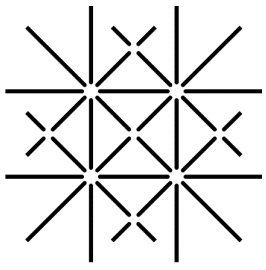
Dissection of *cis*-regulatory elements in the *apterous* gene of *Drosophila melanogaster*

Master thesis

Gordian Born
Basel, January 20, 2017

Submitted to:
Prof. Markus Affolter
Dr. Dimitri Bieli

Biozentrum
University of Basel



Abstract

In this project I examined the *ap* early enhancer of *Drosophila melanogaster* in a combined approach with bioinformatic methods, deletion mutations and a RNAi screen. *ap* is a selector gene involved in the wing development, where it causes the compartmentalization of the dorsal and ventral part of the metathoracic disc. Without this segmentation no wing is formed, however, very little is know about the regulation of *ap*. EGFR, *pointed*, *vein*, *homothorax*, *longitudinals lacking* and *ventral veins lacking* were found to have a positive input in *ap* regulation, whereas *armadillo* was found to have a negative input in *ap* regulation.

Contents

List of Figures

1 Introduction

1.1 *Drosophila melanogaster*

The life of most multicellular organism starts by the fertilization of a female haploid gametocyte by a male haploid gametocyte. Independent of how the life started at some time in the beginning there is a diploid unicellular zygote, which proliferates and differentiates into different tissues with distinct purposes. To fulfill the development from a single cell to a complex three-dimensional organism with different tissues, every cell needs a standard input of at least two information. The first one is the place it is in relation to the other cells, the second one is some sort of time measurement or information about the state the surrounding cells are in [1]. Based on this two information, every cell needs to know exactly at what time which genes must be transcribed and in which amount.
Without this complex interplay no multicellular organism could build different tissues in the correct place, to the right time with the fitting size. Thus it is important to understand the correlation between the formation of organs and gene activity. Model organisms are a useful tool for this research. They usually have a short generation time, are economic in keeping, easy to breed and minimize ethical concerns. In this project, I used the common fruit fly *Drosophila melanogaster* as a multicellular eukaryotic model organism. *Drosophila melanogaster* has been used for more than 115 year as a model organism. During this time, an excellent genetic toolbox and a huge knowledge about the developmental processes in this fly has been collected.

Drosophila melanogaster has a life cycle of about 10 days at room temperature (RT, 25°C), which is divided in four distinct parts. In a fertilized egg, embryogenesis takes place. Subsequently, there are three larval stages, which are called instars. Next is the pupal stage, in which a complete metamorphosis from a larve to the fourth stage of life takes place. This last stage is the adult fly, which mates and reproduces (see Fig. 1.1a).
Except for the short life cycle, *Drosophila melanogaster* has an other advantages against most other eukaryotic model organisms. The monolayered epithelial sacs, which form the

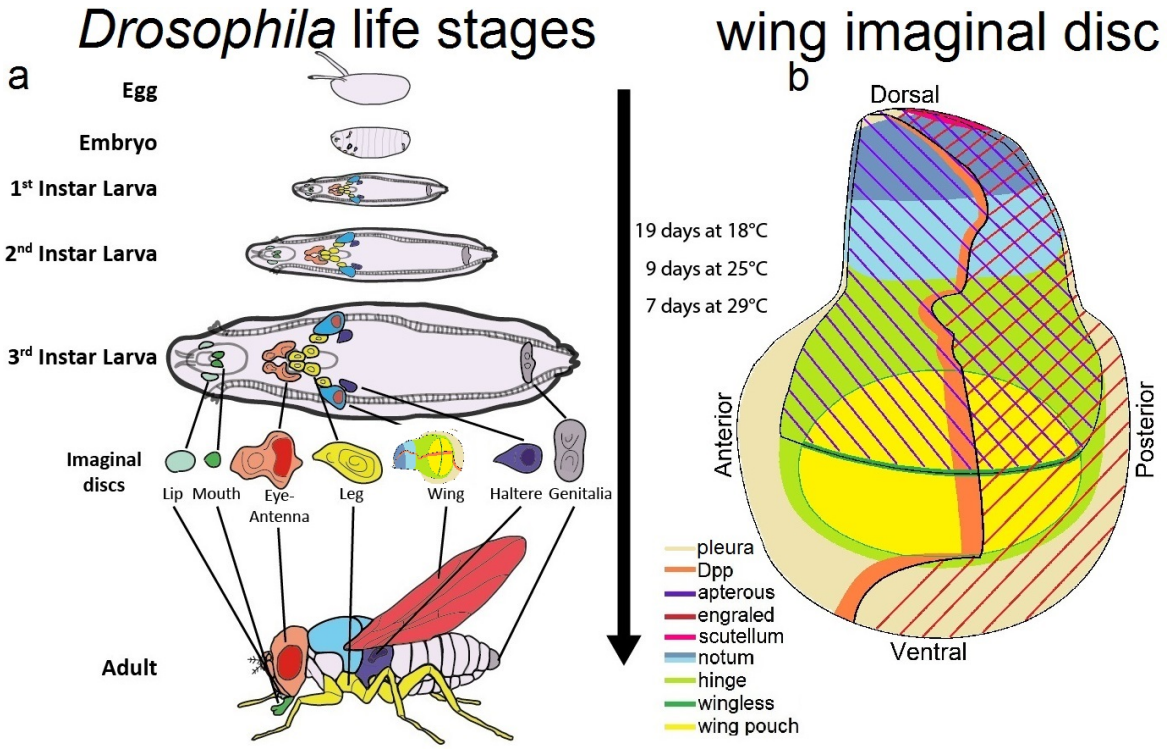


Figure 1.1: **a) *Drosophila melanogaster* at its different life stages**
It takes 24 h for the fertilized egg to end embryogenesis. The larva, which lives in the food, reaches the end of the 3rd instar after four more days then the larvae pupates out of the food and undergoes metamorphosis. After a total of 9 days the adult fly hatches, searches a partner of the other sex, mates and females lay eggs. The imaginal discs do not have a active function in the larva. In the metamorphosis in the pupal stage the larva body changes completely to the adult body. During this stage the imaginal discs will form of the adult parts. [2]
b) Wing disc of third instar larvae of *Drosophila melanogaster*
The areas of gene expression of *ap*, *Dpp*, *wg* and *en* are marked accordingly in striped purple, solid orange, solid green and striped brown. These genes are essential for the formation of the wing. The wing pouch in the center of the wing disc (yellow) folds at the border between dorsal and ventral. The light green area around the pouch will form the wing hinge, which will be between the wing blade, the notum and the pleura. The notum, the scutellum and the pleura will form the respective parts of the body wall **(a)**. To form a functional fly the imaginal discs must be precisely patterend

lips, the mouth, the eye-antennas, the legs, the wings, the halteres and the genitalia during the metamorphosis, are primordial parts of the adult body. The imaginal discs do not fulfill any vital tasks in the larva. Therefore, manipulations in imaginal discs do not affect the survival of the larva [3], but first become an issue during metamorphosis and ultimately into adult animals [4].

Moreover, the expression of certain genes in the imaginal discs of the third instar larvae and the shape of the discs indicate the rough shape and functionality of the adult counterpart [5]. In the case of the wing disc (see Fig. 1.1b) the transformation to the adult organs is well known, so the development of every part can be mapped. The wing disc is also called dorsal metathoracic disc, because it does not only form the wing blade and the wing margin, but also the notum, scutellum, wing hinge, and pleura. The wing blade and the wing margin are made from the wing pouch, which extends orthogonally to the plane of the disc. During this process, the mono-layered wing pouch folds along the dorsal-ventral compartment boundary to a two layered wing blade, while the other disc parts and other imaginal discs also form complex three-dimensional structures. Finally, the former ventral part of the wing pouch forms the rear layer of the wing blade, which is closer to the pleura. Whereas the dorsal part of the wing pouch forms the front wing layer, which is closer to the notum. So the primordial dorsal and the primordial ventral cells end up connected at their basal side in the final wing.

This precise development is archieved by compartmentalization. During development, several functional units of cells seperate from the sourrounding tissue via strictly defined compartment boundaries [6]. This is an unusual event in due consideration of the amount of cells in the animal. Genetic mosaics usually manifest themselves in erratic frayed patterns in the adult cuticle [7]. However the mosaics show a clear straight border at certain points, across which no expansion is possible [8]. At this border no special features or morphologically landmarks are discernible. Further research of the areas showed, that each of these compartment has its own properties, for example a special surface protein composition, or the secretion of a certain protein [9]. Since the specialized cells within the compartments would cause catastrophic chaos at other places in the organism, cells from

a compartment do not mingle with cells outside the sharp boundaries [9]. Short-range signaling events, which mainly appear close to the compartment boundary, mark the cells as border cells. Most of these cells at the boundary, which are termed organizers, give information to the surrounding cells by secretion of long-range signaling molecules, also called morphogens [10]. Due to the fact that the morphogen concentration decreases with the distance to its source, every cell knows its approximate distance to the compartment boundaries [11]. This distance information can be used for further compartmentalization or differential gene expression at certain thresholds.

The start of this compartmentalization process is marked by the activation of selector genes [8]. Selector genes are a group of transcription factors, that are only active within certain compartments. In most cases, several selector genes are active at the same time, to enable complex subcompartmentalization. An example is, that *en* and *ap* are among other selector genes active in the dorsal posterior part of the wing disc (see Fig. 1.1b). However, there are three more criteria to define a selector gene. Firstly, they change the adhesion properties, so a mingling of cells within and outside of the compartment is hampered. Secondly, they must, as mentioned above, provide a flow of information inside the functional unit and outside close to the compartment border via signaling events. Thirdly, their function must be cell-autonomously [12]. These principles hold true for invertebrates as well as for vertebrates such as mice [13] [14].

The segmentation of the wing disc is a classic example of compartmentalization. The wing disc is divided by two borders into four compartments. From the beginning of wing disc development, the wing disc is already divided in a posterior (p) and an anterior (a) compartment. *en* is the responsible selector gene for this first division. *en* is expressed in the whole posterior compartment. At the A/P-border on the anterior side, where cells, which do not produce *en*, are in proximity to *en* producing cells, *decapentaplegic* (*dpp* morphogen) is produced by the organizer cells [15] [16]. At the L2 stage in development *ap* starts to be expressed in the future dorsal part of the wing disc causing a second division of the wing disc into a dorsal and a ventral compartment. However, the corresponding organizer protein *wingless* (*wg*) is produced in the dorsal and in the ventral compartment at the border

[16] [17]. The exact expression pattern is shown in fig. 1.1b. *wg* and *dpp* are responsible for the correct development of the wing to wing blade [16]. Indeed, the compartments as well as the organizer protein gradient must be stable and strictly regulated, otherwise the organ development would become uncontrollable for the organism.

There are different mechanisms contributing to compartment stability. Four of the most important are listed here. Firstly, similar surface proteins usually attach stronger to each other than to unlike proteins. Here, the group of cadherins are crucial, because they are primarily responsible for the differences in the surface properties. Since the cells in every compartment have their own special cadherin protein mix, the cell adhesion within a compartment is stronger than the cell adhesion between compartments [18]. This causes a straight and smooth border to minimize the interacting surface area and therefore the energy. An analogy is the separation of water and oil, due to strong homotypic interaction and weak heterotypic interaction. Secondly, the border is maintained by a controlled growth. Tissue growth is ordinarily caused by cell proliferation, whereby the extent of cell proliferation is determining the speed of the growth. At the compartment boundary, a slow growth is needed hence the border is one dimensional. The compartment on the other hand needs to grow in two dimensions and therefore a faster growth is observed. Thus, a strong decrease in cell proliferation at the compartment boundary stabilizes the boundary shape [19]. Thirdly, actomyosin-based filaments can cause a mechanical tension, which keeps tissues together and helps sorting cells. Although an enrichment of actomyosin-based filaments are found at the two borders in the imaginal wing disc, this mechanism is only known in *Drosophilas* parasegment boundaries of the embryonic epidermis [20] [21] [22]. Fourthly, an extracellular matrix mainly made of fibronectin is used to resist tissue and cellular rearrangements [23].

In summary, *Drosophila melanogaster* has shown to be simple enough for easy manipulations of its genome, but complex enough to be a good model organism for most biological questions. Its wing discs are paradigms for compartmentalization, which is a very important mechanism for multicellular organisms. The *ap* gene is necessary for the D/V compartment boundary, but its regulation by transcription factors is mostly unknown.

1.2 *apterous*

ap is the selector gene expressed in the dorsal part of the wing disc [24] [25]. It is responsible for the establishment of the D/V compartment boundary [3]. Besides, *ap* is necessary for the maintenance and integrity of the compartment via the transmembrane proteins Tartan (Trn) and Capricious (Caps), and the integrins PS1 and PS2 [26] [27]. The Ap protein consists of two LIM-domains, which can interact with other proteins, and a C-terminal homeodomain, which can bind to specific DNA sequences [28][29] [76]. However, additional proteins are necessary for the function of Ap and the activation of the target gene. The LIM-domain binding protein Chip is needed to bind to the two LIM-domains. Chip consists of a LIM-interaction domain, a transactivation domain and a dimerization domain. In the functional complex, the dimerization domains form a tetrameric Ap-Chip complex, where Ap provides the DNA binding domain and Chip provides the transactivation domain. This transactivation domain interacts with the general transcription factor II A to activate the target gene. Without Chip, Ap is not able to activate its target genes, because there is no transactivation domain. Chip on its own can not attach to the DNA, hence it can not activate the DNA either [30].

