Alma Mater Studiorum - Università degli Studi di Bologna
Dottorato di Ricerca in Biochimica
CICLO XX
Coordinatore Prof. Giorgio Lenaz

Identification of *Drosophila* heart-specific *Cis*-Regulatory Modules under *Hox* control

Tesi di Dottorato della Dott.ssa Maria Florencia TEVY

Relatore: Dr.ssa Maria CAPOVILLA

SUMMARY

Cardiac morphogenesis is a complex process governed by evolutionarily conserved transcription factors and signaling molecules. The *Drosophila* cardiac tube is linear, made of 52 pairs of cardiomyocytes (CMs), which express specific transcription factor genes that have human homologues implicated in Congenital Heart Diseases (CHDs) (NKX2-5, GATA4 and TBX5). The *Drosophila* cardiac tube is linear and composed of a rostral portion named aorta and a caudal one called heart, distinguished by morphological and functional differences controlled by *Hox* genes, key regulators of axial patterning. Overexpression and inactivation of the *Hox* gene *abdominal-A* (*abd-A*), which is expressed exclusively in the heart, revealed that *abd-A* controls heart identity. The aim of our work is to isolate the heart-specific *cis*-regulatory sequences of *abd-A* direct target genes, the realizator genes granting heart identity. In each segment of the heart, four pairs of cardiomyocytes (CMs) express *tinman* (*tin*), homologous to NKX2-5, and acquire strong contractile and automatic rhythmic activities. By tyramide amplified FISH, we found that seven genes, encoding ion channels, pumps or transporters, are specifically expressed in the Tin-CMs of the heart. We initially used online available tools to identify their heart-specific *cis*-regutatory modules by looking for Conserved Non-coding Sequences containing clusters of binding sites for various cardiac transcription factors, including *Hox* proteins. Based on these data we generated several reporter gene constructs and transgenic embryos, but none of them showed reporter gene expression in the heart. In order to identify additional *abd-A* target genes, we performed microarray experiments comparing the transcriptomes of aorta *versus* heart and identified 144 genes overexpressed in the heart. In order to find the heart-specific *cis*-regulatory regions of these target genes we developed a new bioinformatic approach where prediction is based on pattern matching and ordered statistics. We first retrieved Conserved Non-coding Sequences from the alignment between the *D.melanogaster* and *D.pseudobscura* genomes. We scored for combinations of conserved occurrences of ABD-A, ABD-B, TIN, PNR, dMEF2, MADS box, T-box and E-box sites and we ranked these results based on two independent strategies. On one hand we ranked the putative *cis*-regulatory sequences according to best scored ABD-A biding sites, on the other hand we scored according to conservation of binding sites. We integrated and ranked again the two lists obtained independently to produce a final rank. We generated nGFP reporter construct flies for *in vivo* validation. We identified three 1kb-long heart-specific enhancers. By *in vivo* and *in vitro* experiments we are determining whether they are direct *abd-A* targets, demonstrating the role of a *Hox* gene in the realization of heart identity. The identified *abd-A* direct target genes may be targets also of the NKX2-5, GATA4 and/or TBX5 homologues *tin*, *pannier* and *Doc* genes, respectively. The identification of sequences coregulated by a *Hox* protein and the homologues of transcription factors causing CHDs, will provide a mean to test whether these factors function as *Hox* cofactors granting cardiac specificity to *Hox* proteins, increasing our knowledge on the molecular mechanisms underlying CHDs. Finally, it may be investigated whether these *Hox* targets are involved in CHDs.

INTRODUCTION

Homeotic transformations

The mutant fly with four wings and no halteres was first identified by Calvin Bridges in Thomas Hunt Morgan laboratory in 1915 (**Figure 1**).

Figure 1.

This now famous little "monster" was a starting point for Edward Lewis in his research on homeotic transformations. In 1978, Lewis published his work on the *bithorax* complex (Lewis, 1978; Lewis; 1998). In this paper he reviewed a series of *Drosophila* mutations that affect the thoracic and abdominal segmental identities. These mutations were called “homeotic” (from the greek *homoios* or “similar”) because one single mutation transforms a whole segment into another (towards an anterior one). Remarkably, the different homeotic mutations mapped to the third chromosome in an order that corresponds to the anterior to posterior order of the segments in which they act. These collinear genes responsible for the anterior to posterior patterning along the body axis are now called *Hox* genes, because they share a 180bp sequence called “homeobox”, which encodes a 60 amino acids DNA-binding domain called “homeodomain”. Still more astonishing was the fact that these genes are evolutionarily conserved in function and organization from *Hydra* to *Homo*. In the genome of *Drosophila*, there are two clusters called *Antennapedia* and *bithorax*, both located on chromosome III. Mammals possess 39 *Hox* genes ordered in four clusters that reflect their temporal order of expression during embryogenesis and along the anterior-posterior (AP) axis (**Figure 2**). This means that the genetic control mechanisms of body planning have been preserved roughly unchanged through 650 million years of evolution. (reviewed in Lawrence and Morata, 1994; Carroll, 1995; Carroll, 2005; Maeda & Karch, 2006; Iimura & Pourquie, 2007).

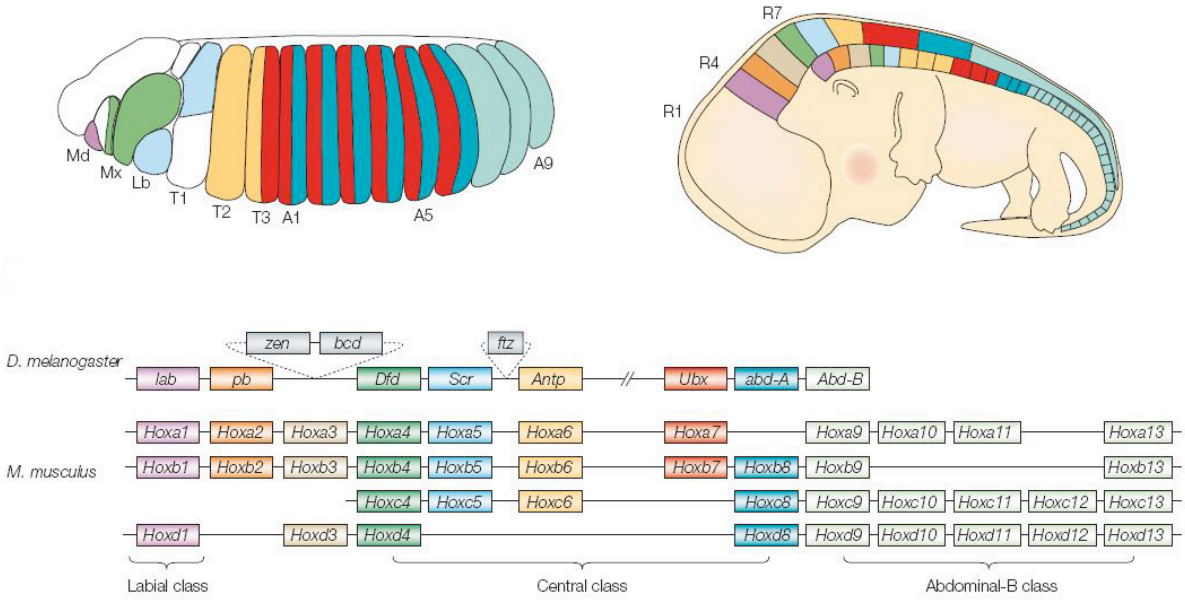


Figure 2 (modified from Pearson *et al.*, 2005). ***Hox* expression and genomic organization in *Drosophila melanogaster* and *Mus musculus*.** The diagram on the left shows a stage 13 *Drosophila* embryo colored to indicate the approximate domains of expression for all *Drosophila Hox* genes except *proboscipedia* (*pb*), which is not expressed in the embryo. The segments are labeled: Md, mandibular; Mx, maxillary; Lb, labial; T1-T3, thoracic segments; A1-A9, abdominal segments. The diagram on the right shows a mouse embryo at embryonic day 12.5, with the approximate *Hox* expression domains depicted on the head to tail axis. The positions of the hindbrain rhombomeres are labeled R1, R4, R7. In both diagrams, the domains of expression of the *Hox* transcripts are color-coded as the corresponding genes in the *Hox* cluster diagram below. The diagrams below the embryos show the *Hox* clusters in the genomes of *Drosophila* and mouse with the genes in the order in which they are found in the chromosome. Genes are colored to differentiate between *Hox* paralogous members. Genes that are orthologous between clusters and specie are labeled in the same color. The position of three non-*Hox* homeodomain genes *zen*, *bcd* and *ftz* are shown in the fly *Hox* cluster in grey boxes. Gene abbreviations: *lab, labial*; *pb*, *proboscipedia*; *zen*, *zerknullt*; *bcd*, *bicoid*; *Dfd*, *Deformed*; *Scr*, *Sex combed reduced*; *ftz*, *fushi tarazu*; *Antp*, *Antennapedia*; *Ubx*, *Ultrabithorax*; *abd-A*, *abdominal-A*; *Abd-B*, *Abdominal-B*. Anterior is to the left and dorsal to the top.

Hox genes in morphogenesis

The homeotic function is the trademark by which *Hox* genes became known. Nevertheless, the function of *Hox* genes is that of regulating morphogenesis and organogenesis. This is the function we have to study if we want to answer the question of how differences arise from sets of genes so widely shared.

Hox genes encode homeodomain transcription factors (reviewed in McGinnis, 1994). One target of *Hox* genes are *Hox* genes themselves. To control organogenesis, *Hox* proteins act at a local cellular level within the segment they specify and regulate groups of target genes. The *Hox* protein function here is to get integrated into an enhanceosome at a given moment of development and confer AP positional information to this enhanceosome (Castelli-Gair Hombría & Lovegrove, 2003). Garcia Bellido (1977) coined the term "realizator genes" to describe these sets of genes subordinated to *Hox* control that mediate segment-specific morphogenesis.

Hox target genes

As Capovilla *et al.* stated in their 1994 paper: "Understanding how genes of the HOM/*Hox* family control morphogenesis requires the identification and characterization of their target genes [...]. Yet very little is known about the nature of these target genes and their roles in segment-specific morphogenesis."

There are characteristics inherent to *Hox* genes that make the identification and characterization of their target realizators difficult. To start with, almost all *Hox* proteins bind *in vitro* to a short core consensus 5'-TAAT-3' sequence. This sequence occurs in the genome approximatively every 1 kilobase. *In vitro* promiscuous binding is very common and in consequence, of all the *in vitro* detected binding sites most have no *in vivo* significance (Ekker *et al*., 1991, 1992). Second, *Hox* protein action may not be cell autonomous, acting through signaling cascades which can in turn lead to effects outside the *Hox* expression domain. This is the case where the *Hox* gene *abd-A* activates only the target gene *rhomboid* (*rho*) in the C1 cell lineage to promote the differentiation of oenocytes in the dorsal ectoderm (Brodu *et al.*, 2002). At another level, in some tissues and organs, *Hox* genes have overlapping expression domains and cross-regulate each other, therefore the target may require inputs from multiple *Hox* genes. For example, the activities of the *Hox* genes *Ubx* and *abd-A* are integrated at the *decapentaplegic* (*dpp*) visceral mesoderm enhancer, where UBX acts as an activator and ABD-A as a repressor by binding to the same sites in the sequence (Capovilla *et al.*, 1994; Capovilla and Botas, 1998). UBX and ABD-A instead appear to act both as repressors of the limb promoting target gene *Distal-less* (*Dll*) apparently also binding through the same HOX binding sites (Gebelein *et al.*, 2004). Moreover, the control region of the putative target may be far from the transcription unit of the target gene, and thus difficult to find.

In spite of all these experimental difficulties, a few *Hox* target genes (direct or indirect) are known, although the exact number is still a matter of debate. A direct target provides compelling evidence that a specific *Hox* protein is binding to a specific enhancer *in vivo* (Mahaffey, 2005; Pearson *et al*., 2005). The definitive demonstration that a given gene is a direct target of a given HOX is the mutation of the HOX binding sites of its enhancer towards BICOID (BCD) binding sites. *Bcd* is not a *Hox* gene and its protein binds to a different consensus site than *Hox* proteins. HOX can be made to bind to this BCD site by changing only one amino acid in their homeodomain (Treisman *et al*., 1992; Schier & Gehring, 1992). So, if BCD sites are introduced in place of the HOX sites and the enhancer looses activity, and if a compensatory mutation in the *Hox* protein, allowing it to bind to the mutant BCD site, restores the activity of the enhancer, we are sure that that HOX transcription factor binds *in vivo* to the enhancer sequence of the given gene. Nevertheless, such rigorous test has only been performed in four cases: to demonstrate that *dpp* is a direct target of *Ubx* (Capovilla *et al*., 1994) and of *abd-A* (Capovilla and Botas, 1998), that *apterous* is a direct target of *Antp* (Capovilla *et al.,* 2001)*,* and that the intronic sequence of *Hoxa-4* is regulated by *Ubx* when assayed in *Drosophila* (Haerry & Gehring, 1997).

Moreover, most of the known targets are either transcription factors or signaling molecules, therefore not true realizators. On the other hand, we still have little or no knowledge of targets regulated at the same time and in the same place by the same *Hox* protein. Finding groups of co-regulated direct target realizator genes within similar cell types during development will help us assess the question of how is it that a *Hox* gene triggers specification, differentiation and organogenesis. From a broader point of view, *Hox*-dependent regulatory networks will help us to understand how these master genes create cell diversity in order to make serial structures different.

HOX specificity

There are only eight *Hox* genes in *Drosophila* and probably thousands of realizator genes in order to make and organ. How do *Hox* proteins selectively regulate such broad spectrum of target genes? The segments in *Drosophila* share plenty similarities so one could think that similar sets of target genes are shared by different *Hox* proteins or by different combinations of *Hox* proteins. But if we think the number of target genes in a co-regulated set is variable from one in the precursors of Keilin's organs (Vachon *et al.*, 1992) or the oenocytes (Brodu *et al.*, 2002) to possible more than a hundred in the halteres (Akam, 1998; Weatherbee *et al.*, 1998), we then realize such explanation does not account for so many clear morphological differences that are *Hox*-dependent.

On the other hand, at the protein level, *Hox* proteins bind to DNA through the homeodomain. The three alpha helices that conform the homeodomain are shared by all these transcription factors (Gehring *et al.*,1994, Mann, 1995). Moreover, *Hox* proteins (except for ABD-B) have indistinguishable *in vitro* binding sites. Then, how is HOX functional specificity achieved *in vivo*? How do individual *Hox* proteins selectively activate or repress target gene expression?

HOX specificity appears to depend on several non-exclusive mechanisms. One possibility is that the binding of a cofactor to a *Hox* protein confers the selectivity (and higher affinity of the *Hox* protein for its target). Unfortunately, only a couple of HOX cofactors constituted by *extradenticle* (*exd*) and *Homothorax* (*Hth*), the homologs of mammalian PBX and MEIS proteins respectively (Moens & Selleri, 2006), have been identified to date (Chan *et al*., 1994; Mann & Affolter, 1998; Ryoo *et al*., 1999). On the other hand, many processes involving the function of *Hox* genes have been found to be *exd*-independent (Pinsonneault *et al.*, 1997; Galant *et al.*, 2002). The failure of traditional genetic and biochemical screens to identify new HOX cofactors suggests that *Hox* genes might not always function exactly how we previously expected. This view is reinforced by the recent and unforeseen finding that *Hox* and segment-polarity genes cooperate to control *Dll* expression. Thus, another possibility is that *Hox* proteins collaborate (cooperate and/or recruit sets of activators and repressors) with other transcription factors to control gene expression. Collaboration does not imply a physical interaction between the transcription factors integrated in the enhanceosome (Walsh & Carroll, 2007). In other words, a combination of different proteins including one or more HOX would regulate transcription. In fact, there is an expanding pool of transcription factors that are putative collaborators in the modulation of *Hox* target genes. Some evidence comes from the collaboration between UBX and SMADs in the selective repression of the target gene *spalt* (*sal*) in the *Drosophila* haltere disc (Walsh & Carroll, 2007). Furthermore, new insights on target selection come from the way a *Hox* protein binds to the DNA structure. Recent experiments suggest that *Hox* proteins recognize "generic" binding sites through the homeodomain-major groove interactions while the N-terminal arm and the linker residues would select among these sites by "reading" the structure and electrostatic potential in the minor groove (Joshi *et al*., 2007).

A new model for old questions

From all the mentioned characteristics of *Hox* genes and *Hox* proteins it is evident that their functions are difficult to analyze. Since the 1980s new techniques have been developed and with them new questions regarding *Hox* genes have been raised. Instead, the number of model systems to assess *Hox* function and HOX specificity has remained mainly the same and new ones need to be introduced. Cardiogenesis in *Drosophila* constitutes such a system.

The *Drosophila* cardiac tube

The embryonic cardiac system of *Drosophila* is formed by a simple linear tubular structure at the dorsal midline below the epidermis of the embryo and it is called "dorsal vessel" or "cardiac tube" (Campos-Ortega & Hartenstein, 1997). Just like the vertebrate primitive cardiac tube, the *Drosophila* dorsal vessel presents AP polarity (Lo & Frasch, 2003). The anterior portion of the tube is called "aorta" and the posterior one "heart" (**Figure 3**). The heart is the autonomous beating organ (Wessells *et al.*, 2004) that pumps the haemolymph ("insect blood") in a posterior to anterior manner (Rizki, 1978), comparable to the direction of blood flow in the vertebrate primitive cardiac tube (Forouhar *et al.*, 2006). In the fly open circulatory system, the haemolymph reenters the heart through valve-like cells named *ostia* (Rizki, 1978).

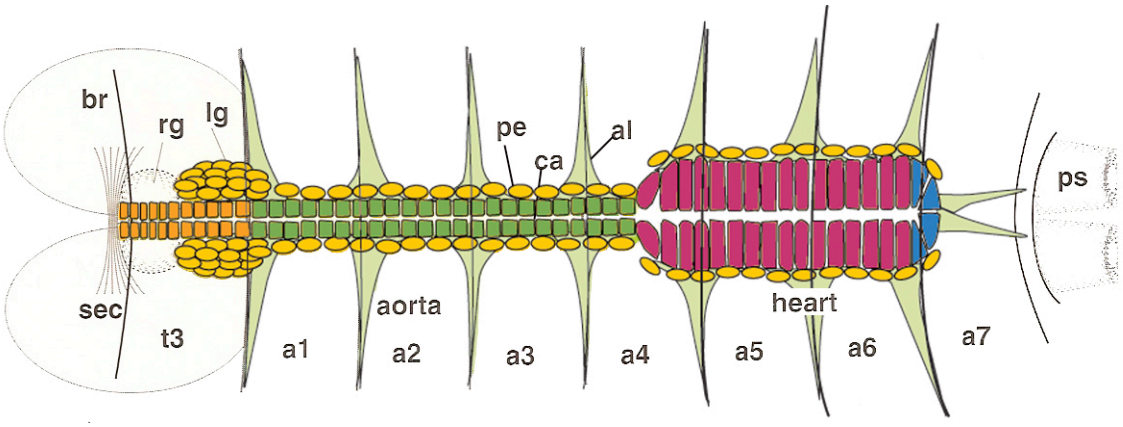


Figure 3 (modified from Campos-Ortega & Hartenstein, 1997). **Dorsal view of the cardiac tube of an embryo at the end of embryogenesis.** The dorsal vessel extends from segment T1 (not shown in the diagram) to segment A7. Alary muscles (*al*) attach the dorsal vessel to the apodemes. The dorsal vessel, consisting of the heart posteriorly (middle of segment A4 to segment A7) and the aorta anteriorly (segment T1 to middle of segment A4), is formed by cardiomyocytes (*ca*), in association with a bilateral row of pericardial cells (*pc*). In its anterior portion, the cardiac tube is flanked by the lymph glands (*lg*) and the ring gland (*rg*). Abbreviations: brain, *br*; supraoesophageal commissure, *sec*; posterior spiracle, *sp*; thoracic segment, *t*; abdominal segment, *a*.

The formation of the cardiac tube

The dorsal vessel, like the somatic and visceral musculature and the fat body, has mesodermal origin. The mesoderm is derived from the most ventral cells of the blastoderm (Campos-Ortega & Hartenstein, 1997). The cells of the mesoderm express *twist* (*twi*) and later *snail* (*sna*), both encoding transcription factors. *twi,* in turn, activates *tinman* (*tin*), the gene encoding a homeodomain transcription factor of the NK class (reviewed in Ruiz-Gomez, 1998). At gastrulation, these cells invaginate along the ventral furrow and spread dorsally to form a monolayer beneath the ectoderm. During dorsal migration important regulatory events that will ultimately lead to specification of heart progenitors are triggered (Frasch, 1999). The cardioblasts become specified thanks to inductive signals that come from the neighbouring ectoderm (Zaffran & Frasch, 2002), which comprehend the cascades of the signaling factors *dpp* and *wingless* (*wg*), homologous to the vertebrate BMP and WNT molecules in vertebrates, respectively. Upon spreading of the mesoderm beneath the ectoderm, the mesoderm utilizes the pattern information from the ectoderm to acquire its distinct dorsal-ventral polarity. *dpp*

transmits dorsal positional information from the ectoderm to the mesoderm (Frasch, 1995; Lockwood & Bodmer, 2002). A consequence of *dpp* function is the restriction of *tin* expression to the dorsal mesoderm (Frasch, 1999) and possibly the induction of *pannier* (*pnr*), encoding a Zinc finger transcription factor (Klinedinst & Bodmer, 2003). As development proceeds, *wg* plays a key role in segmental regulation in cooperation with *sloppy paired* (*slp*) (Lee & Frasch, 2000).

The cells from the dorsal-most edge of the mesoderm form the so-called mesodermal crests on either side of the embryonic germ band. The heart progenitors derive from segmentally restricted clusters of cells from the dorsal mesodermal crests (Campos-Ortega & Hartenstein, 1997) (**Figure 4**). The presumptive heart progenitors rearrange and the clusters come into contact. These cells undergo a mesenchymal-epithelial transition and as a result of these movements, two rows of cells on either side of the embryo are formed: a dorsal row of cardioblasts and a ventrally adjacent row of pericardial cells (Fremion *et al*., 1999) (**Figure 4**).

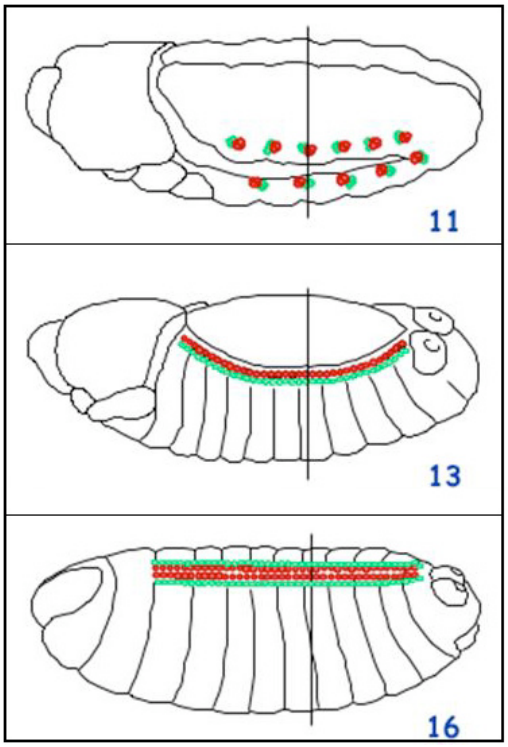


Figure 4 (from Campos-Ortega & Hartenstein, 1997). **Embryonic heart development.** (A) Stage 11 embryo showing segmentally restricted clusters of cells from the dorsal mesoderm, which will give rise to the heart progenitors. (B) Stage 13 embryo where the heart progenitors rearrange and the clusters come into contact after mesenchymal-epithelial transition. Two rows of cells on either side of the embryo are formed: a dorsal row of cardioblasts and a ventrally adjacent row of pericardial cells. (C) Stage 17 embryo where the two rows of cardioblasts meet and fuse at the dorsal midline to form the dorsal vessel, which becomes a hollow tube.

During germ band elongation the two rows of cardioblasts meet and fuse at the dorsal midline to form the dorsal vessel, which becomes a hollow tube extending from the first thoracic segment (T1) to seventh abdominal segment (A7) (**Figures 3 and 4**). At this moment, upon dorsal closure, all the differentiation programs have already been triggered so that the cardioblasts become cardiomyocytes. Part of these differentiation programs are activated by *tin* and *pnr* (Gajewski *et al.,* 2001). Two differentiation transcription factors activated by *tin* and *pnr*

are *muscle enhancer factor 2* (*mef2*) (Gajewski *et al.,* 2001) and the basic helix loop helix transcription factor *dHand* (Han & Olson, 2005).

The mature cardiac tube is constituted by 52 pairs of cardiomyocytes, which become connected by adherent junctions. The dorsal vessel becomes surrounded in the thorax by the lymph glands (of mesodermal origin) and the ring gland (of ectodermal origin) and also by loosely arranged, non-myogenic pericardial cells (thought to act as stationary macrophages) (Rizki, 1978). The cardiac tube is attached to the dorsal ectoderm by seven pairs of alary muscles (**Figure 3**). As *Drosophila* is a holometabolous insect, this morphogenetic patrimony will remain the same until metamorphosis when all the structures will be eliminated or remodeled (Monier *et al*., 2005).

Anterior-posterior organization of the dorsal vessel

Like the primitive heart tube in vertebrates, the mature cardiac tube of *Drosophila* presents morphological and functional axial differentiation. Anterior to posterior polarity is manifested not only in the subdivision of the dorsal vessel into the two domains, aorta and heart (intersegmental polarity), but also, within each segment that forms the cardiac tube (intrasegmental polarity) (Ponzielli *et al.*, 2002; Lo *et al.*, 2002).

Rostral-caudal polarity can be observed within each segment that forms the cardiac tube. Each segment is composed of six pairs of cardiomyocytes. The four most anterior pairs of cells of each segment that forms the cardiac tube express the gene encoding the homeodomain transcription factor *tin* (Bodmer, 1993). In addition to *tin*, the two most anterior pairs of cells express the transcription factor *ladybird* (*lb*) (Jagla *et al*., 1997). The two most posterior pairs of cells within each segment express the gene encoding the transcription factor *seven up* (*svp*), the *Drosophila* homolog of the COUP-TF orphan receptors (Lo & Frasch, 2001).

Moreover, as mention previously, the entire dorsal vessel presents two distinct domains: an anterior one (from segment T1 to middle of segment A4) termed “aorta” and a posterior one (from middle of segment A4 to segments A7) called “heart” (Rizki, 1978) (**Figure 3**). The heart portion of the cardiac tube presents a wider diameter and lumen compared to the aorta (Ponzielli *et al.*, 2002). The three pairs of *svp*-expressing cells of the heart differentiate into valve-like cells called *ostia*, which allow the inflow of haemolymph into the cavity (Lo & Frasch, 2001). The *tin*-expressing cardioblasts of the heart differentiate into strong contractile beating cardiomyocytes, which propel this hemolymph, through the aorta, into an open circulatory system in a caudal to rostral direction (Molina & Cripps, 2001). Thus, the aorta constitutes the outflow tract of the fly cardiac tube. The SVP- positive cells of the aorta do not differentiate into functional *ostia* and the TIN-positive cells of this domain differentiate into cardiomyocytes that are smaller and poorly contractile compared to those of the heart (Ponzielli *et al.*, 2002).

The correlation between *Hox* expression domains and axial polarity of the dorsal vessel prompted several laboratories to study the possible involvement of *Hox* genes in cardiac diversity. Four *Hox* genes were found to be expressed in the cardiac tube. *Antp* is expressed in the anterior portion of the aorta, *Ubx* in the posterior portion of the aorta, and *abd-A* is expressed in all the cells of the heart except in the last four cells where *Abd-B* is expressed (**Figure 5**).

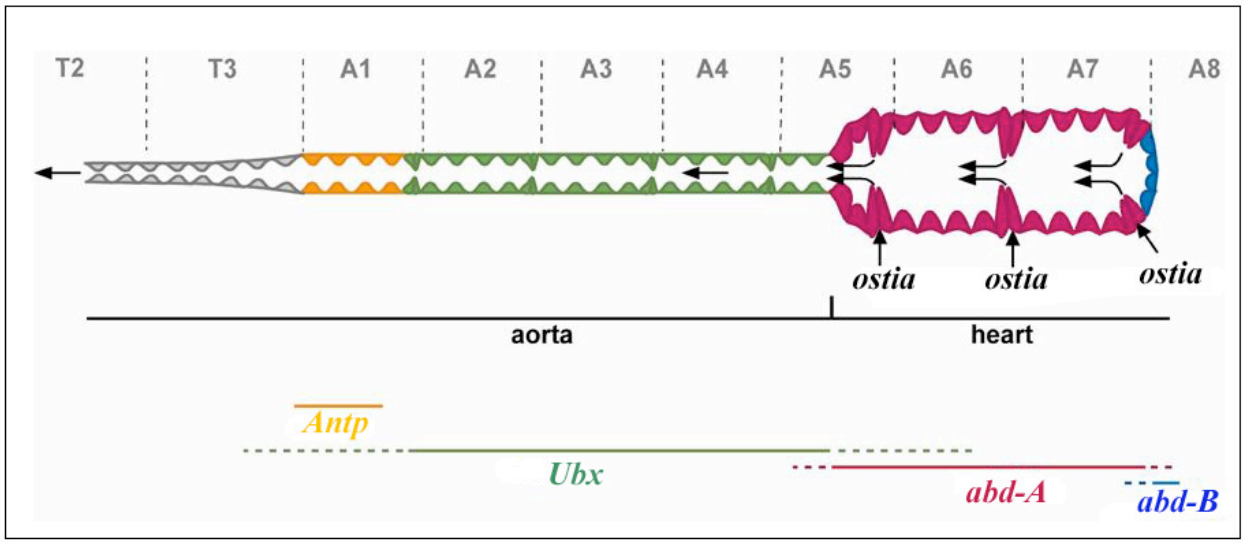


Figure 5 (from Ponzielli *et al.*, 2002). **Rostral-caudal polarity in the *Drosophila* cardiac tube.** The diagram shows the expression of the *Hox* gene *Antp* and the three genes of the *bithorax* complex (*Ubx*, *abd-A* and *Abd-B*) along the AP axis of the dorsal vessel. Anterior is to the left. The black arrows indicate the direction of the haemolymph flow.

In the quintuple mutant *Scr*$^{-}$, *Antp*$^{-}$, *Ubx*$^{-}$, *abd-A*$^{-}$, *abd-B*$^{-}$, all cardioblasts adopt anterior thoracic cardiac cell identity (*i.e.*, cardiomyoblasts that express *Antp*) and thus no posterior aorta cardiomyocytes and heart cardiomyocytes differentiate. *Antp* affects *svp* expression. Loss of *Antp* leads to a loss of *svp* expression in the first pair of abdominal cardiomyocytes, while ectopic *Antp* expression activates *svp* expression ectopically in the thorax (Perrin *et al.*, 2004).

Abd-B is known to be a repressor of myogenesis (Michelson, 1994). Early overexpression of *Abd-B* in the mesoderm inhibits cardiogenesis. Consistently, in loss of function experiments for this gene, the cardiac tube is formed by 116 cells instead of the wild-type number 104 (Lo *et al*., 2002).

The double mutant *Ubx*$^{-}$, *abd-A*$^{-}$ embryos phenocopy the quintuple mutant, that is, there is no cell diversity along the AP axis of the cardiac tube, indicating the crucial requirement of these two genes in the control of thoracic versus abdominal cell identity during early stages of cardiac development (Ponzielli *et al.*, 2002).

In *Ubx* loss of function embryos, the cells of the posterior aorta (segments A1 to middle of segment A4), where *Ubx* is normally expressed, have some defects in polarization and pericardial cells of this domain are disorganized. On the other hand, the heart of *Ubx* loss of function embryos is similar to that of wild-type embryos. Ectopic expression of *Ubx* impairs morphology of the heart cardioblasts and the *ostia* do not become functional (Ponzielli *et al.*, 2002).

In *abd-A* loss of function embryos, the cells of the heart are transformed into posterior aorta-like cells (*i.e.*, those cells which express *Ubx*) and thus *ostia* function and heartbeat are not observed in living embryos (**Figure 6**). Ectopic expression of *abd-A* transforms the aorta morphologically and physiologically into heart (Ponzielli *et al.*, 2002; **Figure 6**). Thus, *abd-A* controls heart identity.

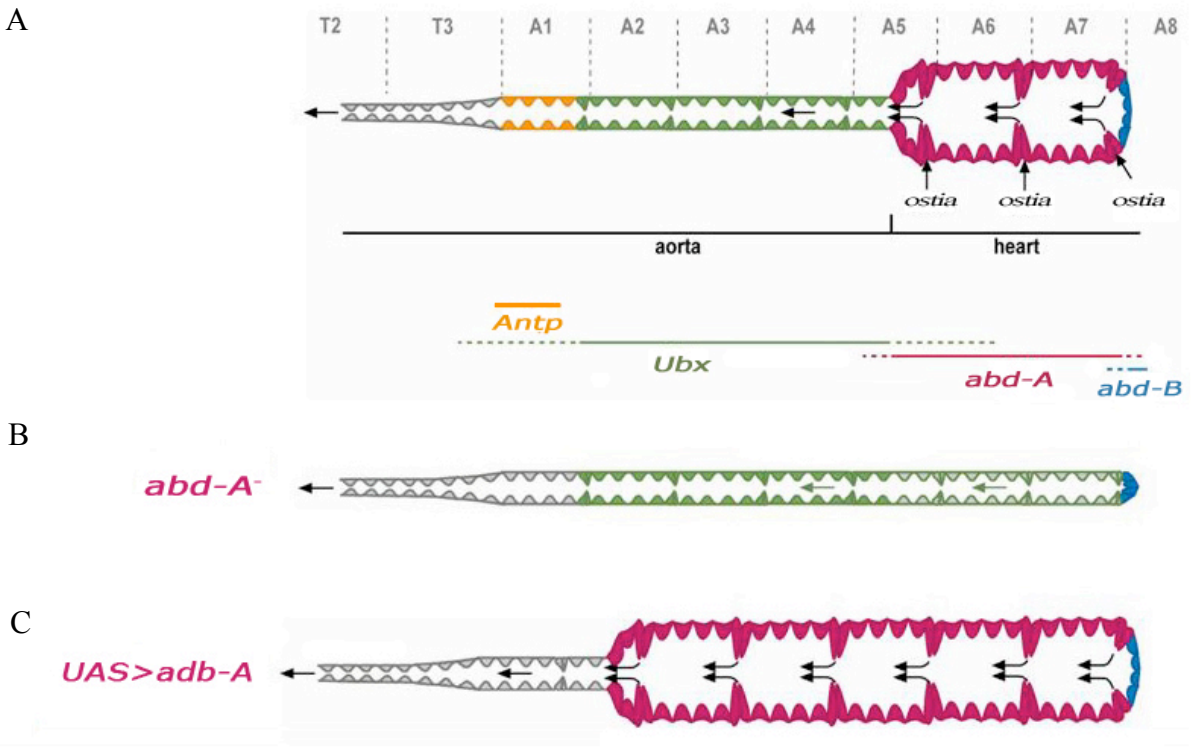


Figure 6 (from Ponzielli *et al.*, 2002). **Control of heart identity by *abd-A*.** (A) Expression of *Hox* genes in a wild-type embryo. (B) Transformation of the heart into aorta in an *abd-A*$^-$ embryo. (C) Transformation of the aorta into heart in an embryo overexpressing *abd-A* (*UAS>abd-A*).

