5)

Pole cell determination and differentiation: a molecular screen for germ cell markers in *Drosophila melanogaster*

by
Anne Williamson

B.A. Chemistry
B.A. Molecular Biology
University of California, Berkeley

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
at the
Massachusetts Institute of Technology
September, 1995

Signature of Author _____
Department of Biology
August 11, 1993

Certified by _____
Ruth Lehmann
Associate Professor of Biology
Thesis supervisor

Accepted by _____
Frank Solomon
Chair of the Graduate Committee
Department of Biology

Pole cell determination and differentiation: a molecular screen for germ cell markers in *Drosophila melanogaster*

by

Anne L. Williamson

Submitted to the Department of Biology

on August 11, 1995 in partial fulfillment of the requirements for

the Degree of Doctor of Philosophy in Biology

Abstract

This thesis describes a subtractive hybridization screen I conducted to isolate genes expressed in the primordial germ cells, or "pole cells" of the developing *Drosophila* embryo. As soon as they are formed the pole cells initiate a unique developmental program fundamentally different from that of the somatic cells. Most of what is known about the genetic requirements for pole cell development derives from analysis of maternal effect mutations that affect abdominal segmentation as well as pole cell formation in the early embryo. However, zygotically expressed genes required for differentiation of the primordial germ cells during late embryogenesis have yet to be identified. From pole cell formation at 1.5 hours post fertilization, until late larval and pupal development, 6-8 days after fertilization, there is a gap in our understanding of the process of embryonic and larval germ line differentiation. In this thesis I describe the isolation and characterization of two gonad-specific genes by a direct molecular approach to identify genes on the basis of their specific expression in the embryonic gonads. Both of these genes are expressed zygotically in germ cells of the late embryonic gonads. I describe the isolation of these genes by subtractive hybridization and the analysis of their gene products in the context of embryonic germline development.

Thesis Supervisor: Dr. Ruth Lehmann

Title: Associate Professor of Biology

Dedication

This thesis is dedicated to my mother

Nancy Tufts Leenheer
April 22, 1937-November 22, 1991

Table of Contents

CHAPTER 1:

Introduction:

Primordial germ cell Determination:

A central question in developmental biology is: How are unique developmental fates conferred upon individual cells that are initially identical in makeup and potential to their neighbors? The problem is even more complex in embryos such as *Drosophila* that develop through a syncytial stage, in which the nuclei divide without cytokinesis in a single cellular environment. Early cell formation leads to differentiation of the primordial germ cells, whereas the somatic nuclei are programmed by morphogens asymmetrically distributed within the syncytial environment before they are sequestered into separate cellular environments (Foe and Alberts, 1983). Although the timing of primordial germ cell formation is distinct from that of the somatic cells, it is unclear whether the pole cell nuclei are, like somatic nuclei, already determined before incorporation into separate cytoplasmic environments.

The embryo develops from the fertilized egg through a series of synchronized, rapid nuclear divisions in a multi-nucleate syncytium (see **figure 1**). The first nuclei to become incorporated into cells are those that migrate to the posterior tip of the embryo (Huettner, 1923; Sonnenblick, 1941) . The microtubule-dependent migration of these nuclei begins late in the seventh nuclear cycle and ends by cycle nine when they reach the posterior cortical region or "pole plasm" (Foe and Alberts, 1983; Karr and Alberts, 1986; Zalokar and Erk, 1976) Cell formation around these polar nuclei depends upon this specialized cytoplasm that accumulates at the posterior pole of the embryo, called "pole plasm" (Allis et al., 1979; Illmensee et al., 1976; Mahowald, 1962; Mahowald, 1968; Mahowald, 1971; Turner and Mahowald, 1976; Underwood et al., 1980) .

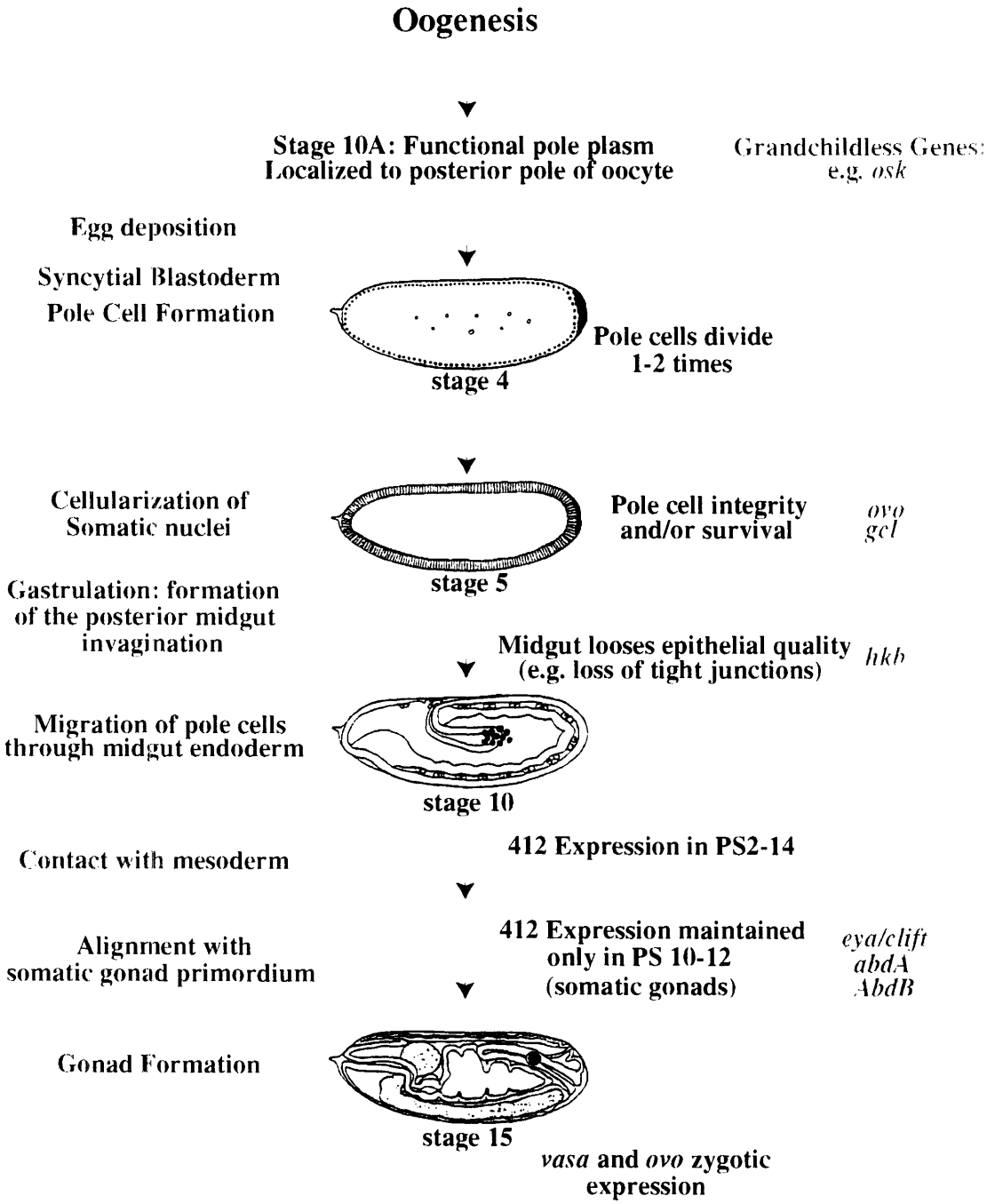


Figure 1: This figure illustrates embryonic germ cell determination and differentiation in the context of embryonic development. The major events in the development of the germline throughout embryogenesis are indicated on the left in red. Genes expressed and important pole cell- or gonad-specific events are indicated to the right in black. The genes required for the processes are indicated are shown on the far right in purple.

Mutations which disrupt the formation of pole plasm prevent incorporation of these posteriorly-migrating nuclei into pole cells. In these mutants the posteriorly migrating nuclei cellularize later, with the rest of the somatic nuclei, and take on endodermal cell fate, eventually contributing to formation of the embryonic midgut. Thus, early segregation of the posterior polar nuclei into pole plasm-containing cells shunts them permanently into a unique developmental program. Transplantation of newly formed pole cells results in colonization of the host's germline and no other tissues, demonstrating that from the time of their formation these cells are determined (Illmensee and Mahowald, 1974; Technau, 1986) .

Germ cells versus somatic cells: nuclear division and cell formation

The majority of somatic nuclei begin their cortical migration at telophase of nuclear cycle eight, penetrate the cortex by early interphase of cycle 10 (Foe and Alberts, 1983; Hatanaka and Okada, 1991; Raff and Glover, 1989) , and undergo four more synchronous nuclear division cycles before cellularization. By the end of nuclear cycle fourteen the embryo is completely cellularized (Zalokar and Erk, 1976) . During these last four somatic nuclear divisions the pole cells divide asynchronously 1-2 times to produce from 20 to 60 cells before their mitotic phase ends, roughly 3.25 hours post fertilization. (Counce, 1963; Sonnenblick, 1950; Turner and Mahowald, 1976) . There is no evidence for any further pole cell division until after formation of the embryonic gonads at late stage 16 , 16 hours post fertilization (Allis et al., 1979; Campos-Ortega and Hartenstein, 1985; Sonnenblick, 1941) .

In contrast to the mitotically arrested pole cells, by the end of cycle fourteen the somatic nuclei lose their synchronous mitotic behavior and initiate normal G1 and G2-containing cell cycles as small groups or "Mitotic Domains" throughout the embryo (Foe, 1989; Foe and Alberts, 1983) . Clonal analysis studies using transplantation of labeled cells suggest that for the most part, somatic cells of all three germ layers undergo an average of three post-blastoderm mitoses. Exceptions to this general rule include neuroblasts, which are thought to divide from five to nine times during embryogenesis (Bate and Martinez Arias, 1993; Poulson, 1950) , the progenitors of epidermal sensilla, which divide after germ band retraction (Campos-Ortega and Hartenstein, 1985) , and possibly some mesodermal derivatives, which appear to divide more than three times after the blastoderm stage (Beer et al., 1987; Campos-Ortega and Hartenstein, 1985; Technau and Campos-Ortega, 1986; Technau and Campos-Ortega, 1986) .

Polar granules: Cytological descriptions:

Numerous experiments have demonstrated that pole plasm is necessary and sufficient to confer germ cell-precursor fate on cells that contain it (Illmensee and Mahowald, 1974; Illmensee et al., 1976; Okada et al., 1974) . Electron microscopy has shown that pole plasm contains spherical electron dense masses, termed "polar granules" (Mahowald, 1962) . The polar granules contain both RNA and protein (Counce, 1963; Mahowald, 1971; Mahowald, 1971) . In particular, vasa protein, oskar protein (Dickinson and Lehmann, unpublished observations; Hay, 1990) and the mitochondrial 16S large rRNA, or "mtlrRNA," (Kobayashi et al., 1993) have been shown to be present in polar granules. However, the exact structure and composition of these organelles and of the pole plasm that surrounds them remains unknown.

Electron microscopy has revealed electron-dense structures in the ovarian nurse cell nuclei, called "nuclear bodies." Small, putative precursors of cytoplasmic polar granules are first visible during oogenesis at the posterior pole in stage 9 oocytes, and by stage 10a, polar granules of a typical appearance can be seen (Mahowald, 1962; Mahowald, 1968; Mahowald, 1971) . At this stage of oogenesis cytoplasm from the posterior pole, or 'Pole plasm' of the developing oocytes is competent to induce pole cell formation when injected into host blastoderm embryos (Illmensee et al., 1976). Immediately after egg activation the polar granules appear to fragment and are associated with polysome-like clusters (Mahowald, 1968) . Visible polar granules persist until just prior to pole cell formation when they appear to fragment and are succeeded by the re-appearance of nuclear bodies, as well as small particles surrounding the nuclear envelope called "Nuage" (Counce, 1963; Mahowald, 1971) . By stage 9 of embryogenesis, roughly 3.5 hours after egg-laying, the nuclear bodies are no longer visible in the pole cells (Campos-Ortega and Hartenstein, 1985; Mahowald, 1971) .

Extensive efforts have been made to isolate pole plasm components biochemically, and molecular identification of some embryonic pole plasm components has contributed to the knowledge of specific molecules included in these structures (Kobayashi and Okada, 1989) . For example, the *mtlrRNA* was identified on the basis of its ability to restore pole cell formation to UV-irradiated eggs (Kobayashi and Okada, 1989) . This RNA has been shown to be a component of polar granules, and although its presence correlates with pole

cell formation (Kobayashi et al., 1993) , it is not yet clear whether the mtlrRNA is required for pole cell formation (Ding et al., 1994; Kobayashi et al., 1995).

Another gene that was identified by a molecular approach is *germcell-less*, which was identified on the basis of its localization to the posterior pole and incorporation into pole cells (Jongens et al., 1992) . Heat shock promoter-driven expression of antisense *gcl* RNA during oogenesis causes a decrease in the number of germ cells formed at the posterior pole (see **figure 1**). In addition, in embryos with reduced levels of maternal *gcl* product, the pole cells that do form tend to sink beneath the somatic cellular layer into the yolk and become degraded (Jongens et al., 1992) . These experiments demonstrate a clear requirement for the *germcell-less* gene product in pole cell formation; however, mutations in the *germcell-less* locus have not yet been isolated. Immunofluorescence antibody staining shows that gcl protein is localized to the nuclear membrane, possibly to nuclear pores of pole cells at the time of their formation (Jongens et al., 1994) . Further, the germcell-less protein shows homology to nuclear lamins and is thought to be associated with the "Basket" structure surrounding the nuclear pore (Jongens et al., 1994) .

The genetics of germ cell determination: identifying components of polar granules, via the "Grandchildess-knirps" class of mutants.

The most detailed information about the composition of the pole plasm has come from identification of maternal effect genes required for abdominal segmentation in the early embryo. Genetic screens have identified nine genes required for formation of posterior pole plasm: *oskar, vasa, valois, tudor, cappucino, spire, staufen, pipsqueak and mago nashi* (Boswell and Mahowald, 1985; Boswell et al., 1991; Lehmann and Nüsslein-Volhard, 1986; Manseau and Schüpbach, 1989; Schüpbach and Wieschaus, 1986a; Siegel et al., 1993) . The products of *cappucino, spire, staufen* and probably *mago nashi* are required for transport of pole plasm components to the posterior pole, while *oskar, tudor, vasa, valois*, and *pipsqueak* appear to be required to tether and/or maintain the posterior localization of pole plasm components. In addition, it appears that continuing interaction between the *staufen* and *oskar* gene products is required for maintenance of posteriorly-localized pole plasm during oogenesis and early embryogenesis (Ephrussi et al., 1991; Rongo et al., 1995; St. Johnston, 1991). Mutations in the posterior group genes also result in an abdominal defects due to the lack of proper *nanos* localization to the posterior pole in the absence of functional pole plasm. Therefore these genes have been called the

"Grandchildless-knirps" class of mutants to reflect this dual effect on germ cell formation and abdominal segmentation (see **figure 1** and **table 1,** below)

Mislocalization of *oskar* RNA to the anterior pole induces the formation of functional germ cells at this ectopic location (Ephrussi and Lehmann, 1992) . In addition *oskar* RNA induces the formation of polar granules at this ectopic location which have been shown by electron microscopy to contain tudor protein and *mtlrRNA* (Kobayashi et al., 1995). *In situ* hybridization to whole mount embryos reveals that vasa protein, *nanos* RNA (Ephrussi and Lehmann, 1992) and *gcl* RNA are also mis-localized to the anterior pole in these embryos (P. Zamore, unpublished observations). Formation of these anterior pole cells is dependent upon *vasa* and *tudor* function(Ephrussi and Lehmann, 1992). However, it is not known whether either *mtlrRNA* or *gcl* RNA is required for formation of the ectopic anterior pole cells (Jongens et al., 1994) . These experiments demonstrate that *oskar* RNA is sufficient to nucleate polar granule formation and thereby recruit all of the factors required for pole cell formation.

In the absence of the true null phenotype of *gcl,* it is not possible to rule out a role for this gene in abdomen formation. However, if the complete absence of functional germcell-less product does not cause abdominal defects, then it will be the first gene isolated so far that is required for embryonic germ cell formation but not abdomen formation. The other mutations that produce this phenotype are weak, conditional alleles of *tudor*, *oskar*, *vasa*, *valois* and *staufen*. Although these genes clearly play a role in both abdomen formation and germ cell determination, at 18°C (the permissive temperature) they show normal abdominal segmentation patterns but lack polar granules and pole cells completely (Ephrussi et al., 1991; Lehmann and Nüsslein-Volhard, 1986; Lehmann and Nüsslein-Volhard, 1991) .

Embryonic Germ Cell Differentiation: early pole cell migration

In between pole cell formation and terminal differentiation of the ovaries and testes, embryonic germ cells go through a number of developmental phases, including migration to the site of gonad formation and interactions with different somatic tissues before they are incorporated into the embryonic gonads (Poulson, 1950) . Although the process of pole cell migration and gonad formation has been morphologically described in detail, very little is known about the genes required in germ cells and the surrounding tissues they interact with during this time interval. Examining germ cell development in the context of

mesoderm differentiation may provide clues about the cells, tissues and genes likely to influence the differentiation of the primordial germ cells (see **figure 1** and **table 1**).

During germband extension the pole cells adhere closely to the prospective endodermal cells of the anal plate and are carried dorsally as the germ band extends. These cells are then incorporated into the posterior-midgut invagination (Poulson, 1950; Sonnenblick, 1941). The fate of the endodermal cells in the posterior midgut is determined by the terminal group of maternal effect genes including *torso* and *trunk* (Nüsslein-Volhard et al., 1987). In the absence of patterning information from these and downstream zygotic genes, the gut endoderm does not differentiate properly (Jürgens and Weigel, 1988; Klingler et al., 1988; Weigel and Jäckle, 1990). For example, the gap gene *huckebein* plays an important role in determining this tissue. In *huckebein*-mutant embryos the posterior midgut epithelium fails to differentiate, blocking normal migration of the primordial germ cells through this cell-layer and into the body cavity of the embryo (Jaglarz and Howard, 1994; Warrior, 1994).

In wild type embryos, at 7 hours post fertilization, the majority of the pole cells lose their round shape and extend pseudopodia as they migrate through the midgut epithelial layer (Hay et al., 1990; Warrior, 1994). The pole cells then arrest, sitting on the basal side of the midgut endoderm cells, beneath the syncytial yolk sac membrane (Poulson, 1950). By the time the pole cells become ameboid and begin to pass through the midgut epithelium, the cells in this primordium have lost the apical-basal polarity typical of epithelial cell-layers to become solid clusters of mesenchymal cells (Poulson, 1950). However, before this point the midgut provides a barrier to pole cell migration. Unless the normal polarity of this tissue layer is relaxed, the pole cells are unable to migrate to their mesodermal destination. Mutations that cause cellularization defects in the endoderm seem to permit the passage of the primordial germ cells, (Degelmann et al., 1986; Klingler et al., 1988; Schüpbach and Wieschaus, 1986a) presumably due to lack of tight-junctions and other epithelial-specific structures that may be required for prevention of inappropriate pole cell migration (Jaglarz and Howard, 1994; Warrior, 1994).

As described above, the embryonic gonads are made up of both germ line-derived pole cells and mesodermally-derived somatic cells. The primordial somatic gonadal cells derive from the somatic mesoderm (or somatopleura), when this tissue splits from the visceral mesoderm (or splanchnopleura), during late stage ten, roughly 5 hours after egg-laying (Campos-Ortega and Hartenstein, 1985; Poulson, 1950). By stage 11, subdivision of the

somatic mesoderm generates large ventrolateral and smaller dorsolateral somatic mesoderm anlage in each segment. The ventrolateral precursors give rise to the body musculature, while the dorsolateral mesoderm differentiates into fat bodies, gonadal mesoderm and the circulatory system (Bate and Martinez Arias, 1993; Campos-Ortega and Hartenstein, 1985; Poulson, 1950) .

The somatic gonad precursor cells are distributed in the posterior compartments of abdominal segments 4 through 7 (Brookman et al., 1992; Warrior, 1994) . From late stage 11 to early stage 12, after traversing the endodermal cell layer, the pole cells contact mesoderm cells on both sides of the embryonic gut. Finally, 27-37 somatic mesodermal cells migrate anteriorly along with the primordial germ cells to form rounded embryonic gonads located in the fifth abdominal segment on either side of the embryo (Sonnenblick, 1950) . Though the exact sequence of cellular and tissue contacts required for lateral distribution of germ cells and coalescence of the gonads is unclear, a number of genes, including *abd A*, *Abd B* and the recently identified gene, *clift/eyes-absent* (Bonini et al., 1993) (H.Broihier, L. Moore, personal communication), are known to be required.

Molecular markers specific for the somatic gonads include the *412* retrotransposon which is expressed in the posterior compartment of segments 4 through 7 (Brookman et al., 1992) . The *412* retrotransposon was originally identified as a gonadal mesoderm marker as the result of experiments to identify downstream DNA target sites of the homeobox protein, Ultrabithorax (Gould et al., 1990). Immnuoprecipitation of embryonic chromatin DNA with a monoclonal antibody directed against Ubx resulted in the isolation of genomic clones containing *412* retrotransposon sequences. The expression pattern of the *412* genome includes repeated parasegmental stripes in the mesoderm of extended germ-band embryos (stages 10 and 11). This early expression then fades in most of the embryo except for parasegments 10 through 12, in which expression is maintained at high levels (Brookman et al., 1992)

Fate mapping: origins of the gonadal mesoderm in the blastoderm

Gynandromorph studies suggest the somatic gonad primordium includes a single segment-wide group of roughly ten cells located in the prospective mesoderm of abdominal segments four and five (Szabad and Nöthinger, 1992) . This primordium appears to be spatially separated from the segmentally repeated gonadal mesoderm primordia, defined by *412* expression, and may provide the signal used to locate the gonads in the fifth abdominal

segment. If the *412* expression pattern can be regarded as a true manifestation of gonadal mesoderm cell-fate, then estimates of the size and spatial derivation of the somatic gonadal primordium are at odds with the expression of the *412* mesodermal marker. However, in principle, the primordial gonadal mesoderm cells could be determined early, then migrate to their final locations in parasegments 10 through 12 (Brookman et al., 1992) . Alternatively, the segmentally repeated primordia may not carry the intrinsic information to become gonad tissue, but instead receive that information from a unique instructive center in A5. In support of this notion, mesodermal cells appear to be equivalent and to differentiate according to the type of ectoderm and other embryonic tissues they contact (Bate and Martinez Arias, 1993) .

There are a number of mutations that affect development of the somatic gonad including *abd A*, *Abd B*, *clift/eyes absent* and *schnurri* (Bonini et al., 1993; Boyle and DiNardo, 1995; Cumberledge et al., 1992) , (H. Tarczy-Broihier and L. Moore, personal communication). All of these genes appear to be required during the later stages of embryonic development for specification of gonadal mesoderm and/or coalescence of the somatic gonads. The maintenance of *412* expression in the prospective gonadal mesoderm of parasegments 10-12 is dependent upon both *abd A* and *Abd B* (Brookman et al., 1992; Cumberledge et al., 1992) . An ongoing genetic screen to isolate genes required for embryonic germ cell migration and/or gonad formation, has identified mutations which appear to affect germ cells during or just after migration through the midgut epithelium (Heather Broihier, Lisa Moore, personal communication). Interestingly, this migration occurs just after the earliest detectable zygotic transcription begins in the pole cells (Zalokar, 1976) .

The iab-4 mutation in *abdA* is a regulatory mutation, affecting the upstream regulatory sequences controlling expression the abdA protein, and not the transcription unit of the homeobox protein itself (all mutations disrupting the open reading frame are homozygous lethal (Lewis, 1978)). In iab-4 mutants the fourth abdominal segment develops cuticle structures like those of the third, and embryos homozygous for this allele demonstrate aberrant gonad formation leading to sterility in adults. The inference from this result is that *abd A* is required for proper segment identity in mesodermal tissues as well as in the ectoderm (Lewis, 1978) .

The origin of iab-4 sterility is that although the primordial germ cells traverse the midgut epithelium and associate with ventral mesoderm cells normally, by late stage 13, when the

somatic mesodermal cells normally begin to encapsulate the germ cells, gonad formation arrests. Antibody staining of *iab-4* mutant embryos reveals no detectable reduction in the level of abdA protein in the nuclei of the somatic mesodermal cells. This result is puzzling since the mutation is not thought to affect the coding sequences of the protein (Cumberledge et al., 1992) . However, given the combinatorial functioning of homeobox-containing transcription factors (Struhl and White, 1985) , it may be that fractional reductions in the level of *abd A* protein in the gonadal mesoderm primordia are sufficient to cause dramatic effects on transcription and/or repression of downstream target genes necessary for gonad formation (Cumberledge et al., 1992) . Alternatively, it has been proposed that a threshold level of abd A protein is required in the presumptive somatic gonadal mesoderm to specify formation of embryonic gonads (Cumberledge et al., 1992) and that the decrease in levels of the abd A protein in iab-4 mutants is too subtle to be detected by standard whole-mount antibody staining techniques.

The homeobox genes *Abd A* and *Abd B* are both required for normal expression of *412* in the gonadal mesoderm as well as for proper formation of embryonic gonads (Boyle and DiNardo, 1995; Cumberledge et al., 1992) . In embryos lacking the bithorax complex genes, *Ubx, abd A*, and *Abd B*, the *412*-expressing somatic gonad primordia do not form and the pole cells fail to migrate laterally as in wild type. In addition, both *abdA* and *AbdB* are expressed in gonadal mesoderm (Brookman et al., 1992; Delorenzi and Bienz, 1990; Karch et al., 1990) . Brookman and colleagues have shown that in *extra sex combs*-mutant embryos, in which *abd A* and *Abd B* are derepressed and expressed in overlapping domains throughout the embryo, the gonads still coalesce specifically in the fifth abdominal segment. Therefore, overlap of *abd A* and *Abd B* expression is not sufficient to provide the signal localizing the embryonic gonads.

Although somatic mesodermal tissues are clearly required to direct the proper assembly of the embryonic gonads, it has been shown that the germ cells are not required for the normal determination and differentiation of the somatic components of the gonads. In mutant embryos lacking primordial germ cells, the somatic gonadal mesoderm cells still migrate as in wild type, and coalesce into gonad-like structures in the embryo (Brookman et al., 1992). These agametic "gonads" then develop into ovary or testis-like organs in adults, lacking all germ line tissues. Given these results it is clear that germ cells do not play a role in specifying the dorsolateral cluster of *412*-expressing somatic mesoderm cells that will make up the somatic gonads.

Early manifestations of embryonic germ cell differentiation:

Evidence for germ cell differentiation can be examined indirectly through early differences between male and female embryos, such as pole cell number, embryonic manifestation of P-cytotype, and expression of early genes such as *ovo*. Estimates for the number of pole cells that actually populate the embryonic gonads range from 5 to 20 (Underwood et al., 1980; Wieschaus et al., 1981; Wieschaus and Gehring, 1976) . Pole cell counting studies suggest that there is a sex-specific difference in the number of primordial germ cells incorporated into embryonic gonads (Poirié et al., 1995; Sonnenblick, 1941) . The difference was detected as two distinct classes of ten hour-old embryos (stage 13). At gonad coalescence, embryonic gonads either contain 5-7 or 9-13 primordial germ cells (Sonnenblick, 1941) . This is one of the first visible manifestations of zygotic germ cell identity in the embryo. The number of germ cells per gonad increases to 8-12 versus 36-38 between 16 hour after egg laying and hatching, 20 hours after egg laying (Sonnenblick, 1941) . During this last four hours of embryogenesis the germ cells are thought to divide one or two times. In freshly hatched first instar larve, male gonads are three to three and a half times the size of female gonads (Kerkis, 1931) .

The issue of sexual dimorphism is important to the differentiation of embryonic germ cells because it raises the question of when these cells begin the inherently zygotic process of sexual differentiation. As mentioned above, the earliest visible sign of sexual dimorphism is during gonad formation, when the male somatic gonads accommodate a larger number of primordial germ cells than their female counterparts (Sonnenblick, 1941) . This process is likely to depend upon sex-specific signaling between somatic and primordial germ cells, that is required for the two types of cells to productively coordinate during the process of ovary or testis differentiation. One of the clearest examples of embryonic interactions between primordial germ cells and somatic mesoderm cells is the signal transmitted from the soma to the germ cells during embryonic germline sexual differentiation. Transplantation experiments suggest that female embryos produce a somatic factor that is required for female determination in germ cells (Granadino et al., 1993; Poirié et al., 1995)

Gonadal dysgenesis and the embryonic phenotype of *ovo* are two of the earliest known defects affecting male and female germ cells differently. Female progeny of "M", or non-P-element carrying females crossed to "P," or P-element carrying males, display a temperature-dependent sterility traceable to the degeneration of the primordial germ cells in female embryos prior to stage 16. Female germ cells appear to be uniquely susceptible to

degeneration at 27.5°C due to P-element induced gonadal dysgenesis (Engels, 1983; Brigliano and Kidwell, 1983). This female-specific sensitivity to P-element inheritance can be partially overcome if the female progeny inherit a dominant allele of the *ovo* gene from their P-element carrying father (Wei et al., 1991) . The recessive, loss of function phenotype of *ovo* is that the germ cells of homozygous embryos appear to die during embryogenesis and adult females produce normal somatic ovarian structures in which no egg chambers are visible (Oliver et al., 1987) .

The *ovo* transcript is provided maternally to the developing oocyte and is ubiquitous throughout the early embryo. Ovo protein is also distributed throughout blastoderm embryos, and disappears from all cells except the primordial germ cells during germ band retraction (stage 12; Mével-Ninio et al., 1991; Mével-Ninio et al., 1995). Using *lacZ* fusions with *ovo* genomic sequences to generate transgenic animals, the authors examined zygotic transcription of this gene. In contrast to the genetic data indicating an early role for the zygotic *ovo* product during embryogenesis (Oliver et al., 1987) , zygotic expression of the *ovo-lacZ* fusion RNA is not detectable until embryonic stage 17 (just before hatching), and does not appear to be expressed differently in male and female embryos (Mével-Ninio et al., 1995).

Although the molecular data appear to contradict the genetic evidence, it is nevertheless possible that the fusion construct used does not allow complete recapitulation of endogenous *ovo* regulation. In support of this, the *ovo-lacZ* fusion construct (which does not contain all of the sequences contained in the original genomic rescue fragment) does not complement either the loss of function phenotype, nor does it completely rescue the dominant *ovo* defect. This fusion construct may therefore be lacking promoter sequences that either enhance the early zygotic expression of this gene and/or repress transcription at this locus in male germ cells. If the apparent delay in zygotic transcription of the *ovo-lacZ* transgene is due to an experimental artifact, and the RNA is present in the early embryonic germline, then the sex-specificity could be contributed by female germline-specific factors that promote post-transcriptional activation of *ovo* specifically in female embryonic germ cells, or 3'UTR-mediated instability of this RNA in male primordial germ cells.

Germline sex determination

In the soma, the sex of each cell is determined autonomously by the X-chromosome to autosome, (X:A) ratio and its downstream effect on transcription and splicing of *Sxl* mRNAs during the blastoderm stage. Although sex determination of the germ cells is integral to their differentiation, it is unclear when the sex of the germ cells is first established. Splicing control of *Sxl* RNA to prevent the translation of Sxl protein occurs in the male germ cells, but the regulation depends upon a different set of genes from those that act in the soma. For example, *sisterless a*, *sisterless b*, *runt* and *daughterless* are all required for this process in somatic cells, but have no detectable function in the germ line (Cronmiller and Cline, 1987; Granadino et al., 1993; Schüpbach, 1982; Steinmann-Zwicky, 1994) . Instead, a number of female-sterile genes in the so called "ovarian tumor " class, (including *ovo*), act upstream of *Sxl* to control female-specific differentiation of the germline. *Sxl* RNA is not present in pole cells during the blastoderm stage when it is first activated in the soma. This protein is apparently present only in the female germline cells of late third-instar larvae just prior to ovarian reorganization and differentiation. It is not known whether distinct zygotic promoter elements exist to direct this larval-specific germline expression of *Sxl* (Bopp et al., 1993; Keyes et al., 1992) .

In addition to *ovo*, *ovarian tumor* (*otu*), *female lethal (2)d* (*fl(2)d*), *sans fille* (*snf*), *bag of marbles* (*bam*), *fused* (*fu*), and *orb* all appear to participate in the control of sex-specific expression of *Sxl* in XX germ cells (see **Table 3**). For example, aberrant expression of the male-specific form of *Sxl* RNA can be detected in the germline of XX flies mutant for *snf, otu, ovo, bam* or *fused* (Bopp et al., 1993; Keyes et al., 1992; Oliver et al., 1993) . These data suggest that these five genes may be involved directly or indirectly, in the splicing of *Sxl* RNA in the germline. *Bam* and *fused* have been shown to act not at the level of RNA synthesis or splicing but at the level of nuclear-cytoplasmic localization of the Sxl protein (Bopp et al., 1993) . Whether these factors function in the germline autonomously, inductively from the soma, or both remains to be determined.

Sex determination in the germline, unlike in the soma, does not appear to be completely cell autonomous. Inductive interactions with the soma are critical to ensure terminal differentiation of the wild type female and male germlines. The genes *transformer* , *transformer-2* , and *doublesex* are all required in the surrounding female soma for XX germ cells to undergo normal ovarian differentiation (Nöthinger et al., 1989; Steinmann-Zwicky et al., 1989) . Similarly, although XY pole cells express male-specific markers

regardless of the sexual identity of the surrounding somatic cells, they require a male somatic environment to fully differentiate according to the normal male-specific pathway. Thus, neither XX nor XY germ cells can undergo normal gametogenesis in the absence of sex-specific cues from the surrounding somatic tissues (Nöthinger et al., 1989; Steinmann-Zwicky et al., 1989) .

Both *ovo* and *otu* are thought to be required for survival of germ cells during late embryonic and early larval female germline development, before Sxl protein is expressed in these cells (Oliver et al., 1987) . *Ovo* is unique among the genes in this class because it is apparently required in female pole cells at the time of formation during blastoderm stage as well. Loss of function mutations give rise to female specific germ cell death starting before gastrulation and continuing until 14 hours after egg-laying (AEL), when coalescence of the embryonic gonads occurs (Oliver et al., 1987) , thereby earning it the label of a "Germline maintenance" mutant. *Ovo* may be required in female embryonic germ cells to assess the somatic sex of the early embryo and if it does not match that of the germ cells, the theory is that they then degenerate, explaining why they are often lost even before incorporation into the midgut pocket (Mahowald and Wei, 1994)

Germline-specific splicing: Embryonic splicing of P-transposase third intron.

Another one of the earliest reported germ cell differentiation events is splicing of the third intron of the P-element transposase gene third intron. The transposase RNA was originally shown to be spliced to produce active transposase only in the germline-derived cells of ovaries and testes (Laski et al., 1986; Rio et al., 1986). This tissue specific splicing event has been shown to occur in germ cells as early as four hours after egg laying, reaching a maximum at five to six hours of embryonic development (Kobayashi et al., 1993) . In addition, double-labeling experiments with anti-vasa antibodies and histochemical detection of beta-galactosidase activity reveal a strong correlation between pole cells that are capable of productively splicing the P-element third intron and those that populate the embryonic gonads. These experiments raise the possibility that there exists an intrinsic mechanism responsible for the regulation of germline-specific differentiation events and that this machinery may have been co-opted by the transposition machinery to ensure efficient propagation of P-elements to the next generation. The tissue-specific splicing of the transposase transcript may therefore reflect an intrinsic mechanism for determining which of the numerous primordial germ cells is competent to populate the embryonic gonads (Kobayashi et al., 1993) .

Oogenesis, spermatogenesis, and male/female sterile mutants:

Mutations in genes required for germline differentiation before the sexual dimorphism in the developing embryonic germline is manifest should cause sterility in both sexes. In practice screening for male/female sterility has yielded a large number of genes that affect fertility in one sex or in both, but few so far that have demonstrable effects on embryonic germ cell differentiation (Castrillon et al., 1993; Schüpbach and Wieschaus, 1991). It is possible that a large proportion of the genes required for zygotic differentiation of the embryonic germline are required elsewhere in the developing embryo, larva, or pupa for viability and therefore have not been isolated in screens requiring viability of adults. Mutations in genes that act later in differentiation may also cause male/female sterility because of shared factors that act during homologous germline differentiation processes; for example, proliferation of germline stem cells (King, 1970; Szabad et al., 1979) and the four rounds of incomplete cell division required to generate a sixteen cell clusters in both ovaries and testes.

The *diaphanous* gene fits the criteria of a mutant with potential effects on embryonic germline development. Hypomorphic alleles of the locus are homozygous viable and exhibit both male and female sterility (Castrillon et al., 1993) . *Diaphanous* males have late-stage cysts at eclosion, however the testes are empty in five-day-old flies. Females carrying the P-element over deficiency are semi-sterile and their ovaries contain few egg chambers. These phenotypes are, however, more consistent with *diaphanous* playing a role in stem cell divisions than in embryonic germline differentiation.(It is notable that the null phenotype of this gene is lethality, since no viable male female-sterile mutants with embryonic germ cell-specific phenotypes have been identified.)

Parallels between Drosophila, Xenopus and C. elegans: Germline determination

There are numerous parallels between germline cells in C. elegans, Drosophila and Xenopus leavis. The germ cells in all three organisms differ from their somatic counterparts not only in size and cleavage pattern, but also by the presence of unique, electron-dense, cytoplasmic organelles, called "P-granules" in C. elegans, "Germ plasm" in Xenopus, or "Polar granules" in Drosophila (Eddy, 1975) .

In C. elegans "P-granules" have been described as electron-dense granules resembling polar granules in Drosophila and germinal granules in Xenopus (Krieg et al., 1978) . P-granules are uniformly distributed in the nematode egg before they become localized to the posterior pole, prior to the first embryonic division (Strome and Wood, 1983) . During embryogenesis the P-granules are asymmetrically partitioned into the P1-P4 daughter cells in succession during the first four blastomere divisions. The germline founder cell P4 has been found to give rise exclusively to germ line tissues (Deppe et al., 1978). No determinative role for P-granules has yet been established but they nevertheless serve as reliable markers for the germ cell lineage throughout embryonic, larval and adult stages (Strome, 1993; Strome and Wood, 1983)

In a different nematode species, Ascaris megalocephala, the asymmetric distribution of germline granules correlates with protection against chromosomal diminution (Eddy, 1975). Blastomeres that do not inherit germinal granules develop as somatic cells and undergo dramatic and permanent chromosomal rearrangement. If the first cleavage division is disrupted in such a way that the germinal granules are distributed evenly between the first two blastomeres, chromosomal diminution does not take place in either daughter cell (Eddy, 1975). By comparison, mutations in *par-4* cause unequal cell division in early C. elegans embryos in which P-granules are divided equally between early blastomeres. Mutations in this gene result in the absence of maternal RNA degradation that normally occurs only in the somatic cells, and repression of embryonic transcription normally found in somatic cells but not germ cells in early embryos (Seydoux and Fire, 1994) .

Genetic screens to identify P-granule components in C. elegans have been conducted in which embryos laid by homozygous mutant females were screened for disruptions in germ cell formation. Although numerous mutations causing defects in sterility of the progeny were isolated, none were seen to be due to defects in P-granule structure or composition (Capowski et al., 1991) . As in Drosophila, it may be difficult to recover or identify mutations in genes that are not only required for germ cell determination but also for survival of the homozygous female germline cells themselves.

Xenopus oocytes also contain P-granule-like material or "Germ plasm". The "Germ plasm" in Xenopus has been described as "Electron-dense granulofibrillar material" that originates in the mitochondrial cloud (Heasman et al., 1984) . Germ plasm is restricted to the primordial germ cells during embryogenesis and is thought to act as the determinant of these cells (Bounoure, 1939; Mahowald and Hennen, 1971; Smith and Williams, 1975).

Transplantation studies reveal that single blastomeres containing germ plasm are capable of populating the genital ridges of host neurula embryos, and are therefore thought to be capable of populating the germline (Ikenishi, 1987) . However, these cells do not appear to be irreversibly determined in Xenopus as they are in Drosophila (Wylie et al., 1985) .

Transcription in early embryos:

Unlike in Drosophila and Xenopus, where general zygotic transcription does not start until the early rapid divisions are complete, embryonically transcribed RNAs are already detectable in C. elegans at the four cell stage (Seydoux and Fire, 1994) . However, as in Drosophila, zygotic transcription in the developing germline appears to be controlled by a different mechanism than in the soma and is not detected early during embryogenesis. In Xenopus, generalized zygotic transcription does not occur until the 4000 cell-stage when there is a shift from synchronous to asynchronous cell divisions (Gerhart, 1980; Newport and Kirschner, 1982; Yasuda and Schubiger, 1992) . Although transcription of specific zygotic RNAs has been detected in the somatic cells at the 32 cell-stage, it is not yet known when zygotic transcription begins in the embryonic germline.

In Drosophila the earliest detectable zygotic transcription is found in the soma at nuclear cycle 10, and possibly earlier (Edgar and Schubiger, 1986). In somatic cells, the increase in the ratio of nuclei to cytoplasm and/or the length of interphase plays a key role in the activation of zygotic transcription during embryogenesis (Edgar and Schubiger, 1986) . The mechanism for controlling the onset of transcription may also involve titration of repressor molecules by increasing amounts of DNA (Almouzni and Wolffe, 1995) and/or lack of active transcriptional activators prior to 3.5 hours of development. One of the unique characteristics of germ cells in Drosophila is that from the time of their formation until their incorporation into the midgut pocket, they remain transcriptionally silent. The first stage at which transcription has been detected in pole cells is not until ~3.5 hours post fertilization, just before the pole cells migrate through the midgut epithelium (Zalokar, 1976) . This transcriptional silencing could be due to specific inactivation of transcription factors in early pole cells or to inaccessibility of the DNA to transcription factors at this time in development. The nuclear-cytoplasmic ratio model for activation of zygotic transcription in the soma does not correlate with the fact that these cells are not mitotically active until 14 hours after egg-laying, at least 10 hours later than generalized transcription is first detectable in the pole cells (Zalokar, 1976) . In fact, repression of transcription in the germ cells may be linked to or depend upon their mitotic quiescence.

The timing of nuclear divisions in the soma of both Drosophila and Xenopus embryos appears to be controlled by the nuclear to cytoplasmic ratio. This ratio in turn appears to dictate the timing between nuclear divisions and S-phases, i.e., the length of G2, which dictates the onset of transcriptional activation, first detectable at cycle 11 in Drosophila (McKnight and Miller, 1976; McKnight and Miller, 1979) . The zygotic genome appears to become transcriptionally competent during cycle 10; when Edgar and Schubiger (1986), propose that proteins required for transcriptional activation are synthesized. These authors speculate that the transcriptional silencing of the pole cells is due to their cellularization at cycle 10, when somatic transcription factors are synthesized in the rest of the cytoplasm from maternal RNAs, which may not be translated in pole cells or are excluded from them when these cells form.

As in C. elegans, maternal factors segregated unevenly between soma and germline in Drosophila may cause repression, directly or indirectly, of transcription in germline cells. The repression of zygotic transcription in the pole cells correlates with their inheritance of polar granules. Polar granules have been shown to contain *nanos* protein (L. Dickinson and R. Lehmann, unpublished observations), which plays a key role in translational repression of hunchback RNA. In fact, the *nanos* protein is present in pole cells throughout embryonic development and has recently been shown to be required for the differentiation of female germ cells during oogenesis (D.Curtis unpublished observations) Therefore, *nanos* itself could play a role in translational repression of germline transcriptional activators or cell-cycle regulators that prevent the pole cells from entering G2 and/or initiating zygotic transcription (Andéol, 1994; Edgar et al., 1986)

Almouzni and colleagues have shown (Almouzni et al., 1991) that although class II gene basal transcription machinery, including the TATA binding factor, is fully competent in the cleavage-stage Xenopus embryo, there is a mechanism in place before the mid-blastula transition to inhibit class II basal transcription machinery from stable association with promoter elements. In addition, these authors postulate that one component of transcriptional quiescence in Xenopus embryos prior to MBT is the absence or functional constraint of transcriptional activators (Almouzni et al., 1991) . In fact, both chromatin assembly and lack of transcriptional activators may be responsible for transcriptional quiescence of class II and III genes in the pre-MBT embryo (Almouzni and Wolffe, 1995) .

Searching for Genes controlling embryonic germ cell differentiation:

Polar granules are visible in various forms beginning before pole cell formation and continuing through the onset of terminal gonad differentiation (Mahowald, 1971) . These granules are therefore likely to play some role in the differentiation and/or maintenance of the germ line during development. However, the biochemical approach of directly isolating pole plasm components has proven difficult, because of the small amount of polar granule material in the embryo (Waring et al, 1978). Classical genetic screens have identified genes that are maternally required for pole cell determination (Lehmann and Nüsslein-Volhard, 1986) and terminal differentiation of the ovaries and testes (Castrillon et al., 1993; Schüpbach and Wieschaus, 1989; Schüpbach and Wieschaus, 1991) . However, such screens have not yet led to the identification of genes required for early pole cell differentiation.

There is a large body of information about the specific events that occur during embryonic germ cell differentiation, including migration through the midgut epithelium, specific association with somatic mesodermal cells, and coalescence into embryonic gonads. A number of mutations in genes that affect the somatic tissues involved in these processes have been isolated, but the programming required in the germ cells themselves, and the real nature of the cellular interactions required for migration of the primordial germ cells from the posterior of the embryo to the coalesced gonads remain a mystery. The number of genes required for development of germ cells in the embryo appears to be either very small, or many of them may also be required in other tissues, and therefore for viability of the developing embryo or adult. In addition it is likely that a number of genes are refractory to identification by classical screening techniques because they are expressed and required not only zygotically, but maternally as well. The products of such loci may be amply provided to the freshly laid egg in the form of maternal transcripts whose protein products can perdure well into late embryonic life of the developing progeny and in some cases into adult tissues (e.g. *nanos*, D. Curtis, unpublished observations).

Primordial germ cell formation is but the first in a series of developmental processes leading ultimately to terminal differentiation of the germline. If germ plasm components are derived from genes that are required in the embryonic germ cells for survival, then their role in germ cell determination may have gone undetected (Mahowald and Wei, 1994). Homozygous mutant embryos would contain necrotic pole cells that would not survive through to oogenesis or spermatogenesis to produce the fertilized oocytes in the next

generation. Moreover, if the gene products are required for oogenesis itself at the same time or after they are required for germ plasm biogenesis, then it may be impossible to examine the pole plasm of the next generation since oogenesis will be disrupted.

The aim of this work is to isolate genes expressed in the embryonic germ cells during the various phases of their differentiation. The hope is that by studying the expression of germ-cell specific genes and the factors that control their expression, we will identify molecules controlling the complex, coordinated interactions that occur in the embryo to form the specialized embryonic gonad. Given the difficulty of isolating genes required for differentiation of the embryonic germ cells by genetic means, I have undertaken a molecular screen to identify RNAs specifically expressed or stably maintained in the pole cells during embryogenesis. The molecular approach circumvents the difficulties of screening for genes required at multiple stages of development and for survival of the animal.

The feasibility of isolating developmentally important genes by molecular screening approaches, such as the subtractive hybridization, depends upon the level of expression of the genes of interest in the tissues or cell types being studied. In order to target genes expressed in germ cells I have generated a cDNA library enriched for clones expressed in germ cells by subtracting cDNA derived from agametic embryos from the cDNA of normal, pole cell-forming animals. In principle, pole cell-specific transcripts may include maternally provided messages stabilized in the primordial germ cells by specific interactions with germ plasm components, or zygotic transcripts, activated in primordial germ cells during the embryonic gonad formation process.

Table 1. This table provides a summary of the genes known to be required during embryonic germ cell formation and differentiation (as described above). Although there are many genes known to be required for pole cell determination, and many others known to be required in somatic tissues for embryonic gonad formation, there is only one gene that is known to be required zygotically for differentiation of the embryonic germ cells (ovo). The table is divided into genes required for pole cell determination, genes required for pole cell differentiation, genes required in the soma for gonad formation and markers, expressed either in germline or somatic gonad tissues during embryogenesis. In each case the tissue in which the gene is required is indicated, as well as a brief description of its function.

Genes Required for Pole Cell Determination		
Maternal Genes:	**Tissue:**	**Function:**
oskar (osk)	germline RNA and protein localized to posterior pole	pole plasm assembly pole cell determination (Lehmann and Nüsslein-Volhard, 1986)
vasa (vas)	germline pole plasm component: protein (soma?)	pole plasm assembly pole cell determination oogenesis (cellularization?) (Schüpbach and Wieschaus, 1986a)
tudor (tud)	germline pole plasm component: protein	pole plasm assembly pole cell determination (Boswell and Mahowald, 1985)
valois (val)	germline soma?	transport of pole plasm to posterior pole (cellularization?) (Schüpbach and Wieschaus, 1986a)
mago-nashi (mago)	germline	transport of pole plasm to posterior pole (Boswell et al., 1991)
cappucino (cappu)	germline	transport of pole plasm to posterior pole dorsal-ventral patterning (Manseau and Schüpbach, 1989)
spire (spir)	germline	transport of pole plasm to posterior pole dorsal-ventral patterning (Manseau and Schüpbach, 1989)
staufen (stau)	germline protein localized to posterior pole during oogenesis and to both poles during embryogenesis	pole plasm assembly at posterior pole (maintenance of oskar RNA and protein localization?). Localization of bicoid RNA to anterior pole (Schüpbach and Wieschaus, 1986a; St. Johnston, 1991)
pipsqueak (psq)	germline	pole plasm assembly: vasa mRNA expression oogenesis (Siegel et al., 1993)
germcell-less (gcl)	germline	pole cell formation (Jongens et al., 1992)
Genes Required for Pole Cell Differentiation		
Sex lethal (Sxl)	female germline and soma	Required for normal oogenesis (loss of function mutations result in tumorous ovaries (Schüpbach, 1985))
ovo (maternal and zygotic functions)	germline	**post-blastoderm pole cell survival (?)**, oogenesis: sex-specific expression of Sxl in germ cells(Bopp et al., 1993; Oliver et al., 1993)
ovarian tumor (otu)	germline	oogenesis: sex-specific expression of Sxl in germ cells (Bopp et al., 1993; Oliver et al., 1993)

female lethal (2)d (fl(2)d)	germline	oogenesis: sex-specific expression of Sxl in germ cells (Granadino et al., 1992)
sans fille (snf)	germline	oogenesis: sex-specific expression of Sxl in germ cells(Bopp et al., 1993; Oliver et al., 1993)
bag of marbles (bam)	germline	oogenesis: sex-specific expression of Sxl in germ cells(Bopp et al., 1993)
fused (fu)	germline	oogenesis: sex-specific expression of Sxl in germ cells(Bopp et al., 1993; Oliver et al., 1993)
orb (orb)	germline	sex-specific expression of Sxl in germ cells of developing ovaries, and assymteric distribution of oskar and gurken mRNAs during late oogenesis (D/V and A/P axis formation) (Christerson and McKearin, 1994)
transformer (tra)	soma	oogenesis: inductive signal from soma to germline: required for normal Sxl expression in XX germ cells (Oliver et al., 1993)
transformer-2 (tra-2)	soma	oogenesis: inductive signal from soma to germline: required for normal Sxl expression in XX germ cells (Oliver et al., 1993)
doublesex (dsx)	soma	oogenesis: inductive signal from soma to germline: required for normal Sxl expression in XX germ cells (Oliver et al., 1993)
Genes Required in the Soma for Embryonic Gonad Formation		
Zygotic Genes:	**Tissue:**	**Function:**
Ultrabithorax (Ubx)	mesoderm: PS 6-12 ectoderm: PS 6-12(Bate and Martinez Arias, 1993; Tremml and Bienz, 1989)	Ubx can functionally substitute for abdA to allow normal encapsulation of the gonads. (Greig and Akam, 1995)
abdominal A (abdA)	mesoderm: PS 8-12 ectoderm: PS 7-13 (Bate and Martinez Arias, 1993; Tremml and Bienz, 1989)	gonad coalescence, maintenance of *412* expression in PS 10-12. Specification of anterior somatic gonad cells and, in combination with AbdB, specification of posterior somatic gonad cells (Boyle and DiNardo, 1995; Cumberledge et al., 1992)
Abdominal B (AbdB)	mesoderm: PS 10-14(Bate and Martinez Arias, 1993; Tremml and Bienz, 1989) ectoderm: PS 10-14 (Akam, 1987)	gonad coalescence, maintenance of *412* expression in PS 10-12, in combination with abdA, specification of posterior somatic gonads and expression of eya/clift in these cells (Boyle and DiNardo, 1995; Brookman et al., 1992)
clift/eyes absent (cli/eya)	somatic gonad primordia?	gonad coalescence maintenance of *412* expression in PS 10-12 (Bonini et al., 1993)

huckebein (hkb)	endoderm	determination of endoderm: required for differentiation of posterior midgut epithelium, allowing migration of pole cells into the body cavity (Jaglarz and Howard, 1994; Warrior, 1994).
Markers: germline	**Tissue:**	**Time of expression:**
faf-lacZ	germline: fat facets-lacZ fusion protein localized to posterior pole and pole cells	Faf is required for cell fate determination of non-photoreceptor cells in the eye-imaginal disc, and during oogenesis for somatic cellularization during embryogenesis (Fischer-Vize et al., 1992)
OvoB	germline: ovo-lacZ fusion protein expressed in pole cells	maternal product: RNA and protein localized to pole cells upon formation zygotic expression: stage 17 in germ cells of embryonic gonads (Mével-Ninio et al., 1995)
Markers: somatic gonad primordia		
68-77	somatic gonad primordia 6.8kb of abx regulatory region from BX-C fused to lacZ	stage 11: patches in mesoderm from PS 2-14, higher levels in posterior regions of PS10-12. Also expressed in ectoderm. Stage 12: expression maintained in anteriorly moving cells and coalesced gonads in PS 10. (Boyle and DiNardo, 1995; Simon et al., 1990)
412 retrotransposon	Somatic gonad primordia	stage 11: dorso-lateral cell clusters in PS2-14, stage 12-13: high levels in PS 10--12, stage 14-16: somatic gonad cells, coalesced in PS10 (Brookman et al., 1992)
Blue-tail	posterior somatic gonad primordia, 1 kb of the iab-7 regulatory region fused to lacZ (P-insert in BX-C)	stage 11: somatic gonad cells in PS12-14, stage 14: posterior-most somatic gonad cells (Boyle and DiNardo, 1995; Galloni et al., 1993)

References

Akam, M. 1987. The molecular basis for metameric pattern in the Drosophila embryo. *Development.* 101:1-22.

Allis, C. D., E. M. Underwood, J. H. Caulton, and A. P. Mahowald. 1979. Pole cells of Drosophila melanogaster in culture. Normal metabolism, ultrastructure and functional capabilities. *Dev Biol.* 69:451-465.

Almouzni, G., M. Méchali, and A. P. Wolffe. 1991. Transcription complex disruption caused by transition in chromatin structure. *Mol. Cell. Biol.* 11:655-665.

Almouzni, G., and A. P. Wolffe. 1995. Constraints on transcriptional activator function contribute to transcriptional quiescence during early Xenopus embryogenesis. *The EMBO Journal.* 14:1752-1765.

Andéol, Y. 1994. Early transcription in different animal species: implication for transition from maternal to zygotic control in development. *Roux's Arch Dev Biol.* 204:3-10.

Bate, M., and A. Martinez Arias. 1993. The Development of Drosophila melanogaster. 746.

Beer, J., G. M. Technau, and J. A. Campos-Ortega. 1987. Lineage analysis of transplanted individual cells in embryos of Drosophila melanogaster. IV. Commitment and proliferative capabilities of mesodermal cells. *Wilhelm Roux's Archives of Developmental Biology.* 196:222-230.

Bonini, N. M., W. M. Leierson, and S. Benzer. 1993. The *eyes absent* gene: genetic control of cell survival and differentiation in the developing Drosophila eye. *Cell.* 72:379-395.

Bopp, D., J. I. Horabin, R. A. Lersch, T. W. Cline, and P. Schedl. 1993. Expression of the *Sex-lethal* gene is controlled at multiple levels during *Drosophila* oogenesis. *Development.* 118:797-812.

Boswell, R. E., and A. P. Mahowald. 1985. *tudor*, a gene required for assembly of the germ plasm in *Drosophila melanogaster*. *Cell*. 43:97-104.

Boswell, R. E., M. E. Prout, and J. C. Steichen. 1991. Mutations in a newly identified *Drosophila melanogaster* gene, *mago nashi*, disrupt germ cell formation of mirror-image symmetrical double abdomen embryos. *Development*. 113:373-384.

Boyle, M., and S. DiNardo. 1995. Specification, migration and assembly of the somatic cells of the Drosophila gonad. *Development*. 121(6):1815-1825.

Brookman, J. J., A. T. Toosy, L. S. Shashidhara, and R. A. H. White. 1992. The 412 retrotransposon and the development of gonadal mesoderm in *Drosophila*. *Development*. 116:1185-1192.

Campos-Ortega, J. A., and V. Hartenstein. 1985. The embryonic development of Drosophila melanogaster. Springer-Verlag, Heidelberg.

Capowski, E. E., P. Martin, C. Garvin, and S. Strome. 1991. Identification of Grandchildless Loci Whose Products Are Required for Normal Germ-line Development in the Nematode Caenorhabditis elegans. *Genetics*. 129:1061-1072.

Castrillon, D. H., P. Cönczy, S. Alexander, R. Rawson, C. G. Eberhart, S. Viswanathan, S. DiNardo, and S. A. Wasserman. 1993. Toward a molecular genetic analysis of spermatogenesis in *Drosophila melanogaster*: characterization of male-sterile mutants generated by single P element mutagenesis. *Genetics*. 135:489-505.

Christerson, L. B., and D. M. McKearin. 1994. *orb* is required for anteroposterior and dorsoventral patterning during Drosophila oogenesis. *Genes & Development*. 8:614-628.

Counce, S. J. 1963. Developmental morphology of polar granules in Drosophila. *J. Morphol*. 112:129-145.

Cronmiller, C., and T. W. Cline. 1987. The Drosophila sex determination gene *daughterless* has different functions in the germ line versus the soma. *Cell*. 48:479-487.

Cumberledge, S., J. Szabad, and S. Sakonju. 1992. Gonad formation and development requires the *abd-A* domain of the bithorax complex in *Drosophila melanogaster*. *Development*. 115:395-402.

Degelmann, A., P. A. Hardy, N. Perrimon, and A. P. Mahowald. 1986. Development analysis of the torso-like phenotype in Drosophila produced by a maternal-effect locus. *Developmental Biology*. 115:479-489.

Delorenzi, M., and M. Bienz. 1990. Expression of Abdominal-B homeoproteins in Drosophila embryos. *Development*. 108:323-329.

Deppe, U., E. Schierenberg, T. Cole, C. Krieg, D. Schmitt, B. Yoder, and G. von Ehrenstein. 1978. Cell lineages of the embryo of the nematode Caenorhabditis elegans. *Proc. natn. Acad. Sci., U.S.A.* 75:376-380.

Ding, D., K. L. Whittaker, and H. D. Lipshitz. 1994. Mitochondrially encoded 16S large ribosomal RNA is concentrated in the posterior polar plasm of early Drosophila embryos but is not required for pole cell formation. *Dev. Biol.* 163:503-515.

Eddy, E. M. 1975. Germ plasm and the differentiation of the germ cell line. *International Review of Cytology (New York, NY)*. 43:229-280.

Edgar, B. A., C. P. Kiehle, and G. Schubiner. 1986. Cell cycle control by the nucleo-cytoplasmic ratio in early Drosophila development. *Cell*. 44:365-372.

Edgar, B. A., and G. Schubiger. 1986. Parameters controlling transcriptional activation during early Drosophila development. *Cell*. 44:871-877.

Ephrussi, A., L. K. Dickinson, and R. Lehmann. 1991. Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell*. 66:37-50.

Ephrussi, A., and R. Lehmann. 1992. Induction of germ cell formation by *oskar*. *Nature*. 358:387-392.

Fischer-Vize, J., G. M. Rubin, and R. Lehmann. 1992. The *fat facets* gene is required for *Drosophila* eye and embryo development. *Development*. 116:985-1000.

Foe, V. E. 1989. Mitotic domains reveal early commitment of cells in Drosophila embryos. *Development.* 107:1-22.

Foe, V. E., and B. M. Alberts. 1983. Studies of nuclear and cytoplasmic behavior during the five mitotic cycles that precede gastrulation in Drosophila embryos. *Journal of Cell Science (Cambridge).* 61:31-70.

Galloni, M., H. Gyurkovics, P. Schedl, and F. Karch. 1993. The bluetail transposon: evidence for independent cis-regulatory domains and domain boundaries in the bithorax complex. *EMBO Journal.* 12:1087-1097.

Gerhart, J. G. 1980. Mechanisms regulating pattern formation in the amphibian egg and early embryo., p. 133-315. *In* Biological Regulation and Development (Vol. 2). R. F. Goldberger, R. F. Goldbergers. Plenum Press, New york. 133-315.

Granadino, B., A. San Juan, P. Santamaria, and L. Sanchez. 1992. Evidence of a dual function in fl(2)d, a gene needed for Sex-lethal expression in Drosophila melanogaster. *Genetics.* 130:597-612.

Granadino, B., P. Santamaria, and L. Sánchez. 1993. Sex determination in the germ line of *Drosophila melanogaster:* activation of the gene *Sex-lethal. Development.* 118:813-816.

Greig, S., and M. Akam. 1995. The Role of Homeotic Genes in the Specification of the Drosophila Gonad. *Current Biology:*. In Press:

Hatanaka, K., and M. Okada. 1991. Retarded nuclear migration in Drosophila embryos with aberrant F-actin reorganization caused by maternal mutations and by cytochalasin treatment. *Development.* 111:909-920.

Hay, B., L. H. Jan, and Y. N. Jan. 1990. Localization of *vasa*, a component of Drosophila polar granules, in maternal-effect mutants that alter embryonic anteroposterior polarity. *Development.* 109:425-433.

Heasman, J., J. Quarmby, and C. C. Wylie. 1984. The mitochondrial cloud of Xenopus oocytes: the source of germinal granule material. *Developmental Biology*. 105:458-469.

Huettner, A. F. 1923. The origin of the germ cells in *Drosophila melanogaster*. *Journal of Morphology*. 39:249-265.

Ikenishi, K. 1987. Functional gametes derived from explants of single blastomeres containing the "germ plasm" in *Xenopus laevis:* a genetic marker study. *Developmental Biology*. 122:35-38.

Illmensee, K., and A. P. Mahowald. 1974. Transplantation of posterior polar plasm in Drosophila. Induction of germ cells at the anterior pole of the Egg. *Proceedings of the National Academy of Science USA*. 71:1016-1020.

Illmensee, K., A. P. Mahowald, and M. R. Loomis. 1976. The ontogeny of germ plasm during oogenesis in *Drosophila*. *Developmental Biology*. 49:40-65.

Jaglarz, M. K., and K. R. Howard. 1994. Primordial germ cell migration in *Drosophila melanogaster* is controlled by somatic tissue. *Development*. 120:83-89.

Jongens, T. A., L. D. Ackerman, J. R. Swedlow, L. Y. Jan, and Y. N. Jan. 1994. *Germ cell-less* encodes a cell type-specific nuclear pore-associated protein and functions early in the germ-cell specification pathway of *Drosophila*. *Genes & Development*. 8:2123-2136.

Jongens, T. A., B. Hay, L. Y. Jan, and Y. N. Jan. 1992. The *germ cell-less* gene product: a posteriorly localized component necessary for germ cell development in Drosophila. *Cell*. 70:569-584.

Jürgens, G., and D. Weigel. 1988. Terminal versus segmental development in the *Drosophila* embryo: the role of the homeotic gene *fork head*. *Wilhelm Roux's Archives of Developmental Biology*. 197:345-354.

Karch, F., W. Bender, and A. B. Weiffenbach. 1990. abdA expression in Drosophila embryos. *Genes Dev*. 4:1573-1587.

Karr, T. L., and B. M. Alberts. 1986. Organization of the cytoskeleton in early Drosophila embryos. *Journal of Cell Biology.* 102:1494-1509.

Kerkis, J. 1931. The growth of the gonads in Drosophila melanogaster. *Genetics.* 16:212-244.

Keyes, L. N., T. W. Cline, and P. Schedl. 1992. The primary sex determination signal of Drosophila acts at the level of transcription. *Cell.* 68:933-943.

King, R. C. 1970. Ovarian Development in *Drosophila melanogaster.* Academic Press, New York.

Klingler, M., M. Erdelyi, J. Szabad, and C. Nüsslein-Volhard. 1988. Function of torso in determining the terminal anlagen of the Drosophila embryo. *Nature.* 335:275-277.

Kobayashi, S., R. Amikura, A. Nakamura, S. H., and M. Okada. 1995. Mislocalization of oskar product in the anterior pole results in ectopic localization of mitochondrial large ribosomal RNA in Drosophila embryos. *Developmental Biology.* 169:384-386.

Kobayashi, S., R. Amikura, and M. Okada. 1993. Presence of mitochondrial large ribosomal RNA outside mitochondria in germ plasm of *Drosophila melanogaster. Science.* 260:1521-1524.

Kobayashi, S., T. Kitamura, H. Sasaki, and M. Okada. 1993. Two types of pole cells are present in the *Drosophila* embryo, one with and one without splicing activity for the third P-element intron. *Development.* 117:885-893.

Kobayashi, S., and M. Okada. 1989. Restoration of pole-cell-forming ability to u.v.-irradiated Drosophila embryos by injection of mitochondrial lrRNA. *Development.* 107:733-742.

Krieg, C., T. Cole, U. Deppe, E. Schierenberg, D. Schmitt, B. Yodler, and G. von Ehrenstein. 1978. The cellular anatomy of embryos of the nematode Caenorhabditis elegans. *Devl Biol.* 65:193-215.

Laski, F. A., D. C. Rio, and G. M. Rubin. 1986. Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. *Cell.* 44:7-19.

Lehmann, R., and C. Nüsslein-Volhard. 1986. Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of *oskar*, a maternal gene in Drosophila. *Cell.* 47:141-152.

Lehmann, R., and C. Nüsslein-Volhard. 1991. The maternal gene *nanos* has a central role in posterior pattern formation of the *Drosophila* embryo. *Development.* 112:679-691.

Lewis, E. B. 1978. A gene complex controlling segmentation in Drosophila. *Nature.* 276:565-570.

Mahowald, A. P. 1962. Fine structure of pole cells and polar granules in Drosophila melanogaster. *J. Exp.Zool.* 151:201-205.

Mahowald, A. P. 1968. Polar granules of Drosophila. II. Ultrastructural changes during early embryogenesis. *J.Exp.Zool.* 167:237-262.

Mahowald, A. P. 1971. Fine structure of pole cells and polar granules in Drosophila melanogaster. *Journal of Experimental Zoology.* 176:329-344.

Mahowald, A. P. 1971. Polar Granules of Drosophila. III. The continuity of polar granules during the life cycle of Drosophila. *Journal of Experimental Zoology.* 176:329-344.

Mahowald, A. P., and G. Wei. 1994. Sex determination of germ cells in Drosophila. 193-209.

Manseau, L. J., and T. Schüpbach. 1989. cappuccino and spire: two unique maternal-effect loci required for both the anteroposterior and dorsoventral patterns of the Drosophila embryo. *Genes & Development.* 3:1437-1452.

McKnight, S. L., and O. L. Miller Jr. 1976. Ultrastructural patterns of RNA synthesis during early embryogenesis in Drosophila melanogaster. *Cell.* 8:305-319.

McKnight, S. L., and O. L. Miller Jr. 1979. Post replicative non-ribosomal transcription units in D. melanogaster embryos. *Cell.* 17:551-563.

Mével-Ninio, M., r. Terracol, C. Salles, A. Vincent, and F. Payre. 1995. ovo, a Drosophila gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning. *Mechanisms of Development.* 49:83-95.

Newport, J., and M. Kirschner. 1982. A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. *Cell.* 30:687-696.

Nöthinger, R., M. Jonglez, M. Leuthold, P. Meier-Gerschwiler, and T. Weber. 1989. Sex determination in the Germ Line of Drosophila depends on Genetic Signals and Inductive Somatic Factors. *Development.* 107:505-518.

Nüsslein-Volhard, C., H. G. Frohnhöfer, and R. Lehmann. 1987. Determination of anteroposterior polarity in Drosophila. *Science.* 238:1675-1681.

Okada, M., I. A. Kleinman, and H. A. Schneiderman. 1974. Restoration of fertility in sterilized *Drosophila* eggs by transplantation of polar cytoplasm. *Developmental Biology.* 37:43-54.

Oliver, B., Y.-J. Kim, and B. S. Baker. 1993. *Sex-lethal*, master and slave: a hierarchy of germ-line sex determination in *Drosophila. Development.* 119:897-908.

Oliver, B., N. Perrimon, and A. P. Mahowald. 1987. The *ovo* locus is required for sex-specific germ line maintenance in *Drosophila. Genes & Development.* 1:913-923.

Poirié, M., E. Niederer, and M. Steinmann-Zwicky. 1995. A sex-specific number of germ cells in embryonic gonads of Drosophila. *Development.* 121:1867-1873.

Poulson, D. F. 1950. Histogenesis, organogenesis, and differentiation in the embryo of Drosophila melanogaster (Miegen)., p. 168-274. *In* Biology of Drosophila. M. Demerec, M. Demerecs. Wiley, New York. 168-274.

Raff, J. W., and D. M. Glover. 1989. Centrosomes, and not nuclei, initiate pole cell formation in Drosophila embryos. *Cell.* 57:611-619.

Rio, D. C., F. A. Laski, and G. M. Rubin. 1986. Identification and immunolochemical analysis of biologically active Drosophila P element transposase. *Cell.* 44:21-32.

Rongo, C., E. R. Gavis, and R. Lehmann. 1995. Localization of oskar RNA regulates oskar translation and requires oskar protein. *Development.* in press:

Schüpbach, T. 1982. Autosomal mutations that interfere with sex determination in somatic cells of *Drosophila* have no direct effect on the germline. *Developmental Biology.* 89:117-127.

Schüpbach, T. 1985. Normal female germ cell differentiation requires the female X chromosome to autosome ratio and expression of sex-lethal in *Drosophlia melanogaster. Genetics.* 109:529-548.

Schüpbach, T., and E. Wieschaus. 1986a. Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo. *Wilhelm Roux's Archives of Developmental Biology.* 195:302-317.

Schüpbach, T., and E. Wieschaus. 1989. Female sterile mutations on the second chromosome of *Drosophila melanogaster.* I. Maternal effect mutations. *Genetics.* 121:101-117.

Schüpbach, T., and E. Wieschaus. 1991. Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics.* 129:1119-1136.

Seydoux, G., and A. Fire. 1994. Soma-germline asymmetry in the distributions of embronic RNAs in *Caenhorhabditis elegans. Development.* 120:2823-2834.

Siegel, V., T. A. Jongens, L. Y. Jan, and Y. N. Jan. 1993. *pipsqueak*, an early acting member of the posterior group of genes, affects *vasa* level and germ cell-somatic cell interaction in the developing egg chamber. *Development.* 119:1187-1202.

Simon, J., M. Peifer, W. Bender, and M. O'Connor. 1990. Regulatory elements of the bithorax complex that control expression along the anterior-posterior axis. *EMBO J.* 9:3945-3956.

Sonnenblick, B. P. 1941. Germ cell movements and sex differentiation of the gonads in the Drosophila embryo. *Proc.Natn.Acad.Sci.USA.* 26:373-381.

Sonnenblick, B. P. 1950. The early embryology of Drosophila melanogaster., p. 62-167. *In* The early embryology of Drosophila melanogaster. M. Demerec, M. Demerecs. Wiley, New York. 62-167.

St. Johnston, D., Beuchle, D., and Nüsslein-Volhard C. 1991. *Staufen*, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell.* 66:51-63.

Steinmann-Zwicky, M. 1994. Sex determination of the Drosophila germ line: tra and dsx control somatic inductive signals. *Development.* 120:707-716.

Steinmann-Zwicky, M., H. Schmid, and R. Nöthinger. 1989. Cell-autonomous and inductive signals can determine the sex of the germ line of Drosophila by regulating the gene *Sxl*. *Cell.* 57:157-166.

Strome, S. 1993. Determination of cleavage planes. *Cell.* 72:3-6.

Strome, S., and W. B. Wood. 1983. Generation of asymmetry and segregation of germ-line granules in early *C. elegans* embryos. *Cell.* 35:15-25.

Struhl, G., and R. A. H. White. 1985. Regulation of the *ultrabithorax* gene of Drosophila by other bithorax complex genes. *Cell.* 43:507-519.

Szabad, J., and R. Nöthinger. 1992. Gynandromorphs of *Drosophila* suggest one common primordium for the somatic cells of the female and male gonads in the region of abdominal segments 4 and 5. *Development.* 115:527-535.

Szabad, J., T. Schüpbach, and E. Wieschaus. 1979. Cell lineage and development in the larval epidermis of Drosophila melanogaster. *Developmental Biology.* 73:256-271.

Technau, G. M. 1986. Lineage analysis of transplanted individual cells in embryos of Drosophila melanogaster. I. The method. *Wilhelm Roux's Archives of Developmental Biology*. 195:389-398.

Technau, G. M., and J. A. Campos-Ortega. 1986. Lineage analysis of transplanted individual cells in embryos of Drosophila melanogaster. II. Commitment and proliferative capabilities of neural and epidermal cell progenitors. *Roux's Arch. Dev. Biol.* 195:445-454.

Technau, G. M., and J. A. Campos-Ortega. 1986. Lineage analysis of transplanted individual cells in embryos of Drosophila melanogaster. III. Commitment and proliferative capabilities of neural and epidermal cell progenitors. *Roux's Arch. Dev. Biol.* 195:489-498.

Tremml, G., and M. Bienz. 1989. Homeotic gene expression in the visceral mesoderm of Drosophila embryos. *EMBO Journal.* 8:2677-2685.

Turner, F. R., and A. P. Mahowald. 1976. Scanning electron microscopy of Drosophila embryogenesis. I. The structure of the egg envelopes and the formation of the cellular blastoderm. *Developmental Biology*. 50:95-108.

Underwood, E. M., J. H. Caulton, and C. D. M. Allis A.P. 1980. Developmental fate of pole cells in Drosophila melanogaster. *Developmental Biology*. 77:303-314.

Warrior, R. 1994. Primordial germ cell migration and the assembly of the Drosophila embryonic gonad. *Developmental Biology*. 166:180-194.

Wei, G., B. Oliver, and A. P. Mahowald. 1991. Gonadal dysgenesis reveals sexual dimorphism in the embryonic germline of Drosophila. *Genetics*. 129:203-210.

Weigel, D., and H. Jäckle. 1990. The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell*. 63:455-456.

Wieschaus, E., C. Audit, and M. Masson. 1981. A clonal analysis of the roles of somatic cells and germ line duringoogenesis in *Drosophila. Developmental Biology*. 88:92-103.

Wieschaus, E., and W. Gehring. 1976. Clonal analysis of primordial disc cells in the early embryo of *Drosophila melanogaster. Developmental Biology*. 50:249-263.