After a certain time, the transcription of the target gene must be stopped. Therefore, Beadex, a protein with two LIM-domains, is exprssed. The LIM domains of Beadex compete with the LIM-domains of Ap. Since most Chip is bound to Beadex when high concentrations of Beadex are present, no Ap-Chip complex can be formed and Ap is transcriptionally inactive [31].

During wing disc development, Ap starts to be expressed in early second instar [32]. At that time, its expression in the wing disc causes the D/V compartmentalization, whereas the A/P border is already established during embryonic development [33]. Once *ap* is expressed, it starts cell-cell adhesion and cell-cell signaling processes. Firstly, it initiates the expression of *serrate* in dorsal cells, which is involved in short-range signaling. Serrate is a Notch ligand signaling to ventral cells at the D/V compartment boundary to start expressing *wg* and *Dl* [34]. Delta is also a Notch ligand, which induces *wg* and *Ser* in dorsal

cells. So a self-regulating feedback loop at the D/V compartment is established without further input of *ap* [35]. In summary, in dorsal cells *ap* activates *Ser*, which induces *wg* and *Dl* expression in ventral cells, subsequently Delta ensures further signaling of Serrate from dorsal cells [36]. *ap* also induces the protein Fringe, which makes the dorsal cells resistant against their own Serrate signal. So no *Dl* is expressed in the dorsal part [37]. Finally, the result of this signaling cycle is a constant expression of *wg* at the border. Wingless is a ligand of the Wnt family and responsible for the wing identity by the activating the expression of further target genes [17].

Ap also activates a self-regulating mechanism. After the early second instar, *Beadex* expression is activated by Ap. As mentioned above, Beadex competes with Ap for Chip and rendering Ap transcriptional inactive.

The *ap* gene consists of four main parts, the protein coding sequence, *ap* Polycomb Response Element (apPRE), and the two enhancers apE and apDV. apPRE, apE and apDV are the functional most important sections of the 27kb long *ap* wing enhancer region. All three are absolutely essential for the proper development of the *Drosophila* wing. The PRE is closest to the coding sequence. It functions as a cellular memory and keeps the expression of Ap at a constant level. Besides, it makes sure that Ap is absent in cells where *ap* has not been activated [38]. apDV is furthest away from the protein coding sequence. It is activated by Ap, Vestigial and Scalloped proteins in dorsal cells close to the D/V boundary. The early *ap* enhancer (apE) is activated in early second instar in all future D cells. Its activity gets reduced during development, whereby the notum and the hinge have high and consistent apE activation [39]. There are four conserved regions in this enhancer, m1, m2, m3 and m4 [2]. m1 and m4 mediate repressive input, both containing an Ets domain and Zink finger transcription factor binding sites. Additionally, m4 has a homeodomain transcription factor binding site. m2 and m3 integrate activation input. Each one contains a TCF helper site, a TCF site and a homeodomain transcription factor binding site [2]. Deciphering the regulatory input into apE is the main issue of this thesis.

1.3 Genetic methods

1.3.1 Bioinformatic

Simultaneously with the development of new procedures and methods in biology, a growing amount of data was produced which also needed processing. Therefore data analysis became more complex and time-consuming, or even impossible. An example for this is the discovery of the double helix of the DNA. It took years and hundreds of scientists failed to solve the structure before the breakthrough was achieved [40]. Around 1970, the first advanced help was used. A computer program predicted the binding site of antibodies [41]. This was the start of bioinformatic.

Bioinformatic has several typical assignments such as sequence analysis, prediction of three dimensional structures, prediction of interactions, development of new visualization tools and compilation of simulations [42]. However, the major task is data management, which consists of handling gene, protein and drug databases and making the information easily available to the public [43].

Over time, the importance of bioinformatic grow exponentially due to the progress in computing power, growing databases and advancing algorithms. Besides it became more powerful, cheaper, capable of more tasks and available for everyone [41] [44]. This growth was only possible, because more and more projects were done with the use of computers. Missing functions of programming languages were written in the projects and published, contributing to the open source libraries. So with most projects using bioinformatics, the possible applications of its programs grew. A very popular example of a bioinformatic work is the human genome project. Its four main aims were the identification of all human genes, the sequencing and alignment of all 3 billion base pairs in the human DNA, the development of the algorithms needed to deal with the data and to discuss ethical questions, which were encountered along the way. It took form 1990 to 2003 and around 3 billion Dollar until all four main goals were considered achieved [45] [46]. However, the ethical consequences are still heavily discussed today [47]. Nevertheless, the human genome

project was a huge success, not just for genetics, but also for bioinformatic. During the 13 years the project was running, the capabilities of sequence analysis and data management were enhanced on and on. Furthermore, the existing data network was extended tremendously to distribute all information around the world. An important part of this development are the algorithms assembling the small pieces of sequenced DNA into the complete genome. In modern whole genome analysis, slightly modified programms are still in use [48]. The established networks are likewise in use today. Today big data analyses such as the interaction between remote DNA segments and the genome-wide association studies looking for illnesses [49] are based on this foundation [50].
In addition pharmacy companies have been using bioinformatics to reduce the cost of the design of new drugs [51]. In silico analyses like predictions of three dimensional structures, predictions of drug-receptor interactions and compilations of simulations are used regularly. Hence also pharmacy companies support the development of the corresponding programs. Over time, more and more libraries, modules and other extensions for the most common programming languages were made [46].
The most popular programming language, especially for scientific issues, is python [52]. Due to thousands of modules and libraries all-purpose high-level coding is possible with a few lines of code [53].

1.3.2 Reporter analysis

A popular way to analyze gene regulatory elements is via classical reporter analysis. For this method, a putative gene enhancer region or a part of it is isolated from of the genome, might be mutated and combined with a reporter gene. This reporter gene construct is then reinserted into the genome of the organism. The activation of the reporter shows then directly the activity of the investigated enhancer fragment. A typical reporter gene for this purpose is *LacZ*. The *LacZ* gene is part of *Escherichia colis Lac operon*, which is responsible for the metabolization of lactose [73]. *LacZ* encodes the protein βGalactosidase. In *E. coli*, β-Galactosidase hydrolyzes the disaccharide lactose into glucose and galactose. How-

ever, it can also turn the colourless dye 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactopyranosid (X-Gal) into its blue form. Therefore, all cells with an active *LacZ* reporter gene will turn blue if stained with X-Gal [74] [73]. Nevertheless, the sequence containing the reporter and the enhancer must be inserted in the genome first.

For this issue, P-elements have been used for a long time. P-elements are class II transposons, which are 2907 bp long. They contain two 11 bp long direct repeats and two 31 bp long terminal inverted repeats which function as binding sites for the transposase [68] [69]. Next, the transposase sequence is replaced with the gene sequence of interest. The P-element without the transposase sequence is immobile on its own, since transposase is needed for cutting out the P-element at the two recognition sites and inserting it at a random place. Finally, the purified immobilized P-element is injected in the organism along with the sequence for transposase. Transposase is produced once and the P-element is inserted at a random place. After insertion the P-element is immobile, because the sequence of transposase was not inserted, hence no transposase is expressed anymore [68] [69].

However, the random insertion of P-elements also means that different constructs are inserted into different genomic environments. This can lead to so-called position effects. To avoid different position effects, which can influence the expression of the inserted sequence, a more specific method is used, where different reporter constructs can be inserted into the same genomic locus. The φC31 integrase system enables a targeted integration [70]. φC31 integrase comes initially from the bacteriophage φC31, but has been adapted for other species like *D. melanogaster* [71]. It catalyses the reaction between the bacterial attachment site (attB) and the phage attachment site (attP). To use the φC31 integrase system, a plasmid containing attB and the desired sequence, a attP landing site somewhere in the genome and an some φC31 integrase are necessary. When the φC31 integrase and the plasmid are in contact to the attP site, the integrase cuts the attB and the attP and inserts the plasmid. The parts of attB and attP are combining during the insertion forming attR and attL, hence φC31 integrase can not cut the plasmid out or a new plasmid in again [70] [71]. This is shown in figure 1.2.

However, reporter analysis is not as meaningful as the manipulation of the endogenous

region itself.

1.3.3 In situ rescue system

In situ rescue systems are a way of manipulating the endogenous region. In the following, an in situ rescue system at the *ap* locus is explained. Firstly, the region of interest is removed from the genome and an attP site is inserted instead. This can, for example, be done by flipase- mediated recombination [2]. Secondly, a plasmid containing the mutated region of interest, a marker (for example *yellow*) and an attB site and φC31 integrase are injected. This leads to a site-specific recombination, bringing the new version of the sequence of interest back to the endogenous locus. This sequence is typically an enhancer region, so the expression conditions of a gene can be observed. The process of site-specific recombination can be seen in figure 1.2

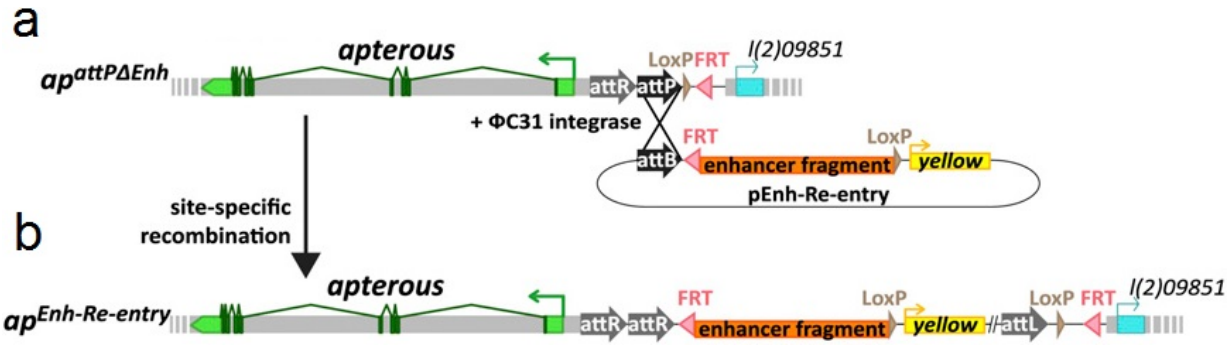


Figure 1.2: **Insertion of the rescue sequence**
(a) φC31 integrase and a plasmid containing a modified version of the *ap* enhancer region, a *yellow* marker and an attB site are injected in a fly egg. The *ap* enhancer region was previously removed from the genome of the fly. Instead of the enhancer region it has an attP site.
(b) The φC31 integrase catalysis a site-specific recombination, so the modified version of the *ap* enhancer region and the yellow marker are in the place, where the original *ap* enhancer region was. (modified from [2])

1.3.4 RNAi and its control systems

There have been several other projects similar to the human genome project for species like mice, *C. elegans* and *Drosophila melanogaster* [54]. By now, all of the named projects resulted in the completely sequenced genome of the according species. Despite the fact that all the sequences of every gene in the genomes are known, respective function of most genes remain unknown. One way of investigating the function of these genes are loss-of-function screens. To ascertain this information the genes can be silenced. A well known method for this task is RNA interference.

RNA interference (RNAi) or Post-Transcriptional Gene Silencing (PTGS) was first discovered in 1990 in plants [55]. Eight years later the first paper targeting RNAi in *C. elegans* was published [56]. It showed that double-stranded RNA (dsRNA) but not single-stranded RNA (ssRNA) can lead to degradation of endogenous mRNA. Later on, it was discovered that the dsRNAs were cut by a special protein called Dicer to small interfering RNAs (siRNAs), which are around 20-25 base pairs long. Subsequently, the siRNAs form a RNA-induced silencing complex (RISC) with argonaute proteins that have endonuclease activity. This RISC binds to the mRNA via its antisense siRNA, leading to the degradation or blockage of the respective mRNA [57] [58]. Thus the affected protein will not be produced.

Originally this mechanism is a natural defense, degrading the genetic material of intruded viruses [56]. By now, RNA interference is established as a valuable genetic tool to generate loss-of-function phenotypes [58]. RNAi has been used for functional analysis of the genes in the genome databases by depletion of chosen transcripts. Thousands of genes have been characterized with this technique [58] [59] [60]. Certainly the technique, in which RNAi is used, has changed drastically since its first use. New, more effective, types of RNA are used for silencing, like small hairpin RNA (shRNA) [61] [62]. Combining RNAi with other genetic methods, no injection of dsRNA is needed anymore, additionally a spatial and temporal control of the knockdown area is possible.

