Identification and functional analysis of Hox downstream genes in *Drosophila*

der Fakultät für Biologie
der EBERHARD KARLS UNIVERSITÄT TÜBINGEN

zur Erlangung des akademischen Grades eines Doktors
der Naturwissenschaften

von

Stefanie Diana Hueber
aus
Frankfurt am Main

vorgelegte
D i s s e r t a t i o n
2009

Tag der mündlichen Prüfung:	17.03.2009
Dekan:	Prof. Dr. Hanspeter A. Mallot
1. Berichterstatter:	Prof. Dr. Detlef Weigel
2. Berichterstatter:	Prof. Dr. Rolf Reuter
3. Berichterstatter:	Prof. Dr. Friedrich Schöffl
4. Berichterstatter:	Prof. Dr. Hanspeter A. Mallot

Curriculum vitae

***Name*:** Stefanie Diana Hueber
Date of Birth: 06/26/1978
Place of Birth: Frankfurt/Main, Germany
Contact Address: Max-Planck-Institute for Developmental Biology,
Spemannstr. 37-39 / VI, 72076, Tübingen, Germany
Contact Phone: [49] (7071) 6011415
E-mail: Stefanie.Hueber@tuebingen.mpg.de

Education:

2004 – 2009	Ph.D., "Identification and functional analysis of Hox downstream genes in *Drosophila melanogaster*", Ingrid Lohmann group, Max-Planck-Institute for Developmental Biology, Tübingen, Germany.
2002 – 2003	Diploma, "Regulation of Regulators of G-Protein Signaling (RGS) Proteins in Yeast", Henrik G. Dohlman group, Department of Biochemistry and Biophysics, University of North Carolina, USA Grade: 1,5 (very good)
2001 – 2003	Continuing Biology (Diploma), University of Dresden, Germany Majors: Molecular Genetics Minors: Microbiology, Biochemistry
2000 – 2001	Continuing Education, Biology, Ohio State University, Columbus, USA Dresdner Bank Scholarship
1999 – 2000	Begin of Biology (Diploma), University of Dresden, Germany
1997 – 1999	"Prediploma": top 10%
1997	Abitur (University Entrance Exam): Schiller-Gymnasium, Frankfurt/Main, Germany Grade: 1,5 (very good)
1995 – 1997	High School: Schiller-Gymnasium, Frankfurt/Main, Germany
1994 – 1995	11th grade, Roedean-School, Brighton, UK

Professional experience:

2004 – 2009	Ph.D., "Identification and functional analysis of Hox downstream genes in *Drosophila melanogaster*", Ingrid Lohmann group, Max-Planck-Institute for Developmental Biology, Tübingen, Germany
2002 – 2003	Diploma, "Regulation of "Regulators of G-Protein Signaling (RGS) Proteins in Yeast", Henrik G. Dohlman group, Department of Biochemistry and Biophysics, The University of North Carolina, Chapel Hill, USA
2001, 09-12	Paid practical, "Stress response of the RGS protein Sst2p in

	Saccharomyces cerevisiae", Henrik G. Dohlman group, Department of Biochemistry and Biophysics, The University of North Carolina, Chapel Hill, USA
2001	Student Research Assistant, H.-O. Park group, Ohio-State-University, Columbus, USA
1999	Student Research Assistant, Department of Yeast Microbiology, University of Dresden, Germany

Skills:

Microscopy: light, LSM and SEM
Microarray-technology (Affymetrix)
Biochemical, cell- and molecular-biological methods

Languages:

German and English (fluent), French (moderate)

Other interests:

badminton, alpine skiing, aquaristics, voyaging, diving, climbing.

Prizes and Awards:

1st prize Biology-Olympics, 1997, Hessen, Germany
Dresdner Bank Scholarship, 2000-2001, Dresden, Germany

Presentations at conferences:

Poster-presentation, Crossroads in Biology, Cologne, Germany, 2007
Talk, 12th Regional Drosophila Meeting, Vienna, Austria, 2006
Poster-presentation, Santa-Cruz Conference on Developmental Biology, Santa-Cruz, USA, 2006
Poster-presentation, Biannual Meeting of the German Society for Developmental Biology, Göttingen, Germany, 2004

Publications:

Multifactorial Regulation of a Hox Target Gene
PLOS Genetics 5 (3), e1000412 (2009)
Petra Stöbe, Matthias A. S. Stein, Anette Habring-Müller, Daniela Bezdan, Aurelia Fuchs, Stefanie D. Hueber, Haijia Wu and Ingrid Lohmann

Shaping segments: Hox gene function in the genomic age
BioEssays 30, 965-979 (2008)
Stefanie D. Hueber and Ingrid Lohmann

Comparative analysis of Hox downstream genes in *Drosophila*
Development 134, 381-392 (2007)
Stefanie D. Hueber, Daniela Bezdan, Stefan R. Henz, Martina Blank, Haijia Wu and Ingrid Lohmann

Preface

The thesis has been organized in a manner that recapitulates key aspects of Hox protein research and presents the relevant findings to enable:

a) The audience to follow the content of this thesis regardless of whether they have an in-depth knowledge of *Drosophila melanogaster* development or Hox protein function.

b) The significance of the research to be understood without the necessity to also read the peer-reviewed publications that underlie this thesis.

Abstract (Deutsch)

Eine Vielzahl unterschiedlicher morphologischer Strukturen ist im Verlauf der Evolution entstanden. Eine signifikante Anzahl dieser Strukturen benötigt die Expression einer hoch konservierten Gen-Familie, der Hox Gene. Diese Hox Gene kodieren Transkriptionsfaktoren und werden in unterschiedlichen Regionen entlang der anterior-posterior Körperachse eines Embryos exprimiert. Die Anwesenheit oder Abwesenheit spezifischer Hox Proteine bestimmt, welche Art von Strukturen in der entsprechenden Region entstehen. Der Mechanismus, mit dem Hox Transkriptionsfaktoren die Entwicklung solch unterschiedlicher Strukturen bewirken, ist bis heute noch nicht bekannt, obwohl sich Wissenschaftler nunmehr bereits seit Jahrzehnten mit dieser Frage beschäftigen. Die Ursache dafür ist die Eigenschaft von Hox Proteinen *in vitro* an fast identische DNA-Sequenzen zu binden.

Zur Zeit werden zwei Theorien diskutiert, auf welche Weise Hox Proteine *in vivo* die unterschiedliche Entwicklung morphologischer Strukturen auslösen. Eine Theorie geht davon aus, dass Hox Proteine einige wenige Gene, so genannte “Schlüssel”-Gene, spezifisch regulieren. Eine andere Theorie nimmt an, dass Hox Proteine die Expression vieler verschiedener Gene auf einmal regulieren. Leichte Veränderungen in der Expression dieser Gene würden akkumulieren und so einen größeren Effekt bewirken. Dadurch würde eine Erklärung für die großflächigen Effekte von Hox Proteinen auf morphologische Strukturen geliefert.

Um einen fundierten Beitrag zur Klärung dieser Streitfrage zu leisten, haben wir eine quantitative Analyse durchgeführt, welche die Auswirkungen der Überexpression der unterschiedlichen Hox Proteine auf die Genregulation vergleicht. Hierzu haben wir die embryonalen Stadien 11 und 12 von *Drosophila melanogaster* untersucht. Zudem haben wir ein bioinformatisches Programm entwickelt, um Gene zu identifizieren, die von dem Hox Protein **Dfd** reguliert werden. Mit Hilfe der ermittelten Daten haben wir den Grad, in dem sich Hox Proteine in der Regulation der Gene unterscheiden, determiniert und konnten zeigen, dass Hox Proteine nicht, wie zunächst angenommen, primär die Expression weiterer Transkriptionsfaktoren regulieren, sondern eine Vielzahl unterschiedlicher Gen-Klassen.

Abstract (English)

A wide variety of morphological features have arisen over the course of evolution of bilateral organisms. A significant number of these features depend on the expression of Hox genes, a set of highly conserved genes encoding transcription factors. Different Hox genes are expressed along the anterior-posterior axis of the embryo. The presence or absence of specific Hox transcription factors in a region determines what kind of morphological structures will form. The mechanism by which Hox transcription factors control the development of such vastly different structures has eluded researchers for decades, especially in light of their ability to bind to nearly identical DNA-sequences *in vitro*.

This has sparked a debate over how Hox proteins effect the development of different structures. It was proposed that a select few target genes were differentially regulated by Hox proteins, thereby leading to the observed morphological differences. Alternatively, it was proposed that Hox proteins influence the expression of many downstream genes at a time. Many minor changes in expression would then accumulate to a greater effect and explain how Hox proteins achieve the major effects on morphology observed along the body axis.

To elucidate the mechanism of Hox protein function, we carried out a comparative analysis of Hox downstream gene regulation at a quantitative level. *Drosophila melanogaster* embryonic stages 11 and 12 were examined and the number of downstream genes under Hox protein control was determined in these two stages. We then used the set of Hox downstream genes to create a bioinformatic tool for direct target gene prediction for the Hox protein **Dfd**. With the data collected, we were able to establish the degree to which Hox proteins differ in the regulation of their downstream genes and were able to show that Hox transcription factors act on a variety of different downstream genes. The latter result is a significant finding, as, in contrast to previous views, our results indicate that Hox proteins do not primarily act on other transcription factors.

Table of Contents

Curriculum vitae ..... **i**
Preface ..... **iv**
Abstract (Deutsch) ..... **v**
Abstract (English) ..... **vi**
Table of Contents ..... **vii**
1 Introduction **1**
1.1 Development of Drosophila melanogaster 1
1.2 Evolution of Hox clusters 2
1.3 Hox function 5
1.3.1 Hox molecular function 5
1.3.2 Hox molecular specificity 6
1.3.3 Hox developmental function 15
1.4 Hox downstream genes 18
1.4.1 known Hox downstream genes 18
1.4.2 Methods for identification of Hox downstream genes 21
1.5 Aim of the Ph.D. thesis 24
1.6 Experimental Approach 27
2 Results and Discussion **32**
2.1 Data generation and validation 32
2.1.1 Validation of over-expression 32
2.1.2 Downstream gene validation 33
2.1.3 Direct target gene prediction for Dfd 34
2.2 Data analysis 35
2.2.1 Number of Hox downstream genes 35
2.2.2 Number of direct Hox targets 36
2.2.3 Context dependence 38
2.2.4 Specificity of Hox downstream gene regulation 42
2.2.5 Classification of Hox downstream genes 45
2.2.6 Hox regulatory networks 46
2.3 Hox dependent morphogenesis 48
3 Conclusion **52**
4 Outlook **57**
Contributions **61**
Appendix ..... **I**
Tables: I
Table 1: Identified direct Hox target genes I
Table 2: large-scale Hox downstream gene identification III
Bibliography ..... **IV**
Publications ..... **XVIII**

1 Introduction

1.1 Development of *Drosophila melanogaster*

A multitude of strategies evolved to subdivide and pattern the bodies of multicellular organisms and allow the formation of different body structures. All bilateral animals examined to date use a highly conserved set of genes, the Hox genes, to pattern their body along the anterior-posterior axis. Hox genes are expressed at specific, partially overlapping regions along the anterior-posterior body axis. Each body region experiences a unique Hox protein composition, defining a unique positional identity[1,2,3]. In regards to Hox protein function, the fruit fly *Drosophila melanogaster* is one of the best analysed model organisms. Its development is, compared to other species, fairly well understood. For this reason, the emphasis in the subsequent sections of the manuscript will predominantly be placed on the Hox genes as key players in establishing developmental differences along the anterior-posterior axis in the model organism *Drosophila melanogaster*.

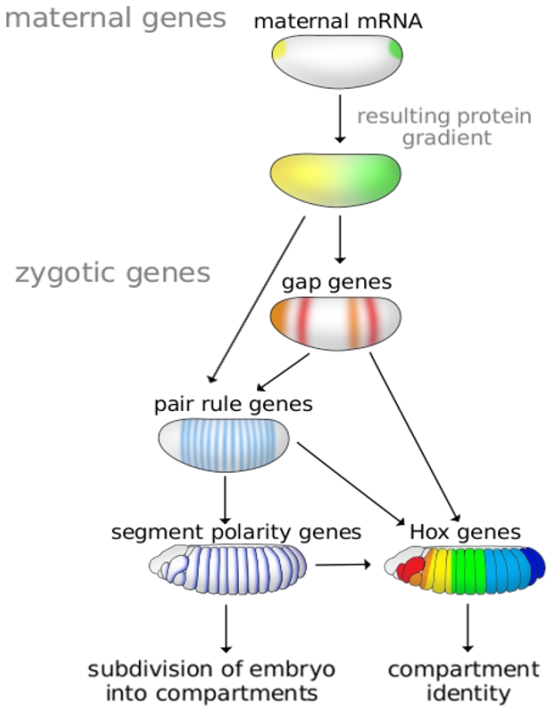

Figure 1: Development of *Drosophila melanogaster.* Schematic representation of segmentation genes, adapted from Developmental Biology 8th edition by Gilbert[4].

The first step in the development of a *Drosophila melanogaster* embryo is to divide its main body axis into serially repeating units (***Figure 1***). The primary positional information for patterning of the embryo is provided by the mother. When producing the egg, the required information is placed, in form of mRNA, in different regions of the egg. Shortly after fertilization this maternal mRNA is translated into protein[4]. These proteins then diffuse from the point of their translation and create a gradient across the embryo. The combination of these proteins as well as their relative concentration controls the expression of so called zygotic segmentation genes, encoded by the embryonic genome[5]. Gap genes are the first set of zygotic genes to be expressed. Their transcription starts approximately two hours post-fertilization and is used to determine the first

broad regions within the embryo[6]. Further subdivision of these regions into smaller sub-sections is carried out by the pair-rule genes, a second set of zygotic segmentation genes[6]. Interestingly, the positions of alternating subsections are defined by different genes, e.g. *even-skipped* (*eve*) for seven odd-numbered and *futshi-tarazu* (*ftz*) for seven even-numbered sections of the embryo[7,8]. The resulting fourteen stripes of expression define the parasegments[9]. The border between one parasegment and the next will later correspond to the central area of morphological boundaries called segments. These segments correspond to the body segments present, later on, in the adult fly. Each segment is further subdivided into an anterior and a posterior region by segment polarity genes[6], a set of genes transcribed at the same relative position within each segment.

Having established the number of segments and their anterior-posterior polarity, the zygotic genes (gap genes, pair-rule and segment polarity genes) act together to regulate the expression of genes that, in contrast to further subdividing the embryo, now determine the developmental identity of the established segments. These are the Hox genes[10,11] (***Figure 1***).

1.2 Evolution of Hox clusters

Hox genes encode transcription factors that provide the cells with information about their location along the anterior-posterior axis[12,13] (***Figure 2***). Hox transcription factors are shared among all bilateral animals and are used to establish morphological diversity between different body segments[10,11]. Remarkably, although orthologous Hox proteins produce considerably different morphological structures in different species, they are conserved to a degree that these proteins can be interchanged between species and still lead to the development of a morphologically and functionally correct structure[14,15].

Hox proteins have been defined as a set of homeodomain containing transcription factors whose genes are located in a cluster on the chromosome[16]. It is believed that the prototypic Hox gene cluster evolved from a single Hox gene via tandem duplication and subsequent divergence, as the homeodomains of individual Hox proteins usually exhibit greater similarity to orthologous homeodomains in other species than to adjacent genes within their own Hox cluster[17]. The ancestral vertebrate Hox cluster is believed to have contained 14 genes[18,19] and is probably most closely resembled by the cluster of Hox genes present in Amphioxus (*Branchiostoma lanceolatum*) (***Figure 2***). The ancestral Hox cluster is thought to have

duplicated at least twice in mammals, as evidenced by four separate Hox clusters, comprising a total of 39 genes, being present in many mammalian genomes (***Figure 2***)[20].

Determining the specific function of individual Hox genes in vertebrate model organisms has been complicated by the presence of multiple and partially redundant Hox clusters. The vertebrate model organisms mouse (*Mus musculus*) and zebrafish (*Danio rerio*), for example, have four and seven Hox clusters respectively[20]. The seven Hox clusters present in *Danio rerio* are thought to have arisen by an additional Hox cluster duplication or whole genome duplication in Osteichthyes (bony fish). Consequently, some phenotypes can only be observed when multiple, paralogous, Hox genes are removed at the same time. A close relative to vertebrates, *Branchiostoma lanceolatum*, provides a potential alternative model organism. Because this organism contains only a single Hox cluster[20], it is a promising model for analysis of Hox dependent phenotypes, as it should allow easier manipulation of these genes as well as observation of their effects on development. While reproduction of *Branchiostoma lanceolatum* under laboratory conditions was achieved only recently, genetic manipulation, though possible, has proven difficult (personal communication, Peter Holland). However promising, these difficulties make this organism not quite suitable to be used as a fully-fledged model organism for functional analysis of Hox proteins. Another model organism, *Drosophila*, is much more distantly related to vertebrates. *Drosophila* has a single Hox cluster comprising 8 Hox genes (***Figure 2***). The presence of only one Hox cluster in a bilateral animal that can be easily manipulated and its presence in large numbers across a wide variety of laboratories, may explain why the first Hox gene mutations were observed in *Drosophila*[21] and, why *Drosophila* has

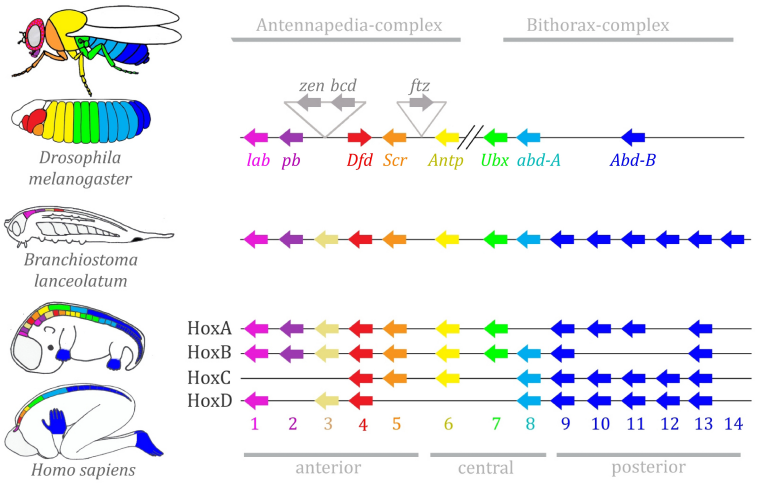

Figure 2: Evolution of *Hox* clusters. Schematic representation of *Drosophila melanogaster*, *Branchiostoma lanceolatum*, and *Homo sapiens* orthologs. The orthologs for the *Drosophila* genes *Antp*, *Ubx* and *abd-A* and vertebrate Hox genes *6* to *8* have not yet been determined, although the diagram does represent the order in which they are located on the chromosome and the pattern of their expression within the organisms.

proven to be an excellent model organism in which to perform Hox gene research.

In *Drosophila*, the Hox cluster is located on the third chromosome and can be divided into two sub-clusters (***Figure 2***): The first sub-cluster, the Antennapedia-complex, specifies the head and the anterior part of the thorax. It consists of the genes *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*) and *Antennapedia* (*Antp*). The second sub-cluster, the Bithorax-complex, specifies the posterior part of the thorax and the abdomen. It comprises of *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*)[21,22,23,24]. Although these two sub-clusters are usually regarded as one Hox cluster, it has rightly been argued that there is a considerable distance between the two sub-clusters in the genome[20]. In addition, the Antennapedia-complex is interrupted by other homeobox containing genes, probably derived from Hox genes[19]. Whether viewed as a single cluster or two, the genes within the complex are positioned in the same order as their orthologs in the vertebrate clusters and, with the exception of *Dfd*, are aligned in the same orientation on the chromosome. Furthermore, the *Drosophila* Hox genes, just as their vertebrate counterparts, are expressed along the anterior-posterior axis in correspondence with their relative position on the chromosome (spatial collinearity). More importantly, a number of Hox genes have been shown to be functionally equivalent between vertebrates and *Drosophila*[14,15].

Various questions arise considering the high level of conservation of Hox genes and the apparent differences in morphology they determine:

1) What are the molecular functions of Hox transcription factors?
2) What morphologies are determined by Hox proteins along the anterior-posterior axis?
3) How do different Hox proteins, despite their high conservation at the molecular level, achieve their different developmental effects?
4) What are the downstream genes responsible for Hox dependent developmental effects?
5) Are any of the former, Hox molecular and developmental function, Hox specificity and Hox downstream genes conserved across organisms and, if so, to what degree?