How does *abd-A* control heart identity?

Hox genes encode transcription factors, so in order to control morphogenesis they regulate downstream effector genes. These effector genes can be other transcription factors, signaling molecules or realizator genes. In order to find out which are the *Hox* target realizator genes and in order to determine if these are direct or indirect targets we need to find and study the *cis*-regulatory sequence of such genes.

Three genes, *Ndae1, Ih* and *Ork1* have been found to be expressed in the TIN-positive cells of the heart portion of the cardiac tube (Perrin *et al.*, 2004; Monier *et al.*, 2007). These encode channel proteins: *Ndae1* encodes a Na^+ driven anion exchanger belonging to the NBC family (Romero *et al.*, 2000), *Ih* encodes a Na^+/K^+ hyperpolarization activated channel belonging to the HCN family, and *Ork1* a two pore domain K^+ channel (Perrin *et al.*, 2004; Monier *et al.*, 2007). The latter one, is involved in heart rate and heartbeat activities in the fly (Lalevée *et al.*, 2006) and *Ih* possibly generates the cardiac pacemaker I_f or I_h currents (Occor *et al.*, 2007). Thus, these genes constitute part of the realizators of the heart. Moreover, these genes have concomitant expression to that of *abd-A* in wild-type as well as in gain and loss of function experiments (Perrin *et al.*, 2004). To find the heart *cis*-regulatory sequences of these putative *abd-A* target realizator genes constitutes the main aim of our project. The importance of understanding how realizator genes are regulated in normal heart development will help us understand the mechanisms underlying Congenital Heart Diseases (CHDs) From these experiments we will also gain insight on *Hox* specificity during development.

The late role of *abd-A* and *Ubx* in cardiogenesis

There is only indirect evidence, from loss and gain of function experiments, for the role of *abd-A* as an activator of putative target genes such as *Ndae1, Ih* and *Ork1*. The possible role of *Ubx* as a repressor in late cardiogenesis comes from the observation of a slight loss of heart expression of *Ndae1* (Perrin *et al*., 2004), *Ih* and *Ork1* (Bruno Monier, Laurent Perrin and Michel Sémériva, unpublished) upon ectopic expression of *Ubx*. Therefore, in such embryos the loss of expression of the target genes in the heart may be due to the forced posterior expression of the repressor. On the other hand, in the same embryos, the target genes may be activated by normal presence of *abd-A*. In wild-type embryos, *abd-A* normally represses *Ubx*. In *abd-A*$^{-}$ embryos, *Ubx* is de-repressed posteriorly, and therefore the lack of expression of the target gene in these mutant embryos could be due either to repression of *Ubx* or to lack of activation by *abd-A*. Given the fact that *abd-A* and *Ubx* are involved in lineage specification at earlier stages of cardiogenesis (Perrin *et al*., 2004), it is not possible to asses the expression of putative target genes in double mutant *Ubx*$^{-}$, *abd-A*$^{-}$ embryos, since these embryos lack cardiomyocyte diversity and all the cells of the cardiac tube resemble those cells of the anterior aorta that normally express *Antp* (Perrin *et al*., 2004).

We designed one experiment to resolve this issue. *UAS>abd-A*^{Hx} are flies that carry an inducible mutated hexapeptide variant of *abd-A* (Merabet *et al.*, 2003) capable of inducing normal lineage choice in early cardiogenesis and repressing *Ubx* in the heart. Such protein variant does not lead to ectopic expression of the putative target gene *Ih* (BM, LP and MS, unpublished). We propose here to study *in vivo* the late role of *abd-A* and *Ubx* in the activation/repression of realizator genes by assaying expression in embryos which bear the *UAS>abd-A*^{Hx} transgene in a double mutant *Ubx*$^{-}$, *abd-A*$^{-}$ background. Such embryos will lack *Ubx* and *abd-A* expression but will have appropriate early lineage specification and thereby abdominal cardiomyocytes will be formed. Nevertheless, the abdominal cardiomyocytes will acquire posterior aorta morphology. A lack of *Ih* expression in these embryos indicates a late role of *abd-A* as an activator, since in these double mutant embryos the repressor function of *Ubx* is absent.

A core network of transcription factors to develop a heart

Although *abd-A* controls heart identity, it does not account alone for the regulation of a myriad of target genes that are needed to form a heart and that are also expressed in other tissues or organs. Indeed, heart development is governed by a core network of evolutionarily conserved transcription factors (Olson, 2006; summarized in **Figure 7**).

The first gene to play a role in mesodermal subdivisions encodes the homeodomain transcription factor TIN (Bodmer, 1993). In *tin* mutants the heart fails to form (Bodmer, 1993; Azpiazu & Frasch, 1993), but on the other hand, overexpression of *tin* does not cause ectopic heart induction (Yin & Frasch, 1998; Lockwood & Bodmer, 2002). *Nkx2-5* is the vertebrate homologue of *tin*, but it does not rescue completely the phenotype of *tin* mutants (Park *et al*., 1998; Ranganayaculu *et al.,* 1998). Nevertheless, the repression function of the protein is equivalent in late stages of development (Zaffran *et al*., 2006) meaning that some of their functions during heart development have diverged while others have been conserved.

pnr (Ramain *et al.,* 1993; Winick *et al.*,1993), encoding the ortholog of the vertebrate Zn-finger GATA-4 transcription factor (Gajewski *et al.,* 1999), is expressed only in the cardiogenic region during mesoderm differentiation where it mediates *dpp* signaling and acts in concert with *tin* (Klinedinst & Bodmer, 2003). In *pnr* mutant embryos, *tin* expression is dramatically reduced in the clusters that correspond to the cardiac precursors (Klinedinst & Bodmer, 2003). Ectopic expression of *pnr* can only activate early mesodermal ectopic expression of *tin*, meaning that *pnr* is insufficient to maintain *tin* expression later on (Klinedinst & Bodmer, 2003). Like NKX 2-5

and GATA-4 in vertebrates, TIN and PNR act synergistically (Durocher *et al.*, 1997) to activate differentiation genes like *D-mef2* (Gajewski *et al.*, 2001) and *dHand* (Han & Olson, 2005).

Possibly the oldest of these core heart transcription factors is the MADS-box protein MEF2 (Olson, 2006). Vertebrates have three different *mef2* genes (Potthoff & Olson, 2007) while there is only one in *Drosophila*. The single *mef2* gene in *Drosophila* (*D-mef2*) is expressed in all muscle lineages (Nguyen *et al.*, 1994; Lilly *et al.*, 1994), thus it is expressed in all the cardiomyocytes. *D-mef2* mutant embryos have proper heart specification but cardioblasts fail to differentiate and contractile protein genes fail to be activated (Lilly *et al.*, 1995). Recently, it has been demonstrated that *D-mef2* activates its targets in a temporal dose-dependent manner (Elgar *et al*., 2008).

dHAND is a basic helix-loop-helix transcription factor that, like *D-mef2*, is a direct target of *tin* and *pnr* during cardioblast differentiation (Han & Olson, 2005). *dHand* is expressed in all the cardiomyocytes, pericardial cells as well as in the haematopoietic precursors of the lymph glands (Kolsch & Paululat, 2002; Han & Olson, 2005). *dHand* mutant embryos display abnormal cardiac morphology and a deficient number of pericardial cells is observed (Han *et al.*, 2006). HAND factors are also involved in vertebrate cardiogenesis (Yamagishi *et al.*, 2001). Moreover, human *Hand* genes rescue *Drosophila dHand* mutant embryos, suggesting an evolutionarily conserved role of these factors in mammalian cardiogenesis (Han *et al.*, 2006).

The TBX family of transcription factors is also necessary to build a heart. There are eight T-box genes in the *Drosophila* genome. Five of them, *Dorsocross 1-3* (*Doc*) (Reim *et al*., 2003, 2005), *H15* and *midline* (*mid*) (Miskolczi-McCallum *et al*., 2005; Qian *et al.*, 2005; Reim *et al.*, 2005) are expressed in the cardiac tube. The three *Doc* genes are expressed in the *svp*-expressing cells of the heart (Reim & Frasch, 2005) while *H15* and *mid* are expressed in all the cardiomyocytes of the cardiac tube (Miskolczi-McCallum *et al*., 2005; Qian *et al.*, 2005; Reim *et al.*, 2005). *Doc2* and *Doc3* mutants display severe defects in cardiac cell specification as shown by the absence of multiple markers for the cardioblast population (Reim & Frasch, 2005). *Doc* interacts with *tin* and *pnr* in cardiac cell specification (Reim & Frasch, 2005). Ectopic expression of *Doc* in conjunction with *tin* and *pnr* results in cardioblast differentiation, suggesting these three factors act synergistically during dorsal vessel development (Reim & Frasch, 2005), just like the vertebrate TBX20 directly interacts with NKX2-5 and GATA factors (Stennard *et al.*, 2003).

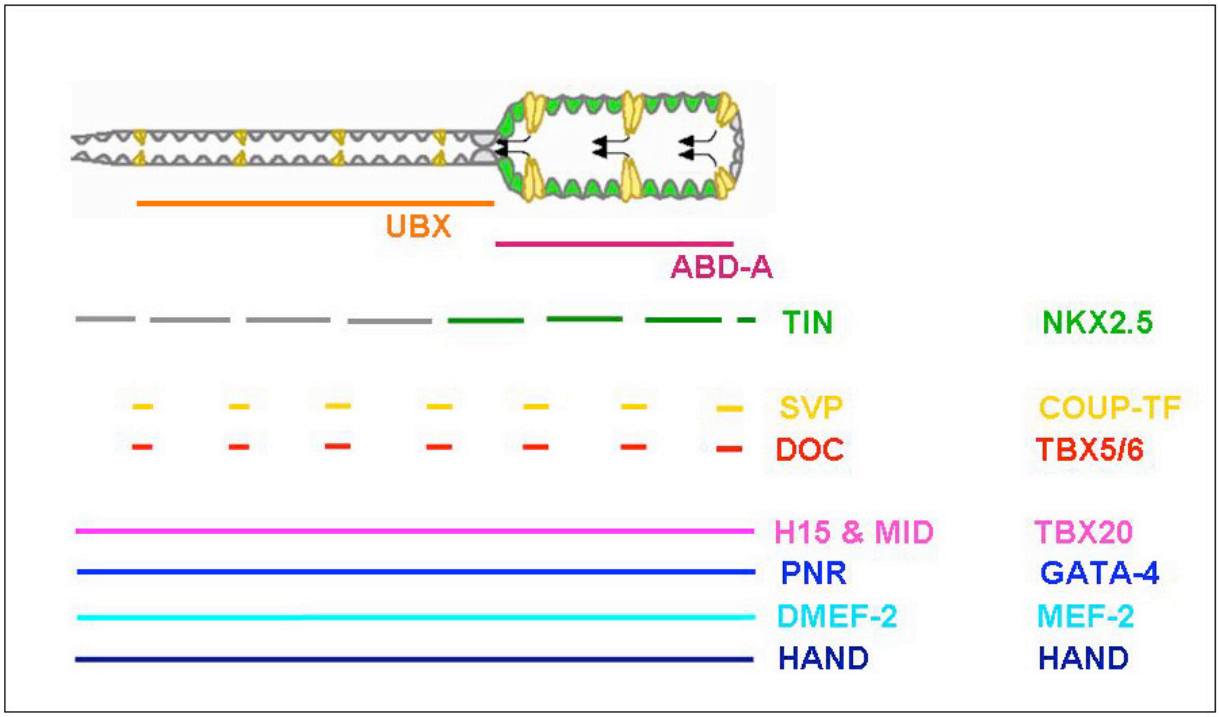


Figure 7. Cardiac expression of transcription factors involved in cardiac development. (A) Diagram representing the cardiac tube divided in aorta and heart. Below are shown the domains of expression of *Ubx* and *abd-A*. (B) The colored horizontal bars indicate the domains of expression in the cardiac tube of the transcription factors indicated to their right. On the right column are indicated the vertebrate homologs of the fly genes.

All these transcription factors involved in heart development bind to DNA through characteristic binding sites. The binding site of a given transcription factor can be represented in the form of a Positional Weight Matrix (PWM). A PWM is a matrix that is created from a collection of *in vitro* footprints of a given RNA or DNA-binding protein. Such biding sites are aligned, and a score is assigned to the probability of occurrence of a certain base in a certain position (Hertz & Stormo, 2002). In this way, binding sites that slightly diverge from the consensus sequences can be also represented. Due to the sequencing of *Drosophila* genomes and the availability of information of the sequence of binding sites of many transcription factors, it is possible to assess through bioinformatic methods the presence of putative *in vivo* binding sites of transcription factors in a given DNA sequence (Stormo, 2000; Osada *et al*., 2004). In the past, *in vitro* inmunoprecipitation experiments were used to look for putative *in vivo* binding sites of a transcription factor to the DNA sequence of a putative target gene. For example, the visceral mesoderm enhancer of *dpp* was found by inmunoprecipitating a genomic fragment containing the enhancer with the *Hox* protein *Ubx* (Capovilla *et al.*, 1994). Bioinformatic tools allow to assess the presence of a number of transcription factor binding sites simultaneously and moreover to assess their spatial distribution in order to determine if such transcription factor binding sites form a cluster which could stand *in vivo* for a *cis*-regulatory motif of a given gene (Halfon *et al*., 2002; Markstein *et al.*, 2002; Grad *et al*., 2004; Berman *et al.*, 2004). Using bioinformatic tools, the binding sites of the mentioned evolutionarily conserved transcription factors in the core network can be found in the sequence of the heart realizator genes.

In first instance, we analyzed through bioinformatics if the binding sites of these conserved transcription factors are present in the putative *cis*-regulatory sequences of the putative *abd-A* target genes expressed in the heart portion of the cardiac tube. We looked for conserved clusters of binding sites of the mentioned transcription factors that might be forming

part of the enhanceosome regulating realizator gene expression in the heart. On second place, we are currently studying which of these transcription factors might cooperate or collaborate *in vivo* with *abd-A* in the regulation of the putative *Hox* heart targets in order to realize a heart with its morphological and physiological distinct cell types. As some of the human homologues of these transcription factors are involved in congenital heart diseases (in particular, NKX2-5, TBX5 and GATA4) the identification of the *cis*-regulatory sequences of their target genes may help elucidate what are the molecular abnormalities of the altered proteins in the patients and help device eventual therapeutic strategies.

MATERIALS AND METHODS

A. Bioinformatics

The bioinformatic approaches applied in this project have been developed through time as more informatic tools evolved from 2004 to present. All bioinformatic procedures applied here are the product of a collaboration with experts in this research area (David Martin and Bruno Zeitouni at the IBDML of CNRS in Marseille, France; Stein Aerts at the Laboratory of Neurogenetics, Dept. of Molecular and Developmental Genetics, University of Leuven, Belgium).

1 - First bioinformatic approach based on available online tools

Most methods used up to 2004 identified *cis*-regulatory sequences from interspecies sequence comparison. They identified Conserved Non coding Sequences (CNSs), operationally defined as islands of non coding sequence with relatively high conservation flanked by regions of low conservation and assumed that this conservation reflected regulatory function (Bergman *et al*, 2002). The benchmark set by the new toolkits developed in 2004 was that of searching for conserved clusters of known transcription factor binding sites within conserved regions of the *Drosophila* genome.

The sequence of a given transcription factor binding site is expressed as a Positional Weight Matrix (PWM). The source of the PWMs in this boinformatic approach is FlyReg/pollard [http://rana.lbl.gov/~dan/matrices.html]. These PWMs were created by Dan Pollard (Eisen's Lab) from a DNAseI footprint database. The PWMs used from this database are those of the following proteins: DMEF-2, TIN, GATA, TBX, UBX. We used the UBX PWM instead of the ABD-A one because the former was more precise (*i.e.*, was created from a larger set of DNAseI footprints) and in any case UBX and ABD-A bind *in vitro* to the same sequences (Capovilla *et al.*, 1994).

For the transcription factor SVP, which was absent in the mentioned database, and no footprints were available in literature, we created a PWM *de novo* from a few published binding sites of the SVP vertebrate homolog COUP-TF.

In 2004, Berman *et al*. published the eCIS-ANALYST toolkit, which searches in the *Drosophila melanogaster* and *Drosophila pseudoobscura* genomes conserved clusters of Transcription Factor (TF) Binding Sites (BSs). More information about how this tool works can be found at http://rana.lbl.gov/cis-analyst/. At the same time, were published the VISTA tools (Frazer *et al.*, 2004), which are a comprehensive suite of programs and databases for comparative analysis of genomic sequences. More information about this toolkit can be found at http://genome.lbl.gov/vista/index.shtml. We used the VISTA Browser to retrieve CNSs between the gene of interest in the *D. melanogaster* genome and its orthologue in the *D. pseudoobscura* genome sequence and to visualize these CNSs in an alignment. We also used rVISTA (Loots *et al.*, 2002) to search for conserved clusters of TF BSs.

A third approach, based always on the same principle, was the one we set up in collaboration with Dr. David Martin (unpublished). This method applied two strategies. One strategy was to look first throughout the *D. melanogaster* gene of interest for TF BSs using PWMs. We took the genomic sequence including the query gene from the previous Computed Gene (CG) to the next one. In a second step, we aligned this sequence to the orthologue in *D. pseudobscura*, using pre-computed global alignments. We then tested if the TF BS score fell into CNSs. A second strategy of this same approach was designed to circumvent the problem that a few conserved bases inside a non-conserved region of the genome might be missed in the general alignment. We manually introduced in the previously aligned sequence all TF BSs found previously. This allowed us to distinguish aligned sites from conserved sites. We also allowed the program to find TF BSs that slightly deviated from the consensus (*i.e.*, we changed the *p*-value during the scoring step). Moreover, with this method we could analyze different percentages of sequence identity (*i.e.*, 50% conservation and 70% conservation).

The outcome of these three tools used - eCIS-ANALYST, VISTA and our own - were

merged and compared in order to give the selected regions the priority to be cloned. The overlapping was made taking the query gene sequence aligned (either to the *D. pseudobscura* or the *D. virilis* orthologue) and manually drawing all the different outputs. We did not use any statistical methods.

2 - Second bioinformatic approach based on a novel pattern matching based method

Throughout the last three years, our collaborators in Dr. Semeriva's lab found seven genes to be differentially expressed in the heart portion of the cardiac tube, through a candidate gene approach (Monier *et al.*, 2007). Moreover, recently, they undertook a genome-wide approach to gain more knowledge on the heart transcriptome. They performed microarray experiments comparing the larval heart transcriptome with that of the aorta. With this experiment, 144 genes were found to be overexpressed in the heart with respect to the aorta. These constituted a new set of putative downstream target genes under specific *abd-A* regulation. Following these new results came the necessity to improve our bioinformatic methodology for heart *cis*-regulatory discovery.

Summary

A pattern matching based approach was used to search for motifs with PWMs of known TF BSs in CNSs. This approach consists of two strategies, which we named A and B.

In strategy A, we first searched for the CNSs with the best multiple high-scoring ABD-A BS. We then looked for clusters of BS of heart TFs except ABD-A. The same two searches were applied to the CNSs of *D. pseudoobscura*. We then integrated these results and produced a first scoring ranking which is after refined in a final ranking according to the presence of at least three high scoring TF BSs besides that of ABD-A. We called this last ranking list "TopA".

Instead in strategy B, we first searched the best CNSs with clusters of BSs of heart TFs (ABD-A included). We did this separately for the *D. melanogaster* and *D. pseudoobscura* CNSs. We kept the CNSs with the best motif conservation score from alignments. Then, we integrated the results for CNSs of both species and produced a ranking which we called "TopB" list.

We posteriorly overlapped and ranked the lists TopA and TopB to produce a final list containing the best 20 scores that we called "Top20". For these best 20 putative *Cis*-Regulatory Modules (CRM) we used Toucan to draw inside each of them the predicted binding sites.

Our new approach in detail

We used a pattern matching based approach to search for motifs with PWMs of known TF BSs in CNSs. This approach consists of two strategies, which we named A and B. Our first gene dataset from which we started off to validate the approach was composed of 7 genes, expressed in the heart and not in the aorta, which were found by candidate gene approach (*Na^+ driven anion exchanger 1*: *Ndae1; Ca^{++} channel protein β subunit*: *Ca-β; I_h channel*: *Ih; Calcium ATPase at 60A*: *CaP60A; Open rectifier K^+ channel 1*: *Ork1; seizure*: *sei; painless*: *pain*) and 2 genes (*Dro-myosupressin*: *Dms* and *CG15537*) found in the microarray data analysis to have the highest fold change expression in the heart compared to the aorta. The expression of these latter two genes was further corroborated through fluorescent *in situ* hybridization.

This is our first gene set:

- CG4675 (*Ndae1*)
- CG6320 (*Ca-β*)
- CG8585 (*Ih*)
- CG3725 (*Ca-P60A*)
- CG1615 (*Ork1*)
- CG3182 (*sei*)
- CG15860 (*pain*)
- CG15537 (*CG15537*)
- CG6440 (*Dms*)

We then decided to cover all the intergenic region to avoid the problem of missing regulatory elements distantly positioned with respect to the transcription unit of the gene. We delimited the regions containing the query gene plus 20kb upstream and 20kb downstream the transcription unit. We aligned these sequences to their corresponding orthologues in the *D. pseudobscura* genome using SLAGAN and we retrieved all CNSs using CNScan (Bruno Zeitouni, unpublished). The CNSs retrieved match the following parameters:

- 65% identity, 100bp windows
- not exons
- not UTR

We obtained 386 CNSs. We then decided to extend the CNSs with flanking sequence so that they are minimally 300bp long because it is generally acknowledged that an enhancer element is around 300-1000 bp. For example, the cardiac tube enhancer of *dHand* is 300 bp (Han & Olson, 2005) as the *dSur* cardiac tube enhancer (Akasaka *et al.,* 2006; Hendren *et al.*, 2007).

Since VISTA only returns the *D. melanogaster* sequences of the CNSs, the *D. pseudoobscura* sequences are obtained from the UCSC "net" alignments (http://genome.ucsc.edu/index.html?org=D.+melanogaster)

In strategy A, we first searched for multiple high-scoring ABD-A BSs within the CNSs. For doing this, we used MotifLocator (http://homes.esat.kuleuven.be/~thijs/help/help_motiflocator.html) with the ABD-A PWM from FlyReg/pollard (http://rana.lbl.gov/~dan/matrices.html) using a threshold of 0.8. From the output we retained the top 3 scoring hits for each CNS. We then scored each CNS according to the formula:

$$cnsscore \sum_{1}^{3} PMWscore$$

CNSs were then ranked accordingly. The same scoring method is used to score and rank the *D. pseudobscura* CNSs found for these genes. We integrated these ranks using order statistics as described in Aerts *et al.* (2006). The second step of strategy A was to find clusters of BSs of TFs known to be involved in heart development. Below are listed the selected TFs and the source from where the corresponding PWM was taken. The website addresses are cited at the end of the list.

- ABD-B (from Pollard matrices)
- NKX25.01 (high affinity, from TRANSFAC): homologous to the *D. melanogaster* TIN
- NKX25.02 (low affinity, from TRANSFAC): homologous to the *D. melanogaster* TIN
- GATA (from TRANSFAC): orthologous to the *D. melanogaster* PNR
- MEF2A (MADS box, from Jaspar): homologous to the *D. melanogaster* D-MEF2
- SRF (MADS box, from ClusterBuster): homolog to the *D. melanogaster* D-MEF2
- TBX core motif "GGTGT" (from Gosh *et al.*, 2001): same binding site than the *D. melanogaster* TBX factors H15, MID and DOC (Frasch lab, unpublished).
- EBOX (CANNTG): orthologous to the *D. melanogaster* dHAND

***TRANSFAC:** http://www.biobase-international.com/pages/index.php?id=transfac
***Jaspar:** http://jaspar.genereg.net/
*** ClusterBuster:** http://zlab.bu.edu/cluster-buster/cbust.html

We then used ClusterBuster setting the gap parameter at 20 and the motif threshold equal to 6 in order that predicted motifs will have relative high specificity and thus reduce the number of false positives. We ranked all CNSs of *D. melanogaster* according to the ClusterBuster score. We treated *D. pseudobscura* CNSs in the same way and finally we integrated these rankings using order statistics. We then integrated the ABD-A ranking obtained in first place to that obtained with ClusterBuster ranking and again using order statistics, we obtained an overall ranking from these previous two. We then selected those sequences with clusters with binding sites for a minimum 3 different PWMs excluding ABD-A and from these we selected the top 30 best scored in the overall ranking. We called these 30 predictions "TopA".

In strategy B, we first looked for clusters of binding sites for all the above mentioned PWMs, including ABD-A. For this, again, we used ClusterBuster, but we ranked them according to the CNS score as defined in strategy A. We treated the CNSs obtained from *D. pseudobscura* in the same way. Once more, we integrated these rankings using order statistics. In a second step we looked at the evolutionary conservation of the binding sites in a cluster. We retrieved all the binding sites found in *D. melanogaster* CNSs. From a LAGAN alignment of each CNS we retrieved all the aligned sequence in *D. pseudobscura* and looked for the binding sites of *D. melanogaster* in these aligned sequences. If the *D. pseudobscura* aligned site was of the same length as that of *D. melanogaster*, we counted the number of identical nucleotides. The sum of all conserved base pairs within all binding sites of a cluster constituted our conservation score. We ranked separately all the CNSs of *D. melanogaster* and *D. pseudobscura* according to this conservation score and we integrated the obtained rankings with order statistics. In the same way as we had done in strategy A, we selected those CNSs with clusters of binding sites containing a minimum of three different PWMs including ABD-A and from these we selected the top 30 best scored in the overall ranking. We called these 30 predictions "TopB".

We then proceeded to integrate the "TopA" and "TopB" lists according to order statistics thus creating a final "Top20" list with the best 20 putative CRMs. We used Toucan (www.esat.kuleuven.be/~saerts/software/toucan.html) to draw inside each of these 20 CNSs the binding sites predicted to be in each CRM.

For *in vivo* validation of these CRMs we arbitrarily selected further parameters as follows:

1) distance of the putative CRM to the transcription start site
2) quality and quantity of the TF BSs found in the CRM:
   - number of ABD-A sites (the more the better)
   - at least a TIN site and a MEF2 site
   - spatial configuration of the BS inside the cluster
3) GO annotations

For cloning of the predicted CRMs, whenever possible, CRMs belonging to same gene were cloned into the same construct in order to reduce the amount of transgenic flies and increase the number of predictions validated. Further details of the transgenic reporter constructs selected for *in vivo* validation are explained in "Results".

Application of this novel approach using as dataset co-regulated genes differentially overexpressed in the heart found through microarray experiments:

After *in vivo* validation of this "Top20", we applied the same strategy using as initial dataset those genes found to be differentially expressed in heart with respect to the aorta through the microarray experiments. The initial dataset contains 144 significant differentially expressed genes ranked according to their expression Fold-Change (FC) in the heart with respect to the aorta. *Dms* and *CG15537* are first in the FC ranking. From this new enlarged dataset we obtained a list of 40 CRMs, which we called "Top40". Further details of the transgenic reporter constructs selected for *in vivo* validation are explained in "Results".

B. Cloning

The molecular biology techniques described below have been adjourned from "Molecular Cloning" (Sambrook *et al.*, 1989).

To obtain the fragments to be cloned, either restriction or polymerase chain reaction (PCR) methods were used according to the availability of restriction sites present in the fragment of interest.

1 - Vectors used

The vectors used for this project are:

Blue/white selection vectors

pBluescript KS^+ (Stratagene)
pUC19

Transformation reporter vectors

CHAB (Capovilla *et al.,*1994)
pStinger (Barolo *et al*., 2004)
pH-Stinger (Barolo *et al*., 2004)

2 - Polymerase Chain Reaction (PCR)

Primer design

For amplification by PCR, primers were designed using the Invitrogen online tool "OligoPerfect Designer" (http://www.invitrogen.com/content.cfm?pageid=9716).

To each primer, we added a restriction site not present in the fragment to be amplified, but present in the polylinker of the vector, to enable ligation of such fragment to the vector.

Afterwards, primers were checked for primer-dimers and secondary structure using the online tool "Oligo Calc" (http://www.basic.northwestern.edu/biotools/oligocalc.html).

PCR reactions

The DNA polymerase used in all reactions is KOD XL polymerase (Novagen), a proofreading enzyme isolated from the extreme thermophile *Thermococcus kodakaraensis* KOD1, which possesses superior processivity and fidelity that enables faster and more accurate PCR amplification than that which can be achieved with conventional enzymes, including *Pfu* DNA polymerase.

The thermocycler used in every case is Px2 Thermal Cycler (Thermo Scientific).

3 - Restriction Reactions

Restrictions with the appropriate restriction enzyme/s were carried out using one unit of enzyme for each microgram of DNA in the restriction buffer recommended by the company in a final volume according to the amount of DNA to be cut. Each restriction was carried out at 37^{o}C for at least one hour.

4 - Purification of vector and insert

Restricted vectors were always purified from TAE 1% agarose gel using the QIAquick Gel Extraction Kit (QIAGEN). The same procedure was utilized for inserts when we had not obtained a unique PCR product. Instead, inserts where only the desired PCR product had been obtained where restricted and purified using the QIAquick Gel Extraction Kit without loading them on gel.

5 - Ligation

The nanograms of insert to put in the ligation reaction are the product of the proportion in size of insert to vector by the nanograms of vector by the excess number of molecules of insert with respect to vector, according to the following formula:

(kb of insert : kb of vector) x ng of vector x 5-10 (excess number of molecules) = ng of insert

The mixture with the appropriate amounts of insert and vector was heated up at 65^{o}C for 20 minutes. An equal volume containing a mix of the T4 ligase and buffer (Takara Ligation kit) was added to the mixture. This final ligation mixture was incubated for at least one hour at 18^{o}C.

6 - Transformation

The plasmids constructed were transformed into Calcium-competent cells of the *Escherichia coli* DH5α strain lacking ampicillin resistance.

Bacterial cells were incubated on ice with the DNA of interest for 30 minutes. Transformation was performed by heatshock at 42^{o}C for 1 minute and subsequent incubation on ice for 5 minutes. Recovery of the cells and resistance expression were allowed by shaking the cells in 1ml of Luria Broth (LB) for 20 minutes.

Recovered cells were plated on LB agar plates containing 100 μg/ml of ampicillin.

7 - Minipreps

Screening of the colonies was done using Qiagen miniprep protocol (http://www1.qiagen.com), without column purification.

8 - Midipreps

Amplification of the quantity of the plasmid of interest was done growing 50ml of plasmid-containing bacterial culture and purifying the DNA with the QIAgen tip100 midiprep kit. In every case, the DNA was resuspended in 100 μl of TE (10mM Tris-HCl pH 7.5, 1mM EDTA).

9 - Glycerol stocks

Long-term conservation of the plasmid of interest at -80°C was achieved by adding 1 ml of 50% sterile glycerol to 0.5 ml of midiprep cell culture.

C. Transgenesis

1 - Preparation of DNA for injection

The midiprep DNA to be injected was centrifuged for 20 minutes to eliminate any insoluble particles that might have remained, which could clog the needle. The supernatant was transferred to a new tube and quantified with the spectophotometer.

The plasmids to be injected bear a transformation vector containing a *P* element and were co-precipitated with the so called "helper plasmid" (pIChIID2-3, Ken Irvine, personal communication) bearing the transposase necessary for the integration of the *P* element and thus of the construct of interest into the *Drosophila* genome. The construct to be injected was mixed with helper plasmid in a ratio 6:1 (12 ug +2 ug) and co-precipitated in a final volume of 50 μl by addition of 0.3M NaOAc and 2.5 volumes of absolute ethanol.

After centrifugation the pellet was washed in 70% ethanol for 3 minutes and resuspended in 20 ml of injection buffer (5mM KCl, 0.1mM $NaPO_4$, pH 6.8). The DNA was recentrifuged, transferred to a clean new tube and only 2 μl were loaded into a Femptoneedle.

2 - Egg laying and preparation of the embryos to be injected

Egg laying

The flies were put to lay in cages on a molasses agar medium (see Appendix). Flies were submitted to a normal photoperiod, thus eggs were usually laid from the beginning of the afternoon until late in the evening.

As the transformation vectors used carry the mini *white* gene, the strain of flies used for egg laying were mutant for this gene. This strain is w^{1118} carrying a deletion of the whole endogenous white gene.

Dechorionation, dissection and alignment of the embryos

Embryos were recovered from the molasses agar plate with a brush and washed with milliQ water. Dechorionation of the embryos was done using 40% bleach for 1 minute, followed by washing with milliQ water.

After dechorionation, embryos were aligned on a piece of agar plate with the same anterior to posterior direction for 7-10 minutes, since embryos must be injected before blastoderm cellularization, a developmental stage that begins 45-50 minutes after eggs are laid at 22°C. Then embryos were transferred to a double-sticky tape on a coverslip. The embryos aligned on the coverslip were put to dessicate for 4 minutes in a chamber containing silica gel. After dessication, embryos were covered with Halocarbon oil, and the coverslip was placed on a slide under a Nikon SMZ 645 microscope, as shown in **figure 8** below.

The injector used was an Eppendorf Femtojet Easy.

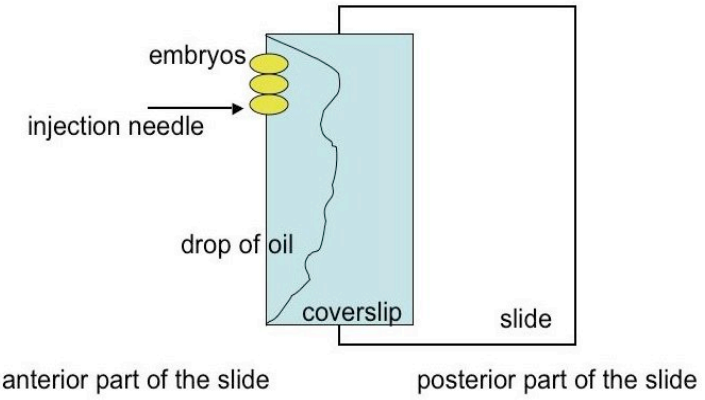


Figure 8. Diagram of embryos to be injected. The posterior part of the embryo, where the needle is introduced is towards the anterior part of the slide.

