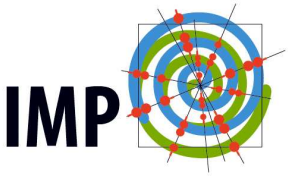


DISSERTATION

Titel der Dissertation

Pax3 defines a new pathway at pericentric heterochromatin

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr. rer.nat.)

Verfasserin / Verfasser:	Manuela Scaranaro
Matrikel-Nummer:	0603175
Dissertationsgebiet (lt. Studienblatt):	Molekulare Biologie
Betreuerin / Betreuer:	Prof. Dr. Thomas Jenuwein
Wien, am Mai 2009	

TABLE OF CONTENTS

SUMMARY

The association between DNA and histones gives rise to chromatin. In interphase cells there are two different kinds of chromatin: euchromatin and heterochromatin. Pericentric heterochromatin represents the vast majority of the heterochromatin and several features allow to distinguish it from the rest of the chromatin. Pericentric heterochromatin is characterized by a repetitive DNA sequence: in mouse a 234 bp motif (major satellite) is present around 10´000 times in every chromosome. One of the first events that takes place there is the transcription of the major satellites and the subsequent formation of dsRNA. Thereafter, some proteins (such as Suv39h enzymes and HP1α) and histone posttranslational modifications (H3K9me3 and H4K20me3) are enriched there.

Pax3 is a protein so far very well characterized for its developmental role. By EGFP-tagging, it was observed that, like a few other proteins, Pax3 localizes at pericentric heterochromatin in mouse fibroblasts and its accumulation there is Suv39h-independent: Pax3 recognizes, by its homeodomain, a palindromic sequence of the second subrepeat of the major satellite. Fibroblasts, where only a truncated Pax3 is expressed have been generated: this truncated protein cannot bind the major satellite and therefore it mimics a knockout. In Pax3 null cells repressive histone methylation is reduced and the amount of transcripts (dsRNA and ssRNA) generated from the major satellites increases. The data generated suggest that it is indeed the local increase of the amount of ssRNA and dsRNA to be the cause of the reduction of H3K9me3, H4K20me3 and H3K27me1 that usually decorate pericentric heterochromatin.

Pax3 pericentric localization defines a novel pathway: transcription of the major satellites needs to be repressed by developmental regulators in order to establish heterochromatin (Scaranaro et al., submitted).

In this thesis in addition we will ask whether dsRNA can be a modulator of HMTase activity and it will be shown that mapping new Su(var) genes in Drosophila allows to discover new pericentric components in other species too.

1. INTRODUCTION

1.1 Epigenetics and chromatin

The possibility of reading the sequence of entire genomes has been one of the most important tools of the recent years (Morozova et al., 2008; Venter et al., 2001). But, most likely, even to an expert, every genome appears nothing more than a combination of 4 different nucleotides: which of them is in a certain position is hard to predict and difficult to explain. The reading of 3.2 billions of nucleotides cannot explain why we are as we are and why the keratinocytes of my finger that in this moment writes this text are different from the neurons of my brain that at the same time think of it, although both of them contain exactly the same succession of nucleotides: my genome.

Some differences make my keratinocytes proper keratinocytes and my neurons real neurons. These differences are not only in the DNA sequence but are mainly in the proteins associated with the DNA, the histones, and in the identity of the transcription factors activated in each cell type.

The combination between 147 nucleotides and 4 different pairs of histones (H2A, H2B, H3 and H4) gives rise to the nucleosome (Luger et al., 1997), the basic unit of the chromatin (reviewed in Quina et al., 2006). Histones are among the smallest and the most conserved proteins; their C-terminal domain is used for assembling the nucleosome, the N-terminal domain has less structural constrains and it is more prone to be post translational modified. The modifications taking place mainly on the N-terminal domains of the different histones can explain why a keratinocyte is different from a neuron: the nucleosomes associated with a certain part of the genome have some modifications in the keratinocyte and other modifications in the neuron. The keratinocyte and the neuron have the same genome, but two different epigenomes. So, in contrast to Genetics, that considers only the nucleotide sequence of the genome, Epigenetics relies on the chromatin.

Transcription factors have a fundamental role in establishing which modifications will be present on the histones associated with a certain DNA sequence according to the cell type we deal with. The different histone modifications regulate the status of the chromatin and the transcription profile and lead in this way to the stabilization of the expression programmes.

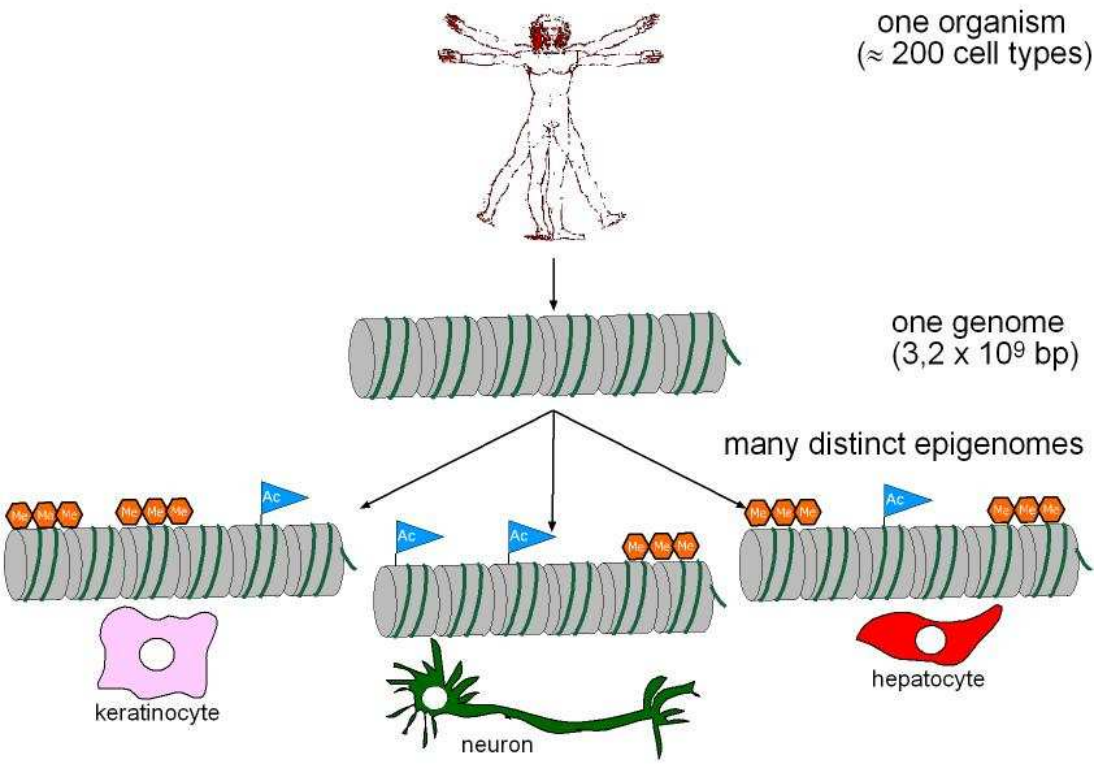


Figure 1 One genome, several epigenomes. Although almost all the cells of an organism contain exactly the same genome, their epigenomes can be extremely different. This is also what Epigenetics deals with: the post-translational modifications on the histones that render a cell (for example a keratinocyte) different from others (for example a neuron or a hepatocyte).

1.2 Distinct types of chromatin

A number of different dyes can stain chromatin. Among them, one of the most widely used is DAPI (4',6-Diamidino-2-phenylindole), which preferentially stains especially AT-rich DNA fragments.

When interphase cells are stained with DAPI, mainly two different levels of intensity can be observed. The parts having a very weak staining form the so-called euchromatin; others are stained very deeply and for this reason the chromatin there is considered heterochromatin. The different staining reflects a different structure and anticipates a difference in the function and in the accessibility: while the euchromatin is more relaxed and transcriptionally active, the heterochromatin has a very compact structure that makes the transcription not completely impossible but rather difficult.

Some parts of the genomes, such as the regions flanking the chromocenters, are always heterochromatic and for this reason they form the constitutive heterochromatin.

In mammalian female interphase cells an additional heterochromatic region can be observed: the inactive X-chromosome. Since this chromosome is inactive only in female cells it constitutes an example of facultative heterochromatin (reviewed in Starmer and Magnuson, 2009).

The highest level of chromatin structure organization is achieved during mitosis, when all the chromosomes are condensed throughout their entire lengths.

Independently of the cell cycle phase and of the level of chromatinization, chromatin always needs to be very flexible and ready to be condensed and decondensed: all the chromatin is transcribed, even the heterochromatic regions, and in addition the entire genome is faithfully replicated with every round of cell division. The flexibility making chromatin a very organized structure that becomes relaxed at the right moments is what renders it also very fascinating.

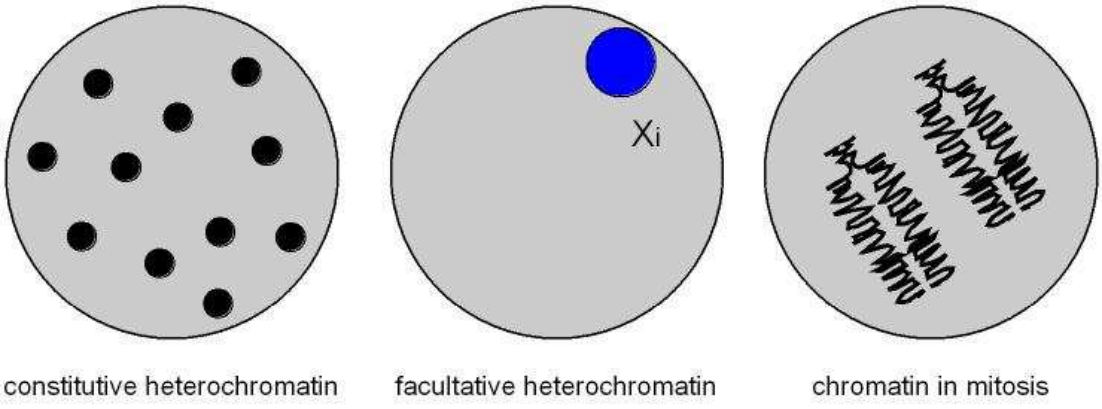


Figure 2 Heterochromatin. Heterochromatin can be either constitutive (like the pericentric heterochromatin) or facultative (like the inactive X in mammalian female organisms). During mitosis all the chromatin is very condensed.

1.3 Hallmarks of mouse heterochromatin

Pericentric heterochromatin can be observed in several organisms, and in recent years it has become clear that its features are largely conserved from yeast to mammals. Unlike X, which replicates early in S-phase, pericentric heterochromatin is always replicated either in the middle or at the end of the S-

phase and it is often transcriptionally inactive, although not completely silent as had been thought for a long time. Heterochromatic proteins and the characteristic histone modifications are also very well conserved (Martin et al., 2005).

The DNA sequence within pericentric heterochromatin is always repetitive, but there is no specific sequence conservation among different species. In S. pombe two different sequences (termed dh and dg) are present and they can be in two different orientations (Grewal et al., 2007). In D. melanogaster several types of AT-rich repeats can be found: their length and their sequences are very different and the shortest repeat is just 5 nt long (Bosco et al., 2007; Ugarkovic, 2005). In mouse pericentric heterochromatin the major satellite, also called γ-satellite, represents the basic repeat (Joseph et al., 1989; Kuznetsova et al., 2006). The major satellite is 234 bp long and has a very low degree of polymorphism. In humans the basic repeat is the α-satellite, which consists of 171 bp and whose sequence can be slightly different from chromosome to chromosome (Waye et al., 1986; Warburton et al., 2008).

Although they are always AT-rich, all the repeats contain around 30% of GC: at pericentric heterochromatin methylation of cytosine residues at carbon 5 within CpG dinucleotides can be observed (Bird et al., 1979; reviewed in Bird, 2002). DNA methylation (MeCpG) is the only known epigenetic modification of DNA in mammalian cells and has a fundamental role for chromatin organization (Ma et al., 2005). The enzymes responsible for DNA methylation are Dnmt1, Dnmt3a and Dnmt3b. Through Dnmt1 activity, newly-replicated DNA can be methylated on the daughter strand, using the parental methylated strand as a template. In the absence of Dnmt1, DNA remains hemimethylated after replication, and the structure of chromatin is severely compromised (Li et al., 1992; Ma et al., 2005). DNA methylation is also dependent on a specific modification on histone H3: the trimethylation of lysine 9 (H3K9me3). In cells where H3K9me3 is abrogated by deleting both enzymes that catalyze this reaction, DNA methylation at pericentric heterochromatin is significantly altered (Lehnertz et al., 2003). The exact link between DNA methylation and H3K9me3 is still, however, poorly understood.

H3K9me3 is an important hallmark of pericentric heterochromatin. In mouse, Suv39h1 and Suv39h2 (Aagaard et al., 1999; O´Carroll et al., 2000) catalyze the

conversion of H3K9me1 into H3K9me3 (Rea et al., 2000). H3K9me3 creates a binding site for HP1 (Lachner et al., 2001; Bannister et al., 2001), a small characteristically heterochromatic protein conserved in several species (Lachner et al., 2003).

Mouse heterochromatin is also characterized by the presence of another two histone lysine methylation marks: H4K20me3 and H3K27me1. H4K20me3 is catalyzed by the enzymes Suv420h1 and Suv420h2 and can be found, like H3K9me3, in all the nuclei of cultured cells. Pericentric H4K20me3 is dependent on H3K9me3: mutants lacking H3K9me3 lack H4K20me3 as well (Schotta et al., 2004). On the other hand, H3K27me1 can be found only in about half of the nuclei of cultured cells and it is also dependent on H3K9me3 (Rice et al., 2003; Lachner et al., 2004; Schotta et al., 2004).

Histone lysine methylation can be removed by enzymes containing the jmjC domain, a group of thirty proteins in mammals. Five of these enzymes comprise the Jmjd2 group, and antagonize Suv39h activity at pericentric heterochromatin by removing H3K9me3 (Fodor et al., 2006; Klose et al., 2006). Little is known about the in vivo regulation of the balance between methylation and demethylation at mammalian heterochromatin.

A recently discovered feature of heterochromatin is transcriptional activity that produces non-coding RNA. Work initially done in S. pombe showed that repeats are transcribed (Volpe et al., 2002). The transcripts are then processed and this phenomenon is closely related to histone lysine methylation: in mutants where RNA processing is not possible, H3K9me3 decreases (Morey et al., 2004; Grewal et al., 2007; Huisinga et al., 2006). The enzymes responsible for the transcription and processing of the non-coding RNA are the key players of the siRNA pathway. The siRNA pathway allows post-transcriptional gene regulation. Its main steps are very well conserved: the silencing RNA is first transcribed and then amplified by an RNA-dependent RNA polymerase (Chen et al., 2008). During this reaction, the single-stranded RNA (ssRNA) is converted into double-stranded RNA (dsRNA). The dsRNA is then processed in 20 nt fragments by Dicer and the small RNA is then loaded into the RISC complex (RNA-induced silencing complex). One of the proteins in the RISC complex is Argonaute. This

protein is responsible for the degradation of the target mRNA, when this is bound by the short silencing RNA (Iida et al., 2008).

Non-coding RNA transcripts generated from heterochromatin are required to maintain heterochromatin structure: in cultured mouse (L929) and human cells (HeLa), the treatment with RNaseA causes the loss of both H3K9me3 and HP1α at pericentric heterochromatin, (Maison et al., 2002). In addition, the transcription of the major satellites specifically generates dsRNA: dsRNA is not generated at other interdispersed repeats (Martens et al., 2005).

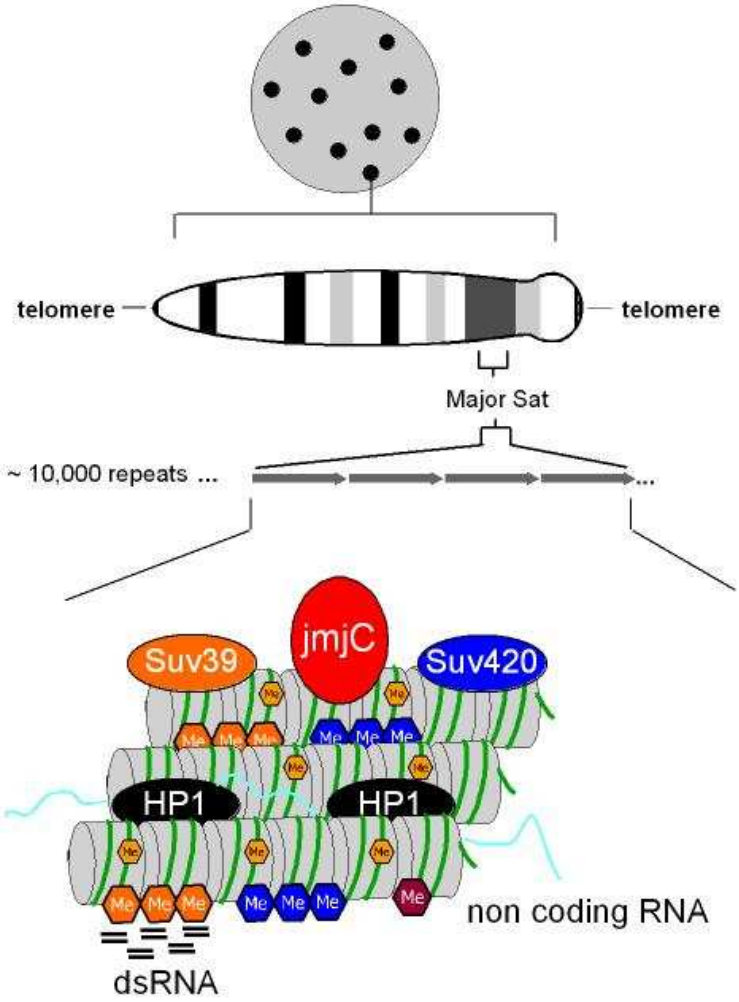


Figure 3 Hallmarks of mouse pericentric heterochromatin. Major satellites are responsible for the formation of pericentric heterochromatin. There, several enzymes localize: among them, histone methyltransferases (Suv39h and Suv420h) and histone demethylases (jmjC proteins) regulate, by coordinating the writing and the erasing, the amount of H3K9me3 and H4K20me3. No demethylase responsible for the removal of H4K20me3 has been identified so far. H3K9me3 provides a binding site for HP1 and regulates DNA methylation. Heterochromatin is also transcribed and is enriched in dsRNA.

1.4 DNA sequence of the mouse heterochromatin

In mouse, the major satellite is 234 bp long and can be divided into four different subrepeats whose length is 58-60 bp (Vissel et al., 1989). The four subrepeats share a common motif in their sequence. A number of single nucleotide polymorphisms (SNPs) have been reported within major satellite sequences, but they present in general a minor degree of polymorphism in comparison to repeats in other species. Major satellites are particularly enriched in AT and they can be tandemly repeated in a chromosome around 10´000 times (Garagna et al., 2002).

Although the sequence of individual major satellites has been well characterized, very little is known about their organization into longer structures. The reason is the impossibility of sequence them properly, as it has been done for the non repetitive parts of the genome. Their repetitive nature makes it very difficult to assemble sufficiently long stretches of sequence information (Martens et al., 2005).

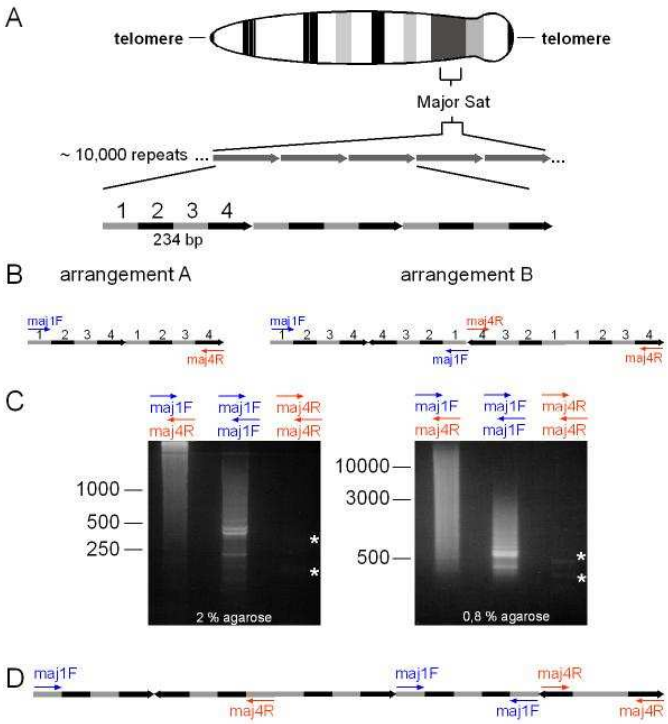
A very simple experiment can be performed to understand the complex organization of the basic unit of the major satellites: genomic DNA can be amplified with forward and reverse primers designed against the major satellite repeat (234 bp), or only with a forward or a reverse primer.

Distinct DNA fragments can be amplified when PCR is performed with only one primer. When the amplification uses only the forward primer, there are three bands very prominent whose sizes are respectively 250 bp, 400 bp and 450 bp, approximately. The longest fragments, which are also the more abundant, must be the amplification of two repeats with opposite orientation, while the shortest could be the amplification of two truncated repeats (having for example subrepeat 1 and subrepeat 2) with opposite orientation. When only the reverse primer is used for the amplification, two very weak bands can be observed. The size of the bigger one (≈400 bp) is comparable to the size of the two repeats with opposite orientation. The small fragment, as observed previously, could be the product of the amplification of two truncated repeats with opposite orientation.

We cannot explain why the amount of amplified DNA is so low when only the reverse primer is used in the PCR and 35 cycles are done.

When genomic DNA is amplified using a forward and a reverse primer, a smear of fragments whose length is 400-10000 bp can be observed. The range covered by the smear supports the idea that a group of satellites can have the same orientation for very long parts. The amount of DNA amplified with these two primers seems less than the amount amplified just by using the forward primer: an organization where a group of satellites have different orientation seems to be more common.

From these results, we can conclude that at pericentric heterochromatin the major satellites are arranged in a combination of two dispositions: a tandem array where the basic subrepeat always has the same orientation, and a sequence of repeats (some of them truncated) with alternate forward and reverse orientation.



<u>Figure 4 Mouse major satellites.</u> (A) Mouse pericentric heterochromatin consists of the repetition of the same sequence (major satellite) around 10'000 times. This sequence is 234 bp long and can be divided into four different subrepeats. (B) The exact organization of the repeats is not known. (C) By performing PCR with only one primer, we observed DNA amplification especially when the forward primer was used. We cannot explain the lower amount of amplified DNA when we used the reverse primer, but we can suggest (D) that the major satellites are a combination of repeats with forward and reverse orientation.

1.5 Mouse heterochromatin is transcribed

In S. pombe pericentric repeats are transcribed by RNA Polymerase II (RNAPol II). The RISC complex then associates with the nascent RNA. Subsequently, Chp1, one of the proteins forming the RISC, recruits Swi6 and the histone methyltransferase Clr4 (Morey et al., 2004; Grewal et al., 2007; Huisinga et al., 2006). Thus, at the heterochromatic repeats the RISC complex and the classical heterochromatic components are both present (reviewed in Cam et al., 2009). The chromatin starts to be more condensed and, as a consequence, it becomes impossible for RNAPol II to transcribe the repeats. In other words, the initial activity of RNAPol II allows the formation of heterochromatin, which, once formed, inhibits the enzyme.

No RNA-dependent RNA polymerase has been found in mammals. However, we can surmise that such an enzyme is not absolutely required to induce silencing using RNA: if the RNAPol II can transcribe bidirectionally, the two transcripts that are synthesized will be complementary and will then anneal, thus forming dsRNA, already proven to be a key molecule of pericentric heterochromatin (Martens et al., 2005).

Transcription of repeats can be regulated in different ways: certain repeats are highly transcribed under certain conditions, whereas for others the transcription is coupled to the cell cycle.

The analysis of the transcripts of human cells that have been either heat shocked or stressed showed that satellite III (5 nt long repeats) transcription is enhanced in these circumstances (Jolly et al., 2006; Sengupta et al., 2009). The low level of satellite III RNA detectable in untreated cells is increased ten thousand-fold after heat shock, through the activation of heat shock factor 1 (Rizzi et al., 2004; Jolly et al., 2004). Stressing agents, like UV-C, heavy metals and oxidative and hyper-osmotic stresses, also increase the transcription of satellite III with different extents of induction (>100- to 80´000-fold) and kinetics (Lakhotia et al., 2006; Valgardsdottir et al., 2008).

In mouse, the transcription of the major satellites (Gaubatz et al., 1990; Rudert et al., 1995) is coupled to the cell cycle (Wu et al., 2006; Lu et al., 2007). By RNA-FISH using major satellite specific probes it was observed that small groups of

transcripts surround the pericentric heterochromatin and their transcription is strictly regulated during cell cycle (Lu et al., 2008): major satellites are transcribed mainly during interphase and the beginning of the S-phase. Transcription decreases during G2 and is completely absent during mitosis.

1.6 Transcription factors at pericentric heterochromatin

10% of the genes encode for transcription factors. Transcription factors can recognize a specific DNA sequence by using one or more DNA-binding domains. The most common DNA-binding motifs are Zn-fingers, helix-loop-helix, helix-turn-helix and leucine zippers. The consensus sequence for many transcription factors has been characterized but it is more difficult to determine the in vivo DNA-binding.

In different organisms several transcription factors have been found to localize at pericentric heterochromatin and to bind repeats.

One of the strongest Su(var) in D. melanogaster is Su(var)3-7 (Cleard et al., 1997), a protein with 7 Zn-fingers that have been shown to bind different DNA sequences (Cleard and Spierer, 2001). In Drosophila Su(var)3-7 forms a strong complex with Su(var)3-9 and HP1 (Delattre et al., 2000; Delattre et al., 2004). Although these last two proteins are conserved in other species, no orthologues have been reported for Su(var)3-7. In D. melanogaster there is just another protein similar to Su(var)3-7, that has been called ravus (Delattre et al., 2002), since the gene that encodes for it localizes very close to Su(var)3-7 but is transcribed with opposite orientation.

In mouse more than 500 Zn-finger proteins have so far been described (Carninci et al., 2005) and one of them with 6 Zn-fingers, Gfi1b, has also been shown to localize at pericentric heterochromatin in B-cells (Vassen et al., 2006). Gfi1b is a transcription factor specific of B-cells and by band shift assay it was observed that it recognizes a 5 nt sequence in subrepeat 4 of major satellite. In Gfi1b-null cells, H3K9me3 at pericentric heterochromatin is reduced but no formal link has so far been established between the absence of a transcription factor binding the major satellite and the impairment of H3K9me3 at pericentric heterochromatin.

One third of the Zn-finger proteins in mammals have a KRAB domain (Huntley et al., 2006), and their repressive action upon transcription is possible only when the KRAB domain interacts with KAP1. This protein is 834 aminoacids long and contains a RING finger, B-boxes, a coiled-coil region, a PHD finger and a bromodomain. It has recently been shown that KAP1 binds HP1 and that their interaction is fundamental for removing RNAPol II from gene bodies, reducing H3K9Ac and H3K4me3 and increasing H3K9me3 and H4K20me3 at regulatory elements of genes (Sripathy et al., 2006).

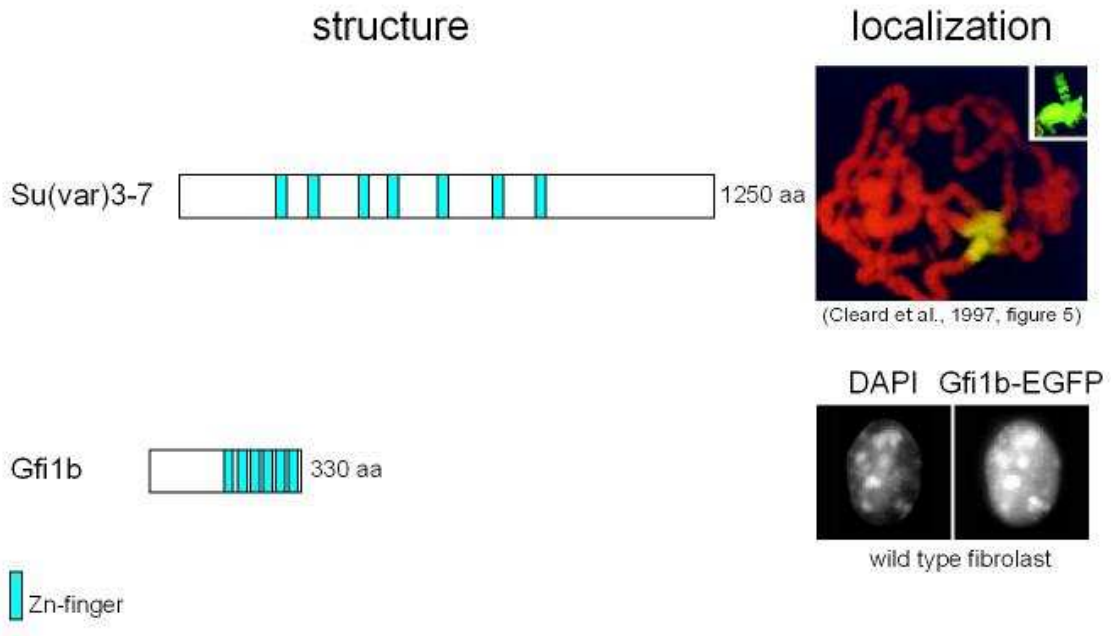


Figure 5 Examples of transcription factors at pericentric heterochromatin. Endogenous Su(var)3-7 (green) localizes at pericentric heterochromatin in polytene chromosomes (stained in red). Gfi1b also localizes at pericentric heterochromatin when overexpressed in mouse fibroblasts.

1.7 Aim of this study

Many of the proteins localizing at pericentric heterochromatin still have to be identified, and very little is known about transcription factors regulating RNA transcription of the major satellites. In order to characterize key players in this process, we made use of two different approaches. Many proteins have been shown in several publications to interact with HP1 (reviewed in Kwon and Workman, 2008) and among them we concentrated on bona fide transcription

factors. We then asked which of the HP1-interacting transcription factors have a binding site embedded into the major satellite.

By combining these two different approaches we observed that, among the HP1-interacting transcription factors, at least one could potentially bind the major satellite: Pax3.

The role of this protein in regulating development of the tissues derived from the neural crest has been well characterized and in a recent publication it was observed that the EGFP-tagged murine Pax3, transfected in HEK cells, localizes at DAPI-dense foci through its homeodomain (Hsieh et al., 2006). We first observed by band shift assay that both Pax3 and Pax7, a protein whose DNA-binding domains are similar to Pax3 ones, can bind the major satellite through the homeodomain. By using Pax3 null cells, we observed that Pax3 is a repressor of transcription at the major satellites: the absence of Pax3 is responsible for loss of heterochromatic histone methylation and for the destabilization of mouse heterochromatin.

On the basis of these data, we suggest a novel role for Pax3 and describe novel and important mechanisms for heterochromatin maintenance.

2. MANUSCRIPT: TRANSCRIPTION FACTOR Pax3 SAFEGUARDS RNA OUTPUT AT MOUSE HETEROCHROMATIN

2. Transcription factor Pax3 safeguards RNA output at mouse heterochromatin

Manuela Scaranaro,[1] Borbala Gerle,[1] Inti Alberto De La Rosa-Velazquez,[1,2] Johannes Popow,[1] Valentina Perrera,[1,2] Mario Richter,[1] Susanne Opravil,[1] Michaela Pagani[1] and Thomas Jenuwein[1,2]

[1] Research Institute of Molecular Pathology (IMP), The Vienna Biocenter, A-1030 Vienna, Austria, [2] Max Planck Institute for Immunobiology, Stübeweg 51, D-79108 Freiburg, Germany

2.1 Abstract

Pax3 is a well characterized transcription factor having an important function in development and here we show that, in addition, Pax3 works as a repressor of transcription at the major satellite. By using a candidate approach focused on bona fide transcription factors interacting with HP1 and having a heterochromatin-embedded binding site, we identified Pax3 as a key component of heterochromatin. Pax3 recognizes a palindromic sequence within the major satellite with its homeodomain and represses their transcription. In Pax3 null cells the increased amount of non-coding transcripts allows to identify some 30-40 bp long and leads to the loss of H3K9me3 and H4K20me3 at pericentric heterochromatin. These data show that transcription of the major satellites needs to be repressed in order to form heterochromatin and that this process can be regulated by transcription factors, also involved in other pathways, that can recognize the pericentric repeats.

2.2 Introduction

Pericentric heterochromatin has a fundamental role in different organisms and it has several characteristic features that distinguish it from the rest of the chromatin. The DNA at pericentric heterochromatin consists of repeat elements, whose sequence depends on the species: in mouse the basic repeat is the major satellite (Joseph et al., 1989; Kuznetsova et al., 2006). This AT-rich sequence is 234 bp long and can be divided into four different subrepeats (Vissel et al., 1989). The major satellite is repeated 10'000 times in every chromosome and it therefore represents 10% of the mouse genome (Garagna et al., 2002).

At pericentric heterochromatin, C5 methylation of cytosine residues within CpG dinucleotides can be observed (Bird et al., 1979). 70% of all CpG sites are methylated. Dnmt1 is the enzyme responsible for cytosine methylation after every DNA replication and its activity is fundamental for chromatin organization: Dnmt1 knockout mice show significant loss of DNA methylation and embryonic lethality (Li et al., 1992; Ma et al., 2005). DNA methylation is coregulated in part by a specific modification on histone H3: the trimethylation of lysine 9. In cells where Suv39h1 and Suv39h2, the enzymes that convert H3K9me1 into H3K9me3 (Rea et al., 2000), are knocked out, DNA methylation at pericentric heterochromatin is altered (Lehnertz et al., 2003).

H3K9me3 is one of the characteristic histone modifications of pericentric heterochromatin and it regulates other pericentric modifications, like H4K20me3 and H3K27me1. In Suv39h-dn cells pericentric heterochromatin loses not only H3K9me3 (Peters et al., 2001), but also H4K20me3 and H3K27me1 (Schotta et al., 2004; Lenhertz et al., 2003). In addition, H3K9me3 creates a binding site for HP1 (Lachner et al., 2001; Bannister et al., 2001), a structural component of heterochromatin.

Heterochromatin can also be transcribed, giving rise in this way to non-coding RNA. This phenomenon has been observed in several organisms. In humans the transcription of satellite III repeats, always hard to detect, is enhanced with different extent of induction (>100- to 80.000-fold) and kinetics upon heat shock and stress (Rizzi et al., 2004; Jolly et al., 2004; Valgardsdottir et al., 2008). Work initially done in S. pombe showed that heterochromatic transcripts are processed

and in RNAi mutant strains H3K9me3 is altered (Volpe et al., 2002). In S. pombe pericentric heterochromatin consists of two sequences, dh and dg, repeated in the outer repeats, two regions flanking the centromeres (Volpe et al., 2003; reviewed in Grewal and Jia, 2007).

RNA also plays a fundamental role in mouse heterochromatin: cells treated with RNaseA lose heterochromatic marks (Maison et al., 2002). Heterochromatin is also a template for the synthesis of dsRNA, generated by the bidirectional transcription of the major satellites (Martens et al., 2005). The synthesis of heterochromatic transcripts in mouse (Gaubatz et al., 1990; Rudert et al., 1995) is coupled to the cell cycle (Lu et al., 2007). This is also the case in S. pombe (Chen et al., 2008; Kloc et al., 2008; reviewed in Kloc and Martienssen, 2008). In mouse it was shown by in situ hybridization that the transcripts surround the pericentric heterochromatin. It is not known which RNAPol is responsible for this transcription and which transcription factors regulate this process.

Particularly in mammals, the findings highlight how little we know about the relationship between transcription of the repeats and repressive histone lysine methylation and about the regulation of this non-coding transcription. It is intriguing to think that the transcription factors regulating this phenomenon could also be structural components of heterochromatin, like HP1 (reviewed in Maison et Almouzni, 2004). In order to identify transcription factors regulating pericentric non-coding transcription, we used two different approaches: we first took into consideration transcription factors that have been shown to interact with HP1. Then we made a DNA sequence interrogation, for finding transcription factors with a heterochromatin-embedded binding site. In this way we identified transcription factor Pax3 as a potential new component of mouse pericentric heterochromatin.

The murine Pax family has nine members (Robson et al., 2006; Lang et al., 2007) that are very important developmental regulator (Nutt et al., 2001; Busslinger, 2004). Among these Pax3 and Pax7 encode for the two most complete proteins: both of them contain a paired domain, a homeodomain and the octapeptide. The other 7 members of the family lack either the octapeptide or the homeodomain. The homeodomain is 60 aminoacids long and consists of a

helix-loop-helix-turn-helix. The homeodomain is used to interact with other proteins and for binding DNA, which is mainly recognized by the third helix. This domain usually recognizes palindromic sequences and can then form a dimer with another homeodomain, reinforcing in this way the binding to the DNA (Lang et al., 2007).