1.3 Hox function

1.3.1 Hox molecular function

Transcription factors are defined as proteins exerting their function by binding DNA and activating or repressing a target gene[4]. Hox proteins are a special set of helix-turn-helix transcription factors[4]. In Hox proteins, the helix-turn-helix motif is located in a region of about 60 highly conserved amino acids, the homeodomain, which forms a well-defined three dimensional structure that binds DNA[27,28] (***Figure 3***). The helix-turn-helix motif is formed by helix 2 and 3. A third helix, helix 1, stabilizes the helix-turn-helix motif. The most conserved part of the homeodomain is helix 3. Helix 3, together with the N-terminal arm of the homeodomain (residues 5 and 3), recognizes the cis-regulatory Hox-binding motif on the DNA, consisting of the core sequence TAAT[17,29,30,31,32,33,34].

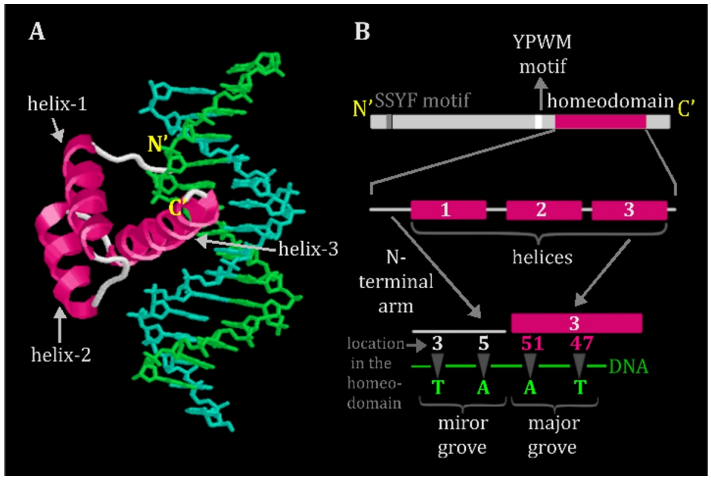

Figure 3: Molecular structure of Hox proteins. A structure of the Hox protein DNA-binding domain, the homeodomain, shown here for **Antp** (PDB "9ant") in RasMol[25] and **B** schematic representation of Hox protein domains, lower part adapted from Svingen and Tonissen, 2006[26].

While the DNA-binding domain within Hox proteins is extremely well defined, a domain indisputably responsible for activation or repression of transcription has not been identified. Some Hox proteins contain a conserved sequence region towards the N-terminus called the SSYF or NSYF motif. For the Hox protein **Ubx**, this motif was shown to be able to increase the transcription of a target gene[35]. Other regions towards the N-terminus of Hox proteins were also found to be conserved across subsets of Hox proteins and appear to have similar effects[36].

Apart from the homeodomain, the only other part conserved across nearly all Hox proteins is the hexapeptide (I)**YPWM**(K)[37] located N-terminally to the homeodomain. Only the most posterior Hox protein, **Abd-B** and its vertebrate homologs (paralogous groups 9–13) do not possess a canonical YPWM motif. Instead, only the core tryptophane (W) is conserved and is thus referred to as the W-motif. The YPWM motif is assumed to primarily function as a region allowing interaction between a Hox protein and a TALE (three amino acid loop extension) class homeodomain protein such as **Exd**[38]. The region in-between

the hexapeptide and the homeodomain is called “linker” and can vary in length between Hox sequence homologs (3-53 amino acids)[39].

1.3.2 Hox molecular specificity

Homeodomain containing proteins generally bind to a core TAAT sequence on the DNA. A subclass of these proteins contains a glutamine residue at position 50 within the homeodomain. This class is referred to as Q50 homeodomain proteins. Q50 homeodomain proteins, such as Hox proteins, **Eve**, **Ftz** and **Engrailed** (**En**), are involved in patterning the anterior-posterior axis[31]. The glutamine at position 50 makes direct contact with nucleotides just 5' of the TAAT core sequence, leading to virtually indistinguishable DNA-binding specificities for members of this group. Across these Q50 homeodomain proteins, the binding affinity to DNA-sequences differs up to five fold and the dissociation constants usually vary between 10^{-9} and 10^{-10} M[17,31].

In general, the affinities of this group of proteins to binding sites on the DNA are very similar, but even more so for Hox proteins[40,41]. It is therefore unlikely that Hox proteins can differentially regulate a given target gene on their own. This, of course, raises the question regarding how Hox proteins manage to specify different developmental cell fates along the anterior-posterior axis.

The frequency of the short TAAT recognition DNA-binding site, occurring on average every 128 bp, cannot explain how Hox proteins can regulate some genes, but do not regulate other genes. While it is theoretically possible, it is unlikely that all genes are regulated by all Hox proteins directly. How Hox proteins differentiate between genes or whether they do, is not known. If they do, are only a limited number of DNA sites bound by Hox proteins or do Hox proteins bind to all available TAAT sequences *in vivo*? If the former is the case, how do Hox proteins differentiate between the various TAAT sites in the genome? Alternatively, if Hox proteins bind to all TAAT sequences *in vivo*, are all these binding sites functional, in other words, does binding of a Hox protein result in a significant transcriptional change? Then, what properties determine whether a DNA site bound by a Hox protein is functional or not?

Two main hypotheses have been proposed for how structural differences between Hox proteins can account for their functional effects[42].

A) Additional factors determine whether the Hox protein binds to a specific stretch of DNA *in vivo* and thereby whether a gene is regulated or not. The binding of a Hox protein to a specific stretch of DNA in concert with these

factors is therefore the direct cause for the effect on transcription of a target gene. The binding to DNA itself, should be sufficient to predict whether, or how, the transcription of a potential target gene will be affected.

B) Additional factors determine whether the Hox protein activates or represses target gene transcription. The ability of a Hox protein to bind to a particular stretch of DNA is necessary, but not sufficient to predict whether, or how, the transcription of a potential target gene will be affected.

Both hypotheses provide possible explanations for how differential regulation is achieved between Hox proteins and for how some genes are regulated and others are not. I will first focus on how differential expression can be achieved between different Hox proteins via modification of Hox DNA-binding affinity or DNA-binding sequence recognition.

A) Hox DNA-binding affinity and specificity

1. Hox monomers
2. Clusters of Hox monomer sites
3. Hox homodimers
4. Hox heterodimers (co-selective binding/cooperation)

1. Hox monomer: Provided Hox protein monomers are capable of binding to DNA sites in a specific manner, significant differences in their regions responsible for DNA-interaction would have to be present. However, the Hox DNA-binding domain, the homeodomain, is highly conserved and structurally nearly identical across different Hox proteins. Correspondingly, *in vitro* studies have shown that Hox proteins can bind to very similar DNA sequences[17,29,31,32,33,34]. The short TAAT consensus sequence to which basically all Hox proteins can bind, implies that Hox proteins could potentially bind to every gene and regulate it. As it appears unlikely that every single gene is regulated by every single homeodomain protein, Hox monomers (***Figure 4***) per se cannot account for a) only a limited number of genes being under Hox protein control or b) different sets of genes being under control of different Hox proteins.

2. Clusters of Hox monomer sites: A more selective choice of Hox protein targets may be achieved by an accumulation of Hox monomer binding sites on enhancers (***Figure 4***). Such accumulations or clusters occur less often in the genome than Hox monomer sites and this can increase the likelihood of Hox proteins binding to that site. A study showed that the Hox

protein **Ubx** could bind to multiple sites within the same enhancer[31]. Every Hox-binding site within the *spalt major 328* enhancer was able to confer some degree of repression. For complete repression, at least three **Ubx**-binding sites were required[43]. Over the course of evolution, the additive effect of multiple Hox monomer binding sites is likely to have resulted in the formation of clusters of Hox-binding sites when a stronger degree of regulation was advantageous. This may explain how, from thousands of potential binding sites, only a few appear to be regulated in a detectable manner. This potentially makes Hox downstream gene regulation more selective (fewer targets), as genes with only one Hox-binding site in their regulatory element may not be regulated by Hox proteins in a detectable manner. Although potentially capable of accounting for the limited number of genes under Hox protein control, the occurrence of clusters of DNA sites binding Hox proteins requires additional mechanisms to achieve differential regulation between different Hox transcription factors.

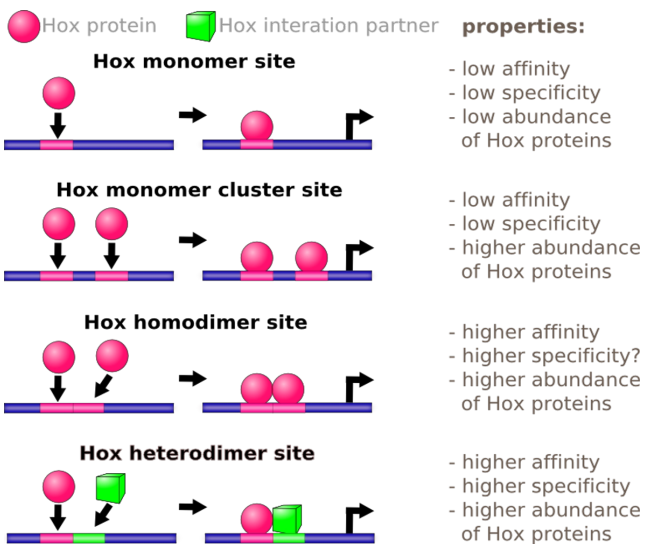

Figure 4: Hox-DNA interaction models. Four different mechanisms of Hox-DNA interactions are depicted. The corresponding properties are listed on the right side. Hypothetical properties are marked with a "?".

3. Hox homodimer: Dimerization (homo- and heterodimerization) appears to be relatively common for transcription factors[44] and the occurrence of Hox DNA-binding site clusters may allow for Hox homodimerization on the DNA. Another homeodomain protein **Paired** (**Prd**), for example, was shown to be able to form homodimers[45]. A further three homeodomain proteins (**Mat alpha**[46], **Eve**[47], **En**[48], and **Pbx3**, a vertebrate homolog of **Hth**[49]) were shown to bind to DNA sites located so close to one another on the DNA, that it is possible they interact with one another and, potentially, form homodimers. As such, it is conceivable that Hox proteins may form homodimers (***Figure 4***). Depending on the strength of the protein-protein and protein-DNA interaction, Hox proteins may first interact to form homodimers and subsequently bind to the DNA or, alternatively, first bind to the DNA independent of one another and thereafter form more stable homodimers. In both cases, protein-protein interactions are capable of

bringing additional amino acid sequences into close proximity to the DNA strand and, thereby, potentially stabilize the protein-complex on the DNA. If such amino acids differ between Hox proteins, they may account for differential binding to certain DNA sequences. Various evidence is present that points to the formation of Hox homodimers: electro mobility shift assays (EMSA) not specifically designed for detection of Hox homodimers, showed two distinct bands in Hox protein-only lanes[49,50.51]. This is sometimes interpreted as the presence of Hox homodimers[50]. In addition, Beachy et al.[52] provided evidence for interaction between Hox proteins. Their study focused on **Ubx** cluster sites and showed that a **Ubx** protein binding to one part of the DNA could stabilize the binding of another **Ubx** protein to a different DNA site *in vitro*. Furthermore, such cluster sites resulted in a synergistic (non-linear) effect on the transcription of a target gene *in vivo*. This synergistic effect can be regarded as a DNA-binding affinity enhancing cooperative effect of a Ubx homodimer or multimer. Strikingly, this remains the only study regarding Hox-Hox homodimer/multimer interactions and, although it was carried out 15 years ago, only three papers in Hox *Drosophila* research[35,43,53] have referred to this paper since the year 2000.

4. Hox heterodimers - co-selective binding / cooperative DNA-binding: The so-called co-selective binding model or cooperative DNA-binding model is the only model supported by experimental evidence showing that Hox proteins can achieve large differences in their binding affinity to DNA[38,54,55,56,57,58] (***Figure 4***). These differences in DNA-binding affinity can explain the differential regulation of downstream genes for different Hox proteins. In this model, the interaction of a Hox protein with an additional factor influences the ability of the multimer to bind to DNA. To explain the differences in DNA-binding affinity observed for Hox heterodimers, the same factor would have to interact in different manners with different Hox proteins. In other words, this interaction requires certain parts to differ between Hox proteins. The interaction is therefore likely to occur via other residues than those present in the third helix of the Hox protein, which is nearly identical between Hox proteins and shown to be responsible for their basic DNA-binding capabilities.

The best-studied example for co-selective binding model or cooperative DNA-binding model is the TALE class homeodomain transcription factor **Extradenticle (Exd)**. The resulting Hox heterodimer has increased specificity to a Hox heterodimer composite response element. A composite response element contains closely situated binding sites, usually adjacent or overlapping, for different transcription factors. Proteins located on the same

composite response elements are thought to interact with each other[59,60]. Being in a complex with another transcription factor such as **Exd**, can have three effects on Hox DNA-binding compared to Hox monomers:

1) Extension of the DNA-recognition site from 6 bp to 10 base pairs[54], leading to a reduction in the number of sites bound by such a complex within the genome and, thereby, to fewer genes being affected by the Hox/**Exd** complex.
2) Increased DNA-binding affinity or stability of the complex[55,56]
3) Increased differential DNA-binding property: While some DNA sites are bound by different Hox/**Exd** complexes with more or less the same affinity[55], other DNA sites are bound with much higher affinity by a certain Hox/**Exd** complexes[38,54,55,56,57,58]. This feature thereby accounts for different genes being regulated by different Hox/**Exd** complexes.

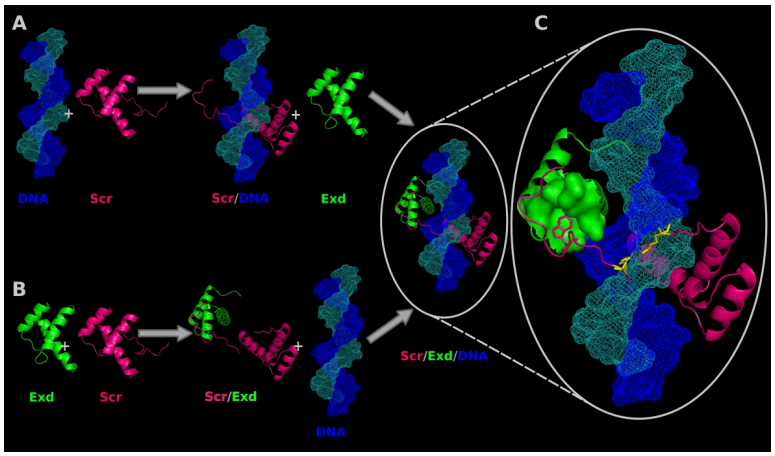

Figure 5: Current DNA-binding specificity enhancing cooperation model. Structure of **Scr**/**Exd** on the *forkhead 250* enhancer[38] created in PyMol[61] . In **A, B** and **C** the DNA is shown in "mesh" structure in dark and light blue. **Scr** and **Exd** are shown in cartoon structures in pink and green respectively. In **C**, the "surface" of the hydrophobic pocket within the homeodomain of **Exd** is shown to highlight its interaction with the amino acid tryptophane of the **Scr** YPWM motif, shown in "stick" structure. The amino acids responsible for "specific" binding of **Scr** (as compared to other Hox proteins) within the minor groove of the DNA in the *forkhead 250* enhancer, arginine 3 and histidine -12, are shown in yellow in "stick" structure. A part of the linker region of the **Scr** protein is not shown as no structural information was available for these residues.

An explanation for how differential binding of Hox proteins is achieved on a molecular level was provided in a recent publication[38]. In the proposed scenario, Hox proteins are first recruited to AT rich regions by the highly conserved homeodomain helix 3, which recognizes such regions within the major groove of the DNA helix. The resulting binding is thought to occur with high affinity, but low specificity. Either once bound to the DNA (***Figure 5A***) or before DNA-binding (***Figure 5B***), the Hox protein and the TALE class homeodomain protein **Exd** interact with each other. This is thought to occur mainly via the tryptophane of the YPMW motif of the Hox protein binding to a hydrophobic pocket within the homeodomain of **Exd**[62] (***Figure 5C***). In the example shown, the Hox protein **Scr** and the protein **Exd** bind at opposite

sides of the DNA on the *forkhead 250* (*fkh*) enhancer. The N-terminal arm of the Hox homeodomain and the linker region is stretched, bringing both regions in close proximity to the DNA minor groove. Both regions of the Hox protein can interact with the minor groove and, potentially, further stabilize the binding of the complex to DNA (***Figure 5C***). As the N-terminal arm of the homeodomain and linker region differ between Hox proteins, the contacts made by these residues to the DNA minor groove are thought to confer a degree of DNA-binding specificity to individual Hox/**Exd** complexes[38,57].

The reality, of course, is more complex than can be explained by DNA-binding specificity enhancing cooperation of Hox and **Exd** alone. First, the Hox-DNA sequence recognition may be modified by formation of higher order complexes. This has been shown for the *Drosophila* homologs of Hox and **Exd** which can form heterotrimers in conjunction with another TALE class homeodomain protein, **Homothorax** (**Hth**) in vertebrates[63]. In *Drosophila*, **Exd** and **Hth** were shown to act in conjunction with Hox proteins[64,65]. Second, Hox-DNA-binding specificity may be enhanced or modified by Hox interacting proteins other than **Exd**, as the YPWM motif may have the potential to interact with any homeodomain and, in particular, with any TALE-class homeodomain protein. More importantly, Hox proteins are capable of creating specific Hox dependent morphological structures in the absence of **Exd** and **Hth**. The interaction between Hox proteins **Exd** and **Hth** can, therefore, not account for all of the observed morphological structures/effects[11].

The search for further Hox interaction partners has, so far, led to the identification of a number of partners[11,66], that do not seem to increase DNA-binding specificity. This includes proteins, such as **Bip2**, which binds to **Antp** via the YPWM motif[67], or **Tsh**, which binds to **Scr**. **Tsh**, in conjunction with **Scr**, determines the identity of the first thoracic segment[68]. Neither of these Hox interacting factors appear to enhance DNA-binding specificity, but they may confer differential regulation of genes via a different mechanism.

B) Hox transcriptional activity

The following models are explained in more detail to establish a uniform nomenclature used throughout this thesis.

1. Additive/Subtractive
2. Combinatorial and Collaborative
3. Cooperation (Reverse cooperation)
4. Modification of Hox proteins

A model that can account for differential regulation of downstream genes by Hox proteins in the absence of factors enhancing the DNA-binding specificity is the “activity regulation” model or “wide-spread binding” model[31,69]. The idea underlying this model is that the effect of a DNA-bound Hox protein on transcription can be modified by cofactors or other factors. Differential regulation of downstream genes between Hox proteins may be achieved by different homeo-proteins interacting with these factors in different manners. Thus, in the “activity regulation” model, the specificity of Hox protein-protein interactions, but not the Hox DNA-binding specificity, explains the differences in morphology induced by Hox proteins[31].

The term “wide-spread binding” model will not be used as it is slightly misleading. Instead, the term “activity regulation” model will be used. The term “wide-spread binding” implies that the ability of Hox proteins to bind to many stretches of DNA is a prerequisite of the model or an underlying feature. This is not the case. The model allows Hox proteins to bind to many and identical sites on the DNA and still achieve differential regulation between Hox proteins by regulation of transcriptional activity instead of the DNA-binding affinity. It also allows Hox proteins to bind to many sites, but potentially regulate only a limited set of genes. However, factors that bind to relatively rare and specific sequences such as Hox/**Exd** complexes, can be affected by “activity regulation” as well. For example, a **Lab/Exd** responsive gene, can be switched to be responsive to **Dfd/Exd** simply by adding a 21 bp sequence to its promoter region[70]. Thus, sites on the DNA strand that are separate and distinct from a homeodomain-binding site, can

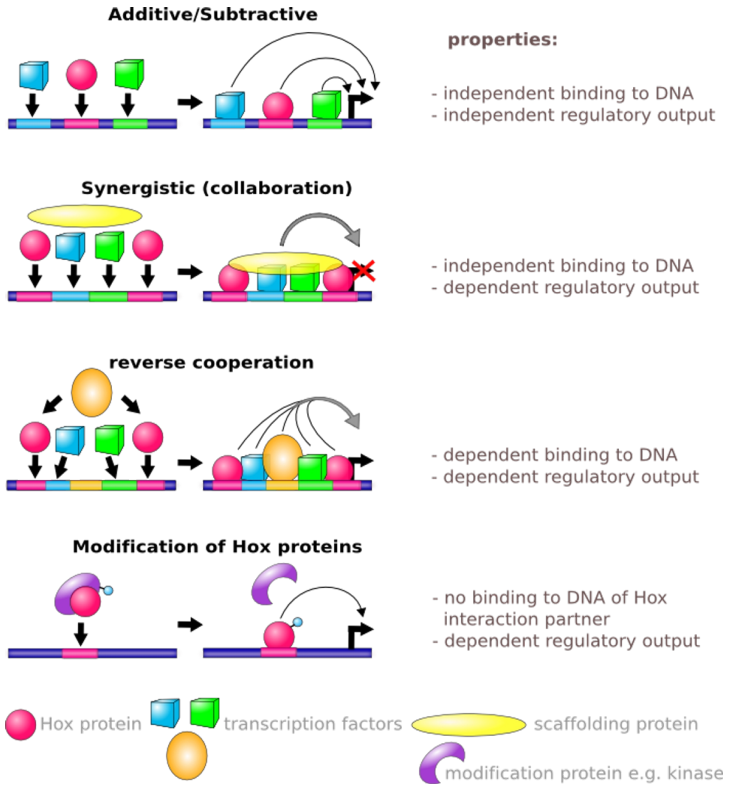

Figure 6: Hox activity regulation models. Four different mechanisms are depicted of how Hox proteins can achieve a transcriptional output on an enhancer, that do not modify the DNA-binding affinity or specificity of the Hox protein itself.

modify the regulatory effect a Hox response element will have. It therefore appears that both unspecific Hox monomers as well as highly specific Hox/**Exd** heterodimers, are subject to “activity regulation”.