In *Drosophila*, the most common way to use RNAi is to insert the RNA of interest with

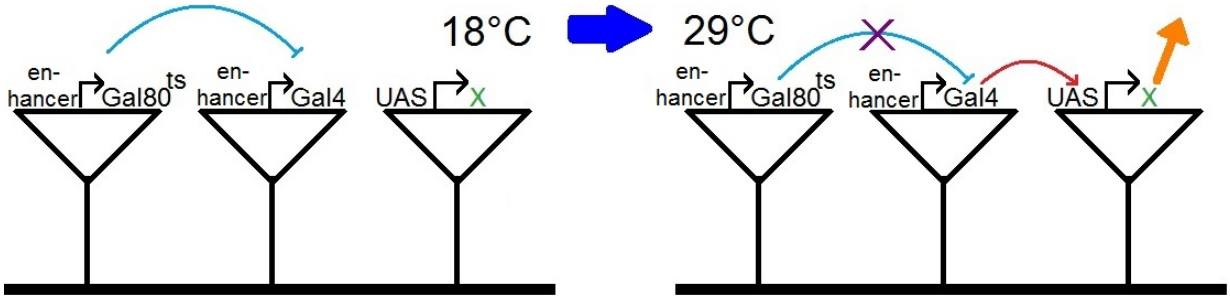


Figure 1.3: **Example of a Gal80^{ts}/Gal4-UAS system**
The left site shows the system at 18°C. Gal80^{ts} is stable at this temperature, so it represses Gal4 in every cell, where its tissue-specific enhancer is active. Therefore at 18°CGal4 can only be active in cells, where its tissue-specific enhancer is active, but the one of Gal80^{ts} is inactive. In those cell Gal4 can activate UAS by binding to it, which in turn activates the expression of x. At 29°Chowever Gal80^{ts} is unstable, hence not able to inactivate Gal4. Every cell, in which the enhancer of Gal4 is active, Gal4, and x gets produced.

an upstream activation sequence (UAS) as a DNA sequence in the genome. Due to the fact, that this system can also be used for the controlled expression of other genes, the RNA is further referred to as x. UASs are enhancers, which are only active, when the yeast transcription factor Gal4 binds specifically to them. Hence, without Gal4 expression, there is no expression of x. The line with the x sequence is called responder line, since it usually gives a visible phenotype, when Gal4 is active. Gal4, on the other hand, drives the responder line, hence it is called driver line. Several different mechanisms are available for the control of the driver line. By choosing a corresponding enhancer, Gal4 can for example just be activated in muscles or neurons or even only in specific cells within these tissues. However, there are more complex possibilities like the necessity for an input of multiple factors at once, or a deactivation if certain conditions are met. These options are analogously to the Boolean operators in microprocessors. The necessity for an input of multiple factors at once can be archived by an intersectional strategy. Split-Gal4 is a possible technique for this. A first line only expresses one half of the Gal4 protein at certain conditions, a second line produces the other half at different conditions. Both half are inactive on their own, but if coexpressed they become active and bind to the UAS [63].

A NOT operator can be made with Gal80. Gal80 is a GAL4 inhibitor by binding to the transcriptional activation domain of Gal4, therefore no activation of the gene under UAS control takes place [65]. Figure1.3 shows such a Gal80/Gal4-UAS system. In this case Gal80 is expressed in the whole organism leading to a complete inactivation of Gal4. However, a temperature-sensitive Gal80 ($Gal80^{ts}$), which is unstable at temperatures around 29°C, can be used. Thus, Gal4 activity is no longer repressed at high temperatures. However, Gal4 is only expressed in cells, where the respective tissue-specific enhancer is active. Hence, *X* is also only expressed in these cells. This system has an additional advantage besides the controlled time and space of the activation of *X*. By actively repressing Gal4, the unwanted expression of it is lower, than in a standby phase [65]. Certainly, there are some disadvantages of these complex control mechanisms, namely an increased amount of work. Several different and independent insertion sites are required, accordingly more crosses must be made to create the desired final organisms [62] [61].
In addition to RNAi, there is an other important way to suppress the function of a gene. Dominant negative mutations (DN) do exactly, what their name suggests. A dominant mutation suppresses the wild type allele by over production, DN mutants produce an enormous amount of functionless but stable mutant polypeptides. Theses mutant polypeptides can prevent the original proteins from functioning correctly via several ways. For example, if the natural protein is a multimeric protein, the mutant polypeptide can take the place of a unit, rendering the whole multimer useless [66].

1.4 Aim of the project

Ap is one of the most important proteins for wing development, therefore its role during development is well studied. Nevertheless very little is known about the regulatory mechanisms of *ap*. However, the understanding of the regulation of *ap* is crucial for the studying of the compartmentalization.
On these grounds *cis*-regulatory elements play a very important role in this understanding. This project aims to narrow down the exact sequences at which the transcription factors

bind to activate or repress the initial *ap* expression and to find the correct transcription factors necessary for the control of *ap* expression. Three complementary genetic and a bioinformatic approach were used to accomplish this task. First, I coded an algorithm, that collected all helpful data from the data bank flybase.org for a preselection of the most probable transcription factors binding at the known binding site areas. Second, seven LacZ reporter constructs with differentially mutated *ap* enhancer sequences were made, to study the effect of the specific deletions. Third, the same deletion mutations were brought back in the *ap* locus via the φC31-integrase-dependent in situ rescue system. Subsequently the results of the LacZ reporter study were compared to the wing phenotype. Based on this results, RNAi-induced knockdowns of the genes known to influence the *ap* expression and the genes of new possible transcription factors were made and the wing phenotypes were examined. For the knockdowns viable to the third instar and showing an significant phenotype antibody and LacZ stainings were made.

2 Material and methods

2.1 Polymerase Chain Reaction (PCR)

All PCR programs are made in a T3 Thermocycler (Biometra). After the PCR the product can be cleaned with a agoarose gel separation.

Standard Phusion PCR protocol:

35.5 μl dH_2O

10 μl Phusion ® HF Reaction Buffer (5X)

1 μl dNTP (10mM)

1 μl Primer I (20 μM)

1 μl Primer II (20 μM)

1 μl Template (1.25ng)

0.5 μl Phusion polymerase (added last)

PCR program:

1. 98°C 2'
2. 98°C 20"
3. 60°C 20"
4. 72°C 30" per 1 kb of fragment $\rightarrow$ Back to 2. (35X)
5. 72°C 7'
6. 4°C ∞

2.2 Single fly PCR

Squishing solution 48.5 μl dH_2O

0.5 μl Tris-HCl (1M, pH 7.5)

0.25 μl EDTA (500mM, pH 8.0)

0.25 μl NaCl (5M)

0.5 μl Proteinase K (10μg)

The fly is placed in an Eppendorf tube and squished with a pipette for 15 seconds. Afterwards it is re-suspended in 50μl of the squishing solution and incubate for 37°Cfor 30'. Subsequently the mixture was heated to 95°Cfor 2 min to deactivate the Proteinase K. Before the DNA solution is used it is centrifuged shortly. The solution can be stored at 4°Cfor some days or several month at -20°C.

2.3 X-Gal staining of imaginal discs

Larvae in the L3 stage were ripped open in the middle, inverted and unimportant tissues were removed. The inverted larvae were stored in PBS on ice until enough samples are collected or the first larvae is stored for maximal time of 10 min. 1% glutaraldehyde (Fluka) in PBS on ice is used to fix the carcasses. After 15 min the fixative was removed and the larvae were washed twice with 1‰Tween 20 (Fluka) in PBS. The larvae were incubated at 37°Cin the dark in the staining solution for exactly 90 min. Afterwards the samples were washed twice with 1‰Tween 20 (Fluka) in PBS and one time in PBS, before the larvae were dissected in 80% glycerol in water on a microscope slide. Finally the sample was sealed with nail polish.

Staining solution:

440μl dH_2O

25μl 200mM NaP_i (pH 7.2)

15μl 5M NaCl (in dH_2O)

0.5μl 1M $MgCl_2$

5μl 333mM $K_4[Fe^{II}(CN)_6]H_2O$

5μl 333mM $K_3[Fe^{III}(CN)_6]$

2.5μl 10% Tween-20

8μl 5% XGal (appliChem) in dimethylformamide

2.4 Fixation with paraformaldehyde

Third instar larvae were ripped open in the middle, inverted and unimportant tissues were removed. The inverted larvae were stored in PBS on ice until enough samples are collected or the first larvae was stored for maximal time of 10 min. 4% paraformaldehyde (PFA) in PBS was used to fix the larvae for 15 min at room temperature. After fixation, the PFA solution was removed and the larvae were washed twice with 1‰Tween 20 (Fluka) in PBS and one time in PBS, before the larvae were dissected in 80% glycerol in water on a microscope slide. Finally the sample was sealed with nail polish.

2.5 Fly crosses for the RNAi constructs

$\varnothing\ \alpha$ w; $\frac{\text{If}}{\text{CyO YFP}}$; $\frac{apRLacZ}{TM3,Sb}$ X $\male\ \beta$ ywhsflp; enGal4 UAS CD8GFPGal80^{ts}; $\frac{\text{MKRs}}{TM6B,Hu}$

$\male$ a $\underset{\rightharpoonup}{\text{w}}$; $\frac{\text{enGal4 UAS CD8GFPGal80}^{\text{ts}}}{\text{CyO YFP}}$; $\frac{\text{apRLacZ}}{\text{TM6B,Hu}}$ X $\varnothing\ \alpha$ w; $\frac{\text{If}}{\text{CyO YFP}}$; $\frac{apRLacZ}{TM3,Sb}$

$\male$ a $\underset{\rightharpoonup}{\text{w}}$; $\frac{\text{enGal4 UAS CD8GFPGal80}^{\text{ts}}}{\text{CyO YFP}}$; $\frac{\text{apRLacZ}}{\text{TM6B,Hu}}$ X $\varnothing\ \alpha$ w; $\frac{\text{If}}{\text{CyO YFP}}$; apRLacZ

w; $\frac{\text{enGal4 UAS CD8GFPGal80}^{\text{ts}}}{\text{CyO YFP}}$; $\frac{\text{apRLacZ}}{\text{TM6B,Hu}}$ and w; $\frac{\text{enGal4 UAS CD8GFPGal80}^{\text{ts}}}{\text{CyO YFP}}$; $\frac{\text{apRLacZ}}{\text{TM3,Sb}}$

Final stock: w; enGal4, UAS-CD8GFP, tubGal80$^{\text{ts}}$; apELacZ

The final stock is homozygous, so after the final cross with the RNAi line everything is heterozygous. Gal80$^{\text{ts}}$ is expressed everywhere in the flies of the final stock, leading to an inactivation of Gal4. At high temperatures however Gal80$^{\text{ts}}$ is unstable and no longer repressing Gal4. So at high temperatures Gal4 is expressed in the posterior part of the wing disc, since its promoter en is active there. Gal4 activates the UAS starting the GFP expression. In the final cross UAS also activates the RNAi construct. Besides on the third chromosome LacZ is expressed depending on the apE activity. In the final cross apE can

be influenced by the RNAi. Final cross:

♀ w; enGal4 UAS CD8GFPGal80^{ts}; apRLacZ X ♂ siRNAThe eggs produced in the final cross kept at 18°Cfor 48 h, then a temperature shift to 28°Cwas made to start the knockdown.

3 Results

3.1 Bioinformatic

The 463 bp long apE enhancer in the C2 region of *ap* is known to consist of four conserved regions. These four cluster have ETS domains (m1 and m4), zinc finger TFs (m1 and m4) and homeodomain TFs binding sites (m2, m3 and m4) [2]. To find the transcription factors (TFs) activating the early *ap* enhancer, an in silico analysis was made by Dimitri Bieli (University of Basel). This analysis predicted around 400 possible TFs.
An algorithm was written to search the database flybase.org for further information about the possible TFs to reduce their number. Besides gathering additional information, the algorithm excluded all genes not expressed in the imaginal discs and sorted the results accordingly to a counter. The collected data are the name, their domains, the expression at L2, L3 and in the imaginal discs, of L3 wandering larvae. In addition, the program searched the known interactions between the 400 proteins. To be incorporated in the final list of candidatesthe genes must be expressed in all three stages. For a high score and therefore a high rating in the list, keywords like "wing disc", "wing disc dorsal", "wing development", "apterous", or "ap" must present. Nevertheless also candidates without the keywords were listed. The final results are shown in figure. 3.1. An overview with the exact binding sequences is shown in figure 3.2(a) and figure 3.3(a).