We demonstrate that Pax3 binds the major satellite by recognizing a palindromic sequence with its homeodomain and represses the RNAPol II-dependent transcription of these repeats. In Pax3 null cells, transcription of the repeats is enhanced and H3K9me3 and H4K20me3 are impaired. Together, these data demonstrate that repression of non-coding transcription is crucial for the integrity of pericentric heterochromatin and that important developmental regulators are structural components of heterochromatin.

2.3 Results

2.3.1 Mouse Pax3 localizes at pericentric heterochromatin in ES and fibroblast cells

Among the many proteins that have been reported to interact with HP1 (reviewed in Kwon and Workman, 2008), only a few classify as bona fide transcription factors. Pax3 is one such example (Hsieh et al., 2006). Pax3 has a sequence motif binding site embedded in major satellite (see below) and in addition it binds KAP1, a transcriptional repressor also interacting with HP1. The interaction between KAP1 and HP1 is fundamental for removing RNAPol II and increasing H3K9me3 and H4K20me3 at genomic sequences that need to be silenced (Sripathy et al., 2006).

We therefore overexpressed Pax3 protein tagged with EGFP in iMEF and examined its subnuclear localization. We also analyzed other Pax members, including Pax5, Pax6 and Pax7. Only Pax3 and Pax7 localize at pericentric heterochromatin among the Pax protein family (Fig. 1A).

Since the pericentric localization could be a consequence of overexpression, the endogenous expression of Pax3 was visualized by immunofluorescence and was compared with the expression of Pax5 (no specific antibody is available for Pax7): Pax3 is expressed both in ES cells and fibroblasts and localize at pericentric heterochromatin. Pax5 is not expressed in these cells (Fig. 1B). The transcription of the Pax gene in the different cell types was analyzed through RT-PCR on cDNA from the different cell types: while Pax7 expression is limited to muscle cells (data not shown), Pax3 is expressed both in fibroblasts and ES cells (Fig. 1C). As a control, we checked the expression in B-cells where no expression of Pax3 was observed, in agreement with the published data: in B-cells the only Pax protein to be expressed is Pax5 (Holmes et al., 2008). Western blotting confirmed what we observed by immunofluorescence and RT-PCR (Fig. 1D).

For all the few proteins that so far have been observed to localize at pericentric heterochromatin, the localization is Suv39h-dependent, meaning that in Suv39h-dn cells the pericentric accumulation is lost and the proteins are redistributed in the nucleus (Lachner et al., 2001; Schotta et al., 2004). We checked whether this

was also the case for Pax3 by overexpressing it in different mutant iMEF (Suv39h-dn, Suv420h-dn and Rb-TKO) where, as previous studies have shown, pericentric heterochromatin is severely impaired (Peters et al., 2001; Schotta et al., 2004; Schotta et al., 2008; Gonzalo et al., 2005). None of the mutant cells showed altered Pax3 pericentric localization (Fig. 1E), suggesting that Pax3 localizes at pericentric heterochromatin by a mechanism that is independent of the activity of Suv39h enzymes.

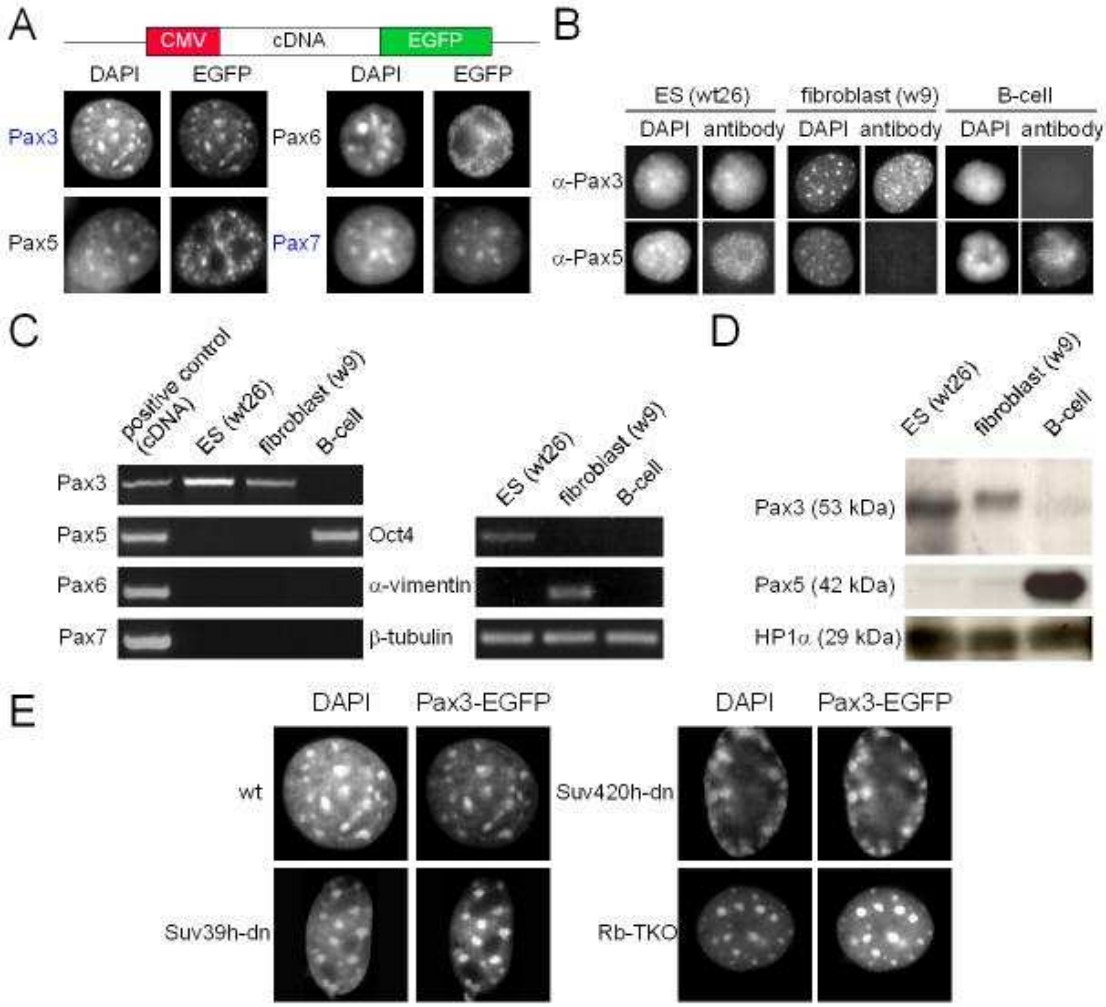


Figure 1 Endogenous Pax3 localizes at pericentric heterochromatin. (A) Overexpression of some Pax proteins, including Pax3, with an EGFP-tag in iMEF. Pax3 and Pax7 accumulate at pericentric heterochromatin. (B) Staining of the endogenous Pax3 in ES cells, iMEF and of Pax5 in B cells. Endogenous Pax3 localizes at pericentric heterochromatin. (C) RT-PCR for different Pax proteins in several cell types. Pax3 is the only Pax protein to be expressed in ES cells and iMEF (D) Western blot for Pax3 and Pax5 in different cell types. (E) Overexpression of Pax3-EGFP in different mutant cell lines. Pax3 always accumulates at pericentric heterochromatin.

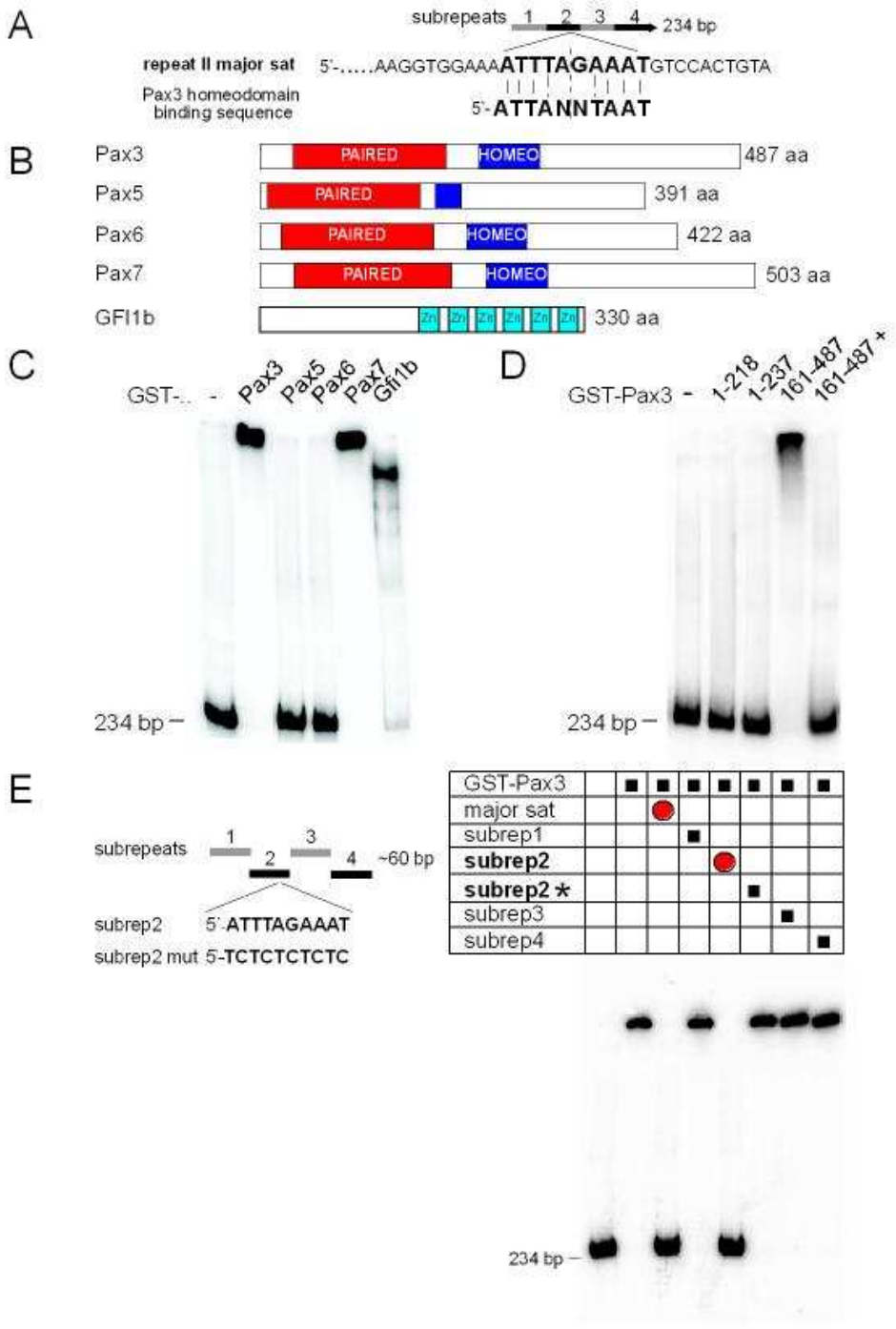
2.3.2 Pax3 binds the second subrepeat of the mouse major satellite through its homeodomain

Pax3 is a protein with two DNA-binding domains. In the second subrepeat of the major satellite there is a palindromic sequence (Fig. 2A) that is very similar to the consensus sequence recognized by the Pax3 homeodomain (Chalepakis et al., 1994). Therefore we tested direct binding by band shift assay.

Recombinant GST-tagged proteins were expressed in bacteria and then purified (Fig. 2B). In addition to the Pax proteins that we tested, we also engineered recombinant GST-tagged Gfi1b, a transcription factor that has recently been shown to bind the major satellite (Vassen et al., 2006). Pax3 and Pax7 reduced the mobility of the major satellite, while Pax5 and Pax6 did not (Fig. 2C).

In order to assess whether the homeodomain was the responsible for the binding to the major satellite, GST-tagged-truncated versions of Pax3, containing either the homeodomain or the paired domain, were tested. Upon incubation with each of the several truncated proteins, major satellite mobility was modified only when the protein contained an intact homeodomain. The mutation of a single amino acid in the recognition helix of the homeodomain (S273A) abolished the binding (Fig. 2D).

We then tried to reveal whether the palindromic sequence in the second subrepeat was the binding site of Pax3 by competition assay. Each of the four different subrepeats of the major satellite was incubated as a cold competitor during the binding reaction between Pax3 and labelled major satellite. Only the second subrepeat, bearing the palindromic candidate Pax3-binding site, successfully competed out the binding of Pax3 to the major satellite. Moreover, when a second subrepeat with a mutation in the palindromic site (ATTTAGAAAT→TCTCTCTCTC) was used as the cold competitor, the competition was completely abrogated, indicating that the palindromic sequence located in the second subrepeat is the specific sequence that Pax3 recognizes for binding to the major satellite (Fig. 2E).



<u>Figure 2 Pax3 homeodomain binds the major satellite.</u> (A) Alignment between the major satellite sequence and the sequence recognized by Pax3 homeodomain. (B) Tagged proteins used in the band shift assays. (C) Full length proteins used in band shift assay with radioactively labelled major satellite. Pax3, Pax7 and Gfi1b bind the major satellite. (D) Different Pax3 truncations used in the band shift assay. Pax3 shifts the major satellite only when it has an intact homeodomain. (E) Competition assay using the different subrepeats as cold competitors for the binding with labelled major satellite. Pax3 binds the palindromic sequence on the second subrepeat.

2.3.3 In Pax3 null cells H3K9me3 and H4K20me3 are lost at pericentric heterochromatin

For understanding the in vivo role of Pax3, we generated Sp2H fibroblast (that we denote as 'Pax3 null'). Sp2H is a radiation-induced mutation that deletes 32 bp in the open reading frame of Pax3 (Epstein et al., 2001; Chan et al., 2004). The frameshift forms a premature stop codon a few nucleotides downstream of the site of the deletion (Fig. 3A), and the resulting truncated protein consists of 237 amino acids rather than 487 (Fig. 3B). Homozygous embryos die at E14.5. Fibroblasts have been generated at E13.5 from wild type and from Pax3 null female embryos. Immunofluorescence with antibodies recognizing H3K9me3 and H4K20me3 revealed that these marks were reduced at pericentric heterochromatin in the Pax3 null cells (Fig. 3C). Immunofluorescence was also performed for H3K27me3, a mark that is distributed all over the chromatin but especially enriched at the inactive X chromosome in female cells. No difference for Xi was observed between wild type and Pax3 null cells (Supplemental Figure S1).

We also observed reduction of H3K9me3 and H4K20me3 at pericentric heterochromatin in Sp2G ES cells (Mansouri et al., 2001). In these ES cells a homologous recombination introduced the lacZ gene (with a stop codon) upstream of the first exon of Pax3, thus disrupting its expression.

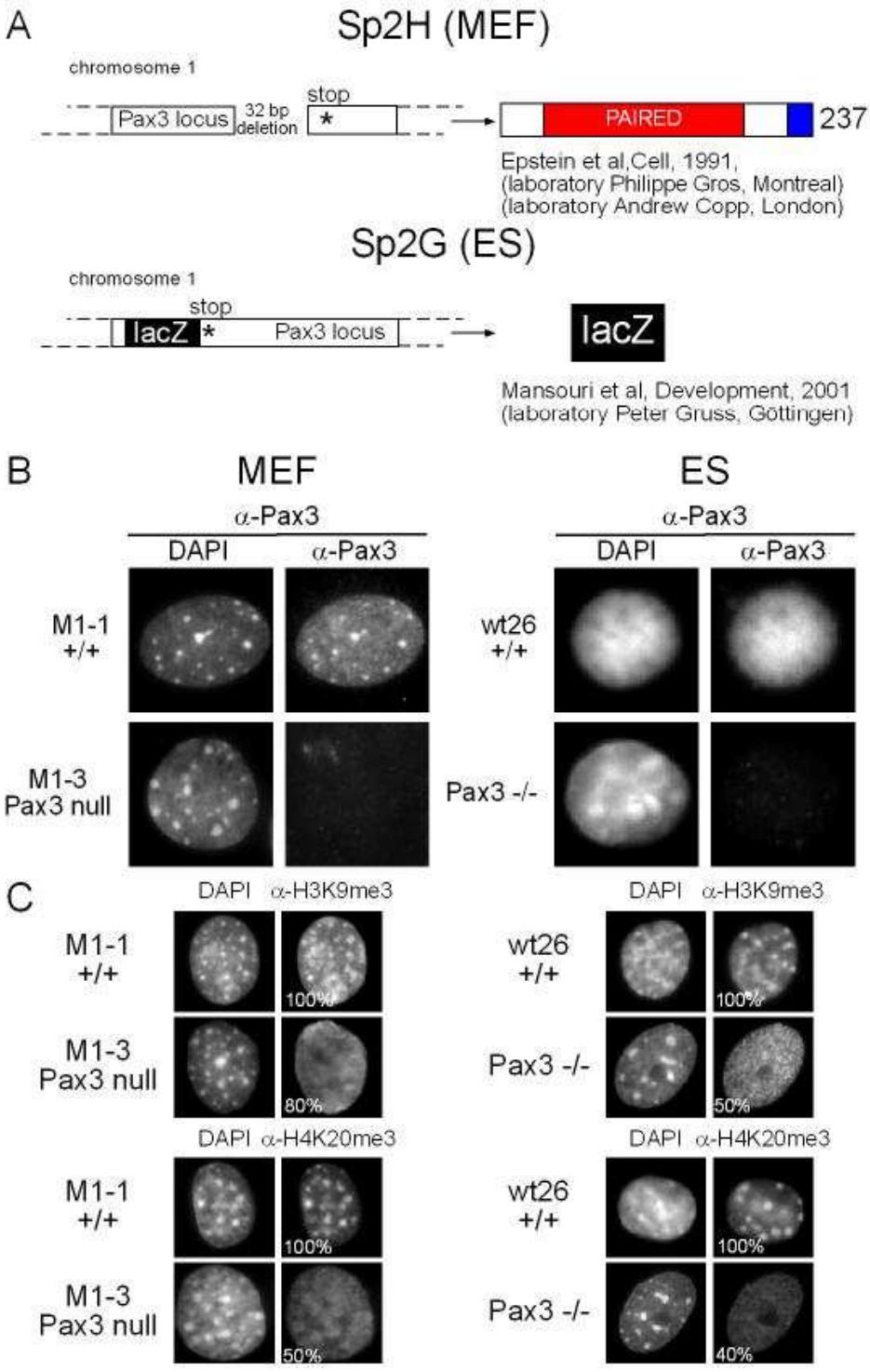


Figure 3 Pericentric heterochromatin is impaired in Pax3 null cells. (A) Schematic representation of the two Pax3 mutant cells used. Sp2H are considered to be full knock-out cells, due to the strong phenotype of the homozygote embryo. (B) Immunofluorescence for Pax3 in Pax3 null cells. The antibody recognizes an epitope in the C-terminal part of the protein; therefore no staining is present in Pax3 null fibroblast. (C) Immunofluorescence for H3K9me3 and H4K20me3 in Pax3 null cells. The two pericentric marks are severely reduced in absence of Pax3.

2.3.4 Major satellite histones are not decorated with H3K9me3 and H4K20me3 in the absence of Pax3

We subsequently confirmed the reduction of histone lysine methylation at pericentric heterochromatin, by chromatin immunoprecipitation: primers specific for the major satellite (Lenhertz et al., 2003; Martens et al., 2005) were used (Fig. 4A). As observed by immunofluorescence, in Pax3 null cells H3K9me3 and H4K20me3 are reduced severely at pericentric heterochromatin (Fig. 4B).

One possible explanation for the reduction of heterochromatic methyl marks could be the depletion of the histones, rather than a loss of specific methylation marks. In order to test this possibility, we immunoprecipitated chromatin from wild type and Pax3 null cells with antibody recognizing H3 histone. The antibody recognizing H3 histone is very specific. More H3 histone is present at pericentric heterochromatin in Pax3 null cells (Fig. 4C): this datum supports the idea that the decrease of the repressive marks at pericentric heterochromatin is not due to depletion of the histones, but it could be related either to delocalization of the Suv39h enzymes or to impairment of the HMTase activity.

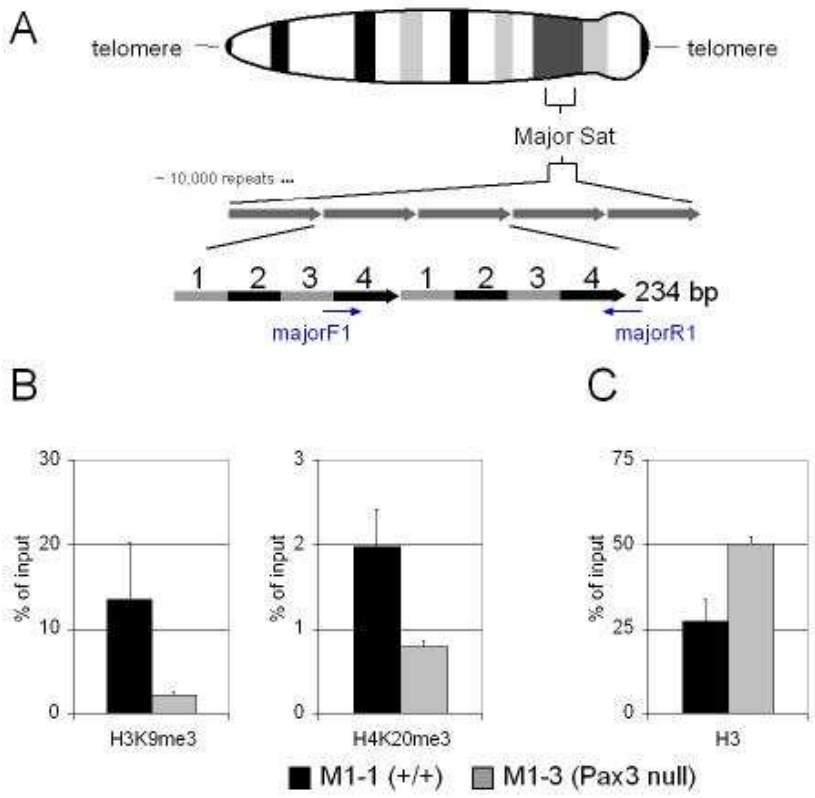


Figure 4 In Pax3 null cells major satellites are not decorated with H3K9me3 and H4K20me3. (A) Scheme of the primers used for major satellite amplification after chromatin immunoprecipitation. (B) Chromatin immunoprecipitation in wild type and Pax3 null fibroblast using α-H3K9me3 and α-H4K20me3. Both marks are reduced in the absence of Pax3. (C) Chromatin immunoprecipitation with α-H3. Histone H3 is present at major satellite in absence of Pax3: the decrease of H3K9me3 is not due to absence of histones.

2.3.5 Major satellite transcription is increased in Pax3 null cells

Pax3 is a transcription factor; therefore it could in theory be an activator or a repressor of transcription at pericentric heterochromatin. In order to elucidate its function we made use of total RNA isolated from wild type and Pax3 null cells and of major satellite specific primers and RNA probes (Fig. 5A).

When total RNA isolated from wild type, Pax3 null and Suv39h-dn cells was converted into cDNA and then amplified using major satellite-specific primers (Fig. 5A), we observed that in Pax3 null cells the signal coming from the major satellite was much higher than the signal in wild type and Suv39h-dn cells (Fig. 5B).

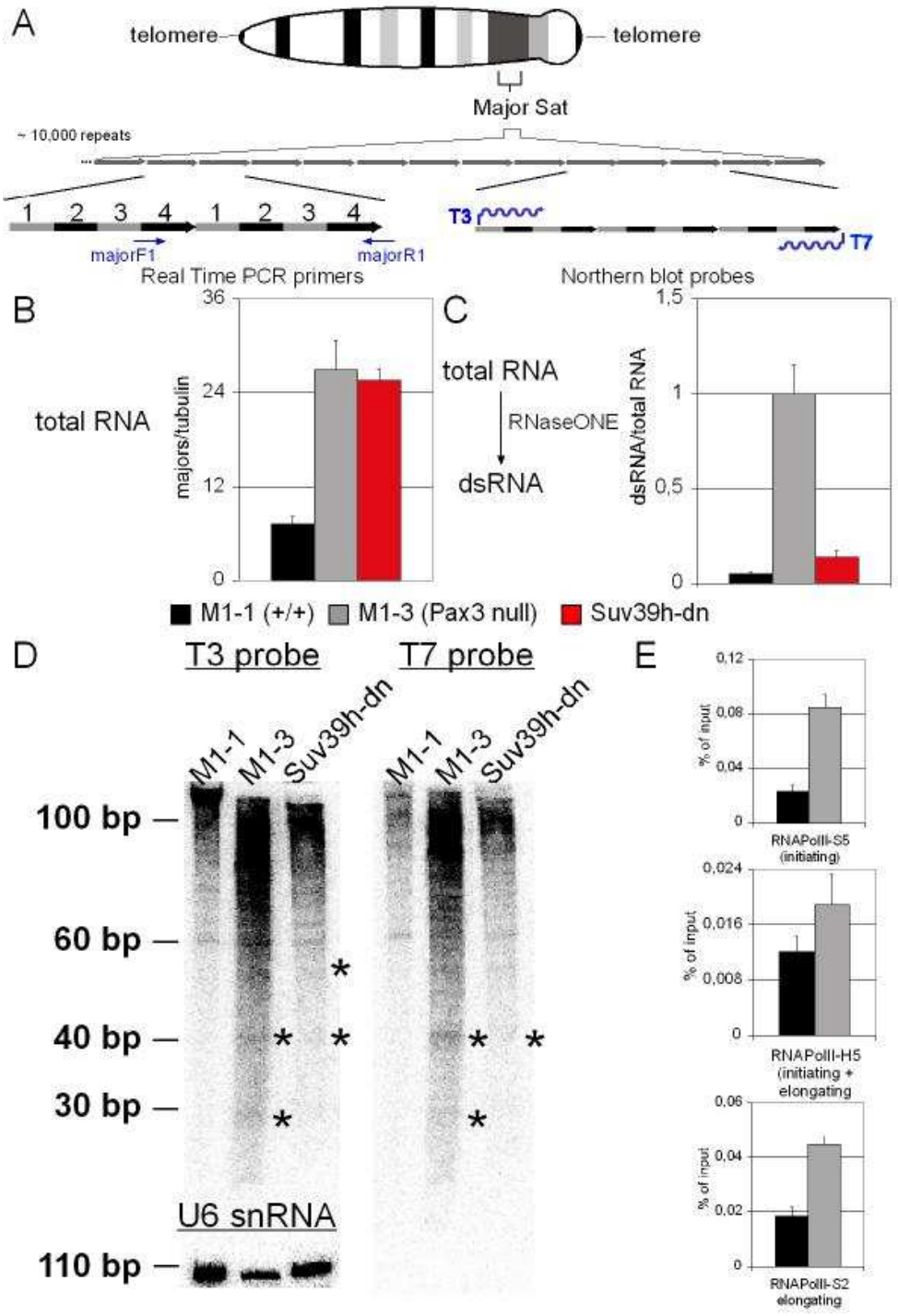
Major satellites can be distinguished from interdispersed repeats by the abundance of double-stranded RNA and we tried to understand whether in Pax3

null cells dsRNA was also altered; we analyzed dsRNA by digesting total RNA with RNaseONE before reverse transcription (Martens et al., 2005): RNaseONE preferentially cleaves ssRNA, leaving a pool of dsRNA. The amount of dsRNA generated in Pax3 mutant cells is much higher than in wild type and in Suv39h-dn cells (Fig. 5C).

For directly detecting RNA, we performed a Northern blot using strand specific probes that were transcribed from pCR4-Maj1-3 by using either T7 or T3 RNA polymerase (Fig. 5A). Both forward and reverse transcripts are much more abundant in Pax3 null cells. Furthermore, very short transcripts, whose length is 30-40 nt, not present in the wild type cells, can be detected in Pax3 null cells and also, to lesser degree, in Suv39h-dn cells (Fig. 5D).

It has recently been shown in A. thaliana that RNAPol IV and RNAPol V are responsible for the transcription of repeats (Wierzbicki et al., 2008). These enzymes have no mouse orthologues, therefore the transcription of repeats in mouse must be carried out by other RNAPol enzymes. We immunoprecipitated chromatin from wild type and Pax3 null cells with three different antibodies recognizing RNAPol II. We observed that this enzyme is associated with the major satellite and a relatively greater enrichment can be found in Pax3 null cells (Fig. 5E).

Together, these data indicate that Pax3 represses transcription of the major satellites and that, in the absence of Pax3, RNAPol II is not prevented from transcribing the major satellites.



<u>Figure 5 Pax3 represses transcription of the major satellites.</u> (A) Primers used for Real Time PCR and strand specific probes for the Northern Blot. (B, C) Real Time PCR for the major satellite using total RNA (B) or dsRNA (C) as starting material. In the absence of Pax3 there is an increased transcription of the major satellites. (D) Northern Blot with strand-specific probes. When Pax3 is impaired, more and shorter transcripts from the major satellite are generated. (E) Chromatin immunoprecipitation in wild type and Pax3 null cells using three different RNAPol II antibodies. More RNAPol II is found on the major satellite when Pax3 is absent.

2.3.6 Characterization of the transcription start site of the major satellite

Major satellites are organized in a complex way, although we always represent them as an array of repeats with the same orientation. When genomic DNA is amplified using major satellite specific primers, amplification can be observed also when only one primer is used, which shows that repeats in the genome can have two different orientations (Fig. 6A). Surprisingly, the amount of DNA amplified is more when PCR is performed only with a forward primer than when it is done by using only a reverse one, although one would expect the same amount.

It is not known where the promoter for transcribing these repeats is but the transcription start site can be mapped and in order to do this we made use of S1 nuclease digestion and 5' RACE.

S1 nuclease digestion was performed on wild type and Pax3 null cells by an RNA preparation from nuclear extracts and, for every direction we made use of 4 different probes, whose 5´-ends where 60 bp apart (Fig. 6B). Signal from undigested probe could be detected with some of the probes. The signal was stronger in Pax3 null cells than in wild type, even though the same amount of nuclear RNA was used for the hybridization (Fig. 6C). The strong signal obtained with the F1 probe suggests that reverse transcription starts in the second subrepeat close to the binding site of Pax3 (Fig. 6C): the undigested probe is roughly 80 nt long and is labelled at its 5' end. This probe anneals with reverse transcripts and the undigested part was covered from the transcript, meaning that the transcript and the probe overlapped in these 80 bp and that the 5' end of the transcript was 80 bp apart from the 5' end of the probe, which is very close to the Pax3 binding site (Fig. 6D). The S1 nuclease experiment also suggested that forward transcription could start at subrepeat 1, but the result was not as clear as for the reverse transcription. Therefore we performed 5' RACE on an RNA preparation from nuclear extracts and used the maj3R primer for the reverse transcription. PCR products were then cloned and sequenced and we observed that in both wild type and Pax3 null cells transcription starts at nucleotide 20 of the major satellite repeats (Fig. 6E). The amount of positive colonies obtained from Pax3 null cells was 3-fold the amount obtained from wild type cells. More

detailed information will in future be obtained by deep sequencing of the small RNAs.

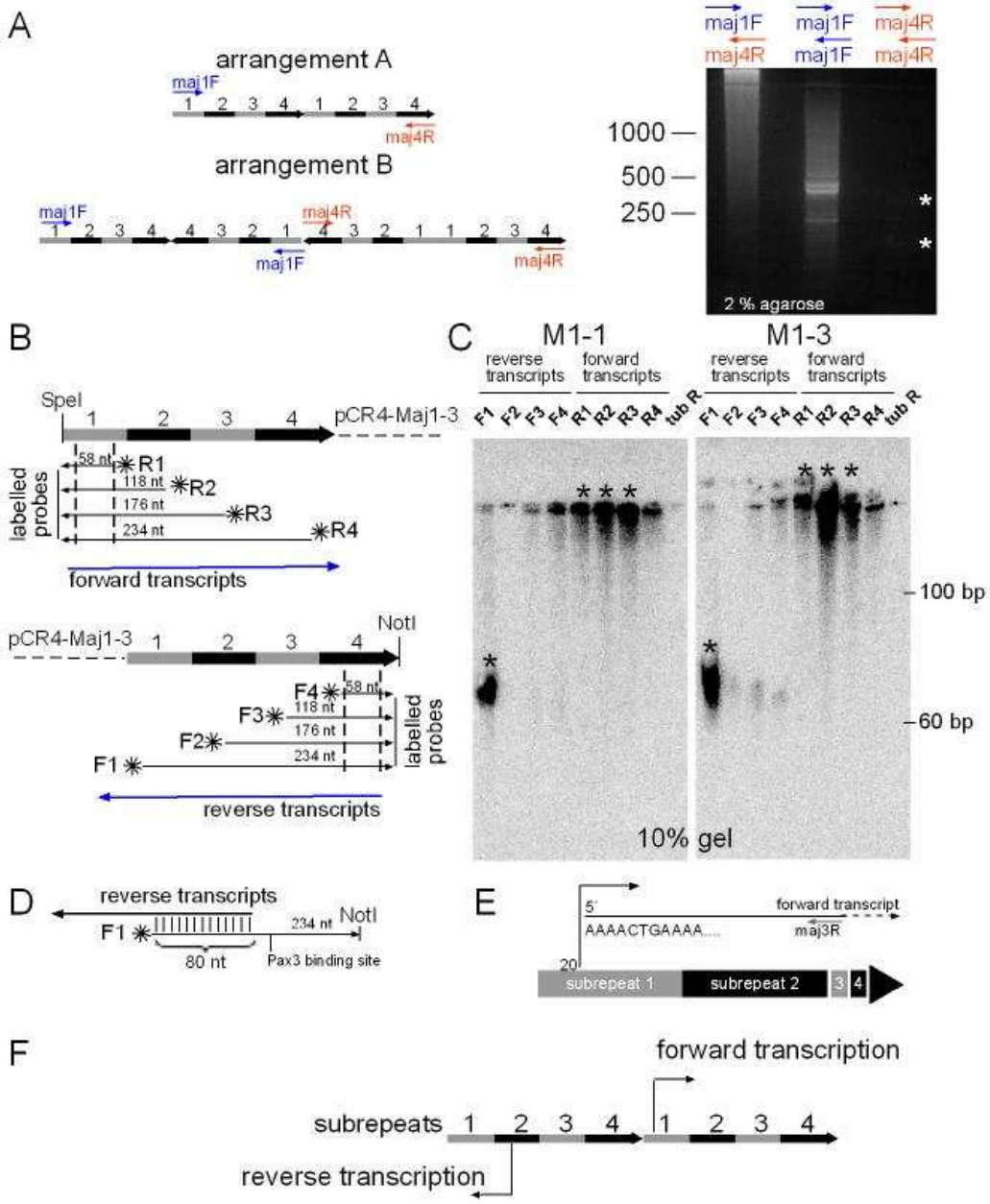


Figure 6 Transcription start site at major satellite. (A) PCR performed on genomic DNA with major satellite primers. DNA can be amplified also with one primer, suggesting that major satellites can have two opposite orientations in the genome. (B) Schematic representation of the probes used for S1 nuclease assay. (C) Result of S1 nuclease digestion in wild type and Pax3 null cells. In Pax3 null cells more signal from the major satellite is detected. (D)The strong and clear signal given by the probe F1 allows to conclude that the TSS of the reverse transcription is 80 nt apart from the 5' end of this probe, so in the second subrepeat. (E) 5' RACE for forward transcripts: the TSS is located in subrepeat 1 and is the same in wild type and Pax3 null cells. (F) Schematic representation of the TSS at major satellite.

2.3.7 Pax3 repression is cell cycle regulated

Some of the proteins with which Pax3 interacts, like Daxx, localize at pericentric heterochromatin only during a certain phase of the cell cycle (Hollenbach et al., 1999). We tried to understand whether this was also the case for Pax3 and performed immunofluorescence on synchronized iMEF. Endogenous Pax3 is never excluded from the pericentric heterochromatin, as happens to the vast majority of transcription factors, but it accumulates there especially during G2 (Fig. 7A). G2 is the cell phase when transcription of the major satellite is blocked (Lu et al., 2007; Lu et al., 2008) and Pax3 could be one of the transcription factors responsible for coupling major satellite transcription and cell cycle. In order to better understand the role of Pax3 in relation to the cell cycle we performed RNA-FISH by using a mixture of four different major satellite specific probes. We observed that the major satellite transcripts localize at pericentric heterochromatin both in wild type and Pax3 null cells. In Pax3 null cells the amount of these transcripts increases in every phase of the cell cycle and it becomes possible to detect them even in mitosis, although to a much lower extent in comparison to interphase and G1/S (Fig. 7B).