Wylie, C. C., J. Heasman, A. Snape, A. O'Driscoll, and S. Holwill. 1985. Primordial germ cells of *Xenopus laevis* are not irreversibly determined early in development. *Developmental Biology*. 112:66-72.

Yasuda, F. K., and G. Schubiger. 1992. Temporal regulation in the early embryo: is MBT too good to be true? *Trends Genet*. 8:124-127.

Zalokar, M. 1976. Autoradiographic Study of Protein and RNA Formation during Early Development of Drosophila Eggs. *Developmental Biology*. 49:425-437.

Zalokar, M., and I. Erk. 1976. Division and Migration of Nuclei during Early Embryogenesis of *Drosophila melanogaster*. *J Microsc. Biol Cell*. 25:97-106.

Specific Aims:

The work described in this thesis is directed toward understanding the process of embryonic germ cell differentiation in Drosophila. Chapter two describes the subtractive hybridization screen designed to isolate genes expressed specifically in the germ cells of developing embryos. The technique is described in detail and data reflecting the progress and effectiveness of the subtractive hybridization is presented. Preliminary characterization of cDNA clones in the subtracted cDNA library is also presented. Chapter three describes the full characterization of two germ-cell-specific cDNAs isolated from the subtracted library. This chapter includes analysis of the RNA expression patterns for these two genes, complete sequence of full length cDNAs, and analysis of their protein products *in vivo* and *in vitro*. In addition the results of genetic localization of these genes is presented, with preliminary analysis of the possible function of these genes *in vivo*. The epilogue contains a discussion of the questions remaining and the next logical phase of experiments to pursue in order to fully understand the role of these genes in germline differentiation. Appendix I describes the analysis of a non-germline specific cDNA clone isolated from the subtracted library. RNA expression, sequencing and identification of the genomic locus encoding this cDNA are described. These results are discussed in light of the homology to known proteins. Appendix II contains the aligned sequence data generated during the random-sonnication sequencing method used for all three cDNAs.

Chapter 2:

Introduction

In the absence of genetic information about the requirements for determination and differentiation of a specific cell or tissue type, one effective way of identifying genes required for these processes is to isolate RNAs uniquely expressed in these cells or tissues. Molecular screening techniques are unbiased by choice of mutagenesis method or prediction of phenotype and result in cDNAs which can be easily cloned, sequenced, expressed and analyzed by *in situ* hybridization and RNA blot or RNAse protection. The usefulness of the molecular screening approach has been demonstrated in many different circumstances by the identification of loci not previously detected in classical genetic screens (Alt et al., 1978; Ding and Lipshitz, 1993; Hedrick et al., 1984; Palazzolo et al., 1989). With judicious choice of starting materials, rigorous criteria for screening clones, and some knowledge of the characteristics expected of the target RNAs, molecular screens can be used very successfully to isolate genes required in many different biological processes.

The first experiments to pave the way for molecular isolation of sequences encoded by specific genes were published in 1961 (Hall and Spiegelman, 1961). These authors demonstrated that viral-encoded RNA could pair with viral DNA. Subsequently, development of techniques for immobilizing single stranded DNA made it feasible to measure the extent of annealing of radioactively labeled single stranded DNA or RNA with specific pools of target DNA (Bautz and Hall, 1962; Bolton and McCarthy, 1962; Gillespie and Spiegelman, 1965; Hall and Spiegelman, 1961; Nygaard and Hall, 1964). The advent of these technologies opened a new field of inquiry into the properties of specific populations of DNA and RNA molecules. One of the earliest applications of this technology was the measurement of the half-lives of different abundance populations of RNA isolated from Xenopus embryos (Brown and Gurdon, 1966).

Another major line of inquiry made possible by the discovery of techniques to study DNA-DNA hybridization was the estimation of the total amount of non-redundant sequence (or "Complexity") in the genomes of various organisms by measuring the kinetics of genomic DNA renaturation (Britten and Kohne, 1968). Reassociation kinetics of genomic DNA from various different organisms is one of the earliest methods used not only to characterize genome sizes but also the amount of repetitive sequences they contain. This

technique was further adapted to study hybridization of single-copy genomic DNA to excess mRNA (by "Saturation hybridization"). The percentage of unique DNA sequences that form double stranded hybrids with mRNA is used to calculate the amount of unique mRNA sequence in the population, and thereby obtain minimum estimates of the number of mRNAs in a given cell-type (Bantle and Hahn, 1976; Galau et al., 1974).

Kinetic analysis of mRNA:cDNA hybridization is another method of determining the overall number of messenger RNA's in different cell types. The advantage of hybridization kinetics over saturation hybridization is that this technique can also be used to determine the relative abundance of various mRNA populations in the cell, e.g. what percentage of mRNAs are very abundant and what percentage are very rare? (Anderson et al., 1976; Axel et al., 1976; Bishop et al., 1974; Hereford and Rosbash, 1977; Lewin, 1974; Perlman and Rosbash, 1978). The combination of these two methods has been applied to the study of mRNA complexity in a wide variety of organisms. Such experiments reveal that a typical mammalian cell contains a total of 2-5x 10^5 mRNA molecules, comprising approximately 10-20,000 unique mRNA molecules (Lewin, 1974). If all of these mRNAs were present in equal abundance, each one would be present at from 4-50 copies per cell, which is 0.002-0.01% of the total mRNA. However, cellular mRNAs range from 1 copy per cell (~0.0002%) to at least 5,000 copies per cell (or ~1% of the total). In fact, 10% of the mRNA molecules in a typical eukaryotic cell belong to the "very abundant" class, present at the level of 5,000 copies per cell (Soares et al., 1994). The rarest RNAs, each represented by only 1-15 copies per cell, make up 40-45 % of the total, and the rest of the transcripts typically range in abundance between these two extremes, depending upon the cell type (Lewin, 1974).

Comparisons of the number of unique mRNAs expressed in sea urchins at various developmental stages provided some of the first evidence that distinct sets of structural genes are active during different stages of oocyte, embryonic and adult differentiation (Galau et al., 1976; Hough-Evans et al., 1977). These early studies provided a wealth of information about the number and types of sequences expressed during embryonic development of this important model system and led to the adoption of hybridization techniques for identification of differentially regulated RNAs in many different systems.

Depending on the degree of differentiation of a given cell, mRNAs unique to that cell may encode products with a variety of different specific functions. Early in the differentiation process one might expect to find that key regulatory molecules are specifically synthesized

in response to exogenous or endogenous cues. The abundance of such molecules is likely to be relatively low, especially if these factors play a role in regulating expression of downstream differentiation factors. Later in the differentiation process, one might expect to find more abundant mRNAs encoding structural proteins specific to the function of the terminally differentiated cell. Therefore, depending upon where they fall in the regulatory hierarchy the abundance of differentially regulated genes will range from very rare to very abundant mRNAs.

One of the first strategies used to look for differentially expressed genes is "Plus-minus" or "Differential" screening. In this method a cDNA library is hybridized with two different cDNA probes derived from, for example, two cell types, only one of which expresses the target gene(s) of interest (Sambrook et al., 1989). However, the utility of this technique is limited by the high complexity of cDNA probes, making detection of rare sequences difficult and restricting the usefulness of this approach to the identification of relatively abundant RNA transcripts, which comprise 0.05-0.1% of an mRNA pool. The complexity of the probes and the cDNA library can be decreased by subtractive hybridization. This technique enables detection of less abundant messages by eliminating common sequences from the cDNA pools, which increases the concentration of unique sequences in both pools, and thereby increasing the signal to noise ratio.

The earliest use of subtractive hybridization took place before the advent of DNA cloning techniques. Messenger RNA molecules encoded by the rIIA and rIIB cistrons of bacteriophage T4 were isolated by hybridization of labeled wild-type T4 RNA to a rIIA-rIIB deletion-mutant strain. Only those molecules encoded by deleted sequences remained single stranded after hybridization to the deletion strain (Bautz and Reilly, 1966). Subsequently, subtractive hybridization has been used successfully to isolate genes based on their differential expression in a number of different biological contexts (Fornace and Mitchell, 1986; Hedrick et al., 1984; Sargent and Dawid, 1983; Timberlake, 1980). Starting with two mRNA populations that differ by only a small fraction of their expressed RNAs, eliminating the vast quantity of common sequences from the starting cDNA population can cause target species to be enriched from 50 to 1000-fold. Therefore, even transcripts present at only a few copies per cell, corresponding to as few as 1 in 10^5 mRNA molecules, can be isolated (Sargent and Dawid, 1983).

Conventional methods of subtractive hybridization involve hybridization of tens of micrograms of polyA$^+$ driver mRNA with at least 10-fold less cDNA target. The

RNA:cDNA hybrids are isolated by hydroxyapatite chromatography and the resultant single stranded material is either cloned to make a subtracted library or labeled to make a subtracted probe. Because unique target sequences are usually relatively rare, elimination of common sequences often results in elimination of greater than 95% of the original cDNA. Therefore, not only does the method require large amounts of driver mRNA, in addition, typically only a few nanograms of target sequences remain after subtraction, making cloning and isolation of rare species in the subtracted library difficult.

The subtractive hybridization technique has been adapted and modified in a myriad of different ways to allow isolation of important DNA and RNA sequences, including key reagents for the identification of disease loci in the human genome. For example, isolation of DNA sequences linked to the Duchenne muscular dystrophy gene was accomplished by using a method called "phenol emulsion reassociation technique" (or PERT,(Kohne et al., 1977)). This method increases the rehybridization rate of complementary DNA sequences enough to allow isolation of rare DNA fragments corresponding to deleted sequences between normal and deficiency-carrying human chromosomes. The increased speed of hybridization compensates for the high complexity of the two DNA pools that normally makes it impossible to achieve hybridization of very rare sequences (Kohne et al., 1977; Kunkel et al., 1985).

PERT has been combined with a number of other modifications to the original subtractive hybridization and cloning procedures. For example, Zeng et al. combine this technique with specific enzymatic degradation following hybridization of driver cDNA with thio-nucleotide-modified tracer cDNA. Their technique, called "enzymatic degrading subtraction" or EDS uses hybridization between tracer cDNA synthesized with thionucleotides and unmodified driver cDNA, followed by enzymatic degradation by exonucleases. Only molecules in the modified tracer pool are resistant to degradation, therefore all driver and driver:tracer hybrids are destroyed, leaving only the annealed tracer duplexes intact (Zeng et al., 1994).

Differential display is another method used to isolate specific gene products based on their expression in one cell or tissue type but not another (Liang et al., 1993; Liang and Pardee, 1992). In this method representative cDNA pools are amplified by the polymerase chain reaction (PCR) from templates generated by reverse transcription of the two mRNA populations of interest. The PCR primers are designed to generate an array of amplified products (reflecting the starting mRNA pools) that can be visualized by separation of

radioactively labeled PCR fragments on polyacrylamide gels. Amplified fragments present in one pool but not the other can thereby be identified and isolated directly from the dried gel matrix (Liang et al., 1993; Liang and Pardee, 1992).

The subtractive hybridization screen described in this work is based on the method developed by Wang and Brown (Wang and Brown, 1991), and combines the high sensitivity of using a subtracted probe to screen a subtracted library with the use of PCR amplification of the subtracted pool after every round (see **figure 3**). Unlike traditional subtractive hybridizations, in which the population of molecules in the subtracted pool becomes vanishingly small after a limited number of rounds, this PCR-based method allows regeneration of the enriched material after every subtractive hybridization. The quantity of material produced by PCR at each stage is sufficient to allow multiple rounds of subtraction to be performed until the population has been depleted of known common sequences and enriched for any known differential cDNAs. In addition, regeneration of workable amounts of material after every round of hybridization, allows both cloning of subtracted cDNA sequences to make subtracted libraries as well as ample template for probe synthesis and further positive or negative selection of specific sequences in the subtracted population(Wang and Brown, 1991). Further successful modifications of this and other protocols are continuously being developed. For example, Hakvoort et al. (Hakvoort et al., 1993) have successfully combined this method with autoradiographic display of PCR amplified fragments adapted from Liang and Pardee (Liang et al., 1993) to allow direct visualization of different cDNA fragments during the course of the subtraction (Hakvoort et al., 1993).

One of the most dramatic examples of the successful use of subtractive hybridization is the cloning of the *MyoD* gene (Davis et al., 1987). These authors describe a subtractive hybridization between RNA from undifferentiated mouse embryonic fibroblast cells and cDNA prepared from proliferating myoblasts of two different myogenic cell lines. This screen resulted in the isolation of a single myoblast-specific gene, named "*MyoD,*" whose expression converts undifferentiated embryonic fibroblast cells into differentiated myoblasts (Davis et al., 1987). The *MyoD* gene was isolated at a frequency of 4×10^{-4} clones, making it 0.04% of the cDNAs in the library screened and a rare species in the mRNA population.

The key factor in the success of the *MyoD* screen was the stringent criteria used in screening cDNAs of the subtracted pool for qualities expected of the target gene(s) of

interest. The design of the screen drew upon extensive immunological and biochemical studies which provided evidence for the expression of a repertoire of lineage-specific markers in proliferating myoblasts. These studies suggested that proliferating myoblasts continually express mRNAs encoding regulatory factors which activate myoblast-specific markers and contribute to the differentiation of these muscle precursor cells by making them competent to express other muscle-specific genes (Davis et al., 1987).

The effectiveness of any subtractive hybridization or related technique depends upon the methods devised for screening the resultant clones for those that are relevant to the process being studied. In many developmental systems, information about the kinds of cells or tissues expected to express the genes of interest, as well as lineage-specific markers, are extremely limited or non-existent. Such is the case for embryonic germ cell differentiation, so in the screen described here I focus on identification of RNAs expressed specifically in germ cells without initial regard to their actual role in determination and/or differentiation.

The rationale for the subtractive hybridization was that any mRNA uniquely expressed and/or maintained in the embryonic germ cells would be present in wild-type embryos but not in embryos that never form pole cells. On the other hand, all transcripts expressed in somatic tissues should be present in both types of embryos and would therefore be depleted during the course of the subtractive hybridization, leaving only those RNAs unique to or highly enriched in the embryonic germ line cells.

Results:

Generation of Template cDNAs for subtractive hybridization

Embryos laid by *oskar*^{301}/*oskar*^{CE4} mutant females at 18°C are completely wild-type except that they do not form pole cells and therefore lack germline tissues completely (Lehmann and Nüsslein-Volhard, 1986). These embryos were used to isolate germ-cell-minus RNA for subtraction (see **figure 1**). However, given that there are no genes known to be expressed in wild-type but not in embryos of *oskar*$^{301/CE4}$ mutant mothers, it was critical to include an exogenous RNA as a positive control, present as a transgene in the wild-type embryos but not in those lacking pole cells. A transgenic strain carrying a P-element-encoded *lacZ* transcript in which the bacterial *LacZ* marker gene was cloned upstream of the *nanos* 3' untranslated region (UTR) sequences was therefore used as the source of wild-type RNA (see **figure 1**). The *nanos* RNA is localized to the posterior pole in early

embryos through its 3'UTR sequences, taken up into the pole cells when they form at the posterior pole, and stably maintained in them throughout embryonic development. The mRNA produced from the *lacZ/nos* transgene has been shown to behave just like the endogenous *nanos* RNA. This fusion transcript is maintained in the embryonic germ cells throughout embryogenesis and can be detected in the embryonic gonads as late as stage 16, just before cuticle deposition (E. Gavis, personal communication).

Subtraction Starting Materials:

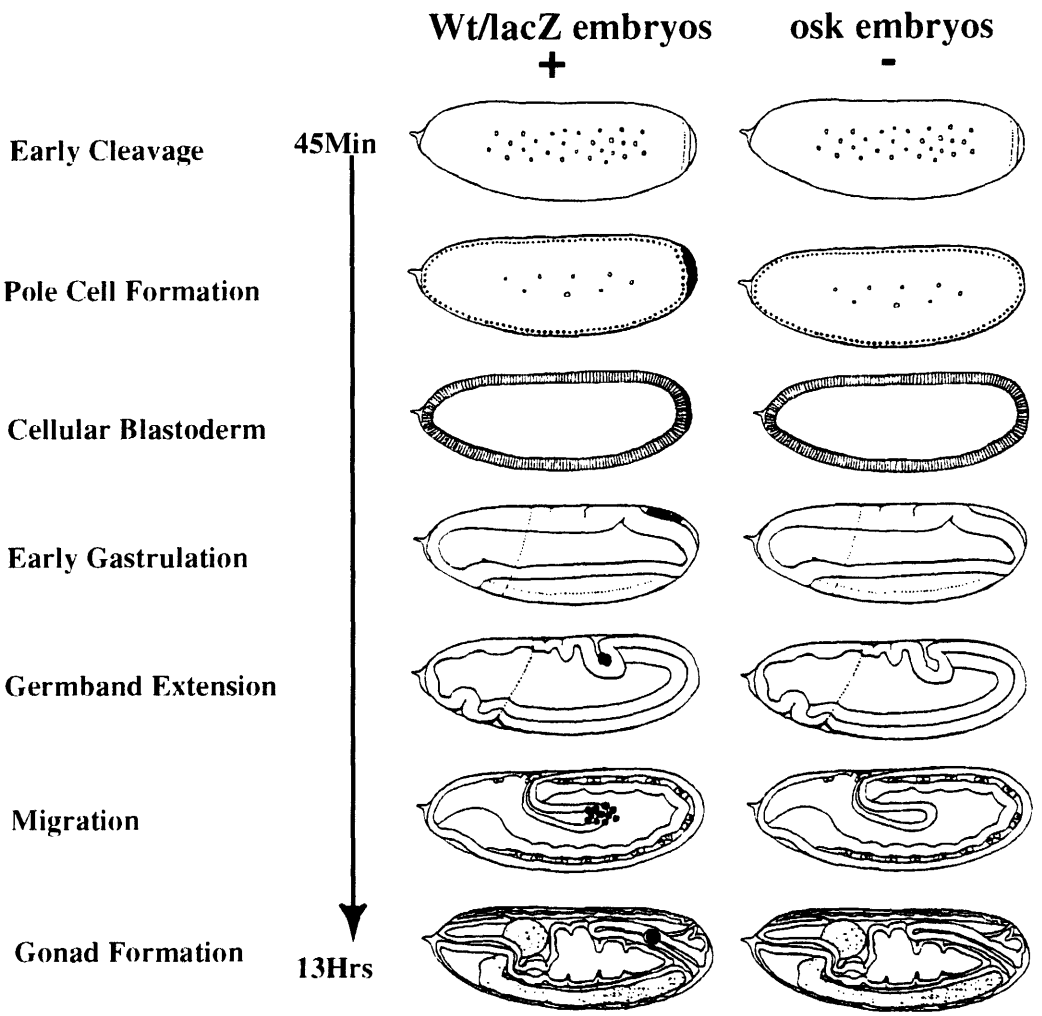


Figure 1 This figure shows the stages of development included in collections for isolation of RNA from wild-type/LacZ embryos and those laid by *oskar*-mutant females. On the far left are descriptions of the developmental events occuring during the embryonic stages pictured to the right. The left-hand column depicts wild-type/ LacZ embryos. These transgenic embryos carry a P-element transposon that encodes a *lacZ-nanos 3'UTR* fusion RNA which is incorporated into the pole cells when they form, and remains stabilized in these cells throughout embryonic development (pole cells are shaded in black). The column on the right depicts embryos from the same developmental stages as the wild-type/LacZ embryos. These embryos develop identically to wild-type, except that they lack pole cells.

The *lacZ/nos* mRNA serves not only as a control for cross-contamination between the two starting RNA pools and subsequent cDNA pools, but also as a means of monitoring the sensitivity of the screen. The abundance of the *LacZ/nos* RNA was expected to reflect the extent to which a transcript present only in the germ cells of the embryo would be detectable in the background of all of the transcripts present in the somatic cells throughout embryogenesis. At cellular blastoderm the pole cells number from 20-50 cells in contrast to the roughly 5000 somatic cells present (Campos-Ortega and Hartenstein, 1985; Sonnenblick, 1941) . This ratio shrinks progressively throughout embryogenesis until by late embryonic gonad formation, the primordial germ cells number less than 20 in comparison to the ~24,000 somatic cells present at this stage.

Zygotic transcription in the pole cells is undetectable until approximately 3.5 hours after fertilization (Zalokar, 1976) . In addition, only two genes, *vasa* and *ovo* (see Chapter 1), are known to be transcribed zygotically in the late embryonic germ cells. Therefore, in light of the limited evidence for zygotic transcription in pole cells, embryos from a broad range of ages were collected. In order to include embryos that contain significant levels of maternal transcripts enriched and/or stabilized in pole cells, as well as late stage embryos containing substantial amounts of zygotically transcribed RNAs, embryos ranging in age from 45 minutes to 13 hours after egg laying were collected (see **figure 1**). The embryos in these collections include all stages of embryonic pole cell development: pole cell formation, gastrulation, migration through the midgut epithelium, and finally, gonad formation.

Roughly 600 micrograms of embryos of each type, was used as starting material for the isolation of ~30 micrograms of polyA$^+$ mRNA. An aliquot of the mRNA isolated from these embryos was analyzed by Northern blot hybridization. Labeled cDNAs encoding *lacZ* (**figure 2A**), *hunchback* (**figure 2B**), and *vasa* (**figure 2C**) were used as probes to check for the expected distribution of these three transcripts between the two starting mRNA populations and to assess the quality and relative quantities of RNA in the two pools. *LacZ* transcripts were detected only in the wild-type mRNA pool, whereas the *hunchback* and *vasa* transcripts were present in roughly equivalent amounts in both starting pools, although there appears to be a very small excess of *vasa* RNA in the "-" lane of **figure 2C**.

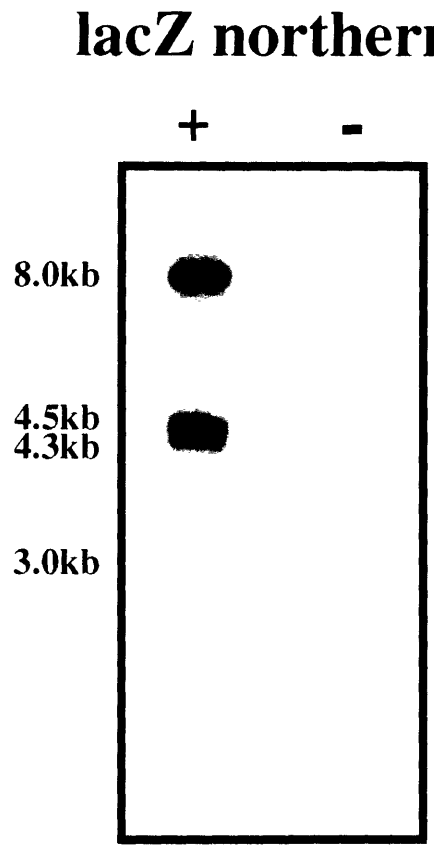


Figure 2A: RNA blots of starting mRNA pools: 10µg aliquots of mRNA from the original pools used to generate "+" and "-" cDNA pools for subtractive hybridization were analyzed by RNA blot hybridizations using probes for the *lacZ* marker and two endogenous genes. This figure shows the result of probing the blot with a *lacZ* probe, revealing that transcripts from the *LacZ-nos3'UTR* transgene are present only in the "+" pool, as expected. The fact that this transgene appears to give rise to four different transcripts is presumably due to incomplete termination of properly initiated transcripts as well as, or in addition to, products of internal transcription initiation. However, the precise origin of these different transcripts has not been determined.

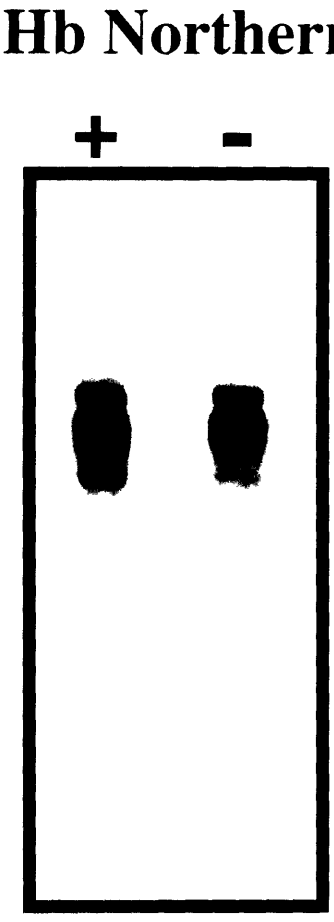


Figure 2B. The same RNA blot shown in **2A** was stripped and re-probed with a cDNA derived from the *hunchback* gene. Although the levels of the *hunchback* messages are roughly equivalent, there appear, to be a slight excess of the smallest transcript in the wild-type or "+" lane relative to the "-" lane.

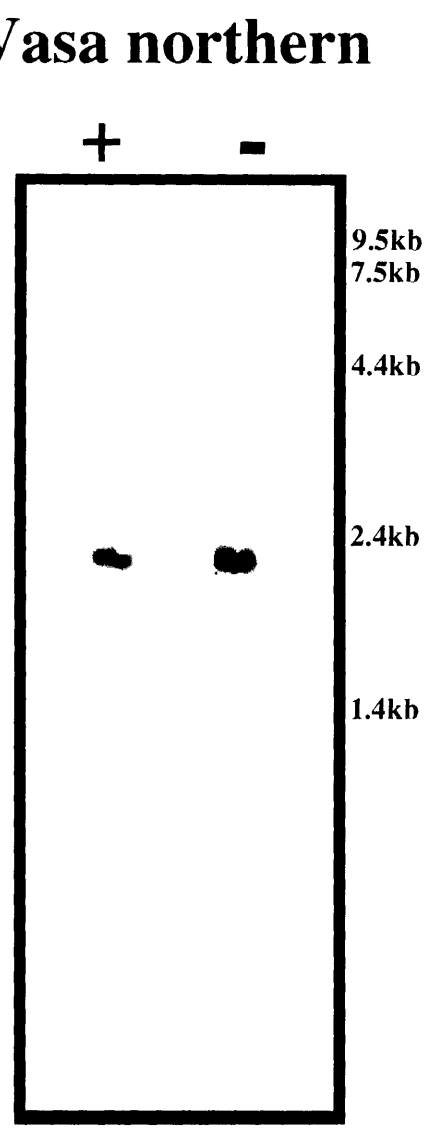


Figure 2C. The same RNA blot shown in A and B was stripped and re-probed with a cDNA probe derived from the *vasa* gene. As in the case of *hunchback*, this gene is present in roughly the same amounts in "+" and "-" pools, but in contrast to the *hunchback* messages, there appears to be a very slight excess of the vasa transcript in the "-" pool.

Five micrograms of mRNA was used for cDNA synthesis from the two RNA populations. Oligo dT was chosen as the primer for first strand synthesis as a further means of selecting polyA$^+$ RNAs as the substrate for library construction, and the second strand was synthesized using RNAseH/polymerase I (see materials and methods). Trace amounts of α^{32}PdCTP were added to the first strand synthesis to allow monitoring of the size of the cDNAs and to facilitate analysis of yields during subsequent digestion and cloning steps. Based on alkaline agarose electrophoresis followed by autoradiography, the two cDNA populations appear to extend to the length of greater than 8kb (data not shown). This shows that the RNA used for cDNA synthesis was not significantly degraded and suggests that the majority of transcripts therefore would be represented in the cDNA population.

Preparation of cDNA pools:

In any PCR-based approach, there is a likelihood that PCR bias in favor of smaller or more easily amplified sequences (i.e. those with little or no tendency to form problematic secondary structures) will be established and enhanced during repeated rounds of PCR amplification. To decrease the potential PCR bias, and to ensure that the PCR-amplified starting material represented the original full length starting library as much as possible, the two cDNA pools were sheared semi-randomly by digestion with two different restriction enzymes that have four-base recognition sites. By combining the two pools of digested cDNA, a semi-random distribution of fragments in the size range of 150 to 800 bp was generated (Wang and Brown, 1991). Adapters were ligated onto the ends of the fragments to provide specific sites for reamplification of the pool. Adapter-ligated cDNA populations were subsequently size selected by agarose gel-electorphoresis to include those cDNAs in the range from 150-500bp. Restriction of cDNAs to this small size range helps to increase the efficiency with which the entire population can be amplified to give high yields of cDNA pools for use during each round of subtraction.

Subtractive hybridization:

The subtractive hybridization protocol is diagrammed in **figure 3A-C**. Once the cDNAs for each pool have been isolated, and linkers ligated to both ends for PCR amplification, both pools are amplified by PCR to generate the 100-200 μg necessary to begin the subtractive hybridization (**figure 3A**). Both wild-type and mutant cDNA pools are treated identically throughout, and the subtraction is carried out in two parallel reactions. In one case, the sequences unique to the wild-type pool, but not present in the mutant pool are

selected. Simultaneously, cDNAs present in the mutant pool, but not in the wild-type pool are selected. The "Driver" in each hybridization consists of cDNA that has been modified by cutting off the ends of each fragment to prevent re-amplification of this material. In addition, the "Driver" cDNA is biotinlyated to allow hybrids formed during the annealing reaction to be removed after addition of streptavidin followed by phenol extraction.

Each hybridization reaction involves mixing the "Tracer", a small amount of unmodified cDNA from one pool, with twenty-fold excess of biotinylated "Driver" derived from the opposite pool. These populations are allowed to anneal for either two or twenty hours ("Short" or "Long" hybridizations). The subtraction technique consisted of alternating long and short hybridizations. The "Driver" cDNA is present in vast excess during the hybridization so that any sequences common to both pools will be likely to form hybrids with biotinylated molecules and therefore be removed by extraction in the presence of streptavidin. Any excess, unhybridized driver will also be removed at this stage. The long hybridizations are designed to allow rare species present in both populations to hybridize, while the short hybridizations favor the annealing of molecules that are abundant in both pools. The subtraction protocol is designed so that the cDNAs in tracer pools that remain single-stranded after short hybridization against the appropriate driver pool are then reamplified by PCR and subjected to a long hybridization against driver derived from the opposite cDNA pool. These long and short hybridizations are done in succession (see **figure 3C**) until the cDNAs in both pools no longer contain significant amounts of common cDNAs.

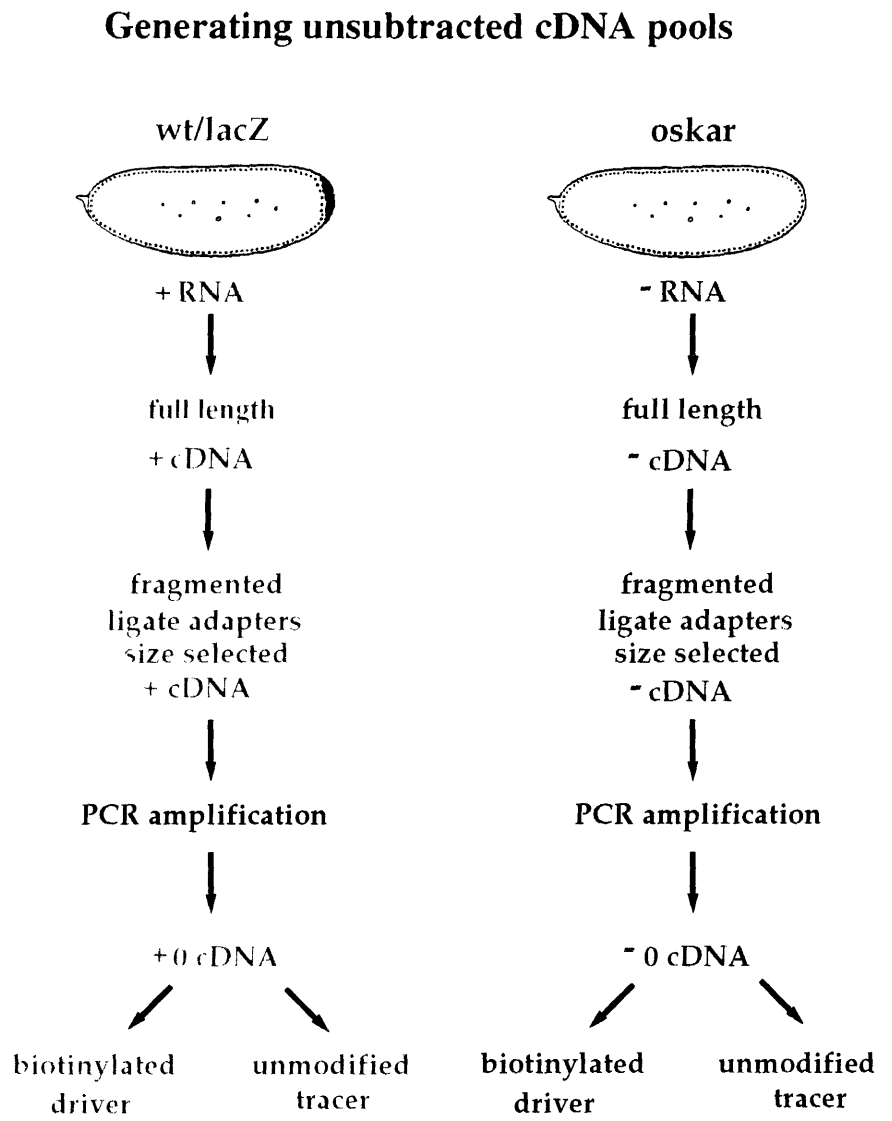


Figure 3A. Generating unsubtracted cDNA pools: This figure diagrams the isolation of RNA from wild-type/lacZ-transformant flies and embryos laid by *oskar*-mutant mothers. The RNA is polyA+ selected, transformed into full length cDNA by olgo dT-primed first strand synthesis followed by RNAseH/polymerase I-second strand synthesis. The resultant full length cDNA is then fractionated by digestion with four-cutter restriction enzymes, ligated with adapters, and size selected to generate fragments in the range of 150 to 500 base pairs. The pools were then amplified by polymerase chain reaction using primers specific for sequences in the adapters and purified. Aliquots of both pools are subsequently digested with Eco RI to remove sequences homologous to the PCR primers, photobiotinylated, and purified (see materials and methods) for use as driver in reciprocal subtractive hybridizations to produce the + and - 1cDNA pools (see **figure 1B**).

Generating +/-1 cDNA pools

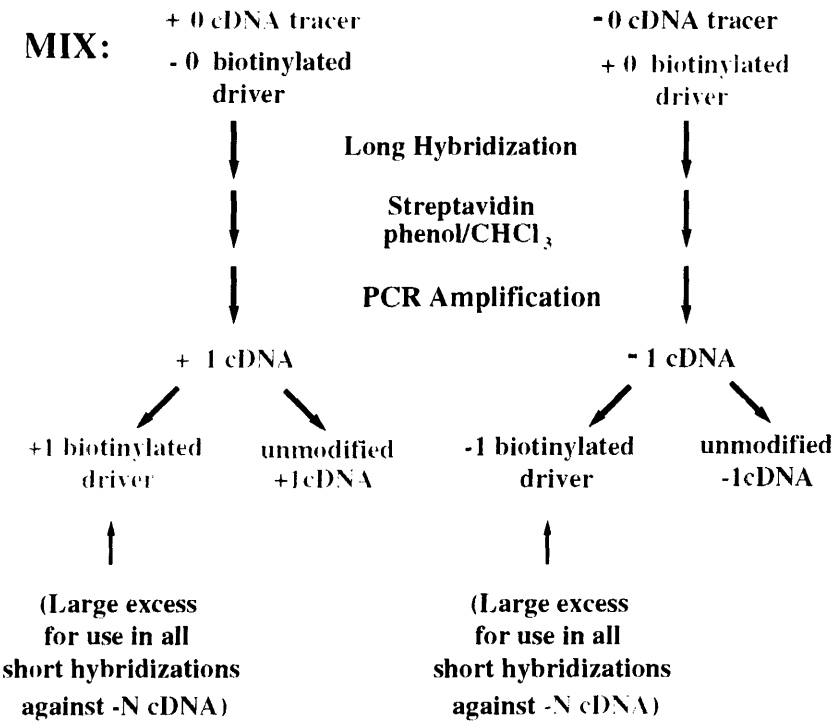
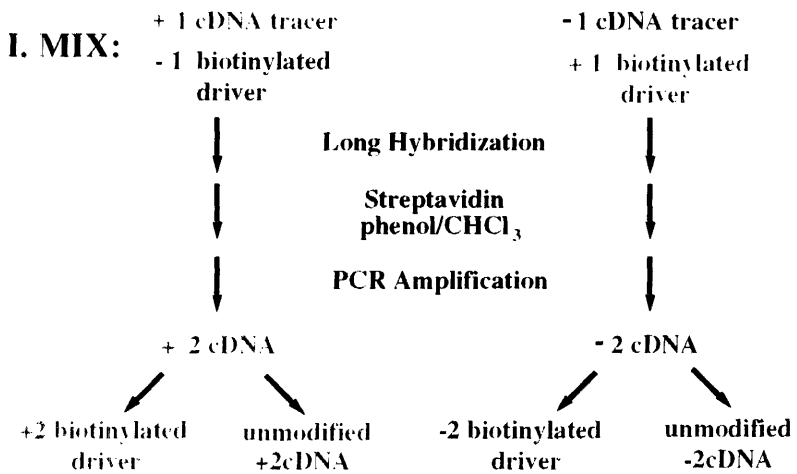


Figure 3B. Generating +/-1 cDNA pools: Trace amounts of + or -0cDNAs were mixed with excess - or +0 biotinylated cDNA respectively. The two sets of cDNAs were allowed to hybridize for 20 hours, followed by incubation with streptavidin and subsequent phenol extraction to remove all biotinylated cDNAs. Single stranded materials remaining in the aqueous phase are presumed to be depleted of rare, common sequences with respect to the unsubtracted starting materials. The unsubtracted tracer pools are then re-amplified to generate the + and -1cDNA pools. As in **figure 1A**, an aliquot of each pool is set aside to use as tracer and the rest is photo-biotinylated to produce the +/-1 biotinylated driver pools for use in the next round of long hybridization and in all subsequent rounds of short hybridization.

Long Hybridizations



Short Hybridizations

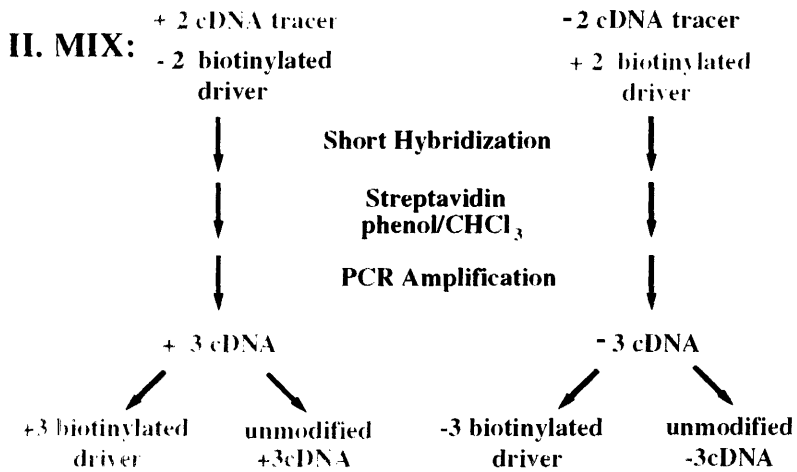


Figure 3C. Long and short hybridizations: For long hybridizations, +1cDNA tracer is mixed with 20-fold excess -1 biotinylated cDNA driver and hybridized for 20 hours (as described in figure 1B legend). Resultant + and -2cDNAs are purified, and prepared for the next round of subtraction as described in figure legend 1B. Short hybridizations: tracer pools derived from long hybridizations are mixed with excess +/-1 biotinylated driver (synthesis of which is described in figure 1B) and allowed to hybridize for two hours. Following short hybridizations, +/-3cDNA is purified and made ready for the next round of long hybridization as described above. Long and short hybridizations are done in alternation until the two subtracted pool are at least 95% different as judged by cross-hybridization.

Long Hybridizations:

The long hybridizations were set up using a 20-fold excess of Eco RI-digested, biotinylated cDNA from embryos lacking germ cells (designated "-" cDNA) with trace amounts of cDNA from wild-type embryos (designated "+" cDNA). In parallel, trace amounts of "-" cDNA were mixed with a 20-fold excess of Eco RI-digested, biotinylated "+" cDNA (**figure 3B** and **C**). The key to the success of this technique it that two symmetrical subtractive hybridizations are done in parallel to allow simultaneous generation of "+" and "-" pools which, aside from the differences they may exhibit in the presence or absence of certain unique sequences, resemble each other closely in the extent to which abundant, common cDNAs have been depleted during the course of the subtraction process.

Doing subtractions in parallel and using subtracted pools as drivers results in decreased complexity of the driver pools which increases the efficiency with which sequences present in driver and tracer pools will hybridize and therefore be depleted. If the complexity of the driver pool is too high, then there is a certain subset of rare, common species that will remain too dilute in the driver pool to effectively deplete their corresponding sequences in the tracer pool (Lisitsyn et al., 1993). As the subtraction of abundant cDNAs from both pools is achieved, cDNAs that were originally rare in both populations become enriched. Therefore, it is more effective to hybridize the subtracted "+" cDNA with "-" cDNA that has been subtracted to an equal extent so that the shared sequences in the two populations are present in similar proportions, optimizing the chance that homologous sequences will pair and therefore be subtracted from both pools (see **figure 3**) (Wang and Brown, 1991).

Short Hybridizations:

Whereas the long hybridizations were designed to deplete rare, common cDNAs in both pools, short hybridizations were performed in the presence of driver cDNA that had been generated not from the parallel or equally subtracted cDNA pool, but from the original, "+/- 1" cDNA pools that have been through one round of long hybridization to deplete rare, shared sequences. These short hybridizations were designed to target sequences that were abundant in the starting material and present in both "+" and "-" pools . The vast excess of these cDNAs in the +/- 1 driver population over their amounts in the two tracer populations at each stage ensured that through successive subtractive hybridizations the corresponding cDNAs in the two + and - pools would be depleted regardless of whether these cDNAs had become anomalously enriched in one or the other of the + or - cDNA pools.

Removing biotinylated cDNAs:

Following hybridization, streptavidin was added to the hybridizations, followed by phenol extraction which removed virtually all biotinylated material from the aqueous phase. In theory, biotinylation of driver cDNAs will allow removal of all common sequences from the hybridization mix, including unhybridized, excess driver after biotin-streptavidin complex formation and phenol extraction. In addition, removal of PCR primer-binding sites from the ends of driver cDNAs helps to inhibit re-amplification of driver cDNAs not removed during the streptavidin binding and ensuing phenol/chloroform extractions, although not entirely (data not shown, (Wang and Brown, 1991)). The remaining single stranded "+" or "-" cDNA was then re-amplified by PCR and used to make both driver and tracer for the next round of long or short hybridizations.(see **figure 3B-C**).

Measuring the efficiency of extraction of biotinylated sequences:

The efficiency of biotinylation was measured by incorporation of α^{32}P-dCTP into the PCR-amplified cDNA prior to photobiotinylation, then after binding of streptavidin and phenol extraction, the amount of radioactive material remaining in the aqueous phase was determined. Photobiotinylation was somewhat variable in that two rounds of photobiotinylation (materials and methods) resulted in from 20 to 90% of the radioactive material being removed from the aqueous phase. However, after 3 rounds of photobiotinylation, the labeled cDNA remaining in the aqueous phase was consistently less than 4% of the input. Whether labeled cDNA was retained in the aqueous phase because it was un-biotinylated, it was not bound to streptavidin and extracted, or it had remained at the interface after phenol extraction is not clear. However, since the binding is done in the presence of excess streptavidin, it seems unlikely that the material remaining is there because it did not bind to streptavidin. Therefore, it is likely that either 4% of the cDNA was not biotinylated and/or the phenol extraction was incomplete.

Monitoring the progress of subtraction :

The progress of successive subtractions was followed at all stages by slot blot hybridizations in which small amounts of cDNA from each pool was cross-linked to nylon membranes and hybridized with α^{32}P-dCTP-labeled probes for specific genes. **Figure 4A** shows one such slot blot in which both *lacZ* and *nanos* cDNAs have already begun to be

significantly enriched in the "+" pool but are indistinguishable from background in the "-" pool by rounds 4 and 8 respectively. Simultaneously, the "-" and "+" cDNA populations were depleted of both abundant, shared transcripts such as actin 5C (see **figure 4B**) as well as of rare, shared transcripts such as *germcell-less, r16S, hunchback* and *pumilio* (see **figure 4C**).

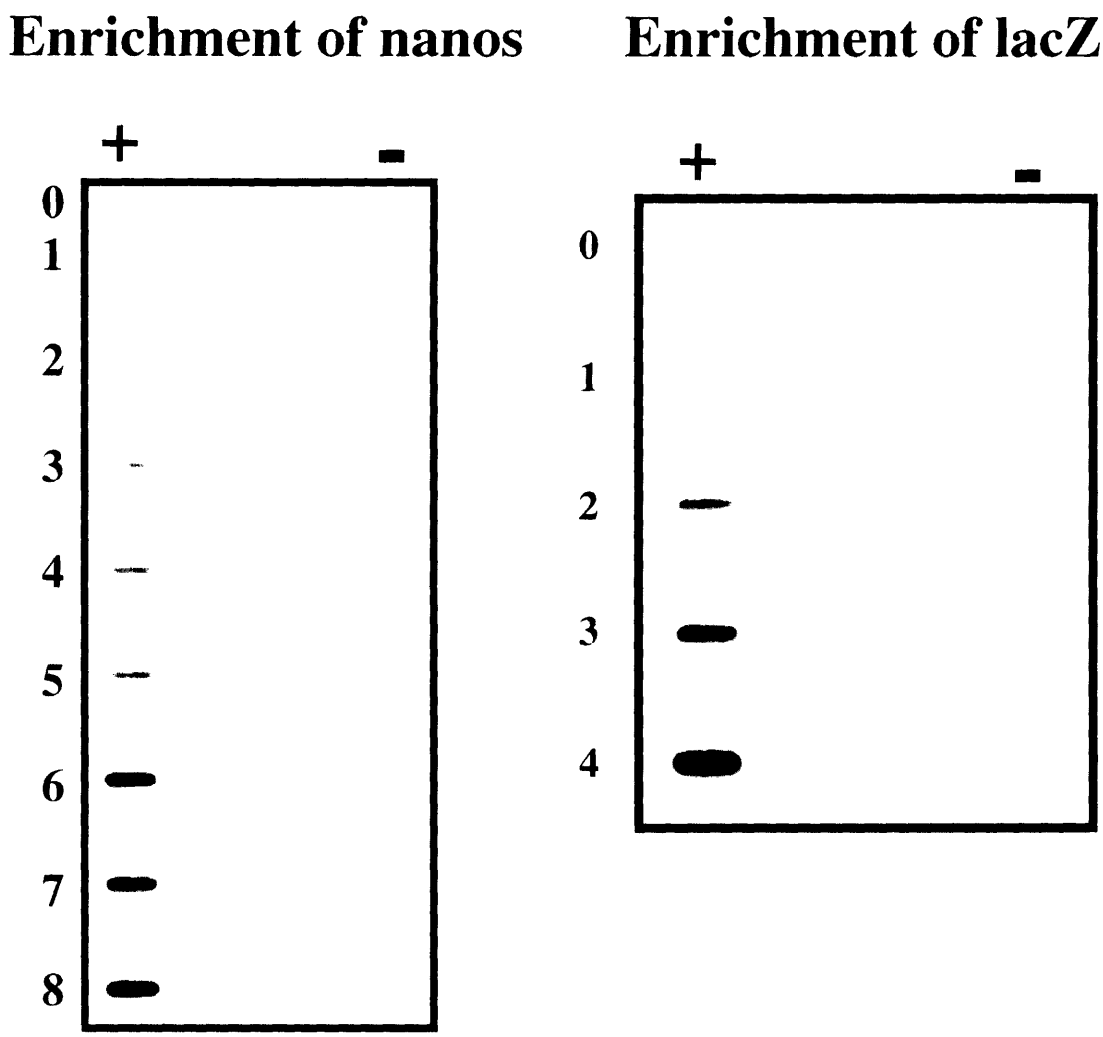


Figure 4A. Positive controls enriched: Slot blots showing the enrichment of endogenous *nanos* and and exogenous *lacZ* cDNAs from subtraction rounds 0 to 8 and 0 to 4, respectively. On the left is a slot blot in which equivalent amounts of cDNA from "+" and "-" cDNA pools from 0 to 8 have been cross-linked to nylon membrane (see materials and methods). Probing the blot with a *nanos* cDNA reveals that the endogenous *nanos* cDNA becomes progressively enriched in the "+" pools and remains below background levels in the "-" pools. On the right is pictured another slot blot, generated in the same manner as described above using cDNA pools 0 through 4. This blot is probed with a labeled *LacZ* cDNA, revealing that the exogenous control cDNA is detectable only in the "+" pool (and that there has been no detectable cross-contamination) and that this cDNA, like the endogenous *nanos* cDNA, becomes enriched in the "+" pool as expected.

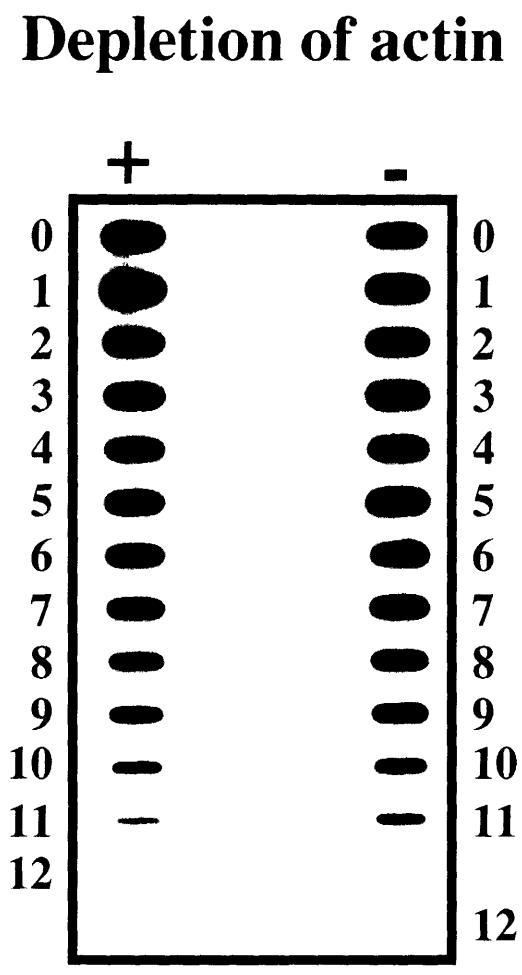


Figure 4B. *Actin* depletion: Slot blot showing the depletion of *actin* during subtraction rounds 0 to 12. This blot is generated in the same manner described in figure 4A. The probe in this case is derived from the *actin 5C* gene (see materials and methods). This blot reveals that the *actin* cDNA, which is abundantly expressed in both "+" and "-" pools, is depleted gradually from both pools. Finally, the *actin* cDNA is undetectable in +/-14cDNA pools (data not shown).

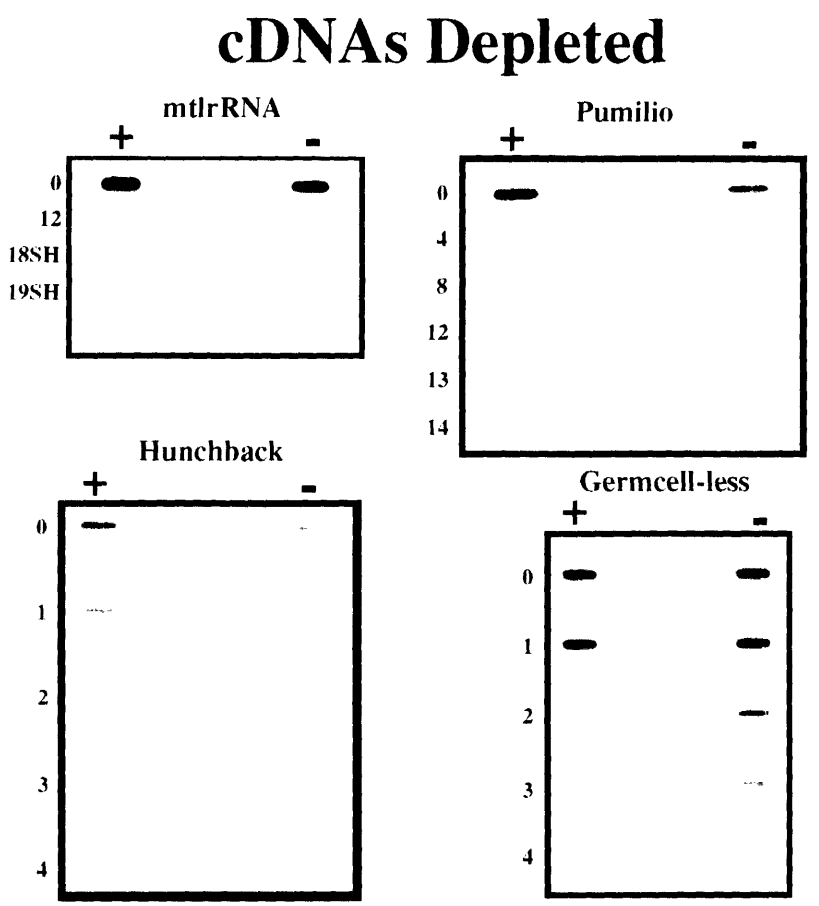


Figure 4C. Slot blots showing the depletion of four different endogenous cDNAs, *mtlrRNA*, *pumilio*, *hunchback* and *germcell-less* from rounds 0 to 19SH, 0 to 14, 0 to 4 and 0 to 4, respectively. The transcripts for *pumilio* and *hunchback* are uniformly distributed throughout the embryo and are therefore not expected to be enriched in either pool. The *mtlrRNA* is enriched at the posterior pole of syncytial blastoderm embryos, but this RNA is not detectable in fully formed pole cells, and is therefore expected to be depleted from both pools as well. The *germcelll-less* RNA is enriched at the posterior pole of syncitial embryos, incorporated into pole cells upon formation, and is detectable in these cells until stage 9, prior to pole cell migration through the midgut epithelium (see **figure 1**). After this stage, *germcell-less* is no longer detectable in pole cells, and zygotic expression of this RNA becomes apparent in somatic tissues. Despite the apparent stabilization of the *germcell-less* message in pole cells during invagination of the migut, the cDNA appears to be depleted, perhaps because of the somatic expression which presumably occurs both in wild-type and *oskar*-mutant embryos. The cDNA pools loaded on each of these blots are indicated to the left of each panel. All of the cDNAs shown in this figure are present in both "+" and "-" starting cDNA pools and are eventually depleted. The uneven depletion of the *hunchback* cDNA is particularly noticea ble, however, and is probably due to the slight excess of *hunchback* cDNA in the "+0" cDNA pool (also see **figure 3B**).

Eight of the ten genes examined by slot blot hybridization show the expected distribution between the "+" and "-" pools throughout the subtraction process (see **figure 4A-C**). The two cDNAs that gave unexpected results are the *oskar* and *vasa* cDNAs (**figure 4D**). In the case of *oskar* there is no prior evidence for decreased stability of this transcript in embryos of $osk^{301/CE4}$ females, at 18°C (C. Rongo, personal communication), nevertheless, *oskar* cDNA was present in the "+" cDNA pool and not detectable in the "-" cDNA pool. It is therefore not surprising that this cDNA then became highly enriched in the "+" pool after many rounds. Clarification of this unexpected result awaits the probing of developmental northern blots of RNA from wild-type and *oskar* mutant embryos raised at 18 and 22°C to determine whether there is indeed an effect on RNA stability. A different but equally unexpected result was obtained with *vasa* cDNA (**figure 4D**), which was thought to be present in roughly equivalent amounts in wild-type and *oskar* mutant embryos, but appears to have become at least transiently enriched in the "-" pool before it was subtracted from both. Apparently the small excess of *vasa* RNA present in the "-" pool (see **figure 3C**) was sufficient to allow this cDNA to become enriched in the "-" pool instead of being subtracted from both populations as expected.

cDNAs Anomalously Enriched

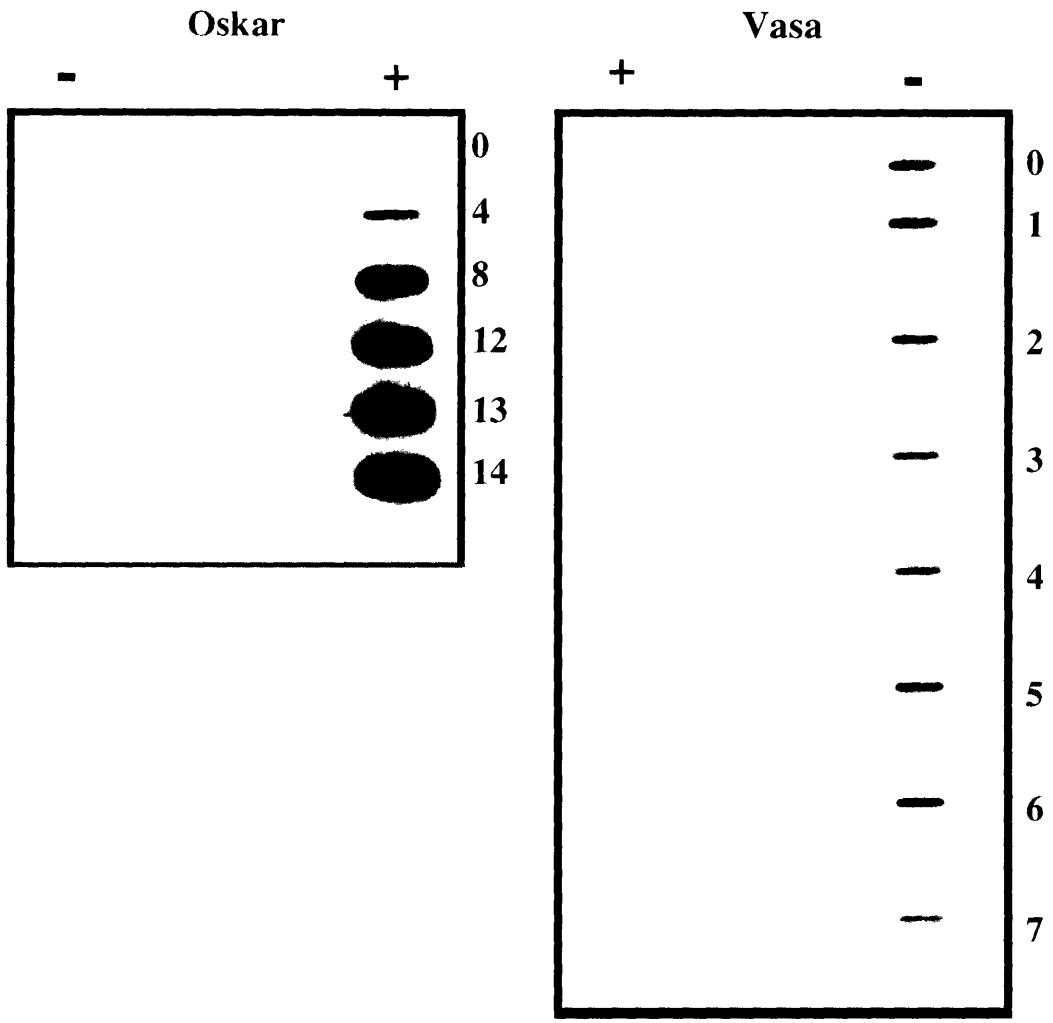


Figure4D. Anomalously enriched cDNAs: On the left is a slot blot showing enrichment of the *oskar* cDNA in the "+" pool from rounds 0 to 14. The fact that the *oskar* cDNA is readily detectable in the +0 cDNA pool, whereas it appears to be just above background in the -0cDNA pool does not correlate with any information on the stability of *oskar* transcripts in wild-type versus mutant embryos. However, given it's unequal distribution between the two starting cDNA pools, it is not surprising that the cDNA became further enriched during the course of subtraction. On the right is shown a slot blot in which the enrichment of *vasa* cDNA in the "-" pools from rounds 0-7 can be seen. In contrast to the result shown for *oskar*, the *vasa* cDNA appears to be enriched in the "+" cDNA pool relative to the "-" cDNA pool. This enrichment can most likely be attributed to the slight enrichment of the *vasa* mRNA in the "-" pool shown on the RNA blot in figure **3C**.

By round 14 *actin 5C* cDNA was no longer detectable in either pool (data not shown), however, there was still significant cross-hybridization between the two populations and it appeared that the long hybridizations to +/-n driver were not helping to significantly decrease the amount of common sequences between the two populations (data not shown). At this stage the protocol was modified and two sets of hybridizations were performed, the first being the normal 20 hour hybridization between +/-14 cDNA tracer and -/+14 biotinylated driver. In parallel, the same sets of driver and tracer cDNA pools were mixed, but the hybridization was only allowed to continue for two hours (pools resulting from these parallel short hybridizations are designated "+/-NSH" to distinguish them from pools generated by the original long hybridization protocol). These two types of hybridizations were conducted in parallel for comparison and the results can be seen in **figure 4E**. By +/-18SH there is a clear difference in cross-hybridization between the two types of subtracted pools, the long hybridizations did not appear to be necessary and after +/-18, only short hybridizations were performed.

Long hyb versus short hyb

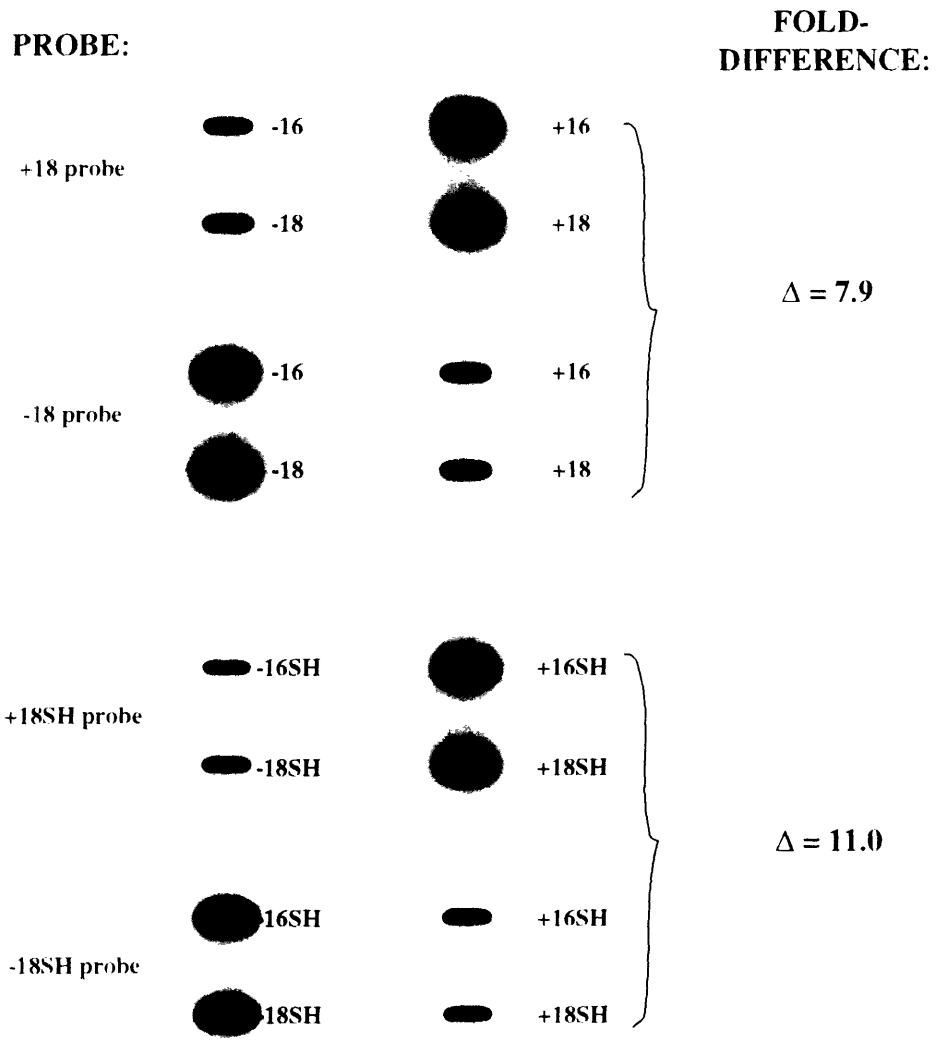


Figure 4E. This figure shows a slot blot in which equal amounts of cDNA from the pools indicated were probed with labeled cDNA pools corresponding to the opposite pool, and for comparison, to cDNA from the same pool. These cross-hybridizations reveal that by round 16 the long hybridizations of + or - NcDNA tracer against - or + NcDNA driver were no longer more effective in depleting common sequences between the two pools than short hybridizations were. Although rare, common cDNAs present in both pools were still being preferentially subtracted after the long hybridzations, at this stage in the experiment, the goal was to obtain the maximum % Difference between the two pools.

Once the difference in cross-hybridization signals between the two populations reached 11-fold (or 90.91% different, see **table 1**), the known cDNAs, *oskar, nanos*, and *lacZ*, which had been enriched in the "+" library were subtracted by addition of purified full-length cDNA fragments encoding these genes to the "-18SH" driver pool. The progress of their depletion from the "+"cDNA was monitored by slot blot hybridization (data not shown), and the percent difference between the two pools, as measured by their cross-reactivity, is recorded in **table 1**.

Table 1: This table shows the calculated "% Carry-over" frequencies for the +/-14 through +/-21SH subtractive hybridizations. The % Difference values were calculated by quantifying the hybridization signal resulting from the ^{32}P-labeled +cDNA pool hybridized to the -cDNA pool in comparison to the signal for ^{32}P-labeled +cDNA pool hybridized to the +cDNA pool itself. For example, if the + on + hybridization signal is 10-fold stronger than the + on - signal, this is taken to mean that 10% of the + pool is present in the - pool as well, leading to a % Difference of 90%. "% Carry-over" values were calculated using the equation below:

All % Difference and "% Carry-over" values were calculated using the following equation:

$$100x \quad \frac{\%\text{Diff.(before)}}{(\%\text{Diff.(before)}+\%\text{Carry-over})} = \%\text{Diff.(after)}$$

The key variables are the percent difference before "%Diff.(before)," and after "%Diff.(after)," subtractive hybridization, and the amount of common sequences that are retained in the subtracted pool after hybridization or "% Carry-over." The effect of different levels of "% Carry-over" can be calculated using the equation above. Such calculations reveal that , assuming a "Carry-over" value of 5%, using the equation above, it would take 5 subtractions to reach the maximal value of 95% specific cDNAs in the final pool. For comparison, with a "% Carry-over" value of 10%, it would take seven rounds before the maximal percent difference of 90% could be reached. The "Carry-over" value presumably depends on a number of factors including the extent of hybridization achieved and the fold-excess of driver over tracer. Counteracting these variables is the frequency with which driver cDNAs are capable of being re-amplified after subtraction.

Entries in the "Subtraction" column denote the subtraction round, e.g. "+/-14-+/-16" indicates that the % Carry-over values on that line were calculated using the %

Difference between the +14probe hybridized to the + and -14cDNA pools versus the +16probe hybridized to the + and -16 cDNA pool. The "Hybridizations" column contains entries that explain the kinds of subtractions performed between successive pools. The last three hybridization reactions had additional sequences added to the driver pool in each case (see materials and methods).

The values in this table allow direct comparison between the original subtraction scheme, consisting of alternating long and short hybridizations and the modified scheme, in which the long hybridizations were shortened. Comparing the +/-16-+/-18 and +/-16SH-+/-18SH values (indicated by "*") reveals that the % Carry-over for the modified protocol is less than that for the unmodified protocol, suggesting that the short hybridizations are just as effective a long hybridizations and may result in subtraction of more common sequences than the long hybridization performed in parallel.

Table 1 The calculated "% Carry-over" for a sample of six of the subtractive hybridizations ranging from +/-14 to +/-21SH.

Subtraction:	**Hybridizations:**	**%Difference:**	**%Carry-over:**
+/-14-+/-16	Long, Short (+/-1)	+/-14: 84.61%	10.14%
		+/-16: 89.31%	
+/-16-+/-18	Long, Short (+/-1)	"	12.76%*
		+/-18: 87.5%	
+/-16SH-+/-18SH	Short (+/-16), Short (+/-1)	+/-16SH: 91.03%	9.13%*
		+/-18SH: 90.91%	
+/-18SH-+/-19SH	Short (+/-18SH,lacZ,nos,osk)	"	5.68%
		+/-19SH: 94.12%	
+/-19SH-+/-20SH	Short (+/-18SH,lacZ,nos,osk)	"	6.07%
		+/-20SH: 93.94%	
+/-20SH-+/-21SH	Short (+/-20SH,lacZ,nos,osk)	"	6.06%
		+/-21SH: 93.94%	

Cloning of putative pole cell specific cDNAs:

Since the aim of this work was to identify genes that play a role in determination and/or differentiation of the germline, I chose to focus on isolating genes present in the "+" cDNA population and not in the "-" cDNA pool. The rationale for this was that any genes unique to the "-" pool could not, by design, be derived from germ cells. Although the isolation of genes required in somatic tissues for proper differentiation of embryonic germ cells will be interesting and critical to formulating a complete understanding of the process of gonad development in the embryo, this process was not addressed in these experiments.

The "+"cDNA pool from round 21 was cloned into a plasmid vector and transformed into bacteria to make the +21SH subtracted library. Colonies from the +21SH library were screened by hybridization to probes made from the final "+" and "-" cDNA pools. Colonies hybridizing strongly to the "+" cDNA probe but not to the "-" cDNA probe were picked and plasmid DNA was isolated from each one.

Classification of cDNAs:

The +21SH cDNA library was hybridized with both +21SH and -21SH probes, roughly 10% of the colonies were picked on the basis of their strong hybridization to +21SH cDNA probe but not to -21SH-derived probe. The number of colonies that gave moderate hybridization signals but were not picked is ~10% of the total, and those that gave weak hybridization comprised another ~10%. When the +21SH library was probed with the -cDNA probe, roughly 10% gave strong hybridization signals whereas 30% of the signals were very weak. The remaining 30% of the colonies showed no hybridization to either probe. The lack of detection could be because these colonies contained re-ligated vector without inserts or because the amount of the corresponding cDNA in both the + and - libraries was too low to give a detectable hybridization signal. In order for a given sequence in a mixed probe to give rise to a signal recognizable above background it must be present at a level of 0.1% or more in the population (Sambrook et al., 1989) . The 10% of the original colonies that were picked comprised 146 "positive" clones.

Cross hybridization between PCR-amplified inserts from eight randomly chosen clones from the 146 positives, suggested that there were a small number of unique classes to which these clones could be assigned (data not shown). The identity of cross-hybridizing inserts was confirmed by sequencing. Subsequently, by using a combination of cross

hybridization between inserts from all 146 positives and sequencing it was established that the majority of the 146 clones could be definitively assigned to one of 9 distinct classes. Sequencing revealed that the inserts within each of these groups were identical, suggesting that through the process of subtraction and amplification the initially heterogeneous populations of cDNAs containing overlapping sequences were narrowed to include just one representative sequence that out-competed the others during PCR amplification. **Table 2** provides information from cross-hybridization between the products of PCR amplified inserts from each of the 146 inserts.

Table 2: Table showing frequency of each clone in the final pool of 146 cDNAs analyzed. In the column labeled "Clone" are listed the names given to the nine original cDNAs. "# of copies" connotes the actual number of isolates for each cDNA. The "% population" table entries represent the percentage of the total each clone contributes. Cross-hybridizations between all 146 library inserts and each of the 9 cDNAs identified by sequencing were performed (see materials and methods). Note that 21 of the inserts gave "Multiple signals," which indicates that strong cross-hybridization signals were obtained with more than one probe, suggesting that these clones contained multiple inserts. Sequencing these inserts would determine unequivocally whether they are in fact, hybrid clones. The "No hybridization" class of inserts is most likely due to simple variability in efficiency of PCR amplification of inserts. The PCR amplifications of each clone were conducted in parallel, and equivalent amounts of each PCR reaction were loaded directly onto slot blots (see materials and methods).

Clone:	**# of copies**	**% of population:**
MA	5	3.4
MB	9	6.2
MC	8	5.5
ME	2	1.4
MF	2	1.4
MG	8	5.5
MH	5	3.4
THE5	39	26.7
Toll	18	12.3
Multiple signals	21	14.4
No hybridization	29	19.9

Sequencing of the 9 distinct cDNAs, followed by BLAST (Atschul et al., 1990) database homology searches revealed that one of them corresponded to the *Toll* gene, which is involved in dorso-ventral patterning of the early embryo and is not differentially expressed in wild-type versus germ-cell-less embryos. The remaining clones were named M (for "mystery") followed by A, B, C, E, F, G, H and "The 5". The MH clone showed significant homology to the *Drosophila* sarco/endoplasmic reticulum-type Ca^{2+}-ATPase gene, and sequence from the near-full length clone confirms that this cDNA derives from the previously identified Ca^{2+}-ATPase gene (Magyar and Varadi, 1990) ,see Appendix I for sequence of MH. The remaining seven cDNAs showed no sequence homologies to known genes in the databases and therefore define new genes.

Discussion:

I have performed a subtractive hybridization to screen for genes expressed specifically in the developing embryonic gonads. The two populations of starting material were RNAs derived from wild-type embryos and RNA from embryos laid by females homozygous mutant for a weak allele of the grandchildless gene, *oskar*. The embryos produced by these mutants do not make embryonic germ cells but are otherwise wild-type. The principle behind this screen was that any RNA molecules that are uniquely present in wild-type embryos but not in embryos lacking germ cells would become enriched as the common sequences were depleted. After twenty one rounds of subtraction I identified seven unique sequences which are enriched in the cDNA library derived from embryos with germ cells, but not in that derived from those lacking germ cells.

Hybridization:

One of the primary considerations for any subtractive hybridization is the size of the target RNA population. In the case of the pole cell subtraction the percent difference is small because the complexity of the somatic RNA is so large in comparison to that of the germline-specific RNA. The smaller the population of differential clones relative to background, the greater the potential enrichment of target sequences will be (see below and **table 3**). However, the extent of this enrichment after each round of subtractive hybridization depends critically upon the extent of the hybridization achieved. Repeated cross-subtraction allows one to reach the high levels of hybridization between both rare and abundant common sequences necessary to enrich for unique sequences in such a complex background.

Another argument in favor of using an iterative subtractive hybridization has been proposed by Alt (1989). The author describes subtractive hybridization experiments between two cell types that exhibit a small percent difference in the RNAs they express. Tracer cDNA was synthesized from the RNA of one cell type and RNA from another cell type was used as driver during two successive hybridizations, using 5-fold excess driver RNA relative to the tracer cDNA. In parallel, a single round of subtractive hybridization was performed, in which the driver was present in 10-fold excess over tracer. Comparison of the subtracted pools revealed that two successive subtractions against the 5-fold mass-excess of driver was much more effective in enriching sequences unique to the tracer pool than the single subtractive hybridization of this same cDNA pool against 10-fold excess of RNA. In fact,

for very complex starting pools, iterative subtractions may be absolutely required to achieve sufficient enrichment of certain differential sequences. However, for two populations with a large percent difference to begin with, there is less benefit from repeated subtractions (see **table 3**, below).