In the following, four models are described by which a Hox/**Exd** complex or a Hox monomer can achieve either spatial restriction or differential regulation of target gene expression: the additive/subtractive model, the synergistic combinatorial model, the synergistic cooperation (reverse cooperation) model and modification of Hox proteins such as phosphorylation.

1. Additive/Subtractive: transcription factors bind independently of each other to the DNA. Once bound to DNA, they jointly determine the final expression pattern of a gene. The key premise in this model is that each transcription factor acts in parallel to the others. Some effects may be additive, where one activating factor A slightly enhances transcription on its own, another transcription factor B does so as well and their combined effect results in a stronger enhancement (***Figure 6***). Other effects may be exclusive, e.g. a repressive factor C that binds to the same enhancer and by inducing conformational changes in the DNA, inhibits either binding of one or both activating factors A and B. This model should be viewed as a general mechanism applicable to all transcription factors.

This additive/subtractive model can easily account for the spatially and temporally restricted expression patterns of Hox target genes as temporally or spatially restricted factors may bind to the enhancer of a given Hox target. As the factors in this model neither directly nor indirectly interact with Hox proteins, the model cannot account for the differential expression of downstream genes observed between Hox proteins.

In contrast to the additive/subtractive model above, the following models describe synergistic effects.

2. Combinatorial and Collaborative: In the combinatorial model, similar to the model above, transcription factors bind to DNA independently of Hox proteins. In contrast to the above model, the combined effect of the transcription factors cannot be explained by additive and subtractive effects alone. Some kind of direct or indirect interaction has to take place between the various factors[71,72,73]. A good example for this is the regulation of *wingless* (*wg*) expression in the gut[74]. Binding of **Mothers against dpp** (**Mad**)/**Medea** (**Med**) and **Abd-A** to the enhancer is absolutely essential for *wg* expression. If any of the three is missing, no *wg* is expressed[71].

A specific case of the combinatorial model is the collaboration model as defined by Walsh et al. 2007[72]. By their definition, the Hox protein and at

least one other transcription factor have to bind in close proximity to one another to the DNA without direct cooperative interaction occurring between the factors. A DNA region to which transcription factors bind in close proximity, adjacent or overlapping, is often referred to as a composite response element[60]. The key point in the collaboration model is that transcription factors binding to the same composite element are thought to act as one unit and the distance between the DNA-binding sites is crucial. As a result, the mode of expression (activation or repression) and area of Hox action is determined by the entire unit while the individual components on their own may have no, only minor, or even a completely opposite effect. Experimental evidence in support for the collaboration model was provided when the developmental differences between wing and haltere were analysed. Two **Ubx** sites and two **Smad** sites are located within one composite element on the enhancer of the Hox target gene *spalt major* (*salm*). While each transcription factor binds to its DNA target site independently of the other factor, the presence of each factor in the same composite response element and the distance of each individual transcription factor site to the other is significant for *salm* repression. Interestingly, no interaction between **Ubx** and **Smad** proteins was detected[72]. A simple explanation for the requirement of these binding sites to be in close proximity to each other is that **Ubx** and **Smads** proteins collaborate in the recruitment of an additional factor, the co-repressor **Schnurri**. **Schnurri** is also required for *salm* repression and known to interact with **Mad**[75]. If this model is correct, the Hox protein **Ubx** would interact with **Mad** and **Med** indirectly via **Schnurri** as visualized in ***Figure 6***. Since repression of *salm* occurs only in the haltere disc and not in the wing disc, **Antp** apparently does not fulfil the requirements for *salm* repression, possibly by failing to interact with and therefore failing to recruit the unknown repressor.

3. Cooperation (Reverse cooperation): As mentioned in chapter 1.3.2a, cooperation enhances the DNA-binding specificity of either or both of the interaction partners significantly. For DNA-binding specificity enhancing cooperation to be regarded as part of the “activity regulation” model, one condition has to be fulfilled: the DNA-binding affinity of the Hox protein cannot be affected significantly by the interaction partner. As the model described in the following is effectively the inverse of the co-selective binding or DNA-binding specificity enhancing cooperation model and to distinguish it from this kind of cooperation, I have termed it “reverse cooperation”. In the “reverse cooperation” model, the DNA-binding affinity of the Hox interaction partner is altered. The segment polarity protein **En**, for example, barely binds to the *Dll* repressor element on its own. In conjunction with **Exd** and **Hth**, two

Hox proteins, **Ubx** and **Abd-A**, are able to enhance the affinity of **En** to the *Dll* repressor (***Figure 6***). Although not yet analysed at a molecular level, this model can potentially account for differential transcriptional output between Hox proteins, because another Hox protein, **Antp**, does not seem to contain the structural requirements for **En** recruitment as it does not enhance **En** binding to this site, and does not repress *Dll*[66].

4. Modification of Hox proteins: This model requires the presence of a factor with the ability to modify a Hox protein, for example, a factor with the ability to phosophorylate or myristylate the Hox protein. It does not require the factor to be bound to DNA. To enable differential function of a Hox protein, the modification site has to be specific to the Hox protein or a subset of Hox proteins. Phosphorylation sites of Hox proteins have been shown to be able to switch Hox proteins from repressors to activators[76]. The above mentioned ability of the Hox protein **Ubx** to repress *Dll*, for example, can be inhibited by phosphorylation of **Ubx**[77] (***Figure 6***). The precise mechanism by which this change in activity occurs has not yet been determined.

All of the above models enabling differential expression between Hox proteins (combinatorial model, reverse cooperation model and modification of Hox proteins) can be used by the same Hox protein on the same enhancer. This was shown by the example of **Ubx** dependent repression of *Dll*, necessary for repression of appendage formation. **Ubx** requires a specific state of modification (unphosphorylated) to repress *Dll*. Furthermore, **Ubx** is capable of enhancing the binding of another transcription factor, **En**, thereby leading to *Dll* repression. Finally, binding of the **Ubx** protein to DNA is enhanced by cooperation of **Ubx** with **Exd** and **Hth**. Using these three mechanisms assures that **Ubx** executes one of its developmental functions only when the correct factors are present in conjunction at the right place at the right time.

1.3.3 Hox developmental function

The first conclusive analysis of Hox mutant phenotypes was carried out in *Drosophila* by Ed Lewis in 1978[21]. Major morphological malformations were observed when the expression of Hox genes was altered. A reoccurring phenotype were homeotic transformations, a term used to describe the transformation of one structure to resemble another, homologous structure present in the body[80]. The two most prominent examples are the four winged fly and a fly with an extra pair of legs on its head. In the case of the four winged fly, the expression of the Hox gene *Ubx* is decreased. Under normal

circumstances, a wild type fly will form two wings and two balancing organs, the halteres. Due to the phenomenon of posterior repression[25], the absence of the **Ubx** protein causes the expression domain of the **Antp** Hox protein to be extended to encompass the third thoracic segment[71]. As a result, the cells usually forming the halteres develop into a second set of wings. Formation of legs on the head of a fly, in contrast, is caused by ectopic expression of the thoracic Hox gene *Antp* in the head (***Figure 7A***). When expressed in the head, **Antp** alters the developmental program of the cells forming antenna (an olfactory sensory organ) to form a pair of legs in their place[83].

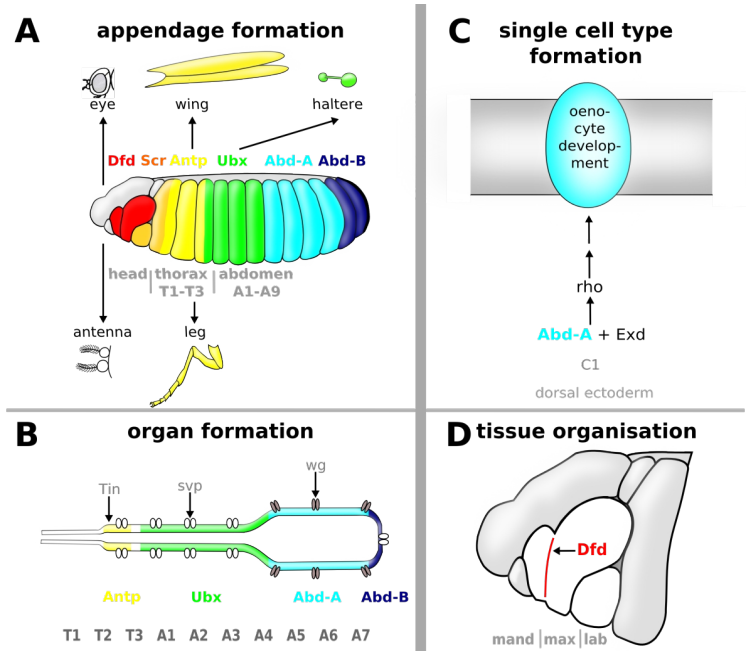

Figure 7: Hox developmental functions. A appendage formation **B** organ formation, the heart **C** specification of a single cell type, the oenocyte. **D** boundary maintenance between the mandibular and maxillary segment. B adapted from Perrin et al., 2004[78] and C adapted from Brodu et al., 2002[79].

As Hox proteins are capable of switching the development of one type of appendage to another, scientists started to speculate about what kind of appendage would form in absence of Hox proteins. This is referred to as "ground state"[82,83,84]. While much progress has been made in Hox protein research to identify which Hox protein is responsible for which appendage (see ***Figure 7A***), two common misconceptions may have hampered research in this field.

The first misconception relates to the idea of which ventrally derived appendage would form in the absence of Hox proteins. The literature repeatedly states, that antennae* are formed in absence of Hox proteins[84,85,86] and often refers to a supposedly complete *Hox* knock-out in *Tribolium*[84,86]. The original paper actually states, that they did not remove the Hox gene homolog of *proboscipedia* or *labial*, called *maxillopedia*[82] and, consequently,

* the antennae are ventrally derived and only located dorsally after head involution in Drosophila development.

does not represent a true *Hox* null mutant. Another argument, stating that the eye-antennal disc is outside the Hox expression domain, thus antennae form without Hox input[84], is also invalid. *pb*, together with other the Hox genes *lab*, *Dfd* and *Scr* is expressed in the outer rim of the eye-antennal disc[87] (for details see larval stages, ***Figure 15***). Thus, a Hox independent state or "ground state" for ventral appendage formation has not yet been determined.

The second misconception relates to wing formation which is regarded as the Hox independent appendage or "ground state" for dorsal appendage formation[88]. Early research showed that wings are malformed in the absence of the **Antp** protein[89,90]. Yet, ectopic expression of **Antp** does not induce wing formation[88]. As such it was thought that the phenotype "wing to haltere" in *Antp* mutants was due to an expanded expression of **Ubx** in that region. Consequently, it was believed that **Antp** acts as "place holder" to prevent the expression of other Hox proteins, which would induce fates other than the supposed "ground state" wing. Recent evidence, however, suggests, that **Antp** is capable of inducing wing formation in place of eyes in the head, but this induction depends on the use of the right promotor[67,91]. Similarly to the discussion in the last paragraph, the argument that no Hox gene is expressed in the wing disc is also invalid as **Antp** is expressed in the outer rim of the wing disc[87]. It is therefore quite possible that **Antp** may actively induce wing formation in wild type flies.

Due to the large area affected, appendage formation is the most noticeable and hence the best described effect of Hox proteins. However, less obvious morphological effects are also determined by Hox proteins, including organ formation, time-frame dependent generation and maintenance of tissue-formations as well as determination of cell-type specific differentiation. An example for Hox dependent organ formation is the development of the heart (***Figure 7B***). The *Drosophila* heart tube can be subdivided into two morphologically and functionally distinct areas, the anterior aorta and the posterior heart. Again, the expression pattern of the Hox genes correspond to their relative position on the chromosome and determines the polarity of the *Drosophila* heart tube in the mesoderm. **Antp** and **Ubx** are expressed in the aorta, where they are required for normal aorta formation in the most posterior part. **Abd-A** is expressed in the heart where it is required and sufficient for heart formation and **Abd-B** is expressed in the four most posterior heart cells, where it determines the end of the heart and represses cardioblast formation[78,92,93,94] (***Figure 7B***).

An even less noticeable Hox dependent phenotype is the determination of

the identity of a single cell type, the oenocyte. The oenocyte is a large secretory cell below the abdominal epidermis of the larvae. Its differentiation is governed by the Hox protein **Abd-A** and cannot be substituted by the closely related Hox protein **Ubx**[79] (***Figure 7C***). Another such example is the organization of the tissue between the mandibular and maxillary head lobes in the embryo. To maintain the boundary between the two segments, the Hox protein **Dfd** induces highly localized apoptosis along this border[95] (***Figure 7D***).

There are various other known developmental functions determined by Hox proteins, e.g. determination of the positional identity in various processes of hair formation[21,96], formation of the salivary glands[97], the gut[98,99], the peripheral and central nervous system[100,101], and morphogenesis of the tracheal pits and the spiracles (the *Drosophila* breathing apparatus)[21,102]. Most likely more Hox developmental functions exist in *Drosophila* that remain to be discovered. Angiogenesis, for example, is known to be under Hox protein control in vertebrates[103,104] and potentially could also be under Hox protein control in *Drosophila*. In summary, Hox genes are required for an enormous amount of features specific to individual segments varying form large and visible phenotypes, such as wing formation, to the formation of single cell types, such as the oenocyte.

1.4 Hox downstream genes

1.4.1 known Hox downstream genes

Knowledge about the Hox protein itself, its binding sites and the Hox downstream genes is required to understand how the positional information is provided by Hox proteins and translated into a morphological effect.

So far, the Hox mediated information regarding which appendage should be formed, appears to be due to relatively few key direct Hox target genes. These key targets include a remarkable number of transcription factors (***Appendix Table 1***). The *Distal-less* (*Dll*) gene is one of those key downstream genes and its expression is required for the development of ventral appendage or appendage-like structures (including anal plates which are derivatives of the genital discs)[106,107,108]. In the abdomen, expression of *Dll* is repressed by the Hox transcription factors **Ubx**, **Abd-A** and **Abd-B** and, as a result, no appendages are formed[40]. In contrast to the abdomen specific Hox proteins, Hox transcription factors expressed in the head and thorax, do not repress *Dll* expression - at least not in the amount or time frame required to repress appendage formation. Formation of ventrally derived appendages

such as legs and antennae depends on the co-expression of *Dll* with *hth* in the appendage primordium. If **Dll**, but no **Hth** is present, legs will form. If **Dll** and **Hth** are co-expressed in the same cells, antennae will form[109,110]. Nearly all Hox proteins repress *hth* in *Dll* expressing cells[43] (***Figure 8***). Only in the eye antennal disc, Hox genes are not expressed in sufficient amounts to repress *hth* in *Dll* expressing cells and thus, only the eye-antennal disc forms antennae. Therefore, only two Hox downstream genes, *Dll* and *hth* seem to be required to distinguish between the antennal, leg and "no appendage" fate in an appendage primordium.

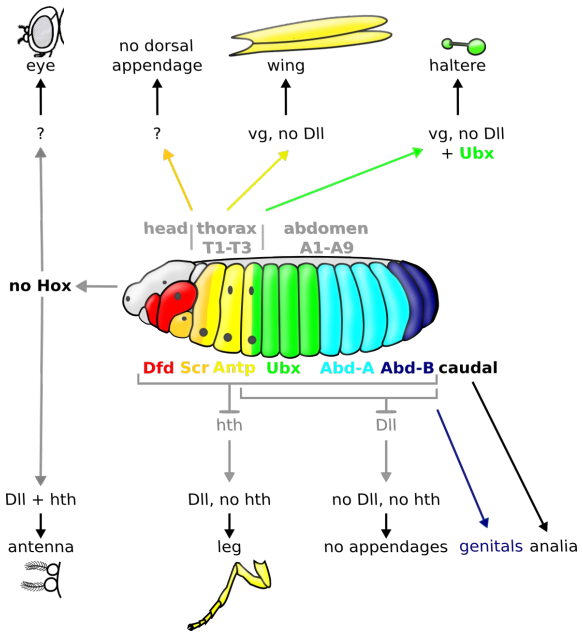

Figure 8: "Key" Hox downstream genes in appendage formation. "no Hox" is put in quotation marks as the eye-antennal disc does express Hox genes in the outer rim[87] and although these appendages are assumed to form independent of Hox gene expression, experimental validation of this assumption is still missing. The homeotic gene *caudal* is shown here, as it contains a homebox and appears to confer a Hox-like function by defining the identity of a segment[105]. *Caudal*, however, it is not part of the *Hox* cluster.

How Hox protein govern which type of dorsal appendage to form is not well understood (***Figure 8***). Although it is known that the Hox protein **Antp** induces wing formation[67] and **Scr** suppresses wing formation[111], the responsible downstream genes are not known. For the **Ubx** dependent switch from wing to haltere, potential key downstream genes were identified. For example, the transcription factor gene *vestigial* (*vg*) is required for wing formation and repressed by **Ubx**[112,113]. Expression of *vg* in the haltere, however, is only capable of transforming the haltere into a wing-like structure to a limited degree[112,114] and indicates that additional Hox dependent factors required for the formation of the wing remain to be identified.

Hox proteins were termed master regulators due to the tremendous effect mis-expression of these proteins has on morphological structures along the body axis. As master regulators, Hox proteins were thought to regulate other molecules at the top of regulatory cascades, mainly other transcription factors[2,114,42]. In regards to appendage formation, Hox transcription factors seem to act by regulating other transcription factors and thereby determining

the type of appendage that will form. As a consequence, certain Hox dependent loss-of-function phenotypes can easily be rescued, provided the key downstream transcription factors are known. How these key regulators themselves achieve the Hox dependent morphological effects via their own downstream genes remains largely elusive. One of the major causes is that only a few Hox downstream genes have been identified and even fewer that directly affect morphogenesis. Yet, to understand the differentiation of segments, Hox proteins need to be linked to the genes implementing differential morphogenesis. Such morphology implementing proteins have been termed realisators which, by definition, directly control processes such as cell division, cell death, cell shape, cell-cell adhesion and cell migration[115]. Ideally, such realisators are located within a regulatory network that only leads to one morphological output as opposed to molecules at the top of signalling cascades. Molecules at the top of signalling cascades enhance a signal leading to transcription of a wide variety of genes, including transcription factors and, thereby, exhibit many different effects. Consequently, to identify the missing links between Hox transcription factors and their morphological outputs, it would be helpful to identify Hox downstream realisator genes. The development of the posterior spiracle was successfully linked to a number of realisator genes[116]. The Hox protein **Abd-B** controls at least four genes within a network of a wide variety of morphology-shaping genes. Simultaneous expression of the four downstream genes, three transcription factors and a signalling molecule, proved to be sufficient to induce expression of several realisator genes coding for proteins involved in cell adhesion, cell polarity and cell shape, which led to the development of posterior-spiracle like structures in the absence of **Abd-B**.

So far, basically all Hox downstream genes linked to a morphological effect either code for transcription factors or signalling molecules. It is possible that Hox transcription factors primarily control the expression of other transcription factors directly, however, the previously mentioned maintenance of the mandibular and maxillary segment boundary points to the contrary. The maintenance of this segment boundary is due to the Hox protein **Dfd**, which directly binds to the enhancer and induces the expression of the gene *reaper* (*rpr*)[95]. The apoptosis promoting protein **Rpr** maintains the segment boundary by removing cells[117]. Therefore *rpr* is an example for a realisator gene directly controlled by a Hox protein.

To identify the missing links between Hox proteins and their morphological outputs, either via transcription factors or direct control of realisator genes,

and to understand how developmental specificity of Hox proteins is achieved, we require more information about Hox downstream genes.