Treatment of the embryos after injection

The slide with the injected embryos covered with oil was placed on a molasses agar plate and covered with a lid to keep the necessary humidity. The humid chambers were placed at 18°C until the eclosion of the larvae 48 hours later.

Isolation of transformants

First instar larvae were collected from the humid chamber and out of the oil to be placed in a food vial containing a drop of fresh yeast. The vial was placed at 25°C. At this temperature, after 10 to 12 days the new adults emerged from the pupae and were crossed individually to flies of the strain w^{1118} in order to evidence in the next generation the presence of the transgene bearing the w^+ gene and thus the transformants (colored eyed flies) obtained from injections.

The sucessive treatment given to transformants is described on this same chapter "Materials and Methods" under the subtitle "Fly Genetics".

D. Fly Genetics

Summary of symbols

♀ : female
♂: male
w^-: white strain w^{-1118}
Hmz: homozygote
Htz: heterozygote
MASS: mass cross
R!: recombination event
G_n: generation
+: wild-type chromosome

Notes:

* The chromosomes of *Drosophila* are written in the folllowing order: first the X, next the second (II) and last the third (III). Each pair of chromosomes is separated from the other with a

semicolon while the homologous are separated by a slash. Different alleles in the same homologous are separated by a comma.

* *Drosophila* males are achiasmatic.

* In every cross and throughout the generations the genotype of interest and to be selected is highlighted in yellow.

1 - Mapping of transformants

In order to map in which chromosome the transgene landed we crossed a colored eyes male to virgin females of the double balancer line *SM5/Xa/TM3*. *SM5* and *TM3* are balancer chromosomes of the second and third chromosome, respectively. They avoid meiotic recombination because they bear inversions. *Xa* is a translocation between the second and third chromosomes with a breakpoint in the gene *apterous*. *SM5* is characterized by the dominant marker *Curly of Oster* (*CyO*) whose phenotype is Curly wings while *TM3* is characterized by the dominant marker *Stubble* (*Sb*) whose phenotype is short bristles. In G1, males with colored eyes (due to the mini w^- present in the P-element) and with both balancers will be selected to cross to w^- females. In this cross the balancer chromosomes will segregate independently from each other and the P-element will segregate opposite to the homolog where it is inserted.

G0 ♀ w^-/w^-; SM5/Xa/TM3 x ♂ w; **P?**/+

G1 ♂ w/ w^-/ Y; SM5/**P?**/ TM3 x ♀ w^-/w^-; +/+; +/+

G2 Possible genotypes and phenotypes obtained depending on chromosome linkage:

w^- / w^- or Y; **P**/ +; +/ TM3: colored eye flies with Stubble bristles
w^- / w^- or Y; **P**/ +; +/ +: colored eye flies
w^- / w^- or Y; +/ SM5; +/ TM3: white eye flies with Curly wings and Stubble bristles
w^- / w^- or Y; +/ SM5; +/ +: white eye flies with Curly wings

P on the II

or

w^- / w^- or Y; +/ SM5; **P**/+: colored eye flies with Curly wings
w^- / w^- or Y; +/ +; **P**/ +: colored eye flies
w^- / w^- or Y; +/ SM5; +/ TM3: white eye flies with Curly wings and Stubble bristles
w^- / w^- or Y; +/ +; +/TM3: white eye flies with Stubble bristles

P on the III

or

P/ w-; +/+; +/+: colored eye females
w-/Y; +/+; +/+: white eye males

P on the X

2 - Making balanced stocks of transformants

In order to generate balanced stocks of each transgene, virgin females with genotype *w*$^{-}$/*w*$^{-}$; *SM5*/**P?**/*TM3* (obtained in the G1 of the mapping cross described above) are crossed to males of the same genotype (their brothers) if the P-element insertion maps to the second or third chromosome. Instead, if no color eye males where obtained in the G1 of the "mapping cross", the virgin females of the mentioned genotype are crossed to males bearing an X balancer chromosome (FM7i).

A) If the P-element insertion is on the II chromosome:

G0 ♀♀ w^{-}/w^{-}; **P**/SM5; +/TM3 x ♂♂ w w^{-}/w^{-}; **P**/SM5; +/TM3

G1 (STOCK LINE) w^{-}/w^{-}; **P**/ SM5; +/TM3 or w^{-}/w^{-}; **P**/ **P** ; +/TM3

B) If the P-element insertion is on the III chromosome:

G0 ♀♀ w^{-} /w^{-}; +/SM5; **P**/TM3 x ♂♂ w/w^{-}; +/SM5; **P**/TM3

G1 (STOCK LINE) w^{-} /w^{-}; +/ SM5; **P**/TM3 or w^{-} /w^{-}; +/ SM5; **P**/**P**

C) If the P- element insertion is on the X chromosome:

FM7i is a *w*$^{-}$ balancer X chromosome which bears the dominant marker *Bar* whose phenotype is "bar" shape eyes.

G0 ♀♀ **P**/w^{-}; +/SM5; +/TM3 x ♂♂ FM7i/Y; +/+; +/+

G1 ♀♀ **P**/FM7i x ♂♂ **P**/Y

G2 (STOCK LINE) **P**/**P** or Y; +/+; +/+

3 - Putting together two insertions that are on the same chromosome

These crosses are done to obtain individuals that carry in the same chromosome two independent insertions of the same reporter construct or of two different constructs (*e.g.*, a reporter and a UAS construct). In the first case, the aim of this cross was to obtain a stronger signal of the reporter construct since individuals with only one copy of the insertion reported a weak expression of the transgene. In all the transformants obtained for the Ork1-seq9 reporter construct (pMC055), the insertion mapped to the third chromosome. 055-M11 and 055-M9m1 are the names of two independent transformant lines that were recombined through the following crosses:

G0 ♀♀ 055-M11/055-M11 x ♂♂ 055-M9m1/TM3

*From the progeny of this cross collect darker eye virgins not TM3

G1 ♀♀ **(R!)** +/+; 055-M11/055-M9m1 x ♂♂ w^-/w^-; +/+; +/+
(MASS)

*From the progeny of this cross collect 10 males with very dark eyes.
*Cross individually these 10 males:

G2 ♂ 1-10 055-M11, 055-M9m1/+ x ♀♀ SM5/Xa/TM3

*From this cross collect colored eye males and females that are also SM5 and TM3.

G3 (STOCK LINE) ♀♀ x ♂♂ SM5/+; 055-M11, 055-M9m1/TM3

*** In paralell to the G2 cross, we crossed individually the two original parental males carrying only one insertion:**

♂ 055-M9m1/TM3 x ♀♀ w^-/w^-; +/+; +/+

♂ 055-M11 (Hmz viable) x ♀♀ w^-/w^-; +/+; +/+

These two parallel crosses were performed in order to allow comparison of the eye color between the progeny of G2 and the original parental lines.

4 - Putting an allele in a mutant context

In order to analyze *in vivo* how the heart *cis*-regulatory sequences identified behave with respect to the presence of the *Hox* gene *abd-A,* we crossed each reporter construct lines into an *abd-A*$^-$ background. For each reporter construct we crossed two independent lines.
abd-A is located in the third chromosome. *abdA*^{M1} (later indicated as *abd-A*$^-$) is a recessive amorph EMS lethal allele and the line we used is balanced with *TM6b*, a balancer of the third chromosome whose markers are *Humeral* (*Hu*) in adults and *Tubby* (*Tb*) in pupae. *TM6bZ* carries also the *Ubx-lacZ* marker, which makes β-galactosidase in the *Ubx* expression pattern. All the flies used carry the *w*^{1118} mutation on the X chromosome.

A) <u>Putting the *Ork1* heart enhancer (055) in an *abd-A*$^-$ mutant background through recombination</u>:

055-M11 and 055-M9 m1 are the names of two independent transformant lines obtained for Ork1-seq9 reporter construct (055). Both lines mapped the insertion to the third chromosome.

G0 ♀♀ 055-M9m1/TM3
or ♀♀ 055-M11/055-M11 x ♂♂ abdA$^-$/TM6b
(MASS)

* From this cross collect colored eye virgins not *Hu*.

G1 ♀♀ **(R!)** 055/abd-A$^-$ x ♂♂ abd-A$^-$/TM6b
(MASS)

* From this cross collect colored eye *Hu* males.
* Cross individually 100 colored eye *Hu* males.

G2 ♂$_{1\text{-}100}$ 055, abd-A$^-$**?**/TM6b x ♀♀ abd-A$^-$/TM6bZ

Possible genotypes in the progeny:

Recombinant genotypes:
055, abd-A$^-$**/TM6bZ: colored eye flies, TM6b** ⟸ **desired genotype**
055, abd-A$^-$/ abd-A$^-$: lethal
Not recombinant genotypes:
055/abd-A$^-$: colored eye flies, not *Hu*
055/TM6bZ: colored eye flies, *Hu*
Other genotypes:
abd-A$^-$/TM6b: white eye flies
TM6b/TM6bZ: lethal

If the not-recombinant genotypes were missing in the individual cross, we proceeded to make the stock line.

G3 (STOCK LINE) ♀♀ x ♂♂ 055, abd-A$^-$/TM6bZ

B) <u>Putting the *Ih* heart enhancer in an *abd-A*$^-$ mutant background</u>

For the Ih-seq34 reporter construct (077), we obtained transformant lines that mapped the insertion either to the second or third chromosomes. For the line 077-F10, which mapped to the third chromosome, we proceeded with the crosses to obtain recombinant individuals in the same way as described above for the *Ork1* heart enhancer.

For the line 077-M2, which mapped to the second chromosome, in order to obtain a final stock line with the genotype 077-2M/*CyOZ; abdA*$^-$*/TM6Z* containing the two chromosomes of interest, we proceeded in the following way.

***Dichaete* (*D*) is a dominant mutation that results in wings extended at 45 degrees from the body axis and elevated 30 degrees above and missing the alulae. *Lobe* (*L*) is a dominant mutation resulting in smaller eyes.

i) We balanced the allele on the second chromosome to the line *CyOZ*/*Xa*/*TM6bZ*:

G0 ♀♀ 077-2M/077-2M; TM3/+ x ♂♂ CyOZ/Xa/TM6bZ

*From this cross, collect virgins 077-2M/CyOZ; TM6bZ/+ (Curly wings and Humeral bristles)

G1 ♀♀ 077-2M/CyO; TM6bZ/+

ii) In parallel, we crossed the allele on the third chromosome to the line *CyOZ*/*L*; *TM6bZ*/*D*:

G0 ♀♀ +/CyOZ; abd-A$^-$/TM6bZ x ♂♂ CyOZ/L; TM6bZ/D

*From this cross collect males +/L; abd-A$^-$/D (with white Lobe eyes and Diachaete wings, with bristles not Humeral)

G1 ♂♂ +/L; abd-A$^-$/D

iii) We crossed the differentially balanced lines:

G1 ♀♀ 077-2M/CyOZ; TM6bZ/+ **x** ♂♂ +/L; abd-A$^-$/D

*From this cross, collect males 077-2M /L; abd-A$^-$/TM6bZ (with colored Lobe eyes, Humeral bristles, not Dichaete or Curly wings)

G2 ♂♂ 077-2M/L; abd-A/TM6bZ

iv) We rebalanced the males that bear the two alleles of interest:

G2 ♂♂ 077-2M/L; abd-A$^-$/TM6bZ x ♀♀ CyOZ/Xa/TM6bZ

* From this cross, we collected individuals 077-2M/*CyOZ*; *abd-A$^-$/TM6bZ* (with Curly wings and Humeral bristles)

G4: (STOCK LINE) ♀♀ x ♂♂ 077-2M/CyOZ; abd-A$^-$/TM6bZ

C) <u>Putting the *Dms* heart enhancer in an *abd-A$^-$* mutant background</u>

For the Dms-seq5 reporter construct (078) we obtained transformant lines that mapped the insertion either to the second or third chromosomes. For the line 078-F10, which mapped to the third chromosome, we proceeded with the crosses to obtain recombinant individuals in the same way as described above for the *Ork1* heart enhancer. For the line 078-M1, which mapped to the second chromosome, in order to obtain a final stock line with the genotype 078-M1*/CyOZ; abdA$^-$ /TM6Z,* we proceeded in the same way as for the line Ih-seq34 reporter construct line 077-M2, which also mapped to the second chromosome.

5 - Analyzing the heart specific enhancers with respect to the ectopic expression of *Hox* genes

We crossed each reporter construct lines found to drive expression in the heart to flies bearing a *UAS>abd-A* transgene and to flies bearing a *UAS>Ubx* transgene, to analyze *in vivo* the behavior of heart specific *cis*-regulatory sequences with respect to overexpression of each one of the *Hox* genes.

The UAS/GAL4 system

GAL4 encodes a protein of 881 amino acids, identified in the yeast *Saccharomyces cerevisiae* as a regulator of genes induced by galactose. GAL4 activates transcription by directly binding to four related 17 bp sites upstream the transcription start site. These sites are termed Upstream Activating Sequences (UAS) and function similar to a higher eukaryotic enhancer element. In 1988, Fischer *et al.* demonstrated that GAL4 was capable of stimulating transcription of a reporter gene under UAS control in *Drosophila*. GAL4 can be expressed under the control of *D. melanogaster*-specific promoters with little effect upon the organism.

In 1993, Brand and Perrimon published the development of the UAS/GAL4 system, for targeting *in vivo* gene expression in *Drosophila* in a temporal and spatial fashion. In this system, expression of the gene of interest, the responder, is controlled by the presence of the UAS element upstream of the cDNA of interest. Because transcription of the responder requires the presence of GAL4, the absence of GAL4 in the responder lines maintains them in a transcriptionally silent state. To activate their transcription, responder lines are mated to flies expressing GAL4 in a pattern corresponding to the expression pattern of the gene in which the GAL4 construct is inserted. Each GAL4 line is called "driver" line. The resulting progeny then expresses the responder in a transcriptional pattern that reflects the GAL4 pattern of the respective driver (**Figure 9**).

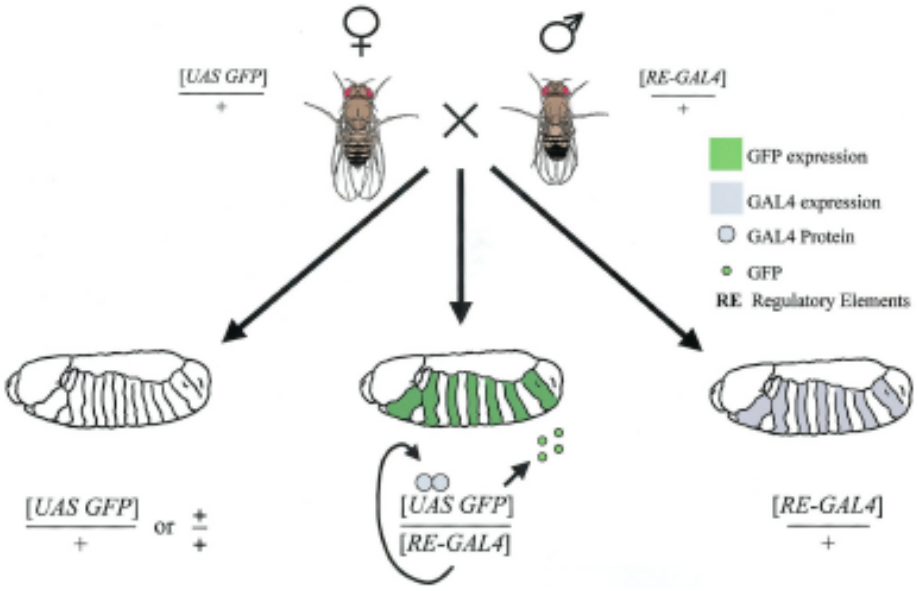


Figure 9. (from Duffy, 2002) **The bipartite UAS/GAL4 system in D. melanogaster**. When females carrying a UAS reponder (*e.g.*, *UAS>GFP*) are mated to males carrying a *RE-GAL4* driver, part of the progeny will carry both *UAS>GFP* and the *RE-GAL4* driver. The presence of *RE-GAL4* in an alternating segmental pattern in the embryos drives expression of the *UAS>GFP* responder gene in a corresponding pattern (depicted in green).

The UAS/GAL4 ectopic expression system is widely used in *D. melanogaster* for the overexpression of transgenes. The *UAS>abd-A* and *UAS>Ubx* transgenes we used are both inserted in the second chromosome. The *24B-Gal4* late mesodermal driver (Zaffran *et al.*, 1997) was used in all cases to trigger expression in the whole mesoderm. This *24B-Gal4* transgene is inserted in the third chromosome.

A) Analyzing the *Ork1* heart enhancer after ectopic *Hox* expression (*abd-A* and *Ubx*):

We crossed the 055-M11, 055-M9 m1/TM3 recombinants to flies *UAS>abd-A* and *UAS>Ubx*. We expected the transheterozygote to have concomitant ectopic expression of nuclear GFP reporter (nGFP) to that of *abd-A*, thus *abd-A* to act as an activator of the target enhancer. We expected a reduction of expression of the nGFP signal upon ectopic expression of *Ubx*, due to the possible role of *Ubx* as a repressor of the target enhancer.

i) Balance the UAS transgene (*UAS>abdA* and *UAS>Ubx*) on the second chromosome with a *lacZ* balancer:

G0 ♀♀ CyOZ/Xa/TM6bZ x ♂♂ UAS>abd-A/CyO; TM6b/MKRS
or
UAS>Ubx/CyO; TM6b/MKRS

*From this cross collect virgins *UAS>abd-A* or *UAS>Ubx /CyO*; *TM6b/MKRS* (with Curly wings and Humeral bristles).

G1 ♀♀ UAS>abd-A/CyO; TM6bZ/MKRS
and
♀♀ UAS>Ubx/CyO; TM6bZ/MKRS

ii) In parallel, cross the reporter on the third chromosome:

G0 ♀♀ CyOZ/L; TM6bZ/D x ♂ +/+; 055-M11, 055-M9 m1/TM3

*From this cross, collect males +/L; III/D (with Lobe eyes and Diachaete wings)

G1 ♂♂: +/L; 055-M11, 055-M9 m1/D

iii) Cross the progeny of the newly balanced flies:

G1 ♀♀: UAS>abd-A/CyO; TM6bZ/MKRS x ♂♂: +/L; 055-M11, 055-M9 m1/D
and
♀♀: UAS>Ubx/CyO; TM6bZ/MKRS x ♂♂: +/L; 055-M11, 055-M9 m1/D

*From this cross collect males UAS>abd-A or UAS>Ubx /L; 055-M11, 055-M9 m1/TM6bZ (with Lobe eyes and Humeral bristles)

G2 ♂♂: UAS *abdA*/L; 055-M11, 055-M9 m1/TM6bZ
and
♂♂ UAS>Ubx/L; 055-M11, 055-M9 m1/TM6bZ

iv) Rebalance the new line containing the two alleles of interest:

G3 ♂♂ UAS>abd-A/L; 055-M11, 055-M9 m1/TM6bZ x ♀♀ CyOZ/Xa/TM6bZ
and
♂♂ UAS>Ubx/L; 055-M11, 055-M9 m1/TM6bZ

*From this cross, collect individuals *UAS>abd-A* or *UAS>Ubx /CyOZ*; 055-M11, 055-M9 m1/ *TM6bZ* (with Curly wings and Humeral).

G4 (STOCK LINE) ♀♀ x ♂♂ UAS>abd-A/CyOZ; 055-M11, 055-M9 m1/TM6bZ
and
UAS>Ubx/CyOZ; 055-M11, 055-M9 m1/TM6bZ

v) Cross the new stock balanced lines from G4 to trigger ectopic expression:

G5 ♀♀ UAS>abd-A/CyOZ; 055-M11, 055-M9 m1/TM6bZ x ♂♂ +/+; 24BG4/24BG4
and
♀♀ UAS>Ubx/CyOZ; 055-M11, 055-M9 m1/TM6bZ x ♂♂ +/+; 24BG4/24BG4

* The embryos of this cross were fixed and stained using both classical immunohistochemistry and FISH. The protocols used are described in this chapter of "Materials and Methods" under the subtitle "Immunohistochemistry" and "Fluorescent *in situ* Hybridization".

B) Analyzing *Ih* heart enhancer under *Hox* ectopic expression (*abd-A* and *Ubx*)

For the Ih-seq34 reporter construct (077) line 077-F10, which mapped to the third chromosome, we proceeded with the crosses in the same way as described above for the *Ork1* heart enhancer to put the two chromosomes of interest in the same individual and cross the new line to the 24B-GAL4 driver.

For the line 077-M2, which mapped to the second chromosome, in order to obtain a final recombinant lines with the genotype *UAS>abd-A,* 077-M2/CyOZ and *UAS>Ubx,* 077-M2/CyOZ we proceeded in the following way:

G0 ♀♀ 077-M2 / 077-M2; +/+ x ♂♂ UAS>abd-A/CyO; MKRS/Tb
and
♂♂ UAS>Ubx/CyO; MKRS/Tb

* From the progeny of this cross collect darker eye virgins not CyO.

G1 ♀♀ **(R!)** 077-M2/UAS>abd-A x ♂♂ w^-/w^-; +/+; +/+
or (MASS)
♀♀ **(R!)** 077-M2/UAS>Ubx x ♂♂ w^-/w^-; +/+; +/+

*From the progeny of this cross collect 10 males with very dark eyes and cross them individually.

G2 ♂ 1-10 077-M2, UAS>abd-A/+ x ♀♀: CyOZ/L
or
♂ 1-10 077-M2, UAS>Ubx/+

* From this cross, collect colored eye males and females which are also SM5 and TM3.

G3 (STOCK LINE) ♀♀ x ♂♂ 077-M2, UAS>abd-A/CyOZ

or

♀♀ x ♂♂ 077-M2, UAS>Ubx/CyOZ

*** In paralell to the G2 cross, we crossed individually the two original parental males carrying only one insertion,** in order to allow comparison of the eye color between the progeny of G2 and the original parental lines.

1♂ 077-M 2/077-M 2 x ♀♀ w^-/w^-; +/+; +/+

1♂ UAS>abd-A/CyO x ♀♀ w^-/w^-; +/+; +/+

1♂ UAS>Ubx/CyO x ♀♀ w^-/w^-; +/+; +/+

G4: ♀♀ 077-M2, UAS>abd-A/CyOZ x ♂♂ +/+; 24BG4/24BG4

or

♀♀ 077-M2, UAS>Ubx/CyOZ x ♂♂ +/+; 24BG4/24BG4

* The embryos of this cross were fixed and stained using both classical immunohistochemistry and FISH. The protocols used are described in this chapter of "Materials and Methods" under the subtitle "Immunohistochemistry" and "Fluorescent *in situ* Hybridization", respectively.

C) <u>Analyzing the *Dms* heart enhancer after *Hox* ectopic expression (*abd-A* and *Ubx*)</u>

For the Dms-seq5 reporter construct (078) line 078-F10, which mapped to the third chromosome, we proceeded with the crosses in the same way as described above for the *Ork1* heart enhancer to put the two chromosomes of interest in the same individual and cross the new line to the 24B-GAL4 driver. Instead, for the line 078-M1, which mapped to the second chromosome, in order to obtain a final stock line with the genotype *UAS>abd-A,* 078-M1/*CyOZ* and *UAS>Ubx,* 078-M1/*CyOZ*, we proceeded as for the Ih-seq34 reporter construct line 077-M2 in order to obtain recombinant lines and cross them to the 24B-GAL4 driver.

6 - Determining the late *in vivo* role of *abd-A* during heart development

The following fly crosses where done to gather information on the late role of *abd-A* during heart development. As explained in the Introduction, there is only indirect evidence for the role of *abd-A* as an activator and *Ubx* as a repressor in late cardiogenesis and this is due to the problem that they are also involved in lineage specification at earlier stages of cardiogenesis. To overcome this situation we recombined *UAS>abd-A*^{Hx} flies (Merabet *et al.*, 2003) with flies carrying the so called "Deficiency 109" (*Df109*), which removes both *abd-A* and *Ubx*. *UAS>abd-A*^{Hx} are flies which carry a UAS construct carrying a mutated hexapeptide variant of *abd-A* capable of promoting normal lineage choice in early cardiogenesis and repressing *Ubx*, but does not lead to ectopic expression of the putative target gene *Ih* (BM, LP and MS, unpublished). *Df 109* flies carry a deletion on the third chromosome that removes *abd-A* and *Ubx* simultaneously. Recombinant *UAS>abd-A*^{Hx}, *Df109* flies where then crossed to recombinant *24B-Gal4*, *Df109* flies. The latter line can force expression of *UAS>abd-A*^{Hx} in the mesoderm and is also deficient in *abd-A* and *Ubx*. The homozygous *Df109* progeny of the cross between *UAS>abd-A*^{Hx}, *Df109* and *24B-Gal4*, *Df109* flies have appropiate lineage choice and should show the late role of *abd-A* in the absence of endogenous *abd-A* and *Ubx*.

A) <u>In order to obtain UAS>*abd-A*^{Hx}, *Df109* flies the following crosses were made</u>:

The *UAS>abd-A*^{Hx} flies where courtesy of Y. Graba (LGPD, Marseille, France). *UAS>abd-A*^{Hx5} and *UAS>abd-A*^{Hx21} are two independent lines for which the P-element insertion was mapped to the third chromosome. The two independent lines where crossed separately in this experiment.

Df109 is a deletion on the third chromosome. As *abd-A*$^{-}$ is a recessive allele the presence of the deletion will be detected by the presence of a *Ubx* haploinsufficient phenotype (big haltere).

G0: ♀♀ Df109/TM6b x ♂♂ UAS>abd-A^{Hx5} or UAS>abd-A^{Hx21}

* From this cross collect virgins with bristles not Humeral

G1: ♀♀ **(R!)** Df109/ UAS>abd-A^{Hx} x ♂♂ abd-A^{-}/TM6b (MASS)

*Cross individually 100 males colored eye males with Humeral bristles and Ubx^{-}.

G2: ♂$_{1-100}$ Df109**?**, UAS>abd-A^{Hx}/TM6b x ♀♀ *abd-A*/TM6bZ

*From this cross collect flies that are colored eye, with Humeral bristles and Ubx^{-} phenotype.

G3: (STOCK LINE) ♀♀ x ♂♂ Df109, UAS>abd-A^{Hx}/TM6bZ

B) <u>In order to obtain *24B-Gal4*, *Df109* flies the following crosses were made</u>:

G0 ♀♀ Df109/TM6b x ♂♂ 24B-G

* From this cross collect virgins with bristles not Humeral.

G1 ♀♀ **(R!)** Df109/24B-G4 x ♂♂ abd-A^{-}/TM6b (MASS)

*Cross individually 100 males colored eye males with Humeral bristles and Ubx^{-} phenotype.

G2 ♂$_{1-100}$ Df109, 24B-G4/TM6b x ♀♀ abd-A/TM6bZ

*From this cross collect colored eye, with Humeral bristles and Ubx^{-} phenotype.

G3 (STOCK LINE) ♀♀ x ♂♂ Df109, 24B-G4/TM6bZ

C) <u>In order to obtain *UAS>abd-A*^{Hx}, *Df109/24B-Gal4*, *Df109* embryos the following cross was made</u>:

The embryos of the cross below were fixed and stained using both classical immunohistochemistry and FISH. The protocols used are described in the chapter "Materials and Methods" under the subtitles "Immunohistochemistry" and "Fluorescent *in situ* Hybridization" respectively.

G4 ♀♀ Df109, UAS>abd-A^{Hx}/TM6bZ x ♂♂ Df109, 24B-G4/TM6bZ

E. Immunohistochemistry

1 - Fixing of embryos

Embryos were collected from the agar plate of the cage with a brush and transferred into a basket with a bottom made of nylon mesh. The embryos in the basket were rinsed with water and then were dechorionated using 50% bleach for 2 minutes and rinsed again with water. The dechorionated embryos on the mesh were immersed in 800 µl of heptane. Fixing was done by addition of 300 µl of fix buffer (see Appendix) and 200 µl of 10% formaldehyde to each tube with the embryos in heptane. The embryos were left to rotate for 30 minutes. After fixation, embryos were devitellinized with 100% methanol, rinsed twice in 100% ethanol and conserved at -20^{o}C in 100% ethanol.

2 - Staining of fixed embryos

Antibody preabsortion

For antibody preabsorption, antibodies were incubated overnight at 4^{o}C in a 1:10 dilution in a solution of BBT (see Appendix) with an amount of embryos equal to one fifth of volume of antibody added. The antibody was recuperated avoiding the embryos and an equal volume of 100% glycerol was added to preserve the protein at -20^{o}C.

Antibody staining

Embryos in 100% ethanol were re-hydrated by washing them for 5 minutes in a solution 1:1 100% ethanol: PBT (see Appendix). Afterwards, embryos were washed twice in BBT-250 (see Appendix) and incubated overnight at 4^{o}C in 500 µl of BBT-250 with preabsorbed antibody diluted to a final concentration of 1:1000.

The next morning embryos were washed two times ten minutes in BBT-250 and other two times ten minutes in blocking solution of BBS-250 (see Appendix). After these washes embryos were incubated for 2 hours at room temperature in a BBS-250 solution containing preabsorbed biotinylated secondary antibody in a final concentration of 1:500.

Following the secondary antibody embryos were washed thrice in 1 ml of PBT and were further incubated for one hour in a solution of 500 µl PBT containing avidin and biotinylated peroxidase (Vectastain HRP Elite kit, Vectorlabs). This step was required to amplify the signal. After incubation, embryos were washed another three times in 1 ml PBT.

Embryos staining was done with a solution containing 1 ml PBT, 100 µl DAB (5mg/ml in PBS) and 5-10 µl of 0,3% H_2O_2. Staining was blocked by rinsing in PBT.

Mounting of stained embryos

Stained embryos in PBT were dehydrated by rinsing them first in 50% ethanol and then in 100% ethanol. A posterior wash in methyl salicylate was necessary to digest the yolk, after what the salt was rinsed with 100% ethanol. Mounting of the embryos was made on a slide with Canada Balsam. Embryos were studied in a Nikon SMZ 645 microscope equipped with Hoffman optics.

F. Fluorescent *in situ* Hybridization (FISH)

Below are schematically described the steps undertaken to perform FISH on embryos :

Hybridization

1) Incubate devitellinized embryos in 500µl 100% ethanol and 500µl xylene for 30 minutes
2) Wash the embryos five times in 100% ethanol
3) Wash the embryos twice in methanol
4) Wash the embryos three times in PBT for five minutes each wash
5) Inactivate the endogenous peroxidase with: 3% H_2O_2 (50µl 30% H_2O_2 + 450µl PBT) in PBT washing for 10 minutes
6) Wash the embryos three times in PBT for five minutes each wash
7) Postfix embryos by incubating 25 minutes in 4% formaldehyde diluted in PBT
8) Wash the embryos three times in PBT for five minutes each wash
9) Incubate embryos at room temperature for 5-10 minutes in 500µl PBT + 1µl 2mg/ml proteinase-K
10) Wash the embryos five times in PBT for five minutes each wash
11) Postfix embryos by incubating 25 minutes in 4% formaldehyde diluted in PBT
12) Wash the embryos five times in PBT for five minutes each wash
13) Wash embryos in 500 µl PBT + 500µl Hybe A for ten minutes
14) Replace the above solution with 500 µl of Hybe A and wash for 10 minutes
15) Replace the above solution with 500 µl of Hybe B and prehybridyze in agitation at 55°C for 1 hour
16) Replace the above prehybridization solution with 75µl of Hybe B and add 3µl of RNA-DIG probe. Hybridize overnight at 55°C without agitation in a water bath.

---

17) Wash embryos three times in 500 µl of Hybe B at 55°C for fifteen minutes each wash
18) Add 500µl of PBT and place embryos at room temperature
19) Rinse embryos twice in 1ml PBT
20) Wash embryos for 5 minutes in 1ml PBT
21) Wash embryos for 15 minutes in 1ml PBT
22) Wash embryos for 20 minutes in 1ml PBT
23) Incubate embryos at room temperature for 1hr in 1:500 biotinylated α-DIG (preabsorbed at 1:10) in PBT
24) Rinse rapidly embryos twice in 1ml PBT
25) Wash embryos for 5 minutes in 1ml PBT
26) Wash embryos for 15 minutes in 1ml PBT
27) Wash embryos for 20 minutes in 1ml PBT

Amplification with the Tyramide StreptAvidin Biotin System (Perkin Elmer-Life Sciences)

28) Incubate embryos for 30 minutes in TNB
29) Incubate embryos for 30 minutes in 1:100 SA-HRP in TNB
30) Wash embryos three times in PBS with 0,1% Triton-X for 5 minutes each wash.
31) Incubate embryos for 10' in Tyramide biotinylated (1/50) in amplification solution
32) Wash embryos three times in PBX for 5 minutes each wash
33) Incubate embryos for 30 minutes in TNB
34) Incubate embryos for 30 minutes in 1:100 SA-FITC or SA-TexasRed in TNB

Immunofluorescence

35) Wash embryos three times in PBX for 5 minutes each wash
36) Incubate embryos for 30 minutes in TNB
37) Incubate embryos with primary antibody/ies overnight at 4°C

---

38) Wash embryos four times in PBS with Triton-X 0,1% for 15 minutes each wash
39) Saturate embryos by washing 30 minutes in TNB
40) Incubate embryos with the secondary antibody in TNB, for 1hr at room temperature
41) Wash embryos four times in PBX for 15 minutes each wash
42) Mount embryos in a slide with fluoromount media or other mounting media

Confocal microscopy

Embryos stained by TSA amplified FISH were analyzed at a Zeiss confocal microscope in order acquire only the plane where the heart is observed.

G. DNAseI binding assay

In vitro DNAseI binding assays were performed as in Capovilla *et al*., 1994, using ABD-A and UBX proteins.

RESULTS

First bioinformatic analysis and *in vivo* validation for *Ndae1* and *Ih*

SUMMARY

We started our project working on three putative *Hox* target genes: *Ndae1, Ih* and *Ork1*. We first performed *in silico* analysis to search for candidate heart *cis*-regulatory modules of the genes *Ndae1* and *Ih.* If this bioinformatic analysis had proven useful after *in vivo* validation, we would have proceeded in the same way to study *Ork1*. The tools used are described in "Materials and Methods: Bioinformatics: First bioinformatic approach based on available online tools".

