the mouse. In particular regions, expressed sequences can be cloned by direct cDNA selection (Lovett, 1994) or by exon-trapping methods (Buckler et al., 1991). Currently, obtaining cDNA fragments remains an crucial method towards the construction of transcript maps.

### *E. Candidate Gene Testing*

Ultimately, one must identify the sequence defect associated with the mutant phenotype and present evidence that it is causal to the phenotype. Mutations may range from the straightforward (e.g. 10 kb deletions leading to

a truncated gene product) to the difficult (e.g., single base-pair polymorphism of a missense mutation or a single base-pair polymorphism in a non-coding regulatory region). If the parental chromosome on which the mutation arose is available, then the mutation is likely to be the only sequence polymorphism. However, additional evidence, as discussed below, would still be desirable since spontaneous mutations could have arisen which are non-causal to the mutant phenotype.

Loss-of-function mutations can be complemented by the introduction of a wild-type allele of the gene. This can be achieved by chromosomal translocations (classic method), or more commonly now by transgenic techniques. If a cellular phenotype exists, then transfection into mutant somatic cell lines may provide an excellent *in vitro* complementation assay. If no cellular phenotype has been noted, then it is necessary to generate germ-line transgenic organisms through microinjection. Introduced DNA can either be large genomic clones containing the gene in question, or consist of a construct and appropriate promoter. Whether a cDNA construct or genomic clone is used, it is always possible that insufficient sequences were introduced, and that the expression pattern of the artificial construct will not completely mimic that of the wild type gene. In addition, since transgenic insertions often can result in more than one copy per cell, overexpression is a potential problem as well. The addition of wild type sequences into mutants, in the attempt to restore normal gene expression and compensate the mutant phenotype, can be invaluable "proof" that a candidate gene is involved in a loss-of-function mutation.

Alternatively, specific gene candidates can be tested by inducing a mutation in the wild type allele. Activity can be eliminated or reduced by homologous recombination methods, but generating the desired allele may not

be as simple as "knocking" out a gene. The phenotype observed with the mutation being positionally cloned may result from a specific loss-of-function or degree of loss, and this phenotype may not be reproduced by an induced mutation. For example, targeted mutations of the *Wnt3a* gene produce a much more severe phenotype than the allelic *vestigial tail* mutation (Dickinson and McMahon, 1992; Greco et al., 1996; Takada et al., 1994). Given the long periods of time required to generate a knockout and breed to homozygosity, this method may be difficult to test more than one or two candidates. Traditional alteration of the genome, through mutagenic chemical or irradiation treatments followed by mutant screen, is another alternative that is seldom used in the mouse due to the difficulty of carrying out such a screening. An exception is the isolation of the mouse *clock* mutation which was isolated through a large-scale screen of ethyl-nitrosourea mutagenized mice (Vitaterna et al., 1994).

## *IV. Overview: Positional Cloning of the pudgy Locus*

In this doctoral thesis, I outline my efforts to positionally clone a classic developmental mutation in the mouse. The *pudgy* mutation was initially described by Hans Grüneberg in 1961 as a mutant with defective somitogenesis, and highly deformed vertebrae and ribs. Since the original report, no further work on the mutation has been reported. I will describe herein our efforts to identify the *pudgy* gene, involving:

1. The detailed analysis of the *pudgy* mutant phenotype, including documentation of defects in neonatal and adult skeleton, mutant histology, and altered patterns of gene expression by whole mount *in situ* hybridization.
2. The construction of the first publicly distributed large-insert mouse YAC library for positional cloning efforts.
3. The integration of genetic, physical, transcriptional, and genomic sequence information in the ~600 kb genomic region containing the *pudgy* gene.
4. The screening for the gene responsible for the *pudgy* defect. This screen is ongoing, and the evidence for the key *pudgy* candidate, a novel relative of the p21-activated serine-threonine kinase (PAK) family, will be presented. In addition, ongoing efforts to rescue the *pudgy* phenotype by BAC and P1 transgenic complementation will be described.
5. Finally in an appendix, I describe the genetic mapping of three other developmental mutations, *motor neuron degeneration*, *ducky*, and *curly whiskers*.

## References

Balling, R., Deutsch, U., and Gruss, P. (1988). undulated, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of Pax 1. Cell *55*, 531-5.

Bellefroid, E. J., Lecocq, P. J., Benhida, A., Poncelet, D. A., Belayew, A., and Martial, J. A. (1989). The human genome contains hundreds of genes coding for finger proteins of the Kruppel type. Dna *8*, 377-87.

Bettenhausen, B., de Angelis, M. H., Simon, D., Guenet, J. L., and Gossler, A. (1995). Transient and restricted expression during mouse embryogenesis of Dll1, a murine gene closely related to Drosophila Delta. Development *121*, 2407-18.

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.

Candia, A. F., Hu, J., Crosby, J., Lalley, P. A., Noden, D., Nadeau, J. H., and Wright, C. V. (1992). Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially expressed during early mesodermal patterning in mouse embryos. Development *116*, 1123-36.

Carroll, S. B. (1995). Homeotic genes and the evolution of arthropods and chordates. Nature *376*, 479-85.

Chalepakis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M., and Gruss, P. (1991). The molecular basis of the undulated/Pax-1 mutation. Cell *66*, 873-84.

Conlon, R. A., Reaume, A. G., and Rossant, J. (1995). Notch1 is required for the coordinate segmentation of somites. Development *121*, 1533-45.

del Amo, F. F., Smith, D. E., Swiatek, P. J., Gendron-Maguire, M., Greenspan, R. J., McMahon, A. P., and Gridley, T. (1992). Expression pattern of Motch, a mouse homolog of Drosophila Notch, suggests an important role in early postimplantation mouse development. Development *115*, 737-44.

Dickinson, M. E., and McMahon, A. P. (1992). The role of Wnt genes in vertebrate development. Curr Opin Genet Dev *2*, 562-6.

Dietrich, W. F., Miller, J., Steen, R., Merchant, M. A., Damronboles, D., Husain, Z., Dredge, R., Daly, M. J., Ingalls, K. A., Oconnor, 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., Page, D. C., and Lander, E. S. (1996). A comprehensive genetic map of the mouse genome. Nature *380*, 149-152.

Doolittle, D. P., Davisson, M. T., Guidi, J. N., and Green, M. C. (1996). Catalog of mutant genes and polymorphic loci. In Genetic Variants and Strains of the Laboratory Mouse, M. F. Lyon, S. Rastan and S. D. M. Brown, eds. (Oxford: Oxford University Press), pp. 1807.

Duband, J. L., Dufour, S., Hatta, K., Takeichi, M., Edelman, G. M., and Thiery, J. P. (1987). Adhesion molecules during somitogenesis in the avian embryo. J Cell Biol *104*, 1361-74.

Fan, C. M., and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. Cell *79*, 1175-86.

Fickett, J. W. (1996). Finding genes by computer: the state of the art. Trends Genet. *12*, 316-320.

Gilbert, S. F. (1994). Developmental Biology, 4th Edition (Sunderland, MA: Sinauer Associates, Inc.).

Goodrich, L. V., Johnson, R. L., Milenkovic, L., McMahon, J. A., and Scott, M. P. (1996). Conservation of the hedgehog/patched signaling pathway from flies to mice: induction of a mouse patched gene by Hedgehog. Genes Dev *10*, 301-12.

Greco, T. L., Takada, S., Newhouse, M. M., McMahon, J. A., McMahon, A. P., and Camper, S. A. (1996). Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. Genes Dev *10*, 313-24.

Grüneberg, H. (1961). Genetical studies on the skeleton of the mouse: XXIX. Pudgy. Genet. Res., Camb. *2*, 384-393.

Haldi, M. L., Strickland, C., Lim, P., Vanberkel, V., Chen, X. N., Noya, D., Korenberg, J. R., Husain, Z., Miller, J., and Lander, E. S. (1996). A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome. Mamm Genome *7*, 767-769.

Herrmann, B. G., and Kispert, A. (1994). The T genes in embryogenesis. Trends Genet *10*, 280-6.

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.

Hogan, B., Beddington, R., Constantini, F., and Lacy, E. (1994). Manipulating the Mouse Embryo:A Laboratory Manual, 2nd Edition ( Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Jacobson, A. G. (1988). Somitomeres: mesodermal segments of vertebrate embryos. Development *104*, 209-20.

Johnson, R. L., Laufer, E., Riddle, R. D., and Tabin, C. (1994). Ectopic expression of Sonic hedgehog alters dorsal-ventral patterning of somites. Cell *79*, 1165-73.

Johnson, R. L., Rothman, A. L., Xie, J., Goodrich, L. V., Bare, J. W., Bonifas, J. M., Quinn, A. G., Myers, R. M., Cox, D. R., Epstein, E. H., Jr., and Scott, M. P. (1996). Human homolog of patched, a candidate gene for the basal cell nevus syndrome. Science *272*, 1668-71.

Kent, G. C. (1987). Comparative Anatomy of the Vertebrates (Dubuque, IA: William C. Brown, Publishers).

Keynes, R. J., and Stern, C. D. (1988). Mechanisms of vertebrate segmentation. Development *103*, 413-29.

Kingsley, D. M. (1994). What do BMPs do in mammals? Clues from the mouse short-ear mutation. Trends Genet *10*, 16-21.

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.

Kispert, A., Koschorz, B., and Herrmann, B. G. (1995). The T protein encoded by Brachyury is a tissue-specific transcription factor. Embo J *14*, 4763-72.

Krumlauf, R. (1993). Mouse Hox genetic functions. Curr Opin Genet Dev *3*, 621-5.

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.

Larin, Z., Monaco, A. P., and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc Natl Acad Sci U S A *88*, 4123-7.

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. (New York: Current Protocols).

Marigo, V., Scott, M. P., Johnson, R. L., Goodrich, L. V., and Tabin, C. J. (1996). Conservation in hedgehog signaling: induction of a chicken patched homolog by Sonic hedgehog in the developing limb. Development *122*, 1225-33.

Meyer, B. I., and Gruss, P. (1993). Mouse Cdx-1 expression during gastrulation. Development *117*, 191-203.

Noll, M. (1993). Evolution and role of Pax genes. Curr Opin Genet Dev *3*, 595-605.

Ostrovsky, D., Cheney, C. M., Seitz, A. W., and Lash, J. W. (1983). Fibronectin distribution during somitogenesis in the chick embryo. Cell Differ *13*, 217-23.

Pourquie, O., Fan, C. M., Coltey, M., Hirsinger, E., Watanabe, Y., Breant, C., Francis-West, P., Brickell, P., Tessier-Lavigne, M., and Le Douarin, N. M. (1996). Lateral and axial signals involved in avian somite patterning: a role for BMP4. Cell *84*, 461-71.

Rawls, A., Burgess, R., Brown, D., Bradley, A., and Olson, E. N. (1996). Requirement of the bHLH gene *paraxis* for somite formation and epithelialization. In Mouse Molecular Genetics, R. Beddington, A. Bradley, R. Krumlauf and L. Robertson, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory), pp. 334.

Ruppert, J. M., Vogelstein, B., Arheden, K., and Kinzler, K. W. (1990). GLI3 encodes a 190-kilodalton protein with multiple regions of GLI similarity. Mol Cell Biol *10*, 5408-15.

Ruppert, J. M., Vogelstein, B., and Kinzler, K. W. (1991). The zinc finger protein GLI transforms primary cells in cooperation with adenovirus E1A. Mol Cell Biol *11*, 1724-8.

Schimmang, T., Lemaistre, M., Vortkamp, A., and Ruther, U. (1992). Expression of the zinc finger gene Gli3 is affected in the morphogenetic mouse mutant extra-toes (Xt). Development *116*, 799-804.

Schimmang, T., Oda, S. I., and Ruther, U. (1994). The mouse mutant Polydactyly Nagoya (Pdn) defines a novel allele of the zinc finger gene Gli3. Mamm Genome *5*, 384-6.

Schuler, G. D. (1996). A gene map of the human genome. Science *274*, 540-546.

Schulte-Merker, S., van Eeden, F. J., Halpern, M. E., Kimmel, C. B., and Nusslein-Volhard, C. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. Development *120*, 1009-15.

Serbedzija, G. N., Fraser, S. E., and Bronner-Fraser, M. (1990). Pathways of trunk neural crest cell migration in the mouse embryos as revealed by vital dye labelling. Development *108*, 605-612.

Sidow, A. (1996). Gen(om)e duplications in the evolution of early vertebrates. Curr. Opinion Genet. Dev. *6*, 715-722.

Subramanian, V., Meyer, B. I., and Gruss, P. (1995). Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. Cell *83*, 641-53.

Swiatek, P. J., Lindsell, C. E., del Amo, F. F., Weinmaster, G., and Gridley, T. (1994). Notch1 is essential for postimplantation development in mice. Genes Dev *8*, 707-19.

Takada, S., Stark, K. L., Shea, M. J., Vassileva, G., McMahon, J. A., and McMahon, A. P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. Genes Dev *8*, 174-89.

Tam, P. P., and Tan, S. S. (1992). The somitogenetic potential of cells in the primitive streak and the tail bud of the organogenesis-stage mouse embryo. Development *115*, 703-15.

Vitaterna, M. H., King, D. P., Chang, A. M., Kornhauser, J. M., Lowrey, P. L., McDonald, J. D., Dove, W. F., Pinto, L. H., Turek, F. W., and Takahashi, J. S. (1994). Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. Science *264*, 719-25.

Wallin, J., Wilting, J., Koseki, H., Fritsch, R., Christ, B., and Balling, R. (1994). The role of Pax-1 in axial skeleton development. Development *120*, 1109-21.

Winnier, G., Blessing, M., Labosky, P. A., and Hogan, B. L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes Dev *9*, 2105-16.

Wolf, C., Thisse, C., Stoetzel, C., Thisse, B., Gerlinger, P., and Perrin-Schmitt, F. (1991). The M-twist gene of Mus is expressed in subsets of mesodermal cells and is closely related to the Xenopus X-twi and the Drosophila twist genes. Dev Biol *143*, 363-73.

Yamaguchi, T. P., Harpal, K., Henkemeyer, M., and Rossant, J. (1994). fgfr-1 is required for embryonic growth and mesodermal patterning during mouse gastrulation. Genes Dev *8*, 3032-44.

Yamaguchi, T. P., and Rossant, J. (1995). Fibroblast growth factors in mammalian development. Curr Opin Genet Dev *5*, 485-91.

**Chapter Two**

# The Mouse *pudgy* Mutation Causes Defects in Somitogenesis and Early Embryonic Lethality

KENRO KUSUMI,*† EILEEN S. SUN* AND ERIC S. LANDER*†

** Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, 02142;*

*†Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139*

## INTRODUCTION

In vertebrates, somitogenesis is a temporally and spatially regulated process which results in the formation of highly regular mesodermal segments. Formation of somites follows an anterior-to-posterior progression through embryonic development, with the production of an epithelial sphere of a somite on average every 1.5 hours (reviewed in Hogan et al., 1986). In the mouse, the most anterior somites coalesce beginning on embryonic day 8 and the most posterior somites coalesce ending embryonic day 12. From the initial stage of an epithelial sphere which lasts approximately 10 hours, the ventral sections of the somites dissociate, producing mesenchymal cells which migrate to form the sclerotome. The sclerotome is the precursor to adult vertebrae and ribs, and the remaining laminar cells form the dermomyotome, the precursor to the adult axial musculature and dermis.

Identification of genes involved in the process of vertebrate somitogenesis has come from three sources—the cloning of vertebrate homologs of *Drosophila* patterning genes, the identification of somitic expression in other

cloned genes, and the identification of murine genes disrupted by mutations. The number of gene families involved in the development of the somitic mesoderm has grown considerably. This list presently includes the *Brachyury* (*T*) related genes (reviewed in Herrmann and Kispert, 1994), the fibroblast growth factors (FGF) (reviewed in Yamaguchi and Rossant, 1995), the *Wnt* family (reviewed in Dickinson and McMahon, 1992), the BMP/TGFβ family of cytokines (reviewed in Hogan, 1996; Kingsley, 1994), and the basic helix-loop-helix (bHLH) genes (Burgess et al., 1995; Cserjesi et al., 1995) which were first identified in vertebrates. Genes first identified in *Drosophila* include the homeobox genes (*Hox*) (reviewed in Krumlauf, 1993), the paired-rule (*Pax*) (reviewed in Noll, 1993), the *hedgehog* related genes (Fan and Tessier-Lavigne, 1994; Johnson et al., 1994), and the EGF repeat/*Notch* related genes (Conlon et al., 1995; del Amo et al., 1992; Lardelli and Lendahl, 1993; Reaume et al., 1992; Swiatek et al., 1994).

Genes involved in somitogenesis have been identified by positional cloning of mouse mutations disrupting normal vertebral formation. During the past century, 94 mutations with tail and limbs phenotypes have been identified (Doolittle et al., 1996). Three factors aid in the isolation of tail defects: 1. almost half of the vertebrae in the mouse are in tail, 2. the tail is not required for animal viability, 3. the tail is external and spontaneous mutations are easily identified. Of the numerous mutations, *Brachyury, fused, vestigial tail,* and *undulated* have been positionally cloned.

From the time the *Brachyury* (*T*) mutant was identified in 1927 until the gene was positionally cloned in 1990, the locus and its many alleles affecting notochord development have been an intense focus of study. When cloned, the *Brachyury* (*T*) gene was predicted to be a DNA binding transcription factor (reviewed in Herrmann and Kispert, 1994; Herrmann et al., 1990). The gene is

expressed throughout the mesoderm during gastrulation and gradually becomes restricted in expression to the notochord and presomitic mesoderm. Subsequently, homologs have been identified in other vertebrates (the frog *Xenopus laevis* and the zebrafish *Brachydanio rerio*), with the *no tail* mutant in zebrafish closely resembling the original mouse *Brachyury* mutant (Schulte-Merker et al., 1994). Loss of *Brachyury* function leads to loss of notochordal segments, which secrete factors such as *Sonic hedgehog* that are essential for the patterning of the somite.

Several vertebral mutants were found to be allelic with the *Pax1* gene, which is expressed in early somites and becomes restricted to sclerotome during somite maturation. An amino acid substitution was found in the paired-box domain in the *undulated* (*un*) mutation (Balling et al., 1988; Chalepakis et al., 1991). In addition, two other mutants were found to have *Pax1* defects including *extensive* ($un^{ex}$) and *short-tail* ($Un^{s}$), with the $Un^{s}$ allele representing a null allele. The *un* mutant is characterized by a loss of the posterior portion of vertebrae, and a distortion of vertebrae leading to the "kinky tail" phenotype. The null allele $Un^{s}$ mutant had a more severe phenotype, with absent proximal costae, vertebral bodies, and intervertebral discs (descended from notochord). Only the neural arch of vertebrae, which is the most dorsal structure, appeared unaffected. These mutant alleles help to identify the role of *Pax1* in the ventralization of somites during development.

One classic mouse mutation that has remained uncloned is *pudgy*. The *pudgy* mutation arose during the course of X-ray mutagenesis experiments by W. and L. Russell, designed to find additional mutations in vertebral patterning (Grüneberg, 1961). In the *pudgy* mutation, somitic coalescence is irregular and lamination is defective. These defects are noticeable at the very onset of somitogenesis on embryonic day 8. No deletion of embryonic structures such

as the notochord was noted in the original description, making the *pudgy* mutant distinct from defects such as *Brachyury* in which there is loss of notochord structure. In the adult, the defects appear limited to the vertebrae and ribs, without affecting cranial or limb skeletogenesis. In overall appearance, the mutants have a shortened vertebral column and stunted or absent tail, giving them a "toad-like" appearance.

To further understand the role of the *pudgy* gene during somitogenesis, we studied the mutant phenotype which included: 1. neonatal and adult skeletal defects, including deformities severe enough to cause hindlimb paralysis, 2. defects in somitic structure, by histological sections, 3. poorly segmented somites in E9.5 embryos, by whole mount *in situ* hybridization with somitic and other mesodermal probes, 4. a loss of homozygous *pudgy* mutant embryos, due to early lethality detectable at embryonic day 9.5, by genetic studies.

## MATERIALS AND METHODS

*Mouse skeletal preparation*

Preparations for the differential staining of cartilage and bone were carried out essentially as described previously (McLeod, 1980). Prior to staining, neonatal mice were carefully dissected to remove all internal organs and muscle and skin tissue. Neonates were fixed in 95% ethanol for 5 days, placed in acetone for 2 days, and stained for 3 days at 37°C in an opaque container with the following solution: 1 volume 0.1% Alizarin red S (Sigma Chemical, St. Louis, MO) in 70% ethanol, 1 volume 0.3% Alcian blue 8GS (Sigma) in 70% ethanol, 1 volume glacial acetic acid (EM Science) and 17 volumes 70% ethanol. After staining, skeletons were washed in distilled water and cleared in 1% aqueous potassium hydroxide for 48 hours. Preparations were gradually placed into glycerol with 1 week incubations in the following

solutions: 20% glycerol/0.8% potassium hydroxide, 50% glycerol/0.5% potassium hydroxide, 80% glycerol/0.2% potassium hydroxide. Finally, skeletons were transferred to 100% glycerol and stored in the dark.

For preparations of adult skeletons, all soft tissue was carefully removed and skeletons stained only with Alizarin red solution. Adult skeletons were left in potassium hydroxide solution until soft tissue was mostly dissolved.

*Histological Examination of Mouse Tissues*

Embryos for histological studies were fixed overnight in 4% paraformaldehyde solution in PBS at 4°C. Samples were dehydrated and fixed in paraffin wax, and 5 μm paraffin embedded sections were prepared by standard histological protocols (Kaufman, 1992). Sections were stained in hematoxylin and counterstained with eosin.

*In situ hybridization experiments*

Templates for RNA probes were prepared by PCR using T7-chimeric primers. Underlined sequences represent non-gene sequences containing T7 promoter site, and sequence numbers in the primer name represent the starting position relative to GENBANK sequence. *Brachyury* probe was made using primers T1584 (TGTTCTACAGCCTCTTGTTTGA) and T7-T2008 (GCGTAATACGACTCACTATAGGGAGATTTCTGCAGATTGTCTTTGGC). *Mox1* probe was made using primers MOX1191 (AGTTGCCCAGTATGTGGGAG) and T7-MOX1628 (GCGTAATACGACTCACTATAGGGAGATTGGAGAACACAAGACGCTG). *Pax1* probe was made using primers PAX1920 (AGCAGCCACAGTCCCAAG) and T7-PAX2321 (GCGTAATACGACTCACTATAGGGAGAAACCTCACCACCCTGAAGC). *Dll1* probe was made from clone p15-6 (gift from A. Gossler). *Evx1* probe was made from clone pAB11 (gift from M. Dush and G. Martin). *Shh1* probe was made

from clone Hh-16.1 (gift from D. Epstein, B. St.-Jacques, and A. MacMahon). Templates were purified by gel purification in Seaplaque GTG agarose (FMC Bioproducts, Inc., Rockland, ME). DNA was freed from agarose using β-agarase (New England Biolabs, Beverly, MA) and purification using Ultrafree-MC® 30,000 NMWL filters (Millipore Corporation, Bedford, MA) and resuspended in 1x TE. RNA probes for in situ hybridization were made using digoxigenin-labeled UTP (Boehringer-Mannheim, Indianapolis, IN) using T7 or Sp6 MAXIscript™ in vitro transcription kits (Ambion, Inc., Austin, TX).

Homozygous *pu/pu* and wild-type heterozygous *pu/+* embryos were generated by crosses of *pu/pu* male mutants with *pu/+* females. Females were checked daily for evidence of copulation plugs, and embryos were designated as 0.5 day at the time of plug identification. E8.5, E9.5, and E10.5 embryos were dissected for in situ analysis, with embryos immediately fixed in 4% paraformaldehyde in PBS overnight at 4°C with gentle shaking. Visceral yolk sacs were saved and frozen at -80°C in TE. Genomic DNA was extracted from yolk sacs as described previously (Laird et al., 1991). Yolk sac DNAs were genetically typed to determine if the corresponding embryos were *pu/pu* or *pu/+*, expected to be in a 1:1 ratio from the cross. DNAs were typed with a single stranded conformation polymorphism (SSCP) marker generated from the promoter region of *Zfp36*, found to be recombinationally inseparable from the *pudgy* mutation in over 2,000 meioses, as described previously (Kusumi et al., 1997).

Whole mount *in situ* hybridization with digoxigenin-labeled probes was carried out essentially as described previously (Wilkinson, 1992). Sections of *in situ* hybridized tissues were prepared from frozen samples embedded in Tissue-Tek O.C.T. (Miles Diagnostic Division, Elkhart, IN) sectioned at 7μm.

## RESULTS

*Skeletal Pathology*

The external *pudgy* phenotype is quite striking in the adult and neonatal mouse (Figure 1). The torso of the *pudgy* mouse is significantly shortened compared to its wild-type littermate, and the tail is always stunted and malformed. Due to the shortened tail, affected animals are unable to right themselves when placed on their backs, and they show reduced ability to balance when placed on a pedestal. The most severely affected adults manifest an extremely short torso and hindlimb motor paralysis. Despite these defects, *pudgy* males breed, and *pudgy* females are able to bear litters, although with reduced fecundity.

Alizarin Red/ Alcian Blue preparations of neonatal *pudgy* and heterozygous wild-type sib showed that the *pudgy* mice have defective costae (ribs), vertebral malformations, but no apparent defects in the formation of limbs or cranium (Figure 2). Cervical vertebrae are malformed, although less severely than thoracic vertebrae (Figure 3). In the lumbar and sacral regions, it is difficult to distinguish one vertebra from another. The position of vertebrae relative to the hindlimbs and internal organs is displaced anteriorly, due to the decreased size of the *pudgy* mutant vertebrae. This affects the formation of both surrounding bone and soft tissue structures. Due to the foreshortening of all vertebrae in the affected *pudgy* animals, the entirety of the vertebral column is shifted anteriorly, i.e. tail vertebrae would be found in the pelvic location, lumbar vertebrae shifted to a thoracic location, etc. As a consequence, the egress of the sciatic nerve in the *pudgy* animals is located in the thoracic region vertebrae rather than the normal lumbar region (data not shown).

In addition to the vertebral defects, *pudgy* animals show significant costal abnormalities (Fig. 4). Costal malformations are variable and include fusion, bifurcation, and ectopic condensations. Thus we see severe developmental defects in the descendants of the sclerotome—the vertebrae and ribs—but have not observed any gross defects in the dermomyotomal descendants, the spinal, limb, and costal musculature, or the dermis of the back.

*Histological Examination*

Grüneberg's original histological analysis of the *pudgy* phenotype showed no detectable differences in the presomitic mesoderm of wild-type and mutant E11.5 embryos (Grüneberg, 1961). Examining the first three somites, he noted both irregular somitic outline and inner organization. We have examined hematoxylin and eosin stained histological sections of E9.5 and E11.5 embryos in order to complement his original observations.

We find irregular outline of the somites in mutant E9.5 embryos as compared to wild-type embryos, but no evidence of any gross deviation of epithelial cells from the general columnar somite organization (Fig. 5). Although the dermomyotome maintains a laminar organization, the disaggregation of the sclerotome makes it difficult to detect any organizational differences, since there is less clear organization seen under hematoxylin and eosin stained sections. As shown in Figure 5, the somitic marker *Mox1* shows no major difference in expression pattern in homozygous mutant versus heterozygous wild-type sib.

*Expression of Somitic Marker Genes in pudgy Embryos*

One technology not available in the original 1961 description of the *pudgy* mutant is the ability to examine the distribution of gene expression in mutant animals by *in situ* hybridization. We examined the expression of several mesodermal markers and other developmental markers in order to further dissect the *pudgy* phenotype.

We chose to examine E9.5 embryos, as that stage would reflect the intermediate period of somitogenesis. To confirm the histological observation of a normal notochord in *pudgy* mutants, we used a *Sonic hedgehog* probe for whole mount *in situ* hybridization and observed an intact and uninterrupted notochord in *pudgy* and heterozygous wild-type sib embryos (Figure 6 c & d). To examine the presomitic mesoderm, we used a probe against the mesodermal marker *Brachyury* which is expressed predominantly in the tailbud at E9.5. The pattern or expression in the mutants was indistinguishable from the heterozygous wild-type embryos (Figure 6 a & b).

To examine the organization of the somites themselves, we used three markers that are expressed in different patterns within the somites at E9.5: *Mox1* which is expressed throughout the dermomyotome and sclerotome, *Pax1* which is restricted in expression to the sclerotome in more mature somites, and *Dll1* which has a complex pattern of somitic expression. *Dll1* is a homolog of the *Drosophila Delta* gene, a ligand of the cellular determination gene, *Notch*. At E9.5, *Dll1* is expressed in regions including the caudal half of the somites, the presomitic mesoderm, the mesonephric tubules, the central nervous system (Bettenhausen et al., 1995). With the *Dll1* gene, we see similar regions of presomitic mesoderm in *pudgy* and wild-type embryos, and very weak caudal somitic hybridization (Fig. 6 e & f).

With the *Mox1* marker, we note the irregular outer shape of the somites in the *pudgy* embryos (Fig. 7 c & d), and indistinct invaginations in the somites adjacent to the presomitic mesodermal plate (Fig.7 a & b). The somites also appear irregularly sized as compared to their wild-type correlates. Similarly with the *Pax1* expression, the irregularity of the *pudgy* somites is indicated by the arrows (Fig. 7 e & f). In summary, the three somitic markers examined show morphological abnormalities consistent with the original description of the mutant.

*pudgy mutant early lethality*

The *pudgy* mutant has been previously described to have a high rate of embryonic lethality, estimated to be 18% in the single cross examined. (Grüneberg, 1961) We have found an under-representation of the *pu/pu* genotypic class in each of the two separate intercrosses (Kusumi et al., 1997). In the (TKDU/J-*du/du* x PU/J-*pu/pu*)F2 intraspecific cross, we found 126 affected animals out of 691 progeny, yielding only 18% *pu/pu* homozygous animals in the cross (rather than the expected 25% from a recessively acting mutation in an intercross), or a 27% lethality ($p<0.001$, $X^2=16.9$). In the intersubspecific cross with *M. musculus molossinus*, we find 44 affected animals out of 349 progeny, yielding only 13% homozygous mutant animals, or a 50% lethality ($p<0.001$, $X^2=28.6$). All *pudgy* affected animals were homozygous for PU/J alleles at nearby flanking markers, and all *pu/pu* animals were affected. Thus, the *pudgy* mutation was fully penetrant, with a loss of the *pu/pu* homozygous class in these crosses.

We further investigated the step at which this lethality can be observed. As part of our efforts to generate embryos for in situ hybridizations, we used mice from our (PU/J-*pu/pu* x MOLF/Ei)F2 lineage and set up matings with

*pu/pu* homozygous males and *pu/+* heterozygous females. Females were checked for plugs, and the pregnant mothers were sacrificed 9.5 days after conception and the embryos were dissected. Yolk sacs from embryos were saved and used for the preparation of genomic DNA for typing. Decidua which appeared to contain no extant embryos were not saved for further experimentation. Whereas the expected ratio of *pu/pu* vs. *pu/+* embryos for a closely linked marker would be 1:1, we observed 57 *pu/pu* embryos out of 148 embryos examined, as assayed by the marker *Zfp36* which is genetically inseparable from *pu*. The mean litter size of the 19 litters examined was 8, which is within normal range. 39% of all embryos were typed as *pu/pu* (41% divided by litter), which translated into a 23% lethality rate, similar to that seen in the intraspecific crosses ($p<0.01$, $X^2=7.8$). The backcross used in this experiment differs from the MOLF/Ei F2 intercrosses which yielded 50% lethality, thus we are not able to make a direct comparison, due to possible effects of MOLF background loci. However, we still observed a significant under representation of *pu/pu* genotypic embryos at E9.5 (23% lethality) making the step at which *pudgy* mutation exerts a lethal effect prior to embryonic day 9.5.

## DISCUSSION

The *pudgy* mutation has a severe effect on the formation of skeletal structures in the mouse. The defects are primarily limited to the vertebrae and ribs, with no effects on limb or cranial formation and minor effects of the pelvis. While the cervical vertebrae appear the least affected, the thoracic, lumbar, sacral, and caudal vertebrae are so severely affected that it is difficult to distinguish each vertebra, and spinous processes are misaligned and occasionally fused. The entire spinal column is affected, yet there is no

reduction in the number of vertebrae, as is observed in the skeletons of *Brachyury* mice. This also distinguishes the *pudgy* mutation from mutations in *Hox* cluster genes, such as Hoxc8 which transforms the morphology of one vertebrae into another, or *Pax1*, which has subtle effects such as the loss of a specific segment of each vertebra (Krumlauf, 1993; Wallin et al., 1994).

The effects of the *pudgy* mutation on the ribs are striking. The defects encompass fusions, bifurcations, and truncations of ribs. Combined with the thoracic vertebral defects, the overall effect is the widening and shortening of the thoracic cage. The specificity of defects to the ribs and vertebrae are instructive, given their descent from a common mesodermal structure, the sclerotome.

We can contrast this with the absence of major defects in other descendants of the somite. The dermotome develops into the dermis of the back. Dermis induces and maintains hair follicles, and we did not observe any gross abnormalities in the hair of *pudgy* mutants . The myotome forms the axial musculature of the spinal column. We did not observe any gross pathology of the axial musculature, and *pudgy* mutants are able to move normally and even mate, making major defects in axial musculature unlikely.

Effects of the *pudgy* mutation are evident from the earliest stages of somitogenesis. Histological examination using hematoxylin and eosin stained embryos displays subtle defects in the shape of the somitic border. Whole mount *in situ* hybridization with the mesodermal markers *Pax1* and *Mox1* show that epithelial spherical somites appear to be indistinctly separated from the presomitic mesoderm. As the somites mature, their spherical shape appears irregular in outline as well as size relative to their neighbors. Transverse sections of whole mount *in situs* do not show any significant reduction or alteration in the pattern of expression of the *Mox1* or *Pax1* genes.

Hybridizations with mesodermal markers such as *T*, *Dll1*, and *Shh* do not reveal any major abnormalities in pattern of expression in *pudgy* embryos. Taken collectively, the whole mount *in situ* hybridization findings do not show any major loss of somitic structures during any point of somitogenesis.

Significant intrauterine embryonic lethality is another key finding relating to the *pudgy* mutation. Examining the results from two separate intercrosses, we see a significant underrrepresentation of expected *pu/pu* homozygotes even at E9.5. Given that the mutant phenotype is first observable at E8, when somitogenesis begins (Grüneberg, 1961), we would expect that the important period of *pudgy* gene expression begins prior to E9.5. It is possible that the *pudgy* gene may have a role in developmental events prior to E8, such as mesodermal induction during gastrulation, given that there is a significant loss of *pu/pu* embryos at E9.5 and that defects in somitogenesis would not normally lead to rapid degeneration of embryos in the time between E8 and E9.5. Since we have not noted grossly deformed embryos of the 149 dissected at E9.5, the early embryonic lethality *pudgy* phenotype is not fully penetrant, such that embryos either die early in development or survive with somitic defects.

We have not noted any infanticide amongst the *pu/+* mothers, making is unlikely that *pu/pu* pups are selectively being killed and eaten. Once born, the *pudgy* pups appear to suckle normally and do not suffer from significant runting or growth defects, other than skeletal defects described. Examination of embryos at E8 at the beginning of somitogenesis or even E6 during gastrulation may yield further information as to the timing of the premature embryonic lethality.

In order to better understand the possible action of the *pudgy* gene, we can examine mutants with phenotypes similar to the *pudgy* phenotype. To

date, no report has been published of a targeted mutation whose phenotype is similar to *pudgy*. Of the spontaneously generated mutations, the *stub* mutant was reported in 1942 and showed a phenotype indistinguishable from *pudgy,* but unfortunately the stock was lost before allelism tests could be conducted with *pudgy*. More recently, an autosomal recessive mutation (nm1883) has been identified at the Jackson Laboratory which has an adult external appearance of the *pudgy* mutation, but tests for allelism have indicated that these two mutations complement and are not allelic (K. Johnson, personal communication).

The *Rib fusions* (*Rf*) semi-dominant mutation was reported by Mackeson and Steven in 1960 and displayed a remarkable phenotype in the homozygous affected animals. In homozygotes, there was a total absence of somite formation, resulting in extensive rib and vertebral fusion in the adult (Theiler and Stevens, 1960). The vertebral column was almost completely fused, and the homozygous state was lethal. In the heterozygote, there is a subtle fusion of ribs, with occasional vertebral fusions. The *Rf* mutation has remained unmapped, but the phenotype is evocative of a more severe allele of the *pudgy* locus.

What is the primary molecular defect of the *pudgy* mutation? Hypotheses consistent with the *pudgy* phenotype include the following:

• *Defective cell adhesion.* Selective adhesivity may play a crucial role in the transition from mesenchymal cell to tightly bound epithelial cells. Embryo dissociation experiments have demonstrated that somitic cells are selectively more adhesive than cells in the segmental plate (reviewed in Keynes and Stern, 1988). Blocking β-integrin or N-cadherin by monoclonal antibody injection in chick embryos disrupted somitic cells and sclerotome migration (Drake et al., 1992; Duband et al., 1987).

• *An extracellular matrix defect.* Although defects in extracellular matrix may be expected to produce abnormalities in many tissues, the somites may be particularly sensitive to inhomogeneities in the matrix. Heat shock experiments in *Xenopus* embryos produced altered patterns of fibronectin and laminin which disrupted normal somitic patterning (Danker et al., 1992). Prior to somite formation, the presomitic mesoderm is patterned in loose circular collections of mesenchyme, termed somitomeres, which may affect the distribution of extracellular matrix (Jacobson, 1988).

• *Defective cytoskeletal regulation.* In order to make the transition between vastly different cytoskeletal architectures such as those of mesenchyme and epithelial cells, a complex cellular machinery exists to carry signals from the receptor through the signal transduction cascade onwards to the numerous actin and microtubule binding proteins. Disruption of MacMARCKS gene, a substrate of protein kinase C which regulates the actin cytoskeleton, results in defects in the closure of the neural tube and eventually anencephaly and spina bifida (Chen et al., 1996; Karpova et al., 1993; Li and Aderem, 1992; Zhu et al., 1995). The closure of the neural tube involves major cytoskeletal rearrangements in epithelial neural ectoderm that is comparable to changes that occur in the somitic mesoderm.

Ultimately, the identification of the *pudgy* gene itself will be necessary for further investigation. Positional cloning methods will be most likely to identify the gene, and such efforts are currently in progress.

## ACKNOWLEDGMENTS

We would like to thank Dow-Chung Chi, Peggy Kolm, Laura Gammill, Ruth Curry, Jesse Dausmann, Laurie Jackson-Grusby, Paul Bain, Monique Bulotsky, David Wilcomb, Carolyne Dunbar, and Jeanne Rheis for their use of equipment, protocols, and technical assistance. Humphrey Gardner provided consultation for gross pathological examinations of *pudgy* mice. We also thank Arend Sidow, Martin de Angelis, and Achim Gossler for their helpful discussions.

## REFERENCES

Balling, R., Deutsch, U., and Gruss, P. (1988). undulated, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of Pax 1. Cell *55*, 531-5.

Bettenhausen, B., de Angelis, M. H., Simon, D., Guenet, J. L., and Gossler, A. (1995). Transient and restricted expression during mouse embryogenesis of Dll1, a murine gene closely related to Drosophila Delta. Development *121*, 2407-18.

Burgess, R., Cserjesi, P., Ligon, K. L., and Olson, E. N. (1995). Paraxis: a basic helix-loop-helix protein expressed in paraxial mesoderm and developing somites. Dev Biol *168*, 296-306.

Chalepakis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M., and Gruss, P. (1991). The molecular basis of the undulated/Pax-1 mutation. Cell *66*, 873-84.

Chen, J., Chang, S., Duncan, S. A., Okano, H. J., Fishell, G., and Aderem, A. (1996). Disruption of the MacMARCKS gene prevents cranial neural tube closure and results in anencephaly. Proc Natl Acad Sci U S A *93*, 6275-9.

Conlon, R. A., Reaume, A. G., and Rossant, J. (1995). Notch1 is required for the coordinate segmentation of somites. Development *121*, 1533-45.