1.4.2 Methods for identification of Hox downstream genes

Identification of more downstream genes is one way to gain more information about Hox downstream genes and a prerequisite for more detailed analysis of their features. Numerous approaches were used to identify Hox downstream genes before the advent of large-scale analyses[118,119,120,121,122]. These approaches have greatly advanced our understanding of the DNA-binding properties of Hox proteins. However, a greater number than the 24 direct target genes identified are needed to further advance our knowledge in this area. The limited number of identified downstream genes is largely due to the rather unspecific manner in which Hox proteins bind to DNA *in vitro*[17,29,31,32,33,34]. This poor DNA-binding specificity has made it difficult to identify downstream genes using *in vitro* techniques. Possible reasons for the lack of DNA-binding specificity observed *in vitro* could be the absence of Hox-protein interacting factors that would be present *in vivo*. This includes factors such as the Hox-binding protein **Exd** known to modify Hox DNA-binding specificity. Alternatively, Hox proteins may bind relatively unspecific to DNA *in vivo* and differential target gene regulation is achieved by differences in the interaction with further proteins that modify transcriptional activity. In either case, the additional factors would be present *in vivo*. The relevance of such additional factors is highlighted by a variety of *in vivo* studies that were carried out and, accordingly, proved to be more successful.

The candidate gene approach is an example for an *in vivo* approach that was used to identify Hox downstream genes. Candidate genes are genes considered likely to be Hox downstream genes either because their expression pattern corresponded to the Hox expression pattern or because their developmental function corresponded to a developmental function under Hox protein control. Since the candidate gene approach was first used, an enormous number of gene expression profiles have been made available thanks to the Berkeley *Drosophila* Genome Project (BGDP). The BGDP carried out a systematic approach to determine patterns of gene expression in *Drosophila* embryos by high-throughput RNA *in situ* hybridization[123]. The number of genes with expression profiles overlapping those of Hox genes has increased correspondingly. Using conventional methods, this increase has made it difficult to analyse all these genes within a reasonable time-frame. Consequently, matching expression patterns to identify further candidate

genes, nowadays, can only be used in conjunction with additional methods. Furthermore, the matching of expression patterns is likely to miss genes regulated by multiple Hox proteins, as it actively searches for genes with expression profiles corresponding to a particular Hox protein. The candidate gene approach matching developmental functions, however, is likely to miss genes with additional, Hox independent functions. If those functions occur earlier in development than the Hox dependent functions, they may result in malformations that can obscure the Hox dependent effects occurring later on. An alternative *in vivo* approach to the search for candidate genes that may help circumvent the above mentioned difficulties is subtractive hybridisation. Subtractive hybridisation is the measurement of differential gene expression in wild type tissues compared tissues from Hox gene mutants. Such expression profiling methods are able to analyse Hox downstream genes in a relatively unbiased manner, especially when larger sets of genes are used, that were not preselected for certain features. They can also measure in which direction Hox downstream genes are regulated. However, this method does not differentiate between direct and indirect downstream genes.

For identification of Hox direct targets, *in vivo* methods can be used, such as:

- one-hybrid analysis
- immuno-precipitation of DNA fragments bound by Hox proteins (ChIP)
- methylation of DNA fragments bound by Hox proteins (DamID)

The above mentioned DNA-protein analyses allow identification of the DNA sites bound by Hox proteins. Provided whole genomes are analysed, this method is fairly unbiased. However, these approaches cannot elucidate, whether a DNA site bound by a Hox protein will have an effect on transcription. Furthermore, should a given protein bound to this site have an effect on transcription, these approaches cannot predict whether the protein will promote or suppress transcription. The importance of understanding which DNA sites bound by Hox proteins are functional was highlighted when studies identified a large number of homeodomain-binding DNA sites and some Hox-binding DNA sites for which no *in vivo* function was observed[31,32]. Surprisingly, some DNA sites with high affinity for Hox proteins were identified with no apparent function *in vivo*[34]. It is possible that these high-affinity sites are bound by Hox proteins and are, indeed, non-functional. Alternatively, the effect of those sites might have remained undetected due to redundancy, regulatory compensation or due to spatially or temporally limited effects.

When large-scale analyses became available, microarray expression profiling was employed to identify Hox downstream genes. The success of the individual approaches varied (***Appendix Table 2***)[124,125,126,127,128,129,130,131,132,133,134,135,136,137], which may partly be due to differences in the ability of the methods to retain *in vivo* conditions.

Various expression profiling approaches, in particular those carried out in vertebrates, used cell culture assays[127,131,132,134,137]. While cell culture assays generate data that have less variations than assays using whole organs or embryos, they are not particularly suitable for retaining the natural *in vivo* conditions in which Hox proteins act. In cell culture assays cells are taken out of their spatial and temporal context. This causes similar problems as for the *in vitro* methods, as relevant factors, located only in specific areas or specific temporal contexts, may be missing. However, in contrast to *in vitro* assays, some more general factors modifying Hox binding to DNA or modifying the effect of the Hox protein on transcription, such as **Exd** and **Hth** may still be present. This makes cell culture assays probably more reliable in identifying functional Hox downstream genes than *in vitro* techniques, though not quite as reliable as other *in vivo* techniques using whole organs or whole embryos. One additional concern is that, in cell culture assays, neither the spatial nor the temporal context is known in which the Hox downstream gene is under Hox protein control. As a result, this uncertainty complicates the assessment of which out of the many possible functions a downstream gene has, is subject to Hox regulation. Even some *in vivo* approaches, using whole organs or whole embryos, can create similar problems if, for example, the chosen time frame is very large[124] (***Appendix Table 2***). As the regulation of a downstream gene may occur only at a particular stage, choosing a large time-frame may dilute the regulatory effect by pooling with stages where no regulation occurs. Time-constrained Hox dependent effects may therefore easily be missed and mis-interpreted as false negatives. In addition, as mentioned before, using a large time-frame makes it difficult to ascertain which of the many possible functions a downstream gene has, is subject to Hox regulation. Such complications can be avoided, provided the context in which the downstream gene regulation occurs in is retained and a tightly constrained time frame is used. This kind of approach has been used by a number of apparently more successful studies[125,126,135,136,138] (***Appendix Table 2***).

Detailed analysis of the Hox downstream genes identified so far, has shed some light on the mechanisms of how Hox proteins achieve developmental

specificity. Some important developmental fates determined by Hox proteins could be linked to the responsible downstream genes. However, many aspects of Hox molecular and developmental function are still poorly understood. For this reason, a more detailed knowledge of the sets of downstream genes regulated by different Hox proteins, combined with a method to distinguish direct from indirect targets, is required.

1.5 Aim of the Ph.D. thesis

Aim of this Ph.D. thesis is to shed light on how Hox proteins to create vastly different structures over the course of development. Q50 homeodomain proteins, including Hox proteins, have virtually indistinguishable DNA-binding properties on their own and yet exhibit remarkable differences in their developmental function. Two main hypotheses have been proposed to explain how Hox proteins achieve this[42]. Although not always stated in such a specific manner, the main difference between the hypotheses boils down to whether Hox proteins achieve their differential morphological output via 1) large differential regulation of a few key downstream genes or via 2) minute differential regulation of many different downstream genes with similar morphological functions (see ***Figure 9***).

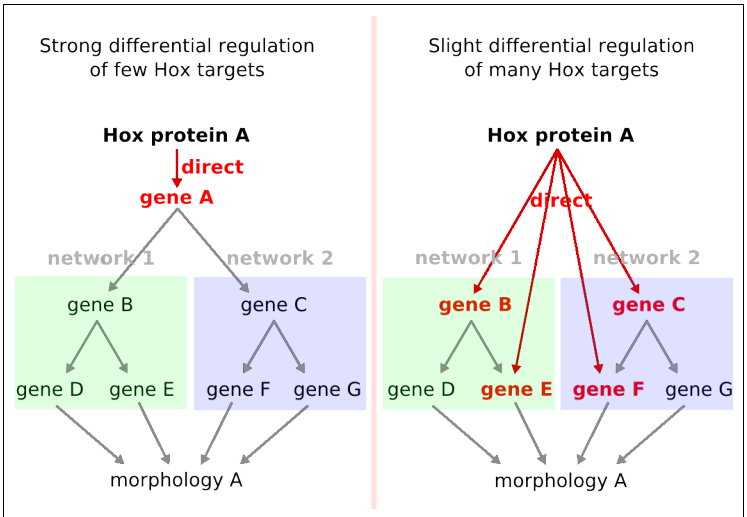

Figure 9: Two models for how Hox proteins determine different morphologies. The model to the left proposes regulation of a select few direct target genes by Hox protein while the model to the right proposed regulation of many direct downstream genes.

The first hypothesis requires large differential regulation of a few key downstream genes to achieve differences in morphology. In addition, as only a few downstream genes are regulated, these key downstream genes will have to be at the top of signalling cascades, such as transcription factors or signalling molecules, as they allow for amplification of the signal. Scientists supporting the first hypothesis tend to defend the idea that Hox proteins achieve their differential effect on target gene transcription through

increasing their DNA-binding specificity by direct interaction with other factors such as **Exd**. Modification of DNA-binding specificity is one way to achieve large differential regulation of a few key downstream genes. Differential regulation between Hox proteins, however, can also be achieved by the "activity regulation" model, as some Hox proteins may be capable of interacting with other factors to modify target gene expression, while others may not. Some Hox proteins may therefore activate expression of the target gene, while others may repress it. Thus, modification of DNA-binding specificity or regulation of transcriptional activity can be employed to achieve large differences in expression of key downstream genes. Consequently, to create different morphologies, these few targets would have to be regulated primarily by one Hox protein or shared targets would have to be regulated in a different direction primarily by one Hox protein (see ***Table 1***).

The second hypothesis allows Hox proteins to achieve differences in morphology via slight differential regulation of many downstream genes at a time. It takes into account that minor changes can accumulate and have major effects on morphology. In addition, only slight differences in the preferences of Hox proteins to certain DNA sequences would be required, as regulation of many genes by Hox proteins allows for accumulation of small changes in the pathway. A Hox protein could therefore be able to bind to many sites in the genome, directly regulate many target genes and share many downstream genes with other Hox proteins. In addition, as no preference for a particular type of direct target genes, such as transcription factors, is required, Hox proteins could potentially regulate a wide variety of different types of genes (see ***Table 1***).

Strong differential regulation of few Hox targets "Key" Hox targets / co-selective binding	**Slight differential regulation of many Hox targets** Many Hox targets /activity regulation
- few direct target genes	- many direct target genes
- regulation of a Hox target gene primarily by one Hox protein only	- regulation of a Hox target gene is primarily shared by multiple Hox proteins
- regulation of primarily other genes encoding transcription factors or signaling molecules	- regulation of all kinds of target genes possible
- either specific binding to the DNA (cooperation) or highly differential activity regulation	- unspecific binding to the DNA, some degree of specificity achieved alone or by activity regulation

Table 1: Two models for how Hox proteins determine different morphologies.

As a minor degree of difference in binding to certain DNA sequences is inherent to different Hox proteins[17,29], no further mechanism enhancing differential regulation of direct target genes would be needed.

Differentiating between these two hypotheses is essential. Should the second hypothesis hold true, we may need to drastically modify our methods for identifying Hox target genes. For some morphologies, there may be no "key" direct or downstream target capable of inducing a functional morphology in absence of the relevant Hox protein. As actual regulation may occur via multiple medium to low affinity sites, the search for strongly regulated genes or genes with high affinity sites may be misleading. Rather than continued search for "key" target genes, a new approach for identifying Hox downstream genes and direct targets would be required, with more emphasis on lower affinity sites and identification of Hox controlled regulatory networks affecting the same morphological structure. Most likely, the focus for the search of Hox protein interacting factors would also have to shift. Methods, so far, have been focussing on DNA-binding specificity enhancing factors such as **Exd** may not be required for all Hox dependent features. Instead, factors of the "activity regulation" model may hold the key. Should this be the case, the parameters for identification of DNA sites binding Hox proteins would have to shift away from the focus on Hox/**Exd**-binding sites and be modified to include potential other interaction partners.

No statistically significant statement capable of differentiating between the two hypotheses can be made up to this point, as the number of known Hox downstream genes and direct targets is too low. To discriminate between the two hypotheses, it is essential to increase the number of identified Hox downstream and direct target genes. To achieve this, we chose to carry out a comparative analysis of Hox over-expression in *Drosophila* and examined the effects on gene-expression. This is the first time that such a quantitative and comparative approach for Hox downstream gene identification was carried out. The data generated by this analysis may provide us with answers to a number of questions. These are:

- How many Hox downstream genes are uniquely regulated by one Hox protein and how many are shared with other Hox proteins?
- How many downstream genes and direct targets are regulated by Hox proteins in general?
- What kind of downstream genes do Hox proteins preferentially regulate?

In addition, the increased number of newly identified Hox downstream genes, may help us to:

- link new downstream genes to specific morphological features
- examine the possibility of identifying new or extending known Hox regulatory networks based on our data

A number of RNA *in situ* hybridizations were generated to verify the Hox dependent regulation of downstream genes identified in the microarray analysis. As two temporally consecutive stages, embryonic stage 11 and 12, were examined, the RNA *in situ* hybridizations will enable us to carry out one additional analysis and visualize to what extent the spatio-temporal context influences the regulation of Hox downstream genes.

1.6 Experimental Approach

Overview:

1. **Comparative analysis of Hox proteins**
    1. Microarray-experiment
    2. Validation of over-expression
        1) Antibody staining
        2) Degree of over-expression
    3. Downstream gene validation
        1) RNA *in situ* hybridization
        2) QRT-PCR
    4. Data analysis

2. **Detailed analysis of the Hox protein Dfd**
    1. **Dfd** dependent morphological structures
        1) SEM => cell shape and morphology
        2) BrdU => cell division
        3) apoptosis
    2. Direct target prediction
    3. Validation of direct target prediction

A gene expression profiling experiment was carried out over-expressing six of the eight *Drosophila* Hox genes at embryonic stages 11 and 12 to quantitatively and comparatively identify Hox downstream genes. The Hox genes used were: *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), a*bdominal-A* (a*bd-A*) and *Abdominal-B* (*Abd-B*). All of these Hox genes exhibit strong mutant phenotypes and specify different morphological structures[10]. The two genes not analysed in this study were

probposcipedia (*pb*), which does not show a loss-of-function phenotype in embryos[139] and *labial* (*lab*), which has already been studied using microarray technology[124]. With the help of the *UAS-GAL4* system[140], the six Hox genes of our study were placed under the control of the promoter *armadillo* (*arm*)[141]. The chosen *arm*-promotor was used to ectopically express the Hox proteins or the **LacZ** reference in the whole embryo. Ectopic expression of **LacZ** was used as a control as it mimics the effect of expressing an additional protein and is thought to have no effect on *Drosophila* development. Over-expression analysis was chosen rather than loss-of-function analysis for two main reasons:

Primarily, the effect of a loss-of-function experiment on the whole embryo could be expected to be relatively small and hence difficult to identify, due to the relatively small expression domains of some Hox proteins[10,16]. An ectopic expression experiment, in contrast, could be expected to amplify weak or locally restricted changes in expression by expanding the expression domain to the entire embryo. The regulatory effect of the Hox protein would thereby be amplified and become more easily detectable.

Secondarily, Hox genes are known to regulate each other's expression[10,142,143]. As a result, a loss of expression of one Hox gene often results in ectopic expression of another. This may obscure whether the observed effects are due to the loss of the Hox protein of interest or are a secondary effect caused by ectopic expression of another Hox protein. One might argue that this phenomenon also occurs vice-versa, as an over-expression of one Hox protein may cause another to be repressed. This is correct, however, ectopic expression of the Hox protein of interest can be achieved within the entire embryo. The effect of the repressed Hox protein, consequently, would only be observed in its own native and significantly smaller expression domain. The effect of the Hox gene of interest should therefore be observable in a significantly larger expression domain, which should provide a beneficial signal-to-noise ratio. Concerns about whether a Hox protein is capable of carrying out its function in regions of the embryo in which it is usually not expressed are unsubstantiated as over-expression of Hox proteins has been shown to result in ectopic morphological structures specific to that Hox protein[67,81,91,144,145].

To confirm over-expression via the *arm*-promotor, antibody stainings were performed at embryonic stage 11 and 12. As different Hox protein concentrations are known to have different effects on downstream gene regulation, it was important to use expression intensities similar to those that we intended to study. This could not be guaranteed for certain promoters

used in previous studies, such as heat shock promoters[124]. Quantitative real-time polymerase chain reaction (QRT-PCR) was carried out to analyse the degree of over-expression of the Hox mRNA and examine, whether the expression levels are comparable across all used *UAS-Hox* constructs. Ectopic expression of Hox proteins may result in non-physiological concentrations of these transcription factors within the nuclei of cells and cause gene expression to be affected in an unspecific manner. To address this concern, the degree of Hox protein expression was visualized by fluorescent antibody staining. The average fluorescent intensity of 20 nuclei at different locations was measured in wild type cells and cells ectopically expressing the Hox protein via confocal microscopy and compared to one another.

As the expression of genes is highly variable throughout development, we wanted to know to what extent the temporal environment affects Hox downstream gene regulation. For that purpose, embryos of two consecutive stages, stage 11 and 12, were collected (***Figure 10 A***). In addition, two adjacent stages can be used as an internal control for target gene prediction. The overlap of Hox target genes in adjacent stages is expected to be much higher than by pure chance as gene expression is not an instantaneous on and off switch. Stage 11 and 12 were chosen, as these are the stages where the first segments become visible and additional morphological structures are formed for the first time[117,139]. Stages 11 and 12 are therefore the earliest stages at which morphological defects of Hox gene mis-expression can be observed. After collecting the embryos, the RNA of whole embryos was extracted and transcribed into double stranded cDNA. Biotinylated cRNA probes were then generated, fragmented and hybridized to the Affymetrix *Drosophila* Genome array 1.0 (www.affymetrix.com) (***Figure 10 B***).

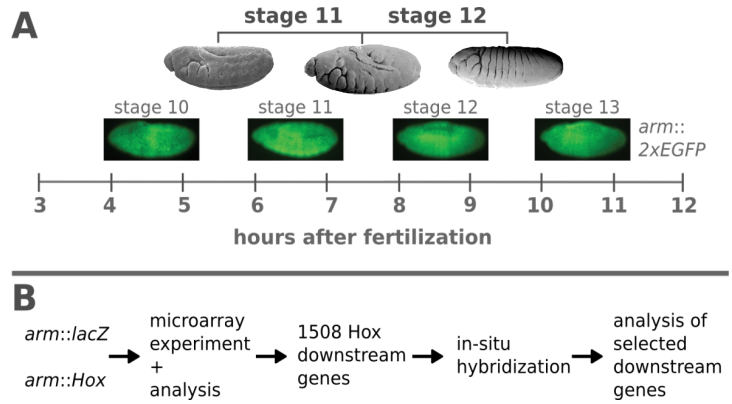

Figure 10: Experimental approach. A depicts the stages chosen for the microarray-experiment. Expression of the *armadillo* (*arm*) promotor is visualized by the **EGFP** reporter in these stages. **B** Chosen experimental approach.

The quality of the microarray-experiment was assessed by examining the recovery rate of known Hox downstream genes in embryonic stages 11 and

12. Cut-off parameters for the analysis were chosen so that the number of recovered “true-positive” target genes was maximized. In addition, the percentage of false positive Hox downstream genes in the microarray-experiment was evaluated by QRT-PCR and RNA *in situ* hybridizations of randomly selected downstream genes.

To assess the types of downstream genes regulated by Hox proteins, we utilized GO-term definitions provided by Flybase[146]. We also aspire to identify potential Hox downstream networks within the newly identified set of Hox downstream genes. Thus, we assessed whether multiple genes taking part in the same regulatory network described in the literature were present within the newly identified Hox downstream gene sets.

We focused on morphologies determined by the Hox protein **Dfd** to analyse the functional role of Hox downstream genes in more detail. Some **Dfd** dependent morphological structures may have been missed in previous analyses of embryonic stages 11 and 12, in particular those structures that are only present in highly constrained time-frames. We, therefore, reassessed what morphological structures are under **Dfd** control by using scanning electron microscopy (SEM) to examine wild type and *Dfd*$^{w21/r11}$ mutant embryos. Effects of **Dfd** on cell division and apoptosis was examined in wild type and *Dfd*$^{w21/r11}$ mutant embryos by BrdU staining and activated caspase-3 antibody staining respectively.

A new cell shape phenotype was identified in *Dfd* mutants. As **Dfd** affects the Jun kinase (JNK) pathway and the JNK pathway is known to affect cell shapes in other morphological processes, such as embryonic dorsal closure and adult thoracic closure[147,148], we determined, whether the JNK pathway could be responsible for the **Dfd** dependent change in cell shape. To this effect, we ectopically expressed a ubiquitously active form of **Hemipterous** (**Hep**)[149], **Hep**^{act},with the *arm-GAL4* driver construct. Thereby, the JNK pathway was activated across the entire embryo, including the maxillary segment. Using SEM, we then compared the cell shapes of wild type to those embryos ectopically expressing **Hep**^{act}.