To take a concrete example, let us examine a subtractive hybridization between RNA pools isolated from cellular blastoderm-stage, wild-type and *oskar* mutant embryos, keeping in mind that of the approximately 5000 cells cellular blastoderm, at most 50 are pole cells (Campos-Ortega and Hartenstein, 1985; Zalokar and Erk, 1976). Roughly one percent of the cells in wild-type embryos are pole cells, as compared to 0% in *oskar* mutants. If pole cell-specific RNAs represent ~1% of the total (probably a large over-estimate), all others being present in somatic cells, there would be a 0.01% difference between the two starting pools (assuming for simplicity that cDNA synthesized from this RNA is 100% representative of the original RNA population). The efficiency of subtraction can be expressed in terms of the amount of common cDNA sequences that are not removed during hybridization or "Carry over" which is equivalent to one minus the efficiency of the subtraction (see legend of **table 1**). Assuming a carry-over value of 5%, the resultant percent difference after one round of subtraction would be 0.2%. After two rounds this number increases 19-fold to 3.85%, and finally reaches 95% after five rounds. If, however, the percent difference is already 20% to begin with, then after one round of subtraction this number only increases four-fold to 80%. After two rounds it is up to 94%, and after only 3 rounds the percent difference is already up to the maximal value of 95%. (see **table 3**).

Table 3: Table showing the effect of repeated subtractions on populations with low and high % differences to begin with. This table contrasts the theoretical results of subtractive hybridizations that begin with the percent difference between the two populations being 0.01% and 20% . In the column labeled "subtraction" is the number of subtractive hybridizations performed to achieve the % Difference seen in the "% Difference" column. "Fold Increase" indicates the increase in the % Difference achieved after that particular round of subtractive hybridization relative to what it was before that round.

Comparison of the "Fold increase" numbers for the 0.01% Difference versus 20% Difference subtractions reveals that the benefit of doing more rounds of subtraction is much

greater in the case where the starting pools have a very small % Difference (e.g. 0.01%), than if the % Difference between the starting pools is already 20%.

Table 3: Demonstrating the relative advantage of iterative subtractions: Depending upon the initial % Difference.

Subtraction:	% Difference: starting with **0.01%**	Fold-increase:	% Difference: starting with **20%**	Fold-increase:
0	0.01%	NA	20%	NA
1	0.2%	20	80%	4
2	3.8%	380	94%	1.2
3	43%	4300	95%	1.01
4	90%	9000	95%	1
5	95%	9500	95%	1

The example outlined in table 3 demonstrates that part of the reason the pole cell subtraction took twenty one rounds to reach the point where the two pools were 95% different from one another is probably due to the small percent difference (probably less than 0.01%) between the mRNA pools. The wide distribution of embryonic stages used to isolate the starting mRNA pools increased the complexity of these populations since they not only contain all of the early maternal RNAs but also zygotically transcribed sequences. Other factors contributing to the slow progress of the subtraction were the amount of common sequences left in the tracer pool (% Carry-over, see below), and the fact that the driver was double stranded cDNA. With double stranded driver the hybridization between tracer and driver is always competed by reassociation of driver duplexes. Therefore, more of the common tracer remains unhybridized to driver after every round than would be true if the driver were single stranded.

The percent difference between the two starting populations could be increased by narrowing the embryo collection time to include only embryos from ~3-6 hours after egg laying. Collecting embryos within this time window would increase the proportion of embryos in which early maternal RNA has begun to be degraded and zygotic transcription in the pole cells has clearly begun (Zalokar, 1976). Therefore, the abundance of any given

germ cell-specific RNA in this population would be greater than in a 45 minute-13 hour collection which contains extremely high amounts of maternal RNAs that effectively dilute any sequences specifically contributed by the germ cell population.

The reassociation of complementary DNA or RNA strands depends upon their collision in solution and therefore the rate of annealing depends upon the concentration of complementary strands. The parameter most frequently used to describe the level of completion of reannealing reactions is "C_0t:" the product of the nucleic acid concentration (C_0) and the time (t) allowed for hybridization (Britten and Kohne, 1968). The units of this parameter are typically moles of nucleotides times seconds per liter. The C_0t calculated for the long hybridizations in this experiment is 545 moles base pairs liter^{-1} seconds. However, given that the driver cDNA is present in 20-fold excess over the tracer cDNA molecules, the probability that any given cDNA driver molecule will hybridize to a tracer molecule is 20 times greater than the probability that the tracer pool homologue will hybridize to it. Therefore, the C_0t value calculated here is not strictly comparable to C_0t's calculated for hybridizations between DNA strands present in equal proportion, but resembles more "R_0t" values obtained when excess RNA is used as driver instead of cDNA. In this case R_0 represents the concentration of RNA used to drive hybridization of cDNA tracer into hybrids. The key difference between R_0t and C_0t values is that in excess RNA reassociation experiments the RNA driver is incapable of inter-strand hybridization and therefore the rate of annealing of tracer to driver depends completely upon the concentration of the driver RNA (assuming the driver is in sufficient excess that the complementary cDNA strand of the tracer is negligible in comparison). It has been shown that R_0t values of 200 moles bases liter^{-1} seconds are sufficient to allow hybridization of even very rare cDNAs in a population (Zebrowski et al., 1994). However, the effective C_0t value achieved in these experiments is probably somewhat less than the calculated value given above since the effective concentration of the driver is exponentially decreasing with time due to reannealing of biotinylated driver to itself.

Although the effective C_0t for hybridization of tracer to driver will be decreased by self-association of driver, the cumulative extent of hybridization between driver and tracer will increase during subsequent rounds of subtractive hybridization in which the most abundant shared molecules present in the driver and tracer are depleted and the complexity of the driver pool decreases. The iterative nature of this technique therefore allows one to reach final additive C_0t values much greater than would otherwise be achievable in one round with double stranded driver. If single stranded driver were used, the effective C_0t would

actually be very close to the value of 545 moles base pairs liter^{-1} second calculated above, and therefore the subtractive hybridizations would be expected to be much more effective, likely decreasing the number of rounds necessary to achieve complete subtraction of common sequences and enrichment of unique clones.

Separation of hybrids from target

After hybridization, the key to any subtractive hybridization method is separation of tracer target molecules from driver. As described above, one of the original methods used for separation of single from double-stranded DNA was HAP chromatography (Bernardi, 1965; Miyazawa and Thomas, 1965; Walker and McLaren, 1965). This technique provides clean separation of single-stranded from double-stranded nucleotides, but requires the construction of a water-jacketed column or similar device to provide a 60°C constant-temperature environment. The advantage of HAP chromatography is that with it one can isolate not only the single stranded material that initially flows through the column but the double stranded material as well. This property makes it extremely useful for positive selection protocols designed to isolate cDNAs present in another population of RNA or cDNA. However, besides the logistical difficulties involved in setting up the chromatography column, HAP carries the disadvantage that any RNAs or single stranded cDNAs that have a high tendency to form stable secondary structures will bind tightly to the column and therefore contaminate the double stranded pool (Welcher et al., 1986).

HAP chromatography has been used extensively in subtraction schemes using RNA as the primary source of driver sequences, since large excess of RNA driver can be used during hybridization and then completely removed from singe stranded cDNA tracer pool by incubation at high pH, without the need for chemical modification of either driver or tracer. In principle, one could use HAP to separate single stranded cDNAs from cDNA:cDNA hybrids, and any contaminating driver in the tracer pool could be selected against by using tracer-specific adapter sequences for PCR amplification. However, in this subtraction I used a single adapter sequence for both driver and tracer pool, obviating the use of HAP chromatography for separation of hybrids from target cDNAs.

Among the methods currently in use for separating driver from tracer, biotinylation followed by complex formation with streptavidin and repeated phenol/chloroform extraction (biotin/streptavidin/phenol) is arguably among the simplest and most versatile (Sive and St. John, 1988). Any population of double or single stranded DNA or RNA can

be biotinylated either by photocross-linking or by enzymatic incorporation of biotinylated nucleosides and hybridized with the unlabeled tracer pool of choice (M. Patel, personal communication). The mixture can then be incubated with streptavidin and simply extracted with phenol:chloroform, leaving single stranded, non-biotinylated cDNAs in the aqueous phase. Alternatively, the streptavidin-bound material can be passed over a copper chelate-agarose column that reversibly binds streptavidin-biotin-nucleic acid complexes (Porath et al., 1975; Welcher et al., 1986). This chromatographic step allows one to specifically isolate the enriched single stranded material isolate double stranded material through elution of the bound material with EDTA. Therefore, for the iterative subtraction approach used in this work, biotinylation provides the flexibility in driver synthesis necessary to allow repeated rounds of subtractive hybridization without the need to isolate or synthesize large amounts of mRNA, and although biotin/streptavidin/phenol-driven removal of driver may not be as stringent as alkaline hydrolysis of driver RNA after HAP chromatography, biotinylation allows one to specifically target one population of cDNAs (or RNAs) in a mixture for removal, even if the two pools contain identical or no adapter.

Efficient biotinylation:

The problem with having inefficient biotinylation is that any driver cDNAs not extracted from the aqueous phase after hybridization will remain in the tracer pool and (since they contain the same primer annealing sequences) can be re-amplified along with the tracer if the primer annealing sites were not removed by the restriction digestion. Unfortunately, neither EcoRI digestion nor photobiotinylation completely prevent re-amplification of cDNA pools (data not shown) (Kaufman and Evans, 1990). However, one could decrease the amount of common material re-amplified along with the tracer during each round of re-amplification by using different adapters for the "+" and "-" pools (Hakvoort et al., 1993). As mentioned above, incorporation of biotinylated nucleotides into the cDNA during amplification of the driver cDNA pool is another alternative to photobiotinylation, as is chemical synthesis of singly biotinylated primers. A single biotin molecule per oligo is sufficient to allow efficient streptavidin binding and therefore allows purification of full length biotinylated oligos by virtue of their ability to bind reversibly to streptavidin-agarose. (Multiple biotinylation causes the oligos to bind irreversibly to avidin-agarose, making them difficult to work with). Using such oligos allows one to eliminate the problem of inefficient incorporation of biotin into the driver cDNA.

Selection of differentially expressed clones:

The fact that the pole cell subtraction resulted in a library containing relatively few clones, ranging from 1.4 to 26.7 percent of the population (see **table 2**), suggests that further screening of the +21SH library with + and -21SH probes to identify clones that gave weaker hybridization signals would result in the isolation of more cDNAs of interest. Alternatively, the nine cDNAs identified in this first screen could be added to the driver pools, as was done for *oskar*, *lacZ*, and *nanos*, and subtracted out of the +21SH library. The resulting further-subtracted library could then be screened in the same manner as described above to isolate cDNAs that were too rare to be easily detected in the +21SH library.

Even in the subtracted population some species will be very abundant and other, equally interesting species will be rare, due to initial abundance of certain clones in the original cDNA and to preferential amplification of certain clones during the subtraction. One method that should be incorporated into future subtractive screens is to normalize the final +21SH library using hybridization to genomic DNA to decrease the number of sequences that are very abundant in the final subtracted pool. Normalization can also be accomplished by the use of reassociation kinetics: based on data that cDNA reannealing follows second order kinetics (Wetmur and Davidson, 1968), rare species in the population will anneal less rapidly and will become progressively normalized during the course of the reaction. Based on these principles, protocols have been designed in which primer extension is used to create short complementary copies of a population of cDNAs which can then be hybridized at low C_0t to selectively deplete abundant species which will re-anneal more quickly, decreasing the amount of repetition of the very abundant clones and thereby increasing the concentration of rare cDNAs in the population (Soares et al., 1994).

Alternatively, there are a number of different schemes which can be used in combination with subtractive hybridization to select for or exclude certain species from the subtracted population. For example, Davis et al. (Davis et al., 1987) used a number of criteria to screen candidate cDNAs for the possibility of encoding a gene with the expected properties. Their approach included subtractive hybridization of cDNA from two different tracer populations, both of which were expected to express the gene of interest but which had very different origins, against an undifferentiated cell type which was not expected to express the determining gene. These two rounds of negative selection were followed by positive selection for cDNAs in the subtracted pools that were also present in a third differentiated cell type. Strategies like these can be employed to further increase the

specificity of a pool of subtracted cDNAs using secondary and tertiary screens based on the expected unique properties of the genes of interest.

In the case of the subtractive hybridization described here, although positive selection could have been accomplished by hybridizing the final subtracted pool with an excess of mRNA from 3-6 hour embryos, to a high R_0t, followed by isolation by hydroxyapatite chromatography to isolate cDNAs capable of forming duplexes with molecules in the mRNA pool. Not only does this provide another selection for cDNAs of interest but it provides a convenient means of eliminating "false positive" cDNAs that may have been carried through the entire series of subtractions, but which are not differentially expressed. The *Toll* cDNA is and example of a false positive that may have been carried through due to intra-strand annealing, preventing hybridization with biotinylated driver. Carry-over of false positives can also be due to single stranded molecules that are unable to form stable duplexes under the experimental conditions, due to high AT content, for example (Fargnoli et al., 1990).

Finally, the sensitivity of the differential screening in which double stranded cDNA probes made from the final library pool were used to screen the final subtracted "+21SH" library may have been improved by the use of single stranded RNA or DNA probes. Therefore it may have be beneficial to include directional cloning in the library synthesis by using a modified oligo-dT primer for first strand synthesis that contains a T7 polymerase promoter sequence flanked by a restriction site. It is possible that increasing the signal to noise ratio for this final screening step may have prevented isolation of clones that appeared to hybridize differentially to the two probes, but which in fact were not differentially represented. Although this method would only allow isolation of the 3' ends of mRNAs, this experimental design would allow one to transcribe antisense cRNA probes directly from the final + and - cDNA pools to make RNA probes for screening the cloned subtracted library. In addition, after directional cloning and digestion with a defined restriction site 5' to the inserts, one could synthesize anti-sense riboprobes for *in situ* hybridization to embryos.

Defining function of subtracted cDNAs:

One important disadvantage of molecular screening is that the function of genes identified in this manner can be difficult to establish. However, there are a number of techniques that can be used to address this question. For example, *in situ* hybridization to cells or tissues

can be used to examine when and where the corresponding RNAs are expressed in the animal. Information of this sort is extremely valuable not only for screening purposes but also in establishing function. Sequencing and identification of the predicted protein sequences can also provide hints about function, especially if the genes encode proteins with significant similarity or identity to polypeptides of known function. However, even when compelling sequence identities allow the formulation of logical and testable experimental models, proof of functional significance ultimately depends upon isolation of mutations in the gene(s) of interest and the examination of the loss of function phenotype *in vivo*.

In the absence of mutations in the gene of interest, misexpression of the wild-type or specifically-mutated protein *in vivo* can also provide information about function. If sequence analysis reveals significant identity to well-studied classes of molecules, then clear predictions about the changes necessary to create dominant negative versions of the protein can be made. These specifically mutated proteins can then be expressed in the animal followed by assessment of the resulting phenotype. For example, proteins with strong sequence identity to small GTPases provide excellent subjects for making changes in known active sites that will disrupt the function of the protein in predictable ways. Of course the ultimate aim of this kind of analysis is to determine what the real requirement for the gene is *in vivo*. Toward this end, isolation of mutations by various means can be undertaken, including, homologous recombination in mice and generation of deletions by x-ray mutagenesis in Drosophila.

In the case of *MyoD*, *in situ* hybridization confirmed its expression during skeletal muscle foundation and maturation *in vivo*. Sequencing this gene revealed homology to *c-myc* and gene products of the Drosophila achaete-scute locus, leading to the speculation that *MyoD* was a transcription factor responsible for activating expression of downstream muscle-specific differentiation genes. Further analysis revealed that MyoD is one of at least four members of a sub-group of the helix-loop-helix class of transcription factors including, *myogenein*, *myf-5* and *MRF4*.

Given that all four of the myogenic genes are capable of stimulating myogenic conversion in non-muscle tissue-culture cells (Weintraub et al., 1991), and that they can often substitute for each other to activate downstream genes encoding muscle proteins, it is perhaps not surprising that *in vivo* knock-out experiments show that *MyoD* function alone results in no dramatic or even detectable morphological defects in skeletal muscles

(Rudnicki et al., 1992). Therefore, although identification of the loss-of-function phenotype often provides the key to understanding function *in vivo*, in the case of *MyoD*, functional redundancy makes the genetic analysis of this locus much more complicated. Ironically, this redundancy in the muscle-determination pathway makes it unlikely that this important class of developmental regulators would ever have been identified by classical genetic approaches alone. This example reveals the power of molecular screening approaches, but also underscores the point that a combination of molecular and genetic approaches is often required for dissection of complex biological processes.

Materials and Methods:

Embryos staged from 45 minutes to 13 hours post egg-deposition were collected on apple juice-agar plates from population cages containing either *osk*$^{301/CE4}$ or wt/P(*lacZnos3'UTR*) flies (Gavis and Lehmann, 1992; Lehmann and Nüsslein-Volhard, 1986), and were treated identically in parallel reactions throughout the subtraction protocol. The embryos were washed with water, dechorionated in 50% Bleach (2.5% sodium hypochlorite in water), frozen in liquid nitrogen and stored at -80°C until RNA extraction. PolyA$^+$ mRNA was extracted using the QuickPrep mRNA Purification Kit (Pharmacia). Briefly, ~400µl frozen embryos of each type were homogenized in a buffered solution containing a high concentration of guanidinium thiocyanate and the resultant extract was diluted three-fold with Elution buffer, homogenized again and clarified by centrifugation. The supernatant is then incubated with oligodT-cellulose, washed sequentially with High and Low salt buffers and finally the polyA$^+$ RNA is eluted from the spin column according to the manufacturers instructions (Pharmacia, cat.# 27-9254-01). mRNA from the initial oligodT-cellulose spin columns was further purified by a second round of oligo-dT cellulose selection using Stratagene Oligo-dT Push columns, according to manufacturer's instructions.

For RNA blot analysis, ten micrograms of poly(A)$^+$ RNA was fractionated on a 1.3% agarose gel containing formaldehyde, transferred to nylon membrane (Hybond-N, Amersham), cross-linked in Stratalinker according to manufacturer's instructions, hybridized at 42°C in 50% formamide-containing hybridization buffer according to standard protocols (Sambrook et al., 1989) . After hybridization blots were washed sequentially as follows: 45 minutes at 42°C in 50% formamide, 5X SET, 0.5%SDS; 45 minutes in 2XSET, 0.5% SDS at 65°C; 30 minutes, twice, in 0.5X SET, 0.5%SDS, and finally 30 minutes, twice in 0.1XSET, 0.5% SDS. 20X SET consists of: 3M NaCl, 20mMEDTA, 200mM Tris, pH 7.8. Labeled blots were exposed on Fuji phospho-image plates, and the resultant hybridization signals were quantified directly. [α-^{32}P]dCTP-labeled probes were removed from the original blot by incubation in 50% formamide, 10mM $NaPO_4$, pH 6.8-7 for 1 hour at 65°C, shaking.

LacZ probe for hybridization to RNA blots was synthesized by random hexamer-primed [α-^{32}P]dCTP labeling (Sambrook et al., 1989).of the 3.4 kb EcoRI-SalI fragment of the E. coli *LacZ* gene from pGEM4lacZ, (gift of E. Gavis), which is derived from 2.4 kb Bam HI-XbaI fragment from pC4βgal, (Thummel et al., 1988) . The *actin* probe was

synthesized from whole plasmid DNA containing the Drosophila *actin 5C* gene (gift of D. Curtis). *Nanos* probes were generated by labeling an Hind III, Eco RI 2.3 kb fragment of pN5nos (Wang and Lehmann, 1991) . *Hunchback* cDNA fragments were derived from Bam HI, HindIII-digest of .... to generate a ~2.4 kb cDNA fragment. *Germcell-less* probes were produced by Bam HI, NruI digest of *gcl* cDNA (P. Zamore, personal communication). *Vasa* probes were generated from Bam HI, Xba I fragments derived from pARV-11 (Rongo et al., 1995)

Blunt ended cDNA was synthesized from 5µg polyA$^+$ RNA using an Invitrogen "Bidirectional Copy Kit Version 3.0" cDNA synthesis kit with oligo dT as first strand synthesis primer according to manufacturer's instructions. First strand synthesis was performed in the presence of 1 µl 3,000 Ci/mmol [α-^{32}P]dCTP and reverse transcriptase, to trace label the synthesized cDNA for ease of detection during subsequent manipulations. A small fraction of the final cDNA, was run on a 1% agarose gel in the presence of 1XTBE running buffer (Sambrook et al., 1989), dried and exposed to XAR-5 film (Kodak) to check the length distribution of the cDNA populations.

cDNA synthesis reaction products were resuspended in a final volume of 32µl after precipitation. Eight µl of each cDNA pool was digested in the presence of the 1X Low salt digestion buffer: 50mM Tris, pH7.5, 10mM $MgCl_2$, in a 10µl reaction with AluI and in parallel, 8µl of each cDNA pool was also digested under the same conditions with x units of AluI plus RsaI. All digests were carried out at 37°C for two hours, then heated to 65°C for 10 minutes to inactivate the enzymes and stored on ice until adapter ligation or stored at 20°C. One µl from each digestion was analyzed gel electophoresis on a 1.4% agarose/1X TBE, followed by autoradiography to check extent of digestion.

Gel purification of the two oligos annealed to make the adapters was critical to ensure that the majority of the oligos were full length so that after annealing the two would pair correctly to form one blunt end and one with a non-palindromic 3' overhang. The failure to form the correct adaptor with a blunt end would prevent efficient ligation of adapters to the cDNA fragments. Newly synthesized oligos in NH_4OH were extracted with n-butanol repeatedly until all aqueous phase was removed. Final pellets were resuspended in formamide/dyes sample buffer and fractionated on 15% Acrylamide 7.7M Urea gels. Bands were visualized by UV shadowing with long wavelength monitor (nm=320), excised and shaken overnight in H_2O at 37°C to elute. Eluate was filtered through 0.2µm

filter (Aerodisc) to remove acrylamide, extracted with n-butanol to reduce sample volume and ethanol-precipitated in the presence of 10mM $MgCl_2$ and 0.3M NaOAc.

Adaptor-kinase and ligation reactions: 45µg of the following 25mer: designated "Oligo A", TAGTCCGAATTCAAGCAAGAGCACA was incubated in the presence of 1X Kinase buffer: 66mM Tris, pH 7.5, 1mM ATP, 10mM MgCl2, 15mM DTT, 1mM spermidine, 0.2 mg/ml acetylated BSA, with 50 unitsT4 polynucleotide kinase (New England Biolabs) in a final volume of 250µl. The reaction was incubated at 37°C for 1 hour. 38.4µg (an equimolar amount) of the following unphosphorylated 21 mer: designated "Oligo B", CTCTTGCTTGAATTCGGACTA was then added to the reaction, which was then incubated at 45°C for 10 minutes to allow formation of non-self-complementary adapters with one blunt end and one 4-base 3'-overhang end containing an Eco RI site 6 bases 3' of the blunt end (Wang and Brown, 1991).

Adaptor Ligation reaction: 55µl of the freshly prepared duplex phosphorylated adaptor was then incubated in the presence of 30µl H_2O, 9µl 5X Kinase buffer (see above), and 5 Weiss units T4 DNA ligase (or 5µl Boehringer Mannheim 1U/µl) with 10µl Alu I or Alu I + Rsa I-digested cDNA, over-night (20 hours) at 15°C followed by 2 hours at room temperature (~22°C). Reactions were stopped by addition of 5µl 0.5M EDTA, pH 8.0 followed by two phenol:Cholorform:Iso-amyl-alcohol extraction (1:1:48), and a single Choroform:Iso-amyl-alcohol extraction. The cDNA was then precipitated in the presence of 0.3M NaOAc, air-dried and resuspended in 12µl TE (Sambrook et al., 1989) .

Gel Purification of Adaptor-ligated cDNA: 6µl (half) of each of the adaptor-ligated cDNAs (Alu I and Alu I + Rsa I samples were treated identically) were fractionated on 1.4% FMC SeaPlaque GTG low melt agarose (FMC, Rockland, ME) in 1X TAE. Sample buffer was run to a distance of 2 cm from wells, lanes containing pBR322-Bst NI markers were removed and stained with 0.05% ethidium bromide to visualize bands. Marker lanes were aligned with un-stained sample lanes and the region of the gel containing cDNA in the range from 150 bp up to 1.5 kb was cut out and the gel slices from Alu I and Alu I + Rsa I digested samples were mixed to give final volumes of 200-300 µl low-melt sample for each "+" and "-" cDNA pools. Low-melt samples were stored at 4°C.

Low melt samples were melted at 42°C and 1µl was used as template in 100µl PCR reactions containing 1.2µg 21mer oligo (see above), 1X PCR buffer (Promega), 200µM dNTPs, 1.5mM $MgCl_2$, 1 µl Taq polymerase (Promega). All amplifications was

performed according to the following protocol: 94°C 2 minutes, 1 cycle, 94°C 1 minute, 60°C 1 minute, 72°C, 2 minutes, 30 cycles, 72°C 10 minutes, one cycle. PCR products were purified following amplification by phenol:chloroform:isoamyl-alcohol extraction, followed by chloroform:isoamyl alcohol extraction, precipitated in the presence of 2M NH_4OAc, and 2.5 volumes 100% ethanol. Pellets were resuspended in a final volume of 100µl and purified over Sephadex® G-50 spin columns (Quick spin High Capacity Columns, Cat.#: 100-965, Pharmacia) according to manufacturer's instructions. Spin column eluates were ethanol precipitated with 0.3M NaOAc and 10µg glycogen and pellets were resuspended in sterile water. All PCR products were examined by 1.2% agarose gel electrophoresis to check purity and yield of cDNA amplifications after each round of subtractive hybridization. "No-template" controls were run in parallel with all PCR amplification and on the few occasions when amplified material was detected in these control reactions, all buffers were discarded and fresh aliquots were used in repeat amplifications to prevent any cross-contamination. Typical yields from PCR reactions were 12-13µg per 100µl reaction.

Preparation of Biotinylated Driver: 115µg purified cDNA was digested in a 150µl reaction volume of 1X EcoRI restriction buffer (New England Biolabs) with 1,500 units EcoRI for 3 hours at 37°C. After digestion cDNAs were purified by phenol:chloroform:isoamyl-alcohol extraction, twice, chloroform:isoamyl-alcohol once, precipitation in the presence of 2M NH_4OAc and resuspended in 100µl 1mM EDTA, pH 8.0. This purified cDNA is mixed with 100µl Photobiotin (1mg/ml in sterile water, Vector Labs, Burlingame CA) in a 1.5ml eppendorf tube, irradiated for 25 minutes at a distance of 10cm from a 275 W sunlamp (Vector Labs) with the cap open, in an ice bath. After photo-crosslinking DNA was purified by the addition of 30µl 1M Tris pH 9.1 followed by extraction four times with 200µl H_2O-saturated sec-butanol, once with 200µl chloroform and ethanol precipitated with 0.3M NaOAc. Pellets were resuspended in 100µl 1mM EDTA pH 8.0, mixed with photobiotin and the process was repeated twice more for a total of 3 times (Sive and St. John, 1988; Welcher et al., 1986); see Results for discussion of measuring biotinylation efficiency.

Efficiency of Photobiotinylation was measured as follows: PCR amplification of cDNA pools was conducted as described above with the addition of 0.1µl α^{32}PdCTP (3000Ci/mmol) to one of ten 100µl reactions. Resulting amplified products were pooled and purified by G-50 spin column followed by ethanol precipitation as described above.

Roughly one quarter of the resulting trace labeled pool of cDNA (or 20μg) was initially diluted to 100μl H_2O, and 10μl were removed as
"Pre-biotin" sample for scintillation counting. The remaining cDNA was ethanol precipitated in the presence of NH_4OAc and 20μg glycogen, resuspended in 10μl 1mM EDTA, pH8.0 and mixed with 10μl 1mg/ml Photobiotin (see above). Scaled down photobiotinylation was carried out as described above and 10% of the remaining aqueous samples were removed as "Post-biotin" samples after each round of biotinylation. Efficiency of biotinylation was determined according to the percent of the starting material remaining in the aqueous phase after extraction.

Long Hybridization: 2.5μg tracer cDNA was mixed with 50μg biotinylated driver cDNA (50μl of 1mg/ml solution, see above), 7.5μl sterile water, 6μl 3M NaOAc and 120μl 100% ethanol, incubated on ice 30 minutes, spun in microfuge at 14,000 rpm 15 minutes at 4°C, air-dried and resuspended in 10μl HE (10mM Hepes, pH 7.3, 1mM EDTA). Resuspended pellets were place in boiling water bath for 3 minutes, chilled on ice, and spun again briefly before the addition of 10μl pre-warmed 2X Hybridization buffer (1.5M NaCl, 50mM Hepes pH 7.3, 10mM EDTA, 0.2% SDS, this buffer should be pre-warmed to 68°C briefly to solubilize SDS prior to use), vortexed to mix, and spun briefly before the addition of several drops of mineral oil to cover the solution and prevent evaporation. Samples were then heated for 3 minutes at 100°C and transferred to 68°C for 20 hours.

Removal of biotinylated (single and double stranded) molecules: After hybridization was complete, 130μl HE (pre-warmed to 55°C) was added to the 20 μl hybridization and incubated at 55°C for 5 minutes. The aqueous phase was then transferred to a fresh 1.5 ml tube and allowed to cool to room temperature whereupon 15μl of 2mg/ml Streptavidin (Gibco-Bethesda Research Laboratories), in 10mM Hepes, 1mM EDTA, 0.15M NaCl) was added, the solution was vortexed gently to mix, spun briefly and incubated at room temperature 20 minutes. After streptavidin complex formation the solution was extracted once with phenol:chloroform:isoamyl-alcohol and the aqueous phase was transferred to a fresh tube. 10μg streptavidin was then added, mixed and incubated 20 minutes, and phenol:chloroform:isoamyl-alcohol extracted as above. This procedure was repeated twice more with 10 μl streptavidin and after the final phenol extraction the aqueous phase was extracted once with chloroform:isoamyl-alcohol, (15μl were removed, diluted to 200μl TE and 1-3μl was used as template for PCR amplification to generate +/-1,3,5,etc..(odd numbered) cDNA pools),and the remainder of the cDNA was used immediately for the subsequent short hybridization (following the 8^{th} long hybridization, which resulted in the

+/-17 pool, all subsequent "Long hybridizations" were only incubated for 2 hours, see results and figure 4E).

Short hybridization: The non-biotinylated cDNA remaining in the aqueous phase after long hybridization (~180µl) was mixed with 25µl of 1mg/ml biotinylated "+1" or "-1" cDNA and ethanol precipitated in the presence of 0.3M NaOAc on ice for 30 minutes, spun 15 minutes at 4°C, and air dried. pellets were resuspended in 10µl pre-warmed HE, boiled and covered with mineral oil as for Long hybridizations and incubated for 2 hours at 68°C. Removal of biotinylated driver and duplex cDNAs was carried out as above and the resultant cDNA was ethanol precipitated in the presence of 0.3M NaOAc and 10µg glycogen, washed once in 70% ethanol, air dried resuspended in 70µl TE and ≤1µl was used as template in PCR amplifications to generate cDNA material for the subsequent subtractions.

Slot Blot hybridization: DNA samples (from 4 nanograms for purified PCR amplified library inserts, up to 5 µg cDNA) were diluted to a final volume from 200-400µl in TE. One tenth volume 3M NaOH was added and samples were incubated at 65°C 15 minutes to denature. Reactions were cooled to room temperature and neutralized by the addition of an equal volume of 2M NH_4OAc, pH 7.0 and applied directly to Hybond N filter cut to fit Schleicher and Schuell Minifold II® slot blotter according to manufacturer's instructions (Schleicher and Schuell). Filters were cross-linked in the Stratalinker (Stratagene) and pre-hybridized for at least 1 hour in 5X SSCP, 5X Denhardt's, 0.1% SDS, 0.2mg/ml sheared herring sperm DNA, followed by incubation overnight in the same buffer with 0.5-1 x 10^6 Cpm/ml labeled probe according to standard Southern blot procedures (Sambrook et al., 1989).

EcoRI digestion of +21SH cDNA pool was carried out as follows: +21SH cDNA was amplified after subtraction of +20SH against -1 biotinylated driver with an added extension time of 1 hour at 72°C after the last cycle to ensure that the majority of the PCR products were full length. After purification of the resulting PCR products, 40µg of cDNA was digested in a reaction volume of 100µl with 4000 units Eco RI in EcoRI 10X restriction buffer (New England Biolabs) for 3 hours at 37°C, phenol:chloroform:isoamyl-alcohol extracted, ethanol precipitated, resuspended in a TE and quantified.

Ligation of "+21SH cDNA" library into pSP72: 90 nanograms of EcoRI-digested, +21SHcDNA was mixed with 600 nanograms EcoRI-digested, Calf-Intestinal phosphatase

treated pSP72 (Promega) in a volume of 30µl and ligated overnight at 16°C (Sambrook et al., 1989) .

Two sets of transformations were performed: in one case, 2µl of the resultant plasmid library was diluted to a final volume of 10µl in TE and 5µl of this diluted ligation were transformed into 100µl super-competent cells (Gibco-Bethesda Research Laboratories) according to manufacturer's instructions. In the second case, 1µl of the 1:5 dilution of the ligation from above was added to 100µl supercompetent cells in a final volume of 5 µl, and treated exactly as described above. Heat shocked cells were allowed to recover at 37°C for 1 hour after the addition of 900µl SOC (2% (w/v) bactotryptone, 0.5% yeast extract, 2.5 mM KCl, 10 mMNaCl, 20 mM Mg^{2+}, 10 mMCl^- , 10 mMSO_4^-,20 mM Glucose, pH 7.0) and plated on 7, 9 cm LB-agar plates containing 100µg/ml ampicillin for a total of 14 plates and 1418 colonies. Replicas were made on Hybond N (Amersham) filters and prepared for hybridization as described in Sambrook (1989). Filters were pre-hybridized in 6X SSC, 10X Denhardt's, 1% SDS, 100µg/ml sheared herring sperm DNA, for two hours, then incubated in the presence of 6X SSC, 5X Denhardt's, 1% SDS, 100µg/ml sheared herring sperm DNA with 0.5 x 10^6 cpm/ml of α^{32}P]dCTP-labeled probe synthesized from 50 nanograms of purified +21SH cDNA as described above for Northern blot probes, in 15 mls overnight at 65°C. The filters were washed 3 times at room temperature in 2X SSC (Sambrook et al., 1989) and exposed to XAR-5 (Kodak) over-night. The filters were then stripped by incubation in 0.4M NaOH, 30 minutes at 65°C, followed by 30 minutes in 0.2M Tris pH 7.5, 0.1X SSC, 0.1% SDS at 65°C, and hybridized with probe derived from -21SH probe as described above. Differentially expressed recombinants were identified on the filters by aligning autoradiograms from exposures to both probes. 146 colonies were picked on the basis of their strong hybridization to the +21SH but not to the -21SH probes and used to start 2ml overnight cultures in LB broth containing 100µg/ml ampicillin. Miniprep plasmid DNA was prepared from each clone individually by alkaline lysis according to standard protocols (Sambrook et al., 1989).

Sequence analysis: Plasmid DNA prepared as described above was sequenced from both ends using primers designed to flanking vector sequences in the pSP72 polylinker by the dideoxynucleotide strategy (Sambrook et al., 1989), using modified phage T7 DNA polymerase (Sequenase, United States Biochemical, Cleveland).

Single stranded digoxygenin-labeled RNA probes were synthesized from both strands by using a Boehringer Mannheim Genius kit and either T7 or SP6 RNA polymerase after

digestion with sites flanking inserts in the pSP72 polylinker to linearize template. Probes were made according to manufacturer's instructions (Boehringer Mannheim). Digoxygenin-labeled riboprobes were used in whole mount *in situ* hybridizations to fixed embryos (Tautz and Pfeifle, 1989) with modifications as described in Gavis and Lehmann 1992 (Gavis and Lehmann, 1992).

Library inserts were amplified by PCR using the following primers flanking the EcoRI site in the pSP72 polylinker: BSKSFWD: 5'GATCCCCGGGTACCGAGCTCG3' and BECREV: 5'GCAGATCTGATATGATCGATG3'. Amplification reactions were as described above, in which 1µl of a 1:50 dilution of each miniprep DNA sample was added as template. The PCR reactions were then purified from 1.2% agarose/1XTBE gels by excising bands, freezing in liquid nitrogen, thawed gel slices were then spun at 14,000 rpm 2-5 minutes in "spinex" columns (Integrated Separation Systems). DNA was then ethanol precipitated with 0.3M NaOAc and 20µg glycogen on ice for 15 minutes, washed once in 70% ethanol air-dried and resuspended in a final volume of 50µl TE.

References:

Alt, F. W., R. E. Kellems, J. R. Bertino, and R. T. Schimke. 1978. Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. *The Journal of Biological Chemistry.* 253:1357-1370.

Anderson, D. M., G. A. Galau, R. J. Britten, and E. H. Davidson. 1976. Sequence complexity of the RNA accumulated in oocytes of Arbacia punctulata. *Dev. Biol.* 51:

Atschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

Axel, R., P. Feigelson, and G. Schultz. 1976. Analysis of the complexity and diversity of mRNA from chicken liver and oviduct. *Cell.* 7:247-254.

Bantle, J. A., and W. E. Hahn. 1976. Complexity and characterization of polyadenylated RNA in the mouse brain. *Cell.* 8:139-150.

Bautz, E. K. F., and B. D. Hall. 1962. The isolation of T4-specific RNA on a DNA-cellulose column. *Proc. Nat. Acad. Sci. U.S.* 48:400.

Bautz, E. K. F., and E. Reilly. 1966. Gene-specific messenger RNA: Isolation by the deletion method. *Science.* 151:328-330.

Bernardi, G. 1965. Chromatography of nucleic acids on hydroxyapatite. *Nature.* 206:779.

Bishop, J. O., J. G. Morton, M. Rosbash, and M. Richardson. 1974. Three abundance classes in HeLa cell mRNA. *Nature.* 250:199-204.

Bolton, E. T., and B. J. McCarthy. 1962. A General Method for the Isolation of RNA Complementary to DNA. *Proc. Nat. Acad. Sci.* 48:1390.

Britten, R. J., and D. E. Kohne. 1968. Repeated sequences in DNA. *Science.* 161:529-540.

Brown, D., and J. B. Gurdon. 1966. Size distribution and stability of DNA-like RNA synthesized during development of anucleate embryos of Xenopus laevis. *J. Mol. Biol.* 19:399-422.

Campos-Ortega, J. A., and V. Hartenstein. 1985. The embryonic development of Drosophila melanogaster. Springer-Verlag, Heidelberg.

Davis, R. L., H. Weintraub, and A. B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell.* 51:987-1000.

Ding, D., and H. D. Lipshitz. 1993. A molecular screen for polar-localized maternal RNAs in the early embryo of Drosophila. *Zygote.* 1:257-271.

Fargnoli, J., H. N. J., and A. Fornace. 1990. Low-ratio hybridization subtraction. *Analytical Biochemistry.* 187:364-373.

Fornace, A. J. J., and J. B. Mitchell. 1986. Induction of B2 RNA polymerase III transcription by heat shock: enrichment for heat shock induced sequences in rodent cells by hybridization subtraction. *Nucleic Acids Research.* 14:5793-5811.

Galau, G. A., R. J. Britten, and E. H. Davidson. 1974. A measurement of the sequence complexity of polysomal messenger RNA in sea urchin embryos. *Cell.* 2:9-20.

Galau, G. A., W. H. Klein, M. M. Davis, B. J. Wold, R. J. Britten, and E. H. Davidson. 1976. Structural Gene Sets Active in Embryos and Adult Tissues of the Sea Urchin. *Cell.* 7:487-505.

Gavis, E. R., and R. Lehmann. 1992. Localization of *nanos* RNA controls embryonic polarity. *Cell.* 71:301-313.

Gillespie, D., and S. Spiegelman. 1965. A Quantitative Assay for DNA-RNA Hybrids with DNA immobilized on a membrane. *J. Mol. Biol.* 12:829-842.

Hakvoort, T. B. M., A. C. J. Leegwater, F. A. M. Michiels, R. A. F. M. Chamuleau, and W. H. Lamers. 1993. Identification of enriched sequences from a cDNA subtraction-hybridization procedure. *Nucleic acids Research.* 22:878-879.

Hall, B. D., and S. Spiegelman. 1961. Sequence Complementary of T2-DNA and T2-Specific RNA. *Proc. Nat. Acad. Sci. U.S.* 47:137.

Hedrick, S. M., D. I. Cohen, E. A. Nielsen, and M. M. Davis. 1984. The isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature.* 308:149.

Hereford, L., and M. Rosbash. 1977. Number and distribution of polyadenylated RNA sequences in yeast. *Cell.* 10:453-462.

Hough-Evans, B. R., R. J. Wold, S. G. Ernst, R. J. Britten, and E. H. Davidson. 1977. Appearance and Persistence of Maternal RNA sequences in Sea Urchin Development. *Developmental Biology.* 60:258-277.

Kaufman, K. L., and G. A. Evans. 1990. Restriction endonuclease cleavage at the termini of PCR products [published erratum appears in Biotechniques 1990 Dec; 9(6):720]. *BioTechniques.* 9:304-306.

Kohne, D. E., S. A. Levinson, and M. F. Byers. 1977. Room temperature method for increasing the rate of DNA reassociation by many thousand fold: the phenol emulsion reassociation technique. *Biochemistry.* 16:5329.

Kunkel, L. M., A. P. Monaco, W. Middlesworth, H. D. Ochs, and S. A. Latt. 1985. Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. *Proc. Natl. Acad. Sci. U.S.A.* 82:4778.

Lehmann, R., and C. Nüsslein-Volhard. 1986. Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of *oskar*, a maternal gene in Drosophila. *Cell.* 47:141-152.

Lewin, B. 1974. Units of transcription and translation: Sequence components of heterogeneous nuclear RNA and messenger RNA. *Cell.* 4:77-95.

Liang, P., L. Areboukh, and A. B. Pardee. 1993. Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. *Nucleic Acids Research.* 21:3269-3275.

Liang, P., and A. B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science.* 257:967-971.

Lisitsyn, N., N. Lisitsyn, and M. Wigler. 1993. Cloning the differences between two complex genomes. *Science.* 259:946-951.

Magyar, A., and A. Varadi. 1990. Molecular cloning and chromosomal localization of a sarco/endoplasmic reticulum-type Ca2+-ATPase of Drosophila melanogaster. *Biochemical and Biophysical Research Communications.* 173:872-877.

Miyazawa, Y., and C. A. J. Thomas. 1965. Nucleotide Composition of Short Segments of DNA Molecules. *J. Mol. Biol.* 11:223-237.

Nygaard, A. P., and B. D. Hall. 1964. Formation and Properties of RNA-DNA Complexes. *J. Mol. Biol.* 9:125-142.

Palazzolo, M. J., D. R. Hyde, K. VijayRaghavan, K. Mecklenburg, S. Benzer, and E. Meyerowitz. 1989. Use of a new strategy to isolate and characterize 436 Drosophila cDNA clones corresponding to RNAs detected in adult heads but not in early embryos. *Neuron.* 3:527-539.

Perlman, S., and M. Rosbash. 1978. Analysis of Xenopus leavis ovary and somatic cell polyadenylated RNA by molecular hybridization. *Developmental Biology.* 63:197-212.

Porath, J., J. Carlsson, I. Olsson, and G. Belfrage. 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature.* 258:598-599.

Rongo, C., E. R. Gavis, and R. Lehmann. 1995. Localization of oskar RNA regulates oskar translation and requires oskar protein. *Development.* in press:

Rudnicki, M. A., T. Braun, S. Hinuma, and R. Jaenisch. 1992. Inactivation of myoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell.* 71:383-390.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sargent, T. D., and I. B. Dawid. 1983. Differential gene expression in the gastrula of Xenopus leavis. *Science.* 222:135-139.

Sive, H. L., and T. St. John. 1988. A simple subtractive hybridization technique employing photoactivatable biotin and phenol extraction. *Nucleic Acids Research.* 16:10937-10938.

Soares, M. B., M. Bonaldo, P. Jelene, L. Su, L. Lawton, and A. Efstratiadis. 1994. Construction and characterization of a normalized cDNA library. *Proceedings, National Academy of Sciences, USA.* 91:9228-9232.

Sonnenblick, B. P. 1941. Germ cell movements and sex differentiation of the gonads in the Drosophila embryo. *Proc.Natn.Acad.Sci.USA*. 26:373-381.

Tautz, D., and C. Pfeifle. 1989. A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. *Chromosoma*. 98:81-85.

Thummel, C. S., A. M. Boulet, and H. D. Lipshitz. 1988. Vectors for Drosophila P element mediated transformation and tissue culture transformation. *Gene*. 74:445-456.

Timberlake, W. E. 1980. Developmental gene regulation in Aspergillus nidulans. *Developmental Biology*. 78:497-510.

Walker, P. M. B., and A. McLaren. 1965. Fractionation of Mouse Deoxyribonucleic Acid on Hydroxyapatite. *Nature*. 208:1175.

Wang, C., and R. Lehmann. 1991. *Nanos* is the localized posterior determinant in Drosophila. *Cell*. 66:637-648.

Wang, Z., and D. D. Brown. 1991. A gene expression screen. *Proceedings of the National Academy of Science USA*. 88:11505-11509.

Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, and R. H. Rupp S. 1991. The myoD gene family: nodal point during specification of the muscle cell lineage. *Science*. 251:761.

Welcher, A. A., A. R. Torres, and D. C. Ward. 1986. Selective enrichment of specific DNA, cDNA and RNA sequences using biotinylated probes, avidin and copper-chelate agarose. *Nucleic Acids Research*. 14:10027-10044.

Wetmur, J. G., and N. Davidson. 1968. Kinetics of renaturation of DNA. *J. Mol. Biol.* 31:349-370.

Zalokar, M. 1976. Autoradiographic Study of Protein and RNA Formation during Early Development of Drosophila Eggs. *Developmental Biology*. 49:425-437.

Zalokar, M., and I. Erk. 1976. Division and Migration of Nuclei during Early Embryogenesis of *Drosophila melanogaster*. *J Microsc. Biol Cell*. 25:97-106.

Zebrowski, R., J. Thorman, T. Norred, and R. Hurst. 1994. Isolation and identification of rare and differentially expressed genes using subtractive hybridization. *Analytical Biochemistry*. 222:285-287.

Zeng, J., R. A. Gorski, and D. Hamer. 1994. Differential cDNA cloning by enzymatic degrading subtraction (EDS). *Nucleic Acids Research*. 22:4381-4385.

Chapter 3

Screening and analysis of germ cell-specific cDNAs:

Introduction:

Using subtractive hybridization I generated a cDNA library enriched for sequences present in the RNA of embryos with pole cells and whose expression was reduced or absent in agametic embryos. The final subtracted library pool contained at least nine different cDNAs. Sequencing each of these clones revealed that one of them encoded the dorsal-ventral patterning gene, *Toll*, while the other eight were derived from unknown genes. Full length cDNAs were isolated for seven of the eight novel cDNAs and analyzed further.

The most stringent test to determine whether these newly isolated genes were germ cell-specific was to examine the expression patterns of their corresponding RNAs in embryos. Among the seven full-length cDNAs, two (called "ME and MA"), showed specific RNA expression in the embryonic gonads. One cDNA, (called "MH"), although not expressed in the embryonic germ cells, showed a striking tissue-specific pattern of RNA expression in the amnioserosa and somatic musculature of the developing embryo (further analysis of this clone is presented in Appendix I). The remaining five cDNAs showed no compelling germline or other discernible tissue-specific expression and were not analyzed further.

The next task was to determine what role the two gonad-specific genes play in the development of the embryo, and in particular, whether they are required for the differentiation of the embryonic germ cells. The first step towards determining their functions was to sequence full-length cDNAs and determine whether they encode proteins with compelling similarity to proteins of known structure or function. Both gonad-specific cDNAs as well as the amnioserosa-specific clone were fully sequenced and their encoded open reading frames determined. I then raised antibodies against bacterially-expressed MA and ME proteins to examine the expression of the endogenous proteins by protein blot as well as on whole-mount embryo preparations.

In situ hybridizations to wild-type and deficiency-carrying salivary gland chromosomes were conducted to identify the cytological map positions of ME and MA. By this method, deficiencies uncovering ME as well as lethal P-element enhancer-trap insertions in the

genomic intervals surrounding these two genes were identified. Such transposable elements will be useful for generating new deficiencies uncovering these two genes.

Results:

Cloning Full length cDNAs:

In situ hybridizations to whole mount embryos using RNA probes generated from the eight original 200 to 500 base-pair cDNA fragments were inconclusive because no specific staining patterns were discernible above background. Control experiments performed in parallel, using 200 and 500 nucleotide probes derived from the *nanos* cDNA (see materials and methods), revealed that these short probes only gave strong hybridization to *nanos* RNA in early embryos, when this RNA is most abundant. In later embryos when *nanos* RNA is present only in the pole cells, none of the specific staining normally detected with probes of greater than 1kb was visible. It is therefore not surprising that the initial *in situ* hybridizations using the library clones were inconclusive. Therefore I set out to clone full length cDNAs for each of the eight novel clones.

The original full length "+" cDNA pool was cloned into a bacteriophage lambda vector to make the unsubtracted +cDNA library, of which 400,000 plaques were screened (see materials and methods). Larger or full-length clones were isolated for five of the eight clones from this library. Subsequent screening of 528,000 colonies from a pre-existing 4-8 hour embryonic cDNA library (Brown and Kafatos, 1988), resulted in the isolation of multiple cDNAs for seven of the eight unique clones (full length clones for MB were not detected in any existing libraries, probably due to exceptionally low abundance). PCR analysis with clone-specific primers confirmed the identity of the cDNAs isolated from full-length libraries (materials and methods). Finally, the full-length clones were used to make digoxygenin-labeled RNA probes for *in situ* hybridizations to whole mount embryos. Two of the clones showed gonad-specific expression during late embryogenesis and the results for each are described below. The amnioserosa-specific clone is described in Appendix I.

Analysis of MA:

Expression of MA RNA throughout embryogenesis:

The expression pattern of the MA RNA in wild-type embryos is shown in **figure 1**. This RNA is expressed throughout the syncytial embryo at pole bud formation (embryonic stage 3 (Campos-Ortega and Hartenstein, 1985)) (**figure 1A**). The pole buds can be seen at the posterior of this embryo (arrow). This uniform RNA distribution is maintained during pole

cell formation, (**figure 1B**), and the RNA can be seen in the pole cells (see arrow). At stage 5, just before cellularization of the somatic nuclei, the transcript becomes more concentrated in the cortical layer beneath the cellularizing nuclei (**figure 1C**). By the time the germ band begins to elongate, during embryonic stage 8, the transcript is much less abundant in the embryo as a whole, although it is still detectable in the pole cells and surrounding cells of the posterior midgut invagination (arrow), as well as in the cells undergoing ventral (large arrow head) and cephalic (small arrow head) furrow formation (**figure 1D**). However, because these are highly folded cell-layers, the abundance of the transcript may appear deceptively high in these tissues. By stage 11, in the fully extended germband embryo (**figure 1E**), the RNA level appears to increase again, presumably due to zygotic transcription, and is visible in the developing central nervous system (CNS) (arrow) and in the cells of the hindgut and posterior-midgut invagination (arrowhead). Figure **1F** shows the RNA distribution pattern at stage 16, roughly fourteen hours after egg-laying (AEL). High levels of expression are present in the supra- and suboesophageal ganglia of the CNS (small arrow heads), as well as in the ventral nerve chord (large arrowhead). In the dorsal, posterior domain, the embryonic gonads can be seen to express high levels of the MA RNA (arrow). **Figure 1G** shows a dorsal view of a similarly-staged embryo (gonads are indicated by arrows). **Figure 1H** shows high magnification (63X) of a single stage 16 embryonic gonad. Expression of the MA RNA appears to be in the cytoplasm of the germ cells. The germ-cell-specific expression of this RNA was confirmed by *in situ* hybridization to agametic embryos (laid by *oskar*$^{301/CE4}$-mutant females), in which strong CNS expression was detectable, but no gonad staining was observed (data not shown). In addition, *in situ* hybridizations with sense strand controls give no discernible staining (above background) on whole mount embryo preparations (data not shown).

MA in situ Hybridizations

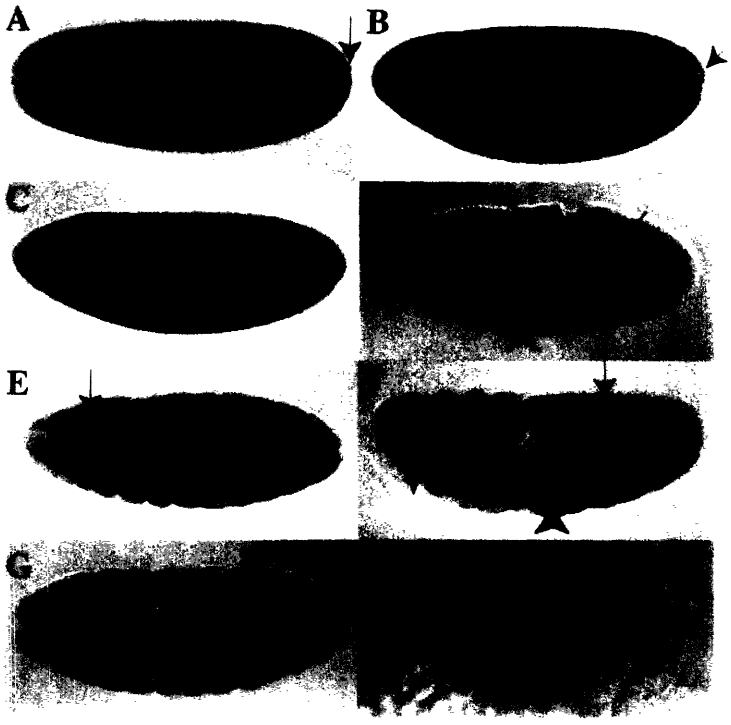


Figure 1. MA RNA expression in embryos: **A.** Stage 3: MA RNA is uniformly distributed in early cleavage-stage embryos. The RNA is incorporated into the pole buds when they form (arrow). **B.** Stage 4: Uniform RNA expression is maintained through pole cell formation. The RNA can be detected in the fully formed pole cells (see arrow). **C.** Stage 5: Just before cellularization of the somatic nuclei, the MA transcript becomes more concentrated in the cortical layer beneath the cellularizing nuclei. **D.** Stage 8: During germband extension the level of MA RNA in the embryo as a whole decreases but remains concentrated in the invaginating mesoderm (large arrow head), the posterior midgut pocket (possibly including the pole cells (arrow)), and slighlty enriched in the cephallic furrow (small arrow head). **E.** Stage 11: At full germband extension (dorsolateral view) the RNA level increases again and is expressed in the developing CNS (arrow), and in the cells of the invaginating hindgut and posterior midgut (arrow head). **F.** Stage 16: High levels of expression are present in the supra- and suboesophageal ganglia of the CNS (small arrowheads), as well as in the ventral nerve chord (large arrow head). The germ cells of the embryonic gonads, located in the dorsal, posterior region of the embryo express MA RNA strongly at this stage (arrow). **G.** Stage 16: (dorsal view) Embryonic gonads are indicated by arrows. **H.** 63X magnification of a single gonad of the embryo in **G.**

The MA gene is differentially expressed:

Aliquots of the original pools of mRNA used to make the "+" and "-" cDNA populations were examined by RNA blots with a mixed probe containing random-hexamer-labeled DNA from the original MA library insert as well as the rp49 gene as a control (see materials and methods). The result is shown in **figure 2**. The MA gene appears to undergo multiple splicing events to produce at least four different transcripts of approximately 3.9, 2.8, 2.2, and 2.1 kb. The 2.8 kb mRNA alone appears to be differentially expressed, being enriched 6-8 fold in "+" versus "-" RNA pools. Comparison of the hybridization signals for MA and rp49 reveals that the MA transcripts are relatively abundant, compared to the ubiquitously expressed rp49 RNA, and that the 2.8 kb MA transcript is indeed differentially expressed between the *oskar* and wild-type pools.

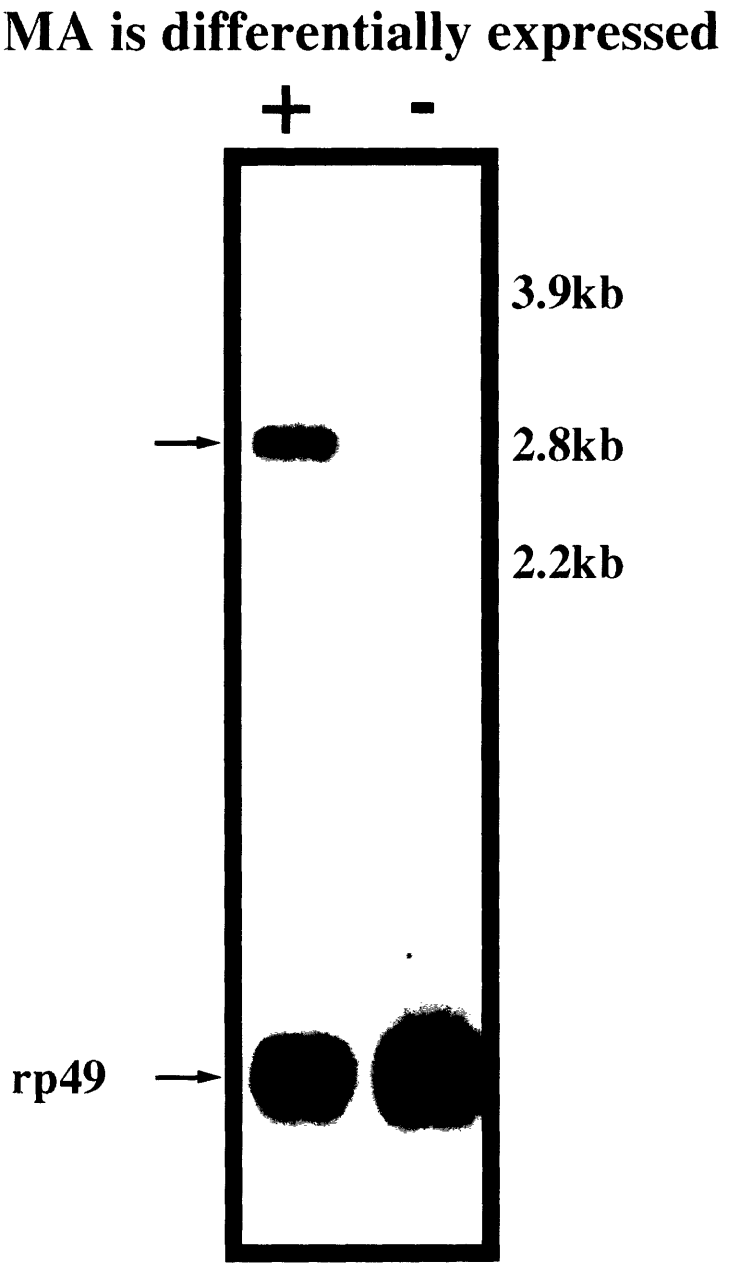


Figure 2. MA RNA distribution between the two starting RNA pools. The same RNA blot probed in figures **2A**, **2B**, and **2C** of chapter 2 was stripped and re-probed with labeled DNA derived from the original MA cDNA clone isolated from the+21SH subtracted library. For comparison, labeled *rp49* cDNA was included in the hybridization mixture to serve as a control for loading and as a reference to compare the relative abundance of the MA cDNA to that of a known, abundantly expressed RNA.

MA encodes a novel protein:

A cDNA of the size corresponding to the 2.8 kb, differentially expressed transcript was sequenced using the random-sonnication method of Bankier et al. (1987). The sequence data are presented in Appendix II. In addition, cDNAs of 1.3 , 2.1, 2.8, 3.5 and 3.9 kb, were isolated from the two cDNA libraries screened. In order to isolate a cDNA of the size corresponding specifically to the 2.8 kb transcript seen on the Northern, colonies that gave positive hybridization signals in the primary screen of 400,000 colonies with the original MA +21SH library-insert probe, were screened by PCR using the two internal, sequence-specific primers, with two vector-specific primers from either side of the cDNA inserts (see materials and methods). The MA cDNA sequenced was 2,836 bp (see **figure 3**). The sequence obtained contains 10 potential initiation methionines (highlighted in blue), all continuous with the same open reading frame that begins with base pair 199 and extends to base pair 1767. This open reading frame encodes a protein of 523 amino acids with a predicted mass of 57,902 daltons. The sequence is rich in proline residues (8%) but shows no homologies, discernible structural folds.

Bacterial expression of MA protein:

The predicted coding sequences of MA, beginning with the third in-frame methionine (see **figure 3**), were expressed as a fusion protein with glutathione S-transferase (see materials and methods). After solubilization of inclusion body material in urea followed by dialysis, soluble GST-MA fusion protein was used to raise rabbit polyclonal antisera. The resultant antibodies specifically detect bacterial-expressed fusion protein and embryonic extract-derived proteins at dilutions as high as 1:20,000 on protein blots (data not shown). The antiserum reacts with two sets of doublets in 2-14 hour embryo extracts, one cluster from ~72-76 kilodaltons and a second cluster at ~65-68 kDa. However, depending upon the blot conditions, the 65 kilodalton band seems to be the predominant species. Neither of these size ranges corresponds with the theoretical molecular weights of the MA protein derived from the cDNA sequence, however, the proline richness of the primary amino acid sequence may be expected to produce an observed molecular weight as much as 20 kilodaltons larger by SDS-PAGE (Carroll and Scott, 1986). It is also possible that the protein product encoded by the 2.8 kb transcript is present primarily in embryos after 14 hours of development, and/or is expressed at very low levels between 2 and 14 hours AEL. If this is the case, then the proteins detected by protein blot may be derived from the higher molecular weight RNAs detected on RNA blots (see **figure 2**).

MA Protein Sequence

MA cDNA 2836 bp **Longest ORF bp199-1767**

523 Amino acids **Predicted mass 57,902Da**

MVQMLRVVELQKILSFLNISFAGRKTDLQSRILSFLRTNLELLAPKVQEVYAQS
VQEQNATLQYIDPTRMYSHIQLPPTVQPNPVGLVGSGQGVQVPGGQMNVVGG
APFLHTHSINSQLPIHPDVRLKKLAFYDVLGTLIKPSTLVPRNTQRVQEVPFYFT
LTPQQATEIASNRDIRNSSKVEHAIQVQLRFCLVETSCDQEDCFPPNVNVKVNN
KLCQLPNVIPTNRPNVEPKRPPRPVNVTSNVKLSPTVTNTITVQWCPDYTRSYC
LAVYLVKKLTSTQLLQRMKTKGVKPADYTRGLIKEKLTEDADCEIATTMLKVS
LNCPLGKMKMLLPCRASTCSHLQCFDASLYLQMNERKPTWNCPVCDKPAIYD
NLVIDGYFQEVLGSSLLKSDDTEIQLHQDGSWSTPGLRSETQILDTPSKPAQKVE
VISDDIELISDDAKPVKRDLSPAQDEQPTSTSNSETVDLTLSDSDDDMPLAKRLPP
AKQAVASSTSNGSGGGQRAYTPAQQPQQSGDFDPFLQ.

Figure 3. MA protein sequence: The MA cDNA sequenced was 2,836 bp. The sequence contains ten potential initation methionines (indicated in blue). all continuous with same long open reading frame that begins with base pair 199 and extends to base pair 1,767 for a total of 1,568 base pairs. This open reading frame would encode a protein of 523 amino acids with a predicted mass of 57,902 daltons. Upstream of the first methionine there are nonsense codons in all three reading frames. In combination with the fact that the cDNA matches the expected size of one of the transcripts encoded by the MA gene, it appears that the cDNA is full length. In red is shown a potential nuclear localization signal, which does not fit the criteria for a classical bipartite SV40 nuclear localization signal but nevertheless could mediate nuclear localization (A. Lammond, personal communication).

Antibody staining of whole mount embryos:

The anti-GST-MA antiserum was affinity purified against the bacterial-expressed GST-MA fusion protein, followed by absorption against GST-Sepharose (see materials and methods) to remove antibodies specific to the GST portion of the fusion protein. The resultant serum was used to examine the distribution of the MA protein in whole mount embryo preparations. The results of these whole mount antibody stainings are shown in **figure 4**. MA protein appears to be in the nuclei of early syncytial embryos (**figure 4A**, arrows), and possibly visible in the oocyte nucleus and polar bodies, data not shown). As the nuclei multiply and migrate to the blastoderm cortex, MA protein is clearly evident in the pole cells at the posterior pole (**figure 4B**, arrow, and see figure **4G** for 40X magnification). **Figure 4C** shows that the protein is stably maintained at high levels in the pole cell nuclei during germband extension (long arrow). This late stage 8 embryo (roughly 3.5 hours after egg-laying (AEL)), also shows nuclear staining in the ventral neuroblasts (arrow heads), cephalic mesoderm (arrow) as well as in the proctodeal and posterior midgut primordia (small arrow heads). Meanwhile, the level of MA protein in the somatic cells of the embryo appears to decrease. At this point it is unclear whether the observed decrease in the soma is caused by dilution of this protein during cell division or whether the amount of MA protein in each nucleus actually decreases (see discussion). After germ band retraction, (**figure 4D**), in stage 13 embryos (roughly 10 hours AEL), MA protein can still be detected in the pole cell nuclei as they align with the somatic mesoderm cells on either side of the embryo (short arrow, **figure 4D**). In addition, strong protein expression is visible in the hindgut (long arrow) and supraoesophageal ganglion (arrowhead). During stage seventeen, just before hatching (**figure 4E**), the supra- and suboesophageal ganglia and ventral chord of the CNS are clearly expressing this protein, as are the germ cells in the embryonic gonads (arrow). This expression pattern matches the late embryonic RNA expression pattern seen in **figures 1F** and **G**. A dorsal view of a slightly younger embryo is shown in **figure 4F**, in which the protein can be detected in the nuclei of a number of different cell types including the embryonic gonads (arrow), the malpighian tubules (small arrowhead), the visceral mesoderm (long arrow), and the CNS. The appearance of MA protein in the nuclei can appear particulate or slightly speckled, especially in cortical blastoderm nuclei (see **figure 4H**). The appearance of MA in these nuclei is similar to that seen for the Krüppel protein, which is described as showing punctate nuclear staining only in interphase nuclei (see discussion), consistent with its role as a transcription factor. Whether MA protein actually reflects chromosomal localization during S-phase in blastoderm nuclei remains to be seen (see discussion). Additional control antibody

stainings on agametic embryos reveal that the gonad-specific staining is in fact in the germline cells of these structures (data not shown).

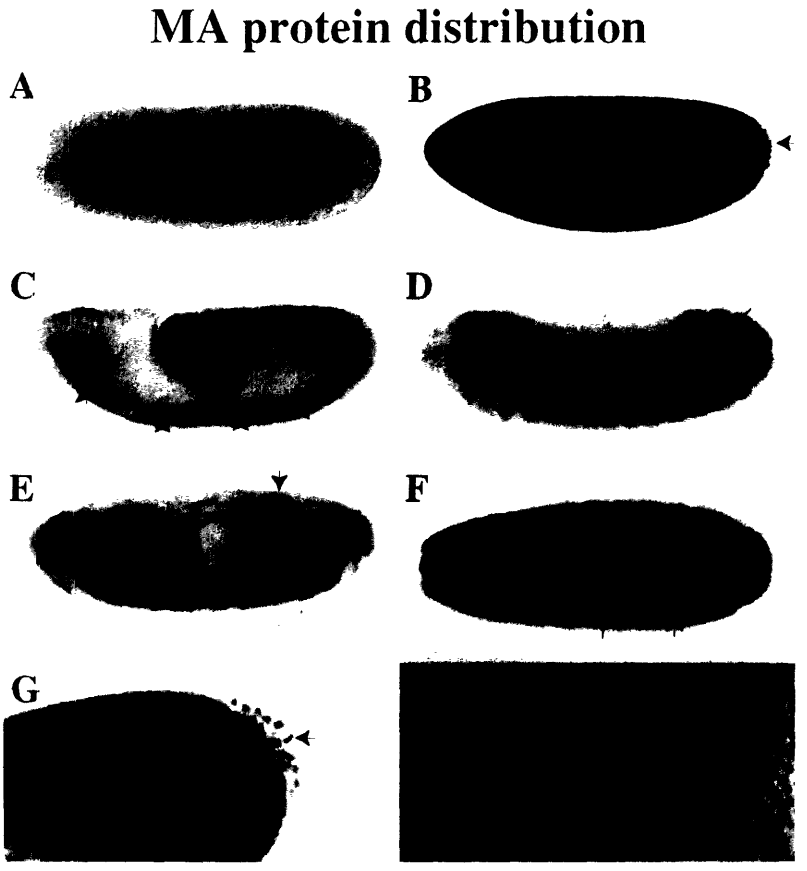


Figure 4. MA protein expression pattern during embryogenesis: **A.** Stage 2: Early cleavage nuclei show expression of MA protein (arrows). **B.** Stage 4: As the nuclei multiply and migrate to the cortex of the syncytial blastoderm, MA protein is evident in the nuclei of the pole cells at the posterior pole (arrow). **C.** Stage 8 (late): During germband extension the protein is stably maintained at high levels in the pole cell nuclei (long arrow). There is also MA protein expression in the nuclei of the ventral neuroblasts (arrow heads), cephalic mesoderm (short arrow), as well as in the proctodeal and posterior midgut primordia (small arrow heads). **D.** Stage 13: After germ band retraction MA proteinis still detectable in the pole cells (short arrow). Strong expression of the protein is also detectable in the supraoesophgeal ganglion (arrow head) and in the hindgut (long arrow).**E.** Stage 17: After germband retraction MA protein expression is detectable in the same tissues as the RNA (see **figure 1F** for comparison. Arrow indicates germ cells of the embryonic gonad). **F.** Stage 16: (dorsal view) Protein is detected in the nuclei of a number of different cell types at this stage, including the embryonic gonads (short arrow), the visceral mesoderm (long arrow), the malpighian tubules (small arrow head), and the CNS (see **figure 1G** for comparison). H. Close-up on nuclei of a cellular blastoderm stage embryo. The appearance of MA in these nuclei is similar to that seen for the Krüppel protein, which is a known transcription factor that also gives punctate nuclear antibody staining.

Cytological location of MA:

In situ hybridizations to salivary gland chromosomes reveal that MA lies in the proximal region of 2R at 45C-D. **Figure 5** shows hybridization of the MA cDNA probe to two different chromosomal squash preparations. Unfortunately, although the 45C-D region of 2R is not haploinsufficient (Lindsley and Sandler, 1972), and deficiencies in the region have been described previously, (Lindsley and Zimm, 1992), these stocks have been lost, and no new deficiencies uncovering MA have been identified so far.

MA *in situ* hybridization to wild-type salivary gland chromosomes:

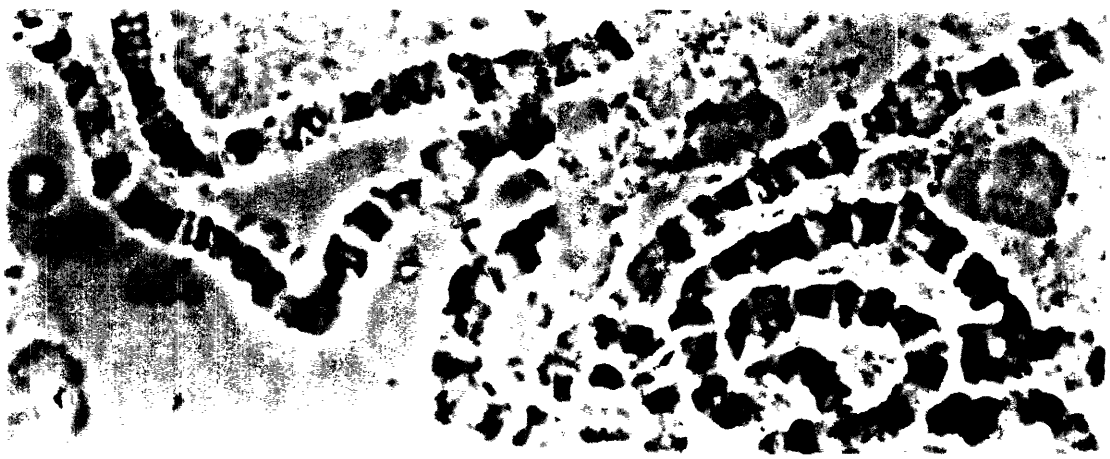


Figure 5. *In situ* hybridizations to wild-type salivary gland chromosomes reveal that MA maps to the centromere-proximal region of 2R, in the interval 45C-D. Unfortunately there are no available deficiencies that uncover this genomic region. However, there are a number of P-element insertions in the region that will be useful for isolation of deficiencies in the interval by gamma-ray and/or transposase-induced P-element excision. It may also be possible to generate P-element-induced mutations in the MA gene by local hopping of nearby P-elements into the MA transcription unit.

Analysis of ME:

Expression of ME RNA throughout embryogenesis

The RNA distribution pattern for the ME clone is shown in **figure 6**. This RNA is ubiquitously expressed in early cleavage embryos, shown at stage 2 in **figure 6A**. After pole cell formation, by early stage 4, the RNA is excluded from the pole cells (arrow) as well as the somatic nuclei (arrow head) as they begin to cellularize (see **figure 6B**). The expression pattern for the ME RNA during early germband extension resembles that for MA, (compare **figures 6C** and **1D**), although the embryo in figure **6C** is from a slightly earlier stage. At stage 11 (**figure 6D**), the RNA is enriched in the anterior (arrow head) and posterior (arrow) midgut primordia. This enrichment remains in these two tissues after germband retraction, (stage 13, **figure 6E**), as the anterior and posterior midgut primordia extend toward each other before forming a continuous tube by stage 14 (Campos-Ortega and Hartenstein, 1985) . **Figure 6F** shows a dorsal view of a stage 16 embryo, in which RNA expression is restricted to the proventriculous (arrow), possibly the salivary glands (small arrowheads), as well as the hindgut (large arrow head). Finally, the transcript appears to be expressed specifically in the embryonic gonads (arrows) as well as in the labrum (arrow head) of very late embryos, just before cuticle deposition (stage 17, **figure 6G**). A 63X magnification of the gonads is shown in **figure 6H**, where it can be seen that the RNA is expressed in the germline cells of the late embryonic gonads. As for MA, germline-specific expression of ME was confirmed by *in situ* hybridization to embryos laid by $osk^{301/CE4}$ mutant females. In these agametic embryos, although maternally supplied transcripts were clearly present, gonad-specific RNA expression was not detectable (data not shown). As for MA, further control hybridizations using sense-strand probes show no specific staining above background.