3.2 LacZ staining and in situ rescue system

The four conserved regions and the approximate location of potential TF binding sites in the apE are known [39] [2] [72]. However, the exact sequence important for the interaction is unknown. Therefore, a classical LacZ assay was performed to narrow down the possible binding sequences, by inserting smaller deletions in the conserved clusters. Therefore, a plasmid containing a mutated apE region and the LacZ gene is inserted via the attP landing site at zh-86Fb. zh-86Fb is known to have very little to no position effects in wing disc

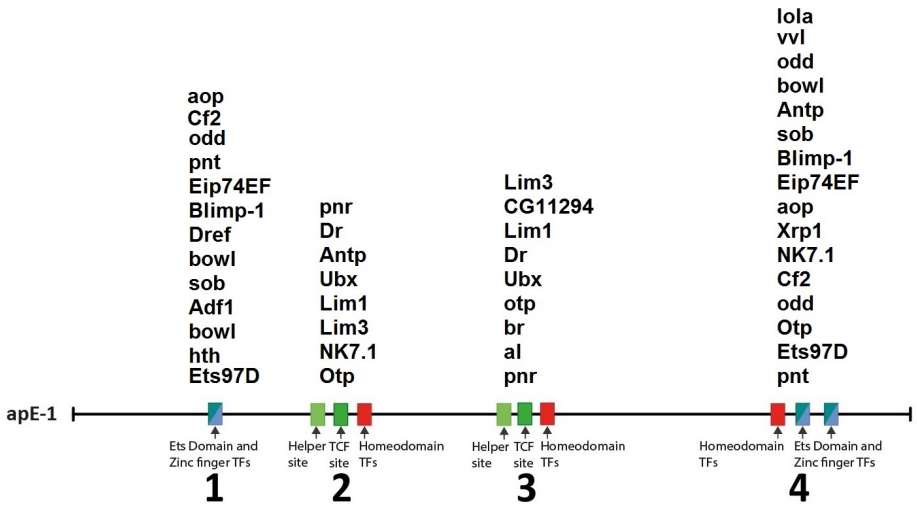


Figure 3.1: **Overview of the proposed transcription factor binding sides, after the selection** The candidates are above their suggested binding site. 1 and 4 are are repressor regions, 2 and 3 are activator regions [2]

tissue [2].

Ten differently mutated apE region were inserted this way in the already known clusters, trying to narrow down the TF binding sites. Three constructs with a mutated region 1 each in a m4 background for higher sensibility and four constructs with a mutated region3 were made. In addition, three controls were made, the first one had the clusters 1, 2 and 4 deleted, the second one 1, 3 and 4 and the third one had none of the clusters left. For reference, four additional fly stocks, which had the apE-LacZ plasmid inserted at the same place, were used. The negative controls are a wild type for the m3 deletions and a construct with a m4 deletion for the insertions of a m1 deletion in the m4 background. For positive controls stocks with deletions in the m3 or the m1 m4 region of the plasmid were used.

Figure 3.2(a) and figure 3.3(a) show the different mutations of apE inserted via the apE/LacZ plasmid. The respective X-Gal stainings are shown in figure 3.2(b)-(f) and figure 3.3(b)-(g). The differences between the X-Gal stainings shown in fig. 3.2 and fig. 3.3 and figure. 3.4 are caused by different staining solutions and incubation times. Figure. 3.2(b)-(f) are the L3 wing discs of m1 deletions in m4 background and the controls. m1.3m4 has a similar activation pattern like m4, hence the deletion of m1.3 had no effect of the apE

activity. m1.2m4 and m1m4 also have a similar activation pattern. Only the ventral part of the wing pouch is completely free of blue dye. The binding sequence of the repressive TF is likely in the m1.2 part. m1.1m4, however, shows a intermediate activation between m4 and m1m4. The anterior part looks like m4, the posterior part looks like m1m4.
Figure 3.3(b)-(g) show the LacZ stainings of the m3 deletions and the corresponding controls. m3, m3.1 and m3.3 show no visible apE activation, so the complete TF binding site for the activation is removed. m3.4 has strong but small-area apE activation in the hinge, especially in the posterior part. m3.2 shows a broader expression, also mainly in the hinge.
Figure 3.5 show the LacZ stainings of the multiple deletions, which are an overall control. m1m2m3m4 is expressed as expected by showing no apE activation. m1m2m4 and m1m3m4, however, show completely new patterns. In figure 3.5(a), the apE activity describes three-quarter of a circle around the wing pouch, leaving the posterior part next to the wing pouch without activation. In figure 3.5(b), no activation is visible, except for two to three dots in the wing pouch at the D/V boundary.
To verify of the LacZ staining results, additional in situ rescue constructs with the same deletions like the ones in the LacZ assay were made (see fig. 3.2(g)-(i) and fig. 3.3(h)-(k)). For the manipulation of the endogenous locus, a fly stock was used, which had an attP site instead of the *ap* enhancer region (see fig. 1.2). At this attP site, a plasmid containing the mutated forms apE, the wild type DV enhancer and a yellow marker were inserted.
The phenotype of the in situ rescue systems are displayed in figure 3.2(g)-(i) and figure 3.3(h)-(k). All phenotypes are rather strong, every wing has large fluid retentions, is just a stump or even completely missing. However, the results are similar to the ones of the LacZ constructs. m1.2m4 has the most drastic phenotype of all the m1 deletions. Also m1.3m4 shows strongly modified wings, despite the fact that the corresponding LacZ stainings display no aberrations. m3.1 and m3.3 reentries have no wings. The m3.2 and m3.4 reentries looks similar to the reentries of m1.1m4 and m1.3m4. The posterior parts of the wings, including the alula, are nearly completely gone. Besides only the veins L1, L2 and L3 are left. L1, L2 and L3 are 3-8 times broader then in wild type wings. This might be caused by the to the bending of the wings due to the large fluid retentions.

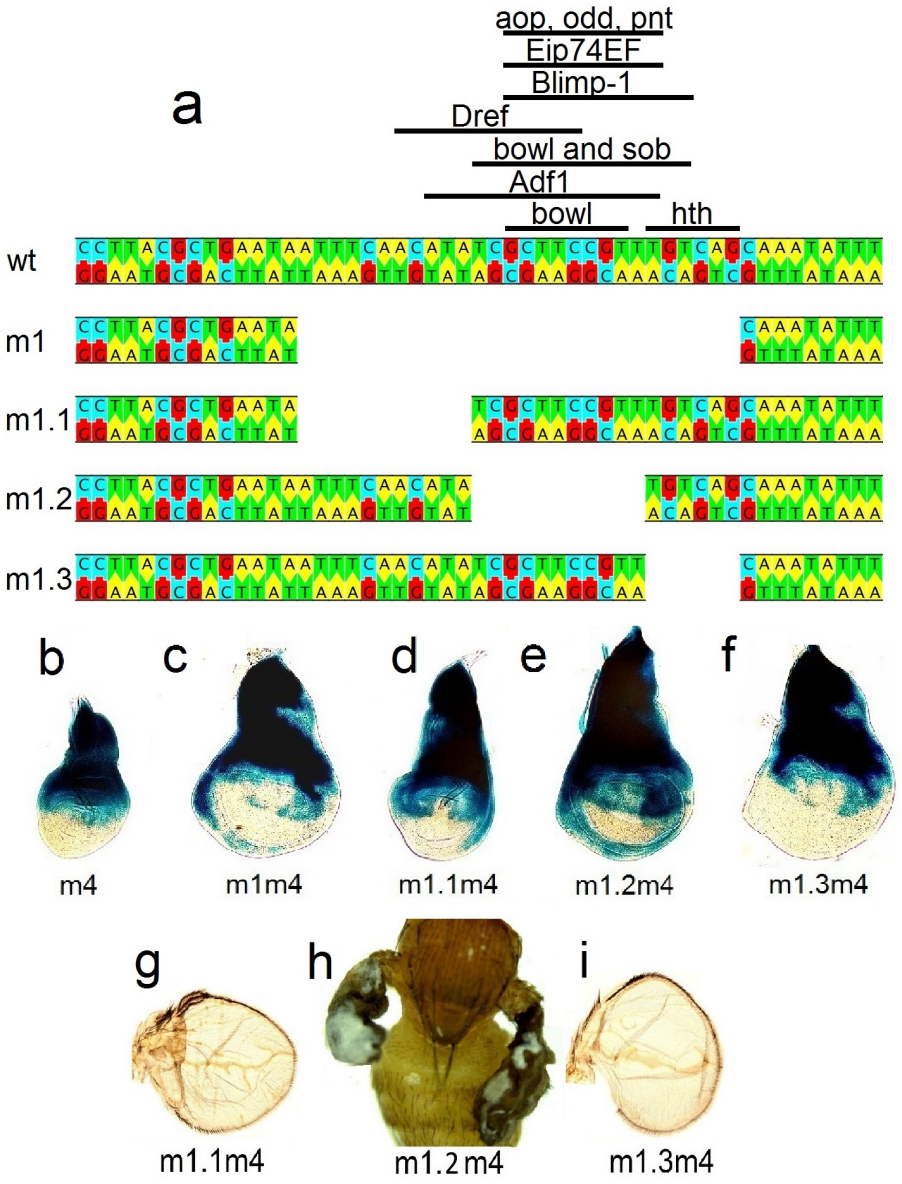


Figure 3.2: **TFs, sequences, X-Gal stainings and reentries of m1 deletions**
(a) shows the sequences of the different deletions in the m1 region with the corresponding TF binding sites. The binding sites are symbolized with a black line. The deletion of the complete m1 part removes all possible binding sites. All induced m1 deletions are in a m4 background, not shown here.
(b)-(f) are the X-Gal stainings of the deletions shown in (a). m1.2m4 has a phenotype as strong as the one of the control m1m4. m1.3m4 shows no phenotype. m1.1m4 however shows a phenotype between the positive and the negative control.
(g)-(i) are the reentry constructs with the same deletions shown in (a)-(f). Again m1.2m4 has the strongest phenotype, however m1.1m4 as well as m1.3m4 show strong phenotypes.

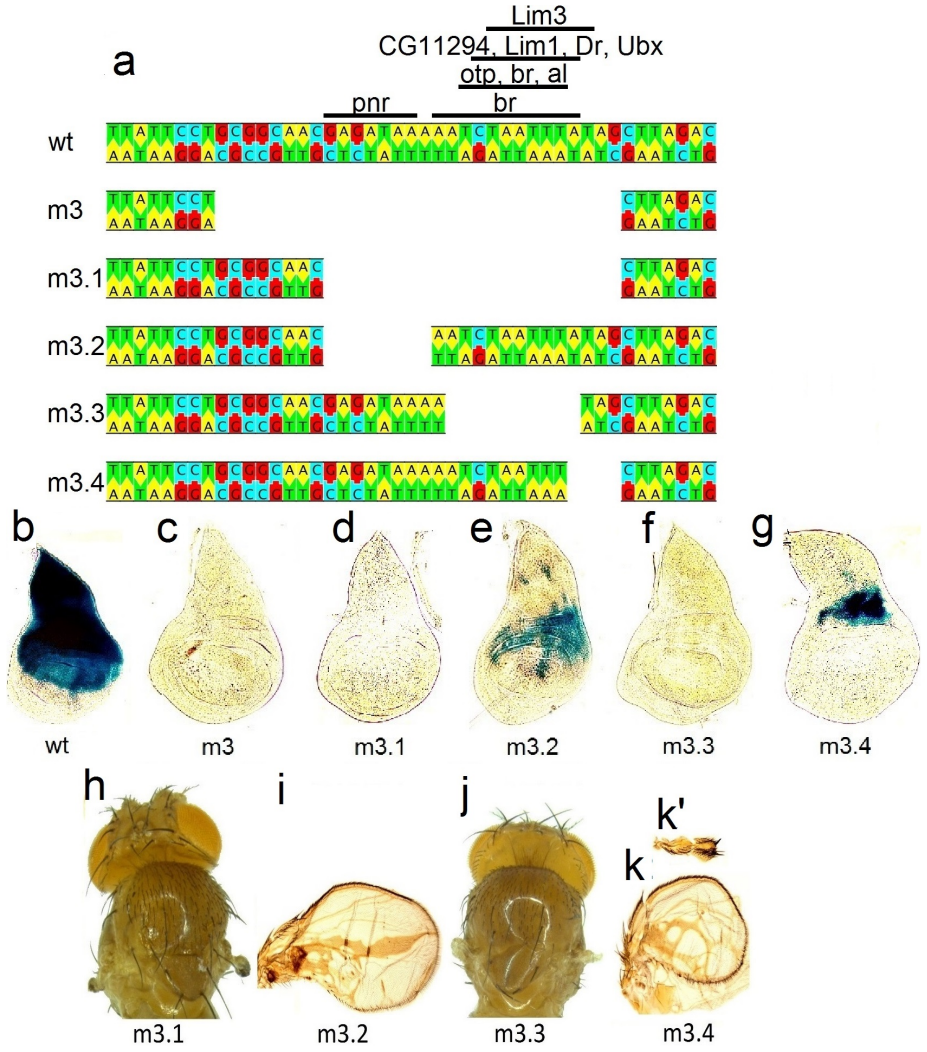


Figure 3.3: **TFs, sequences, X-Gal stainings and reentries of m3 deletions**
(a) shows the sequences of the different deletions in the m3 region with the corresponding TF binding sites. The binding sites are symbolized with a black line. The deletion of the complete m3 part removes all possible binding sites. All induced m3 deletions were made in a wild type apE.
(b)-(g) are the X-Gal stainings of the deletions shown in (a). m3, m3.1 and m3.3 show no apE activation at all. In m3.2 and m3.4 the apE activity is virtually reduced to the hinge region.
(h)-(k) are the reentry constructs with the same deletions shown in (b)-(g). Like expected flies with m3.1 and m3.3 deletions in the endogenous locus have no wings. m3.2 and m3.4 have crippled wings. k' is an extreme example of an m3.4 phenotype.