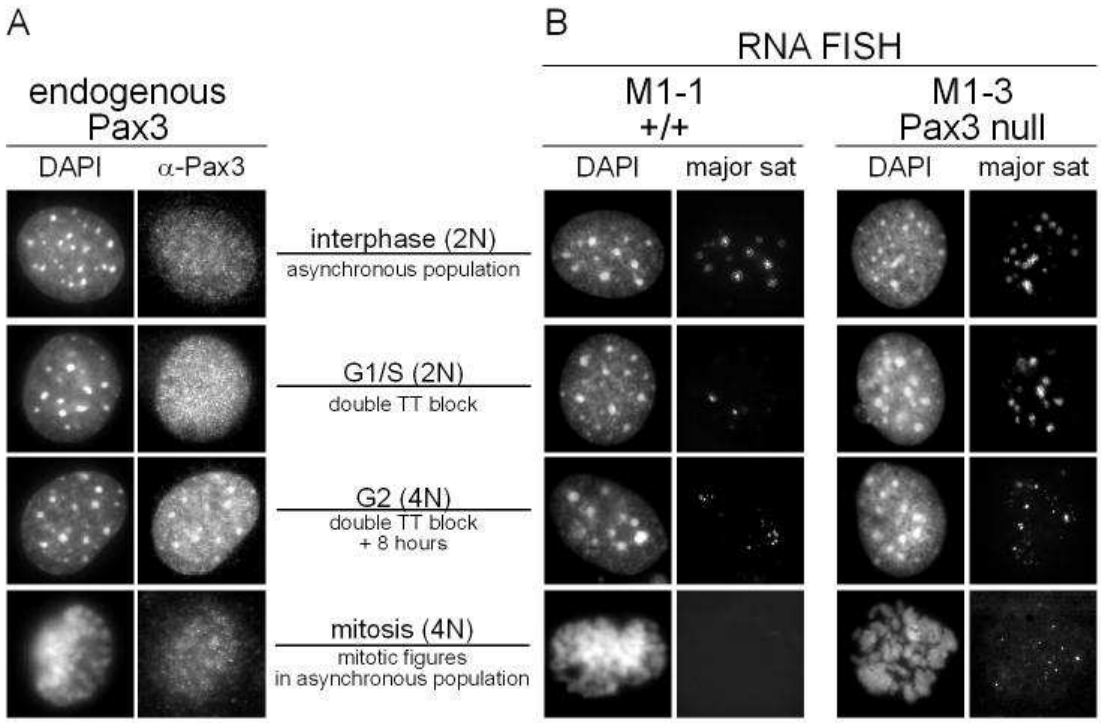


Figure 7 Pax3 accumulates at pericentric heterochromatin during G2 and represses transcription of major satellite in a cell cycle dependent manner. (A) Endogenous Pax3 localization at different cell phases in synchronized wild type fibroblast. (B) RNA FISH on synchronized wild type and Pax3 null fibroblast using a mixture of four different probes (see material and methods) specific for the major satellite. In Pax3 null cells the number of signals for every cell phase is higher than in wild type.

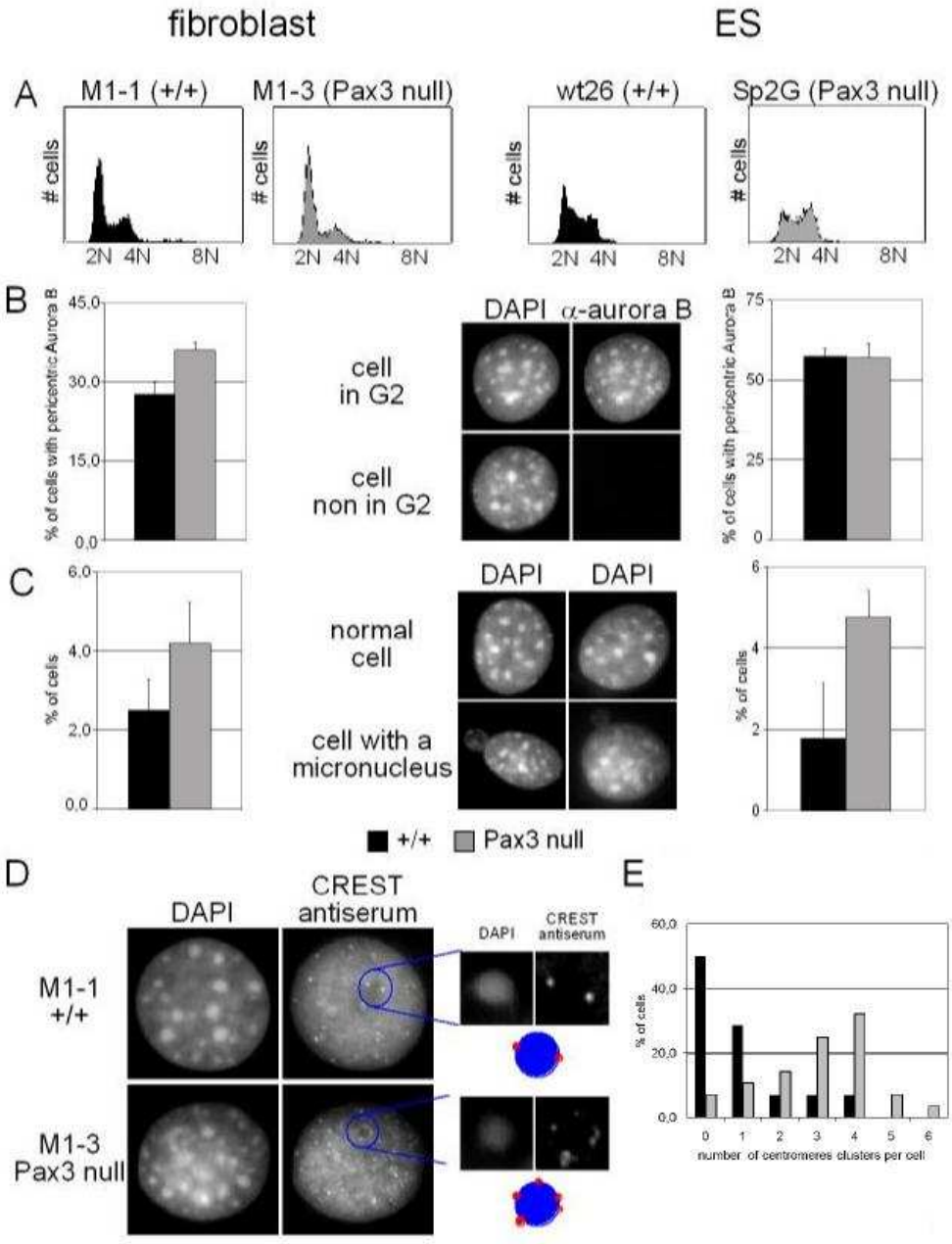
2.3.8 Pax3 affects chromosome stability

Since Pax3 localization is coupled to the cell cycle and it remains on the chromosome during mitosis, we wanted to check whether Pax3 could be involved in chromosome stability.

We first measured the amount of DNA in wild type and Pax3 null fibroblast and ES cells. In ES cells we observed a clear difference: in the absence of Pax3 the amount of DNA in the cells increases and more cells contain 4N (Fig. 8A). This fact cannot be linked to a major number of cells in G2 phase: by staining cells with an antibody recognizing Aurora B and counting the percentage of cells where Aurora B was pericentric, we observed no difference in the percentage of G2 cells between wild type and Pax3 null cells (Fig. 8B). In fibroblasts the amount of cells in G2 in Pax3 null cells was slightly increased when compared to wild type, in ES cells the percentage was exactly the same.

We also analyzed the formation of micronuclei: the formation of micronuclei is a consequence of chromosome instability (reviewed in Iarmarcovai et al., 2007). Micronuclei form because after mitosis lagging chromosomes are excluded from the reforming nucleus and they are therefore incorporated into micronuclei separated from the main genome. In S. pombe it has been reported that heterochromatin is fundamental for cohesion at centromeres and therefore for chromosome stability (Bernard et al., 2001). We observed that in Pax3 null cells the amount of cells with a micronucleus is higher than in wild type cells (Fig. 8C), suggesting chromosome instability in absence of Pax3.

To test whether there is a difference in the distribution and number of centromeres in the absence of Pax3, we performed a staining for centromeres by using CREST antiserum. CREST is an autoimmune disease that leads to the development of antibodies recognizing the centromeres (Van Hooser et al., 1999). We observed a difference in the distribution of the centromeres between wild type and Pax3 null cells: in wild type cells two centromeres usually surround a pericentric region and in general centromeres are equally distributed throughout the nucleus. In Pax3 null cells, centromere distribution is not uniform: centromeres aggregate, and more than two centromeres can be found around a single pericentric heterochromatin region (Fig. 8D). We also counted the numbers of aberrant centromere clusters at DAPI-dense foci, where aberrant is more than 2 CREST signals per DAPI focus. We observed that in wild type cells half of the cells have always two centromeres around one DAPI focus, while in Pax3 null cells the vast majority of the population has three or four clusters of centromeres per cell (Fig. 8E). In addition, we counted the number of centromeres in cells: we observed both in wild type and Pax3 null cells a Gaussian distribution but in Pax3 null cells the number of centromeres is on average higher than in wild type: wild type cells have on average 35 centromeres while Pax3 null cells have at least 40 centromeres (Supplementary Figure S2). All these data show that Pax3 is important for chromosome stability.



<u>Figure 8 Pax3 affects chromosome stability.</u> (A) DNA content was measured with propidium iodide in 2x10^6 wild type and Pax3 null fibroblast and ES cells. In ES cells there is clear difference in the amount of DNA. (B) Aurora B staining in wild type and Pax3 null fibroblast and ES cells. 1000 cells were counted for each cell type. Pax3 does not influence the cell cycle progression. (C) Micronuclei formation was evaluated in wild type and Pax3 null fibroblasts and ES cells. 1000 cells were counted for each cell type. More micronuclei form in absence of Pax3. (D) CREST antiserum was used for staining centromeres in wild type and Pax3 null fibroblasts. In Pax3 null cells, more than two centromeres surround a single region of pericentric heterochromatin. (E) In Pax3 null fibroblasts half of the population has three or four centromeres clusters. The number of clusters was counted in at least 100 cells.

2.4 Discussion

The overlap between bona fide transcription factors interacting with HP1 and DNA binding proteins with a binding site embedded into major satellite repeats allowed us to identify Pax3 as a novel factor that regulates mammalian pericentric heterochromatin.

In both mouse and humans, Pax3 has been well characterized as a developmental regulator: in mouse the homozygous mutation is lethal and in humans PAX3 mutations are responsible for Waardenburg syndrome I and III (Ishikiriyama et al., 1989), an autosomal dominant disease. Several independent point mutations can generate Waardenburg syndrome and very recent studies reported that each of them modifies the subnuclear localization of Pax3 in a different way (Corry et al., 2008). Extensive studies regarding the alteration of the DNA-binding properties of the Pax3 protein synthesized in mutant mice and in Waardenburg syndrome patients have been carried out (Chalepakis et al., 1994): in all the cases Pax3 DNA-binding properties are altered. Therefore some of the mutants, including Sp2H, are considered loss of function mutations.

In addition to its developmental function, Pax3 has also been shown to interact with proteins localizing at pericentric heterochromatin: Pax3 co-immunoprecipitates with HP1α, HP1γ and KAP1. These three proteins bind the Pax3 paired domain in a competitive manner, supporting the idea that Pax3 mainly interacts with corepressor (Hsieh et al., 2006).

Pax3 also interacts with the Pocket proteins—Rb, p107, p130—by using helices I and II of the homeodomain (Wiggan et al., 1998). The Pocket proteins are very important cell cycle regulators and are known to influence heterochromatinization: in cells where all of them are knocked out, pericentric heterochromatin is altered and H4K20me3 is reduced (Gonzalo et al., 2005). The homeodomain is used also for interacting with HIRA (Magnaghi et al., 1998), a histone chaperone involved in chromatin assembly, and with Daxx (Hollenbach et al., 1999). Daxx is involved in anti-apoptotic pathways and it accumulates at pericentric heterochromatin at the middle-late S-phase, in coincidence with the replication of the major satellite repeats. This raises the possibility that Daxx could be used as a chaperone during the DNA replication of the pericentric

heterochromatin. We also observed that Pax3 localization changes during the cell cycle and its accumulation at pericentric heterochromatin takes place when the cell is in G2 phase.

Transcription factors can have a heterochromatin-embedded binding site. Here we demonstrated that Pax3 is one of them. In mouse it has previously been reported that Gfi1b, a B-cell specific transcription factor bearing six Zn-fingers, localizes at pericentric heterochromatin. Gfi1b has a binding site in the fourth subrepeat of the major satellite and in Gfi1b-null cells alteration of H3K9me has been observed, but no connection has been made between the absence of a transcription factor binding the major satellite and the impairment of H3K9me3 at pericentric heterochromatin (Vassen et al., 2006). α-satellite in humans also contains binding sites for transcription factor: the protein having the highest probability of binding there is ZEB1, a protein that has 7 Zn-fingers and a homeodomain. Another example of a transcription factor with a repeat-embedded binding site is Su(var)3-7, one of the strongest Su(var) in D. melanogaster (Cleard et al., 1997). This protein has 7 Zn-fingers and it has been shown to bind different DNA sequences embedded in repeats (Cleard and Spierer, 2001) and to form a strong ternary complex with Su(var)3-9 and HP1/Su(var)2-5 (Delattre et al., 2000; Delattre et al., 2004).

In mouse 10% of the genes encode for transcription factors, which are thus the most abundant class of proteins (Carninci et al., 2005). A lot of transcription factors bind DNA by using Zn-fingers. So far, more than five hundred different Zn-finger proteins have been described and many of them, characterized as repressors, require KAP1 in order to function properly. It is likely that some of the transcription factors, in addition to Pax3 and Gfi1b, has a heterochromatin-embedded binding site and, in addition to a developmental role, are important for the regulation of the major satellite transcription.

The data we presented in this manuscript showed that the inhibition of major satellite transcription and of dsRNA formation is crucial for the establishment of heterochromatic marks: in Pax3 null cells the higher amount of transcripts induces loss of pericentric H3K9me3 and H4K20me3. Previous publications reported the opposite situation (Maison et al., 2002): cells treated with RNaseA,

an endonuclease that cleaves single-stranded RNA, lose every known heterochromatic mark. We are now in a paradox: does RNA facilitate or inhibit heterochromatin formation?

Transcription of repeats has so far been better characterized in S. pombe. It has been shown that the initial transcription of the repeats is required for their subsequent silencing (reviewed in Cam et al., 2009): dh and dg are transcribed from RNAPol II and the RISC complex is recruited. One of the proteins within the RISC complex, Chp1, in its turn recruits Swi6 and Clr4 (Morey et al., 2004; Grewal et al., 2007): the initial transcription of the repeats allows the establishment of pericentric methyl marks. Transcription and processing of the repeats are fundamental for heterochromatinization: in strains where argonaute, dicer, and RNA-dependent RNA polymerase genes are deleted the transcription of transgenes integrated at the centromeres is derepressed (Volpe et al., 2002). Once heterochromatin is formed the transcription of the repeats decreases significantly and there will be just a particular window in the cell cycle when this phenomenon will still be possible: the vast majority of transcripts generated from dh and dg repeats will appear during S phase, when H3S10 is still phosphorylated and Swi6 is not bound to the chromosomes (Kloc et al., 2008).

What has been described in S. pombe could also be the case for mouse heterochromatin but, contrary to S. pombe, in mouse the transcription does not determine its own repression by simply recruiting the RNAi machinery and Suv39h enzymes: transcription needs to be properly repressed by transcription factors in order to establish heterochromatin. Transcription factor localization at pericentric heterochromatin allows the removal of RNAPol II and the subsequent establishment of heterochromatic marks. In this situation RNA works both as an activator and as an inhibitor of the heterochromatinization process. RNA is an activator because, as already observed, impairment of RNA, for example by digesting it, prevents the activation of heterochromatin formation (Maison et al., 2002), as happens in S. pombe. On the other hand, RNA can be an inhibitor: too much non-coding RNA impairs heterochromatinization. If RNAPol II is not blocked and impeded from transcribing the major satellite, the chromatin will

always be relaxed and appropriate formation of heterochromatin will not be possible.

In this apparently paradoxical situation, the threshold of RNA is an important parameter for heterochromatin formation and developmental regulators with a major-satellite-embedded binding site could be the key molecules for controlling the appropriate amount of transcripts.

2.5 Materials and methods

Pax3 mutant cells and tissue culture: Pax3 null fibroblast: Sp2H heterozygous mice were obtained from Andrew Copp's laboratory (Chan et al., 2004) and crossed to homozygosity. Embryos were isolated at E13,5 and primary fibroblasts derived from them. Fibroblasts were cultured at 37℃, 5% CO_2 in pMEF medium [403 mL High glucose Dulbecco's Modified Eagle Medium, 75 mL FCS, 5 mL Penicillin Streptomycin solution (Invitrogen) (100x), 5 mL L-glutamine (200mM), 5 mL MEM Non essential amino acids (Invitrogen) (100x), 5 mL MEM Sodium pyruvate (Invitrogen) (100mM), 1 mL β-Mercaptoethanol (50mM)]. Genomic DNA was used for genotyping with the primers Sp-1168: 5´-CCTCGGTAAGCTTCGCCCTCTG and Sp1042: 5´-CAGCGCAGGAGCAGAACCACCTTC. In wt fibroblasts the amplification product is 127 bp, in Pax3 null fibroblast is 95 bp (Epstein et al., 1991). Pax3 null ES cell: Pax3 null ES cell were obtained from Ahmed Mansouri (Mansouri et al., 2001) and culture on a feeder layer with the appropriate medium [403 mL High glucose Dulbecco's Modified Eagle Medium (DMEM), 75 mL FCS ES cell culture tested, 5 mL Penicillin Streptomycin solution (Invitrogen) (100x), 5 mL L-glutamine (200mM), 5mL MEM Non essential amino acids (Invitrogen) (100x), 5 mL MEM Sodium pyruvate (Invitrogen) (100mM), 1 mL β-Mercaptoethanol (50mM), 1.25 mL filter-sterilized supernatant of Leukemia inhibitory factor (LIF) producing cells]; Confluent feeder cells were irradiated two hours before culturing ES cells on the top of them. Wild type w9 fibroblast: w9 fibroblast were grown in iMEF medium [440 mL High glucose Dulbecco's Modified Eagle Medium (DMEM), 50 mL Fetal calf serum (FCS), 5 mL Penicillin Streptomycin solution (Invitrogen) (100x), 5 mL L-glutamine (200mM)]. Cell transfection: Cells plated on glass chamber slides were transfected with PLUS-Reagent and Lipofectamine (Invitrogen) following standard procedures and fixed 48 hours after transfection.

Immunofluorescence: Immunofluorescence analysis was performed on cells plated on glass chamber slides (Nunc, #154534), using the following antibodies: H3K9me3, H4K20me3 (Peters et al., 2003), Aurora B (Sigma Aldrich, #WH0009212M3), CREST antiserum (kindly gifted by Jan Michael Peters), Pax3 (The Developmental Studies Hybridoma Bank, Iowa University) and Pax5 (kindly

gifted by Meinrad Busslinger). Cells were first fixed with 2% paraformaldehyde, then permeabilized with 0,1% sodium citrate-0,1% Triton X-100 and later incubated overnight at 4℃ with the primary antibod y and 1 hour at 25℃ with the secondary antibody.

Western Blot: proteins were extracted from cultured cells using a cell lysis buffer (50 mM TRIS-HCl pH7,5, 500 mM NaCl, 1% Triton X-100) and the extract was then sonicated 3 times for 15 seconds. 20 µg of protein extract were used in every well of a western blot. Antibodies used for Western blot are the same reported for immunofluorescence.

Expression profiling by RT-PCR**:** Total RNA was extracted with TRIzol (Invitrogen), reverse transcribed with random hexamers and SuperScript II reverse transcriptase (Invitrogen). First-strand cDNA was used for PCR. Primer pairs for RT-PCR were**:**

gene	**forward primer**	**reverse primer**
Pax3	5'-TGCCCTCAGTGAGTTCTATCAGC	5'-GCTAAACCAGACCTGCACTCGGGC
Pax5	5'-TCAACAGGATCATTCGGACA	5'-TGGTCTGTTCTGGCTTGATG
Pax6	5'-TCTAATCGAAGGGCCAAATG	ATAACTCCGCCCATTCACTG
Pax7	5'-TGGAAGTGTCCACCCCTCTTGGC	5'-ATCCAGACGGTTCCCTTTGTCGCC
tubulin	5'-TCTTAGAACCTTCCTGCGGTC	5'-ATCACCTCCCAGAACTTAGCACC
oct4	5'-GGCGTTCGCTTTGGAAAGGTGTTC	5'-CTCGAACCACATCCTTCTCT
vimentin	5'-TGTTTGTTCCAGGGTTAGGC	5'-AATGCCCAGTGACTGAAAGG**.**

Recombinant protein expression and purification: BL21 bacteria were transformed with pGEX plasmids containing cDNA of the protein fused to GST. 10mL of an overnight 100 mL culture were then grown at 37℃ in 1L LB medium till OD_{600} was 0,6. Protein expression was induced with IPTG (at 1 mM final concentration) and bacteria were cultured for another 3 hours at 25℃. Bacterial lysis was performed with 1 mg/mL lysozyme in 40 mL of lysis buffer (40 mM Tris-HCl pH8, 500 mM KCl, 9% glycerol); the supernatant was sonicated 3 times for 15 seconds and incubated for 3 hours under fast rotation with Glutathione-Sepharose 4B beads (GE Healthcare, #17-0756-01) previously equilibrated with PBS. Beads were washed 6 times with 10 mL washing solution (40 mM Tris-HCl

pH8, 500 mM KCl, 9% glycerol). Protein elution was performed for 15 minutes on ice using glutathione elution buffer (20 mM Tris-HCl pH8, 500 mM KCl, 9% glycerol, 10 mM reduced glutathione).

Band shift assay: major satellite was amplified using pCR-Maj1-3 (Lehnertz et al., 2003) as a template with the following primers: MajF: 5´-GGACCTGGAATATGGCGAGAAAACTGAAAATAACG, MajR: 5´-TTCAGTGTGCATTTCTCATTTTTCACG. PCR was done with Taq polymerase and the extension time of the PCR was 15 seconds, in order to amplify only a single subrepeat. The amplification product was subsequently run on a 2% agarose gel, isolated (Qiagen, Gel extraction kit) and radioactively labeled with γ^{32}-ATP. Recombinant expressed proteins were purified and incubated with labeled major satellites. Every binding reaction was performed in a total volume of 20 µl that includes: 4 µl 5X binding buffer (20% glycerol, 5 mM $MgCl_2$, 2,5 mM EDTA, 250 mM NaCl, 50 mM TRIS-HCl pH 7,5), 0,2 µl 0,1 M DTT, 1 µl poly-dIdC (300 ng/µl), 20000 Cpm, 200 ng protein. Binding reactions were then loaded on a 5% polyacrylamide native gel and run for 5 hours, 80 V at 4℃. The gel was then dried on Whatmann paper and the paper exposed to Phosphor Imager overnight.

Chromatin immunoprecipitation: Cells were fixed with 1% formaldehyde and lysed with lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 0.1% sodium deoxycholate, with protease inhibitors). The cell lysate was sonicated with a Bioruptor machine for a total of 20 min with 30-second 'on' and 1-minute 'off' cycles, to shear the DNA to a final size of 200–500 base pairs. After preclearing with Protein A or G beads (Upstate), antibody was added and incubated overnight at 4 ℃. The next day , Protein A-agarose or Protein G-agarose beads were added and incubated for a further 1–2 h. The complex was washed twice with lysis buffer, once with high salt buffer (50 mM HEPES-KOH pH 7.5, 500 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 0.1% sodium deoxycholate), twice with LiCl buffer (10 mM Tris-HCl pH 8.0, 0.25 M LiCl, 0.5% Nonidet P40, 0.5% sodium deoxycholate, 1 mM EDTA) and once with TE buffer, followed by elution in TE buffer containing 1% SDS. The crosslinks were reversed, and the DNA was purified with a QIAquick 8 PCR purification kit (Qiagen) and subjected to analysis by quantitative real-time PCR.

Chromatin from wt and Pax3 null fibroblast was precipitated with H3K9me3, H4K20me3 (Peters et al. 2003), H3K9me1, H3K27me1, H3K27me3 and three different RNAPol II antibodies (RNAPol II-phospho S5, Abcam #5131, RNAPol II-phospho S2-H5 Abcam #24758 and RNAPol II-CTD repeat-phospho S2, Abcam #5095). Immunoprecipitated DNA was analyzed by quantitative PCR using primers specific for major satellites:

MajF1: 5´-GACGACTTGAAAAATGACGAAATC,

MajR1: 5´-CATATTCCAGGTCCTTCAGTGTGC (Lehnertz et al., 2003).

dsRNA analysis: 5 µg RNA in 15 µL water were digested with 2U RNaseONE (Promega, #M4261) for 30 seconds. The reaction was then stopped by adding 93 µL water and 6 µL 10% SDS. RNA was then extracted with RNeasy (Qiagen) and converted into cDNA that was amplified with primers specific for the major satellites (see Chromatin Immunoprecipitation).

Northern Blot: 10 µg RNA purified with MirVana (Ambion, #AM1560) were charged on a 15% polyacrylamide-urea gel (National Diagnostic) and run for 1h at 30 mA. Transfer was done on a nylon membrane (GeneScreen Plus HybridizationTransfer Membrane, Perkin Elmer). RNA probes were synthesized through in vitro transcription, using pCR4-Maj1-3 (Lehnertz et al., 2003) as a template and either T3 or T7 RNA polymerase. The probes were then digested for 40 minutes in alkaline buffer to generate 15 nt-long fragments. Prehybridization (2 hours) and hybridization (overnight) were performed at 50℃ with 5X SSC, 20mM Na_2HPO_4, 7% SDS, 1X Denhardt solution, 30% formamide and 1 mg sheared salmon sperm (Ambion, #AM9680). Washing was done first (10 minutes, 30 minutes) with 2X SSC, 1%SDS and then (2 times 10 minutes) with 0.5X SSC, 1% SDS. The membrane was then exposed to Phosphor Imager overnight.

S1 nuclease: S1 nuclease probes were synthesized using pCR4-Maj1-3 as a template. The plasmid was linearized with NotI for the synthesis of the forward probes and with SpeI for the synthesis of the reverse probes. Primers used for the synthesis of the probes were the following:

primer	sequence
major 1F	5’-GGACCTGGAATATGGCGAGAAAACTGAAAATCACGGAAAA
major 2F	5’-GGACGTGAAATATGGCGAGGAAAACTGAAAAAGGTGGAAA
major 3F	5’-GGACGTGGAATATGGCAAGAAAACTGAAAATCATGGAAAA
major 4F	5’-CGACTTGAAAAATGACGAAATCACTAAAAAACGTGAAA
major 1R	5’-TAAAGTGTGTATTTCTCATTTTCCGTGATTTTCAGTTT
major 2R	5’-TACAGTGGACATTTCTAAATTTTCCACCTTTTTCAGTTTT
major 3R	5’-TCAAGTGGATGTTTCTCATTTTCCATGATTTTCAGTTT
major 4R	5’-TTCAGTGTGCATTTCTCATTTTTCACGTTTTTTAGTGA

For the synthesis of ssDNA probes, Taq polymerase was incubated with 0,5 µg of cut plasmid and 1 µL primer 10 µM in a PCR machine (40 cycles: 30 seconds 94℃, 30 seconds 55℃, 1 minute 72℃). The ssDNA wa s then labelled at the 5’ end with γ^{32}-ATP, by using T4-PNK. 50000 Cpm of every single probe were used for the hybridization reaction. Hybridization, S1 nuclease digestion (Sigma Aldrich, #N5661-50KU) and recovery of the undigested DNA probe were performed according to a standard protocol (Nature Methods 2, 397-398). Dried gel was incubated for 48 hours with Phosphor Imager cassette.

5’RACE: Nuclei were isolated from cells and RNA was purified with MirVana (Ambion, #AM1560). cDNA was synthesized with major 3R primer: 4 µL of primer 10^{-12} µM were used for the reverse transcription. cDNA was polyadenylated with terminal transferase (Roche, #03 333 566 001) and amplified using major 3R and adaptor-polyT primer (5’- GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT). PCR product was then phosphorylated and cloned with Zero Background/Kan cloning kit (Invitrogen, #K260001).

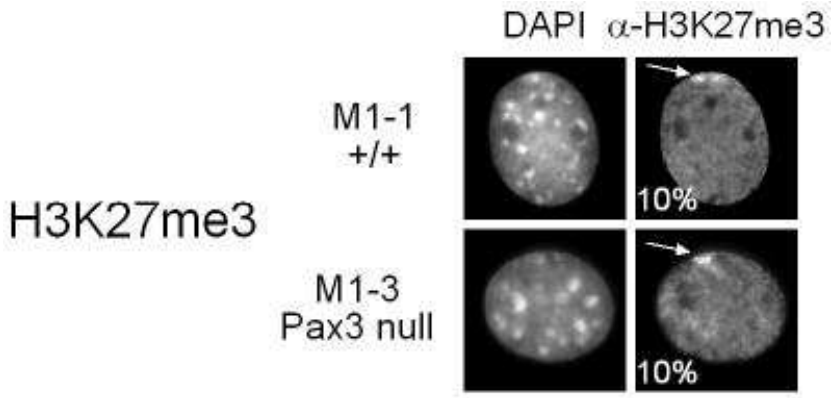
RNA-FISH: For RNA-FISH cells were plated on chamber slides and fixed with 4% PFA for 10 minutes, permeabilized with Triton for 5 minutes and fixed again with 4% PFA for 10 minutes. The slides were then washed for 3 minutes each time with 70% EtOH, 80% EtOH, 95% EtOH and 100% EtOH. A mixture of probes specific for the major satellites was used. The sequence of the probes is the following:

probe	sequence (capitals are LNA)
oMajL_1	5'-DSB-108Carb-AtTtAgAaAtGtCcAcTgTaGgAc
oMajL_2	5'-DSB-108Carb-AaTaTgGcAaGaAaAcTgAaAat
oMajL_3	5'-DSB-108Carb-AaTgAgAaAcAtCcAcTtGaCgAct
oMajL_4	5'-DSB-108Carb-TgAaAaAtGaGaAaTgCaCaCtg

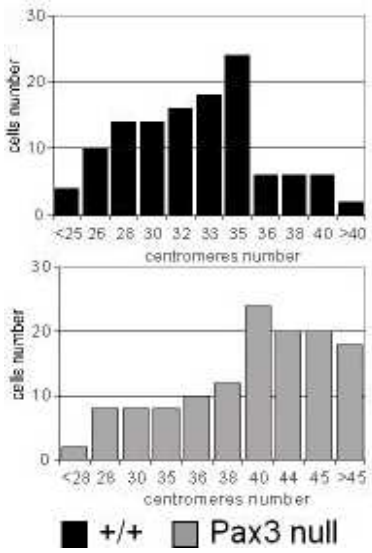
The probes were hybridized overnight at 37℃ in 2XS SC/formamide 50%. The next day chamber slides were washed first twice with warm 2XSSC/ formamide 50% and then twice in 0.1 % SSC. The chamber slides were blocked with 2.5% BSA diluted in 4X SSC/0.1% Tween, incubated with Avidin-Alexa 488 for 30 minutes at 37℃ and then washed twice with PBS/0.1% Tween.

Propidium iodide analysis: 2 x 10^6 were collected and resuspended in 0,8 mL PBS. 2.2 mL cold methanol was added and cells were incubated for 1 hour at 4℃. Cells were then centrifuged, resuspended in PI buffer (10mM TRIS pH 7,5, 5 mM $MgCl_2$, 200 µg/mL RNaseA, 50 µg/mL PI) and incubated at 37℃ for 30 minutes before analysis.

2.6 Supplemental data



Supplemental Figure S1 H3K27me3 on the inactive X is not affected by the absence of Pax3. Immunofluorescence using H3K27me3 antibody in wild type and Pax3 null fibroblasts. The arrow indicates the inactive X.



Supplemental Figure S2 In Pax3 null cells there are more centromeres. The number of centromeres was counted in at least 100 wild type and Pax3 null iMEF. In wild type cells, the average of centromeres is 35 and very rarely cells have more centromeres while in Pax3 null cells the number of centromeres increases and the vast majority of the cells have 40 or more centromeres.

3. CAN dsRNA MODULATE HMT ACTIVITY?

3.1 Introduction

RNA has a fundamental role at pericentric heterochromatin. It has been observed that mouse and human cells treated with RNaseA, an endonuclease that cleaves single-stranded RNA, lose heterochromatic marks: H3K9me3 and HP1α and are no longer enriched there (Maison et al., 2002).

Major satellites, like many other repeats, are transcribed. Transcription of pericentric repeats, both in S. pombe and in mouse, has been coupled to the cell cycle (Kloc et al., 2008; Lu et al., 2007); in humans, satellite III transcription is enhanced upon stress.

In S. pombe, S phase starts just after mitosis, when H3S10 is still phosphorylated and Swi6 is not bound to the chromosomes; therefore the vast majority of transcripts generated from dh and dg repeats appear during S phase: the forward strand transcript appears first, at much lower levels than the reverse strand. In mouse cells, transcription of the major satellite repeats starts during G1 and has a peak in G1/S transition. The transcripts then surround the heterochromatin (Lu et al., 2007) and also give rise to dsRNA, a unique feature of major satellite repeats (Martens et al., 2005).

In humans, satellite III repeats are 5 nt long sequences (5'-GGAAT) and are present in the centromeric region of some, but not all, chromosomes. These repeats are always transcribed at a low level and therefore it is very difficult to detect the transcripts, but upon stress and heat shock the transcription increases (>100- to 80´000-fold), depending on the source of stress (Rizzi et al., 2004; Jolly et al., 2004; Valgardsdottir et al., 2008).

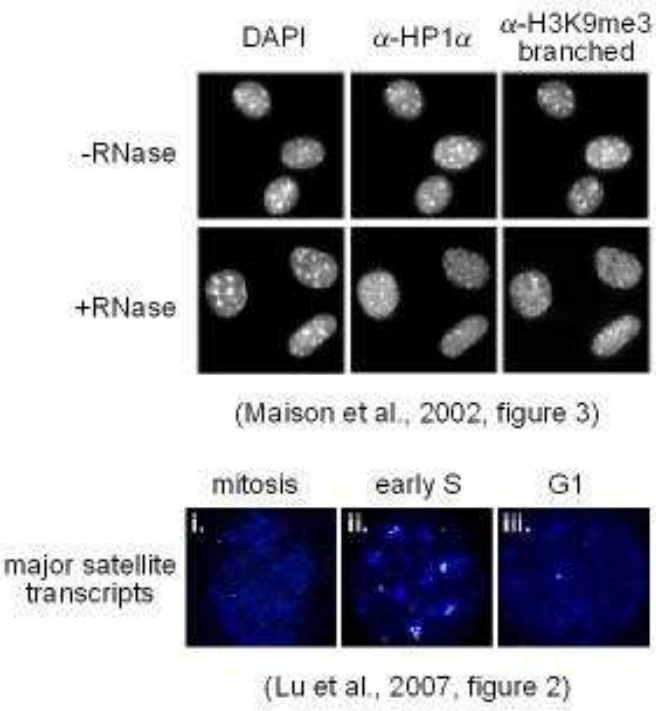


Figure 1 Major satellite transcripts are components of heterochromatin. When cells are treated with RNaseA, heterochromatic marks, like H3K9me3 and HP1α, are lost (top of the figure), suggesting that there should be a RNA structural component in the heterochromatin. This can be observed by in situ hybridization, using major satellite specific probes (bottom of the figure): major satellite transcripts surround pericentric heterochromatin.

In Pax3 null cells the amount of major satellite transcripts increases, due to the lack of Pax3 repressional activity. In addition, Pax3 null cells lose heterochromatic marks. The lost of heterochromatic H3K9me3 and H4K20me3 can be linked neither to depletion of the histones nor to delocalization of Suv39h and Suv420h enzymes. By performing chromatin immunoprecipitation with antibodies recognizing H3 histone, we observed that more histone H3 accumulate at pericentric heterochromatin in Pax3 null cells in comparison to wild type (see Figure 4, chapter 2). By ectopically expressing Suv39h and Suv420h enzymes tagged with EGFP in wild type and Pax3 null fibroblasts, we observed that the enzymes localize at pericentric heterochromatin even in Pax3 null cells: only the accumulation of Suv39h2 in Pax3 null cells seems to be slightly impaired.