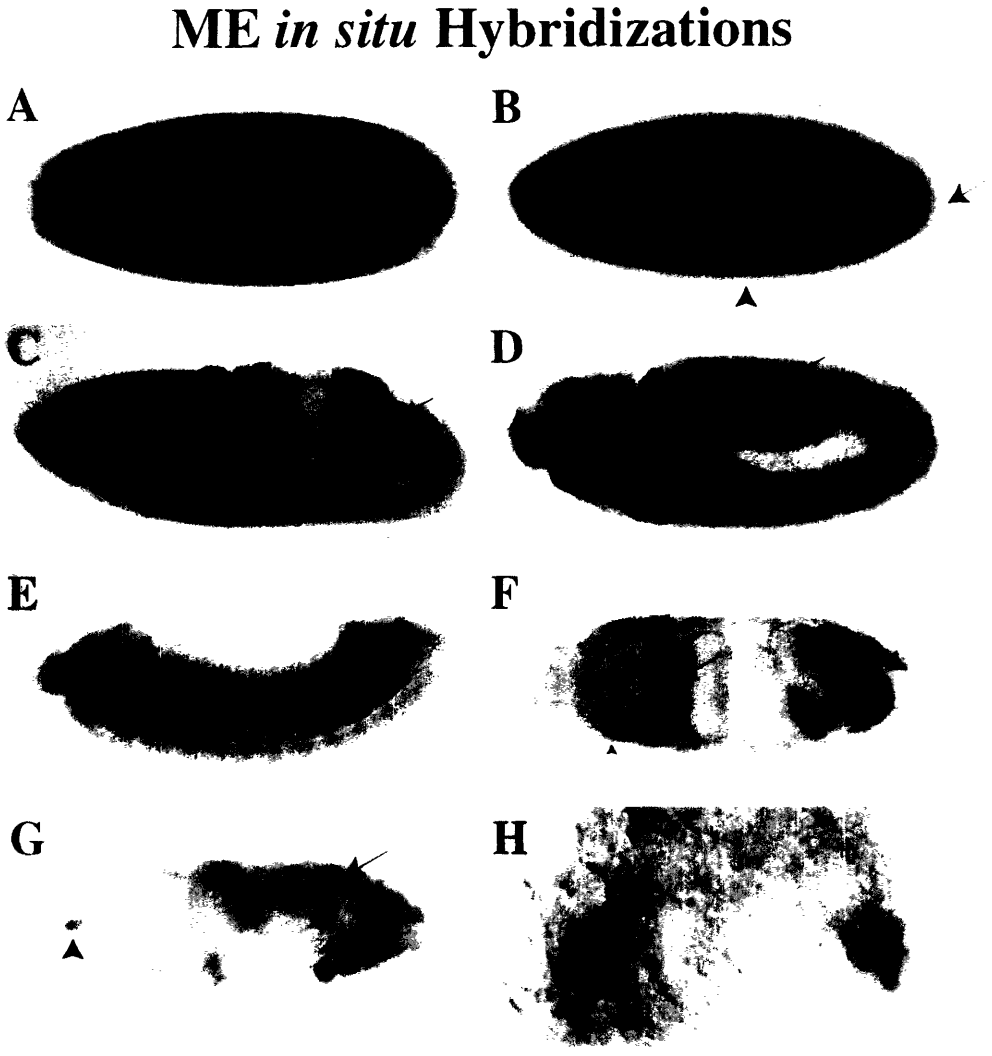


Figure 6. ME RNA expression in embryos: **A**. Stage 1: ME RNA is uniformly distributed in syncytial blastoderm stage embryos. **B.** Early stage 4: By the time pole cells form the ME RNA has been excluded from both somatic and pole cells. **C.** Early stage 8: The level of ME RNA decreases throughout the embryo during germband extension, although the infolded tissues of the cephalic furrow (large arrow), the posterior midgut invagination (small arrow), and the invaginating mesodermal primordia (arrow head) appear to express the RNA more strongly than the rest of the somatic cells. **D.** Stage 11: During germband extension the RNA becomes enriched in the anterior (arrow head) and posterior (arrow) midgut primordia. **E.** Stage 13: The enrichment in the gut primordia continues after germband retraction. **F.** Stage 16: This embryo is shown in dorsal view. ME expression is present in the hindgut (large arrow head), the proventriculus (arrow) and in the salivary glands (small arrow heads). **G.** Stage17: Finally, the RNA expression becomes restricted to the germ cells of the embryonic gonads (arrows) and the labrum (arrowhead). H. A 63X magnification of the embryonic gonads shown in **G**.

ME mRNA is differentially expressed

The number and distribution of transcripts seen for ME is shown in **figure 7**. This blot shows that there are two transcripts encoded this gene, one of 6.5 kb, that appears to be equally distributed between wild-type and *oskar* mutant RNA pools, and another RNA of 5.5 kb that is approximately 12-fold more abundant in RNA from wild-type embryos as compared to RNA from embryos lacking germ cells.

ME RNA is differentially expressed

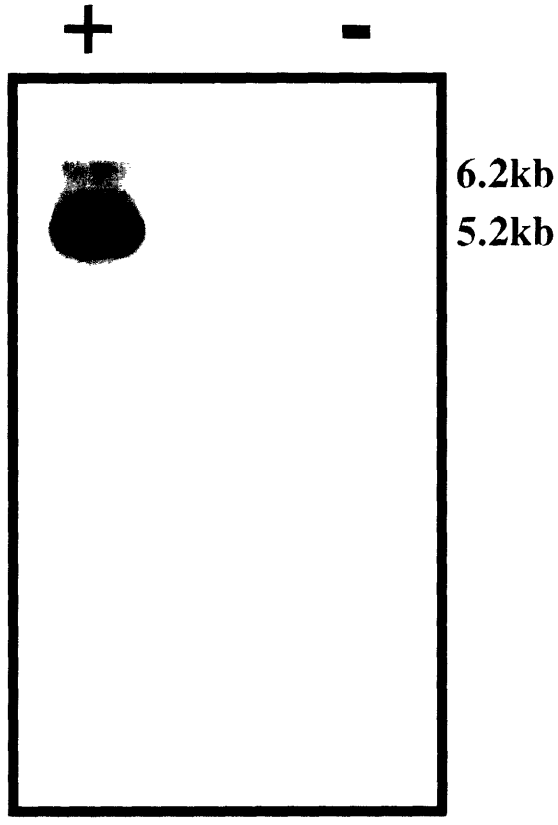


Figure 7. The number and distribution of transcripts encoded by ME is shown on the original RNA blot, stripped and re-probed with +21SH library ME clone. Two transcripts are detected with all ME cDNA probes used. The smaller 5.2 kb transcript appears to be differentially expressed between the starting RNA pools. There appear to be at least two splice forms of the ME RNA, but the longest cDNA recovered was roughly 300 nucleotides shorter than the differentially expressed transcript, and therefore it is not known what the differences are between the two splice forms.

Sequence of ME cDNA:

The longest ME cDNA cloned was 4,924 bp. This cDNA was isolated by PCR screening of primary colony pools hybridizing to a 1.2 kb N-terminal probe derived from a 3.5 kb cDNA (see materials and methods). Thirteen other cDNAs identified in the same library included clones of 1.8 kb, 3.3 kb, 3.5 kb, 3.6 kb, 3.8 kb, 4.2 kb, and 4.3 kb. Although the 4929 bp cDNA clone is probably not full length (as evidenced by the size of the transcripts on the RNA blot), sequencing this clone (see **figure 8** and Appendix II) revealed that the longest open reading frame was from base pair (bp) 44 to 3559: or 3516 bp, encoding 1172 amino acids for a predicted mass of 128,195 daltons. In total there are 19 in-frame methionines (highlighted in blue) that would initiate proteins of predicted molecular weights ranging from 17 kilodaltons up to the largest open reading frame of 128 kilodaltons.

The library from which this clone was isolated was generated by oligo dT-primed first strand synthesis (Brown and Kafatos, 1988), and therefore, as is common for cDNAs from this library, this clone contains a polyA tail, although it does not appear to contain all of the sequences of the 5'UTR. The missing ~300 bases (as estimated by RNA blot) could, in principle, contain additional coding sequences, since the open reading frame starts with the first codon of the cDNA. However, identification of these upstream sequences awaits future cloning of the genomic DNA encoding the transcript (see Epilogue).

The ME sequence is rich in glutamine and proline residues (8% for each) but contains no obvious structural folds. Three regions in the C-terminus of the protein, indicated in **figure 8**, show 34 percent identity, and 59% similarity over a total of 290 amino acids to the Ubiquitin-specific protease III (Ubp3) enzyme of Saccharomyces cerevisiae (outlined in green) see **figure 8B** and discussion).

ME Protein Sequence

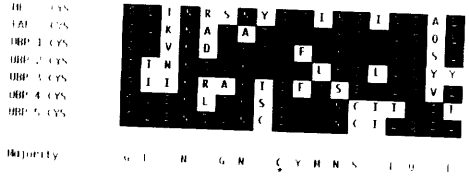
ME cDNA 4,924 bp **LONGEST ORF bp 44-3,559**

1,172 Amino acids **PREDICTED MASS 128,195 Da**

MTQIWHQGPMYAEDFEALQQQQQQQQVAVHPNGGAVSVADELNHNSSSLPSS
ETSSMISPNYPIYDPQMHEMQHQMGVMQIFDDGQMAALQPIHPGAGPLPQYED
ELAECGAPPPGPIPVAWTAAMPLPPDAVQPALLPPPPHHILQQTSPLLIEQPAQL
VQAQPPQPPTPTSQQQQQQPHQEQTADKSLTQNNNEVVALDEKLANEKQQQID
FHQQQQEQQQQQQQQQQQLHPLAQQQQQEEQQQPQQLQPLQTEKPPTKVPN
LVVATVAPQLQQPPLQQHQQQQTSPVVVSPKQQQQQSQAAHQNSYHDQHPLP
QQQQPQQQQPQPQRKPSQQYNQAQQQQVRRKYSSEYNHHHHQHGSSDTGIQT
TKTMSWTNSAQHKKSTQSVSVTASPNSVNNGPGVGAASGPAGKANYSHTTKTF
TNQQHYQQQHYHTQNSYNNNSSGGGSSSNSSSSSSNNQPQVRNYGTMKMPSSPV
AWTAPTPERKNSQQAAAAPAAATASAAPVAITTVTTSTASITATPNQFQPETGSS
SASPAVAAAEQQQPAAQQHAQAAQTGASAAPSKSHSNYASSKKHQSYEPVVLSS
VVATSSSANPQPQLNLAPPAPPASQNSSDSSQLSWASLFASNKPVAKVAPYEASKI
AQQQPSHPVLQLAPPQPAQVSAAAPVAQLSSPPQNLPLPTAHQQSQLPAPVPAP
TAPLVTPGALSYSAASAQAVPAPSPASASVKPLKPEPPRPVQQQLDEWTSKYAE
YLTRHKTNLASISLRPRGLTNRSNYCYINSILQALLGCSPFYNLLRSIPKQAAVLS
EVKTPTVNAMMSFMTNFSSLPSGLRLRLNNLNKGSQSKGKDDFVGSDLQCDMA
FEPTEIYKLWNDSREEHVEGRQEDAEEFLGYVLNKLNDEMLEVIKLIDKPTPQQ
NGQEPAEPEDGGDVWQMICNNRNKGSVTRQTDFGRTPVSDIFRGELRSRLQRE
GEHSTDVIQPFFTLQLNIEKAASVKEALEILVGRDQLEGVTGSKTKQEVVAWQQ
MTLEKLPVVLILHLKYFDYRSDGCTKILKKVDFPVELKIDAKILGSKKTSQKQR
AYRLFAVVVYHDGKEASKGHYITDVFHTGYSSWLRYDDSSVKPVSEKHVLQPHT
PRVPYLLVYRRSDTLPPQQQQTQQQNGGGSGGVVGSSSSSSNAGDNK.

Figure 8A. The conceptual translation of the largest continuous open reading frame in the ME cDNA is shown above. Ninteen potential initiation methionines are shown in blue. It is likely that the five prime end of the protein is missing since the cDNA sequenced is roughly 300 base pairs shorter than the size predicted by RNA blot hybridization and the open reading frame begins at the first residue of the cDNA. The three regions underlined in the C-terminal portion of the protein indicate the 290 amino acids that show 34% identity and 59% similarity to the Ubiquitin-specific protease III of Saccharomyces cerevisiae (see discussion). The highly conserved Cysteine and Histidine domains are shown in green, with critical active site cysteine and histidine residues shown in red (see **figure 8B** for alignment).

Cysteine Domain:



Histidine Domain:

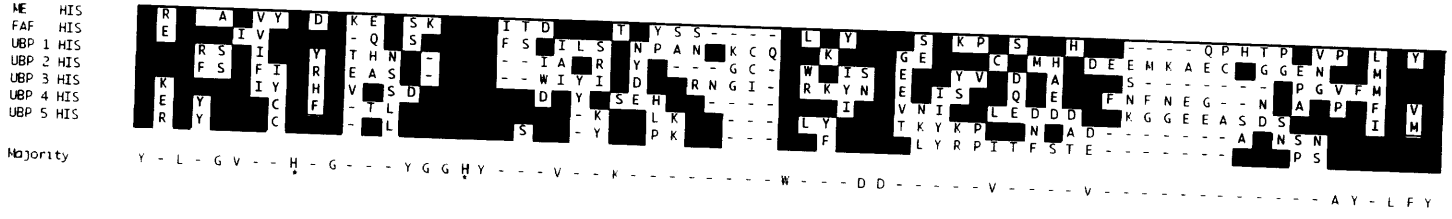


Figure 8B. Above are shown the two domains most highly conserved between members of the Deubiquitinating enzyme family. This family includes over sixteen unique sequences from a wide variety of different organisms including yeast, humans, Arabadopsis, and Drosophila. Many of these sequences consist of partial open reading frames identified by random sequencing of cDNAs corresponding to expressed RNAs in various cell types (Expressed Sequence Tags, or ESTs). Shown above are the alignments ME with five yeast deubiquitinating enzymes called Ubiquitin dependent proteases or Ubp1-5 with the only known Drosophila deubiquitinating enzyme, fat facets (faf). Highlighted in pink are the active site cysteine and histidine residues. Amino acids conserved in four out of the seven sequences are indicated in the row labeled "Majority." Ubp's 1 through 5 and faf have all been shown to have deubiquitinating function when co-expressed in E. coli with the appropriate ubiquitin protein-fusion substrates.

Bacterial Expression of ME Protein:

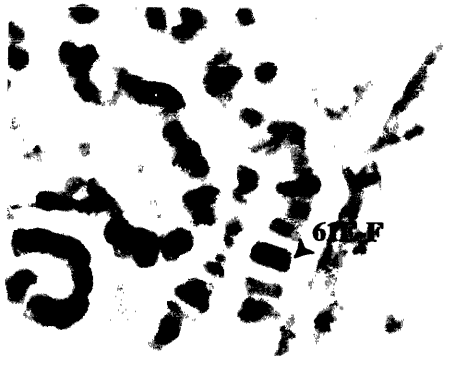
A 176 amino acid fragment of ME, from base pairs 2170-2697 was fused to glutatione S-transferase (see materials and methods) as described for MA. The fusion protein produced was not soluble under any of the conditions tried. Therefore, the over-expressed protein was purified from inclusion bodies by preparative SDS-PAGE, followed by electro-elution of the fusion protein band from the gel slice and injection into a rabbit and a Chicken (see materials and methods). The resultant antisera were affinity purified against the same source of gel-purified protein used for injection, absorbed against bacterial-expressed and purified GST alone (as described for MA antisera, see materials and methods) and tested both on protein blots and against whole mount embryos.

Although the antiserum reacted specifically on protein blots against bacterial expressed fusion protein and 2-14 hour embryo extracts (data not shown), there was no discernible specific cross reactivity against the endogenous protein by antibody staining of whole-mount embryos. The species detected in 2-14 hour embryo extracts by protein blot were two bands centered at 55kDa, and a single species of approximately 125kDa (data not shown). As in the case of the endogenous MA protein, there are multiple methionines encoded in the cDNA that if utilized for initiation of protein synthesis would lead to production of protein products in a variety of sizes, including those detected by these protein blot experiments. Coupled with the high proline content of this protein as well as the potential for post-translational modification, rigorous proof that the species being detected on protein blots are in fact encoded by endogenous ME will depend upon micro-protein sequencing of these proteins and/or blot analysis of protein extracts made from embryos homozygous for deficiencies uncovering the ME locus.

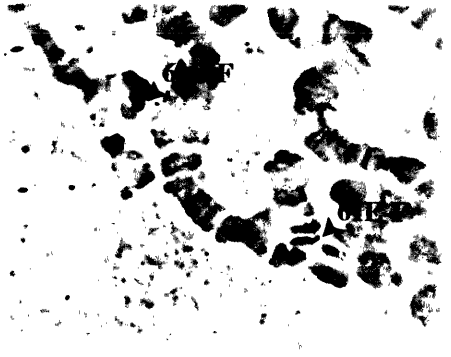
Cytological location of ME:

In situ hybridizations to wild-type salivary gland chromosomes reveal that the ME gene lies at the tip of chromosome 3L, in the interval 61E-F (see **figure 9A**). Further *in situ* hybridization analysis to deficiences within the region confirm that ME lies between 61C-3 and 61F-3 on the basis of inclusion in both Df(3L)Ar14-8 (see **figure 9B**), which spans from 61C-3 to 62A-8, and Df(3L)Ar12-1 (see **figure 9C**), which spans from 61C to 61F-3. Given that ME clearly lies within these two deficiencies, important control experiments will be conducted to demonstrate that the ME RNA expression detected by *in situ* hybridization to whole mount embryos is specific for the gene cloned.

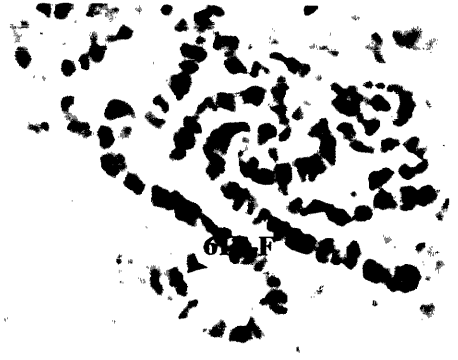
Cytological Location of ME



A. ME *in situ* hybridization to wild type chromosomes



B. ME *in situ* hybridization to: Df(3L)Ar14-8



C. ME *in situ* hybridization to: Df(3L)Ar12-1

Figure 9A. *In situ* hybridizations to wild-type salivary gland chromosomes with probes derived from the ME cDNA reveal that ME lies at the end of the left arm of chromosome 3, in the interval 61E-F. **B.** Further *in situ* hybridization analysis to a large deficiency in the region, Df(3L)Ar14-8, which spans from 61C-3 to 62A-8, confirms the initial location of ME. **C.** Df(3L)Ar12-1, which spans from 61C to 61F-3 can also be seen to contain the ME gene. Combining the data from the two deficiencies defines the chromosomal location of ME to the region between 61C-3 and 61F-3.

Given that both of these deficiencies are large and therefore contain numerous genes, the morphology of the homozygous progeny is expected to be visibly disrupted compared to their heterozygous siblings. Nevertheless, it will be helpful to use balancer chromosomes that carry a known lacZ fusion protein expressed in specific tissues during embryogenesis to allow definitive identification of embryos homozygous for the deficiency chromosome. Balancer-carrying embryos can be easily identified during the *in situ* hybridization experiment by including a *lacZ* riboprobe with the ME-specific probe. In addition, it may be advantageous to examine the ME RNA expression pattern in embryos transheterozygous for the two deficiencies (in addition to embryos homozygous for each deficiency separately) to avoid the added effects of recessive lethals and/or other disruptive mutations that may exist on the deficiency chromosomes.

Discussion:

In situ hybridization to whole mount embryos using probes from the original eight novel cDNAs revealed that two of these clones are derived from genes expressed specifically in the embryonic gonads. However, in addition to the germline-specific expression, these two genes were abundantly expressed in other embryonic tissues as well. Therefore the question was whether the original RNAs for these two clones were in fact differentially distributed between the starting "+" and "-" RNA pools. RNA blot analysis revealed that both ME and MA are in fact differentially expressed between the two starting transcript populations. Furthermore, these RNA blots revealed that the MA gene, encodes at least five different transcripts, one of which appears to be enriched at least six-fold in RNA from embryos with pole cells versus those without. The ME transcription unit, on the other hand, encodes two different RNAs, one of which appears to be enriched at least twelve-fold in wild-type relative to mutant RNA.

Given the significant levels of expression of these two transcripts in tissues other than germ cells it was initially surprising that these genes were identified in this screen. However, as the RNA blots show, there is a clear quantitative difference in the level of both of these transcripts in the wild-type RNA pool relative to the *oskar* mutant embryo RNA population. This difference appears to be sufficient to allow enrichment in the wild-type pool. In fact, any RNA that is significantly enriched in one pool relative to the other, whether this differential representation is biologically relevant or not, is likely to become further enriched during the process of subtractive hybridization. Indeed, the *oskar* transcript, which by *in situ* hybridization to whole mounts embryos appears to be equally

well expressed in wild-type and in embryos of *oskar*$^{301/CE4}$ embryos (C. Rongo, personal communication), was significantly enriched in the +14cDNA pool, due to its initial enrichment in the +0 cDNA starting pool (see chapter 2, **figure 4D**).

In the case of the MA gene there appears to be differential expression of certain exon sequences between the "+" and "-" RNA pools. Fragmentation of the cDNAs before subtractive hybridization appears to allow the enrichment of exon sequences from differentially spliced RNAs. The small (~250base pair) MA probe hybridizes to at least three transcripts, only one of which appears to be differentially expressed. Thus, even though this small cDNA fragment contains enough sequences to hybridize to all three splice forms on RNA blots, these sequences are not sufficient to cause this germ cell-specific exon to be subtracted from the pool. Apparently the requirements for hybridization in solution (and therefore for subtraction) are more stringent than those for hybridization and signal detection on slot blots. This may be due, in part, to higher non-specific binding of nucleic acids in the probe to sequences cross-linked to nylon membranes as opposed to hybridizations occurring in solution.

Sequencing full or near-full-length cDNAs for both MA and ME reveals that these two genes encode novel proteins. The MA protein has no significant similarity or homology to known proteins, however antibody staining of the MA protein in embryos reveals that it is localized to nuclei (see **figure 4**). Although parallel antibody staining reactions with pre-immune sera showed no detectable tissue-specific expression, an important further control will be to use expressed GST-MA protein-resin to absorb out the MA-specific antibodies from the sera used in the stainings described above. The expected result is that this absorption step should deplete the MA specific antibodies from the serum and therefore decrease or possibly eliminate the specific staining patterns seen with the affinity purified antibody described above.

Formal proof that the protein recognized by the affinity purified anti-MA antiserum would require that the serum be tested for cross-reactivity with embryos lacking MA protein expression. In the absence of mutants, the ability of these antibodies to recognize the primary translation product of the MA cDNA would provide additional evidence to support the specificity of this antibody for MA *in vitro*. For this experiment the cDNA would be transcribed and translated in cell free translation extracts in the presence of ^{35}S-methionine. If the antibody is specific it would be expected that the affinity purified serum, but not the pre-immune serum would be able to immunoprecipitate the labeled MA protein.

The fact that the MA protein appears to be so much stronger in the germ cells could be because they are not dividing during most of embryogenesis. This hypothesis could be tested by examining MA protein distribution in *string* mutant embryos in which the somatic cells do not divide after stage 14 (Edgar and O'Farrell, 1989). If the level of MA protein in the somatic cells of these mutant embryos appears to be more like that in the pole cells (which are unaffected by this mutation), this would suggest that the proposed model may be valid.

Although the sequence of MA does not contain any zinc fingers, HLH motifs or canonical nuclear localization signals, it is localized to nuclei and the question remains as to what kind of molecule this gene encodes. The appearance of MA protein, especially in cortical blastoderm nuclei (see **figure 4H)** is similar to the distribution seen for the Krüppel protein, a known transcription factor that shows punctate nuclear staining only in nuclei in interphase (Gaul et al., 1987) . If the punctate staining seen in embryonic nuclei stained with the anti-MA antibody is not simply an artifact of nickel-enhanced horseradish peroxidase-based detection system (see materials and methods), this staining pattern is consistent with MA protein being a transcription factor or other S-phase chromosome-associated protein.

Examination of the sub-cellular distribution of MA during the cell-cycle could be achieved in a number of ways including the use of Hoechst, or other DNA-specific dye to visualize the chromosomes, along with α–tubulin to reveal the mitotic spindle, and anti-MA antibodies. In addition, F-actin and spectrin provide excellent markers for the periodic cytoskeletal re-organizations that take place in the blastoderm embryo before and after cellularization (Karr and Alberts, 1986) . Co-localization of MA protein with chromosomes during interphase would be consistent with a role for this protein in transcriptional regulation and maintenance of higher levels of MA in germ cell nuclei than in somatic nuclei could be part of the mechanism for transcriptional silencing in these cells.

The carboxy-terminal region of ME contains three stretches, for a total of 290 amino acids (underlined), that show 34 % identity and 59 % similarity to the Saccharomyces cerevisiae Ubiquitin-specific-protease 3 (Ubp3) at the amino acid level (see **figure 8A**). Further examination of the predicted protein sequence reveals that ME contains both of the putative active site regions critical for the functioning of the deubiquitinating enzyme family. These putative active site sequences have now been identified in over sixteen different cloned and

expressed-sequence-tag open reading frames in yeast, C. elegans, Arabidopsis, humans, mouse and Drosophila (Papa and Hochstrasser, 1993) . The regions in ME that show homology to Ubp3 are underlined in **figure 8A**, with key histidine and cysteine residues highlighted in red. The more N-terminal of the two domains is the "Cysteine domain," which includes a crucial cysteine residue. The other highly conserved domain is referred to as the "Histidine domain" because of the presence of two critical histidine residues (see **figure 8B**). In addition to the overall homology to Ubp3 in the carboxyl terminus, ME contains the important cysteine and histidine residues, and therefore likely encodes a member of the eukaryotic deubiquitinating enzyme family.

The other Drosophila protein shown to contain this putative active site sequence is the *fat-facets* gene (Fischer-Vize et al., 1992; Papa and Hochstrasser, 1993) . *Fat facets (faf)* is required during oogenesis for cellularization of the early embryo and the protein product of this gene is localized to the posterior pole of the developing oocyte. In addition, faf protein appears to be a component of the pole plasm and it's posterior localization is mediated by *oskar* (Fischer-Vize et al., 1992) . Embryos laid by *fat facets*-mutant females display cellularization defects during early embryogenesis; the pole cells are reduced in number and appear to be positioned differently from those in wild-type embryos. Finally, *fat facets* is also required during eye development to prevent particular cells in the eye-imaginal disc from becoming photoreceptors (Fischer-Vize et al., 1992) .

Ubiquitin-dependent proteolysis plays a key role in many different cellular processes including regulation of gene expression and control of the cell cycle (Papa and Hochstrasser, 1993). Degradation of proteins by the ubiquitin-dependent proteolysis pathway occurs in four stages (see **figure 10**). First, the α-carboxyl group of ubiquitin is activated by formation of a high energy thioester bond. This is accomplished by enzymes in the ubiquitin activating enzyme (Uba) class in an ATP-dependent manner. Next, the activated ubiquitin molecule is transferred to a ubiquitin conjugating enzyme (Ubc), which catalyzes covalent linkage of multiple ubiquitin molecules to the protein substrate. Finally, the target protein is degraded by ubiquitin-specific proteases (Ubp's, highlighted in blue in **figure 10**). These enzymes appear to have a number of different functions in the ubiquitin-dependent protein degradation pathway (Papa and Hochstrasser, 1993) . Although all of the different roles played by deubiquitinating enzymes in the cell have yet to be discovered, these enzymes are known to be required for removal of ubiquitin polypeptides from protein targets prior to degradation. Moreover, Ubp's not only play key roles in the degradation of

specific targets by the 26S protease but are also thought to maintain intracellular stores of free ubiquitin (see **figure 10**) (Papa and Hochstrasser, 1993) .

Ubiquitin Dependent Proteolysis:

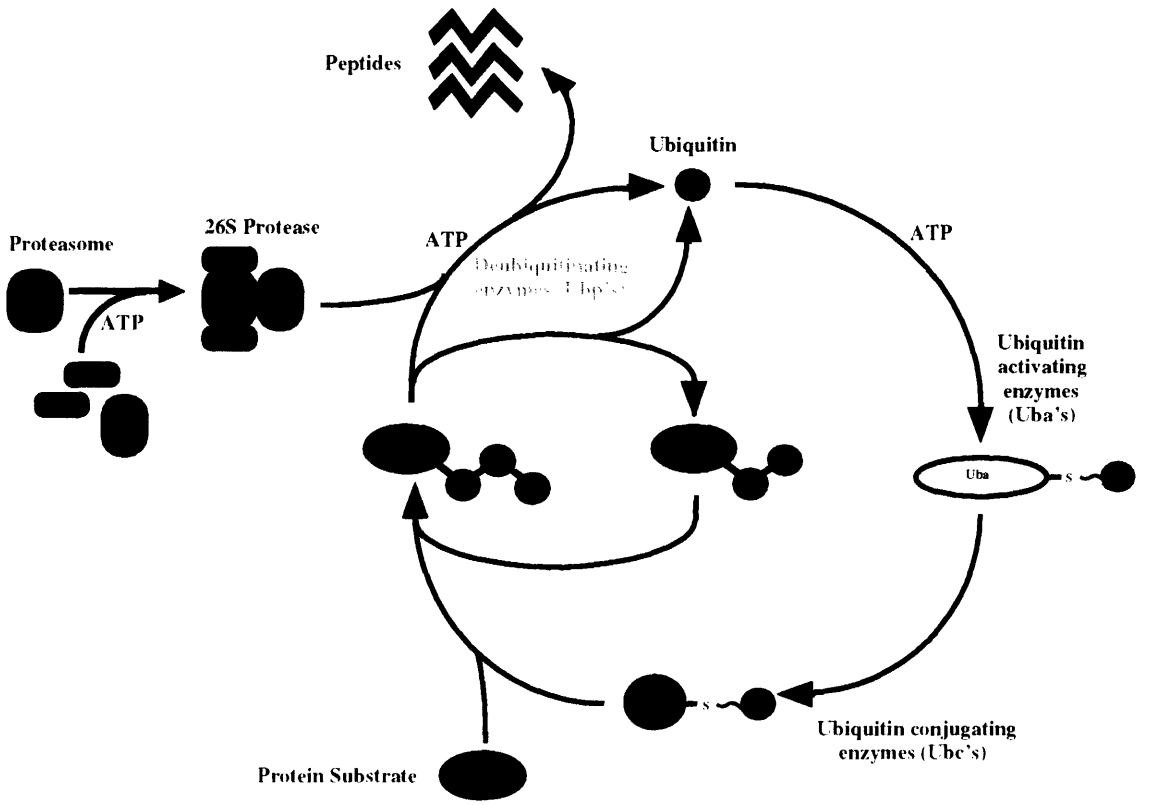


Figure 10. This figure illustrates the major features of the ubiquitin dependent proteolysis pathway in Saccharomyces cerevisiae. The role of deubiquitinating enzymes (Ubp's) such as Doa4 is to remove ubiquitin moieties from proteolytic targets prior to degradation. It is not yet known where in the pathway each of the different Ubp's act, but it has recently been shown that DOA4 co-purifies with the 26S Proteosome and appears to be a part of this enzyme complex (figure modified from Hochstrasser, 1992)

Six closely related members of this family have been isolated in Sacharromyces cerevisiae, Yuh1 and Ubp1-5 (Baker et al., 1992; Liu et al., 1989; Papa and Hochstrasser, 1993; Tobias and Varshavsky, 1991) . Ubp4 (or Doa4), has been shown to be closely related to the human tre-2 oncogene (Papa and Hochstrasser, 1993) . Analysis of tre-2 reveals that this human protein is capable of functioning as a deubiquitinating enzyme *in vitro* and mutations in the yeast homologue, Doa4 cause growth-control phenotypes similar to those caused by the oncogenic version of tre-2 in vertebrate cells. These data suggest that the ubiquitin system also plays a role in mammalian growth control (Papa and Hochstrasser, 1993) . Given that the expression of ME in germ cells correlates (roughly) with the time during embryogenesis that these cells are exiting their mitotically quiescent stage and beginning mitosis, it is tempting to speculate that, like tre-2, ME may play a role in growth control, regulating mitosis in the embryonic germline. Is the mitotic arrest of the germ cells due to interference with known cell-cycle regulators like cdc2/28-cyclinB (also known as M-phase promoting factor, or MPF, see **figure 11A**)? For example, by repression of specific activators like cdc25/string (Edgar and O'Farrell, 1989; Edgar et al., 1994), or maintenance of MPF kinase activity inhibitors like wee1 (Lohka et al., 1988).

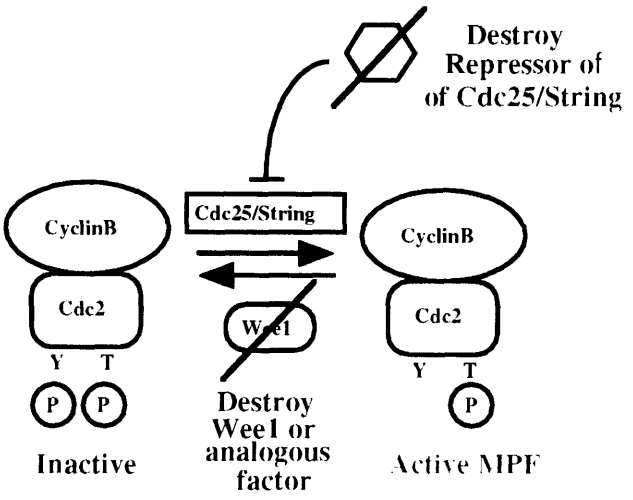


Figure 11A. Potential cell cycle targets for ME in germ cells: controlling re-entry into mitosis by eliminating repressors of Cdc25/String, an MPF activator or Wee1, an inhibitor of MPF activity. Homologues of the cell-cycle regulators in the figure above have been identified in Drosophila. Mechanisms controlling entry into mitosis in the embryonic germ cells could include some or all of these factors. If ME acts as a deubiquitinating enzyme in the embryonic germline, it could be responsible for targetted proteolysis of mitotic regulatory molecules such as Wee1 and/or repressors of Cdc25/String, allowing the embryonic germ cells to exit G2 and begin their mitotic proliferation phase. (figure adapted from Murray and Hunt (1993))

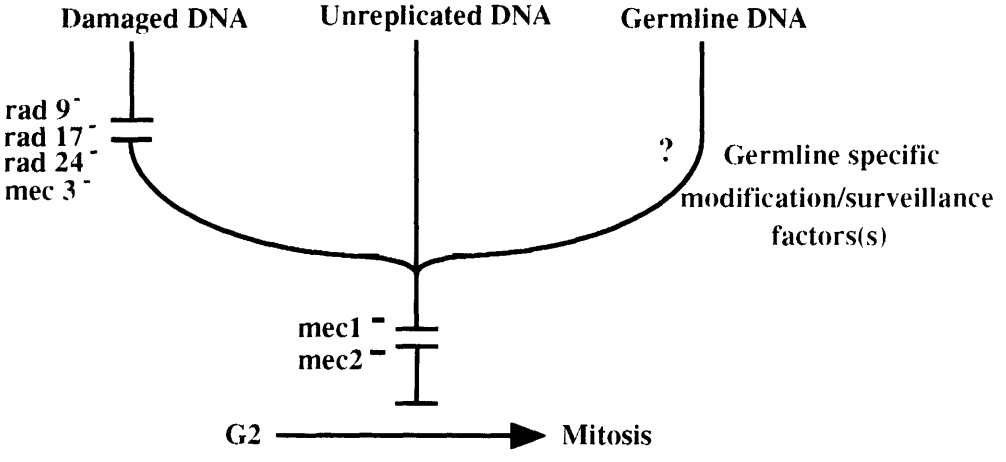


Figure 11B. Damaged and Unreplicated DNA: Controlling the exit from G2 by germ cells. Is there a mechanism to protect germ cell DNA content from damage during their embryonic quiescent state until resumption of mitosis during gonad differentiation? Targets for degredation by ME may include germline specific modification and/or surveillance factors analogous to Rad9 which is required to prevent cells from entering mitosis when their DNA has sustained damage due to UV or Gamma rays (figure adapted from Murray and Hunt (1993))

Given that the genomic DNA content of the embryonic germ cells will later give rise to the next generation, maintenance of the integrity of this DNA is critical for the propagation of the species. Therefore it is interesting to speculate that there may exist a mechanism whereby the genomic DNA in the germ cells is specifically modified and/or protected from the time the pole cells form until they re-enter their mitotic phase. Such a surveillance mechanism could, in principle, serve as a cue, like DNA damage and un-replicated DNA for inhibitory factors preventing cells from entering mitosis(Weinert, 1992; Weinert and Hartwell, 1988).

The Rad9 class of yeast mutants are all required in the budding yeast to prevent cells with damaged DNA from entering mitosis (see **figure 11B**). Another class of mutants in this pathway includes mec1 and 2, both of which prevent cells from entering mitosis before completion of DNA replication. It may be that the G2 arrest of embryonic germ cells (P. O'Farrell, unpublished observations) mimics the inhibition of mitosis seen in irradiated yeast and mammalian cells which cannot enter mitosis until DNA damage is repaired and/or replication of the genome is complete. Results from phenotypic analysis of Doa4 mutants in yeast reveal that cells lacking this ubp-component of the 26S proteosome show multiple defects including slow growth and sensitivity to UV and gamma rays associated with defects in DNA repair mechanisms (Papa and Hochstrasser, 1993). These same defects can be seen in other proteosome component mutants as well as in ubiquitin conjugating enzyme mutants, consistent with a role for ubiquitin-dependent proteolysis in keeping cells from entering mitosis because of a DNA surveillance mechanism.

Since there are no existing mutations in the ME gene, generation of a dominant negative version of the protein would be one way of determining what its normal function is and, by extension, what role the ubiquitin-specific proteolysis pathway plays in the embryo as a whole, and in the germ cells in particular. As described in chapter two, any specific information about structure and/or function of a gene can facilitate the design of targeted mutations. For example, since the putative active site of deubiquitinating enzymes is well established, this domain will be the target for specific changes in the most highly conserved residues. In particular, a dominant negative mutation that disrupts the function of Doa4 is a change of the active site cysteine at position 571 to a serine. A single copy of this mutant gene completely abolishes deubiquitinating activity of the mutant protein as well as the wild-type copy *in vivo* and *in vitro* (Papa and Hochstrasser, 1993). Therefore, one of the most logical mutations to make in the ME coding sequence is the analagous cysteine to serine change. Transgenic flies carrying this mutant gene under an inducible promoter may

allow me to determine what the mutant phenotype of an ME loss of function mutation would be, without the need to isolate mutations in the endogenous wild-type gene. Alternatively, if a truncated protein can be engineered that is capable of folding and interacting with its normal proteolytic partners *in vivo* but is lacking the proteolytic active site completely, this protein product might also be expected to act in a dominant negative fashion, by sequestering important proteolysis regulators in unproductive complexes.

The simplest approach to finding out what roles ME and MA play *in vivo* is to identify the genetic map location of these genes, and thereby identify potential mutants in loci already described genetically. The hope is that this approach might lead to the identification of previously isolated mutant strains whose genetic lesions map to the same chromosomal locations as the genes of interest. If the existing mutations can be shown to disrupt expression of ME or MA RNA or protein, then the analysis of *in vivo* function for these clones is relatively straight-forward. On the other hand, in the absence of obvious candidate mutants, concrete evidence to tie the expression of these two genes to known developmental processes may result from identification of mutations in genes known to affect germline determination and/or differentiation that disrupt the normal expression of RNA or protein encoded by these two genes.

The process of embryonic gonad formation is disrupted in a number of zygotically acting mutants including the gap gene *Krüppel* (Wieschaus et al., 1984) . This gene has a role in determining the metameric pattern of the larval abdomen. In particular, *Krüppel* mutant embryos are lacking all three thoracic segments as well as abdominal segments 1-5, while the remaining tissues appear to be normal. *Krüppel* is therefore placed in the segmentation "Gap" gene class along with *hunchback*, *knirps*, *giant*, (Pachter et al., 1987; Petschek et al., 1987) and *tailless* (Pignoni et al., 1990) . Since embryos homozygous mutant for *Krüppel* are missing abdominal segments 1-5 , they therefore lack the anlage of the somatic component of the gonads (Wieschaus et al., 1984) .

I have examined the expression if ME and MA in *Krüppel*-mutant embryos to address the question of whether zygotic transcription of any of these germ cell markers may be activated only when the pole cells form normal gonads. A lack of pole cell-specific transcription in these scattered cells would suggest the existence of a differentiation signal produced only upon proper association with somatic mesoderm cells and/or coalescence of the gonads.

In homozygous *Krüppel*-mutant embryos, pole cell migration is normal until the time of germband shortening, after which time the pole cells fail to aggregate, scattering throughout the posterior half of the embryo. I have examined the RNA expression pattern of both ME and MA in *Krüppel* mutant embryos to determine whether the germ cell-specific expression of either one of these genes is prevented in pole cells that are scattered throughout the mutant embryos instead of forming normal, coalesced gonads, encapsulated by somatic mesodermal cells. The result is that although not all of the scattered pole cells appear to express these transcripts in addition to vasa protein, there are subsets of the germ cells that do indeed express these two markers (data not shown). Therefore it appears that expression of both ME and MA in pole cells is independent of the formation of embryonic gonads. What remains to be determined is whether there are requirements for contact with mesodermal (or other) germ layers or tissues in order for the germ cells to differentiate normally.

It is possible that the real requirement is for contact with certain subsets of mesodermal cells which still exist in these mutant embryos, thereby allowing the expression of these gonad-specific genes. The most logical experiment to test this model is to examine zygotic ME and MA gene expression in *twist* or *snail* mutant embryos, which lack all mesodermal tissues (Leptin, 1991) . Another possibility is that the intrinsic or extrinsic signals controlling the gonad-specific transcription of the ME RNA acts independently from and/or before the *Krüppel* mutant defect is manifested.

Materials and Methods

Digoxygenin-labeled RNA probes were synthesized as described in Chapter 2 from +21SH library clones by digestion with BamHI for T7 polymerase-derived transcripts and with BglII for SP6 polymerase-derived transcripts.

The unsubtracted full length "+"cDNA was generated by ligating EcoRI adapters (Stratagene) to the blunt cDNAs followed by ligation into pre-cut EcoRI λZAP (Stratagene) arms and packaged, titered (1.02×10^6 pfu/ml) and plated according to manufacturer's instructions (Gigapack, Sratagene, La Jolla, CA) at a density of approximately $30\text{-}40 \times 10^3$ pfu per 150 x 15 mm plate. Filter lifts were made according to standard protocols (Sambrook et al., 1989) on Hybond N (Amersham) filters, cross-linked as described above and pre-hybridized and hybridized as described for colony lifts (see chapter 2, materials and methods). Probes were synthesized from purified, PCR-amplified +21SH library inserts as described above.

After pBluescript plasmid rescue using the EXASSIST/SOLR System (Stratagene) Plasmids were digested 5' and 3' of insertion sites, and single stranded digoxygenin-labeled RNA probes were synthesized as described in materials and methods of chapter 2.

nanos cDNA fragments of 502 and 205 bp derived from pN5 (Wang and Lehmann, 1991) were generated by digestion with PvuII and PstI and cloned into pSP72. Templates for riboprobe synthesis were generated by digestion with Xho I to generate a T7-derived transcripts and with Xba I to generate SP6 transcripts.

Both the plasmid library (Brown and Kafatos, 1988) and the phage library generated from the unsubtracted +cDNA were screened using ^{32}P dCTP-labeled probes derived by random hexamer labeling (Sambrook et al., 1989) . Template for these syntheses was generated by PCR and gel purified as described in chapter 2.

Screening for 2.8 kb MA cDNA and larger ME cDNAs: Primary pools of colonies were picked based on hybridization to ME or MA-specific probes, grown in overnight cultures and used for miniprep DNA isolation by alkaline lysis. One microliter of a 1:10 dilution of these plasmid DNA pools was then used in each of two PCR reactions, using one vector primer with the corresponding internal primer to amplify either the N or C-terminal portions of the cDNA. Primary cultures that gave rise to DNA pools containing positive(s) with

cDNA inserts of ~2.8 kb (for MA) and >4 kb (for ME) were plated and screened as described above for secondary and tertiary screens of the original library.

Nucleic acid and protein database searches were performed using the National Center for Biotechnology Information Server using the BLAST algorithm (Atschul et al., 1990) .

Bacterial expression of ME and MA proteins: To express the coding sequences of the MA gene (see **figure 3**), the cDNA was modified for cloning into the pGEX 2T vector (glutathione S transferase or "GST" vector, Pharmacia) by the insertion of an in-frame Bam HI site adjacent to the third methionine codon and a TAA-TGA double stop EcoRI site after the last codon. The complete sequence of this expression clone was verified by double stranded sequencing with specific MA primers (see materials and methods, chapter 2).

The GST-MA construct was expressed in DH5α by addition of IPTG to a final concentration of 0.6 mM, once freshly inoculated cultures had reached an OD_{600} ~0.2. Total E.coli samples were analyzed by SDS-PAGE to confirm expression. Soluble and inclusion body preparations confirmed that the expressed protein was in the inclusion body fraction. Inclusion bodies were re-solubilized. Briefly, inclusion bodies were resuspended with a dounce homogenizer in 20 mM tris pH 8.0, 50 mM NaCL, 0.2 mM EGTA, 1 mM DTT and 1 mM azide containing 8M urea. This solution was then rapidly diluted 20 fold into either PBS or 20 mM tris pH 8.0, 500 mM NaCL, 0.2 mM EGTA, 1 mM DTT and 1 mM azide. The resulting suspension was then spun at 20,000 rpm at 4°C for 30 minutes. The supernatant was dialyzed against a 60-fold volume of the respective dilution buffer over-night at 4°C and clarified at 20,000 rpm at 4°C for 30 minutes and filtered through a 0.2 μm filter. SDS-PAGE confirmed that the expressed GST-MA fusion protein was fully soluble in both buffers, and greater than 90% pure.

Antisera were affinity purified as follows:

NaCl and Carbonate buffer were added to soluble GST-MA protein to bring the solution to 0.1M $NaHCO_3$, 0.5M NaCl pH 8.3, to make it compatible with CNBr coupling. CNBr-activated Sepharose 4B (Sigma) was activated by washing with 3 volumes 1M HCl 3 times, followed by 3 washes in 1X coupling buffer (0.1M $NaHCO_3$, 0.5M NaCl, pH 8.3), according to manufacturer's protocol (Pharmacia). Two milligrams protein and activated sepharose were incubated, rocking at 4°C for 2 hours, supernatant was removed and saved for efficiency calculations and unreacted moieties on the protein-bound sepharose were blocked by incubation in 1M Tris pH8.0 for 1-2 hours at 4°C. Tris

buffer was removed, sepharose was washed five times with alternating low pH buffer: 0.1M NaOAc, pH4, followed by high pH Tris buffer: 0.1M Tris, 0.5M NaCl, pH 8.0., to remove non-covalently bound protein from the resin (Pharmacia). Finally, the coupled resin was washed three times in 1X PBS and stored in 1X PBS, 0.01% Thimerosal at 4°C.

Two mls of antiserum were incubated with 1ml of GST-MA-coupled Sepharose 4B (~1 mg of protein), 2-5 hours rocking at 4°C. The resin plus buffer was then poured into a small column (Bio-rad). The unbound antiserum was allowed to flow through and the resin was then washed extensively in 1X PBS until the absorbence at 280 nm (A_{280})of the follow-through was at baseline. Specifically bound antibody fractions were eluted in 0.1M NaOAc, 10mM NaCl, pH 2.5. 500 µl fractions were collected directly into tubes containing 38 µl 2M Tris pH 8 to neutralize the fractions upon collection. (all columns were run at room temperature). Fractions were assayed for protein content by measuring the A_{280} for each one. Peak and side-peak fractions were pooled separately, brought to 0.1% BSA (extra-special crystallized BSA, immunology-grade), 0.02% Azide and stored at 4°C or frozen in small aliquots at -80°C.

GST was expressed and purified from pGEX2T in DH5α cells according to protocols supplied by the manufacturer (Pharmacia, GST Gene Fusion System). GST-containing soluble fractions were incubated in the presence of glutathione-sepharose in 1X PBS, pH 7.5, 4 hours at 4°C. After GST binding, the resin and supernatant was poured into a column (Bio-rad), unbound material was allowed to flow through and the GST-bound resin was washed extensively with 1X PBS pH 7.5 until A_{280} measurements were down to baseline. The resin was then scooped into a tube and incubated with affinity-purified antiserum (1 ml) at 4°C overnight to absorb out GST-specific antibodies from the serum. After binding, the mixture was again poured into a column to facilitate collection of the unbound antiserum. This final pool of affinity purified, GST-absorbed material was stored at 4°C until use.

Antibody stainings to whole mount embryos were carried out according to the method described in Barker et al, (1992) using the avidin/biotin ABC system (Vector labs).

The dialyzed inclusion body material was loaded on to an Sephacryl-200 gel filtration column pre-equilibrated in 1X PBS, pH7.5 (Pharmacia). The material eluted from this column was analyzed by SDS-PAGE and the peak fractions pooled and concentrated to 0.5mg/ml prior to injection into one rabbit for production of polyclonal antisera (Hazelton Research Products, North Carolina).

Protein Blots: Blocking was accomplished by incubation in BLOTTO: 10mM Tris pH 7.5, 150mM NaCl, 5% w/v Nonfat Dry Milk, 0.1% Tween-20, 0.01% Thimerosal. Developer Buffer: 100mM Tris pH9.5, 100mM NaCl, 5mM $MgCl_2$, using 0.165 mg/ml 5-bromo-4-indolyl phosphate toluidinium salt (BCIP) and 0.280mg/ml nitroblue tetrazolium salt (NBT) (Vector Labs) in developer buffer. SDS-PAGE gels: 10-12% acrylamide:bis (29:1) were run according to standard protocols, electroblotted onto Imobilon-P membranes according to manufacturer's instructions.

Primary and secondary antibodies were diluted into BLOTTO, incubated at room temperature for 1-5 hours at room temperature or over night at 4°C. Washes after both primary and secondary antibody incubations were done in BLOTTO (4-6 times, 30 minutes each). Finally, membranes were washed twice, for five minutes in 1X Developer buffer prior to signal detection.

In situ hybridizations to salivary gland chromosomes were performed by the method of Rubin et al. (Rubin lab manual II, 1990).

EPILOGUE. Unanswered Questions

The Search for function: Are ME and MA required for germ line differentiation?

The ultimate goal of this project is to understand the process of embryonic gonad differentiation in general, and specifically, what role the ME and MA genes play in this process. As with any molecular screen, the challenge that remains after identification of genes expressed in the tissues or during the developmental time period of interest is to establish the relevance of this gene expression to the process being studied. What are the mechanisms underlying the differentiation and what is the role these genes play in it? The next step is to take the initial observation that ME and MA show expression patterns during embryogenesis consistent with their being markers of germ cell differentiation in embryos and extend the analysis to find evidence for or against a requirement for these genes in this process.

In addition to examining the expression of ME and MA in existing mutants, another approach to finding out what role these genes may play in germline differentiation is to generate phenotypes for these genes by expressing heat shock-driven anti-sense cDNAs at different stages during gonad development. The advantage of using the heat shock promoter is that I can induce expression of the anti-sense cDNA at defined points during development. However, the problem with heat shock-induced expression is that this promoter is not activated in a tissue-specific manner. Since both ME and MA are ubiquitously expressed in early embryos and are present in non-germline tissues during late embryogenesis, global reduction in the level of these proteins may cause lethality during late embryogenesis, larval, or pupal phases. The ideal situation would be to use a germline-specific promoter to drive zygotic antisense RNA expression only in the embryonic gonads. Unfortunately, there are no embryonic germline-specific GAL4 UAS lines that are known to be active in the germline in embryos. Moreover, although the *vasa* gene is transcribed zygotically in these tissues during late stage 11 to early stage 12 of embryogenesis, this may not be early enough to see a clear effect on germline differentiation.

Strategies to generate flies lacking ME and MA expression

The next step is to use the genetic map location information to establish what the mutant phenotype of embryos lacking ME and MA would be. As described in chapter 3, *in situ* hybridizations to salivary gland chromosomes using probes derived from ME cDNAs revealed that ME lies within at least two existing deficiencies, Ar14-8 and Ar12-1. Although these two deficiencies are large, including many different loci, they will provide useful material for examining smaller deficiencies in the region, nearby lethal P-element insertions, and other mutations thought to reside in the ME transcription unit. For example, by using balancer chromosomes that carry well-defined *lacZ* enhancer trap markers (e.g. *Ubx-lacZ*), embryos homozygous for deficiencies can be unequivocally identified during in situ hybridizations to test for the expression of ME RNA, and to examine these mutant embryos for the lack of pole cells.

Another approach will be to locate P-elements in the ME and MA chromosomal regions which can either be mobilized *in trans* to deficiencies, thereby creating small deletions uncovering the transcription units, or to hop nearby P-elements locally into the transcription units and thereby disrupt expression of the two genes. Alternatively, x-rays can be used to target the P-elements in the region to make small deletions that can then be tested for lethality over existing deficiencies. The X-ray deletion-containing chromosomes could first be examined by *in situ* hybridization to salivary gland chromosomes to confirm the absence of both the nearby P-element and the ME or MA transcription units. Additional confirmation of successful deletions could be obtained by looking for the lack of ME or MA RNA or protein expression, either by whole mount embryo *in situs* or by single-embryo westerns.

Identification of upstream control genes:

The primary reason for examining various mutants to look for lack of expression of these putative germ line differentiation factors is to identify the genes controlling this germline-specific expression of ME and MA. A complementary approach to these studies will be to isolate and sequence genomic DNA conferring this tissue-specific regulation of these two RNAs. The process of dissecting sequences 5' of these transcription units can be accomplished in relatively straight forward fashion by the construction of a set of overlapping *lacZ* promoter fusion constructs. These sequences can then be used to

generate germline transformant animals and examined for the expression of the *lacZ* RNA and/or protein.

Once a construct carrying upstream regulatory regions from either of these genes has been established which recapitulates the normal RNA expression pattern, the genomic DNA sequences within this construct can, in theory, be systematically pared down. In the simplest case, if the factors that bind to these sequences to specifically activate transcription are few, one may be able to identify them biochemically. However, another possible and desirable outcome from such experiments would be the identification of regulatory sequences responsible for the germline specific expression of these two genes.

In particular, there may be unique regulatory sequences responsible for the gonad-specific expression of MA, independently controlled and/or separable from the expression of this RNA in the central nervous system. In addition, both of these late embryonic expression features may be under control of zygotic control elements, distinct from the maternal transcriptional control sequences that are likely to be activated during oogenesis.

References:

Atschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

Baker, R. T., J. W. Tobias, and A. Varshavsky. 1992. Ubiquitin-specific proteases of Saccharomyces cerevisiae. Cloning of UBP2 and UBP3, and functional analysis of the UBP gene family. *Journal of Biological Chemistry.* 267:23364-23375.

Brown, N. H., and F. C. Kafatos. 1988. Functional cDNA libraries from Drosophila embryos. *Journal of Molecular Biology.* 203:425-437.

Campos-Ortega, J. A., and V. Hartenstein. 1985. The embryonic development of Drosophila melanogaster. Springer-Verlag, Heidelberg.

Carroll, S. B., and M. P. Scott. 1986. Zygotically active genes that affect the spatial expression of the fushi tarazu segmentation gene during early Drosophila embryogenesis. *Cell.* 45:113-126.

Edgar, B. A., and P. H. O'Farrell. 1989. Genetic control of cell division patterns in the Drosophila embryo. *Cell.* 57:177-187.

Edgar, B. A., F. Sprenger, R. J. Duronio, P. Leopold, and P. H. O'Farrell. 1994. Distinct molecular mechanisms regulate cell cycle timing at successive stages of *Drosophila* embryogenesis. *Genes & Development.* 8:440-452.

Fischer-Vize, J., G. M. Rubin, and R. Lehmann. 1992. The *fat facets* gene is required for *Drosophila* eye and embryo development. *Development.* 116:985-1000.

Gaul, U., E. Seifert, R. Schuh, and H. Jäckle. 1987. Analysis of Krüppel protein distribution during early Drosophila development reveals posttranscriptional regulation. *Cell.* 50:639-647.

Hochstrasser, M. 1992. Ubiquitin and intracellular protein degradation. *Current Opinion in Cell Biology.* 4:1024-1031.

Karr, T. L., and B. M. Alberts. 1986. Organization of the cytoskeleton in early Drosophila embryos. *Journal of Cell Biology.* 102:1494-1509.

Leptin, M. 1991. *twist* and *snail* as positive and negative regulators during Drosophila mesoderm development. *Genes & Development.* 5:1568-1576.

Lindsley, D. L., and L. Sandler. 1972. Segmental aneuploidy and the genetic gross structure of the Drosophila genome. *Genetics.* 71:157-184.

Lindsley, D. L., and G. G. Zimm. 1992. The genome of *Drosophila melanogaster.* Academic Press, Inc., San Diego.

Liu, C. C., H. I. Miller, W. J. Kohr, and J. I. Silber. 1989. Purification of a ubiquitin protein peptidase from yeast with efficient in vitro assays. *J. Biol. Chem.* 264:20331-20338.

Lohka, M. J., M. D. Hayes, and J. L. Maller. 1988. Purification of maturation-promoting factor, an intracellular regulator of early mitotic events. *Proc. Natl. Acad. Sci.* 85:3009-3013.

Murray, A., and T. Hunt. 1993. The cell cycle: an introduction. W.H. Freeman and Company, New York. 251.

Pachter, J. S., T. J. Yen, and D. W. Cleveland. 1987. Autoregulation of tubulin expression Is achieved through specific degradation of polysomal tubulin mRNAs. *Cell.* 51:283-292.

Papa, F. R., and M. Hochstrasser. 1993. The yeast DOA4 gene encodes a deubiquitinating enzyme related to a product of the human tre-2 oncogene. *Nature.* 366:313-319.

Petschek, J. P., N. Perrimon, and A. P. Mahowald. 1987. Region-specific defects in (1)giant embryos of Drosophila melanogaster. *Developmental Biology.* 119:175-189.

Pignoni, F., R. M. Baldarelli, E. Steingrimsson, R. J. Diaz, A. Patapoutian, J. R. Merriam, and J. A. Lengyel. 1990. The Drosophila gene *tailless* is expressed at the embryonic termini and is a member of the steroid receptor superfamily. *Cell.* 62:151-163.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Tobias, J. W., and A. Varshavsky. 1991. Cloning and functional analysis of the ubiquitin-specific protease gene UBP1 of Saccharomyces cerevisiae. *Journal of Biological Chemistry.* 266:

Wang, C., and R. Lehmann. 1991. *Nanos* is the localized posterior determinant in Drosophila. *Cell.* 66:637-648.

Weinert, T. A. 1992. Dual cell cycle checkpoints sensitive to chromosome replication and DNA damage in the budding yeast Saccharomyces cerevisiae. *Radiation Research.* 132:141-143.

Weinert, T. A., and L. H. Hartwell. 1988. The RAD9 gene controls the cell cycle response to DNA damage in Saccharomyces cerevisiae. *Science.* 241:317-322.

Wieschaus, E., C. Nüsslein-Volhard, and H. Kluding. 1984. Krüppel, a gene whose activity is required early in the zygotic genome for normal embryonic segmentation. *Developmental Biology.* 104:172-186.

Appendix I

Characterization of MH, a gene expressed specifically in the amnioserosa and somatic musculature

Introduction

The advantage of doing molecular screens for genes that show tissue-specific expression in *Drosophila* is that one can rigorously screen against false positives by examining the RNA distribution pattern of all clones *in vivo*. In addition, if any of the isolated clones that do not fit the original criteria nevertheless show interesting RNA expression patterns, the cDNAs are already in hand, facilitating further analysis of the unexpected clones. In the process of examining expression patterns of the 7 full length cDNA positives I discovered that one of the clones (called "MH"), although not expressed in the germ cells, gave a striking tissue-specific expression pattern (see **figure 1**). This clone gave such a remarkable expression pattern that I decided to characterize it further. Therefore I mapped the gene by *in situ* hybridization to salivary gland chromosomes and sequenced most of a 3.6 kb cDNA. The results of these experiments are detailed below.

Results

Tissue-specific expression of MH *in vivo*:

Figure 1 shows the embryonic RNA expression pattern of the MH clone. Expression is ubiquitous early (**figure 1A**), and decreases somewhat by stage four (~2 hours AEL), when the RNA can be seen in the cortex underlying the somatic nuclei (**figure 1B**). MH RNA appears to be excluded from both the germ line and somatic cellular regions in these embryos (**figure 1B**). At stage 9 (**figure 1C**), the amnioserosa can be seen to specifically express this RNA. After germ-band retraction, the amnioserosa-specific expression is still predominant (**figure 1D**). **Figure 1E** shows a dorsal view of a stage 14 embryo (~12 hours AEL), giving a clear view of the RNA in the cytoplasm of the polyploid amnioserosa cells. Figure **1F** shows the expression of this RNA in the somatic musculature of a stage 1 embryo. This muscle-specific RNA expression is shown from a dorsal view in a stage 1 embryo (**figure 1G**),~14.5 hours AEL. At this stage RNA expression starts to be discernible in the pharyngeal musculature (arrow). By the end of embryogenesis, in late

stage 1 to early stage 17 embryos (**figure 1H**), the RNA is expressed strongly in the majority of the somatic musculature in addition to the dorsal pharyngeal musculature, the pharyngeal maxillary muscles (arrows) and pericardial cells (arrow head), as well as remnants of the amnioserosa (see **figure 1H**, Campos-Ortega and Hartenstein 1985).

MH *in situ* Hybridizations

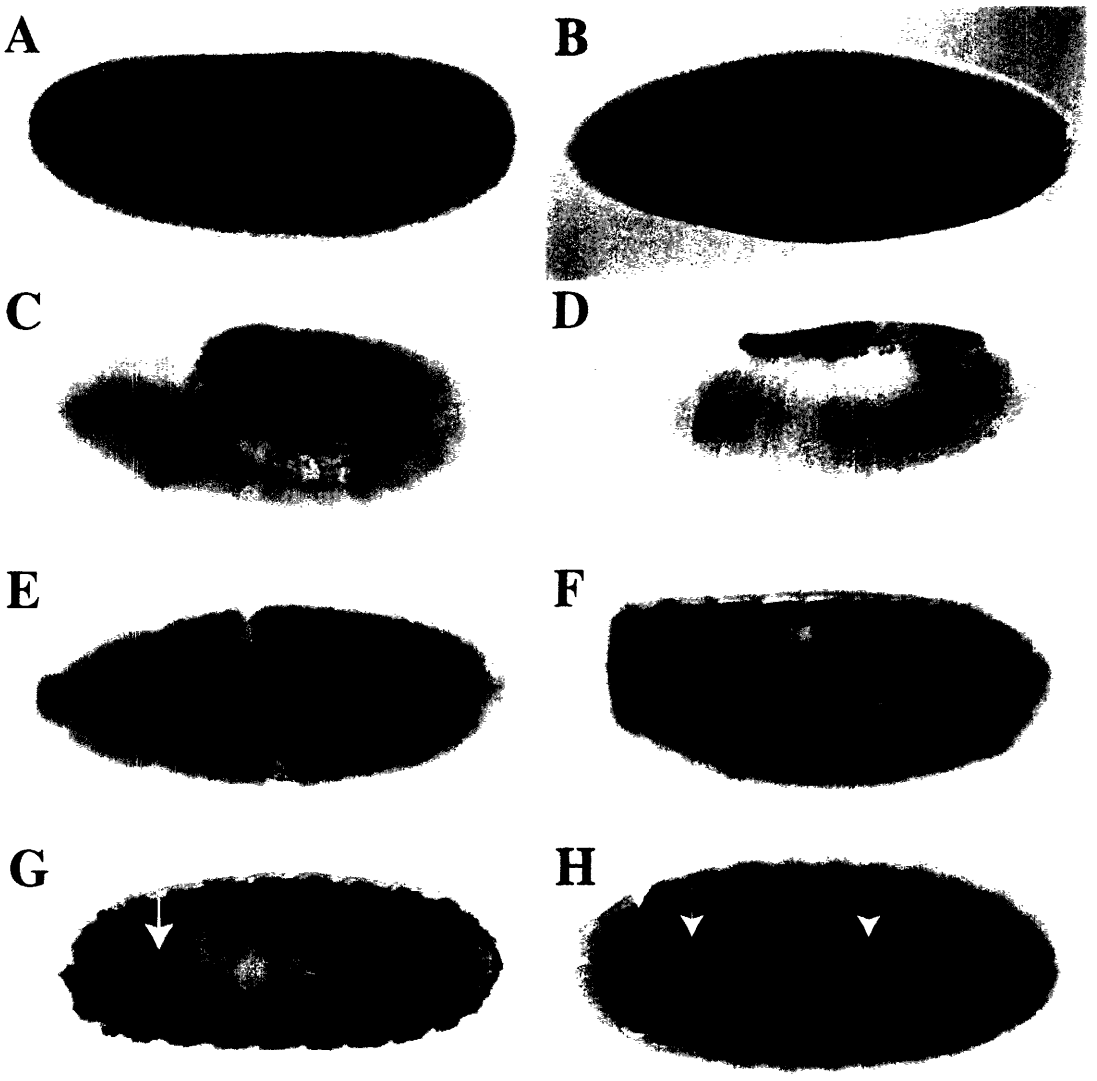


Figure 1. MH RNA expression pattern in embryos: A. Stage 1: The RNA is distubuted uniformly throughout freshly laid eggs. B. Stage 4, The overall level of expression decreases during the syncytial blastoderm stage. C. Stage 9: By the middle of germband extension the RNA is strongly expressed in the amnioserosa (dorso-lateral view). D. Stage 13: after germband retraction the RNA is still expressed specifically in the amnioserosa (lateral view). E. Stage 14: Dorsal view of MH expression in the amnioserosa following germ band retraction. F. Stage 16: Staining becomes apparent in a subset of somatic muscle precursors. G. Stage 16: Dorsal view of MH expression shown in F. Expression starts to be detectable in the developing pharyngeal musculature H. Stage 17: MH expression in the somatic musculature, including the dorsal pharyngeal and pharyngeal maxillary muscles (arrows), and the pericardial cells (arrow head), remnants of the amnioserosa

MH is not differentially expressed:

Not surprisingly, given the RNA expression pattern of this gene, RNA blot hybridization with the original library insert as well as with near full length cDNA probes revealed that none of the 5 transcripts of approximately 4.4, 4.0, 3.6, 3.0, and 2.0 kb, encoded by this gene are differentially expressed in wild type verses pole cell-minus embryos (see **figure 2**). The 2.0 kb transcript however, is clearly under zygotic transcriptional and/or splicing control as it is present in 45 minute-13 hour embryo RNA pools but not in RNA from 0-4 hour embryos (data not shown).

MH is not differentially expressed

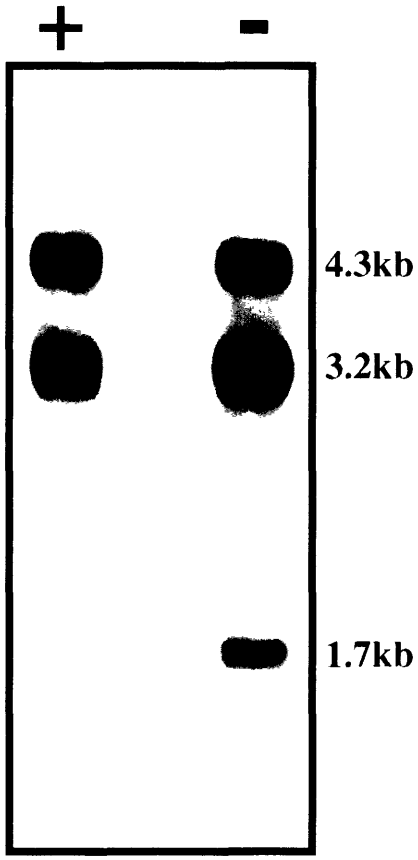


Figure 2. MH RNA blot hybridization: The same blot as shown in all previous figures was probed with the original MH +21SH cDNA library insert as well as near full-length cDNA probes. These blots reveal that none of the transcripts of 4.4, 4.0, 3.6, 3.0 and 2.0kb encoded by this gene are differentially expressed in wild-type versus agametic embryos.

MH encodes a sarco/endoplasmic reticulum type Ca-ATPase:

MH: cDNAs of 2.5, 3.0 and 3.6 kb were isolated from the cDNA library described above (Brown and Kafatos, 1988) . The cDNA sequenced was approximately 3.6 kb. Virtually complete sequence of this cDNA confirmed the identity of MH as a cDNA encoding a *Drosophila* sarco/endoplasmic reticulum-type Ca^{2+}-ATPase. The correspondence between the sequence of the 3.6 kb MH clone and that of the 3.3 kb cDNA described in Magyar et al., 1990, (Magyar and Varadi, 1990) is almost 100% and is shown in **figure 3**.

MH cDNA Sequence

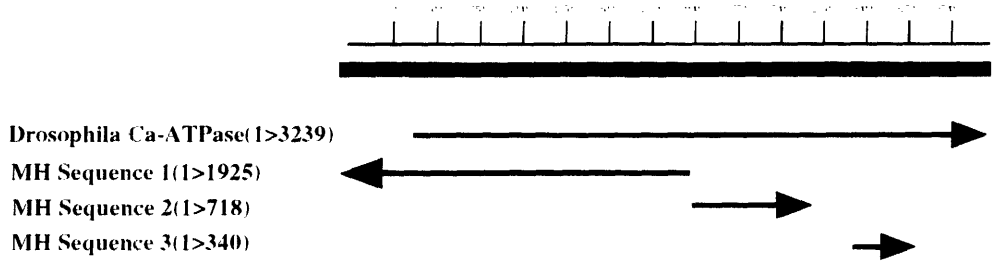


Figure 3. Sequence from the 3.6 kb MH clone isolated from a full length 4-8 hour embryonic cDNA library (see chapter 3, materials and methods) using a probe derived from the +21SH subtracted library. The sequence derived from this cDNA is shown in alignment with that of the 3.3 kb cDNA described in Magyar et al., (1990). The correspondence between the two sequences is virtually 100% (indicated in blue). Sequences present only in MH or in the previously cloned cDNA but not both are indicated in red. Although the cDNA sequenced in this work appears to contain more 5' sequences than that of Magyar et al., the open reading frame is the same in both cDNAs.

Further confirmation that this clone is derived from the same gene that had been previously identified was provided by *in situ* hybridization to salivary gland chromosomes which revealed that MH maps to interval 60A, at the tip of chromosome 2R (see **figure 4**), as does the gene described by Magyar and colleagues (1990).

MH *in situ* hybridization to wild-type salivary gland chromosomes

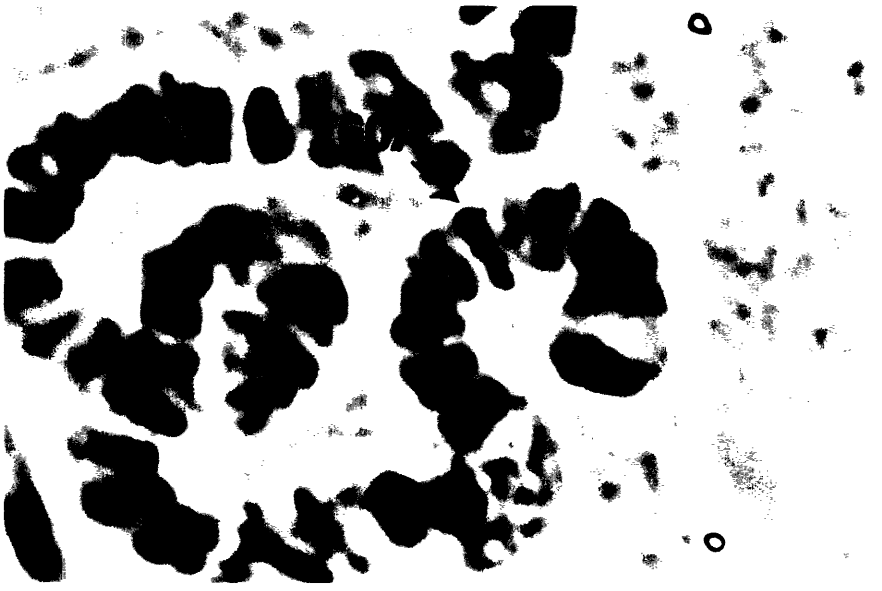


Figure 4. *In situ* hybridization to wild-type salivary gland chromosomes reveals that MH maps to the interval 60A, on the tip of chromosome 2R.This is the same cytological map location found for the Drosophila sarco/endoplasmic reticulum-type Ca2+-ATPase isolated by Magyar and Varadi (1990).

Discussion

Despite the fact that MH does not fit the original criteria for the targets of the molecular screen, it nonetheless has an interesting embryonic expression pattern. The RNA is maternally supplied to the oocyte and is homogeneously distributed throughout the syncytial blastoderm. During germ band extension the RNA is expressed specifically in the amnioserosa, the extra embryonic membrane that normally provides a barrier between the yolk and the external environment before dorsal closure is complete (Campos-Ortega and Hartenstein, 1985) . After germ band retraction, as the amnioserosa beings to degenerate, the MH RNA begins to show specific expression in the embryonic somatic muscle precursors. The RNA for this gene is present in roughly equivalent amounts in "+" and "-" RNA pools, and therefore likely represents an anomalous enrichment event due, perhaps to preferential amplification of cDNA fragments from this gene in the "+" pool at some point during the subtraction process (as was also seen for *Toll*, see chapter 2).

BLAST homology searches revealed that the MH cDNA is virtually identical to a previously identified gene in the database. The sequence encodes a protein with all of the conserved sequences expected in a sarco/endoplasmic reticulum-type Ca^{2+}ATPase. This gene, named "DRSERCA," was isolated in a PCR screen conducted using primers designed to hybridize to the highly conserved phosphorylation site/flourescein isothiocyanate binding site of vertebrate membrane-bound calcium ATPases (Magyar and Varadi, 1990) .

The sarco/endoplasmic reticulum-type Ca^{2+} ATPases are membrane bound calcium transport proteins localized to the membrane of the sarco/endoplamsic reticulum. These proteins function in the retrieval of intracellular calcium, pumping it into the sarcoplasmic reticulum in an ATP-dependent fashion. The fact that DRSERCA is expressed in the amnioserosa and somatic musculature may be merely coincidental; this Ca-ATPase may function in a completely different manner in the two tissues, but it is possible that intracellular mechanisms may be shared between the early extra-embryonic membrane and the embryonic somatic musculature.

The role of calcium release in triggering muscle contraction in response to nerve stimulation is well established, and therefore it is not terribly surprising that a calcium pump protein of this sort is found to be specifically expressed in the embryonic somatic musculature. Less obvious, however, is the role that such a molecule might play in the amnioserosa. The

expression of DRSERCA RNA in the amnioserosa during germ band extension raises the possibility that the cell shape changes taking place in this tissue are regulated by intracellular calcium. Indeed, the regulation of actin filament length is critical to the accomplishment of actin-mediated cell shape changes in many cell and tissue types, and is known to be exquisitely sensitive to calcium concentration.

There have recently been a number of genetic screens initiated to isolate genes with a role in differentiation of the amnioserosa and there are a number of existing zygotic mutations that affect the differentiation of this tissue. In *zerknüllt, twisted gastrulation, decapentaplegic,* and *short gastrulation* mutants, the amnioserosa cells do not undergo their normal shape change from columnar to cuboidal to squamous and finally to form a thin sheet of cells covering the dorsal side of the embryo. Instead, these mutations cause ventralization of the dorsal epidermis that is associated with a failure of the amnioserosa cells to flatten and move laterally to make way for the extending germ band (Bacharova, 1992; Bate and Martinez Arias, 1993) . It is possible that the cell shape changes that occur in the amnioserosa cells during gastrulation require calcium regulation in the same way that muscle cells do during contraction.