Cserjesi, P., Brown, D., Ligon, K. L., Lyons, G. E., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., and Olson, E. N. (1995). Scleraxis: a basic helix-loop-helix protein that prefigures skeletal formation during mouse embryogenesis. Development *121*, 1099-110.

Danker, K., Hacke, H., and Wedlich, D. (1992). Effects of heat shock on the pattern of fibronectin and laminin during somitogenesis in Xenopus laevis. Dev Dyn *193*, 136-44.

del Amo, F. F., Smith, D. E., Swiatek, P. J., Gendron-Maguire, M., Greenspan, R. J., McMahon, A. P., and Gridley, T. (1992). Expression pattern of Motch, a mouse homolog of Drosophila Notch, suggests an important role in early postimplantation mouse development. Development *115*, 737-44.

Dickinson, M. E., and McMahon, A. P. (1992). The role of Wnt genes in vertebrate development. Curr Opin Genet Dev *2*, 562-6.

Doolittle, D. P., Davisson, M. T., Guidi, J. N., and Green, M. C. (1996). Catalog of mutant genes and polymorphic loci. In Genetic Variants and Strains of the Laboratory Mouse, M. F. Lyon, S. Rastan and S. D. M. Brown, eds. (Oxford: Oxford University Press), pp. 1807.

Drake, C. J., Davis, L. A., Hungerford, J. E., and Little, C. D. (1992). Perturbation of beta 1 integrin-mediated adhesions results in altered somite cell shape and behavior. Dev Biol *149*, 327-38.

Duband, J. L., Dufour, S., Hatta, K., Takeichi, M., Edelman, G. M., and Thiery, J. P. (1987). Adhesion molecules during somitogenesis in the avian embryo. J Cell Biol *104*, 1361-74.

Fan, C. M., and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. Cell *79*, 1175-86.

Grüneberg, H. (1961). Genetical studies on the skeleton of the mouse: XXIX. Pudgy. Genet. Res., Camb. *2*, 384-393.

Herrmann, B. G., and Kispert, A. (1994). The T genes in embryogenesis. Trends Genet *10*, 280-6.

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.

Hogan, B., Beddington, R., Constantini, F., and Lacy, E. (1994). Manipulating the Mouse Embryo:A Laboratory Manual, 2nd Edition ( Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Hogan, B. L. (1996). Bone morphogenetic proteins: multifunctional regulators of vertebrate development. Genes Dev *10*, 1580-94.

Jacobson, A. G. (1988). Somitomeres: mesodermal segments of vertebrate embryos. Development *104*, 209-20.

Johnson, R. L., Laufer, E., Riddle, R. D., and Tabin, C. (1994). Ectopic expression of Sonic hedgehog alters dorsal-ventral patterning of somites. Cell *79*, 1165-73.

Karpova, T. S., Lepetit, M. M., and Cooper, J. A. (1993). Mutations that enhance the cap2 null mutant phenotype in Saccharomyces cerevisiae affect the actin cytoskeleton, morphogenesis and pattern of growth. Genetics *135*, 693-709.

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

Keynes, R. J., and Stern, C. D. (1988). Mechanisms of vertebrate segmentation. Development *103*, 413-29.

Kingsley, D. M. (1994). What do BMPs do in mammals? Clues from the mouse short-ear mutation. Trends Genet *10*, 16-21.

Krumlauf, R. (1993). Mouse Hox genetic functions. Curr Opin Genet Dev *3*, 621-5.

Kusumi, K., Frankel, W. N., Sun, E. S., Birren, B., Chi, D. C., Hawkins, T. L., Reeve-Daly, M. P., and Lander, E. S. (1997). A candidate gene map of the mouse *pudgy* region: Integration of genetic, physical, transcriptional, expression and genomic sequence maps: Massachusetts Institute of Technology.

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.

Lardelli, M., and Lendahl, U. (1993). Motch A and motch B--two mouse Notch homologues coexpressed in a wide variety of tissues. Exp Cell Res *204*, 364-72.

Li, J., and Aderem, A. (1992). MacMARCKS, a novel member of the MARCKS family of protein kinase C substrates. Cell *70*, 791-801.

Noll, M. (1993). Evolution and role of Pax genes. Curr Opin Genet Dev *3*, 595-605.

Reaume, A. G., Conlon, R. A., Zirngibl, R., Yamaguchi, T. P., and Rossant, J. (1992). Expression analysis of a Notch homologue in the mouse embryo. Dev Biol *154*, 377-87.

Schulte-Merker, S., van Eeden, F. J., Halpern, M. E., Kimmel, C. B., and Nusslein-Volhard, C. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. Development *120*, 1009-15.

Swiatek, P. J., Lindsell, C. E., del Amo, F. F., Weinmaster, G., and Gridley, T. (1994). Notch1 is essential for postimplantation development in mice. Genes Dev *8*, 707-19.

Theiler, K., and Stevens, L. C. (1960). The development of Rib fusions, a mutation in the house mouse. Am. J. Anat. *106*, 171-184.

Wallin, J., Wilting, J., Koseki, H., Fritsch, R., Christ, B., and Balling, R. (1994). The role of Pax-1 in axial skeleton development. Development *120*, 1109-21.

Wilkinson, D. G. (1992). *In situ* hybridization: A practical approach (Oxford: Oxford University Press).

Yamaguchi, T. P., and Rossant, J. (1995). Fibroblast growth factors in mammalian development. Curr Opin Genet Dev *5*, 485-91.

Zhu, Z., Bao, Z., and Li, J. (1995). MacMARCKS mutation blocks macrophage phagocytosis of zymosan. J Biol Chem *270*, 17652-5.

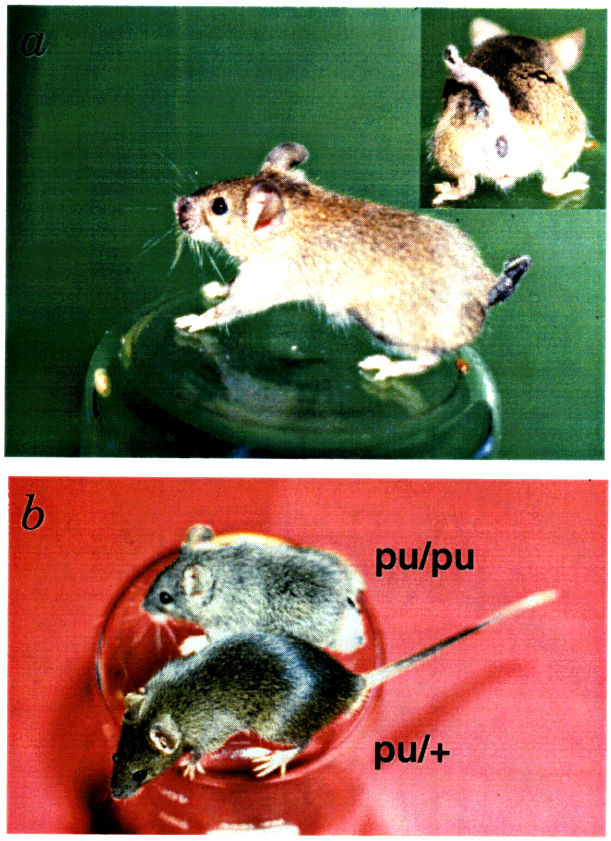



**Figure 1** External appearance of a *pudgy* mutant mouse *a*. Side and rear views of a typical *pudgy* mouse. The torso is shortened, and the tail is hardened and crumpled. *b*. Comparison of a *pudgy* mouse to a heterozygous wild-type sibling. There is no tail visible in this mutant mouse, and the torso is shortened and wider, as compared to its sib.

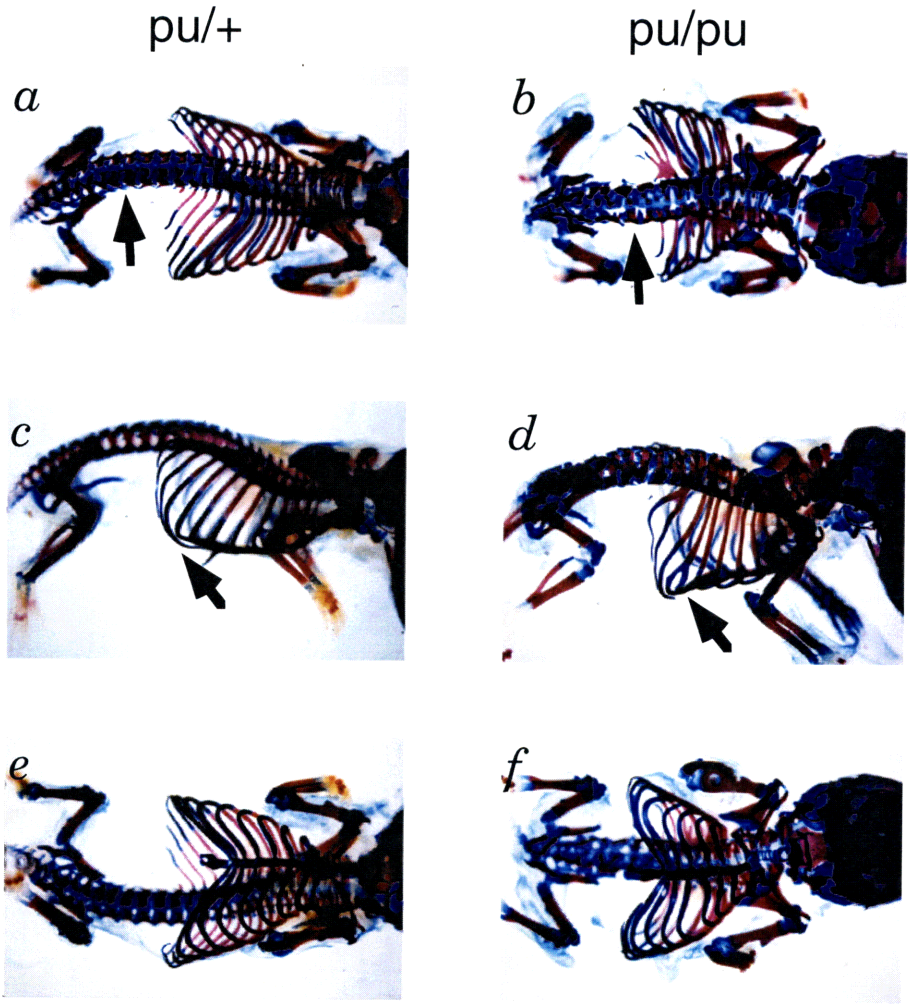



**Figure 2** Skeletal preparations of neonatal *pu* / + (*a,c,e*) and *pu* / *pu* (*b,d,f*) mice. *a, b* are from a dorsal view. Limbs appear normal in the mutant, but lumbar vertebrae and ribs are highly deformed. *c, d* are from a lateral view. Bifurcations and fusions of ribs are evident in the mutant. *e, f* are from a ventral view. In the mutant, the thoracic cage is wider and shorter, with fusion of the normal 13 pairs of ribs.

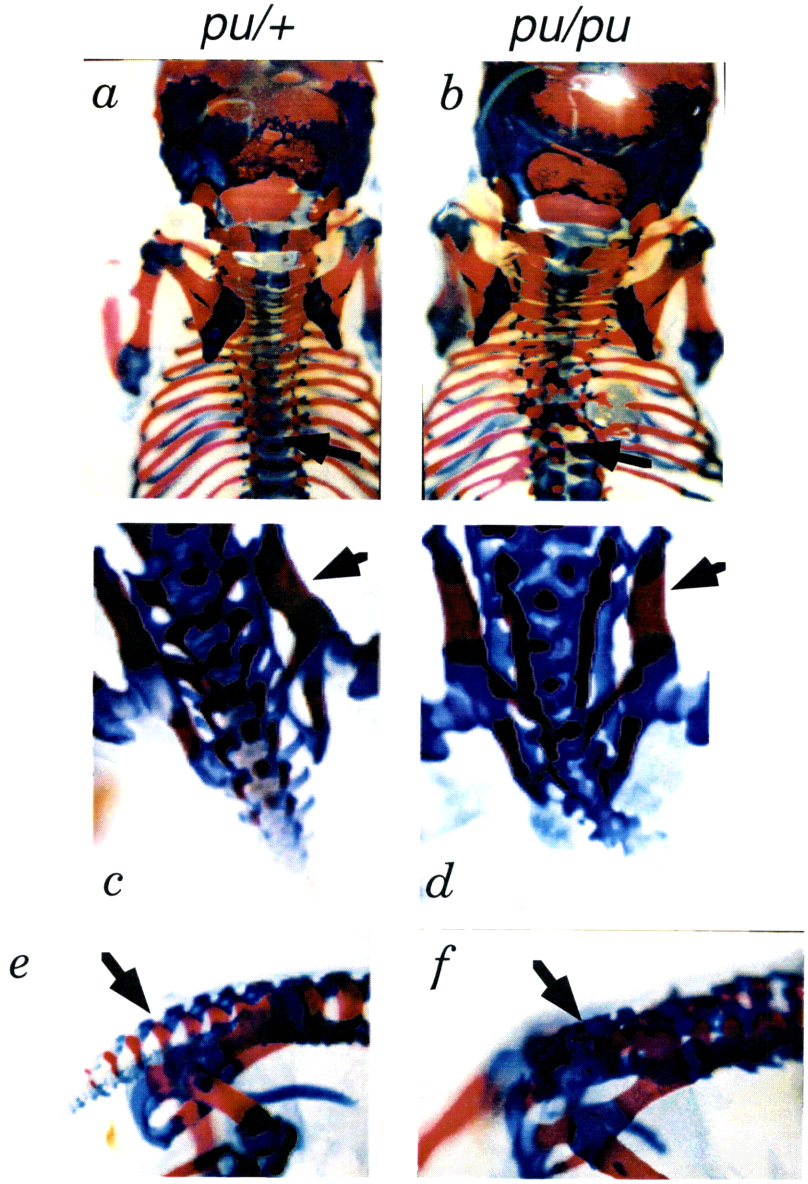



**Figure 3** Vertebrae from *pu* / + (*a,c,e*) and *pu* / *pu* (*b,d,f*) mice. *a,b* Cervical and thoracic vertebrae. Unlike the regular vertebrae of the wild type, the *pudgy* mutant thoracic vertebrae have irregularly spaced neural arches and spinous processes. *c,d* Lumbar and sacral vertebrae. Irregular vertebrae also result in pelvic girdle deformations in the mutant. *e,f* Lateral view of sacral and caudal vertebrae. In the mutant, vertebrae are indistinguishable and the vertebral column ends at the pelvic girdle.

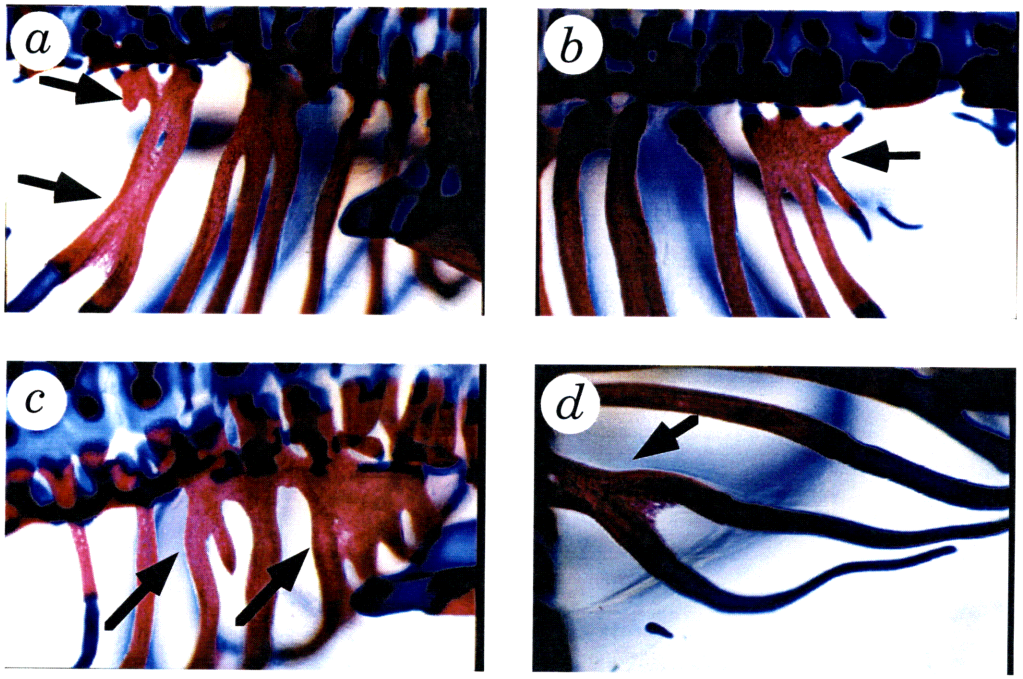



**Figure 4** Deformities in *pudgy* mutant ribs. Several types of abnormalities can be seen in *pudgy* mutant ribs including fusions of two, three, or even four ribs in *a,b,c*, and bifurcations in *c,d*. In *a*, truncation of the last rib can be seen.

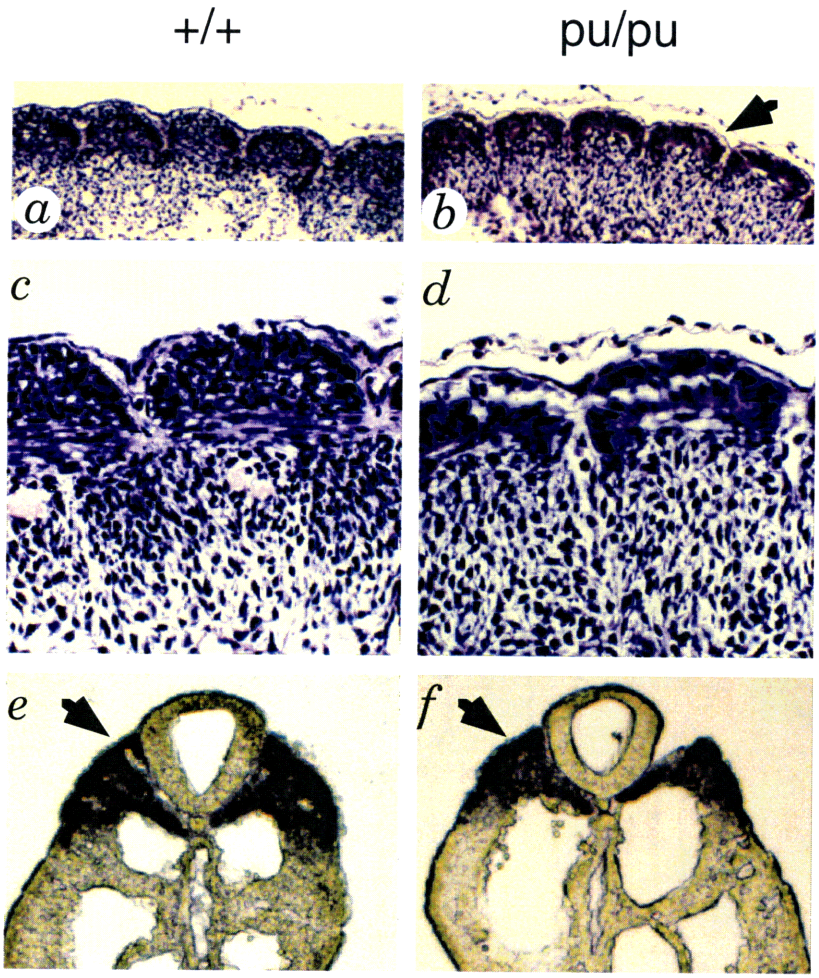



**Figure 5** Irregular borders of E9.5 *pu/pu* somites. *a,c* are E9.5 +/+ BALB/cJ wild type embryos. *b,d,f* are E.95 *pu/pu* mutant embryos. *f* is from a E9.5 *pu/+* wild type embryo. *a,b* 100x magnification of somite sections, stained with hematoxylin and eosin. Slightly irregular somitic borders can be seen in the *pu/pu* embryos, as compared to the circular borders of wild type somites. *c,d* 1000x magnification somite sections, hematoxylin and eosin stain. *e,f* Sections of embryos assayed by whole mount *in situ* hybridization with *Mox1* probe. The area of *Mox1* expression is decreased in mutant embryos.

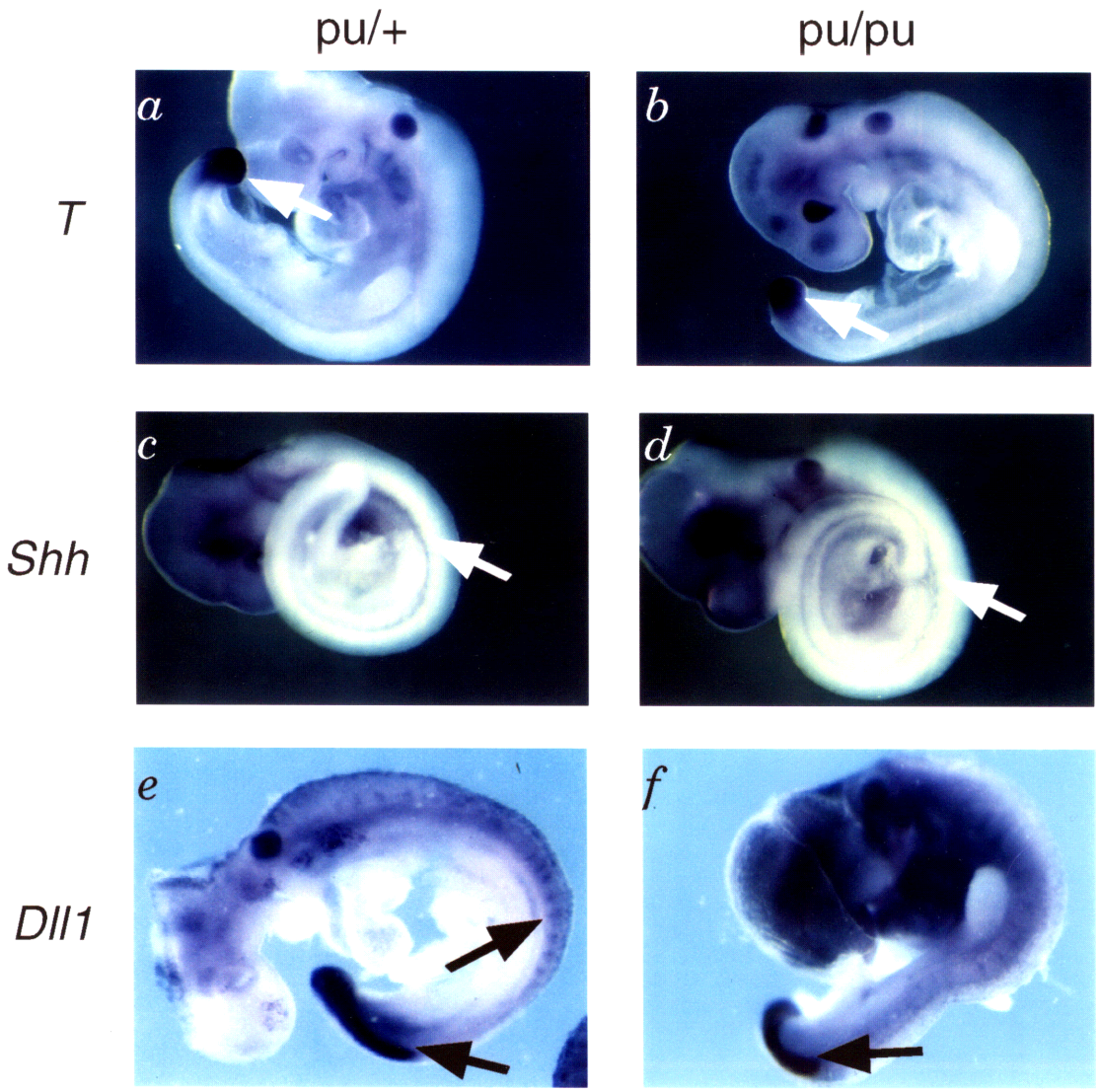



**Figure 6** Normal notochord and presomitic mesoderm *in situ* hybridization in wild-type *pu*/+ (*a,c,e*) and mutant *pu*/*pu* (*b,d,f*) E9.5 embryos. *a,b Brachyury (T)* probe shows expression in the tailbud in both mutant and wild type. *c,d Sonic hedgehog (Shh)* probe displays expression all along the uninterrupted notochord in the mutant. *e,f Delta-like-1 (Dll1)* probe shows expression primarily in the presomitic mesoderm, with weak expression in the caudal somite.

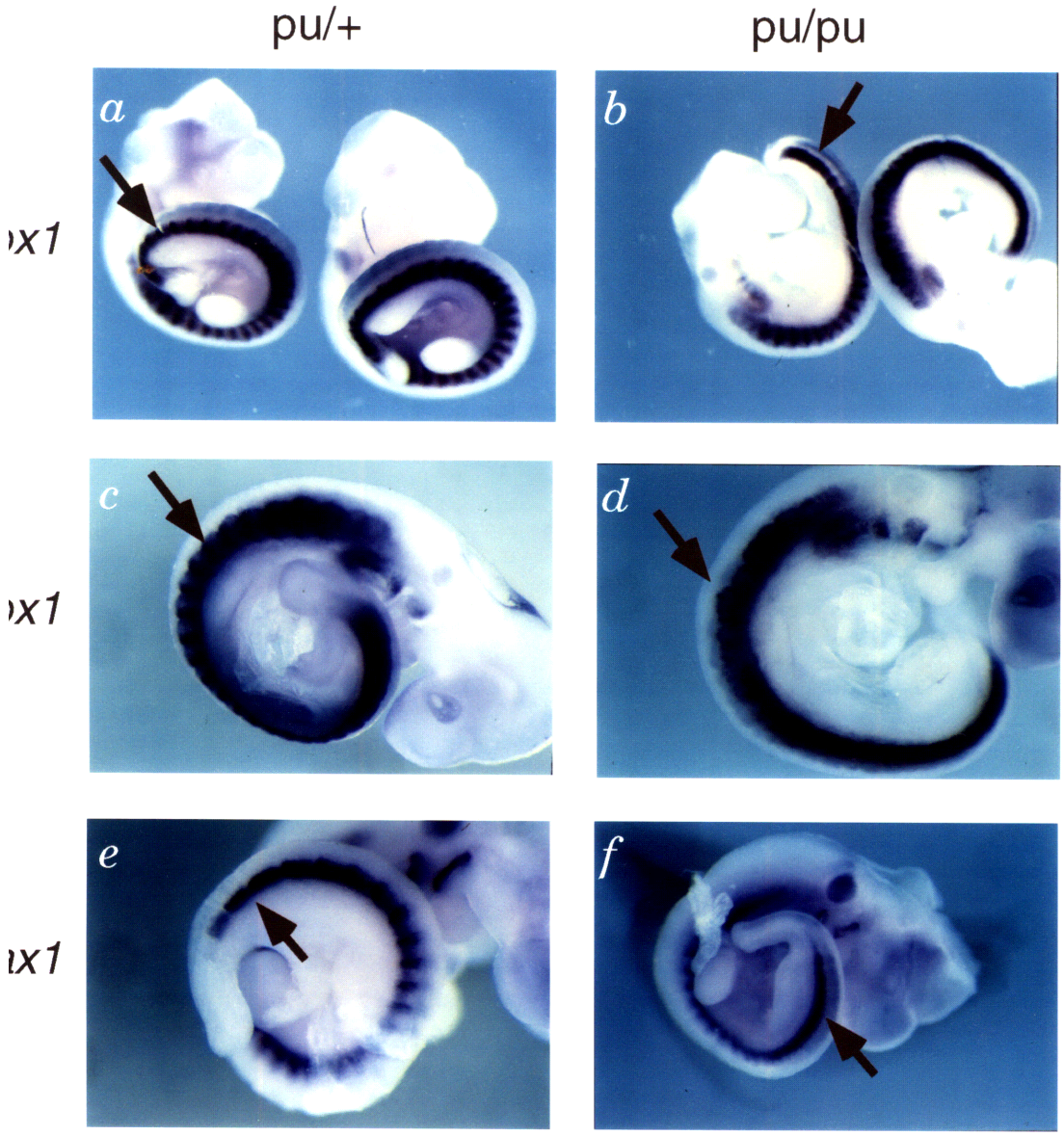



**Figure 7** Irregular somitic segmentation in *pu/pu* mutant (*b,d,e*) as compared to *pu/+* wild-type (*a,c,e*) E9.5 embryos. *a,b,c,d Mox1* probe shows expression in the entire somite, and highlights the indistinct early somitic invaginations in the mutant in *b* and the irregular somitic borders in later somites in *d*. *e,f Pax1* probe displays expression restricted to the sclerotome, and also shows the lack of clear somitic segmentation in early mutant somites, as compared to wild type embryos.

## Chapter Three

# The Construction of a Large-Insert Mouse Yeast Artificial Chromosome Library for Positional Cloning Efforts

Kenro Kusumi,*† Jennifer S. Lanni*, Julia A. Segrè,*†
David S. Koos,‡ and Eric S. Lander*†

* Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, 02142; †Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139; and ‡Howard Hughes Medical Institute and Department of Molecular Biology, Princeton University, Princeton, New Jersey, 08544

## INTRODUCTION

The mammalian genome is relatively large: 3 billion base pairs versus only 12 million for the unicellular eukaryote *Saccharomyces cerevisiae* (Sherman, 1991). This enormous relative size of the mammalian genome is thought to be due to an increase in gene number as well as the spreading out of genes over a larger genomic region. Unlike organisms such as the nematode *Caenorhabditis elegans* for which multicopy plasmids (15-20 kb upper limit insert size) and cosmids (47 kb maximum insert size) sufficed for chromosomal walks and transgenic experiments, larger clones are needed for mammalian studies (Sambrook et al., 1989)

In positional cloning, candidates genes are localized by physical proximity to genetic markers. In the case of the mouse, over 6,000 simple sequence length polymorphisms have been collected, yielding on average one marker every 0.3 cM (Dietrich et al., 1996). With the approximation of 1 cM equal to 2

Mbp in the mouse, this yields on average of one marker every 600 kb, a distance not easily spanned in smaller-insert clones like P1s and cosmids.

For transgenic studies where entire genes are inserted *de novo* into the genome, it is essential to include the entire gene, including regulatory regions, within the clone used. The size of mammalian genes ranges from 5S tRNAs contained within 120 bp to mega-genes such as dystrophin, which is mutated in the clinical syndromes of Becker's and Duchenne's muscular dystrophy and is spread out amongst 2.4 million base pairs (Den Dunnen et al., 1992; Watson et al., 1987).

The necessity to work with large, genomic regions of mammalian DNA has driven the challenge to introduce larger and larger inserts into DNA clones. A major breakthrough in this effort was the development of larger DNA clones by yeast biologists. Yeast chromosomes are linear, have defined chromosomal elements, and are relatively large, from 200 to 2,200 kb (Sherman, 1991). In 1983, J.W. Szostack and colleagues demonstrated that it was possible to create a new type of artificial recombinant chromosome in yeast (Murray and Szostak, 1983). In 1987, M. Olsen and colleagues reported a genomic library made in an artificial chromosome, or YAC, or Yeast Artificial Chromosome (Burke et al., 1987). The YAC vector contained the necessary elements of a yeast chromosome: a CEN sequence to act as a centromere, and ARS (autonomous replication sequence) to activate the DNA replication machinery of the host, and TEL sequences to serve as telomeres in the host. In addition, the YACs featured a functional *TRP1* gene on one arm and a *URA3* gene on the other to serve as selectable positive markers in the *trp*$^-$*ura*$^-$ host, as well as a cloning site within a *SUP4* ochre suppressor in the vector to serve as an inactivatable marker signaling the presence of insert, as assayed by numerous phenotypic markers including suppression of *ade2* phenotype of the host.

Although the initial sizes of inserts reported was only 100-200 kb (Burke et al., 1987), later efforts have pushed the insert sizes possible beyond the 1 million base-pair mark (Chumakov et al., 1992; Chumakov et al., 1995).

## CONSTRUCTING A LARGE-INSERT MOUSE YAC LIBRARY

In 1991, only three mouse YAC libraries were available, none suitable for positional cloning due to small insert size or low coverage a 2.2-fold coverage library with average insert size 265 kb (Burke et al., 1987), a 3.5-fold coverage library with average insert size 240 kb (Chartier et al., 1992), and a 3-fold coverage library with average insert size 700 kb which was not readily available for public use and provided insufficient genomic coverage (Larin et al., 1991). Accordingly, we set out to construct a large-insert YAC library suitable for positional cloning work. Our new YAC library provides a total of 4.3-fold coverage in two parts: part A providing 0.7-fold coverage in clones of average size 480 kb and part B providing 3.6-fold coverage in clones of average size 680 kb. Together, the four available libraries now provide 13-fold coverage of the mouse genome in YAC clones.

The YAC library was prepared using genomic DNA from C57BL/6J female mice, essentially as described previously (Foote et al., 1992) and detailed in (Foote, 1994). Genomic DNA was prepared from kidney nuclei as described previously (Strauss et al., 1992) and partially digested to the desired molecular weight by *EcoR* I restriction endonuclease. The extent of digestion was regulated by competition with *EcoR* I methylase. The restriction-digested DNA was subjected to two rounds of size fractionation by pulsed-field gel electrophoresis. In the first round, DNA fragments greater than 460 kb were selected and then ligated to vector arms from the plasmid pYAC4 (Burke et al., 1987). In the second round, the ligated products were size-fractionated greater

than 460 kb (for part A) or 600 kb (for part B). The resulting size-selected ligation mixture was then transformed into the yeast host strain AB1380 (*MAT*a, *ura3-52, trp1, can1-100, lys2-1, his5, ade2-1, thr*$^-$, $\Psi^+$). The pYAC4 vector carries the markers URA3 and TRP1, but the TRP1 gene lacks a complete promoter element, and thus clones only weakly expressed the TRP1 gene. Therefore, transformants were initially selected for a Ura$^+$ phenotype and then screened for a Ura$^+$Trp$^+$ phenotype. In addition, the pYAC4 cloning site lies within the SUP4 suppressor which acts on the host *ade2-1* color marker, so only red colonies which indicate an insertion were selected. Colonies were not streaked to single clones before inclusion in the library, and it is possible that a microtiter well may contain more than one YAC. Clones were transferred to standard 96-well microtiter plates containing AHC yeast minimal medium and grown to saturation at 30°C. To make frozen stocks, YPD yeast complex medium, AHC medium, and glycerol were added to saturated cultures to make a final concentration of 15% glycerol, 0.1x YPD, and 1.35x AHC. Frozen stocks were maintained at -85°C. In total, the preparation of the library required 17 ligations and 56 separate transformations.

Essential quality control procedures were used throughout library construction. To confirm the presence of mouse DNA in the transformants, a random sample of 5% of the transformants from each ligation was screened using a PCR assay detecting the presence of the B2-repeat element, which is found at an average spacing of 30 kb in the mouse genome (Bennett et al., 1984). (The primers used were 5'-CCGAATTCGATGGCTCAGTGGTTAAGAG-3' and 5'-CCGAATTCCACTGTAGCTGTCTTCAGAC-3' where the underlined sequences represent non-B2 sequence; PCR products were analyzed using agarose gel

electrophoresis and visualized by ethidium bromide.) To measure the insert size of the YACs, a random sample of 1% of the transformants from each ligation was examined by pulsed-field gel electrophoresis. Specifically, DNA was prepared from spheroplasted yeast essentially as described by Carle and Olson (Carle and Olson, 1985) and Schwartz and Cantor (Schwartz and Cantor, 1984), subjected to pulsed-field gel electrophoresis in 1% Seaplaque GTG agarose (FMC Bioproducts, Rockland, ME), transferred to nylon membrane, and hybridized with probes for mouse repeats B1, B2, and L1 to visualize the location of the YAC band. Of approximately 4,100 clones in Part I, we examined 68 clones and estimated a mean size of 480 kb and a median size of 440 kb. Of approximately 15,840 clones in part II, we have examined 218 and estimated a mean size of 680 kb and a median size of 640 kb as shown in Figure 1. Based on an estimated genome length of 3 billion bp, the library provides a total of 4.3-fold coverage of the haploid mouse genome.

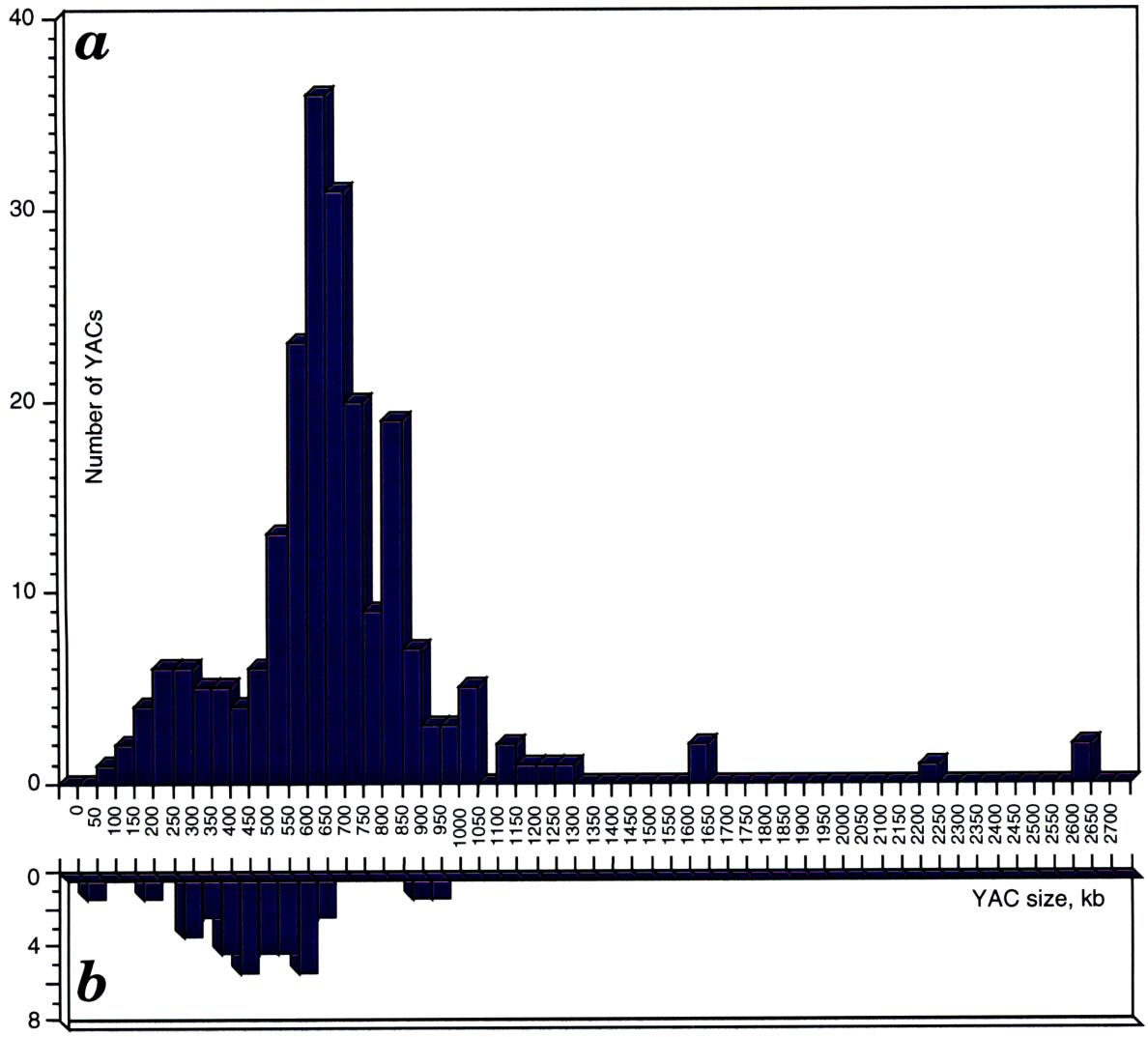



**Figure 1** Mouse YAC Library
*a* A histogram showing the molecular weights of 218 randomly selected YAC clones examined, from 15,840 clones in part II of the YAC library.
*b* A histogram showing the molecular weights of 68 randomly selected YAC clones examined, from 4,100 in part I of the YAC library

The YAC library has been prepared for PCR-based screening (Green and Olson, 1990), using a two-level pooling scheme based on screening of "super-pools" and "sub-pools". Specifically, the library has been partitioned into 26 blocks, each consisting of 8 microtiter plates. From each block, we prepared a single DNA super-pool (containing all 8 x 96 = 768 clones in the block) and 28 sub-pools (with 8 consisting of those clones in the block contained in the same microtiter plate, 8 consisting of those clones in the same row, 12 consisting of those clones in the same column). There is thus a total of 26 super-pools and 26 x 28 = 728 sub-pools. In preparing the pools, we grew each of the roughly 20,000 clones separately to avoid competition and subsequently pooled them before DNA preparation.