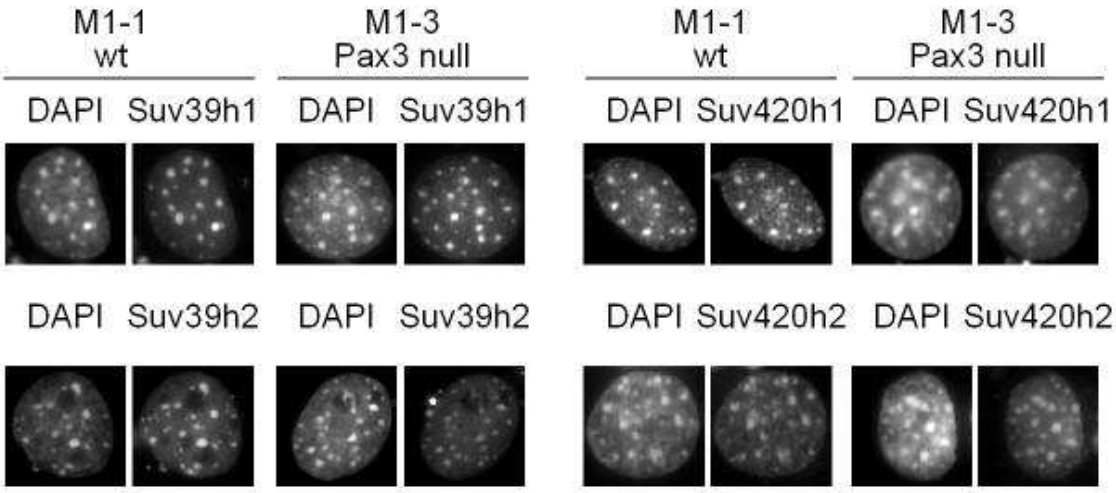


Figure 2 Suv39h and Suv420h localize at pericentric heterochromatin when ectopically expressed in Pax3 null cells. Localization of Suv39h and Suv420h tagged with EGFP upon overexpression in wild type and Pax3 null iMEF. The absence of Pax3 does not influence the localization of the Suv39h and Suv420h enzymes. Full length cDNA was used for Suv39h1, Suv420h1 and Suv420h2. For Suv39h2 the fragment 157-477 was expressed.

For explaining the paradox we conceived three different hypotheses:

a. RNA could inhibit HMTase activity: in Pax3 null cells the high amount of RNA present at pericentric heterochromatin is responsible for the inhibition of the enzymes;

b- RNA-capture mechanism: it has already been observed that certain RNA structure can bind SAM (Winkler et al., 2003; Loenen, 2006); therefore, at pericentric heterochromatin in Pax3 null cells, the increased amount of RNA present could sequester SAM and prevent histone lysine methylation;

c- RNA covers the nucleosomes and occlude the epitope. The interaction between RNA and nucleosomes would be facilitated by their opposite charges: in Pax3 null cells many nucleosomes are occupied by RNA and the activity of Suv39h enzyme is therefore prevented.

In order to find the correct explanation for solving the paradox, we performed HMTase assays by using different HMTase proteins and RNA transcribed from different repeats.

3.2 Results

3.2.1 Strand specific RNA synthesis

We transcribed strand specific RNA, using a plasmid previously generated in the lab (pCR4-Maj1-3), by cloning some repeats of the major satellites into the plasmid pCR4 (Lehnertz et al., 2003). Thanks to its particular backbone, when the plasmid is linearized with the restriction enzyme NotI, the forward transcripts can be generated by using T3 RNA polymerase (Fig. 2A). Reverse transcripts can be obtained by linearizing the plasmid with the restriction enzyme SpeI and using instead the T7 RNA polymerase (Fig. 2A). The RNA can then be annealed and digested with RNaseONE, an endonuclease that cuts preferentially single stranded RNA, leaving a pool of dsRNA. The quality of ssRNAs and dsRNA can be checked by using a Bioanalyzer device. Forward and reverse single-stranded RNA have the same length and can be visualized as a single band in the gel; compared to the size standards (or ladders), their size appears a little larger than what would be expected and this can be related to the complete absence of secondary structures. dsRNA appears instead like a smear: due to the repeats there are several positions where forward and reverse transcripts can overlap (Fig. 2B), and thus the size of the fragments generated after RNaseONE digestion is variable.

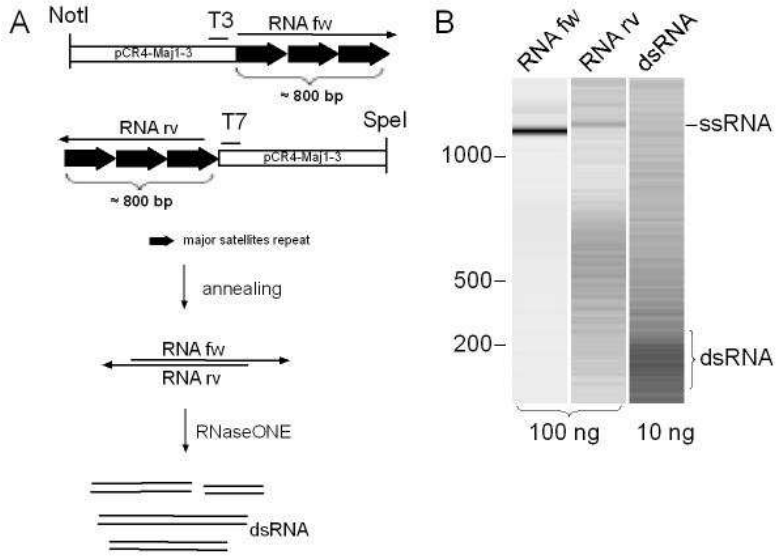


Figure 3 Preparation of ss and dsRNA and quality control. (A) Major satellite ssRNAs are transcribed using pCR4-Maj1-3 as a template. After annealing and digestion with RNaseONE, dsRNA is formed. (B) Quality control of ss and dsRNA using a Bioanalyzer.

3.2.2 Enzyme and RNA titration

It is very intriguing to think that RNA can inhibit HMTase proteins but on the other hand the alteration of pericentric marks we observed could be due to interaction of RNA either with SAM or with histones. Therefore for performing the HMTase assay, we introduced RNA before adding recombinant histone H3.1 and SAM to the reaction: if the RNA binds one of these two molecules, it should do so as soon as they are added into the reaction, preventing SET domain protein activity (Fig. 3A). In a first experiment we used 2 µg of Suv39h2 and different amounts of major satellite ss and dsRNA. The enzyme we used (aa157-477) contains an intact SET domain and a partial chromodomain (Rea et al., 2000). The amount of dsRNA used was 10-fold less than the amount of ssRNA, although at pericentric heterochromatin dsRNA is more abundant than ssRNA (Martens et al., 2005). We observed that Suv39h2 HMTase activity is inhibited by both forward and reverse major satellite transcripts and is completely blocked when dsRNA is also present during the HMTase reaction (Fig. 3B).

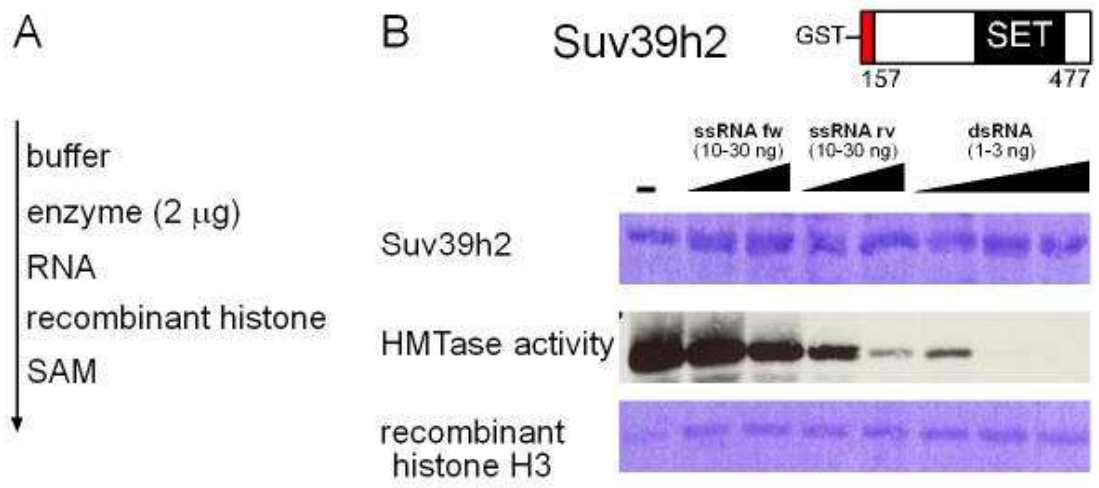


Figure 4 Titration HMTase assay. (A) The different reagents of the HMTase assay are always used in the order indicated. (B) HMTase activity of Suv39h2 towards recombinant histone H3.1 when different amounts of major satellite RNA are used. Major satellite RNA modulates Suv39h2 HMTase activity.

3.2.3 Is full length Suv39h1 inhibited by major satellite RNA?

By using several nanograms of major satellite RNA we observed inhibition of Suv39h2. Therefore we extended the experiment to other SET domain proteins and also made use of RNA transcribed from different repeats: the minor satellites. We observed that Suv39h1 is inhibited by both ds and ssRNA only when a proper chromodomain is present: indeed, the activity of the truncated protein (aa 82-412) previously characterized by our lab (Rea et al., 2000) is not influenced by the presence of RNA (Fig. 4A).

The inhibition by ds and ssRNA is RNA specific and protein specific: the activity of G9a, another SET-domain containing protein methylating H3K9 (Tachibana et al., 2001), is not influenced by RNA (Fig. 4A). In addition, the activity of the full length Suv39h1 and of the truncated Suv39h2 is not similarly perturbed when the reaction is performed with either minor satellite ss or dsRNA (Fig. 4B). Only Suv39h1 seems to be very slightly inhibited by forward RNA minor satellites transcripts, but this phenomenon is quite irrelevant in comparison to the inhibition with the major satellite transcripts.

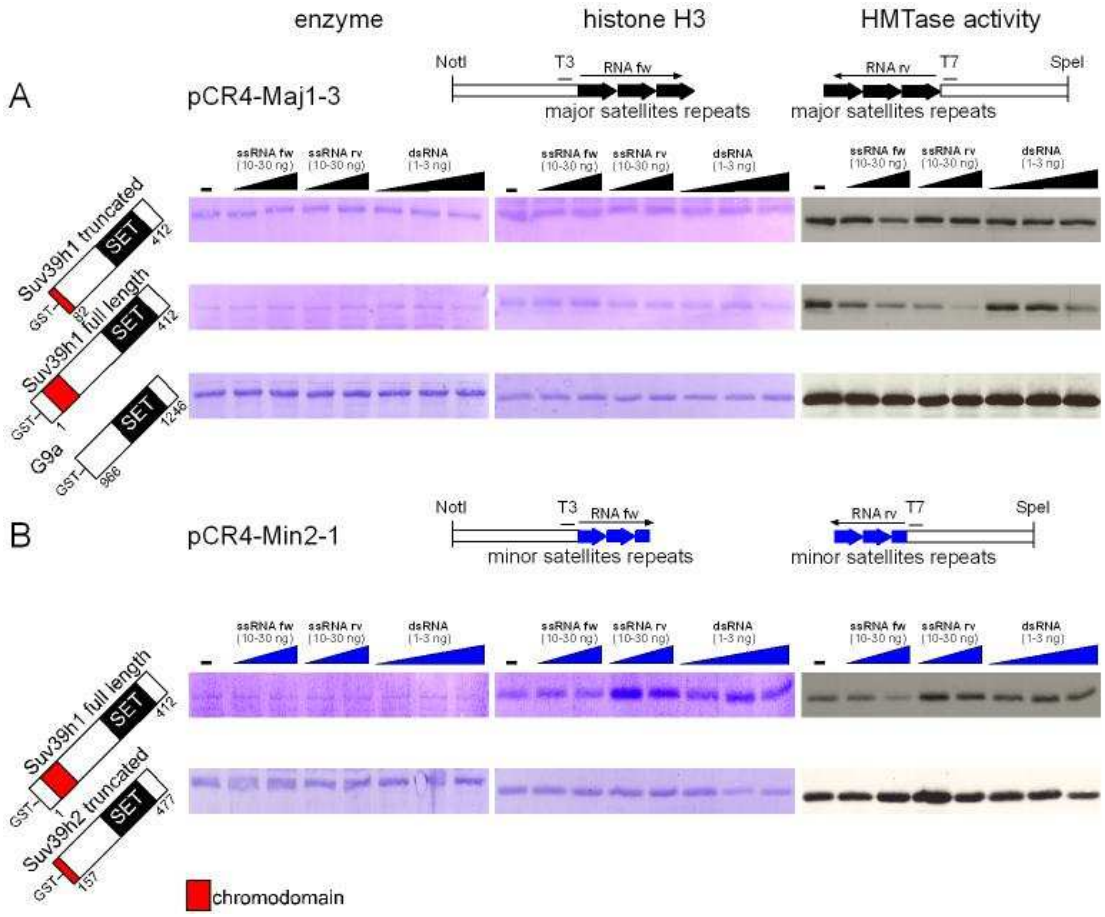
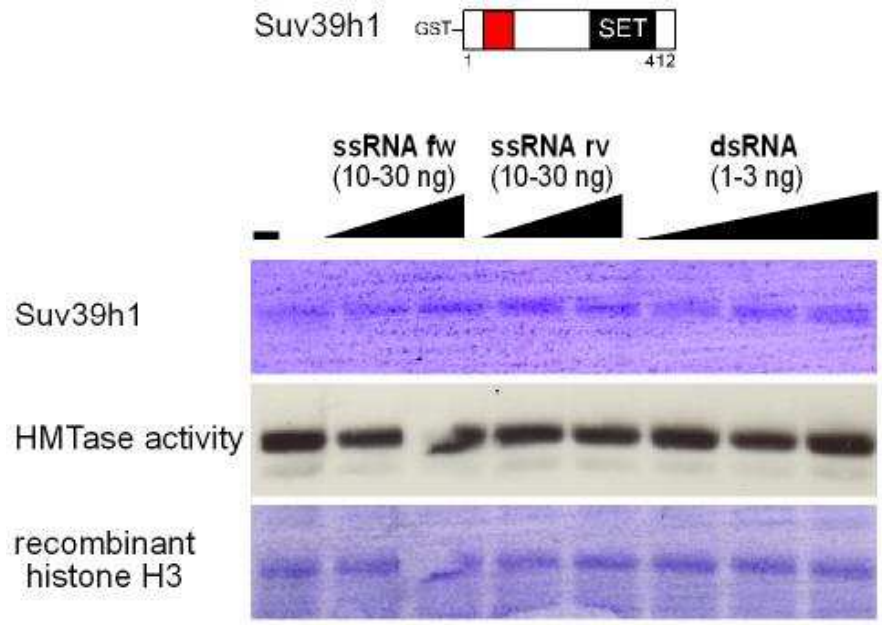


Figure 5 Suv39h1 HMTase activity is modulated by ss and dsRNA transcribed from the major satellite. Coomassie staining of the enzyme and histone H3 are reported respectively on the left and in the centre; the HMTase activity is reported on the right. RNA species and amount used are indicated on the top. (A) HMTase assay for full length and truncated Suv39h1and G9a using recombinant histone H3 as a substrate and RNA transcribed from the major satellite. Full length Suv39h1 is inhibited by major satellite transcripts. (B) HMTase assay for full length Suv39h1 and truncated Suv39h2 with minor satellite transcripts incorporated into the reaction. The activity of the two enzymes is not inhibited by minor satellite transcripts.

RNA inhibition of SET domain proteins could be an important mechanism and in order to understand better how this phenomenon could be regulated we tried to recapitulate this experiment using full length Suv39h1 and RNA transcribed from the major satellite. The enzyme was active but unfortunately we did not observe inhibition by RNA.



<u>Figure 6 Can dsRNA inhibit Suv39h1?</u> HMTase assay was performed with full length Suv39h1 and RNA transcribed from the major satellite in order to recapitulate what was observed in the previous experiment. No inhibition from the RNA was observed.

3.3 Discussion

In a first experiment we observed that Suv39h1 and Suv39h2 are specifically inhibited by major satellite transcripts. Suv39h1 is inhibited only when it contains an intact SET domain and a full length chromodomain. The enzymatic activity of Suv39h1 containing a partial chromodomain (aa82-412) is not affected by major satellite RNA. For Suv39h2 we can report only the result for the truncated protein: despite several effort in our lab, unfortunately we still do not have a full length clone. In contrast to what we observed for Suv39h1, major satellite transcripts inhibit Suv39h2 even if it does not carry a full-length chromodomain. The inhibition observed in the two enzymes is paradoxical and it is not clear whether the presence of the homeodomain promotes major satellite RNA inhibition.

Nevertheless, one of the most intriguing hypotheses is that the RNA binds the chromodomain in order to inhibit the SET domain. Chromodomains binding non-coding RNA have been observed in Drosophila (Akhtar et al., 2000): MOF, a histone acetyltransferase involved in dosage compensation, specifically binds through its chromodomain to roX2 RNA in vivo; in vitro analysis has shown that the MSL-3 chromodomain can also bind RNA. For Suv39h1 and Suv39h2 we observed that inhibition of the HMTase activity is RNA specific but unfortunately we have not any data supporting the binding of major satellite RNA to the chromodomain.

Unfortunately, when we tried to reproduce these data, we could not observe modulation of the HMT activity by RNA. The data are difficult to interpret and it is not clear whether dsRNA modulates the activity of HMT. Specific experiments will be done in future to solve this contradiction and eventually to find the exact mechanism of inhibition.

3.4 Material and methods

RNA in vitro transcription: RNA from major and minor satellites was in vitro transcribed using pCR4-Maj1-3 and pCR4-Min2-1 (Lehnertz et al., 2003) as templates. The plasmids were first linearized with either NotI (for the forward transcription with T3 RNA polymerase) or SpeI (for the reverse transcription with T7 RNA polymerase). Transcription was performed for 1h at 37°C. Complementary ssRNA were annealed at 70°C for 10 minutes, immediately cooled on ice and then digested with 2U RNaseONE. RNaseONE digestion was stopped by adding 93 µL water and 6 µL 10% SDS. RNA was then extracted with RNeasy (Qiagen). The quality of the RNA was evaluated with Bioanalyzer (Agilent Technologies).

HMTase assay: For radioactive assays tritium labeled S-adenosyl methionine was used (S-Adenosyl-l-methyl-^{3}H]methionine (^{3}H-SAM); GE Healthcare Lifesciences). 2µg of recombinant enzyme were mixed with ss or dsRNA, 1µg of recombinant histone H3.1 (New England Biolab, #M2503S) and 1.5µL of ^{3}H-SAM (1µCi/µL) in a 50mM Tris-HCl pH 8.5 buffer containing 5mM DTT. The total volume of the reaction was 25µL. The mixture was incubated for 1 hour at 37°C and run on a 15% SDS-PAGE gel. The gel was blotted to Immobilon-P membrane (Millipore) using a TE70 semi-dry transfer unit (GE Healthcare Lifesciences) and sprayed 3 times with EN3HANCE spray (Perkin Elmer) according to the manufacturer's instructions before exposure to Amersham Hyperfilm MP. The radiography film was incubated for 2 days at -80°C before being developed.

4. Su(var)

4.1 Introduction

4.1.1 PEV is a tool for finding new Su(var) genes

For a long time it had been believed that chromosome structure had no effect on gene expression: some chromosomal rearrangement had been characterized in Drosophila and no mutant phenotype had been observed.

In 1930 Muller described a new phenotype: in certain flies, after an X-ray induced chromosome rearrangement, the eye color was variegated: the different ommatidia forming the single eye expressed a different amount of pigment. This phenotype could also be stably inherited. It was discovered later that the cause of this phenotype was an inversion, defined $In(I)w^{m4}$. X-rays had induced two breakpoints in the first chromosome: one breakpoint was 25 kb far from the *white* gene and the other in the pericentric heterochromatin. As a consequence, the *white* gene, usually very distant from pericentric heterochromatin, was now in proximity of it and its expression was variegated, due to spreading of the heterochromatic structure from the breakpoint over the gene. It became clear that the position of a gene on a chromosome has an effect on its expression, which in some case may be variegated. For this reason the phenomenon was called position effect variegation (PEV) (reviewed in Girton and Johansen, 2008).

The variegation of the expression of the white gene was the first to be observed and it is also one of the easiest, since the eye of the fly is very large, but potentially all the genes can variegate when they are in the appropriate location. Variegation can be observed when euchromatic genes are inserted into pericentric regions or when heterochromatic sequences move into euchromatin.

Position effect variegation can be used as a tool for discovering components of pericentric heterochromatin (Reuter et al., 1992). Mutations can be induced in flies for example by feeding them with EMS (Ethylmethane Sulphonate); when a strain in which the *white* gene is variegated, is mutagenized and a gene encoding for a heterochromatic protein is mutated, heterochromatin will be impaired and the *white* gene will be expressed. The eye pigmentation of the offsprings will be much darker than the phenotype typical of the variegation. A gene, that, when mutated, suppresses position effect variegation, is a *Su(var)* (Reuter et al., 1982).

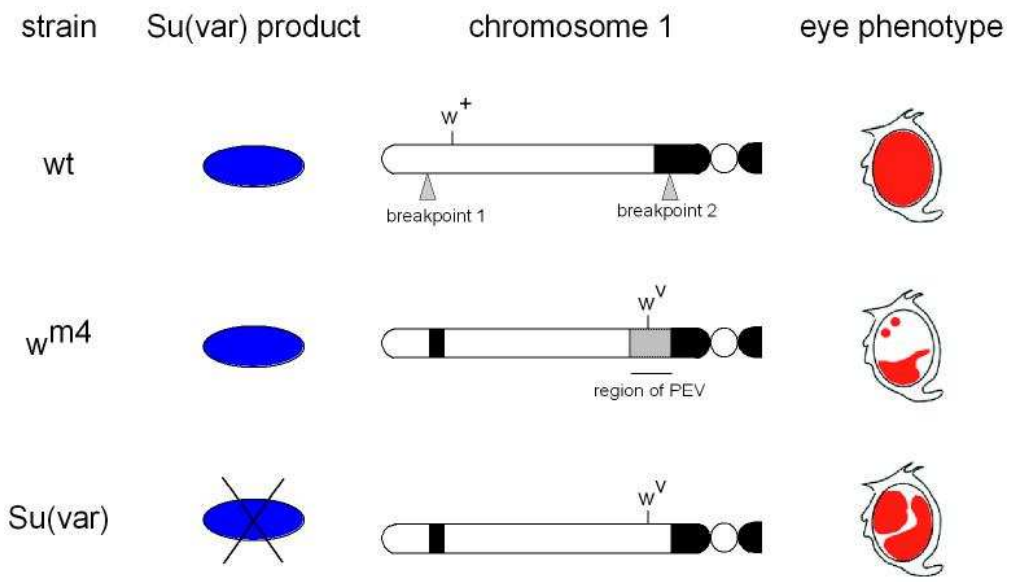


Figure 1 w^{m4} flies as a tool for finding new Su(var). When in the wt strain rearrangements are induced, the white gene, usually very far from heterochromatin, can be relocated in its proximity. As a consequence, the color of the eye will not be uniform red but rather variegated, due to the repressive influence of heterochromatin on the expression of the white gene. When in this second (w^{m4}) strain a Su(var) gene is mutated, heterochromatin is impaired and the white gene is expressed at levels comparable to the wild type. The mapping of the mutated gene thus identifies a new Su(var).

In 1967 the first Su(var) genes were isolated and characterized (Spofford, 1967). Later, large screens were performed, mainly by Gunter Reuter. Around 120 independent mutations have been isolated so far, but the number of Su(var) genes could potentially be much larger, considering that the screen has not been saturated and to date has only focused on dominant suppressors (Reuter et al., 1981; Schotta et al., 2003; Schulze et al., 2007).

Less than 50 out of around 120 *Su(var)* genes in Drosophila have been characterized. They encode either for structural components of heterochromatin, like HP1/Su(var)2-5 (James et al., 1996) and Su(var)3-7 (Reuter et al., 1990), or for heterochromatin modifying enzymes, like Su(var)3-9 (Schotta et al., 2003) and Rpd3 (De Rubertis et al., 1996).

The number of Su(var) lines already isolated and the possibility of finding many new ones gives an idea of how complex heterochromatin must be. The Su(var) collection represents the best tool for finding proteins playing a role in heterochromatin. Although based on Drosophila, results from this approach are

not limited to this species, since the key players of heterochromatinization are found to be strongly conserved (Grigliatti, 1991).

4.1.2 SNP mapping is an approach for mapping new Su(var)

The forward genetics can make use of several approaches for mapping a mutation: one of them is SNP mapping, which is based on the existence of SNPs among the different Drosophila melanogaster strains isolated during the years (Dodgson et al., 1997).

It is fundamental to determine in which genomic positions there are SNPs while comparing the Su(var) and P-element strains. It is better to perform this analysis using heterozygous flies that for every chromosome couple carry one chromosome of the Su(var) strain and one chromosome of the P-element one. By comparing their genomic sequence with the one from homozygous P-element carrying flies, it is possible to understand which nucleotides are specific to the P-element strain and which are characteristic of the Su(var) strain. Several regions covering one chromosome arm can be analyzed, by amplifying and then sequencing 1 kb regions in both F1 heterozygous and P-element carrying flies. The identified SNPs are then used as anchor point during the successive analysis.

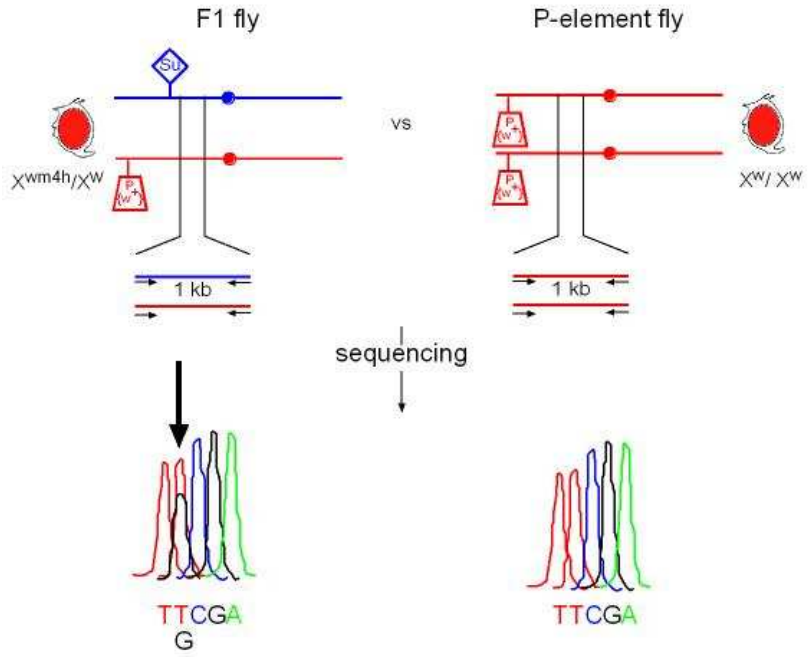


Figure 2 SNP detection. This approach is based on the fact that at certain genomic positions the nucleotides present in the strain in which mutations were induced are different from those present in the P-element carrying strain. For this reason the flies of the F1 generation (generated by crossing the Su(var) strain and a P-element carrying strain) will show two waves (corresponding to two different nucleotides) in that position.

A fundamental requirement for the SNP mapping approach is the chromosome recombination that happens in females. Recombination can take place at several locations on the genome and when it occurs in the region between the P-element and the Su(var) genes, it gives rise to two distinct chromosomes that will carry either both the P-element and the Su(var) mutation or none of them. The flies of the F2 that receive from the mother the former chromosome have a red eye, because of the P-element; the ones that will inherit the latter will have variegated eye. This is visible just in the male F2 progeny (50%) that carry the X^{wm4h} chromosome, thanks to the combination of the inversion in the X chromosome and of the chromosome without P-element and Su(var) mutation.

The further the P-element and the Su(var) mutation are from each other, the higher the frequency of recombination and so the number of male flies that in the F2 generation have variegated eye colour.

The analysis of the recombinants makes it possible to localize Su(var) genes: the flies have variegated eyes because the Su(var) mutation was replaced, through

recombination, with a functional Su(var) gene. This means that in a certain region close to the P-element position they are heterozygous, but in the proximity of the Su(var) gene they should be homozygous. The localization of this border region of heterozygosity/homozygosity in different recombinant flies, using the known SNP position, allows the identification of the region where the Su(var) gene localizes.

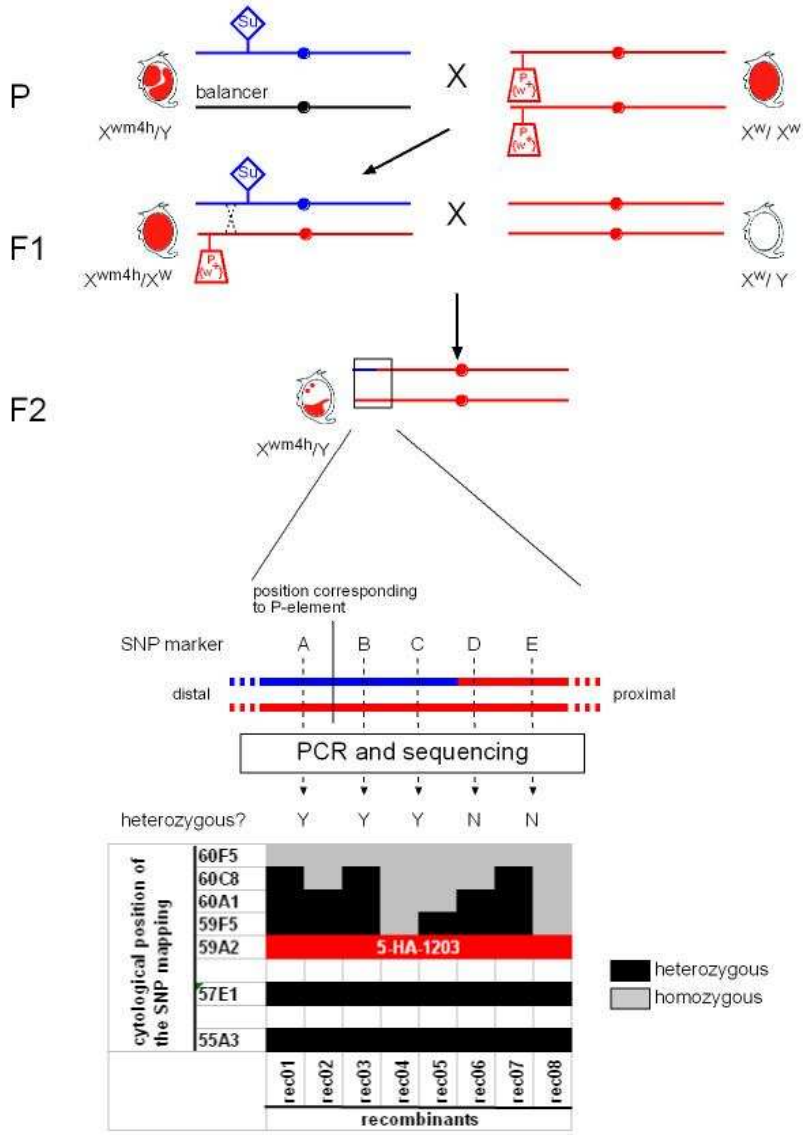


Figure 3 Su(var) mapping by using SNPs. Flies from the F2 generation that have the variegated eye, do not have the Su(var) mutation because a recombination event took place in the genomic region between the P-element and the Su(var). By knowing the position of the SNPs and by sequencing those genomic regions from the flies, it is possible to know whether a certain fly in a certain position is heterozygous or homozygous. After analyzing several recombinants, as illustrated on the bottom of the figure, the border between homozygosity and heterozygosity suggests where the Su(var) gene is located.

4.2 Results

4.2.1 Su(var)2-21 localizes at the tip of 2R and two genes localized at the tip of 2R encode for pericentric proteins

Preliminary data obtained from the analysis of a few recombinants generated by recombination of Su(var)2-21 with several P-elements distributed all over the second chromosome showed that Su(var) 2-21 is located at the tip of 2R (cytological region 55-60). For this reason, P-elements localized in the tip of 2R were used in a large-scale generation of recombinants. The analysis of the recombinants suggested that Su(var)2-21 is located at the tip of 2R, in a space of 700kb that covers the cytological region 60D4-60F5.

Some Su(var) genes have been mapped by recombining them with P-elements located on both sides along the chromosome. For Su(var)2-21 this was difficult: although the frequency of recombination is higher in the proximity of the telomeres, there are not many strains available with a P-element inserted in 60D4-60F5. Moreover, few SNPs were present in that part of the chromosome. Deletion mapping could have been performed. This technique was very useful for mapping the first Su(var) genes (Reuter et al., 1987) but requires a very fine distribution of the deletions available in the region of interest and this is not the case for the strains carrying a deletion at the tip of 2R. With one exception all these deletion strains are lethal in homozygosity, suggesting that some important gene is located there, for instance Su(var) 2-21, which is also lethal in homozygosity. Considering all these reasons, we decided to look at the annotated genes in that region and to assess which one could have been a Su(var). In 60D4-60F5 there are 97 annotated genes that encode for proteins that cover a variety of functions, spanning from metabolic activity to membrane signaling (Table S1). Not all the Drosophila genes have a mammalian orthologue. Excluding genes that possess domains known not to participate in heterochromatin formation and maintenance, 22 genes can be involved in heterochromatin organization and half of them have a mammalian orthologue (Table S2). The mammalian orthologues of these genes were cloned and fused to EGFP and their subnuclear localization visualized in fibroblasts. There were two mammalian orthologues of the cloned genes that localized at pericentric

heterochromatin: Pax3 and Pax7. These two proteins are the mammalian counterparts of, respectively, gsb-n and gsb in Drosophila.

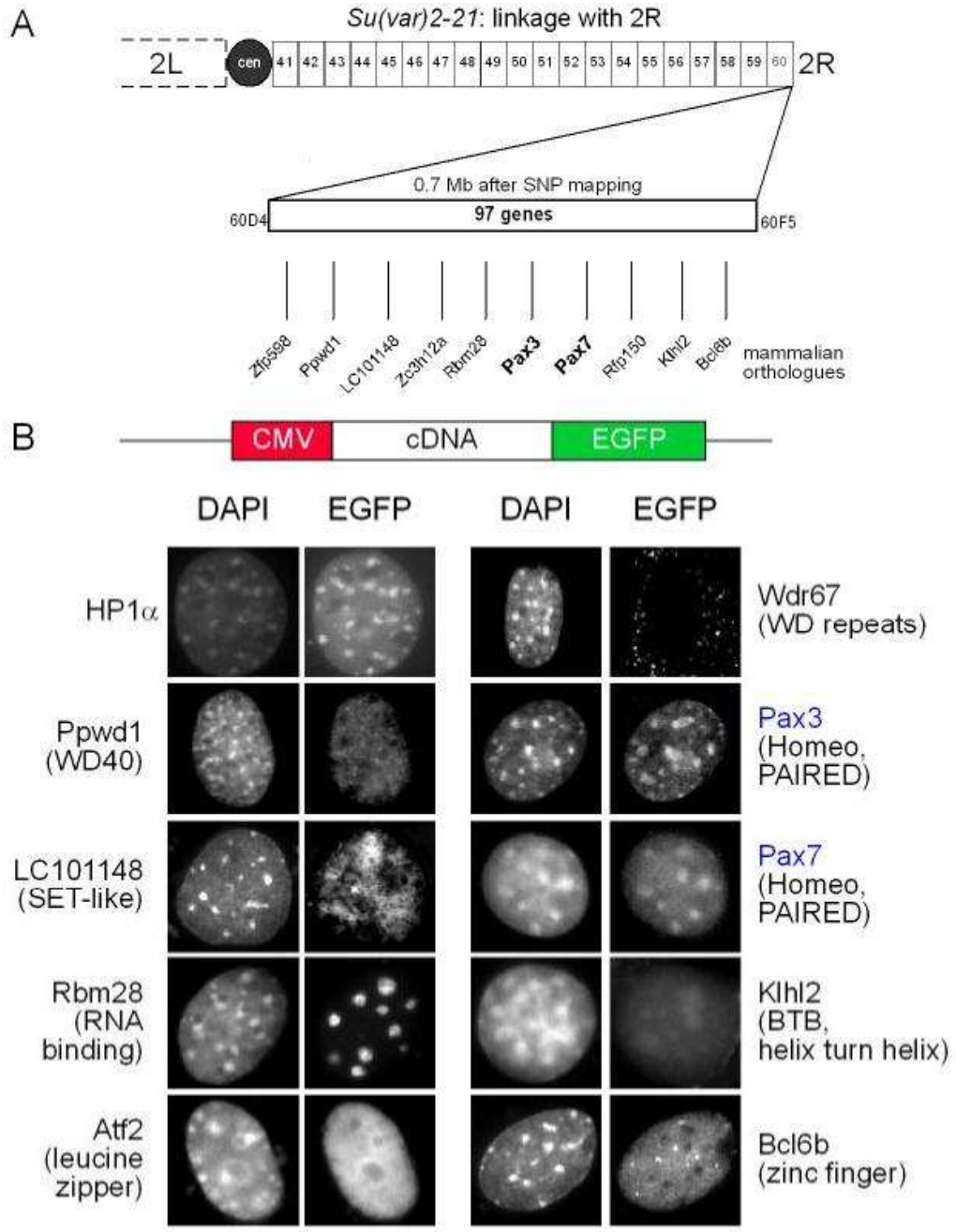


Figure 4 Two mammalian orthologues of the candidates in 60D4-60F5 localize at mouse heterochromatin. (A) By using the SNP mapping the region where Su(var)2-21 is located could be narrowed down from 20 Mb to 700 kb. The mammalian orthologues of the Su(var) candidate genes are indicated. (B) The cDNA of the mammalian orthologues was fused with EGFP and the localization was observed in mouse wild type female cells (w9). Among the several candidates only two proteins, after fusion to EGFP, localized at pericentric heterochromatin: Pax3 and Pax7.