A number of other genes have been identified that are specifically expressed in the amnioserosa, including *Krüppel*, which is also expressed during germ-band retraction in the lateral muscle precursor cells (Gaul et al., 1987) . Mutations in genes that affect germband elongation may be expected to alter the expression patterns of downstream genes expressed in the amnioserosa. Examination of a battery of ventralizing mutants would be a logical first step in determining whether transcription of this Ca-ATPase is regulated by genes in the dorso-ventral pathway.

DRSERCA is another example of a gene that would be amenable to generation of dominant negative mutants by changing or eliminating residues in known active or regulatory sites. In the case of this calcium ATPase, an obvious choice is to mutate the highly conserved phosphorylation site/flourescein isothiocyanate binding site. Construction of transgenes containing gal4 UAS binding sites would allow one to drive expression of the dominant negative mutant protein specifically in the amnioserosa and somatic muscle tissues during embryogenesis and (Brand and Perrimon, 1993) thereby gain insight into the function of this gene during embryogenesis.

Appendix II

Sequencing Data

The method used for sequencing is described in detail in Bankier et al., (Bankier et al., 1987) . The amassed sequence information is presented in this appendix as the original Seqman contig documents containing the alignment of sequence fragments generated by random-sonnication. Briefly, the cDNAs to be sequenced were excised from the vector and self-ligated to form circles, then sonnicated to produce a random distribution of fragments representing the entire cDNA. These fragments are repaired to generate blunt ends and cloned into a convenient vector for sequencing (in this case, pBluescriptKS(-), Stratagene) and sequenced using vector-specific primers (e.g. SP6 and T7). The random sequences are aligned using Seqman software to produce the "Contigs" found in Appendix II.

References:

Bacharova, R. F. 1992. A maternal tail of Poly(A): The long and the short of It. *Cell.* 69:895-897.

Bankier, A. T., K. M. Weston, and B. G. Barrell. 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods Enzymol.* 155:51.

Bate, M., and A. Martinez Arias. 1993. The Development of Drosophila melanogaster. 746.

Brand, A. H., and N. Perrimon. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development.* 118:401-415.

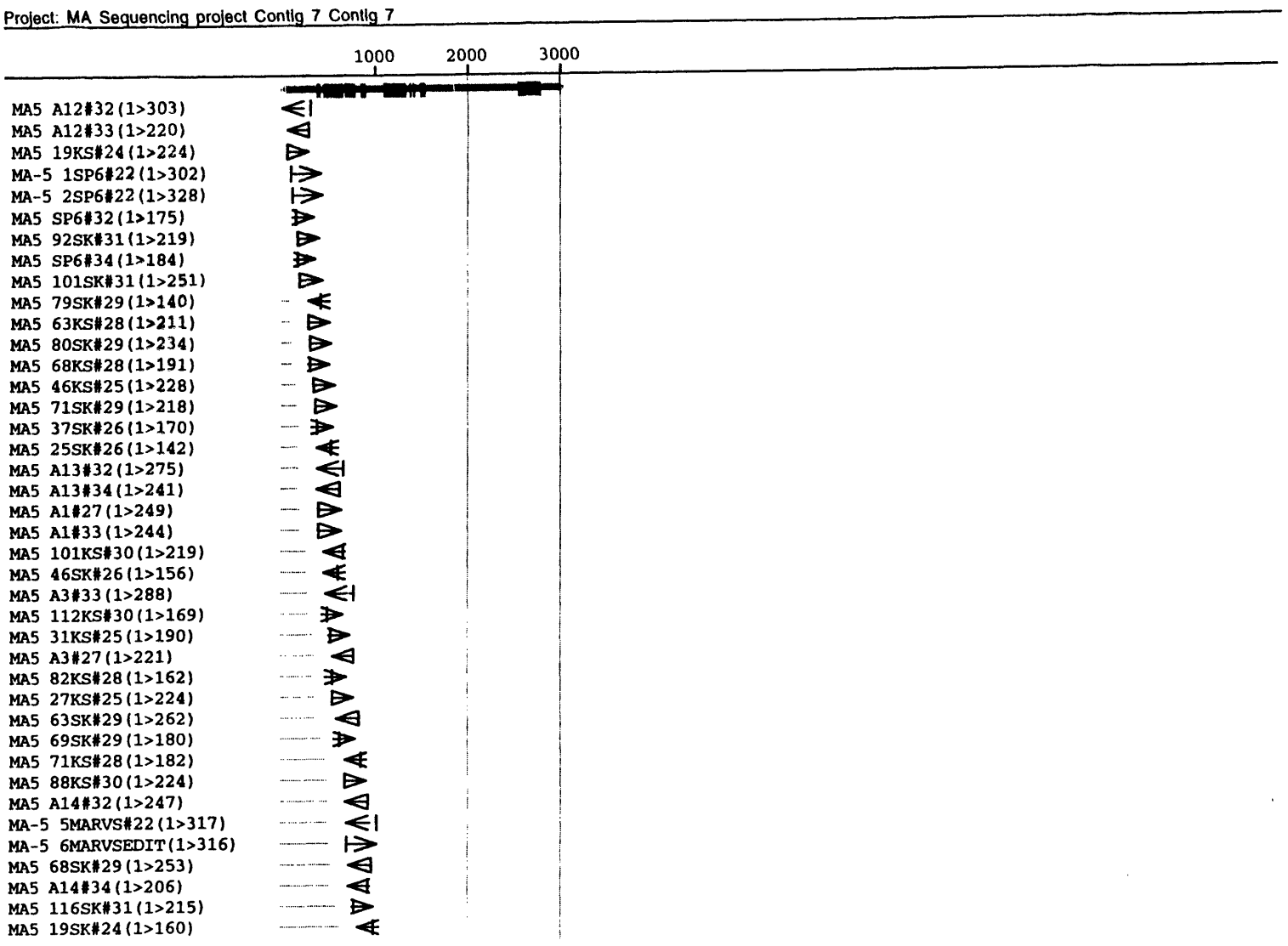
Brown, N. H., and F. C. Kafatos. 1988. Functional cDNA libraries from Drosophila embryos. *Journal of Molecular Biology.* 203:425-437.

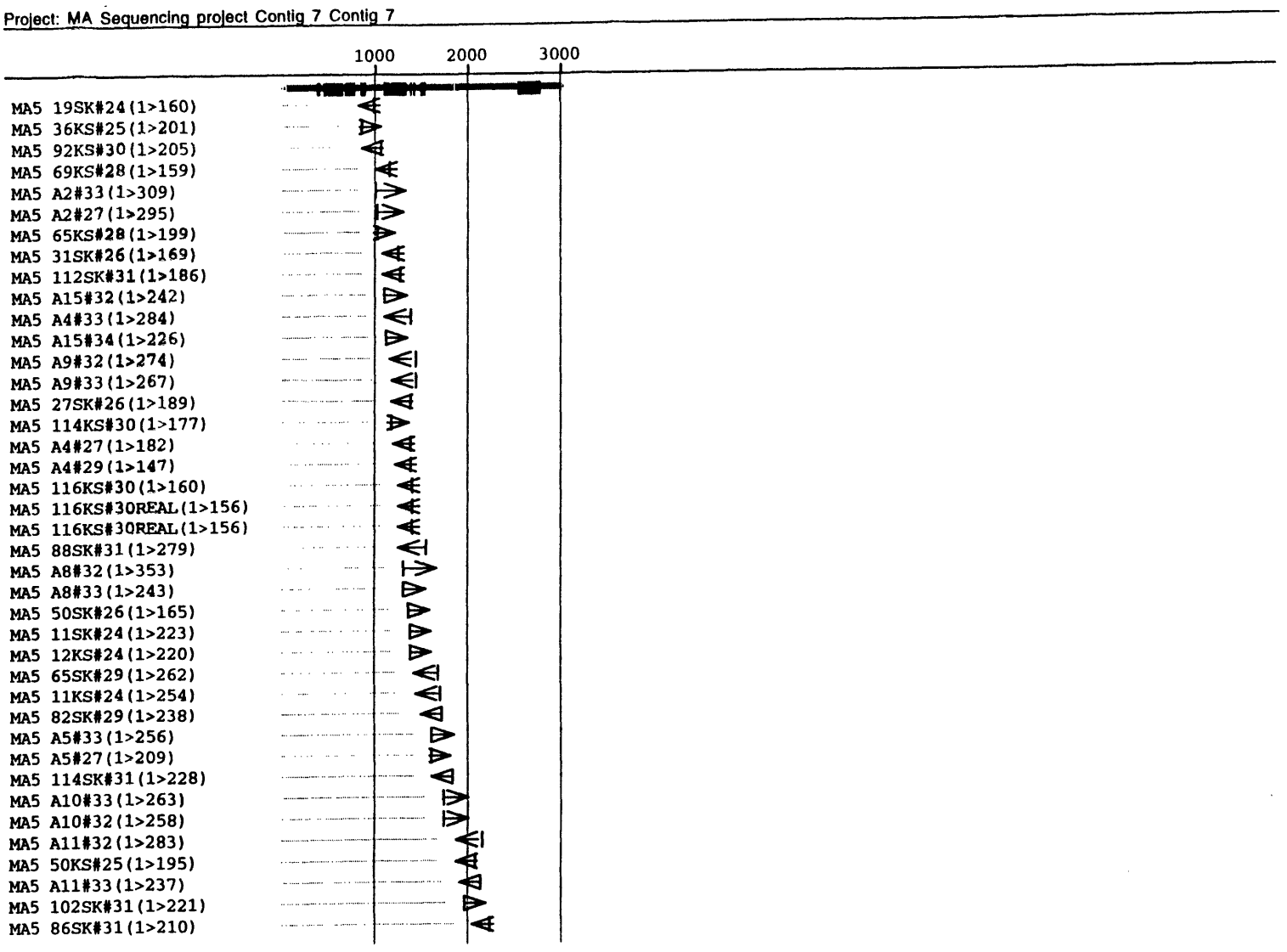
Campos-Ortega, J. A., and V. Hartenstein. 1985. The embryonic development of Drosophila melanogaster. Springer-Verlag, Heidelberg.

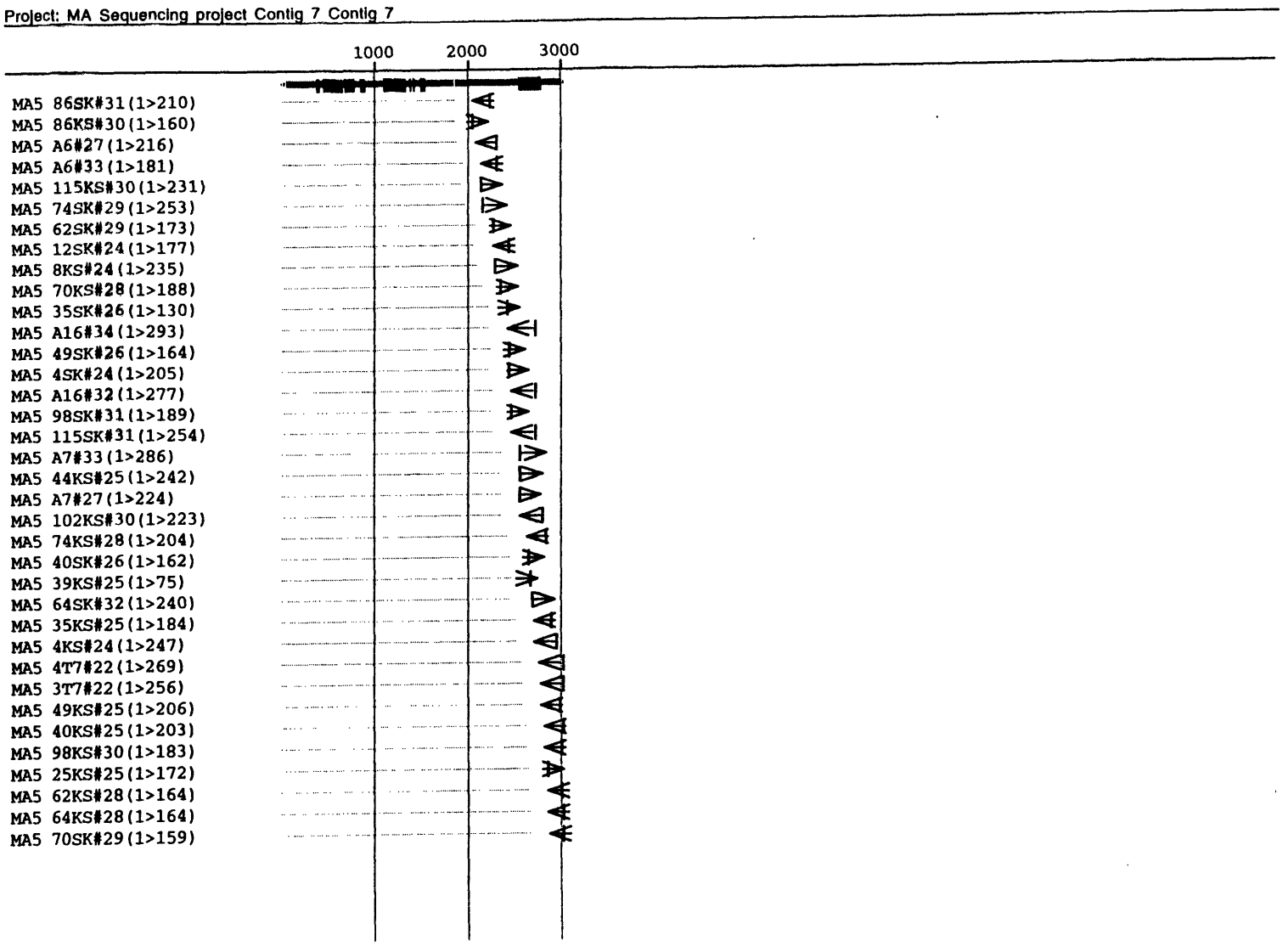
Gaul, U., E. Seifert, R. Schuh, and H. Jäckle. 1987. Analysis of Krüppel protein distribution during early Drosophila development reveals posttranscriptional regulation. *Cell.* 50:639-647.

Magyar, A., and A. Varadi. 1990. Molecular cloning and chromosomal localization of a sarco/endoplasmic reticulum-type Ca2+-ATPase of Drosophila melanogaster. *Biochemical and Biophysical Research Communications.* 173:872-877.

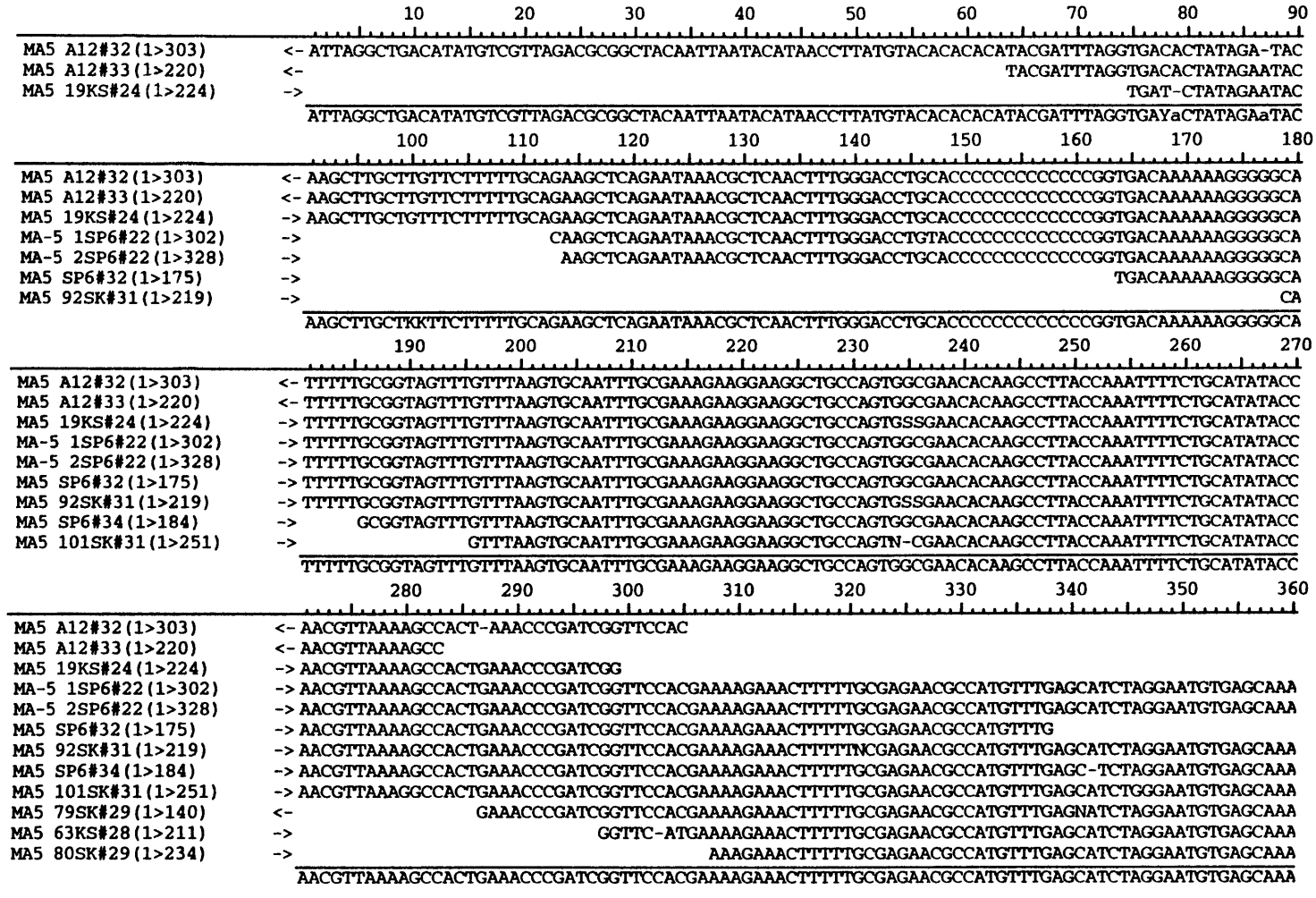
MA Sequencing project: Contig 7: Strategy View:

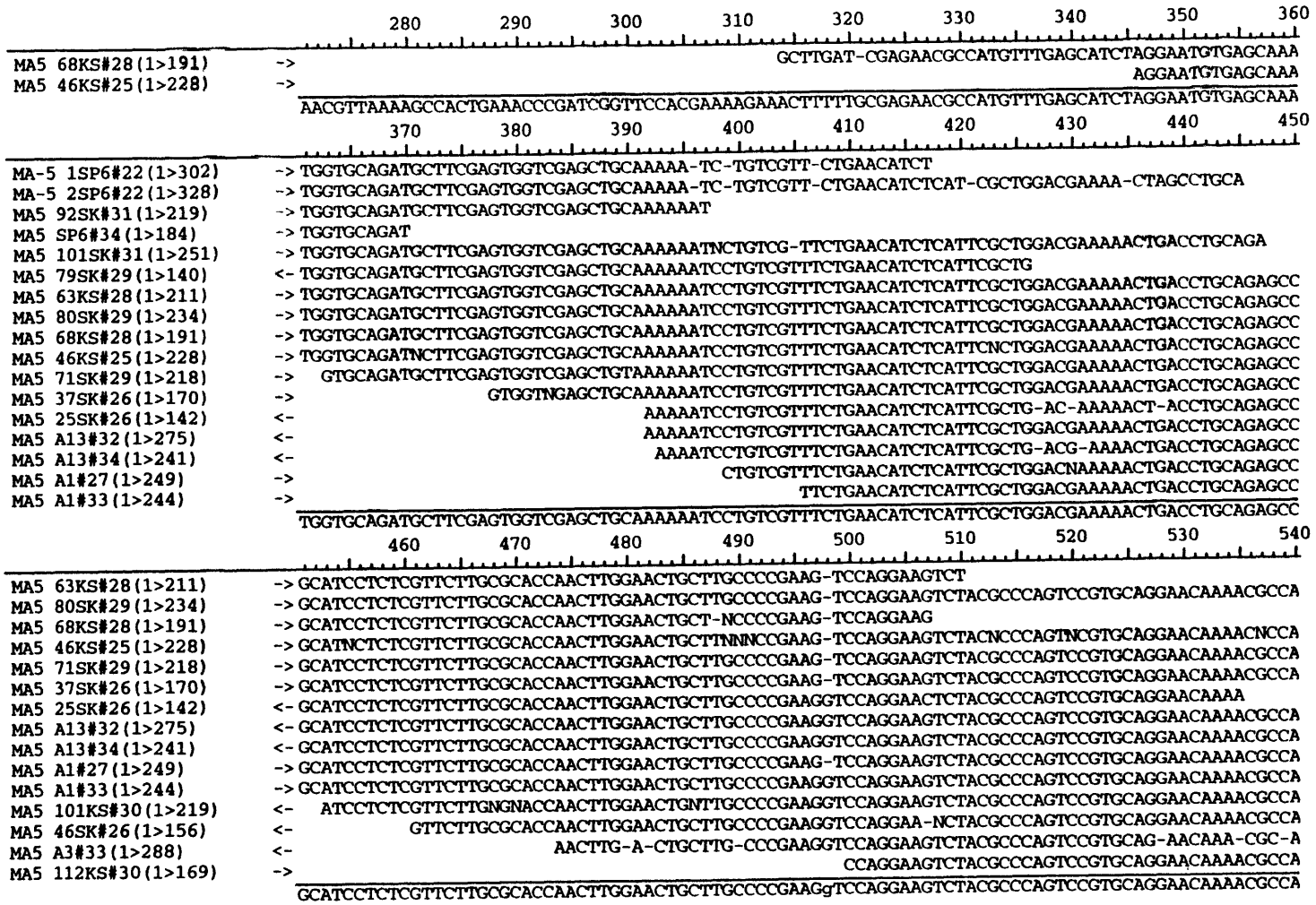

Project: MA Sequencing project Contig 7 Contig 7
1000
2000
3000
MA5 A12#32(1>303)
MA5 A12#33(1>220)
MA5 19KS#24(1>224)
MA-5 1SP6#22(1>302)
MA-5 2SP6#22(1>328)
MA5 SP6#32(1>175)
MA5 92SK#31(1>219)
MA5 SP6#34(1>184)
MA5 101SK#31(1>251)
MA5 79SK#29(1>140)
MA5 63KS#28(1>211)
MA5 80SK#29(1>234)
MA5 68KS#28(1>191)
MA5 46KS#25(1>228)
MA5 71SK#29(1>218)
MA5 37SK#26(1>170)
MA5 25SK#26(1>142)
MA5 A13#32(1>275)
MA5 A13#34(1>241)
MA5 A1#27(1>249)
MA5 A1#33(1>244)
MA5 101KS#30(1>219)
MA5 46SK#26(1>156)
MA5 A3#33(1>288)
MA5 112KS#30(1>169)
MA5 31KS#25(1>190)
MA5 A3#27(1>221)
MA5 82KS#28(1>162)
MA5 27KS#25(1>224)
MA5 63SK#29(1>262)
MA5 69SK#29(1>180)
MA5 71KS#28(1>182)
MA5 88KS#30(1>224)
MA5 A14#32(1>247)
MA-5 5MARVS#22(1>317)
MA-5 6MARVSEDIT(1>316)
MA5 68SK#29(1>253)
MA5 A14#34(1>206)
MA5 116SK#31(1>215)
MA5 19SK#24(1>160)

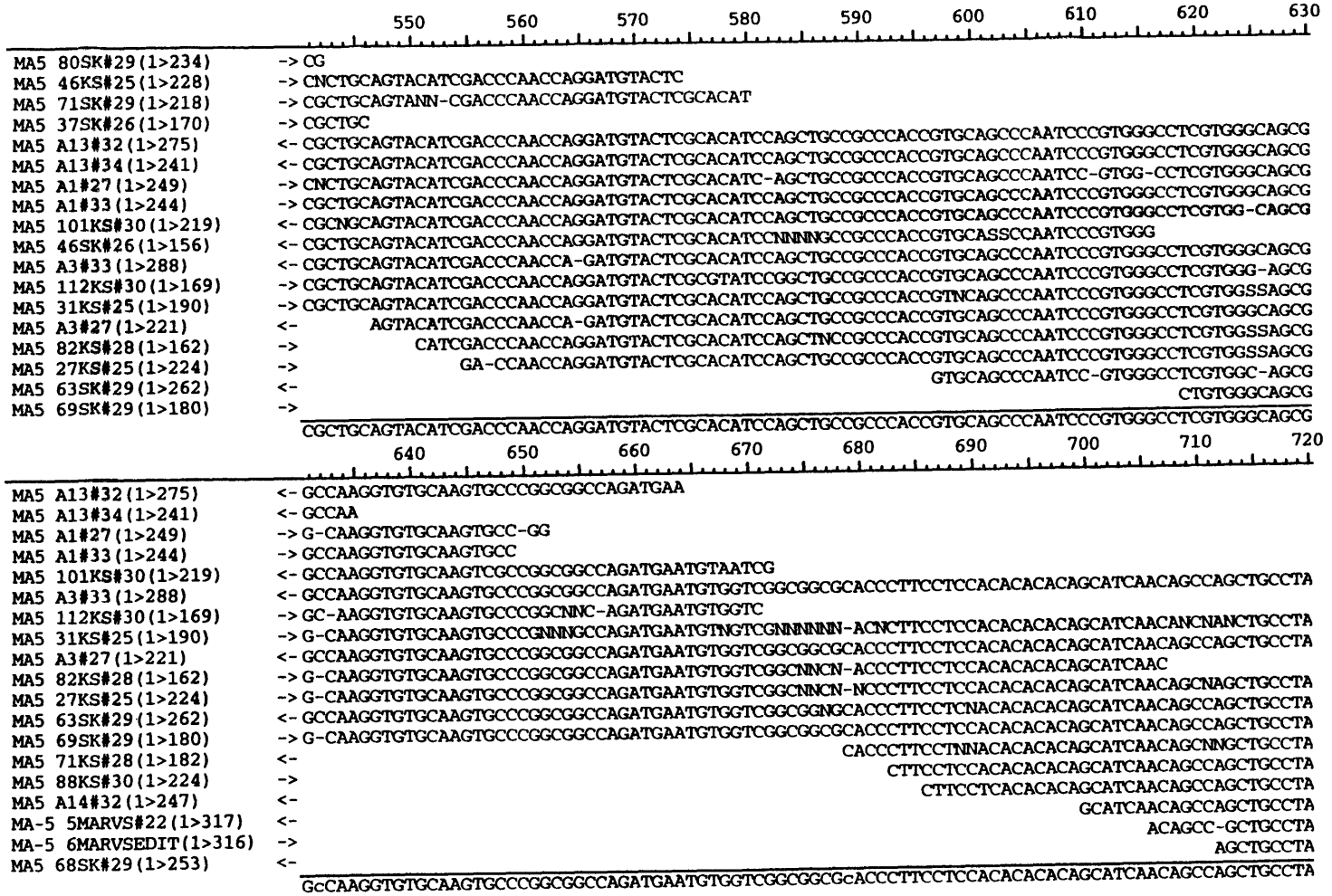

Project: MA Sequencing project Contig 7 Contig 7
1000
2000
3000
MA5 19SK#24(1>160)
MA5 36KS#25(1>201)
MA5 92KS#30(1>205)
MA5 69KS#28(1>159)
MA5 A2#33(1>309)
MA5 A2#27(1>295)
MA5 65KS#28(1>199)
MA5 31SK#26(1>169)
MA5 112SK#31(1>186)
MA5 A15#32(1>242)
MA5 A4#33(1>284)
MA5 A15#34(1>226)
MA5 A9#32(1>274)
MA5 A9#33(1>267)
MA5 27SK#26(1>189)
MA5 114KS#30(1>177)
MA5 A4#27(1>182)
MA5 A4#29(1>147)
MA5 116KS#30(1>160)
MA5 116KS#30REAL(1>156)
MA5 116KS#30REAL(1>156)
MA5 88SK#31(1>279)
MA5 A8#32(1>353)
MA5 A8#33(1>243)
MA5 50SK#26(1>165)
MA5 11SK#24(1>223)
MA5 12KS#24(1>220)
MA5 65SK#29(1>262)
MA5 11KS#24(1>254)
MA5 82SK#29(1>238)
MA5 A5#33(1>256)
MA5 A5#27(1>209)
MA5 114SK#31(1>228)
MA5 A10#33(1>263)
MA5 A10#32(1>258)
MA5 A11#32(1>283)
MA5 50KS#25(1>195)
MA5 A11#33(1>237)
MA5 102SK#31(1>221)
MA5 86SK#31(1>210)


Project: MA Sequencing project Contig 7 Contig 7
1000
2000
3000
MA5 86SK#31(1>210)
MA5 86KS#30(1>160)
MA5 A6#27(1>216)
MA5 A6#33(1>181)
MA5 115KS#30(1>231)
MA5 74SK#29(1>253)
MA5 62SK#29(1>173)
MA5 12SK#24(1>177)
MA5 8KS#24(1>235)
MA5 70KS#28(1>188)
MA5 35SK#26(1>130)
MA5 A16#34(1>293)
MA5 49SK#26(1>164)
MA5 4SK#24(1>205)
MA5 A16#32(1>277)
MA5 98SK#31(1>189)
MA5 115SK#31(1>254)
MA5 A7#33(1>286)
MA5 44KS#25(1>242)
MA5 A7#27(1>224)
MA5 102KS#30(1>223)
MA5 74KS#28(1>204)
MA5 40SK#26(1>162)
MA5 39KS#25(1>75)
MA5 64SK#32(1>240)
MA5 35KS#25(1>184)
MA5 4KS#24(1>247)
MA5 4T7#22(1>269)
MA5 3T7#22(1>256)
MA5 49KS#25(1>206)
MA5 40KS#25(1>203)
MA5 98KS#30(1>183)
MA5 25KS#25(1>172)
MA5 62KS#28(1>164)
MA5 64KS#28(1>164)
MA5 70SK#29(1>159)

MA Sequencing project: Contig 7: Sequence Alignment:

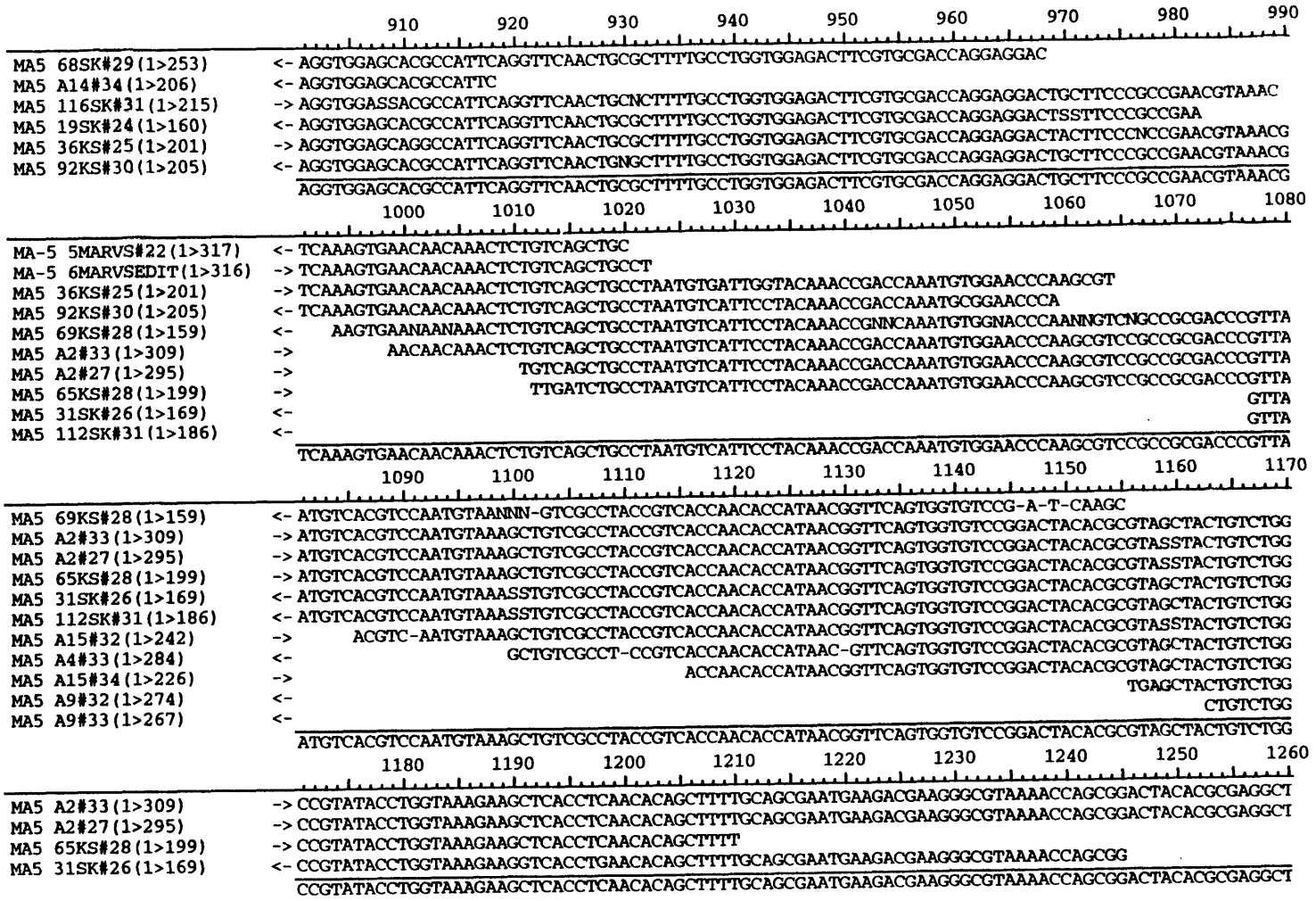

10 20 30 40 50 60 70 80 90
MA5 A12#32(1>303) <- ATTAGGCTGACATATGTCGTTAGACGCGGCTACAATTAATACATAACCTTATGTACACACACATACGATTTAGGTGACACTATAGA-TAC
MA5 A12#33(1>220) <- TACGATTTAGGTGACACTATAGAATAC
MA5 19KS#24(1>224) -> TGAT-CTATAGAATAC
ATTAGGCTGACATATGTCGTTAGACGCGGCTACAATTAATACATAACCTTATGTACACACACATACGATTTAGGTGAYaCTATAGAaTAC
100 110 120 130 140 150 160 170 180
MA5 A12#32(1>303) <- AAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGGACCTGCACCCCCCCCCCCCCCGGTGACAAAAAAGGGGGCA
MA5 A12#33(1>220) <- AAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGGACCTGCACCCCCCCCCCCCCCGGTGACAAAAAAGGGGGCA
MA5 19KS#24(1>224) -> AAGCTTGCTGTTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGGACCTGCACCCCCCCCCCCCCCGGTGACAAAAAAGGGGGCA
MA-5 1SP6#22(1>302) -> CAAGCTCAGAATAAACGCTCAACTTTGGGACCTGTACCCCCCCCCCCCCCGGTGACAAAAAAGGGGGCA
MA-5 2SP6#22(1>328) -> AAGCTCAGAATAAACGCTCAACTTTGGGACCTGCACCCCCCCCCCCCCCGGTGACAAAAAAGGGGGCA
MA5 SP6#32(1>175) -> TGACAAAAAAGGGGGCA
MA5 92SK#31(1>219) -> CA
AAGCTTGCTKKTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGGACCTGCACCCCCCCCCCCCCCGGTGACAAAAAAGGGGGCA
190 200 210 220 230 240 250 260 270
MA5 A12#32(1>303) <- TTTTTGCGGTAGTTTGTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTGGCGAACACAAGCCTTACCAAATTTTCTGCATATACC
MA5 A12#33(1>220) <- TTTTTGCGGTAGTTTGTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTGGCGAACACAAGCCTTACCAAATTTTCTGCATATACC
MA5 19KS#24(1>224) -> TTTTTGCGGTAGTTTGTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTGSSGAACACAAGCCTTACCAAATTTTCTGCATATACC
MA-5 1SP6#22(1>302) -> TTTTTGCGGTAGTTTGTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTGGCGAACACAAGCCTTACCAAATTTTCTGCATATACC
MA-5 2SP6#22(1>328) -> TTTTTGCGGTAGTTTGTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTGGCGAACACAAGCCTTACCAAATTTTCTGCATATACC
MA5 SP6#32(1>175) -> TTTTTGCGGTAGTTTGTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTGGCGAACACAAGCCTTACCAAATTTTCTGCATATACC
MA5 92SK#31(1>219) -> TTTTTGCGGTAGTTTGTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTGSSGAACACAAGCCTTACCAAATTTTCTGCATATACC
MA5 SP6#34(1>184) -> GCGGTAGTTTGTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTGGCGAACACAAGCCTTACCAAATTTTCTGCATATACC
MA5 101SK#31(1>251) -> GTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTN-CGAACACAAGCCTTACCAAATTTTCTGCATATACC
TTTTTGCGGTAGTTTGTTTAAGTGCAATTTGCGAAAGAAGGAAGGCTGCCAGTGGCGAACACAAGCCTTACCAAATTTTCTGCATATACC
280 290 300 310 320 330 340 350 360
MA5 A12#32(1>303) <- AACGTTAAAAGCCACT-AAACCCGATCGGTTCCAC
MA5 A12#33(1>220) <- AACGTTAAAAGCC
MA5 19KS#24(1>224) -> AACGTTAAAAGCCACTGAAACCCGATCGG
MA-5 1SP6#22(1>302) -> AACGTTAAAAGCCACTGAAACCCGATCGGTTCCACGAAAAGAAACTTTTTGCGAGAACGCCATGTTTGAGCATCTAGGAATGTGAGCAAA
MA-5 2SP6#22(1>328) -> AACGTTAAAAGCCACTGAAACCCGATCGGTTCCACGAAAAGAAACTTTTTGCGAGAACGCCATGTTTGAGCATCTAGGAATGTGAGCAAA
MA5 SP6#32(1>175) -> AACGTTAAAAGCCACTGAAACCCGATCGGTTCCACGAAAAGAAACTTTTTGCGAGAACGCCATGTTTG
MA5 92SK#31(1>219) -> AACGTTAAAAGCCACTGAAACCCGATCGGTTCCACGAAAAGAAACTTTTTNCGAGAACGCCATGTTTGAGCATCTAGGAATGTGAGCAAA
MA5 SP6#34(1>184) -> AACGTTAAAAGCCACTGAAACCCGATCGGTTCCACGAAAAGAAACTTTTTGCGAGAACGCCATGTTTGAGC-TCTAGGAATGTGAGCAAA
MA5 101SK#31(1>251) -> AACGTTAAAGGCCACTGAAACCCGATCGGTTCCACGAAAAGAAACTTTTTGCGAGAACGCCATGTTTGAGCATCTGGGAATGTGAGCAAA
MA5 79SK#29(1>140) <- GAAACCCGATCGGTTCCACGAAAAGAAACTTTTTGCGAGAACGCCATGTTTGAGNATCTAGGAATGTGAGCAAA
MA5 63KS#28(1>211) -> GGTTC-ATGAAAAGAAACTTTTTGCGAGAACGCCATGTTTGAGCATCTAGGAATGTGAGCAAA
MA5 80SK#29(1>234) -> AAAGAAACTTTTTGCGAGAACGCCATGTTTGAGCATCTAGGAATGTGAGCAAA
AACGTTAAAAGCCACTGAAACCCGATCGGTTCCACGAAAAGAAACTTTTTGCGAGAACGCCATGTTTGAGCATCTAGGAATGTGAGCAAA

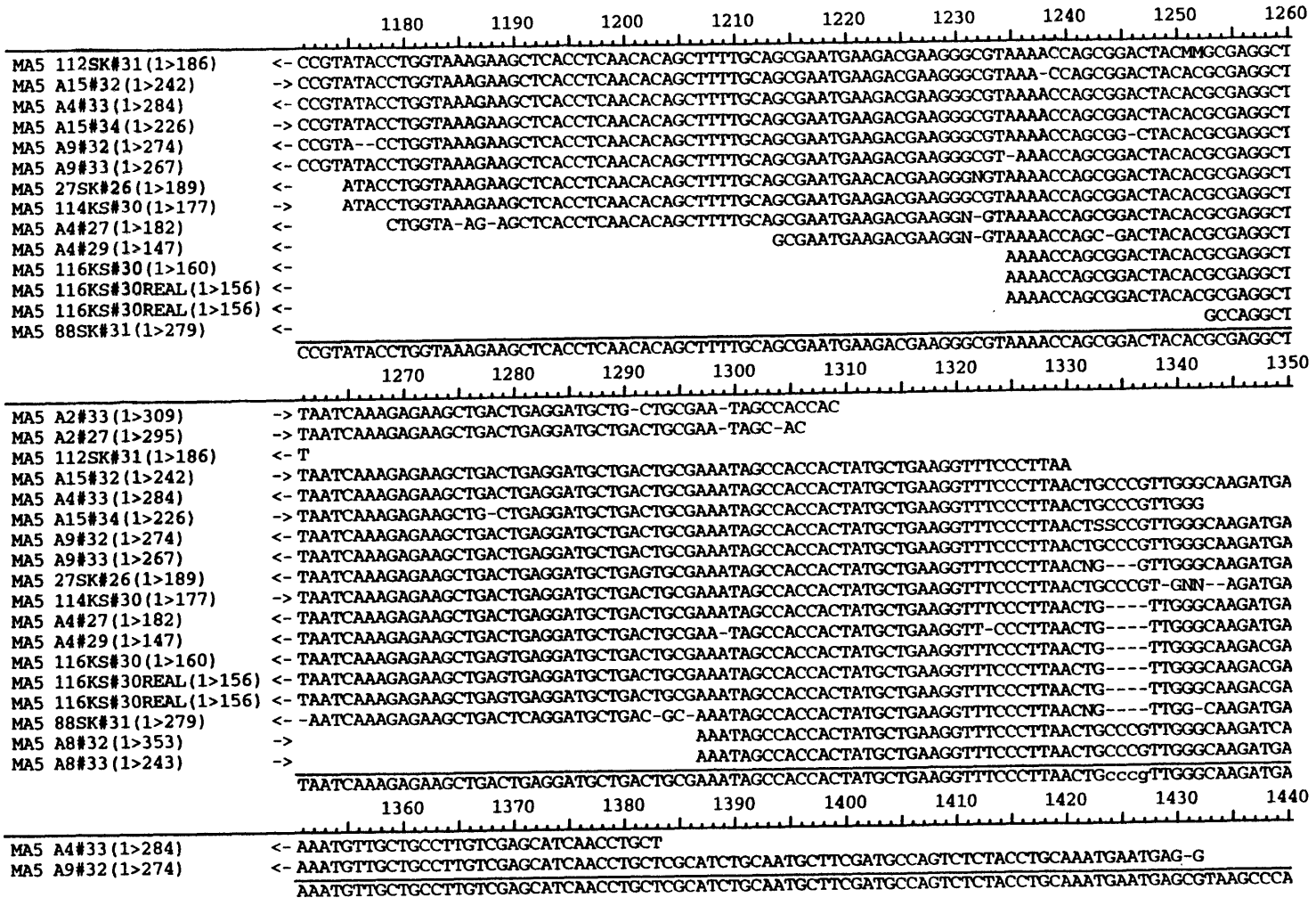

280 290 300 310 320 330 340 350 360
MA5 68KS#28(1>191) -> GCTTGAT-CGAGAACGCCATGTTTGAGCATCTAGGAATGTGAGCAAA
MA5 46KS#25(1>228) -> AGGAATGTGAGCAAA
AACGTTAAAAGCCACTGAAACCCGATCGGTTCCACGAAAAGAAACTTTTTGCGAGAACGCCATGTTTGAGCATCTAGGAATGTGAGCAAA
370 380 390 400 410 420 430 440 450
MA-5 1SP6#22(1>302) -> TGGTGCAGATGCTTCGAGTGGTCGAGCTGCAAAAA-TC-TGTCGTT-CTGAACATCT
MA-5 2SP6#22(1>328) -> TGGTGCAGATGCTTCGAGTGGTCGAGCTGCAAAAA-TC-TGTCGTT-CTGAACATCTCAT-CGCTGGACGAAAA-CTAGCCTGCA
MA5 92SK#31(1>219) -> TGGTGCAGATGCTTCGAGTGGTCGAGCTGCAAAAAAT
MA5 SP6#34(1>184) -> TGGTGCAGAT
MA5 101SK#31(1>251) -> TGGTGCAGATGCTTCGAGTGGTCGAGCTGCAAAAAATNCTGTCG-TTCTGAACATCTCATTCGCTGGACGAAAAACTGACCTGCAGA
MA5 79SK#29(1>140) <- TGGTGCAGATGCTTCGAGTGGTCGAGCTGCAAAAAATCCTGTCGTTTCTGAACATCTCATTCGCTG
MA5 63KS#28(1>211) -> TGGTGCAGATGCTTCGAGTGGTCGAGCTGCAAAAAATCCTGTCGTTTCTGAACATCTCATTCGCTGGACGAAAAACTGACCTGCAGAGCC
MA5 80SK#29(1>234) -> TGGTGCAGATGCTTCGAGTGGTCGAGCTGCAAAAAATCCTGTCGTTTCTGAACATCTCATTCGCTGGACGAAAAACTGACCTGCAGAGCC
MA5 68KS#28(1>191) -> TGGTGCAGATGCTTCGAGTGGTCGAGCTGCAAAAAATCCTGTCGTTTCTGAACATCTCATTCGCTGGACGAAAAACTGACCTGCAGAGCC
MA5 46KS#25(1>228) -> TGGTGCAGATNCTTCGAGTGGTCGAGCTGCAAAAAATCCTGTCGTTTCTGAACATCTCATTCNCTGGACGAAAAACTGACCTGCAGAGCC
MA5 71SK#29(1>218) -> GTGCAGATGCTTCGAGTGGTCGAGCTGTAAAAAATCCTGTCGTTTCTGAACATCTCATTCGCTGGACGAAAAACTGACCTGCAGAGCC
MA5 37SK#26(1>170) -> GTGGTNGAGCTGCAAAAAATCCTGTCGTTTCTGAACATCTCATTCGCTGGACGAAAAACTGACCTGCAGAGCC
MA5 25SK#26(1>142) <- AAAAATCCTGTCGTTTCTGAACATCTCATTCGCTG-AC-AAAAACT-ACCTGCAGAGCC
MA5 A13#32(1>275) <- AAAAATCCTGTCGTTTCTGAACATCTCATTCGCTGGACGAAAAACTGACCTGCAGAGCC
MA5 A13#34(1>241) <- AAAATCCTGTCGTTTCTGAACATCTCATTCGCTG-ACG-AAAACTGACCTGCAGAGCC
MA5 A1#27(1>249) -> CTGTCGTTTCTGAACATCTCATTCGCTGGACNAAAAACTGACCTGCAGAGCC
MA5 A1#33(1>244) -> TTCTGAACATCTCATTCGCTGGACGAAAAACTGACCTGCAGAGCC
TGGTGCAGATGCTTCGAGTGGTCGAGCTGCAAAAAATCCTGTCGTTTCTGAACATCTCATTCGCTGGACGAAAAACTGACCTGCAGAGCC
460 470 480 490 500 510 520 530 540
MA5 63KS#28(1>211) -> GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAG-TCCAGGAAGTCT
MA5 80SK#29(1>234) -> GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCT-NCCCCGAAG-TCCAGGAAG
MA5 68KS#28(1>191) -> GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTNNNNCCGAAG-TCCAGGAAGTCTACNCCCAGTNCGTGCAGGAACAAAACNCCA
MA5 46KS#25(1>228) -> GCATNCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAG-TCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
MA5 71SK#29(1>218) -> GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAG-TCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
MA5 37SK#26(1>170) -> GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAG-TCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAA
MA5 25SK#26(1>142) <- GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAGGTCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
MA5 A13#32(1>275) <- GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAGGTCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
MA5 A13#34(1>241) <- GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAGGTCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
MA5 A1#27(1>249) -> GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAG-TCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
MA5 A1#33(1>244) -> GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAGGTCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
MA5 101KS#30(1>219) <- ATCCTCTCGTTCTTGNGNACCAACTTGGAACTGNTTGCCCCGAAGGTCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
MA5 46SK#26(1>156) <- GTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAGGTCCAGGAA-NCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
MA5 A3#33(1>288) <- AACTTG-A-CTGCTTG-CCCGAAGGTCCAGGAAGTCTACGCCCAGTCCGTGCAG-AACAAA-CGC-A
MA5 112KS#30(1>169) -> CCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA
GCATCCTCTCGTTCTTGCGCACCAACTTGGAACTGCTTGCCCCGAAGgTCCAGGAAGTCTACGCCCAGTCCGTGCAGGAACAAAACGCCA

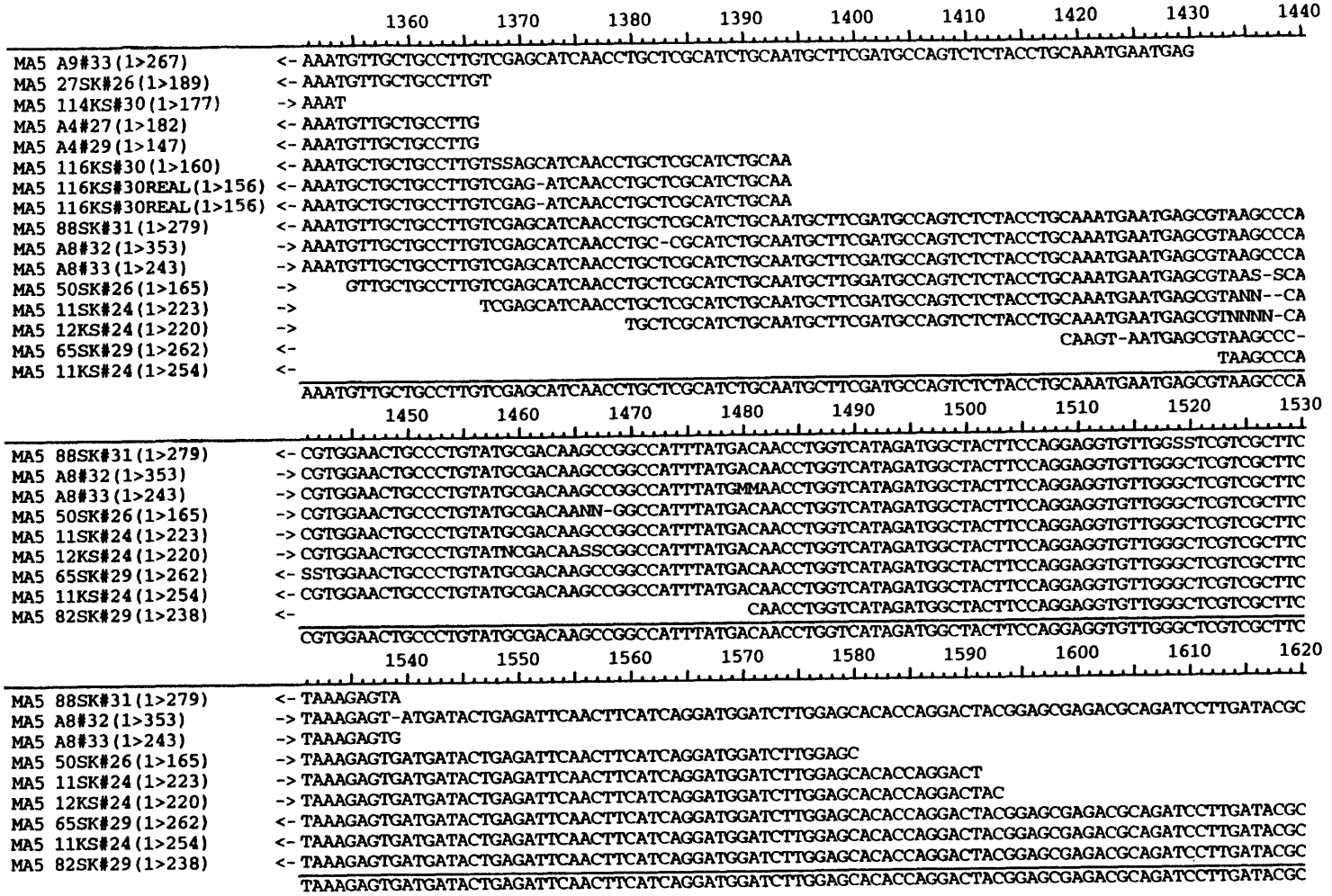

550 560 570 580 590 600 610 620 630
MA5 80SK#29(1>234) -> CG
MA5 46KS#25(1>228) -> CNCTGCAGTACATCGACCCAACCAGGATGTACTC
MA5 71SK#29(1>218) -> CGCTGCAGTANN-CGACCCAACCAGGATGTACTCGCACAT
MA5 37SK#26(1>170) -> CGCTGC
MA5 A13#32(1>275) <- CGCTGCAGTACATCGACCCAACCAGGATGTACTCGCACATCCAGCTGCCGCCCACCGTGCAGCCCAATCCCGTGGGCCTCGTGGGCAGCG
MA5 A13#34(1>241) <- CGCTGCAGTACATCGACCCAACCAGGATGTACTCGCACATCCAGCTGCCGCCCACCGTGCAGCCCAATCCCGTGGGCCTCGTGGGCAGCG
MA5 A1#27(1>249) -> CNCTGCAGTACATCGACCCAACCAGGATGTACTCGCACATC-AGCTGCCGCCCACCGTGCAGCCCAATCC-GTGG-CCTCGTGGGCAGCG
MA5 A1#33(1>244) -> CGCTGCAGTACATCGACCCAACCAGGATGTACTCGCACATCCAGCTGCCGCCCACCGTGCAGCCCAATCCCGTGGGCCTCGTGGGCAGCG
MA5 101KS#30(1>219) <- CGCNGCAGTACATCGACCCAACCAGGATGTACTCGCACATCCAGCTGCCGCCCACCGTGCAGCCCAATCCCGTGGG-CAGCG
MA5 46SK#26(1>156) <- CGCTGCAGTACATCGACCCAACCAGGATGTACTCGCACATCCNNNNGCCGCCCACCGTGCASSCCAATCCCGTGGG
MA5 A3#33(1>288) <- CGCTGCAGTACATCGACCCAACCA-GATGTACTCGCACATCCAGCTGCCGCCCACCGTGCAGCCCAATCCCGTGGGCCTCGTGGGCAGCG
MA5 112KS#30(1>169) -> CGCTGCAGTACATCGACCCAACCAGGATGTACTCGCGTATCCGGCTGCCGCCCACCGTGCAGCCCAATCCCGTGGGCCTCGTGGSSAGCG
MA5 31KS#25(1>190) -> CGCTGCAGTACATCGACCCAACCAGGATGTACTCGCACATCCAGCTGCCGCCCACCGTNCAGCCCAATCCCGTGGGCCTCGTGGSSAGCG
MA5 A3#27(1>221) <- AGTACATCGACCCAACCA-GATGTACTCGCACATCCAGCTGCCGCCCACCGTGCAGCCCAATCCCGTGGGCCTCGTGGSSAGCG
MA5 82KS#28(1>162) -> CATCGACCCAACCAGGATGTACTCGCACATCCAGCTNCCGCCCACCGTGCAGCCCAATCCCGTGGGCCTCGTGGSSAGCG
MA5 27KS#25(1>224) -> GA-CCAACCAGGATGTACTCGCACATCCAGCTGCCGCCCACCGTGCAGCCCAATCCCGTGGGCCTCGTGGSSAGCG
MA5 63SK#29(1>262) <- GTGCAGCCCAATCC-GTGGGCCTCGTGGC-AGCG
MA5 69SK#29(1>180) -> CTGTGGGCAGCG
CGCTGCAGTACATCGACCCAACCAGGATGTACTCGCACATCCAGCTGCCGCCCACCGTGCAGCCCAATCCCGTGGGCCTCGTGGGCAGCG
640 650 660 670 680 690 700 710 720
MA5 A13#32(1>275) <- GCCAAGGTGTGCAAGTGCCCGGCGGCCAGATGAA
MA5 A13#34(1>241) <- GCCAA
MA5 A1#27(1>249) -> G-CAAGGTGTGCAAGTGCC-GG
MA5 A1#33(1>244) -> GCCAAGGTGTGCAAGTGCC
MA5 101KS#30(1>219) <- GCCAAGGTGTGCAAGTCGCCGGCGGCCAGATGAATGTAATCG
MA5 A3#33(1>288) <- GCCAAGGTGTGCAAGTGCCCGGCGGCCAGATGAATGTGGTCGGCGGCGCACCCTTCCTCCACACACACAGCATCAACAGCCAGCTGCCTA
MA5 112KS#30(1>169) -> GC-AAGGTGTGCAAGTGCCCGGCNNC-AGATGAATGTGGTC
MA5 31KS#25(1>190) -> G-CAAGGTGTGCAAGTGCCCGNNNGCCAGATGAATGTNGTCGNNNNNN-ACNCTTCCTCCACACACACAGCATCAACAGCCAGCTGCCTA
MA5 A3#27(1>221) <- GCCAAGGTGTGCAAGTGCCCGGCGGCCAGATGAATGTGGTCGGCGGCGCACCCTTCCTCCACACACACAGCATCAAC
MA5 82KS#28(1>162) -> G-CAAGGTGTGCAAGTGCCCGGCGGCCAGATGAATGTGGTCGGCNNCN-ACCCTTCCTCCACACACACAGCATCAACAGCNAGCTGCCTA
MA5 27KS#25(1>224) -> G-CAAGGTGTGCAAGTGCCCGGCGGCCAGATGAATGTGGTCGGCNNCN-NCCCTTCCTCCACACACACAGCATCAACAGCCAGCTGCCTA
MA5 63SK#29(1>262) <- GCCAAGGTGTGCAAGTGCCCGGCGGCCAGATGAATGTGGTCGGCGGNGCACCCTTCCTCNACACACACAGCATCAACAGCCAGCTGCCTA
MA5 69SK#29(1>180) -> G-CAAGGTGTGCAAGTGCCCGGCGGCCAGATGAATGTGGTCGGCGGCGCACCCTTCCTCCACACACACAGCATCAACAGCNNGCTGCCTA
MA5 71KS#28(1>182) <- CACCCTTCCTNNACACACACAGCATCAACAGCCAGCTGCCTA
MA5 88KS#30(1>224) -> CTTCCTCCACACACACAGCATCAACAGCCAGCTGCCTA
MA5 A14#32(1>247) <- CTTCCTCACACACAGCATCAACAGCCAGCTGCCTA
MA-5 5MARVS#22(1>317) <- GCATCAACAGCCAGCTGCCTA
MA-5 6MARVSEDIT(1>316) -> ACAGCC-GCTGCCTA
MA5 68SK#29(1>253) <- AGCTGCCTA
GcCAAGGTGTGCAAGTGCCCGGCGGCCAGATGAATGTGGTCGGCGGCGcACCCTTCCTCCACACACACAGCATCAACAGCCAGCTGCCTA

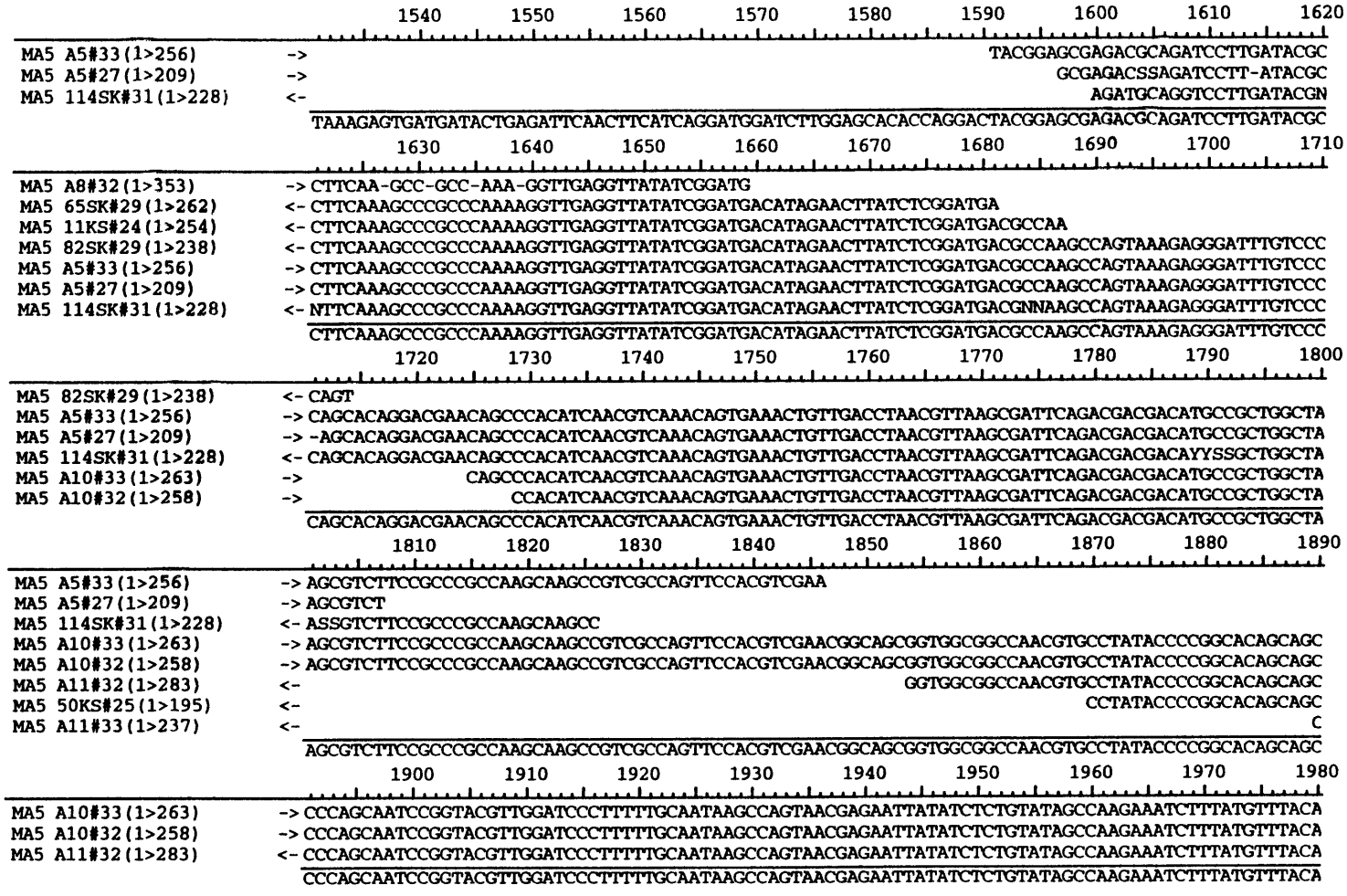
Project: MA Sequencing project Contig 7 Contig 7

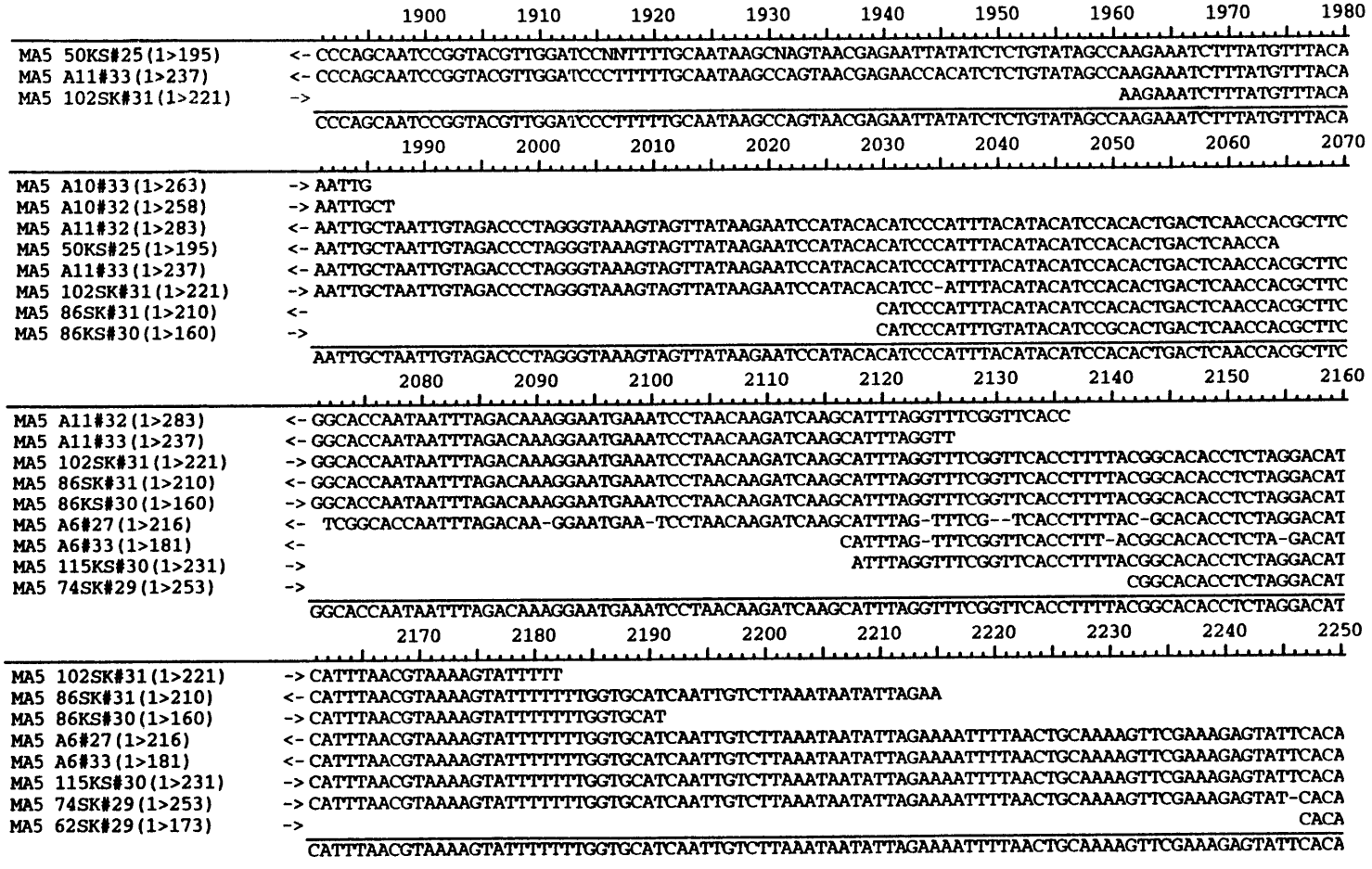
```
                                640       650       660       670       680       690       700       710       720
MA5 A14#34(1>206)          <-                                                                              AGCTGCCTA
                           GcCAAGGTGTGCAAGTGCCCGGCGGCCAGATGAATGTGGTCGGCGGCGcACCCTTCCTCCACACACACAGCATCAACAGCCAGCTGCCTA
                                730       740       750       760       770       780       790       800       810
MA5 A3#33(1>288)           <- TTCACCCCGATGTGCGGCTGAAAAAGCTAGCCTTCTACGATGTACTCG
MA5 31KS#25(1>190)         -> TTCACCCCGAT
MA5 A3#27(1>221)           <- TTCACCCCGATGTGCGGCTGAAAAAGCTAGCCTTCTACGATGTACTC
MA5 27KS#25(1>224)         -> TTCACCCCGATGTGCGGCTGAAAAAG-TAGCCTTCTACGATGTACTCGGAACG-TAATTAA
MA5 63SK#29(1>262)         <- TTCACCCCGATGTGCGGCTGAAAAAGCTAGCCTTCTACGATGTACTCGGAACGCTAATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
MA5 69SK#29(1>180)         -> TTCACCCCGATGTGCGGCTGAAAAAGCTAGCCTTCTACGATGTACTCGGAACGCTAATTAAGCCTTCAACTCTGGTGC
MA5 71KS#28(1>182)         <- TTCACCCCGATGTGCGGCTGAAAAAGCT-GCCTTCTACGATGTACTCGGAACGCTAATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
MA5 88KS#30(1>224)         -> TTCACCCCGATGTGCGGCTGAAAAAGCTAGCCTTCTACGATGTACTCGGAACGCTAATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
MA5 A14#32(1>247)          <- TTCACCCCGATGTGCGGCTGAAAAAGCTAGCCTTCTACGA-GTACTCGG-ACGCTAATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
MA-5 5MARVS#22(1>317)      <- TT-ACCCCGATGTGCGGCTG-AAAAGCTAGCCTTCTACGATGTACTCGGAACGCTAATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
MA-5 6MARVSEDIT(1>316)     -> TTCACCC-GATGTGCGGCTGAAAAAGCTAGCCTTCTACGATGTACTCGGAACGCTAATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
MA5 68SK#29(1>253)         <- TTCACCC-GATGTGCGGCTGAAAA-GCTAGCCTTCT-CGATGTACTCGGAACGCTAATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
MA5 A14#34(1>206)          <- TTCA-CCCGATGTGCGGCTGAAAAAGCTAGCCTTCTACGATGTACTCGGAACGCTAATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
MA5 116SK#31(1>215)        ->                                                                  AATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
                           TTCACCCCGATGTGCGGCTGAAAAAGCTAGCCTTCTACGATGTACTCGGAACGCTAATTAAGCCTTCAACTCTGGTGCCACGCAACACTC
                                820       830       840       850       860       870       880       890       900
MA5 63SK#29(1>262)         <- AGCGCGTCCAAGAGGTGCCTTTCTACTTCAC
MA5 71KS#28(1>182)         <- AGCGCGTCCAAGAGGTGCCTTTCTACTTCAC-C-SACGCCGCAGC
MA5 88KS#30(1>224)         -> AGCGCGTCCAAGAGGTGCCTTTCTACTTCACGCTCACGCCGCAGCAGGCCACCGAGATTGCCTCCAATCGCGACATTCGCAACAG-TCCA
MA5 A14#32(1>247)          <- AGCGCGTCCAAGAGGTGCCTTTCTACTTCACGCTCACGCCGCAGCAGGCCACCGAGATTGCCTCCAATCGCGACATTCGCAACAGCTCCA
MA-5 5MARVS#22(1>317)      <- AGNGCGTCCAAGAGGTGCCTTTCTACTTCACG-TCACGCCGCAGCAGGCCACCGAGATTGCCTCCAATCGCGACAT-NGCAACAS-TCCA
MA-5 6MARVSEDIT(1>316)     -> AGNGCGTCCAAGAGGTGCCTTTCTACTTCACG-TCACGCCGCAGCAGGCCACCGAGATTGCCTCCAATCGCGACATTCGCAACAGCTCCA
MA5 68SK#29(1>253)         <- AGCGCGTCCAAGAGGTGCCTTTCTACTTCACG-TCACGCCGCAGCAGGCCACCGAGATTGCCTCCAATCGCGACATNNGCAACAGCTCCA
MA5 A14#34(1>206)          <- AGCGCGTCCAAGAGGTGCCTTTCTACTTCACGCTCACGCCGCAGCAGGCCACCGAGATTGCCTCCAATCGCGACATTCGCAACAGCTCCA
MA5 116SK#31(1>215)        -> AGCGCGTCCAAGAGGTGCCTTTCTACTTCACGCTCACGCCGCAGCAGGCCACCGAGATTGCCTCCAATCGCGACATTCGCAACAGCTCCA
MA5 19SK#24(1>160)         <-          AGAGGTGCCTTTCTCATTCACGC--NSS-CGCAGCAGGCCACCGAGATTGCCTCCAATCNNGACATNNGCAACASSTCCA
MA5 36KS#25(1>201)         ->                                                  CCGCAGCAGGCCACCGAGATTGCCTCCAATCGCGACATTCGCAACAGCTCCA
MA5 92KS#30(1>205)         <-                                                    CAGGCCACCGAGATTGCCTCCAATCGCGACATTCGCAACASSTCCA
                           AGCGCGTCCAAGAGGTGCCTTTCTACTTCACGcTCACGCCGCAGCAGGCCACCGAGATTGCCTCCAATCGCGACATTCGCAACAGCTCCA
                                910       920       930       940       950       960       970       980       990
MA5 88KS#30(1>224)         -> AGGTGGA
MA5 A14#32(1>247)          <- AGGTGGAGCACGCCATTCAGGTTCAACTGCGC
MA-5 5MARVS#22(1>317)      <- AGGTGGAGCACGCCATTCAGGTTCAACTGCGCTTTTGCCTGGTGGAGACTTCGTGCGACCAGGAGGACTGCTTCCCGCCGAACGTAAACG
MA-5 6MARVSEDIT(1>316)     -> AGGTGGAGCACGCCATTCAGGTTCAACTGCGCTTTTGCCTGGTGGAGACTTCGTGCGACCAGGAGGACTGCTTCCCGCCGAACGTAAACG
                           AGGTGGAGCACGCCATTCAGGTTCAACTGCGCTTTTGCCTGGTGGAGACTTCGTGCGACCAGGAGGACTGCTTCCCGCCGAACGTAAACG
```