Hox target genes to be used for further study had to pass a set of filtering criteria and be associated with a known morphological effect, for example regulation of cell morphology and size, apoptosis, cell cycle or cell adhesion. To analyse the effect of the downstream genes at stage 11 abd 12, knock-out alleles were either obtained from one of the stock centres (Bloomington, Tuscon, Szeged) or were mimicked by generated RNAi-fly strains carrying **GAL4**-inducible, inhibitory RNA (RNAi) constructs[150,151]. However, it is possible that loss-of-function mutations in many of these genes are either lethal or

have pleiotropic effects. In addition, many could have no obvious phenotype as other genes with similar function may be able to compensate for the loss of a gene. Thus, conditional alleles, gain-of-function alleles and knock-out combinations of presumably redundant genes were obtained whenever possible.

We performed our experiment approximately one stage after first gene expression by the *arm*-GAL4 promotor construct was observed. Using such a short time-frame post-induction minimizes the amount of activation of downstream gene cascades. Thus, the chosen time-frame is likely to yield a higher proportion of direct Hox target genes.

The experiments described in the following paragraph were carried out by Daniela Bezdan and Stefan Henz. Individual contributions are specified in the section “contributions”.

To identify direct downstream genes within the identified set of Hox downstream genes, we developed a bioinformatic tool to detect conserved Hox cluster sites on the DNA for the Hox protein **Dfd**, based on consensus sequences taken from the literature. The percentage of direct Hox targets is expected to be higher in our selected set of Hox downstream genes than over the genome as a whole. Therefore, we used our set of **Dfd** downstream genes identified in the microarray experiment as a calibration for our bioinformatic tool. The accuracy of the direct target gene prediction was assessed by carrying out multiple EMSA as described[95].

2 Results and Discussion

2.1 Data generation and validation

2.1.1 Validation of over-expression

Ectopic expression of each of the six Hox proteins of interest in embryonic stages 11 and 12 was confirmed via antibody staining (***Figure 11A***). We carried out additional validation experiments, as differences in Hox protein concentration can have a considerable effects on Hox downstream gene regulation[152,153]. For a comparative approach it was important to ensure that the expression levels of the ectopically expressed Hox proteins are similar. QRT-PCR confirmed similar Hox mRNA expression levels in all Hox over-expression fly lines (***Figure 11B***). A slightly lower expression intensity of the *abd-A* gene is possibly due to either a slightly less efficient *UAS*-driver construct or due to differences in the experimental approach, as the *abd-A* gene expression profiling experiment was carried out at a later time.

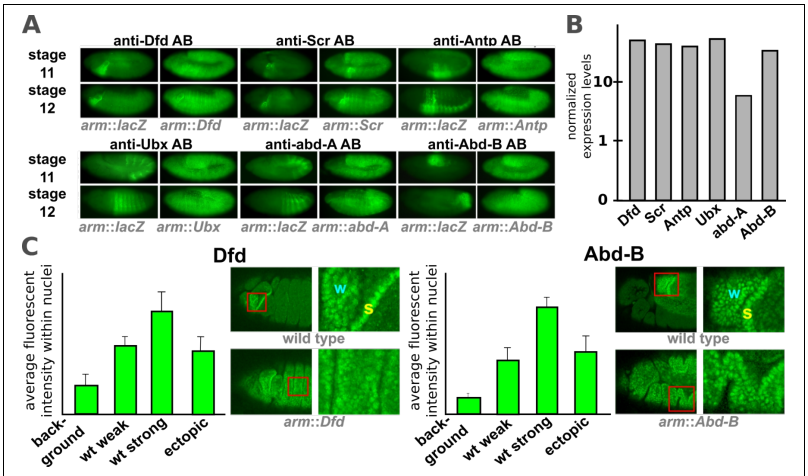

Figure 11: Validation of Hox over-expression. A validation of over-expression of wild type and Hox ectopic expressing embryos (arm::*Hox* constructs) at stage 11 and 12 via fluorescent antibody stainings to the corresponding Hox protein. **B** Comparable levels of Hox ectopic expression via QRT-PCR. **C** Analysis of the level of over-expression of wild type compared to Hox ectopic expressing embryos. The level of ectopic Hox protein expression corresponds to the lower level of the Hox wild type expression for both **Dfd** and **Abd-B** protein constructs.

The above mentioned concentration dependent effects, also made it necessary to determine whether the expression intensity of the ectopically expressed Hox proteins were comparable to the wild type. This is important as differences in concentration levels would lead to transcription artefacts. To account for this, we measured the expression intensities of the **Dfd** and **Abd-B** protein within the nuclei of wild type, *armadillo*::*Dfd* and *armadillo*::*Abd-B* driven embryos via fluorescent antibody staining and confocal microscopy. The level of ectopic Hox protein expression corresponded to the lower level of wild type Hox protein expression at embryonic stage 11 and 12 (***Figure 11C***).

2.1.2 Downstream gene validation

A total of 1508 Hox downstream genes were identified for the Hox proteins **Dfd**, **Scr**, **Antp**, **Ubx**, **Abd-A** and **Abd-B** by first normalizing the raw data and calculating expression intensities using gcRMA[123]. Subsequently, a combination of per-gene (Logit-T ≤0.025)[154] and common variance (>1.5 fold change) filtering was applied. Six[40,95,155,156,157,158,159] of the 18 genes previously shown to be under Hox control during stages 11 and 12[40,58,79,95,119,120,156,157,158,159, 160,161,162,163,164,165,166,167,168,169] were recovered by filtering with the above mentioned cut-off criteria. RNA *in situ* hybridizations for three of these transcripts confirmed the microarray results (***Figure 12A***). 12 of the known Hox target genes not recovered in the gene expression profiling experiment were a) either not regulated in Hox ectopic expression embryos[160], b) tested in loss-of-function mutant embryos only[58,79,169], c) regulated in such a small region in ectopic Hox expressing embryos, that they remained undetected in the microarray experiment[120,162,166,167,168], or d) their actual ectopic expression was induced in earlier stages, while the analysis of the embryos itself was carried out in embryonic stage 11 or 12[97,120,165,166.167]. Some of these 12 genes may therefore represent Hox downstream genes not regulated at the stages we examined.

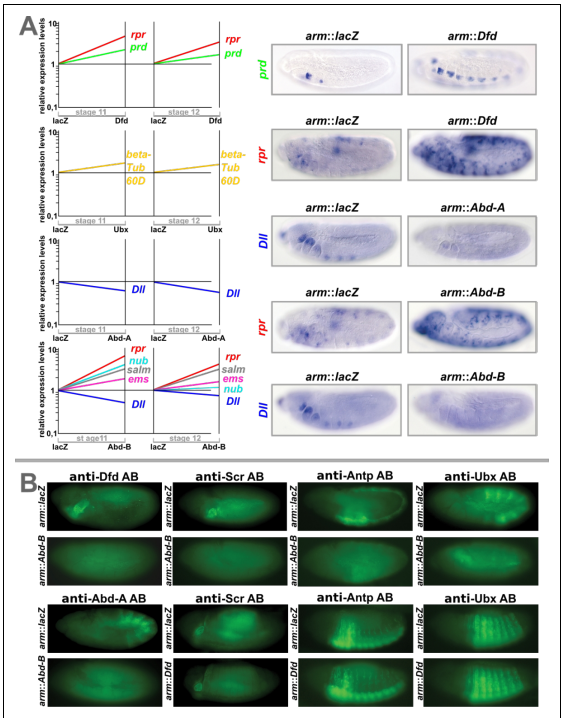

Figure 12: Recovering known downstream genes. A Downstream genes previously identified in stage 11 and 12. **B** Hox cross regulation.

Another set of genes regulated by Hox proteins is the group of Hox genes themselves. The phenomenon of cross-regulation between Hox genes is termed posterior suppression, because the expression of more posterior located Hox genes usually represses expression of more anterior located Hox genes[143]. This phenomenon was also observed for the Hox downstream genes identified in our microarray experiment and was confirmed by antibody staining (***Figure 12B***).

After confirming the ability of our gene expression profiling experiment to

identify known Hox downstream genes, we determined the accuracy by which we identified the new downstream genes. For this, we used RNA *in situ* analysis in wild type embryos and embryos ectopically expressing different Hox genes for 25 randomly selected genes. 24 of 25 genes behaved as detected in the gene expression profiling microarray. For genes with the most prominent expression patterns and lowest backgrounds in the over-expression RNA *in situ* hybridizations, additional RNA *in situ* hybridizations were carried out in Hox loss-of-function embryos. The evaluation of results in loss-of-function approaches was challenging due to considerable variation in expression patterns at slightly different stages. Therefore, to ensure the validity of the comparison, we only compared embryos showing exactly the same expression pattern and intensity outside the Hox expression domain. Seven of these genes showed Hox dependent regulation in the loss-of-function RNA *in situ* analyses (six of them are shown in ***Figure 13***).

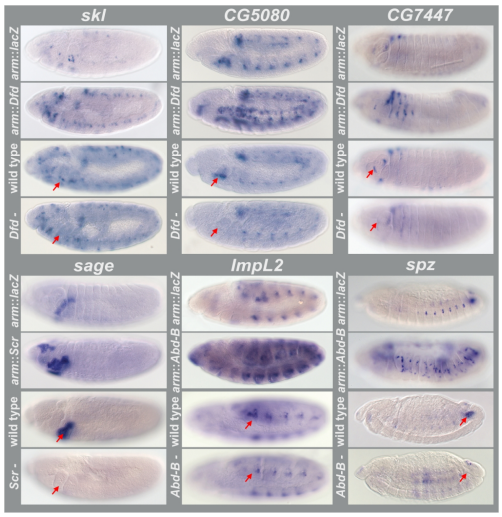

Figure 13: Newly identified Hox downstream genes.

Furthermore, within our set of newly identified downstream genes, we expected to identify genes regulating the same morphological structure as the corresponding Hox protein. For example, **Scr** and **Sage** are transcription factors required for normal salivary gland morphogenesis[97,170]. As such, it was not surprising that we identified *sage* as a downstream gene of **Scr** in our microarray analysis. The result was confirmed in both gain-of-function and loss-of-function RNA *in situ* hybridizations (***Figure 13***).

In conclusion, these results make us believe that the majority of the newly identified Hox downstream genes are true Hox downstream genes.

2.1.3 Direct target gene prediction for Dfd

Using a bioinformatic approach, Daniela Bezdan and Stefan Henz were able to identify a total of 70 Hox cluster sites within the 240 newly identified downstream genes for **Dfd**. 46 of these sites were conserved in at least two out of four *Drosophila* species we examined. To evaluate the quality of the

prediction, Daniela Bezdan selected six of the predicted and conserved **Dfd**-binding sites, and used EMSA to verify that all of them specifically bind **Dfd**.

2.2 Data analysis

2.2.1 Number of Hox downstream genes

To understand how Hox proteins shape different morphologies and to differentiate between the two mechanisms suggested, i.e. large differential regulation of a few key downstream genes or minute differential regulation of many different downstream genes with similar morphological functions, it is essential to understand some basic mechanisms by which Hox proteins effect downstream gene regulation:

- how many genes and networks are under Hox protein control?
- to which degree Hox proteins differ in their regulation of downstream genes?
- what kind of downstream genes are regulated by Hox proteins?

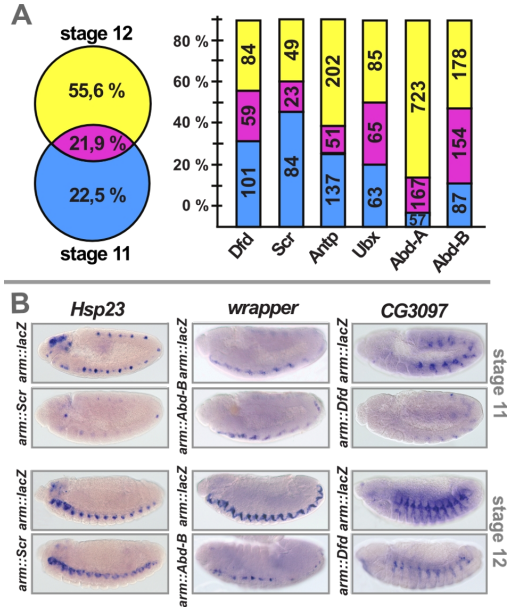

Figure 14: Hox time dependent gene regulation. **A** Overlap of Hox downstream genes between stage 11 and 12. **B** Genes predicted to be down-regulated at either stage 11, but not 12 (Hsp23), at stage 12, but not at stage 11 (wrapper) and in both stage 11 and 12 (CG3097).

The number Hox downstream genes identified in our gene expression profiling experiment at embryonic stages 11 and 12 consists of a total of 1508 genes. The individual Hox proteins exhibited an average overlap of 22% in their sets of downstream genes regulated in the temporally consecutive developmental stages. Such an overlap has a huge impact on estimations of the overall number of Hox downstream genes and evaluations of the dependence of Hox proteins on time dependent factors. Therefore, to to validate the stage dependent regulation of Hox downstream genes observed in our microarray analysis, we carried out RNA *in situ* hybridizations of genes that were regulated in only one of the two stages examined and a gene that was regulated in both: The heat shock gene *Hsp23* was was identified as being regulated at stage 11 by **Scr**, but not at stage 12. The gene *wrapper*

was identified as being down-regulated by **Abd-B** in stage 12, but not in stage 11, while *CG3097* (unknown function) was was identified as being down-regulated by **Dfd** in both developmental stages. The results of the RNA *in situ* hybridizations support our findings (***Figure 14***).

Of all the genes we identified as Hox downstream genes, the percentage of genes regulated in both embryonic stages tested, was about 30% (463 out of 1508). It can be speculated that Hox downstream genes will similarly differ between other, non-consecutive developmental stages. This idea is supported by the observation that there is no overlap between the 30 validated **Ubx** downstream genes identified in larval stages (third instar) [125,126] and the **Ubx** downstream genes we identified in embryonic stages (stage 11 and 12). As a result, I would expect an increase in the number of Hox downstream genes as further experiments identify Hox downstream genes in additional stages. Considering that Hox genes are expressed nearly throughout the development of the fly (***see Figure 15***), it could indicate that nearly all genes are under Hox protein control at one time or another, provided a roughly comparable number of genes is regulated by Hox proteins in other developmental stages.

2.2.2 Number of direct Hox targets

The possibility that a majority of genes are under Hox protein control, fits with the hypothesis that most or all cells in homeodomain protein expression domains are under homeodomain protein control[31]. However, considering the enormous amount of downstream genes, it is essential to differentiate between how many of the downstream genes are directly controlled by Hox proteins and how many are downstream of direct Hox target genes. Two different mechanisms for how Hox dependent morphogenesis occurs have been proposed. First mechanism: Hox dependent morphogenesis occurs by Hox genes affecting regulation of a few key Hox target genes. Those targets would require a strong differential regulation between Hox proteins. Estimates for the number of direct Hox target genes vary from around 85 to more than 100 direct target genes per Hox protein[121,122,171]. Second mechanism: Hox dependent morphogenesis occurs by Hox genes affecting regulation of multiple genes in the same developmental pathway. In this scenario, minor differences in binding affinity between Hox proteins would be sufficient to result in a morphological change. No precise number of direct Hox targets was proposed for this scenario, but it was implied that the majority of genes could be under direct homeo-protein control[172].

Our analysis of the embryonic stages 11 and 12 indicates that 75 **Dfd** downstream genes are directly regulated by **Dfd**, with 46 of the 75 DNA sites predicted to bind **Dfd** being conserved in two out of four *Drosophila* species. Assuming similar increases in numbers of direct target genes with analysis of further developmental stages, the number of direct target genes for **Dfd** is likely to exceed the predicted number of 85 to ≥100 proposed for the first mechanism[121,122,171]. Again, this holds true in particular when considering that Hox genes are expressed almost throughout the entire life-span of a fly (***Figure 15***). However, the 46 to 75 **Dfd** direct target genes identified at embryonic stages 11 and 12 make it unlikely that the number of direct Hox target encompass the majority of genes, as proposed in the second mechanism[31]. The actual number of direct Hox target genes per Hox protein appears to lie in-between the values proposed in the two mechanisms above. I suggest, that both mechanisms of how Hox proteins achieve differential regulation are used *in vivo* and that this explains the observed number of Hox downstream and direct target genes.

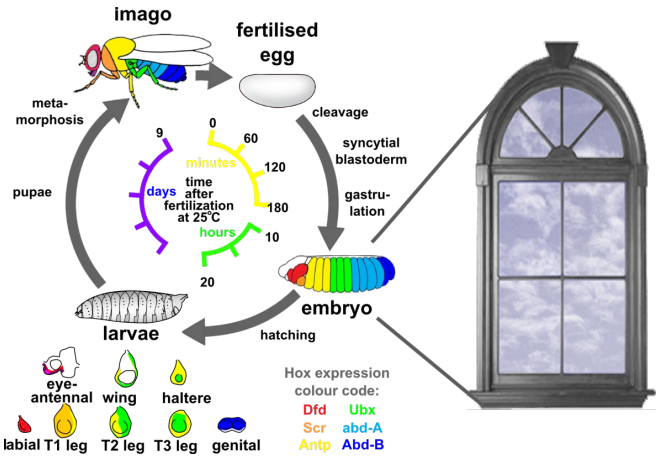

Figure 15: stages of Hox gene expression. We analyzed a small window of the total stages within the life of a fly expressing Hox genes.

To evaluate this assessment, information gathered from additional studies is reviewed. As mentioned in ***chapter 1.4.1***, we know of some key Hox downstream genes for which induction or repression is sufficient to induce or rescue certain Hox dependent phenotypes. Some of these were analysed in detail and explained at the molecular level. These detailed studies, for example, demonstrated drastically different DNA-binding specificities between different Hox protein/**Exd** complexes[38]. These results show a mechanism by which strong differential regulation of direct Hox target genes can be achieved and thus support the use of the first mechanism by Hox proteins. However, this mechanism is unable to provide an explanation for two properties exhibited by homeodomain proteins such as Hox proteins. They can regulate a majority (up to 87% of the expressed genes) and similar sets of genes[172], and, more importantly, they can bind to a majority of

genes[32]. The second hypothesis can easily account for both properties. Unfortunately, there are far fewer detailed examples that support the second hypothesis. However, evidence has recently become available for the ability of Hox proteins to regulate many downstream genes at once within the same pathway, thereby affecting morphogenesis[116]. This is encouraging data supporting the second mechanism and as it typically is harder to identify multiple genes required for the formation of a morphology in absence of the Hox protein, it is likely that many examples supporting the second mechanism remain to be identified.

Both mechanisms for Hox dependent regulation of morphological structures do not exclude each other. Therefore, in the light of recent findings, our data of the number of Hox downstream and direct target genes indicates that both mechanisms are employed by Hox proteins to induce different features along the anterior-posterior axis.

2.2.3 Context dependence

As demonstrated in ***chapter 2.1.1***, the ability of a Hox protein to regulate a target gene depends, to a considerable degree, on the time-frame the Hox protein is placed in. In addition to the temporal context, the spatial context appears to be significant, as the same Hox protein is capable of determining the fate of many different structures. For example, **Abd-A** can specify a secretory cell, the oenocyte, part of the heart and gut or specific hair structures in the larvae. While it is likely that different sets of downstream genes are regulated by the Hox protein in each of these morphologies, this dependence on a particular environment has often been overlooked. Many approaches to identify new Hox downstream genes have taken Hox proteins out of their regulatory context, either spatially, by choosing cell culture assays and taking the cells out of their surrounding tissue[127,131,132,134,137] or temporally, by choosing a time-frame that spans several different developmental stages[124].

While some degree of downstream gene regulation can be achieved by timing, location and concentration of the Hox protein itself[35,152,173,174], not a single Hox downstream gene, recapitulates the exact same expression pattern of the Hox protein by which it is controlled.

We analysed the Hox dependent expression pattern of various Hox downstream genes identified in the gene expression profiling experiment by RNA *in situ* hybridization. All Hox downstream genes show a regional or a

segmentally repeated pattern distinct from the Hox expression pattern (***Figure 16***). Furthermore, ubiquitous expression of a Hox protein does not lead to ubiquitous expression of the downstream gene. The expression patterns visible in ***Figure 16***, imply that other factors determining the expression pattern of a Hox target occur either regionally or in segmentally repeated patterns. Segmentation genes such as gap, pair-rule and segment polarity genes used in parallel or in combination with Hox proteins, may be responsible for such regional, highly localized Hox target gene expression.