To screen the library to find clones containing a sequence-tagged site (STS) of interest, one first determines the positive super-pools and then the appropriate sub-pools. If a single plate, row and column sub-pool within a block are positive, the location of the clone is uniquely specified. If multiple plates, rows or columns are found to be positive, then there is some limited degeneracy in the clone address. (Specifically, the number of addresses to expect within a block would equal the number of plate sub-pool positives x row sub-pool positives x column sub-pool positives.)

The YAC library is provided without license fee to any companies willing to provide services (such as selling PCR pools or carrying out STS screening) at affordable prices to the scientific community. Currently, several commercial distributors offer these services and include:

- Research Genetics, Inc. (Huntsville, AL)
- Clontech Laboratories, Inc. (Palo Alto, CA)
- Genome Systems, Inc. (St. Louis, MO)

## DISCUSSION

The availability of large-insert DNA clones are essential for the production of physical maps used in positional cloning efforts. Indeed, our YAC library was the first large-insert mouse YAC library publicly available and has greatly facilitated positional cloning. It was critically used in the mapping and cloning of the following mutations.

| ***Symbol*** | ***Mutation*** | ***Gene Name*** | ***Gene Function*** | ***Reference(s)*** |
|---|---|---|---|---|
| *bg* | beige | Lyst | lysosomal function | (Barbosa et al., 1996) |
| *db* | diabetic | leptin receptor | receives satiety signal | (Chua et al., 1996; Tartaglia et al., 1995) (Lee et al., 1996) |
| *eed* | embryonic ectoderm development | 17Rn5 | gastrulation | (Holdener et al., 1995) |
| *fu* | fused | N.A. | RGS domain | (Perry et al., 1995; Sutherland et al., 1995) |
| *Hst1* | hybrid sterility 1 | N.A. | N.A. | (Trachtulec et al., 1994) |
| *Mtm1* | X-linked myotubular myopathy | N.A. | N.A. | (Degouyon et al., 1996) |
| *nu* | nude | winged-helix nude | transcription factor | (Nehls et al., 1994; Segre et al., 1995) (Kurooka et al., 1996) |
| *ob* | obese | leptin | satiety regulation | (Zhang et al., 1994) |
| *rl* | reeler | reelin | extracellular matrix protein | (Montgomery et al., 1994) (D'Arcangelo et al., 1995; Hirotsune et al., 1995) |
| *scid* | severe combined immunodeficiency syndrome | DNA-PKcs | protein kinase | (Miller et al., 1995) |
| *sg* | staggerer | RORα | nuclear receptor | (Hamilton et al., 1996) |
| *sh1* | shaker 1 | myosin type VII | auditory transduction | (Gibson et al., 1995) |
| *tg* | tottering | $\alpha_{1A}$ | $Ca^{2+}$ channel | (Fletcher et al., 1996) |
| *wr* | wobbler | N.A. | N.A. | (Wedemeyer et al., 1996) |
| *XIST* | XIST | X-inactivation RNA | chromosomal inactivation | (Lee et al., 1996) |

Subsequent to the release of this library, other mouse genomic YAC libraries have been reported and made publicly available, bringing the total genomic coverage to over 20 fold (Haldi et al., 1996; Larin et al., 1991).

What role do YACs have in future genomic research, and what directions should YAC engineering efforts take? The availability of other smaller-insert vector system clones such as BACs (Bacterial Artificial Chromosomes) and P1s have provided complementary cloning resources but there is still a clear role and room for improvement of large-insert YACs. Below, these issues are addressed:

1.) *Towards greater genomic fidelity*

Several deficiencies prevent YACs from being an ideal DNA clone for accurately representing the mammalian genome. As a host, yeast permits large chromosomal elements, but yeast is also an efficient host for recombination. Active recombination in the insert DNA can lead to numerous problems including:

- deletions—Repeats common in human and mouse DNA (LINE and SINE elements) can serve as sites for recombination, leading to internal deletions. This can make certain regions of the genome very difficult to maintain stably in yeast.
- chimeras—During the vector-insert ligation reaction for YAC construction, ligated products with only one type of selectable marker can be formed. Normally, these one-marker YACs are not viable due to selection for $trp^+ura^+$ inserts in the $trp^-ura^-$ host. However, co-transformation of the yeast host by both *trp* only YACs and *ura* only YACs will select for recombination between different constructs. The recombination products will be chimeric for different regions in the genome.

- DNA isolation—Whereas bacterial plasmids and lambda-based vectors enable specific DNA purification due features in their propagation, YAC DNA is indistinguishable from host yeast chromosomes in a general purification. Size exclusion purifications based on pulsed-field gel methods is complicated if the size of the YAC is similar to that of a host chromosome. In addition, the process to achieve high-purity DNA is more complicated.

These problems have been addressed to a limited extent by efforts to alter the yeast host strain used in YAC library construction. As recombination deficient strains of bacterial hosts are used with plasmids, several recombination specific loci including *rad1*and *rad52* have been tested for their effects on recombination based problems in YACs. A mouse library has been constructed in a *rad52*$^-$ strain (Chartier et al., 1992), and examination of megabase-pair sized human YACs in a *rad52*$^-$ host shows a significantly reduced rate of chimerism (Haldi et al., 1994).

2.) *Larger YACs?*

Large numbers of genetic markers and STS markers have been generated in both the human and mouse, currently over 6,000 in the mouse and 10,000 in the human (Dietrich et al., 1996; Schuler, 1996). This results in a marker density of one every 500 kb or less, which falls within the insert size of the current generation of large-insert YACs. This makes YACs larger than 1 Mbp unnecessary for positional cloning efforts or even attempts to make contiguous physical maps of the genome. While the production of even larger YAC clones (>1Mbp) might be useful, in reality the problems of deletions and chimerism limit this usefulness. At this point, the production of large-insert YACs from

strains other than C57BL/6J in the mouse would represent a more valuable effort for the scientific community.

3.) *Improved transgenic reagents*

Increasingly, transgenic experiments have shown the limitations of cosmid-sized clones which do not contain the entire genomic regions necessary for proper tissue-limited expression. Indeed, YACs have been necessary for the study of *Pax6* and in the study of the large X-inactivation region (Lee et al., 1996).

Most current YAC libraries have not been constructed with transgenesis in mind. As a result, extensive reengineering of the YAC vector arms to add cell selectable markers is required for each effort, adding time for each study. Several methods have been published in order to introduce eukaryotic selectable markers such as neomycin using targeting vectors (Emanuel et al., 1995; Fairhead et al., 1995; Lee et al., 1996; Riley et al., 1992; Srivastava and Schlessinger, 1991). Future vector systems could include 2 or more eukaryotic markers on the two arms to aid tracking YACs during transgenesis.

4.) *YAC reagents for other organisms*

While YAC library resources have been developed for systems such as the human and mouse, such reagents are only now being reported for agriculturally important systems such as cattle, swine, chicken, barley, *Arabidopsis*, and rice (Creusot et al., 1995; Eyers et al., 1992; Grill and Somerville, 1991; Kleine et al., 1993; Leeb et al., 1995; Libert et al., 1993). Many agricultural organisms have yet to have any DNA clone resources developed. As the need for increased crop yield and disease resistance increases with the growing human population, transgenic techniques will be essential for transferring

desired alleles into the genetic stock, saving the many years it takes to genetically cross in such traits without losing other alleles. Towards this end, continued refinement of YAC techniques to improve clone stability and ease of library construction would be useful. If YAC libraries were available for tens or hundreds of different agricultural species, the number of new transgenic strains could increase to meet the varied conditions worldwide.

In summary, large-insert YAC resources are an essential component of positional cloning efforts in mammals. We have constructed the first publicly distributed, large-insert, mouse YAC library which has been essential in the cloning of numerous mouse mutations. Large-insert YAC clones will continue to find use in transgenesis experiments and physical mapping in mouse and other model organisms, e.g. the zebrafish, where genetic resources are currently being developed.

## ACKNOWLEDGMENTS

We would like to thank Adrian Argento, Hillary Katz, Asaf Presente, Nga Tang, and Marta Velez-Stringer for valuable technical assistance. In addition, we are grateful to Simon Foote, Adrienne Hilton, Maryann Haldi, Armand MacMurray, and Alix Weaver for their instruction and protocols.

## REFERENCES

Barbosa, M. D. F. S., Nguyen, Q. A., Tchernev, V. T., Ashley, J. A., Detter, J. C., Blaydes, S. M., Brandt, S. J., Chotai, D., Hodgman, C., Solari, R. C. E., Lovett, M., and Kingsmore, S. F. (1996). Identification of the homologous beige and Chediak-Higashi-Syndrome genes. Nature *382*, 262-265.

Bennett, K. L., Hill, R. E., Pietras, D. F., Woodworth-Gutal, M., Kane-Haas, C., Houston, J. M., Heath, J. K., and Hastie, N. D. (1984). Most highly repeated dispersed DNA families in the mouse genome. Mol Cell Biol *4:*, 1561-1571.

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.

Carle, G. F., and Olson, M. V. (1985). An electrophoretic karyotype for yeast. Proc Natl Acad Sci USA *82*, 3756-3760.

Chartier, F. L., Keer, J. T., Sutcliffe, M. J., Henriques, D. A., Mileham, P., and Brown, S. D. (1992). Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. Nat Genet *1*, 132-6.

Chua, S. C., Jr., Chung, W. K., Wu-Peng, X. S., Zhang, Y., Liu, S. M., Tartaglia, L., and Leibel, R. L. (1996). Phenotypes of mouse diabetes and rat fatty due to mutations in the OB (leptin) receptor. Science *271*, 994-6.

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.

Chumakov, I. M., Rigault, P., Le Gall, I., Bellanne-Chantelot, C., Billault, A., Guillou, S., Soularue, P., Guasconi, G., Poullier, E., Gros, I., and et al. (1995). A YAC contig map of the human genome. Nature *377*, 175-297.

Creusot, F., Fouilloux, E., Dron, M., Lafleuriel, J., Picard, G., Billault, A., Le Paslier, D., Cohen, D., Chaboue, M. E., Durr, A., and et al. (1995). The CIC library: a large insert YAC library for genome mapping in Arabidopsis thaliana. Plant J *8*, 763-70.

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.

Degouyon, B., Chatterjee, A., Monaco, A., Quaderi, N., Brown, S. D. M., and Herman, G. E. (1996). Comparative Mapping On The Mouse X-Chromosome Defines A Myotubular Myopathy Equivalent Region. Mamm Genome *7*, 575-579.

Den Dunnen, J. T., Grootscholten, P. M., Dauwerse, J. G., Walker, A. P., Monaco, A. P., Butler, R., Anand, R., Coffey, A. J., Bentley, D. R., Steensma, H. Y., and et al. (1992). Reconstruction of the 2.4 Mb human DMD-gene by homologous YAC recombination. Hum Mol Genet *1*, 19-28.

Dietrich, W. F., Miller, J., Steen, R., Merchant, M. A., Damronboles, D., Husain, Z., Dredge, R., Daly, M. J., Ingalls, K. A., Oconnor, 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., Page, D. C., and Lander, E. S. (1996). A comprehensive genetic map of the mouse genome. Nature *380*, 149-152.

Emanuel, S. L., Cook, J. R., O'Rear, J., Rothstein, R., and Pestka, S. (1995). New vectors for manipulation and selection of functional yeast artificial chromosomes (YACs) containing human DNA inserts. Gene *155*, 167-74.

Eyers, M., Edwards, K., and Schuch, W. (1992). Construction and characterisation of a yeast artificial chromosome library containing two haploid Beta vulgaris L. genome equivalents. Gene *121*, 195-201.

Fairhead, C., Heard, E., Arnaud, D., Avner, P., and Dujon, B. (1995). Insertion of unique sites into YAC arms for rapid physical analysis following YAC transfer into mammalian cells. Nucleic Acids Res *23*, 4011-2.

Fletcher, C. F., Lutz, C. M., O'Sullivan, T. N., Shaughnessy, J. D. J., Hawkes, R., Frankel, W. N., Copeland, N. G., and Jenkins, N. A. (1996). Absence epilepsy in tottering mutant mice is associated with calcium channel defects. Cell *87*, 607-617.

Foote, S. (1994). Large-insert YAC cloning. In *Current Protocols in Human Genetics*, S. Bonitz, ed. (New York: John Wiley & Sons).

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.

Gibson, F., Walsh, J., Mburu, P., Varela, A., Brown, K. A., Antonio, M., Beisel, K. W., Steel, K. P., and Brown, S. D. (1995). A type VII myosin encoded by the mouse deafness gene shaker-1. Nature *374*, 62-4.

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.

Grill, E., and Somerville, C. (1991). Construction and characterization of a yeast artificial chromosome library of Arabidopsis which is suitable for chromosome walking. Mol Gen Genet *226*, 484-90.

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.

Haldi, M. L., Strickland, C., Lim, P., Vanberkel, V., Chen, X. N., Noya, D., Korenberg, J. R., Husain, Z., Miller, J., and Lander, E. S. (1996). A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome. Mamm Genome *7*, 767-769.

Hamilton, B. A., Frankel, W. N., Kerrebrock, A. W., Hawkins, T. L., FitzHugh, W., Kusumi, K., Russell, L. B., Mueller, K. L., van Berkel, V., Birren, B. W., Kruglyak, L., and Lander, E. S. (1996). Disruption of the nuclear receptor hormone receptor RORα in *staggerer* mice. Nature *379*, 736-739.

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.

Holdener, B. C., Thomas, J. W., Schumacher, A., Potter, M. D., Rinchik, E. M., Sharan, S. K., and Magnuson, T. (1995). Physical localization of eed: a region of mouse chromosome 7 required for gastrulation. Genomics *27*, 447-56.

Kleine, M., Michalek, W., Graner, A., Herrmann, R. G., and Jung, C. (1993). Construction of a barley (Hordeum vulgare L.) YAC library and isolation of a Hor1-specific clone. Mol Gen Genet *240*, 265-72.

Kurooka, H., Segre, J. A., Hirano, Y., Nemhauser, J. L., Nishimura, H., Yoneda, K., Lander, E. S., and Honjo, T. (1996). Rescue of the hairless phenotype in nude-mice by transgenic insertion of the wild-type Hfh11 genomic locus. Int Immunology *8*, 961-966.

Larin, Z., Monaco, A. P., and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc Natl Acad Sci U S A *88*, 4123-7.

Lee, G. H., Proenca, R., Montez, J. M., Carroll, K. M., Darvishzadeh, J. G., Lee, J. I., and Friedman, J. M. (1996). Abnormal splicing of the leptin receptor in diabetic mice. Nature *379*, 632-5.

Lee, J. T., Strauss, W. M., Dausman, J. A., and Jaenisch, R. (1996). A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell *86*, 83-94.

Leeb, T., Rettenberger, G., Hameister, H., Brem, G., and Brenig, B. (1995). Construction of a porcine YAC library and mapping of the cardiac muscle ryanodine receptor gene to chromosome 14q22-q23. Mamm Genome *6*, 37-41.

Libert, F., Lefort, A., Okimoto, R., Womack, J., and Georges, M. (1993). Construction of a bovine genomic library of large yeast artificial chromosome clones. Genomics *18*, 270-6.

Miller, R. D., Hogg, J., Ozaki, J. H., Gell, D., Jackson, S. P., and Riblet, R. (1995). Gene for the catalytic subunit of mouse DNA-dependent protein kinase maps to the scid locus. Proc Natl Acad Sci U S A *92*, 10792-5.

Montgomery, J. C., Guarnieri, M. H., Tartaglia, K. E., and Flaherty, L. A. (1994). High-resolution genetic map and YAC contig around the mouse neurological locus reeler. Mamm Genome *5*, 756-61.

Murray, A. W., and Szostak, J. W. (1983). Construction of artificial chromosomes in yeast. Nature *305*, 189-93.

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.

Perry, W. r., Vasicek, T. J., Lee, J. J., Rossi, J. M., Zeng, L., Zhang, T., Tilghman, S. M., and Costantini, F. (1995). Phenotypic and molecular analysis of a transgenic insertional allele of the mouse Fused locus. Genetics *141*, 321-32.

Riley, J. H., Morten, J. E., and Anand, R. (1992). Targeted integration of neomycin into yeast artificial chromosomes (YACs) for transfection into mammalian cells. Nucleic Acids Res *20*, 2971-6.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2 Edition, A. Irwin, ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schuler, G. D. (1996). A gene map of the human genome. Science *274*, 540-546.

Schwartz, D. C., and Cantor, C. R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell *87*, 67-75.

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.

Srivastava, A. K., and Schlessinger, D. (1991). Vectors for inserting selectable markers in vector arms and human DNA inserts of yeast artificial chromosomes (YACs). Gene *103*, 53-9.

Strauss, W. M., Jaenisch, E., and Jaenisch, R. (1992). A strategy for rapid production and screening of yeast artificial chromosome libraries. Mamm Genome *150-157*.

Sutherland, H. F., Pick, E., Francis, F., Lehrach, H., and Frischauf, A. M. (1995). Mapping around the Fused locus on mouse chromosome 17. Mamm Genome *6*, 449-53.

Tartaglia, L. A., Dembski, M., Weng, X., Deng, N., Culpepper, J., Devos, R., Richards, G. J., Campfield, L. A., Clark, F. T., Deeds, J., and et al. (1995). Identification and expression cloning of a leptin receptor, OB-R. Cell *83*, 1263-71.

Trachtulec, Z., Vincek, V., Hamvas, R. M., Forejt, J., Lehrach, H., and Klein, J. (1994). Physical map of mouse chromosome 17 in the region relevant for positional cloning of the Hybrid sterility 1 gene. Genomics *23*, 132-7.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. M. (1987). Molecular Biology of the Gene, 4th Edition (Menlo Park, CA: The Benjamin/Cummings Publishing Company, Inc.).

Wedemeyer, N., Lengeling, A., Ronsiek, M., Korthaus, D., Baer, K., Wuttke, M., and Jockusch, H. (1996). YAC contigs of the Rab1 and wobbler (wr) spinal muscular-atrophy gene region on proximal mouse chromosome-11 and of the homologous region on human-chromosome 2p. Genomics *32*, 447-454.

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 Four**

# A Candidate Gene Map of the Mouse *pudgy* Region: Integration of Genetic, Physical, Transcriptional, Expression, and Genomic Sequence Maps

Kenro Kusumi[1,2], Wayne N. Frankel[3], Eileen S. Sun[1], Bruce Birren[1], Dow C. Chi[2], Trevor L. Hawkins[1], Mary P. Reeve-Daly[1], and Eric S. Lander[1,2]

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts, 02142; [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139; [3]The Jackson Laboratory, Bar Harbor, Maine, 04609

## INTRODUCTION

Over 2,000 mutations have been identified in the mouse, but only a small fraction have been positionally cloned (Doolittle et al., 1996). When isolated, these positionally cloned genes have provided an enormous insight into the genetic basis of developmental processes (Berta et al., 1990; Brannan et al., 1991; D'Arcangelo et al., 1995; Hamilton et al., 1996; Herrmann et al., 1990; Nehls et al., 1994; Segre et al., 1995; Zhang et al., 1994). A limiting factor in the identification of these mutations is the paucity of well-characterized gene sequences. In the human, fewer than 6,000 genes have been cloned and annotated in GenBank out of the estimated 70,000 total genes (Boguski and Schuler, 1995). The situation in the mouse is even less advanced.

In the human, efforts are underway to create a transcript map by mapping expressed sequence tags (EST) onto the physical map (Schuler and

al., 1996). For the mouse, EST sequencing efforts are underway at several centers (Washington U., Kyoto U.), but large-scale mapping efforts not as advanced. In the interim before completion of a genome-wide transcript map, regional transcript maps must be constructed for positional cloning efforts, using previously mapped genes and ESTs and identifying new transcripts by direct cDNA selection or exon-trapping methods (Buckler et al., 1991; Lovett, 1994).

Out of the more than 2,000 mouse mutations, one of the most striking skeletal phenotypes is seen in the *pudgy* mutant, which displays a severely shortened torso due to vertebral dysgenesis and rib defects, resulting from irregular somitic segmentation (Grüneberg, 1961; Kusumi et al., 1997). The phenotype is first observed on embryonic day 8, where defects can be seen in the condensation of segmental plate to form somites, collections of mesodermal cells which will eventually form adult vertebrae and ribs, dermis, and axial musculature.

Previous genetic mapping experiments have localized the mutation to the proximal region of Chromosome 7 (Grüneberg, 1961). Grüneberg noted that the mutation was linked and proximal to the pink-eye and albino loci on Chromosome 7 (Grüneberg, 1961) Attempts were made to produce closer proximal markers on chromosome 7 using microdissection library (Greenfield and Brown 1987). However, no further work has been reported on this developmental mutation, and the genetic mapping location has not been refined since the original description.

Here, we describe our efforts to positionally clone the *pudgy* gene. We narrowed the region where the *pudgy* mutation must be located through genetic and physical mapping. By genetic crosses yielding more than 2,000 meioses, we mapped *pudgy* within a 1.8 cM interval, and by physical mapping

with YAC, BAC, and P1 clones, this region corresponded to ~600 kb. We isolated transcript fragments by direct cDNA selection and localized them on the physical map to create a transcriptional map. By genomic sequencing of ~250kb of the region and integration of expressed sequences, we generated a base-pair resolution transcriptional map. Transcript fragments were screened by *in situ* hybridization for expression in embryonic day 9.5 mouse somites. Through this screening, we identified a single transcript, a novel serine-threonine kinase expressed in ventral somites, as a prime candidate for the *pudgy* gene.

## MATERIALS AND METHODS

### *Animals*

The *pudgy* mutation arose during the course of an X-ray mutagenesis experiment carried out by W. and L. Russell at the Oak Ridge National Laboratory, TN. In the original X-ray mutagenesis experiments, hybrid males (101/RlxC3H/Rl) were irradiated with 600 Roentgens of radiation and mated to non-inbred *Brachyury* stock females. The pudgy mutation has been subsequently maintained as its own strain through brother-sister sib intercrosses. (Russell, 1995) Previously, the pudgy stock has been referred to as $pu + c^{ch}/pu + +$ but will be referred to subsequently as PU/J. Inbred PU/J-*pu*/*pu* stock mice were purchased from the Jackson Laboratory (Bar Harbor, ME). MOLF/Ei, TKDU/J-*du*/*du*, and B6CBACa-$A^{w\text{-}j}$/*A-wv* were obtained from the Jackson Laboratory.

The crosses were as follows: an intersubspecific cross of 349 F2 progeny with *M. musculus molossinus* (PU/J-*pu*/*pu* x MOLF/Ei)F2, an intraspecific cross of 691 F2 progeny with an inbred stock containing the *ducky* mutation (PU-*pu*/*pu* x TKDU-*du*/*du*)F2, and an intraspecific cross of 92 F2 progeny

with an inbred stock containing the *weaver* mutation (PU-*pu*/*pu* x B6C3H/J-*wv*/*wv*).

The crosses were set up primarily with male homozygous mutant animals, since the affected males breed, whereas the females have difficulty achieving pregnancy and do not tend their litters well. Animals found to be recombinant in the *pudgy* region by genotype analysis (see below) were test crossed to *pudgy* homozygous or heterzygous animals. Test cross litters were examined for phenotype, and also assayed for genotype as described below.

*Genotype analysis*

DNAs used for genetic mapping analysis were prepared from ear or tail biopsy samples as described previously (Laird et al., 1991). Mouse simple sequence length polymorphism (SSLP) genetic markers were used as described previously (Dietrich et al., 1992; Dietrich, 1995; Dietrich et al., 1996; Dietrich et al., 1994; Dietrich et al., 1994; Dietrich et al., 1995). Novel primers developed using PRIMER version 0.5 (Center for Genome Research, Whitehead Institute, Cambridge, MA) for single-stranded conformational polymorphisms (SSCP) analysis are shown in Table 1 and were used in PCR assays using 35 cycles (94°C for 30 seconds, 55°C for 1 minute, 72°C for 1.5 minute). SSCP samples were separated by electrophoresis in 0.5x or 0.7x MDE™ Hydrolink gels (FMC Bioproducts, Rockland, ME) for 16 to 20 hours at 10 V/cm, according to manufacturer's recommendations. Gels were wrapped in Saran Wrap™ (Dow Chemical), exposed to Reflection™ film (Dupont/ NEN Products, Boston, MA) or X-OMAT™ AR film (Kodak, Rochester, NY) with photointensifying screens or BAS2000™ phosphorescent imaging plates (Fuji, Norwalk, CT) overnight.

*YAC, P1, and BAC screening*

YAC, P1, and BAC libraries were screened by radioactive-labeled PCR as described for genotype analysis and by nonradioactive PCR using the following conditions (reaction buffer: 50μM KCl; 1.25 mM dATP, dTTP, dCTP and dGTP; 1.5 mM $MgCl_2$; 0.5 μM forward primer, 0.5 μm reverse primer, 0.4 unit Taq polymerase (Perkin Elmer) Cycling conditions 35 cycles of (94°C for 30 seconds, 55°C for 1 minute, 72°C for 1.5 minute). YAC libraries screened included the Whitehead Institute/MIT mouse YAC library (Kusumi et al., 1993), the Princeton mouse YAC library (Burke et al., 1991), and the Imperial Cancer Research Fund Center mouse YAC library (Larin et al., 1991). A mouse 129/SvJ P1 library was screened through Genome Systems, Inc. (St. Louis, MO) and a mouse 129/SvJ BAC library was screened at the Center for Genome Research, Whitehead Institute (Birren, 1996).

YACs were sized as described previously (Kusumi et al., 1993). P1s and BACs were digested and linearized with NotI and analyzed by pulsed-field gel electrophoresis using 1% Seakem ME agarose (FMC Bioproducts, Rockland, ME) with 0.1 to 20 sec. switch time for a 16 to 20 hour run on a CHEF II™ apparatus (Bio-Rad Laboratories, Hercules, CA).

*Cloning of YAC, P1, and BAC ends*

YAC and P1 insert end-sequences were subcloned using inverse-PCR methods as described previously (Haldi et al., 1994; Segre et al., 1995). In summary for YACs, purified DNA was digested with the following restriction endonucleases: HaeIII, Sau3a, and TaqI for the TRP1/ARS1/CEN4 vector arm side, and AluI, HaeIII, and HhaI for the URA3 vector arm side (enzymes from New England Biolabs, Beverly, MA). For P1s, purified DNA was digested with the following restriction endonucleases: HhaI and AluI for the T7 vector arm and NlaIII and

RsaI for the Sp6 vector arm. Digested DNA samples were ligated at a concentration of 2.5 ng/μl with T4 DNA Ligase, and samples were amplified with inverse-PCR primers as described previously (Haldi et al., 1994; Segre et al., 1995). YAC end assays were tested against a mouse chromosome 7 polychromosomal hybrid panel (gift of C. Kozak).

*Direct cDNA selection*

Directed selection of cDNAs against template P1 DNA was carried out in order to obtain transcripts from the pudgy region. cDNA was prepared using poly(A) RNA from BALB/cJ E10.5 embryos using Superscript™ Choice cDNA Synthesis System (Life Technologies, Gaithersburg, MD) and selected against P1 clones using methods described previously (Lovett, 1994; Segre et al., 1995). In addition, an adult C57BL/6J testis cDNA library (Segre et al., 1995) and a pooled, embryonic cDNA library with normalized samples from each day from E0 to E19 (Takahashi and Ko, 1994) were used for direct selection. Following two rounds of hybridization, selected cDNAs were subcloned into the CLONEAMP™ pAMP10 System using uracil containing linkers (E10.5 and testis libraries CUACUACUACUACTGAGCGGAATTCGTGAGAC; Takahashi & Ko library CUACUACUACUAATTGACGTCGACTATCCAGG with the underlined sequences representing the uracil deglycosylase target sections) using manufacturer's methods (Life Technologies, Gaithersburg, MD).

*Screening of cDNA libraries*

Several libraries were screened to obtain further clones of the *pudgy* region genes. Oligo-dT primed embryonic library from E8.5 embryos (Fahrner et al., 1987), E10.5 embryos (Novagen, Inc., Madison, WI), and mixed random-primed and oligo dT-primed libraries from adult brain and spleen (Stratagene

Cloning Systems, La Jolla, CA) were screened using standard procedures (Sambrook et al., 1989). Primary plates were made at a density of $10^6$ plaques/175 $cm^2$ and transferred onto Genescreen Plus™ nylon membranes (Dupont/NEN Products, Boston, MA). Membranes were hybridized in Church-Gilbert buffer with radioactively labelled probes at a concentration of $10^6$ cpm/ml in buffer at 65°C. Membranes were washed in 1x SSC, 0.1% SDS at 65°C for a total of 20 minutes and exposed to X-ray film overnight. Individual phage inserts were isolated from secondary or tertiary screening phage dilutions by PCR using either -21M13 (TGTAAAACGACGGCCAGT) and M13rev (CAGGAAACAGCTATGACC) primers for λZAPII brain and spleen libraries; M13 chimeric λgt10 forward and reverse primers for the E8.5 library (M13Forλgt10F TGTAAAACGACGGCCAGTAGCAAGTTCAGCCTGGTTAAG , M13Revλgt10R CAGGAAACAGCTATGACCATGAGTATTTCTTCCAGGGTA where the underlined sequences are M13For or M13Rev sequences); and M13 chimeric T7 and Sp6 primers for the Novagen λSH*lox* vector E10.5 library (M13ForT7 TGTAAAACGACGGCCAGTCGACTCACTATAGGGAGCTCG, M13RevSp6 CAGGAAACAGCTATGACCGGTGACACTATTAGAATATGCATCAAG).

*DNA sequencing*

Plasmid DNA was prepared using Wizard™ miniprep DNA purification system, according to manufacturer's directions (Promega, Madison, WI). PCR amplified DNA samples were gel purified in 0.8% SeaPlaque GTG™ (FMC Bioproducts, Rockland, ME) and excised agarose bands were melted at 65°C, treated with β-agarase (New England Biolabs, Beverly, MA) according to manufacturer's specifications. Agarased samples were purified by dilution with 1x TE and concentration on Ultrafree-MC® 30,000 NMWL filters (Millipore Corporation, Bedford, MA).

DNAs were sequenced using fluorescently-labelled dye-primer or dye-terminator based sequencing methods using an ABI370A automated sequencer (Perkin-Elmer/Applied Biosystems, Foster City, CA) using the manufacturer's suggested protocol. For dye-primer based sequencing, M13 sequences for priming were added to templates by PCR reamplification with M13 chimeric primers (see Table 1). Dye terminator reaction products were purified from unincorporated dye terminator nucleotides using CENTRI-SEP™ columns (Princeton Separations, Adelphia, NJ). Sequence chromatogram data was collected using ABI Data Collection version 1.2.0. Automated basecalling of sequences was carried out using the ABI50 or ABI100 base-calling software modules to ABI Prism™ Sequence Analysis version 2.1.0. DNA sequences were assembled and analyzed using Sequencher™ version 3.0 (Gene Codes Corp., Ann Arbor, MI) and SeqMan™ version 3.05 (DNASTAR, Inc., Madison, WI).

*Large-scale shotgun genomic sequencing*

BAC and P1 clone DNA preparations were carried out by alkaline lysis as described previously (Hamilton et al., 1996; Hawkins, 1996). Clones were subcloned into M13mp18 vectors and prepared for dye-primer based sequenced as described previously. In summary, subclones of average insert size 2kb were generated and automated fluorescent sequencing reads of ~400-500bp were generated from the ends using -21M13 and M13Rev primers. A ratio of 5:1 forward to reverse reads was used, and sequence and clone information were analyzed using the Gap4 application within the Staden DNA Analysis Suite on UNIX platforms. Remaining gaps were closed using dye terminator-based sequencing off primers generated from the gap edge sequences.

Sequence analysis of contigs was conducted with BLASTN, BLASTX, and TBLASTX against the dBEST expressed sequence tag database. Sequence alignments were carried out using Sequencher™ version 3.0 (Gene Codes Corp., Ann Arbor, MI) and Seqman application within the DNA* suite (DNASTAR, Inc., Madison, WI) on MacOS platforms.

*Southern and Northern hybridization*

Southern and Northern blots were prepared according to standard methods onto Genescreen Plus™ or Hybond-N™ nylon membranes (Sambrook et al., 1989). Probes were labelled with $\alpha^{32}$P-dCTP (Dupont/NEN Products, Boston, MA) using PCR or by random oligo priming (Feinberg and Vogelstein, 1983). Labelled product was then purified away from unincorportated nucleotide using NucTrap™ push columns (Stratagene Cloning Systems, La Jolla, CA), using manufacturer's instructions.

*Whole mount embryo in situ hybridization*

Direct cDNA selection fragments subcloned into pAMP10 were used as templates for RNA probes. Inserts were amplified by PCR using -21M13 and M13Rev primers, whose sites flank the insert. As control probes for somite specific markers, we used *Mox1* and *Pax1* gene fragments amplified by PCR from whole embryos E10.5 polyA selected RNA. Underlined sequences represent non-gene sequences containing T7 promoter site, and sequence numbers in the primer name represent the starting position relative to GenBank sequence. *Mox1* probe was made using primers MOX1191 (AGTTGCCCAGTATGTGGGAG) and T7-MOX1628 (GCGTAATACGACTCACTATAGGGAGATTGGAGAACACAAGACGCTG). *Pax1* probe was made using primers PAX1920 (AGCAGCCACAGTCCCAAG) and T7-PAX2321

(GCGTAATACGACTCACTATAGGGAGAAACCTCACCACCCTGAAGC). Products were purified by gel purification in Seaplaque GTG agarose (FMC Bioproducts, Inc., Rockland, ME). DNA was freed from agarose using β-agarase (New England Biolabs, Beverly, MA) and purification using Ultrafree-MC® 30,000 NMWL filters (Millipore Corporation, Bedford, MA) and resuspended in 1x TE. RNA probes for in situ hybridization were made using digoxigenin labelled UTP (Boehringer-Mannheim, Indianapolis, IN) using T7 or Sp6 MAXIscript™ in vitro transcription kits (Ambion, Inc., Austin, TX).

BALB/cJ, C3H/HeJ, or DB A/2J embryos were used for *in situ* screening, and embryos were designated as 0.5 day at the time of coital plug identification. E9.5 embryos were dissected for *in situ* analysis, with embryos immediately fixed in 4% paraformaldehyde in PBS overnight at 4°C with gentle shaking. Whole mount *in situ* hybridization with digoxigenin-labelled probes was carried out essentially as described previously (Wilkinson, 1992).

## RESULTS

*Genetic mapping of pudgy to a 1.8 cM interval*

The pudgy mutation was genetically mapped through the use of three intercrosses with a total of 1,132 F2 progeny or 2,264 meioses (Materials the Methods). Out of 698 F2 intercross meioses in the MOLF/Ei intersubspecific cross, the pudgy mutation was recombinationally inseparable from markers D7MIT72, D7MIT266, and *Zfp36* (Figure 1b). The genetic region was bounded proximally by PY61R, a SSCP-based marker generated from a YAC end, which is 0.7 cM distant from the mutation and distally by PP13Hin, a SSCP marker generated from a P1 clone end, which is 1.1 cM distant. The minimal genetic region for pudgy in the MOLF cross is thus 1.8 cM.

Although on average 1 cM of genetic distance can be approximated by 2.0 Mb in the mouse, it should be noted that markers D7MIT72 and PP13Hin map to the same set of 3 P1 clones, which have an average insert size of 80-85 kb. This would imply that 1.1 cM distance in this region falls within a small physical region and that the recombination breakpoints are not spaced along the physical distance map in an average fashion.

The intraspecific cross with TKDU was less informative proximally, since D7MIT77 and PY61R were not polymorphic in this cross. The closest proximal marker, D7MIT23 is 1.8 cM distant. The pudgy mutation is recombinationally inseparable from D7MIT266 and *Zfp36*, with D7MIT72 not polymorphic in this cross. The closest distal marker was PP13Hin, which is 0.1 cM distal. This distance is much less that the corresponding distance in the MOLF intersubspecific cross. Thus, the minimal genetic region for the pudgy mutation in the TKDU cross is 1.9 cM, and the corresponding interval in the MOLF cross is 2.0 cM.

The intraspecific cross with the *weaver* hybrid stock, due to its small size, was not typed as extensively as the previous two genetic mapping crosses. However, the results confirm the previous mapping results. The pudgy mutation is 1.9 cM distal to the SSLP marker D7MIT23 and 1.7 cM proximal to D7MIT55. This extended genetic interval is 3.6 cM, which is intermediate between the equivalent interval in the MOLF cross (2.4 cM) and the larger region in the *ducky* cross (4.9 cM).

The pudgy mutant has been previously described to have a high rate of embryonic lethality, estimated to be 18% in the single cross examined. (Grüneberg, 1961) We have found a statistically significant loss of the *pu/pu* genotypic class in two crosses examined, with a 50% lethality in the MOLF/Ei cross (Kusumi et al., 1997). All animals homozygous for PU/J alleles near the

*pu* locus were affected, and all affected animals were PU/J homozygous at one or both flanks. Thus, the *pudgy* mutation is fully penentrant, but there is a underrepresentation of *pudgy* homozygous animals in these crosses.

*Physical mapping of pudgy to a ~600kb region*

The genetic mapping efforts reduced the *pudgy* minimal interval to 1.8 cM in the MOLF/Ei cross. One percent of recombination frequency averages 2 Mbp of physical distance in the mouse, giving an expected 3.6 Mbp region. To span these large distances, we screened several large-insert YAC libraries: WI/MIT Mouse YAC Library I (Kusumi et al., 1993), Imperial Cancer Research Fund Mouse Library (Larin et al., 1991), and the WI/MIT Mouse YAC Library II (Haldi et al., 1996). We initiated a YAC chromosomal walk with simple sequence length polymorphism-based markers D7MIT72, D7MIT77, D7MIT266, and STS markers based on the previously identified gene *Zfp36* (Table I).

Due to a high rate of chimerism inherent in the present generation of available YAC libraries, we confirmed that the genomic clones were located in the *pudgy* region. We cloned YAC ends by inverse-PCR methods, and selected primers for PCR-based assays from end sequences. Markers were considered to be located within the *pudgy* region if at least one of the following conditions was met

- an assay detected more than one clone within the region (YAC, BAC, P1) and was determined to be chromosome 7 positive by polychromosomal hybrid panel
- an assay was used as a genetic marker to confirm location

To generate internal STS assays for further mapping efforts, we subcloned two YACs (FFX-C12, FAE-A8) and obtained internal fragment sequences. These

assays were also used for the isolation of smaller insert DNA clones.

Several factors make YACs unsuitable for further physical mapping efforts and resources for gene identification. A high rate of internal rearrangements (deletions, inversions) and chimerism make for a low accuracy representation of the genomic DNA region (Haldi et al., 1994). In addition, it is difficult to purify YAC DNA away from yeast genomic DNA, complicating procedures which necessitate homogeneous sample, e.g. subcloning for generation of STS assays. Beyond low resolution physical mapping with YACs to cover megabase-pair sized regions of genome, DNA clones are needed for the following efforts: large-scale sequencing of the region, template for direct cDNA selection, reagents for transgenic complementation, confirmatory evidence for the genomic distances and coverage in the region.

Using end and internal STS assays generated from the YACs, we have generated a fine-scale physical map using bacterial artificial chromosome (BAC) and P1 based clones. BAC clones (generated by a HindIII partial digest of 129/SvJ DNA) were of average insert size 100-200kb (Birren, 1996). P1 clones (generated by a partial Sau3a digest of 129/SvJ DNA) were of insert size 80-100kb, due to limitations of the phage packing limits. STS assays were used to integrate the low-resolution YAC map with a higher resolution BAC and P1 map (Fig. 3). This allowed us to refine the size estimates of portions of the *pudgy* interval, and provided DNA clone templates for direct cDNA selection and genomic sequencing efforts. Based on BAC and P1 insert sizes and coverage, we estimate that instead of a 3.6 Mbp region, the *pudgy* interval is only ~0.6 Mb (Figure 2).