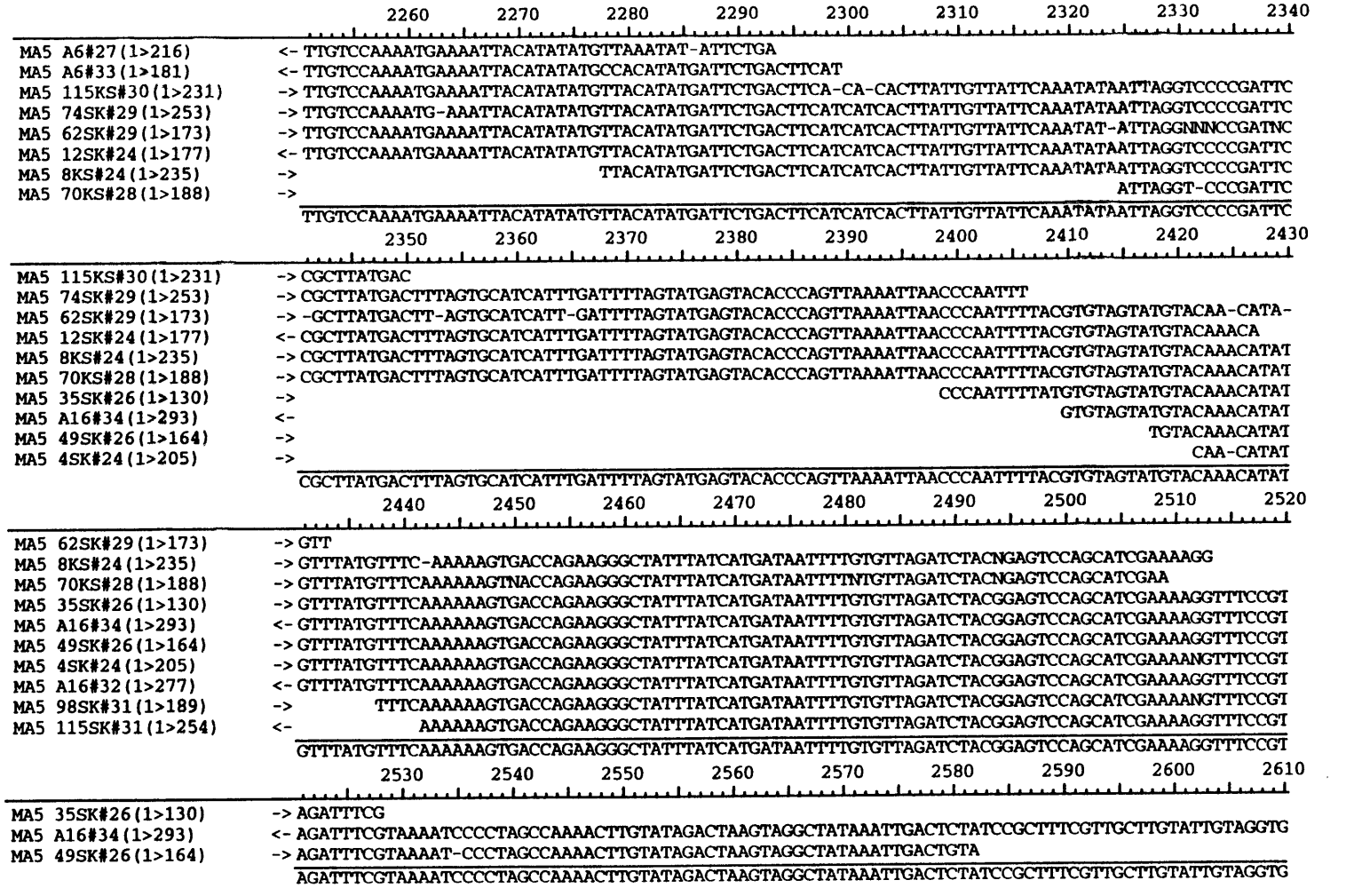

910 920 930 940 950 960 970 980 990
MA5 68SK#29(1>253) <- AGGTGGAGCACGCCATTCAGGTTCAACTGCGCTTTTGCCTGGTGGAGACTTCGTGCGACCAGGAGGAC
MA5 A14#34(1>206) <- AGGTGGAGCACGCCATTC
MA5 116SK#31(1>215) -> AGGTGGASSACGCCATTCAGGTTCAACTGCNCTTTTGCCTGGTGGAGACTTCGTGCGACCAGGAGGACTGCTTCCCGCCGAACGTAAAC
MA5 19SK#24(1>160) <- AGGTGGAGCACGCCATTCAGGTTCAACTGCGCTTTTGCCTGGTGGAGACTTCGTGCGACCAGGAGGACTSSTTCCCGCCGAA
MA5 36KS#25(1>201) -> AGGTGGAGCAGGCCATTCAGGTTCAACTGCGCTTTTGCCTGGTGGAGACTTCGTGCGACCAGGAGGACTACTTCCCNCCGAACGTAAACG
MA5 92KS#30(1>205) <- AGGTGGAGCACGCCATTCAGGTTCAACTGNGCTTTTGCCTGGTGGAGACTTCGTGCGACCAGGAGGACTGCTTCCCGCCGAACGTAAACG
AGGTGGAGCACGCCATTCAGGTTCAACTGCGCTTTTGCCTGGTGGAGACTTCGTGCGACCAGGAGGACTGCTTCCCGCCGAACGTAAACG
1000 1010 1020 1030 1040 1050 1060 1070 1080
MA-5 5MARVS#22(1>317) <- TCAAAGTGAACAACAAACTCTGTCAGCTGC
MA-5 6MARVSEDIT(1>316) -> TCAAAGTGAACAACAAACTCTGTCAGCTGCCT
MA5 36KS#25(1>201) -> TCAAAGTGAACAACAAACTCTGTCAGCTGCCTAATGTGATTGGTACAAACCGACCAAATGTGGAACCCAAGCGT
MA5 92KS#30(1>205) <- TCAAAGTGAACAACAAACTCTGTCAGCTGCCTAATGTCATTCCTACAAACCGACCAAATGCGGAACCCA
MA5 69KS#28(1>159) <- AAGTGAANAANAAACTCTGTCAGCTGCCTAATGTCATTCCTACAAACCGNNCAAATGTGGNACCCAANNGTCNGCCGCGACCCGTTA
MA5 A2#33(1>309) -> AACAACAAACTCTGTCAGCTGCCTAATGTCATTCCTACAAACCGACCAAATGTGGAACCCAAGCGTCCGCCGCGACCCGTTA
MA5 A2#27(1>295) -> TGTCAGCTGCCTAATGTCATTCCTACAAACCGACCAAATGTGGAACCCAAGCGTCCGCCGCGACCCGTTA
MA5 65KS#28(1>199) -> TTGATCTGCCTAATGTCATTCCTACAAACCGACCAAATGTGGAACCCAAGCGTCCGCCGCGACCCGTTA
MA5 31SK#26(1>169) <- GTTA
MA5 112SK#31(1>186) <- GTTA
TCAAAGTGAACAACAAACTCTGTCAGCTGCCTAATGTCATTCCTACAAACCGACCAAATGTGGAACCCAAGCGTCCGCCGCGACCCGTTA
1090 1100 1110 1120 1130 1140 1150 1160 1170
MA5 69KS#28(1>159) <- ATGTCACGTCCAATGTAANNN-GTCGCCTACCGTCACCAACACCATAACGGTTCAGTGGTGTCCG-A-T-CAAGC
MA5 A2#33(1>309) -> ATGTCACGTCCAATGTAAAGCTGTCGCCTACCGTCACCAACACCATAACGGTTCAGTGGTGTCCGGACTACACGCGTAGCTACTGTCTGG
MA5 A2#27(1>295) -> ATGTCACGTCCAATGTAAAGCTGTCGCCTACCGTCACCAACACCATAACGGTTCAGTGGTGTCCGGACTACACGCGTASSTACTGTCTGG
MA5 65KS#28(1>199) -> ATGTCACGTCCAATGTAAAGCTGTCGCCTACCGTCACCAACACCATAACGGTTCAGTGGTGTCCGGACTACACGCGTASSTACTGTCTGG
MA5 31SK#26(1>169) <- ATGTCACGTCCAATGTAAAGCTGTCGCCTACCGTCACCAACACCATAACGGTTCAGTGGTGTCCGGACTACACGCGTAGCTACTGTCTGG
MA5 112SK#31(1>186) <- ATGTCACGTCCAATGTAAASSTGTCGCCTACCGTCACCAACACCATAACGGTTCAGTGGTGTCCGGACTACACGCGTAGCTACTGTCTGG
MA5 A15#32(1>242) -> ACGTC-AATGTAAAGCTGTCGCCTACCGTCACCAACACCATAACGGTTCAGTGGTGTCCGGACTACACGCGTASSTACTGTCTGG
MA5 A4#33(1>284) <- GCTGTCGCCT-CCGTCACCAACACCATAAC-GTTCAGTGGTGTCCGGACTACACGCGTAGCTACTGTCTGG
MA5 A15#34(1>226) -> ACCAACACCATAACGGTTCAGTGGTGTCCGGACTACACGCGTAGCTACTGTCTGG
MA5 A9#32(1>274) <- TGAGCTACTGTCTGG
MA5 A9#33(1>267) <- CTGTCTGG
ATGTCACGTCCAATGTAAAGCTGTCGCCTACCGTCACCAACACCATAACGGTTCAGTGGTGTCCGGACTACACGCGTAGCTACTGTCTGG
1180 1190 1200 1210 1220 1230 1240 1250 1260
MA5 A2#33(1>309) -> CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGGACTACACGCGAGGCT
MA5 A2#27(1>295) -> CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGGACTACACGCGAGGCT
MA5 65KS#28(1>199) -> CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTT
MA5 31SK#26(1>169) <- CCGTATACCTGGTAAAGAAGGTCACCTGAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGG
CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGGACTACACGCGAGGCT

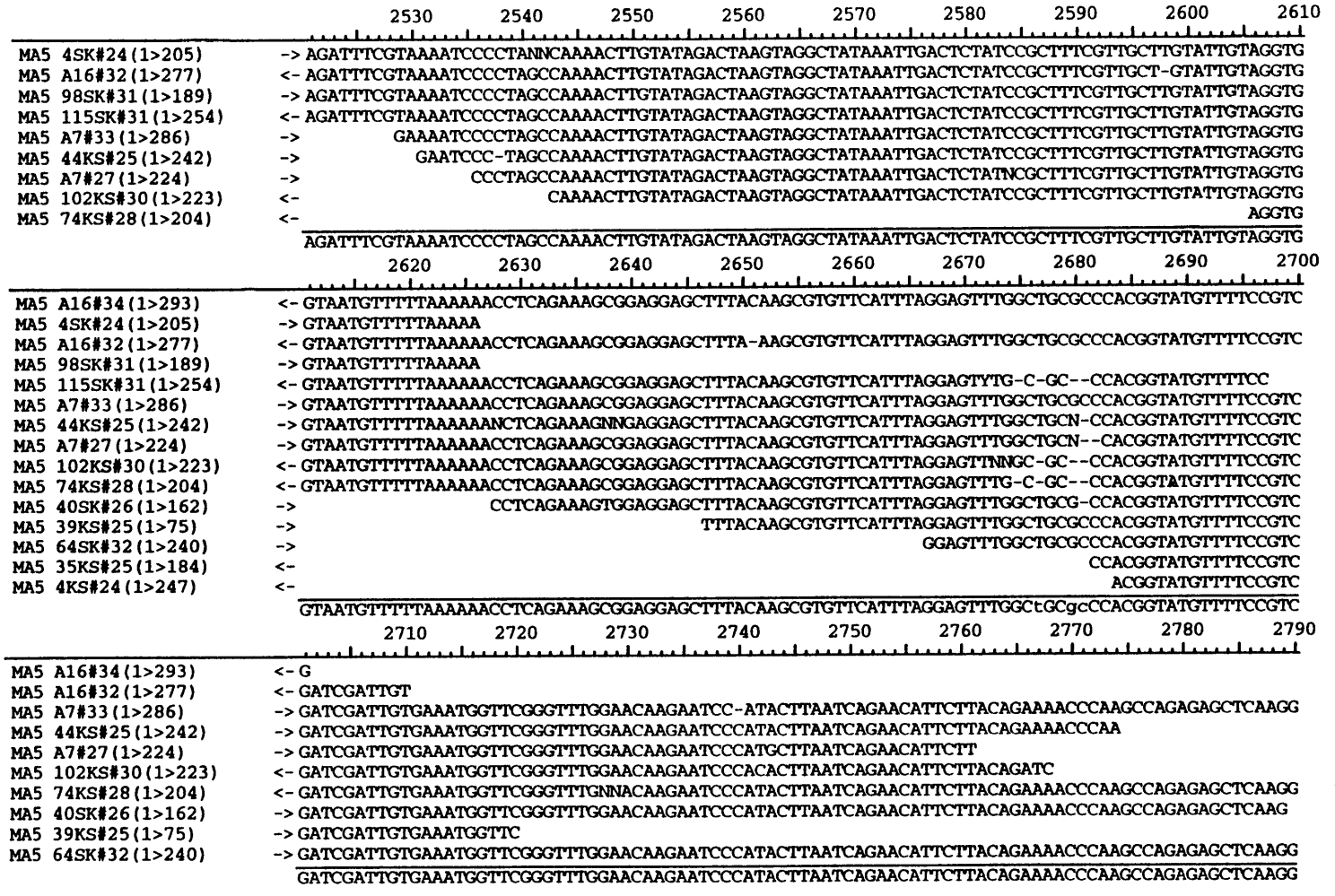

1180 1190 1200 1210 1220 1230 1240 1250 1260
MA5 112SK#31(1>186) <- CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGGACTACMMGCGAGGCT
MA5 A15#32(1>242) -> CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAA-CCAGCGGACTACACGCGAGGCT
MA5 A4#33(1>284) <- CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGGACTACACGCGAGGCT
MA5 A15#34(1>226) -> CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGGACTACACGCGAGGCT
MA5 A9#32(1>274) <- CCGTA--CCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGG-CTACACGCGAGGCT
MA5 A9#33(1>267) <- CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGT-AAACCAGCGGACTACACGCGAGGCT
MA5 27SK#26(1>189) <- ATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAACACGAAGGGNGTAAAACCAGCGGACTACACGCGAGGCT
MA5 114KS#30(1>177) -> ATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGGACTACACGCGAGGCT
MA5 A4#27(1>182) <- CTGGTA-AG-AGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGN-GTAAAACCAGCGGACTACACGCGAGGCT
MA5 A4#29(1>147) <- GCGAATGAAGACGAAGGN-GTAAAACCAGC-GACTACACGCGAGGCT
MA5 116KS#30(1>160) <- AAAACCAGCGGACTACACGCGAGGCT
MA5 116KS#30REAL(1>156) <- AAAACCAGCGGACTACACGCGAGGCT
MA5 116KS#30REAL(1>156) <- AAAACCAGCGGACTACACGCGAGGCT
MA5 88SK#31(1>279) <- GCCAGGCT
CCGTATACCTGGTAAAGAAGCTCACCTCAACACAGCTTTTGCAGCGAATGAAGACGAAGGGCGTAAAACCAGCGGACTACACGCGAGGCT
1270 1280 1290 1300 1310 1320 1330 1340 1350
MA5 A2#33(1>309) -> TAATCAAAGAGAAGCTGACTGAGGATGCTG-CTGCGAA-TAGCCACCAC
MA5 A2#27(1>295) -> TAATCAAAGAGAAGCTGACTGAGGATGCTGACTGCGAA-TAGC-AC
MA5 112SK#31(1>186) <- T
MA5 A15#32(1>242) -> TAATCAAAGAGAAGCTGACTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAA
MA5 A4#33(1>284) <- TAATCAAAGAGAAGCTGACTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTGCCCGTTGGGCAAGATGA
MA5 A15#34(1>226) -> TAATCAAAGAGAAGCTG-CTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTGCCCGTTGGG
MA5 A9#32(1>274) <- TAATCAAAGAGAAGCTGACTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTSSCCGTTGGGCAAGATGA
MA5 A9#33(1>267) <- TAATCAAAGAGAAGCTGACTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTGCCCGTTGGGCAAGATGA
MA5 27SK#26(1>189) <- TAATCAAAGAGAAGCTGACTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACNG---GTTGGGCAAGATGA
MA5 114KS#30(1>177) -> TAATCAAAGAGAAGCTGACTGAGGATGCTGAGTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTGCCCGT-GNN--AGATGA
MA5 A4#27(1>182) <- TAATCAAAGAGAAGCTGACTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTG----TTGGGCAAGATGA
MA5 A4#29(1>147) <- TAATCAAAGAGAAGCTGACTGAGGATGCTGACTGCGAA-TAGCCACCACTATGCTGAAGGTT-CCCTTAACTG----TTGGGCAAGATGA
MA5 116KS#30(1>160) <- TAATCAAAGAGAAGCTGAGTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTG----TTGGGCAAGACGA
MA5 116KS#30REAL(1>156) <- TAATCAAAGAGAAGCTGAGTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTG----TTGGGCAAGACGA
MA5 116KS#30REAL(1>156) <- TAATCAAAGAGAAGCTGAGTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACNG----TTGG-CAAGATGA
MA5 88SK#31(1>279) <- -AATCAAAGAGAAGCTGACTCAGGATGCTGAC-GC-AAATAGCCACCACTATGCTGAAGGTTTCCCTTAACNG----TTGG-CAAGATGA
MA5 A8#32(1>353) -> AAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTGCCCGTTGGGCAAGATCA
MA5 A8#33(1>243) -> AAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTGCCCGTTGGGCAAGATGA
TAATCAAAGAGAAGCTGACTGAGGATGCTGACTGCGAAATAGCCACCACTATGCTGAAGGTTTCCCTTAACTGcccgTTGGGCAAGATGA
1360 1370 1380 1390 1400 1410 1420 1430 1440
MA5 A4#33(1>284) <- AAATGTTGCTGCCTTGTCGAGCATCAACCTGCT
MA5 A9#32(1>274) <- AAATGTTGCTGCCTTGTCGAGCATCAACCTGCTCGCATCTGCAATGCTTCGATGCCAGTCTCTACCTGCAAATGAATGAG-G
AAATGTTGCTGCCTTGTCGAGCATCAACCTGCTCGCATCTGCAATGCTTCGATGCCAGTCTCTACCTGCAAATGAATGAGCGTAAGCCCA

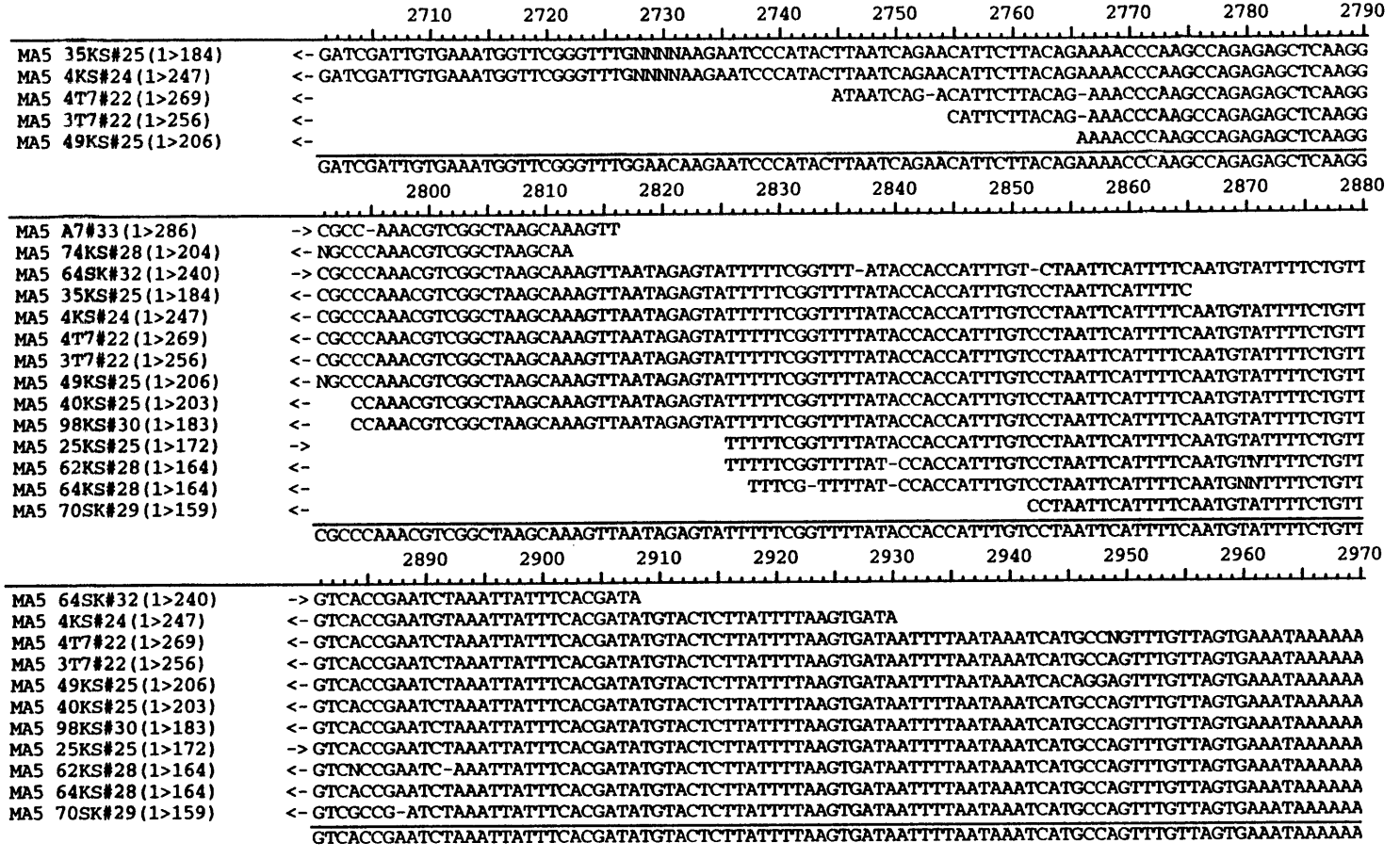

1360 1370 1380 1390 1400 1410 1420 1430 1440
MA5 A9#33(1>267) <- AAATGTTGCTGCCTTGTCGAGCATCAACCTGCTCGCATCTGCAATGCTTCGATGCCAGTCTCTACCTGCAAATGAATGAG
MA5 27SK#26(1>189) <- AAATGTTGCTGCCTTGT
MA5 114KS#30(1>177) -> AAAT
MA5 A4#27(1>182) <- AAATGTTGCTGCCTTG
MA5 A4#29(1>147) <- AAATGTTGCTGCCTTG
MA5 116KS#30(1>160) <- AAATGCTGCTGCCTTGTSSAGCATCAACCTGCTCGCATCTGCAA
MA5 116KS#30REAL(1>156) <- AAATGCTGCTGCCTTGTCGAG-ATCAACCTGCTCGCATCTGCAA
MA5 116KS#30REAL(1>156) <- AAATGCTGCTGCCTTGTCGAG-ATCAACCTGCTCGCATCTGCAA
MA5 88SK#31(1>279) <- AAATGTTGCTGCCTTGTCGAGCATCAACCTGCTCGCATCTGCAATGCTTCGATGCCAGTCTCTACCTGCAAATGAATGAGCGTAAGCCCA
MA5 A8#32(1>353) -> AAATGTTGCTGCCTTGTCGAGCATCAACCTGC-CGCATCTGCAATGCTTCGATGCCAGTCTCTACCTGCAAATGAATGAGCGTAAGCCCA
MA5 A8#33(1>243) -> AAATGTTGCTGCCTTGTCGAGCATCAACCTGCTCGCATCTGCAATGCTTCGATGCCAGTCTCTACCTGCAAATGAATGAGCGTAAGCCCA
MA5 50SK#26(1>165) -> GTTGCTGCCTTGTCGAGCATCAACCTGCTCGCATCTGCAATGCTTGGATGCCAGTCTCTACCTGCAAATGAATGAGCGTANN--CA
MA5 11SK#24(1>223) -> TCGAGCATCAACCTGCTCGCATCTGCAATGCTTCGATGCCAGTCTCTACCTGCAAATGAATGAGCGTNNNN-CA
MA5 12KS#24(1>220) -> TGCTCGCATCTGCAATGCTTCGATGCCAGTCTCTACCTGCAAATGAATGAGCGTNNNN-CA
MA5 65SK#29(1>262) <- CAAGT-AATGAGCGTAAGCCC-
MA5 11KS#24(1>254) <- TAAGCCCA
AAATGTTGCTGCCTTGTCGAGCATCAACCTGCTCGCATCTGCAATGCTTCGATGCCAGTCTCTACCTGCAAATGAATGAGCGTAAGCCCA
1450 1460 1470 1480 1490 1500 1510 1520 1530
MA5 88SK#31(1>279) <- CGTGGAACTGCCCTGTATGCGACAAGCCGGCCATTTATGACAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGSSTCGTCGCTTC
MA5 A8#32(1>353) -> CGTGGAACTGCCCTGTATGCGACAAGCCGGCCATTTATGACAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGGCTCGTCGCTTC
MA5 A8#33(1>243) -> CGTGGAACTGCCCTGTATGCGACAAGCCGGCCATTTATGMMAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGGCTCGTCGCTTC
MA5 50SK#26(1>165) -> CGTGGAACTGCCCTGTATGCGACAANN-GGCCATTTATGACAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGGCTCGTCGCTTC
MA5 11SK#24(1>223) -> CGTGGAACTGCCCTGTATGCGACAAGCCGGCCATTTATGACAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGGCTCGTCGCTTC
MA5 12KS#24(1>220) -> CGTGGAACTGCCCTGTATNCGACAASSCGGCCATTTATGACAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGGCTCGTCGCTTC
MA5 65SK#29(1>262) <- SSTGGAACTGCCCTGTATGCGACAAGCCGGCCATTTATGACAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGGCTCGTCGCTTC
MA5 11KS#24(1>254) <- CGTGGAACTGCCCTGTATGCGACAAGCCGGCCATTTATGACAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGGCTCGTCGCTTC
MA5 82SK#29(1>238) <- CAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGGCTCGTCGCTTC
CGTGGAACTGCCCTGTATGCGACAAGCCGGCCATTTATGACAACCTGGTCATAGATGGCTACTTCCAGGAGGTGTTGGGCTCGTCGCTTC
1540 1550 1560 1570 1580 1590 1600 1610 1620
MA5 88SK#31(1>279) <- TAAAGAGTA
MA5 A8#32(1>353) -> TAAAGAGT-ATGATACTGAGATTCAACTTCATCAGGATGGATCTTGGAGCACACCAGGACTACGGAGCGAGACGCAGATCCTTGATACGC
MA5 A8#33(1>243) -> TAAAGAGTG
MA5 50SK#26(1>165) -> TAAAGAGTGATGATACTGAGATTCAACTTCATCAGGATGGATCTTGGAGC
MA5 11SK#24(1>223) -> TAAAGAGTGATGATACTGAGATTCAACTTCATCAGGATGGATCTTGGAGCACACCAGGACT
MA5 12KS#24(1>220) -> TAAAGAGTGATGATACTGAGATTCAACTTCATCAGGATGGATCTTGGAGCACACCAGGACTAC
MA5 65SK#29(1>262) <- TAAAGAGTGATGATACTGAGATTCAACTTCATCAGGATGGATCTTGGAGCACACCAGGACTACGGAGCGAGACGCAGATCCTTGATACGC
MA5 11KS#24(1>254) <- TAAAGAGTGATGATACTGAGATTCAACTTCATCAGGATGGATCTTGGAGCACACCAGGACTACGGAGCGAGACGCAGATCCTTGATACGC
MA5 82SK#29(1>238) <- TAAAGAGTGATGATACTGAGATTCAACTTCATCAGGATGGATCTTGGAGCACACCAGGACTACGGAGCGAGACGCAGATCCTTGATACGC
TAAAGAGTGATGATACTGAGATTCAACTTCATCAGGATGGATCTTGGAGCACACCAGGACTACGGAGCGAGACGCAGATCCTTGATACGC

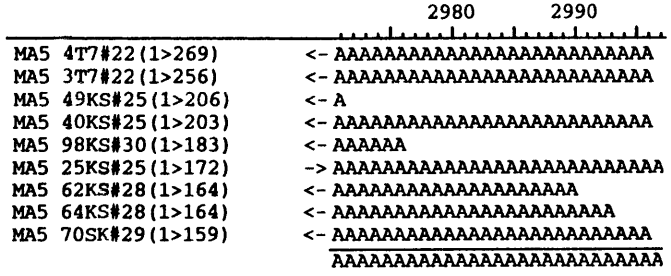

1540 1550 1560 1570 1580 1590 1600 1610 1620
MA5 A5#33(1>256) ->
MA5 A5#27(1>209) ->
MA5 114SK#31(1>228) <-
TAAAGAGTGATGATACTGAGATTCAACTTCATCAGGATGGATCTTGGAGCACACCAGGACTACGGAGCGAGACGCAGATCCTTGATACGC
1630 1640 1650 1660 1670 1680 1690 1700 1710
MA5 A8#32(1>353) -> CTTCAA-GCC-GCC-AAA-GGTTGAGGTTATATCGGATG
MA5 65SK#29(1>262) <- CTTCAAAGCCCGCCCAAAAGGTTGAGGTTATATCGGATGACATAGAACTTATCTCGGATGA
MA5 11KS#24(1>254) <- CTTCAAAGCCCGCCCAAAAGGTTGAGGTTATATCGGATGACATAGAACTTATCTCGGATGACGCCAA
MA5 82SK#29(1>238) <- CTTCAAAGCCCGCCCAAAAGGTTGAGGTTATATCGGATGACATAGAACTTATCTCGGATGACGCCAAGCCAGTAAAGAGGGATTTGTCCC
MA5 A5#33(1>256) -> CTTCAAAGCCCGCCCAAAAGGTTGAGGTTATATCGGATGACATAGAACTTATCTCGGATGACGCCAAGCCAGTAAAGAGGGATTTGTCCC
MA5 A5#27(1>209) -> CTTCAAAGCCCGCCCAAAAGGTTGAGGTTATATCGGATGACATAGAACTTATCTCGGATGACGCCAAGCCAGTAAAGAGGGATTTGTCCC
MA5 114SK#31(1>228) <- NTTCAAAGCCCGCCCAAAAGGTTGAGGTTATATCGGATGACATAGAACTTATCTCGGATGACGNNAAGCCAGTAAAGAGGGATTTGTCCC
CTTCAAAGCCCGCCCAAAAGGTTGAGGTTATATCGGATGACATAGAACTTATCTCGGATGACGCCAAGCCAGTAAAGAGGGATTTGTCCC
1720 1730 1740 1750 1760 1770 1780 1790 1800
MA5 82SK#29(1>238) <- CAGT
MA5 A5#33(1>256) -> CAGCACAGGACGAACAGCCCACATCAACGTCAAACAGTGAAACTGTTGACCTAACGTTAAGCGATTCAGACGACGACATGCCGCTGGCTA
MA5 A5#27(1>209) -> -AGCACAGGACGAACAGCCCACATCAACGTCAAACAGTGAAACTGTTGACCTAACGTTAAGCGATTCAGACGACGACATGCCGCTGGCTA
MA5 114SK#31(1>228) <- CAGCACAGGACGAACAGCCCACATCAACGTCAAACAGTGAAACTGTTGACCTAACGTTAAGCGATTCAGACGACGACAYYSSGCTGGCTA
MA5 A10#33(1>263) -> CAGCCCACATCAACGTCAAACAGTGAAACTGTTGACCTAACGTTAAGCGATTCAGACGACGACATGCCGCTGGCTA
MA5 A10#32(1>258) -> CCACATCAACGTCAAACAGTGAAACTGTTGACCTAACGTTAAGCGATTCAGACGACGACATGCCGCTGGCTA
CAGCACAGGACGAACAGCCCACATCAACGTCAAACAGTGAAACTGTTGACCTAACGTTAAGCGATTCAGACGACGACATGCCGCTGGCTA
1810 1820 1830 1840 1850 1860 1870 1880 1890
MA5 A5#33(1>256) -> AGCGTCTTCCGCCCGCCAAGCAAGCCGTCGCCAGTTCCACGTCGAA
MA5 A5#27(1>209) -> AGCGTCT
MA5 114SK#31(1>228) <- ASSGTCTTCCGCCCGCCAAGCAAGCC
MA5 A10#33(1>263) -> AGCGTCTTCCGCCCGCCAAGCAAGCCGTCGCCAGTTCCACGTCGAACGGCAGCGGTGGCGGCCAACGTGCCTATACCCCGGCACAGCAGC
MA5 A10#32(1>258) -> AGCGTCTTCCGCCCGCCAAGCAAGCCGTCGCCAGTTCCACGTCGAACGGCAGCGGTGGCGGCCAACGTGCCTATACCCCGGCACAGCAGC
MA5 A11#32(1>283) <- GGTGGCGGCCAACGTGCCTATACCCCGGCACAGCAGC
MA5 50KS#25(1>195) <- CCTATACCCCGGCACAGCAGC
MA5 A11#33(1>237) <- C
AGCGTCTTCCGCCCGCCAAGCAAGCCGTCGCCAGTTCCACGTCGAACGGCAGCGGTGGCGGCCAACGTGCCTATACCCCGGCACAGCAGC
1900 1910 1920 1930 1940 1950 1960 1970 1980
MA5 A10#33(1>263) -> CCCAGCAATCCGGTACGTTGGATCCCTTTTTGCAATAAGCCAGTAACGAGAATTATATCTCTGTATAGCCAAGAAATCTTTATGTTTACA
MA5 A10#32(1>258) -> CCCAGCAATCCGGTACGTTGGATCCCTTTTTGCAATAAGCCAGTAACGAGAATTATATCTCTGTATAGCCAAGAAATCTTTATGTTTACA
MA5 A11#32(1>283) <- CCCAGCAATCCGGTACGTTGGATCCCTTTTTGCAATAAGCCAGTAACGAGAATTATATCTCTGTATAGCCAAGAAATCTTTATGTTTACA
CCCAGCAATCCGGTACGTTGGATCCCTTTTTGCAATAAGCCAGTAACGAGAATTATATCTCTGTATAGCCAAGAAATCTTTATGTTTACA


1900 1910 1920 1930 1940 1950 1960 1970 1980
MA5 50KS#25(1>195) <- CCCAGCAATCCGGTACGTTGGATCCNNTTTTGCAATAAGCNAGTAACGAGAATTATATCTCTGTATAGCCAAGAAATCTTTATGTTTACA
MA5 A11#33(1>237) <- CCCAGCAATCCGGTACGTTGGATCCCTTTTTGCAATAAGCCAGTAACGAGAACCACATCTCTGTATAGCCAAGAAATCTTTATGTTTACA
MA5 102SK#31(1>221) -> AAGAAATCTTTATGTTTACA
CCCAGCAATCCGGTACGTTGGATCCCTTTTTGCAATAAGCCAGTAACGAGAATTATATCTCTGTATAGCCAAGAAATCTTTATGTTTACA
1990 2000 2010 2020 2030 2040 2050 2060 2070
MA5 A10#33(1>263) -> AATTG
MA5 A10#32(1>258) -> AATTGCT
MA5 A11#32(1>283) <- AATTGCTAATTGTAGACCCTAGGGTAAAGTAGTTATAAGAATCCATACACATCCCATTTACATACATCCACACTGACTCAACCACGCTTC
MA5 50KS#25(1>195) <- AATTGCTAATTGTAGACCCTAGGGTAAAGTAGTTATAAGAATCCATACACATCCCATTTACATACATCCACACTGACTCAACCA
MA5 A11#33(1>237) <- AATTGCTAATTGTAGACCCTAGGGTAAAGTAGTTATAAGAATCCATACACATCCCATTTACATACATCCACACTGACTCAACCACGCTTC
MA5 102SK#31(1>221) -> AATTGCTAATTGTAGACCCTAGGGTAAAGTAGTTATAAGAATCCATACACATCC-ATTTACATACATCCACACTGACTCAACCACGCTTC
MA5 86SK#31(1>210) <- CATCCCATTTACATACATCCACACTGACTCAACCACGCTTC
MA5 86KS#30(1>160) -> CATCCCATTTGTATACATCCGCACTGACTCAACCACGCTTC
AATTGCTAATTGTAGACCCTAGGGTAAAGTAGTTATAAGAATCCATACACATCCCATTTACATACATCCACACTGACTCAACCACGCTTC
2080 2090 2100 2110 2120 2130 2140 2150 2160
MA5 A11#32(1>283) <- GGCACCAATAATTTAGACAAAGGAATGAAATCCTAACAAGATCAAGCATTTAGGTTTCGGTTCACC
MA5 A11#33(1>237) <- GGCACCAATAATTTAGACAAAGGAATGAAATCCTAACAAGATCAAGCATTTAGGTT
MA5 102SK#31(1>221) -> GGCACCAATAATTTAGACAAAGGAATGAAATCCTAACAAGATCAAGCATTTAGGTTTCGGTTCACCTTTTACGGCACACCTCTAGGACAT
MA5 86SK#31(1>210) <- GGCACCAATAATTTAGACAAAGGAATGAAATCCTAACAAGATCAAGCATTTAGGTTTCGGTTCACCTTTTACGGCACACCTCTAGGACAT
MA5 86KS#30(1>160) -> GGCACCAATAATTTAGACAAAGGAATGAAATCCTAACAAGATCAAGCATTTAGGTTTCGGTTCACCTTTTACGGCACACCTCTAGGACAT
MA5 A6#27(1>216) <- TCGGCACCAATTTAGACAA-GGAATGAA-TCCTAACAAGATCAAGCATTTAG-TTTCG--TCACCTTTTAC-GCACACCTCTAGGACAT
MA5 A6#33(1>181) <- CATTTAG-TTTCGGTTCACCTTT-ACGGCACACCTCTA-GACAT
MA5 115KS#30(1>231) -> ATTTAGGTTTCGGTTCACCTTTTACGGCACACCTCTAGGACAT
MA5 74SK#29(1>253) -> CGGCACACCTCTAGGACAT
GGCACCAATAATTTAGACAAAGGAATGAAATCCTAACAAGATCAAGCATTTAGGTTTCGGTTCACCTTTTACGGCACACCTCTAGGACAT
2170 2180 2190 2200 2210 2220 2230 2240 2250
MA5 102SK#31(1>221) -> CATTTAACGTAAAAGTATTTTT
MA5 86SK#31(1>210) <- CATTTAACGTAAAAGTATTTTTTTGGTGCATCAATTGTCTTAAATAATATTAGAA
MA5 86KS#30(1>160) -> CATTTAACGTAAAAGTATTTTTTTGGTGCAT
MA5 A6#27(1>216) <- CATTTAACGTAAAAGTATTTTTTTGGTGCATCAATTGTCTTAAATAATATTAGAAAATTTTAACTGCAAAAGTTCGAAAGAGTATTCACA
MA5 A6#33(1>181) <- CATTTAACGTAAAAGTATTTTTTTGGTGCATCAATTGTCTTAAATAATATTAGAAAATTTTAACTGCAAAAGTTCGAAAGAGTATTCACA
MA5 115KS#30(1>231) -> CATTTAACGTAAAAGTATTTTTTTGGTGCATCAATTGTCTTAAATAATATTAGAAAATTTTAACTGCAAAAGTTCGAAAGAGTATTCACA
MA5 74SK#29(1>253) -> CATTTAACGTAAAAGTATTTTTTTGGTGCATCAATTGTCTTAAATAATATTAGAAAATTTTAACTGCAAAAGTTCGAAAGAGTAT-CACA
MA5 62SK#29(1>173) -> CACA
CATTTAACGTAAAAGTATTTTTTTGGTGCATCAATTGTCTTAAATAATATTAGAAAATTTTAACTGCAAAAGTTCGAAAGAGTATTCACA

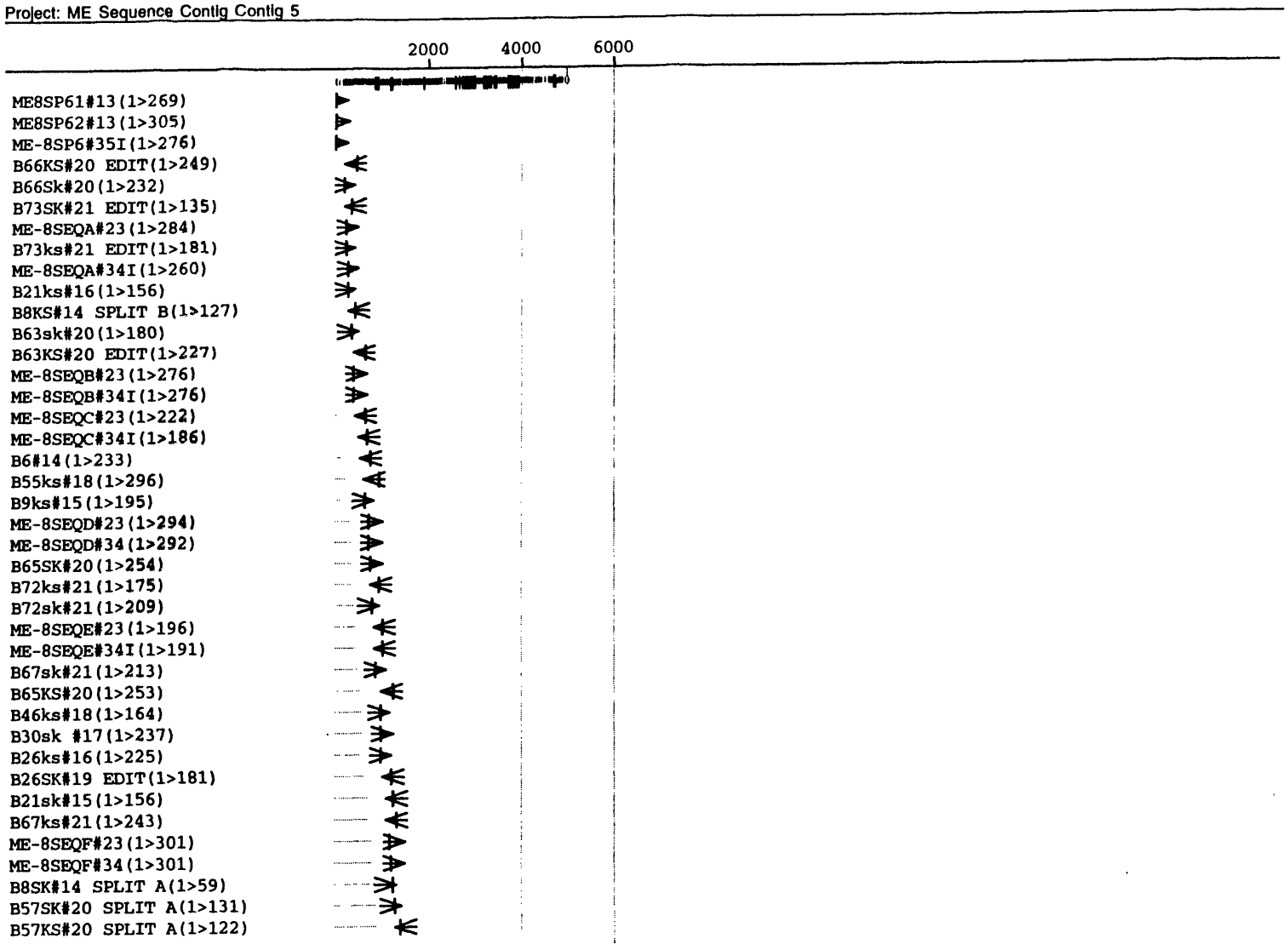

2260 2270 2280 2290 2300 2310 2320 2330 2340
MA5 A6#27(1>216) <- TTGTCCAAAATGAAAATTACATATATGTTAAATAT-ATTCTGA
MA5 A6#33(1>181) <- TTGTCCAAAATGAAAATTACATATATGCCACATATGATTCTGACTTCAT
MA5 115KS#30(1>231) -> TTGTCCAAAATGAAAATTACATATATGTTACATATGATTCTGACTTCA-CA-CACTTATTGTTATTCAAATATAATTAGGTCCCCGATTC
MA5 74SK#29(1>253) -> TTGTCCAAAATG-AAATTACATATATGTTACATATGATTCTGACTTCATCATCACTTATTGTTATTCAAATATAATTAGGTCCCCGATTC
MA5 62SK#29(1>173) -> TTGTCCAAAATGAAAATTACATATATGTTACATATGATTCTGACTTCATCATCACTTATTGTTATTCAAATAT-ATTAGGNNNCCGATNC
MA5 12SK#24(1>177) <- TTGTCCAAAATGAAAATTACATATATGTTACATATGATTCTGACTTCATCATCACTTATTGTTATTCAAATATAATTAGGTCCCCGATTC
MA5 8KS#24(1>235) -> TTACATATGATTCTGACTTCATCATCACTTATTGTTATTCAAATATAATTAGGTCCCCGATTC
MA5 70KS#28(1>188) -> ATTAGGT-CCCGATTC
TTGTCCAAAATGAAAATTACATATATGTTACATATGATTCTGACTTCATCATCACTTATTGTTATTCAAATATAATTAGGTCCCCGATTC
2350 2360 2370 2380 2390 2400 2410 2420 2430
MA5 115KS#30(1>231) -> CGCTTATGAC
MA5 74SK#29(1>253) -> CGCTTATGACTTTAGTGCATCATTTGATTTTAGTATGAGTACACCCAGTTAAAATTAACCCAATTT
MA5 62SK#29(1>173) -> -GCTTATGACTT-AGTGCATCATT-GATTTTAGTATGAGTACACCCAGTTAAAATTAACCCAATTTTACGTGTAGTATGTACAA-CATA-
MA5 12SK#24(1>177) <- CGCTTATGACTTTAGTGCATCATTTGATTTTAGTATGAGTACACCCAGTTAAAATTAACCCAATTTTACGTGTAGTATGTACAAACA
MA5 8KS#24(1>235) -> CGCTTATGACTTTAGTGCATCATTTGATTTTAGTATGAGTACACCCAGTTAAAATTAACCCAATTTTACGTGTAGTATGTACAAACATAT
MA5 70KS#28(1>188) -> CGCTTATGACTTTAGTGCATCATTTGATTTTAGTATGAGTACACCCAGTTAAAATTAACCCAATTTTACGTGTAGTATGTACAAACATAT
MA5 35SK#26(1>130) -> CCCAATTTTATGTGTAGTATGTACAAACATAT
MA5 A16#34(1>293) <- GTGTAGTATGTACAAACATAT
MA5 49SK#26(1>164) -> TGTACAAACATAT
MA5 4SK#24(1>205) -> CAA-CATAT
CGCTTATGACTTTAGTGCATCATTTGATTTTAGTATGAGTACACCCAGTTAAAATTAACCCAATTTTACGTGTAGTATGTACAAACATAT
2440 2450 2460 2470 2480 2490 2500 2510 2520
MA5 62SK#29(1>173) -> GTT
MA5 8KS#24(1>235) -> GTTTATGTTTC-AAAAAGTGACCAGAAGGGCTATTTATCATGATAATTTTGTGTTAGATCTACNGAGTCCAGCATCGAAAAGG
MA5 70KS#28(1>188) -> GTTTATGTTTCAAAAAAGTNACCAGAAGGGCTATTTATCATGATAATTTTNTGTTAGATCTACNGAGTCCAGCATCGAA
MA5 35SK#26(1>130) -> GTTTATGTTTCAAAAAAGTGACCAGAAGGGCTATTTATCATGATAATTTTGTGTTAGATCTACGGAGTCCAGCATCGAAAAGGTTTCCGT
MA5 A16#34(1>293) <- GTTTATGTTTCAAAAAAGTGACCAGAAGGGCTATTTATCATGATAATTTTGTGTTAGATCTACGGAGTCCAGCATCGAAAAGGTTTCCGT
MA5 49SK#26(1>164) -> GTTTATGTTTCAAAAAAGTGACCAGAAGGGCTATTTATCATGATAATTTTGTGTTAGATCTACGGAGTCCAGCATCGAAAAGGTTTCCGT
MA5 4SK#24(1>205) -> GTTTATGTTTCAAAAAAGTGACCAGAAGGGCTATTTATCATGATAATTTTGTGTTAGATCTACGGAGTCCAGCATCGAAAANGTTTCCGT
MA5 A16#32(1>277) <- GTTTATGTTTCAAAAAAGTGACCAGAAGGGCTATTTATCATGATAATTTTGTGTTAGATCTACGGAGTCCAGCATCGAAAAGGTTTCCGT
MA5 98SK#31(1>189) -> TTTCAAAAAAGTGACCAGAAGGGCTATTTATCATGATAATTTTGTGTTAGATCTACGGAGTCCAGCATCGAAAANGTTTCCGT
MA5 115SK#31(1>254) <- AAAAAAGTGACCAGAAGGGCTATTTATCATGATAATTTTGTGTTAGATCTACGGAGTCCAGCATCGAAAAGGTTTCCGT
GTTTATGTTTCAAAAAAGTGACCAGAAGGGCTATTTATCATGATAATTTTGTGTTAGATCTACGGAGTCCAGCATCGAAAAGGTTTCCGT
2530 2540 2550 2560 2570 2580 2590 2600 2610
MA5 35SK#26(1>130) -> AGATTTCG
MA5 A16#34(1>293) <- AGATTTCGTAAAATCCCCTAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCTTGTATTGTAGGTG
MA5 49SK#26(1>164) -> AGATTTCGTAAAAT-CCCTAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTGTA
AGATTTCGTAAAATCCCCTAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCTTGTATTGTAGGTG


2530 2540 2550 2560 2570 2580 2590 2600 2610
MA5 4SK#24(1>205) -> AGATTTCGTAAAATCCCCTANNCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCTTGTATTGTAGGTG
MA5 A16#32(1>277) <- AGATTTCGTAAAATCCCCTAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCT-GTATTGTAGGTG
MA5 98SK#31(1>189) -> AGATTTCGTAAAATCCCCTAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCTTGTATTGTAGGTG
MA5 115SK#31(1>254) <- AGATTTCGTAAAATCCCCTAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCTTGTATTGTAGGTG
MA5 A7#33(1>286) -> GAAAATCCCCTAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCTTGTATTGTAGGTG
MA5 44KS#25(1>242) -> GAATCCC-TAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCTTGTATTGTAGGTG
MA5 A7#27(1>224) -> CCCTAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATNCGCTTTCGTTGCTTGTATTGTAGGTG
MA5 102KS#30(1>223) <- CAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCTTGTATTGTAGGTG
MA5 74KS#28(1>204) <- AGGTG
AGATTTCGTAAAATCCCCTAGCCAAAACTTGTATAGACTAAGTAGGCTATAAATTGACTCTATCCGCTTTCGTTGCTTGTATTGTAGGTG
2620 2630 2640 2650 2660 2670 2680 2690 2700
MA5 A16#34(1>293) <- GTAATGTTTTTAAAAAACCTCAGAAAGCGGAGGAGCTTTACAAGCGTGTTCATTTAGGAGTTTGGCTGCGCCCACGGTATGTTTTCCGTC
MA5 4SK#24(1>205) -> GTAATGTTTTTAAAAA
MA5 A16#32(1>277) <- GTAATGTTTTTAAAAAACCTCAGAAAGCGGAGGAGCTTTA-AAGCGTGTTCATTTAGGAGTTTGGCTGCGCCCACGGTATGTTTTCCGTC
MA5 98SK#31(1>189) -> GTAATGTTTTTAAAAA
MA5 115SK#31(1>254) <- GTAATGTTTTTAAAAAACCTCAGAAAGCGGAGGAGCTTTACAAGCGTGTTCATTTAGGAGTYTG-C-GC--CCACGGTATGTTTTCC
MA5 A7#33(1>286) -> GTAATGTTTTTAAAAAACCTCAGAAAGCGGAGGAGCTTTACAAGCGTGTTCATTTAGGAGTTTGGCTGCGCCCACGGTATGTTTTCCGTC
MA5 44KS#25(1>242) -> GTAATGTTTTTAAAAAANCTCAGAAAGNNGAGGAGCTTTACAAGCGTGTTCATTTAGGAGTTTGGCTGCN-CCACGGTATGTTTTCCGTC
MA5 A7#27(1>224) -> GTAATGTTTTTAAAAAACCTCAGAAAGCGGAGGAGCTTTACAAGCGTGTTCATTTAGGAGTTTGGCTGCN--CACGGTATGTTTTCCGTC
MA5 102KS#30(1>223) <- GTAATGTTTTTAAAAAACCTCAGAAAGCGGAGGAGCTTTACAAGCGTGTTCATTTAGGAGTTNNGC-GC--CCACGGTATGTTTTCCGTC
MA5 74KS#28(1>204) <- GTAATGTTTTTAAAAAACCTCAGAAAGCGGAGGAGCTTTACAAGCGTGTTCATTTAGGAGTTTG-C-GC--CCACGGTATGTTTTCCGTC
MA5 40SK#26(1>162) -> CCTCAGAAAGTGGAGGAGCTTTACAAGCGTGTTCATTTAGGAGTTTGGCTGCG-CCACGGTATGTTTTCCGTC
MA5 39KS#25(1>75) -> TTTACAAGCGTGTTCATTTAGGAGTTTGGCTGCGCCCACGGTATGTTTTCCGTC
MA5 64SK#32(1>240) -> GGAGTTTGGCTGCGCCCACGGTATGTTTTCCGTC
MA5 35KS#25(1>184) <- CCACGGTATGTTTTCCGTC
MA5 4KS#24(1>247) <- ACGGTATGTTTTCCGTC
GTAATGTTTTTAAAAAACCTCAGAAAGCGGAGGAGCTTTACAAGCGTGTTCATTTAGGAGTTTGGCtGCgcCCACGGTATGTTTTCCGTC
2710 2720 2730 2740 2750 2760 2770 2780 2790
MA5 A16#34(1>293) <- G
MA5 A16#32(1>277) <- GATCGATTGT
MA5 A7#33(1>286) -> GATCGATTGTGAAATGGTTCGGGTTTGGAACAAGAATCC-ATACTTAATCAGAACATTCTTACAGAAAACCCAAGCCAGAGAGCTCAAGG
MA5 44KS#25(1>242) -> GATCGATTGTGAAATGGTTCGGGTTTGGAACAAGAATCCCATACTTAATCAGAACATTCTTACAGAAAACCCAA
MA5 A7#27(1>224) -> GATCGATTGTGAAATGGTTCGGGTTTGGAACAAGAATCCCATGCTTAATCAGAACATTCTT
MA5 102KS#30(1>223) <- GATCGATTGTGAAATGGTTCGGGTTTGGAACAAGAATCCCACACTTAATCAGAACATTCTTACAGATC
MA5 74KS#28(1>204) <- GATCGATTGTGAAATGGTTCGGGTTTGNNACAAGAATCCCATACTTAATCAGAACATTCTTACAGAAAACCCAAGCCAGAGAGCTCAAGG
MA5 40SK#26(1>162) -> GATCGATTGTGAAATGGTTCGGGTTTGGAACAAGAATCCCATACTTAATCAGAACATTCTTACAGAAAACCCAAGCCAGAGAGCTCAAG
MA5 39KS#25(1>75) -> GATCGATTGTGAAATGGTTC
MA5 64SK#32(1>240) -> GATCGATTGTGAAATGGTTCGGGTTTGGAACAAGAATCCCATACTTAATCAGAACATTCTTACAGAAAACCCAAGCCAGAGAGCTCAAGG
GATCGATTGTGAAATGGTTCGGGTTTGGAACAAGAATCCCATACTTAATCAGAACATTCTTACAGAAAACCCAAGCCAGAGAGCTCAAGG

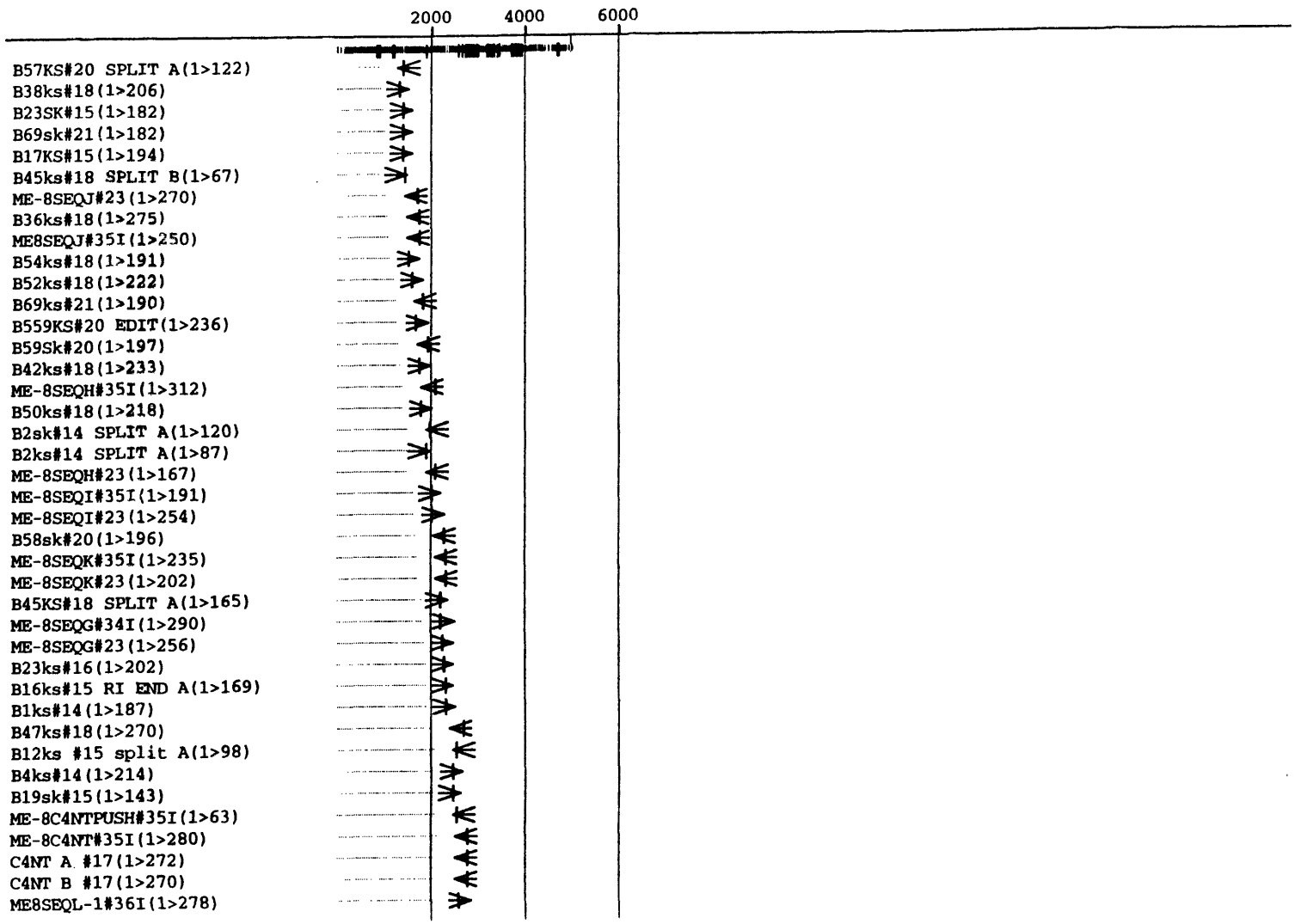

2710 2720 2730 2740 2750 2760 2770 2780 2790
MA5 35KS#25(1>184) <- GATCGATTGTGAAATGGTTCGGGTTTGNNNNAAGAATCCCATACTTAATCAGAACATTCTTACAGAAAACCCAAGCCAGAGAGCTCAAGG
MA5 4KS#24(1>247) <- GATCGATTGTGAAATGGTTCGGGTTTGNNNNAAGAATCCCATACTTAATCAGAACATTCTTACAGAAAACCCAAGCCAGAGAGCTCAAGG
MA5 4T7#22(1>269) <- ATAATCAG-ACATTCTTACAG-AAACCCAAGCCAGAGAGCTCAAGG
MA5 3T7#22(1>256) <- CATTCTTACAG-AAACCCAAGCCAGAGAGCTCAAGG
MA5 49KS#25(1>206) <- AAAACCCAAGCCAGAGAGCTCAAGG
GATCGATTGTGAAATGGTTCGGGTTTGGAACAAGAATCCCATACTTAATCAGAACATTCTTACAGAAAACCCAAGCCAGAGAGCTCAAGG
2800 2810 2820 2830 2840 2850 2860 2870 2880
MA5 A7#33(1>286) -> CGCC-AAACGTCGGCTAAGCAAAGTT
MA5 74KS#28(1>204) <- NGCCCAAACGTCGGCTAAGCAA
MA5 64SK#32(1>240) -> CGCCCAAACGTCGGCTAAGCAAAGTTAATAGAGTATTTTTCGGTTT-ATACCACCATTTGT-CTAATTCATTTTCAATGTATTTTCTGTT
MA5 35KS#25(1>184) <- CGCCCAAACGTCGGCTAAGCAAAGTTAATAGAGTATTTTTCGGTTTTATACCACCATTTGTCCTAATTCATTTTC
MA5 4KS#24(1>247) <- CGCCCAAACGTCGGCTAAGCAAAGTTAATAGAGTATTTTTCGGTTTTATACCACCATTTGTCCTAATTCATTTTCAATGTATTTTCTGTT
MA5 4T7#22(1>269) <- CGCCCAAACGTCGGCTAAGCAAAGTTAATAGAGTATTTTTCGGTTTTATACCACCATTTGTCCTAATTCATTTTCAATGTATTTTCTGTT
MA5 3T7#22(1>256) <- CGCCCAAACGTCGGCTAAGCAAAGTTAATAGAGTATTTTTCGGTTTTATACCACCATTTGTCCTAATTCATTTTCAATGTATTTTCTGTT
MA5 49KS#25(1>206) <- NGCCCAAACGTCGGCTAAGCAAAGTTAATAGAGTATTTTTCGGTTTTATACCACCATTTGTCCTAATTCATTTTCAATGTATTTTCTGTT
MA5 40KS#25(1>203) <- CCAAACGTCGGCTAAGCAAAGTTAATAGAGTATTTTTCGGTTTTATACCACCATTTGTCCTAATTCATTTTCAATGTATTTTCTGTT
MA5 98KS#30(1>183) <- CCAAACGTCGGCTAAGCAAAGTTAATAGAGTATTTTTCGGTTTTATACCACCATTTGTCCTAATTCATTTTCAATGTATTTTCTGTT
MA5 25KS#25(1>172) -> TTTTTCGGTTTTATACCACCATTTGTCCTAATTCATTTTCAATGTATTTTCTGTT
MA5 62KS#28(1>164) <- TTTTTCGGTTTTAT-CCACCATTTGTCCTAATTCATTTTCAATGTNTTTTCTGTT
MA5 64KS#28(1>164) <- TTTCG-TTTTAT-CCACCATTTGTCCTAATTCATTTTCAATGNNTTTTCTGTT
MA5 70SK#29(1>159) <- CCTAATTCATTTTCAATGTATTTTCTGTT
CGCCCAAACGTCGGCTAAGCAAAGTTAATAGAGTATTTTTCGGTTTTATACCACCATTTGTCCTAATTCATTTTCAATGTATTTTCTGTT
2890 2900 2910 2920 2930 2940 2950 2960 2970
MA5 64SK#32(1>240) -> GTCACCGAATCTAAATTATTTCACGATA
MA5 4KS#24(1>247) <- GTCACCGAATGTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATA
MA5 4T7#22(1>269) <- GTCACCGAATCTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCATGCCNGTTTGTTAGTGAAATAAAAAA
MA5 3T7#22(1>256) <- GTCACCGAATCTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCATGCCAGTTTGTTAGTGAAATAAAAAA
MA5 49KS#25(1>206) <- GTCACCGAATCTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCACAGGAGTTTGTTAGTGAAATAAAAAA
MA5 40KS#25(1>203) <- GTCACCGAATCTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCATGCCAGTTTGTTAGTGAAATAAAAAA
MA5 98KS#30(1>183) <- GTCACCGAATCTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCATGCCAGTTTGTTAGTGAAATAAAAAA
MA5 25KS#25(1>172) -> GTCACCGAATCTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCATGCCAGTTTGTTAGTGAAATAAAAAA
MA5 62KS#28(1>164) <- GTCNCCGAATC-AAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCATGCCAGTTTGTTAGTGAAATAAAAAA
MA5 64KS#28(1>164) <- GTCACCGAATCTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCATGCCAGTTTGTTAGTGAAATAAAAAA
MA5 70SK#29(1>159) <- GTCGCCG-ATCTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCATGCCAGTTTGTTAGTGAAATAAAAAA
GTCACCGAATCTAAATTATTTCACGATATGTACTCTTATTTTAAGTGATAATTTTAATAAATCATGCCAGTTTGTTAGTGAAATAAAAAA

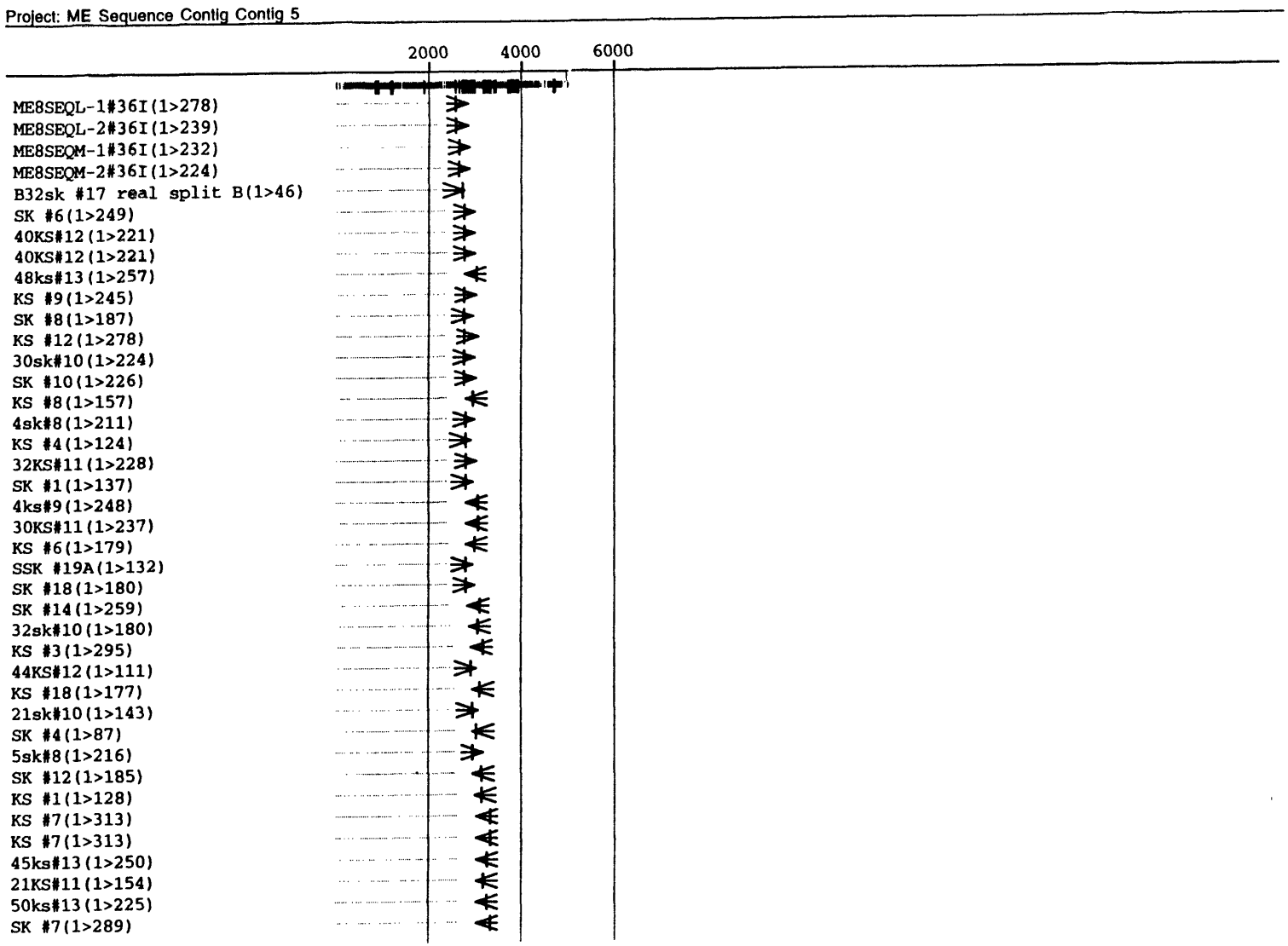

2980
2990
MA5 4T7#22(1>269) <- AAAAAAAAAAAAAAAAAAAAAAAAAA
MA5 3T7#22(1>256) <- AAAAAAAAAAAAAAAAAAAAAAAAAA
MA5 49KS#25(1>206) <- A
MA5 40KS#25(1>203) <- AAAAAAAAAAAAAAAAAAAAAAAAAA
MA5 98KS#30(1>183) <- AAAAAA
MA5 25KS#25(1>172) -> AAAAAAAAAAAAAAAAAAAAAAAAAA
MA5 62KS#28(1>164) <- AAAAAAAAAAAAAAAAAAAA
MA5 64KS#28(1>164) <- AAAAAAAAAAAAAAAAAAAAAAAA
MA5 70SK#29(1>159) <- AAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAA

ME Sequencing project: Contig 5: Strategy View:

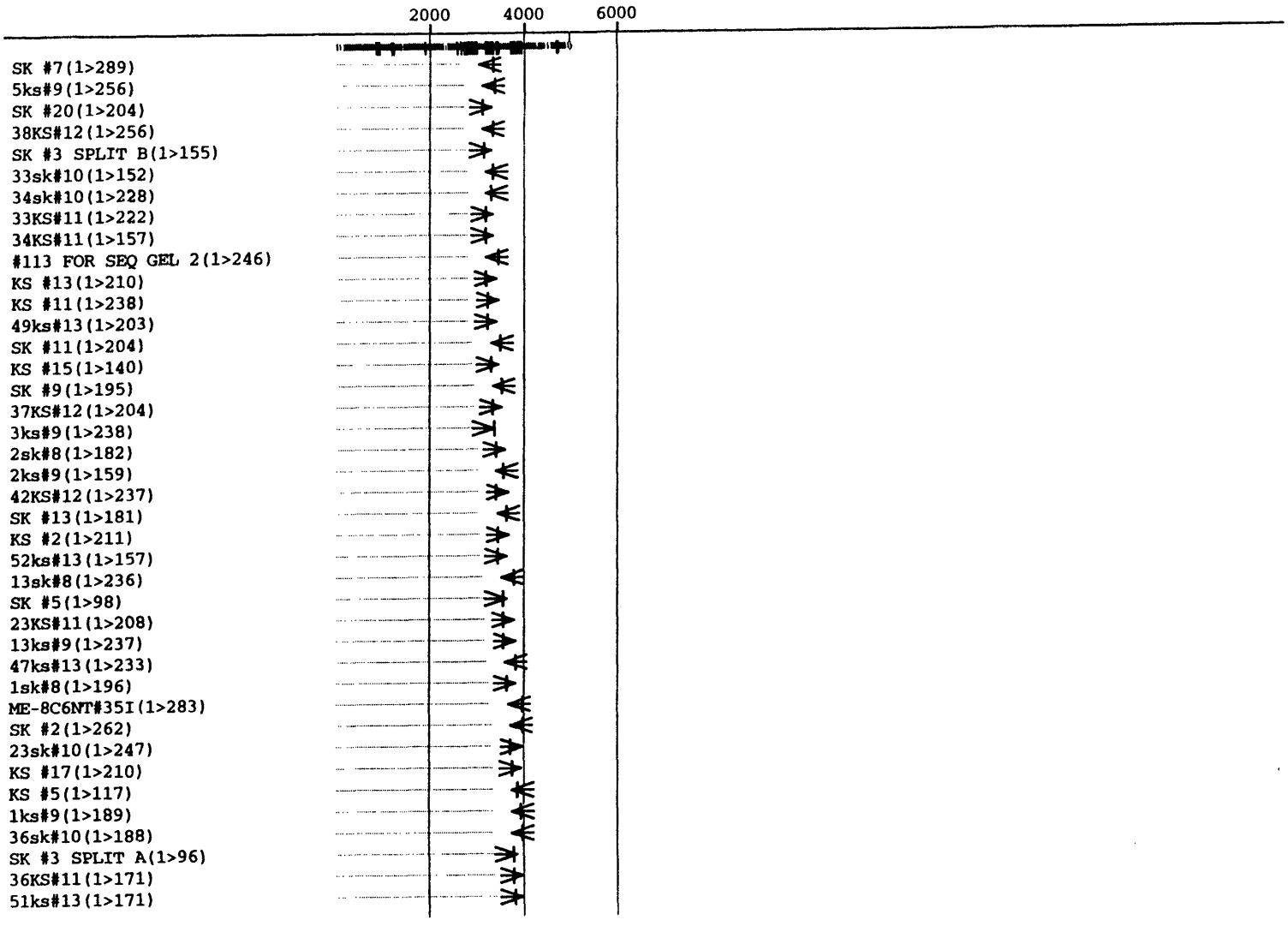

Project: ME Sequence Contig Contig 5
2000
4000
6000
ME8SP61#13(1>269)
ME8SP62#13(1>305)
ME-8SP6#35I(1>276)
B66KS#20 EDIT(1>249)
B66Sk#20(1>232)
B73SK#21 EDIT(1>135)
ME-8SEQA#23(1>284)
B73ks#21 EDIT(1>181)
ME-8SEQA#34I(1>260)
B21ks#16(1>156)
B8KS#14 SPLIT B(1>127)
B63sk#20(1>180)
B63KS#20 EDIT(1>227)
ME-8SEQB#23(1>276)
ME-8SEQB#34I(1>276)
ME-8SEQC#23(1>222)
ME-8SEQC#34I(1>186)
B6#14(1>233)
B55ks#18(1>296)
B9ks#15(1>195)
ME-8SEQD#23(1>294)
ME-8SEQD#34(1>292)
B65SK#20(1>254)
B72ks#21(1>175)
B72sk#21(1>209)
ME-8SEQE#23(1>196)
ME-8SEQE#34I(1>191)
B67sk#21(1>213)
B65KS#20(1>253)
B46ks#18(1>164)
B30sk #17(1>237)
B26ks#16(1>225)
B26SK#19 EDIT(1>181)
B21sk#15(1>156)
B67ks#21(1>243)
ME-8SEQF#23(1>301)
ME-8SEQF#34(1>301)
B8SK#14 SPLIT A(1>59)
B57SK#20 SPLIT A(1>131)
B57KS#20 SPLIT A(1>122)