4.2.2 No point mutation is present in the gooseberry locus

The gooseberry locus in Drosophila is around 30 kb (Bopp et al., 1986; Baumgartner et al., 1987) and is located in 60F1, very close to the tip of 2R. gsb-n and gsb are transcribed in opposite directions without sharing the promoter. The structure of the two genes is different from one another: gsb-n has 5 exons and 2 of its introns are rather long, while gsb has only 2 exons separated by a relatively short intron. Knowing that the mammalian orthologues localize at pericentric heterochromatin, we sequenced the full locus both in Su(var)2-21 and in a randomly selected Su(var) strain, Su(var)2-18, and no point mutation was found.

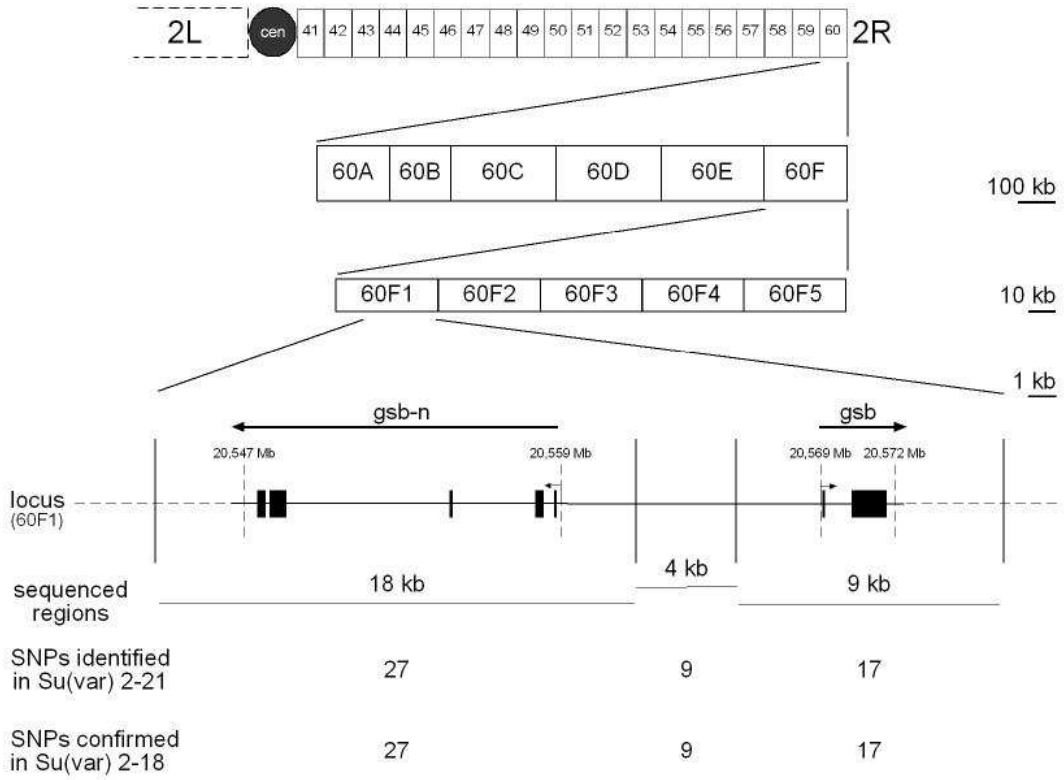
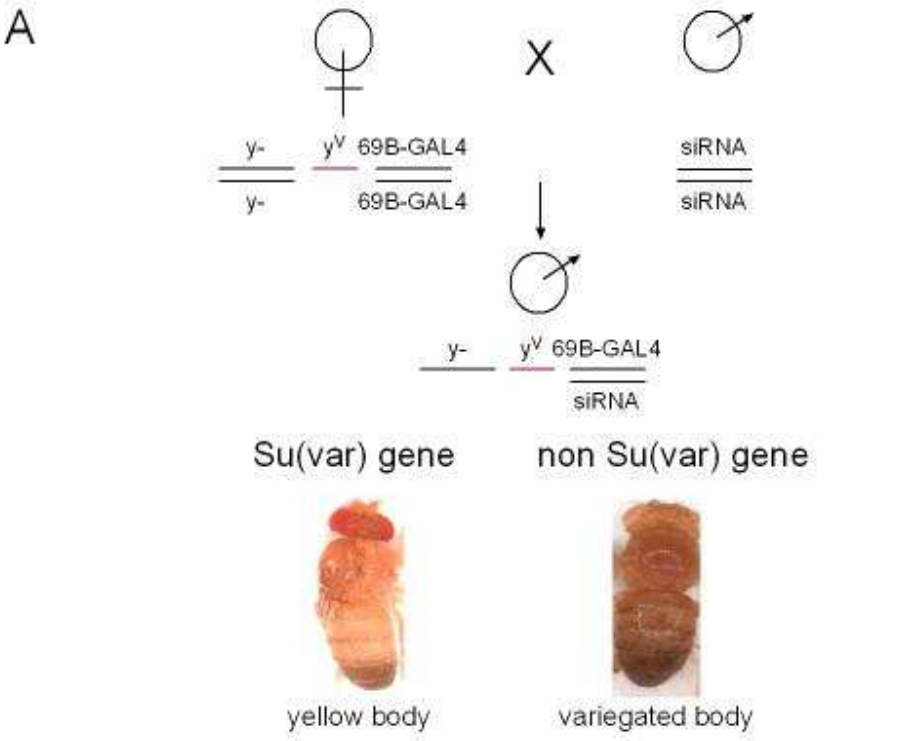


Figure 5 Structure of the gooseberry locus. 53 different SNPs are present in this locus and all of them can be found both in the Su(var)2-21 strain and in another randomly picked Su(var) strain, Su(var)2-18. This means that none of these SNPs is the Su(var) mutation that we aimed to map but it does not rule out the possibility that gooseberry is a Su(var).

4.2.3 Could RNAi library be used for finding new Su(var)?

The identification of genes belonging to a certain pathway has been greatly accelerated in Drosophila thanks to the generation of a RNAi library (Dietzl et al., 2007): 80% of the known genes in Drosophila can be silenced in vivo by crossing RNAi flies and GAL4-lines. By using specific GAL4 lines it is possible to silence a certain protein only in a specific tissue and this is very helpful when the mutation of a protein is lethal very early in development. The RNAi library was one method we could have used for identifying Su(var)2-21, once we narrowed down the region where it localizes from 20 Mb to 700 kb and we knew that this mutation is located in the tip of 2R. Unfortunately there were some problems connected to the use. The Su(var) phenotype can be observed in males only when they have the inverted X chromosome X^{wm4h}. The flies from the RNAi library carry a *white* gene close to the RNAi gene that is normally used for checking which flies in the progeny have received the silencing chromosome. Therefore, the offspring that will inherit the silencing RNA will also inherit the *white* gene, whose expression makes it impossible to understand whether the silenced gene is a Su(var). A different variegated phenotype could also have been used, like bw (brown) variegation, affecting also the eye color, and yellow variegation, affecting the body color. The bw variegation (bw^{v}) can be visualized only when flies have a recessive allele for the *v* (vermilion) gene, another gene involved in the eye pigmentation. *bw* and *v* are located on two different chromosome: *bw* is on chromosome 2, while *v* is on the X chromosome. Drosophila has 4 chromosomes and for screening by using bw^{v} only RNAi constructs inserted into the chromosome 3 could have been used, limiting in this way the number of screened genes. An extra chromosome is responsible for the yellow variegation that can be visualized if the male flies carry a recessive allele for the yellow gene on the first chromosome. Female yellow variegated flies carrying also the GAL4 gene can be crossed with males from the RNAi library (Fig.6A). A small attempt has been made by silencing Su(var)3-9 and Su(var)2-5 but no suppression of the yellow variegation was observed (Fig. 6B). This could have been due either to the low efficiency of the silencing constructs or to the mild activity of the promoter driving the expression of GAL4.



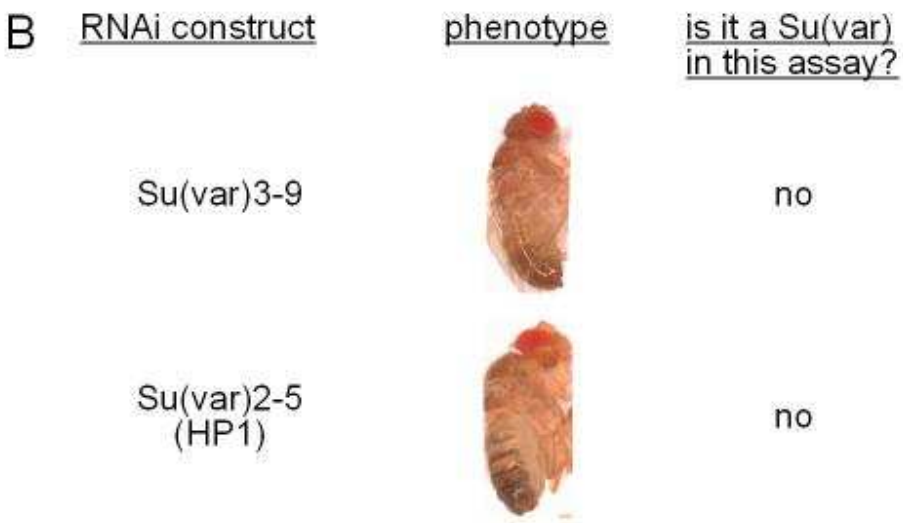


Figure 6 An attempt to use RNAi flies for finding Su(var) genes. (A) Female flies carrying yellow variegation and 69B-GAL4 are crossed with males from the RNAi library. The phenotype of the progeny can be used for finding new Su(var) genes: the male offsprings will have a yellow body if the silenced gene is a Su(var). 69B is a ubiquitously active promoter. (B) Fly lines expressing silencing constructs for Su(var)3-9 and Su(var)2-5 were used but these two genes did not suppress the yellow variegation in this assay.

4.3 Discussion

Our goal of mapping Su(var)2-21 was not reached, although several approaches were tried: we could observe that Su(var)2-21 localizes in a region of 700 kb in the tip of 2R (60D4-60F5). In this gene-poor region, some important one must be located, since all the strains carrying a deletion in this part of the genome are lethal in homozygosity.

The experiments we performed on mouse fibroblasts suggested that, among the 97 genes present in the tip of 2R, it is likely that either gsb or gsb-n is Su(var)2-21, since their mammalian orthologues, Pax3 and Pax7, localize at pericentric heterochromatin. No other mammalian orthologues of the Drosophila genes located in 60D4-60F5 localized at pericentric heterochromatin, and since we know from previous studies that mammalian orthologues of Su(var) genes localize at pericentric heterochromatin we suspected that the gsb locus encodes for a Su(var) gene.

But, although Pax3 and Pax7 localize at pericentric heterochromatin, we were not able to find any point mutation either in the open reading frame or in the untranslated regions of gsb-n and gsb. These two genes are transcribed from two distinct promoters located between the two gene bodies: they were also sequenced and no mutation was found.

There are several interpretations that can explain why we cannot claim, on the strength of the present results, that the gsb locus contains a Su(var) gene.

In comparison to mouse, the structure of individual genes in Drosophila is less complex: Drosophila genes in general have a limited number of introns and they are usually relatively short. If we compare for example gsb-n and Pax3, we notice that their extension and structures are completely different, although the two genes encode for two orthologues and the sequences of the introns at the exon-intron boundaries are very similar: the gsb-n locus is less than 15 kb long, while the Pax3 locus is over 100 kb. The fact the structure of a gene is not so sophisticated, when compared to more complex species, does not imply that the organization of the gene is simple too: when we take into consideration regulatory elements in addition to exons and introns, we can observe that also in Drosophila the gene organization is very complex and we cannot exclude that a

point mutation could be located in a regulatory element whose position is to date unknown.

Another possibility is represented by the interference of an unidentified locus on the gsb locus. It is very difficult in this case to define the identity, the position and the precise effect of the locus. The SNP data obtained by analyzing recombinants pinpointed a point mutation to position 60D4-60F5. The interference from an unknown locus could reinforce or weaken the Su(var) effect. In this case, since no point mutation is present in the gene, we could argue that the mutation, wherever it is, does not completely abolish the activity of the protein, and the overall Su(var) effect is reinforced by the interference of another locus.

Concerning this last aspect we should also mention that gsb mutations are homozygous lethal, like Su(var)2-21, but at the same time we know that Su(var)2-21 is not a strong modifier of PEV, although it is a dominant one. The mutation could therefore still leave a residual activity of the transcription factor and this would support the idea that the mutation is not in the gene body, as we observed for gsb and gsb-n.

For understanding whether gsb and gsb-n are Su(var) genes, the most straightforward approach is to destroy them selectively and to analyze the heterozygous phenotype: as mentioned above, the homozygous mutation is lethal, meaning that the gsb locus has a fundamental role in fly development. In mouse, Pax3 is also an important developmental regulator and in this thesis we showed that it is also a repressor of the transcription of the major satellite. Therefore there is no reason for excluding the possibility that gsb-n could also be a developmental as well as a heterochromatic regulator.

The identification of a new Su(var), as depicted here, can be a very long and complex process but it allows us to discover and characterize key components that are conserved in different species and that take part in the very well orchestrated process that is heterochromatinization.

4.4 Supplemental data

Table S1: Analysis of the genes located in Drosophila melanogaster in the cytological region 60D4-60F.

cytological region	Drosophila gene	mammalian orthologue	domains in the protein
60D5	itp		crust neurohorn
	nurf 38	ppa2	pyrophosphatase
	CG11414	zfp598	RING Zn-finger
	origami		prefoldin
	CG12252	ctdp1	Ctd like phosphatase
	CG3511	ppdw1	WD40, cyclophillin
	start1	stard	START
60D8	CG4681	tes 9	Smc
	pio		Zp
60D9	CG3570	gene for hypothetical protein LC101148	
	CG13587		CLECT
	CG4692		
	CG3548		
	CG4707		
	CG13588	zinc finger CCCH type containing 12C	
	CG3608	adck1	ABC1, S/T kinase
	CG13594		
	CG4741		
	CG3565		
	CG3579		
60D10	Cyp9c1	cyp3a16	CytP450
	CG3640		Scp
	CG4781	gene for hypothetical protein	LRR
60D13	CG30161		
	CG3683	ndufa8	CHCH
	CG33228		
	pof		RRM
	mmp1	mmp14	Zn-dependent metalloprotease
	CG4806	RNA-binding motif protein 28	RRM
	CG3663	isoc1	
	CG30163		
60D14	ST6Gal	st6gal2	glycotransferase
	CG33988		
60D16	CG30421	usp43	peptidase
60 E1	CG9196		
	CG3880	gene for hypothetical protein	
	CG12848	chchcd1	
	CG3894	neurl2	NEUZ, SOCS

	CG16936	GST theta 3	GST
	eps15	epidermal growth factor	
	CG30426	setDB1	SET
	CG30422		
	lcp9		
	eap		RRM
	CG16912	yars2	tyrosil-tRNA synthetase
	CG3589	tysnd1	trypsin like protease, degQ
	key		Smc
	ETH		
60 E2	orc4	orc4l	CDC,ATPase, ABC multidrug pump
	CG12849		DM8
	CG3611		
	Dll	distal lex homeobox 1	homeodomain
60 E3	CG3650	trypsinogen 7	trypsin like serine protease
60 E4	CG15859		
	atf2	atf2	
	CG16896	wdr67	WD repeats
	CG30423	leprotl1	Vsp55
60 E5	CG9047	gene for hypothetical protein	DUF798
	CG32836		
	ance-5	angiotensin converting enzyme	peptidase M2
	CG9083		
	CG2857	solute carrier family 25	
	CG30427	mlstd2	
	CG3770	lhfpl2	
	phk3		
	CG2736	scavenger receptor class B	CD36
	zip	myh10	myosin motor and myosin tail
	rpl41	ribosomal protein	
	CG18506		
	pain	trpa1	ankirin repeats
	CG3760		
	CG2811	gene for hypothetical protein	UPF0131
60 E8	tina-1		UPF0131
	CG12851		
	CG15861		
60 E11	CG30424		
	rpl19	rpl19	
	CG3776	gene for hypothetical protein	tim44
	CG2790	gene for hypothetical protein	Zn fingers
	epithelial membrane	epithelial membrane protein	

	protein		
	CG2765		spot14
60F1	uzip		
	nplp1		
	gsb-n	pax3	PAIRED, homeodomain
	gsb	pax7	PAIRED, homeodomain
60F3	gol	ring finger protein 150	RING finger
60F5	tkr	klhl2	BTB, helix turn helix
	CG30430		
	CG34038		
	peb		
	CG9380		
	CG30429	gene for hypothetical protein	
	CG33680		DUF223
	CG30428		
	reg-5		
	kr	bcl6b	Zn finger, Sfp1

Table S2: List of genes located in Drosophila melanogaster in the cytological region 60D4-60F5 that cannot be excluded from being Su(var) genes for any obvious reasons.

	Drosophila protein	mammalian protein	size in Drosophila	size in mouse
Drosophila genes WITH a mammalian orthologue	CG11414	Zfp 598	867 aa	908 aa
	CG3511	Ppwd1	637 aa	646 aa
	CG3570	hypothetical protein LC101148	302 aa	343 aa
	CG4806	Rbm 28	657 aa	750 aa
	Atf2	Atf2	536 aa	487 aa
	CG16896	Wdr67	988 aa	996 aa
	gsb-n	pax3	449 aa	488 aa
	gsb	pax7	427 aa	503 aa
	gol	ring finger protein 150	286 aa	437 aa
	tkr	Klhl2	1046 aa	593 aa
	Kr	Bcl6b	502 aa	474 aa
			size	
Drosophila genes WITHOUT a mammalian orthologue	CG3548	/	453 aa	
	CG4707	/	673 aa	
	CG13588	/	430 aa	
	CG30161	/	251 aa	
	CG30422	/	315 aa	
	CG32836	/	381 aa	
	CG18506	/	313 aa	
	CG12851	/	1047 aa	
	CG30424	/	459 aa	
	CG9380	/	521 aa	
	CG30428	/	263 aa	

5. Appendix

5.1. Buffer composition

Phosphate buffered saline (PBS) per litre (pH to be adjusted at 7,4):

- 8 g Sodium chloride
- 0.2 g Potassium chloride
- 1.44 g Na_2HPO_4
- 0.24 g KH_2PO_4

SDS Running buffer (10x) per litre:

- 30.2g Tris-Base
- 144 g Glycine
- 100mL10% SDS (sodium dodecyl sulphate)

4x SDS loading dye per 50mL:

- 4 g SDS
- 23 mL Glycerol (87%)
- 0.77 g DTT
- 12 mL 1M Tris pH6.8
- 2 mg Bromophenol blue

Tris buffer saline with Tween 20 (10x) per 1 litre:

- 500 mL 1M Tris pH8
- 200mL 5M Sodium chloride
- 10mL Tween 20

SDS-gel staining buffer:

- 50% Methanol
- 10% Acetic acid
- 0.25% Coomassie brilliant blue R-250

SDS-gel destaining buffer:

- 50% Methanol
- 10% Acetic acid

Annealing buffer (10x):

- 100mM Tris-HCl pH8.0
- 10mM EDTA pH8.0

- 1M NaCl

Tris acetate EDTA buffer (10x):

- 0.4M Tris Base
- 0.2M Acetic acid
- 0.01M EDTA

Washing solution for immunofluorescence:

- 0.25% Bovine serum albumine
- 0.1% Tween 20
- in 1x PBS

Blocking solution for immunofluorescence:

- 0.25% Bovine serum albumin
- 0.1% Tween 20
- 10% normal goat serum (Jackson Immuno)
- in 1x PBS

5.2. Media composition and concentration of antibiotics

Luria-Bertani (LB) medium per litre:

- 10g Tryptone
- 5g Yeast extract
- 10g NaCl
- dissolved in distilled water and autoclaved

Antibiotics for bacterial cell culture:

Ampicillin 100 µg/mL

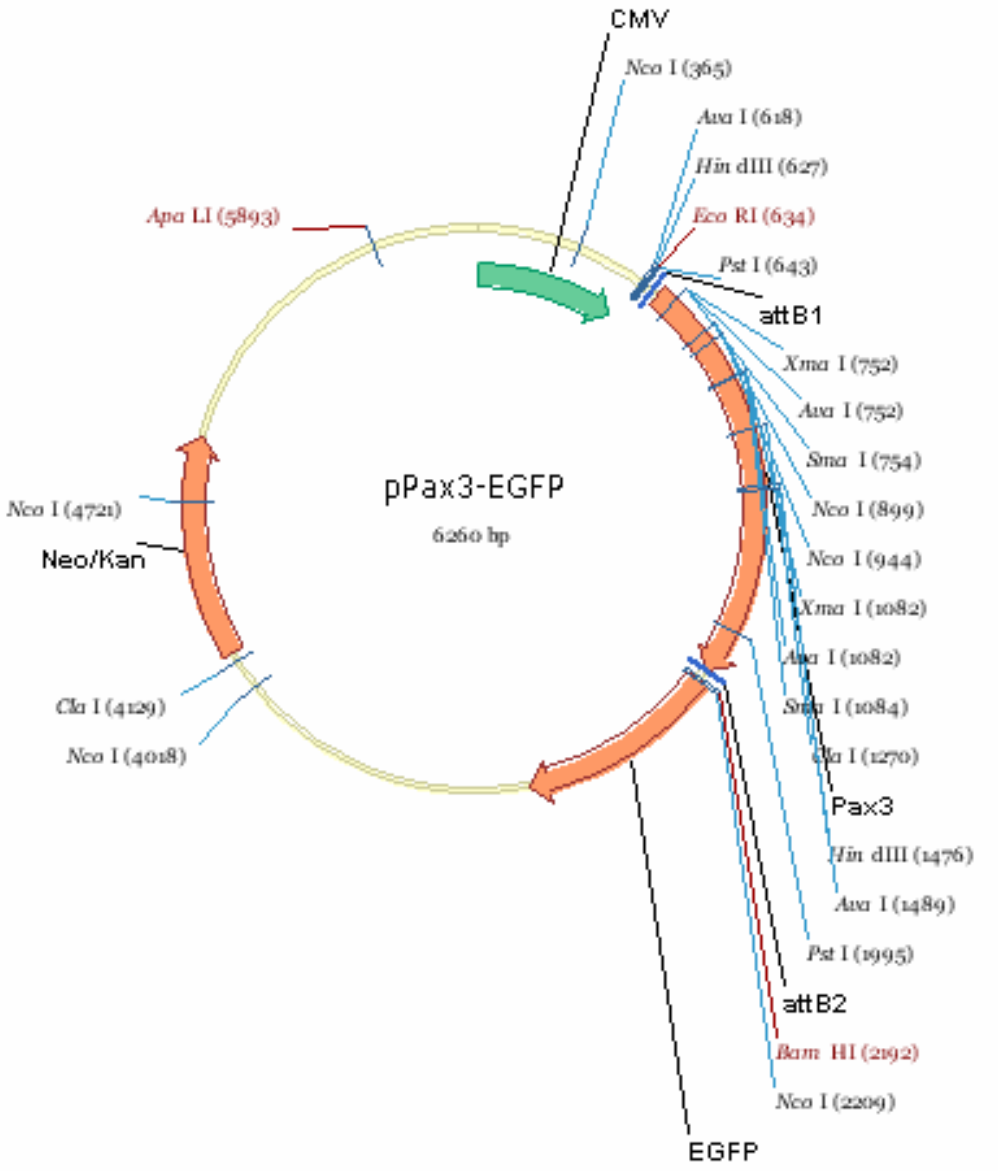
Chloramphenicol 170 µg/mL

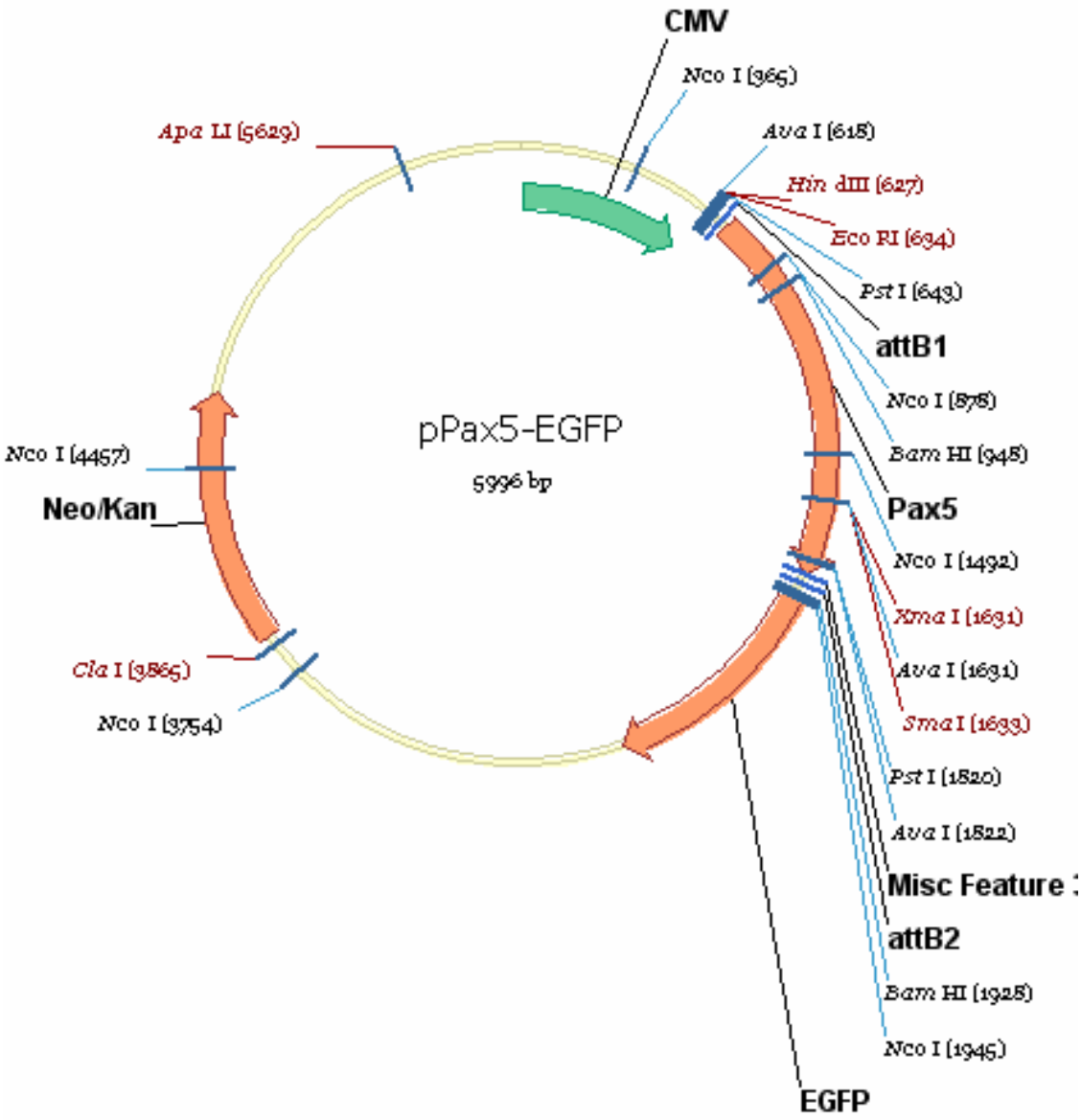
Kanamycin 50 µg/mL

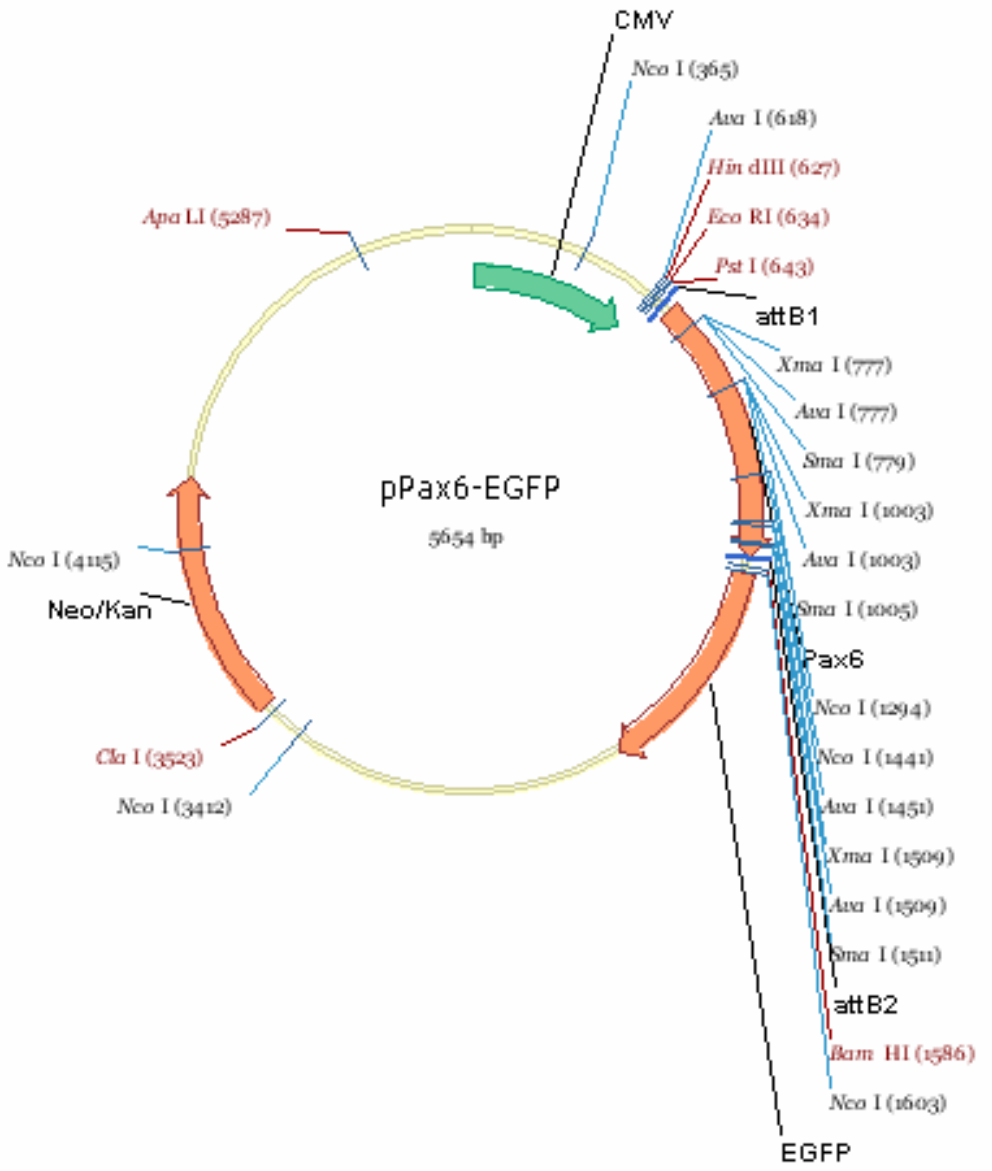
Zeocin 50 µg/mL

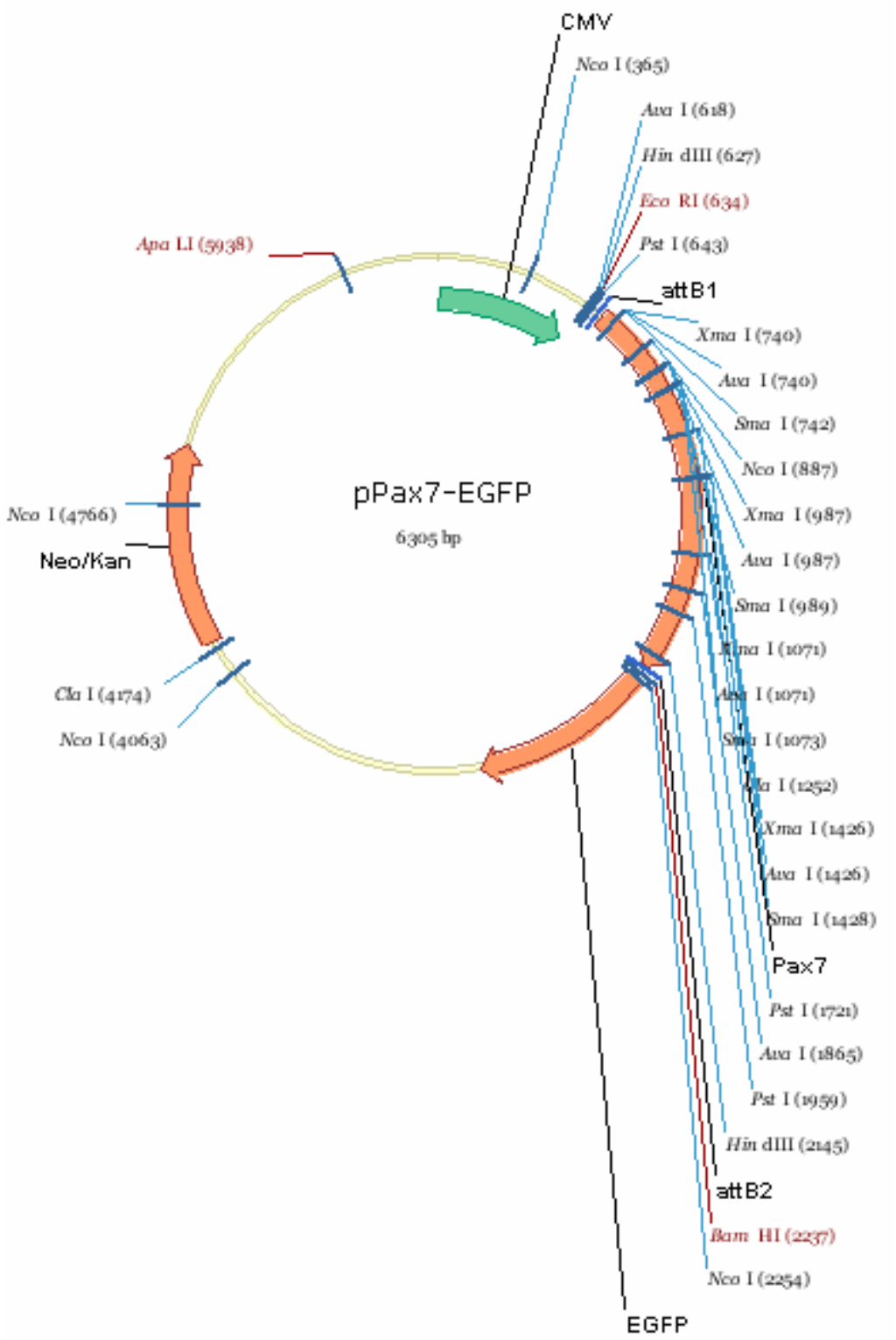
5.3. Plasmid maps

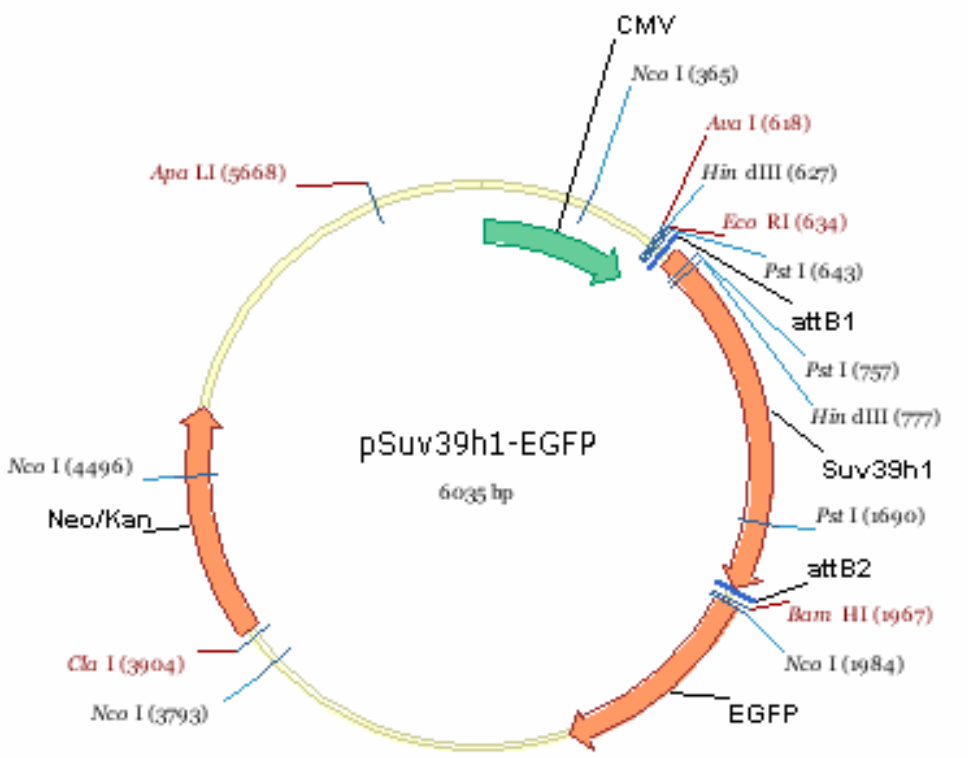
Plasmids for the mammalian expression of the proteins described in the thesis. Clones used for cDNAs of the different genes are the following: Pax3 RIKEN AK012493, Pax6 RIKEN AK139054, Atf2 IMAGE 6839572, Bcl6b RIKEN AK156733, LC1011148 RIKEN AK046745, Ppwd1 RIKEN AK028532, Rbm28 IMAGE 4015830, Wdr67 IMAGE 30016266. Pax5 was kindly given by Meinrad Busslinger. cDNA for Pax7 and Klhl2 was isolated by PCR using respectively cDNA prepared from muscle cells and w9. Constructs expressing Suv39h1, Suv39h2, Suv420h1 and Suv420h2 have been previously cloned from our laboratory (Rea et al., 2000; Schotta et al., 2004).