*Generating a transcript map from direct cDNA selection*

Previous studies have shown that the *pudgy* phenotype is first detectable at E8.5 and that somitogenesis is disrupted from E8 to E12 (Grüneberg, 1961; Kusumi et al., 1997). We used cDNA selection methods to selectively look for only those genes expressed at those times. The method of direct cDNA selection will only identify transcripts that are expressed in a particular sample tissue. We targeted the following tissue as possible sources of a *pudgy* transcript:

- cDNA from murine embryonic day 10.5 whole embryos (BALB/cJ)
- cDNA from murine embryonic day 0 to day 19 whole embryos, sampled daily, normalized, and pooled (courtesy of M. Ko, Northwestern University)
- cDNA from mouse adult testis, as a control for nonembryonic transcripts.

We selected E10.5 embryos instead of E8.5 embryos, since more somites are present in this midpoint stage in somitogenesis.

Due to PCR steps involved in direct cDNA selection methods, exon fragments identified were usually limited in size to 200-300bp. Direct cDNA selection fragments were sequenced and aligned for redundancy or overlap. In addition, cDNA selection fragments with BLASTN alignments to repetitive genomic sequences or ribosomal sequences were eliminated as by-products of insufficient selection blocking of repeats. PCR assay primers were generated from each fragment and were tested against a clone panel of the *pudgy* region to confirm STS content. Clones with assays which either did not map to the *pudgy* nonrecombinant region or mapped to clones which were inconsistent with the physical map were eliminated from consideration.

Direct cDNA selection fragments sequences were analyzed by BLASTN, BLASTX, and TBLASTX. Fragments with significant homology (less than $10^{-10}$ probability of being identified randomly in the database) to GenBank

sequences were noted (Table 2). BLASTN and BLASTX matches for 2 or more fragments (probabilities < $10^{-50}$) consistent with identification of known mouse genes or orthologs of human and rat genes were as follows, with clones indicated in parentheses:

• *Ech1*, peroxisomal enoyl hydratase (Fitz Patrick et al., 1995) (A202, 6b66)

Peroxisomal enoyl hydratase is expressed ubiquitously and involved in the degradation of cholesterol.

• a heterogeneous nuclear protein (Biamonti et al., 1989; Buvoli et al., 1988) (13a13, 4b2, 4b9, A16, A19, A20, B102, C205, C207, C91)

HNRPs are components of ribonuclear protein complexes that interact with RNA polymerase II.

BLAST and BLASTX matches (probabilities < $10^{-20}$) consistent with identification of possible paralogs of sequences from mouse, human and rat genes were as follows, with cloned fragments indicated in parentheses:

• the *Pak/STE20* family of serine-threonine kinase genes (D212, D42, (Bagrodia et al., 1995; Bagrodia et al., 1995; Benner et al., 1995; Brown et al., 1996; Harden et al., 1996; Manser et al., 1995; Martin et al., 1995)

• a histone H2A gene (Ivanova et al., 1994) (A209)

• 5E5 antigen, a putative neural DNA binding protein (Suzuki et al., 1995) (A19)

• elongation factor EF1A, a translation machinery component (Brands et al., 1986) (F37)

• zinc finger protein *Mfg3*, a ubiquitously expressed transcription factor (Passananti et al., 1989) (J76)

*Mfg3* was found to be located outside the *pudgy* genetic interval by physical mapping, and is not a candidate for *pudgy*.

• alpha-actinin, a skeletal muscle expressed actin binding protein (Youssoufian et al., 1990) (C50)

• galectin, a putative ligand for β-galactosidase lectin and secreted by macrophages as part of the inflammatory response. (Madsen et al., 1995; Rosenberg et al., 1991) (C62)

For certain clones, no significant BLAST and BLASTX matches were found, but BLASTN searches against the dBEST expressed sequence tag database resulted in significant matches (probabilities < $10^{-20}$). These are likely to be either sequences from genes not entered in GenBank or uncloned regions of known genes (e.g. 5' UTR). These clones included 14a8, 14a3, A15, A24, D32, G210a, G72, and H7. However, 57 of the 113 cDNA selection fragments did not show any significant BLAST matches, suggesting a paucity of known genes in this region.

Given that less than 6,000 genes have been well-annotated in GenBank out of the 70,000 genes in the human genome (Boguski and Schuler, 1995), homology information alone is not a sufficient basis on which to evaluate gene candidates. Initially, cDNA selection fragments were binned into groups based on PCR assay-based localization on the physical mapping clone panel. For E10.5 transcript fragments, 64 clones were binned as shown in Figure 4a, with coverage evenly distributed across the region selected, except for the most proximal region which was not as extensively selected. For E0-19 and testis transcript fragments, 39 clones were binned as shown in Figure 4b. Interestingly, we were unable to select a fragment of the *Zfp36* zinc finger gene from any of the three RNA sources selected, despite the expression of the gene within the embryonic stage RNAs.

*Genomic sequence analysis*

We have used shotgun genomic sequencing methods to obtain the genomic sequence of the *pudgy* region. Two BAC clones were selected for sequencing: 19k5 (insert size 125kb, project L44) and 315c18 (insert size 115kb, project L45) which were located as shown in Figure 3 & 4. These clones were selected since they contained genes of interest as candidates—a novel serine-threonine kinase gene located, but not limited to, BAC 315c18 and a zinc-finger gene *zfp36* on BAC 19k5. Sequencing is still in progress on these two projects, but a substantial amount of preliminary genomic sequence has been obtained which approximately 10 large contigs (5-50kb) and numerous small contigs (<1kb) per clone. This sequence can be obtained via the web site (`http://www-genome.wi.mit.edu/cgi-bin/seqrelease/public/seqdata.cgi?database=release`). Although this sequence is preliminary and has yet to be completely edited with 99% base-pair accuracy, combination of this genomic sequence with cDNA selection sequence produces a base-pair scale map of up to ~250 kb of the *pudgy* interval. With the availability of large-scale genomic sequence, we significantly enhanced our localization of cDNA selection fragments, as shown in Figure 5 for BAC 315c18 and Figure 6 for BAC 19k5.

*Previously known genes and genomic sequence*

Two GenBank entry genes were clearly localized to the *pudgy* region—*Ech1*, peroxisomal enoyl hydratase, and *Zfp36*, a zinc finger protein. Both have human orthologs on chromosome 19q13.1, and combined with general synteny conservation to human 19q13.1 based on flanking genes such as *Mag*, myelin associated glycoprotein, and *Cea*, carcinembryonic antigen, the *pudgy* region corresponds to 19q13.1 in human. In addition to our genomic sequencing

efforts, we anticipate future comparison with human chromosome 19 genomic sequencing efforts underway at Lawrence Livermore National Laboratory (http://www-bio.llnl.gov/bbrp/genome/genome.html).

Genomic sequence also immediately permits resequencing efforts for mutation identification. Given that *Zfp36* as a transcription factor is an attractive candidate for the *pudgy* gene, we have resequenced the entire RNA encoding sequence of the *Zfp36* gene and found no polymorphism between *pudgy* homozygous and parental strain (C3H/Rl and 101/Rl) DNA sequences. In addition, we have sequenced 5' upstream regulatory sequences and found an interesting polymorphism in the 5' flanking region of *Zfp36* at -180bp upstream of the transcriptional start site, with 7 C nucleotides instead of 6 found in both C3H and *pu/pu* but not observed in C57BL/6J, AKR/J, DBA/2J, BALB/cJ, 129/SvJ, TKDU/J, CAST/Ei, MOLF/Ei, and *M. spretus*. This C3H/Rl and *pu/pu* associated polymorphism is a useful marker for the *pu* chromosome. Although we were able to find this useful polymorphism, the failure to detect a polymorphisms between *pudgy* and its parental chromosomes give us evidence against *Zfp36* as the *pudgy* gene. We were not able to detect *Zfp36* expression in E10.5 embryonic RNA by Northern hybridization (data not shown).

*Integration of cDNA fragments and genomic sequence*

Genomic sequence was used to position the numerous direct cDNA selection fragments on a precise sequence map, giving clues to intron/exon boundaries and which fragments are portions of the same transcript. For example, several cDNA selection fragments had significant homology to the *Pak* serine-threonine kinase gene family (Table 2). Alignment of all such fragments against genomic sequence contigs (Fig. 5 e, f, h) which have been

ordered by STS mapping indicates that this homology is organized over contigs totalling 56,840 base-pairs. Given that cDNA fragments without serine-threonine kinase matches representing divergent coding sequence or untranslated sequence may in fact correspond to cDNA fragments mapping to adjacent contigs (Fig. 5c, d), this gives us a tool to identify more sections of a widely dispersed gene without extensive cDNA library screening.

When sequences from cDNA library clones are included, we are able to expand gene bins even further. Cloning of cDNA library fragments links direct selection clones D19, D219 and D23 shown in Fig. 5c to the *Pak* related serine-threonine kinase transcript (Kusumi et al., 1997). Alignment of 3.0 kb of this kinase transcript with genomic sequence encompasses contigs covering more than 97,941 base-pairs.

*Open reading frames in genomic sequence correspond to ESTs*

Open reading frame information can also be used to identify transcripts. We examined the L44.3 genomic sequence contig, which contains the previously annotated *Zfp36* gene. L44.3 covers 47,664 bp, and contains the genomic region of *Zfp36* (Nup475/Tristetraproline/TIS) zinc finger gene on the BAC 19k5. Both the 1789bp mouse cDNA and 7493 bp of genomic region have been identified, divided into 83bp and 1706bp exons (Lai et al., 1995; Taylor et al., 1995). We have independently confirmed this organization and added 40kb of sequence downstream of the *Zfp36* gene. In the sequences downstream to *Zfp36*, we have identified 5 B2 and 15 B1 repeats, as shown in Figure 6b. We noted several large open reading frames in the preliminary sequence, and alignment of direct selection fragments (G223, G230, G72, G76, G214, G210, G62, G84) indicates that they align to several of these open reading frames. Given that the 3' end of the *Zfp36* gene likely represents the 5'

boundary to these transcript fragments, genomic sequence immediately gives us access to likely 5' regulatory regions to this novel transcript.

*In situ hybridization pattern based screening*

Given the large percentage of cDNA fragments without significant GenBank homology, we approached the screening of candidates in a homology-neutral fashion. The *pudgy* phenotype has a striking effect on the segmentation of somites. Although many alternate models can be proposed, the *pudgy* gene is likely to be expressed within somites. We have selected to examine candidates by whole mount *in situ* hybridization, since this method provides the greatest amount of information regarding spatial pattern. The method is comparable in sensitivity to RNA blot analysis (detection of more than 0.001% of total RNA present) but not as sensitive as RT-PCR (potential detection of one molecule of RNA) (Sambrook et al., 1989).

We carried out a whole mount *in situ* hybridization screen using 50 nonoverlapping cDNA selection clones distributed across the *pudgy* physical region. We examined E9.5 stage embryos, since that stage is optimal for scoring expression in somites. Both strands were generated for testing, since directional information is lost during the process of direct cDNA selection. Thus, 100 probes were tested in total.

Clones could be categorized in three general classes: ubiquitously hybridizing probes, probes which hybridized to selective regions in the embryo, and probes that did not hybridize beyond background levels. Examples of these classes are illustrated in Figure 7. We added additional expression pattern information against *Zfp36* as the *pudgy* gene, since the transcript was expressed predominantly in the neural folds and developing neural tube. We selected 10 candidate clones which appeared to selectively hybridize to somitic

tissue. These included the fragments D42, D48, D212, D218, D51, D19, D23, F49, G214, and F36. Of these 10, cDNA library screening along with RNA blot analysis has linked 8 of the above fragments together, leaving currently only G214 and F36 unlinked (Kusumi et al., 1997). Further discussion of the nature of this cDNA group, including members with significant homology to the *Pak* serine-threonine kinase family, will be provided in Chapter 5.

*Summary of Results*

Due to the complexity of transcript fragments and integration of physical and sequence mapping data, the results are summarized in the table below:

| ***gene homology/alignment*** | ***cDNA fragments*** | ***total number*** | ***in situ expression*** |
|---|---|---|---|
| cDNA fragments with no homology | | 57 | |
| cDNA fragments with only dbEST homology | | 8 | |
| cDNA fragments with nonspecific homology | | 5 | |
| Zfp36 | none | 0 | cns |
| serine-threonine kinase | D8<br>D19<br>D23<br>D42<br>D43<br>D48<br>D60<br>D102<br>D209<br>D212<br>D218<br>D101<br>D38<br>4b22<br>4b2<br>4b7<br>4a2 | 17 | somitic |
| HNRP | A104<br>A16<br>A20<br>A210<br>B102<br>C205<br>C207<br>13a13 | 8 | ubiquitous |
| peroxisomal enoyl hydratase | A202<br>6b66 | 2 | background |
| 5E5 antigen | A19 | 1 | background |
| histone H2A | A209 | 1 | ubiquitous |
| alpha-actinin | C50 | 1 | ubiquitous |
| galectin | C62 | 1 | - |
| elongation factor 1a | F37 | 1 | ubiquitous |
| mfg3 | J76 | 1 | endoderm/gut |

## DISCUSSION

*Genetic mapping*

The localization of the *pudgy* mutation has been significantly refined from previous efforts, due largely to the availability of numerous second generation DNA markers based on simple sequence length polymorphisms. As shown in Figure 1, the work of the 1995 Chromosome 7 Consensus Committee greatly misplaces the location of the pudgy mutation due to the low resolution of previous mapping efforts and incorrect integration of previous maps. Several markers (*Zfp36*, D8MIT72, D8MIT266) have been identified which are recombinationally inseparable from the mutation. Examining the genes which map to the minimal *pudgy* genetic region, the zinc finger protein *Zfp36* warranted close examination as a potential *pudgy* candidate. Significant differences in recombination-based genetic distance exists between the two major crosses examined, but these variance have been observed with other genetic mapping efforts.(Hamilton et al., 1996; Segre et al., 1995)

*Integration of physical, transcriptional, and genomic sequence maps*

We have analyzed the physical region in the *pudgy* interval through three levels of analysis

- megabase-pair level low-resolution mapping with YACs
- 10-100kb scale mapping using BAC and P1 clones, and
- base-pair resolution mapping achieved through genomic sequencing.

The region in the *pudgy* interval contained several areas with low coverage. Despite screening 10x YAC library coverage for the SSLP marker D7MIT72, we were able to identify only one YAC that represented this region of genome. Similarly with P1s and BACs, we identified two regions (between PY20R2/PP13Hha and between PP19Rsa/312n16-Sp6) where there was low

coverage in clones, and termination of clones in a small physical interval. This is likely due to a high frequency of restriction sites used to generate the clones, which were EcoRI for YACs and Sau3A for BACs and P1s. Given that almost all YAC libraries are constructed by partial EcoRI digest and a similar situation with HindIII for BACs and Sau3A for P1s, this illustrates the importance of generating complementary DNA clone resources using different restriction endonucleases in order to maximize coverage of the genome.

In order to identify candidate genes for a mutation, a transcript map for the *pudgy* region was generated by three general methods: the identification of known genes within the region, the identification of exons based on expressed sequences (using direct cDNA selection), and the derivation of exons using conserved gene motifs and sequences (by analysis of genomic sequence or through the use of experimental methods such as exon-trapping). Despite the large region examined, we have identified a large percentage (55%) of cDNA selection fragments without significant homology to previously identified genes in GenBank. This could be due to numerous factors including a paucity of known genes in this region and an abundance of untranslated segments of genes. Since cDNA selection fragments are typically small (200-300bp), a lack of significant database homologies makes it difficult to arrange transcript fragments into an entire gene map. Given a large number (103) of cDNA selection fragments, we first used binning by PCR-based STS mapping on our BAC/P1 physical map, yielding groupings of transcripts on a low-resolution scale of ~20-50kb regions. Given an estimated 70,000 genes in the mouse genome, transcripts would be expected an average genomic frequency of one per 30kb. Thus, binning by physical mapping did not permit us to precisely localize exon fragments to likely gene bins. Using preliminary genomic sequence, we were able to identify and group cDNA selection fragments, and

greatly facilitating the process of cDNA library screening towards whole transcript identification. Given completely edited genomic sequence, remaining cDNA selection fragments which did not map to existing contigs will be mapped to base-pair resolution.

*Evaluation of genes in the pudgy region*

Direct cDNA selection produced a large percentage (55%) of fragments which did not give significant GenBank homology. Given that fewer than 10% of genes have been characterized in detail, this may represent a normal situation. Sequence neutral expression screens using *in situ* hybridization demonstrated somitic expression in a number of fragments which were part of a novel serine-threonine kinase transcript (see below). Of the other significant BLAST matches, many are unlikely candidates for the *pudgy* gene based on biological evidence from homology. Most are ubiquitously expressed genes involved in basic functions. *Ech1* is ubiquitously expressed and involved in a cholesterol breakdown pathway. HNRPs are components of ribonuclear protein complexes that interact with RNA polymerase II. Histones interact with DNA to form chromosomal structure. Other candidates are involved in unrelated biological processes. Galectins are secreted by macrophages as part of the inflammatory reponse, and alpha-actinins are skeletal muscle associated actin-binding proteins. An interesting candidate *Mfg3* is located outside the minimal *pudgy* interval and does not give a somitic *in situ* expression pattern. In summary, the *pudgy* region does not contain many orthologs of previously identified genes, and most genes residing in the interval are unlikely to be involved in somitic development.

*Zfp36 is an unlikely candidate for pudgy*

Due to its role as a transcription factor important during development in *Drosophila* and *C. elegans* (Ma et al., 1994; Mello et al., 1996; Mohler et al., 1992), the *Zfp36* gene was originally a candidate in the *pudgy* region. We used genomic resequencing and *in situ* hybridization-based screens to provide evidence against the *Zfp36* as the *pudgy* gene. Although ubiquitous expression is potentially consistent with the expression of the *pudgy* gene, the pattern for the zinc finger gene *Zfp36* was clearly inconsistent. *Zfp36* is expressed only in the neural tube, with highest expression in the forebrain and leading edge of the closing neural tube. Although neural tube defects can affect the formation of somites as shown in the *curly-tail* mutation or the targeted disruption of the MacMARCKS gene, the type of somitic defects observed in those mutations with the abrupt cessation of normal somitogenesis are distinguishable from the global somitic malformations seen in the *pudgy* mutation (Chen et al., 1996; Kusumi et al., 1997; Zhu et al., 1995). This expression pattern, combined with a lack of identified polymorphism between *pudgy* homozygous and parental strain DNA's within the entire *Zfp36* transcript sequence, makes the gene highly unlikely to be the *pudgy* gene. Given that the *Zfp36* gene is tightly linked to the *pudgy* mutation, the polymorphism associated with C3H/Rl, C3H/HeJ, and *pu/pu* but not other inbred strains examined and CAST/Ei make it likely that the *pudgy* mutation arose on a C3H background.

*Identification of a novel serine-threonine kinase candidate*

Through direct cDNA selection efforts, we isolated a number of fragments with significant homology to the p21 activated kinase family of genes (*Pak, Ste20*). These fragments also were later identified to be part of the same serine-threonine kinase transcript, through cDNA library screens as discussed

in Chapter 5. In addition, sequence-neutral *in situ* hybridization screens revealed that the fragments recognized a transcript expressed in the ventral sections of somites, making the transcript an attractive candidate for the *pudgy* gene. Integration of genomic sequence and cDNA selection data revealed that this transcript was distributed over a large genomic region, being located on genomic sequence contigs totalling almost 100 kb in size. Further work on this candidate is described in Chapter 5.

*Conclusion*

In summary, as part of our positional cloning of the *pudgy* mutation, we have used over 2,000 intercross meioses to localize the mutation to within 1.8 cM in a ~600 kb genomic region. Using direct cDNA selection with E10.5 embryonic cDNA, we isolated over 100 transcript fragments and localized them on the physical map to create a transcriptional map. We have completed genomic sequencing of ~250kb of this interval, and integrated genomic sequence with expressed sequence to generate a base-pair resolution transcript map. Through *in situ* analysis and RT-PCR resequencing, we have been able to eliminate the candidate gene *zfp36* from consideration as the *pudgy* gene. We have screened 48 cDNA fragments by whole mount *in situ* hybridization of E9.5 embryos, and have identified 17 fragments which are part of a novel serine-threonine kinase transcript selectively expressed in somites. Other transcripts fragments with significant homology were shown to be unlikely to be the *pudgy* gene due to ubiquitous expression in embryos. The novel serine-threonine kinase transcript is distributed at least over a 100 kb region, covering an above-average genomic region.

## ACKNOWLEDGMENTS

We would also like to acknowledge Keri Devon, Michael Nguyen, Marta Velez-Stringer, David Wilcomb, and Carolyne Dunbar for valuable technical assistance. In addition, we thank Achim Gossler and Andy MacMahon for helpful comments on *in situ* data. We thank Christine Kozak for the gift of polychromosomal hybrid panel DNA's and Brigid Hogan for her E8.5 embryonic cDNA library. For their computer consultation, we thank Fran Lewitter, Carol Laboissoniere, Carl Rosenberg and Mark Daly. We would like to thank the members of the Lander Lab, in particular Arend Sidow, Julia Segre, Armand MacMurray, Anna Pettersson, Bruce Hamilton, Jennifer Nemhauser, Johanna Hästbacka, Howard Jacob, and Bill Dietrich for helpful discussions. We thank Peggy Kolm and John Kuo for discussion on in situ hybridization screening.

## REFERENCES


Bagrodia, S., Derijard, B., Davis, R. J., and Cerione, R. A. (1995). Cdc42 and PAK-mediated signaling leads to Jun kinase and p38 mitogen-activated protein kinase activation. J Biol Chem *270*, 27995-8.

Bagrodia, S., Taylor, S. J., Creasy, C. L., Chernoff, J., and Cerione, R. A. (1995). Identification of a mouse p21Cdc42/Rac activated kinase. J Biol Chem *270*, 22731-7.

Benner, G. E., Dennis, P. B., and Masaracchia, R. A. (1995). Activation of an S6/H4 kinase (PAK 65) from human placenta by intramolecular and intermolecular autophosphorylation. J Biol Chem *270*, 21121-8.

Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N., and Fellous, M. (1990). Genetic evidence equating SRY and the testis-determining factor. Nature *348*, 448-50.

Biamonti, G., Buvoli, M., Bassi, M. T., Morandi, C., Cobianchi, F., and Riva, S. (1989). Isolation of an active gene encoding human hnRNP protein A1. Evidence for alternative splicing. J Mol Biol *207*, 491-503.

Birren, B. (1996). Mouse BAC Library.

Boguski, M. S., and Schuler, G. D. (1995). ESTablishing a human transcript map. Nature Genet. *10*, 369-371.

Brands, J. H., Maassen, J. A., van Hemert, F. J., Amons, R., and Moller, W. (1986). The primary structure of the alpha subunit of human elongation factor 1. Structural aspects of guanine-nucleotide-binding sites. Eur J Biochem *155*, 167-71.

Brannan, C. I., Lyman, S. D., Williams, D. E., Eisenman, J., Anderson, D. M., Cosman, D., Bedell, M. A., Jenkins, N. A., and Copeland, N. G. (1991). Steel-Dickie mutation encodes a c-kit ligand lacking transmembrane and cytoplasmic domains. Proc Natl Acad Sci U S A *88*, 4671-4.

Brown, J. L., Stowers, L., Baer, M., Trejo, J., Coughlin, S., and Chant, J. (1996). Human Ste20 homologue hPak1 links GTPases to the JNK MAP kinase pathway. Curr. Biol. *6*, 598-605.

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.

Burke, D. T., Rossi, J. M., Leung, J., Koos, D. S., and Tilghman, S. M. (1991). A mouse genomic library of yeast artificial chromosome clones. Mamm Genome *1*, 65.

Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M. T., Ghetti, A., Riva, S., and Morandi, C. (1988). cDNA cloning of human hnRNP protein A1 reveals the existence of multiple mRNA isoforms. Nucleic Acids Res *16*, 3751-70.

Chen, J., Chang, S., Duncan, S. A., Okano, H. J., Fishell, G., and Aderem, A. (1996). Disruption of the MacMARCKS gene prevents cranial neural tube closure and results in anencephaly. Proc Natl Acad Sci U S A *93*, 6275-9.

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.

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. (1995). Genetic Map of the Mouse, Database Release 10: Whitehead Institute/MIT Center for Genome Research).

Dietrich, W. F., Miller, J., Steen, R., Merchant, M. A., Damronboles, D., Husain, Z., Dredge, R., Daly, M. J., Ingalls, K. A., Oconnor, 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., Page, D. C., and Lander, E. S. (1996). A comprehensive genetic map of the mouse genome. Nature *380*, 149-152.

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.

Doolittle, D. P., Davisson, M. T., Guidi, J. N., and Green, M. C. (1996). Catalog of mutant genes and polymorphic loci. In Genetic Variants and Strains of the Laboratory Mouse, M. F. Lyon, S. Rastan and S. D. M. Brown, eds. (Oxford: Oxford University Press), pp. 1807.

Fahrner, K., Hogan, B. L., and Flavell, R. A. (1987). Transcription of H-2 and Qa genes in embryonic and adult mice. Embo J *6*, 1265-71.

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

Fitz Patrick, D. R., Germain-Lee, E., and Valle, D. (1995). Isolation and characterization of rat and human cDNAs encoding a novel putative peroxisomal enoyl-CoA hydratase. Genomics *27*, 457-66.

Grüneberg, H. (1961). Genetical studies on the skeleton of the mouse: XXIX. Pudgy. Genet. Res., Camb. *2*, 384-393.

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.

Haldi, M. L., Strickland, C., Lim, P., Vanberkel, V., Chen, X. N., Noya, D., Korenberg, J. R., Husain, Z., Miller, J., and Lander, E. S. (1996). A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome. Mamm Genome *7*, 767-769.

Hamilton, B. A., Frankel, W. N., Kerrebrock, A. W., Hawkins, T. L., FitzHugh, W., Kusumi, K., Russell, L. B., Mueller, K. L., van Berkel, V., Birren, B. W., Kruglyak, L., and Lander, E. S. (1996). Disruption of the nuclear receptor hormone receptor RORα in *staggerer* mice. Nature *379*, 736-739.

Harden, N., Lee, J., Loh, H. Y., Ong, Y. M., Tan, I., Leung, T., Manser, E., and Lim, L. (1996). A Drosophila homolog of the Rac- and Cdc42-activated serine/threonine kinase PAK is a potential focal adhesion and focal complex protein that colocalizes with dynamic actin structures. Mol Cell Biol *16*, 1896-908.

Hawkins, T. (1996). Large-scale shotgun genomic sequencing. Personal Communication.

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.

Ivanova, V. S., Zimonjic, D., Popescu, N., and Bonner, W. M. (1994). Chromosomal localization of the human histone H2A.X gene to 11q23.2-q23.3 by fluorescence in situ hybridization. Hum Genet *94*, 303-6.

Kusumi, K., Frankel, W. F., Sun, E. S., and Lander, E. S. (1997). Positional cloning of a novel serine-threonine kinase candidate in the mouse *pudgy* region: Doctoral Thesis, Massachusetts Institute of Technology.

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.

Kusumi, K., Sun, E. S., Frankel, W. N., and Lander, E. S. (1997). The mouse *pudgy* mutation causes defects in somitogenesis and early embryonic lethality: Doctoral Thesis, Massachusetts Institute of Technology.

Lai, W. S., Thompson, M. J., Taylor, G. A., Liu, Y., and Blackshear, P. J. (1995). Promoter analysis of Zfp-36, the mitogen-inducible gene encoding the zinc finger protein tristetraprolin. J. Biol. Chem. *270*, 25266-25272.

Larin, Z., Monaco, A. P., and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc Natl Acad Sci U S A *88*, 4123-7.

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. (New York: Current Protocols).

Ma, Q., Wadleigh, D., Chi, T., and Herschman, H. (1994). The drosophila TIS11 homolog encodes a developmentally controlled gene. Oncogene *9*, 3329-3334.

Madsen, P., Rasmussen, H. H., Flint, T., Gromov, P., Kruse, T. A., Honore, B., Vorum, H., and Celis, J. E. (1995). Cloning, expression, and chromosome mapping of human galectin-7. J Biol Chem *270*, 5823-9.

Manser, E., Chong, C., Zhao, Z. S., Leung, T., Michael, G., Hall, C., and Lim, L. (1995). Molecular cloning of a new member of the p21-Cdc42/Rac-activated kinase (PAK) family. J Biol Chem *270*, 25070-8.

Martin, G. A., Bollag, G., McCormick, F., and Abo, A. (1995). A novel serine kinase activated by rac1/CDC42Hs-dependent autophosphorylation is related to PAK65 and STE20. Embo J *14*, 4385.

Mello, C. C., Schubert, C., Draper, B., Zhang, W., Lobel, R., and Priess, J. R. (1996). The PIE-1 protein and germline specification in *C. elegans* embryos. Nature *382*, 710-716.

Mohler, J., Weiss, N., Murli, S., Mohammadi, S., Vani, K., Vasilal, G., Song, C. H., Epstein, A., Kuang, T., English, J., and Cherdak, D. (1992). The embryologically active gene, *unkempt*, of Drosophila encodes a Cys3His finger protein. Genet. *131*, 377-388.

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.

Passananti, C., Felsani, A., Caruso, M., and Amati, P. (1989). Mouse genes coding for "zinc-finger"-containing proteins: characterization and expression in differentiated cells. Proc Natl Acad Sci U S A *86*, 9417-21.

Rosenberg, I., Cherayil, B. J., Isselbacher, K. J., and Pillai, S. (1991). Mac-2-binding glycoproteins. Putative ligands for a cytosolic beta-galactoside lectin. J Biol Chem *266*, 18731-6.

Russell, L. (1995). The pudgy mutation and its origins.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2 Edition, A. Irwin, ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schuler, G. D., and al., e. (1996). A gene map of the human genome. Science *274*, 540-546.

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.

Suzuki, E., Kojima, N., Yoshimura, K., Uyemura, K., Obata, K., and Akagawa, K. (1995). Cloning and sequence analysis of cDNA for a possible DNA-binding protein 5E5 in the nervous system. J Biochem (Tokyo) *118*, 122-8.

Takahashi, N., and Ko, M. S. (1994). Toward a whole cDNA catalog: construction of an equalized cDNA library from mouse embryos. Genomics *23*, 202-10.

Taylor, G. A., Thompson, M. J., Lai, W. S., and Blackshear, P. J. (1995). Phosphorylation of tristetraprolin, a potential zinc finger transcription factor, by mitogen stimulation in intact cells and by mitogen-activated protein kinase in vitro. J. Bio. Chem. *270*, 13341-13347.

Wilkinson, D. G. (1992). *In situ* hybridization: A practical approach (Oxford: Oxford University Press).

Youssoufian, H., McAfee, M., and Kwiatkowski, D. J. (1990). Cloning and chromosomal localization of the human cytoskeletal alpha-actinin gene reveals linkage to the beta-spectrin gene. Am J Hum Genet *47*, 62-71.

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.

Zhu, Z., Bao, Z., and Li, J. (1995). MacMARCKS mutation blocks macrophage phagocytosis of zymosan. J Biol Chem *270*, 17652-5.

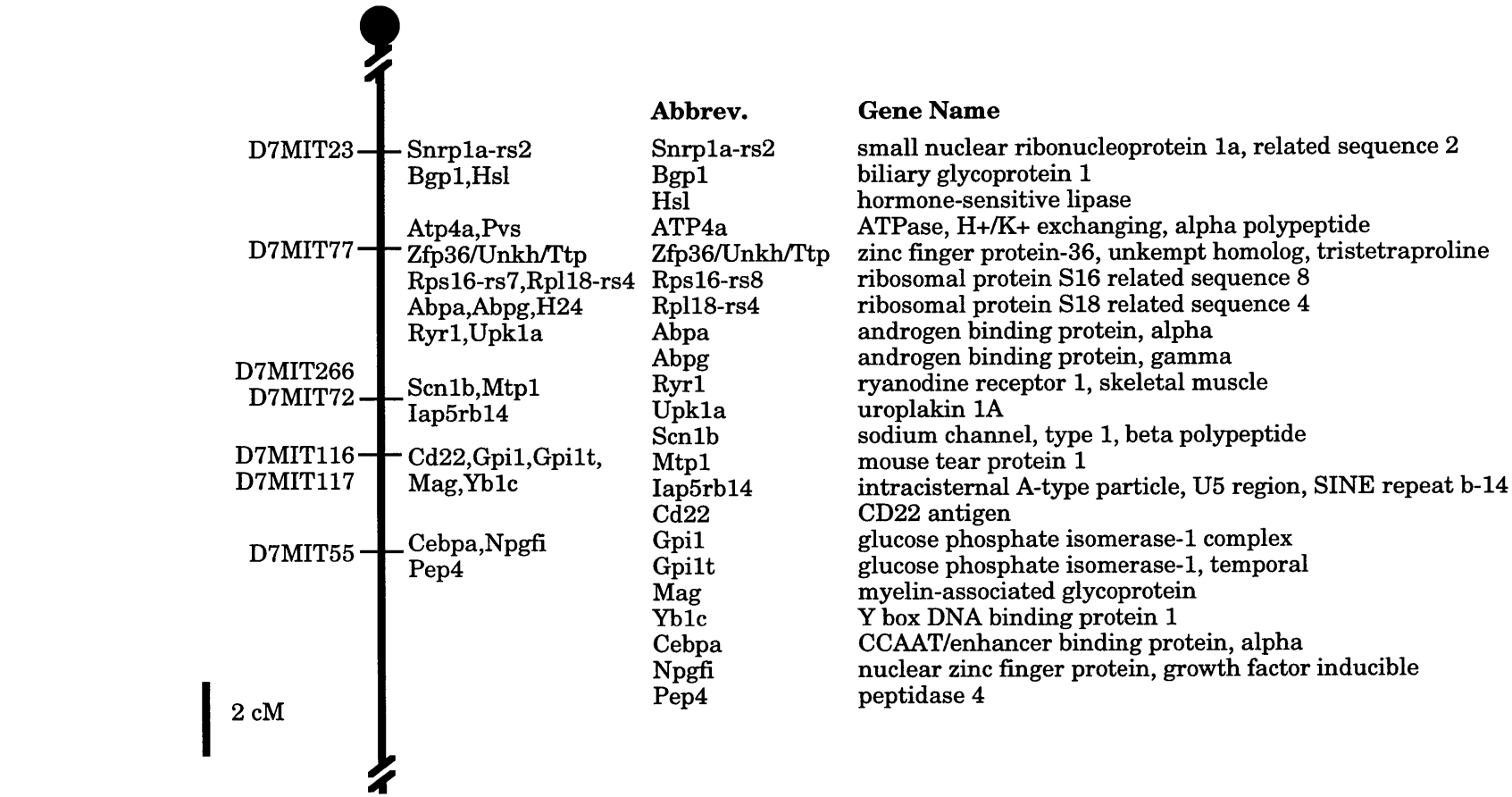



**Figure 1a** Genetic map of the *pudgy* region—SSLP and gene map
SSLP based markers from the WI/MIT Genome Center are shown on the left.
The consensus gene map from the Mouse Chromosome 7 Committee is shown on the right, with abbreviation translations.

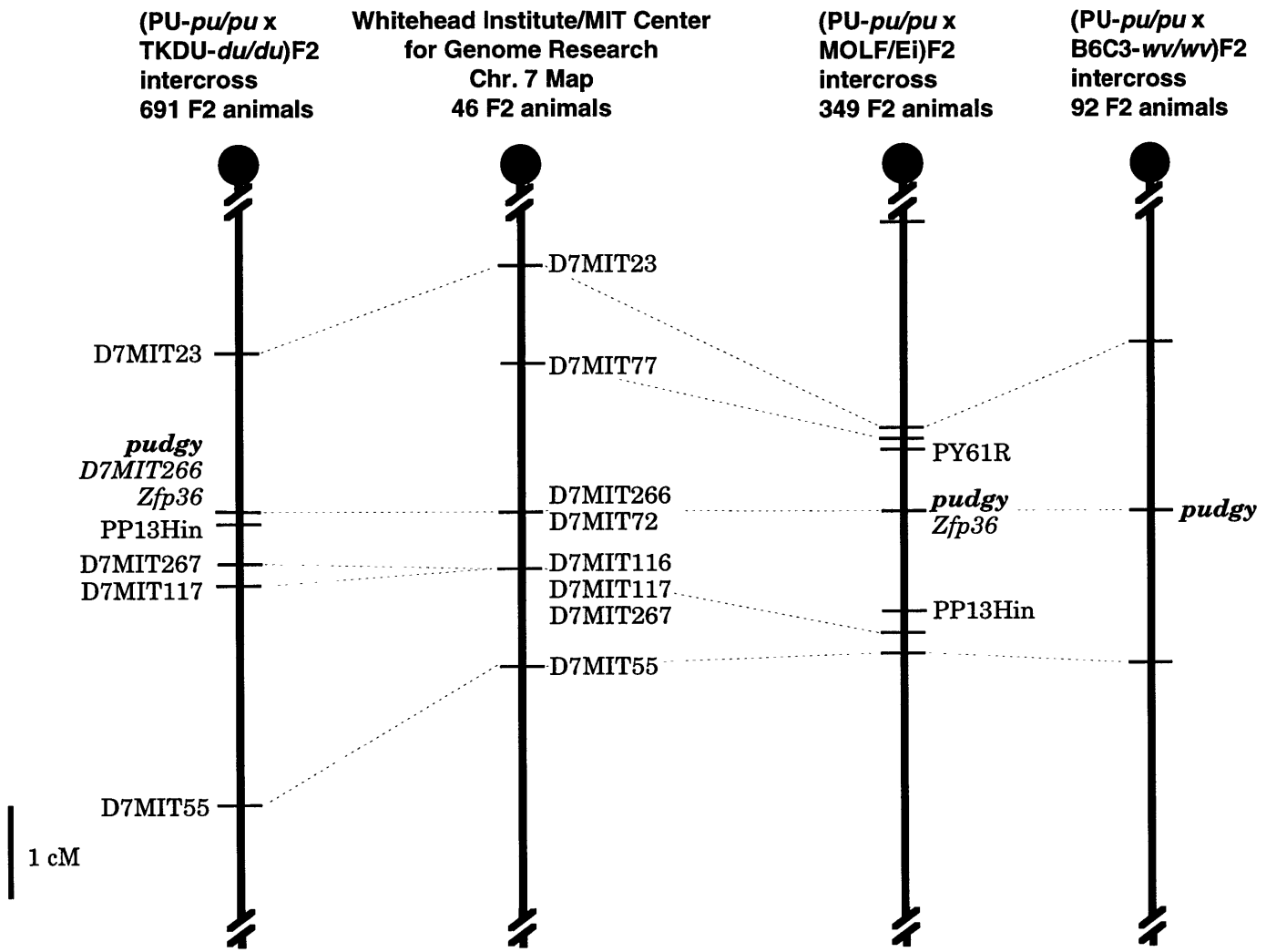



**Figure 1b** Genetic map of the pudgy mutation
The pudgy mutation has been mapped through the three intercrosses, totalling 2264 meioses. Results from the (PU-*pu*/*pu* x TKDU-*du*/*du*)F2 intercross is shown on the left of the WI/MIT Center for Genome Research map. Results from the (PU-*pu*/*pu* x MOLF/Ei)F2 and (PU-*pu*/*pu* x B6C3-*wv*/*wv*)F2 intercross are shown on the right.