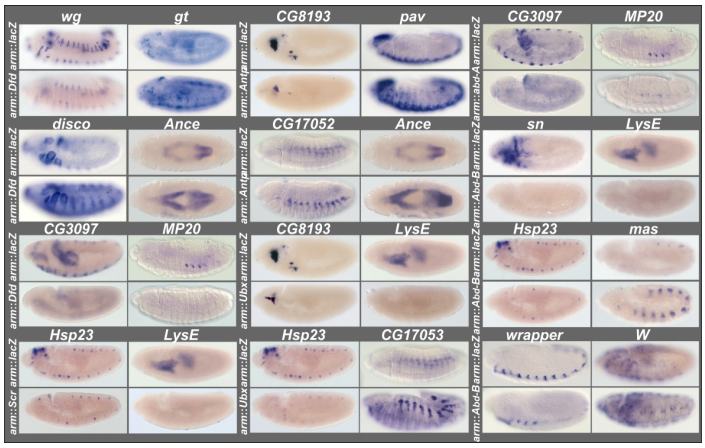

Figure 16: Hox downstream gene expression patterns. RNA *in situ* hybridisations of Hox downstream genes predicted in the array in wild type and Hox over-expression embryos.

There are two ways by which additional factors can define the expression pattern of a Hox downstream gene. Either a single factor has a pre-established expression pattern that corresponds to the expression pattern of the downstream gene, at least within the Hox expression domain, or the occurrence of multiple factors in the Hox expression domain determine a novel expression pattern for the Hox downstream gene (***Figure 17***).

It appears more likely, that the combination of multiple transcription factors is used to achieve a more specific effect on transcription as functional enhancers tend to possess DNA sites for a wide variety of transcription factors[175]. An example for this has been described by Petra Stöbe et al.[73] showing a total of eight different transcription factors acting in combination with the Hox protein **Dfd**. The eight transcription factors and **Dfd** act within the same time-frame and bind independently of each other to a small regulatory unit, the 4S3/3' of the *rpr* enhancer. None of the expression patterns of the factors analysed recapitulated the exact *rpr* expression pattern. In concert with this observation, none of the individual factors were sufficient to re-establish the exact *rpr* expression pattern in conjunction with the Hox protein **Dfd** when additional factors were missing. Beyond the expected additive/subtractive effect of many individual transcription factors on target gene expression, we observed synergistic effects with **Dfd** for some

of these transcription factors. Interestingly, no direct interaction between **Dfd** and the eight proteins was detected, even though both yeast-2-hybrid and GST pull-down assays were used. The observed synergistic effect is therefore unlikely to be due to a Hox heterodimer or heteromultimer formation. In addition, the various transcription factors were able to bind to the DNA-site independently of the Hox protein and the Hox protein independently of the factors as shown by EMSA. To explain the observed synergistic regulation on the enhancer, one needs to look at specific examples.

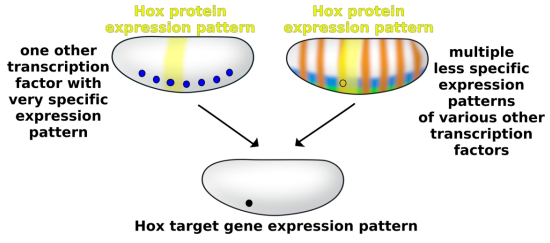

Figure 17: Context dependence. Two potential mechanisms in which Hox proteins receive this information, either by interacting with a factor that already has a pre-established pattern or by acting with a variety of other factors to define a completely new pattern for the downstream gene.

Non-collaborative example: **Empty spiracles** (**Ems**) appears to have no effect on the *rpr* 4S3/3' enhancer when over-expressed without the Hox protein **Dfd**. Simultaneous over-expression of **Ems** with **Dfd**, however, enhances the ability of **Dfd** to activate the expression of *rpr*. Correspondingly, mutation of *ems* does reduce *rpr* expression. Therefore, the **Ems** dependent effect on transcription of the *rpr* 4S3/3' enhancer appears to be dependent on the presence of **Dfd**. The observed synergistic effect cannot be termed collaborative as the DNA site binding **Ems** is not located in a composite response element with a DNA site binding **Dfd**. Furthermore, the observed effect is unlikely to be caused by DNA affinity enhancing cooperation as the two proteins bind to DNA independently of each other and do not appear to enhance each others DNA-binding affinity. A possible explanation for the observed synergistic effect is an indirect interaction taking place between the two factors. For example, a multiprotein complex may require, at least transiently, **Dfd** for its formation and **Ems** could subsequently enhance the recruitment of further factors to the complex. This example would explain why **Dfd** has an effect in absence of **Ems**, but not vice versa.

Potential collaborative example: Petra Stöbe et al.[73] also shows a slightly different example of a synergistic effect for **Dfd** with **Glial cells missing** (**Gcm**). Two aspects differ from the above example. Firstly, **Dfd** and **Gcm** bind to sites on the DNA located immediately next to each other and therefore, by definition, bind to a composite response element. Secondly, even though ectopic co-expression of both transcription factors results in a much stronger activation of the *rpr* enhancer 4S3/3' than can be explained by simply adding

their individual effects on the enhancer, both are quite capable of activating the *rpr* enhancer on their own. As for the non-collaborative example mentioned above, the observed synergistic effect cannot be attributed to a DNA affinity enhancing cooperation as the two proteins do not appear to enhance each others DNA-binding affinity in EMSA. However, unlike the example mentioned above, both proteins are located in a composite DNA site. The observed synergistic effect between **Dfd** and **Gcm** is therefore likely to constitute a case of collaboration[60,72]. A synergistic effect of two proteins located on the same composite response element actually indicates a direct interaction, at least once they are located on the DNA. It is possible that a direct interaction takes place that has been missed in GST-pulldown and yeast-2 hybrid assays. Direct, transient interactions are difficult to identify as the complexes may dissociate during the assay, in particular for physical interaction assays such as GST-pulldown. It is also possible that such a potential interaction between **Dfd** and **Gcm** would require the presence of the composite DNA site to stabilize or initiate the interaction. Neither yeast-two-hybrid nor GST-pulldown assays include the composite DNA site in the approach and would therefore be unable to identify an interaction that requires the presence of a composite DNA site.

Together with other recent results[71,72,176], the potential collaborative example highlights the importance of synergistic composite response elements and offers alternative possibilities for how differential regulation by Hox proteins can be achieved. In addition, as shown in the non-collaborative example, even binding sites that are located further apart from one another are capable of synergistic effects. Such interactions can also be explained, for example, by DNA-bending and multi-protein complexes that pull DNA strands together. Synergistic interactions that are independent of DNA-binding affinity and specificity enhancing interactions, have probably been underestimated in Hox protein research. Much focus has been placed on the best described Hox interaction partner, **Exd**. **Exd** is capable of enhancing Hox-binding specificity to DNA, but **Exd** alone cannot account for the variety of different spatially restricted expression patterns visible in these stages. Hox proteins acting synergistically with a variety of different transcription factors, however, can explain the observable spatially restricted expression patterns and different functional effects.

Non-cooperative (non DNA-binding specificity enhancing) synergistic interactions between different transcription factors are possibly part of a more general mechanism for achieving differential regulation of target genes.

In particular transcription factors that fail to achieve a relevant DNA-binding specificity on their own, such as other homeodomain proteins, Zn-fingers and leucine zippers[177], may use this mechanism as many sites bound by these transcription factors do not appear to have an effect on transcription[177]. Apparently, the regulation of genes by these transcription factors depends on the presence of additional factors. Each transcription factor has the potential to interact with additional factors in a different manner. Thus, besides DNA-binding specificity enhancing cooperation and protein modification, non-cooperative synergistic effects are the only known mechanisms by which differential effects on target gene transcription can be achieved.

2.2.4 Specificity of Hox downstream gene regulation

One of the questions when using a quantitative and comparative approach is whether it is possible to explain the differences in morphology along the anterior-posterior axis by comparing the Hox downstream gene regulation of different Hox genes? To assess how many Hox downstream genes are specific to individual Hox proteins and how many are shared between different Hox proteins, we examined the number of Hox proteins influencing the expression of each identified downstream gene.

Most downstream genes were regulated by only a single Hox protein, with **Abd-A** regulating the majority of the genes (***Figure 18A***). This could indicate that:

a) The regulation of a large proportion of Hox downstream genes is specific to a single Hox protein.

b) At the examined stages, the primary morphological Hox regulator is **Abd-A** and other Hox protein functions are not as important at that particular time point.

c) The **Abd-A** microarray triplicates had less statistical variation, causing more genes to pass the selected cut-off-criteria. This is a possibility we have to take into account, in particular as the **Abd-A** experiments were carried out separately from the others.

To eliminate the last two possibilities, we repeated our analysis excluding **Abd-A**. In this case, 63% of the Hox downstream gene where regulated by one Hox protein only. In contrast, only 1,3% of the downstream genes are common to all Hox proteins examined. This indicates that despite the remarkably similar DNA-binding specificities observed between different Hox

proteins *in vitro*, *in vivo* for the embryonic stages 11 and 12, the majority of downstream genes is regulated primarily by a single Hox protein.

When looking at individual Hox proteins, on average 33% of the downstream genes regulated by a Hox protein are unique to that Hox protein (***Figure 18B***). Similar results are obtained when analysing each stage individually with approximately one third of all downstream genes being unique to one Hox protein. Even when looking at the predicted direct target set for **Dfd**, one third of all predicted direct **Dfd** targets (***Figure 18C***) are regulated only by **Dfd**. Therefore, from the perspective of one Hox protein, the majority of downstream genes are shared with at least one other Hox protein and only approximately one third of its target genes are uniquely regulated by it.

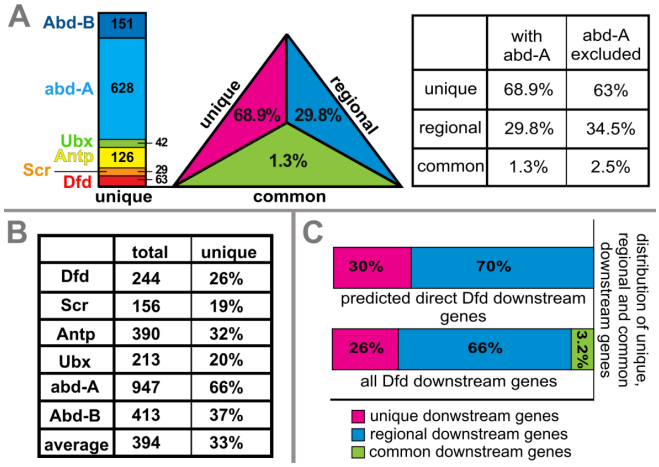

Figure 18: Hox specificity. A left side, numbers of unique downstream genes for individual Hox proteins. Middle, percentage of unique, regional and common downstream genes of all Hox downstream genes identified in the array with the table on the right side depicting the percentages with and without the separately carried out experiment for the Hox protein **Abd-A**. **B** Percentage of unique downstream genes for individual Hox proteins **C** Percentage of unique, regional and common downstream genes of the set of genes predicted to be direct target genes of the Hox protein **Dfd**.

It is plausible that a considerable number of downstream genes of a Hox protein are shared with other Hox proteins, considering that adjacent Hox proteins have to shape similar morphologies in their segments. It also explains why Hox proteins may not always require unique binding to or regulation of certain downstream gene sets. Yet, some degree of difference is required to shape different morphologies, such as head morphology and abdomen. Therefore, genes specific to head morphology and genes specific for abdomen would need to be regulated in a similar manner within the head or within the abdomen, but differentially between the two structures. Consequently, one would expect Hox genes determining more similar morphologies to have more shared downstream genes than Hox genes determining different morphologies. One would also expect them to regulate more shared downstream genes in the same direction than Hox genes determining different morphologies. When looking at the direction of regulation (activation/repression) in our set of shared downstream genes, such a trend can be observed. Adjacent Hox proteins determining similar

morphologies, such as the head expressed Hox proteins **Dfd** and **Scr**, regulated only 4% of their downstream genes in the opposite direction, while the head/thorax Hox protein **Scr** and the abdominal Hox protein **Abd-A** are responsible for very different morphologies and regulate 58% of their shared downstream genes in the opposite direction (***Figure 19***). An example for such a shared downstream gene is *pipe*, which, in our array, is activated by **Scr** and repressed by **Abd-A**. Previous experiments showed it to be expressed exclusively in the **Scr** expression domain in stage 11 and 12[178].

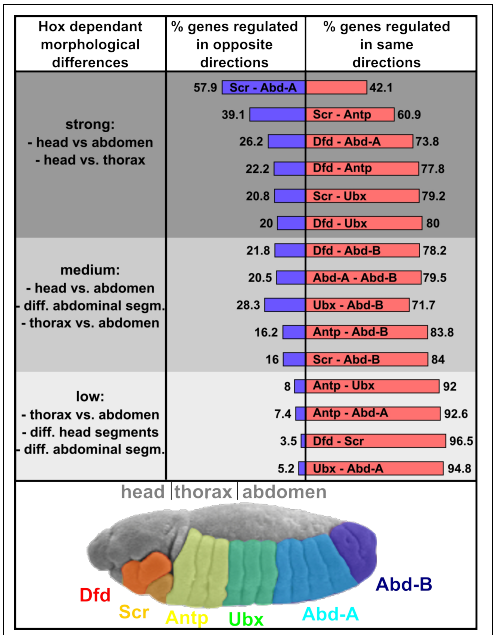

Figure 19: Differential regulation of shared Hox downstream genes.

Even in the small set of common downstream genes, a similar trend can be observed. While most of the common downstream genes are repressed by all six Hox proteins tested, two of the common downstream genes, namely *Paps* and *Ph4alphaSG2* are down-regulated by all Hox proteins tested with the exception of the Hox protein **Scr**. As the only Hox protein required for a specific morphological structure called the salivary glands, it induces these genes, both of which have been described previously as being involved in salivary gland morphogenesis [178,179,180].

Taken together, differential regulation by Hox proteins appears to be achieved by both:

- unique regulation of downstream genes
- regulation in different directions of shared and a few common downstream genes

However, the proportion of downstream genes being uniquely regulated by a single Hox protein, approximately one third of all downstream genes, is higher than previously assumed.

2.2.5 Classification of Hox downstream genes

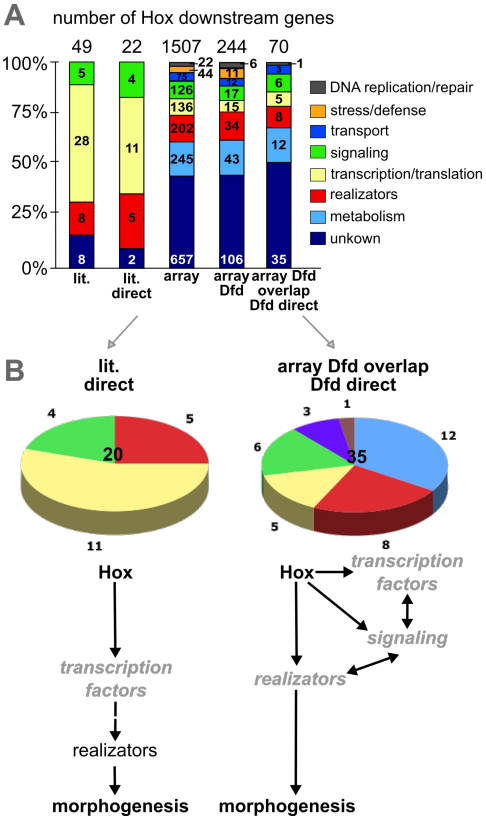

Figure 20: Categories of Hox downstream genes. A Sets of all Hox downstream gene categorized via GO-term definitions. **B** Previously known direct Hox targets and newly predicted direct Hox targets for **Dfd** with known GO-term categories and the resulting concept. Genes of unknown function were not included in B.

The idea that Hox proteins act via a few key downstream regulators is based on the notion that Hox proteins preferentially regulate other transcription factors or signalling molecules directly, thus explaining the vast effects on morphogenesis caused by Hox proteins[42]. The set of Hox downstream genes known before our study was carried out, included 22 identified direct Hox targets, 15 of which were transcription factors or signalling molecules (***Appendix Table 1***). The set of 22 direct target genes did not even include a direct target gene for every Hox protein (Pb, ***Appendix Table 1***) and is as such unlikely to be a representative number for all direct target genes regulated by Hox proteins. However, a percentage of nearly 70% of the 22 direct target genes of Hox transcription factors being transcription factors themselves or signalling molecules, led to the general assumption that Hox proteins preferentially regulate these classes of genes.

To determine the validity of this assumption on direct target genes and to examine, if there are significant differences between the predicted direct target gene set and the downstream gene set, we classified these genes with Gene Ontology (GO) annotation and analysed the distribution of GO-terms within these sets. As can be seen in ***Figure 20A***, the two most prominent groups of downstream genes code for metabolic or realisator functions, followed by transcription or translation, signalling molecules, transport, stress or defence response and DNA repair or replication. Statistical analysis by Fisher's exact test after Bonferoni correction showed no significant bias of Hox downstream genes towards transcription factors or signalling molecules, rather the opposite. The transcription or translation group, but also the transport group were under-represented (P<0.001). More importantly,

analysis of the set of predicted direct target* for **Dfd**, reveals a very similar distribution and correspondingly, the transcription and translation group is not over-represented. In fact, the only group statistically over-represented was the realisator group (P<0.001). This is a significant finding and indicates that realisators and metabolic genes are regulated directly just like transcription/translation, signalling, transport, stress/defence and DNA replication/repair genes. Thus, Hox proteins, unlike previously perceived, do not solely affect the top of regulatory cascades by primarily regulating transcription factors, but instead are capable of directly and indirectly regulating all classes of genes and thereby affect all levels of regulatory networks (***Figure 20B***).

Another observation that may be worth mentioning is that certain subclasses of realisator genes appear to be regulated by Hox proteins in a coordinated fashion. Most apoptosis genes (7 of 8) and cell proliferation genes (18 of 21) were up-regulated, while cell-adhesion (12 of 14) and proteolytic genes (56 of 75) were mostly down-regulated. In fact, a similar effect can be observed in the independently carried out microarray experiment performed for an additional Hox protein, **Lab**. In this experiment, one apoptosis gene and six cell proliferation genes were up-regulated, while three cell adhesion genes were down-regulated[124]. Contrary to our experiment, proteolysis was mostly up-regulated, however, this is possibly due to the experimental set-up inducing **Lab** over-expression via heat shock as heat shock is known to effect proteolysis. We can only speculate about the meaning of this coordinated regulation of realisator-type classes by Hox proteins, but it is possible that certain morphological processes have evolved and consequently need to occur in a coordinated fashion at specific time-points in the course of development for proper and efficient morphogenesis to occur.

2.2.6 Hox regulatory networks

To understand Hox dependent morphogenesis, it is necessary to link Hox proteins to morphogenetic processes and, therefore, to the responsible realisator genes. In the simplest scenario, a realisator gene is directly regulated by a Hox protein, as is the case for the Hox protein **Dfd** directly inducing the expression of the realisator gene *rpr*[95]. It is, however, likely that the majority of such realisator class downstream genes are regulated

* includes those with conserved binding sites and those with binding sites that are not conserved in other *Drosophila* species.

indirectly, as, Hox proteins also directly regulate a considerable number of transcription factors and signalling molecules, which then could regulate further realisator genes. It is, therefore, desirable to elucidate Hox dependent morphological networks.

To this effect we decided to examine whether our data set would allow us to identify novel Hox downstream networks. We first examined whether we were able to detect known Hox dependent networks. One of the best studied Hox dependent networks is the salivary gland network dependent on the Hox protein **Scr**. As can be seen in ***Figure 21 A***, we were able to recover a considerable number of genes within the salivary gland network (*pipe*, *paps*[178], *sage*[181], *PH4alphaSG2*[179], and *hkb*[181]). The genes that were not recovered either did not meet the chosen cut-off parameters, such as the downstream gene *klar* with a fold change of 1.3, or were known to have ceased to be under **Scr** control in these stages, such as *fkh*[58] and, potentially, *CrebA* and *SPCG*.

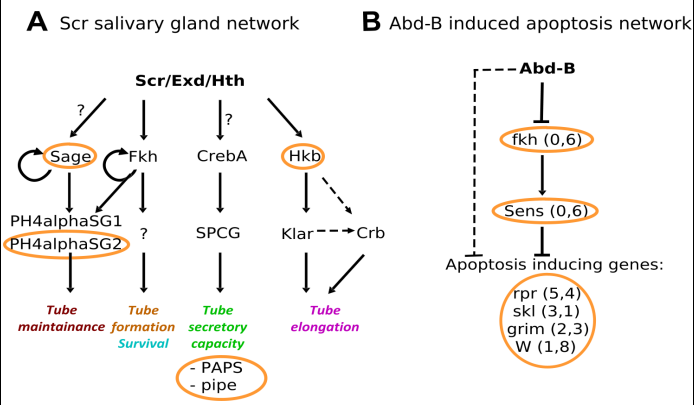

Figure 21. Hox morphological networks. A Scr dependent salivary gland network adapted from Kerman et al. 2006[181] and **B** the known apoptosis cascade now known to be under control of the Hox protein **Abd-B**. Genes in orange rings show genes identified as Hox downstream genes. In B, the numbers in brackets represent the fold change in comparison to wild type expression to indicate the direction in which the gene is regulated.