For each *D. melanogaster* gene of interest (*Ndae1* and *Ih*) the *D. pseudobscura* and the *D. virilis* orthologous sequence was retrieved. We used the VISTA Browser to retrieve a map containing the Conserved Non coding Sequences (CNSs). rVISTA and eCIS-ANALYST were used to search for clusters of Transcription Factor (TF) Binding Sites (BSs). On a further instance, we used our own software to visualize TF BSs in an alignment where CNSs were also delimited.

In every case, we used Positional Weight Matrices (PWMs) to search for TF BSs. The PWMs used alone or in combination are: TIN, SVP, GATA, MEF2, TBX and HOX (either UBX or ABD-A because they bind to the same sequence *in vitro*). We did not obtain reliable results by overlapping the outputs of all the methods used to predict heart *cis*-regulatory sequences. Nevertheless, the few sequences obtained from overlapping results of bioinformatics (for *Ndae1* only) were given priority to begin cloning experiments and produce transgenic reporter flies. For *Ndae1*, we produced a total of 12 reporter constructs covering the whole gene (**Figure 16**). Some of these constructs overlap. Instead, for *Ih*, since we did not obtain overlapping bioinformatic results, we produced 5 reporter constructs (**Figure 21**) regarding only conservation blocks between *D. melanogaster* and *D. pseudobscura*.

None of these transgenic reporter flies drove expression in the heart. All the *Ih* constructs reported an expression pattern in embryos (**Table 2**). Three of the twelve *Ndae1* reporter constructs did not drive any expression pattern in embryos. Without taking into account the overlapping constructs, four *Ndae1* reporter constructs drove an expression pattern in embryos which corresponds to domains of expression previously described for the endogenous *Ndae1* (Romero *et al*., 2000; Sciortino *et al.*, 2001) and one construct drove an expression in the haemocytes, cells which were not described in previous *in situ* experiments done for *Ndae1*.

From these results, we concluded that this first bioinformatic analysis needed refinement, but all in all, this type of examination can give a general idea of the distribution of BSs for the various TFs of interest in the query sequence. On the other hand, using only sequence identity between *D. melanogaster* and *D. pseudobscura* can give a general view of the enhancer elements of a gene, but is not enough when searching for tissue specific enhancers. The bioinformatic tools available still needed development and could not substitute immunoprecipitation studies to search for *cis*-regulatory sequences.

Bioinformatic results on *Ndae1*

The first gene on which we started our bioinformatic analysis was *Ndae1*. This gene had a reliable annotation and its pattern of expression had been published (Romero *et al*., 2000). First, the *Ndae1* gene of *D. melanogaster* was aligned to its orthologues in the *D. pseudobscura* and *D. virilis* shotgun genomes. Annotation of these latter two genomes had not yet been completed at the beginning of 2005, thus the orthologues were retrieved from the database and aligned with AVID, using VISTA Browser, a multiple alignment server from VISTA tools. With the same software we detected 100 bp to 500 bp CNSs among the three species since we hypothesized that short CNSs present in evolutionary related species could stand for regulatory sequences. **Figure 10** shows the output map as shown by VISTA Browser where depicted in pink are the CNSs of *Ndae1* between *D. melanogaster* as base genome and the correspondent orthologues in the shotgun genomic sequence of *D. pseudobscura* and *D. virilis* using a conservation filter of 70%.

From this analysis, we deduced that there is a too high degree of sequence homology between the aligned sequence of *D. melanogaster* and *D. pseudobscura* in order to infer putative *cis*-regulatory sequences from CNSs found in the alignments of these two sequences alone. On the other hand, there are only three highly conserved (> 70%) CNSs between the aligned sequences of *D. melanogaster* and *D. virilis* and although this number could be informative, these CNSs are very short (< 100 bp) and far away from each other in order to constitute a *cis*-regulatory module by themselves or in combination.

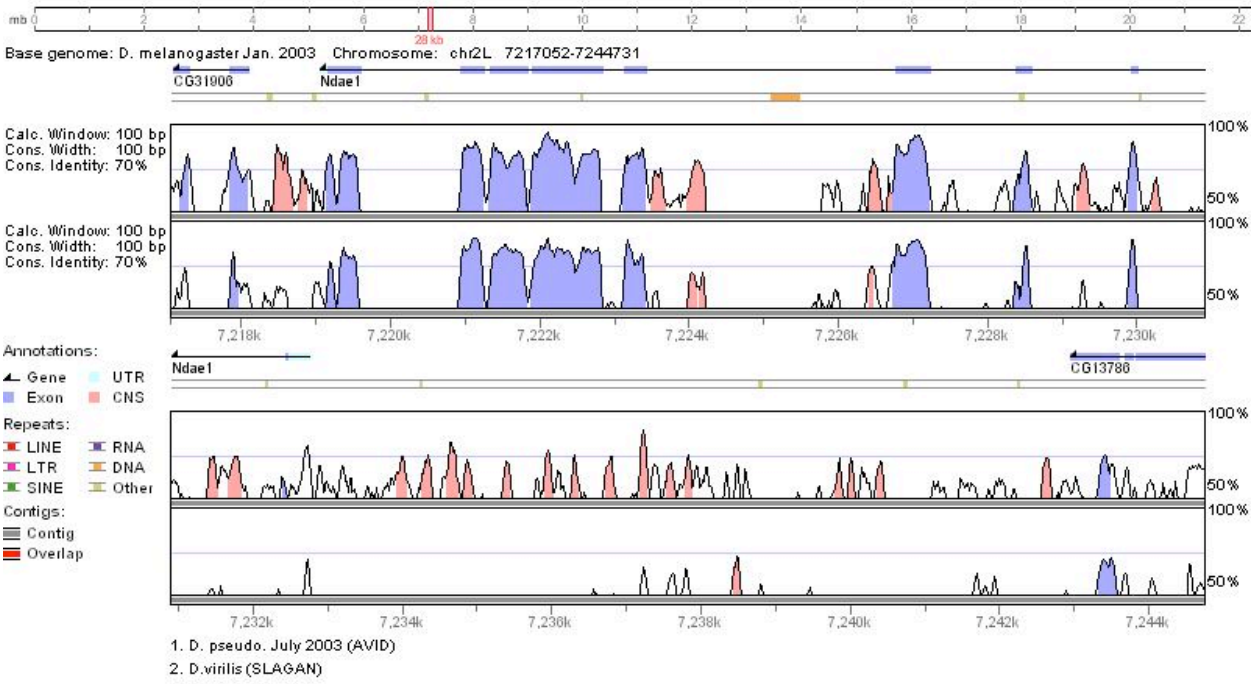


Figure 10. Output map of VISTA Browser for the gene *Ndae1*. Depicted in pink are the CNSs of *Ndae1* between *D. melanogaster* as base genome and the correspondent orthologous in the shotgun genomic sequence of *D. pseudobscura* (top row) and *D. virilis* (bottom row) using a sequence identy filter of 70%. In blue are depicted the coding sequences.

rVISTA (Loots *et al*., 2002) from the VISTA family of computational tools was used to search for conserved clusters of TF BSs. This server allows using user-defined BSs or PWMs to scan the sequence of interest, which is loaded as a FASTA file. The conservation filter and the window size to scan the sequence may be used as variables. To analyze *Ndae1*, we used a fixed conservation filter of 70% and a variable window size (100 and 500 bp). The usefulness of a "non clustering" option is that of evidencing all TF BSs found by the program whether they form or not a cluster. To score for TF BSs we used the PWMs described in "Materials and Methods: Bioinformatics: First bioinformatic approach based on available online tools" as well as combinations of them. The program output distinguishes "aligned sites" from "conserved sites". Aligned sites are those sites that match all nucleotides in an alignment. Conserved sites are those sites that fall into conservation regions of the sequence. **Figure 11**, shows an example of an output of rVISTA. In this example *Ndae1* sequence was scored for the TIN and SVP BSs. The figure shows *Ndae1* transcriptional unit using a conservation filter of 70% between the base genome of *D. melanogaster* and the distant orthologous sequence of *D. virilis*. The window size is 100 bp (automatic default window). The figure shows, in the first and second rows, all the TF BSs found with these two matrices. In the third and fourth rows, it shows all those BSs that are aligned, and in the last two rows those BSs that are also conserved. Notice that only one SVP BS inside intron III is conserved across more than 10 kb.

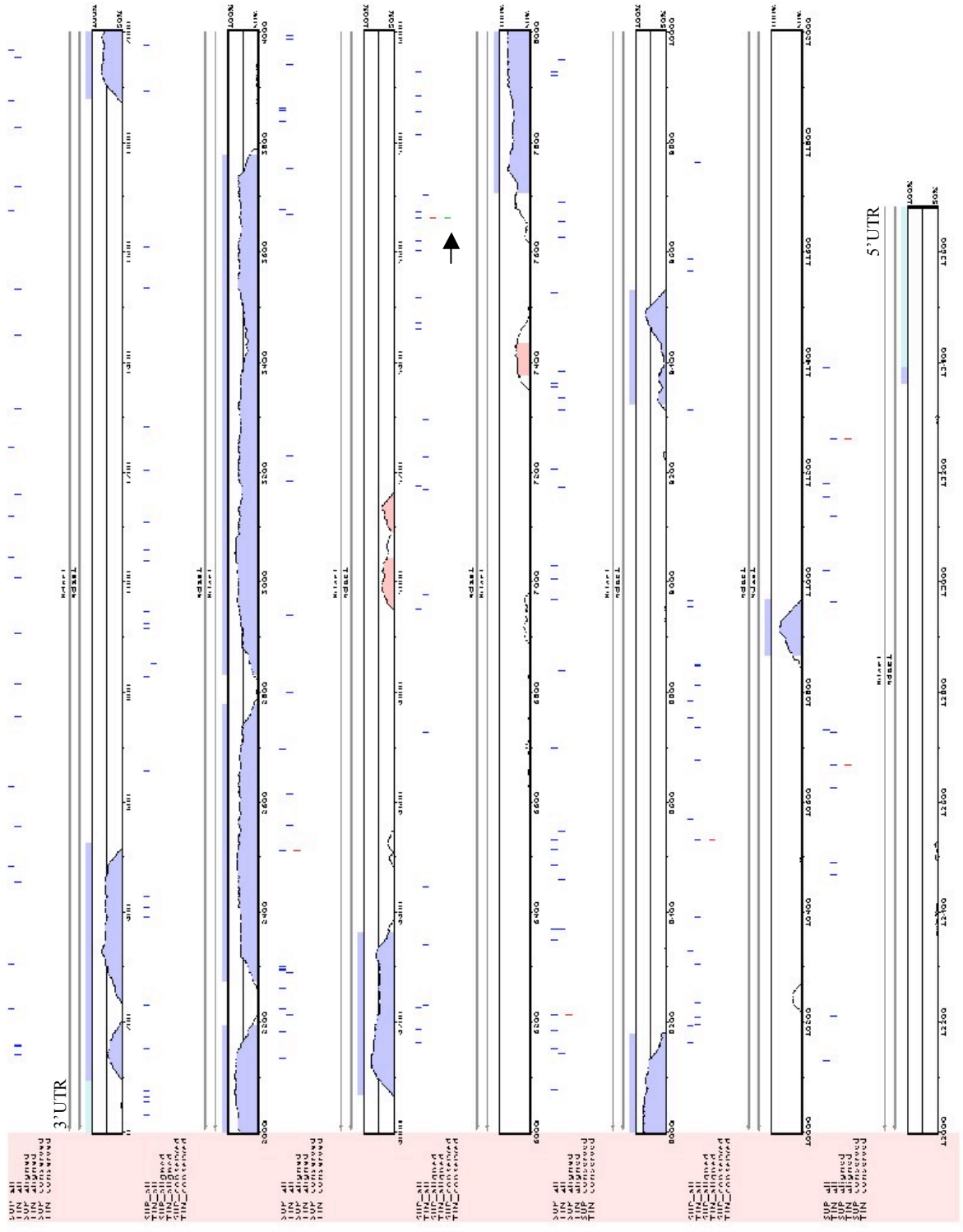


Figure 11. Output of rVISTA when *Ndae1* sequence is scored for TIN and SVP BSs. The figure shows *Ndae1* transcriptional unit using a conservation filter of 70% between the base genome of *D. melanogaster* and the distant orthologous sequence of *D. virilis*. The window size is 100 bp (automatic default window). In the first and second rows, all the TF BSs scored with TIN and SVP PWMs. In the third and fourth rows, it shows all those BSs that are aligned, and in the last two rows those BSs that are also conserved. Notice that only one SVP BS inside intron III is conserved across more than 10 kb (pointed with an arrow). In pink are depicted the CNSs, in blue the coding exons and in pale blue the 3' and 5' UTRs.

As an example of how we compared the different outputs of a program according to the manipulation of the possible variables (*e.g.*, window size or quantity of PWMs loaded), **figure 12** shows part of *Ndae1*, including intron III and the whole 3' UTR and downstream region, allowing a 60% conservation filter between the *D. melanogaster* and *D. pseudobscura* sequences. This time the sequence was scored for the PMWs of TIN, UBX, MEF2 and SVP using a window size of 100 bp. Notice that many more aligned and conserved sites are found. Comparing outputs of the same program gave us a better idea of which TF BSs might be putatively real *in vivo.*

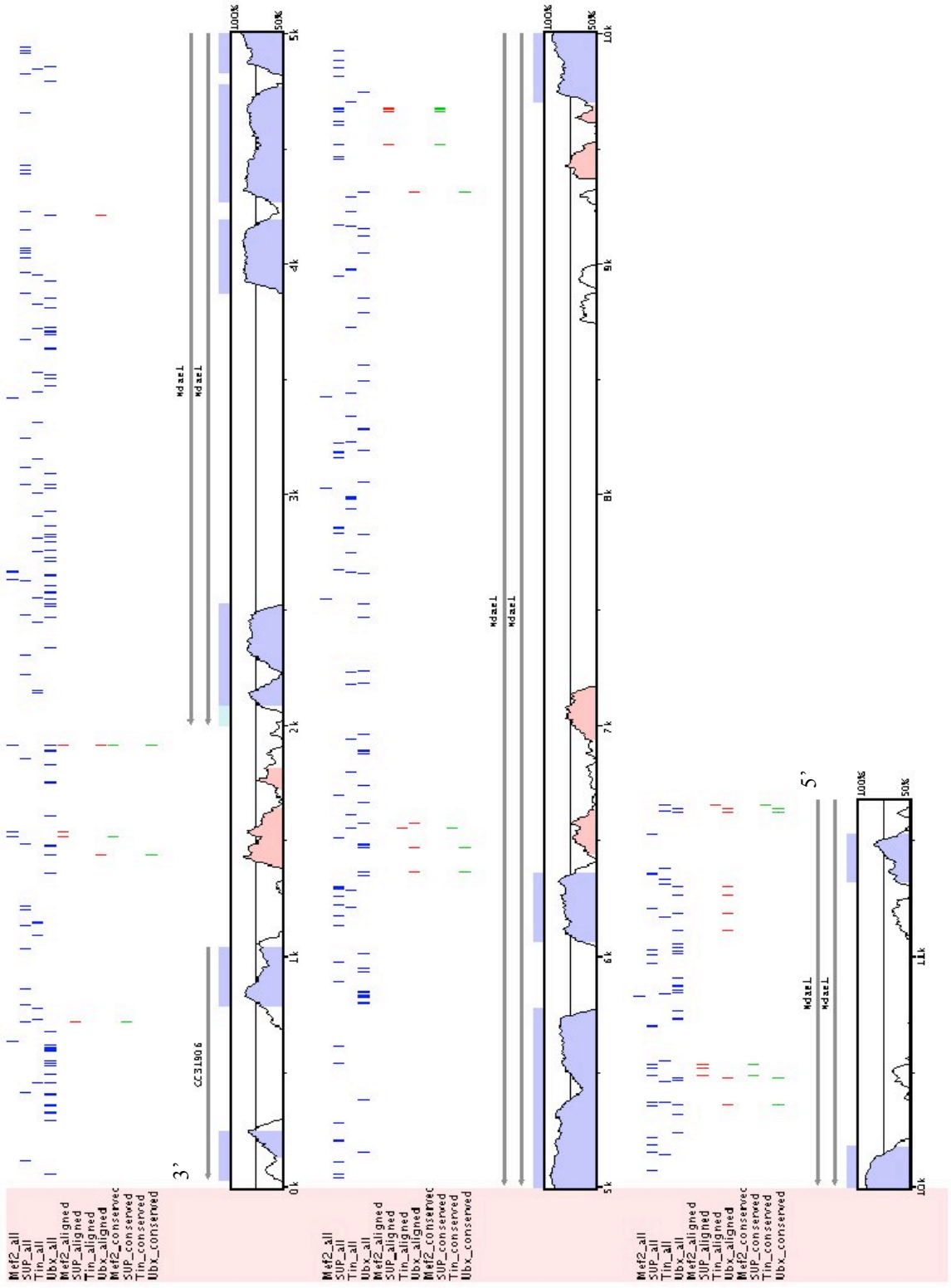


Figure 12. Output of rVISTA when *Ndae1* sequence is scored for TIN, SVP, UBX and MEF2 BSs. The diagram shows part of *Ndae1*, including intron III and the whole 3' UTR and downstream region, allowing a 60% conservation filter between the *D. melanogaster* and *D. pseudobscura* sequences. The sequence was scored using a window size of 100 bp. CNSs are depicted in pink and coding exons in blue.

Furthermore, in order to compare the outputs of rVISTA with other search methods, we performed an analysis using eCIS-ANALYST. **Figures 13A, 13B, and 13C** show an example of the output of this program using the same PWMs (UBX, TIN, SVP and MEF2), a fixed minimum number of scored TF BSs per cluster equal to 3 and a variable window size of 100 (**Figure 13A**), 500 (**Figure 13B**) and 1000 (**Figure 13C**). Notice that there are no clusters when the sequence is scored using a window size of 100 and that a conserved cluster between the sequence of *D. melanogaster* and *D. pseudobscura* is retrieved when using a window size of 500. **Figure 14**, shows part of the alignment with the TF BSs found inside this cluster. Notice in **figure 13C** how many more clusters are found only varying the window size. The window size variable created us a problem at the time to decide which is the optimal window size to reduce false positive results to the minimum. Moreover, the general knowledge we had by then on the size (300 -1000 bp) and position of *cis*-regulatory modules of any given gene did not help us overcome this problem.

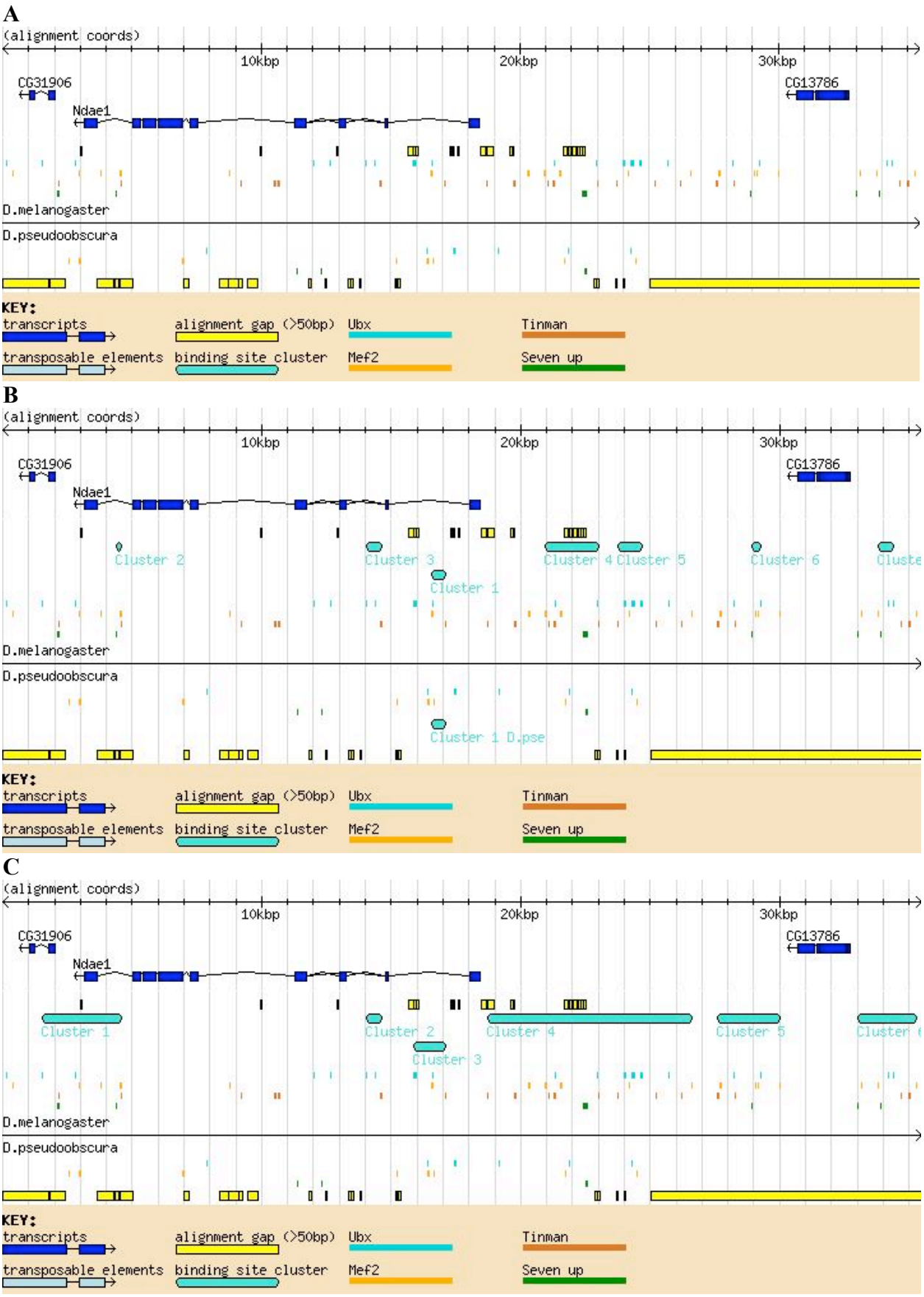


Figure 13. eCIS-ANALYST output score for UBX, TIN, SVP and MEF2 BSs in the whole *Ndae1* genomic DNA. Parametres: fixed minimum number of TF BSs per cluster equal to 3 and a variable window size of 100 (A), 500 (B) and 1000 (C).

Another way to tackle the issue of reducing to the minimum false positive results in *in silico cis*-regulatory sequence discovery was to use our own automatic tool. We proceeded in two ways. First, we scored in the genome region for TF BSs using PWMs, then we produced an alignment with the *D. pseudoobscura* orthologue and retrieved the CNSs with TF BSs. On second place, and to avoid the problem that a few conserved bases inside a non-conserved region of the genome might be missed in the general alignment, we introduced manually in the previously aligned sequence all TF BSs found on the scoring step. This allowed us to distinguish aligned sites from conserved sites. We also allowed the program to find TF BSs that slightly deviate from the consensus. Moreover, with this method we could analyze different percentages of sequence identity. **Figure 14**, shows the alignment of part of intron I of *D. melanogaster* with the corresponding *D. pseudoobscura* ortholog. In this alignment the TF BSs found have been introduced by hand. Part of the shown alignment corresponds to eCIS-ANALYST cluster 1 observed in **figure 13B**.

```
007238766  --------------------CTTTGCTTCTTG--ATGAAAATATCTTTTAATAAG-----  007238798 INTRON I
>>>>>>>>>                      |||||||| | |  ||| || |    ||||||  |       <<<<<<<<<
001372414  GTACTCGTATTTTTGTTGGTCTTTGCTTTTCGTAATGGAACTGCAATTTAATCTGTGCGA  001372473

007238799  ----------------------AAAATGAATATGGTATTGTCAATTTCTTCAAATTCTAA  007238836 INTRON I eCc1
>>>>>>>>>                        | ||   || |  |  ||    | | | | |    ||  <<<<<<<<<
001372474  CGTGTTCCAGGTGTCATCTTCCATAAATGATCTAATGGTGGGTGTGTTTACGATCATTAT  001372533

007238837  GGTGTGACTAAAGAGCTGACTACA--------GAACATGTTTAGTGAAAATATATTGAGA  007238888 INTRON I
>>>>>>>>>    ||    | | ||   || ||||           || || |  | | ||   ||    <<<<<<<<<
001372534  AATGACGTTTAGGAAGCGATTACATTGTTGCCAGGTTTGGTTTGACACATTAATGTGCAG  001372593

007238889  TACGTGAATTCCATTAGTTATTATAAACCTTTTTGCAAAAGATGATAAGTATAAGATTT  007238948 INTRON I eCc1
>>>>>>>>>   |    | || |  ||  ||| |||   ||||||    |   |    |  |||||    |  <<<<<<<<<
001372594  CAGCAAATTTGCTATAAATATAATACTTCTTTTTTGGGATTCTCTCTATATATAATTAAT  001372653

007238949  ATAGACCTTTAGAGTTTTATTTTAATGCTTGAAAATTACAGGGACAT--------ACCTA  007239000 INTRON I eCc1
>>>>>>>>>   |  |  |  |  | | |    ||| | ||  ||||   || || |          ||  <<<<<<<<<
001372654  TTCAATGTCGATTGGTCAACCGGAATTCATGTGAATTTTGGGCACGTTATCTAGGGTCTT  001372713

007239001  ACAATTTCCAAGAGCAGCCATCAATTAAAAACTTAATATATTGTATGTATGTGTACAAAC  007239060 INTRON I eCc1
>>>>>>>>>  |   | |||||  |   |     | |  | | | || || |||||  ||  |||   | || |  <<<<<<<<<
001372714  AGTGTCTCCAAAATAATATTTTATCTGAGAGCTAAACATATTACATTGATGG--AGAAGC  001372771

007239061  TATTTATAAAATATATGGGACGGAATGAATACAATATAGTTTTAATATGTATCAAAAGCA  007239120 INTRON I eCc1
>>>>>>>>>  | | |           | ||||     ||   |   | ||||   |     || ||||    <<<<<<<<<
001372772  TTTGTGGGCTTCGAGTAGGACTTTTGGAGAGCTTAAGAGTTCAGAGGCTGATAAAAATAG  001372831

007239121  AAGACATTTCTTTTGTTGTTTGCACGAGCCG-----------------------------  007239151 INTRON I eCc1
>>>>>>>>>  ||    |  | |||  |    |||| |  |                               <<<<<<<<<
001372832  AATTGGTCACGTTTTATTAACGCACCACTCAAAAAACCACATTTCACAGTCGTGAATAAG  001372891

007239152  ----------GGTAATTTTAGTTATTTCCAAGATT---------GTGGTGGAAAA-TTTG  007239191 INTRON I eCc1
>>>>>>>>>            | | ||||| ||| | |  ||   |         ||| ||||||| ||||  <<<<<<<<<
001372892  TCTAATTTGTGCTGATTTTGGTTTTGTTTAAAGATTAGAAACAAGTGATGGAAAAATTTG  001372951

007239192  CATTGCATAAAGGTTTATTTCTATATAGCCTATCATCTTGAGTCGTTGTCGCCGCTCGTT  007239251 INTRON I eC1
>>>>>>>>>  |||||||||||||||||||||||||||||||||||||||  | ||  |  | ||  ||  <<<<<<<<<
001372952  CATTGCATAAAGGTTTATTTCTATATAGCCTATCATCTTGTTTTGTCTTATCAGCCAGTC  001373011

007239252  TTCTTATCAGCCAGTGAAATGTATAATCTATATGGTACTTAAATATATAGACACCACAAA  007239311 INTRON I eC1
>>>>>>>>>  || ||                   ||||| |||  ||  || |   ||  ||       |  <<<<<<<<<
001373012  TTTTT-----------------TAATCTGTATATTATATAGA---ATTAACCGAGAGGA  001373050

007239312  GTGACAGCCCAGAGGACGGACATCCCAGCCTGTCTGCCAATTGCCGGTCTAAACTGGGTC  007239371 INTRON I eC1
>>>>>>>>>  || |||| |    ||    |||  |||  | |||||||| ||||||||  ||||||||||  <<<<<<<<<
001373051  GTCACAGACAGACGGTGCAACACTCCAAGCCGTCTGCCATTTGCCGGTAAAAACTGGGTC  001373110

007239372  TGGCCAGGGTTTACTGAATATGGTATAT-ACCCATCTACCCACTCGATTTCATGCATTTC  007239430 INTRON I eC1
>>>>>>>>>  |  |||| | ||||   || |||||    |||  ||    || |  || |           <<<<<<<<<
001373111  TATCCAGAGCTTACC--ATTTGGTAAGCCACCTGTCCCTGCATTGAATGTG---------  001373159

007239431  GATGGGCTTATCAACACCTGAGTCACAACAACAAACAATAATTTGGGGCCTGATATTAAC  007239490 INTRON I eC1
>>>>>>>>>   ||    |||||| ||  ||| ||| || ||  ||   |||||||||  |  |||| |||  <<<<<<<<<
001373160  -ATATCTTTATCAGCAA-TGACTCAGAAAAAGCAA-CGTAATTTGGGTTCCAATATAAAC  001373216
```

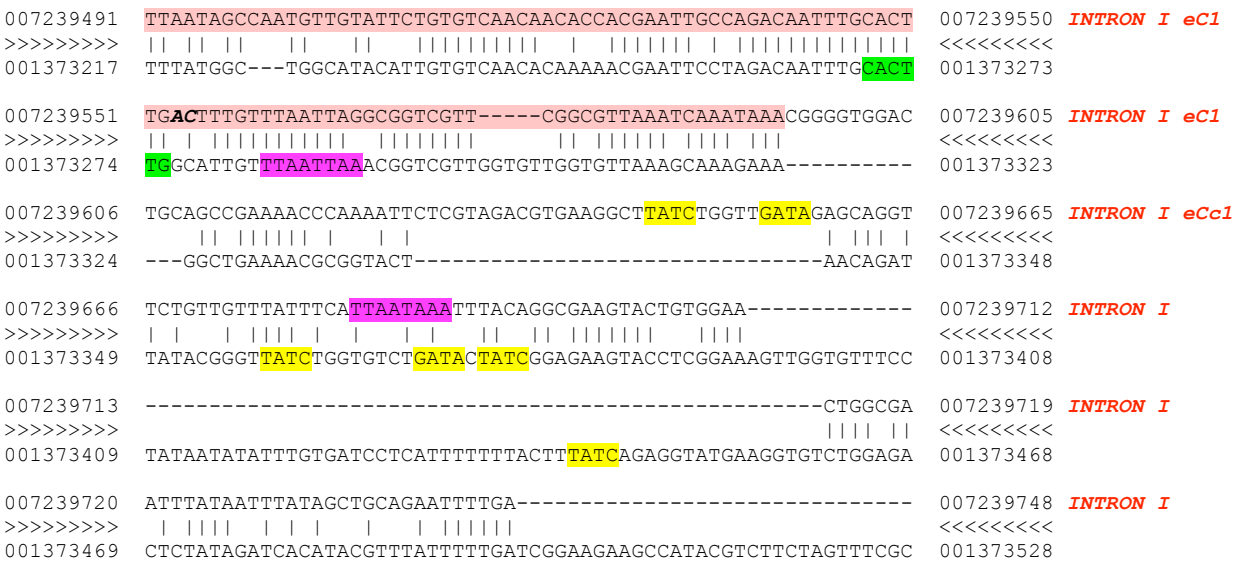
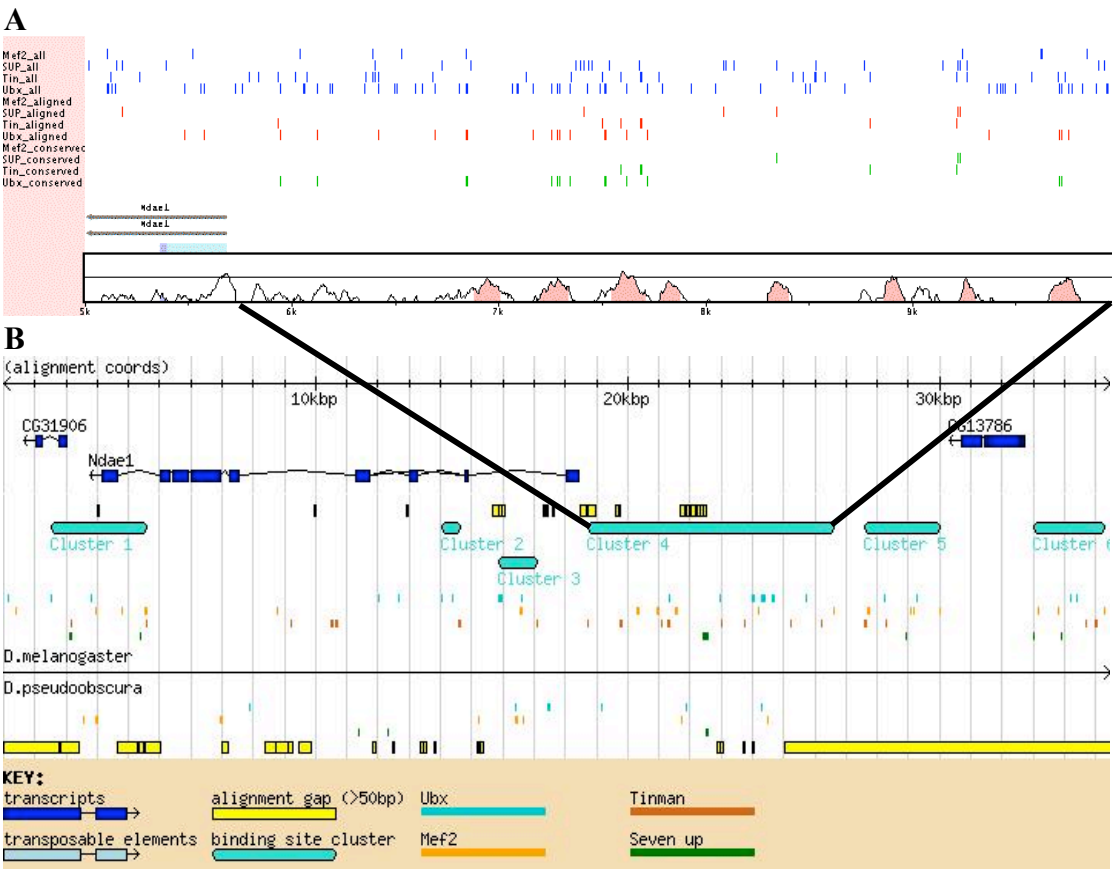


Figure 14. **Alignment of part intron I of *D. melanogaster* with the corresponding *D. pseudoobscura* ortholog.** In this alignment the TF BSs found have been introduced by hand. CNSs retrieved with a 70% conservation filter are highlighted in light pink, HOX sites in pink, TIN sites in green, GATA sites in yellow. In blue writing is depicted a MEF2 site. Part of the shown alignment corresponds to eCIS-ANALYST cluster 1 (eCc1) observed in **figure 13B**.