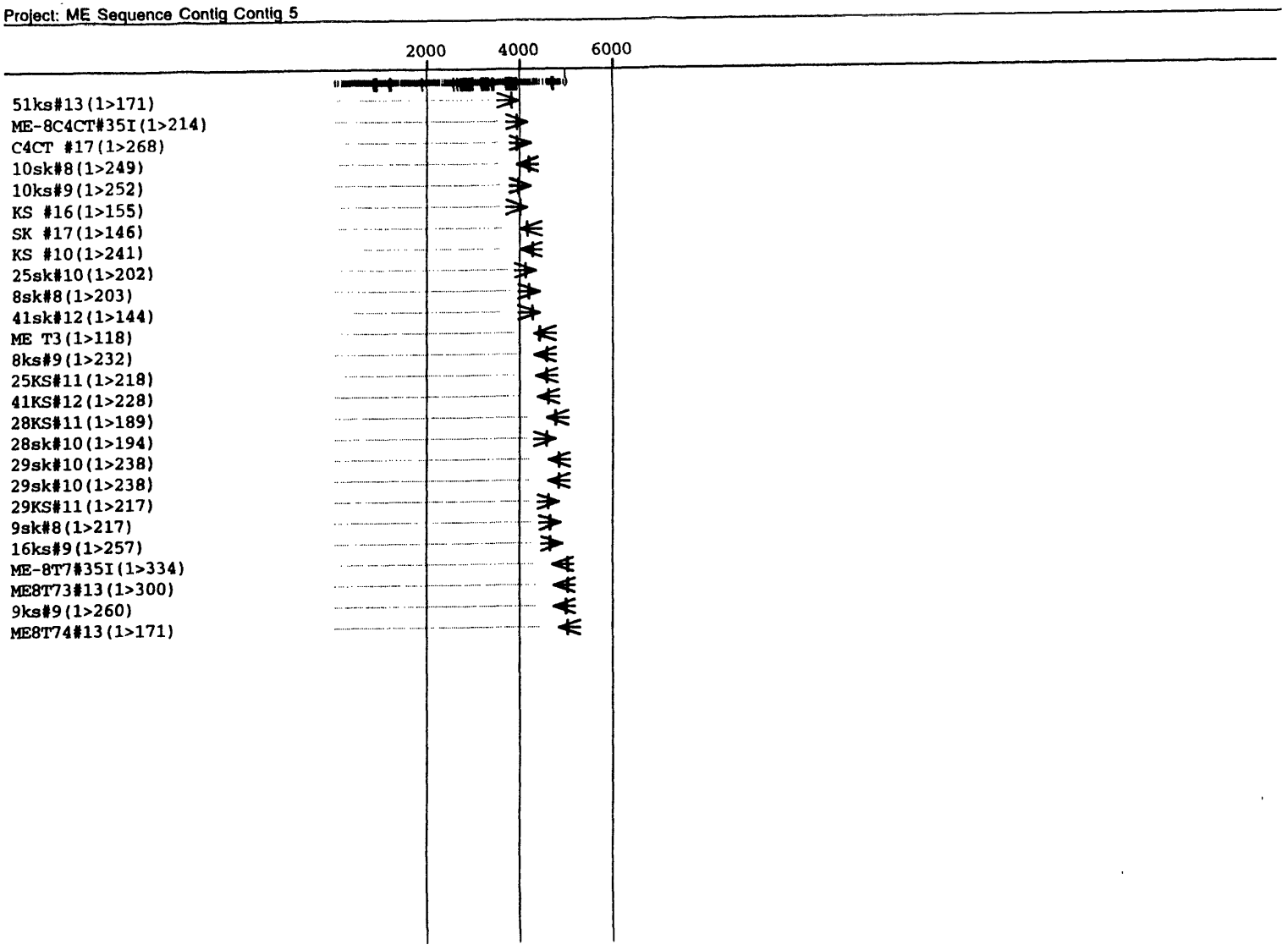
Project: ME Sequence Contig Contig 5



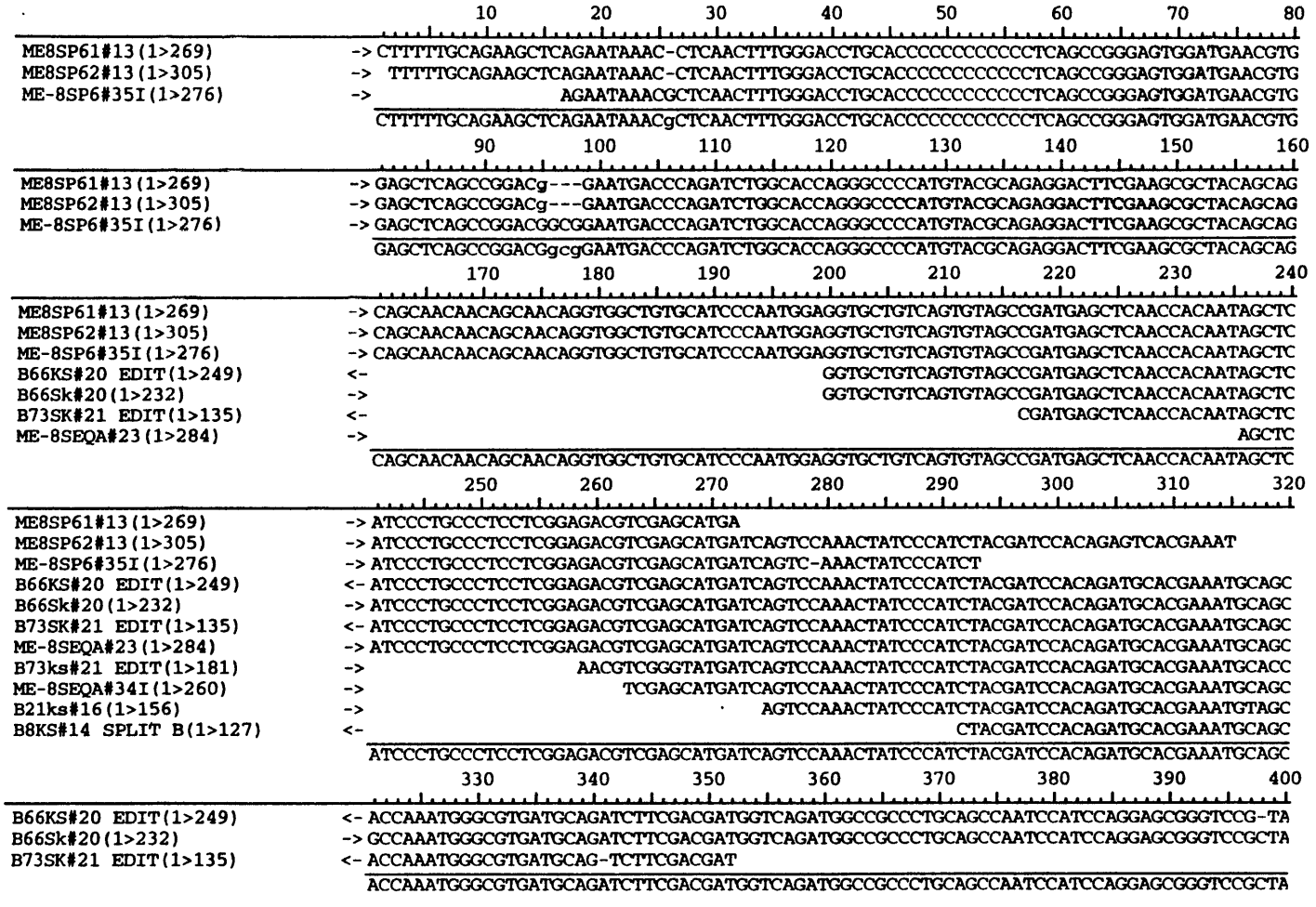

Project: ME Sequence Contig Contig 5
2000
4000
6000
ME8SEQL-1#36I(1>278)
ME8SEQL-2#36I(1>239)
ME8SEQM-1#36I(1>232)
ME8SEQM-2#36I(1>224)
B32sk #17 real split B(1>46)
SK #6(1>249)
40KS#12(1>221)
40KS#12(1>221)
48ks#13(1>257)
KS #9(1>245)
SK #8(1>187)
KS #12(1>278)
30sk#10(1>224)
SK #10(1>226)
KS #8(1>157)
4sk#8(1>211)
KS #4(1>124)
32KS#11(1>228)
SK #1(1>137)
4ks#9(1>248)
30KS#11(1>237)
KS #6(1>179)
SSK #19A(1>132)
SK #18(1>180)
SK #14(1>259)
32sk#10(1>180)
KS #3(1>295)
44KS#12(1>111)
KS #18(1>177)
21sk#10(1>143)
SK #4(1>87)
5sk#8(1>216)
SK #12(1>185)
KS #1(1>128)
KS #7(1>313)
KS #7(1>313)
45ks#13(1>250)
21KS#11(1>154)
50ks#13(1>225)
SK #7(1>289)

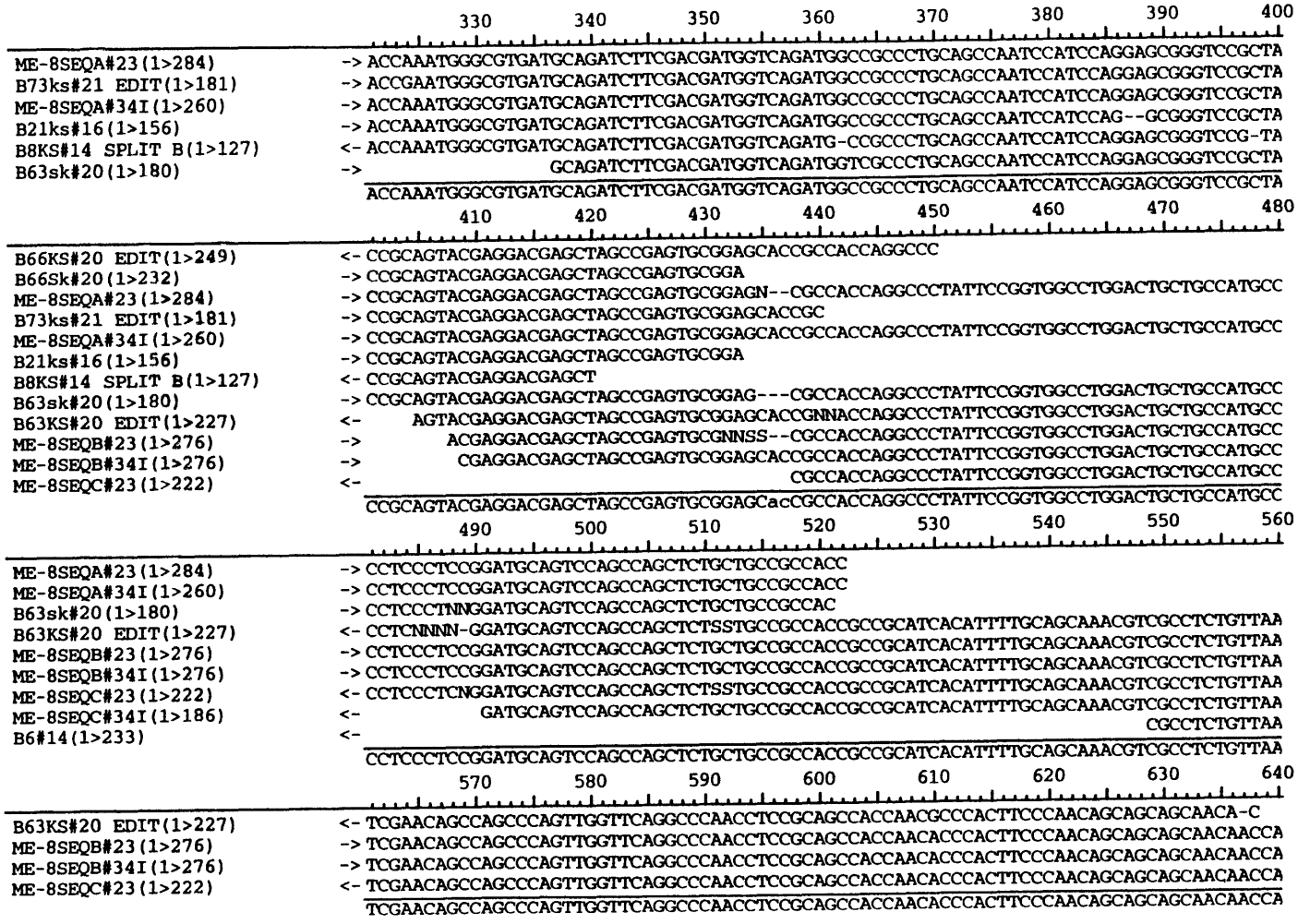
Project: ME Sequence Contig Contig 5

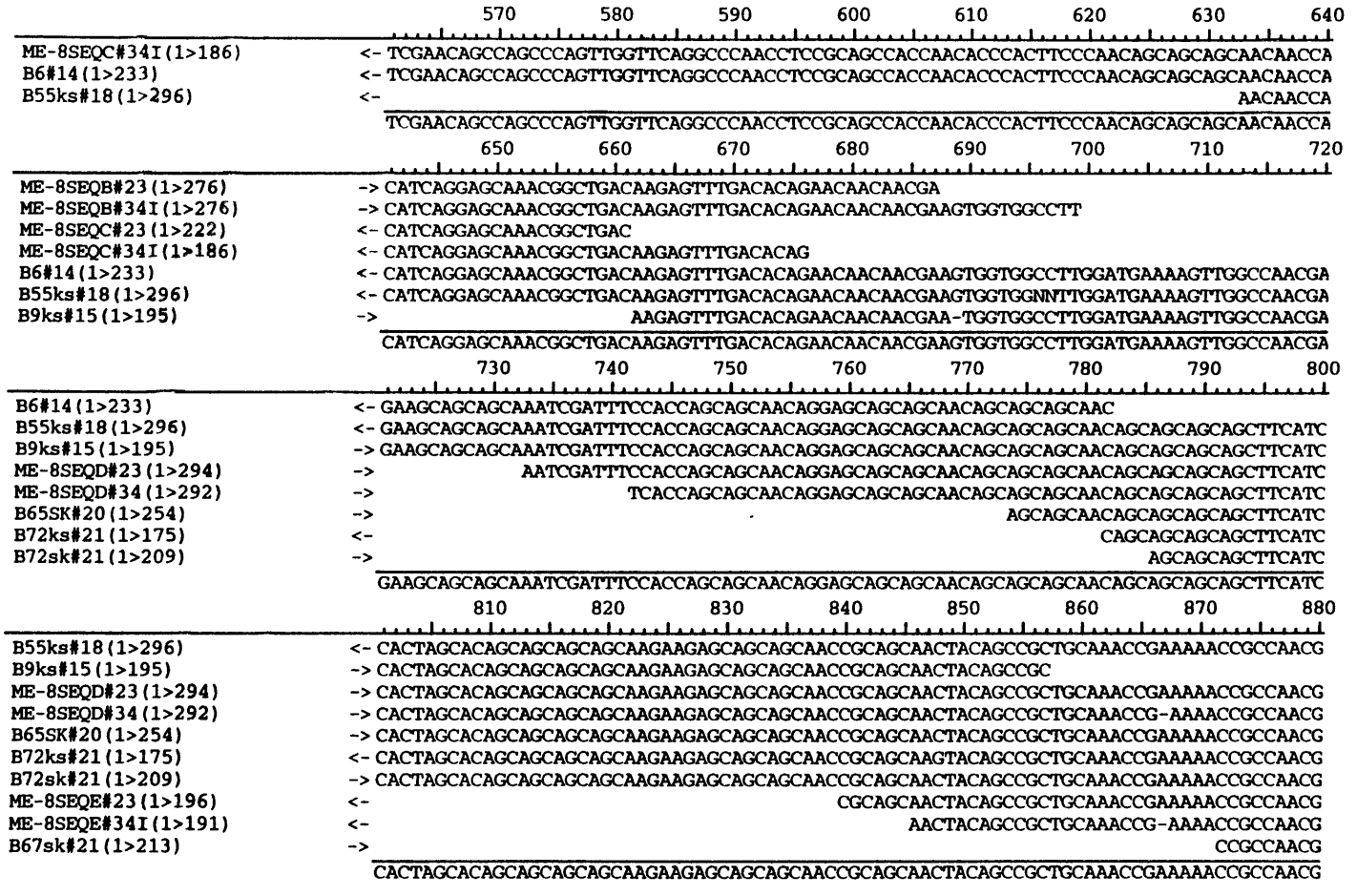



Project: ME Sequence Contig Contig 5
2000
4000
6000
51ks#13(1>171)
ME-8C4CT#35I(1>214)
C4CT #17(1>268)
10sk#8(1>249)
10ks#9(1>252)
KS #16(1>155)
SK #17(1>146)
KS #10(1>241)
25sk#10(1>202)
8sk#8(1>203)
41sk#12(1>144)
ME T3(1>118)
8ks#9(1>232)
25KS#11(1>218)
41KS#12(1>228)
28KS#11(1>189)
28sk#10(1>194)
29sk#10(1>238)
29sk#10(1>238)
29KS#11(1>217)
9sk#8(1>217)
16ks#9(1>257)
ME-8T7#35I(1>334)
ME8T73#13(1>300)
9ks#9(1>260)
ME8T74#13(1>171)

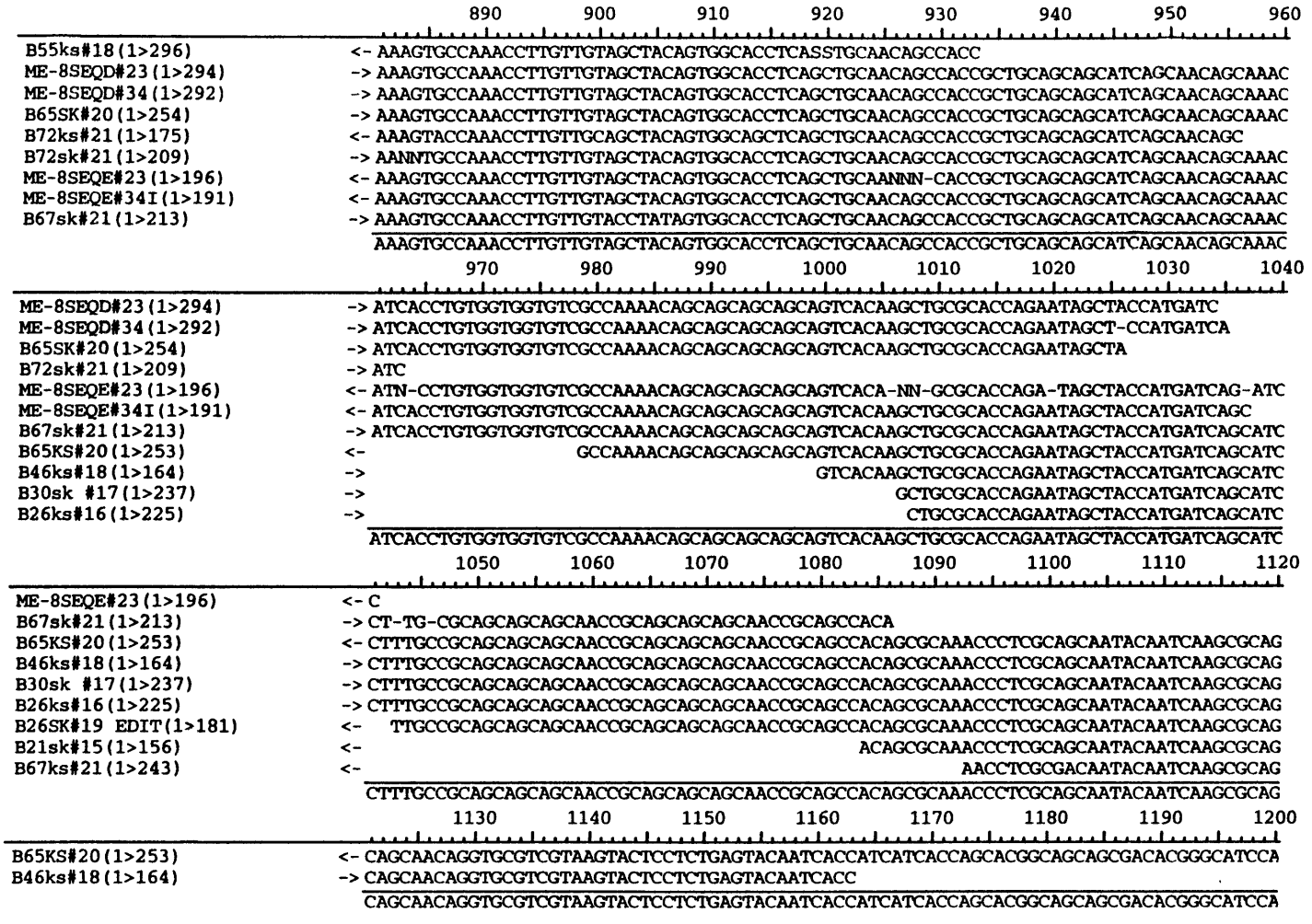
ME Sequencing project: Contig 5: Sequence Alignment:


10 20 30 40 50 60 70 80
ME8SP61#13(1>269) -> CTTTTTGCAGAAGCTCAGAATAAAC-CTCAACTTTGGGACCTGCACCCCCCCCCCCCCCTCAGCCGGGAGTGGATGAACGTG
ME8SP62#13(1>305) -> TTTTTGCAGAAGCTCAGAATAAAC-CTCAACTTTGGGACCTGCACCCCCCCCCCCCCCTCAGCCGGGAGTGGATGAACGTG
ME-8SP6#35I(1>276) -> AGAATAAACGCTCAACTTTGGGACCTGCACCCCCCCCCCCCCCTCAGCCGGGAGTGGATGAACGTG
CTTTTTGCAGAAGCTCAGAATAAACgCTCAACTTTGGGACCTGCACCCCCCCCCCCCCCTCAGCCGGGAGTGGATGAACGTG
90 100 110 120 130 140 150 160
ME8SP61#13(1>269) -> GAGCTCAGCCGGACg---GAATGACCCAGATCTGGCACCAGGGCCCCATGTACGCAGAGGACTTCGAAGCGCTACAGCAG
ME8SP62#13(1>305) -> GAGCTCAGCCGGACg---GAATGACCCAGATCTGGCACCAGGGCCCCATGTACGCAGAGGACTTCGAAGCGCTACAGCAG
ME-8SP6#35I(1>276) -> GAGCTCAGCCGGACGGCGGAATGACCCAGATCTGGCACCAGGGCCCCATGTACGCAGAGGACTTCGAAGCGCTACAGCAG
GAGCTCAGCCGGACGgcgGAATGACCCAGATCTGGCACCAGGGCCCCATGTACGCAGAGGACTTCGAAGCGCTACAGCAG
170 180 190 200 210 220 230 240
ME8SP61#13(1>269) -> CAGCAACAACAGCAACAGGTGGCTGTGCATCCCAATGGAGGTGCTGTCAGTGTAGCCGATGAGCTCAACCACAATAGCTC
ME8SP62#13(1>305) -> CAGCAACAACAGCAACAGGTGGCTGTGCATCCCAATGGAGGTGCTGTCAGTGTAGCCGATGAGCTCAACCACAATAGCTC
ME-8SP6#35I(1>276) -> CAGCAACAACAGCAACAGGTGGCTGTGCATCCCAATGGAGGTGCTGTCAGTGTAGCCGATGAGCTCAACCACAATAGCTC
B66KS#20 EDIT(1>249) <- GGTGCTGTCAGTGTAGCCGATGAGCTCAACCACAATAGCTC
B66Sk#20(1>232) -> GGTGCTGTCAGTGTAGCCGATGAGCTCAACCACAATAGCTC
B73SK#21 EDIT(1>135) <- CGATGAGCTCAACCACAATAGCTC
ME-8SEQA#23(1>284) -> AGCTC
CAGCAACAACAGCAACAGGTGGCTGTGCATCCCAATGGAGGTGCTGTCAGTGTAGCCGATGAGCTCAACCACAATAGCTC
250 260 270 280 290 300 310 320
ME8SP61#13(1>269) -> ATCCCTGCCCTCCTCGGAGACGTCGAGCATGA
ME8SP62#13(1>305) -> ATCCCTGCCCTCCTCGGAGACGTCGAGCATGATCAGTCCAAACTATCCCATCTACGATCCACAGAGTCACGAAAT
ME-8SP6#35I(1>276) -> ATCCCTGCCCTCCTCGGAGACGTCGAGCATGATCAGTC-AAACTATCCCATCT
B66KS#20 EDIT(1>249) <- ATCCCTGCCCTCCTCGGAGACGTCGAGCATGATCAGTCCAAACTATCCCATCTACGATCCACAGATGCACGAAATGCAGC
B66Sk#20(1>232) -> ATCCCTGCCCTCCTCGGAGACGTCGAGCATGATCAGTCCAAACTATCCCATCTACGATCCACAGATGCACGAAATGCAGC
B73SK#21 EDIT(1>135) <- ATCCCTGCCCTCCTCGGAGACGTCGAGCATGATCAGTCCAAACTATCCCATCTACGATCCACAGATGCACGAAATGCAGC
ME-8SEQA#23(1>284) -> ATCCCTGCCCTCCTCGGAGACGTCGAGCATGATCAGTCCAAACTATCCCATCTACGATCCACAGATGCACGAAATGCAGC
B73ks#21 EDIT(1>181) -> AACGTCGGGTATGATCAGTCCAAACTATCCCATCTACGATCCACAGATGCACGAAATGCACC
ME-8SEQA#34I(1>260) -> TCGAGCATGATCAGTCCAAACTATCCCATCTACGATCCACAGATGCACGAAATGCAGC
B21ks#16(1>156) -> AGTCCAAACTATCCCATCTACGATCCACAGATGCACGAAATGTAGC
B8KS#14 SPLIT B(1>127) <- CTACGATCCACAGATGCACGAAATGCAGC
ATCCCTGCCCTCCTCGGAGACGTCGAGCATGATCAGTCCAAACTATCCCATCTACGATCCACAGATGCACGAAATGCAGC
330 340 350 360 370 380 390 400
B66KS#20 EDIT(1>249) <- ACCAAATGGGCGTGATGCAGATCTTCGACGATGGTCAGATGGCCGCCCTGCAGCCAATCCATCCAGGAGCGGGTCCG-TA
B66Sk#20(1>232) -> GCCAAATGGGCGTGATGCAGATCTTCGACGATGGTCAGATGGCCGCCCTGCAGCCAATCCATCCAGGAGCGGGTCCGCTA
B73SK#21 EDIT(1>135) <- ACCAAATGGGCGTGATGCAG-TCTTCGACGAT
ACCAAATGGGCGTGATGCAGATCTTCGACGATGGTCAGATGGCCGCCCTGCAGCCAATCCATCCAGGAGCGGGTCCGCTA

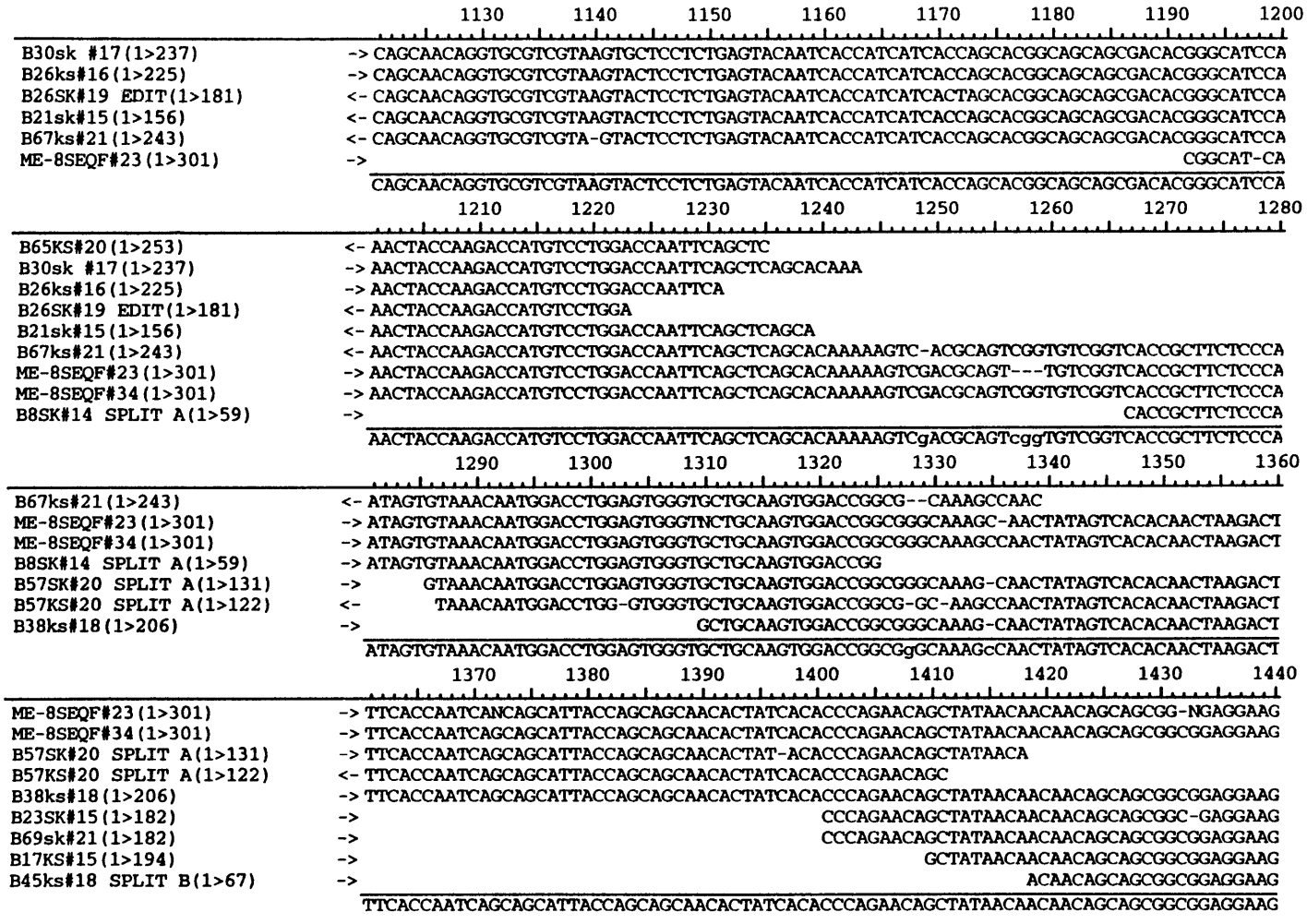

330 340 350 360 370 380 390 400
ME-8SEQA#23(1>284)
B73ks#21 EDIT(1>181)
ME-8SEQA#34I(1>260)
B21ks#16(1>156)
B8KS#14 SPLIT B(1>127)
B63sk#20(1>180)
410 420 430 440 450 460 470 480
B66KS#20 EDIT(1>249)
B66Sk#20(1>232)
ME-8SEQA#23(1>284)
B73ks#21 EDIT(1>181)
ME-8SEQA#34I(1>260)
B21ks#16(1>156)
B8KS#14 SPLIT B(1>127)
B63sk#20(1>180)
B63KS#20 EDIT(1>227)
ME-8SEQB#23(1>276)
ME-8SEQB#34I(1>276)
ME-8SEQC#23(1>222)
490 500 510 520 530 540 550 560
ME-8SEQA#23(1>284)
ME-8SEQA#34I(1>260)
B63sk#20(1>180)
B63KS#20 EDIT(1>227)
ME-8SEQB#23(1>276)
ME-8SEQB#34I(1>276)
ME-8SEQC#23(1>222)
ME-8SEQC#34I(1>186)
B6#14(1>233)
570 580 590 600 610 620 630 640
B63KS#20 EDIT(1>227)
ME-8SEQB#23(1>276)
ME-8SEQB#34I(1>276)
ME-8SEQC#23(1>222)

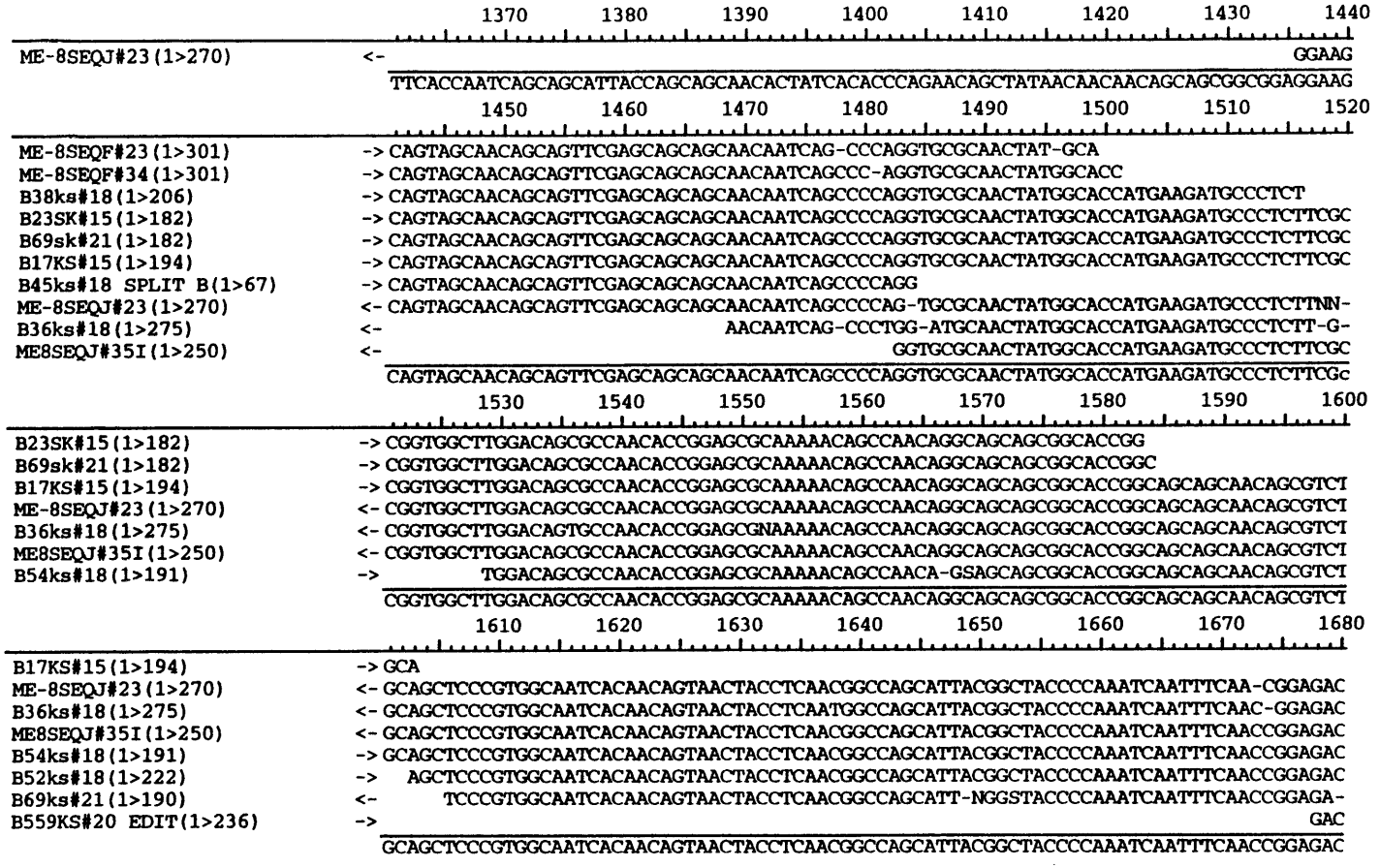

570 580 590 600 610 620 630 640
ME-8SEQC#34I(1>186) <- TCGAACAGCCAGCCCAGTTGGTTCAGGCCCAACCTCCGCAGCCACCAACACCCACTTCCCAACAGCAGCAGCAACAACCA
B6#14(1>233) <- TCGAACAGCCAGCCCAGTTGGTTCAGGCCCAACCTCCGCAGCCACCAACACCCACTTCCCAACAGCAGCAGCAACAACCA
B55ks#18(1>296) <- AACAACCA
TCGAACAGCCAGCCCAGTTGGTTCAGGCCCAACCTCCGCAGCCACCAACACCCACTTCCCAACAGCAGCAGCAACAACCA
650 660 670 680 690 700 710 720
ME-8SEQB#23(1>276) -> CATCAGGAGCAAACGGCTGACAAGAGTTTGACACAGAACAACAACGA
ME-8SEQB#34I(1>276) -> CATCAGGAGCAAACGGCTGACAAGAGTTTGACACAGAACAACAACGAAGTGGTGGCCTT
ME-8SEQC#23(1>222) <- CATCAGGAGCAAACGGCTGAC
ME-8SEQC#34I(1>186) <- CATCAGGAGCAAACGGCTGACAAGAGTTTGACACAG
B6#14(1>233) <- CATCAGGAGCAAACGGCTGACAAGAGTTTGACACAGAACAACAACGAAGTGGTGGCCTTGGATGAAAAGTTGGCCAACGA
B55ks#18(1>296) <- CATCAGGAGCAAACGGCTGACAAGAGTTTGACACAGAACAACAACGAAGTGGTGGNNTTGGATGAAAAGTTGGCCAACGA
B9ks#15(1>195) -> AAGAGTTTGACACAGAACAACAACGAA-TGGTGGCCTTGGATGAAAAGTTGGCCAACGA
CATCAGGAGCAAACGGCTGACAAGAGTTTGACACAGAACAACAACGAAGTGGTGGCCTTGGATGAAAAGTTGGCCAACGA
730 740 750 760 770 780 790 800
B6#14(1>233) <- GAAGCAGCAGCAAATCGATTTCCACCAGCAGCAACAGGAGCAGCAGCAACAGCAGCAGCAAC
B55ks#18(1>296) <- GAAGCAGCAGCAAATCGATTTCCACCAGCAGCAACAGGAGCAGCAGCAACAGCAGCAGCAACAGCAGCAGCAGCTTCATC
B9ks#15(1>195) -> GAAGCAGCAGCAAATCGATTTCCACCAGCAGCAACAGGAGCAGCAGCAACAGCAGCAGCAACAGCAGCAGCAGCTTCATC
ME-8SEQD#23(1>294) -> AATCGATTTCCACCAGCAGCAACAGGAGCAGCAGCAACAGCAGCAGCAACAGCAGCAGCAGCTTCATC
ME-8SEQD#34(1>292) -> TCACCAGCAGCAACAGGAGCAGCAGCAACAGCAGCAGCAACAGCAGCAGCAGCTTCATC
B65SK#20(1>254) -> AGCAGCAACAGCAGCAGCAGCTTCATC
B72ks#21(1>175) <- CAGCAGCAGCAGCTTCATC
B72sk#21(1>209) -> AGCAGCAGCTTCATC
GAAGCAGCAGCAAATCGATTTCCACCAGCAGCAACAGGAGCAGCAGCAACAGCAGCAGCAACAGCAGCAGCAGCTTCATC
810 820 830 840 850 860 870 880
B55ks#18(1>296) <- CACTAGCACAGCAGCAGCAGCAAGAAGAGCAGCAGCAACCGCAGCAACTACAGCCGCTGCAAACCGAAAAACCGCCAACG
B9ks#15(1>195) -> CACTAGCACAGCAGCAGCAGCAAGAAGAGCAGCAGCAACCGCAGCAACTACAGCCGC
ME-8SEQD#23(1>294) -> CACTAGCACAGCAGCAGCAGCAAGAAGAGCAGCAGCAACCGCAGCAACTACAGCCGCTGCAAACCGAAAAACCGCCAACG
ME-8SEQD#34(1>292) -> CACTAGCACAGCAGCAGCAGCAAGAAGAGCAGCAGCAACCGCAGCAACTACAGCCGCTGCAAACCG-AAAACCGCCAACG
B65SK#20(1>254) -> CACTAGCACAGCAGCAGCAGCAAGAAGAGCAGCAGCAACCGCAGCAACTACAGCCGCTGCAAACCGAAAAACCGCCAACG
B72ks#21(1>175) <- CACTAGCACAGCAGCAGCAGCAAGAAGAGCAGCAGCAACCGCAGCAAGTACAGCCGCTGCAAACCGAAAAACCGCCAACG
B72sk#21(1>209) -> CACTAGCACAGCAGCAGCAGCAAGAAGAGCAGCAGCAACCGCAGCAACTACAGCCGCTGCAAACCGAAAAACCGCCAACG
ME-8SEQE#23(1>196) <- CGCAGCAACTACAGCCGCTGCAAACCGAAAAACCGCCAACG
ME-8SEQE#34I(1>191) <- AACTACAGCCGCTGCAAACCG-AAAACCGCCAACG
B67sk#21(1>213) -> CCGCCAACG
CACTAGCACAGCAGCAGCAGCAAGAAGAGCAGCAGCAACCGCAGCAACTACAGCCGCTGCAAACCGAAAAACCGCCAACG

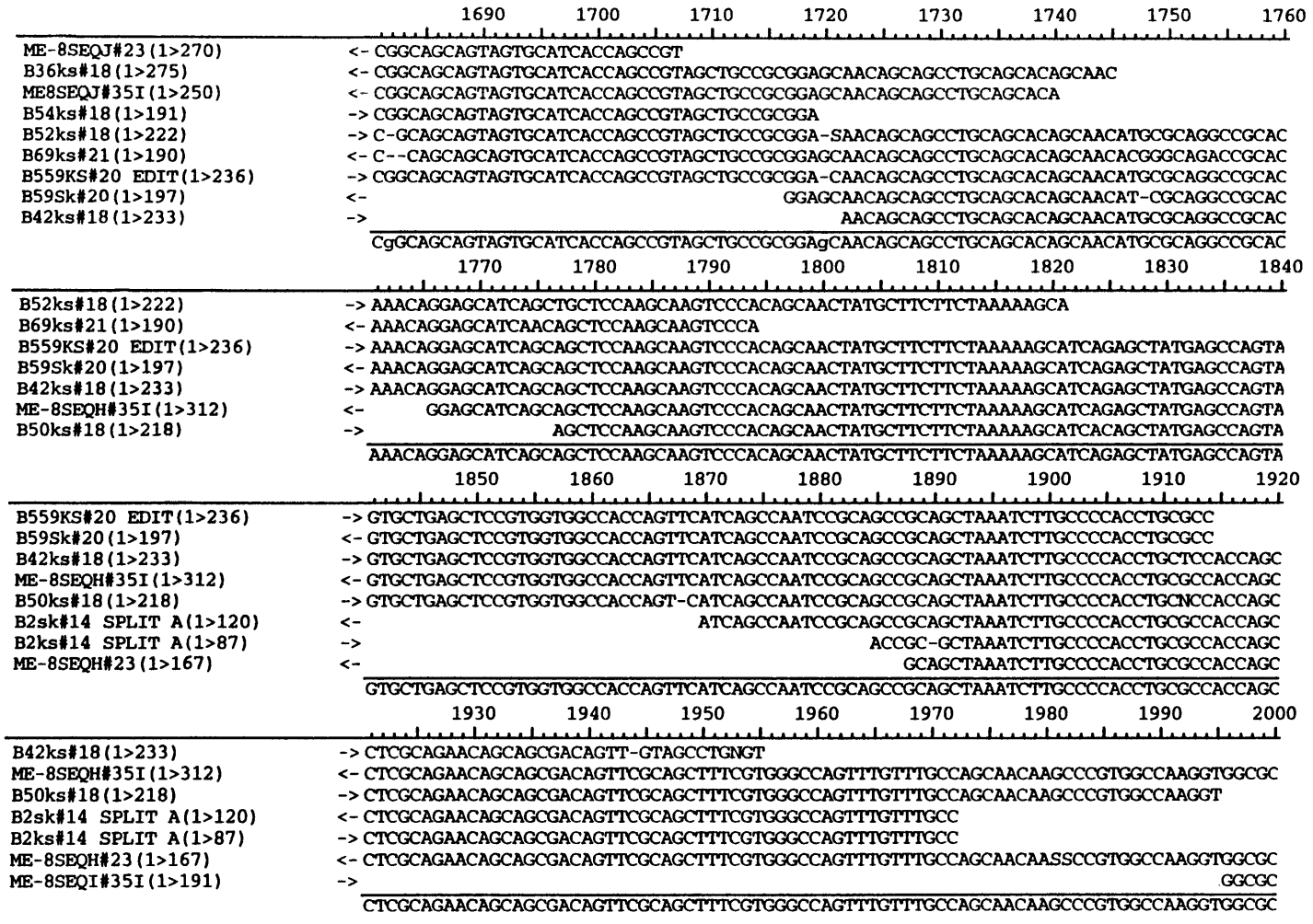
Project: ME Sequence Contig Contig 5



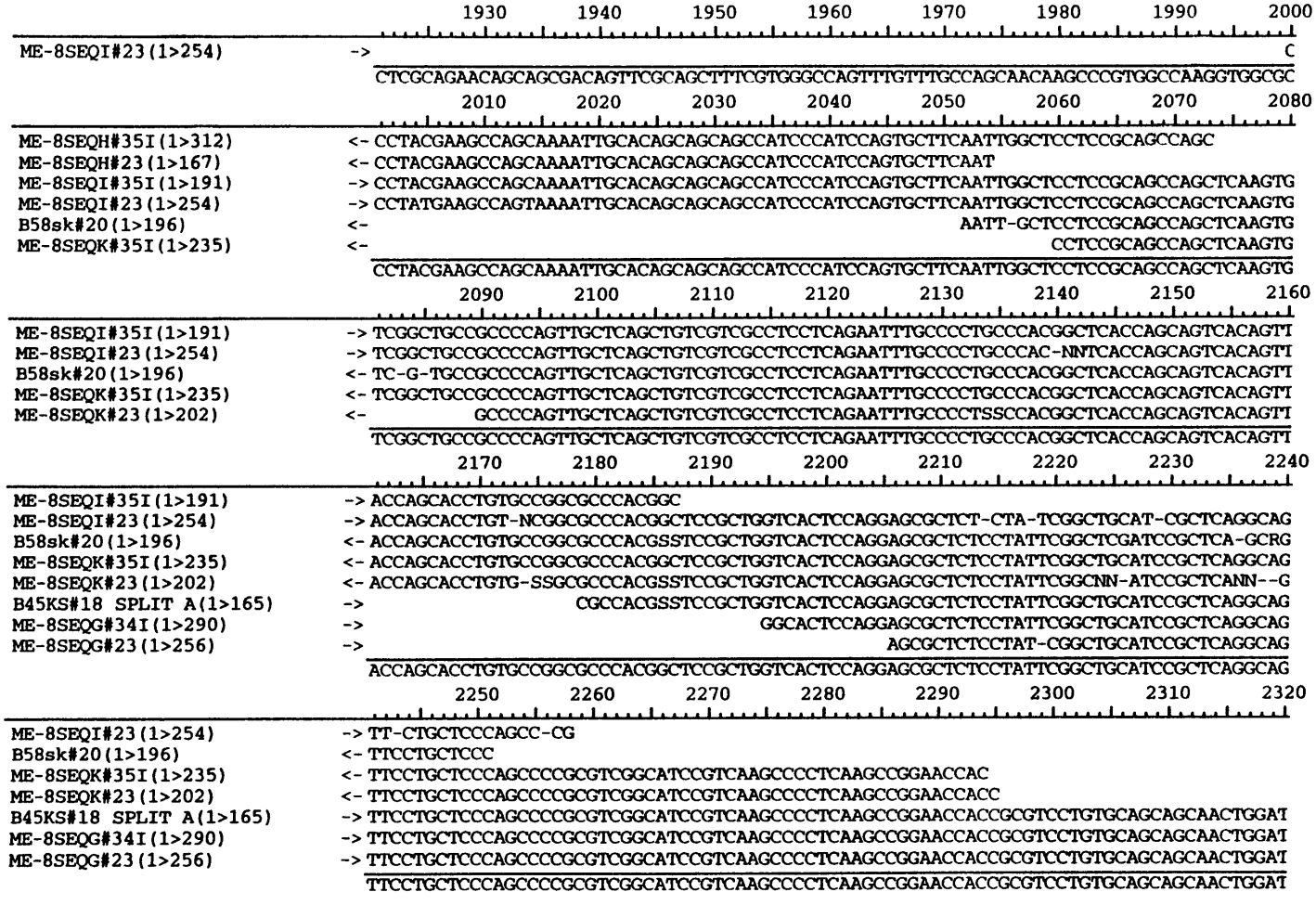
Project: ME Sequence Contig Contig 5

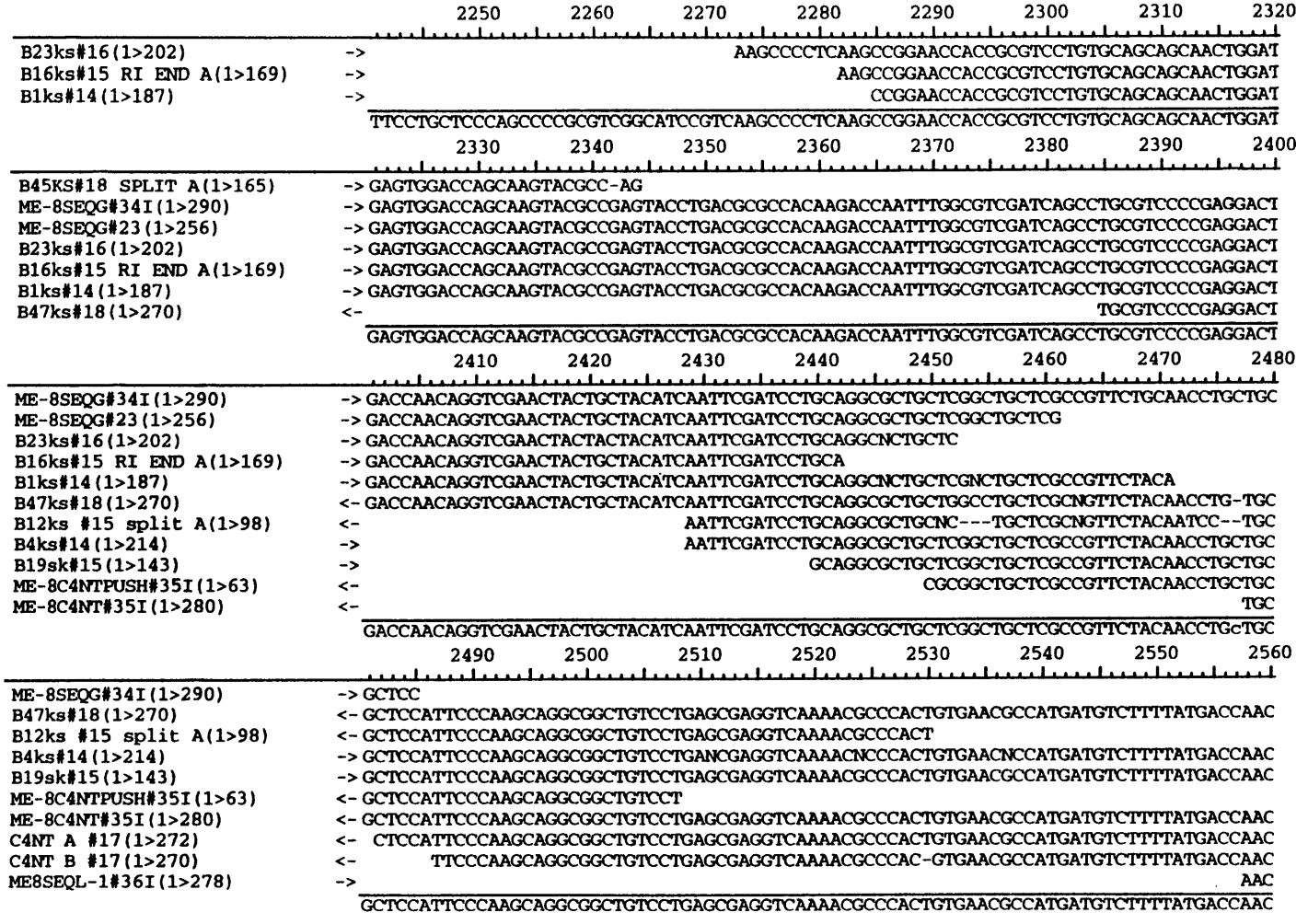


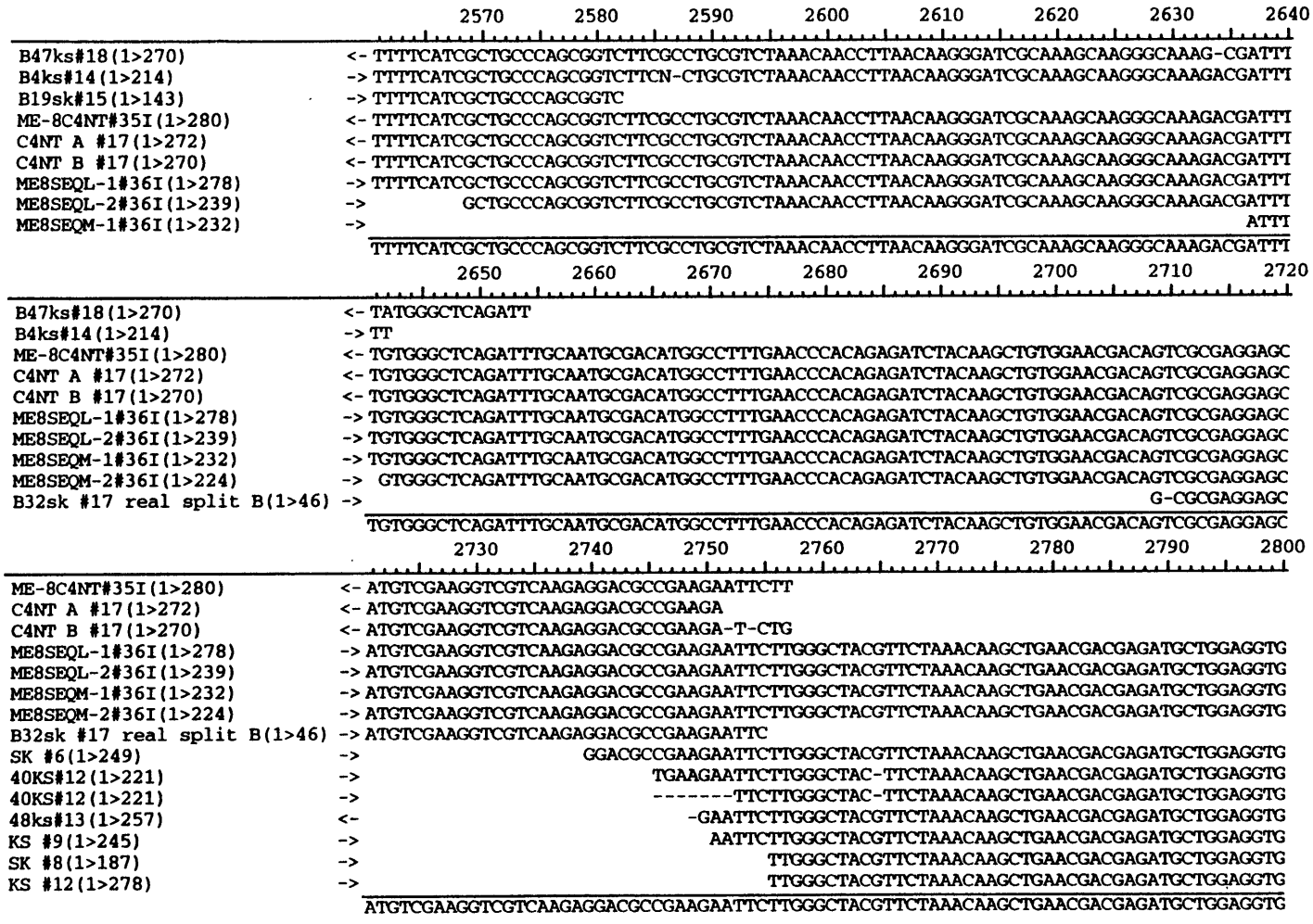

1370 1380 1390 1400 1410 1420 1430 1440
ME-8SEQJ#23(1>270) <- GGAAG
TTCACCAATCAGCAGCATTACCAGCAGCAACACTATCACACCCAGAACAGCTATAACAACAACAGCAGCGGCGGAGGAAG
1450 1460 1470 1480 1490 1500 1510 1520
ME-8SEQF#23(1>301) -> CAGTAGCAACAGCAGTTCGAGCAGCAGCAACAATCAG-CCCAGGTGCGCAACTAT-GCA
ME-8SEQF#34(1>301) -> CAGTAGCAACAGCAGTTCGAGCAGCAGCAACAATCAGCCC-AGGTGCGCAACTATGGCACC
B38ks#18(1>206) -> CAGTAGCAACAGCAGTTCGAGCAGCAGCAACAATCAGCCCCAGGTGCGCAACTATGGCACCATGAAGATGCCCTCT
B23SK#15(1>182) -> CAGTAGCAACAGCAGTTCGAGCAGCAGCAACAATCAGCCCCAGGTGCGCAACTATGGCACCATGAAGATGCCCTCTTCGC
B69sk#21(1>182) -> CAGTAGCAACAGCAGTTCGAGCAGCAGCAACAATCAGCCCCAGGTGCGCAACTATGGCACCATGAAGATGCCCTCTTCGC
B17KS#15(1>194) -> CAGTAGCAACAGCAGTTCGAGCAGCAGCAACAATCAGCCCCAGGTGCGCAACTATGGCACCATGAAGATGCCCTCTTCGC
B45ks#18 SPLIT B(1>67) -> CAGTAGCAACAGCAGTTCGAGCAGCAGCAACAATCAGCCCCAGG
ME-8SEQJ#23(1>270) <- CAGTAGCAACAGCAGTTCGAGCAGCAGCAACAATCAGCCCCAG-TGCGCAACTATGGCACCATGAAGATGCCCTCTTNN-
B36ks#18(1>275) <- AACAATCAG-CCCTGG-ATGCAACTATGGCACCATGAAGATGCCCTCTT-G-
ME8SEQJ#35I(1>250) <- GGTGCGCAACTATGGCACCATGAAGATGCCCTCTTCGC
CAGTAGCAACAGCAGTTCGAGCAGCAGCAACAATCAGCCCCAGGTGCGCAACTATGGCACCATGAAGATGCCCTCTTCGc
1530 1540 1550 1560 1570 1580 1590 1600
B23SK#15(1>182) -> CGGTGGCTTGGACAGCGCCAACACCGGAGCGCAAAAACAGCCAACAGGCAGCAGCGGCACCGG
B69sk#21(1>182) -> CGGTGGCTTGGACAGCGCCAACACCGGAGCGCAAAAACAGCCAACAGGCAGCAGCGGCACCGGC
B17KS#15(1>194) -> CGGTGGCTTGGACAGCGCCAACACCGGAGCGCAAAAACAGCCAACAGGCAGCAGCGGCACCGGCAGCAGCAACAGCGTCT
ME-8SEQJ#23(1>270) <- CGGTGGCTTGGACAGCGCCAACACCGGAGCGCAAAAACAGCCAACAGGCAGCAGCGGCACCGGCAGCAGCAACAGCGTCT
B36ks#18(1>275) <- CGGTGGCTTGGACAGTGCCAACACCGGAGCGNAAAAACAGCCAACAGGCAGCAGCGGCACCGGCAGCAGCAACAGCGTCT
ME8SEQJ#35I(1>250) <- CGGTGGCTTGGACAGCGCCAACACCGGAGCGCAAAAACAGCCAACAGGCAGCAGCGGCACCGGCAGCAGCAACAGCGTCT
B54ks#18(1>191) -> TGGACAGCGCCAACACCGGAGCGCAAAAACAGCCAACA-GSAGCAGCGGCACCGGCAGCAGCAACAGCGTCT
CGGTGGCTTGGACAGCGCCAACACCGGAGCGCAAAAACAGCCAACAGGCAGCAGCGGCACCGGCAGCAGCAACAGCGTCT
1610 1620 1630 1640 1650 1660 1670 1680
B17KS#15(1>194) -> GCA
ME-8SEQJ#23(1>270) <- GCAGCTCCCGTGGCAATCACAACAGTAACTACCTCAACGGCCAGCATTACGGCTACCCCAAATCAATTTCAA-CGGAGAC
B36ks#18(1>275) <- GCAGCTCCCGTGGCAATCACAACAGTAACTACCTCAATGGCCAGCATTACGGCTACCCCAAATCAATTTCAAC-GGAGAC
ME8SEQJ#35I(1>250) <- GCAGCTCCCGTGGCAATCACAACAGTAACTACCTCAACGGCCAGCATTACGGCTACCCCAAATCAATTTCAACCGGAGAC
B54ks#18(1>191) -> GCAGCTCCCGTGGCAATCACAACAGTAACTACCTCAACGGCCAGCATTACGGCTACCCCAAATCAATTTCAACCGGAGAC
B52ks#18(1>222) -> AGCTCCCGTGGCAATCACAACAGTAACTACCTCAACGGCCAGCATTACGGCTACCCCAAATCAATTTCAACCGGAGAC
B69ks#21(1>190) <- TCCCGTGGCAATCACAACAGTAACTACCTCAACGGCCAGCATT-NGGSTACCCCAAATCAATTTCAACCGGAGA-
B559KS#20 EDIT(1>236) -> GAC
GCAGCTCCCGTGGCAATCACAACAGTAACTACCTCAACGGCCAGCATTACGGCTACCCCAAATCAATTTCAACCGGAGAC

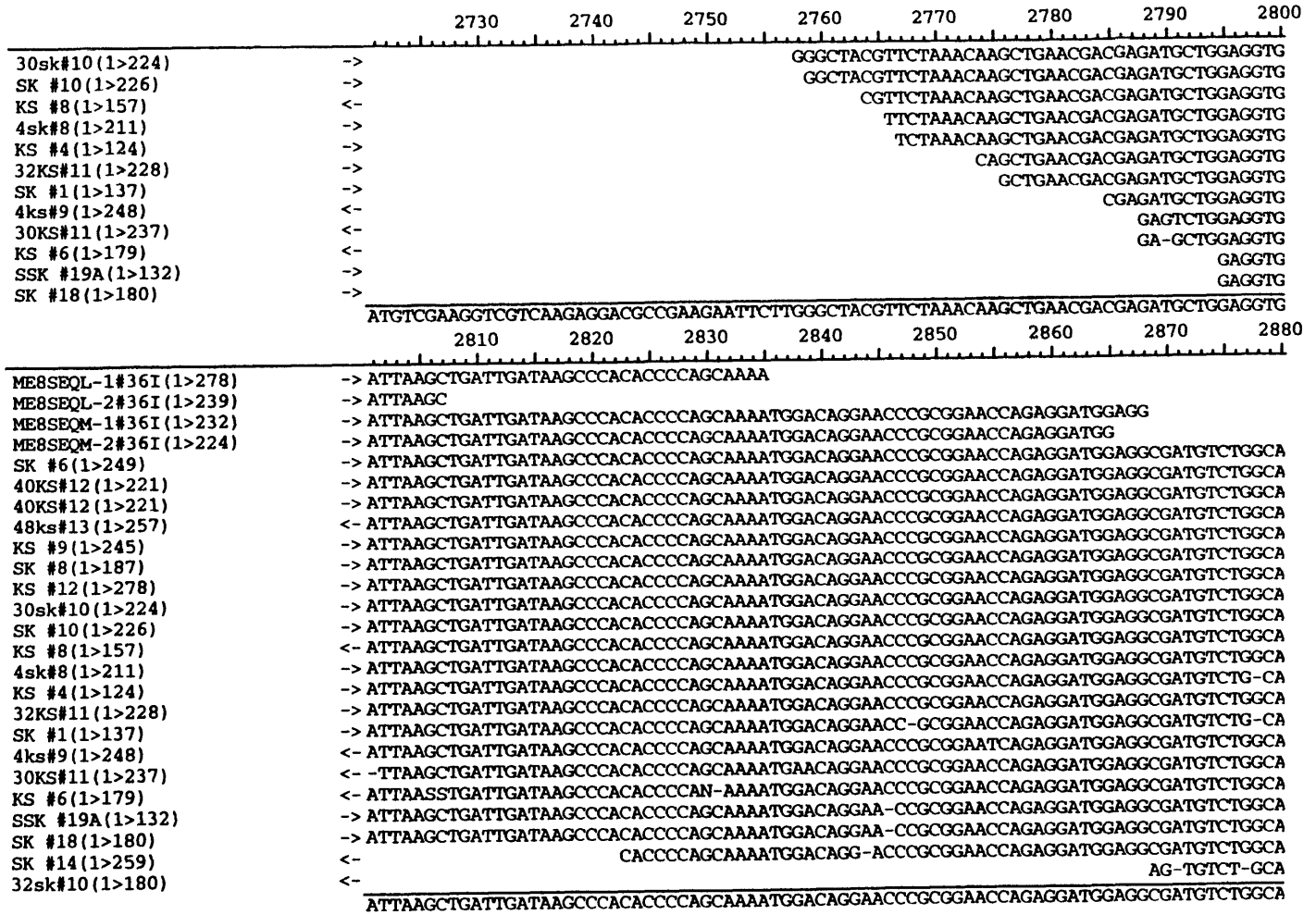

1690 1700 1710 1720 1730 1740 1750 1760
ME-8SEQJ#23(1>270) <- CGGCAGCAGTAGTGCATCACCAGCCGT
B36ks#18(1>275) <- CGGCAGCAGTAGTGCATCACCAGCCGTAGCTGCCGCGGAGCAACAGCAGCCTGCAGCACAGCAAC
ME8SEQJ#35I(1>250) <- CGGCAGCAGTAGTGCATCACCAGCCGTAGCTGCCGCGGAGCAACAGCAGCCTGCAGCACA
B54ks#18(1>191) -> CGGCAGCAGTAGTGCATCACCAGCCGTAGCTGCCGCGGA
B52ks#18(1>222) -> C-GCAGCAGTAGTGCATCACCAGCCGTAGCTGCCGCGGA-SAACAGCAGCCTGCAGCACAGCAACATGCGCAGGCCGCAC
B69ks#21(1>190) <- C--CAGCAGCAGTGCATCACCAGCCGTAGCTGCCGCGGAGCAACAGCAGCCTGCAGCACAGCAACACGGGCAGACCGCAC
B559KS#20 EDIT(1>236) -> CGGCAGCAGTAGTGCATCACCAGCCGTAGCTGCCGCGGA-CAACAGCAGCCTGCAGCACAGCAACATGCGCAGGCCGCAC
B59Sk#20(1>197) <- GGAGCAACAGCAGCCTGCAGCACAGCAACAT-CGCAGGCCGCAC
B42ks#18(1>233) -> AACAGCAGCCTGCAGCACAGCAACATGCGCAGGCCGCAC
CgGCAGCAGTAGTGCATCACCAGCCGTAGCTGCCGCGGAgCAACAGCAGCCTGCAGCACAGCAACATGCGCAGGCCGCAC
1770 1780 1790 1800 1810 1820 1830 1840
B52ks#18(1>222) -> AAACAGGAGCATCAGCTGCTCCAAGCAAGTCCCACAGCAACTATGCTTCTTCTAAAAAGCA
B69ks#21(1>190) <- AAACAGGAGCATCAACAGCTCCAAGCAAGTCCCA
B559KS#20 EDIT(1>236) -> AAACAGGAGCATCAGCAGCTCCAAGCAAGTCCCACAGCAACTATGCTTCTTCTAAAAAGCATCAGAGCTATGAGCCAGTA
B59Sk#20(1>197) <- AAACAGGAGCATCAGCAGCTCCAAGCAAGTCCCACAGCAACTATGCTTCTTCTAAAAAGCATCAGAGCTATGAGCCAGTA
B42ks#18(1>233) -> AAACAGGAGCATCAGCAGCTCCAAGCAAGTCCCACAGCAACTATGCTTCTTCTAAAAAGCATCAGAGCTATGAGCCAGTA
ME-8SEQH#35I(1>312) <- GGAGCATCAGCAGCTCCAAGCAAGTCCCACAGCAACTATGCTTCTTCTAAAAAGCATCAGAGCTATGAGCCAGTA
B50ks#18(1>218) -> AGCTCCAAGCAAGTCCCACAGCAACTATGCTTCTTCTAAAAAGCATCACAGCTATGAGCCAGTA
AAACAGGAGCATCAGCAGCTCCAAGCAAGTCCCACAGCAACTATGCTTCTTCTAAAAAGCATCAGAGCTATGAGCCAGTA
1850 1860 1870 1880 1890 1900 1910 1920
B559KS#20 EDIT(1>236) -> GTGCTGAGCTCCGTGGTGGCCACCAGTTCATCAGCCAATCCGCAGCCGCAGCTAAATCTTGCCCCACCTGCGCC
B59Sk#20(1>197) <- GTGCTGAGCTCCGTGGTGGCCACCAGTTCATCAGCCAATCCGCAGCCGCAGCTAAATCTTGCCCCACCTGCGCC
B42ks#18(1>233) -> GTGCTGAGCTCCGTGGTGGCCACCAGTTCATCAGCCAATCCGCAGCCGCAGCTAAATCTTGCCCCACCTGCTCCACCAGC
ME-8SEQH#35I(1>312) <- GTGCTGAGCTCCGTGGTGGCCACCAGTTCATCAGCCAATCCGCAGCCGCAGCTAAATCTTGCCCCACCTGCGCCACCAGC
B50ks#18(1>218) -> GTGCTGAGCTCCGTGGTGGCCACCAGT-CATCAGCCAATCCGCAGCCGCAGCTAAATCTTGCCCCACCTGCNCCACCAGC
B2sk#14 SPLIT A(1>120) <- ATCAGCCAATCCGCAGCCGCAGCTAAATCTTGCCCCACCTGCGCCACCAGC
B2ks#14 SPLIT A(1>87) -> ACCGC-GCTAAATCTTGCCCCACCTGCGCCACCAGC
ME-8SEQH#23(1>167) <- GCAGCTAAATCTTGCCCCACCTGCGCCACCAGC
GTGCTGAGCTCCGTGGTGGCCACCAGTTCATCAGCCAATCCGCAGCCGCAGCTAAATCTTGCCCCACCTGCGCCACCAGC
1930 1940 1950 1960 1970 1980 1990 2000
B42ks#18(1>233) -> CTCGCAGAACAGCAGCGACAGTT-GTAGCCTGNGT
ME-8SEQH#35I(1>312) <- CTCGCAGAACAGCAGCGACAGTTCGCAGCTTTCGTGGGCCAGTTTGTTTGCCAGCAACAAGCCCGTGGCCAAGGTGGCGC
B50ks#18(1>218) -> CTCGCAGAACAGCAGCGACAGTTCGCAGCTTTCGTGGGCCAGTTTGTTTGCCAGCAACAAGCCCGTGGCCAAGGT
B2sk#14 SPLIT A(1>120) <- CTCGCAGAACAGCAGCGACAGTTCGCAGCTTTCGTGGGCCAGTTTGTTTGCC
B2ks#14 SPLIT A(1>87) -> CTCGCAGAACAGCAGCGACAGTTCGCAGCTTTCGTGGGCCAGTTTGTTTGCC
ME-8SEQH#23(1>167) <- CTCGCAGAACAGCAGCGACAGTTCGCAGCTTTCGTGGGCCAGTTTGTTTGCCAGCAACAASSCCGTGGCCAAGGTGGCGC
ME-8SEQI#35I(1>191) -> GGCGC
CTCGCAGAACAGCAGCGACAGTTCGCAGCTTTCGTGGGCCAGTTTGTTTGCCAGCAACAAGCCCGTGGCCAAGGTGGCGC