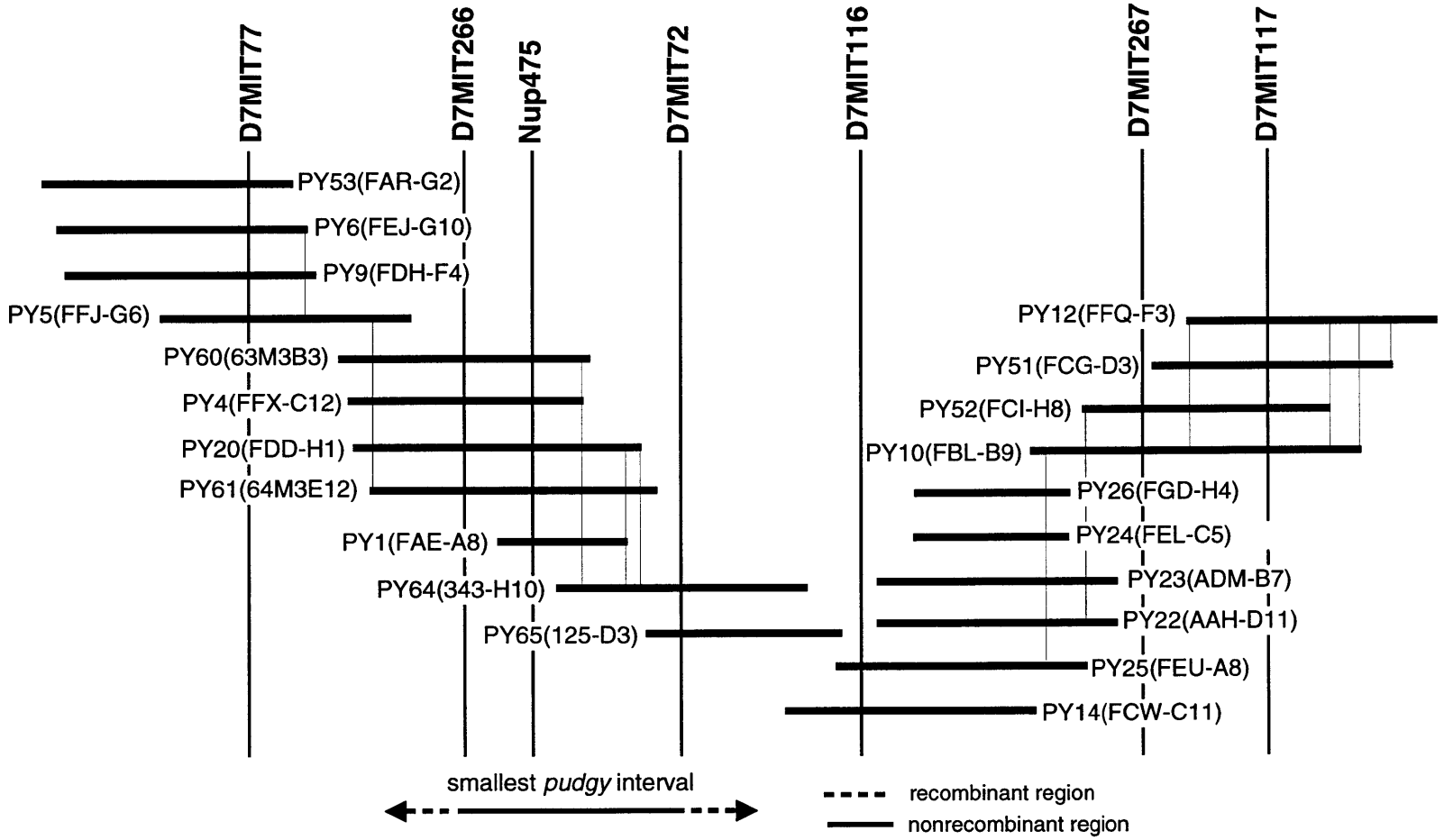



**Figure 2** Large-scale YAC-based physical map of the *pudgy* region.
YACs shown above are from the WI/MIT Mouse YAC Library I (Kusumi et al., 1993), Library II (Haldi et al., 1996), and the ICRF Mouse YAC Library (Larin et al., 1991). The genetic distance between WI/MIT markers D7MIT77 to D7MIT117 corresponds to 2.3 cM on the WI/MIT mapping panel cross. Vertical lines connecting YACs represent clone-end based assays. YACs are drawn to scale where size information is available, otherwise drawn at average size (700kb).

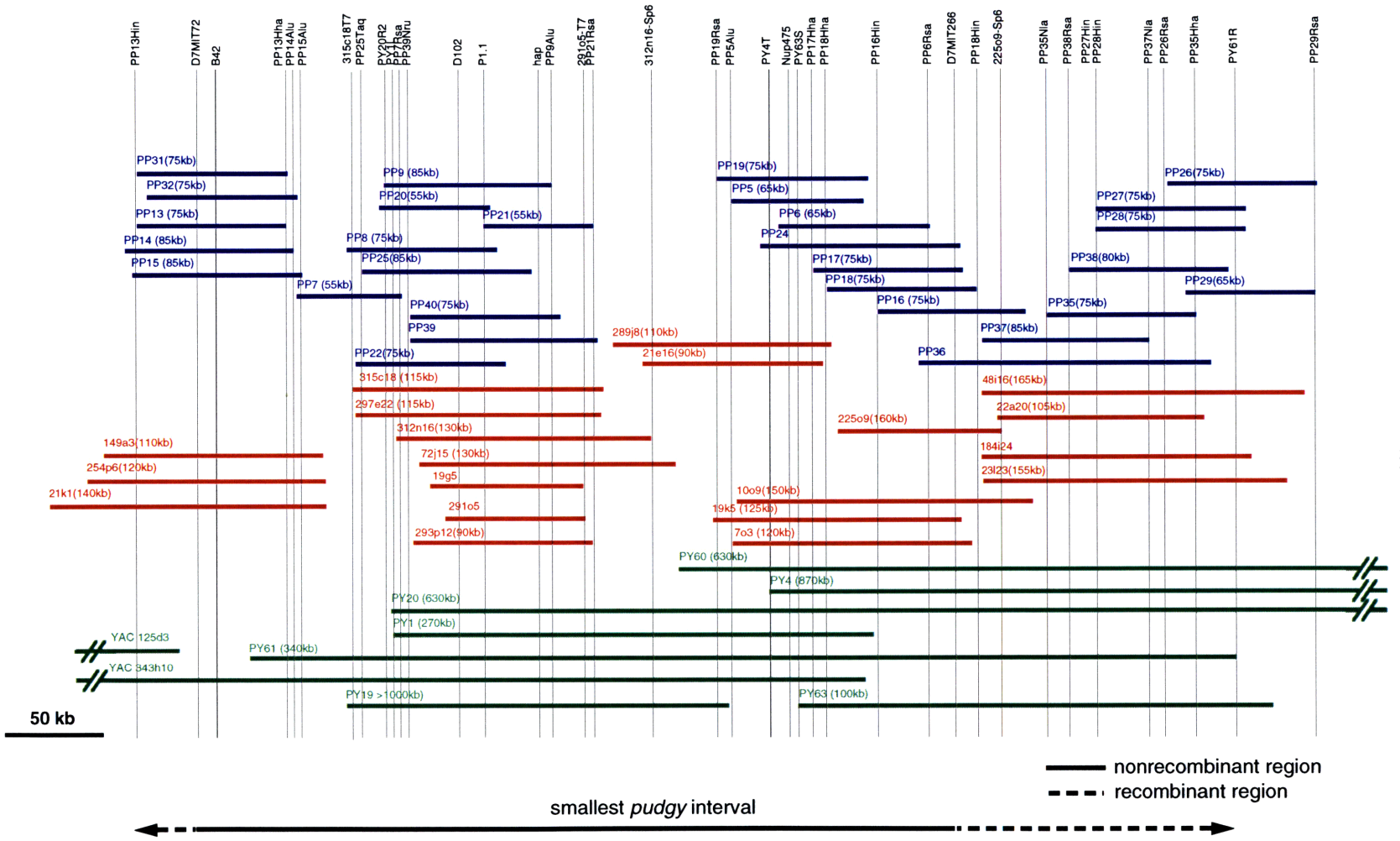



**Figure 3a** Fine-scale map of the mouse pudgy region.
P1 clones (average insert size 80kb) are shown in blue. BAC clones (average insert size 100-200kb) are shown in red. YAC clones (average size 680 kb) are shown in green. Vertical lines indicate STS content of clones as assayed by a selected group of PCR-based markers. Primer sequences for genomic sequence-based markers are listed in Table I and for cDNA selection fragments are listed in Table 2.

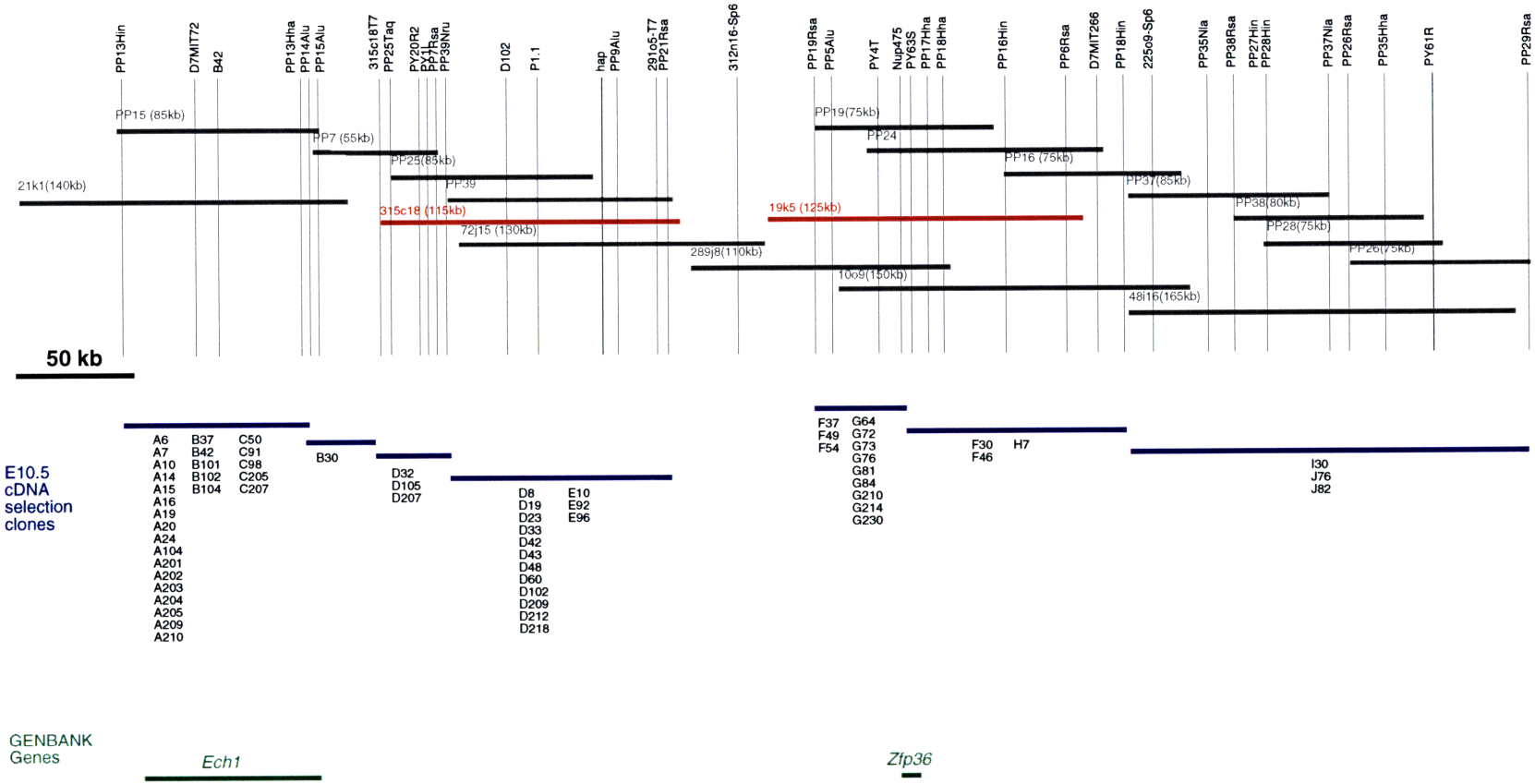



**Figure 4a** Transcriptional map of the mouse pudgy region from E10.5 cDNA selection clones. Minimum tiling P1 and BAC clones are shown above in grey for reference. BACs selected for genomic sequencing (315c18 & 19k5) are shown in red. Vertical lines indicate STS content of clones as assayed by a selected group of PCR-based markers. Primer sequences for genomic sequence-based markers are listed in Table I and for cDNA selection fragments are listed in Table 2. E10.5 cDNA selection fragments which mapped to the *pudgy* region are grouped above. Two GENBANK entry genes, *Zfp36*, a zinc finger protein, *Ech1*, peroxisomal enoyl hydratase, are indicated for reference.

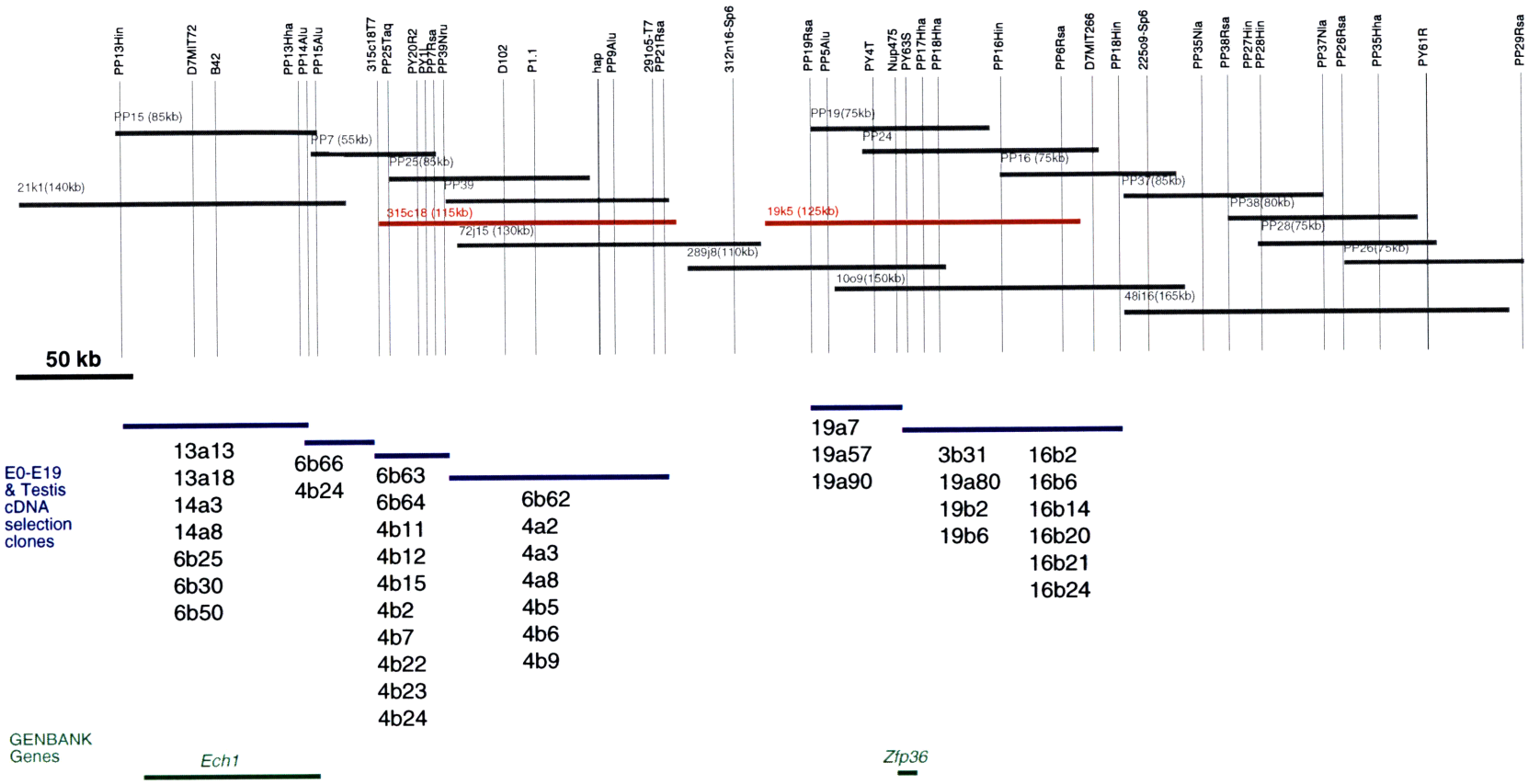



**Figure 4b** Transcriptional map of the mouse pudgy region from E0-E19 and testis cDNA selection clones. Minimum tiling P1 and BAC clones are shown above in grey for reference. BACs selected for genomic sequencing (315c18 & 19k5) are shown in red. Vertical lines indicate STS content of clones as assayed by a selected group of PCR-based markers. Primer sequences for genomic sequence-based markers are listed in Table I and for cDNA selection fragments are listed in Table 2. E10.5 cDNA selection fragments which mapped to the *pudgy* region are grouped above. Two GENBANK entry genes, *Zfp36*, a zinc finger protein, and *Ech1*, peroxisomal enoyl hydratase, are indicated for reference.

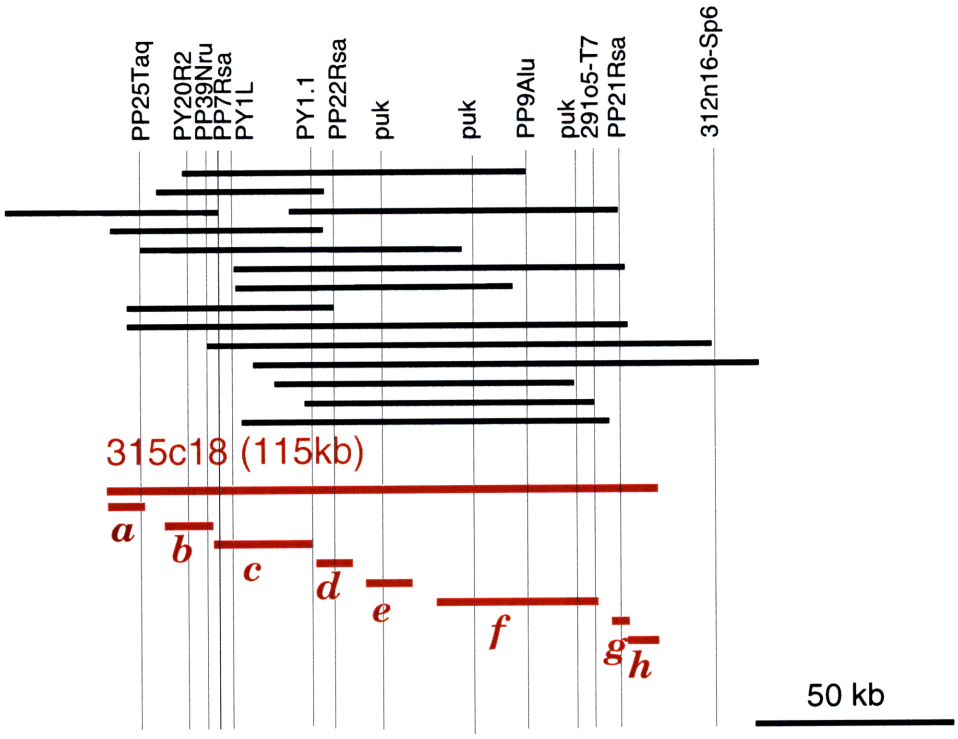



**Figure 5** Alignment of genomic shotgun sequencing contigs from 315c18 with sequence tagged site-based physical map. Sequence fragments are shown according to scale.

**Figure 5** Contigs *a* and *b*. Trancript fragments are shown in red.

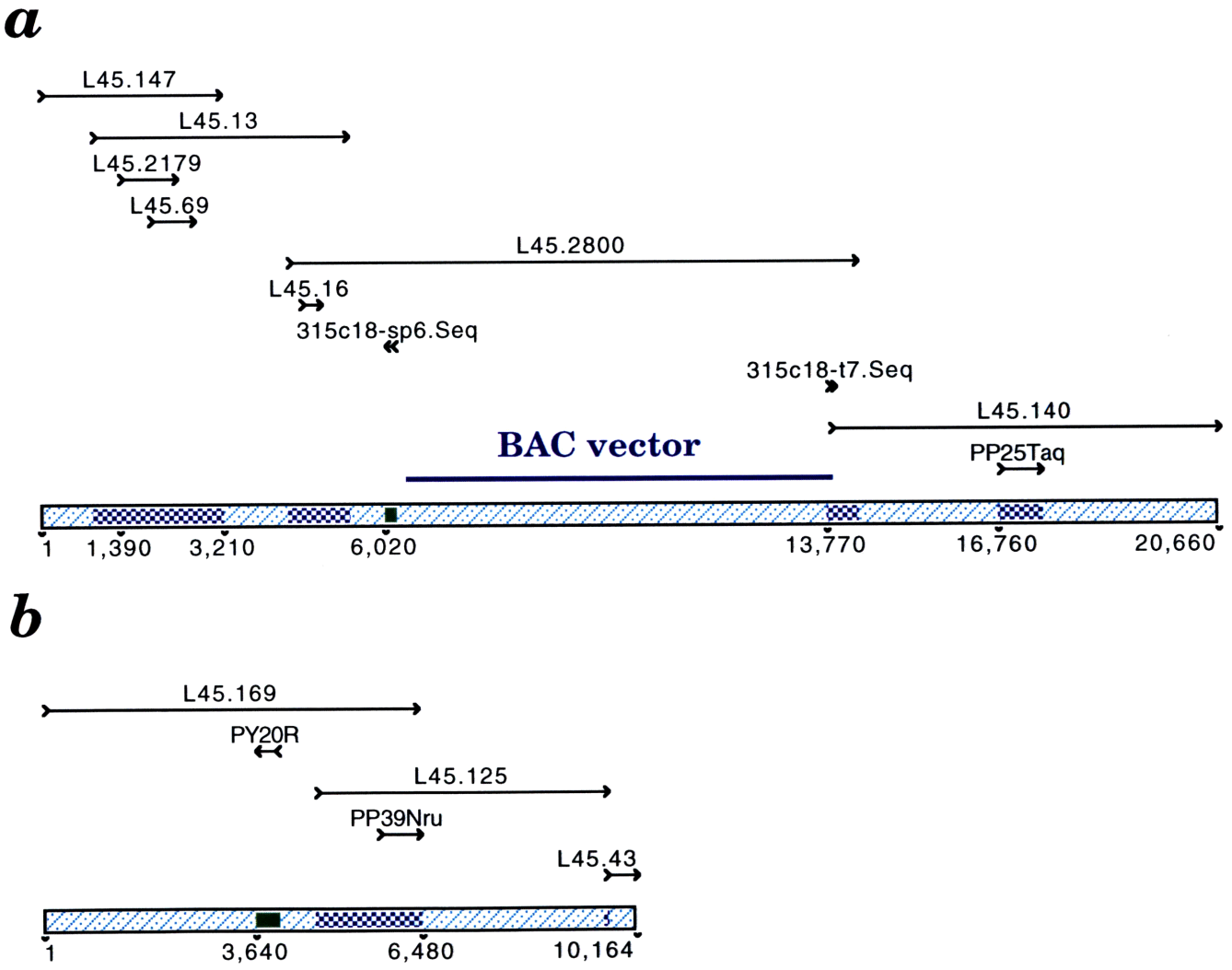
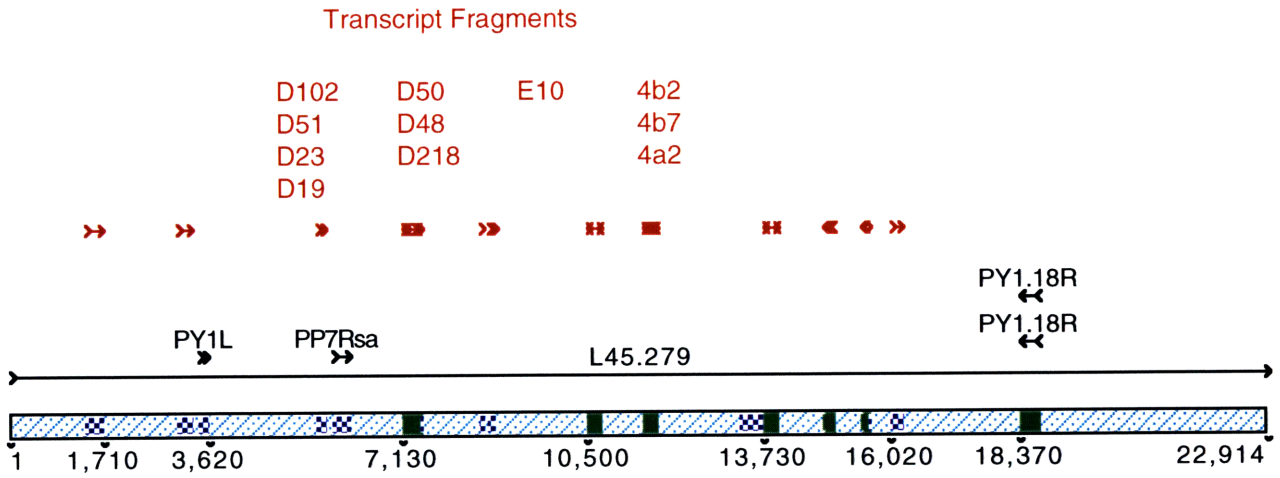


**Figure 5** Contigs *c* and *d*. Transcript fragments are shown in red.

***c***




***d***

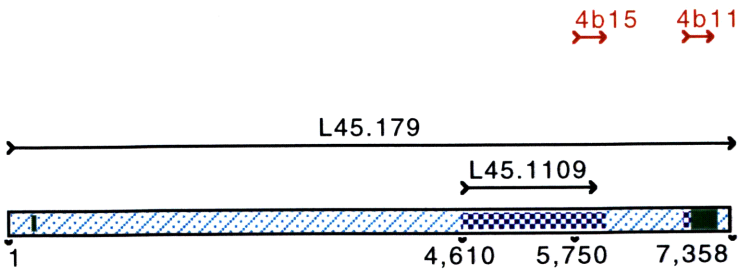


**Figure 5** Contigs *e,f,g*. Transcript fragments are shown in red.

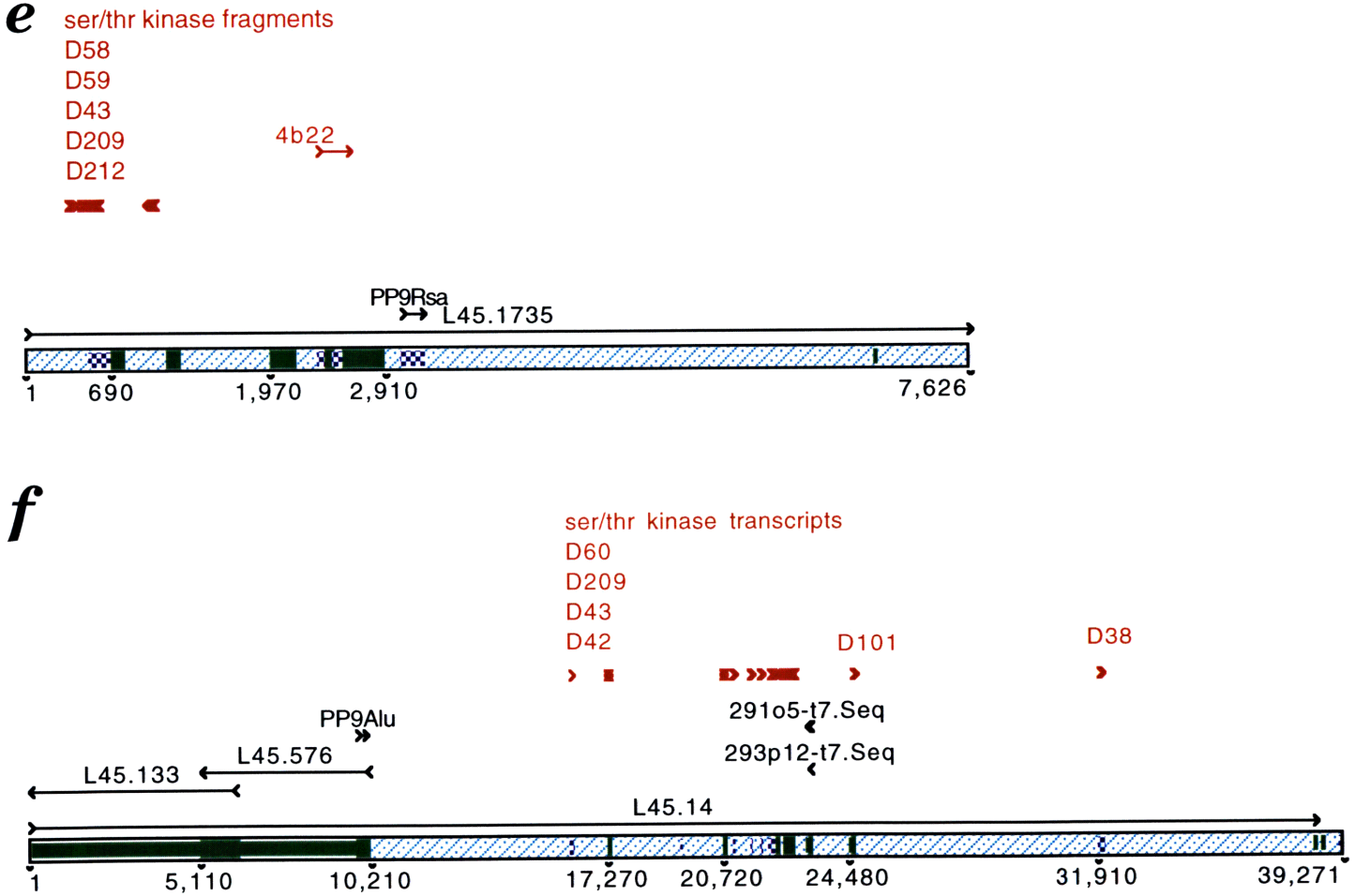


**Figure 5** Contigs *g* & *h*. Transcript fragments are shown in red.

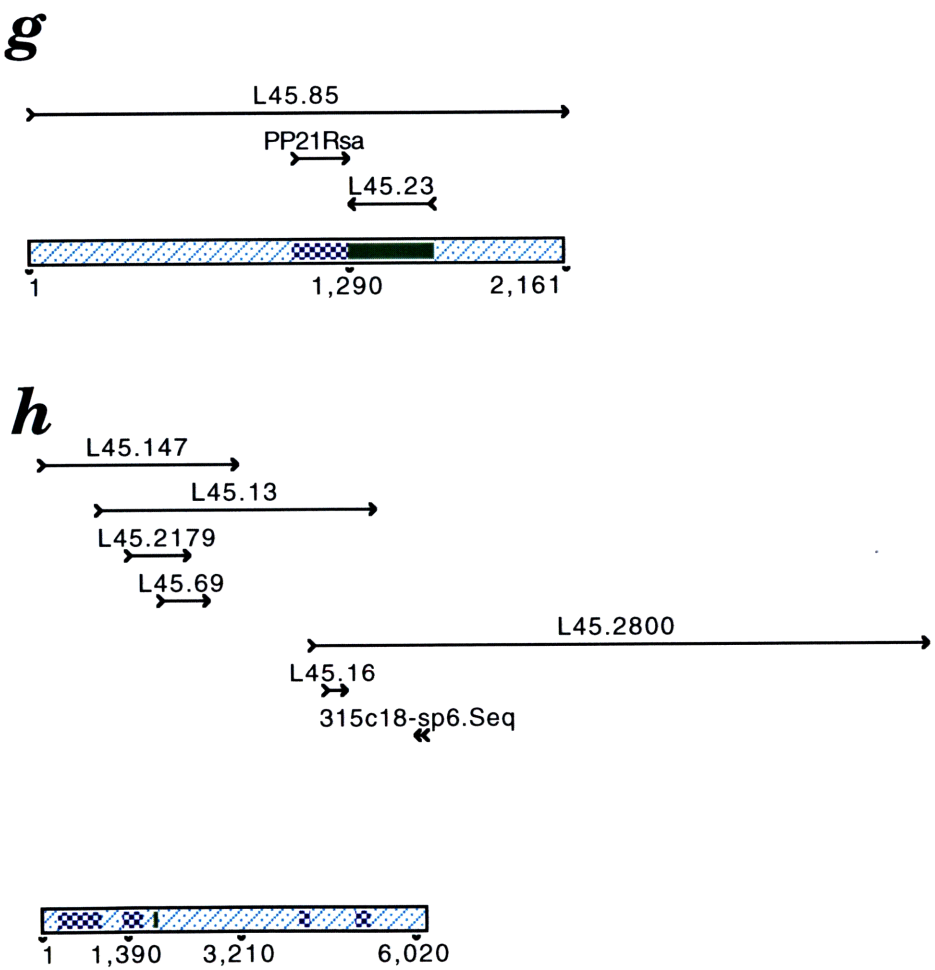
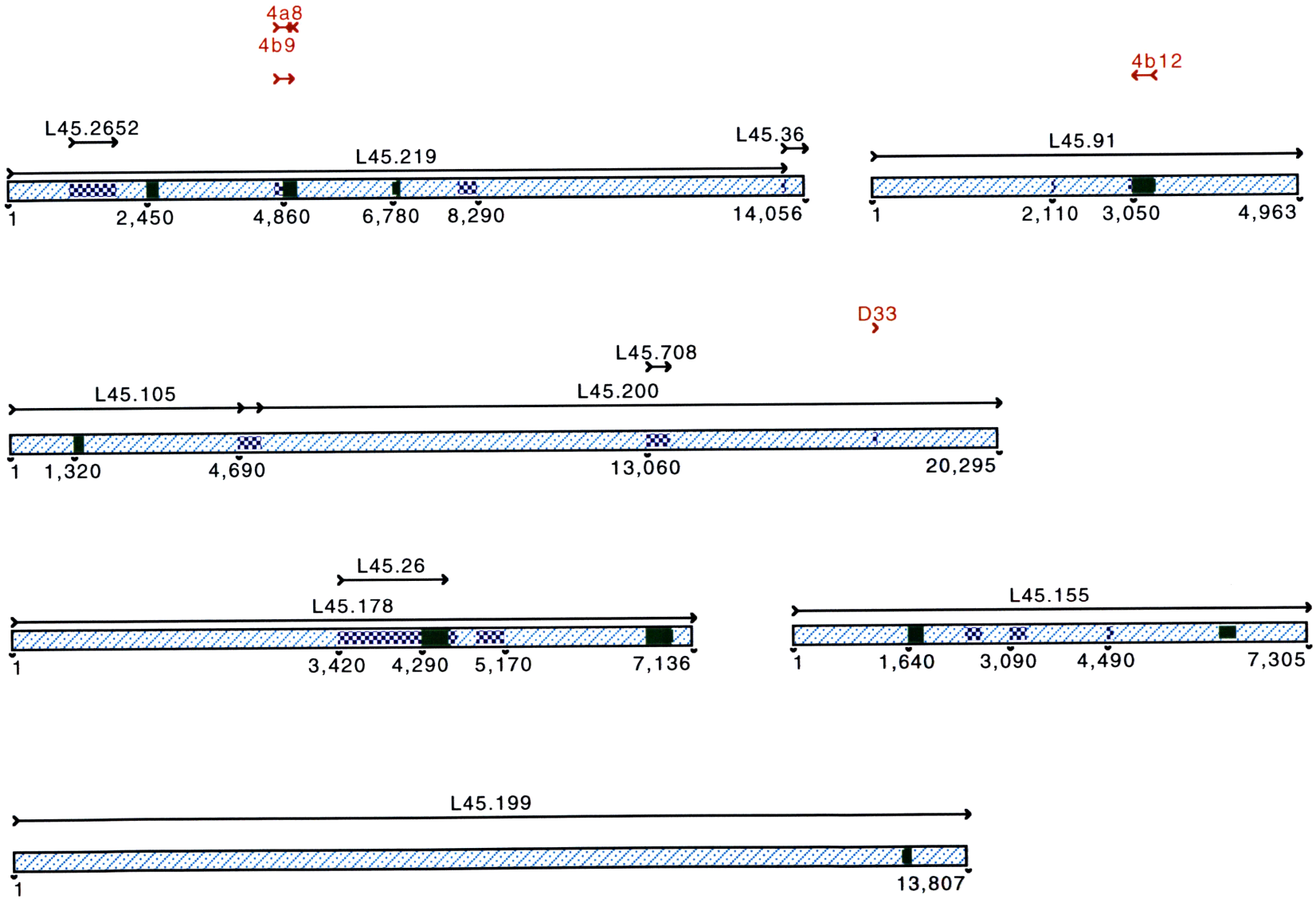


**Fig. 5** Unanchored contigs from BAC 315c18. Transcript fragments are shown in red.



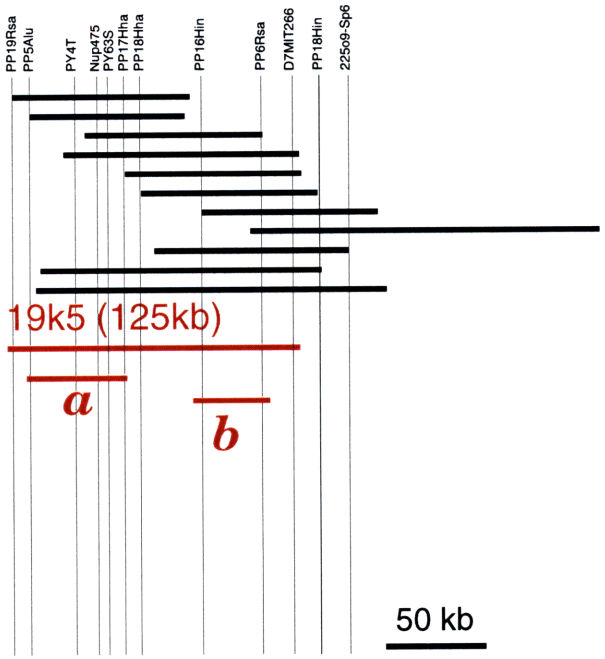



**Figure 6** Alignment of genomic shotgun sequencing contigs from 19k5 with sequence tagged site-based physical map. Sequence fragments are shown according to scale.