On a final step of our analysis, we proceeded to merge the outputs of all the methods used. We believed that by overlapping the results we would obtain a more reliable prediction in order to give priority to regions to produce transgenic reporter constructs. This was not exactly the case. We found more discrepancies than coincidences among approaches. **Figure 15** shows an example of the same sequence analyzed by the three methods. We consider that these discrepancies may be due to various reasons. For example, homology regions differ from program to program because "conservation" intended as sequence identity is not defined in the same way by the different programs. Phylogenetic comparison of BSs present in related species (*i.e.*, looking at the TF BSs in an alignment) proved to be informative due to the possibility that position of BSs within an enhancer might be plastic and therefore positional conservation may not be strictly required. The few overlapping results obtained (part of intron I, and a fragment 6kb upstream the transcription start site containing the fragment we called "Hunch") were given priority to begin *in vivo cis*-regulatory sequence validation.



C

```
D.mel  AACACCCCTTCAATTTGCTATTCAACCCCGCGCTGCTTGCACTTGACATGTTGTTGACCCAGTCTCTAGA--AAA
                                              | ||||||||||||||||||||||||||||    |||   |||
D.pse  ------------------------------------CATGCACTTGACATGTTGTTGACCCAGTC---AGACGAAA

D.mel  AGACAGACA-GACAACACAAGTCATCATCAACAAACCGCAACAAACAGAATAGAAAGTAT------CTATATCAG
        | |||||| ||||   |  | |    || | || |||| |  |||||| ||||||||||        | || ||||
D.pse  CGGCAGACAAGACAGGAACAATTCCCAGCCAC-AACCTCGGCAAACAAAATAGAAAGTAAGCAACCCCATCTCAG

D.mel  T---TAGATACCAAA--------------------GTTCTACAGCCTTTGCTCTGTGATTTCGTGCTGAGTTCTA
           ||||||||  |                    ||||| |||||||| |   |   |    ||||  | |  ||
D.pse  AGAATAGATACCCCATATCTCACAGAATTCCCGAGGTTCTTCAGCCTTTTCAATTAGTCATCGT-ATCATATCAT

D.mel  CGATAATAAAACGAAAACAAACCG------------AAAAATTGGAATCAGCGGATTAATTTGTTTTTCAATTAG
        ||||||  | |  | ||| |||             | |  ||  |||||      |||||||||     |  ||
D.pse  AGATAAT--ATCCCAGACAGACCACCGCTACAGACCAGACGTTCTAATCA------TAATTTGTTGACTATATAT

D.mel  TGTAGCCATTAATTGTATAATTAACGGGCAATATTTAAAGAATTTATAAGCTAAGTACATTGTGCATATAATGAG
       | ||| |||              ||  |||  | |   |||
D.pse  TTTAGGCAT--------------ACTCGCACGAGTAGTAGA----------------------------------

D.mel  AAAGAAAGTTAAGATACAAATCATTTTGACTCAGAATATTCCTCCTTCGCTAACAGCGACTTTTTCAGCGACTTA
                                           |  ||||| ||
D.pse  ----------------------------------------CCTCCCTTCTCT-----------------------

D.mel  AGCACTACGACTATGTAGCGATTGCATATTTAGATAGCGGATTCTATCTATAGCTTTGAGATGTGCGTGTATGGA
            |||||                          ||| |||  ||  |   | | ||||||||  ||||
D.pse  ----------CTATG----------------------ATATTGTATGAATCTCGCCGCG-TGTGCGTG--TGGA

D.mel  A----ATTGTCTATAGAACTGAAATTGCCTATGGGC-------ATAACCAACCTCCATTGTCACTTAATTGGGGG
       |    ||||    | ||| |||| |||   |||  |       ||  ||   |||||||||||  ||||||||||
D.pse  ACTGAATTGAAATTTGAATTGAAGTTGGAAATGTACTACTACGATTCCCGTTCTCCATTGTCATTTAATTGGGGG
```

Figure 15. Overlapping results for part of an *Ndae1* fragment 6kb upstream the transcription start site. VISTA output (A), eCIS-ANALYST output (B), part of the sequence alignment between *D. melanogaster* and *D. pseudobscura* (C), where highlighted in pink are the CNSs, in pale blue HOX BSs and in red TIN BSs.

Homology regions, clusters of TF BSs and comparative studies of program outputs should be taken into account to prioritize regions for functional assays. Nevertheless, nucleotide-to-nucleotide alignment within these regions is still required since, in an enhancer, BSs for different TFs might be conserved, while this might not be the case for the regions between them, which are subject to genetic drift rather than to positive selection. In any case, *in vivo* validation is indispensable.

Expression patterns of *Ndae1* transgenic reporter constructs

Figure 16, shows the position of the sequences that were cloned in order to make transgenic reporter constructs for *Ndae1*. You can observe that we have covered the whole gene, from the previous CG to the next, except for 292 bp upstream the transcription start site.

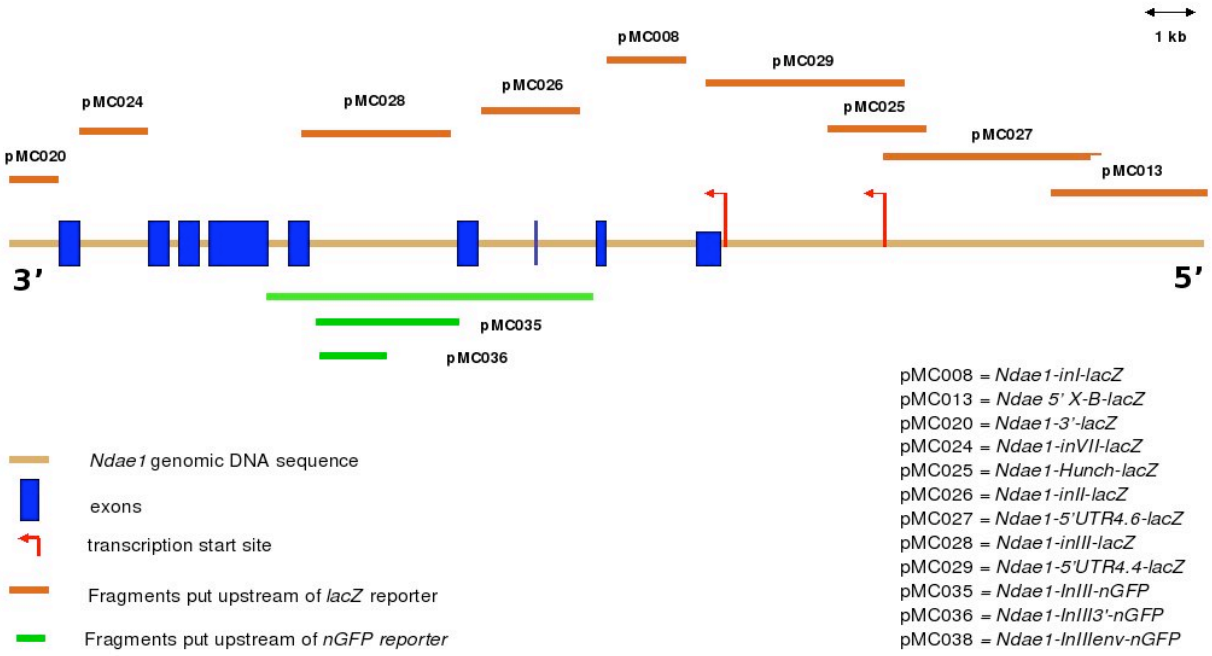


Figure 16. *Ndae1* genomic and cloning map. The corresponding expression patterns are described in **table 2** and shown in **figure 17**.

Table 1 shows the oligonucleotides used to amplif ıe sequences described in **figure 16**. **Table 2** summarizes the independent lines obtained for each reporter construct injection and the HRP expression patterns of embryos as seen for each transgenic reporter construct line. For each tested construct, embryos of at least two independent lines were stained both by HRP and TSA amplified FISH and analyzed at the confocal microscope. **Figure 17** shows the expression patterns in embryos as seen for each *Ndae1* transgenic reporter construct line described in **table 2**. In the case of overlapping constructs driving the same expression pattern we show the staining for only one reporter construct.

In total, we produced transgenic flies for 12 reporter constructs (**Table 2**). Of these, 3 did not drive expression in embryos (pMC026, pMC013 and pMC020). Overlapping constructs showed the same expression pattern except for pMC027 and pMC013. Without taking into account overlapping constructs, 4 *Ndae1* reporter constructs drove an expression pattern in embryos that corresponds to domains of expression previously described for the endogenous *Ndae1* (Romero *et al.*, 2000; Sciortino *et al.*, 2001) (**Table 2**: pMC008, pMC27, pMC29, pMC038; **figure 17**) and one construct (pMC024; **figure 17**) drove an expression pattern in the haemocyte precursors, cells which were not described in previous *in situ* experiments done for *Ndae1*.

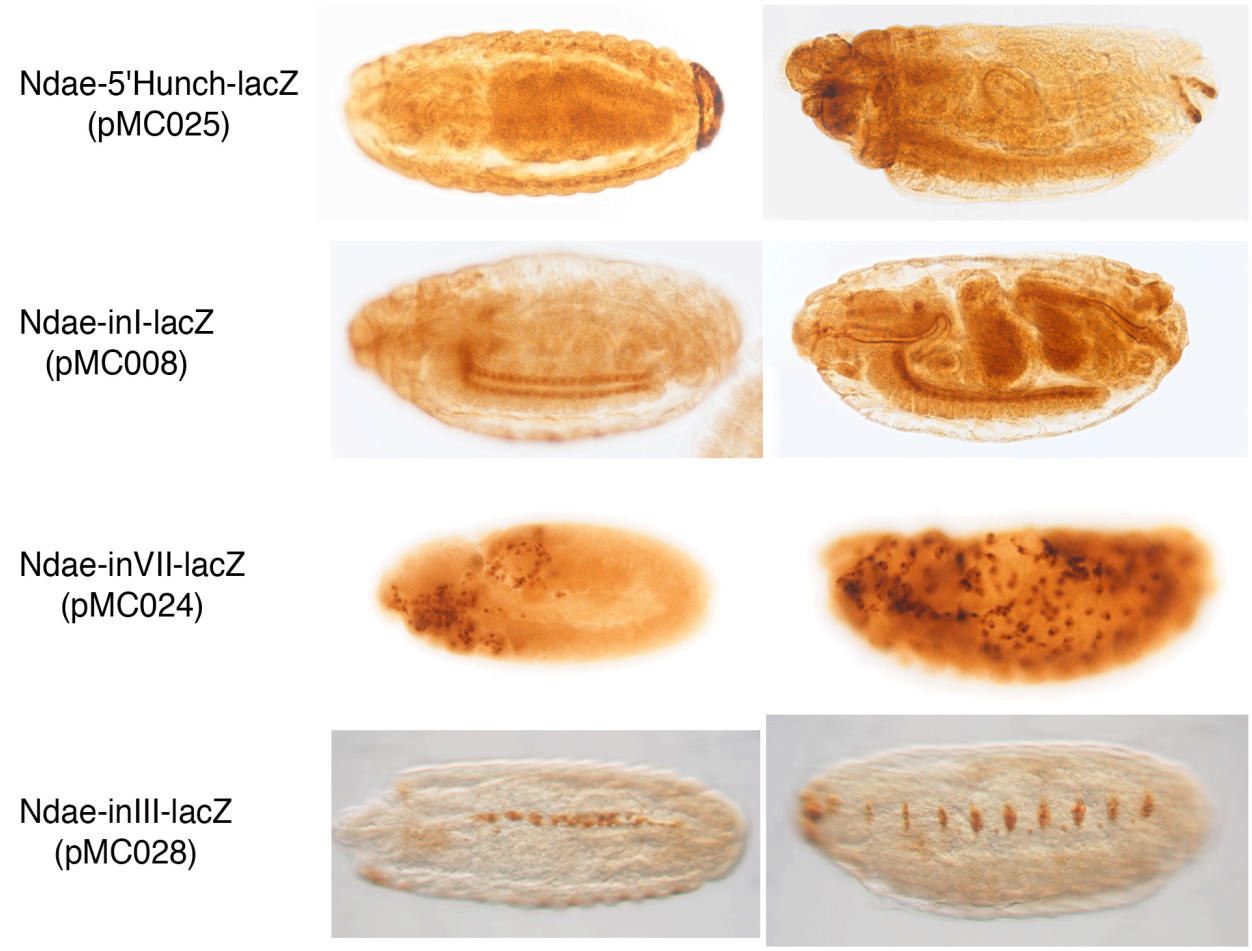


Figure 17. *lacZ* expression patterns of transgenic embryos carrying four different *Ndae1* reporter constructs. The names of the constructs are indicated to the left of the corresponding embryos. The patterns are described in **table 2**. Staining was done by HRP. The anterior part of the embryo is to the left and the dorsal one to the top.

We have covered the complete *Ndae1* gene, from the previous CG to the next, except for 292 bp upstream the transcription start site and we have not found the *Ndae1* heart enhancer. We are only missing to clone and produce a transgenic reporter construct for these 292 bp upstream from the transcription start site. We think such short fragment is unlikely to drive expression in embryos in the heart. It is more likely that a bigger fragment containing these 292 bp is needed in order to detect a signal, as we have observed for the overlapping fragments in the 5' (see **figure 16** and **table 2**): although pMC029 and pMC025 drive the same expression pattern in embryos, pMC029, which is bigger, drives stronger expression.

Bioinformatic results on the *Ih* channel

We thought that comparison of the bioinformatic outputs for a set of co-expressed genes could raise the accuracy of prediction. Therefore, in order to analyze *Ih*, we undertook the same procedure and tools as those used to analyze *Ndae1* (see above). First, the VISTA browser was used to depict the CNSs of *Ih* from whole genome pre-computed alignments among different species of *Drosophila* using *D. melanogaster* as base genome. As these *Drosophila* genomes show a high degree of conservation, different conservation filters were used in order to spot significant short conserved regions where *cis*-regulatory sequences might occur. Nevertheless, aligning the *Ih* sequence of *D. melanogaster* with the putative orthologue sequence of *D.*

pseudobscura and *D. virilis* we obtained the same results as for the respective analysis made on *Ndae1:* a high degree of sequence homology between the genomes of *D. melanogaster* and *D. pseudobscura* proved little informative as well as the scarce conservation between distant related species like *D. melanogaster* and *D. virilis* (**Figure 18**).

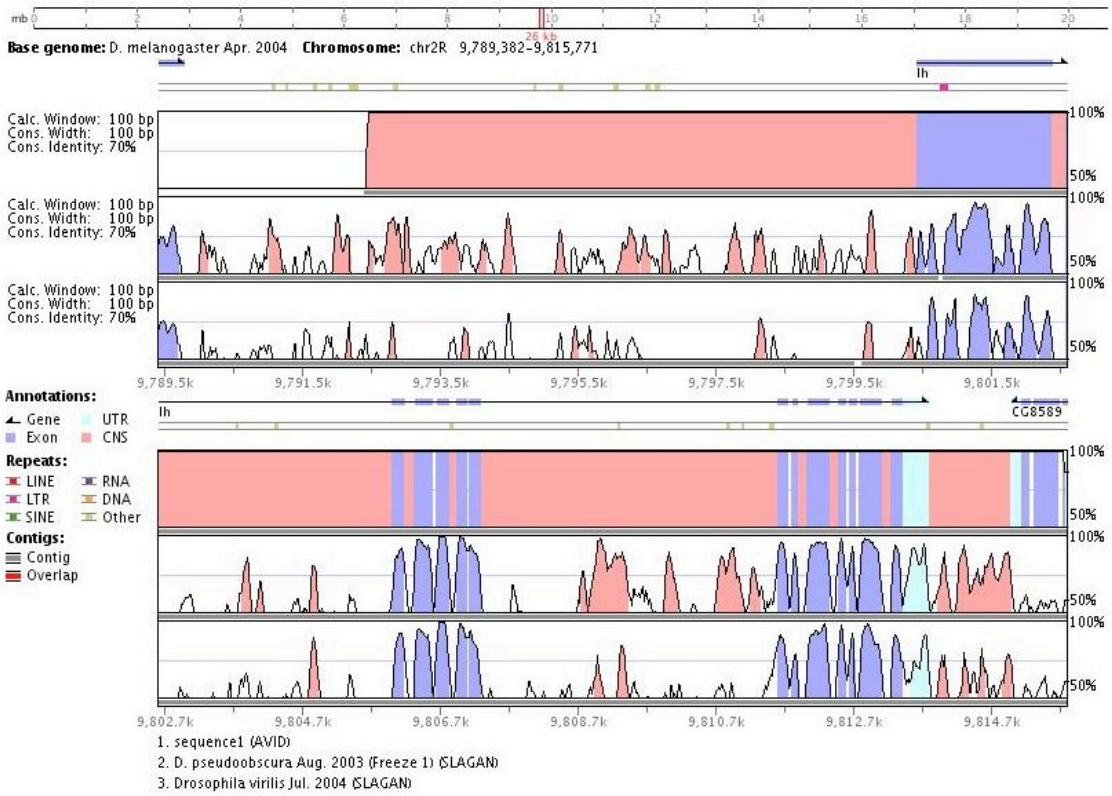
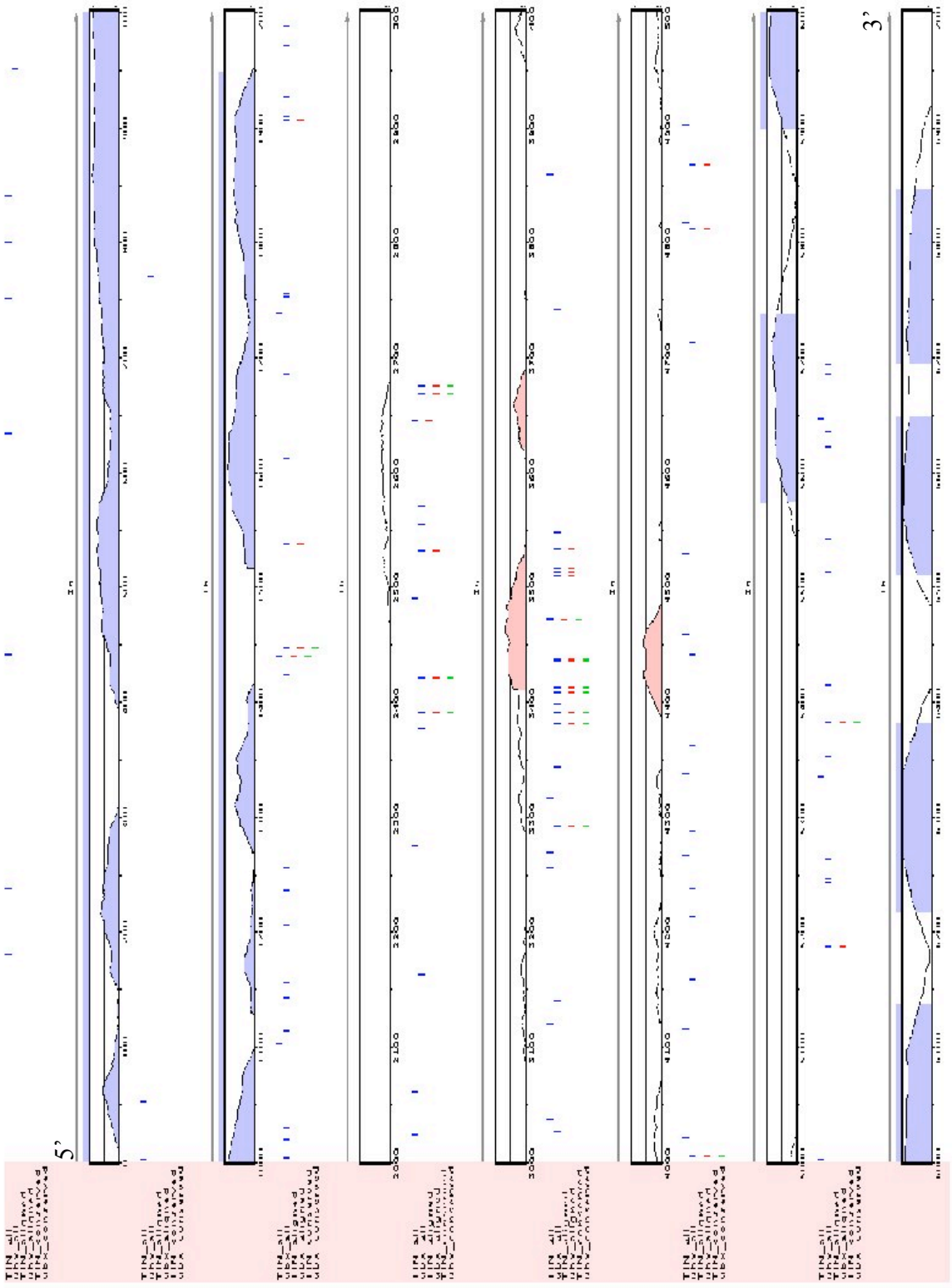


Figure 18. Map of the *Ih* gene as shown in VISTA Browser. Depicted in pink are the CNSs between *D. melanogaster* as base genome and the shotgun genomic sequence of *D. pseudobscura* (second row) and *D. virilis* (third row) using a conservation filter of 70%. The first row, corresponds to the sequence alignment of the same sequence of *D. melanogaster* using *D. melanogaster* as base genome using AVID instead than SLAGAN as whole genome aligment strategy.

As for *Ndae1*, we looked for clusters of binding sites using the rVISTA and eCIS-ANALYST programs. Likewise, we used the same PWMs (described in "Materials and Methods: Bioinformatics: First bioinformatic approach based on available online tools") and combinations of them as input to the programs.

We used rVISTA to analyze *Ih*, using a conservation filter of 70% and a variable window size (100 and 500 bp). Again, we used a "non clustering" option as a method to retrieve all possible BSs, even those that did not form a cluster. **Figure 19** shows an example of the rVISTA output obtained from the comparison of the CNSs of *D. melanogaster* and *D. virilis*, defining a conservation filter of 70% and a window size of 100 bp (default filters of the software). **Figure 19** shows the position of TIN BSs with respect to UBX BSs in the *D. melanogaster* genome, the sites aligned between *D. melanogaster* and *D. pseudobscura*, and the sites conserved from *D. melanogaster* to *D. pseudobscura*.


5'
3'

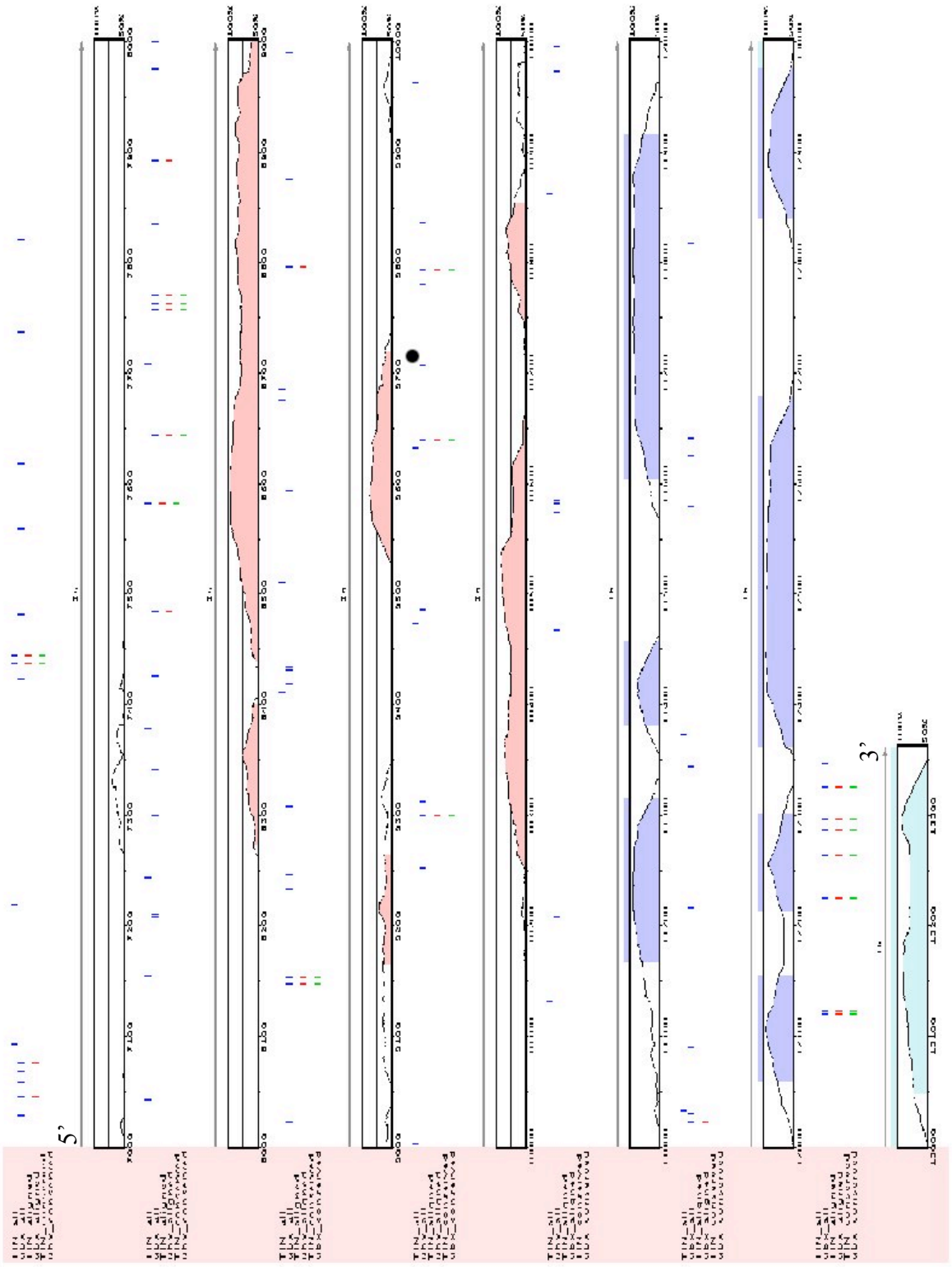


Figure 19. rVISTA output for the 3' half of *Ih* transcriptional unit. Shown are the distributions of the scored TIN and UBX BSs (top rows), the sites aligned between *D. melanogaster* and *D. pseudobscura* (middle rows), and the sites conserved from *D. melanogaster* to *D. pseudobscura* (bottom rows). Depicted in pink are the CNSs retrieved using a 70% conservation filter and in blue the coding sequences.

eCIS-ANALYST was used also as a second method in the search of conserved clusters of TF BSs in an automatic way. We used as variables the window size and the number of TF BSs in a cluster. This latter variable is not an optional in rVISTA software. We also analyzed the distribution of TF BSs without any clustering filter. **Figures 20A and 20B** show how altering the window size varies the amount of clusters. Furthermore, note how by narrowing the window size, cluster 1 of **figure 20A** disappears in **figure 20B** although it appears conserved between the genomes of *D. melanogaster* and *D. pseudobscura*. **Figure 20C** instead illustrates how augmenting the number of TF BSs in a cluster and reducing the window size reduces the number of TF BS clusters, standing for a more stringent analysis. Notice that cluster 1 of **figure 20A** reappears in **figure 20C**.

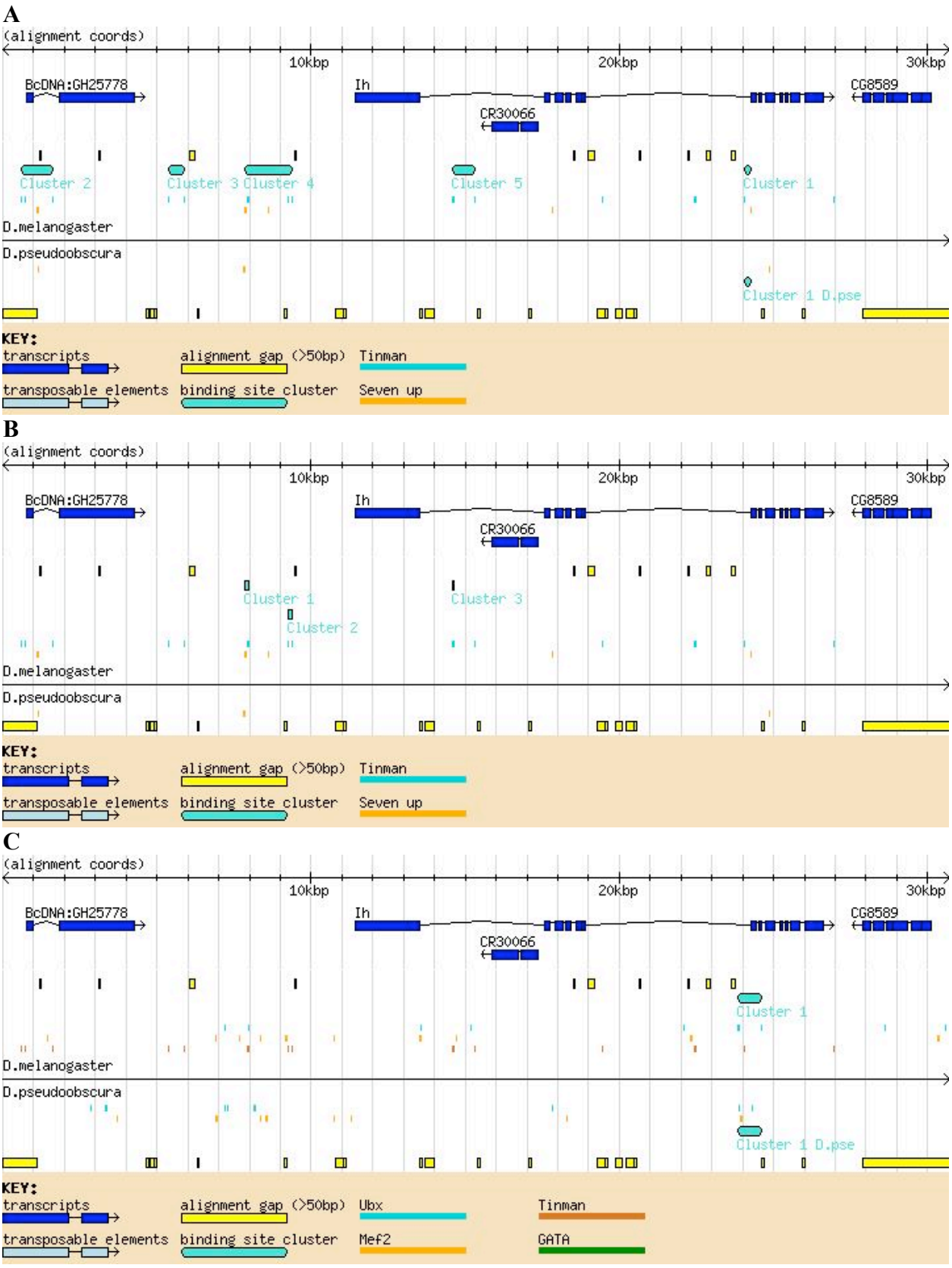


Figure 20. eCIS-ANALYST output for *Ih*. As variable parameters, were the window size of 1000bp (A) and of 100bp (B) and augmenting the number of TF BSs to score for within a fixed window of 100bp (C).

Furthermore, we analyzed with our own software, the complete gene including the genome regions upstream and downstream until the next annotated gene. We first retrieved the CNSs from the whole genome alignment using VISTA tools and scored for TF BSs within these CNSs using PMWs. We also allowed the program to find TF BSs that slightly deviated from the consensus. Similarly to what we had done for *Ndae1*, we introduced manually in the previously aligned sequences all TF BSs found on the scoring step in order to distinguish aligned sites from conserved sites and spot regions more densely populated with the desired TF BSs. **Table 3A** shows the position and quality of the conserved and aligned TF BSs found between *D. melanogaster* and *D. pseudobscura* and **table 3B** shows the position and quality of the conserved and aligned TF BSs found between *D. melanogaster* and *D. virilis.*

A

Position of binding site in Conserved sequences region (CNSs) between D.Melanogaster and D.Pseudobscura

Cluster Position	Ubx	Mef2	Tinman	T-Box	Region	Binding sites preserved and aligned
9796086-9796373	6	0	2	0	5'UTR	9796086 Ubx 90%
9796388-9796500	6	0	2	0	5'UTR	
9796720-9796830	6	0	2	0	5'UTR	9796773 Ubx 90%
9797654-9797861	6	0	2	0	5'UTR	9797662 Ubx 80%, 9797723 Mef2 80%, 9797742 Ubx 90%
9798003-9798188	6	0	2	0	5'UTR	9798143 Ubx 90%, 9798152 Ubx 80%, 9798167 Ubx 80%,9798173 Mef2 90%
9798953-9799046	6	0	2	0	5'UTR	9798977 Ubx 90%, 9799014 Ubx 85%, 9799024 Ubx 90%
9799693-9799791	6	0	2	0	5'UTR	9799693 Ubx 90%, 9799766 Mef2 80%
9800260-9800361	6	0	2	0	5'UTR	
9803783-9803937	6	0	2	0	Intron I	9803788 Ubx 90%, 9803816 Ubx 90%, 9803844 Mef2 80%, 9803928 Ubx 90%
9804021-9804100	6	0	2	0	Intron I	9804048 T-Box 80%(?), 9804063 Ubx 80%, 9804071 Ubx 90%
9804774-9804895	6	0	2	0	Exon II?	9804778 Ubx 80%, 9804787 Ubx 80%, 9804808 Ubx 85%
9808724-9808799	6	0	2	0	Intron VIII	9808753 Ubx 80%
9808839-9809399	6	0	2	0	Intron VIII	9809041Ubx 80%, 9809165 Ubx 80%, 9809264 Ubx 80%
9809914-9810116	6	0	2	0	Intron VIII	
9810688-9811062	6	0	2	0	Intron VIII	9810927 Ubx 80%
9811191-9811295	6	0	2	0	Intron VIII	9811190 Ubx 90%, 9811224 Ubx 80%,
9813903-9814070	6	0	2	0	3'UTR	
9814175-9814962	14	2	1	0	3'UTR	9814230 Ubx 80%, 9814246 Ubx 90%, 9814301 Ubx 80%, 9814313 Ubx 80%, 9814435 Ubx85% 9814450 Ubx 85%, 9814484 Tinman 80%, 9814683 Ubx 80%, 9814920 Ubx 85%

B

Position of binding site in Conserved sequences region (CNSs) between D.Melanogaster and D.Virilis

Position	Ubx	Mef2	Tinman	T-Box	Region	Binding sites preserved and aligned
9799627-9799780	1	2	0	0	5'UTR	
9800264-9800362	1	1	1	0	5'UTR	9800275 Mef2 80%, 9800331 Tinman 80%
9804763-9804883	4	0	1	0	Exon II?	9804787 Ubx 80%, 9804808 Ubx 85%, 9804867 Tinman 80%
9809258-980997	3	0	0	0	Intron VIII	9809288 Ubx 90%
9813901-9814056	0	1	0	0	3'UTR	
9814238-9814348	5	0	0	0	3'UTR	9814303 Ubx 80%, 9814311 Ubx 80%, 9814323 Ubx 80%

Table 3. Summary of output for *Ih* using our own software. (A) Position and quality of the conserved and aligned TF BSs within CNSs found between *D. melanogaster* and *D. pseudobscura.* (B) Position and quality of the conserved and aligned TF BSs within CNSs found between *D. melanogaster* and *D. virilis.*

Like for the *Ndae1* analysis, we found more discrepancies than coincidences by merging the outputs of all three softwares and no consistent information was found in order to consider this bioinformatic analysis as a startpoint for *in vivo* validation. Due to this lack of reliable information obtained from the bioinformatic analysis, for *Ndae1* we had proceeded to clone the whole non-coding regions of the gene, covering in all more than 15kb. This strategy was not possible for *Ih* being the extension of the non-coding region significantly bigger than that of *Ndae1*. Therefore, we proceeded to clone only the non-coding regions of the gene with conserved and aligned blocks (of 1kb approximately) between the genomes of *D. melanogaster* and *D. pseudobscura*. Despite of the fact that sequence identity alone was not the best option, as stated above, it was the most reliable information we had at that moment. Taking into account only sequence identity among related species had been effective in identifying mammalian regulatory sequences (Pennacchio & Rubin, 2001) and some preliminary studies had been published in *Drosophila* (Bergman *et al.*, 2002).