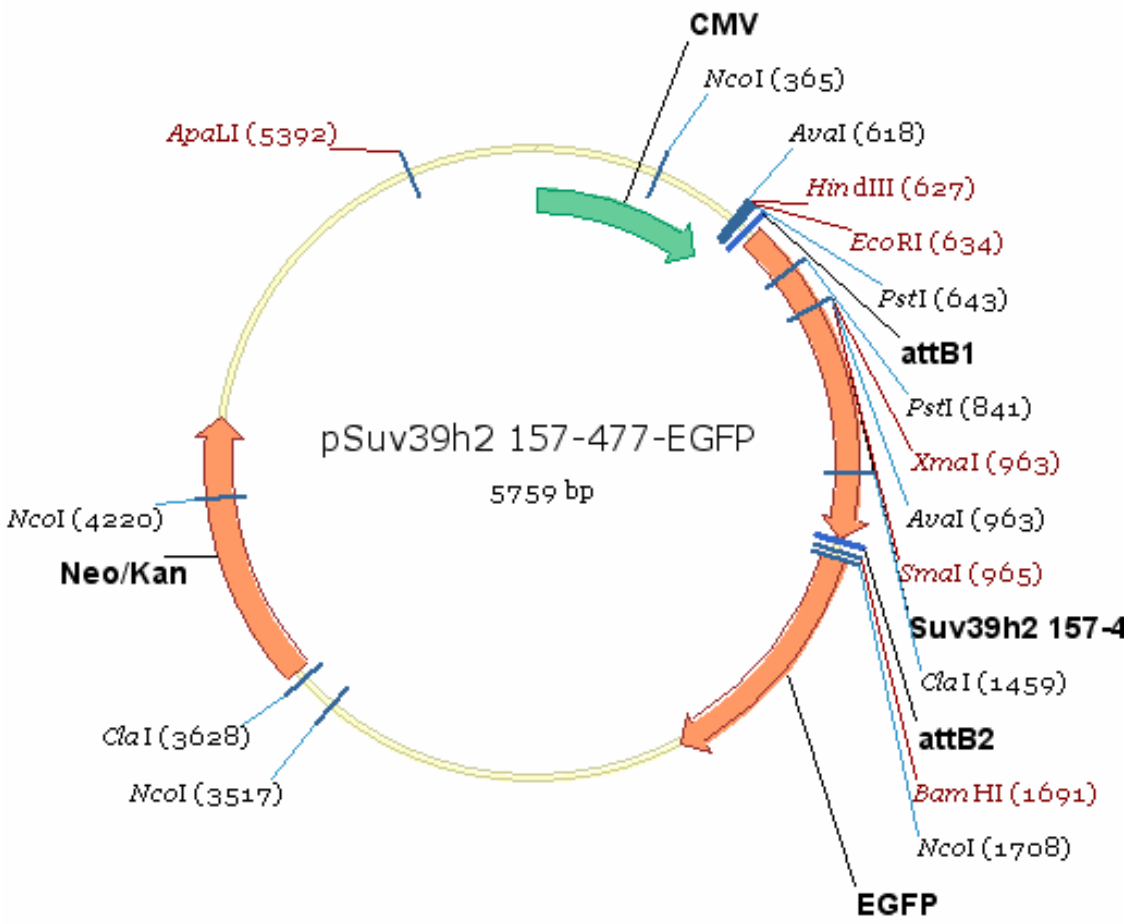


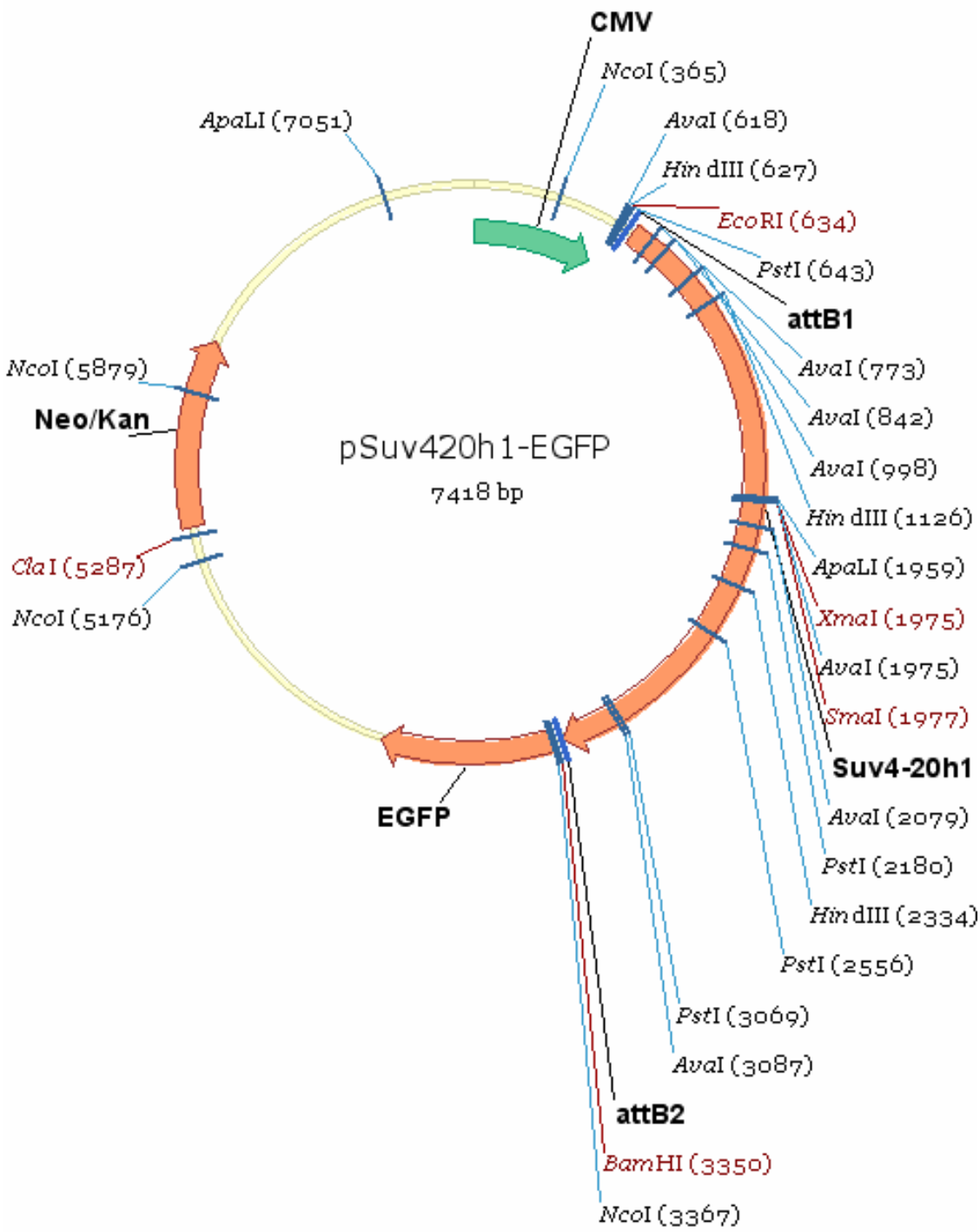

CMV
Nco I (365)
Ava I (618)
Hin dIII (627)
Eco RI (634)
Pst I (643)
attB1
Nco I (878)
Bam HI (948)
Pax5
Nco I (1492)
Xma I (1631)
Ava I (1631)
Sma I (1633)
Pst I (1820)
Ava I (1822)
Misc Feature 1
attB2
Bam HI (1928)
Nco I (1945)
EGFP
Apa LI (5629)
Nco I (4457)
Neo/Kan
Cla I (3865)
Nco I (3754)
pPax5-EGFP
5996 bp

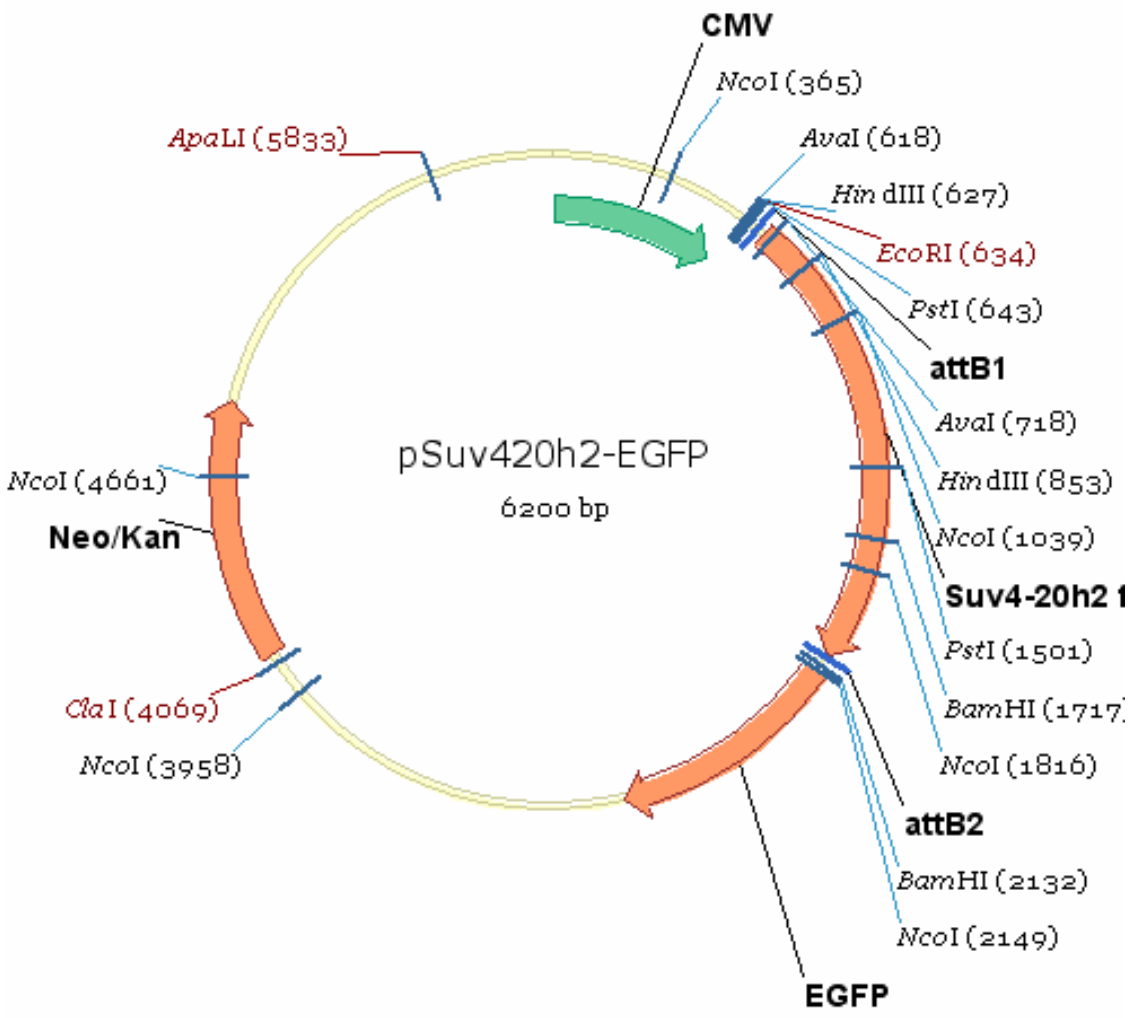

CMV
Nco I (365)
Ava I (618)
Hin dIII (627)
Eco RI (634)
Pst I (643)
attB1
Xma I (777)
Ava I (777)
Sma I (779)
Xma I (1003)
Ava I (1003)
Sma I (1005)
Pax6
Nco I (1294)
Nco I (1441)
Ava I (1451)
Xma I (1509)
Ava I (1509)
Sma I (1511)
attB2
Bam HI (1586)
Nco I (1603)
EGFP
Apa LI (5287)
Nco I (4115)
Neo/Kan
Cla I (3523)
Nco I (3412)
pPax6-EGFP
5654 bp

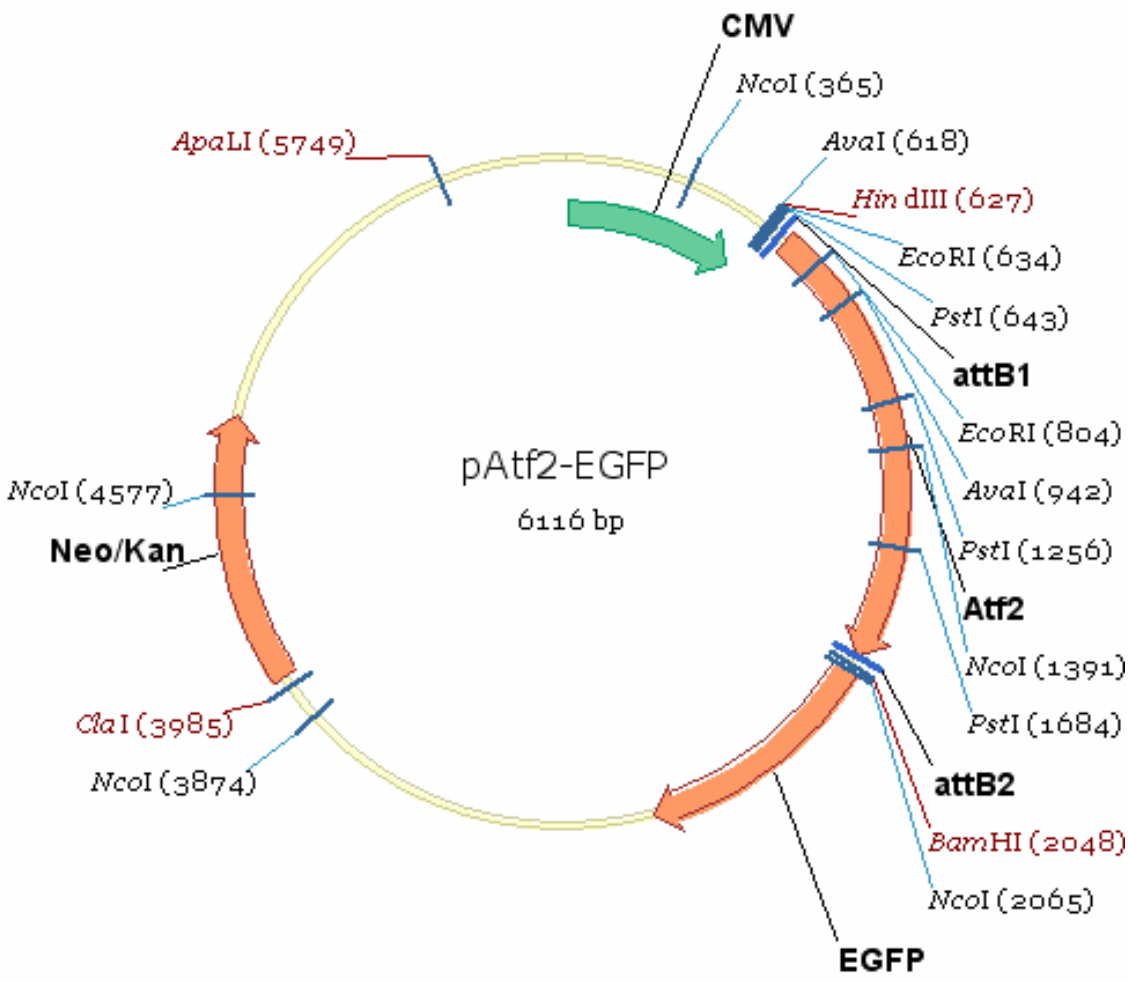

CMV
Nco I (365)
Ava I (618)
Hin dIII (627)
Eco RI (634)
Pst I (643)
Apa LI (5938)
attB1
Xma I (740)
Ava I (740)
Sma I (742)
Nco I (887)
Xma I (987)
Ava I (987)
Sma I (989)
Xma I (1071)
Ava I (1071)
Sma I (1073)
Xba I (1252)
Xma I (1426)
Ava I (1426)
Sma I (1428)
Pax7
Pst I (1721)
Ava I (1865)
Pst I (1959)
Hin dIII (2145)
attB2
Bam HI (2237)
Nco I (2254)
EGFP
Nco I (4766)
Neo/Kan
Cla I (4174)
Nco I (4063)
pPax7-EGFP
6305 bp

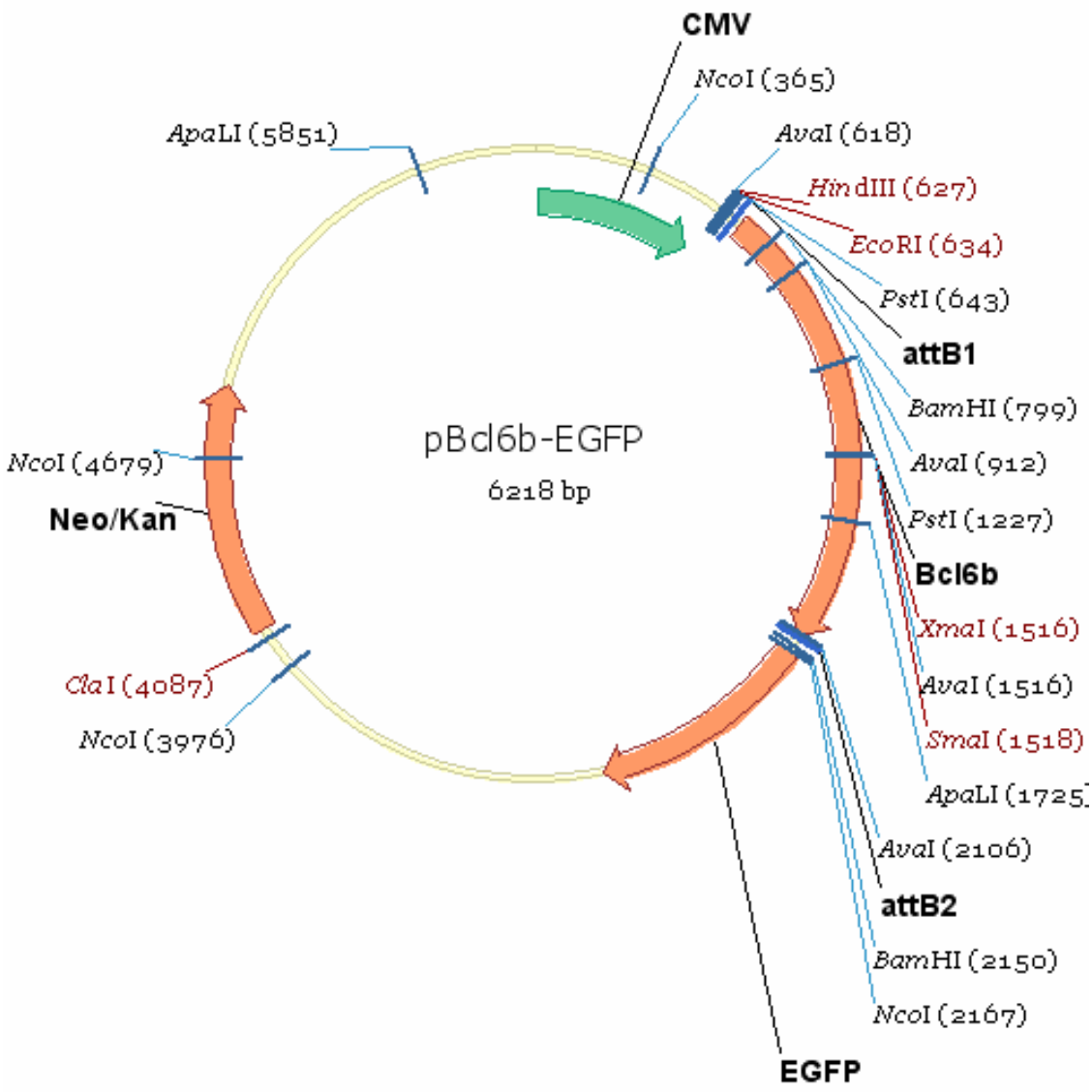

CMV
Nco I (365)
Ava I (618)
Hin dIII (627)
Eco RI (634)
Pst I (643)
attB1
Pst I (757)
Hin dIII (777)
Suv39h1
Pst I (1690)
attB2
Bam HI (1967)
Nco I (1984)
EGFP
Nco I (3793)
Cla I (3904)
Neo/Kan
Nco I (4496)
Apa LI (5668)
pSuv39h1-EGFP
6035 bp

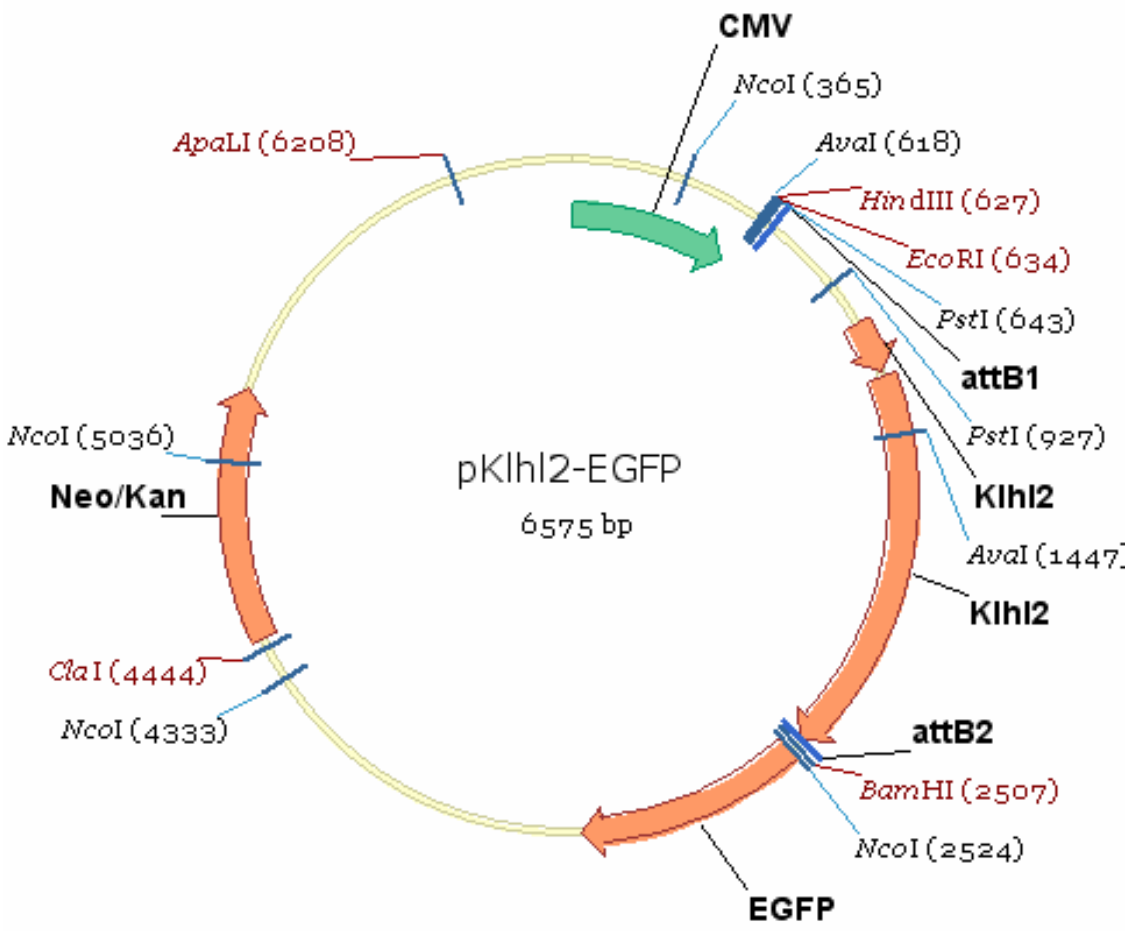

CMV
NcoI (365)
AvaI (618)
HindIII (627)
EcoRI (634)
PstI (643)
attB1
PstI (841)
XmaI (963)
AvaI (963)
SmaI (965)
Suv39h2 157-4
ClaI (1459)
attB2
BamHI (1691)
NcoI (1708)
EGFP
ClaI (3628)
NcoI (3517)
Neo/Kan
NcoI (4220)
ApaLI (5392)
pSuv39h2 157-477-EGFP
5759 bp

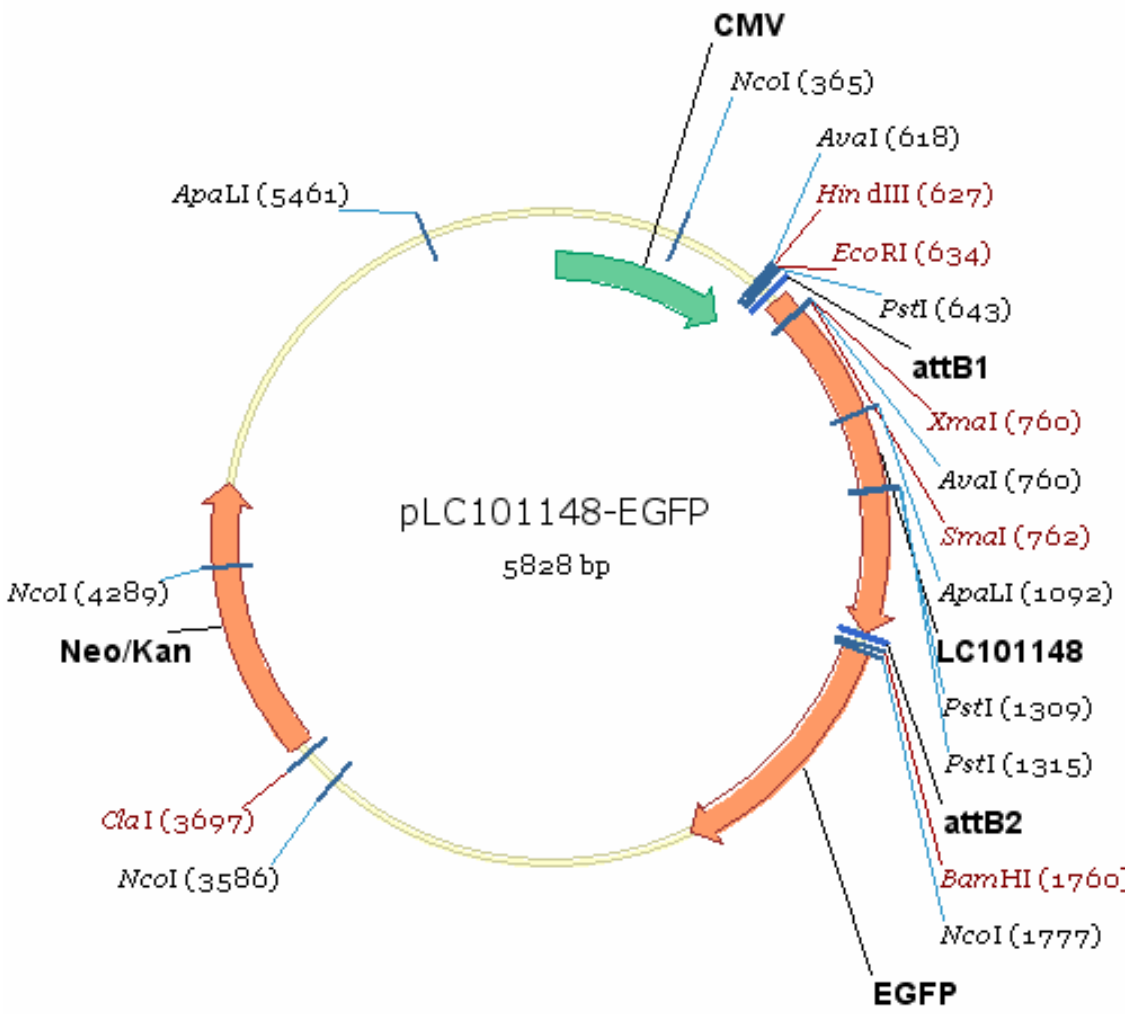

CMV
NcoI (365)
AvaI (618)
Hin dIII (627)
EcoRI (634)
PstI (643)
attB1
AvaI (773)
AvaI (842)
AvaI (998)
Hin dIII (1126)
ApaLI (1959)
XmaI (1975)
AvaI (1975)
SmaI (1977)
Suv4-20h1
AvaI (2079)
PstI (2180)
Hin dIII (2334)
PstI (2556)
PstI (3069)
AvaI (3087)
attB2
BamHI (3350)
NcoI (3367)
EGFP
ClaI (5287)
NcoI (5176)
Neo/Kan
NcoI (5879)
ApaLI (7051)
pSuv420h1-EGFP
7418 bp

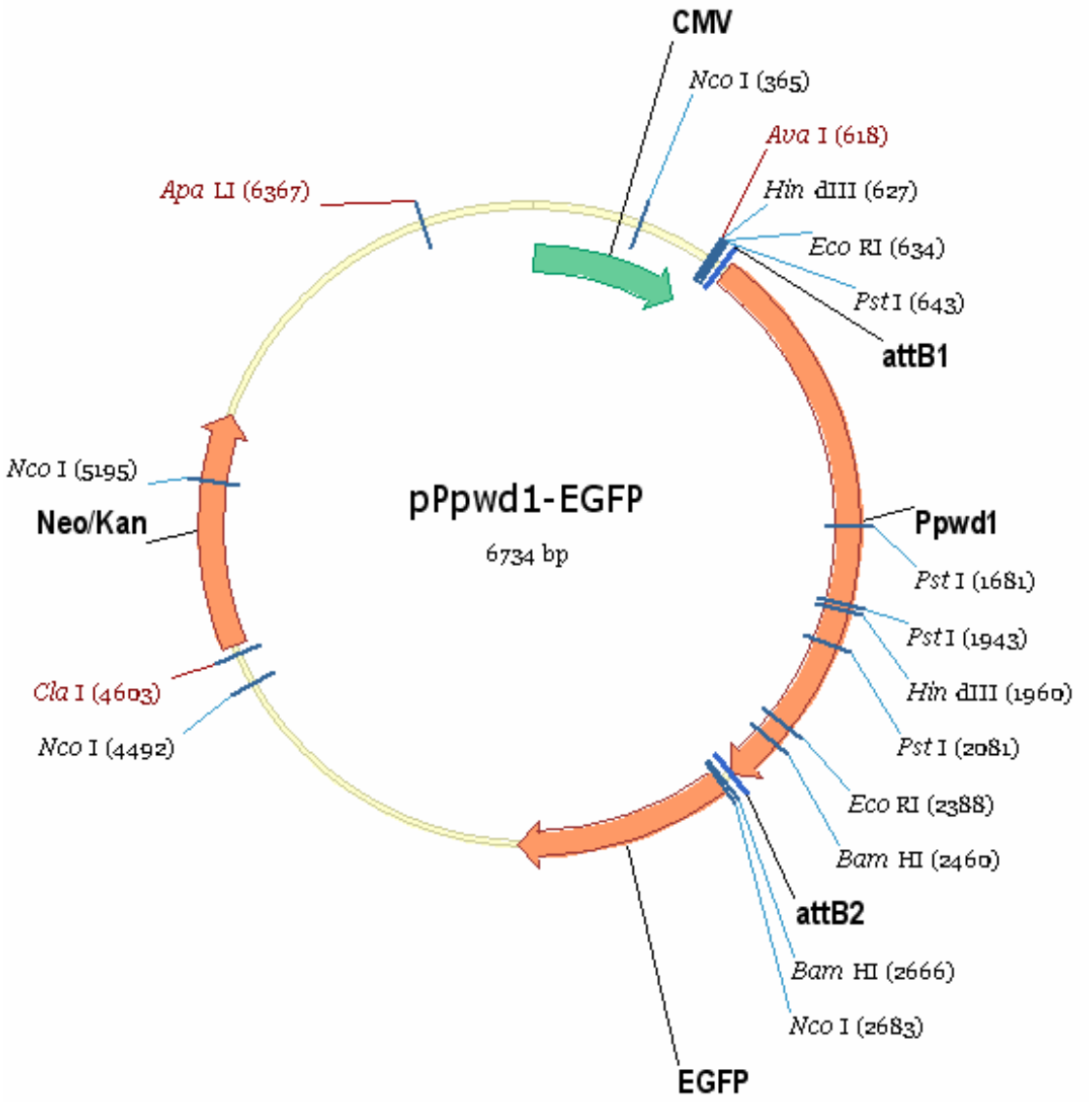

CMV
NcoI (365)
AvaI (618)
HindIII (627)
EcoRI (634)
PstI (643)
attB1
AvaI (718)
HindIII (853)
NcoI (1039)
Suv4-20h2 1
PstI (1501)
BamHI (1717
NcoI (1816)
attB2
BamHI (2132)
NcoI (2149)
EGFP
ApaLI (5833)
NcoI (4661)
Neo/Kan
ClaI (4069)
NcoI (3958)
pSuv420h2-EGFP
6200 bp