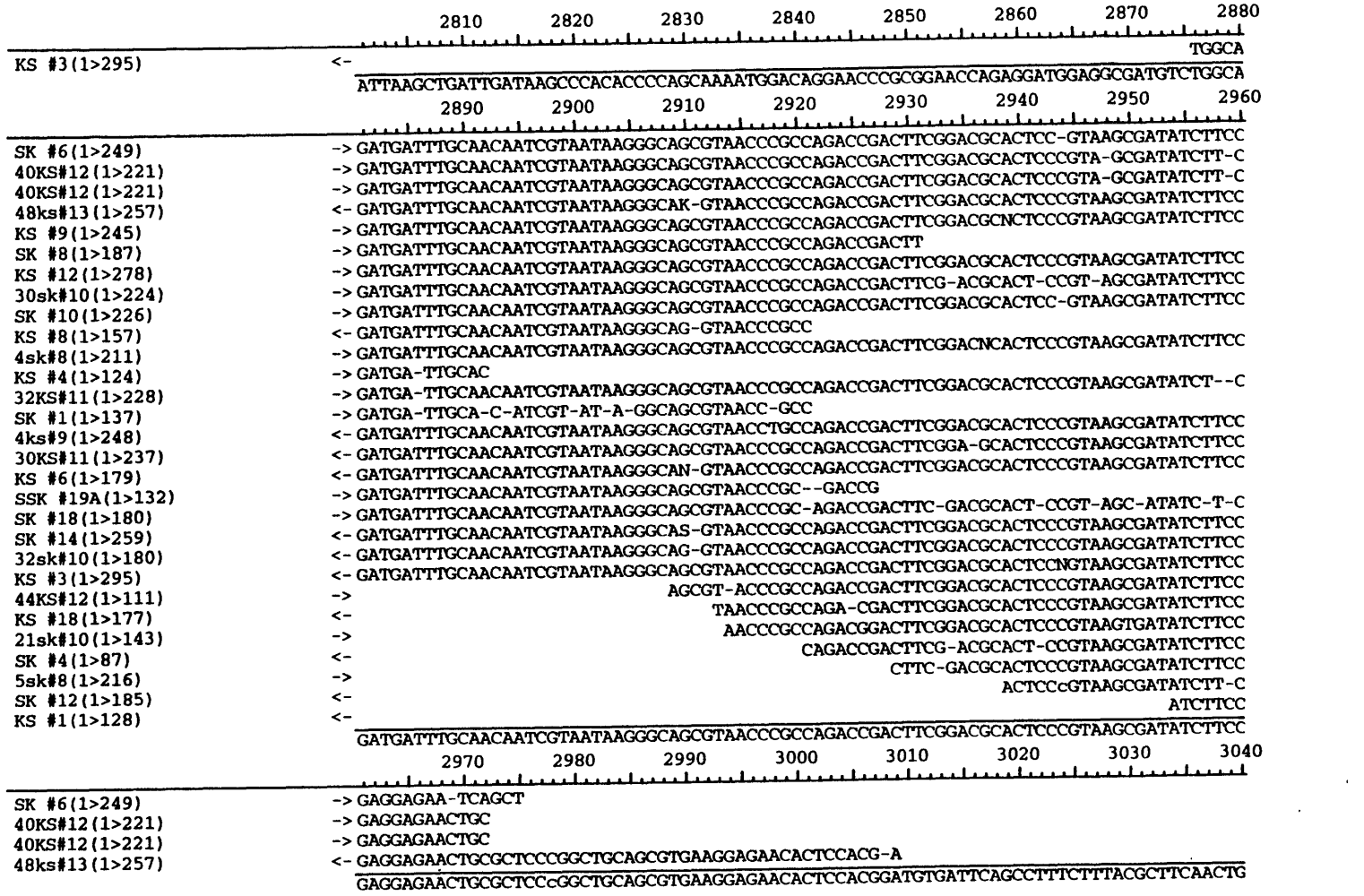
Project: ME Sequence Contig Contig 5

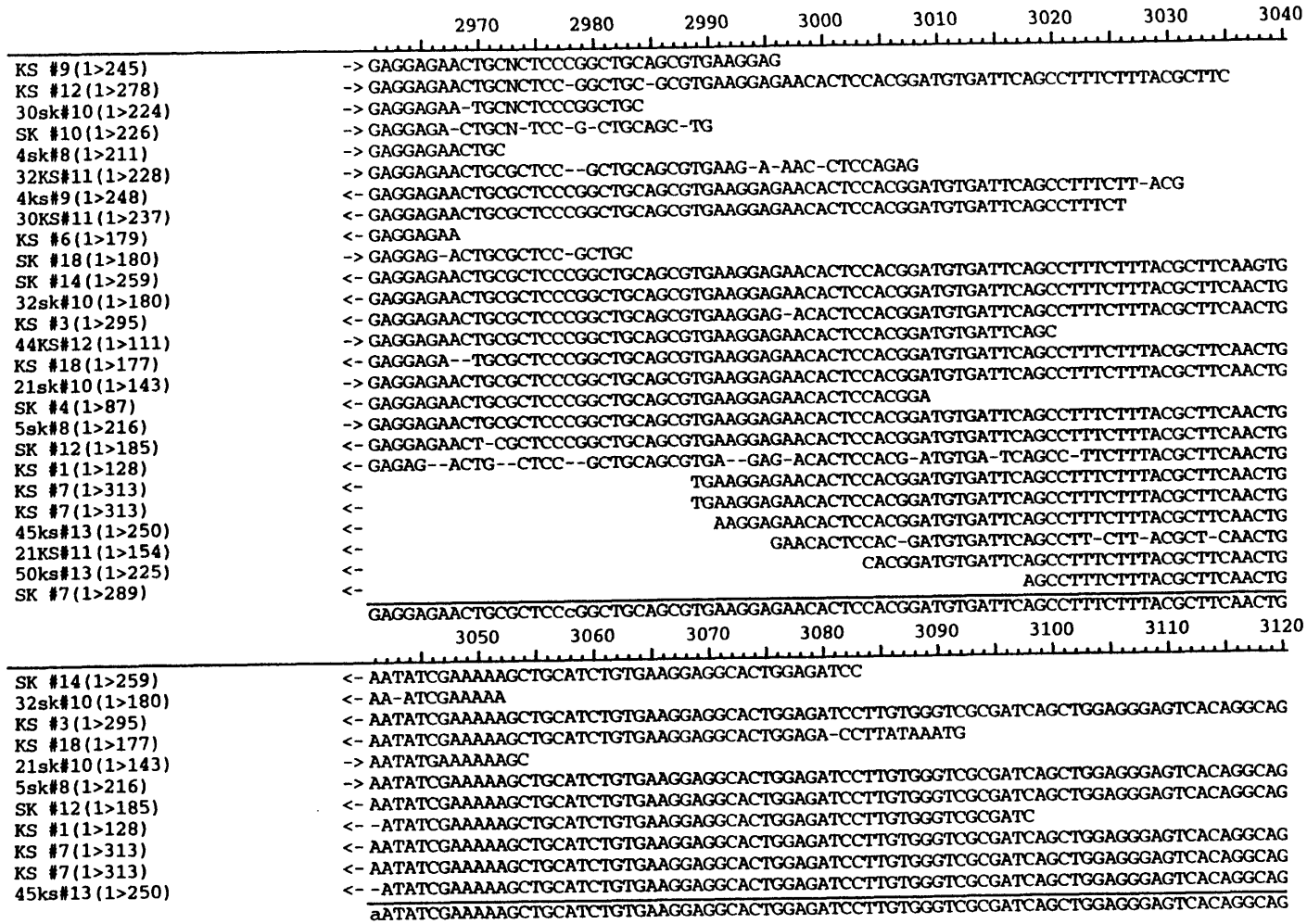


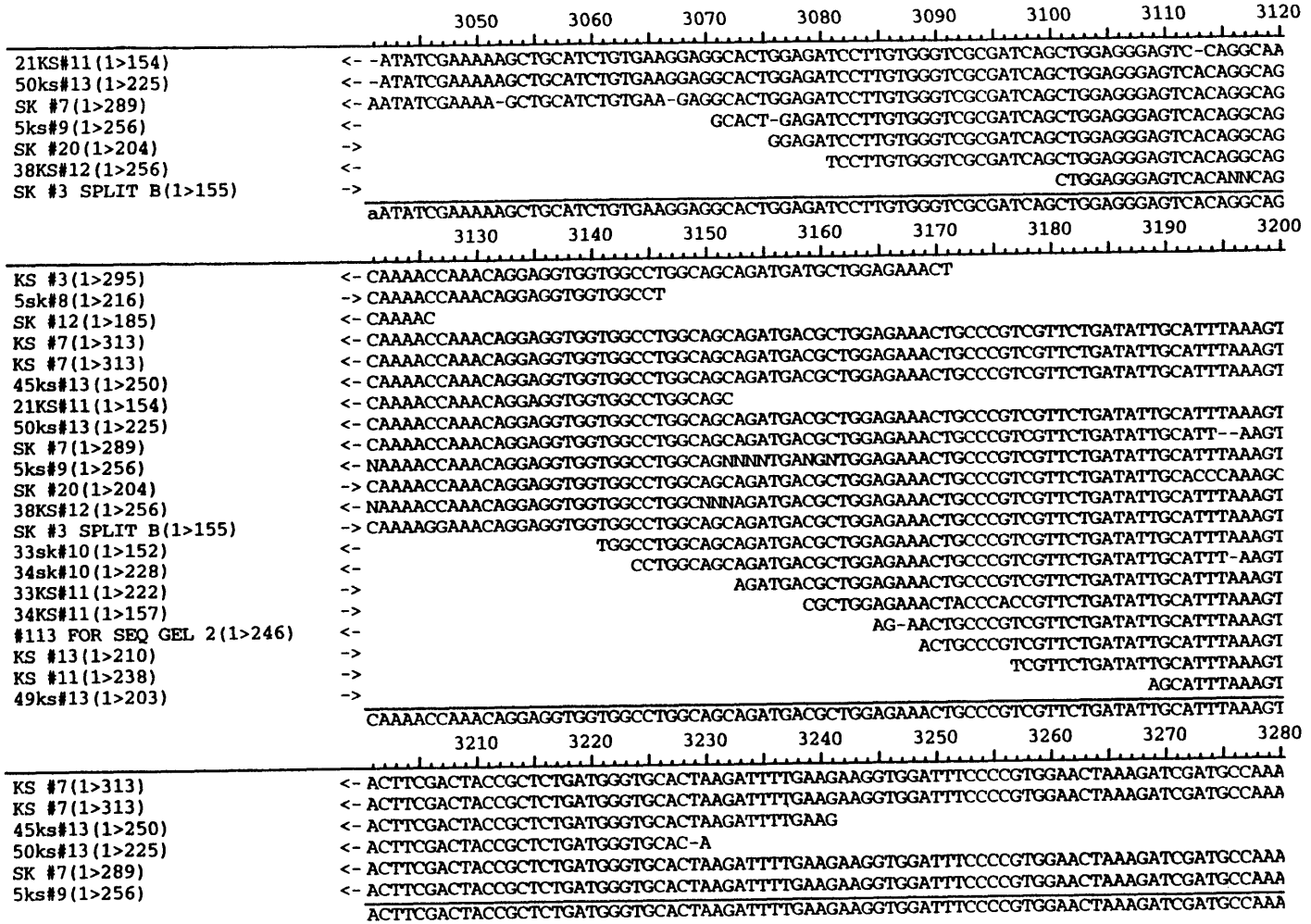

2250 2260 2270 2280 2290 2300 2310 2320
B23ks#16(1>202) ->
B16ks#15 RI END A(1>169) ->
B1ks#14(1>187) ->
TTCCTGCTCCCAGCCCCGCGTCGGCATCCGTCAAGCCCCTCAAGCCGGAACCACCGCGTCCTGTGCAGCAGCAACTGGAT
2330 2340 2350 2360 2370 2380 2390 2400
B45KS#18 SPLIT A(1>165) -> GAGTGGACCAGCAAGTACGCC-AG
ME-8SEQG#34I(1>290) -> GAGTGGACCAGCAAGTACGCCGAGTACCTGACGCGCCACAAGACCAATTTGGCGTCGATCAGCCTGCGTCCCCGAGGACT
ME-8SEQG#23(1>256) -> GAGTGGACCAGCAAGTACGCCGAGTACCTGACGCGCCACAAGACCAATTTGGCGTCGATCAGCCTGCGTCCCCGAGGACT
B23ks#16(1>202) -> GAGTGGACCAGCAAGTACGCCGAGTACCTGACGCGCCACAAGACCAATTTGGCGTCGATCAGCCTGCGTCCCCGAGGACT
B16ks#15 RI END A(1>169) -> GAGTGGACCAGCAAGTACGCCGAGTACCTGACGCGCCACAAGACCAATTTGGCGTCGATCAGCCTGCGTCCCCGAGGACT
B1ks#14(1>187) -> GAGTGGACCAGCAAGTACGCCGAGTACCTGACGCGCCACAAGACCAATTTGGCGTCGATCAGCCTGCGTCCCCGAGGACT
B47ks#18(1>270) <-
GAGTGGACCAGCAAGTACGCCGAGTACCTGACGCGCCACAAGACCAATTTGGCGTCGATCAGCCTGCGTCCCCGAGGACT
2410 2420 2430 2440 2450 2460 2470 2480
ME-8SEQG#34I(1>290) -> GACCAACAGGTCGAACTACTGCTACATCAATTCGATCCTGCAGGCGCTGCTCGGCTGCTCGCCGTTCTGCAACCTGCTGC
ME-8SEQG#23(1>256) -> GACCAACAGGTCGAACTACTGCTACATCAATTCGATCCTGCAGGCGCTGCTCGGCTGCTCG
B23ks#16(1>202) -> GACCAACAGGTCGAACTACTACTACATCAATTCGATCCTGCAGGCNCTGCTC
B16ks#15 RI END A(1>169) -> GACCAACAGGTCGAACTACTGCTACATCAATTCGATCCTGCA
B1ks#14(1>187) -> GACCAACAGGTCGAACTACTGCTACATCAATTCGATCCTGCAGGCNCTGCTCGNCTGCTCGCCGTTCTACA
B47ks#18(1>270) <- GACCAACAGGTCGAACTACTGCTACATCAATTCGATCCTGCAGGCGCTGCTGGCCTGCTCGCNGTTCTACAACCTG-TGC
B12ks #15 split A(1>98) <- AATTCGATCCTGCAGGCGCTGCNC---TGCTCGCNGTTCTACAATCC--TGC
B4ks#14(1>214) -> AATTCGATCCTGCAGGCGCTGCTCGGCTGCTCGCCGTTCTACAACCTGCTGC
B19sk#15(1>143) -> GCAGGCGCTGCTCGGCTGCTCGCCGTTCTACAACCTGCTGC
ME-8C4NTPUSH#35I(1>63) <- CGCGGCTGCTCGCCGTTCTACAACCTGCTGC
ME-8C4NT#35I(1>280) <- TGC
GACCAACAGGTCGAACTACTGCTACATCAATTCGATCCTGCAGGCGCTGCTCGGCTGCTCGCCGTTCTACAACCTGcTGC
2490 2500 2510 2520 2530 2540 2550 2560
ME-8SEQG#34I(1>290) -> GCTCC
B47ks#18(1>270) <- GCTCCATTCCCAAGCAGGCGGCTGTCCTGAGCGAGGTCAAAACGCCCACTGTGAACGCCATGATGTCTTTTATGACCAAC
B12ks #15 split A(1>98) <- GCTCCATTCCCAAGCAGGCGGCTGTCCTGAGCGAGGTCAAAACGCCCACT
B4ks#14(1>214) -> GCTCCATTCCCAAGCAGGCGGCTGTCCTGANCGAGGTCAAAACNCCCACTGTGAACNCCATGATGTCTTTTATGACCAAC
B19sk#15(1>143) -> GCTCCATTCCCAAGCAGGCGGCTGTCCTGAGCGAGGTCAAAACGCCCACTGTGAACGCCATGATGTCTTTTATGACCAAC
ME-8C4NTPUSH#35I(1>63) <- GCTCCATTCCCAAGCAGGCGGCTGTCCT
ME-8C4NT#35I(1>280) <- GCTCCATTCCCAAGCAGGCGGCTGTCCTGAGCGAGGTCAAAACGCCCACTGTGAACGCCATGATGTCTTTTATGACCAAC
C4NT A #17(1>272) <- CTCCATTCCCAAGCAGGCGGCTGTCCTGAGCGAGGTCAAAACGCCCACTGTGAACGCCATGATGTCTTTTATGACCAAC
C4NT B #17(1>270) <- TTCCCAAGCAGGCGGCTGTCCTGAGCGAGGTCAAAACGCCCAC-GTGAACGCCATGATGTCTTTTATGACCAAC
ME8SEQL-1#36I(1>278) -> AAC
GCTCCATTCCCAAGCAGGCGGCTGTCCTGAGCGAGGTCAAAACGCCCACTGTGAACGCCATGATGTCTTTTATGACCAAC

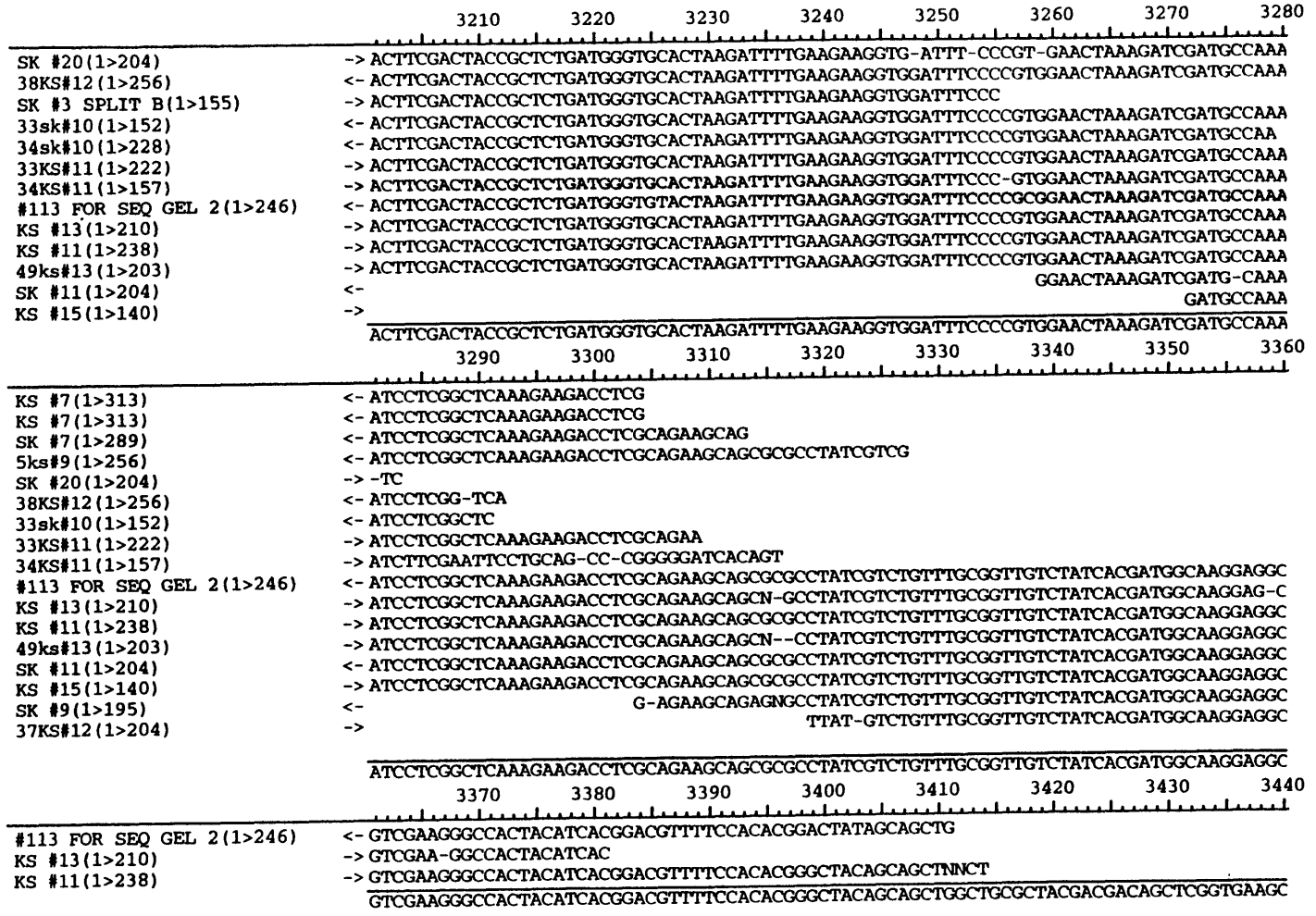

2570 2580 2590 2600 2610 2620 2630 2640
B47ks#18(1>270) <- TTTTCATCGCTGCCCAGCGGTCTTCGCCTGCGTCTAAACAACCTTAACAAGGGATCGCAAAGCAAGGGCAAAG-CGATTT
B4ks#14(1>214) -> TTTTCATCGCTGCCCAGCGGTCTTCN-CTGCGTCTAAACAACCTTAACAAGGGATCGCAAAGCAAGGGCAAAGACGATTT
B19sk#15(1>143) -> TTTTCATCGCTGCCCAGCGGTC
ME-8C4NT#35I(1>280) <- TTTTCATCGCTGCCCAGCGGTCTTCGCCTGCGTCTAAACAACCTTAACAAGGGATCGCAAAGCAAGGGCAAAGACGATTT
C4NT A #17(1>272) <- TTTTCATCGCTGCCCAGCGGTCTTCGCCTGCGTCTAAACAACCTTAACAAGGGATCGCAAAGCAAGGGCAAAGACGATTT
C4NT B #17(1>270) <- TTTTCATCGCTGCCCAGCGGTCTTCGCCTGCGTCTAAACAACCTTAACAAGGGATCGCAAAGCAAGGGCAAAGACGATTT
ME8SEQL-1#36I(1>278) -> TTTTCATCGCTGCCCAGCGGTCTTCGCCTGCGTCTAAACAACCTTAACAAGGGATCGCAAAGCAAGGGCAAAGACGATTT
ME8SEQL-2#36I(1>239) -> GCTGCCCAGCGGTCTTCGCCTGCGTCTAAACAACCTTAACAAGGGATCGCAAAGCAAGGGCAAAGACGATTT
ME8SEQM-1#36I(1>232) -> ATTT
TTTTCATCGCTGCCCAGCGGTCTTCGCCTGCGTCTAAACAACCTTAACAAGGGATCGCAAAGCAAGGGCAAAGACGATTT
2650 2660 2670 2680 2690 2700 2710 2720
B47ks#18(1>270) <- TATGGGCTCAGATT
B4ks#14(1>214) -> TT
ME-8C4NT#35I(1>280) <- TGTGGGCTCAGATTTGCAATGCGACATGGCCTTTGAACCCACAGAGATCTACAAGCTGTGGAACGACAGTCGCGAGGAGC
C4NT A #17(1>272) <- TGTGGGCTCAGATTTGCAATGCGACATGGCCTTTGAACCCACAGAGATCTACAAGCTGTGGAACGACAGTCGCGAGGAGC
C4NT B #17(1>270) <- TGTGGGCTCAGATTTGCAATGCGACATGGCCTTTGAACCCACAGAGATCTACAAGCTGTGGAACGACAGTCGCGAGGAGC
ME8SEQL-1#36I(1>278) -> TGTGGGCTCAGATTTGCAATGCGACATGGCCTTTGAACCCACAGAGATCTACAAGCTGTGGAACGACAGTCGCGAGGAGC
ME8SEQL-2#36I(1>239) -> TGTGGGCTCAGATTTGCAATGCGACATGGCCTTTGAACCCACAGAGATCTACAAGCTGTGGAACGACAGTCGCGAGGAGC
ME8SEQM-1#36I(1>232) -> TGTGGGCTCAGATTTGCAATGCGACATGGCCTTTGAACCCACAGAGATCTACAAGCTGTGGAACGACAGTCGCGAGGAGC
ME8SEQM-2#36I(1>224) -> GTGGGCTCAGATTTGCAATGCGACATGGCCTTTGAACCCACAGAGATCTACAAGCTGTGGAACGACAGTCGCGAGGAGC
B32sk #17 real split B(1>46) -> G-CGCGAGGAGC
TGTGGGCTCAGATTTGCAATGCGACATGGCCTTTGAACCCACAGAGATCTACAAGCTGTGGAACGACAGTCGCGAGGAGC
2730 2740 2750 2760 2770 2780 2790 2800
ME-8C4NT#35I(1>280) <- ATGTCGAAGGTCGTCAAGAGGACGCCGAAGAATTCTT
C4NT A #17(1>272) <- ATGTCGAAGGTCGTCAAGAGGACGCCGAAGA
C4NT B #17(1>270) <- ATGTCGAAGGTCGTCAAGAGGACGCCGAAGA-T-CTG
ME8SEQL-1#36I(1>278) -> ATGTCGAAGGTCGTCAAGAGGACGCCGAAGAATTCTTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
ME8SEQL-2#36I(1>239) -> ATGTCGAAGGTCGTCAAGAGGACGCCGAAGAATTCTTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
ME8SEQM-1#36I(1>232) -> ATGTCGAAGGTCGTCAAGAGGACGCCGAAGAATTCTTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
ME8SEQM-2#36I(1>224) -> ATGTCGAAGGTCGTCAAGAGGACGCCGAAGAATTCTTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
B32sk #17 real split B(1>46) -> ATGTCGAAGGTCGTCAAGAGGACGCCGAAGAATTC
SK #6(1>249) -> GGACGCCGAAGAATTCTTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
40KS#12(1>221) -> TGAAGAATTCTTGGGCTAC-TTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
40KS#12(1>221) -> -------TTCTTGGGCTAC-TTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
48ks#13(1>257) <- -GAATTCTTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
KS #9(1>245) -> AATTCTTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
SK #8(1>187) -> TTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
KS #12(1>278) -> TTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
ATGTCGAAGGTCGTCAAGAGGACGCCGAAGAATTCTTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG


2730 2740 2750 2760 2770 2780 2790 2800
30sk#10(1>224) -> GGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
SK #10(1>226) -> GGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
KS #8(1>157) <- CGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
4sk#8(1>211) -> TTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
KS #4(1>124) -> TCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
32KS#11(1>228) -> CAGCTGAACGACGAGATGCTGGAGGTG
SK #1(1>137) -> GCTGAACGACGAGATGCTGGAGGTG
4ks#9(1>248) <- CGAGATGCTGGAGGTG
30KS#11(1>237) <- GAGTCTGGAGGTG
KS #6(1>179) <- GA-GCTGGAGGTG
SSK #19A(1>132) -> GAGGTG
SK #18(1>180) -> GAGGTG
ATGTCGAAGGTCGTCAAGAGGACGCCGAAGAATTCTTGGGCTACGTTCTAAACAAGCTGAACGACGAGATGCTGGAGGTG
2810 2820 2830 2840 2850 2860 2870 2880
ME8SEQL-1#36I(1>278) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAA
ME8SEQL-2#36I(1>239) -> ATTAAGC
ME8SEQM-1#36I(1>232) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGG
ME8SEQM-2#36I(1>224) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGG
SK #6(1>249) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
40KS#12(1>221) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
40KS#12(1>221) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
48ks#13(1>257) <- ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
KS #9(1>245) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
SK #8(1>187) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
KS #12(1>278) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
30sk#10(1>224) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
SK #10(1>226) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
KS #8(1>157) <- ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
4sk#8(1>211) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTG-CA
KS #4(1>124) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
32KS#11(1>228) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACC-GCGGAACCAGAGGATGGAGGCGATGTCTG-CA
SK #1(1>137) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAATCAGAGGATGGAGGCGATGTCTGGCA
4ks#9(1>248) <- ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGAACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
30KS#11(1>237) <- -TTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAA-CCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
KS #6(1>179) <- ATTAASSTGATTGATAAGCCCACACCCCAN-AAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
SSK #19A(1>132) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAA-CCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
SK #18(1>180) -> ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAA-CCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
SK #14(1>259) <- CACCCCAGCAAAATGGACAGG-ACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
32sk#10(1>180) <- AG-TGTCT-GCA
ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA

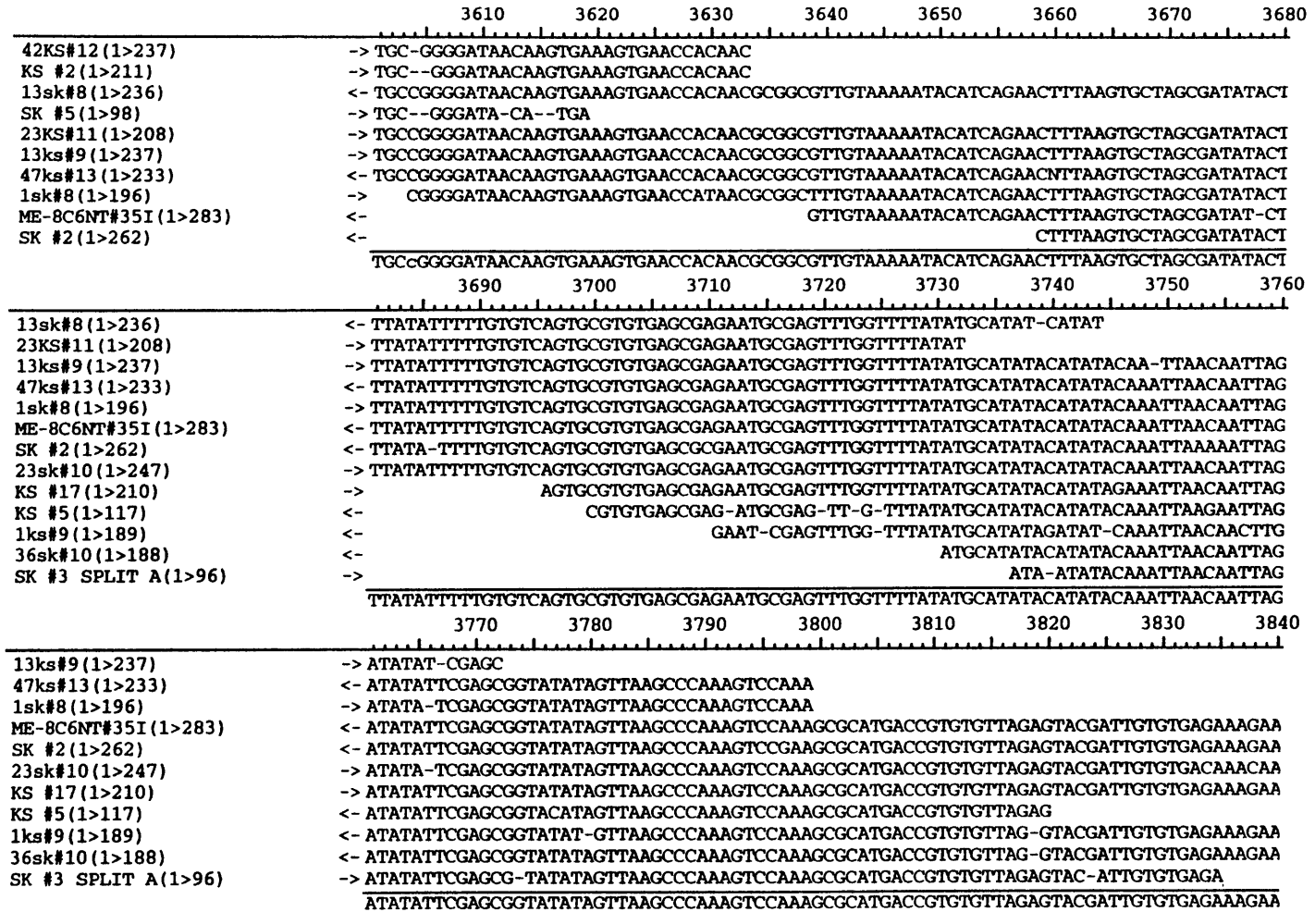

2810 2820 2830 2840 2850 2860 2870 2880
KS #3(1>295) <-
TGGCA
ATTAAGCTGATTGATAAGCCCACACCCCAGCAAAATGGACAGGAACCCGCGGAACCAGAGGATGGAGGCGATGTCTGGCA
2890 2900 2910 2920 2930 2940 2950 2960
SK #6(1>249) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACGCACTCC-GTAAGCGATATCTTCC
40KS#12(1>221) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACGCACTCCCGTA-GCGATATCTT-C
40KS#12(1>221) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACGCACTCCCGTA-GCGATATCTT-C
48ks#13(1>257) <- GATGATTTGCAACAATCGTAATAAGGGCAK-GTAACCCGCCAGACCGACTTCGGACGCACTCCCGTAAGCGATATCTTCC
KS #9(1>245) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACGCNCTCCCGTAAGCGATATCTTCC
SK #8(1>187) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTT
KS #12(1>278) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACGCACTCCCGTAAGCGATATCTTCC
30sk#10(1>224) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCG-ACGCACT-CCGT-AGCGATATCTTCC
SK #10(1>226) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACGCACTCC-GTAAGCGATATCTTCC
KS #8(1>157) <- GATGATTTGCAACAATCGTAATAAGGGCAG-GTAACCCGCC
4sk#8(1>211) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACNCACTCCCGTAAGCGATATCTTCC
KS #4(1>124) -> GATGA-TTGCAC
32KS#11(1>228) -> GATGA-TTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACGCACTCCCGTAAGCGATATCT--C
SK #1(1>137) -> GATGA-TTGCA-C-ATCGT-AT-A-GGCAGCGTAACC-GCC
4ks#9(1>248) <- GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCTGCCAGACCGACTTCGGACGCACTCCCGTAAGCGATATCTTCC
30KS#11(1>237) <- GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGA-GCACTCCCGTAAGCGATATCTTCC
KS #6(1>179) <- GATGATTTGCAACAATCGTAATAAGGGCAN-GTAACCCGCCAGACCGACTTCGGACGCACTCCCGTAAGCGATATCTTCC
SSK #19A(1>132) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGC--GACCG
SK #18(1>180) -> GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGC-AGACCGACTTC-GACGCACT-CCGT-AGC-ATATC-T-C
SK #14(1>259) <- GATGATTTGCAACAATCGTAATAAGGGCAS-GTAACCCGCCAGACCGACTTCGGACGCACTCCCGTAAGCGATATCTTCC
32sk#10(1>180) <- GATGATTTGCAACAATCGTAATAAGGGCAG-GTAACCCGCCAGACCGACTTCGGACGCACTCCCGTAAGCGATATCTTCC
KS #3(1>295) <- GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACGCACTCCNGTAAGCGATATCTTCC
44KS#12(1>111) -> AGCGT-ACCCGCCAGACCGACTTCGGACGCACTCCCGTAAGCGATATCTTCC
KS #18(1>177) <- TAACCCGCCAGA-CGACTTCGGACGCACTCCCGTAAGCGATATCTTCC
21sk#10(1>143) -> AACCCGCCAGACGGACTTCGGACGCACTCCCGTAAGTGATATCTTCC
SK #4(1>87) <- CAGACCGACTTCG-ACGCACT-CCGTAAGCGATATCTTCC
5sk#8(1>216) -> CTTC-GACGCACTCCCGTAAGCGATATCTTCC
SK #12(1>185) <- ACTCCcGTAAGCGATATCTT-C
KS #1(1>128) <- ATCTTCC
GATGATTTGCAACAATCGTAATAAGGGCAGCGTAACCCGCCAGACCGACTTCGGACGCACTCCCGTAAGCGATATCTTCC
2970 2980 2990 3000 3010 3020 3030 3040
SK #6(1>249) -> GAGGAGAA-TCAGCT
40KS#12(1>221) -> GAGGAGAACTGC
40KS#12(1>221) -> GAGGAGAACTGC
48ks#13(1>257) <- GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACG-A
GAGGAGAACTGCGCTCCcGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG


2970 2980 2990 3000 3010 3020 3030 3040
KS #9(1>245) -> GAGGAGAACTGCNCTCCCGGCTGCAGCGTGAAGGAG
KS #12(1>278) -> GAGGAGAACTGCNCTCC-GGCTGC-GCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTC
30sk#10(1>224) -> GAGGAGAA-TGCNCTCCCGGCTGC
SK #10(1>226) -> GAGGAGA-CTGCN-TCC-G-CTGCAGC-TG
4sk#8(1>211) -> GAGGAGAACTGC
32KS#11(1>228) -> GAGGAGAACTGCGCTCC--GCTGCAGCGTGAAG-A-AAC-CTCCAGAG
4ks#9(1>248) <- GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTT-ACG
30KS#11(1>237) <- GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCT
KS #6(1>179) <- GAGGAGAA
SK #18(1>180) -> GAGGAG-ACTGCGCTCC-GCTGC
SK #14(1>259) <- GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAAGTG
32sk#10(1>180) <- GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
KS #3(1>295) <- GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAG-ACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
44KS#12(1>111) -> GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGC
KS #18(1>177) <- GAGGAGA--TGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
21sk#10(1>143) -> GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
SK #4(1>87) <- GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGA
5sk#8(1>216) -> GAGGAGAACTGCGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
SK #12(1>185) <- GAGGAGAACT-CGCTCCCGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
KS #1(1>128) <- GAGAG--ACTG--CTCC--GCTGCAGCGTGA--GAG-ACACTCCACG-ATGTGA-TCAGCC-TTCTTTACGCTTCAACTG
KS #7(1>313) <- TGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
KS #7(1>313) <- TGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
45ks#13(1>250) <- AAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
21KS#11(1>154) <- GAACACTCCAC-GATGTGATTCAGCCTT-CTT-ACGCT-CAACTG
50ks#13(1>225) <- CACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
SK #7(1>289) <- AGCCTTTCTTTACGCTTCAACTG
GAGGAGAACTGCGCTCCcGGCTGCAGCGTGAAGGAGAACACTCCACGGATGTGATTCAGCCTTTCTTTACGCTTCAACTG
3050 3060 3070 3080 3090 3100 3110 3120
SK #14(1>259) <- AATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCC
32sk#10(1>180) <- AA-ATCGAAAAA
KS #3(1>295) <- AATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
KS #18(1>177) <- AATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGA-CCTTATAAATG
21sk#10(1>143) -> AATATGAAAAAAGC
5sk#8(1>216) -> AATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
SK #12(1>185) <- AATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
KS #1(1>128) <- -ATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATC
KS #7(1>313) <- AATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
KS #7(1>313) <- AATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
45ks#13(1>250) <- -ATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
aATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG

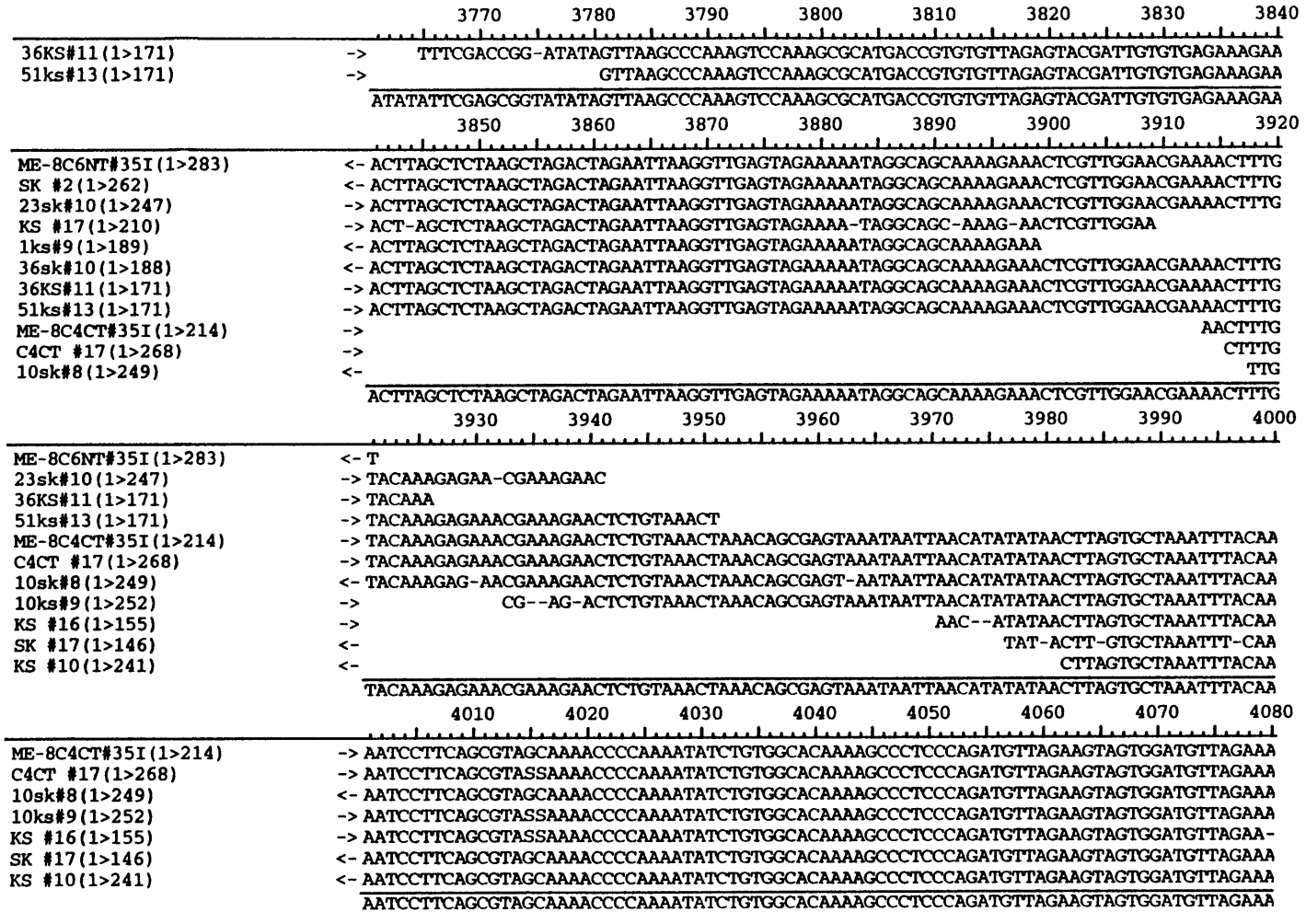

3050 3060 3070 3080 3090 3100 3110 3120
21KS#11(1>154) <- -ATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTC-CAGGCAA
50ks#13(1>225) <- -ATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
SK #7(1>289) <- AATATCGAAAA-GCTGCATCTGTGAA-GAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
5ks#9(1>256) <- GCACT-GAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
SK #20(1>204) -> GGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
38KS#12(1>256) <- TCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
SK #3 SPLIT B(1>155) -> CTGGAGGGAGTCACANNCAG
aATATCGAAAAAGCTGCATCTGTGAAGGAGGCACTGGAGATCCTTGTGGGTCGCGATCAGCTGGAGGGAGTCACAGGCAG
3130 3140 3150 3160 3170 3180 3190 3200
KS #3(1>295) <- CAAAACCAAACAGGAGGTGGTGGCCTGGCAGCAGATGATGCTGGAGAAACT
5sk#8(1>216) -> CAAAACCAAACAGGAGGTGGTGGCCT
SK #12(1>185) <- CAAAAC
KS #7(1>313) <- CAAAACCAAACAGGAGGTGGTGGCCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
KS #7(1>313) <- CAAAACCAAACAGGAGGTGGTGGCCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
45ks#13(1>250) <- CAAAACCAAACAGGAGGTGGTGGCCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
21KS#11(1>154) <- CAAAACCAAACAGGAGGTGGTGGCCTGGCAGC
50ks#13(1>225) <- CAAAACCAAACAGGAGGTGGTGGCCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
SK #7(1>289) <- CAAAACCAAACAGGAGGTGGTGGCCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATT--AAGT
5ks#9(1>256) <- NAAAACCAAACAGGAGGTGGTGGCCTGGCAGNNNNTGANGNTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
SK #20(1>204) -> CAAAACCAAACAGGAGGTGGTGGCCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCACCCAAAGC
38KS#12(1>256) <- NAAAACCAAACAGGAGGTGGTGGCCTGGCNNNAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
SK #3 SPLIT B(1>155) -> CAAAAGGAAACAGGAGGTGGTGGCCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
33sk#10(1>152) <- TGGCCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
34sk#10(1>228) <- CCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTT-AAGT
33KS#11(1>222) -> AGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
34KS#11(1>157) -> CGCTGGAGAAACTACCCACCGTTCTGATATTGCATTTAAAGT
#113 FOR SEQ GEL 2(1>246) <- AG-AACTGCCCGTCGTTCTGATATTGCATTTAAAGT
KS #13(1>210) -> ACTGCCCGTCGTTCTGATATTGCATTTAAAGT
KS #11(1>238) -> TCGTTCTGATATTGCATTTAAAGT
49ks#13(1>203) -> AGCATTTAAAGT
CAAAACCAAACAGGAGGTGGTGGCCTGGCAGCAGATGACGCTGGAGAAACTGCCCGTCGTTCTGATATTGCATTTAAAGT
3210 3220 3230 3240 3250 3260 3270 3280
KS #7(1>313) <- ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
KS #7(1>313) <- ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
45ks#13(1>250) <- ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAG
50ks#13(1>225) <- ACTTCGACTACCGCTCTGATGGGTGCAC-A
SK #7(1>289) <- ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
5ks#9(1>256) <- ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA

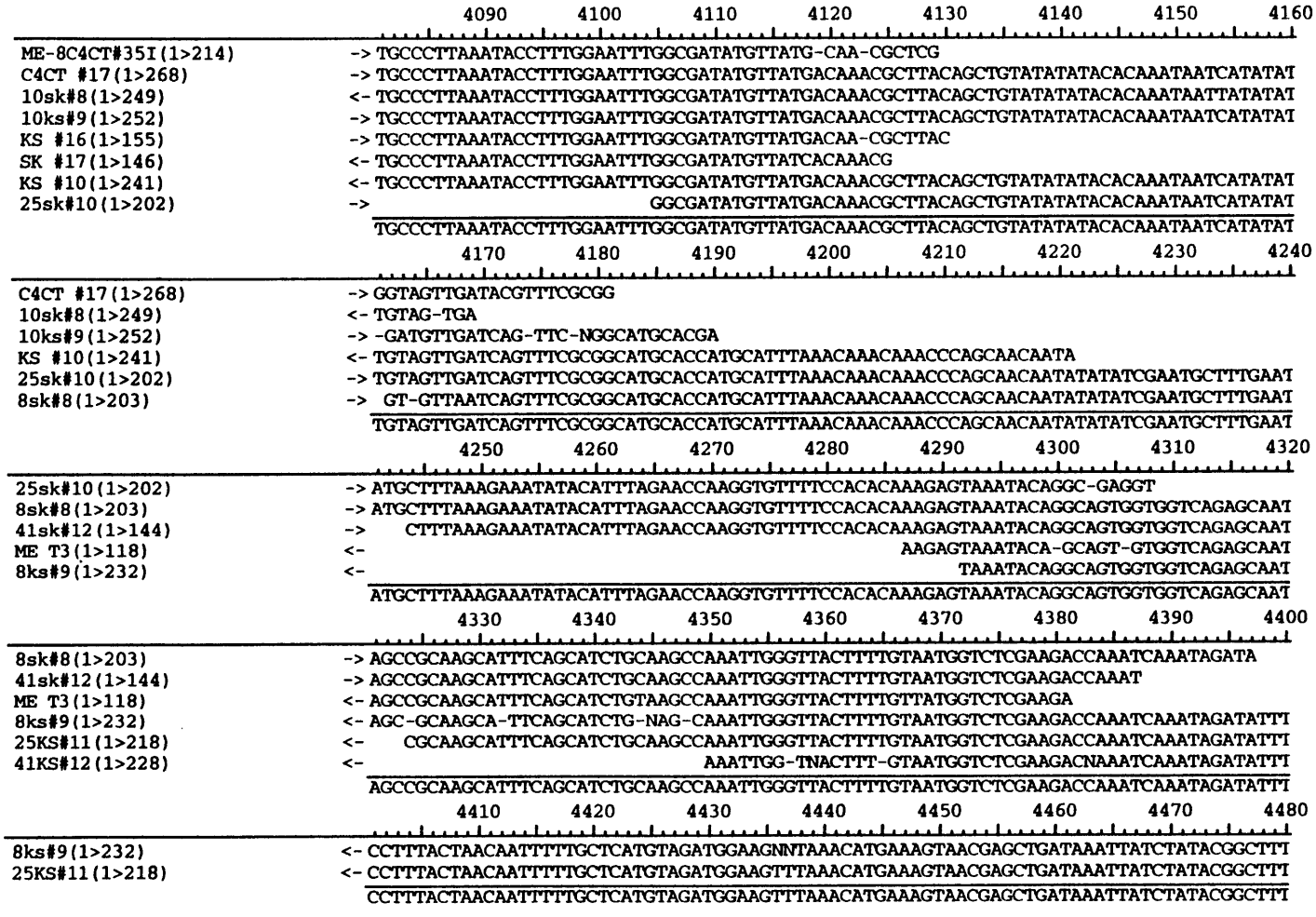

3210 3220 3230 3240 3250 3260 3270 3280
SK #20(1>204)
38KS#12(1>256)
SK #3 SPLIT B(1>155)
33sk#10(1>152)
34sk#10(1>228)
33KS#11(1>222)
34KS#11(1>157)
#113 FOR SEQ GEL 2(1>246)
KS #13(1>210)
KS #11(1>238)
49ks#13(1>203)
SK #11(1>204)
KS #15(1>140)
-> ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTG-ATTT-CCCGT-GAACTAAAGATCGATGCCAAA
<- ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
-> ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCC
<- ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
<- ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAA
-> ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
-> ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCC-GTGGAACTAAAGATCGATGCCAAA
<- ACTTCGACTACCGCTCTGATGGGTGTACTAAGATTTTGAAGAAGGTGGATTTCCCCGCGGAACTAAAGATCGATGCCAAA
-> ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
-> ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
-> ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
GGAACTAAAGATCGATG-CAAA
<-
GATGCCAAA
->
ACTTCGACTACCGCTCTGATGGGTGCACTAAGATTTTGAAGAAGGTGGATTTCCCCGTGGAACTAAAGATCGATGCCAAA
3290 3300 3310 3320 3330 3340 3350 3360
KS #7(1>313)
KS #7(1>313)
SK #7(1>289)
5ks#9(1>256)
SK #20(1>204)
38KS#12(1>256)
33sk#10(1>152)
33KS#11(1>222)
34KS#11(1>157)
#113 FOR SEQ GEL 2(1>246)
KS #13(1>210)
KS #11(1>238)
49ks#13(1>203)
SK #11(1>204)
KS #15(1>140)
SK #9(1>195)
37KS#12(1>204)
<- ATCCTCGGCTCAAAGAAGACCTCG
<- ATCCTCGGCTCAAAGAAGACCTCG
<- ATCCTCGGCTCAAAGAAGACCTCGCAGAAGCAG
<- ATCCTCGGCTCAAAGAAGACCTCGCAGAAGCAGCGCGCCTATCGTCG
-> -TC
<- ATCCTCGG-TCA
<- ATCCTCGGCTC
-> ATCCTCGGCTCAAAGAAGACCTCGCAGAA
-> ATCTTCGAATTCCTGCAG-CC-CGGGGGATCACAGT
<- ATCCTCGGCTCAAAGAAGACCTCGCAGAAGCAGCGCGCCTATCGTCTGTTTGCGGTTGTCTATCACGATGGCAAGGAGGC
-> ATCCTCGGCTCAAAGAAGACCTCGCAGAAGCAGCN-GCCTATCGTCTGTTTGCGGTTGTCTATCACGATGGCAAGGAG-C
-> ATCCTCGGCTCAAAGAAGACCTCGCAGAAGCAGCGCGCCTATCGTCTGTTTGCGGTTGTCTATCACGATGGCAAGGAGGC
-> ATCCTCGGCTCAAAGAAGACCTCGCAGAAGCAGCN--CCTATCGTCTGTTTGCGGTTGTCTATCACGATGGCAAGGAGGC
<- ATCCTCGGCTCAAAGAAGACCTCGCAGAAGCAGCGCGCCTATCGTCTGTTTGCGGTTGTCTATCACGATGGCAAGGAGGC
-> ATCCTCGGCTCAAAGAAGACCTCGCAGAAGCAGCGCGCCTATCGTCTGTTTGCGGTTGTCTATCACGATGGCAAGGAGGC
<- G-AGAAGCAGAGNGCCTATCGTCTGTTTGCGGTTGTCTATCACGATGGCAAGGAGGC
-> TTAT-GTCTGTTTGCGGTTGTCTATCACGATGGCAAGGAGGC
ATCCTCGGCTCAAAGAAGACCTCGCAGAAGCAGCGCGCCTATCGTCTGTTTGCGGTTGTCTATCACGATGGCAAGGAGGC
3370 3380 3390 3400 3410 3420 3430 3440
#113 FOR SEQ GEL 2(1>246)
KS #13(1>210)
KS #11(1>238)
<- GTCGAAGGGCCACTACATCACGGACGTTTTCCACACGGACTATAGCAGCTG
-> GTCGAA-GGCCACTACATCAC
-> GTCGAAGGGCCACTACATCACGGACGTTTTCCACACGGGCTACAGCAGCTNNCT
GTCGAAGGGCCACTACATCACGGACGTTTTCCACACGGGCTACAGCAGCTGGCTGCGCTACGACGACAGCTCGGTGAAGC

```
                                3370       3380       3390       3400       3410       3420       3430       3440
                                ....|....  ....|....  ....|....  ....|....  ....|....  ....|....  ....|....  ....|
49ks#13(1>203)                  -> GTCGAAGGGCCACTACATCACGGACN-TTTCCAC
SK #11(1>204)                   <- GTCGAAGGGCCACTACATCACGGACGTTTTCCACACGGGCTACAGCAGCTGgcTGCGCTACGACGACAGCTCGGTGAAGC
KS #15(1>140)                   -> GTCGAAGGGCCACTACATCACGGACNTTTTCCACAC-GGCTACAGCAGCTGGCT
SK #9(1>195)                    <- GTCGAAGGGCCACTACATCACGGACGTTTTCCACACGGGCTACAGCAGCTGSSTGCGCTACGACGACAGCTCGGTGAAGC
37KS#12(1>204)                  -> GTCGAAGGGCCACTACATCACGGACNTTTTCCACACGGGCTACAGCAGCTGGCTSSGCTACGACGACAGCTCGGTGAAGC
2sk#8(1>182)                    ->    GA-GGG-CACTAC-TCACGGACGTTTTCCACACGGGCTACAGCAGCTGGCTSSGCTACGACGACAGCTCGGTGAAGC
2ks#9(1>159)                    <-          CCACTACATCACGGACGTTTTCCACACGGGCTACAGCAGCTGGCTGCGCTACGACGACAGCTCGGTGAAGC
42KS#12(1>237)                  ->                                ACGGGCTACAGCAGCTGGCTGCGCTACGACGACAGCTCGGTGAAGC
SK #13(1>181)                   <-                                                          CTACGACGACAGCTCGGTGAAGC
KS #2(1>211)                    ->                                                          TACGACGACAGCTCGGTGAAGC
52ks#13(1>157)                  ->                                                                GGTGAAGC
                                GTCGAAGGGCCACTACATCACGGACGTTTTCCACACGGGCTACAGCAGCTGGCTGCGCTACGACGACAGCTCGGTGAAGC
                                3450       3460       3470       3480       3490       3500       3510       3520
                                ....|....  ....|....  ....|....  ....|....  ....|....  ....|....  ....|....  ....|
SK #11(1>204)                   <- CGGTCAGCGAGAAGCATGTCCTTC
SK #9(1>195)                    <- CGGTCAGCGAGAAGCGTGTCCTTCAGCC-CACGCGCCTCGTGTGCCTTACCTGCCGTAC
37KS#12(1>204)                  -> CGGTCAGCGAGAAGCATGTC-TTCAGCCGCACACGCCTCG
2sk#8(1>182)                    -> CGGTCAGCGAGAAGTATGTCCTTCAGCCGCACACGCCTCGTGTGCCTTACCTGCTGTACTACCGCCGTTCAGACACCCTG
2ks#9(1>159)                    <- CGGTCAGCGAGAAGCATGTCCTTCAGCC-SACACGCCTCGTGTGCCTTACCTGCTGTACTACCGCCGTTCAGACACCCTG
42KS#12(1>237)                  -> CGGTCAGCGAGAAGCATGTCCTTCAGCCGCACACGCCTCGTGTGCCTTACCTGCTGTACTACCGCCGTTCAGACACCCTG
SK #13(1>181)                   <- CGGTCAGCGAGAAGTATGTCCTTCAGCC-SACACGCCTCGTGTGCCTTACCTGCTGTACTACCGCCGTTCAGACACCCTG
KS #2(1>211)                    -> CGGTCAGCGAGAAGCATGTCCTTCAGCCGCACACGCCTCGTGTSSCTTACCTGCTGTACTACCGCCGTTCAGACACCCTG
52ks#13(1>157)                  -> CGGTCAGCGAGAAGCATGTCCTTCAGCCGCACACGCCTCGCGTG-CTTACCTGCTGTACTACCGCCGTTCAGACACCCTG
13sk#8(1>236)                   <-                                                                TTCAGACACCCTG
SK #5(1>98)                     ->                                                                      CCCTG
                                CGGTCAGCGAGAAGCATGTCCTTCAGCCgCACACGCCTCGTGTGCCTTACCTGCTGTACTACCGCCGTTCAGACACCCTG
                                3530       3540       3550       3560       3570       3580       3590       3600
                                ....|....  ....|....  ....|....  ....|....  ....|....  ....|....  ....|....  ....|
2sk#8(1>182)                    -> CCACCGCAACAGCAGCAAACCTAGCT
2ks#9(1>159)                    <- CCACCGCAA
42KS#12(1>237)                  -> CCACCGCAACAGCAGCAAACCCAGCAGCAGAATGGCGGAGGATCAGGTGGAGTCGTGGGCAGCAGCTCATCATCATCGAA
SK #13(1>181)                   <- CCACCGCAACAGCAGCAAACCCAGCAGCAGAATGGCGGAGGATCAGGTGGAGTCGTGGGCAGCAGCTCATCATCATCG
KS #2(1>211)                    -> CCACCGCAACAGCAGCAAA-CCAGCAGCAGAATGGCGGAGGATCAGGTG-AGTCGTGGGCAGCAGCTCATCATCATCGAA
52ks#13(1>157)                  -> CCACCGCAACAGCAGCAAACCCAGCAGCAGAATGGCGGAGGATCAGGTGGAGTCGTGGGCAGCAGCTCATC
13sk#8(1>236)                   <- CCACCGCAACAGCAGCAAACCCAGCAGCAGAATGGCGGAGGATCAGGTGGAGTCGTGGGCAGCAGCTCATCATCATCGAA
SK #5(1>98)                     -> CCACCGCAACAGCAGCAAACCCAGCAGCAGAATGGCGGAGGATCAGGTGGAGTCGTGGGCAGCAGCT-ATCATCATCGAA
23KS#11(1>208)                  ->      CGCAACAGCAGCAAACCCAGCAGCAGAATGGCGGAGGATCAGGTGGAGTCGTGGGCAGCAGCTCATCATCATCGAA
13ks#9(1>237)                   ->               AGCAAACCCAGCAGCAGAATGGCGGAGGATCAGGTGGAGTCGTGGGCAGCAGCTCATCATCATCGAA
47ks#13(1>233)                  <-                                                    GTGGAGTCGTGGGCAGCAGCTCATCATCATCGAA
                                CCACCGCAACAGCAGCAAACCCAGCAGCAGAATGGCGGAGGATCAGGTGGAGTCGTGGGCAGCAGCTCATCATCATCGAA
```

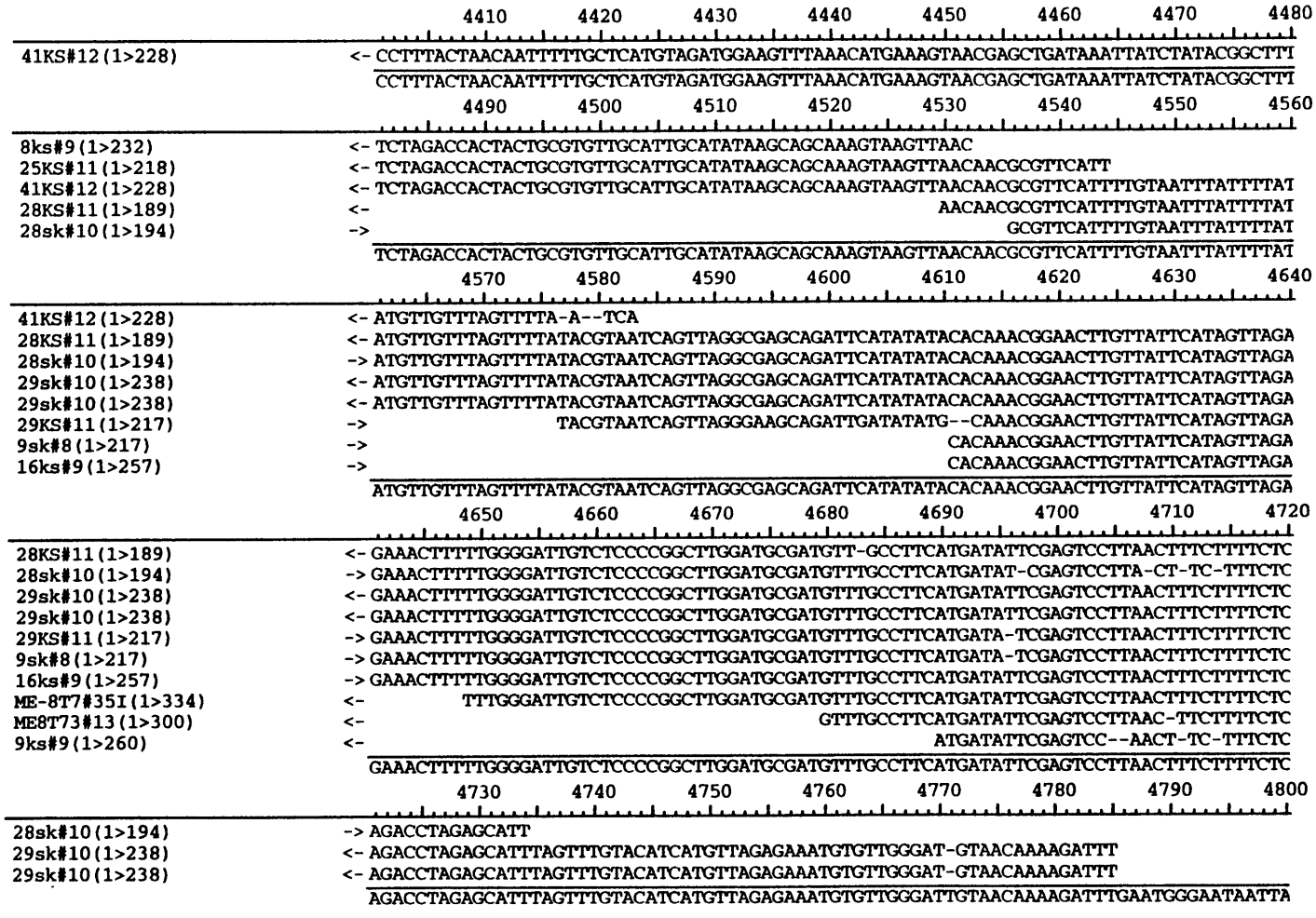

3610 3620 3630 3640 3650 3660 3670 3680
42KS#12(1>237) -> TGC-GGGGATAACAAGTGAAAGTGAACCACAAC
KS #2(1>211) -> TGC--GGGATAACAAGTGAAAGTGAACCACAAC
13sk#8(1>236) <- TGCCGGGGATAACAAGTGAAAGTGAACCACAACGCGGCGTTGTAAAAATACATCAGAACTTTAAGTGCTAGCGATATACT
SK #5(1>98) -> TGC--GGGATA-CA--TGA
23KS#11(1>208) -> TGCCGGGGATAACAAGTGAAAGTGAACCACAACGCGGCGTTGTAAAAATACATCAGAACTTTAAGTGCTAGCGATATACT
13ks#9(1>237) -> TGCCGGGGATAACAAGTGAAAGTGAACCACAACGCGGCGTTGTAAAAATACATCAGAACTTTAAGTGCTAGCGATATACT
47ks#13(1>233) <- TGCCGGGGATAACAAGTGAAAGTGAACCACAACGCGGCGTTGTAAAAATACATCAGAACNTTAAGTGCTAGCGATATACT
1sk#8(1>196) -> CGGGGATAACAAGTGAAAGTGAACCATAACGCGGCTTTGTAAAAATACATCAGAACTTTAAGTGCTAGCGATATACT
ME-8C6NT#35I(1>283) <- GTTGTAAAAATACATCAGAACTTTAAGTGCTAGCGATAT-CT
SK #2(1>262) <- CTTTAAGTGCTAGCGATATACT
TGCcGGGGATAACAAGTGAAAGTGAACCACAACGCGGCGTTGTAAAAATACATCAGAACTTTAAGTGCTAGCGATATACT
3690 3700 3710 3720 3730 3740 3750 3760
13sk#8(1>236) <- TTATATTTTTGTGTCAGTGCGTGTGAGCGAGAATGCGAGTTTGGTTTTATATGCATAT-CATAT
23KS#11(1>208) -> TTATATTTTTGTGTCAGTGCGTGTGAGCGAGAATGCGAGTTTGGTTTTATAT
13ks#9(1>237) -> TTATATTTTTGTGTCAGTGCGTGTGAGCGAGAATGCGAGTTTGGTTTTATATGCATATACATATACAA-TTAACAATTAG
47ks#13(1>233) <- TTATATTTTTGTGTCAGTGCGTGTGAGCGAGAATGCGAGTTTGGTTTTATATGCATATACATATACAAATTAACAATTAG
1sk#8(1>196) -> TTATATTTTTGTGTCAGTGCGTGTGAGCGAGAATGCGAGTTTGGTTTTATATGCATATACATATACAAATTAACAATTAG
ME-8C6NT#35I(1>283) <- TTATATTTTTGTGTCAGTGCGTGTGAGCGAGAATGCGAGTTTGGTTTTATATGCATATACATATACAAATTAACAATTAG
SK #2(1>262) <- TTATA-TTTTGTGTCAGTGCGTGTGAGCGCGAATGCGAGTTTGGTTTTATATGCATATACATATACAAATTAAAAATTAG
23sk#10(1>247) -> TTATATTTTTGTGTCAGTGCGTGTGAGCGAGAATGCGAGTTTGGTTTTATATGCATATACATATACAAATTAACAATTAG
KS #17(1>210) -> AGTGCGTGTGAGCGAGAATGCGAGTTTGGTTTTATATGCATATACATATAGAAATTAACAATTAG
KS #5(1>117) <- CGTGTGAGCGAG-ATGCGAG-TT-G-TTTATATGCATATACATATACAAATTAAGAATTAG
1ks#9(1>189) <- GAAT-CGAGTTTGG-TTTATATGCATATAGATAT-CAAATTAACAACTTG
36sk#10(1>188) <- ATGCATATACATATACAAATTAACAATTAG
SK #3 SPLIT A(1>96) -> ATA-ATATACAAATTAACAATTAG
TTATATTTTTGTGTCAGTGCGTGTGAGCGAGAATGCGAGTTTGGTTTTATATGCATATACATATACAAATTAACAATTAG
3770 3780 3790 3800 3810 3820 3830 3840
13ks#9(1>237) -> ATATAT-CGAGC
47ks#13(1>233) <- ATATATTCGAGCGGTATATAGTTAAGCCCAAAGTCCAAA
1sk#8(1>196) -> ATATA-TCGAGCGGTATATAGTTAAGCCCAAAGTCCAAA
ME-8C6NT#35I(1>283) <- ATATATTCGAGCGGTATATAGTTAAGCCCAAAGTCCAAAGCGCATGACCGTGTGTTTAGAGTACGATTGTGTGAGAAAGAA
SK #2(1>262) <- ATATATTCGAGCGGTATATAGTTAAGCCCAAAGTCCGAAGCGCATGACCGTGTGTTTAGAGTACGATTGTGTGAGAAAGAA
23sk#10(1>247) -> ATATA-TCGAGCGGTATATAGTTAAGCCCAAAGTCCAAAGCGCATGACCGTGTGTTAGAGTACGATTGTGTGACAAACAA
KS #17(1>210) -> ATATATTCGAGCGGTATATAGTTAAGCCCAAAGTCCAAAGCGCATGACCGTGTGTTAGAGTACGATTGTGTGAGAAAGAA
KS #5(1>117) <- ATATATTCGAGCGGTACATAGTTAAGCCCAAAGTCCAAAGCGCATGACCGTGTGTTTAGAG
1ks#9(1>189) <- ATATATTCGAGCGGTATAT-GTTAAGCCCAAAGTCCAAAGCGCATGACCGTGTGTTTAG-GTACGATTGTGTGAGAAAGAA
36sk#10(1>188) <- ATATATTCGAGCGGTATATAGTTAAGCCCAAAGTCCAAAGCGCATGACCGTGTGTTTAG-GTACGATTGTGTGAGAAAGAA
SK #3 SPLIT A(1>96) -> ATATATTCGAGCG-TATATAGTTAAGCCCAAAGTCCAAAGCGCATGACCGTGTGTTTAGAGTAC-ATTGTGTGAGA
ATATATTCGAGCGGTATATAGTTAAGCCCAAAGTCCAAAGCGCATGACCGTGTGTTTAGAGTACGATTGTGTGAGAAAGAA

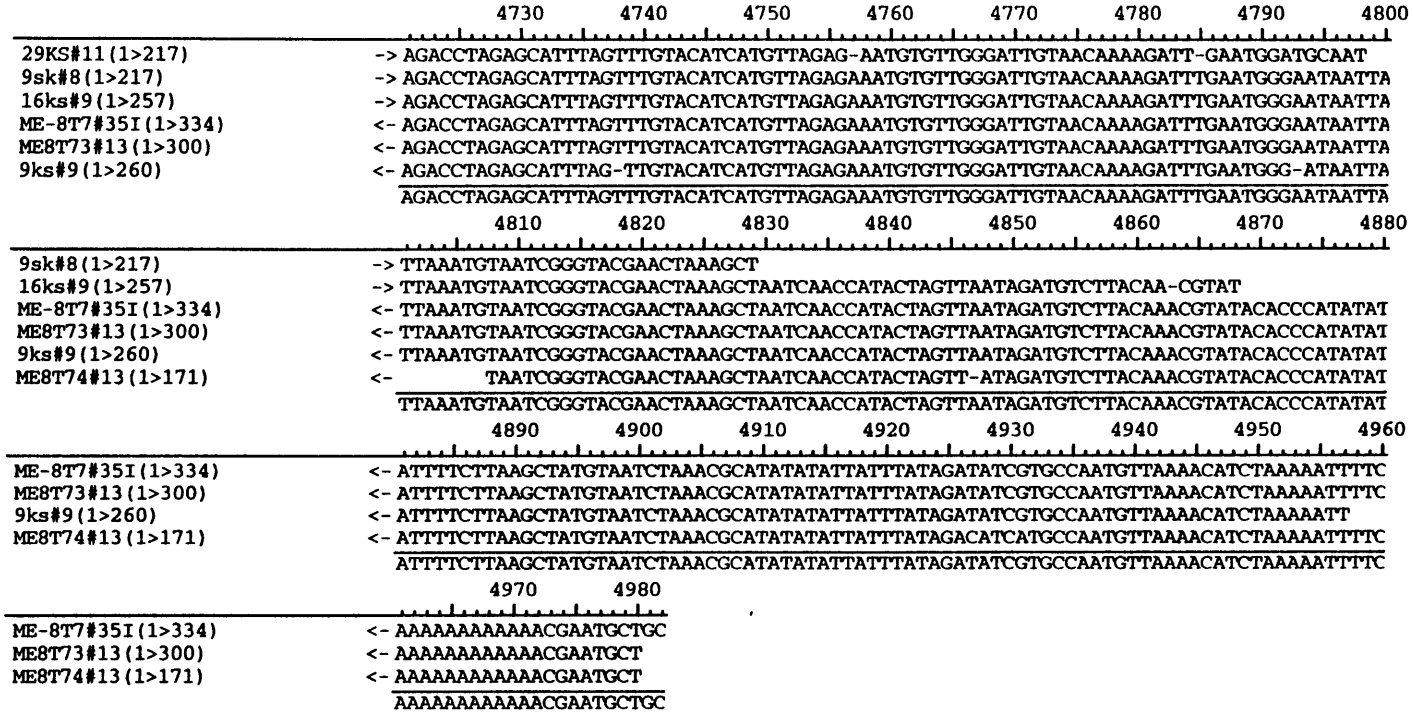
Project: ME Sequence Contig Contig 5




4090 4100 4110 4120 4130 4140 4150 4160
ME-8C4CT#35I(1>214) -> TGCCCTTAAATACCTTTGGAATTTGGCGATATGTTATG-CAA-CGCTCG
C4CT #17(1>268) -> TGCCCTTAAATACCTTTGGAATTTGGCGATATGTTATGACAAACGCTTACAGCTGTATATATACACAAATAATCATATAT
10sk#8(1>249) <- TGCCCTTAAATACCTTTGGAATTTGGCGATATGTTATGACAAACGCTTACAGCTGTATATATACACAAATAATTATATAT
10ks#9(1>252) -> TGCCCTTAAATACCTTTGGAATTTGGCGATATGTTATGACAAACGCTTACAGCTGTATATATACACAAATAATCATATAT
KS #16(1>155) -> TGCCCTTAAATACCTTTGGAATTTGGCGATATGTTATGACAA-CGCTTAC
SK #17(1>146) <- TGCCCTTAAATACCTTTGGAATTTGGCGATATGTTATCACAAACG
KS #10(1>241) <- TGCCCTTAAATACCTTTGGAATTTGGCGATATGTTATGACAAACGCTTACAGCTGTATATATACACAAATAATCATATAT
25sk#10(1>202) -> GGCGATATGTTATGACAAACGCTTACAGCTGTATATATACACAAATAATCATATAT
TGCCCTTAAATACCTTTGGAATTTGGCGATATGTTATGACAAACGCTTACAGCTGTATATATACACAAATAATCATATAT
4170 4180 4190 4200 4210 4220 4230 4240
C4CT #17(1>268) -> GGTAGTTGATACGTTTCGCGG
10sk#8(1>249) <- TGTAG-TGA
10ks#9(1>252) -> -GATGTTGATCAG-TTC-NGGCATGCACGA
KS #10(1>241) <- TGTAGTTGATCAGTTTCGCGGCATGCACCATGCATTTAAACAAACAAACCCAGCAACAATA
25sk#10(1>202) -> TGTAGTTGATCAGTTTCGCGGCATGCACCATGCATTTAAACAAACAAACCCAGCAACAATATATATCGAATGCTTTGAAT
8sk#8(1>203) -> GT-GTTAATCAGTTTCGCGGCATGCACCATGCATTTAAACAAACAAACCCAGCAACAATATATATCGAATGCTTTGAAT
TGTAGTTGATCAGTTTCGCGGCATGCACCATGCATTTAAACAAACAAACCCAGCAACAATATATATCGAATGCTTTGAAT
4250 4260 4270 4280 4290 4300 4310 4320
25sk#10(1>202) -> ATGCTTTAAAGAAATATACATTTAGAACCAAGGTGTTTTCCACACAAAGAGTAAATACAGGC-GAGGT
8sk#8(1>203) -> ATGCTTTAAAGAAATATACATTTAGAACCAAGGTGTTTTCCACACAAAGAGTAAATACAGGCAGTGGTGGTCAGAGCAAT
41sk#12(1>144) -> CTTTAAAGAAATATACATTTAGAACCAAGGTGTTTTCCACACAAAGAGTAAATACAGGCAGTGGTGGTCAGAGCAAT
ME T3(1>118) <- AAGAGTAAATACA-GCAGT-GTGGTCAGAGCAAT
8ks#9(1>232) <- TAAATACAGGCAGTGGTGGTCAGAGCAAT
ATGCTTTAAAGAAATATACATTTAGAACCAAGGTGTTTTCCACACAAAGAGTAAATACAGGCAGTGGTGGTCAGAGCAAT
4330 4340 4350 4360 4370 4380 4390 4400
8sk#8(1>203) -> AGCCGCAAGCATTTCAGCATCTGCAAGCCAAATTGGGTTACTTTTGTAATGGTCTCGAAGACCAAATCAAATAGATA
41sk#12(1>144) -> AGCCGCAAGCATTTCAGCATCTGCAAGCCAAATTGGGTTACTTTTGTAATGGTCTCGAAGACCAAAT
ME T3(1>118) <- AGCCGCAAGCATTTCAGCATCTGTAAGCCAAATTGGGTTACTTTTGTTATGGTCTCGAAGA
8ks#9(1>232) <- AGC-GCAAGCA-TTCAGCATCTG-NAG-CAAATTGGGTTACTTTTGTAATGGTCTCGAAGACCAAATCAAATAGATATTT
25KS#11(1>218) <- CGCAAGCATTTCAGCATCTGCAAGCCAAATTGGGTTACTTTTGTAATGGTCTCGAAGACCAAATCAAATAGATATTT
41KS#12(1>228) <- AAATTGG-TNACTTT-GTAATGGTCTCGAAGACNAAATCAAATAGATATTT
AGCCGCAAGCATTTCAGCATCTGCAAGCCAAATTGGGTTACTTTTGTAATGGTCTCGAAGACCAAATCAAATAGATATTT
4410 4420 4430 4440 4450 4460 4470 4480
8ks#9(1>232) <- CCTTTACTAACAATTTTTGCTCATGTAGATGGAAGNNTAAACATGAAAGTAACGAGCTGATAAATTATCTATACGGCTTT
25KS#11(1>218) <- CCTTTACTAACAATTTTTGCTCATGTAGATGGAAGTTTAAACATGAAAGTAACGAGCTGATAAATTATCTATACGGCTTT
CCTTTACTAACAATTTTTGCTCATGTAGATGGAAGTTTAAACATGAAAGTAACGAGCTGATAAATTATCTATACGGCTTT

Project: ME Sequence Contig Contig 5




4730 4740 4750 4760 4770 4780 4790 4800
29KS#11(1>217) -> AGACCTAGAGCATTTAGTTTGTACATCATGTTAGAG-AATGTGTTGGGATTGTAACAAAAGATT-GAATGGATGCAAT
9sk#8(1>217) -> AGACCTAGAGCATTTAGTTTGTACATCATGTTAGAGAAATGTGTTGGGATTGTAACAAAAGATTTGAATGGGAATAATTA
16ks#9(1>257) -> AGACCTAGAGCATTTAGTTTGTACATCATGTTAGAGAAATGTGTTGGGATTGTAACAAAAGATTTGAATGGGAATAATTA
ME-8T7#35I(1>334) <- AGACCTAGAGCATTTAGTTTGTACATCATGTTAGAGAAATGTGTTGGGATTGTAACAAAAGATTTGAATGGGAATAATTA
ME8T73#13(1>300) <- AGACCTAGAGCATTTAGTTTGTACATCATGTTAGAGAAATGTGTTGGGATTGTAACAAAAGATTTGAATGGGAATAATTA
9ks#9(1>260) <- AGACCTAGAGCATTTAG-TTGTACATCATGTTAGAGAAATGTGTTGGGATTGTAACAAAAGATTTGAATGGG-ATAATTA
AGACCTAGAGCATTTAGTTTGTACATCATGTTAGAGAAATGTGTTGGGATTGTAACAAAAGATTTGAATGGGAATAATTA
4810 4820 4830 4840 4850 4860 4870 4880
9sk#8(1>217) -> TTAAATGTAATCGGGTACGAACTAAAGCT
16ks#9(1>257) -> TTAAATGTAATCGGGTACGAACTAAAGCTAATCAACCATACTAGTTAATAGATGTCTTACAA-CGTAT
ME-8T7#35I(1>334) <- TTAAATGTAATCGGGTACGAACTAAAGCTAATCAACCATACTAGTTAATAGATGTCTTACAAACGTATACACCCATATAT
ME8T73#13(1>300) <- TTAAATGTAATCGGGTACGAACTAAAGCTAATCAACCATACTAGTTAATAGATGTCTTACAAACGTATACACCCATATAT
9ks#9(1>260) <- TTAAATGTAATCGGGTACGAACTAAAGCTAATCAACCATACTAGTTAATAGATGTCTTACAAACGTATACACCCATATAT
ME8T74#13(1>171) <- TAATCGGGTACGAACTAAAGCTAATCAACCATACTAGTT-ATAGATGTCTTACAAACGTATACACCCATATAT
TTAAATGTAATCGGGTACGAACTAAAGCTAATCAACCATACTAGTTAATAGATGTCTTACAAACGTATACACCCATATAT
4890 4900 4910 4920 4930 4940 4950 4960
ME-8T7#35I(1>334) <- ATTTTCTTAAGCTATGTAATCTAAACGCATATATATTATTTATAGATATCGTGCCAATGTTAAAACATCTAAAAATTTTC
ME8T73#13(1>300) <- ATTTTCTTAAGCTATGTAATCTAAACGCATATATATTATTTATAGATATCGTGCCAATGTTAAAACATCTAAAAATTTTC
9ks#9(1>260) <- ATTTTCTTAAGCTATGTAATCTAAACGCATATATATTATTTATAGATATCGTGCCAATGTTAAAACATCTAAAAATT
ME8T74#13(1>171) <- ATTTTCTTAAGCTATGTAATCTAAACGCATATATATTATTTATAGACATCATGCCAATGTTAAAACATCTAAAAATTTTC
ATTTTCTTAAGCTATGTAATCTAAACGCATATATATTATTTATAGATATCGTGCCAATGTTAAAACATCTAAAAATTTTC
4970 4980
ME-8T7#35I(1>334) <- AAAAAAAAAAAAACGAATGCTGC
ME8T73#13(1>300) <- AAAAAAAAAAAAACGAATGCT
ME8T74#13(1>171) <- AAAAAAAAAAAAACGAATGCT
AAAAAAAAAAAAACGAATGCTGC