Expression patterns of *Ih* transgenic *lacZ* reporter constructs

Figure 21, shows the position of the sequences that were cloned in order to make transgenic *lacZ* reporter construcs. **Table 1** shows the oligos used to amplify the sequences described in **figure 21**. For each tested construct, embryos of at least two independent lines were stained both by HRP and TSA amplified FISH and analyzed at the confocal microscope.**Table 2** shows the independent lines obtained for each reporter construct injection and the HRP expression pattern on embryos as seen for each transgenic reporter construct. **Figure 22** shows the expression patterns in embryos as seen for each transgenic reporter construct described in **table 2**.

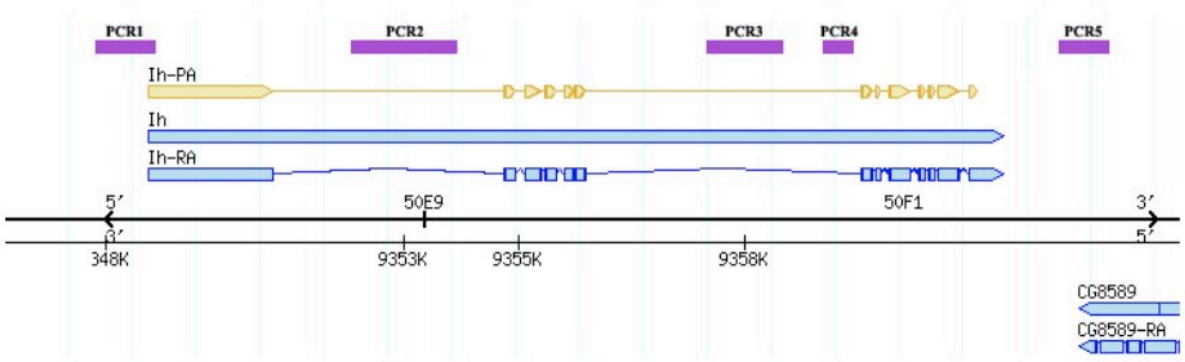


Figure 21. *Ih* genomic and cloning map. The genomic map was taken from FlyBase. Depicted in purple are the fragments of sequences cloned independently upstream a *lacZ* reporter. Each fragment is named as "PCR" followed by a number, which is the name of the reporter construct.

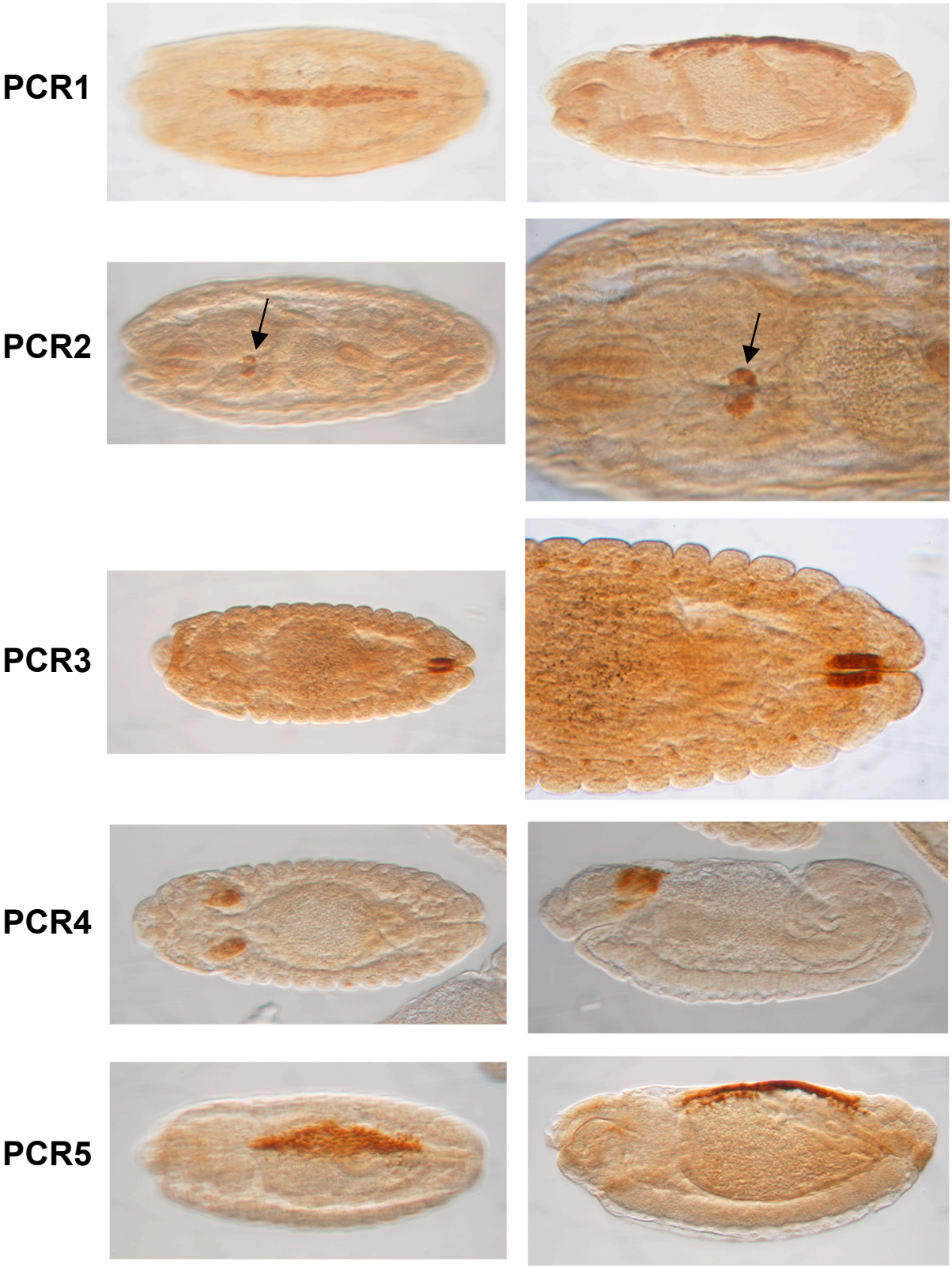


Figure 22. *lacZ* expression patterns of transgenic embryos carrying four different *Ih* reporter constructs. The names of the constructs are indicated to the left of the corresponding embryos. The expression patterns are described in **table 2**. Staining was done by HRP. Anterior is to the left and dorsal is to the top.

The results described in this section show that the bioinformatic methods used are not efficient enough in finding the heart *cis*-regulatory region of *Ih*. Moreover, they show that taking into account only the "conservation" parameter in order to find heart *cis*-regulatory sequences is not sufficient. On the other hand, it can be argued that conservation alone is a good start point for a general view of regions where *cis*-regulatory modules for different tissues in a gene might occur, since, for *Ih,* all the transgenic reporter constructs made taking into account only conservation blocks between *D. melanogaster* and *D. pseudobscura* show an expression pattern in embryos of at least three independent lines.

A novel pattern matching based approach

"Top 20" and *in vivo* validation

We used a pattern matching based approach to search for motifs of known TF BSs using PMWs on CNSs of genes found to be overexpressed in the heart with respect to the aorta. The first dataset used to validate this novel bioinformatic method consisted in seven genes found by candidate gene approach (*Ndae1, Ca-β, Ih, CaP60A, Ork1, sei* and *pain*) and two genes obtained from preliminary results of microarray experiments which compared the heart transcriptome with that of the aorta (*Dms* and *CG15537*). This pattern matching approach consists of two strategies, which we named A and B. These strategies are described in detail in "Materials and Methods: Bioinformatics: Second bioinformatic approach based on a novel pattern matching based method". We overlapped and ranked the outputs of these two strategies to produce a final list that we called "Top20". For these best 20 putative *Cis* Regulatory Modules (CRM) we used Toucan to draw inside the DNA fragments each predicted binding site motif.

Below we show the "Top 20" as ranked in the final scoring. Notice that there are 7 distinct genes in the top 8 and that *seizure* and *Ca-P60* are not in this list. Each putative CRM is called "seq" and named according first to the gene and then with a number. Indicated in bold are putative CRMs assessed for *in vivo* validation and in **table 2** is described the pattern of expression observed in embryos of the transgenic reporter flies. In red bold letters are the CRMs found to drive heart expression in embryos. **Table 1** shows the oligonucleotides used to clone the putative CRMs to produce transgenic reporter constructs for *in vivo* validation. For each tested construct, embryos of at least two independent lines were stained both by HRP and TSA amplified FISH and analyzed at the confocal microscope.

Top 20 ranking:

1. **CG15537_seq_14 (in progress)**
2. **CG15537_seq_11**
3. **Ih_seq_48**
4. **Dms_seq_5**
5. **pain_seq_42**
6. **Ndae1_seq_39**
7. **Ork1_seq_9**
8. Ca-beta_seq_40
9. Ork1_seq_25
10. Ndae1_seq_69
11. **Ca-beta_seq_46**
12. **Ca-beta_seq_55**
13. Ndae1_seq_64
14. Ork1_seq_23
15. **CG15537_seq_1**
16. Ca-beta_seq_41
17. **pain_seq_15**
18. **pain_seq_14**
19. **Ih_seq_34**
20. Ndae1_seq_65

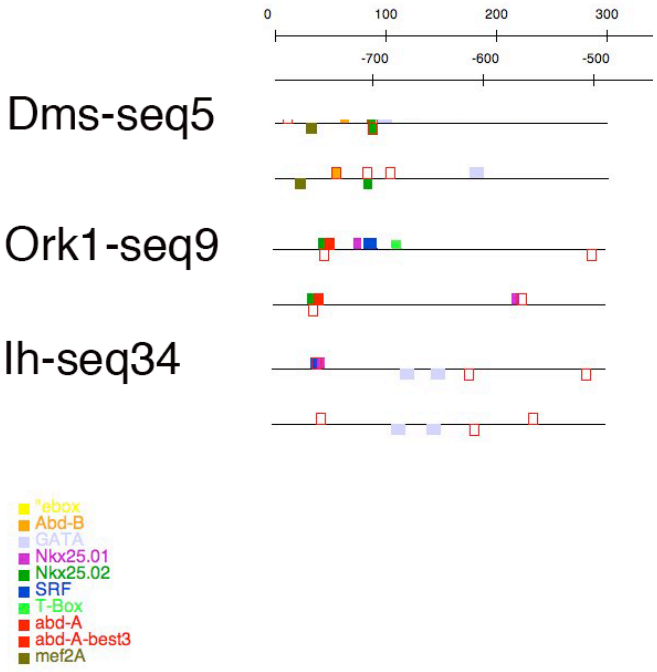


Figure 23. The three heart positive CRMs with predicted TF BSs drawn with Toucan. The diagram shows the three positive heart CRMs with the scored TF BSs represented in colored boxes, drawn with Toucan. Each putative CRM is called "seq" and named according first to the gene and then with a number.

In brief, of this "Top 20" list, thirteen putative CRMs were cloned upstream of the minimal heat shock 43 promoter driving nGFP expression and validated *in vivo*. We made a total of nine reporter constructs (**Table 2**) since some of the constructs bear two putative CRMs (pMC050: *pain-seq14+15-hs43-nGFP*; pMC062: *CG15537-seq1+11-hs43-nGFP*; pMC067: *Ca-beta-seq46+55-hs43-nGFP*). Of these thirteen putative CRMs, one construct containing two predictions (*CG15537-seq-1+11*) drives expression in the CNS in a *Hox*-like regulated pattern (**Figure 24**).

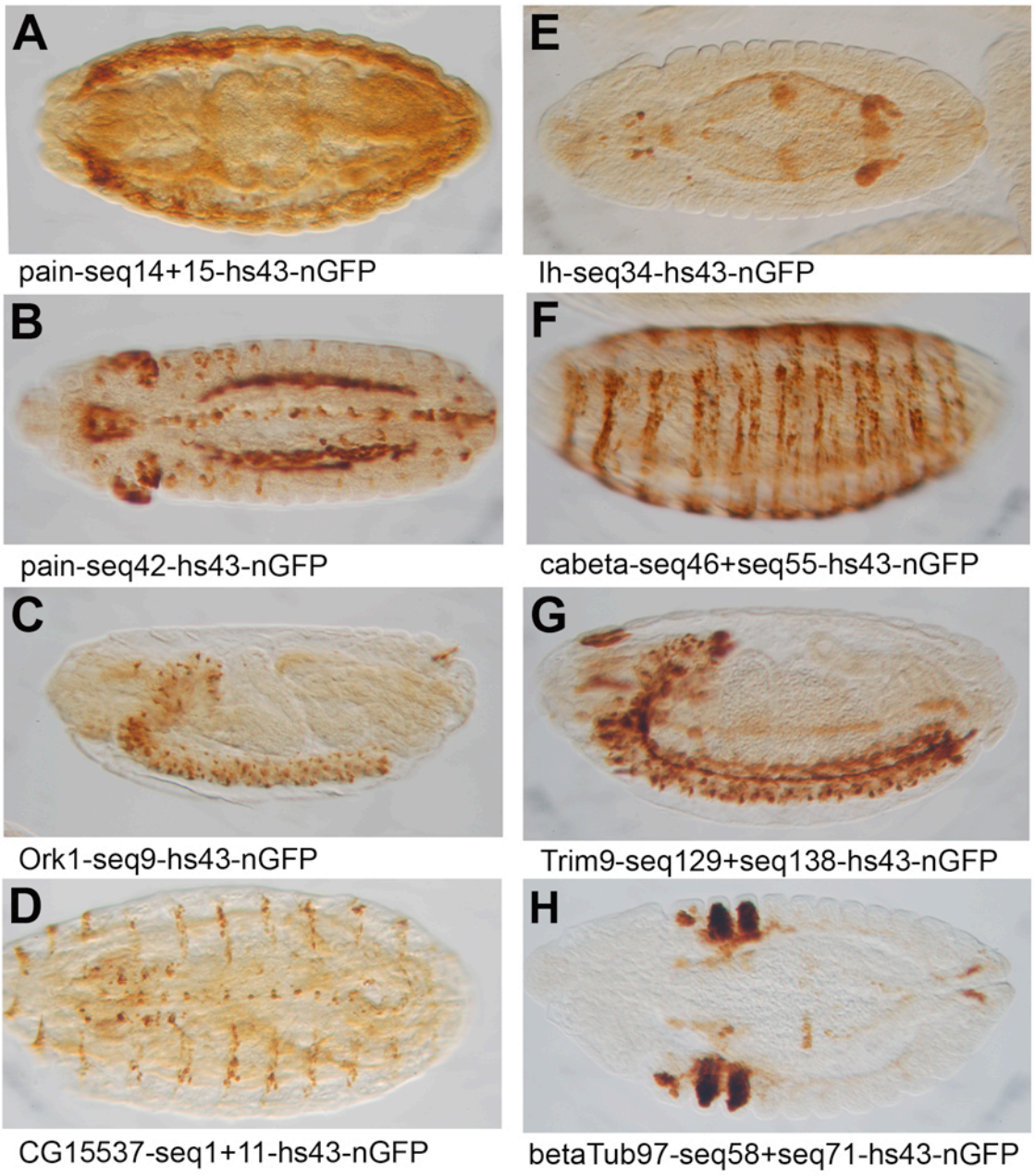


Figure 24. nGFP expression patterns of transgenic embryos carrying reporter constructs of Top 20 and Top 40 CRMs. The name of the reporter construct bearing the respective CRM is indicated below the correspondent embryo and the expression pattern is described in **table 2.** Staining was done by HRP.

Three CRMs (*Ork1-seq9* in pMC055; *Ih-seq34* in pMC077; *Dms-seq5* in pMC078) were found to drive nGFP expression in the heart of transgenic embryos in a pattern coinciding with that of the endogenous genes: exclusively in the *tin*-positive heart cells (**Figure 25).** The

Dms-seq5 and *Ih-seq34* constructs drive expression exclusively in the heart while the construct containing *Ork1-seq9* drives expression also in the central nervous system (**Figure 24**), a tissue known to express *Ork1*. For *Ih-seq34* we saw only in one line staining of the gonads (**Figure 24**) and therefore we consider it a position effect of the insertion of the transgene.

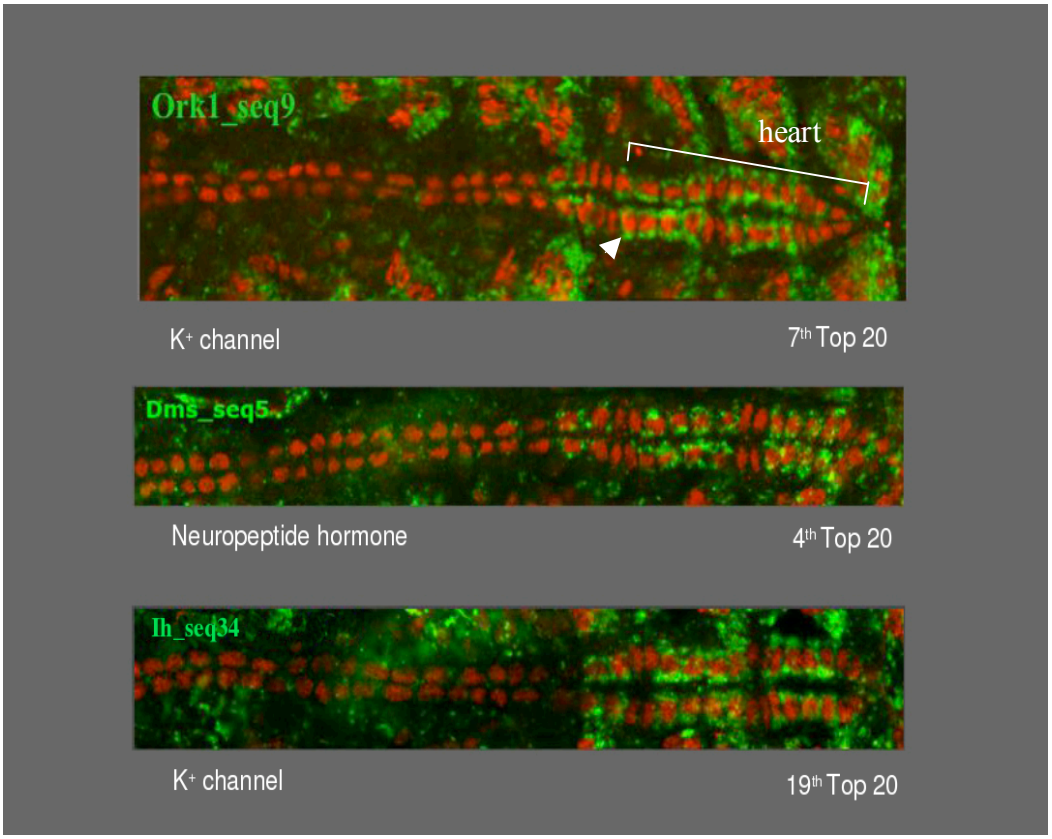


Figure 25. Heart positive enhancers in the Tin-positive cells. nGFP heart expression pattern of the three predicted CRMs as assayed by TSA amplified FISH. In green is shown the FITC labeled GFP probe and in red the TEXAS-RED labeled MEF2 antibody. The white arrow indicates one of the Tin-Positive cells where the enhancer is expressed. Pictures were taken under a confocal microscope using 100X magnification. Anterior is to the left.

"Top 40" and *in vivo* validation

After *in vivo* validation of the first dataset, we proceeded to apply the same bioinformatic method for *cis*-regulatory sequence discovery on the genes that are highly expressed in the heart with respect to the aorta through microarray experiments. The initial dataset consisted of 144 genes whose CNSs were extracted and treated accordingly to strategies A and B of the methodology described above. The final outcome is a "Top40" ranking of putative CRMs (**Table 4**). For *in vivo* validation of these putative CRMs, we set further parameters as explained in "Materials and Methods: Second bioinformatic approach based on a novel pattern matching based method". These parameters were set to reduce the number of false positives. For example, we gave priority to those CRMs nearest to the transcription start site of the gene or inside an intron of the gene).

Table 4 shows the "Top 40" as ranked in the final scoring. As for the "Top20", each putative CRM is called "seq" and named according first to the gene and then with a number. In bold are indicated which putative CRMs were assessed *in vivo*. **Table 1** shows the

oligonucleotides used to clone the putative CRMs to produce transgenic reporter construct for *in vivo* validation. For each tested construct, embryos were stained both by HRP and TSA amplified FISH and analyzed at the confocal microscope.

1. CG10440_seq_80	**11. CG10440_seq_28**	**21. CG15537_seq_14**	31. GlcAT-P_seq_12
2. CG10570_seq_1	**12. betaTub97EF_seq_58**	**22. Trim9_seq_129**	32. CG13907_seq_19
3. CG1443_seq_18	13. jp_seq_59	23. CG33115_seq_5	33. CG1443_seq_55
4. plx_seq_3	14. CG10253_seq_5	24. AP-50_seq_5	34. CG9339_seq_12
5. CG5001_seq_66	**15. betaTub97EF_seq_71**	25. Ance-5_seq_7	35. Tm1_seq_18
6. Trim9_seq_138	16. abd-A_seq_98	26. GlcAT-P_seq_35	36. CG31637_seq_89
7. CG10440_seq_89	17. zormin_seq_56	27. CG5656_seq_32	37. CG7956_seq_61
8. CG3308_seq_28	**18. CG15537_seq_11**	28. CG9326_seq_20	38. fat-spondin_seq_16
9. CG31217_seq_27	19. CG8193_seq_37	29. CG14688_seq_21	39. CG9339_seq_27
10. Sr-CI_seq_78	20. CG7160_seq_13	30. CG7294_seq_3	

Table 4. "Top 40" Ranking. In bold are indicated which putative CRMs were assessed *in vivo*.

Notice that the channel genes found by candidate gene approach are not present in the initial dataset. These genes are very weakly expressed and thus probably missed in the genome wide approach. Nevertheless, the *CG15537-seq14* prediction is present both in the "Top20" and "Top40" lists, as well as the prediction *CG15537-seq 1*. Of the predicted CRMs of the "Top40" assayed *in vivo*, none of them drives heart nGFP expression in embryos, although one (*BetaTub97-seq58+71* in pMC073) drives a *Hox*-like regulated pattern in the thoracic somatic mesoderm of transgenic embryos (**Figure 24**). The embryos, carrying a transgenic reporter construct bearing a Top40 CRM which drives an expression pattern, are shown in **figure 24** and the correspondent expression pattern is described in **table 2**.

Genetic studies for *in vivo* determination of *abd-A* regulation of the heart enhancers

The following experiments are underway. Results are expected to be collected by the end of April 2008.

For the CRMs which drive expression in the heart of embryos (pMC055, pMC077 and pMC078), we proceeded to verify if these sequences are effectively under *abd-A* regulation, since *abd-A* controls heart identity and has been shown to be required for the expression in the heart of at least two of these genes (*Ndae1* and *Ih*; Perrin *et al.*, 2004; Monier *et al.*, 2005). For each construct found to drive expression in the heart of transgenic embryos, we crossed two independent transgenic lines to *abd-A* loss- or gain-of-function backgrounds. In all crosses we used *lacZ* balancers (see "Materials and Methods: Fly genetics") in order to distinguish in the stained embryos those that are homozygote for the loss-of-function *abd-A* mutation or that carry the transgene allowing *abd-A* overexpression.

Loss of *abd-A* function

In order to analyze *in vivo* how the heart *cis*-regulatory sequences behave with respect to the function of the *Hox* gene *abd-A*, we generated flies carrying each reporter construct which drives expression in the heart and a complete loss-of-funtion allele of *abd-A* Thus, we obtained flies that bear the transgene which drives expression in the heart, but that lack *abd-A* function (see "Materials and Methods: Fly Genetics"). We expect the embryos of such a cross to lack the expression of the transgene, since we hypothesize that *abd-A* upregulates the expression of the corresponding endogenous gene.

Gain of *abd-A* function

To analyze if the expression of these heart *cis*-regulatory sequences is concomitant to the expression of *abd-A,* we generated flies that carry both the reporter construct found to drive expression in the heart and a *UAS-abd-A* transgene. We then used the UAS/GAL4 system described in "Materials and Methods: Fly genetics" in order to see if the ectopic expression of *abd-A* in embryos ectopically activates the nGFP reporter driven by the heart *cis*-regulatory sequences, using a late mesoderm driver (*24B-Gal-4*). (see "Materials and Methods: Fly Genetics")

Gain of *Ubx* function

UBX is expressed at high concentration in the posterior aorta and in very low concentration in the heart portion of the cardiac tube (Ponzielli *et al.*, 2002). In *Ubx* mutants, *Ndae1*, *Ork1* and *Ih* are expressed normally. In embryos ectopically expressing *Ubx* in the heart, *Ndae1*, *Ork1* and *Ih* heart expression is slightly reduced, suggesting that *Ubx* functions as a repressor of these genes (BM, LP, MS; unpublished). In order to gain insight about the role of *Ubx* in the regulation of heart specific enhancers, we analyzed the expression in embryos of these heart *cis*-regulatory sequences with respect to the ectopic expression of *Ubx.* We generated flies that carry both the reporter construct found to drive expression in the heart and a *UAS-Ubx* transgene. We then used the UAS/GAL4 system, to activate the ectopic expression of *Ubx*, using a late mesoderm driver (*24B-Gal4*), in order to see in embryos of this cross if the ectopic expression of *Ubx* lowers the nGFP expression in the heart driven by the heart *cis*-regulatory sequences (see "Materials and Methods: Fly Genetics").

DNAseI binding assays on CRMs found to drive heart expression

At the same time, and in order to evaluate if the predicted *abd-A* sites correspond to sites where ABD-A binds *in vitro*, we performed DNAseI binding assays with recombinant ABD-A protein on the enhancers that were found to drive nGFP expression in the heart. **Figure 26** shows a map of the sequence of the *Ork1* heart enhancer with the sites found *in vitro* underlined and the sites predicted by bioinformatics in red bold letters. **Figure 27** shows with red bars, the DNAseI footprints made by ABD-A or UBX proteins on the *Ork1* heart enhancer. Marked with a circle are those sites that coincide with the bioinformatic prediction. Note that not all the sites protected *in vitro* are predicted by bioinformatics. We are performing similar experiments on the other two heart enhancers found.

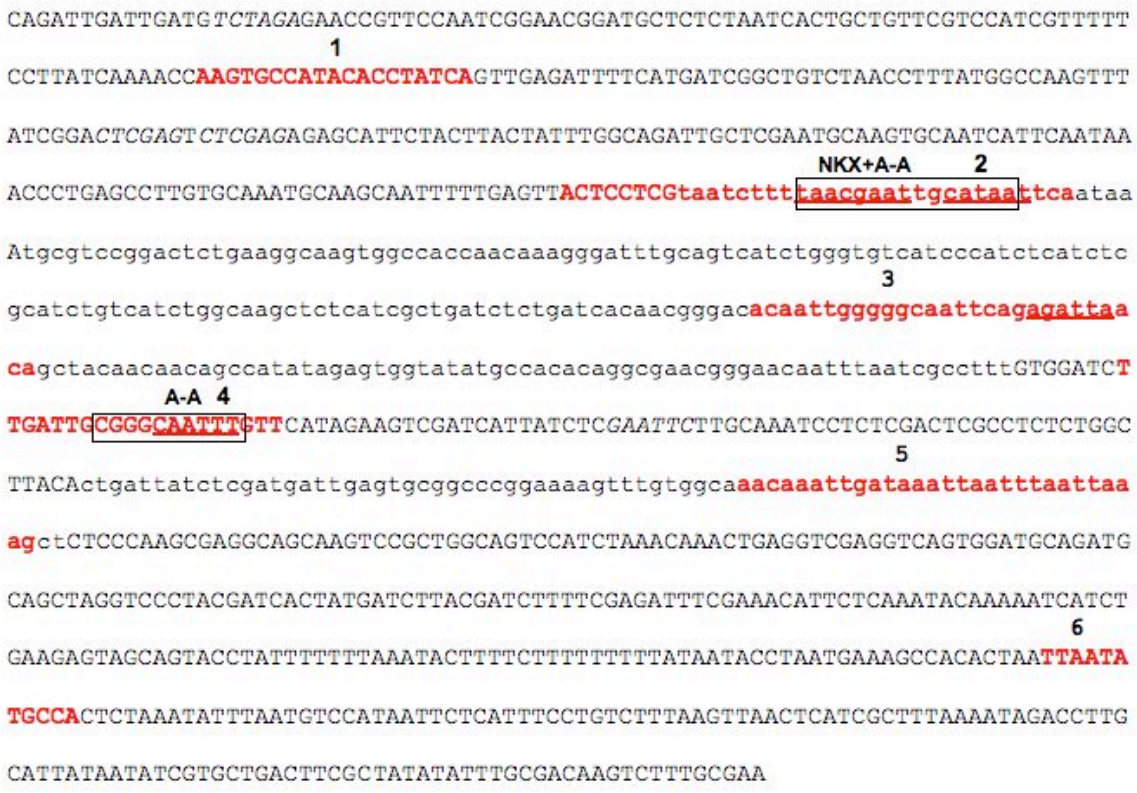


Figure 26. Map of the sequence of the *Ork1* heart enhancer. The sites found *in vitro* are underlined and the sites predicted by bioinformatics in red bold letters.

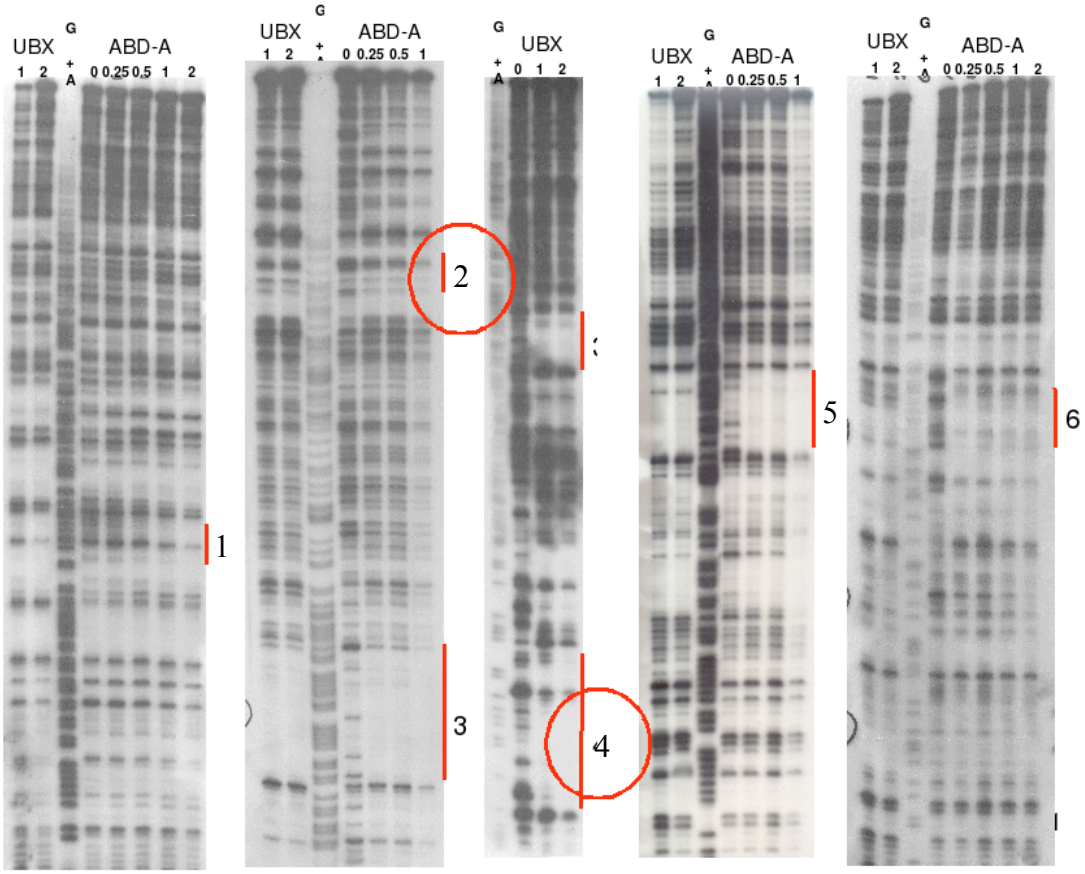


Figure 27. DNAseI footprints made by ABD-A or UBX proteins on the *Ork1* heart enhancer. Marked with red bars are the footprints and with circle are those protected sites that coincide with the bioinformatic prediction.