In addition to the **Scr** dependent salivary gland network, we identified a well defined network, the apoptosis network. Interestingly, this network is dependent on the Hox protein **Abd-B** based on our analysis (***Figure 21B***). Previously, only the apoptosis gene *rpr* was identified as an **Abd-B** downstream gene and associated with boundary maintenance[95]. Now we know that not only *rpr*, but the apoptosis gene cascade is under **Abd-B** control.

In summary, we were able to detect one known and to identify one new Hox dependent network with the data collected from our array. This demonstrates the feasibility of identifying Hox dependent networks using our data. Our data-set should prove particularly useful in combination with large-scale direct target gene prediction and binding site identification approaches such as chromatin immuno-precipitation on a chip/microarray (ChIP on chip)

and DNA adenine methyltransferase identification (DamID). Provided the same stages are used, it may even be possible to determine the hierarchy of the genes within the network in addition to the genes part of the network.

2.3 Hox dependent morphogenesis

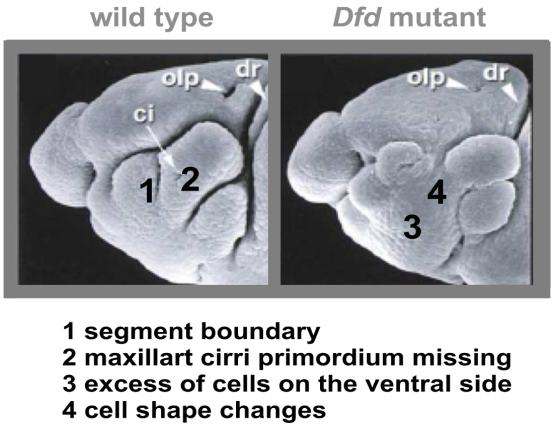

Figure 22: Dfd dependent morphological structures in the maxillary segment. To understand Hox dependent morphogenesis, we focused on one segment, the maxillary segment. The features under control of the Hox protein **Dfd** in stage 12/13 are depicted as numbers. SEM picture taken from Lohmann et al. 2002[95].

To link a specific morphological feature to a specific Hox protein, we focused on the features generated by a single Hox protein, **Dfd**, and on the morphology of a single segment, the maxillary segment. The general function of **Dfd** is to define the identity of both the mandibular and maxillary segments[182]. Thus, **Dfd** is required for formation of components of the feeding apparatus such as the mouth hooks, cirri and ventral organ[14,117]. As over-expression of **Dfd** in the embryo results in ventral labial and thoracic epidermis being transformed into maxillary epidermis[34], **Dfd** and the maxillary segment should provide a suitable model for analysis of morphogenetic processes influenced by **Dfd** downstream genes identified by our ectopic expression microarray approach.

We first examined the **Dfd** dependent features of the maxillary in stage 11 and 12 by scanning electron microscopy to understand which processes are controlled by **Dfd**. As can be seen in ***Figure 22***, four obvious morphological features are under **Dfd** control:

1. maintenance of the boundary between the mandibular and maxillary segment[95],
2. development of the maxillary cirri primordium[183],
3. excess of cells on the ventral side[183] and
4. regulation of cell shape of maxillary cells, a new phenotype that has not been described before.

Of the **Dfd** dependent features, only the maintenance of the segment boundary has been linked to a Hox downstream gene. Lohmann et al., 2002[95]

showed that the realisator gene *rpr* is induced by **Dfd** in the anterior part of the maxillary segment where it is required to maintain the boundary between the mandibular and maxillary segment by inducing apoptosis. In correspondence with this, we recovered *rpr* in our microarray for Hox downstream gene identification as being induced by **Dfd**. The second feature, the generation of the maxillary cirri primordium is known to be under the control of the transcription factor **Prd**[159]. *prd* gene expression in the maxillary segment is completely eliminated in *Dfd* mutants in stage 12 (***Figure 23A***). It is therefore conceivable that the **Dfd** dependent generation of the maxillary cirri is carried out by **Dfd** via **Prd**. For the last two morphological features no candidate genes were known. There are two apparent explanations for the additional cells visible in *Dfd* mutant embryos, additional cell proliferation or missing apoptosis/cell death.

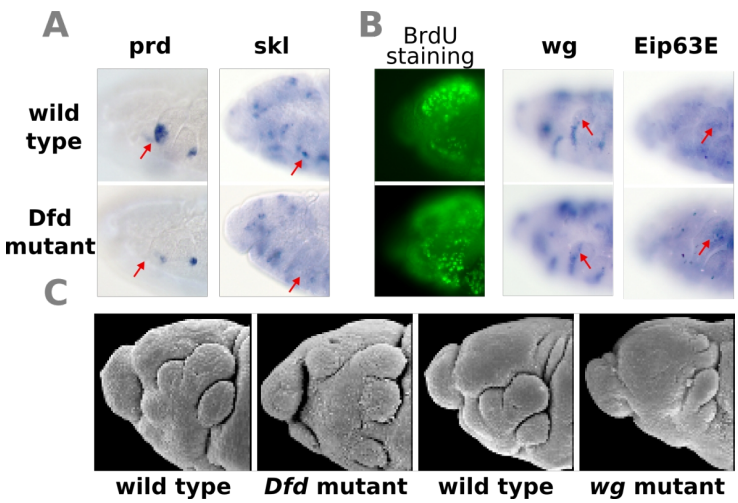

Figure 23: Candidates for the Dfd dependent morphological feature of excess cells on the ventral side of the maxillary segment. A RNA *in situ* hybridization of candidate genes potentially responsible for the **Dfd** dependent formation of the maxillary cirri primordium (*prd*) and regulation of the number of cells at the ventral side of the maxillary segment by apoptosis (*skl*). **B** BrdU stainings of the maxillary segment in wild type and *Dfd* mutant embryos showing **Dfd** dependent regulation of the number of cells at the ventral side of the maxillary segment to be at least partly due to increased cell proliferation. RNA *in situ* hybridization of candidate genes potentially responsible for this **Dfd** dependent phenotype are shown on the right side (*wg*, *Eip63E*). **C** SEM pictures of the maxillary segment in wild type, *Dfd* mutant and *wg* mutant embryos. Compared to wild type, an excess of cells can be seen at the ventral part of the maxillary in *Dfd* mutant embryos, while a loss of cells can be observed in *wg* mutants.

To evaluate, whether a reduction in apoptosis could be responsible for the additional cells in *Dfd* mutant background, we used fluorescent antibody staining for anti-activated caspase 3. Consistent with data published previously[95], we observed a reduction in the number of apoptotic cells (data not shown). In concert with this observation, we identified the **Dfd** downstream gene *sickle* (*skl*) as being induced by **Dfd**. We could confirm **Dfd** dependent regulation of *skl* (***Figure 23A***) and as **Skl** is known to induce apoptosis[184,185,186], this gene is an excellent candidate gene potentially responsible for the additional cells observed in *Dfd* mutant background .

As reduced apoptosis induction and additional cell proliferation do not exclude each other, we used BrdU staining to evaluate whether **Dfd**, also

represses cell proliferation. As can be seen in ***Figure 23B***, this is the case. This de-repression of cell proliferation in *Dfd* mutant embryos can potentially be attributed to two genes, the *Ecdysone induced protein 63E* (*Eip63E*) and *wingless* (*wg*). Both genes were identified in this study as being repressed by **Dfd** and were previously found to be involved in cell proliferation[187,188]. While *Eip63E* encodes a cyclin dependent protein kinase that produces smaller flies when mutated[187], *wg* encodes a signalling molecule capable of not only inducing cell proliferation, but also repressing apoptosis[188]. As can be seen in ***Figure 23B***, we could confirm **Dfd** dependent repression of both, *Eip63E* and *wg,* in the dorsal part of the maxillary as both genes are ectopically expressed in *Dfd* mutant embryos. We were not able to determine the function of **Eip63E** in morphogenesis of the maxillary segment, due to a lack of available mutants for the corresponding gene. However, we were able to confirm the function of *wg* in the maxillary segment and show that *wg* mutants have reduced gnathal lobes[189] (***Figure 23C***).

The fourth and last **Dfd** dependent morphological features described in the stages we analysed, is a cell shape phenotype. Usually roundish cells form in the ventral part of the maxillary, while the same cells appear elongated In *Dfd* mutant embryos (***Figure 24A and B***). This phenotype could be a secondary effect of the above described excess of cells in *Dfd* mutant embryos. An excess of cells at a specific point could induce secondary cell movements causing cell shape changes. However, we had also identified several genes responsive to the JNK pathway (*Ras related protein* (*Rala*) *Angiotensin converting enzyme* (*Ance*) and *CG5080*)[190]. As the JNK-pathway is known to influence cell shape during various morphological processes, such as dorsal closure[147] and adult thorax closure[148], we analysed the effect of the activated JNK pathway on cell shape in the maxillary segment. We ubiquitously activated the JNK pathway in the whole embryo using a constitutively activated form of **Hep**. We then compared the cell shapes to those of wild type, *Dfd* mutant embryos with embryos with a ubiquitously activated the JNK pathway. As can be seen in ***Figure 24B***, wild type embryos have roundish cells, while cells in embryos

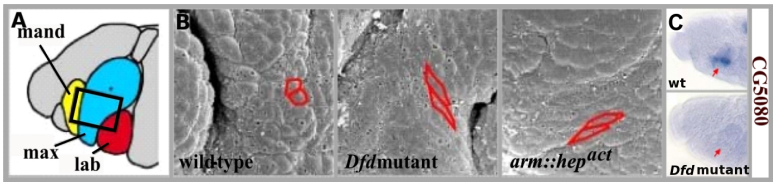

Figure 24: Dfd dependent cell shape phenotype. A the area of the ventral maxillary the SEM pictures were taken is highlighted by the box **B** SEM pictures. Cell shapes were roundish in wild type, elongated in *Dfd* mutants and elongated in embryos with a constitutively active JNK-pathway. **C** *in situ* RNA hybridization of the **Dfd** downstream gene CG5080, a potential candidate responsible for the cell shape phenotype in wild type and *Dfd* mutant embryos.

with an activated JNK pathway had elongated cells in the maxillary segment very similar to those observed in *Dfd* mutant embryos. For one of the JNK responsive genes, *CG5080* has been implicated in cytoskeletal rearrangements[190], and was identified in our microarray as being induced by **Dfd**. As we were able to confirm strong induction of *CG5080* by **Dfd** (***Figure 24C***), *CG5080* is a candidate gene potentially responsible for the cell elongation phenotype.

3 Conclusion

We identified hundreds of new Hox downstream genes that enabled us to analyse the mechanisms of Hox downstream regulation between different Hox proteins (**Dfd**, **Scr**, **Antp**, **Ubx**, **Abd-A** and **Abd-B**) at a quantitative level. In addition, we predicted new direct targets for the Hox protein **Dfd**. For some of these predicted direct targets we were able to confirm binding of **Dfd** to sites within the regulatory region of the gene using EMSA.

One of our aims was to shed light on the mechanism of how Hox proteins manage to create vastly different structures throughout the course of development. We, therefore, tried to differentiate between two prevalent hypotheses in Hox protein research regarding the mechanism of Hox action: strong differences in regulation of a few key downstream genes versus slight differences in the regulation of many different downstream genes (see ***Table 2***).

The first step in this analysis was to look at the number of genes controlled by Hox proteins at certain time-points. Analysis of our microarray data identified 1508 genes to be regulated by Hox proteins at embryonic stage 11 and 12. The stages we analysed encompass only a minute proportion (four hours) within the development of the fly (9 days) that spans morphologically very distinct stages form embryonic stages, over larval and pupal stages to the final adult stage. The Hox dependant regulation changed for approximately two thirds of the 1508 downstream genes we identified within this four hour time frame. Extrapolating to the full 9 days of Drosophila development, it is easy to envisage that, potentially, all of the about 14000 genes in the Drosophila genome[191] could be influenced in their expression by a Hox protein at one time or another.

The situation is slightly different for the genes directly regulated by Hox proteins. Of the 240 genes identified as **Dfd** downstream genes in these stages, we predicted approximately one fifth to one third to be direct **Dfd** target genes. The precise number for **Dfd** direct target genes is 75, 46 of these genes have DNA sites for **Dfd** that are conserved in other *Drosophila* species. The number of direct target genes is likely to increase with the analysis of further stages. Therefore, the probable number of direct targets for Hox proteins appears to lie in-between the amount of direct Hox target genes suggested by scientists supporting the hypothesis of strong differential regulation of a few key downstream regulators, approximately 85 to ≥100[121,122,171] and the amount of Hox targets suggested by scientists defending the hypothesis that Hox proteins regulate many direct target genes

in a slightly different manner, i.e. a majority of all genes are under Hox protein control[172] (***Table 2***).

Strong differential regulation of few Hox targets "Key" Hox targets / co-selective binding	**Slight differential regulation of many Hox targets** Many Hox targets /activity regulation	**Results**
- few direct target genes	- many direct target genes	- somewhere in-between
- regulation of a Hox target gene primarily by one Hox protein only	- regulation of Hox target genes is primarily shared by multiple Hox proteins	- one third of Hox targets regulated primarily by one Hox protein, two thirds shared
- regulation of primarily genes encoding other transcription factors or signaling molecules	- regulation of all kinds of target genes possible	- all kinds of targets are regulated
- either specific binding to the DNA (cooperation) or highly differential activity regulation	- unspecific binding to the DNA, some degree of specificity achieved alone or by activity regulation	- both, cooperative specific binding to DNA and unspecific binding has been observed, as well as activity regulation

Table 2: Two models. Our results indicate that a third hypothesis may be true combining both mechanisms proposed, strong differential regulation of few Hox target genes and slight differential regulation of many Hox target genes. Both mechanisms may be applied to achieve differential regulation of morphological pathways and thus differential development of morphologies between Hox proteins.

The second step was to analyse the specificity of Hox downstream and direct target gene regulation. The most striking finding was that, despite the similar DNA-binding specificity exhibited by Hox proteins *in vitro* and *in vivo*, the majority of the downstream genes appear to be regulated primarily by a single Hox protein *in vivo*. This is the strongest result in favour of the idea of "strong differential regulation of few direct Hox targets". However, when examined from the viewpoint of an individual Hox protein, one third of its downstream genes and presumed direct targets were uniquely regulated by that Hox protein. This is a quite remarkably high proportion when considering the relatively low DNA-binding specificity of Hox monomers. That two thirds of the target genes of an individual Hox protein are shared with other Hox proteins, also points to the alternative mechanism of "slight differential regulation of many downstream genes" being used by Hox proteins for differential regulation *in vivo* (***Table 2***).

The last step in attempting to differentiate between these two hypotheses

was the analysis of the types of genes controlled by Hox proteins. In contrast to the idea that Hox proteins preferentially act on other key regulators, mainly other transcription factors, the newly identified set of Hox downstream genes and the predicted set of direct target genes for the Hox protein **Dfd** did not show an over-representation of transcription factors. Instead, Hox proteins appeared to indiscriminately regulate all levels of the regulatory cascade and, consequently, quite a number of direct target genes of the realisator type. Hox proteins, therefore, appear to have the ability to influence a variety of different processes directly. This results contradicts the previously prevalent idea of preferential action of Hox proteins on transcription factors (***Table 2***).

To recapitulate, our data strongly suggest that, although a number of prior experiments supported the use of strong regulation of few target genes, a new model may apply *in vivo* in which a combination of both mechanisms are used to achieve differential regulation of Hox downstream genes. Our predicted number of Hox direct target genes was in-between the numbers suggested by proponents of either of the two models. In addition, more than 60% of the targets of an individual Hox protein were shared with at least one other Hox protein and Hox proteins could be shown not only to regulate the transcription of key regulators, but also realisator genes as well as a variety of other gene classes. We therefore believe that the mechanism for differential regulation employed by Hox proteins is dependent on the particular Hox DNA-binding sites present in the regulatory region of a Hox target gene. The mechanism for differential regulation may consequently be different from target gene to target gene and morphology to morphology. Hox/**Exd** sites, for example, allow for strong differential regulation. Clusters of Hox monomer sites, in contrast, allow slight differences in the properties of Hox proteins, such as DNA-binding and regulatory activity, to accumulate on the same target, thereby resulting in an effect on Hox target gene transcription. The mechanism relying on the clustering of Hox-binding sites is more likely to require regulation of additional genes with similar morphological effects, an aspect that appears to have been neglected. However, both our experiments and other studies have accumulated data in favour of the theory of “slight differential regulation of many Hox targets”. In other studies, Q50 homeodomain proteins were shown to exhibit not only low DNA-binding specificity *in vitro*, but were also shown to bind to thousands of sites *in vivo*[32,177]. As such, greater emphasis should be placed on elucidating the mechanism for differential regulation of many Hox downstream genes. This mechanism may not only be applicable to all Hox proteins or other Q50 homeodomain proteins, but a general mechanism applicable to many additional transcription factors, in particular those of high DNA-binding

affinity and low specificity. Such transcription factors include Zn-finger and leucine zipper classes, which, apparently, have distinct effects on morphology, yet also bind to thousands of DNA sites *in vivo*[177].

Other analyses we carried out included a linking of identified Hox downstream genes to morphological features. For this, we re-examined morphological phenotypes caused by loss-of-function mutation in the Hox gene *Dfd* in the maxillary segment of *Drosophila* embryonic stages 11 and 12. We confirmed known and identified one new *Dfd* dependent phenotype, a cell shape phenotype on the ventral side of the maxillary segment. We were also able to link the **Dfd** dependent morphological features to known downstream genes and identified new candidate genes. In addition, we examined whether we were able to identify novel or extend known Hox dependent morphological regulatory networks. We were able to identify both one novel and extend upon one known regulatory Hox downstream network. Therefore, our data may be of help to elucidate the genes and regulatory networks participating in Hox dependent morphogenesis. This may prove particularly valuable for morphological features for which no key downstream genes have been identified, such as the dorsal appendage morphogenesis determined by the Hox proteins **Scr**, **Antp** and **Ubx**, respectively leading to the formation of no appendage (notum), wing and haltere.

The final part of our analysis concerns the dependence of Hox downstream gene regulation to the context a Hox protein is placed in. We were able to show that Hox downstream gene regulation is highly spatially as well as temporally context dependent. We suggest that, in particular, segmentally repeated transcription factors play a major role. To some extent, contextual information may be provided by additive or subtractive effects with other transcription factors acting independent of Hox proteins on the regulatory DNA-region of a target gene. However, the high context dependence combined with the fact that many of the total set of Hox downstream genes identified are specific primarily to a single Hox protein suggests, that Hox proteins are likely to act in concert with other factors (directly or indirectly) to achieve differential regulation of target genes. Examples for such concerted action or synergistic effects on Hox target gene regulation are known. While the best described synergistic effect is the cooperation of Hox proteins with DNA-binding specificity enhancing proteins, this type of synergistic effect is unlikely to be used as a mechanism providing contextual information. Only a single Hox DNA-binding specificity enhancing interaction partner has so far been identified (**Exd**), which, in addition, is expressed ubiquitously in the stages we analysed[192]. Thus **Exd** cannot be the sole factor providing the required contextual information. Furthermore, Hox and other homeodomain

proteins have the ability to bind to many sites within the genome. As only one factor using a Hox DNA-binding specificity enhancing mechanism has been identified, it seems likely that this mechanism is not required in many cases. Other, non Hox DNA-binding specificity enhancing synergistic interactions may be used to achieve differential effects on morphology. Correspondingly, quite a number of synergistic effects of Hox proteins with other transcription factors have recently been described. Such factors include the transcription factors **Mad** and **Med**, **Slp** and **En**[72,176]. It is possible that differential effects on transcription are primarily determined by synergistic interactions of Hox proteins with spatially and temporally restricted factors. The observable synergistic effects and recent findings showing Hox proteins to act in conjunction with signalling cascades in the regulation of their target genes[71], caused us to review our premises regarding the mechanism by which Hox proteins achieve differential regulation of their downstream genes. We believe that Hox proteins should no longer be viewed as master regulators in the sense that they act alone and are the sole determinant for the formation of morphological structures, but rather that Hox proteins are master regulators in the sense that they interact with a huge variety of factors in their surrounding environment and on the DNA strand. In fact, it may precisely be their low DNA-binding specificity that makes Hox proteins such efficient master regulators. It allows them to bind to many sites and interact with many factors and potential differences in regulation could be attributed to Hox proteins differing in their ability to create synergistic effects with additional factors. Alternatively, Hox proteins may differ in their ability to make the DNA-strand more accessible to other factors e.g. DNA-bending. Thus, the composition and availability of factors at certain spatial or temporal points, may allow one Hox protein to achieve an effect while another cannot.