*a*

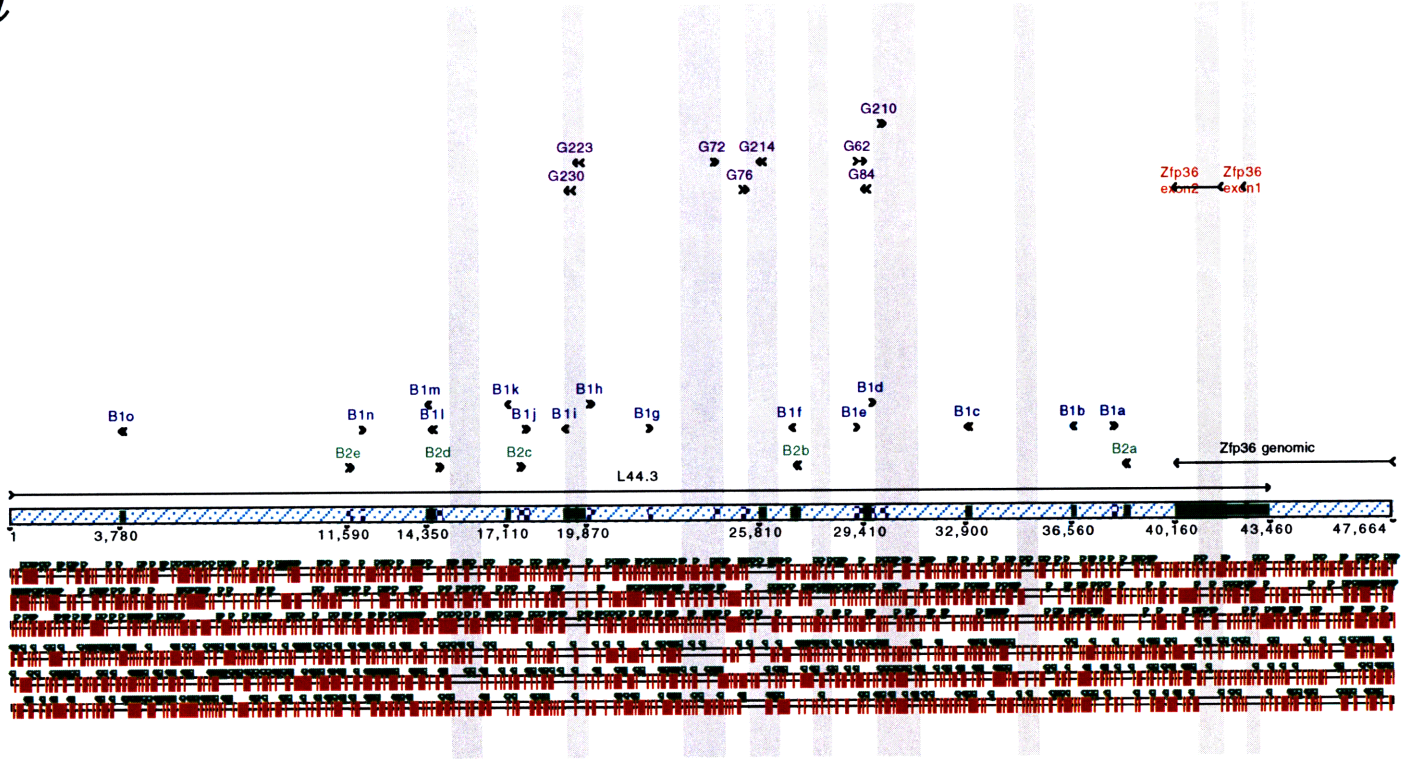
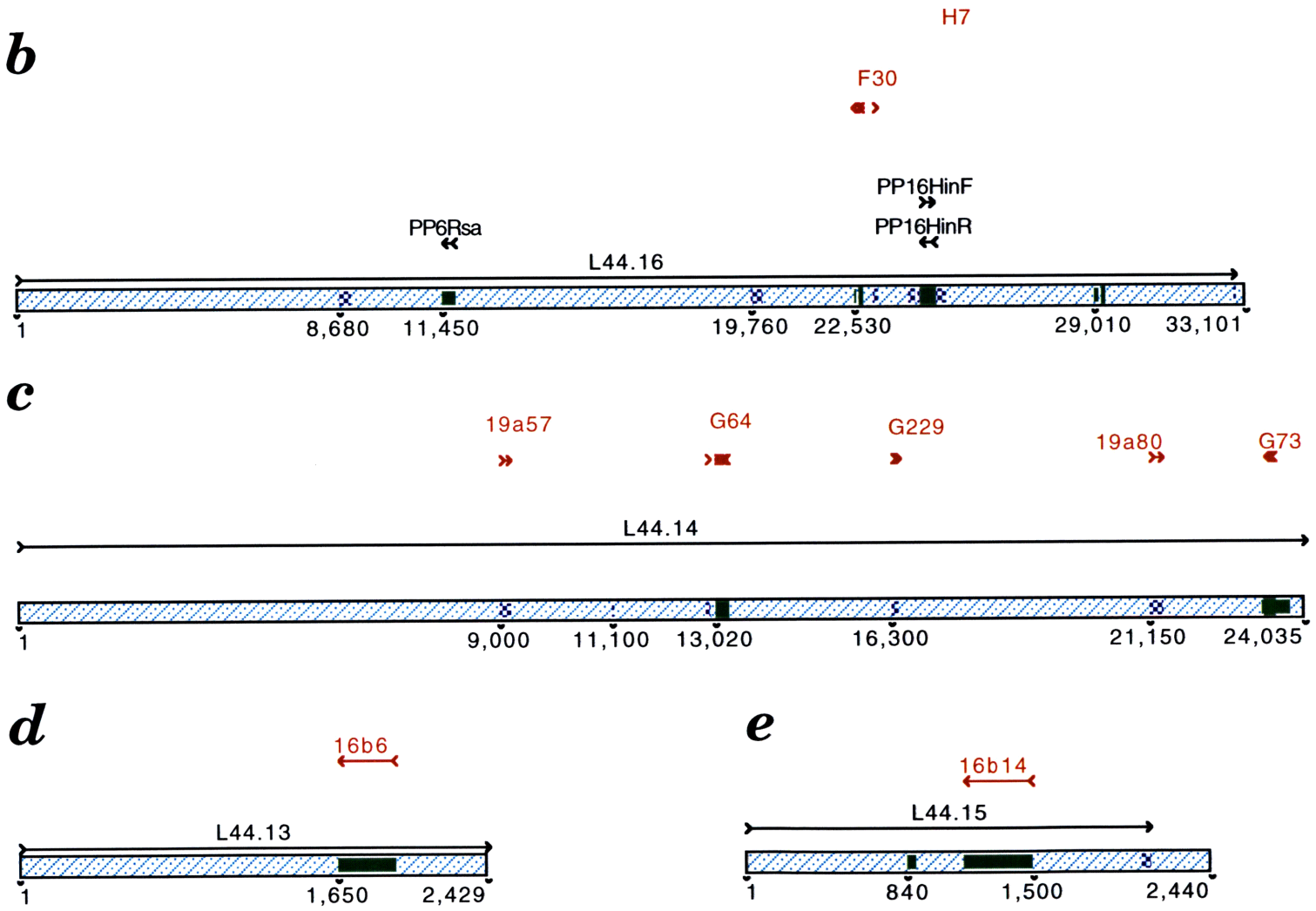



**Figure 6b** Genomic region of *Zfp36* gene
B1 and B2 SINE elements are indicated next to the consensus framework. ORFs in preliminary genomic sequence are indicated by the spaces between stop codons in all six possible reading frames. The largest ORF regions are indicated by light grey bars. Exons identified from *Zfp36* and by cDNA selection fragments are indicated above.

**Figure 6c** Selected contigs from BAC 19k5. Transcript fragments are shown in red.



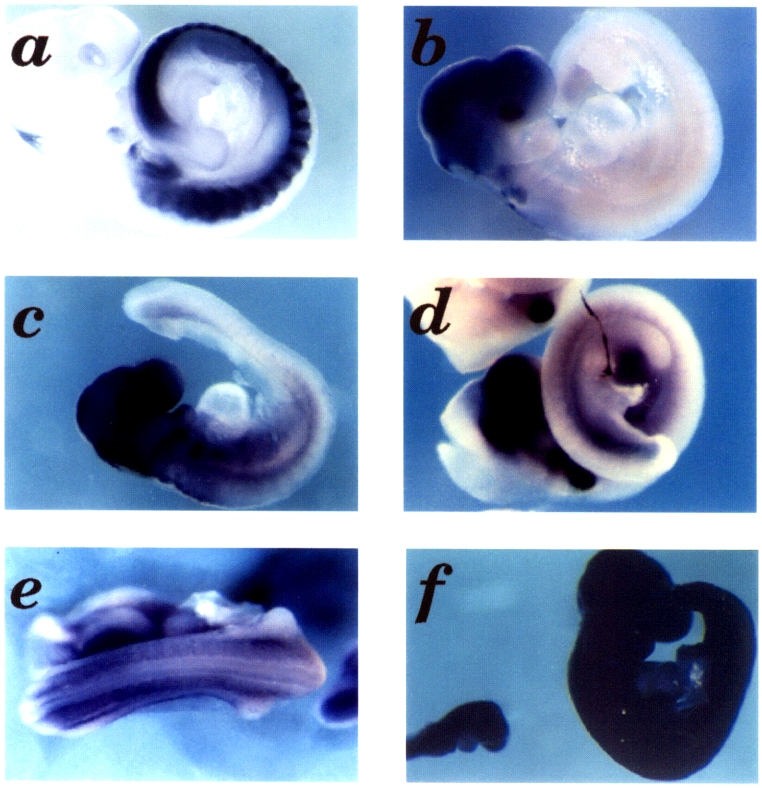



**Figure 7** Large-scale *in situ* hybridization screen, selected examples.
*a* Control *Mox1* probe which hybridizes selectively to somites at E9.5. *b* Previously identified *Zfp36* gene. *c* Probe D42 *d* Probe D51. *e* Probe D212. Direct cDNA selection fragments D42, D51, and D212 were found to be fragments of a larger novel serine-threonine kinase gene, thus displaying similar hybridization patterns. *f* C207, an example of a probe which hybridizes to a ubiquitously expressed gene.

**Table 1**
Primer sequences for sequence tagged sites (STS) in the pudgy region.

| Assay | Forward primer | Reverse primer | size (bp) |
|---|---|---|---|
| PY1.18X | TCGGGACCACAGATAGCAG | TGACCCCATTGCTGTTGTTA | 163 |
| PY1.1X | TGACCCTTCAATGGCTCAG | GAGACGACACTATGAAGAAAGG | 218 |
| PY1.2E | AATAGCAGTCCAGCGAGGAA | TACTCTGGATTGAGTCTTACGG | 149 |
| PY1.1E | GTGTTCCCGCCACATCTC | TCTGACTATCAGGGGGCATC | 208 |
| PY1L | TGTCCCCCTCACTCTGGTAC | GCATTCGAAGTATTCTCACCG | 108 |
| PY20R2 | GGACCTGGAACCCAGGAC | ATTTCCAAGCTGGCTTCCTT | 164 |
| PY20T2 | GAGGAGAGCAAGAGGCCTG | AGCATATCAAATGGAATAGGGG | 194 |
| PP13Hha | GCAAGCTATGGCAGGAAGAG | CTCCCCTGAGCCTCTTCAAT | 131 |
| PP13Hha2 | CTACATCACCCTCTCAACCTCC | ACTTTCTTATCTCCCCTGAGCC | 98 |
| PP13Hin | TCTGTCTCCACTGTCCCACA | GACTTATGGCTGGCCTCATG | 157 |
| PP14Alu | ATGAGCAAATGCGCAAAAC | CCTAAGATGGCTTCCAACTCC | 145 |
| PP14Alu | TGTAGGTCTGTGAACATTGTGC | CAAAGCAAAACCAAGCAACA | 135 |
| PP15Alu | AAGCACAGGGGTTATGCATC | CCCCAGACTTGTTTATGAAGC | 182 |
| PP15Hha | TGGCAGCTAGTGATGACGTC | AAGCGGCTCTGACACACC | 209 |
| PP15Rsa | TCGAGAGGGCTGCAATTACT | GTCCCAGTCCCAAACAGCT | 122 |
| PP15Rsa2 | AAGGTTCACCCCTCCTCACT | GCAAAGTAATTGCAGCCCTC | 92 |
| PP16Hin | GAGGCAGAGGATGAGAGGTG | TGATGTAGAATGTCTTGTGGCC | 163 |
| PP17Hha | AGCAAGTCTGAAGTTAGCACG | GCCTGTGGAAACGGGAAG | 140 |
| PP17Hin | CCACCTCTGAAACCACACCT | GTTCTCTTAAAGGCCAATGTGG | 145 |
| PP17Hin2 | ATATATGAGCCAATTTGGGGC | CCACACCCCATAAACATAATCC | 190 |
| PP18Hha | GAAAAGCCAAGATATCAGGCC | CAGGCCATGGCAGTAGTTTT | 136 |
| PP18Hin | CCTCATACTGTCAAAGCCAGTC | AAAGCTTTCCAAACATGTCTTG | 83 |
| PP19Rsa | TTGCTGTGGCAATTGAGCTA | TCTTTTGTGCTCCAGGCC | 182 |
| PP20Hin | GGACTACGCCACAACAGGAT | CCACAAGTCTACAATAGCACCC | 167 |
| PP21Rsa | TCTGTGACAAGGTCTAAGTGGC | GGTGCAGACATATAACCTCAGC | 91 |
| PP22Rsa | AACAGACACCATGACAAAGGC | AGCTCCTCCTGCACCATG | 136 |
| PP23Rsa | GCCTGTCCTGTATAAAAGCCC | ACCCTTCCACCTATGGGTTC | 131 |
| PP24HpaR | GCTGCCCTTCAGAGCCTTC | GACTATAGGGAGAGGATCCCA | |
| PP24Taq | TGTGGGGTCTCTGTAGCCTC | CTTGAGAGCCTTCAACCCAG | 164 |
| PP25HpaR | CAAATGGAAATGGATGGCAT | CCCATAACCATGCGCTTC | |
| PP25Nla | CATACAGAAGGAGACATAGCGA | CCTCCCCATTGCCTGAATC | 100 |
| PP25NlaR | CCTATGATGTGCCTGGCC | GAAACCAACATAGGTCAAAGGC | |
| PP25Taq1 | ATACCTGGGTGGACATATCAGG | ACTGGAGAGGTTAAAGTGCTGC | 251 |
| PP25Taq2 | TGTGGGGTCTCTGTAGCCTC | AAGCATCTGAAACATCTGGTCA | 104 |
| PP26Nla | AGAAGGAGACATGGACAGGC | GATCTTTCTCCCTCTCCATAGC | 100 |
| PP26Rsa | AACCATGCCAGCTTTCTCC | CATAGCTACCCAGCCTGCTC | |
| PP26TaqR | AAAGAGCTAGCAACTATTGCCA | TATAGGGAGAGGATCCCAATTT | |
| PP27Hin | CACTGTTTTACTGGCTCCCTG | CAGTTCCCACAGCTCCTAGC | |

**Table 1**
Primer sequences for sequence tagged sites (STS) in the pudgy region.

| Assay | Forward primer | Reverse primer | size (bp) |
|---|---|---|---|
| PP27Nla | TGAAGCTGCCCCTAGCTG | TGCCTCTGACAGTGGCTG | 102 |
| PP28Hin | GCGACTCCCACTGTTTTACTG | CAGTTCCCACAGCTCCTAGC | |
| PP28Nla | CACTGTTTTACTGGCTCCCTG | CAGTTCCCACAGCTCCTAGC | |
| PP29Nla | TGGCCTCTACAGTGTTCTTGC | ACCTTCAGATGTGGGCAGAG | 229 |
| PP29Nru | CCACAAGTTCATTCCGCTTT | TCCACAGTTGTGTTTGCCAT | 125 |
| PP29Rsa | TTATTAGCCCCCTTCCAGCT | TCTTGGCATGACTCACAAATG | |
| PP33Taq | CCATAACCATGCACTTCGGT | AAGCATCTGAAACATCTGGTCA | 125 |
| PP33TaqR | CCAGTCACCTCCTTCCAAAA | GAGGATTCCATATAGCATGCG | |
| PP34Taq | TGTGGGGTCTCTGTAGCCTC | CTTGAGAGCCTTCAACCCAG | 164 |
| PP35Hha | CCATAGTGACCAAGCCGG | AGCCCAGGGCTAAGAGTTTC | 171 |
| PP35Nla | CTGAAGGGAGAGACCATTTCA | GTTGGCTCCTCTCCCCTC | 93 |
| PP37Nla | TGTTCTCTCTCCACAGCCCT | ACCACTGAGCACGAAGCC | 158 |
| PP38Rsa | TTCTTCTCCTCTATCCCAGTGC | GTCAGATAGAGGCTTGAATGGG | 150 |
| PP39Hpa | CCTCCTCTAAGTCCCAACTACG | GGCATTTCACAAAGCATCTG | 100 |
| PP39Nru | CCTGGCCCTCACTCTCTTC | GCAGCCTTCCATACTCACTAGC | |
| PP40Hpa | ATGGGCTGTTTGCTGCTC | TGCAACTCGTAGGACAGGTG | 136 |
| PP5Alu | GGGACCAAGTTTGTGGCTTA | ACCTGACCCTGACCCTAACC | 229 |
| PP6Rsa | TGCTCTAGAGGATGCCCACT | CTGTGTGAGAGCCCCACC | 140 |
| PP7Rsa | GCCTGTCCTGTATAAAAGCCC | ACCCTTCCACCTATGGGTTC | 131 |
| PP8Alu | CCATACCTGGCTAGCCAAAA | ATGAAAATGTTCAACAGAGCA | 165 |
| PP8Hin | AAACCCAAACCCAGAGAGGT | CCAATTCTTCAGAAGGTGGC | 181 |
| PP8Rsa | GGCAGGTGGATCTGTTGAGT | TCCTTCCAGAACATCCCAAG | 165 |
| PP9Alu | CACCATCTGGTTCACCCTG | GCATACACTGATTTTGCAGCA | 200 |
| PY60L | ATGCATGAGCTTCTGGATCC | CGATGTCCAAAGAGCTATTTCC | 129 |
| PY60L2 | GTGTTGTTTCAGTTTCTTTGCC | CGATGTCCAAAGAGCTATTTCC | 101 |
| PY60S1 | TGTGATGCTTCTCATGTGAGTT | CCCCTTCATGGATGTATTGG | 110 |
| PY60S2 | ATGAGTGAAAAATGACAGCGG | CCCCTTCATGGATGTATTGG | 241 |
| PY61R | TCCTGTTCCCATCAGAAAGG | AGCAATTCTCACCTCTCTGAGG | 166 |
| PY61S | GGGAAATCATCACTCAGGGA | TGCAGGTCTGTGAACTATTTGC | 112 |
| PY61S2 | ACTCAGGGACTTCAGGCAGA | TGTGAACTATTTGCGATGGC | 93 |
| PY63S | CTCCATTTGGAATCTCCGG | CTGCAGATGCTCGAATTCTG | 90 |
| 19g5sp6-2 | AATGAGGATCCTCTGGCAAG | GGAAGGGACCCAGAATGTTT | |
| 19g5sp6 | CACAGCCATAAACTCTGCCA | TTGTCATTACGGACTCAAGTGG | |
| 19k5sp6 | GGGGTTCCCTCCTTCCTT | GTTTTTCCAAAACCCGCC | |
| 225o9sp6 | CTCCTAACATGTCCTGTTCACA | GAAAAGTCACGGAGCTCAGG | |
| 225o9t7-2 | AGGTATGGGCCTTGAATTCC | AGAAATGCAGGGATGGTTCA | |
| 225o9t7 | TGGGTGTTGGATTTTGTCAA | CCCAGAAATGCAGGGATG | |
| 291o5t7 | ACATGCCTGCTCCTAGGTGT | ATGTCCACTCTGGCCTGTTC | |
| 297c22 | AAGTTTTGGTGGGACCTGG | TTATTGACAGCGGATGGCTT | |

**Table 1**
Primer sequences for sequence tagged sites (STS) in the pudgy region.

| Assay | Forward primer | Reverse primer | size (bp) |
|---|---|---|---|
| 312n16sp6 | AGTCACAGAACTTGATGCTCTG | TCAATAAGTGTGCGTTCAAGC | |
| 315c18sp6-2 | AGTTCTCCGGTCCTTACGGT | TGGATCCACCTGAGTTTGGT | |
| 315c18sp6 | TGCTCCTCTTGAGGGCTAGA | ATGTGCATGAAAAGCCTCG | |
| 315c18t7 | TTTGTGGGACCTGGGACTAG | ATTTATTGACAGCGGATGGC | |
| kk4k3-F | AAGGGACAGACAAGTGGTGG | | |
| kk4k4 | AGAGTGGGTGAAGCTCTGGA | GTACTAGGTGAGGGGCTCCC | |
| kk4k4rev | TCTGACCGTCAGGTCTCATG | AGGCCTAGATAGTTCTCTGGGA | |
| L45.105 | GTTGTTTTCTGACTTCCACAGC | CACTGAGCTATCTCTCCATCCC | |
| L45.105-218F/550R | AACGATGGGGACCTAATTCC | AGCCAGGGGTCCAACTTC | 333 |
| L45.111 | AAAGGCACTTGTTGACAAACA | GGACAGCCAGGGCTATACAG | |
| L45.111-228F/588R | GCCTATGGATGCCAAAAAGA | TTCTCTGTGTAGCCCTGGCT | 361 |
| L45.125-190F/524R | CTGGACCCAGATGTGAGGTT | CCCTTGGCTTTTGTTGTTGT | 435 |
| L45.13-429F/802R | AATGGTCTCCACACTTTGTGC | CACACAGGAGGGTGAAGGTT | 374 |
| L45.133-726F/1002R | TGAGTTCCAGGACAGCCAG | GTGGGAGGAGACCAGATTGA | 277 |
| L45.140-24F/348R | TGGCAGCAAGCACCTTTAC | GGTTATTAGGAGGAGGTCCCC | 325 |
| L45.147-108F/318R | TTGGACTGGGTCTCTTCACA | CTTCCAGAGTCCTGCCATTG | 211 |
| L45.155 | GTAGCCGTACTCCTTCACGC | CATTTTCTTGGAGGCCGC | |
| L45.155-322F/576R | AGAGGACCTAGGTTCAAGTCCC | AAGAGTTCCCACAGGACCCT | 255 |
| L45.169-28F/281R | CTCTCCTTCACTCACTGCCC | GCTACCAGGAATCTGAAATTGC | 254 |
| L45.1735 | GGCCAGCCAGAGCTATGTAG | CAGGCTGGCCTAGAACTCAC | |
| L45.178-586F/909R | TGAGCCAGTGGAAGTGAGG | GTGGCAGGTACACACCAGG | 324 |
| L45.188-1f | GACAGACAGACACGCATACACA | GGACAGATCCCGTGGATG | |
| L45.188-517F/779R | GAATTCACTGTGGAGTTAGGGC | CAAACTAGCTTAAGTGGGGGG | 263 |
| L45.199 | AGAGGAGCTGGACTTTCTTGG | CCCAGATGTTCACCCACAG | |
| L45.199-528F/804R | CACCTGGAGGCTTTAATCCA | ATGCAATTCAGAACATTTCCG | 277 |
| L45.200-152F/432R | GGTTGGTTTGGGAAACCC | GTCACCTTTTGCTCAGAGCC | 281 |
| L45.219-236F/634R | AAAGAAGTGAAAGCAGGCCA | AAGCTTTTAAGGTGCAGCCA | 399 |
| L45.2488-102F/434R | CTTCTGCACAGGTGACAGGA | TTATTCCACGGAGCAAGACC | 333 |
| L45.2488 | TCCACACACGTCTATTCCACA | TGTGCTCCAGTACAACACAGC | |
| L45.279 | GTCCGTAAGAGCACTTGTTGC | TTCTAGGACAGCCAGGGCTA | |
| L45.576-300F/662R | TGTACTTCCTGCCCAGCAG | CCTCCTGCAAGGCTTTACTG | 363 |
| L45.85-431F/568R | CAACCCTGTGTGACATCTGTG | TGTTCATCCCAACCCAAATT | 138 |
| L45.91-278F/629R | CCTCGTAACTGCCATTCTTACC | ACTGCCACACATGCATCATT | 352 |

**Table 2**
Primer sequences for cDNA selection/cDNA derived fragments in the pudgy region.

| Assay | Forward primer sequence | Reverse primer sequence | Product size (bp) | Most significant BLAST score is shown. All scores at probability lower than 10^-10 are listed BLASTN match | BLASTX match | dBEST match |
|---|---|---|---|---|---|---|
| 16B14 | ATGCTCACCTCTGGCAGAGT | CCATTGATTCCTGCCTGG | 185 | none | none | none |
| 16B20 | AGTCCCAGGGTTTGGGAG | TGTTTTGTCTGGGAGCACAC | 151 | none | none | none |
| 16B21 | ATCAAAGCTGAGTTGCAGCA | TATCTCCTGCCTGCTGGTG | 156 | none | none | none |
| 16B24 | TGGCTAAGAAGGTCCGAAGA | GAGCAGAAGTTGTCCTTGGC | 107 | none | none | none |
| 16B2 | ACATTTTTATTGTGTGCCTCTG | GAGTCTACATGACCAGCGGA | 126 | none | none | none |
| 16B6 | ACCACTCTGAGTCCGGGTC | TGAATAGGAAAGCCCAATGG | 134 | none | none | none |
| 19A57 | TGGCCACAGTTGAAAGAAGA | TGTCCCAAACCTACCAGGTG | 80 | none | none | none |
| 19A7 | GCAGCTAGGGATATGGAACC | TCCATTTTTATGCCTGCACA | 142 | none | none | none |
| 19A80 | AGTCTTGGGCTGCTGAATGT | AATGCAAGAACTGTAGACCAGG | 126 | none | none | none |
| 19A90 | AGCCCTCAACAGTTGCTAGG | AGAGGATGATGGGCTGGG | 105 | none | none | none |
| 3b31 | CGACTACCCAGGATCGACAT | CAAGGAAAGACCTGAGGATCC | 106 | none | none | none |
| 4a2 | CTTGCCTTTTTTGCCAGAAG | TCAAGCAAGCACAAGGACC | 131 | none | none | none |
| 4a3 | GGGAGCAGGTGTGTGAGATG | GTTGACAATGACCAGACCCC | 127 | none | none | none |
| 4a8 | CATGCCACACTAACCTCCCT | TCTCTGGCTGCTCTGCTGTA | 106 | none | none | none |
| 4b2 | TTGTTGGCATGCACCATC | AGGCCATGTAACATACACTCCC | 101 | none | none | none |
| 4b11 | AGCTAGTGCATCAACCATTGG | CTGATCTAATTTGATCCTCCGC | 120 | none | none | none |
| 4b12 | GCCAATGAATCAGGGTCTGT | GAAGATTGTGGCTGTGTCTTTG | 115 | none | none | none |
| 4b15 | CTGGAAGCTTTGCTTTCCAC | ATTGAAAGGTCTGCTTCTCCC | 115 | none | none | none |
| 4b24 | GGCAACTAGAATATGGACAGGC | AGCAAGTCAGTGCTCTTAACCC | 113 | none | none | none |
| 4b22 | CCTTTAATCCCAGCACTCCA | TGGCTTCTTTCTGCAAGTTG | 258 | dbjID10914IHAMHG5 Syrian hamster gene for cytochrome ... 371 2.1e-21 | none | gbIAA111730IAA111730 mp54c10.r1 Soares 2NbMT Mus muscu... +2 89 1.2e-14 |
| 4b23 | CTGACGACCCTCTTCTGTCC | CATGGGAGTGAGGCAGATG | 146 | none | none | none |
| 4b24 | GGCAACTAGAATATGGACAGGC | AGCAAGTCAGTGCTCTTAACCC | 113 | none | none | none |
| 4b5 | GCTTCTACCTGCCCAGGTC | ACAAATCTGAAAACTTCCTGGC | 115 | none | none | none |
| 4b6 | TGGCCTTCAAAACTCAAAGG | CCTTGCAAACAAAGGCTCTT | 112 | none | none | none |
| 4b7 | AGCCACCTGTACTCCACACC | GATGGGGATCATCCCTCC | 114 | none | none | none |
| 4b9 | AGATCAGGAGCAGAGAGAGGG | CGCCTGTGCCTTTAACTAGC | 108 | none | none | none |
| A104 | GGTTCTCCACTCCAGAGCAG | CTTTTGCTGTGTCCCCAAAT | 175 | emblX16135IHSHNRNPL Human mRNA for novel heterogeneous nu... 705 1.6e-49 | HETEROGENEOUS NUCLEAR RIBONUCLEOP... +3 314 1.6e-36 | gbIW65179IW65179 md78h06.r1 Soares mouse embryo NbME1... 893 1.2e-67 |
| A10 | TCCCTTCATTGGGAATTACG | CAGCTAAGGATTTTGGGATCC | 140 | none | none | none |
| A14 | TTAGGGGGTGATTGGATTGA | CCCATCAGCTGACAGACAGA | 132 | none | none | none |
| A15 | TGAATGTACCAGAGCGAGCA | TCCCAGGTCACCTCATCAG | 182 | none | none | gbIR44284IR44284 yg34a09.s1 Homo sapiens cDNA clone 34... +1 122 4.7e-24 |

**Table 2**
Primer sequences for cDNA selection/cDNA derived fragments in the pudgy region.

| Assay | Forward primer sequence | Reverse primer sequence | Product size (bp) | Most significant BLAST score is shown. All scores at probability lower than 10^-10 are listed BLASTN match | BLASTX match | dBEST match |
|---|---|---|---|---|---|---|
| A16 | AAATACACCGCCTTCCCTCT | CAGTGATGCCCTGTTCCC | 139 | emblX16135lHSHNRNPL Human mRNA for novel heterogeneous nu... 366 6.5e-21 | HETEROGENEOUS NUCLEAR RIBONUCLEOP... +2 164 1.4e-15 | none |
| A19 | TTTCATCCAGCCCATCGT | TGCTCTGCGCTGCTACAC | 151 | dbjlD37934lRAT5E5ANTE Rat mRNA for 5E5 antigen, complete ... 699 4.5e-57 | none | gblR13142lR13142 yf72f11.r1 Homo sapiens cDNA clon... +3 183 2.4e-19 |
| A201 | GAGTGTGGGAGAGGAGATGC | TATCCCGGTTGCTAAGGTTG | 145 | none | none | none |
| A202 | AACTCATGCGGCCCTCAG | GTGAAGGTCTTCTGGTACTTGC | 85 | gblU08976lRNU08976 Rattus norvegicus Wistar peroxisomal e... 642 7.1e-46 | gil478984 (U08976) peroxisomal enoyl hydratase-like pr... +1 192 1.3e-19 | gblW09817lW09817 ma49f01.r1 Soares mouse p3NMF19.5... +3 163 8. |
| A203 | CTGGAGTAGCCCAGCATCTC | TAAAATGAGGGACAGTGCCC | 123 | none | none | none |
| A204 | CCTCCTGTCTGCACAGATCA | CCCTAAGCGGCCTAGGAG | 125 | none | none | none |
| A205 | TCCCCTCGTTGGGAATTAC | GATCCCTTGTACTGCCGTGT | 145 | none | none | none |
| A209 | AAGTGACGAGGGGTGATACG | ATCTAGGACAACCAGCCACG | 147 | gblU70494lMMU70494 Mus musculus histone H2A.Z (H2A.Z) m... 1287 1.8e-108 | splP08985lH2AV_DROME HISTONE H2A VARIANT. /pirlIS08118... -3 416 7.5e-58 | gblAA002360lAA002360 mg43e08.r1 Soares mouse embryo N... +3 481 1.3e-60 |
| A20 | ACCGATTGTTCCTTGATTGG | CTGTGGACCGAGCCATTACT | 162 | emblX16135lHSHNRNPL Human mRNA for novel heterogeneous nu... 877 2.4e-65 | HETEROGENEOUS NUCLEAR RIBONUCLEOP... -1 371 4.8e-46 | gblW65179lW65179 md78h06.r1 Soares mouse embryo NbME1... 1062 1.5e-83 |
| A210 | ACGCAAAGGTAGGTCTGGG | CACCTTTGTCCCACTGAGGT | 170 | emblX16135lHSHNRNPL Human mRNA for novel heterogeneous nu... 340 5.8e-37 | splP14866lROL_HUMAN HETEROGENEOUS NUCLEAR RIBONUCLEOPR... +1 140 1.9e-26 | none |
| A24 | TGCTGCATGTTAAGAGGGC | CAACCCCAGCACTACCATCT | 155 | none | none | gblAA048889lAA048889 mj44e02.r1 Soares mouse embryo NbME1... 377 4.3e-71 |
| A6 | TCTCTATCAACCCTGGGCC | GCCATTGTTGGGTCTGAGAT | 150 | none | none | none |
| A7 | TCCCTTCATTGGGAATTACG | GATCCCTTGTACTGCCGTGT | 144 | none | none | none |
| B101 | GCAGGCAGGGTATAGATTGC | AAAAGCTATGGACACACCGG | 152 | none | none | none |
| B102 | TGCTGGATGCGGTTCTTAG | AACCAGCCATTATGCCTGTC | 126 | emblX16135lHSHNRNPL Human mRNA for novel heterogeneous nu... 650 4.0e-88 | splP14866lROL_HUMAN HETEROGENEOUS NUCLEAR RIBONUCLEOPR... -2 254 1.3e-59 | gblW65080lW65080 me01e02.r1 Soares mouse embryo NbME1... 636 4.4e-116 |
| B104 | GTTGAGTGGCTGAACGGATT | CCTTGGGAAAGTCTCACCAG | 133 | none | none | none |
| B30 | AGCCAGAACACCATTTCCCT | TCAAGCGAGGCATCTACAAA | 140 | none | none | none |
| B37 | TATGTAATGCCATTGCTCGC | TGCAGCTAAGGACCAGTGG | 146 | none | none | none |
| B42 | CCCCTCTTGCTGTCTCTGTC | TGGCTGTGGTTAGAGCGAG | 146 | none | none | none |
| C205 | TGGGCTTCCTGAACCATTAC | TCCCAGCATCCAATTAGGAG | 106 | emblX16135lHSHNRNPL Human mRNA for novel heterogeneous nu... 394 3.9e-30 | HETEROGENEOUS NUCLEAR RIBONUCLEOP... +1 163 2.0e-18 | gblW48547lW48547 md26d04.r1 Life Tech mouse brain ... -2 153 8.5e-34 |

**Table 2**
Primer sequences for cDNA selection/cDNA derived fragments in the pudgy region.

| Assay | Forward primer sequence | Reverse primer sequence | Product size (bp) | Most significant BLAST score is shown. All scores at probability lower than 10^-10 are listed BLASTN match | BLASTX match | dBEST match |
|---|---|---|---|---|---|---|
| C207 | GACCTTGGGCTTCCTGAAC | TGGAGTCTTCCTTCCCAGC | 125 | emblX16135lHSHNRNPL Human mRNA for novel heterogeneous nu... 509 1.1e-34 | none | gblW48547lW48547 md26d04.r1 Life Tech mouse brain... -2 248 3.0e-28 |
| C50 | AGGCCGAGTACTGCATCG | AGGCCAAGGGAGTCTCTGTT | 150 | gblU19893lRNU19893 Rattus norvegicus alpha actinin mRNA... 714 1.9e-58 | gil1304604 (U41416) alpha-actinin [Mus muscu... +2 264 2.6e-31 | gblAA002732lAA002732 mg45d10.r1 Soares mouse embryo N... +3 298 2.0e-42 |
| C62 | AACCTCGGAAGTGTCCAGG | TGTCTGCTACCCAGCACAAG | 179 | gblL07769lHUMEPLCTN Homo sapiens galectin-7 mRNA, complet... 738 2.1e-54 | spIP47929lLEG7_HUMAN GALECTIN-7 (HKL-14). >gil182132 (... -1 341 4.6e-42 | gblAA104644lAA104644 mo55d06.r1 Life Tech mouse embry... +2 440 7.1e-55 |
| C91 | ACTCCGTTCTTCCGGAAAAT | ACTCCCGGAGCGTAAACAG | 136 | emblX16135lHSHNRNPL Human mRNA for novel heterogeneous ... 918 9.3e-69 | spIP14866lROL_HUMAN HETEROGENEOUS NUCLEAR RIBONUCLEOPR... -1 377 2.9e-55 | gblW98689lW98689 mg12b06.r1 Soares mouse embryo N... +2 365 6.5e-57 |
| C98 | CTGGGTGAAAGGAGATCTGG | CTCCATCTGCACCCATCC | 195 | none | none | none |
| D102 | TCCACACACATGCAGAACAA | AGGCTGTGTGCAAGGACC | 96 | none | none | none |
| D105 | TTGCCAGCTTCTGGCAAAG | TTCTCTTCTTTAAGCATCCTCC | 125 | none | none | none |
| D19 | GGAAGACTGTTCCTCAAACCC | AGCTCTAAGCTAGTGGTGGTGG | 204 | none | none | none |
| D207 | TCACCTGCCTCTGAGGAGTT | GATGGACATCTGCAGGGTG | 96 | gblM88755lMUSTUMSEQC M.musculus mRNA polyA site sequence. 991 4.1e-75 | none | gblAA033432lAA033432 mi36d08.r1 Soares mouse embryo NbME... 1232 1.1e-97 |
| D209 | CTCTGGGTCGTCACCCAC | AGGTATTGTGTGCATTGCCA | 181 | none | pirllS40482 serine/threonine protein kinase -... -3 175 3.6e-17 | none |
| D212 | CACTCTTGATGTCACGGTGG | TCATGGAAGTTCCTGGAAGG | 100 | gblU49446lDMU49446 Drosophila melanogaster serine/threo... 348 1.6e-19 | gil1335890 (U49446) serine/threonine kinase ... -2 162 9.3e-17 | none |
| D218 | TGGATCTGGGGACAATGG | CAAGGACCCTGAAGACCATC | 187 | none | none | none |
| D23 | GGAAGACTGTTCCTCAAACCC | AGTGGTGGGAGGAGCTAGGT | 178 | none | none | none |
| D32 | TGCTCTGAGAGGGGACAGAT | CCTGGGAAATCCAAGAGTCA | 102 | none | none | none |
| D33 | GCACCAAGAAAGTCCAGCTC | AAGGCCACTCTTCAACATTCA | 101 | none | none | gblAA033093lAA033093 mi24f05.r1 Soares mouse embryo NbME1... 475 3.7e-41 |
| D42 | TGAGCAAGGAGGTGCCTC | TGAAGTAAGGGGGCTCCC | 156 | gblU23443lRNU23443 Rattus norvegicus serine/threonine k... 314 4.2e-19 | gil172586 (L04655) serine/threonine protein... -3 244 3.3e-31 | none |
| D43 | CCAACGAGTGTCCCATGAG | GCATGATCACCACCTCATTG | 224 | none | gnllPIDle249065 (Z74029) C45B11.1 [Caenorhabditis... +1 214 1.4e-22 | gblR88460lR88460 ym92f08.r1 Homo sapiens cDNA clone 16640... 305 2.7e-18 |
| D48 | AGTACTCAGAACCTCCCCTCG | TCCACACACATGCAGAACAA | 154 | none | none | none |
| D60 | ACCTGGTAGGCTGTGAATGG | ATTCCTCATGCTGTGATGACC | 238 | gblL03320lMUSH4RPT Mus spicilegus (clone H4) chromosome... 230 3.3e-09 | none | gblW71752lW71752 md38h01.r1 Life Tech mouse brain Mus... 146 4.6e-14 |

**Table 2**
Primer sequences for cDNA selection/cDNA derived fragments in the pudgy region.

Most significant BLAST score is shown.
All scores at probability lower than 10^-10 are listed

| Assay | Forward primer sequence | Reverse primer sequence | Product size (bp) | BLASTN match | BLASTX match | dBEST match |
|---|---|---|---|---|---|---|
| D8 | CTAATGCTGCGACCATTCCT | AGGCCTACAACCCTGGTTCT | 156 | gb\|M80242\|CRUCIR Cricetulus griseus carcinogen-induce... 243 2.5e-10 | none | gb\|W71752\|W71752 md38h01.r1 Life Tech mouse brain Mus... 150 5.3e-13 |
| E10 | CTGAGCCTAGATGGAGCAGG | TATACGTGGTCCTTCCCAGG | 91 | none | none | none |
| E92 | ATTGCGAACCAGAACTGGTC | AACAGCTATCTCCCGGGG | 171 | none | none | none |
| E96 | GCTTTCGTTTCGTCATGGTT | TTGTCCTCCGATTATCAGGG | 136 | none | none | none |
| F30 | CTACCTGCCAGACCATCTCTG | CATGGTGGCTTTGGAAGC | 151 | gb\|L41685\|RATORFC Rattus norvegicus (clone REM3) ORF mRN... 235 3.4e-10 | none | gb\|H32833\|H32833 EST108316 Rattus sp. cDNA 5' end. 467 8.4e-44 |
| F37 | CAGTGTTGTGAAAACCACCG | ATTTGTAGATCAGGTGGCCG | 127 | emb\|X13661\|MMEF1A Mouse mRNA for elongation factor 1-... 925 2.3e-69 | pir\|\|EFMS1 translation elongation factor eEF... +3 262 2.5e-29 | dbj\|D76644\|MUS75B12 Mouse embryonal carcinoma F9 cell c... 925 5.3e-72 |
| F46 | CAAGACTGAGCATCAGGGGT | GACCCTGGCCTCATGACTTA | 145 | none | none | none |
| F49 | ACTTGATGCGTATTCTGTGGC | ACGTCTCCACACAAGGAAGG | 229 | none | none | none |
| F54 | GATGGACATCTGCAGGGTG | TTCCCTGAGCTCACCTGC | 224 | gb\|M88755\|MUSTUMSEQC M.musculus mRNA polyA site sequence. 852 1.4e-77 | none | gb\|AA033432\|AA033432 mi36d08.r1 Soares mouse embryo NbME... 1093 1.1e-102 |
| G210 | CGTGTGGGCATCACATGT | CGAGAGAGGACTTTAGAATGTG | 80 | none | none | none |
| G214 | TTGTGTGCCCACTATAGCCA | TGAAGACTCTCGGCTGGG | 116 | none | none | gb\|W78373\|W78373 me85g09.r1 Soares mouse embryo NbME13.5... 503 1.9e-35 |
| G230 | GACATGCAAGCCTGTACCCT | CCAGAGCCCACTTTCTTGAC | 113 | none | none | none |
| G64 | CAGTCAGCATTCACCCCC | GTGCTGGTGCACACCTTTAA | 129 | none | none | gb\|AA032398\|AA032398 mi33c07.r1 Soares mouse embryo NbME... 221 1.1e-17 |
| G72 | AACTTGGCTTTGGAGCTGG | AGTAGTAACTCCATTCCACCGC | 162 | none | none | gb\|W64895\|W64895 md99a12.r1 Soares mouse embryo NbME13.5 ... 931 1.5e-72 |
| G73 | TGTGGAGGCCAGAAGAGG | CAAGACTGAAAGCATTGGCA | 177 | none | none | none |
| G76 | TGGATGCCCTAGTCTCAGCT | GCTCATCCTGGAGGACTCTG | 155 | none | none | none |
| G81 | ACTTAACGGCGGACAAAATG | ATTTTGTCCTGTACCGCCTG | 207 | none | sp\|P13264\|GLSK_RAT GLUTAMINASE, KIDNEY ISOFORM PRECU... +1 142 1.9e-12 | none |
| G84 | GGGGTGGTTCAGTGGTTAGA | GGATGACCAGGGACTCCC | 151 | none | none | none |
| H7 | TGGATGTTGTTCTGATCCGA | CACCTCTCATCCTCTGCCTC | | none | none | gb\|AA109753\|AA109753 mp32h09.r1 Barstead MPLRB1 Mus musc... 459 3.0e-31 |
| I30 | GTAACCACGGCCTTGTCAGT | GGAGGTGCGTCAACATACCT | 140 | none | none | none |

**Table 2**
Primer sequences for cDNA selection/cDNA derived fragments in the pudgy region.

| Assay | Forward primer sequence | Reverse primer sequence | Product size (bp) | Most significant BLAST score is shown. All scores at probability lower than 10^-10 are listed BLASTN match | BLASTX match | dBEST match |
|---|---|---|---|---|---|---|
| J76 | CACTGGAAAGAAACCCTTCG | TTGCATTCAAACATTTCTGACC | 172 | gb\|M28515\|MUSMFG3A Mouse zinc finger protein (mfg3) mR... 441 2.9e-33 | sp\|P16374\|MGF3_MOUSE ZINC FINGER PROTEIN MFG3. /pir\|\|C... +3 182 2.2e-27 | gb\|H32448\|H32448 EST107556 Rattus sp. cDNA 5' end si... 545 6.0e-57 |
| J82 | ACACATTAGGTCCTCTTCCTCC | GGCGTCATGCAGGAAGAAG | 150 | none | none | none |
| 13a13 | | | | emb\|X16135\|HSHNRNPL Human mRNA for novel heterogeneous nu... 335 1.6e-18 | sp\|...\|..._HUMAN HETEROGENEOUS NUCLEAR RIBONUCLEOP... +2 153 6.4e-14 | none |
| 13a18 | | | | none | none | none |
| 14a3 | TCTGGAGAGCAGCTGGTGC | ACCAACTAGAAATCTGAAGACTGC | | emb\|X59840\|MMPERIPH M.musculus peripherin gene 286 3.7e-14 | none | none |
| 14a8 | CAGAATTGGCCCTCTGGC | CAGCAGTCATGTAATGCCATTG | | none | none | gb\|R44284\|R44284 yg34a09.s1 Homo sapiens cDNA clo... -1 127 4.6e-23 |
| 6b25 | | | | none | none | none |
| 6b30 | | | | none | none | none |
| 6b50 | | | | none | none | none |
| 6b66 | | | | gb\|U16660\|HSU16660 Human peroxisomal enoyl-CoA hydratase-... 208 1.5e-07 | none | gb\|W09815\|W09815 ma49e07.r1 Soares mouse p3NMF19.5 M... 258 2.4e-15 |
| 6b63 | | | | none | none | none |
| 6b64 | | | | none | none | none |
| 6b62 | | | | none | none | none |

## Chapter Five

# Positional Cloning of a Novel Serine-Threonine Kinase Candidate Gene in the Mouse *pudgy* Region

Kenro Kusumi,* Eileen S. Sun*, Wayne N. Frankel,† and Eric S. Lander*

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

The segmentation of mesoderm to form somites is a temporally and spatially regulated process. The development of somites can be perturbed by defects in a number of genes, including members of the *Pax*, *Wnt, FGF,* bHLH, BMP/TGFβ and *Hox* gene families (Burgess et al., 1995; Cserjesi et al., 1995; Dickinson and McMahon, 1992; Hogan, 1996; Kingsley, 1994; Krumlauf, 1993; Noll, 1993; Yamaguchi and Rossant, 1995).