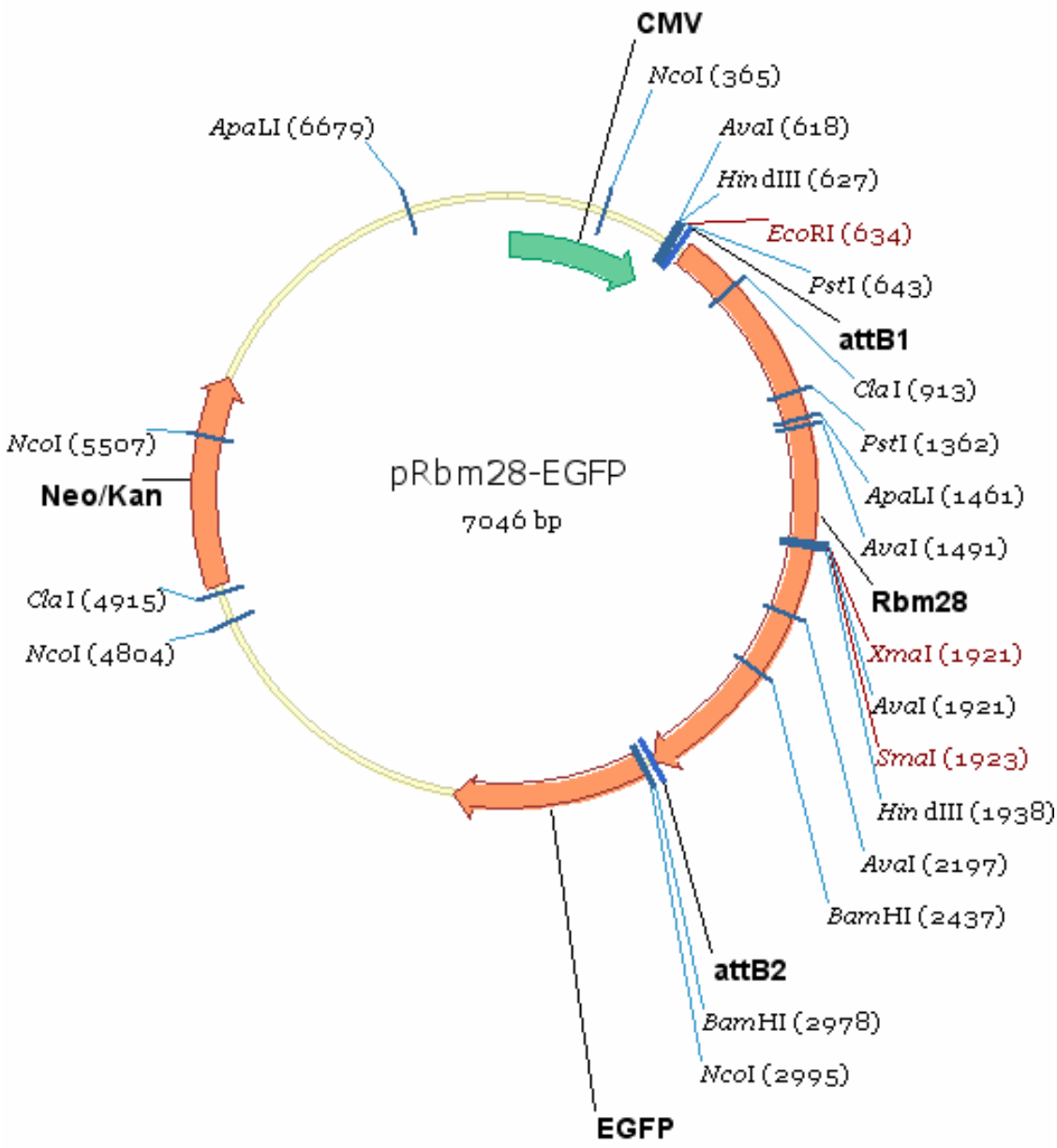

CMV
NcoI (365)
AvaI (618)
HindIII (627)
EcoRI (634)
PstI (643)
attB1
EcoRI (804)
AvaI (942)
PstI (1256)
Atf2
NcoI (1391)
PstI (1684)
attB2
BamHI (2048)
NcoI (2065)
EGFP
ApaLI (5749)
NcoI (4577)
Neo/Kan
ClaI (3985)
NcoI (3874)
pAtf2-EGFP
6116 bp

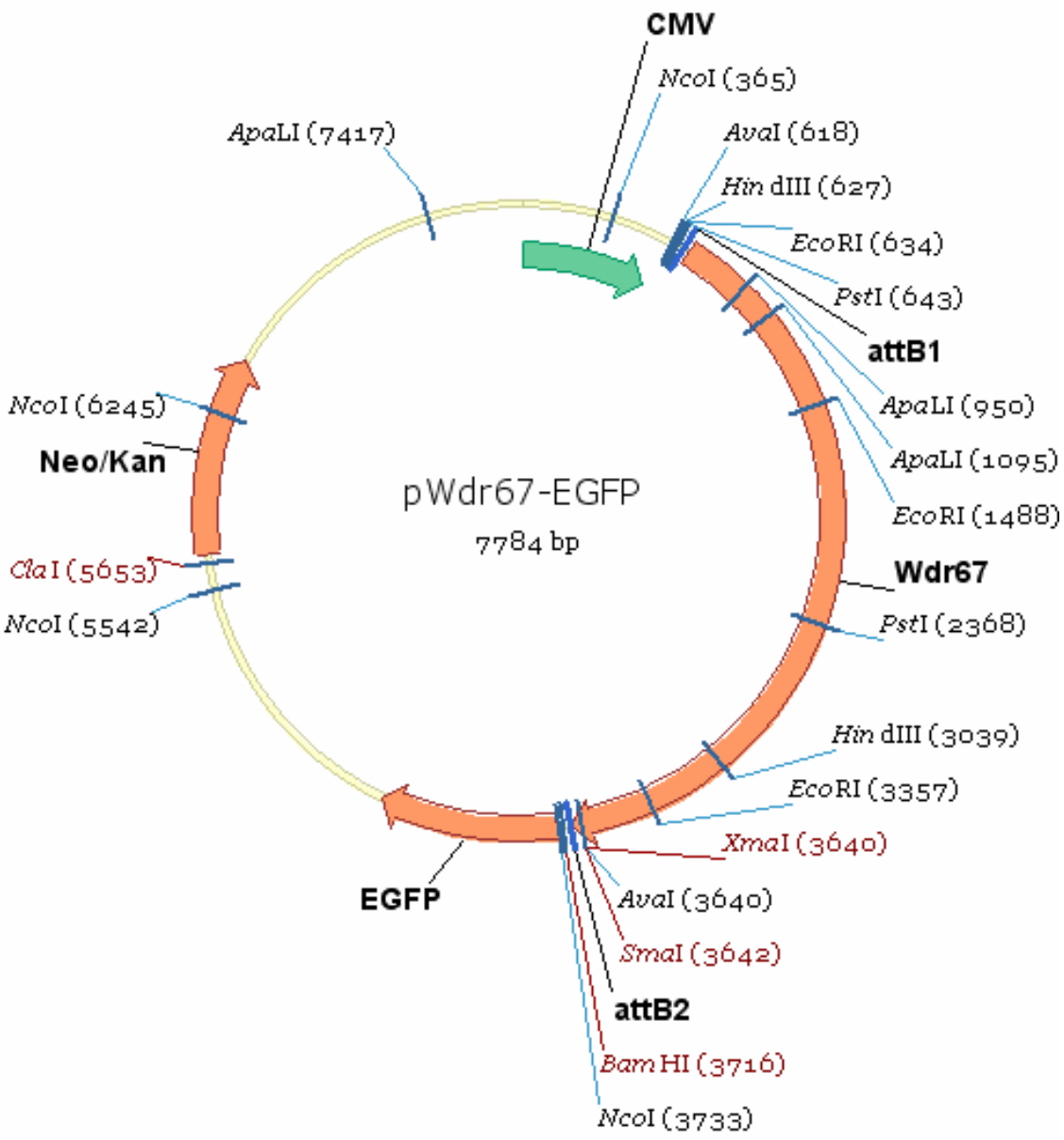

CMV
NcoI (365)
AvaI (618)
HindIII (627)
EcoRI (634)
PstI (643)
attB1
BamHI (799)
AvaI (912)
PstI (1227)
Bcl6b
XmaI (1516)
AvaI (1516)
SmaI (1518)
ApaLI (1725
AvaI (2106)
attB2
BamHI (2150)
NcoI (2167)
EGFP
ApaLI (5851)
NcoI (4679)
Neo/Kan
ClaI (4087)
NcoI (3976)
pBcl6b-EGFP
6218 bp


CMV
NcoI (365)
AvaI (618)
HindIII (627)
EcoRI (634)
PstI (643)
attB1
PstI (927)
KIhl2
AvaI (1447)
KIhl2
attB2
BamHI (2507)
NcoI (2524)
EGFP
ClaI (4444)
NcoI (4333)
Neo/Kan
NcoI (5036)
ApaLI (6208)
pKIhl2-EGFP
6575 bp

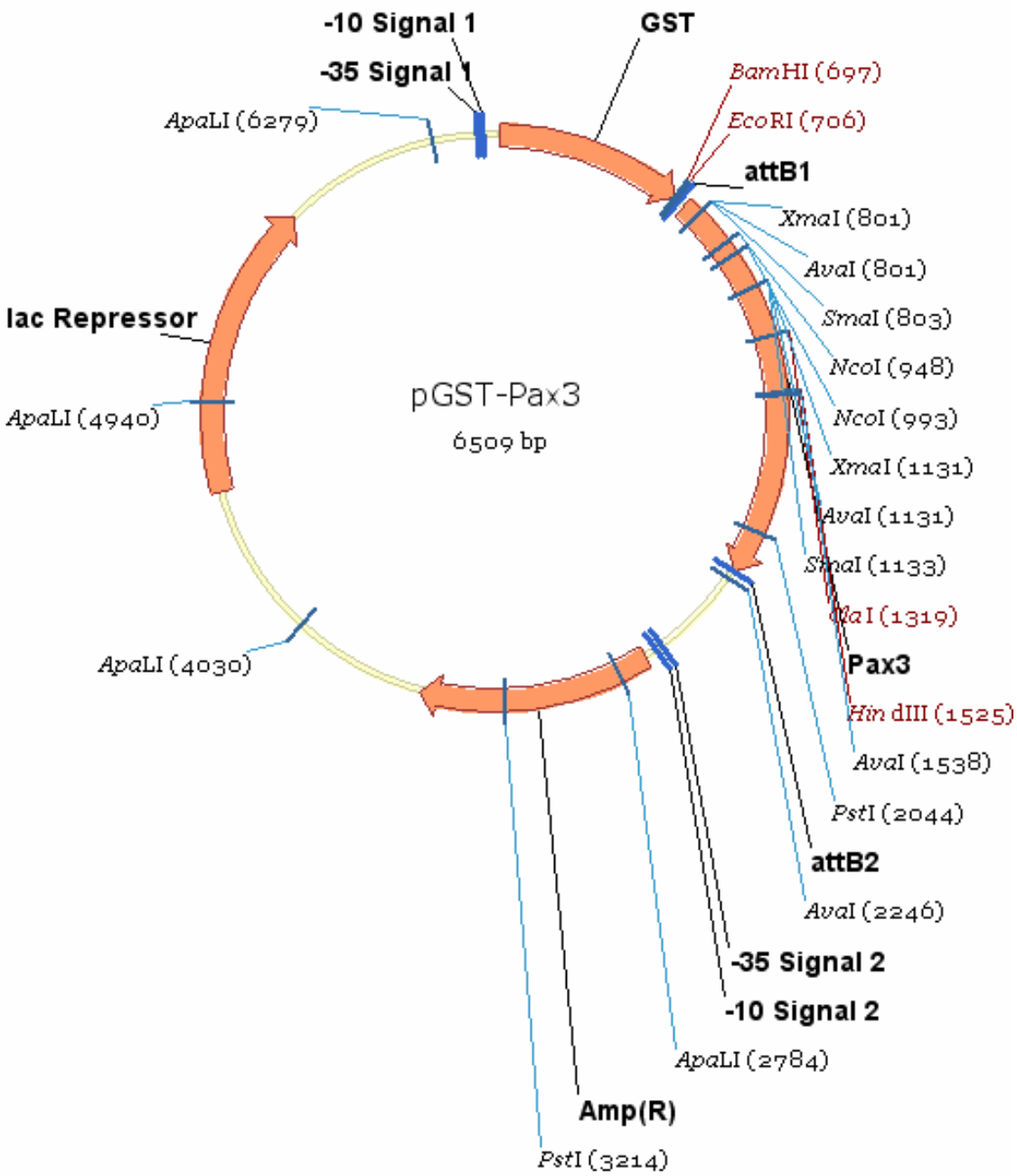

CMV
NcoI (365)
AvaI (618)
HindIII (627)
EcoRI (634)
PstI (643)
attB1
XmaI (760)
AvaI (760)
SmaI (762)
ApaLI (1092)
LC101148
PstI (1309)
PstI (1315)
attB2
BamHI (1760)
NcoI (1777)
EGFP
ApaLI (5461)
pLC101148-EGFP
5828 bp
NcoI (4289)
Neo/Kan
ClaI (3697)
NcoI (3586)

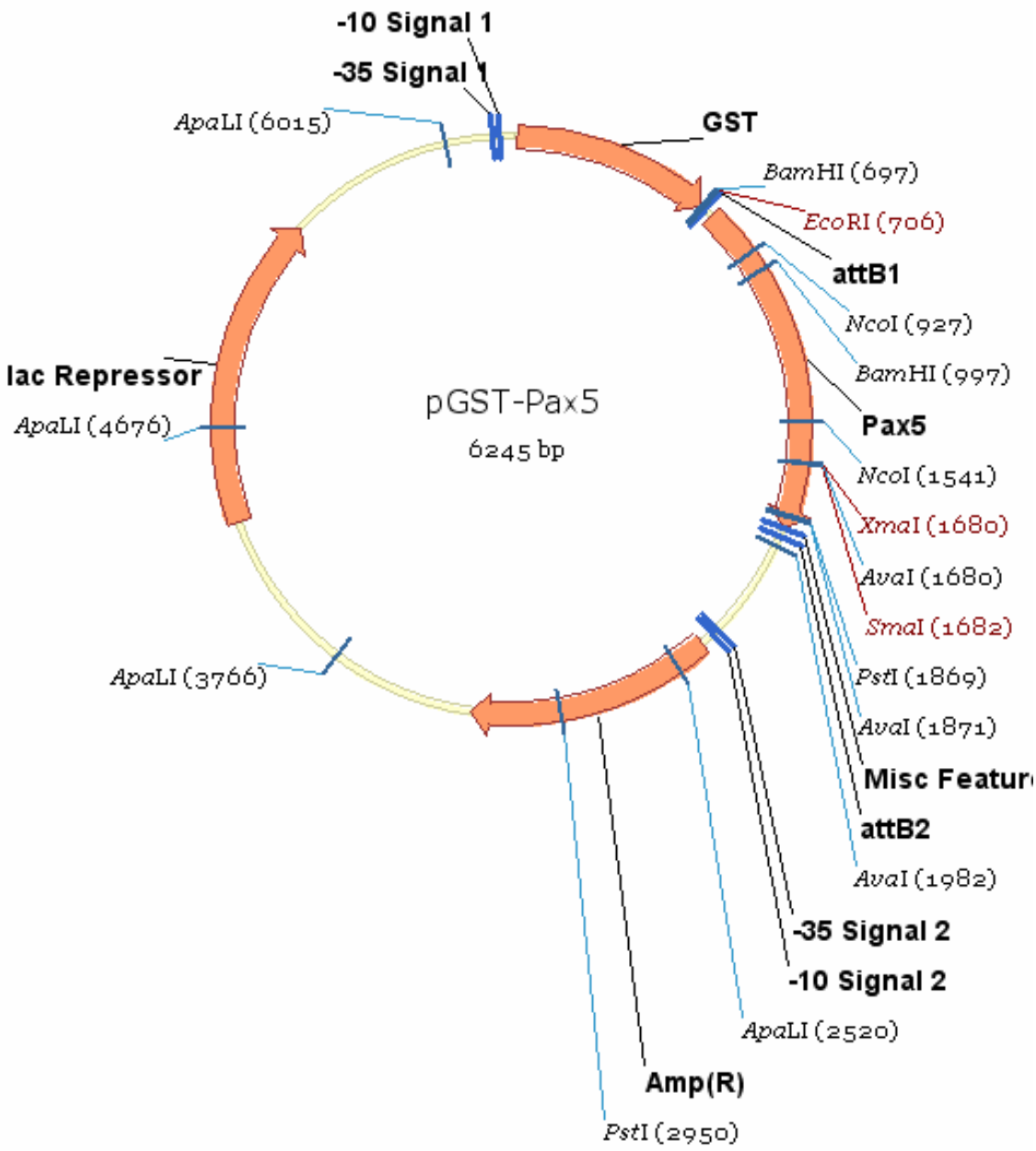

CMV
Nco I (365)
Ava I (618)
Hin dIII (627)
Eco RI (634)
Pst I (643)
attB1
Apa LI (6367)
pPpwd1-EGFP
6734 bp
Ppwd1
Pst I (1681)
Pst I (1943)
Hin dIII (1960)
Pst I (2081)
Eco RI (2388)
Bam HI (2460)
attB2
Bam HI (2666)
Nco I (2683)
EGFP
Nco I (5195)
Neo/Kan
Cla I (4603)
Nco I (4492)

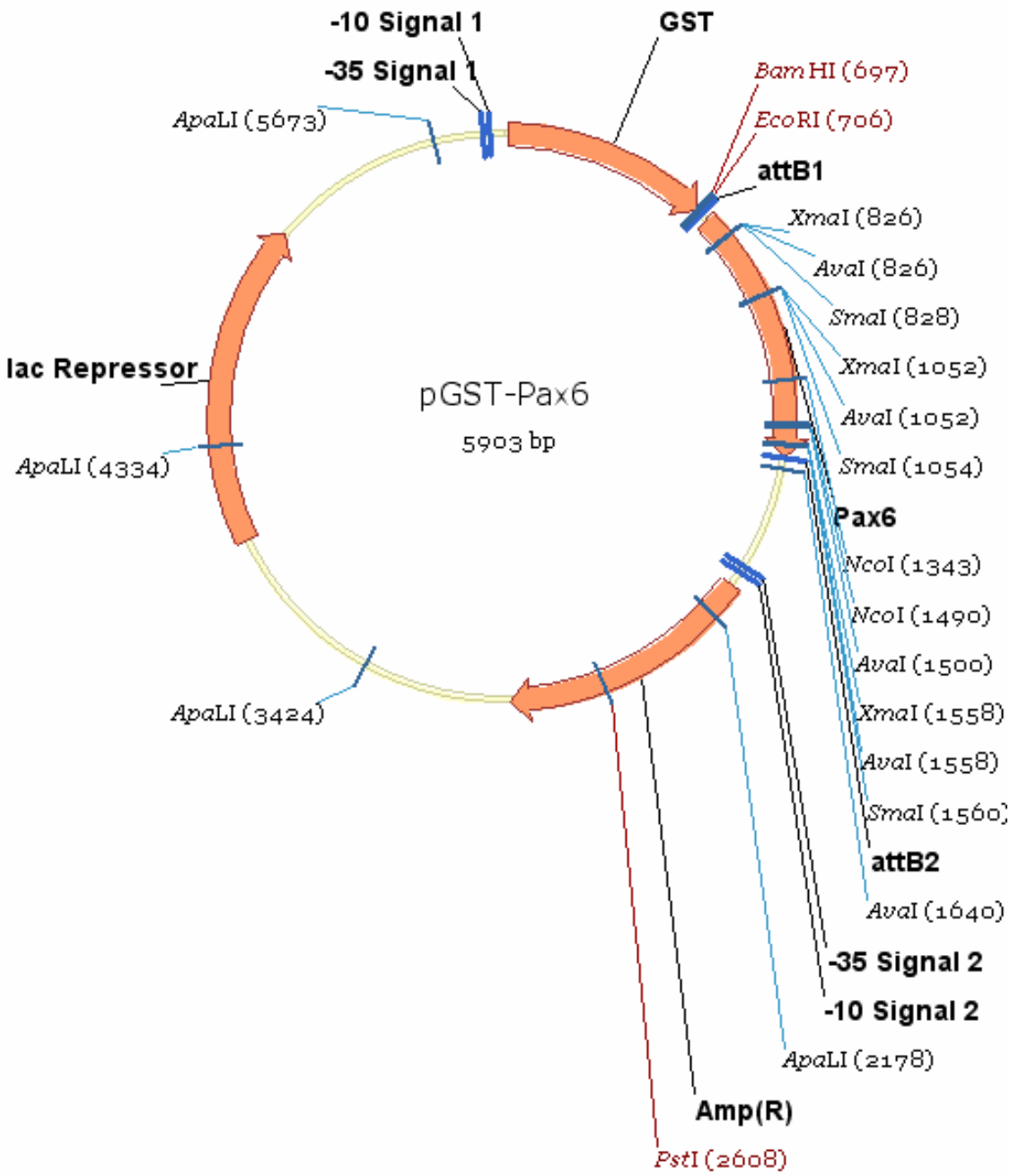

CMV
NcoI (365)
AvaI (618)
HindIII (627)
EcoRI (634)
PstI (643)
attB1
ClaI (913)
PstI (1362)
ApaLI (1461)
AvaI (1491)
Rbm28
XmaI (1921)
AvaI (1921)
SmaI (1923)
HindIII (1938)
AvaI (2197)
BamHI (2437)
attB2
BamHI (2978)
NcoI (2995)
EGFP
ApaLI (6679)
NcoI (5507)
Neo/Kan
ClaI (4915)
NcoI (4804)
pRbm28-EGFP
7046 bp

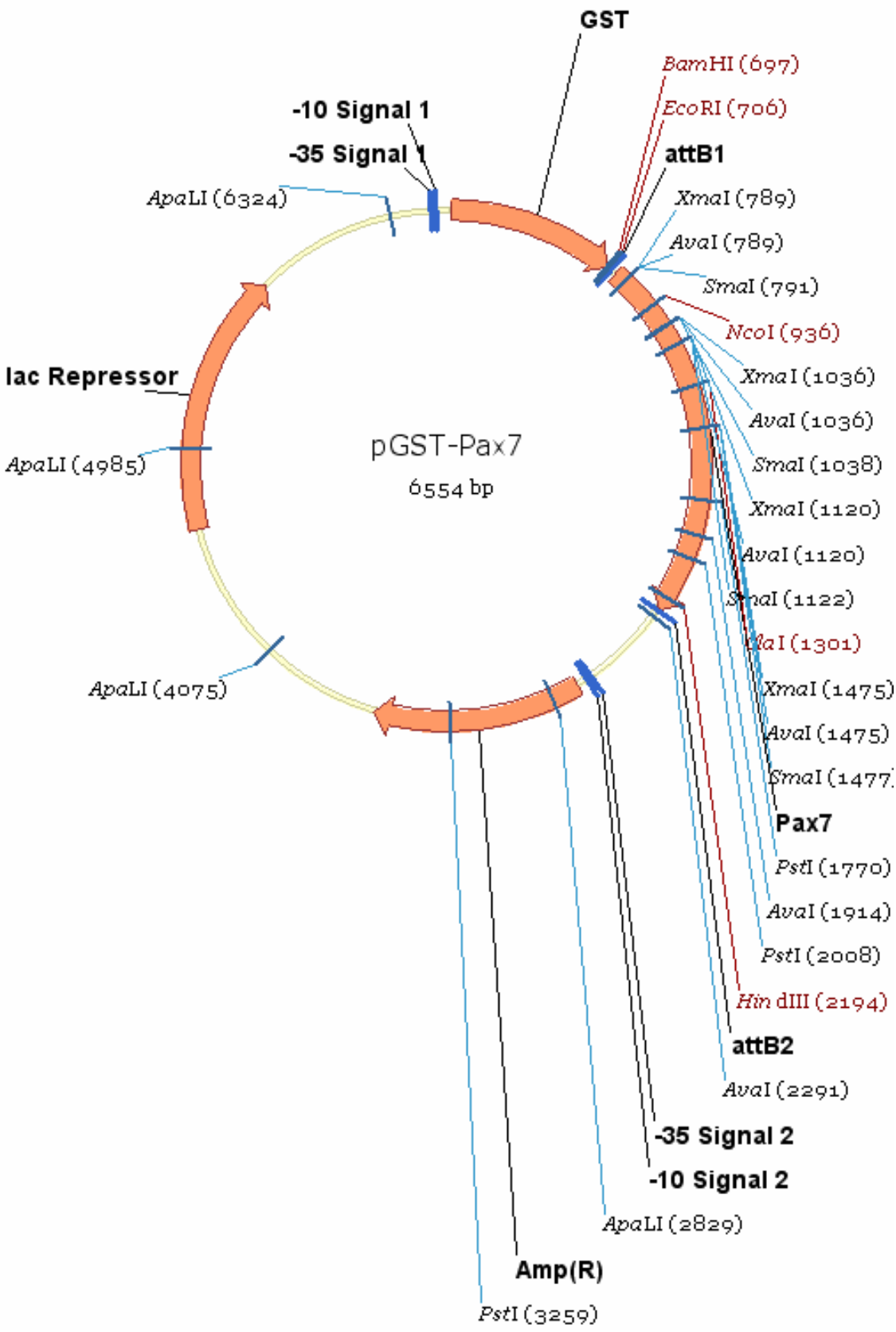

CMV
NcoI (365)
ApaLI (7417)
AvaI (618)
Hin dIII (627)
EcoRI (634)
PstI (643)
attB1
ApaLI (950)
ApaLI (1095)
EcoRI (1488)
Wdr67
PstI (2368)
Hin dIII (3039)
EcoRI (3357)
XmaI (3640)
AvaI (3640)
SmaI (3642)
attB2
BamHI (3716)
NcoI (3733)
EGFP
ClaI (5653)
NcoI (5542)
Neo/Kan
NcoI (6245)
pWdr67-EGFP
7784 bp

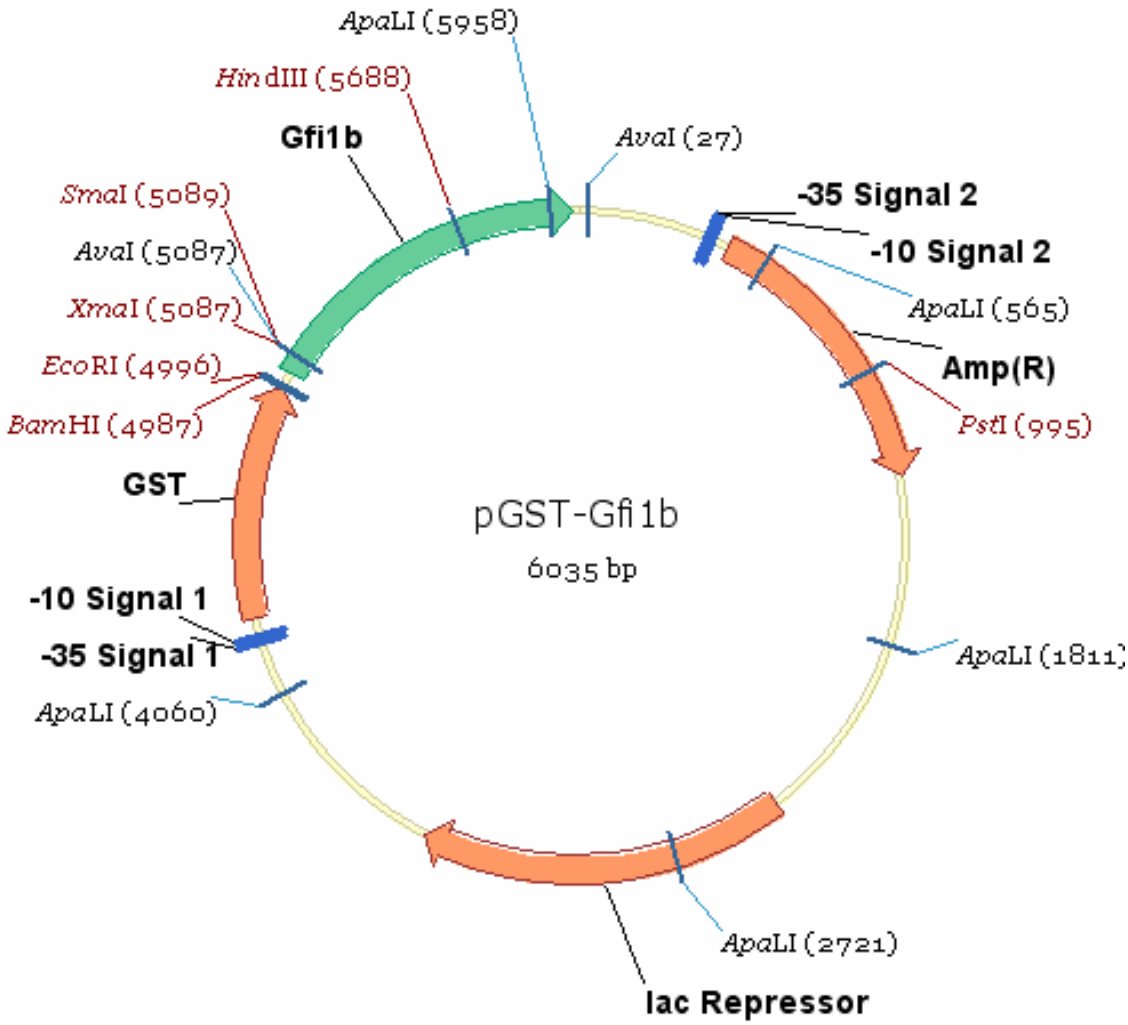
Plasmids for the recombinant expression of the proteins described in this thesis.

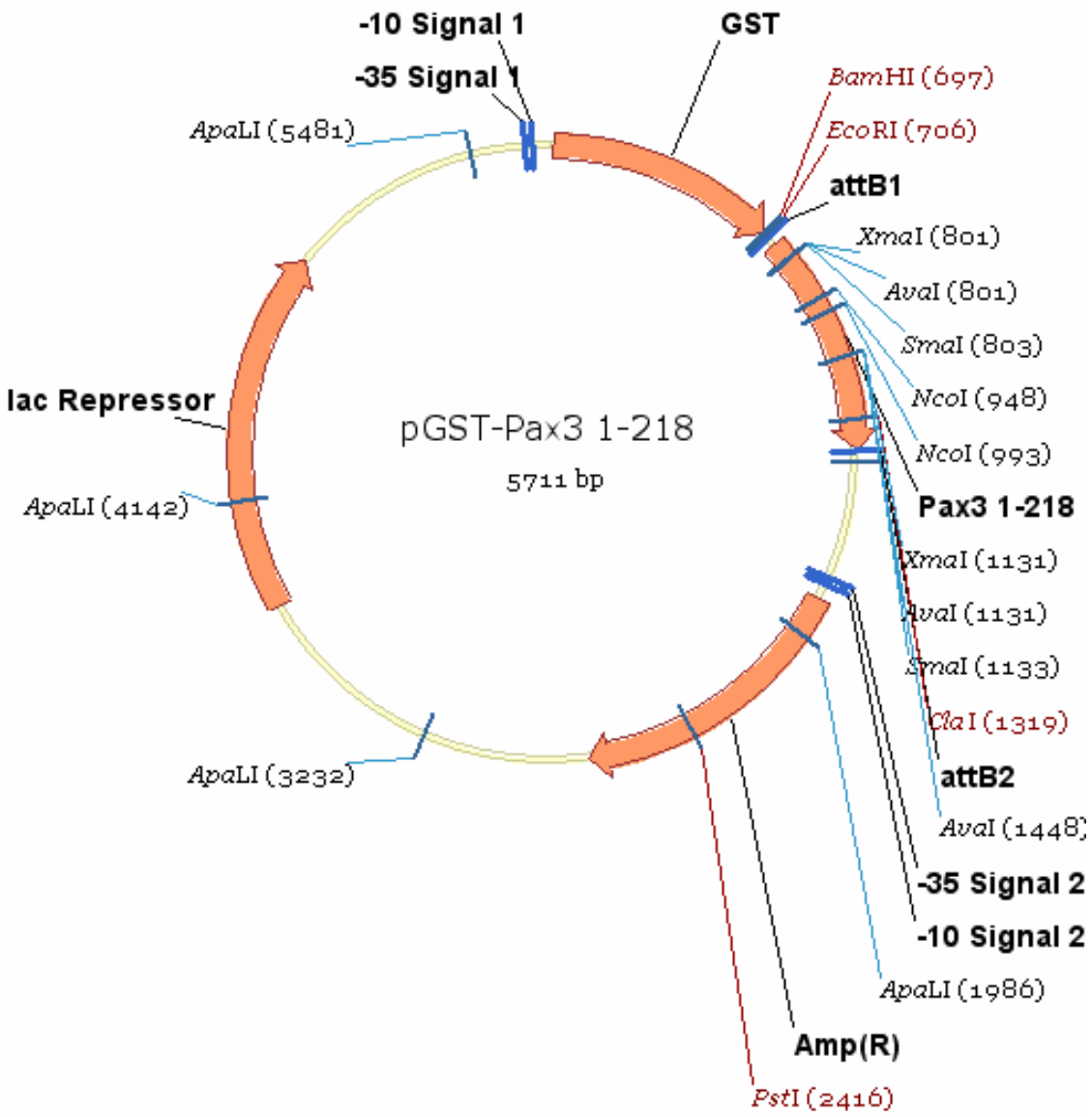


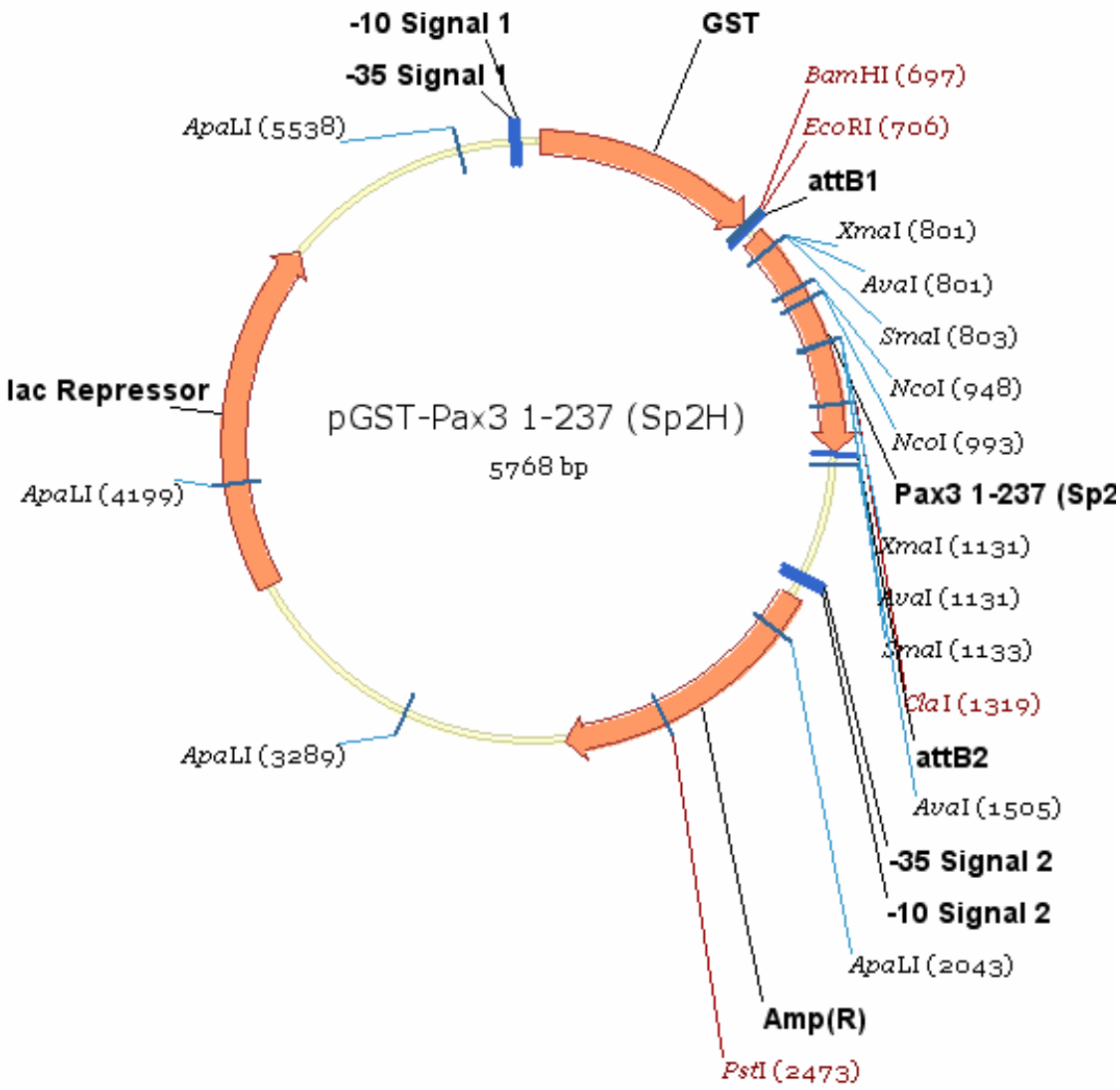

-10 Signal 1
-35 Signal 1
ApaLI (6015)
GST
BamHI (697)
EcoRI (706)
attB1
NcoI (927)
BamHI (997)
Pax5
NcoI (1541)
XmaI (1680)
AvaI (1680)
SmaI (1682)
PstI (1869)
AvaI (1871)
Misc Featur
attB2
AvaI (1982)
-35 Signal 2
-10 Signal 2
ApaLI (2520)
Amp(R)
PstI (2950)
ApaLI (3766)
lac Repressor
ApaLI (4676)
pGST-Pax5
6245 bp