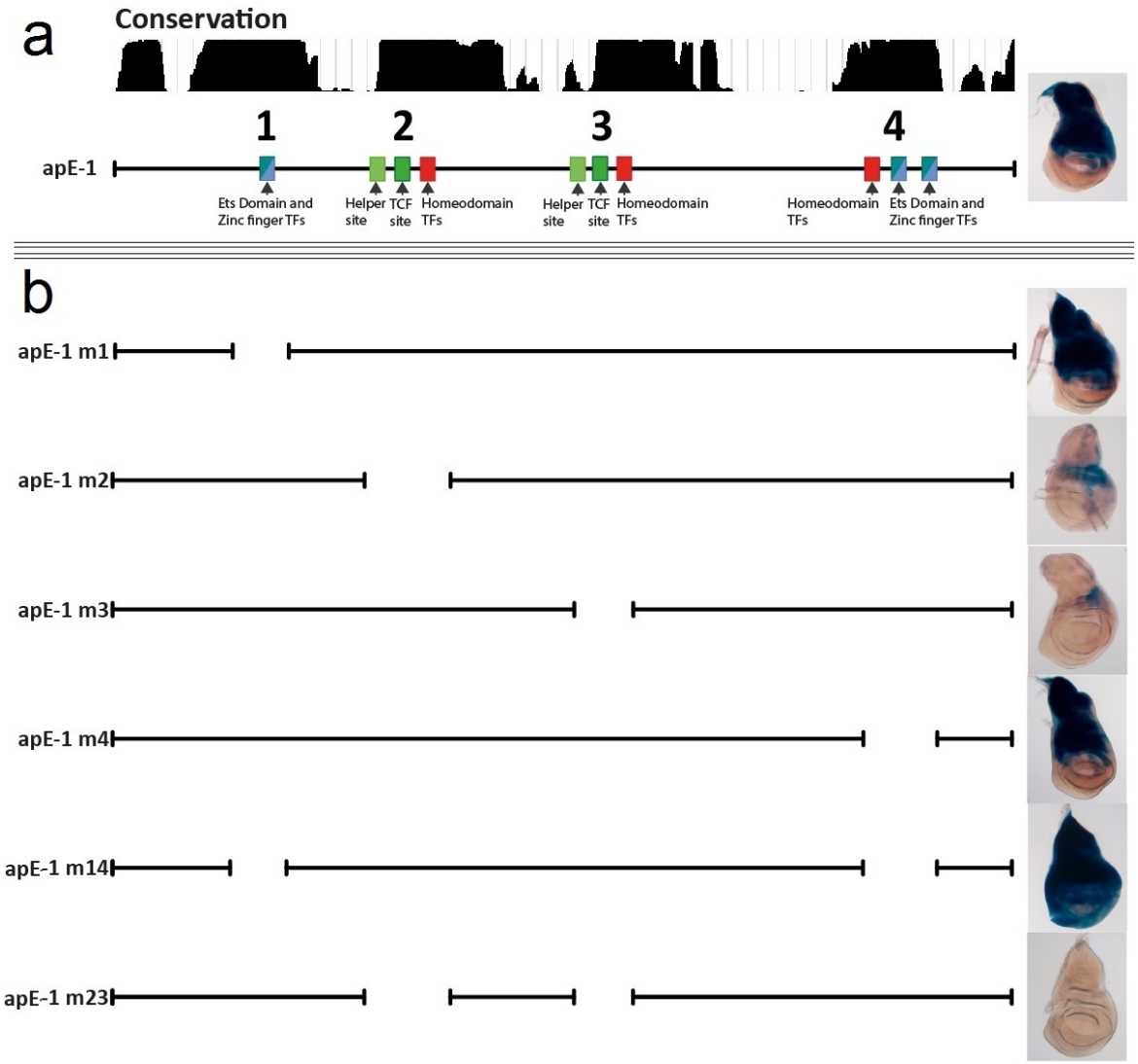


Figure 3.4: (a) **apE with its conserved areas and identified sites**
(b) **Deletions of the located binding sites in LacZ constructs**
Deletion of m1 and m4 show a slight upregulation of the apE activation, whereby the the upregulation of m1 seems stronger. The deletion of both sites lead to a strong up regulation, only sparing the ventral side of the wing pouche. The deletion of m2 and m3 show a strong down regulation of the apE activity, whereby the down regulation caused by the missing of m3 is stronger. The deletion of both sides lead to no apE activity.
Images modified from [2]

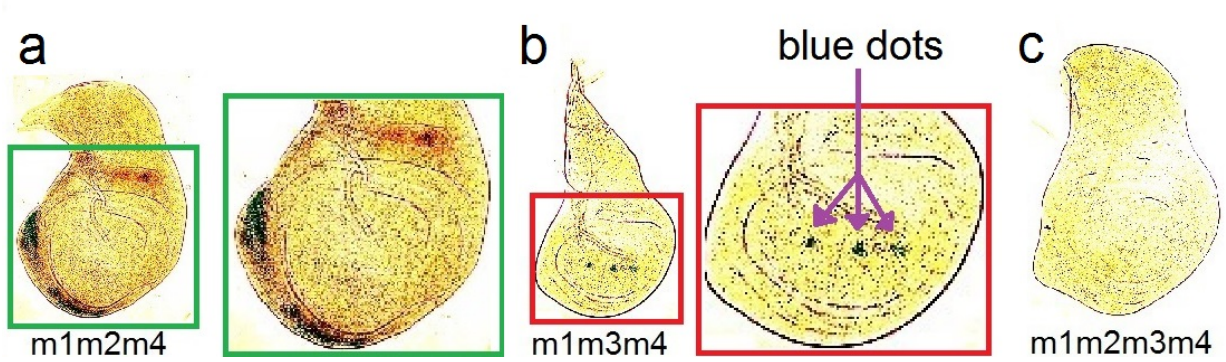


Figure 3.5: **X-Gal stainings of muptiple deletions**
The X-Gal stainings of the wing discs shown here were intended to be controls. m1m2m3m4 has no apE activation at all. However m1m2m4 and m1m3m4 show unexpected phenotypes. m1m2m4 has only m3 as an activating region left, without repressors. The apE activation takes place in three quarters of a circle around the wing pouch. apE activation in the ventral part of the wing disc is uncommon. m1m3m4 activates apE in the wing pouch at the D/V compartment boundary in three points.

3.3 RNAi lines

A RNAi assay in combination with an apE-LacZ reporter was used to investigate the changes on the wing development and apE activation caused by the knockdown of selected genes. This assay is divided into two parts. Therefore, around 55 RNAi, respectively DN, lines are crossed with the final stock "w; enGal4, UAS-CD8GFP, tubGal80^{ts}; apELacZ" (2). The egg from this cross are collected at 18°C, to prevent an activation of Gal4. 48 h after the eggs were laid, they are shifted to 29°C to activate Gal4 and start the knockdown of the selected gene. For the first part, the flies were allowed to grow until they are adult. The resulting wing phenotypes were investigated and a list of candidates with strong phenotypes was made. For the RNAi lines that showed wing phenotypes or lethality, the experiment is repeated, but the larvae were dissected when they started wandering around. One half of the dissected larvae were stained with X-Gal, the other half was stained with antibodies against Wg and Ap.

Before the RNAi assay was performed, the driver line enGal4 was controlled, to check if

there is a leakage of Gal4 at low temperatures, or a missing activation of the UAS at high temperature (see fig. 3.6). The Gal4 expression seen is like expected and LacZ of apE activity is unaffected.

The wing phenotypes are shown in figure 3.7 and figure 3.8. Figure 3.7 includes all lines, which could possibly have an effect on wing development and most likely on Ap expression. However only *arm* (fig. 3.7(d)), *aop* (fig. 3.7(e)), *vvl* (fig. 3.7(f)) and *pnt* (fig. 3.7(j)) show a strong phenotype, while none looks like the controls UASap and ap^{textEy03046} (fig. 3.7(b) and fig. 3.7(c)). The flies of the crossing, whose wings are shown in figure. 3.7(d) (arm), usually died while hatching. Pnt protein could bind at the activating and/or repressing sequence, whereas it is more likely, that it only interacts with the activating regions, because it is a part of the EGFR cascade activating Ap.

Figure 3.8 shows lines, which possibly lead to the activation *ap*. *vein* (fig. 3.8(d)), *EGFRDN* (figure 3.8(e)), *EGFR* (31183) (figure 3.8(f)), and *pnt* (figure 3.8(k)) show the strongest phenotype. The wings of the crossings with *vein* and *EGFR* do not have a posterior part. The wings of the *ap* knockdown look like wild type wings.

Besides the crossings shown, five other crossings, *arms10*, *TCFDN*, *lola* (26714), *Dref* and *hth*, died before hatching.

For the second part of this assay, *arms10*, *arm*, *aop*, *br*, *vvl*, *hth*, *lola*, *pnt*, *vein*, *EGFR* and *EGFRDN* were selected. In the X-Gal stainings, *arms10* and *arm* showed a significant increase in apE activity in the posterior compartment near the wing pouch (see fig. 3.9(f) and (g)). *hth* and *pnt* showed a decrease in apE activity in the posterior hinge region(see fig. 3.9(k) and (m)). *lola* and *vein* also showed a decrease, but in the posterior notum part (see fig. 3.9(l) and (n)). The other crossings seem to have modified apE activities too, but the fluctuations from disc to disc are to big for an exact statement.

The same holds true for the antibody stainings (see fig. 3.10). *arms10* even shows the opposite phenotype, a decreased *ap* expression in the posterior hinge part. The only consistent results are the missing *ap* activity in the posterior hinge of *hth* and *pnt* and the missing *ap* activity in the posterior notum of *lola* and *vein*.

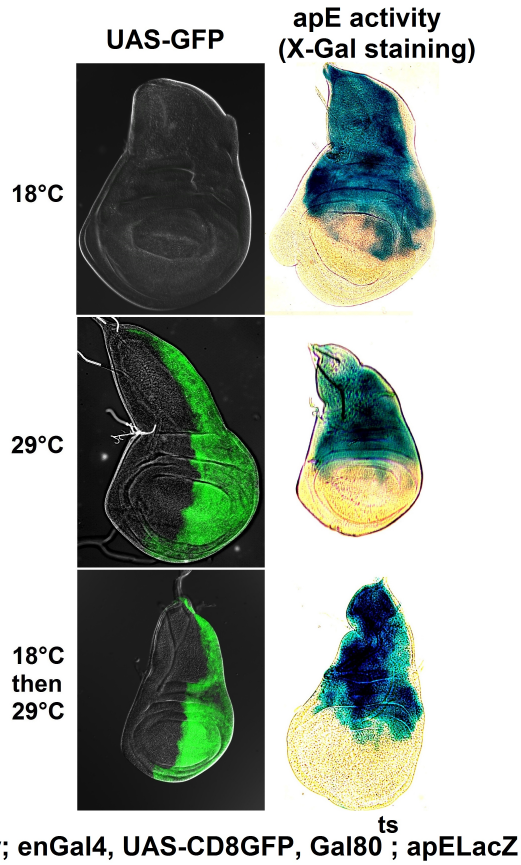


Figure 3.6: **Control of the driver line and the LacZ system**
The knockdown is induced by the instability of Gal80^{ts} at high temperatures. Gal4 is active under the right conditions and the LacZ stainings are not affected.