One uncloned mutation that severely affects somitogenesis is *pudgy*. The *pudgy* (*pu*) mutation was reported in 1961 with the phenotypes of embryonic lethality, defective somitic segmentation, and severely deformed vertebrae and ribs, which descend from somitic sclerotome (Grüneberg, 1961; Kusumi et al., 1997). In the adult, the defects appear limited to the vertebrae and ribs, without affecting cranial or limb skeletogenesis. The single allele of *pu* arose during the course of X-ray mutagenesis efforts at the Oak Ridge National Laboratory, and acts in an autosomal recessive fashion (Russell, 1995).

We have sought to positionally clone the *pudgy* gene by the genetic mapping using over 2,000 informative meioses, physical mapping using YAC, BAC, and P1 DNA clone resources, and transcriptional maps using direct cDNA selection (Figure 1a,b). Meiotic recombinations have limited *pudgy* to a ~600kb interval, completely covered by DNA clones. To understand the genomic organization of the *pudgy* region, we have used shotgun sequencing techniques to determine ~250 kb of genomic sequence, as described previously (Kusumi et al., 1997).

By directed cDNA selection from embryonic day 10.5 cDNA, we have identified a novel serine/threonine kinase transcript with homology to the vertebrate p21-activated kinase (*Pak*), which is involved in cytoskeletal regulation, and yeast STE20 kinase, which is involved in the signal transduction of the pheromone response (Bagrodia et al., 1995; Bagrodia et al., 1995; Brown et al., 1996; Leberer et al., 1992; Manser et al., 1995; Martin et al., 1995). Other transcripts in the *pudgy* region have been described and have been evaluated (Kusumi et al., 1997). We have selected this novel serine-threonine kinase gene as a prime candidate for *pudgy* on the basis of 1.) expression in the embryonic somites, 2.) decreased expression in *pudgy* mutants, as described below, and 3.) role of the closest homologs of this gene in morphological changes during development. Currently, we are working towards a complete comparison of the mutant and wild-type sequences of this transcript and towards phenotypic complementation using BAC and P1 transgenic lines.

A comparison of the kinase domains indicates that this novel gene is not paralogous to the *Pak* family genes; hence we will refer to this novel kinase as *Puk* (*pudgy region kinase*) (Fig. 2a). *Pak*-related serine-threonine kinases are not uninform in transcript size ranging from 8906 bp for the

human trio gene to only 1975 bp for the human "STE20" homologous transcript. Amino acid sequence comparisons of *Puk* with related genes shows close similarity to, but not identity to, *Pak* and *Ste20* throughout the kinase domain (Figure 2b). Northern analysis of *Puk* indicates that there are several isoforms in the adult, as shown in Figure 3b, with the major isoform at 3.6kb and minor isoforms at 6.2 kb and ~11 kb. In E10.5 embryos, we observe ~11 and 3.6 kb isoforms (Figure 3b). We have not yet eliminated the possibility that the larger isoforms represent unprocessed splice forms. Examination of homozygous *pu/pu*, heterozygous *pu/+*, and wild-type E10.5 polyA selected RNA blots hybridized with *Puk* probe indicates a reduction in RNA levels of the heterozygotes down to 44% of wild type levels and mutant homozygotes down to 32% of wild type levels, as compared to *Notch* control probe.

The spatial distribution of *Puk* was examined by whole mount *in situ* hybridization against E9.5 embryos (Figure 4). *Puk* is expressed in ventral somites, with expression also in the cephalic neural tube. We do not observe high levels of *Puk* expression in the presomitic mesoderm, and intensity of signal appears to increase in a rostral direction. This pattern is consistent with expression in the sclerotome, which forms the adult vertebrae and ribs.

Several efforts are underway to identify a mutation in the candidate *Puk* gene. To date, 3.1 kb of the *Puk* transcript has been cloned, and the full transcript is currently being isolated from embryonic cDNA libraries. We have generated primers for reverse transcription-based sequencing of the mutant *Puk* transcript, and have been unable to find a defect in the 3.1 kb examined, based on comparison of the *pu/pu* mutant sequences with wild type strains. The remaining cDNA sequences are being carefully examined, and comparisons of expressed sequence will be aligned to the genomic

sequence that we are currently assembling from ~250kb of the *pudgy* region. Southern blot analysis with *Puk* probes A and B (data not shown) revealed no polymorphisms between the *pu/pu* mutant and C3H/Rl or 101/Rl parental strains. Probes from more 5' regions, once identified, will be used to scan for large-scale mutations, especially since the *pudgy* allele arose during the course of a X-ray mutagenesis screen.

In addition to efforts to find a mutation in the *Puk* gene, we have initiated efforts to complement the mutant phenotype. Currently, we have created transgenic lines with the infection of 4 P1s (PP8, PP9, PP17, and PP19) and 2 BAC clones (315c18 and 19k5) in the *pudgy* region, with one BAC (315c18) and a P1 (PP9) containing currently cloned sections of the *Puk* gene (Figure 1). Clones were injected into C57BL/6J or FVB/J embryos and two or more lines derived from each construct. Transgene positive carriers were crossed to *pu/pu* mice and carrier F1 progeny are currently being backcrossed to *pu/pu*. Progeny were assayed for presence of the transgene and *pudgy* phenotype. However, complementation efforts are contingent on the *Puk* genomic region being contained within currently tested clones.

What could be the function of a *Pak* related serine-threonine kinase gene within the process of somitogenesis? Somitogenesis involves a number of cellular cytoskeleton transitions. During somite formation, mesenchymal cells in the presomitic mesoderm adhere and become epithelial in organization, as shown in Figure 5. Sclerotome is formed from the reconversion of epithelial cells into mesenchymal mesoderm later in somitogenesis. The *Pak* gene expression has been studied in *Drosophila*, where DPAK is a target of the rho family of genes, *Cdc42* and *Rac* (Harden et al., 1996). DPAK protein was found in embryos at the leading edge of epithelial cells undergoing cytoskeletal transformation, where a transient

loss of cytoskeletal structures was observed. Similarly in yeast, *Cdc42* activates *Ste20* which leads to cytoskeletal rearrangements in response to pheromone activation. Although not paralogous to the *Pak / Ste20* genes, an isoform of the *Puk* gene could be involved in similar somitic cell transformations, although binding experiments would be required to confirm if there are interactions with *Rac* and/or *Cdc42*. Downstream targets of STE20 and *Pak* include members of the kinase cascade, *Mekk/Ste11* . A defect in the *regulator of G-protein signaling* (*RGS*) member of this signaling cascade has been associated with the *fused* mutation (F. Constantini, unpublished; Perry et al., 1995). The *fused* mutant displays an overgrowth neural plate folding and a duplication of axes, leading to adult vertebral and rib fusions.

In summary, we have identified a novel serine-threonine kinase, which we term *Puk* for *pudgy* region kinase, with similarity but not paralogy to the *Pak* family of cytoskeletal regulatory genes. This serine-threonine kinase gene is a prime candidate for *pudgy* based on the following findings: 1.) *Puk* is expressed in ventral somites during somitogenesis, and three major isoforms of 3.6kb, 6.2kb, and ~11kb can be found in certain adult tissues and embryos. 2.) The expression level of the ~11kb isoform appears to be decreased in *pu / pu* embryos as compared to wild type levels. The larger isoform is not found in most adult tissues, and we speculate that the larger isoform is required during embryogenesis. 3.) Based on homology to cytoskeletal regulatory kinases, we hypothesize that this novel serine-threonine kinase may play a role in the transition of mesoderm during somitogenesis from mesenchyme to somitic sphere epithelium and back to mesenchyme during sclerotome formation. Currently, we are working towards a complete genomic and cDNA sequence comparison of the mutant

and wild-type and towards complementing the mutant phenotype using BAC and P1 transgenic lines.


## ACKNOWLEDGMENTS

We would also like to acknowledge Anne Kerrebrock, Keri Devon, Michael Nguyen, Marta Velez-Stringer for valuable technical assistance. We thank Christine Kozak for the gift of polychromosomal hybrid panel DNAs and Brigid Hogan for her E8.5 embryonic cDNA library. For their computer consultation, we thank Fran Lewitter, Carol Laboissoniere, Carl Rosenberg and Mark Daly. We would like to thank the members of the Lander Lab, in particular Arend Sidow, Julia Segre, Armand MacMurray, Bruce Hamilton, Jennifer Nemhauser, Johanna Hästbacka, Howard Jacob, and Bill Dietrich for helpful discussions. We thank Hazel Sive, Peggy Kolm and John Kuo for discussion on in situ hybridization screening.

## REFERENCES

Bagrodia, S., Derijard, B., Davis, R. J., and Cerione, R. A. (1995). Cdc42 and PAK-mediated signaling leads to Jun kinase and p38 mitogen-activated protein kinase activation. J Biol Chem *270*, 27995-8.

Bagrodia, S., Taylor, S. J., Creasy, C. L., Chernoff, J., and Cerione, R. A. (1995). Identification of a mouse p21Cdc42/Rac activated kinase. J Biol Chem *270*, 22731-7.

Brown, J. L., Stowers, L., Baer, M., Trejo, J., Coughlin, S., and Chant, J. (1996). Human Ste20 homologue hPak1 links GTPases to the JNK MAP kinase pathway. Curr. Biol. *6*, 598-605.

Burgess, R., Cserjesi, P., Ligon, K. L., and Olson, E. N. (1995). Paraxis: a basic helix-loop-helix protein expressed in paraxial mesoderm and developing somites. Dev Biol *168*, 296-306.

Cserjesi, P., Brown, D., Ligon, K. L., Lyons, G. E., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., and Olson, E. N. (1995). Scleraxis: a basic helix-loop-helix protein that prefigures skeletal formation during mouse embryogenesis. Development *121*, 1099-110.

Dickinson, M. E., and McMahon, A. P. (1992). The role of Wnt genes in vertebrate development. Curr Opin Genet Dev *2*, 562-6.

Fahrner, K., Hogan, B. L., and Flavell, R. A. (1987). Transcription of H-2 and Qa genes in embryonic and adult mice. EMBO J *6*, 1265-71.

Grüneberg, H. (1961). Genetical studies on the skeleton of the mouse: XXIX. Pudgy. Genet. Res., Camb. *2*, 384-393.

Haldi, M. L., Strickland, C., Lim, P., Vanberkel, V., Chen, X. N., Noya, D., Korenberg, J. R., Husain, Z., Miller, J., and Lander, E. S. (1996). A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome. Mamm Genome *7*, 767-769.

Harden, N., Lee, J., Loh, H. Y., Ong, Y. M., Tan, I., Leung, T., Manser, E., and Lim, L. (1996). A Drosophila homolog of the Rac- and Cdc42-activated serine/threonine kinase PAK is a potential focal adhesion and focal complex protein that colocalizes with dynamic actin structures. Mol Cell Biol *16*, 1896-908.

Hogan, B. L. (1996). Bone morphogenetic proteins: multifunctional regulators of vertebrate development. Genes Dev *10*, 1580-94.

Kingsley, D. M. (1994). What do BMPs do in mammals? Clues from the mouse short-ear mutation. Trends Genet *10*, 16-21.

Krumlauf, R. (1993). Mouse Hox genetic functions. Curr Opin Genet Dev *3*, 621-5.

Kusumi, K., Frankel, W. N., Sun, E. S., Birren, B., Chi, D. C., Hawkins, T. L., Reeve-Daly, M. P., and Lander, E. S. (1997). A candidate gene map of the mouse *pudgy* region: Integration of genetic, physical, transcriptional, expression and genomic sequence maps. Doctoral Thesis: Massachusetts Institute of Technology:

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.

Kusumi, K., Sun, E. S., Frankel, W. N., and Lander, E. S. (1997). The mouse *pudgy* mutation causes defects in somitogenesis and early embryonic lethality. Doctoral Thesis: Massachusetts Institute of Technology.

Larin, Z., Monaco, A. P., and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc Natl Acad Sci U S A *88*, 4123-7.

Leberer, E., Dignard, D., Harcus, D., Thomas, D. Y., and Whiteway, M. (1992). The protein kinase homologue Ste20p is required to link the yeast pheromone response G-protein beta gamma subunits to downstream signalling components. Embo J *11*, 4815-24.

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. (New York: Current Protocols).

Manser, E., Chong, C., Zhao, Z. S., Leung, T., Michael, G., Hall, C., and Lim, L. (1995). Molecular cloning of a new member of the p21-Cdc42/Rac-activated kinase (PAK) family. J Biol Chem *270*, 25070-8.

Martin, G. A., Bollag, G., McCormick, F., and Abo, A. (1995). A novel serine kinase activated by rac1/CDC42Hs-dependent autophosphorylation is related to PAK65 and STE20. Embo J *14*, 4385.

Noll, M. (1993). Evolution and role of Pax genes. Curr Opin Genet Dev *3*, 595-605.

Perry, W. L. r., Vasicek, T. J., Lee, J. J., Rossi, J. M., Zeng, L., Zhang, T., Tilghman, S. M., and Costantini, F. (1995). Phenotypic and molecular analysis of a transgenic insertional allele of the mouse Fused locus. Genetics *141*, 321-32.

Russell, L. (1995). The pudgy mutation and its origins.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2 Edition, A. Irwin, ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Wilkinson, D. G. (1992). *In situ* hybridization: A practical approach (Oxford: Oxford University Press).

Yamaguchi, T. P., and Rossant, J. (1995). Fibroblast growth factors in mammalian development. Curr Opin Genet Dev *5*, 485-91.

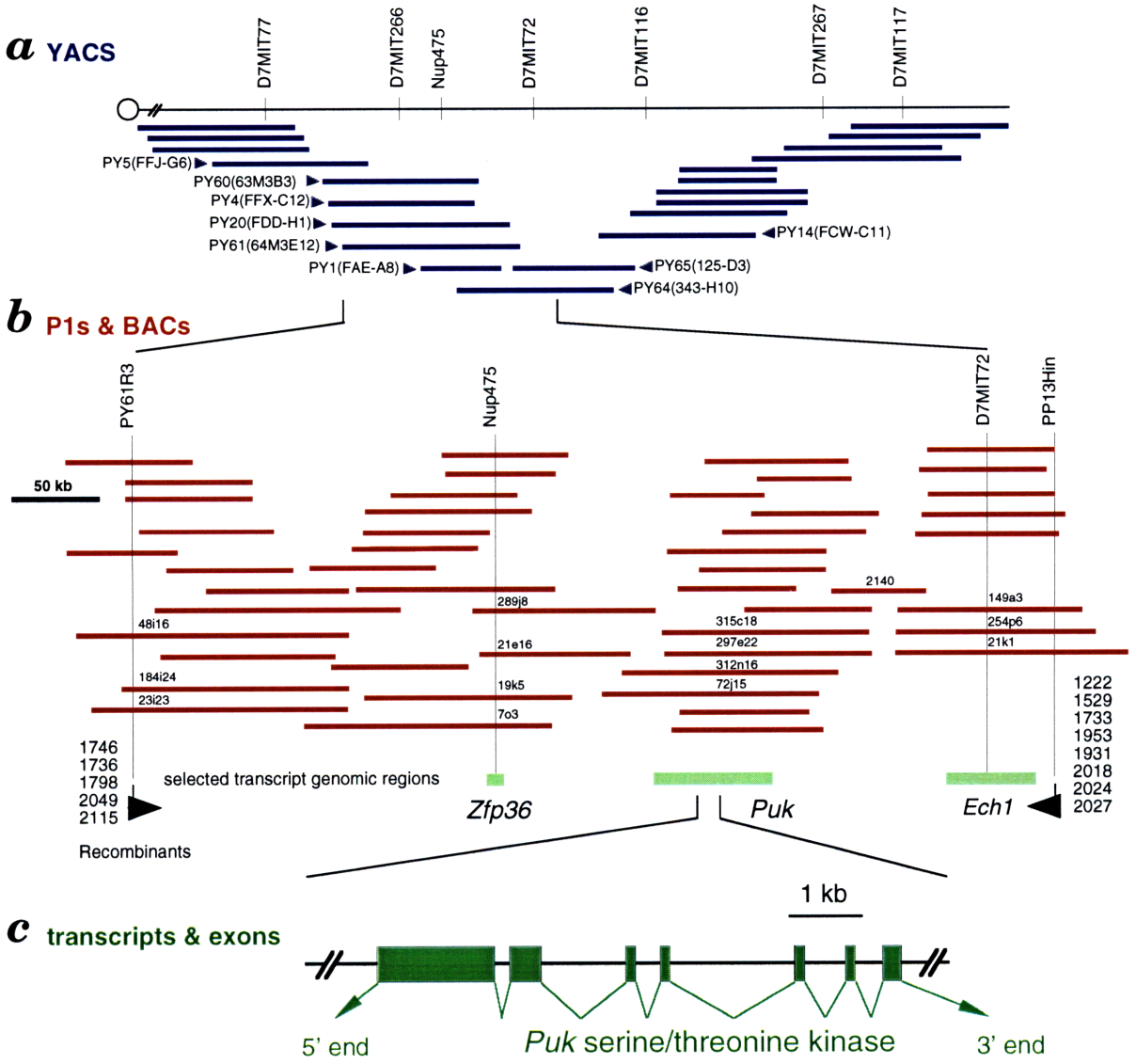



**Figure 1** Physical map of the *pudgy* region and location of the *Puk* gene. *a* YACs from the WI/MIT Mouse YAC Library I (Kusumi et al., 1993) and from Library II (Haldi et al., 1996). The genetic interval between D7MIT77 and D7MIT117 is 2.3 cM on the WI/MIT Mouse Genetic Map. *b* BACs and P1s are shown in red, with relative sizes indicated. Key recombinants delineating the minimal *pudgy* interval are shown on the left and right. DNA clones 315c18, 19k5, PP19, PP17, PP8, and PP9 have been used for transgenic complementation efforts. Two of these clones (315c18 and 19k5) have been sequenced by shotgun genomic sequencing methods. *c* Organization of 3.1 kb of currently available *Puk* gene is shown with intron-exon structure compared against genomic sequence from shotgun sequencing of BACs 19k5 and 315c18. Flanking regions containing possible *Puk* transcript fragments are shown in stippled green.

***a***

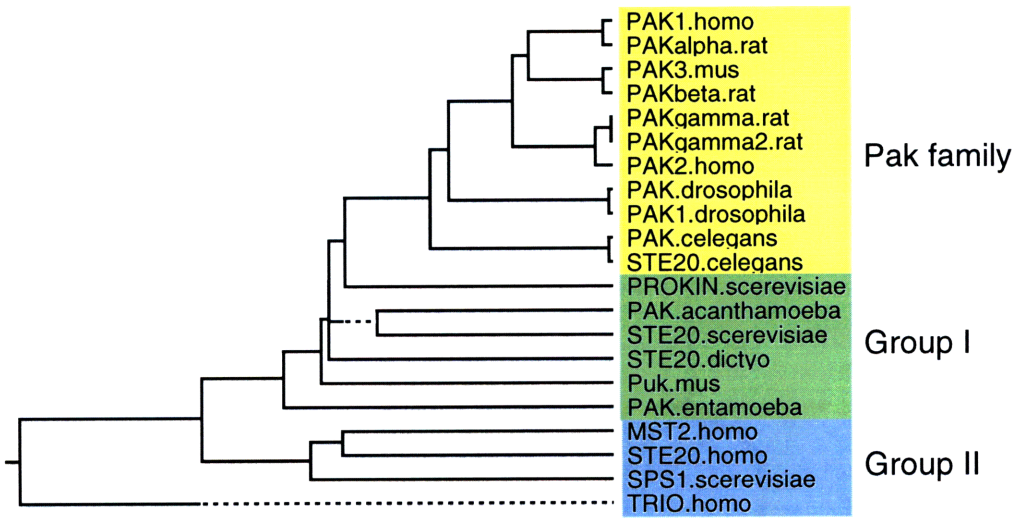



**Figure 2a** Cladogram based on amino acid sequence comparison of *Pak* and *Ste20* related gene entries in GENBANK. Although genes cloned by homology searches are often include *Pak* or *Ste20*in their name, this is not an indicator of their amino acid sequence similarity. Above, the genes have been divided into three general categories: the *Pak* family of genes which includes related genes from *C. elegans*, *Drosophila*, and vertebrates; Group I which includes distantly related genes to *Ste20* from single-celled eukaryotes. This group includes the mouse *Puk* gene; Group II which includes human genes similar to *Ste20* and yeast *Sps1*.

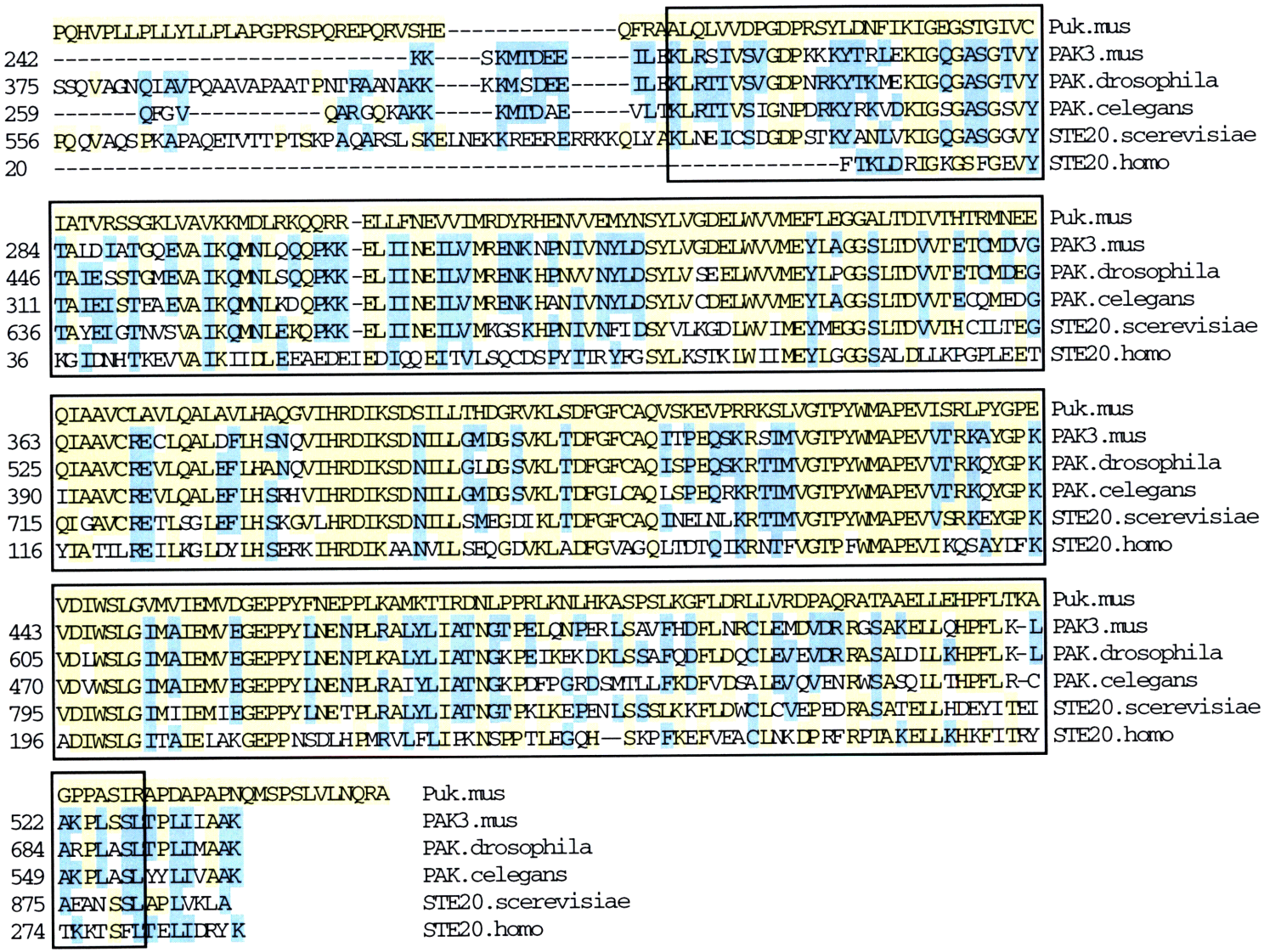



**Figure 2b** Predicted amino acid sequence alignment of the kinase domain region of *Puk* gene with related *Pak*and *Ste20* kinases from mouse (*Mus musculus domesticus*), human (*Homo sapiens*), yeast (*S. cerevisiae*), *Drosophila*, and *C. elegans*. The kinase domain is shown boxed. Amino acids identical to *Puk* are highlighted in yellow. Residues conserved among the other kinases but not with *Puk* are highlighted in blue.

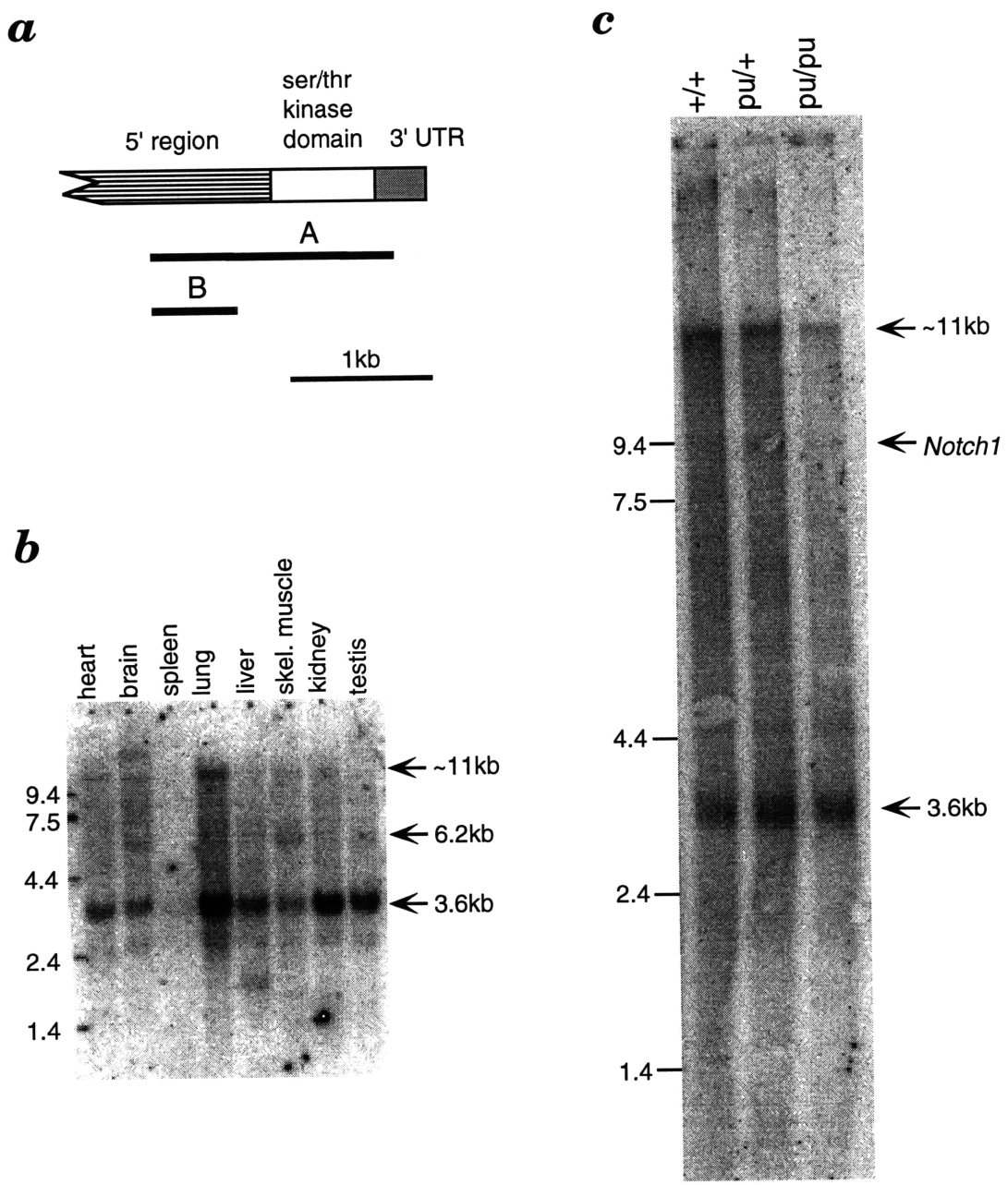



**Figure 3** RNA blot analysis of *Puk* expression.
*a Puk* gene 3' end organization, with regions selected for probes. *b* Adult multi-tissue blot of polyA selected RNAs, hybridized with *Puk* probe A, with a major isoform at 3.6 kb and minor forms at 6.2 kb and ~11 kb. *c* E10.5 embryonic polyA selected RNA blot comparing *pu/pu*, *pu/+*, and +/+. Predominant isoforms are 3.6 kb and ~11kb, with reduction in levels of the ~11kb isoform in mutant embryo sample as compared to *Notch* control.

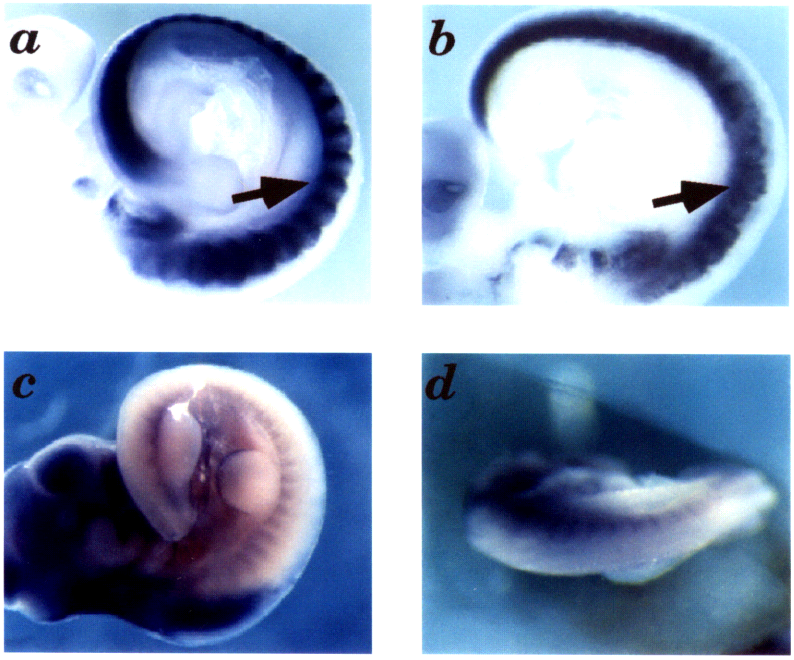



**Figure 4** Whole mount *in situ* hybridization analysis of *Puk* expression.
*a,b* Expression pattern of control *Mox1*which recognizes somites. *a* is a wild type E9.5 embryo while *b* is a *pu/pu* mutant embryo. Note the irregular somitic sphere outline. *c* Expression pattern using *Puk* probe B (shown in Fig. 3). *d* View of *Puk expression* from a dorsal perspective. *In situ* analysis was carried out essentially as described previously (Wilkinson, 1992).

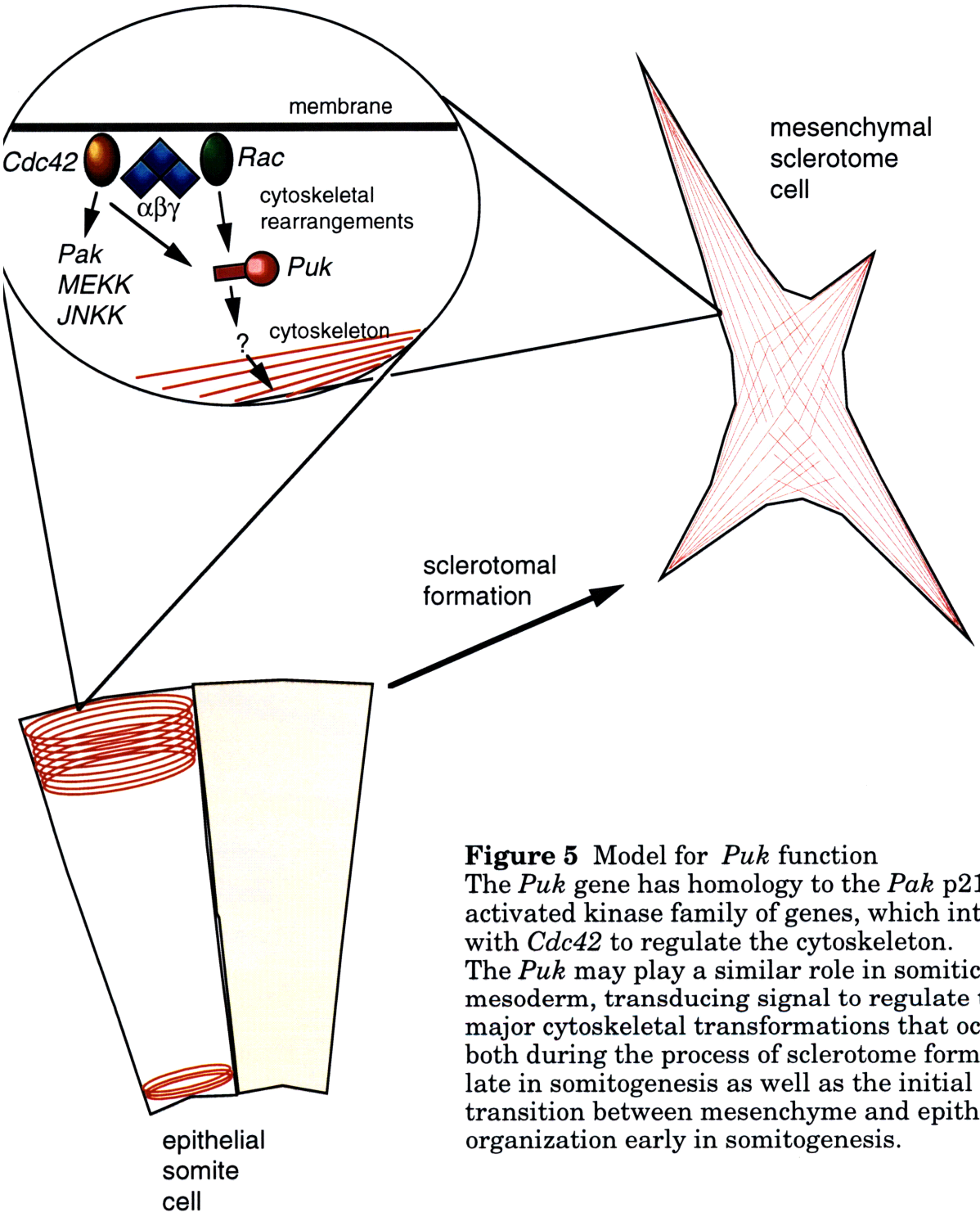



**Figure 5** Model for *Puk* function
The *Puk* gene has homology to the *Pak* p21 activated kinase family of genes, which interact with *Cdc42* to regulate the cytoskeleton. The *Puk* may play a similar role in somitic mesoderm, transducing signal to regulate the major cytoskeletal transformations that occur both during the process of sclerotome formation late in somitogenesis as well as the initial transition between mesenchyme and epithelial organization early in somitogenesis.

## Appendix I

### Overview: Genetic Mapping of Mouse Developmental Mutations

Over 2,000 mutations have been described in the laboratory mouse (Doolittle et al., 1996). As part of my doctoral research, I have initiated several genetic mapping projects using simple sequence length polymorphism markers to examine the following mutations:

- *motor neuron degeneration*, a early-onset retinal degeneration and late-onset motor paralysis syndrome in mouse which is an important animal model for human Batten's disease (Appendix I.A)
- *ducky*, a mutation affecting motor coordination which has not been examined in detail for neuropathology (Appendix I.B)
- *curly whiskers*, a mutation which affects mouse vibrissae, but not coat hair (Appendix I.C) and
- *staggerer*, a classic neurological mutation affecting cerebellar Purkinje cells.