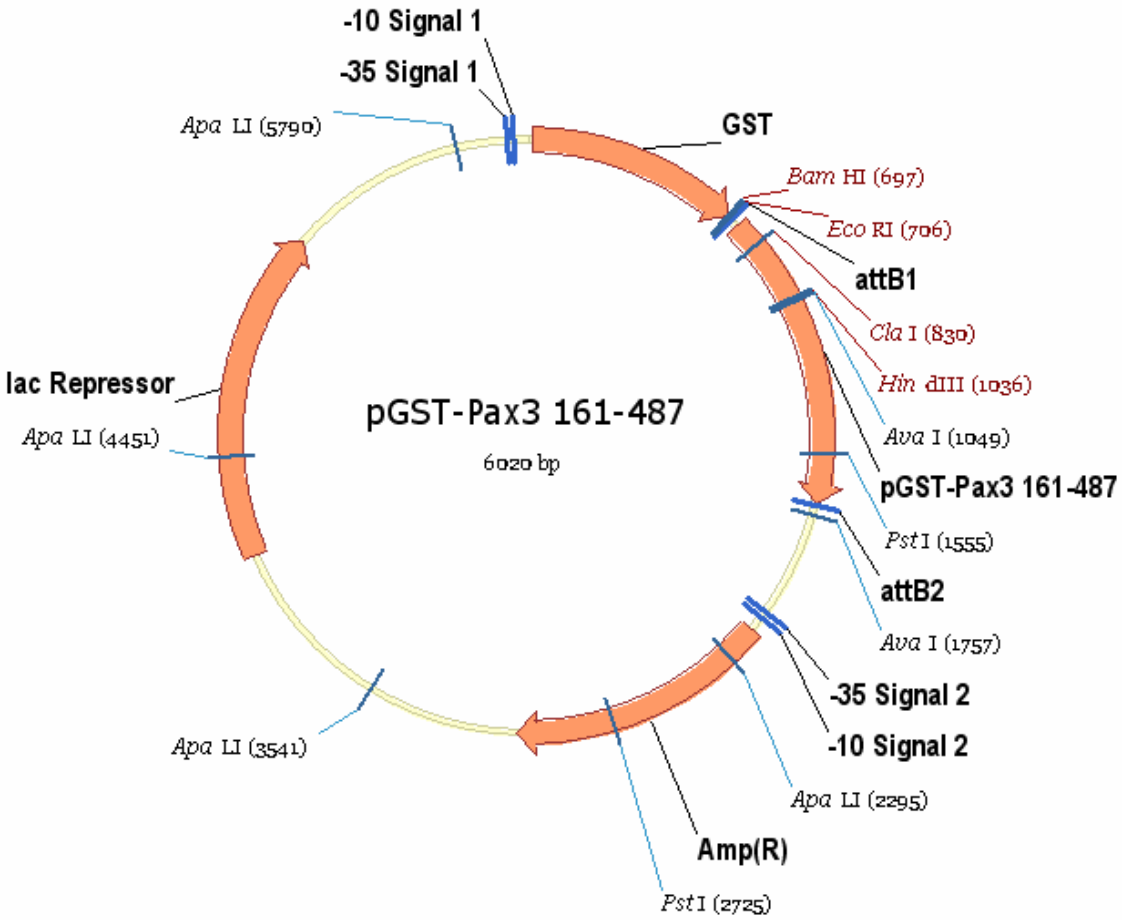

-10 Signal 1
-35 Signal 1
GST
BamHI (697)
EcoRI (706)
attB1
XmaI (826)
AvaI (826)
SmaI (828)
XmaI (1052)
AvaI (1052)
SmaI (1054)
Pax6
NcoI (1343)
NcoI (1490)
AvaI (1500)
XmaI (1558)
AvaI (1558)
SmaI (1560)
attB2
AvaI (1640)
-35 Signal 2
-10 Signal 2
ApaLI (2178)
Amp(R)
PstI (2608)
ApaLI (3424)
ApaLI (4334)
lac Repressor
ApaLI (5673)
pGST-Pax6
5903 bp

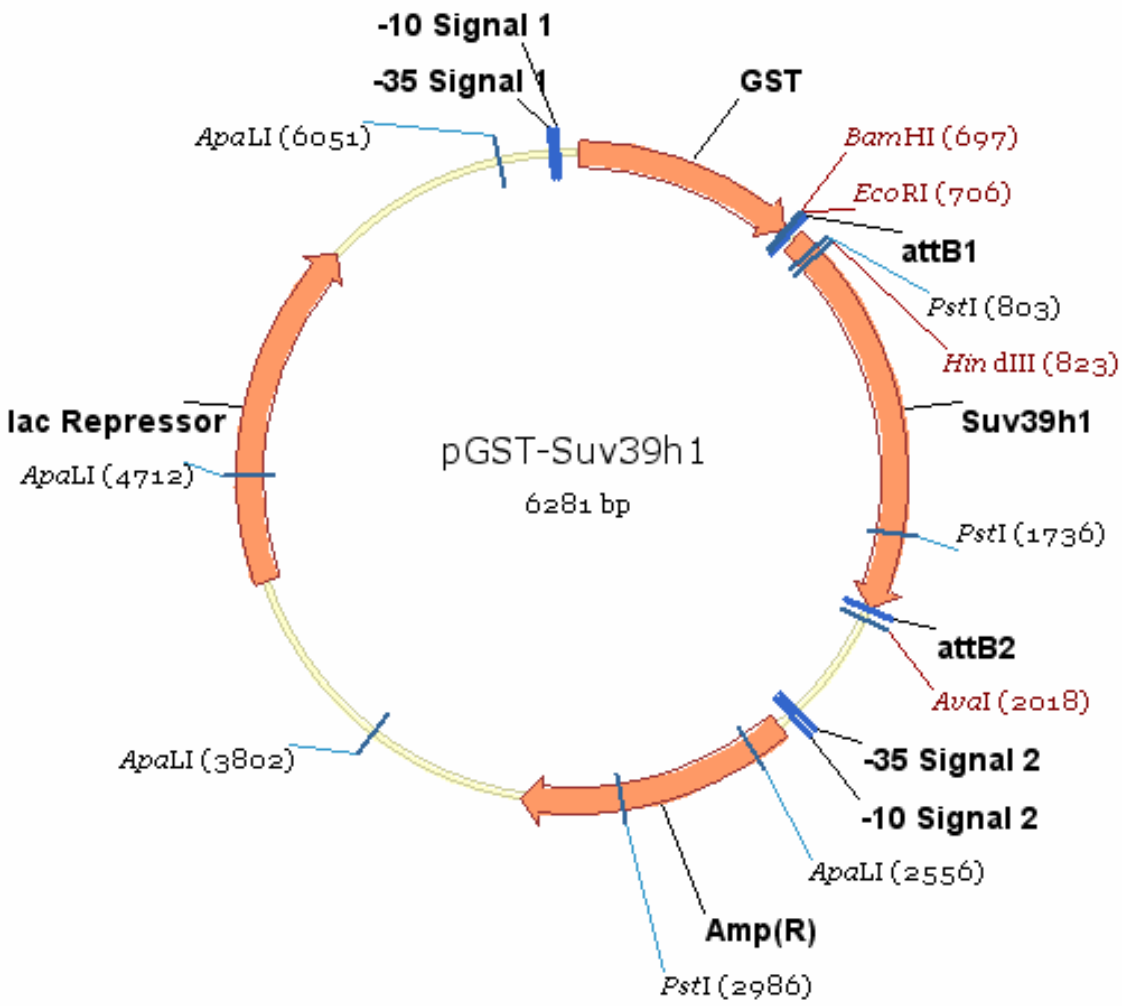

GST
BamHI (697)
EcoRI (706)
-10 Signal 1
-35 Signal 1
attB1
ApaLI (6324)
XmaI (789)
AvaI (789)
SmaI (791)
NcoI (936)
lac Repressor
XmaI (1036)
AvaI (1036)
pGST-Pax7
6554 bp
SmaI (1038)
ApaLI (4985)
XmaI (1120)
AvaI (1120)
SmaI (1122)
XbaI (1301)
XmaI (1475)
AvaI (1475)
SmaI (1477
ApaLI (4075)
Pax7
PstI (1770)
AvaI (1914)
PstI (2008)
HindIII (2194)
attB2
AvaI (2291)
-35 Signal 2
-10 Signal 2
ApaLI (2829)
Amp(R)
PstI (3259)

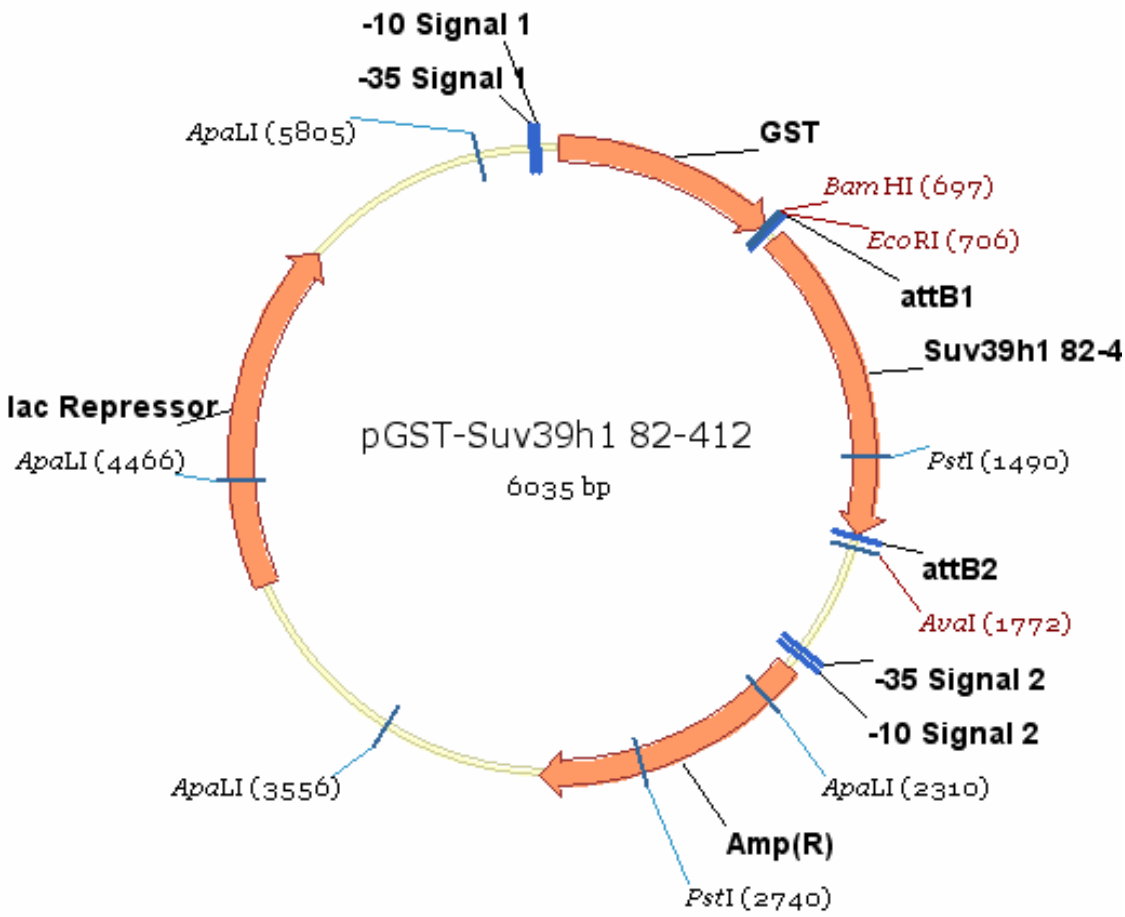

ApaLI (5958)
HindIII (5688)
Gfi1b
AvaI (27)
-35 Signal 2
SmaI (5089)
AvaI (5087)
-10 Signal 2
XmaI (5087)
ApaLI (565)
EcoRI (4996)
Amp(R)
BamHI (4987)
PstI (995)
GST
pGST-Gfi1b
6035 bp
-10 Signal 1
-35 Signal 1
ApaLI (1811)
ApaLI (4060)
ApaLI (2721)
lac Repressor

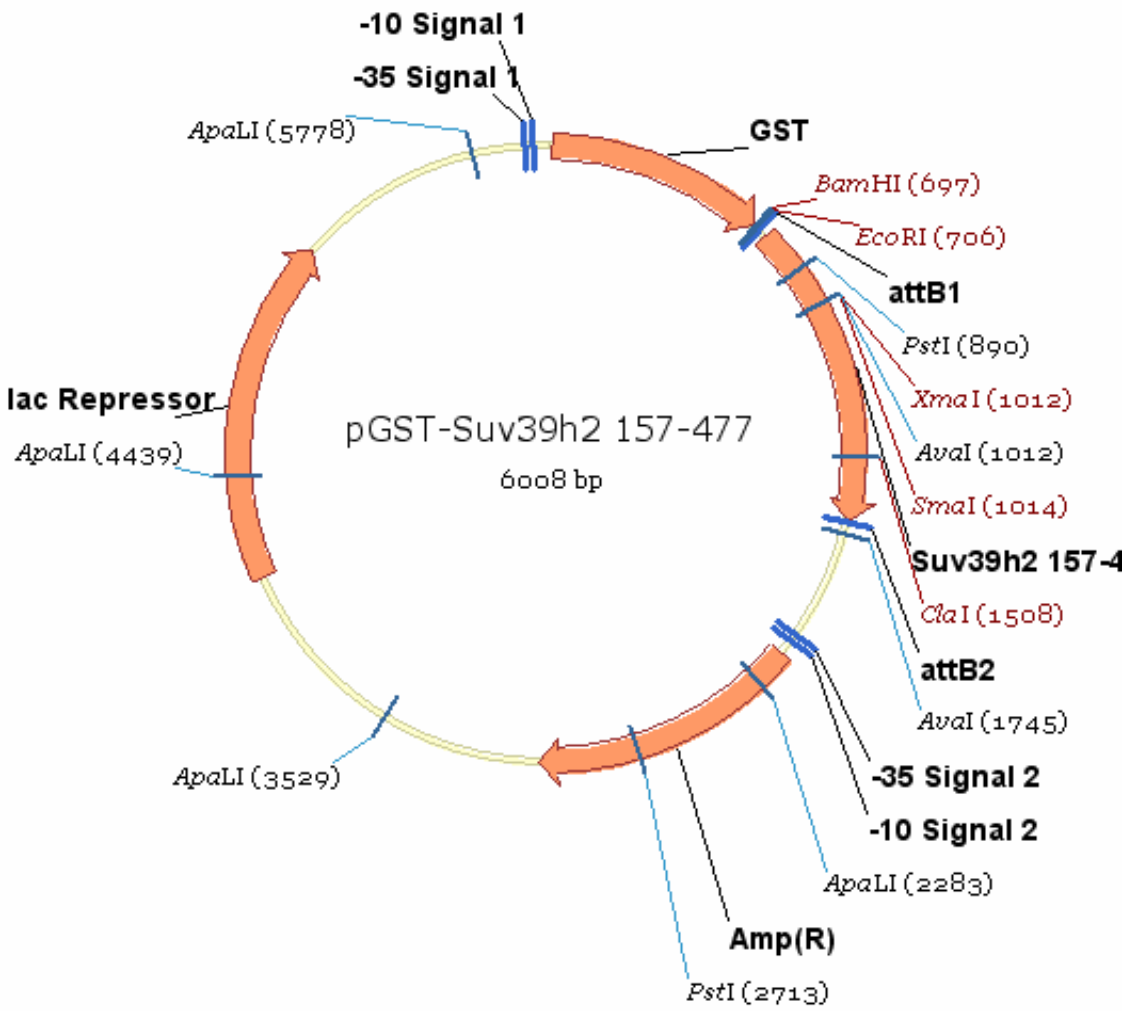

-10 Signal 1
-35 Signal 1
GST
BamHI (697)
EcoRI (706)
attB1
XmaI (801)
AvaI (801)
SmaI (803)
NcoI (948)
NcoI (993)
Pax3 1-218
XmaI (1131)
AvaI (1131)
SmaI (1133)
ClaI (1319)
attB2
AvaI (1448)
-35 Signal 2
-10 Signal 2
ApaLI (1986)
Amp(R)
PstI (2416)
ApaLI (3232)
ApaLI (4142)
lac Repressor
ApaLI (5481)
pGST-Pax3 1-218
5711 bp


-10 Signal 1
-35 Signal 1
GST
BamHI (697)
EcoRI (706)
attB1
ApaLI (5538)
XmaI (801)
AvaI (801)
SmaI (803)
NcoI (948)
NcoI (993)
lac Repressor
pGST-Pax3 1-237 (Sp2H)
5768 bp
Pax3 1-237 (Sp2
ApaLI (4199)
XmaI (1131)
AvaI (1131)
SmaI (1133)
ClaI (1319)
attB2
ApaLI (3289)
AvaI (1505)
-35 Signal 2
-10 Signal 2
ApaLI (2043)
Amp(R)
PstI (2473)


-10 Signal 1
-35 Signal 1
Apa LI (5790)
GST
Bam HI (697)
Eco RI (706)
attB1
Cla I (830)
Hin dIII (1036)
Ava I (1049)
pGST-Pax3 161-487
Pst I (1555)
attB2
Ava I (1757)
-35 Signal 2
-10 Signal 2
Apa LI (2295)
Amp(R)
Pst I (2725)
Apa LI (3541)
lac Repressor
Apa LI (4451)
pGST-Pax3 161-487
6020 bp


-10 Signal 1
-35 Signal 1
GST
ApaLI (6051)
BamHI (697)
EcoRI (706)
attB1
PstI (803)
Hin dIII (823)
Suv39h1
lac Repressor
pGST-Suv39h1
6281 bp
ApaLI (4712)
PstI (1736)
attB2
AvaI (2018)
-35 Signal 2
-10 Signal 2
ApaLI (3802)
ApaLI (2556)
Amp(R)
PstI (2986)


-10 Signal 1
-35 Signal 1
ApaLI (5805)
GST
BamHI (697)
EcoRI (706)
attB1
Suv39h1 82-4
lac Repressor
pGST-Suv39h1 82-412
6035 bp
ApaLI (4466)
PstI (1490)
attB2
AvaI (1772)
-35 Signal 2
-10 Signal 2
ApaLI (3556)
ApaLI (2310)
Amp(R)
PstI (2740)


-10 Signal 1
-35 Signal 1
ApaLI (5778)
GST
BamHI (697)
EcoRI (706)
attB1
PstI (890)
XmaI (1012)
AvaI (1012)
SmaI (1014)
Suv39h2 157-4
ClaI (1508)
attB2
AvaI (1745)
-35 Signal 2
-10 Signal 2
ApaLI (2283)
Amp(R)
PstI (2713)
ApaLI (3529)
lac Repressor
ApaLI (4439)
pGST-Suv39h2 157-477
6008 bp

6. References

Aagaard, L., G. Laible, et al. (1999). "Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31." Embo J 18(7): 1923-38.

Akhtar, A., D. Zink, et al. (2000). "Chromodomains are protein-RNA interaction modules." Nature 407(6802): 405-9.

Bannister, A. J., P. Zegerman, et al. (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature 410(6824): 120-4.

Baumgartner, S., D. Bopp, et al. (1987). "Structure of two genes at the gooseberry locus related to the paired gene and their spatial expression during Drosophila embryogenesis." Genes Dev 1(10): 1247-67.

Berger, J., T. Suzuki, et al. (2001). "Genetic mapping with SNP markers in Drosophila." Nat Genet 29(4): 475-81.

Bernard, P., J. F. Maure, et al. (2001). "Requirement of heterochromatin for cohesion at centromeres." Science 294(5551): 2539-42.

Bird, A. (2002). "DNA methylation patterns and epigenetic memory." Genes Dev 16(1): 6-21.

Bird, A. P., M. H. Taggart, et al. (1979). "Methylated and unmethylated DNA compartments in the sea urchin genome." Cell 17(4): 889-901.

Bopp, D., M. Burri, et al. (1986). "Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of Drosophila." Cell 47(6): 1033-40.

Bosco, G., P. Campbell, et al. (2007). "Analysis of Drosophila species genome size and satellite DNA content reveals significant differences among strains as well as between species." Genetics 177(3): 1277-90.

Busslinger, M. (2004). “Transcriptional control of early B cell development.” Annu Rev Immunol. 22:55-79.

Cam, H. P., E. S. Chen, et al. (2009). "Transcriptional scaffolds for heterochromatin assembly." Cell 136(4): 610-4.

Carninci, P., T. Kasukawa, et al. (2005). "The transcriptional landscape of the mammalian genome." Science 309(5740): 1559-63.

Chalepakis, G., J. Wijnholds, et al. (1994). "Pax-3-DNA interaction: flexibility in the DNA binding and induction of DNA conformational changes by paired domains." Nucleic Acids Res 22(15): 3131-7.

Chan, W. Y., C. S. Cheung, et al. (2004). "Cardiac neural crest of the mouse embryo: axial level of origin, migratory pathway and cell autonomy of the splotch (Sp2H) mutant effect." Development 131(14): 3367-79.

Chen, E. S., K. Zhang, et al. (2008). "Cell cycle control of centromeric repeat transcription and heterochromatin assembly." Nature 451(7179): 734-7.

Cleard, F., M. Delattre, et al. (1997). "SU(VAR)3-7, a Drosophila heterochromatin-associated protein and companion of HP1 in the genomic silencing of position-effect variegation." Embo J 16(17): 5280-8.

Cleard, F. and P. Spierer (2001). "Position-effect variegation in Drosophila: the modifier Su(var)3-7 is a modular DNA-binding protein." EMBO Rep 2(12): 1095-100.

Corry, G. N., M. J. Hendzel, et al. (2008). "Subnuclear localization and mobility are key indicators of PAX3 dysfunction in Waardenburg syndrome." Hum Mol Genet 17(12): 1825-37.

De Rubertis, F., D. Kadosh, et al. (1996). "The histone deacetylase RPD3 counteracts genomic silencing in Drosophila and yeast." Nature 384(6609): 589-91.

Delattre, M., A. Spierer, et al. (2002). "A new gene in Drosophila melanogaster, Ravus, the phantom of the modifier of position-effect variegation Su(var)3-7." Int J Dev Biol 46(1): 167-71.

Delattre, M., A. Spierer, et al. (2004). "Increased expression of Drosophila Su(var)3-7 triggers Su(var)3-9-dependent heterochromatin formation." J Cell Sci 117(Pt 25): 6239-47.

Delattre, M., A. Spierer, et al. (2000). "The genomic silencing of position-effect variegation in Drosophila melanogaster: interaction between the heterochromatin-associated proteins Su(var)3-7 and HP1." J Cell Sci 113 Pt 23: 4253-61.

Dietzl, G., D. Chen, et al. (2007). "A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila." Nature 448(7150): 151-6.

Dodgson, J. B., H. H. Cheng, et al. (1997). "DNA marker technology: a revolution in animal genetics." Poult Sci 76(8): 1108-14.
Epstein, D. J., M. Vekemans, et al. (1991). "Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3." Cell 67(4): 767-74.
Fodor, B. D., S. Kubicek, et al. (2006). "Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells." Genes Dev 20(12): 1557-62.
Garagna, S., M. Zuccotti, et al. (2002). "High-resolution organization of mouse telomeric and pericentromeric DNA." Cytogenet Genome Res 96(1-4): 125-9.
Gaubatz, J. W. and R. G. Cutler (1990). "Mouse satellite DNA is transcribed in senescent cardiac muscle." J Biol Chem 265(29): 17753-8.
Girton, J. R. and K. M. Johansen (2008). "Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila." Adv Genet 61: 1-43.
Gonzalo, S., M. Garcia-Cao, et al. (2005). "Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin." Nat Cell Biol 7(4): 420-8.
Grewal, S. I. and S. C. Elgin (2007). "Transcription and RNA interference in the formation of heterochromatin." Nature 447(7143): 399-406.
Grewal, S. I. and S. Jia (2007). "Heterochromatin revisited." Nat Rev Genet 8(1): 35-46.
Grigliatti, T. (1991). "Position-effect variegation--an assay for nonhistone chromosomal proteins and chromatin assembly and modifying factors." Methods Cell Biol 35: 587-627.
Hollenbach, A. D., J. E. Sublett, et al. (1999). "The Pax3-FKHR oncoprotein is unresponsive to the Pax3-associated repressor hDaxx." Embo J 18(13): 3702-11.
Holmes, M. L., C. Pridans, et al. (2008). "The regulation of the B-cell gene expression programme by Pax5." Immunol Cell Biol 86(1): 47-53.
Hsieh, M. J., Y. L. Yao, et al. (2006). "Transcriptional repression activity of PAX3 is modulated by competition between corepressor KAP1 and heterochromatin protein 1." Biochem Biophys Res Commun 349(2): 573-81.
Huisinga, K. L., B. Brower-Toland, et al. (2006). "The contradictory definitions of heterochromatin: transcription and silencing." Chromosoma 115(2): 110-22.
Huntley, S., D. M. Baggott, et al. (2006). "A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors." Genome Res 16(5): 669-77.
Iarmarcovai, G., S. Bonassi, et al. (2008). "Genetic polymorphisms and micronucleus formation: a review of the literature." Mutat Res 658(3): 215-33.
Iida, T., J. Nakayama, et al. (2008). "siRNA-mediated heterochromatin establishment requires HP1 and is associated with antisense transcription." Mol Cell 31(2): 178-89.
Ishikiriyama, S., H. Tonoki, et al. (1989). "Waardenburg syndrome type I in a child with de novo inversion (2)(q35q37.3)." Am J Med Genet 33(4): 505-7.
James, T. C. and S. C. Elgin (1986). "Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene." Mol Cell Biol 6(11): 3862-72.
Jolly, C., A. Metz, et al. (2004). "Stress-induced transcription of satellite III repeats." J Cell Biol 164(1): 25-33.
Joseph, A., A. R. Mitchell, et al. (1989). "The organization of the mouse satellite DNA at centromeres." Exp Cell Res 183(2): 494-500.
Kloc, A. and R. Martienssen (2008). "RNAi, heterochromatin and the cell cycle." Trends Genet 24(10): 511-7.
Kloc, A., M. Zaratiegui, et al. (2008). "RNA interference guides histone modification during the S phase of chromosomal replication." Curr Biol 18(7): 490-5.
Klose, R. J., K. Yamane, et al. (2006). "The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36." Nature 442(7100): 312-6.
Kuznetsova, I., O. Podgornaya, et al. (2006). "High-resolution organization of mouse centromeric and pericentromeric DNA." Cytogenet Genome Res 112(3-4): 248-55.
Kwon, S. H. and J. L. Workman (2008). "The heterochromatin protein 1 (HP1) family: put away a bias toward HP1." Mol Cells 26(3): 217-27.
Lachner, M., D. O'Carroll, et al. (2001). "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins." Nature 410(6824): 116-20.

Lachner, M., R. J. O'Sullivan, et al. (2003). "An epigenetic road map for histone lysine methylation." J Cell Sci 116(Pt 11): 2117-24.
Lachner, M., R. Sengupta, et al. (2004). "Trilogies of histone lysine methylation as epigenetic landmarks of the eukaryotic genome." Cold Spring Harb Symp Quant Biol 69: 209-18.
Lakhotia (2006). "Human sat III and Drosophila hsr omega transcripts: a common paradigm for regulation of nuclear RNA processing in stressed cells." Nucleic Acids Res 34(19): 5508-14.
Lang, D., S. K. Powell, et al. (2007). "PAX genes: roles in development, pathophysiology, and cancer." Biochem Pharmacol 73(1): 1-14.
Lehnertz, B., Y. Ueda, et al. (2003). "Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin." Curr Biol 13(14): 1192-200.
Li, E., T. H. Bestor, et al. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell 69(6): 915-26.
Loenen, W.A., 2006. “S-adenosylmethionine: jack of all trades and master of everything?”. Biochem Soc Trans. 34(Pt 2): 330-3.
Lu, J. and D. M. Gilbert (2007). "Proliferation-dependent and cell cycle regulated transcription of mouse pericentric heterochromatin." J Cell Biol 179(3): 411-21.
Lu, J. and D. M. Gilbert (2008). "Cell cycle regulated transcription of heterochromatin in mammals vs. fission yeast: functional conservation or coincidence?" Cell Cycle 7(13): 1907-10.
Luger, K., A. W. Mader, et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature 389(6648): 251-60.
Ma, Y., S. B. Jacobs, et al. (2005). "DNA CpG hypomethylation induces heterochromatin reorganization involving the histone variant macroH2A." J Cell Sci 118(Pt 8): 1607-16.
Magnaghi, P., C. Roberts, et al. (1998). "HIRA, a mammalian homologue of Saccharomyces cerevisiae transcriptional co-repressors, interacts with Pax3." Nat Genet 20(1): 74-7.
Maison, C. and G. Almouzni (2004). "HP1 and the dynamics of heterochromatin maintenance." Nat Rev Mol Cell Biol 5(4): 296-304.
Mansouri, A., P. Pla, et al. (2001). "Pax3 acts cell autonomously in the neural tube and somites by controlling cell surface properties." Development 128(11): 1995-2005.
Martens, J. H., R. J. O'Sullivan, et al. (2005). "The profile of repeat-associated histone lysine methylation states in the mouse epigenome." Embo J 24(4): 800-12.
Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysine methylation." Nat Rev Mol Cell Biol 6(11): 838-49.
Morey, C. and P. Avner (2004). "Employment opportunities for non-coding RNAs." FEBS Lett 567(1): 27-34.
Morozova, O. and M. A. Marra (2008). "Applications of next-generation sequencing technologies in functional genomics." Genomics 92(5): 255-64.
Nutt, S.L., D. Eberhard, et al. (2001). “Pax5 determines the identity of B cells from the beginning to the end of B-lymphopoiesis.” Int Rev Immunol. 20(1): 65-82.
Peters, A. H., S. Kubicek, et al. (2003). "Partitioning and plasticity of repressive histone methylation states in mammalian chromatin." Mol Cell 12(6): 1577-89.
Peters, A. H., D. O'Carroll, et al. (2001). "Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability." Cell 107(3): 323-37.
Quina, A. S., M. Buschbeck, et al. (2006). "Chromatin structure and epigenetics." Biochem Pharmacol 72(11): 1563-9.
Rea, S., F. Eisenhaber, et al. (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases." Nature 406(6796): 593-9.
Reuter, G., J. Gausz, et al. (1987). "Modifiers of position-effect variegation in the region from 86C to 88B of the Drosophila melanogaster third chromosome." Mol Gen Genet 210(3): 429-36.
Reuter, G., M. Giarre, et al. (1990). "Dependence of position-effect variegation in Drosophila on dose of a gene encoding an unusual zinc-finger protein." Nature 344(6263): 219-23.
Reuter, G. and P. Spierer (1992). "Position effect variegation and chromatin proteins." Bioessays 14(9): 605-12.
Reuter, G., W. Werner, et al. (1982). "Mutants affecting position-effect heterochromatinization in Drosophila melanogaster." Chromosoma 85(4): 539-51.

Reuter, G. and I. Wolff (1981). "Isolation of dominant suppressor mutations for position-effect variegation in Drosophila melanogaster." Mol Gen Genet 182(3): 516-9.
Rice, J. C., S. D. Briggs, et al. (2003). "Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains." Mol Cell 12(6): 1591-8.
Rizzi, N., M. Denegri, et al. (2004). "Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock." Mol Biol Cell 15(2): 543-51.
Robson, E. J., S. J. He, et al. (2006). "A PANorama of PAX genes in cancer and development." Nat Rev Cancer 6(1): 52-62.
Rudert, F., S. Bronner, et al. (1995). "Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development." Mamm Genome 6(2): 76-83.
Schotta, G., A. Ebert, et al. (2003). "SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing." Genetica 117(2-3): 149-58.
Schotta, G., M. Lachner, et al. (2004). "The indexing potential of histone lysine methylation." Novartis Found Symp 259: 22-37; discussion 37-47, 163-9.
Schotta, G., M. Lachner, et al. (2004). "A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin." Genes Dev 18(11): 1251-62.
Schotta, G., R. Sengupta, et al. (2008). "A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse." Genes Dev 22(15): 2048-61.
Schulze, S. R. and L. L. Wallrath (2007). "Gene regulation by chromatin structure: paradigms established in Drosophila melanogaster." Annu Rev Entomol 52: 171-92.
Sengupta, S., R. Parihar, et al. (2009). "Satellite III non-coding RNAs show distinct and stress-specific patterns of induction." Biochem Biophys Res Commun 382(1): 102-7.
Spofford, J. B. (1967). "Single-locus modification of position-effect variegation in Drosophila melanogaster. I. White variegation." Genetics 57(4): 751-66.
Sripathy, S. P., J. Stevens, et al. (2006). "The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression." Mol Cell Biol 26(22): 8623-38.
Starmer, J. and T. Magnuson (2009). "A new model for random X chromosome inactivation." Development 136(1): 1-10.
Ugarkovic, D. (2005). "Functional elements residing within satellite DNAs." EMBO Rep 6(11): 1035-9.
Valgardsdottir, R., I. Chiodi, et al. (2008). "Transcription of Satellite III non-coding RNAs is a general stress response in human cells." Nucleic Acids Res 36(2): 423-34.
Van Hooser, A. A., M. A. Mancini, et al. (1999). "The mammalian centromere: structural domains and the attenuation of chromatin modeling." Faseb J 13 Suppl 2: S216-20.
Vassen, L., K. Fiolka, et al. (2006). "Gfi1b alters histone methylation at target gene promoters and sites of gamma-satellite containing heterochromatin." Embo J 25(11): 2409-19.
Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science 291(5507): 1304-51.
Vissel, B. and K. H. Choo (1989). "Mouse major (gamma) satellite DNA is highly conserved and organized into extremely long tandem arrays: implications for recombination between nonhomologous chromosomes." Genomics 5(3): 407-14.
Volpe, T. A., C. Kidner, et al. (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science 297(5588): 1833-7.
Volpe. T. A., V. Schramke, et al. (2003). “RNA interference is required for normal centromere function in fission yeast.” Chromosome Res. 11(2):137-46.
Warburton, P. E., D. Hasson, et al. (2008). "Analysis of the largest tandemly repeated DNA families in the human genome." BMC Genomics 9: 533.
Waye, J. S. and H. F. Willard (1986). "Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome." Mol Cell Biol 6(9): 3156-65.
Wierzbicki, A. T., J. R. Haag, et al. (2008). "Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes." Cell 135(4): 635-48.

Wiggan, O., A. Taniguchi-Sidle, et al. (1998). "Interaction of the pRB-family proteins with factors containing paired-like homeodomains." Oncogene 16(2): 227-36.
Winkler, W.C., A. Nahvi, et al. (2003). “An mRNA structure that controls gene expression by binding S-adenosylmethionine.” Nat Struct Biol. 10(9): 701-7.
Wu, R., P. B. Singh, et al. (2006). "Uncoupling global and fine-tuning replication timing determinants for mouse pericentric heterochromatin." J Cell Biol 174(2): 185-94.

Acknowledgments

Thanks to Thomas Jenuwein for the possibility of entering what he used to define paradise, for the possibility of staying there although he moved out for creating another Paradise in Freiburg and of course thanks for the colleagues.

A special thank goes to Bori, Ines, Moni and Johannes.

Thanks to my sister, my mum, my dad, my grandmums and all the rest of my big family, for the constant support and the patience.

Thanks to my granddads for never leaving me.

Thanks to Giovanni for the immense patience during these years and for giving me the reason for going happily back to my lovely country: Italy!!

MANUELA SCARANARO-Curriculum vitae

Klopsteinplatz 1/21
1030, Wien (Austria)
+436765869812 (tel.)

Capo di Sopra 76
45030, Villanova Marchesana, Rovigo (Italy)
+393494554765 (tel.)

manuela.scaranaro@alice.it

Born in Rovigo, on September 22nd 1981

Italian citizenship

EDUCATION

Currently	3rd year PhD student in Thomas Jenuwein laboratory, Research Institute of Molecular Pathology (I.M.P).
2000-2005	Diploma in **Pharmaceutical Biotechnology**, University of Padua, Italy.
	Title of thesis: "The protective role of calcineurin in the keratinocytes UVB response".
	Grade achieved: 110/110 *summa cum laude*.
1995-2000	**Scientific Liceo** "Pietro Paleocapa", Rovigo, Italy.
	High school, 1st Class Honour (100/100).

SKILLS

Linguistic	Fluent in **English, German** and **Italian** (mother tongue).	
	Conversant in **French** and **Spanish**.	
Computer	Operating systems	Windows 2007, XP, Vista, Mac OS X.
	Office Package	Word, PowerPoint, Excel.
	Internet and e-mail	Internet Explorer, Mozilla Firefox, Outlook Express.
	Graphic Programs	Adobe Photoshop, Adobe Illustrator, Macromedia Freehand.