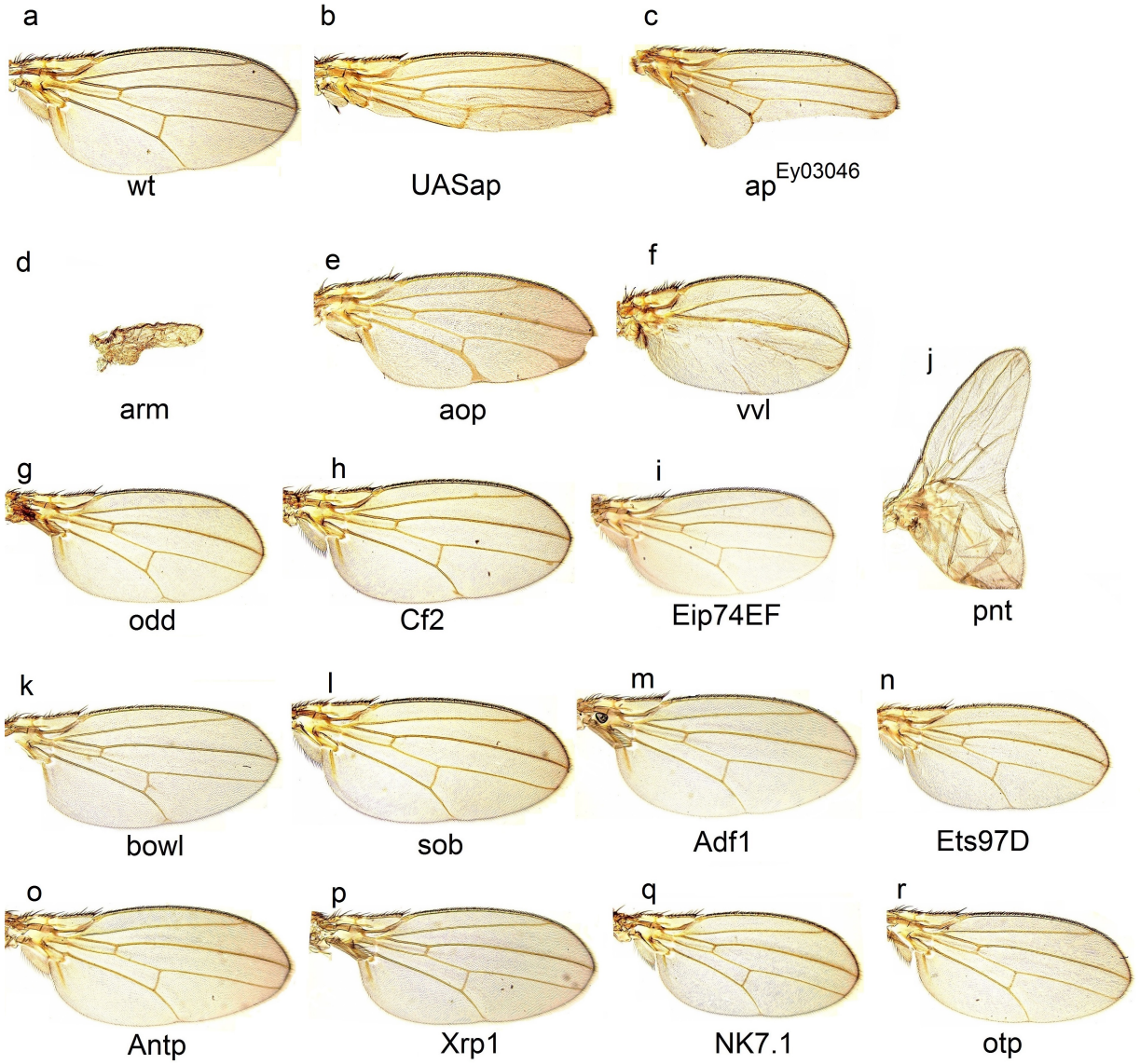


Figure 3.7: **Wing phenotypes of RNAi constructs silencing potential repressing TFs**
All wing phenotypes of TFs, which are supposed to bind to the repressive TF binding sites, are listed here. All of these could possibly have a negative input at apE, hence their knockdown should cause an overexpression like UASap and ap^{textEy03046}. arm, aop, vvl and pnt are selected for the second part of the assay. The shown wings are sorted for the strength of the phenotype.

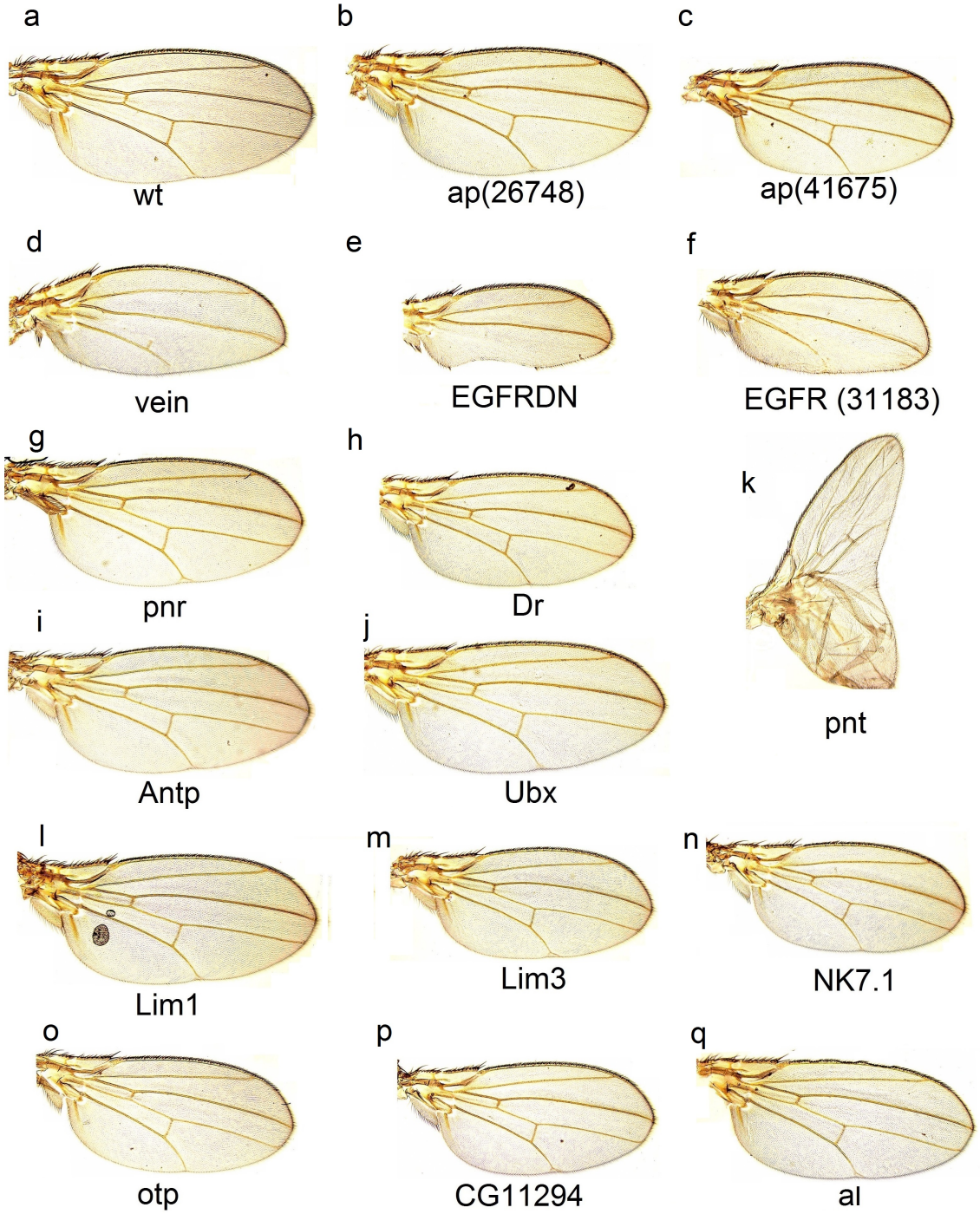


Figure 3.8: **Wing phenotypes of RNAi constructs silencing potential activating TFs**
All wing phenotypes of TFs, which are supposed to bind to the activating TF binding sites, are listed here. All of these could possibly have a activating input at apE, hence their knockdown should cause an reduction in *ap* expression. vein, EGFR and pnt are selected for the second part of the assay. The shown wings are sorted for the strength of the phenotype.

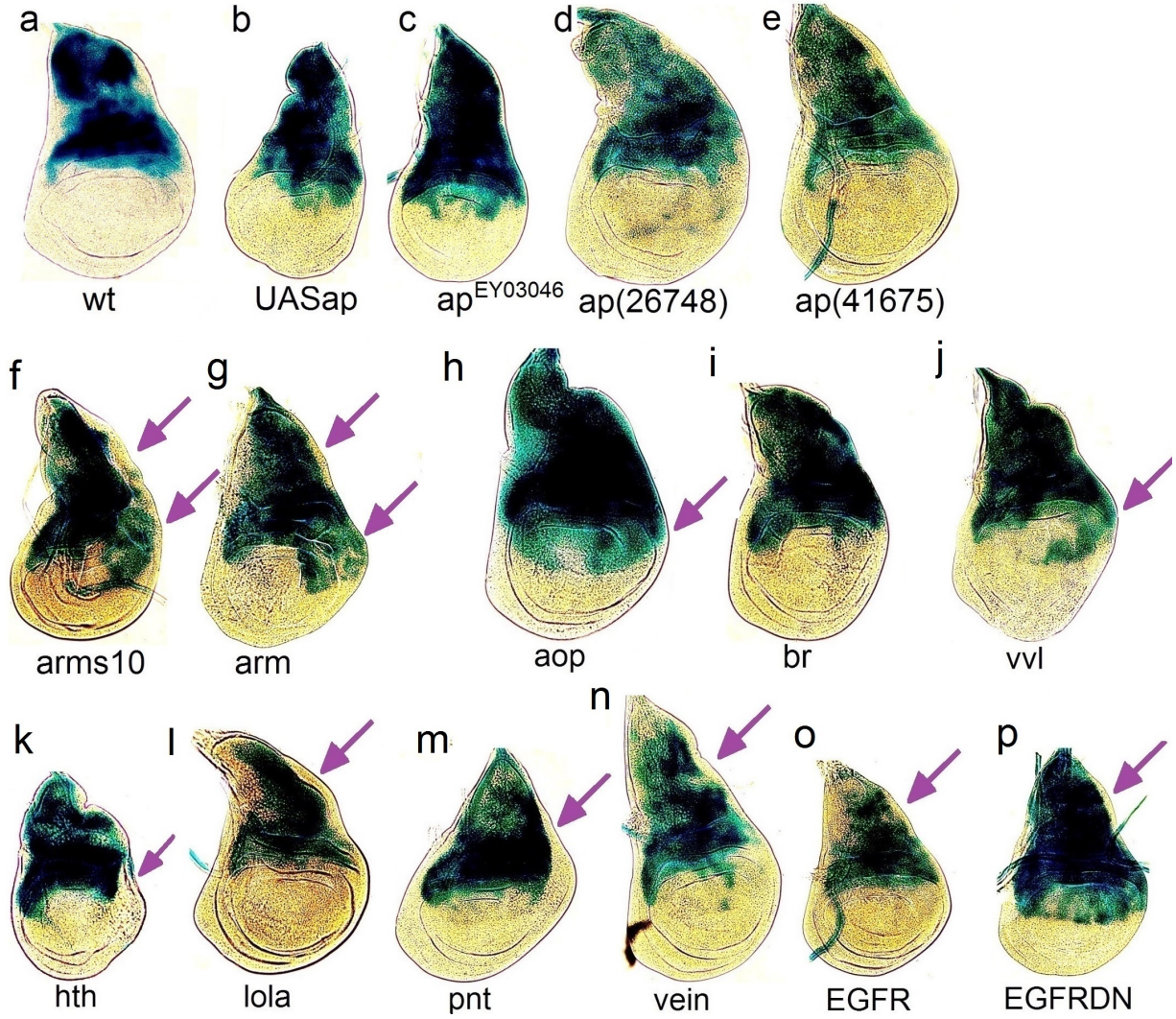


Figure 3.9: **X-Gal stainings of wing discs with knockdowns**
The arrows mark the places, where aberrations could be. (f), (g), maybe (h) and (j) show an increased activity of apE in the posterior part next to the wing pouch. (k)-(p) have areas with less dye.