DISCUSSION

First bioinformatic analysis and *in vivo* validation for *Ndae1* and *Ih*

The bioinformatic tools used in our first *in silico* analysis were developed in 2004 (Grad *et al*., 2004; Frazer *et al.*, 2004; Berman *et al.*, 2004) These tools had already been used for enhancer discovery in *Drosophila* and new enhancers had been found. For example, eCIS-ANALYST validation step retrieved 15 new functional enhancers out of 37 predicted enhancer sequences. These eCIS-ANALYST results had been obtained on early stages of *Drosophila* development, where one of the most extensively described and best understood transcriptional regulatory networks is acting (Berman *et al.*, 2004). The efficiency percentage of *cis*-regulatory module discovery was reported high. On the other hand, the suite of VISTA tools (Frazer *et al.*, 2004) was becoming increasingly used probably due to a user-friendly interface. In particular, the VISTA Browser was the only server to allow viewing pre-computed alignments of many species and the rVISTA tool (Loots *et al.*, 2002), although it had been initially validated in Mammals, had been created on bases similar to eCIS-ANALYST and thus it allowed us to compare outputs among programs. This is the reason of our choice to use these tools and not others.

Different from the validation conditions of eCIS-ANALYST, we performed our search on genes that are expressed late in *Drosophila* development. Another disadvantage with respect to the validation conditions of the tools used (eCIS-ANALYST and rVISTA) was that some of the PWMs for the TFs of interest were degenerate, such as those for HOX and GATA, or created *de novo* from very few experimentally *in vitro* or *in vivo* validated binding sites, like the SVP matrix. With such matrices, a higher percentage of false positives is expected *a priori*. For this matter, we performed analysis with and without a degenerate PMW and compared the results. We found a great deal of irreconcilable variation from the latter results and this can be due to the fact that the short core binding sequence (TAAT) of *Hox* proteins occurs very frequently in the *Drosophila* genome, leading to a higher number of false positives in our analysis. The same reasoning is applied when performing an analysis where the GATA PWM is introduced. Different from the HOX PWM, we could avoid the use of the GATA PWM in further clustering analysis.

Another difficulty we encountered during *in silico* analysis was that of deciding which is the optimum window size in order to score a sequence with a given PWM. We performed various analyses under the same conditions and changing only the window size, in an attempt to define a window size when searching for heart enhancers or *Hox*-regulated enhancers in a sequence. Again, the optimum window size varied from gene to gene and depending on the PWMs we used to perform each analysis. No statistic criteria defined the window size and therefore comparison of outputs for different genes and with different matrices was the chosen method to determine the desirable window size. The lack of a statistic procedure to determine this variable increased the error of prediction, adding up to the problem of specificity of the PWMs.

Moreover, depending on the window size, we also noticed a great deal of variation in the extension in base pairs of the clusters and in the quantity of TF BSs found within a cluster (**Figures 13**). Through the analysis of our sequences using the mentioned online tools we concluded that there were too many variables in order to rule out a significant number of false positives, and therefore we introduced all TF BSs found in an alignment. This artisan way allowed us to analyze where TF BSs found were, if we used different levels of stringency in the scoring step and how these TF BSs were positioned with respect to regions of different conservation percentage. This strategy allowed us to notice sites that were not found in the other studies because they are sites that, for example, are distant from the consensus. Since the *in vivo* tested binding sites to build the PWMs used here were few, this data could have been informative if we could have found a matching pattern between *Ndae1* and *Ih* (*e.g.* if we could have found three GATA BSs for every HOX BS in one gene and in the other). Though this was not the case, DNAseI binding assays would have been useful to assess *in vitro* the information given by the alignment.

To increase the efficiency of prediction, we performed a comparison analysis of the

outputs of the different programs used, keeping constant a small region in the gene in order to identify coincidences between the different tools used. Disappointing was to find only two overlapping results, of which were only for *Ndae1*. More disappointing still was to observe that these fragments, when tested *in vivo*, did not drive expression in the heart. The two overlapping results drove expression in the ventral nerve chord (pMC008: *Ndae1-IntronI-hs43-lacZ*; **table 2; figure 17**) and in the anal plate (pMC025: *Ndae1-Hunch-hs43-lacZ*; **table 2; figure 17**).

None of the tested reporter constructs of *Ndae1* drove expression in the heart. A possible explanation is that the *Ndae1* heart enhancer is very weak and therefore maybe missed even if we have stained the embryos using techniques that amplify the labeling signal during a FISH experiment. Expression in the heart of the endogenous *Ndae1* gene is very weak and hard to see through FISH even if the signal is amplified (BM and LP, personal communication). The transgenic reporter constructs we analyzed for *Ndae1* cover the whole gene, from the previous CG to next, except for 292 bp upstream the transcription start site. Thus, there is a faint possibility that the heart enhancer of *Ndae1* is in these 292bp**.** It is generally acknowledged that an enhancer element is around 300-1000 bp, so this could be the case. For example the cardiac tube enhancer of *dHand* is 300 bp (Han & Olson, 2005), as the *dSur* cardiac tube enhancer (Akasaka *et al.,* 2006; Hendren *et al.*, 2007). Nevertheless, from the experience we have acquired studying *Ndae1*, we think it is more likely that a bigger fragment containing these 292 bp is needed in order to detect a signal, as we have observed for the overlapping fragments in the 5' UTR where, despite the fact that pMC029 and pMC025 drive the same expression pattern in embryos, pMC029 expression is stronger (data not shown). As we have recently learnt from *Ih* (see **figure 22: PCRI, figure 25: Ih-seq34, table2: PCR1 "notes"**), it is possible that by adding these few bases to the construct pMC029 (see **figure 16**) we retrieve heart specificity. It is worth noticing, though, that this short *Ndae1* fragment has not been detected in the different outputs of our *in silico* analysis.

Evaluating our results altogether, a second possibility raised is that elements that are distant from each other need to interact in order to activate transcription, and this is a situation that we could not have detected either through bioinformatics or molecular methods. The next question then would be: which are the fragments that interact?

The transgenic reporter lines for *Ndae1* constructs that drove an expression pattern (**Figures 17**) did so in domains of expression of the gene already described by *in situ* hybridization (Romero *et al.*, 2000), except for the lines *Ndae1-InVII-hs43-lacZ* (pMC024, see **Table 2** and **Figure 17**) which are expressed in the haemocytes, an expression domain not yet been described for *Ndae1*. Furthermore, it is probable that the lateral cells stained in the embryos bearing the construct *Ndae1-InIII-hs43-nGFP* (pMC035; **Table 2** and **Figure 17**) are cells of the peripheral nervous system (PNS), but a double staining with anti-GFP and an antibody for a general marker of the PNS, like 22C10 (Kania *et al*., 1995), is required for confirmation. Nevertheless, and although we made transgenic constructs covering the whole gene, we did not find the enhancer regions for all of the organs/tissues where *Ndae1* is reported to be expressed. For example, we have not found the enhancer of *Ndae1* that drives expression in the Malpighian tubules (MT) nor the one that drives expression in the central nervous system. Again, like in the case of the heart, this raises the possibility that *Ndae1* expression in the MT and CNS is driven by elements that are distant from each other and need to interact in order to activate transcription. Therefore only by assessing bigger reporter constructs or combining fragments in a reporter construct we will be able to identify them. Furthermore, is not to disregard again the possibility that the not-cloned 292bp may contain one or both these enhancers. Again, immunoprecipitation of the whole *Ndae1* genomic DNA would be necessary to provide us with some hints for a start point.

In addition, the expression patterns observed are not reported domains of expression of the combination of two or more of the TFs used in this *in silico* analysis. On the contrary, the choice of the TFs used was made on the assumption that the unique combination of all of them confers heart specificity to any given target gene. For example, *abd-A* confers heart specificity

(Ponzielli *et al*., 2002) and *tin* is expressed exclusively in the mesoderm and becomes restricted to the cardiac cells later in embryogenesis (Bodmer, 1993). One could expect that a cluster containing TIN BSs and ABD-A BSs drives expression in the heart. Nevertheless, we have observed that the fragment where we obtain overlapping results among the bioinformatic approaches (see **Figure 15**), although it contains TIN and ABD-A BSs, drives expression in the anal pads, an expression domain not described for either *tin* or *abd-A*. This gave us a clear example of a false positive bioinformatic output.

As we already referred, the clustering methods were setting a benchmark in 2004. Nevertheless, the validation of these methods had been done either in early stages of development of *Drosophila* (eCIS-ANALYST), where the transcriptional regulatory network governing these stages was well known and characterized, or in humans (rVISTA). In spite of all this, the accuracy of prediction of these methods had not been tested in poorly characterized genes in later developmental stages of *Drosophila*. For this reason, and because of the ambiguous bioinformatic results we had obtained for *Ih* and *Ndae1* using clustering methods, for *Ih* we decided to contemplate only the sequence identity variable as a starting point to clone and produce transgenic reporter flies to test *in vivo* the bioinformatics predictions. Most methods reported up to 2004 used interspecies sequence comparison alone as a method to identify regulatory sequences. These methods had been remarkably effective in identifying mammalian regulatory sequences (Pennacchio & Rubin, 2001) and some preliminar studies had been published in *Drosophila* (Bergman *et al.*, 2002).

For *Ih*, all the transgenic flies show expression in embryos in tissues that we are not certain they belong to the endogenous gene expression pattern (see **table 2** and **figure 22**), since until now there are no data describing it. The *in situ* hybridization experiments we have performed to describe the pattern of expression of the endogenous gene have been done by FISH, and we have obtained high background. For this matter, we are performing ISH with alkaline phosphatase in order to determine if the found enhancer elements actually belong to *Ih*. Nevertheless, none of these constructs was expressed in the heart. *Ih* expression in the heart is slightly stronger than that of *Ndae1,* therefore less likely to be missed by the observer when analyzing embryos stained by amplified FISH. This raised again the possibility that distant elements might be interacting to activate transcription. Another plausible explanation was that the sequences we had cloned to make the reporter constructs were too short, and therefore missing the piece that confers heart specificity to the enhancer. In fact, just recently, we found out this to be the case (see **figure 22: PCR1** and **figure 25: Ih-seq34-hs43-nGFP**, **table 2: PCR1 "notes"**).

The results obtained show that this first bioinformatic analysis we applied is not efficient in finding heart *cis*-regulatory regions. Moreover, we showed that taking into account only the "conservation" parameter in order to find heart *cis*-regulatory sequences is not sufficient, at least in *Drosophila*. On the other hand, it can be argued that conservation alone is a good start point for a general view of regions where *cis*-regulatory modules for different tissues where a gene might be expressed, since all the transgenic reporter constructs made for *Ih* were made taking into account only conservation blocks between *D. melanogaster* and *D. pseudobscura* and they all show an expression pattern in embryos in at least three independent lines. Still, we think that searching for conserved clusters of BSs for all the TFs known to be involved in heart development was the method to use if we needed to assay a large number of genes. Subsequently, with the data obtained from microarray experiments we found the urge to refine our bioinformatic approach, since assessing through transgenic constructs the whole sequence for each putative *abd-A*-regulated heart specific gene would have been an unaffordable enterprise.

The pattern matching based approach and *in vivo* validation of heart specific enhancers

Through the three years of this project, seven genes were found to be differentially expressed along the cardiac tube through a candidate a gene approach. Moreover, the data obtained from microarray experiments showed 144 genes overexpressed in the heart with respect to the aorta. This rendered the pool of putative *abd-A* target genes so large that could only be assessed through bioinformatic approaches. Since there were no bioinfomatic tools available for *Hox* specific targets, we designed our own method combining clustering strategies that gave priority to high affinity ABD-A BSs on one hand and to the conservation of heart expressed TF BSs on the other. This new method was first tried using as initial dataset seven genes found by candidate gene approach plus two genes found through microarray experiments. This analysis lead to the list of 20 best predictions, the "Top20". Using this small initial dataset we have found three heart enhancers for three different genes. The rank position of these three positive CRMs (fifth, ninth and nineteenth) in our "Top20" list lead us to think the method can be used to find *Hox* target enhancers. Taking into account only the validated *in vivo* CRMs, it results that the efficiency of prediction is high (3/13 = 23%). Moreover, when applying the same method to a large initial dataset constituted by the 144 genes found through microarrays, with an outcome of a "Top40" list, we found that the predictions in the "Top20" for the gene *CG15537* (CG15537-seq14 and CG15537-seq11) can be retrieved again in the "Top40". On the other hand, it is to notice, that the CRM found in the "Top20" for the gene *Dms*, which is a true positive prediction, cannot be found in the "Top40". We did not obtain significant scores for the predictions of the genes *seizure* and *Ca-P60,* which can indicate that the parameters used are stringent enough to exclude false positives.

For the "Top40" ranking we have not found yet any positive result in the *in vivo* validation, but it is also true that we have not yet *in vivo* validated enough predictions in order to arrive to any conclusion. Nevertheless, so far, we have found one *Hox*-like regulated enhancer (*BetaTub97-seq58+71* in pMC073) out of nine. The construct made with *CG10440-seq80+89* (pMC069, **table 2**), bearing two predictions ranked positons 1 and 7, respectively, in the "Top40" did not drive expression in the heart. *CG10440* encodes a voltage gated potassium channel protein. *CG10440* has been detected to have a 2.5 fold change expression in the genome wide approach (data not shown) while the other mentioned channel proteins have not been detected in the microarray output. As we have noticed before, channel proteins have a very weak expression in the heart and thus are difficult to detect. In fact, the endogenous expression of these channels may be detected only using the FISH technique with an amplification step by means of TSA and using confocal microscopy. Consequently, the heart enhancers of these channels are difficult to identify. It is possible that the heart enhancer of this channel protein lies within the prediction but we have not detected it due to technical problems. Thus, we are producing a construct that bears the putative CRM *CG10440-seq80* within a bigger piece of genomic DNA, since it is acknowledged that this strategy usually increases the signal of the reporter and we have seen this to be the case for overlapping constructs of *Ndae1*.

In any case, more predicted CRMs of this "Top40" need still to be analyzed. The next sequences that will be validated *in vivo*, will be first validated *in vitro*. Since validation *in vivo* is time consuming, we will perform inmunoprecipitation assays using ABD-A protein, in order to validate more predictions in a shorter time. Those CRMs, which are positive also in *in vitro* assays, will be cloned upstream of the nGFP reporter and injected to produce transgenic flies and examine expression in embryos.

In a second bioinformatic step, we decided to perform a search using the positive results as model targets. We first compared, the enhancers among them and after with the other genes of

the dataset. We have not found similarities among the sequence position of the motifs in the enhancers. Moreover, when comparing these positive results to the genes in the dataset, we have not found similarities. This is leading us to think that there is no repetitive pattern of TF BSs. In other words, that there might be no trace in the DNA sequence in order to predict a prototype heart enhancer. From another point of view, we are comparing the positive results with the orthologous sequences of species other than *D. pseudobscura.* In this analysis the method used is described in Aerts *et al.*, 2007. Finally looking at the initial dataset one can observe that genes belonging to different Gene Ontology (GO) groups have been analyzed in the same way. Instead, one can think, that, for example, *Ih*, encoding a channel protein, has a different structure of its heart enhancer than *Dms,* which encodes a peptide hormone, but probably similar to that of *Ork1*, which also encodes a channel protein. Supposing that different GO groups have different heart enhancers, we are designing new filters to apply to the *cis*-regulatory search in order to take into account the function of the target gene.

In vitro and *in vivo* data will also provide evidence about the trace TFs may leave in the sequence they regulate. Indeed, from the DNAseI footprints on the *Ork1* heart enhancer, it is curious to notice that there are sites like "site 1" (put sequence, **figure 26)** that are not typical ABD-A or UBX binding sites, as it does not have a TAAT or TAAT-like core. We have shown also that not all the binding sites protected *in vitro* are predicted by bioinformatics (**figure 27**), at least for the *Ork1* heart enhancer. To understand which are the *in vivo* functional BSs of ABD-A, mutagenesis experiments are being performed in the following way:

mutate all the ABD-A biding sites found through DNAseI assays;
mutate only the ABD-A binding sites found through bioinformatics.

The mutated enhancers have been cloned upstream of nGFP and will be injected in fly embryos in order to produce mutated reporter construct flies. Again we will assay the expression of these constructs in embryos and perform fly crosses in order to assay expression in a background where *abd-A* is overexpressed. This *in vivo* data will provide the necessary evidence to find out not only which of such ABD-A BSs are functional *in vivo*, but also will determine whether these heart enhancers are direct targets of the *Hox* gene *abd-A*. Moreover, if these "unconventional" BSs found by footprints were to be functional, then we could start studying if such sites are present in all heart enhancers under *abd-A* control to analyze the possibility these "unconventional" BSs are the reason of heart specificity.

Analyzing the expression of the heart enhancers in *abd-A* gain and loss of function backgrounds will provide strong evidence about the key role of *abd-A* in granting heart specificity to these genes. Although *abd-A* confers heart specificity, probably it does not account on its own for specific target gene expression, because the target genes we have found (and thus their enhancers) are expressed only in the four TIN-positive cells of each segment that constitutes the heart. In order to evaluate if there is cooperation of ABD-A and TIN for the regulation of such target genes, we are also performing DNAseI binding assays on the heart enhancer sequences using TIN protein, and we have in program to perform protein-protein interaction studies. Nevertheless, we do not exclude other TFs, like TBX or HAND participating in the regulation of the expression of the heart specific genes found.

Though UBX is found at very low concentrations in the heart cells of the cardiac tube, we have not excluded a possible role of UBX as a repressor in the regulation of these heart enhancers. To gain insight in this matter we will study reporter gene expression in a *Ubx* gain-of-function background. If the expression of the transgene remains unaltered we could conclude *Ubx* is not involved in regulation of genes expressed differentially in the heart with respect to the

aorta. Instead, if expression of the reporter is not observed or is lowered, we will provide evidence for a possible repressor role of UBX in the regulation of heart specific genes.

From a broader point of view, we are analyzing the late role of *abd-A* as an activator. There is only indirect evidence that ABD-A is an activator and UBX is a repressor in late stages of cardiogenesis (BM, LP, and MS, unpublished). This has remained an open question due to the impossibility of assaying expression in double mutant flies, because such flies do not have appropriate abdominal lineage choice, which is the early function of *Ubx* and *abd-A* in cardiogenesis (Perrin *et al.*, 2004). Thus, in these double mutant embryos, all cardiomyocytes are specified as thoracic cardiomyocytes. To learn more about this situation, we have crossed flies deficient for *abd-A* and *Ubx* (*Df 109* flies; see "Materials and Methods: Fly Genetics") to flies which bear the *UAS>abd-A*^{Hx} transgene (Merabet *et al*., 2003). *UAS>abd-A*$^{Hx.}$ are flies which carry an inducible mutated hexapeptide variant of *abd-A* capable of maintaining normal lineage choice in early cardiogenesis and also capable of repressing *Ubx*, but unable to lead to ectopic expression of, for example, the target gene *Ih* (BM, LP and MS, unpublished). Thus, the absence of *Ubx* does not seem to allow expression of *Ih*, which may require positive regulation by *abd-A*. Recombinant *UAS>abd-A*^{Hx}, *Df109* flies where then crossed to recombinant 24BGal4, *Df109* flies (see "Materials and Methods: Fly Genetics"). One fourth of the progeny of this cross will be homozygous for the *Df109* chromosome (lack *Ubx* and *abd-A*) and will express *abd-A*^{Hx} in the whole mesoderm (due to the 24BGal4 driver). By eliminating both *Hox* genes, but leaving intact normal lineage choice towards abdominal fate so that the cells that form the heart and posterior aorta become specified, we will analyze how does differentiation of cardioblasts proceed in the absence of late *Hox* input.

Conclusions

We use cardiogenesis in *Drosophila* as a model to study *Hox* function in organogenesis and *Hox* specificity during development. The aim of this work was that of gaining knowledge on the role of *abd-A* in conferring heart identity. For this we needed to find the heart enhancers of putative direct *Hox* target genes expressed only in the heart portion of the cardiac tube and evaluate if *abd-A* genetically regulates these enhancers and if *abd-A* protein binds to this enhancer sequences. In the process of this project we have learnt that the specific heart *cis*-regulatory sequences of putative *abd-A* heart targets are difficult to assess *in vivo* since we have found out that these can only be detected by the use of FISH techniques using TSA amplification and confocal imaging. Nevertheless, we have found three heart enhancers that belong to three true realizator genes and which we are in the process of analyzing in order to determine if these are direct *abd-A* targets and if they are co-regulated also by other TFs, such as TIN.

From another point of view, we have designed a bioinformatic method to detect heart enhancers of *abd-A* target genes in order to speed up the *in vivo* search of heart enhancers. *Hox* genes are difficult to assess through bioinfomatics or *in vitro* since their proteins bind to a short TAAT core sequence which occurs every 1 kb in the genome and that produces a PWM said to be highly "degenerate". Nevertheless, our method has detected three heart enhancers and two *Hox*-like regulated enhancers, and although this method still needs optimization, the efficiency of enhancer discovery has proven to be high. Finding other heart enhancers and adding filters like grouping by GO will help in the optimization. The ultimate aim of this work in progress is to arrive to construct a prototype heart enhancer that will help us understand abnormalities which can lead to congenital heart diseases. What we have learnt up to the moment from the three heart enhancers found is that they do not have characteristics in common such as the quality, quantity or position of TF BSs. Therefore, we need to determine which are the functional sites *in vivo* to arrive to any conclusion. From an evolutionary point of view, since the larval heart of *Drosophila* is an organ dispensable for viability, it is possible that the turn over rate of heart TF BSs is high and therefore heart enhancers are not highly conserved. Interspecies studies will throw light into this matter. Despite many efforts, we are still far from understanding how *Hox* genes, like *abd-A,* regulate downstream realizators in order to form a functional organ like the heart.

TABLES

Table 1: Oligonucleotides

NAME	SEQUENCE	USE	CONSTRUCT
oLP 001 Ih 5primeF	CGAATCGGTGTGTTTCTGC	clone Ih PCR1 in CHAB	Ih PCR1 lacZ
oLP 002 Ih 5primeR	CCATCGAACACGGTACTCAT	clone Ih PCR1 in CHAB	Ih PCR1 lacZ
oLP 003 Ih IntronIF	GAATCGAACGCGTTAACCG	clone Ih PCR2 in CHAB	Ih PCR2 lacZ
oLP 004 Ih IntronIR	GCCTTTAGTTGGTGCCTTCA	clone Ih PCR2 in CHAB	Ih PCR2 lacZ
oLP 005 Ih IntronVIF	TTCTCAAACCATCTGACGC	clone Ih PCR3 in CHAB	Ih PCR3 lacZ
oLP 006 Ih IntronVIR	CCCCACTTCTCTTGTTTTGTA	clone Ih PCR3 in CHAB	Ih PCR3 lacZ
oLP 007 Ih IntronVI 3primeF	AGCTTCTTGGAGTGACACC	clone Ih PCR4 in CHAB	Ih PCR4 lacZ
oLP 008 Ih IntronVI 3primeR	GCCACCAAAAACACACACAC	clone Ih PCR4 in CHAB	Ih PCR4 lacZ
oLP 009 Ih 3primeF	AAAAGGAACCCCAAGACGC	clone Ih PCR5 in CHAB	Ih PCR5 lacZ
oLP 010 Ih 3primeR	CGGTCCAGTTTGGAGACTTA	clone Ih PCR5 in CHAB	Ih PCR5 lacZ
oMC007 ndaeIF Not	**GGCGGCCG**CCTGAAAAGAAGAGTGGCATT	clone Ndae intron I in CHAB	pMC008: Ndae1-inI-lacZ
oMC008 ndaeIR Bam	**GGGGATCC**GTAAGTAGGACGGTGTGATT	clone Ndae intron I in CHAB	pMC008: Ndae1-inI-lacZ
oMC013	**GGATCC**GGTGTGCTAATCAAGTTTACGTCG	to clone "the hunch" of Ndae in CHAB	pMC025: Ndae1-Hunch-lacZ
oMC014	**CCGCGG**ACAACAGGGCGTATGAATTCG	to clone "the hunch" of Ndae in CHAB	pMC025: Ndae1-Hunch-lacZ
oMC015#2 5'UTR4.4NdaeF	**CGGGATCC**GCCGATCTTTAAACTGAAGCA	to clone 3' of 5' UTR of Ndae in CHAB (fra4.4)	pMC029: Ndae 5'UTR4.4-lacZ
oMC016#2 5'UTR4.4NdaeR	**ATCCGCGG**GACAACAGGGCGTATGAATTC	to clone 3' of 5' UTR of Ndae in CHAB (fra4.4)	pMC029: Ndae 5'UTR4.4-lacZ
oMC017#2 5'UTR4.6NdaeF	**CGGGATCC**CAACCCAACACCCCTTCA	to clone middle of Ndae 5' UTR in CHAB (fra4.6)	pMC027:Ndae 5'UTR4.6-lacZ
oMC031 3'UTRNdaeF	**CCGCTCGAG**CTTCAATATAAAATGGCATATTTGCA	to clone 3'UTR of Ndae in CHAB	pMC020: Ndae1-3'-lacZ
oMC032 3'UTRNdaeR	**CGGAATTC**AAAATCCCACGGACCAGTG	to clone 3'UTR of Ndae in CHAB	pMC020: Ndae1-3'-lacZ
oMC033 InVIINdaeF	**TTGCTCGAGCTG**TAATATAAACCATTGGGTACAGTT	to clone intron VII of Ndae in CHAB	pMC024: Ndae1-inVII-lacZ
oMC034 InVIINdaeR	**CCGGAATTC**GTGAGTCGAATAAATCAATTAAAACAA	to clone intron VII of Ndae in CHAB	pMC024: Ndae1-inVII-lacZ
oMC035 InIIINdaeF	**CCGCTCGAG**CTGCATTAGTCGCGTTTTTTT	to clone intron III of Ndae in CHAB	pMC028: Ndae inIII-lacZ
oMC036 InIIINdaeR	**TTTTTGCGGCCGCG**TGAGTATGGGGGTGTTCAC	to clone intron III of Ndae in CHAB	pMC028: Ndae inIII-lacZ
oMC037 InIINdaeF	**GGCTCGAGGTTC**AAGGATACTTTATGAGAAACAGA	to clone intron II of Ndae in CHAB	pMC026: Ndae-inII-lacZ
oMC038 InIINdaeR	**GGCTCGAGGTTCAAG**GATACTTTATGAGAAACAGA	to clone intron II of Ndae in CHAB	pMC026: Ndae-inII-lacZ
oMC049 Ndae5'UTR4.6Not-F	**AAAGCGGCCGC**CCCAAAGTGGACATGCAG	to clone middle of Ndae 5' UTR in CHAB (fra4.6)	pMC027:Ndae 5'UTR4.6-lacZ
oMC055 Ndae-inV-VI-F	**CCGGATCC**CTGAGGATTTCAAAAAGCAAAAGT	to clone inV+VI in Chab	NOT injected
oMC056 Ndae-inV-VI-R	**GGGAATTC**GTAAGGAATCTTTTTAGGCTTTAAAGA	to clone inV+VI in Chab	NOT injected
oMC057 Ndae-inIII-5'-chopF	**CCGCTAGC**CTGCATTAGTCGCGTTTTTTT	to clone 3' of intron III in pH-Stinger	pMC035: Ndae-InIII nGFP + pMC036: Ndae InIII 3'half-hs43-nGFP
oMC058 Ndae-inIII-5'-chopR	**CCAGATCT**AATCTAGTATGCCTTTTACTCTACGAA	to clone 3' of intron III in pH-Stinger	pMC036: Ndae InIII 3'half-hs43-nGFP
oMC060 Ndae-inIII-3'-chopR	**CCAGATCT**TGAGTATGGGGGTGTTCAC	to clone 5' of intron III in pH-Stinger	pMC035: Ndae-InIII nGFP
oMC72 Ih5'TST-F	**CCTCTAGA**AAGGGGTCATCCGTCACTCCAGT	to clone bigger Ih PCR1	pMC040: BS Ih PCR6
oMC073 Ih5'TST-R	**CCTCTAGA**CAATGAGTACCGTGTTCGATGGC	to clone bigger Ih PCR1	pMC040: BS Ih PCR6
oMC78 NdaeInIII+eIII-VII-F	**GGTCTAGA**CTGGCGCAGGAACATATAGTCTGG	to clone inIII + environment	pMC038: Ndae inIII+environment-hs43-nGFP
oMC79 NdaeInIII+eIII-VII-R	**GGTCTAGA**GCACAGCGAGTTCAATTCATACTTG	to clone inIII + environment in pHStinger	pMC038: Ndae inIII+environment-hs43-nGFP
oMC104 Dms5'-1011-S/E	**CCGAATTC**CTATCTTCTCACTGCATTAGTCAC	to clone Dms CNS 1+2 in pStinger (and pH-Stinger?)	pMC046: Dms-seq5-nGFP + pMC078: Dms-seq5-hs43-nGFP

NAME	SEQUENCE	USE	CONSTRUCT
oMC110 15537-seq14-F/E	**CCGAATTC**GCATGTCGCAGGTTCCAATAAATG	to clone CG15537 seq14 in pStinger	pMC049: CG15537-seq14-nGFP + pMC077: CG15537-seq14-hs43-nGFP
oMC111 15537-seq14-R/B	**CCGGATCC**GCAATCTGAGGAAGGGGTAAAAGC	to clone CG15537 seq14 in pStinger	pMC049: CG15537-seq14-nGFP + pMC077: CG15537-seq14-hs43-nGFP
oMC114 Pain-seq42-F/E	**CGGAATTC**GACCTTTTAAGTACGGATGCCAC	to clone pain seq42 in pH-Stinger	pMC052: pain-seq42-hs43-nGFP
oMC115 Pain-seq42-R/B	**GGGGATCC**GGGAGATGAGTTCAAATTGGGAC	to clone pain seq42 in pH-Stinger	pMC052: pain-seq42-hs43-nGFP
oMC116 Pain-seq14+15-F/E	**CCGAATTC**GTAAACAAACCCGTCAATCAACC	to clone pain seq14+15 in pH-Stinger	pMC050: pain-seq14+15-hs43-nGFP
oMC117 Pain-seq14+15-R/B	**CGGATCCC**TGCACATCACAAAAGTGGATAAT	to clone pain seq14+15 in pH-Stinger	pMC050: pain-seq14+15-hs43-nGFP
oMC118 Ih-seq34-F/E	**GGGAATTC**GCGCCGGTGTTTAGCTATATCAAC	to clone Ih seq34 in pH-Stinger	pMC051: Ih-seq34-nGFP + pMC066: Ih-seq34-hs43-nGFP
oMC119 Ih-seq34-R/B	**GGGGATCC**TGGAGGGAGACGAAACGCAG	to clone Ih seq34 in pH-Stinger	pMC051: Ih-seq34-nGFP + pMC066: Ih-seq34-hs43-nGFP
oMC122 BetaTub97seq58-F/K	**AAGGTACC**AGAGGAGGCGCACTAAACACTAAAT	to clone BetaTub97 seq58 in pHStinger (1688bp)	pMC065: betaTub97-seq58-hs43-nGFP
oMC123 BetaTub97seq58-R/B	**ATGGATCC**GGTTGGGGTGAGAGAGGACTAATTT	to clone BetaTub97 seq58 in pHStinger (1688bp)	pMC065: betaTub97-seq58-hs43-nGFP
oMC0124 BetaTub97seq71-F/Bg	**TTAGATCT**TTCATTGTCGTGTAATTTGTGCAG	to clone BetaTub97 seq71 in pHStinger (1351bp)	pMC072: betaTub97-seq71-hs43-nGFP
oMC0125 BetaTub97seq71-R/Xb	**CCTCTAGA**ACGGAGTGAGCAGATAATTTTGCTT	to clone BetaTub97 seq71 in pHStinger (1351bp)	pMC072: betaTub97-seq71-hs43-nGFP
oMC0126 CG10440seq80-F/K	**AAGGTACC**GTGCAGGTGTGACTGTGTGTGTG	to clone CG10440 seq80 in pHStinger (1068bp)	pMC061: CG10440-seq80-hs43-nGFP
oMC0127 CG10440seq80-R/B	**ATGGATCC**TCCGAATGGCTGCTACGAATAGTAA	to clone CG10440 seq80 in pHStinger (1068bp)	pMC061: CG10440-seq80-hs43-nGFP
oMC0128 CG10440seq89-F/Bg	**CCAGATCT**GTATTTAAACTGCATCGACTTTTGA	to clone CG10440 seq89 in pHStinger (1322bp)	pMC068: CG10440-seq89-hs43-nGFP
oMC0129 CG10440seq89-R/Xb	**TGTCTAGA**TTACAAATTAGCCGTCAATGGAG	to clone CG10440 seq89 in pHStinger (1322bp)	pMC068: CG10440-seq89-hs43-nGFP
oMC130 Trim-seq129-F/K	**TGGGTACC**GATGGATGTGTTTTCATAGCAACCT	to clone Trim seq129 in pHStinger (1079bp)	pMC064: Trim-seq129-hs43-nGFP
oMC131 Trim-seq129-R/B	**TTGGATCC**TGTGCTACGTGTGCGACATTTTC	to clone Trim seq129 in pHStinger (1079bp)	pMC064: Trim-seq129-hs43-nGFP
oMC132 Trim-seq138-F/Sp	**TAGCATGC**CAGACGAGTTGTTACCGGTATTGGT	to clone Trim seq138 in pHStinger (1611bp)	pMC070: Trim9-seq138-hs43-nGFP
oMC133 Trim-seq138-R/Xb	**ACTCTAGA**GGCTTTTAGCAATTTAACGCACAG	to clone Trim seq138 in pHStinger (1611bp)	pMC070: Trim9-seq138-hs43-nGFP
oMC134 CG15537seq1+11-F/K	**GGGGTACC**GTCATTCTCTGCGTCTTAATTTGC	to clone CG15537 seq1+11 in pHStinger (3448bp)	pMC062: CG15537-seq1+11-hs43-nGFP
oMC135 CG15537seq1+11-R/B	**GGGGATCC**GGATTGTCATGCTCTTGTCATCTC	to clone CG15537 seq1+11 in pHStinger (3448bp)	pMC062: CG15537-seq1+11-hs43-nGFP
oMC136 Ork-seq9-F/Sp	**CCGCATGC**CAGATTGATTGATGTCTAGAGAACC	to clone Ork1 seq9 in pHStinger (1053bp)	pMC055: Ork-seq9-hs43-nGFP
oMC137 Ork-seq9-R/SII	**GTCCGCGG**TTCGCAAAGACTTGTCGCAAATATA	to clone Ork1 seq9 in pHStinger (1053bp)	pMC055: Ork-seq9-hs43-nGFP
oMC140 Cabeta-seq55-F/Sp	**AAGCATGC**GTCTTGTCGCTAATTTCATAGGAGT	to clone Cabeta seq55 in pHStinger (1132bp)	pMC063: CaBeta-seq55-hs43-nGFP
oMC141 Cabeta-seq55-R/Sg	**CTCGCCGGCG**AAGACGTGGTTCTATAATTCAATTTG	to clone Cabeta seq55 in pHStinger (1132bp)	pMC063: CaBeta-seq55-hs43-nGFP
oMC142 Cabeta-seq46-F/Nh	**TCGCTAGC**GCACTAAAGGTTCTTGAAAAGCTGC	to clone Cabeta seq46 in pHStinger (1253bp)	pMC059: CaBeta-seq46-hs43-nGFP
oMC143 Cabeta-seq46-R/B	**CCGGATCC**CGTGGCATTTTATGTGTCATCAAC	to clone Cabeta seq46 in pHStinger (1253bp)	pMC059: CaBeta-seq46-hs43-nGFP