The *staggerer* mutation has been cloned and found to be a deletion in the orphan nuclear receptor RORα (Hamilton et al., 1996).

We have initiated these mapping efforts toward the eventual positional cloning of these loci; indeed the *staggerer* mutation was identified based on these genetic mapping efforts.

## REFERENCES

Doolittle, D. P., Davisson, M. T., Guidi, J. N., and Green, M. C. (1996). Catalog of mutant genes and polymorphic loci. In Genetic Variants and Strains of the Laboratory Mouse, M. F. Lyon, S. Rastan and S. D. M. Brown, eds. (Oxford: Oxford University Press), pp. 1807.

Hamilton, B. A., Frankel, W. N., Kerrebrock, A. W., Hawkins, T. L., FitzHugh, W., Kusumi, K., Russell, L. B., Mueller, K. L., van Berkel, V., Birren, B. W., Kruglyak, L., and Lander, E. S. (1996). Disruption of the nuclear receptor hormone receptor RORα in *staggerer* mice. Nature *379*, 736-739.

**Appendix I.A**

# Genetic Mapping of the Mouse *motor neuron degeneration* Mutation

KENRO KUSUMI,*† ANNE MESSER,‡ EILEEN S. SUN,* AND ERIC S. LANDER*†

** Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, 02142; †Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139; and ‡Sony Department of Health, Alexrod Institute, Albany, NY 12208*

## INTRODUCTION

The *motor neuron degneration* (*mnd*) mutation was first reported in 1986 as an autosomally dominant, progressive, late-onset neurological disorder (Messer and Flaherty, 1986). The mutation arose in the C57BL/6.KB2/Rn strain, and was subsequently inbred to form its own stock. During this inbreeding, the manifestation of the clinical phenotype changed from dominant to recessive (Messer, 1996). The reason for this change is not known, but it could be due to several factors, including the elimination of background genetic modifiers through inbreeding and the elimination of infectious agents which accelerate the neurological degeneration.

The *mnd* mutation causes an array of phenotypes:

- A progressive motor dysfunction, presenting initially through hindlimb weakness and eventually progressing to forelimb weakness and eventual death through spastic paralysis. The clinical signs first manifest at approximately 6 months, a late onset time for a mouse phenotype
- An early retinal photoreceptor degeneration in *mnd* homozygous animals, beginning on average postnatal day 15 (Messer et al., 1993).

• The accumulation of an autofluorescent ceroid lipofuscin in numerous neuronal and non-neuronal cell types (Bronson et al., 1993; Messer et al., 1987). Lipofuscin is the component of normal inclusion bodies which accompany cell aging, but the abnormal early accumulation of ceroid lipofuscin is associated with human syndromes such as Batten's Disease (McKusick, 1994).

Previously, the *motor neuron degeneration* mutation has been genetically mapped using the intercross (C57BL/6.KB2-*mnd*/Msr x AKR/J)F2 to a location 7.8 cM proximal to the Polb and Ank-1 genes, and 2.9 cM distal to a retroviral insertion element Xmv-26 (Messer et al., 1992). This does not place the mutation near any known currently identified mouse motor defect mutations.

## GENETIC MAPPING RESULTS

The motor neuron degeneration mutation was genetically mapped through the use of an intercross of 210 F2 progeny with an inbred *M. musculus castaneus* strain (C57BL/6.KB2-*mnd*/Msr x CAST/Ei)F2 and an intercross of 343 F2 progeny with an inbred AKR strain (C57BL/6.KB2-*mnd*/Msr x AKR/J)F2. Thus, we have examined a total of 1,106 meioses through these crosses. The *motor neuron degeneration* phenotype was scored by manifestation of the *mnd* phenotype in the AKR cross. However, with the CAST cross, significant attenuation of the *mnd* motor dysfunction phenotype was noted and made the scoring based on clinical observations difficult. Determination of the *mnd* phenotype in the CAST cross was carried out through examination of autofluorescent lipofuscin accumulation in cross progeny.

We have found in the (C57BL/6.KB2-*mnd*/Msr x AKR/J)F2 that the *mnd* mutation maps 0.2 cM distal to D8MIT124 and 0.8 cM proximal to the cluster

of markers D8MIT3, 16, 62, 143, defining a 1.0 cM minimal genetic region. In the intersubspecific cross (C57BL/6.KB2-*mnd*/Msr x CAST/Ei)F2, the mutation maps 0.8 cM distal to D8MIT124 and 0.8 cM proximal to M14L2, a SSCP-based marker generated from a YAC end, defining a 1.6 cM minimal genetic interval. The corresponding genetic distance in the CAST cross between the equivalent closest AKR cross polymorphic markers is 3.4 cM.

## DISCUSSION

We have refined the genetic localization of the *motor neuron degeneration* mutation to a 1.0 cM region permitting efforts towards positional cloning. The region of mouse Chromosome 8 where *mnd* maps is remarkably poor in published gene candidates. Unfortunately, the proximal region of mouse Chromosome 8 has very discontinuous human synteny conservation with regions of correspondence with human chromosomes 8, 19, and 13. This makes it quite difficult to attempt to find human correlates or other candidate genes.

Although *mnd* was first described as a possible model from human amyotrophic lateral sclerosis, recent findings suggest that *mnd* in fact shares great similarity to the human neuronal ceroid lipofuscinoses (NCL) (Bronson et al., 1993; Messer and Plummer, 1993). Similarities exist in the cellular pathology of the two syndromes and retinal degeneration.

Several forms of NCL exist, including an infantile (Santavuori disease or CLN1), late-infantile (CLN2), juvenile (Batten disease or CLN3), and adult-type (Kufs disease or CLN4) and variant late-infantile (CLN5) (McKusick, 1994). Early onset NCL diseases are characterized clinically by profound seizures, mental retardation due to severe cerebral atrophy and motor paralysis. No such cerebral cortical neuron atrophy has been noticed in the

*mnd* mice. CLN4 is characterized by a later onset (20-30's) with seizures, mental degeneration, and cerebellar ataxias. CLN1 has been cloned and shown to have defects in the palmitoyl-protein thioesterase gene (Vesa et al., 1995). CLN3 has been cloned and identified as a novel gene.(1995) CLN2, 4 and 5 remained uncloned to date, with CLN5 mapping to 13q21.1-q32 (Savukoski et al., 1994).

Although a region of human chromosome 13q34 synteny conservation exists near the mapping location of *mnd*, the patchwork pattern of synteny conservation on the proximal region of mouse chromosome 8 does not permit us to name any key candidates or human homologous diseases at this time. However, given the fine localization of the *mnd* locus with over 1,000 meioses to a 1 cM interval, the genetic mapping basis for a positional cloning effort has been established.

## REFERENCES

(1995). Isolation of a novel gene underlying Batten disease, CLN3. The International Batten Disease Consortium. Cell *82*, 949-57.

Bronson, R. T., Lake, B. D., Cook, S., Taylor, S., and Davisson, M. T. (1993). Motor neuron degeneration of mice is a model of neuronal ceroid lipofuscinosis (Batten's disease). Ann Neurol *33*, 381-5.

McKusick, V. A. (1994). Mendelian Inheritance in Man. Catalogs of Human Genes and Genetic Disorders, 11th Edition (Baltimore: Johns Hopkins University Press).

Messer, A. (1996). The mouse motor neuron degeneration mutation.

Messer, A., and Flaherty, L. (1986). Autosomal dominance in a late-onset motor neuron disease in the mouse. J Neurogenet *3*, 345-55.

Messer, A., Plummer, J., Maskin, P., Coffin, J. M., and Frankel, W. N. (1992). Mapping of the motor neuron degeneration (Mnd) gene, a mouse model of amyotrophic lateral sclerosis (ALS). Genomics *13*, 797-802.

Messer, A., Plummer, J., Wong, V., and Lavail, M. M. (1993). Retinal degeneration in motor neuron degeneration (mnd) mutant mice [letter]. Exp Eye Res *57*, 637-41.

Messer, A., Strominger, N. L., and Mazurkiewicz, J. E. (1987). Histopathology of the late-onset motor neuron degeneration (Mnd) mutant in the mouse. J Neurogenet *4*, 201-13.

Savukoski, M., Kestila, M., Williams, R., Jarvela, I., Sharp, J., Harris, J., Santavuori, P., Gardiner, M., and Peltonen, L. (1994). Defined chromosomal assignment of CLN5 demonstrates that at least four genetic loci are involved in the pathogenesis of human ceroid lipofuscinoses. Am J Hum Genet *55*, 695-701.

Vesa, J., Hellsten, E., Verkruyse, L. A., Camp, L. A., Rapola, J., Santavuori, P., Hofmann, S. L., and Peltonen, L. (1995). Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. Nature *376*, 584-7.

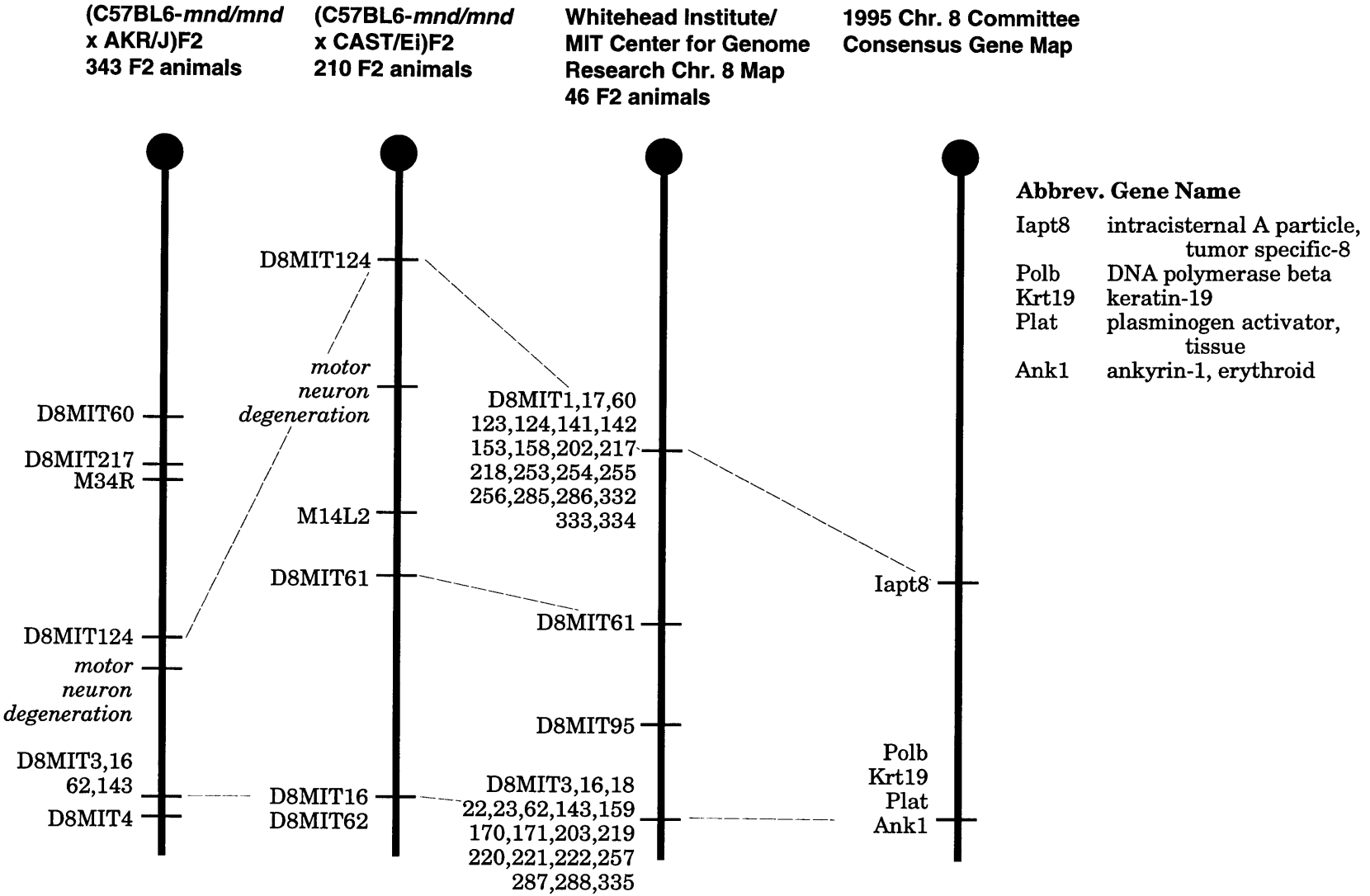



**Figure 1** Genetic map of the *motor neuron degeneration* mutation
The genetic mapping location of *mnd* is shown relative to MIT markers D8MIT124 and 61. The WI/MIT mapping cross and Chromosome 8 Committee consensus gene map are shown for reference. Dotted lines indicate markers which have been mapped on both crosses.

## Appendix I.B

# Genetic Mapping of the *ducky* Mutation

KENRO KUSUMI,*† EILEEN S. SUN*, WAYNE N. FRANKEL,‡
AND ERIC S. LANDER*†

** 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*

## INTRODUCTION

The *du* (ducky) mutation was first isolated in 1946 as having a "peculiar waddling gait" and arose spontaneously on a noninbred ruby-silver stock (Snell, 1955). The mutation is characterized by a number of motor defects first noticeable from 20-40 days postnatally: splaying of the hind feet, waddling gait, and eventual ataxia in older animals. No histology or anatomical pathology has been reported, so the origin of this defective motor behavior has not been determined. Initial mapping efforts linked *du* to short-ear and dilute on mouse Chromosome 9. The ducky stock is referred to as TKDU/J.

## GENETIC MAPPING RESULTS

The ducky mutation was genetically mapped through the use of an intraspecific intercross with 446 F2 progeny with an inbred stock containing the *pudgy* mutation (PU-*pu*/*pu* x TKDU-*du*/*du*)F2. We have scored 84 *du*/*du* affected animals out of 446 F2 progeny, which results in 19% ducky progeny vs. 25% expected, a statistically significant loss of expected homozygotes ($p<0.025$, $X^2=9.18$). All animals with *du*/*du* genotype exhibited the *ducky*

phenotype, thus the loss of expected ducky animals was due to loss of the homozygous *du/du* genotypic class, not due to reduced penetrance of the ducky mutation.

We have mapped the ducky mutation 1.7 cM distal to the SSLP marker D9MIT14. We were unable to find a close marker on the distal side that was polymorphic between TKDU and PU at the time of screening, but we were able to map ducky proximal to marker D9MIT55 by 6.9 cM. We have not characterized markers D9MIT78, 80, 168, 242 and 348 which were added to the WI/MIT map subsequent to our mapping efforts. This gives a minimal genetic interval of 8.6 cM in the TKDU cross. This genetic distance is much larger than the 2.1 cM equivalent region in the WI/MIT (C57BL/6J-*ob/ob* x CAST/Ei)F2 mapping cross.

## DISCUSSION

Although the anatomical pathology of the ducky mutant remains to be established, the clinical phenotype of the defect bears a strong resemblance to the classic cerebellar mutations that have been isolated in the mouse—*weaver, reeler, staggerer, Lurcher,* and *Purkinje cell degeneration*. Of these mutations, several have been cloned and the analysis of the affected genes is expanding our understanding of cerebellar development. The *weaver* mutation has recently cloned and found to encode a G-protein coupled potassium channel (Patil et al., 1995). The reeler gene has been cloned and encodes a novel extracellular matrix protein (D'Arcangelo et al., 1995; Hirotsune et al., 1995). Recently, the staggerer mutation has been found to be a deletion in the nuclear receptor RORα (Hamilton et al., 1996). Members of the nuclear receptor family can serve as transcriptional regulators, and the staggerer gene may serve as a key switch in the developmental transcriptional regulation in the

cerebellum. Further genetic mapping of the *ducky* mutation combined with pathological examination of the *ducky* phenotype could potentially add this mutant to the classic list of neurodevelopmental mutations.

## REFERENCES

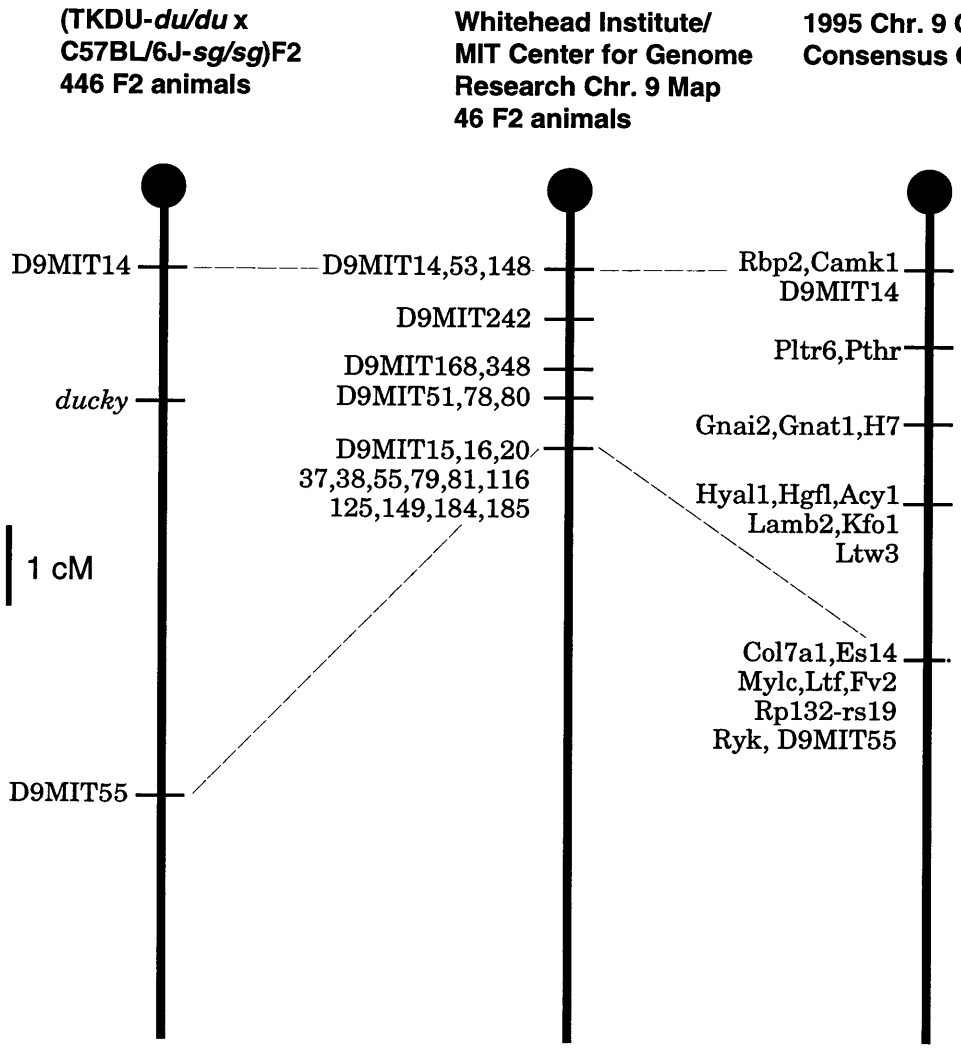
D'Arcangelo, G., Miao, G. G., Chen, S. C., Soares, H. D., Morgan, J. I., and Curran, T. (1995). A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. Nature *374*, 719-23.

Hamilton, B. A., Frankel, W. N., Kerrebrock, A. W., Hawkins, T. L., FitzHugh, W., Kusumi, K., Russell, L. B., Mueller, K. L., van Berkel, V., Birren, B. W., Kruglyak, L., and Lander, E. S. (1996). Disruption of the nuclear receptor hormone receptor RORα in *staggerer* mice. Nature *379*, 736-739.

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.

Patil, 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*, 126-9.

Snell, G. D. (1955). Ducky, a new second chromosome mutation in the mouse. The Journal of Heredity *46*, 27-29.




| Abbrev. | Gene Name |
|---|---|
| Rbp2 | retinol binding protein-2, cellular |
| Camkl | calcium/calmodulin-dependent protein kinase-like |
| Pltr6 | polytropic long terminal repeat-containing proviral locus-6 |
| Pthr | parathyroid hormone receptor |
| Gnai2 | guanine nucleotide binding protein, alpha inhibiting-2 |
| Gnat1 | guanine nucleotide binding protein, alpha transducing-1 |
| H7 | histocompatibility-7 |
| Hyal1 | hyaluronidase-1 |
| Hgfl | hepatocyte growth factor-like |
| Acy1 | aminoacylase-1 |
| Lamb2 | laminin, beta 2 |
| Kfo1 | kidney 40-50 thousand M.Wt. protein-1 |
| Ltw3 | liver 20-30 thousand M.Wt. protein-3 |
| Col7a1 | procollagen, type VII, alpha 1 |
| Es14 | esterase-14 |
| Mylc | myosin light chain, alkali, cardiac ventricles |
| Ltf | lactotransferrin |
| Fv2 | Friend virus susceptibility-2 |
| Rpl32-rs19 | ribosomal protein L32 related sequence 19 |
| Ryk | receptor-like tyrosine kinase |



**Figure 1** Genetic map of the *ducky* mutation
The genetic mapping location of *du* is shown relative to MIT markers D9MIT14 and 55. The WI/MIT mapping cross and Chromosome 9 Committee consensus gene map are shown for reference. Dotted lines indicate markers which have been mapped on both crosses.

**Appendix I.C**

## Genetic Mapping of the *curly whiskers* Mutation

KENRO KUSUMI,*† EILEEN S. SUN*, WAYNE N. FRANKEL,‡
AND ERIC S. LANDER*†

** 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*

### INTRODUCTION

The *cw* (curly whiskers) mutation was identified in 1958 by C.J.W. Smith of the Chester Beatty Research Institute in London and arose spontaneously on a CBA/Cbi inbred background (Falconer and Isaacson, 1966). The mutation is characterized by a number of hair-related phenotypes, including wavy vibrissae noticeable at birth and darkened coat hairs. The curly whisker phenotype is semidominant, and presence of wavy vibrissae is fully penetrant in the homozygote, with lower penetrance in heterozygotes. Previous studies had genetically mapped the mutation to the proximal arm of Chromosome 9, 38 cM proximal to the short ear mutation. These results placed *cw* as one of the most proximal markers on mouse Chromosome 9. The *cw* stock is referred to as CWD/J.

### GENETIC MAPPING RESULTS

The curly whiskers mutation was genetically mapped through the use of an intercross of 176 F2 progeny with an inbred stock containing the staggerer mutation (CWD/J-*cw*/*cw* x C57BL/6J-*sg*/*sg*). Since the unaffected animals of

this cross that were recombinant on proximal Chr. 9 were not progeny tested, this intercross represents a total of 176 meioses. Of the 176 F2 progeny, we have scored 46 curly whisker phenotype positive animals, representing 26% affected in the intercross vs. 25% expected. As shown in Figure 1, the mutation maps 8.3 cM proximal to simple sequence length polymorphism (SSLP) markers D9MIT42, D9MIT63, D9MIT64, and D9MIT65, This compares with the distance of 5.8 cM in the Whitehead Institute/MIT mapping cross between D9MIT42 and the most proximal SSLP marker on chromosome 9, D9MIT59. Although we tested all MIT SSLP genetic markers proximal to D9MIT42, we were unable to find any which were polymorphic between CWD/J and C57BL/6J. Thus, the *curly whiskers* mutation is located close to the centromere on chromosome 9.

## DISCUSSION

Since the isolation of curly whiskers in 1958, the understanding of hair formation and maintenance has been advanced through the cloning of hair-related defect genes, or through the characterization of induced mutants through transgenic insertions or knockouts. For example, an induced mutation in the TGFα gene resulted in abnormally wavy vibrissae and fur, due to the defective orientation of the hair follicles (Luetteke et al., 1993; Mann et al., 1993). TGFα disruption also severely affected normal eye development in these animals. Defects in the TGFα gene where in fact correlated with the classic mouse mutation *waved-1* which has similar hair and eye phenotypes (Luetteke et al. 1993; Mann et al. 1993)

The classic mutant *waved-2* was similarly found to be allelic with the TGFα /EGF receptor (EGFR), thus further establishing the importance of this ligand/receptor interaction for vibrissae development. Recently, the well-

studied *nude* mutation, which affects the normal development of coat hair but does not affect vibrissae, has been identified in a novel gene of the fork-head family (Nehls et al., 1994; Segre et al., 1995). The importance of these genes makes clear that it is difficult to discount genes that affect hair or vibrissae phenotypes as not biologically relevant, but conversely the role of growth factors and transcription factors makes it difficult to predict the type of genes that would be involved.

Of the genes mapping in the *cw* region, several have already been identified as the genes involved in numerous clinical disorders including *Epor* or erythropoeitin receptor in hematopoetic defects (Longmore and Lodish, 1991; Longmore et al., 1994; Longmore et al., 1993). *Ldlr* or low density lipoprotein receptor defects are associated with familial hypercholesterolemia (Horsthemke et al., 1987), and the *Il1bc* or interleukin-1 beta convertase gene has a homolog in *C. elegans*, ced-3, which is a cell-death regulatory gene (Yuan et al., 1993). Further mapping efforts will help to define the candidates for this mutation.

## REFERENCES

Horsthemke, B., Dunning, A., and Humphries, S. (1987). Identification of deletions in the human low density lipoprotein receptor gene. J Med Genet *24*, 144-7.

Longmore, G. D., and Lodish, H. F. (1991). An activating mutation in the murine erythropoietin receptor induces erythroleukemia in mice: a cytokine receptor superfamily oncogene. Cell *67*, 1089-102.

Longmore, G. D., Pharr, P. N., and Lodish, H. F. (1994). A constitutively activated erythropoietin receptor stimulates proliferation and contributes to transformation of multipotent, committed nonerythroid and erythroid progenitor cells. Mol Cell Biol *14*, 2266-77.

Longmore, G. D., Watowich, S. S., Hilton, D. J., and Lodish, H. F. (1993). The erythropoietin receptor: its role in hematopoiesis and myeloproliferative diseases. J Cell Biol *123*, 1305-8.

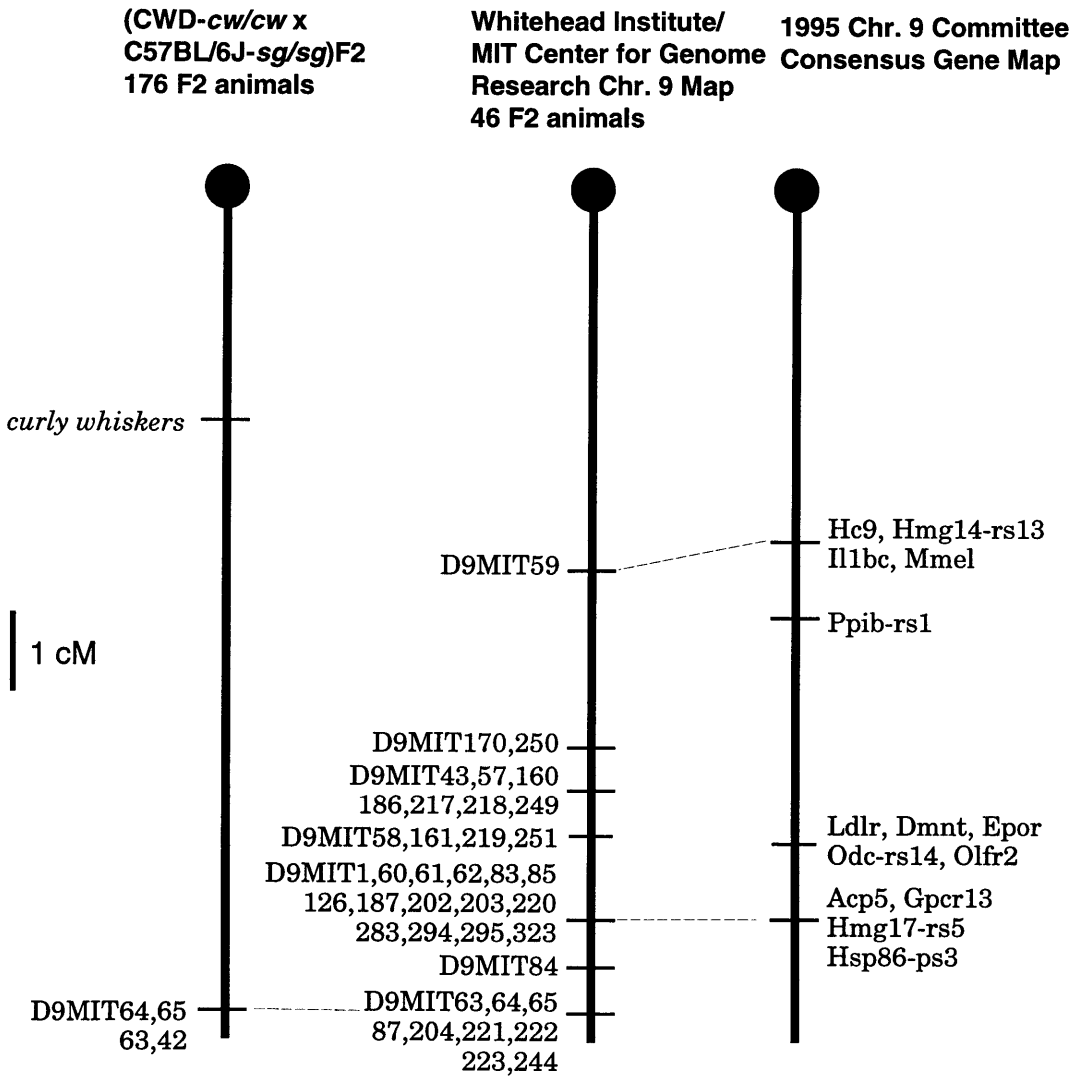
Luetteke, N. C., Qiu, T. H., Peiffer, R. L., Oliver, P., Smithies, O., and Lee, D. C. (1993). TGF alpha deficiency results in hair follicle and eye abnormalities in targeted and waved-1 mice. Cell *73*, 263-78.

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.

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.

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.

Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M., and Horvitz, H. R. (1993). The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme. Cell *75*, 641-52.




| Abbrev. | Gene Name |
|---|---|
| Hc9 | heterochromatin, Chr 9 |
| Hmg14-rs13 | high mobility group protein 14, related sequence 13 |
| Il1bc | interleukin 1 beta convertase |
| Mmel | macrophage metalloelastase |
| Ppibrs1 | peptidylprolyl isomerase B, related sequence 1 |
| Ldlr | low density lipoprotein receptor |
| Epor | erythropoietin receptor |
| Odc-rs14 | ornithine decarboxylase related sequence-14 |
| Olfr2 | olfactory receptor gene 2 |
| Acp5 | acid phosphatase 5, tartrate resistant |
| Gpcr13 | G-protein coupled receptor-13 |
| Hmg17-rs5 | high mobility group protein 17, related sequence 5 |
| Hsp86-ps3 | heat shock protein, 86 kDa, pseudogene 3 |



**Figure 1** Genetic map of the *curly whiskers* mutation
The genetic mapping location of *cw* is shown relative to MIT markers D9MIT42,63,64,65. The WI/MIT mapping cross and Chromosome 9 Committee consensus gene map are shown for reference. Dotted lines indicate markers which have been mapped on both crosses.

Short Communications

Mammalian Genome

Mammalian Genome 4, 391–392 (1993).



# Construction of a large-insert yeast artificial chromosome library of the mouse genome

Kenro Kusumi,[1] Jennifer S. Smith,[1] Julia A. Segre,[1] David S. Koos,[2] Eric S. Lander[1]

[1]Center for Genome Research, Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142 and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[2]Howard Hughes Medical Institute and Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

Received: 1 March 1993 / Accepted: 16 March 1993

We have constructed a large-insert yeast artificial chromosome (YAC) library of the mouse genome to serve as a resource for the mammalian genetics community. The new library is intended to supplement the three mouse YAC libraries that have been previously described: a library with 2.2-fold coverage in clones of average size 265 kb (Burke et al. 1991); a library with 3.5-fold coverage in clones of average size 240 kb (Chartier et al. 1992); and a library with 3-fold coverage in clones of average size 700 kb (Larin et al. 1991). Our new YAC library provides a total of 4.3-fold coverage in two parts: part A provides 0.7-fold coverage in clones of average size 480 kb, and part B provides 3.6-fold coverage in clones of average size 680 kb. Together, the four available libraries now provide 13-fold coverage of the mouse genome in YAC clones.

The YAC library was prepared with genomic DNA from C57BL/6J female mice, essentially as described by Foote and colleagues (1992) and detailed in Foote (1993). Genomic DNA was prepared from kidney nuclei as described by Strauss and coworkers (1992). The DNA was partially digested to the desired molecular weight by *Eco*RI restriction endonuclease/*Eco*RI methylase competition: *Eco*RI sites are cut only if the restriction enzyme cleaves them before the methylase protects them, with the extent of partial digestion controlled by the relative concentration of the two enzymes. The restriction-digested DNA was subjected to two rounds of size fractionation by pulsed-field gel electrophoresis. In the first round, DNA fragments greater than 460 kb were selected and then ligated to vector arms from the plasmid pYAC4 (Burke et al. 1987). In the second round, the ligated products were size-fractionated to yield fragments greater than 460 kb (for part A) or 600 kb (for part B). The resulting size-selected ligation mixture was then transformed into the yeast host strain AB1380 (*MATa, ura3-52, trp1, can1-100, lys2-1, his5, ade2-1, thr*$^-$, $\psi^-$). [N.B. The genotype of AB1380 given in Burke and associates (1987) is incorrect and has been corrected in Burke and Olson (1991).] The pYAC4 vector carries the markers URA3 and TRP1, but the TRP1 gene lacks a complete promoter element. Therefore, transformants were initially selected for a Ura$^-$ phenotype and then screened for a Ura$^-$Trp$^-$ phenotype. In addition, the pYAC4 cloning site lies within the SUP4 suppressor, which affects the host *ade2-1* color marker, so only red colonies which indicate an insertion were selected. Colonies were not streaked to single clones before inclusion in the library, and it is possible that a microtiter well may contain more than one YAC. Clones were transferred to standard 96-well microtiter plates containing AHC (yeast minimal medium) and grown to saturation at 30°C. To make frozen stocks, we added YPD (yeast complex medium), AHC, and glycerol to saturated cultures to a final concentration of 15% glycerol, 0.1× YPD, and 1.35× AHC. Frozen stocks were maintained at −85°C. In total, the preparation of the library required 17 ligations and 56 separate transformations.

Certain quality control procedures were used throughout library construction. To confirm the presence of mouse DNA in the transformants, a random sample of 5% of the transformants from each ligation was screened with a polymerase chain reaction (PCR) assay detecting the presence of the B2-repeat element, which is found at an average spacing of 30 kb in the mouse genome (Bennett et al. 1984). (The primers used were 5′-CCGAATTCGATGGCTCAGTGGTTAAGAG-3′ and 5′-CCGAATTCCACTGTAGCTGTCTTCAGAC-3′, where the underlined sequences represent non-B2 sequence; PCR products were analyzed by agarose gel electrophoresis and visualized by

*Correspondence to:* E.S. Lander

ethidium bromide.) To confirm the insert size of the YACs, a random sample of 1% of the transformants from each ligation was examined by pulsed-field gel electrophoresis. Specifically, DNA was prepared from spheroplasted yeast essentially as described by Carle and Olson (1985) and Schwartz and Cantor (1984), subjected to pulsed-field gel electrophoresis in 0.7% chromosomal grade agarose (Bio-Rad Laboratories, Richmond, Calif.) transferred to nylon membrane, and hybridized with probes for mouse repeats B1, B2, and L1 to visualize the location of the YAC band. Of approximately 4100 clones in Part A, we examined 68 clones and estimated a mean size of 480 kb and a median size of 440 kb. Of approximately 15,840 clones in part B, we have examined 218 and estimated a mean size of 680 kb and a median size of 640 kb, as shown in Fig. 1. On the basis of an estimated genome length of 3 billion bp, the library provides a total of 4.3-fold coverage of the haploid mouse genome.

The YAC library has been prepared for PCR-based screening (Green and Olson 1990), with a two-level pooling scheme based on screening of "super-pools" and "sub-pools". Specifically, the library has been partitioned into 26 blocks, each consisting of 8 microtiter plates. From each block, we prepared a single DNA super-pool (containing all $8 \times 96 = 768$ clones in the block) and 28 sub-pools (with 8 consisting of those clones in the same row, and 12 consisting of those clones in the same column). There are thus a total of 26 super-pools and $26 \times 28 = 728$ sub-pools. In preparing the pools, we grew each of the roughly 20,000 clones separately to avoid competition and subsequently pooled them before DNA preparation.

To screen the library to find clones containing a sequence-tagged site (STS) of interest, one first determines the positive super-pools and then the appropriate sub-pools. If a single plate, row, and column subpool within a block are positive, the location of the clone is uniquely specified. If multiple YACs in a block contain the same STS, then there may be some limited ambiguity in the clone address. (For example, if two YACs in a block contain a given STS, there will typically be two positive plates, rows, and columns. This information does not uniquely identify the locations of the two YACs but narrows it down to eight possible addresses.)

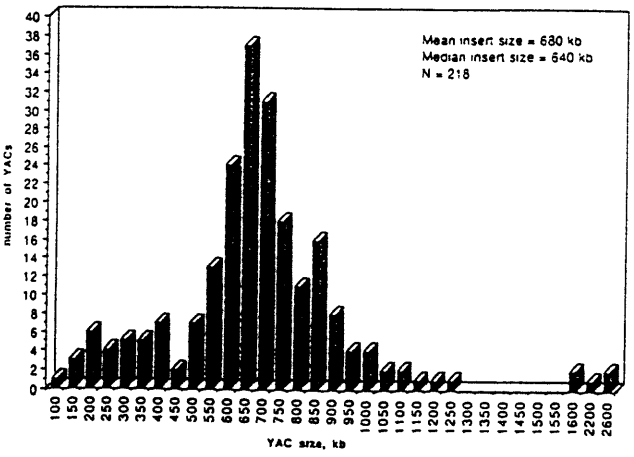



Fig. 1. A histogram showing the molecular weights of 218 YAC clones examined, from approximately 15,840 clones in part B of the YAC library.

The YAC library is available for PCR screening under the auspices of the Center for Genome Research's Core Facility at Princeton University (administered by the laboratory of S. Tilghman; FAX: 609-258-3345), according to the same procedures as described by Burke and coworkers (1991). In addition, the YAC library will be provided without license fee to companies willing to provide services (such as selling PCR pools or carrying out STS screening) at affordable prices to the scientific community.

*Acknowledgments.* We thank Adrian Argento, Hillary Katz, Asaf Presente, Nga Tang, and Marta Velez-Stringer for valuable technical assistance. In addition, we are grateful to Simon Foote, Adrienne Hilton, Maryann Haldi, Armand MacMurray, and Alix Weaver for their instruction and protocols.

## References

Bennett, K.L., Hill, R.E., Pietras, D.F., Woodworth-Gutal, M., Kane-Haas, C., Houston, J.M., Heath, J.K., Hastie, N.D. (1984). Most highly repeated dispersed DNA families in the mouse genome. Mol. Cell. Biol. 4, 1561–1571.

Burke, D.T., Olson, M.V. (1991). Preparation of clone libraries in yeast artificial-chromosome vectors. In *Guide to Yeast Genetics and Molecular Biology*, C. Guthrie and G.R. Fink, eds. (San Diego: Academic Press, Inc.), pp. 251–270.

Carle, G.F., Olson, M.V. (1985). An electrophoretic karyotype for yeast. Proc. Natl. Acad. Sci. USA 82, 3756–3760.

Chartier, F. L., Keer, J. T., Sutcliffe, M. J., Henriques, A., Mileham, P., Brown, S.D.M. (1992). Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. Nature Genet. 1, 132–136.

Foote, S. Large-insert YAC cloning. In *Current Protocols in Human Genetics*, S. Bonitz, ed. (New York: John Wiley & Sons), in press.

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

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

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

Schwartz, D.C., Cantor, C.R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37, 67–75.

Strauss, W.M., Jaenisch, E., Jaenisch, R. (1992). A strategy for rapid production and screening of yeast artificial chromosome libraries. Mammalian Genome 1, 150–157.