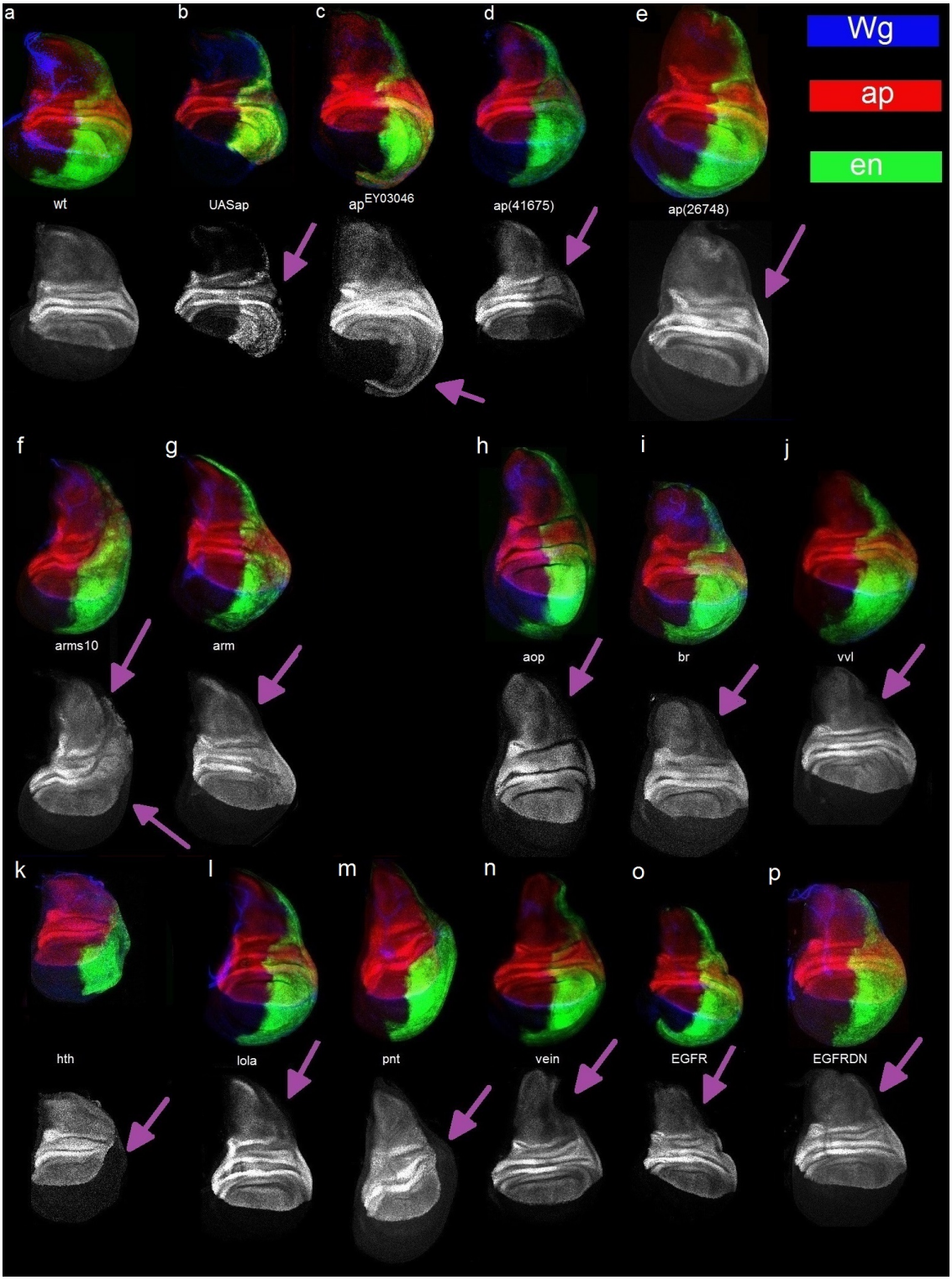


Figure 3.10: **Antibody stainings of wing discs with knockdowns**
The arrows mark the places, where aberrations could be. (i) shows a low concentration of ap in the wing pouch. (k), (m) and (n) show a significant loss of ap at different places.

4 Discussion

The program described in the bioinformatic part reduced the number of selected, possible interaction partners of the apE sequence tenfold. The resulting list was a good basis for the following assays.

The results of the LacZ constructs dealing with the m3 region lead to the conclusion, that the essential region for apE activation is in the 10 bp long m3.3 sequence. However, the surrounding regions m3.2 and m3.4 are not irrelevant. m3.2 and m3.4 are needed for the apE activity in the whole notum and parts of the wing pouch and the hinge. This spatial separation of activation and the dependency of the surrounding area leads to the conclusion, that more than one protein must bind at m3 for the optimal activation of apE. Besides, Pnr can be excluded as a candidate TF, because a deletion of m3.2 would be expected to lead to a more serious decrease of activity.

For the m1 region, similar conclusions can be drawn. Since the staining of the m1.3m4 deletion looks like the one in m4, it can be assumed, that sequences nearby m1.3 are not important. m1.2 seems to be the essential region. The m1.2m4 deletions show the same phenotype like m1m4. Nevertheless, an additional mechanism must be active in the ventral part of the wing pouch, or there is a total lack of the apE activator in the wing pouch, since apE activity is still repressed in these cells. m1.1 plays a minor part in the repression, however, it is necessary to repress apE activity in in the posterior part next to the wing pouch.

The unexpected activation pattern of m1m2m3 and m1m3m4 show, that there are still some unknown mechanisms, at work in the regulation of apE.

A tight control of *ap* expression is indispensable for a proper wing development. The in situ rescue system shows that even slight changes in apE activity have radical effects on the final wing blade structure. In addition the reentry experiments affirm the results of the X-Gal stainings.

Interpreting the results of the RNAi assay is difficult, because there can be several side effects, potential off target effects and fluctuations. Moreover, the *ap* controls figure 3.8(b)

and (c) demonstrate, that the reliability of the used RNAi lines is not ideal. Thus, no final statements can be made based on this data only. Nevertheless, the data seem to support the activating effect of *vn*, *EGFR* and *pnt* on *ap* [2] [72] [39]. The silencing of *EGFR* and *vn* always showed the same effect, a loss in the apE activity and the *ap* expression in the notum appeared (fig. 3.9 fig. 3.10 fig. 3.8). The effect of *pnt*, on the other hand, is way stronger. This fact is in line with the theory suggesting that *vn* activates *EGFR* in the dorsal compartment, which then activates *pnt*. *pnt* finally activates *ap* [2] [72] [39]. The strong repression of *ap* via *arm* (fig. 3.9) is consistent with the repression system in the ventral compartment consisting of wg and arm [2] [72] [39].

Even more striking results were obtained by silencing *hth* (fig. 3.9 fig. 3.10). The area of missing *ap* activity due to *hth* silencing in the posterior hinge is large when compared with the other phenotypes. However, the proposed binding site is m1.3, which does not give a phenotype when deleted. Hence the potential binding site is most likely not required.

The absence of *lola* causes a loss of *ap* expression in the notum (fig. 3.9 fig. 3.10). This is strange, since the predicted binding site for *lola* is m4, therefore no visible change is expected. However, this might be caused by side effects like interactions with other proteins. For better results, the assay should be refined, or repeated several times to reduce the fluctuation noise. A potential error source is a mechanism called transgression. Transgression is the movement of cells from a healthy to a damaged compartment [78]. This works particularly well in newly formed or still forming compartments. To avoid transgression, a driver line active in the whole wing disc should be used. *escargot* (esg) would be suitable for this purpose [77].

The comparison of figure 3.10(b) and figure 3.10(c) shows a sorting phenomenon. One the one hand, UAS-ap and $ap^{EY03046}$ both lead to an over-expression of *ap*. But their different insert locations sites of the UAS sequence lead to different starts and strength of over-expression. $ap^{EY03046}$ is a P-element with a UAS site inserted into the endogenous *ap* locus. UAS-ap, on the other hand, is a transgenic construct leading to a strong *ap* over-expression. This uneven over expressions are reflected in a different surface protein expression. The stronger *ap* over-expression leads to more *ap* specific surface proteins.

More similar surface proteins result in an adhesion force sticking the *ap* expressing cells together and thus sorting them. $ap^{EY03046}$ might have a low enough over expression to only change the cell identity, but not to trigger cell sorting.

5 Conclusion and outlook

The bioinformatic approach helped additionally to avoid days of work and provided a useful list of potentially important genes. The automatically collected information were a good start for further searches. Certainly, these in silico analysis are still too inaccurately to contribute precise solutions. Moreover, the writing and adjusting of the algorithm from scratch took a comparable amount of work to that of a normal search. However, the finished program can be adjusted for other projects within few minutes.

The classical genetic tasks also provided useful results. It was possible to narrow down the transcription factor binding site at m3 to 10 bp. However, the binding side at m1 is more complex. The most important sequence of m1 was defined down to 11 bp, but the surrounding sequences are also not insignificant.

The RNAi assay was problematic. The possibility of non-functional RNAi lines, lethality, transgression and other side effects made it difficult to interpret the results. More samples must be made to counteract the fluctuations by centering the results. Hereby the overall trends can better be detected. In addition, new approaches should be able to illuminate the subject from other angles. A new driver line, such as *escargot* (*esg*), which is expressed in whole wing disc [77], would prevent transgression. A pulsed knockdown at certain important times with regeneration time could be used to minimize side effects and lethality. Nevertheless, *hth*, *pnt*, *vein* and *lola* were not only found to activate the early *ap* enhancer, but also to be essential for the *ap* expression in certain parts of the wing disc. Moreover, an indication was found that *arm* does not directly bind to apE, but at least one more protein is involved in the inhibition process started by *wg*.

Acknowledgments

First and foremost I would like to thank Dr. Dimitri Bieli and Dr. Martin Müller for their help and supervision in the laboratories. Dr. Bieli found the ideal mixture between giving me freedom and motivating me, besides he provided all the explanations and protocols needed to accomplish this project in such a smooth way. Without Dr. Müller help and guidance concerning all fly related work I would have never been able to finish my master in the predefined time.

I truly thank Prof. Markus Affolter for giving me the opportunity to make my Master in his great lab. Besides I want to thank the whole lab for the nice atmosphere and the help.

Bibliography

[1] John Gerhart and Marc Kirschner. **Cells, Embryos, and Evolution: Towards a Cellular and Developmental Understanding of Phenotypic Variation and Evolutionary Adaptability**. *Blackwell book review*, **1999**.

[2] Dr. Dimitri Bieli. **PhD thesis**. *University Basel*, **2016**.

[3] Diaz-Benjumea, F.J., and S.M.Cohen. **Interaction between dorsal and ventral cells in the imaginal disc directs wing development in Drosophila**. *Cell*, 75, 741-752, **1993**.

[4] M. Mandaravally Madhavan, Howard A. Schneiderman. **Histological analysis of the dynamics of growth of imaginal discs and histoblast nests during the larval development of *Drosophila melanogaster***. *Wilhelm Roux' Archiv*, 183-269, Seite-Seite, **25.08.1977**.

[5] Fernando J. Diaz-Benjumea and Stephen M. Cohen. **Interaction between Dorsal and Ventral Cells in the Imaginal Disc Directs Wing Development in Drosophila**. *Cell*, 75, 741-752, **19.11.1993**.

[6] Dahmann C., Basler K. **Compartment boundaries: at the edge of development**. *Trends Genetics*, 15(8), **1999**.

[7] T. Xu, G.M. Rubin. **Analysis of genetic mosaics in developing and adult Drosophila tissues**. *Development*, 117: 1223-1237, **1993**.

[8] A. Garcia-Bellido, P. Ripoll, and G. Morata,. **Developmental compartmentalisation of the wing disk of Drosophila**. *Nat. New Biol.*, 245:251-253., **1973**.

[9] Dahmann C., Oates AC, Brand M. **Boundary formation and maintenance in tissue development**. *Nature reviews genetics*, 43-55, **2011**.

[10] Irvine K. D., Rauskolb C. **Boundaries in development: formation and function**. *Annu. Rev. Cell Dev. Biol.*, 17: 189–214, **2001**.

[11] T. Bollenbach, K. Kruse, P. Pantazis, M. Gonzalez-Gaitan, and F. Juelicher. **Robust Formation of Morphogen Gradients**. *Physical Review Letters*, 94, **2005**.

[12] C. Dahmann, and K. Basler. **Compartment boundaries : at the edge of development**. *Trends Genet.*, 15:320-326, **1999**.

[13] M. Zervas, S. Millet, S. Ahn, A. L. Joyner. **Cell behaviors and genetic lineages of the mesencephalon and rhombomere**. *Neuron*, 43:345-57, **2004**.

[14] L. M. Zeltser, C. W. Larsen. A. Lumsden. **A new developmental compartment in the forebrain regulated by Lunatic fringe**. *Nat. Neuroscience*, 4: 683-4, **2001**.

[15] M. Affolter1 and K. Basler. **The Decapentaplegic morphogen gradient: from pattern formation to growth regulation**. *Nature Reviews Genetics*, 8: 663-674, **2007**.

[16] F. J. Diaz-Benjumea, and S. M. Cohen. **Interaction between dorsal and ventral cells in the imaginal disc directs wing development in Drosophila**. *Cell*, 75:741-752, **1993**.

[17] C. J. Neumann, and S. M. Cohen. **Long-range action of Wingless organizes the dorsal-ventral axis of the Drosophila wing**. *Development*, 124: 871-880, **1997**.

[18] E. Batlle, D. G. Wilkinson. **Molecular mechanisms of cell segregation and boundary formation in development and tumorigenesis**. *Cold Spring Harb. Perspect. Biol.*, 4:a008227, **2012**.

[19] M. E. Stevens and P. J. Bryant. **Apparent genetic complexity generated by developmental thresholds: the apterous locus in Drosophila melanogaster**. *Genetics*, 110: 281-297, **1985**.

[20] B. Monier, A. Pelissier-Monier, A. H. Brand, B. Sanson. **An actomyosin-based barrier inhibits cell mixing at compartmental boundaries in Drosophila embryos**. *Nat. Cell Biol.*,10: 60-65, sup pp 1-9, **2010**.

[21] K. P. Landsberg, R. Farhadifar, J. Ranft, D. Umetsu, T. J. Widmann, T. Bittig, A. Said, F. Jülicher, C. Dahmann. **Increased cell bond tension governs cell sorting at the Drosophila anteroposterior compartment boundary.**. *Curr. Biol.*, 19: 1950-1955, **2009**.

[22] G. W. Brodland. **The Differential Interfacial Tension Hypothesis (DITH): a comprehensive theory for the self-rearrangement of embryonic cells and tissues**. *J. Biomech. Eng*, 124: 188-197, **2002**.

[23] S. Koshida, Y. Kishimoto, H. Ustumi, T. Shimizu, M. Furutani-Seiki, H. Kondoh, S. Takada. **Integrinα5-dependent fibronectin accumulation for maintenance of somite boundaries in zebrafish embryos**. *Dev. Cell*, 8: 587–98, **2005**.