Table 2: Reporter Constructs and Expression Patterns

FIG	CONSTRUCT	ALL LINES	EMBRYOS HRP	EMBRYOS FISH	NOTES
22	lh PCR1-hs43-lacZ	1, 2, 3, 4, 5, 6	amnioserosa	amnioserosa	beside seq34
22	lh PCR2-hs43-lacZ	7, 8, 9, 10	lymph glands?		
22	lh PCR3-hs43-lacZ	11, 12, 13, 14	rectum		
22	lh PCR4-hs43-lacZ	15, 16, 17	brain?		
22	lh PCR5-hs43-lacZ	18, 19, 20, 21	ALL: amnioserosa	amnioserosa	includes seq48
17	pMC008: Ndae1-inI-hs43-lacZ	F11y, F15red, F16or, M8red, F16dk or	CNS only in MIX		
NS	pMC013: Ndae1-5'XB-hs43-lacZ	F11dk or, F11red, F32or, F32red	nothing		most 5' fragment of Ndae1
NS	pMC020: Ndae1-3'-hs43-lacZ	F20, F20jumps #48 #51 #57	nothing		
17	pMC024: Ndae1-inVII-hs43-lacZ	M2, M3, M4, F5	haemocytes	haemocytes	
17	pMC025: Ndae1-Hunch-hs43-lacZ	M1, M3, M6, M8	anals pads		
NS	pMC026: Ndae-inII-hs43-lacZ	F1, M6, M7, M8	nothing		
See pMC 025	pMC027: Ndae 5'UTR4.6-hs43-lacZ	F2, F2jumps: #10red #13 #20dkor	weak anal pads in late embryos		
17	pMC028: Ndae inIII-hs43-lacZ	F4, F11, M1, M9	amnioserosa+ lateral cells (PNS?)		
See pMC 025	pMC029: Ndae 5'UTR4.4-hs43-lacZ	F3, M3, M8, M12	anal plate		
See pMC 028	pMC035: Ndae-InIII-hs43-nGFP	F4, F14, F55, F57, M8, M20, M30	amnioserosa+ lateral cells (PNS?)		
See pMC 028	pMC036: Ndae InIII 3'half-hs43-nGFP	F2, F4, F9, F17, F30, M7, M22	amnioserosa+ lateral cells (PNS?) (stronger than 035 and 038)		
See pMC 028	pMC038: Ndae inIII+env-hs43-nGFP	F3, F9, F16, F31, F49, F52, F58, F61, M6, M21	lateral cells (PNS?)		
NS	pMC046: Dms-seq5-nGFP	M1, F2, F3, F4	nothing		with hs43 P: heart
NS	pMC049: CG15537-seq14-nGFP	F10, M1, M2, M3, M4, M8	nothing		with hs43 P: heart
24	pMC050: pain-seq14+15-hs43-nGFP	F10, M1, M2, M4, M5, M7, M8, M9	M1: head, salivary glands from stg 10, sMD M5: as 1M but weaker	M9, M4	
24	pMC051: lh-seq34-nGFP	M2m1, F6, F5	M2m1, F6, F5: nothing		
24	pMC052: pain-seq42-hs43-nGFP	M1, M2, M3, M4, F5, M6	M6: head, oesophagus, VM longitudinal, midline F5 M2: stronger than M6 + CNS	M2, M4	
24 & 25	pMC055: Ork-seq9-hs43-nGFP	F25, F7, M9or m, M11	F25, F7, M9orm, M11: CNS	M9orm, M11: CNS heart	
24	pMC062: CG15537-seq1+11-hs43-nGFP	F22y, F22r, M3, M5, M27, M30	F22y, F22r, M3, M27, M30: CNS endoderm VM lateral staining	M30, F22: CNS endoderm VM lateral staining	

FIG	CONSTRUCT	ALL LINES	EMBRYOS HRP	EMBRYOS FISH	NOTES
24	pMC066: Ih-seq34-hs43-nGFP	F1y F1y jumps: #1 #6 #13 #15 #28 #30 #36 #39 #40	F1y: Malpighian tubules?	F1#1+15+30 F1#36+39+40	
24	pMC067: CaBeta-seq55 + seq 46-hs43-nGFP	M7 M7 jumps: #1 #13 #15 #27 #34 #39 #46	#46+39+34+27+11: epidermis?	067 #46+39+34+27+11	
NS	pMC069: CG10440-seq80 +89-hs43-nGFP	F9, F10, M1, M2, M3, M4, M5, M6, M7, M8	069 M1+M2+M3 : nothing 069 M4+M5+M6 : idem	M1 + M2+ M3: nothing M4 +M5 +M6: nothing	
24	pMC071:Trim-seq129 + seq 138 -hs43-nGFP	F4, F5, F6, F7, M1, M2, M3, M4	M1+M2+M3: CNS after stg14 F4 +F5+F6: idem	M1+M2 +M3: CNS after stg14 F4 +F5+F6: CNS after stg14	
24	pMC073: betaTub97-seq58 + seq71-hs43-nGFP	M1, M2, M3, M4, F5	M1+M4: thoracic somatic mesoderm after stg13 M2+M3: idem	M1 +M4: thoracic sMD after stg13 M2 +M3: idem	
24 & 25	pMC077: Ih-seq34-hs43-nGFP	F9, F10, M1, M2, M3, M4, M5, M6, M7, M8,	M1+M2+M3: nothing M4+M5+M6: idem M8+F9+F10: idem	M2: weak heart F10: weak heart	
24 & 25	pMC078: Dms-seq5-hs43-nGFP	F10, M1, M2, M3, M4, M5, M6, M7, M8,M9	M1+M2+M3+ M4: nothing M5+M6+M7: idem M8+M9+F10:idem	M1: weak heart F10: heart	

BIBLIOGRAPHY

Bibliography

- Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F, Tranchevent LC, De Moor B, Marynen P, Hassan B, Carmeliet P, Moreau Y. Gene prioritization through genomic data fusion. Nature Biotechnol. 2006 May;24(5):537-44. Erratum in: Nat Biotechnol. 2006 Jun;24(6):719.
- Aerts S, van Helden J, Sand O, Hassan BA. Fine-Tuning Enhancer Models to Predict Transcriptional Targets across Multiple Genomes.PLoS ONE. 2007 Nov 7;2(11):e1115.
- Akam M. Hox genes: from master genes to micromanagers. Curr Biol. 1998 Sep 24;8(19):R676-8.
- Akasaka T, Klinedinst S, Ocorr K, Bustamante EL, Kim SK, Bodmer R. The ATP-sensitive potassium (KATP) channel-encoded dSUR gene is required for Drosophila heart function and is regulated by tinman. Proc Natl Acad USA. 2006 Aug 8;103(32):11999-2004. Epub 2006 Aug 1.
- Azpiazu N, Frasch M. tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila. Genes Dev. 1993 Jul;7(7B):1325-40.
- Barolo S, Castro B, Posakony JW. New Drosophila transgenic reporters: insulated P-element vectors expressing fast-maturing RFP. Biotechniques. 2004 Mar;36(3):436-40, 442.
- Bergman CM, Pfeiffer BD, Rincón-Limas DE, Hoskins RA, Gnirke A, Mungall CJ, Wang AM, Kronmiller B, Pacleb J, Park S, Stapleton M, Wan K, George RA, de Jong PJ, Botas J, Rubin GM, Celniker SE. Genome Biol. 2002;3(12):RESEARCH0086. Epub 2002 Dec 30.
- Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, Rubin GM, Eisen MB. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. Proc Natl Acad USA. 2002 Jan 22;99(2):757-62.
- Berman BP, Pfeiffer BD, Laverty TR, Salzberg SL, Rubin GM, Eisen MB, Celniker SE. Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in Drosophila melanogaster and Drosophila pseudoobscura. Genom Biol. 2004;5(9):R61. Epub 2004 Aug 20.
- Bodmer R. The gene tinman is required for specification of the heart and visceral muscles in Drosophila. Development. 1993 Jul;118(3):719-29. Erratum in: Development 1994 Nov;119(3):969.
- Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 1993 Jun;118(2):401-15.
- Brodu V, Elstob PR, Gould AP. abdominal A specifies one cell type in Drosophila by regulating one principal target gene. Development. 2002 Jun;129(12):2957-63.
- Campos-Ortega, J. A. and Hartenstein, V. The Embryonic development of Drosophila melanogaster Second Edition. Springer-Verlag, Berlin. 1997.
- Capovilla M, Botas J. Functional dominance among Hox genes: repression dominates activation in the regulation of Dpp. Development. 1998 Dec;125(24):4949-57.
- Capovilla M, Brandt M, Botas J. Direct regulation of decapentaplegic by Ultrabithorax and its role in Drosophila midgut morphogenesis. Cell. 1994 Feb 11;76(3):461-75
- Capovilla M, Kambris Z, Botas J. Direct regulation of the muscle-identity gene apterous by a Hox protein in the somatic mesoderm. Development. 2001 Apr;128(8):1221-30.
- Carroll, S.B. Homeotic genes and the evolution of arthropods and chordate. Nature. 1995 376:479-485.
- Carroll. S.B. Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom. W.W. Norton, New York. 2005.

- Chan SK, Jaffe L, Capovilla M, Botas J, Mann RS. The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell. 1994 Aug 26;78(4):603-15.
- Chartier A, Zaffran S, Astier M, Sémériva M, Gratecos D. Pericardin, a Drosophila type IV collagen-like protein is involved in the morphogenesis and maintenance of the heart epithelium during dorsal ectoderm closure. Development. 2002 Jul;129(13):3241-53.
- Duffy JB. GAL4 system in Drosophila: a fly geneticist's Swiss army knife.Genesis. 2002 Sep-Oct;34(1-2):1-15.
- Durocher D, Charron F, Warren R, Schwartz RJ, Nemer M. The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. EMBO J. 1997 Sep 15;16(18):5687-96.
- Ekker SC, von Kessler DP, Beachy PA. Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. EMBO J. 1992 Nov;11(11):4059-72.
- Ekker SC, Young KE, von Kessler DP, Beachy PA. Optimal DNA sequence recognition by the Ultrabithorax homeodomain of Drosophila. EMBO J. 1991 May;10(5):1179-86.
- Elgar SJ, Han J, Taylor MV. mef2 activity levels differentially affect gene expression during Drosophila muscle development. Proc. Natl. Acad. Sci U S A. 2008 Jan 22;105(3):918-23. Epub 2008 Jan 15.
- Fischer JA, Giniger E, Maniatis T, Ptashne M. GAL4 activates transcription in Drosophila. Nature. 1988 Apr 28;332(6167):853-6.
- Forouhar AS, Liebling M, Hickerson A, Nasiraei-Moghaddam A, Tsai HJ, Hove JR, Fraser SE, Dickinson ME, Gharib M. The embryonic vertebrate heart tube is a dynamic suction pump. Science. 2006 May 5;312(5774):751-3.
- Frasch M. Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early Drosophila embryo. Nature. 1995 Mar 30;374(6521):464-7.
- Frasch M. Intersecting signalling and transcriptional pathways in Drosophila heart specification. Semin Dev Cell Biol. 1999 Feb;10(1):61-71.
- Frémion F, Astier M, Zaffran S, Guillèn A, Homburger V, Sémériva M. The heterotrimeric protein Go is required for the formation of heart epithelium in Drosophila. Cell Biol. 1999 May 31;145(5):1063-76.
- Gajewski K, Fossett N, Molkentin JD, Schulz RA. The zinc finger proteins Pannier and GATA4 function as cardiogenic factors in Drosophila. Development. 1999 Dec;126(24):5679-88.
- Gajewski K, Kim Y, Choi CY, Schulz RA. Combinatorial control of Drosophila mef2 gene expression in cardiac and somatic muscle cell lineages. Dev. Genes Evol. 1998 Sep;208(7):382-92.
- Gajewski K, Kim Y, Lee YM, Olson EN, Schulz RA. D-mef2 is a target for Tinman activation during Drosophila heart development. EMBO J. 1997 Feb 3;16(3):515-22.
- Gajewski K, Zhang Q, Choi CY, Fossett N, Dang A, Kim YH, Kim Y, Schulz RA. Pannier is a transcriptional target and partner of Tinman during Drosophila cardiogenesis. Dev Biol. 2001 May 15;233(2):425-36.
- Galant R, Walsh CM, Carroll SB. Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. Development. 2002 Jul;129(13):3115-26.
- Garcia Bellido A. Homeotic and atavic mutation in insects. Am. Zool. **17**:613-629. 1977.
- Gebelein B, McKay DJ, Mann RS. Direct integration of Hox and segmentation gene inputs during Drosophila development. Nature. 2004 Oct 7;431(7009):653-9.
- Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, Affolter M, Otting G, Wüthrich K.Homeodomain-DNA recognition. Cell. 1994 Jul 29;78(2):211-23.

- Ghosh TK, Packham EA, Bonser AJ, Robinson TE, Cross SJ, Brook JD. Characterization of the TBX5 binding site and analysis of mutations that cause Holt-Oram syndrome. Hum Mol Genet. 2001 Sep 1;10(18):1983-94.
- Grad YH, Roth FP, Halfon MS, Church GM. Prediction of similarly acting cis-regulatory modules by subsequence profiling and comparative genomics in Drosophila melanogaster and D.pseudoobscura. Bioinformatics. 2004 Nov 1;20(16):2738-50. Epub 2004 May 14.
- Haerry TE, Gehring WJ. A conserved cluster of homeodomain binding sites in the mouse Hoxa-4 intron functions in Drosophila embryos as an enhancer that is directly regulated by Ultrabithorax. Dev Biol. 1997 Jun 1;186(1):1-15.
- Halfon MS, Grad Y, Church GM, Michelson AM. Computation-based discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. Genom Res. 2002 Jul;12(7):1019-28.
- Han Z, Olson EN. Hand is a direct target of Tinman and GATA factors during Drosophila cardiogenesis and hematopoiesis. Development. 2005 Aug;132(15):3525-36. Epub 2005 Jun 23.
- Han Z, Yi P, Li X, Olson EN. Hand, an evolutionarily conserved bHLH transcription factor required for Drosophila cardiogenesis and hematopoiesis. Development. 2006 Mar;133(6):1175-82. Epub 2006 Feb 8.
- Hendren JD, Shah AP, Arguelles AM, Cripps RM. Cardiac expression of the Drosophila Sulphonylurea receptor gene is regulated by an intron enhancer dependent upon the NK homeodomain factor Tinman. Mech. Dev. 2007 Jul;124(6):416-26. Epub 2007 Mar 12.
- Hertz GZ, Stormo GD. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics. 1999 Jul-Aug;15(7-8):563-77.
- Hombría JC, Lovegrove B. Beyond homeosis--HOX function in morphogenesis and organogenesis. Differentiation. 2003 Oct;71(8):461-76.
- Iimura T, Pourquié O. Hox genes in time and space during vertebrate body formation.Dev Growth Differ. 2007 May;49(4):265-75.
- Jagla K, Frasch M, Jagla T, Dretzen G, Bellard F, Bellard M. ladybird, a new component of the cardiogenic pathway in Drosophila required for diversification of heart precursors. Development. 1997 Sep;124(18):3471-9
- Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, Crickmore MA, Jacob V, Aggarwal AK, Honig B, Mann RS. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell. 2007 Nov 2;131(3):530-43.
- Kania A, Salzberg A, Bhat M, D'Evelyn D, He Y, Kiss I, Bellen HJ. P-element mutations affecting embryonic peripheral nervous system development in Drosophila melanogaster. Genetics. 1995 Apr;139(4):1663-78.
- Klinedinst SL, Bodmer R. Gata factor Pannier is required to establish competence for heart progenitor formation. Development. 2003 Jul;130(13):3027-38.
- Kölsch V, Paululat A. The highly conserved cardiogenic bHLH factor Hand is specifically expressed in circular visceral muscle progenitor cells and in all cell types of the dorsal vessel during Drosophila embryogenesis.Dev Genes Evol. 2002 Nov;212(10):473-85. Epub 2002 Sep
- Lalevée N, Monier B, Sénatore S, Perrin L, Sémériva M. Control of cardiac rhythm by ORK1, a Drosophila two-pore domain potassium channel. Curr Biol. 2006 Aug 8;16(15):1502-8.
- Lawrence PA, Morata G. Homeobox genes: their function in Drosophila segmentation and pattern formation. Cell. 1994 Jul 29;78(2):181-9.
- Lee HH, Frasch M. Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target sloppy paired. Development. 2000 Dec;127(24):5497-508.
- Lewis EB. A gene complex controlling segmentation in Drosophila. Nature. 1978 Dec 7;276(5688):565-70.

- Lewis EB. The bithorax complex: the first fifty years. Int J Dev Biol. 1998;42(3):403-15.
- Lilly B, Galewsky S, Firulli AB, Schulz RA, Olson EN. D-MEF2: a MADS box transcription factor expressed in differentiating mesoderm and muscle cell lineages during Drosophila embryogenesis. Proc Natl Acad Sci U S A. 1994 Jun 7;91(12):5662-6.
- Lilly B, Zhao B, Ranganayakulu G, Paterson BM, Schulz RA, Olson EN. Requirement of MADS domain transcription factor D-MEF2 for muscle formation in Drosophila. Science. 1995 Feb 3;267(5198):688-93.
- Lo PC, Frasch M. A role for the COUP-TF-related gene seven-up in the diversification of cardioblast identities in the dorsal vessel of Drosophila. Mech Dev. 2001 Jun;104(1-2):49-60
- Lo PC, Frasch M. Establishing A-P polarity in the embryonic heart tube: a conserved function of Hox genes in Drosophila and vertebrates? Trends Cardiovasc Med. 2003 Jul;13(5):182-7.
- Lo PC, Skeath JB, Gajewski K, Schulz RA, Frasch M. Homeotic genes autonomously specify the anteroposterior subdivision of the Drosophila dorsal vessel into aorta and heart. Dev. Biol. 2002 Nov 15;251(2):307-19.
- Lockwood WK, Bodmer R. The patterns of wingless, decapentaplegic, and tinman position the Drosophila heart. Mech Dev. 2002 Jun;114(1-2):13-26.
- Loots, G., Ovcharenko,I., Pachter,L., Dubchak,I., Rubin, E. rVISTA for comparative sequence-based discovery of functional transcription factor binding sites. Genome. Res. 2002 12:832-839
- Lovato TL, Nguyen TP, Molina MR, Cripps RM. The Hox gene abdominal-A specifies heart cell fate in the Drosophila dorsal vessel. Development. 2002 Nov; 129(21):5019-27.
- Maeda RK, Karch F. The ABC of the BX-C: the bithorax complex explained. Development. 2006 Apr;133(8):1413-22.
- Mahaffey JW. Assisting Hox proteins in controlling body form: are there new lessons from flies (and mammals)? Curr Opin Genet Dev. 2005 Aug;15(4):422-9.
- Mann RS, Affolter M. Hox proteins meet more partners. Curr Opin Genet Dev. 1998 Aug;8(4):423-9. Review.
- Mann RS. The specificity of homeotic gene function. Bioessays. 1995 Oct;17(10):855-63. Review.
- Markstein, M., Markstein, P., Markstein, V., and Levine, M.S. Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. Proc. Natl. Acad. Sci. USA. 2002 99, 763-768.
- McGinnis W. A century of homeosis, a decade of homeoboxes. Genetics. 1994 Jul;137(3):607-11. Review.
- Merabet S, Kambris Z, Capovilla M, Bérenger H, Pradel J, Graba Y. The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. Dev Cell. 2003 May;4(5):761-8.
- Michelson AM. Muscle pattern diversification in Drosophila is determined by the autonomous function of homeotic genes in the embryonic mesoderm. Development. 1994 Apr;120(4):755-68.
- Miskolczi-McCallum CM, Scavetta RJ, Svendsen PC, Soanes KH, Brook WJ. The Drosophila melanogaster T-box genes midline and H15 are conserved regulators of heart development. Dev Biol. 2005 Feb 15;278(2):459-72.
- Moens CB, Selleri L. Hox cofactors in vertebrate development. Dev Biol. 2006 Mar 15;291(2):193-206. Epub 2006 Mar 3.
- Molina MR, Cripps RM. Ostia, the inflow tracts of the Drosophila heart, develop from a genetically distinct subset of cardial cells. Mech Dev. 2001 Nov;109(1):51-9.
- Monier B, Astier M, Sémériva M, Perrin L.Steroid-dependent modification of Hox function drives myocyte reprogramming in the Drosophila heart. Development. 2005 Dec;132(23):5283-93.

- Monier, B., Tevy, M.F., Perrin, L., Capovilla, M., and Semeriva, M. Downstream of Homeotic genes: in the heart of Hox function. Fly. 2007 1, 59-67
- Nguyen HT, Bodmer R, Abmayr SM, McDermott JC, Spoerel NA. D-mef2: a Drosophila mesoderm-specific MADS box-containing gene with a biphasic expression profile during embryogenesis. Proc Natl Acad USA. 1994 Aug 2;91(16):7520-4.
- Ocorr K, Perrin L, Lim HY, Qian L, Wu X, Bodmer R. Genetic control of heart function and aging in Drosophila. Trends Cardiovasc Med. 2007 Jul;17(5):177-82. Review.
- Olson EN. Gene regulatory networks in the evolution and development of the heart. Science. 2006 Sep 29;313(5795):1922-7.
- Osada R, Zaslavsky E, Singh M. Comparative analysis of methods for representing and searching for transcription factor binding sites. Bioinformatics. 2004 Dec 12;20(18):3516-25. Epub 2004 Aug 5.
- Park M, Lewis C, Turbay D, Chung A, Chen JN, Evans S, Breitbart RE, Fishman MC, Izumo S, Bodmer R. Differential rescue of visceral and cardiac defects in Drosophila by vertebrate tinman-related genes.Proc Natl Acad Sci U S A. 1998 Aug 4;95(16):9366-71. Erratum in: Proc Natl Acad Sci U S A 1998 Oct 27;95(22):13348.
- Pearson JC, Lemons D, McGinnis W. Modulating Hox gene functions during animal body patterning. Nat Rev Genet. 2005 Dec;6(12):893-904.
- Pennacchio LA, Rubin EM. Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet 2001 Feb;2(2):100-9.
- Perrin L, Monier B, Ponzielli R, Astier M, Semeriva M. Drosophila cardiac tube organogenesis requires multiple phases of Hox activity. Dev Biol. 2004 Aug 15;272(2):419-31.
- Pinsonneault J, Florence B, Vaessin H, McGinnis W. A model for extradenticle function as a switch that changes HOX proteins from repressors to activators. EMBO J. 1997 Apr 15;16(8):2032-42.
- Ponzielli R, Astier M, Chartier A, Gallet A, Thérond P, Sémériva M. Heart tube patterning in Drosophila requires integration of axial and segmental information provided by the Bithorax Complex genes and hedgehog signaling. Development. 2002 Oct;129(19):4509-21.
- Potthoff MJ, Olson EN. MEF2: a central regulator of diverse developmental programs. Development. 2007 Dec;134(23):4131-40. Epub 2007 Oct 24.
- Qian L, Liu J, Bodmer R. Tbx20-related genes (H15/midline) promote cell fate specification and morphogenesis of the Drosophila heart. Dev Biol. 2005 Mar 15;279(2):509-24.
- Ramain P, Heitzler P, Haenlin M, Simpson P. pannier, a negative regulator of achaete and scute in Drosophila, encodes a zinc finger protein with homology to the vertebrate transcription factor GATA-1. Development. 1993 Dec;119(4):1277-91.
- Ranganayakulu G, Elliott DA, Harvey RP, Olson EN. Divergent roles for NK-2 class homeobox genes in cardiogenesis in flies and mice. Development. 1998 Aug;125(16):3037-48
- Reim I, Frasch M. The Dorsocross T-box genes are key components of the regulatory network controlling early cardiogenesis in Drosophila. Development. 2005 Nov;132(22):4911-25. Epub 2005 Oct 12.
- Reim I, Lee HH, Frasch M. The T-box-encoding Dorsocross genes function in amnioserosa development and the patterning of the dorsolateral germ band downstream of Dpp. Development. 2003 Jul;130(14):3187-204.
- Reim I, Mohler JP, Frasch M. Tbx20-related genes, mid and H15, are required for tinman expression, proper patterning, and normal differentiation of cardioblasts in Drosophila.Mech Dev. 2005 Sep;122(9):1056-69.
- Ritzki T.M. in The Genetics and Biology of Drosophila (eds Ashburner, M. &. Wright, T.R.F), Academic, London, UK, 1978..

- Romero MF, Henry D, Nelson S, Harte PJ, Dillon AK, Sciortino CM. Cloning and characterization of a Na+-driven anion exchanger (NDAE1). A new bicarbonate transporter. J Biol Chem. 2000 Aug 11;275(32):24552-9.
- Ruiz-Gómez M. Muscle patterning and specification in Drosophila. Int J Dev Biol. 1998;42(3):283-90. Review.
- Ryoo HD, Marty T, Casares F, Affolter M, Mann RS. Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. Development. 1999 Nov;126(22):5137-48.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 1989.
- Schier AF, Gehring WJ. Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. Nature. 1992 Apr 30;356(6372):804-7.
- Sciortino CM, Shrode LD, Fletcher BR, Harte PJ, Romero MF. Localization of endogenous and recombinant Na(+)-driven anion exchanger protein NDAE1 from Drosophila melanogaster. Am J Physil Cell Physiol. 2001 Aug;281(2):C449-63.
- Stennard FA, Costa MW, Elliott DA, Rankin S, Haast SJ, Lai D, McDonald LP, Niederreither K, Dolle P, Bruneau BG, Zorn AM, Harvey RP. Cardiac T-box factor Tbx20 directly interacts with Nkx2-5, GATA4, and GATA5 in regulation of gene expression in the developing heart. Dev Biol. 2003 Oct 15;262(2):206-24.
- Stormo GD. DNA binding sites: representation and discovery. Bioinformatics. 2000 Jan;16(1):16-23. Review.
- Treisman J, Harris E, Wilson D, Desplan C. The homeodomain: a new face for the helix-turn-helix? Bioessays. 1992 Mar;14(3):145-50. Review.
- Vachon G, Cohen B, Pfeifle C, McGuffin ME, Botas J, Cohen SM. Homeotic genes of the Bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less. Cell. 1992 Oct 30;71(3):437-50.
- Walsh CM, Carroll SB. Collaboration between Smads and a Hox protein in target gene repression. Development. 2007 Oct;134(20):3585-92. Epub 2007 Sep 12.
- Weatherbee SD, Halder G, Kim J, Hudson A, Carroll S. Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the Drosophila haltere. Genes Dev.1998 May 15;12(10):1474-82.
- Wessells RJ, Fitzgerald E, Cypser JR, Tatar M, Bodmer R. Insulin regulation of heart function in aging fruit flies. Nat Genet. 2004 Dec;36(12):1275-81. Epub 2004 Nov 21.
- Winick J, Abel T, Leonard MW, Michelson AM, Chardon-Loriaux I, Holmgren RA, Maniatis T, Engel JD. A GATA family transcription factor is expressed along the embryonic dorsoventral axis in Drosophila melanogaster. Development. 1993 Dec;119(4):1055-65.
- Yamagishi H, Yamagishi C, Nakagawa O, Harvey RP, Olson EN, Srivastava D.The combinatorial activities of Nkx2.5 and dHAND are essential for cardiac ventricle formation. Dev Biol. 2001 Nov 15;239(2):190-203.
- Yin Z, Frasch M. Regulation and function of tinman during dorsal mesoderm induction and heart specification in Drosophila. Dev Genet. 1998 ;22(3):187-200.
- Zaffran S, Astier M, Gratecos D, Sémériva M.The held out wings (how) Drosophila gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. Development. 1997 May;124(10):2087-98.
- Zaffran S, Frasch M. Early signals in cardiac development. Circ Res. 2002 Sep 20;91(6):457-69. Review.
- Zaffran S, Reim I, Qian L, Lo PC, Bodmer R, Frasch M. Cardioblast-intrinsic Tinman activity controls proper diversification and differentiation of myocardial cells in Drosophila. Development. 2006 Oct;133(20):4073-83. Epub 2006 Sep 20.

- Zhang Y, Rath N, Hannenhalli S, Wang Z, Cappola T, Kimura S, Atochina-Vasserman E, Lu MM, Beers MF, Morrisey EE. GATA and Nkx factors synergistically regulate tissue-specific gene expression and development in vivo. Development. 2007 Jan;134(1):189-98.

APPENDIX

Solutions

- **BBS-250:** 1x BBT, 250mM NaCl, 0,1% bovine serum
- **BBT-250:** 1x BBT, 250mM NaCl
- **BBT:** 1x PBS, 1% BSA, 0,05% Tween 20
- **Buffer 10´PBS:** 80gr of NaCl, 11,5gr of Na_2HPO_4, 2gr of KH_2PO_4, 2gr of KCl, dH_2O qsp 100ml. Adjust pH to 7,3-7,5 with HCl 10M. Autoclave.
- **Buffer 1xPBT:** 100ml of 10xPBS (final concentration 1x), 1ml of Tween-20 (final (concentration final of 0,1%), dH_2O qsp 1000ml.
- **Buffer 20xSSC:** 175,3gr NaCl, 88,2gr NaCitrate, dH_2O qsp 1000ml. Adjust pH to 7,0. with HCl 10M. First sterilize and then autoclave.
- **Fix buffer:** 25ml of H_2O, 5ml of 10´PBS, 100ml of EGTA 0,5M, 100ml $MGSO_4$ 1M
- **Fly food media :** 170 ml of H_2O, 95 gr of agar, 940gr of sugar, 250gr yeast, 250ml corn flour.
- **Heparine:** stock solution 50 mg/ml in buffer 4xSSC.
- **Hybe A (prehybridization solution):** Final volume: 50 ml. 25 ml 100% formamide, 12,5 ml 20x SSC, 12,5 ml of DEPC water
- **Hybe B (hybridization solution):** Final volume: 100 ml . 50 ml 100% formamide, 25 ml 20x SSC, 2 ml of 10 mg/ml Herring Sperm DNA, 500 ml 20 mg/ ml tRNA, 50 ml 100 mg/ml heparine, DEPC water up to 100 ml. Store at -20°C
- **Molasse agar medium:** mix 95ml mollase, 20g of agar 500ml of water. Autoclave, then transfer to water bath until the temperature goes down to 65°C. Add 16ml of ethanol 100 % and 8ml of acetic acid and stir. Pour the media and stock at 4°C.
- **PBT:** 1x PBS, 0,1% Tween-20
- **PTX:** 1x PBS, 0,1% Triton-X-100

INDEX

Index

TABLES

BIBLIOGRAPHY

APPENDIX

ACKNOWLEDGEMENTS

Acknowledgements

"El orden de los factores no altera el producto"

Per cominciare, devo dire che, durante cuatro anios ha pasado mucha gente y mucha agua bajo el puente, so I am going to write a little bit like this and a little bit like that. E con questo voglio dire scusate por los horrores ortograficos!!!

PS/ los acentos son eccentos, gli acenti sonno assenti, e le doppie sonno a la mia discrezione.

Empecemos por casa,

Gracias a toda mi flia que esta siempre y en todo momento presente pese al Atlantico que nos separa. A la Tata que no se olvido nunca en cuatro de llamarme todos los jueves, a Clara por sus eternas conversaciones telefonicas y a Jose por preguntar siempre "y como va?", a mama por escucharme aun cuando eran las seis de la maniana. Y no enumero a todos porque necesitaria un rollo para escribir, pero sepan que fueron y son mi apoyo.

Passiamo al lab,
Guau!! Ce troppa gente!! Vorrei ringraziare prima di tutto la cheff , ma non solo per avermi insegnato tutto quello che so dei moscherini, ma soprattutto per esserci stata anche quando sembrava che non ce l' avremmo mai fatta, anche quando ero a punto di buttare l'asciugamano. Non e finita ancora, ma vedo piu luce in fondo al tunnel.

Tutti quanti che siete passati per il lab dovete sapere che avete lasciato traccia. A Nadia, per avermi insegnato a insegnare. A Manlio per suo continuo sostegno e amicizia per tutti questi anni, ma ho qualcosa in particolare da dirti: VOGLIO ESSERE LA MADRINA DE G!!!

Per le flygirls attuali, a tutte quante grazie perche m' insegnate, mi ascoltate, e mi accompagnate

Grazie a tutti quanti dei labs di Bologna, particolarmente a tutti quelli della Genetica medica con cui ho condiviso il mio primo periodo in Italia, con qualche birretta di qua e di la.

Je ne peux pas ecrit un francais!!!! Mmmm… peut etre je peux aller a MRS pour etudie francais… alle Calanque…. Thanks to Michel and Laurent for teaching me so much, to Z and Naty for so many FISH and always hosting me and being friends!

Grazie a Diego D. per le mille sessions di microscopia…. e la pazienza!!!

E devo dire anche che non ce l' avrei mai fatta a Ferrara senza Sara, Nicola, Oreste e la flia Taddei!!!! GRAZIE per il sostegno e l' amicizia!!! Anche dei pranzi, delle cene, per Natale, per Pasqua e tanti concerti! Vi devo dire grazie aver reso questa nebbia piu tollllerabile!! A Nico per avermi presentato a tutti questi personaggi. E Ore…. Grazie per tuttte le immmagini!!! Devo dire che tante immmagini di questa tesi sonnno di Bossa productions!

Manu… come ti ringrazio per la serena convivenza e le mille ore di ascolto?…e che ti devo dire… andiamo al Giglio??? GRAZIE!!!!

Fra e Davide, grazie per esserci dentro e fuori del lab, andiamo a cavallo o andiamo a Bologna??!!!

Jules H, thank you for being always. In the good, in the bad, in Bologna, Cartocceto o Ferrara!!

For sure I am missing a million people, but … I didn't do it on purpose.

For all of those who wonder what am I going to do next… the answer is simple: I DON'T HAVE A CLUE!!! MA MI PORTO I MOSCERINI!!!
E stato un piacere abitare in Italia!! Non mi pento!!! Comunque ragazzi… dovete venire in Argentina!!