[24] Seth S. Blair, Danny L. Brower, John B. Thomas and Michael Zavortink. **The role of apterous in the control of dorsoventral compartmentalization and PS integrin gene expression in the developing wing of Drosophila**. *Development*, 120: 1805-1815, **1994**.

[25] Marcus Michel, Maryam Aliee, Katrin Rudolf, Lisa Bialas, Frank Jülicher, Christian Dahmann. **The Selector Gene apterous and Notch Are Required to Locally Increase Mechanical Cell Bond Tension at the Drosophila Dorsoventral Compartment Boundary**. *PLOS ONE*, 11(8): e0161668, **2016**.

[26] M. Milan, U. Weihe, L. Perez, S. M. Cohen. **The LRR proteins capricious and Tartan mediate cell interactions during DV boundary formation in the Drosophila wing**. *Cell*, 106: 785–94, **2001**.

[27] S. S. Blair, D. L. Brower, J. B. Thomas, and M. Zavortink. **1994The role of apterous in the control of dorsoventral compartmentalization and PS integrin gene**

expression in the developing wing of Drosophila. *Development*, 120: 1805-1815, **1994**.

[28] B. M. Cohen, M. E. McGuffin, C. Pfeifle, D. Segal and S. M. Cohen. **apterous, a gene required for imaginal disc development in Drosophila encodes a member of the LIM family of developmental regulatory proteins**. *Genes Dev.*, 6: 715-729, **1992**.

[29] R. Feuerstein, X. Wang, D. Song, N. E. Cooke, and S. A. Liebhaber. **The LIM/double zinc-finger motif functions as a protein dimerization domain**. *Proc. Natl. Acad. Sci. U. S. A.*, 91:106559, **1994**.

[30] J. I. Pueyo, J. P. Couso. **Chip-mediated partnerships of the homeodomain proteins Bar and Aristaless with the LIM-HOM proteins Apterous and Lim1 regulate distal leg development**. *Development*, 131: 3107-3120, **2004**.

[31] R. Bronstein, L. Levkovitz, N. Yosef, M. Yanku, E. Ruppin et al. **Transcriptional regulation by CHIP/LDB complexes**. *PLos Genet.*, 6: e1001063, **2010**.

[32] A. Bachmann, E. Knust. **Positive and negative control of Serrate expression during early development of the Drosophila wing**. *Mech. Dev.*, 76: 67-78, **1998**.

[33] F. J. Diaz-Benjumea and S. M. Cohen. **Interaction between dorsal and ventral cells in the imaginal disc directs wing development in Drosophila**. *Cell*, 75: 741-751, **1993**.

[34] J. Kim, K. D. Irvine, and S. B. Carroll. **Cell recognition, signal induction, and symmetrical gene activation at the dorsal-ventral boundary of the developing Drosophila wing**. *Cell*, 82:795-802, **1995**.

[35] M. Milan, S. M. Cohen. **Temporal regulation of apterous activity during development of the Drosophila wing**. *Development*, 127: 3069-3078, **2000**.

[36] E. J. Rulifson, and S. S. Blair. **Notch regulates wingless expression and is not required for reception of the paracrine wingless signal during wing margin neurogenesis in Drosophila**. *Development*, 121: 2813-2824, **1995**.

[37] V. M. Panin, V. Papayannopoulos, R. Wilson, K. D. Irvine. **Fringe modulates Notch-ligand interactions**. *Nature*, 387: 908–12, **1997**.

[38] Katarzyna Oktaba, Luis Gutierrez, Julien Gagneur, Charles Girardot, Aditya K. Sengupta, Eileen E.M. Furlong, and Jürg Müller. **Dynamic Regulation by Polycomb Group Protein Complexes Controls Pattern Formation and the Cell Cycle in Drosophila**. *Developmental Cell*, 15: 877–889, **December 2008**.

[39] Dimitri Bieli, Oguz Kanca, David Requena, Fisun Hamaratoglu, Daryl Gohl, Paul Schedl, Markus Affolter, Matthew Slattery, Martin Müller, Carlos Estella. **Establishment of a Developmental Compartment Requires Interactions between Three Synergistic Cis-regulatory Modules**. *PLoS Genet*, 11(10): e1005376, **October 2015**.

[40] Leslie A. Pray. **Discovery of DNA Structure and Function: Watson and Crick**. *Nature Education*, 1(1):100, **2008**.

[41] George Johnson and Tai Te Wu. **Kabat Database and its applications: 30 years after the first variability plot**. *Nucleic Acids Research*, 28(1), 214-218, **01.01.2000**.

[42] Peter W. Rose, Andreas Prlic, Chunxiao Bi, Wolfgang F. Bluhm, Cole H. Christie, Shuchismita Dutta, Rachel Kramer Green, David S. Goodsell, John D. Westbrook, Jesse Woo, Jasmine Young, Christine Zardecki, Helen M. Berman, Philip E. Bourne and Stephen K. Burley. **The RCSB Protein Data Bank: views of structural biology for basic and applied research and education**. *Database issue*, 43 (D1): D345-D356, **2015**.

[43] Robert Stevens, Carole Goble, Patricia Baker and Andy Brass. **A classification of tasks in bioinformatics**. *Bioinformatics*, 17: 180-188, **2000**.

[44] Gordon E. Moore. **Cramming more components onto integrated circuits**. *Electronics*, 38(8), **19.04.1965**.

[45] MD Mark P. Sawicki, MD Ghassan Samara, MD Michael Hurwitz, MD Edward Passaro Jr. **Human Genome Project**. *The American Journal of Surgery*, 165: 258-264 , **1993**.

[46] Francis S. Collins, Michael Morgan, Aristides Patrinos. **The Human Genome Project: Lessons from Large-Scale Biology**. *Science*, 300: 286-290, **2003**.

[47] Holly K. Tabor, Benjamin E. Berkman, Sara Chandros Hull, Michael J. Bamshad. **Genomics really gets personal: How exome and whole genome sequencing challenge the ethical framework of human genetics research**. *American journal of medical genetics*, 155: 2916-2924, **2011**.

[48] Christian Gilissen, Jayne Y. Hehir-Kwa, Djie Tjwan Thung, Maartje van de Vorst, Bregje W. M. van Bon, Marjolein H. Willemsen, Michael Kwint, Irene M. Janssen, Alexander Hoischen, Annette Schenck, Richard Leach, Robert Klein, Rick Tearle, Tan Bo, Rolph Pfundt, Helger G. Yntema, Bert B. A. de Vries, Tjitske Kleefstra, Han G. Brunner, Lisenka E. L. M. Vissers and Joris A. Veltman. **Genome sequencing identifies major causes of severe intellectual disability**. *Nature*, 511: 344–347, **17 July 2014**.

[49] International Parkinson Disease Genomics Consortium. **Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies**. *Journal oder Verlag*, 377: 9766, 641-649, **25 February 2011**.

[50] S. Leonelli. **What difference does quantity make? On the epistemology of Big Data in biology**. *Big data and society*, 1-11, **2014**.

[51] Alexey A. Lagunin, Rajesh K. Goel, Dinesh Y. Gawande, Priynka Pahwa, Tatyana A. Gloriozova, Alexander V. Dmitriev, Sergey M. Ivanov, Anastassia V. Rudik, Varvara I.

Konova, Pavel V. Pogodin, Dmitry S. Druzhilovsky and Vladimir V. Poroikov. **Chemo- and bioinformatics resources for in silico drug discovery from medicinal plants beyond their traditional use: a critical review**. *Nat. Prod. Rep.*, 31: 1585-1611, **2014**.

[52] Esther Shein. **Python for beginners**. *Communications of the ACM*, 58: 19-21, **2015**.

[53] `https://www.python.org/about/apps/`, last access **05.01.2017**.

[54] Francis S. Collins, Ari Patrinos, Elke Jordan, Aravinda Chakravarti, Raymond Gesteland, LeRoy Walters, and the members of the DOE and NIH planning groups. **New Goals for the U.S. Human Genome Project: 1998-2003**. *Science*, 282(5389): 682-689, **1998**.

[55] George L. Sen, and Helen M. Blau. **A brief history of RNAi: the silence of the genes**. *The FASEB Journal*, 20(9): 1293-1299, **2006**.

[56] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello. **Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans**. *Nature*, 391: 806-811, **1998**.

[57] S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon. **An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells**. *Nature*, 404: 293-296, **2000**.

[58] Michael Boutros, Amy A. Kiger, et al. **Genome-Wide RNAi Analysis of Growth and Viability in Drosophila Cells**. *Science*, 303 (5659), 832-835, **2004**.

[59] Xavier Belles. **Beyond Drosophila: RNAi In Vivo and Functional Genomics in Insects**. *Annual Review of Entomology*, Volume (Ausgabe), 111-128, **2010**.

[60] Jian-Quan Ni, Rui Zhou, Benjamin Czech, Lu-Ping Liu, Laura Holderbaum, Donghui Yang-Zhou, Hye-Seok Shim, Rong Tao, Dominik Handler, Phillip Karpowicz, Richard Binari, Matthew Booker, Julius Brennecke, Lizabeth A. Perkins, Gregory J. Hannon

and Norbert Perrimon. **A genome-scale shRNA resource for transgenic RNAi in Drosophila**. *Nature Methods*, 8: 405-407, **2011**.

[61] Jian-Quan Ni, Michele Markstein, Richard Binari, Barret Pfeiffer, Lu-Ping Liu, Christians Villalta, Matthew Booker, Lizabeth Perkins, and Norbert Perrimon. **Vector and parameters for targeted transgenic RNA interference in Drosophila melanogaster**. *Nature Methods*, 5(1): 49-51, **2008**.

[62] N. Perrimon, Jian-Quan Ni, and L. Perkins. **In vivo RNAi: Today and Tomorrow**. *Cold Spring Harb Perspect Bio*, 2:a003640, **2010**.

[63] H. Luan, N. C. Peabody, C. R. Vinson, and B. H. White. **Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression**. *Neuron*, 52(3): 425-436, **2006**.

[64] A. H. Brand, N. Perrimon. **Targeted gene expression as a means of altering cell fates and generating dominant phenotypes**. *Development*, 118(2): 401-.415, **1993**.

[65] M. L. Suster, L. Seugnet, M. Bate, and M. B. Sokolowski. **Refining GAL4-driven transgene expression in Drosophila with a GAL80 enhancer-trap**. *Genesis*, 39(4): 240-245, **2004**.

[66] I. Herskowitz. **Functional inactivation of genes by dominant negative mutations**. *Nature*, 329: 219-222, **1987**.

[67] A. GARCIA-BELLIDO and P. RIPOLL. **The number of genes in Drosophila melanogaster**. *Nature*, 273: 399-400, **1978**.

[68] A. C. Spradling, D. M. Stern, I. Kisst, J. Rootet, T. Laverty, and G. M. Rubin. **Gene disruptions using P transposable elements: An integral component of the Drosophila genome project**. *Proc. Natl. Acad. Sci.*, 92: 10824-10830,, **1995**.

[69] A. J. Griffiths. **An introduction to genetic analysis**. *Macmillan*, **2005**.

[70] J. Bischof, R. K. Maeda, M. Hediger, F. Karch, and K. Basler. **An optimized transgenesis system for Drosophila using germ-line-specific φC31 integrases** . *Proc. Natl. Acad. Sci. U. S. A.*, 104(9): 3312-3317, **2006**.

[71] A. C. Groth, M. Fish, R. Nusse, and M. P. Carlos. **Construction of transgenic Drosophila by using the site-specific integrase from phage φC31**. *Genetics*, 166: 1775-1782., **2004**.

[72] D. Bieli, O. Kanca, D. Gohl, A. Denes, P. Schedl, M. Affolter, M. Müller. **The Drosophila melanogaster Mutants apblot and apXasta Affect an Essential apterous Wing Enhancer**. *G3 Genes Genomes Genetics*, 5(6): 1129-1143, **2015**.

[73] D. Kennell, H. Riezman. **Transcription and translation initiation frequencies of the Escherichia coli lac operon**. *Journal of Molecular Biology*, 114(1): 1-21, **July 1977**.

[74] T. Inada, K. Kimata, H. Aiba. **Mechanism responsible for glucose-lactose diauxie in Escherichia coli: challenge to the cAMP model**. *Genes Cells*, 1(3): 293-301, **1996**.

[75] Julius Edgar Lilienfeld. **Method and apparatus for controlling electric current**. *Canadian Intellectual Poperty Office*, CA272437 (A), **22.10.1925**.

[76] W. J. Gehring, Y. Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A. F. Schier, D. Resendez-Perez, M. Affolter, G. Otting, K. Wütherich. **Homeodomain-DNA recognition**. *Cell*, 78: 211-223, **1994**.

[77] S. Hayashi, S. Hirose, T. Metcalfe, A.D. Shirras. **Control of imaginal cell development by the escargot gene of Drosophila**. *Development*, 118, 105-115, **1993**.

[78] S. C. Herrera, G. Morata. **Transgressions of compartment boundaries and cell reprogramming during regeneration in Drosophila**. *Life*, 3:e01831., **2014**.