It is apparent that Hox proteins cannot account for different morphological structures on their own. In many cases, Hox proteins have been shown to require the presence of other transcription factors to exert their function[71,72,176]. These transcription factors usually have effects on transcription independent of Hox proteins and it may be time to abandon the term “cofactor” in favour of the word “co-transcription factor”, emphasizing the equal importance of the Hox protein and the interaction partner.

4 Outlook

Our results indicate that a significant proportion of Hox controlled morphological differences are likely to be due to Hox proteins effecting slight changes in regulation of many downstream genes rather than large regulatory changes of few targets. This leads us to suggest following modifications to future experimental approaches:

1) Future attempts at identifying Hox target genes should also attempt to elucidate potential Hox dependent networks rather than focusing solely on identification of key downstream transcription factors that may or may not exist for a particular morphological feature.
2) Studies attempting identification of Hox protein DNA-binding sites may need to include lower affinity sites. In conjunction with additional factors acting in synergy with Hox proteins, lower affinity sites may be highly relevant.
3) Analyses geared towards examining Hox dependent differential downstream gene regulation have to take into account factors that may act in synergy with the Hox protein of interest, but also have to take into account that such synergistic interactions do not necessarily require changes in DNA-binding specificity or affinity of the Hox protein itself. As the latter mechanism is not very well understood, it would be highly desirable to gain insights specifically into the molecular mechanisms responsible for the synergistic effects that do not enhance Hox DNA-binding properties.

Furthermore, to understand how Hox proteins shape morphologies and to link downstream networks and realisator genes to Hox mutant phenotypes, a thorough re-analysis of Hox dependent phenotypes would be advisable. Especially so, as many older methods had difficulties visualizing the interior of the embryo in real-time and may have missed many important features. Digital Scanned Laser Light-sheet Microscopy (DSLM) was recently used to visualize the development of the much more complex vertebrate model organism *Danio rerio*[193]. This method has considerably increased imaging speed and quality while requiring a lower degree of laser intensity, thereby making prolonged exposure of the entire embryo feasible. Using this approach, the location of each cell was traced throughout the first 24 hours of embryonic development in *Danio rerio*. Similarly thorough analyses have not been carried out for *Drosophila*, but as *Drosophila* embryos are much smaller

than their vertebrate counterparts, this method should be easy to adapt to this organism. If DSLM is used for analysis of wild type and Hox mutant embryogenesis, the results could provide invaluable information about the underlying mechanism as well as the spatial and temporal constraints of Hox dependent phenotypes. It may even be possible to elucidate the morphological basis for some of these phenotypes. For example, “head involution defects” are likely to be a secondary effect caused by various other defects occurring earlier on in development. Using DSLM it may be possible to pinpoint exactly at what time-point and which specific cells fail to undergo a required morphological process, the absence of which leads to the observed defect. Considering that we carried out our re-examination of **Dfd** dependent phenotypes within a very short time-frame and were able to identify one new Hox dependent phenotype, I would expect the use of more sophisticated methods such as DSLM, for analyses over longer time-frames to uncover many more, previously not observed, Hox dependent phenotypes.

I would also like to suggest further analysis of the candidate genes we detected for various **Dfd** dependent morphological structures in loss-of-function and gain-of-function mutant embryos. Some candidate genes may require use of conditional mutants as they are likely to have other, Hox independent, functions in early development. Of the candidate genes with phenotypes that can be correlated to Hox dependent phenotypes, rescue experiments in *Dfd* loss-of-function background should be carried out. In addition to studying specific candidate genes, it may be advantageous to analyse our data-sets specifically in regards to identification of further Hox dependent networks. Potentially more than just one gene may be required to induce the formation of a specific morphology in absence of the corresponding Hox protein. We showed that it is possible to identify new Hox dependent networks using our newly identified sets of Hox downstream genes. Determining the hierarchy of such networks may be facilitated in conjunction with prediction and experimental validation of direct Hox targets. Thus, either by identifying direct realisator Hox target genes or by linking morphological networks to Hox proteins, our set of Hox downstream genes and predicted direct targets may help to provide a link between Hox proteins and their observable morphological downstream effects.

The answer to understanding how Hox proteins achieve morphological differences, however, lies in understanding how Hox proteins affect transcription of their target genes at the molecular level. For this, we need to understand the two basic aspects of Hox proteins in much greater detail.

A) the DNA-binding properties of Hox proteins

B) how Hox proteins achieve their transcriptional effect once bound to DNA

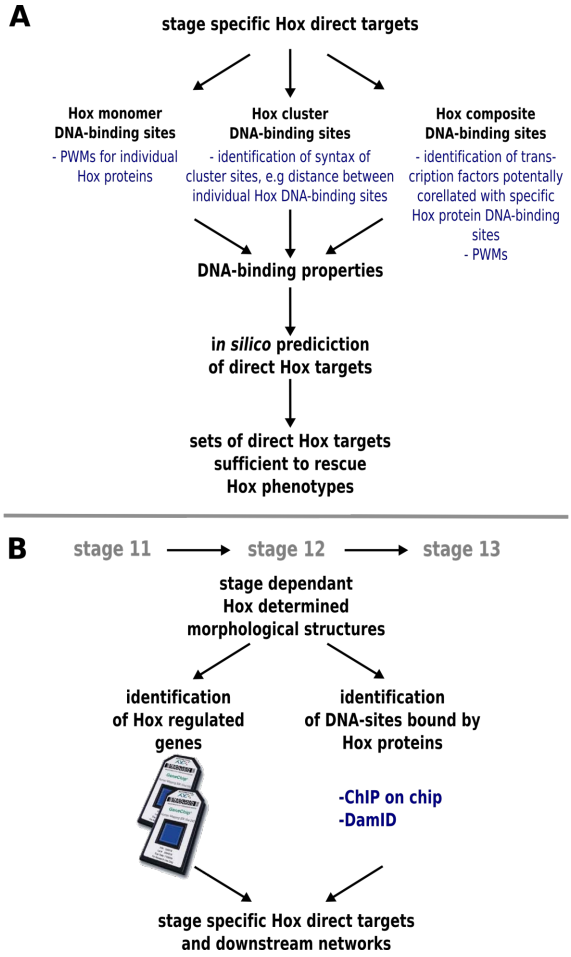

Figure 25: Future directions. A Determination of properties of different Hox proteins to certain DNA sites to be used to predict further direct Hox downstream genes. **B** combining unbiased large scale approaches of gene expression analyzes with DNA-binding site analyzes.

To understand the DNA binding properties of Hox proteins, we need to know the absolute values and relative differences of DNA-binding specificity between individual Hox proteins. These would have to be examined for Hox monomer sites, Hox cluster sites, Hox composite sites such as Hox/**Exd** sites, as well as monomer versus composite sites *in vitro* (***Figure 25A***). The difficulty lies in providing comparable data sets for every Hox protein, regarding their DNA-binding affinities to the different DNA-sites. The relevance of individual nucleotides and the preferred relative distances of Hox-binding DNA sites in relation to each other as well as to other transcription factor binding sites, for example **Exd**, would have to be examined. Obviously, the more accurate, qualitative and, most importantly, the more comparative the analyses are between different Hox proteins, the more precise the subsequent *in silico* prediction for DNA sites capable of binding Hox proteins would become.

To analyse the accuracy of any *in silico* prediction for protein-binding DNA sites, including our own *in silico* prediction for the Hox protein **Dfd**, it will be necessary to identify protein-bound DNA sites *in vivo*, for example using the large scale approaches ChIP on chip or DamID. Moreover, as non-functional sites, i.e. DNA-sites bound by a protein with no transcriptional effect, have repeatedly been suggested for a variety of transcription factors[177], the identified DNA sites bound by Hox proteins would have to be linked to a transcriptional output. As we have already carried out a large-scale approach

for the identification of Hox downstream genes, ideally, the large-scale approach for identification of DNA sites binding Hox proteins would be carried out using the same developmental stages (***Figure 25B***). This would allow a direct comparison of the Hox DNA-binding preferences of the six Hox proteins with their downstream regulatory effects provided by our microarray experiment. The set of potential Hox direct targets could thereby be narrowed down to those genes with DNA sites occupied by Hox proteins coupled with an observable effect on transcription. Combining knowledge about which DNA-sites are likely to be functional with knowledge about the degree to which Hox proteins bind to these sites *in vitro* would facilitate tracing differences in transcriptional activity between Hox proteins back to differences in the ability of Hox proteins to bind to specific sites on the DNA. Those transcriptional differences that cannot be traced back to differences in DNA binding specificity may be caused by specific protein-protein interactions with additional factors. To elucidate the differences in transcriptional activity of Hox proteins, the most promising approach would be to select regulatory DNA sequences that show no preferential binding by a particular Hox protein, but whose genes are regulated differentially between Hox proteins. As protein-protein interaction analyses had some, albeit limited, success in helping to understand the differences in transcriptional activity, it should be taken into account, that the effect on transcription may occur via transient protein-protein interactions. Although transient interactions are difficult to detect and may require the presence of a specific region of DNA, analysis of such transient interactions may be worthwhile. Transient interactions may be detected by analysing DNA sites and associated factors in close proximity of the Hox protein DNA-binding sites. Knowledge of the factors involved may facilitate discriminating between functional and non-functional DNA sites bound by Hox proteins. This would help future *in silico* predictions to distinguish between these two types of sites, and thereby facilitate the identification of further sets of direct Hox target genes.

In future, the insight gained from the above approaches will hopefully allow us to detect the target genes responsible for certain Hox dependent morphologies and thereby elucidate how the expression of different Hox proteins leads to the creation of different morphological structures along the anterior-posterior body axis.

Contributions

The numerous discussions I have had with Ingrid Lohmann throughout the course of this Ph.D. make it difficult to disentangle the individual ideas each of us contributed. Therefore, unless stated otherwise, joint discussions between Ingrid Lohmann and myself are to be regarded as having generated the ideas presented herein.

Comparative analysis of Hox downstream genes in *Drosophila*

The microarray experiment, QPCR experiment and the corresponding analyses, as well as various antibody stainings were carried out by myself. The numerous RNA *in situ* hybridisations necessary for the verification of the identified Hox downstream genes and of candidate genes potentially responsible for **Dfd** dependent phenotypes, were carried out by myself and two student assistants, Martina Blank and Haijia Wu. For analysis of the loss-of-function RNA *in situ* hybridizations, I received extensive help from Ingrid Lohmann. The direct target prediction program was developed by Daniela Bezdan and Stefan Henz and the verification of the direct targets by EMSA was carried out by Daniela Bezdan.

Shaping Segments: Hox gene function in the genomic age

This publication resulted from a joint effort between Ingrid Lohmann and myself. The manuscript was first written by myself and then thoroughly edited by Ingrid Lohmann, making the final, published version very much a joint effort. Figures 2 and 4 were primarily generated by Ingrid Lohmann, with feedback from myself, while Figures 1 and 3 as well as Tables 1 and 2 were primarily generated by myself with corresponding feedback from Ingrid Lohmann.

Multifactorial Regulation of a Hox Target Gene

My contribution to this paper consisted of providing the necessary RNA *in situ* hybridisations showing the time and spatial dependence of Hox downstream genes regulation. This supports the underlying idea that Hox proteins act in concert with other factors to generate the spatial as well as temporal pattern of Hox downstream gene expression.

I would also like to acknowledge extensive feedback form Ingrid Lohmann, Jan Lohmann, Tancred Frickey, Petra Stöbe, Nina Vogt and Karsten Ölkers. Thank you.

Appendix

Tables:

Table 1: Identified direct Hox target genes

adapted from Pearson et al. 2005[16]

Hox target gene	Regulated by Hox protein	Molecular function	Type	Function	Strongest form of validation	Reference
1.28	Dfd	unknown	unknown	unknown	enhancer with mutated Hox sites	Pederson et al., 2000[161]
Antenna-pedia	Antp, Ubx, Abd-A	HD TF	TF	thorax development	enhancers with clusters of mutated Hox sites tested in embryos	Appel and Sakonju, 1993[194]
apterous	Antp	HD TF	TF	muscle identity	footprint + bicoid site swap (K50)	Capovilla et al., 2001[162]
CG11339	Lab	actin-binding	realisator		enhancers with mutated Hox sites tested in embryos	Ebner et al., 2005[195]
CG13222	Ubx	cuticle protein	realisator		band shifts	Hersh et al., 2007[126]
centro-somin	Antp	centrosomal protein (microtubule/centrosome assembly)	realisator	PNS and CNS development	ChIP	Heuer et al., 1995 [196]
connectin	Ubx, Abd-A	GPI linked cell surface protein	realisator	possibly neuromuscular connection	ChIP	Gould and White,1992[197]
decapentaplegic	Ubx, Abd-A	TGF-ß protein	signalling	D/V polarity, imaginal patterning, midgut morphogenesis	bicoid site swap (K50)	Capovilla et al., 1994.[198]; Sun et al., 1995[199]; Capovilla and Botas, 1998[200]
Deformed	Dfd	HD TF	TF	head development	enhancers with mutated Hox sites tested in embryos	Zeng et. al 1994[201]
Distal-less	Ubx, Abd-A	HD TF	TF	limb development	enhancers with mutated Hox sites tested on embryos	Vachon et el., 1992[40]
empty spiracles	Abd-B	HD TF	TF	head development, Filzkörper specification	enhancer study	Jones and McGinnis 1993[155]
forkhead	Scr	forkhead domain TF	TF	specification of the terminal region	enhancer with mutated Hox sites tested in embryos	Ryoo and Mann, 1999[58]; Zhou et al., 2001[164]

knot	Ubx	COE TF (+ cell surface receptor (IPT/TIG) domain)	TF	wing imaginal disc development	enhancers with mutated or deleted Hox sites tested in embryos	Hersh and Caroll, 2005[202]
labial	Lab	HD TF	TF	Head development	enhancers tested with mutated Hox site tested in embryos	Grieder et al., 1997[203]
La-related protein	Scr, Ubx	autophagic cell death	realisator		ChIP Larp DNA-bound to Ubx protein ChIP, no gel shift or reporter lines, nothing tested for direct binding in Scr	Chauvet et al., 2000[165]
reaper	Dfd	apoptosis activator	realisator	Mandibular/ maxillary boundary formation	enhancer study with mutated Dfd sites	Lohmann et al., 2002, Cell[95]
serpent	Ubx	Zn – finger TF	TF		One-hybrid	Mastick, 1995 [121]
scabrous	Ubx	secreted signal transducer	signalling	eye morphogenesis, possibly CNS and PNS development	ChIP	Graba et al., 1992[119]
spalt major	Ubx	Zn finger TF	TF	wing disc development	enhancers with mutated Hox sites tested in larvae	Galant et al. , 2002[43]
Transcript 48	Ubx	transmembrane protein	unknown	unknown, expressed in embryo	ChIP	Gould et al., 1990[204] Strutt and White, 1994[205]
teashirt	Antp, Ubx	Zn-finger TF	TF	trunk segmental identity	enhancers with deleted Hox sites tested in embryos	McCormick et al., 1995[168]
ß-3-tubulin	Ubx	cytoskeletal protein	realisator	possibly SM and VM differentiation	enhancers tested with mutated or deleted Hox sites tested in embryos	Hinz et al., 1992[157]; Kremser et al., 1999[158]
wingless	Abd-A	Wnt signal transducer	signalling	midgut morphogenesis	enhancers with mutated or deleted Hox sites tested in embryos	Grienenberger et al., 2003[71]
Wnt-4	Ubx	Wnt protein	signalling	Possibly epidermal patterning and midgut morphogenesis	ChIP	Graba et al., 1995[206]

TF = transcription factor

HD = homeodomain

Table 2: large-scale Hox downstream gene identification

paper	Hox genes	organism	tissue	stage	# targets found *
Leemanns et al., 2001[124]	*Lab*	*Drosophila melanogaster* (fly)	whole embryo	embryonic stage 10-17	96
Mohit et al., 2006[125]	*Ubx*	*Drosophila melanogaster*	haltere and wing disc	3rd instar larvae	541
Hersh et al., 2007[126]	*Ubx*	*Drosophila melanogaster*	haltere and wing disc	3rd instar larvae	447
Hueber et al., 2007[138]	*Dfd*, *Scr*, *Antp*, *Ubx*, *Abd-A*, *Abd-B*	*Drosophila melanogaster*	whole embryo	embryonic stage 11+ 12	1508
Shen et al., 2000[127]	*HoxA1*	*Mus musculus* (mouse)	cell culture – teratocarcinoma	-	28
Zhao and Potter, 2001[128]	*HoxA13*	*Mus musculus*	uterus and cervix tissue	4.5 weeks old	unknown
Valerius et al., 2002[129]	*HoxA11*	*Mus musculus*	kidney tissue	embryonic stage 18.5	10
Hedlund et al., 2004[130]	*HoxD10*	*Mus musculus*	spinal cord tissue	embryonic stage 12.5	69
Martinez-Ceballos et al., 2005[131]	*HoxA1*	*Mus musculus*	cell culture – embryonic blastocysts	-	145
Lei et al., 2005[132]	*HoxC8*	*Mus musculus*	cell culture – embryonic fribroblasts	-	34
Cobb et al., 2005[133]	entire *HoxD* cluster	*Mus musculus*	mouse tissue of limbs and genitalia	embryonic stage 12.5	16
Williams et al., 2005[134]	*HoxA13*	*Mus musculus*	cell culture – embryonic fribroblasts	-	68
Schwab et al., 2006[135]	*HoxA11* and *HoxD11*	*Mus musculus*	whole embryonic kidneys and urogenital tissue	embryonic stage 11.5, 12.5 13.5, 16.5 + adult	1518 (Affy)
Rohrschneider et al., 2007[136]	*HoxB1a* (*lab* homolog)	*Danio rerio* (zebrafish)	whole embryo	19-20 hours post fertilization	471 (1.3 fold) 113 (1.5 fold)
Ferrell et al., 2005[137]	*HoxA10*	*Homo sapiens* (human)	cell culture – umbilical cord cells	-	115

Bibliography

[1] Botas, J;. Control of morphogenesis and differentiation by hom/hox genes. *Curr Opin Cell Biol* (1993) 5: pp. 1015-1022.

[2] Graba, Y; Aragnol, D; Pradel, J;. Drosophila hox complex downstream targets and the function of homeotic genes. *Bioessays* (1997) 19: pp. 379-388.

[3] Pradel, J; White, RA;. From selectors to realizators. *Int J Dev Biol* (1998) 42: pp. 417-421.

[4] Gilbert, SF;. Developmental biology, 8th edition. SA, 2006.

[5] Driever, W; Nusslein-Volhard, C;. The bicoid protein determines position in the drosophila embryo in a concentration-dependent manner. *Cell* (1988) 54: pp. 95-104.

[6] Nüsslein-Volhard, C; Wieschaus, E;. Mutations affecting segment number and polarity in drosophila. *Nature* (1980) 287: pp. 795-801.

[7] Lawrence, PA;. Pair-rule genes: do they paint stripes or draw lines?. *Cell* (1987) 51: pp. 879-880.

[8] Lawrence, PA; Johnston, P;. Pattern formation in the drosophila embryo: allocation of cells to parasegments by even-skipped and fushi tarazu. *Development* (1989) 105: pp. 761-767.

[9] Martinez-Arias, A; Lawrence, PA;. Parasegments and compartments in the drosophila embryo. *Nature* (1985) 313: pp. 639-642.

[10] McGinnis, W; Krumlauf, R;. Homeobox genes and axial patterning. *Cell* (1992) 68: pp. 283-302.

[11] Mann, RS; Morata, G;. The developmental and molecular biology of genes that subdivide the body of drosophila. *Annu Rev Cell Dev Biol* (2000) 16: pp. 243-271.

[12] Duboule, D; Dolle, P;. The structural and functional organization of the murine hox gene family resembles that of drosophila homeotic genes. *Embo J* (1989) 8: pp. 1497-1505.

[13] Graham, A; Papalopulu, N; Krumlauf, R;. The murine and drosophila homeobox gene complexes have common features of organization and expression. *Cell* (1989) 57: pp. 367-378.

[14] McGinnis, W; Jack, T; Chadwick, R; Regulski, M; Bergson, C; McGinnis, N; Kuziora, MA;. Establishment and maintenance of position-specific expression of the drosophila homeotic selector gene deformed. *Adv Genet* (1990) 27: pp. 363-402.

[15] Zhao, JJ; Lazzarini, RA; Pick, L;. The mouse hox-1.3 gene is functionally equivalent to the drosophila sex combs reduced gene. *Genes Dev* (1993) 7: pp. 343-354.

[16] Pearson, JC; Lemons, D; McGinnis, W;. Modulating hox gene functions during animal body patterning. *Nature Genetics Reviews* (2005) 6: pp. 893-904.

[17] Ekker, SC; von Kessler, DP; Beachy, PA;. Differential dna sequence

recognition is a determinant of specificity in homeotic gene action. *Embo J* (1992) 11: pp. 4059-4072.

[18] Amores, A; Force, A; Yan, YL; Joly, L; Amemiya, C; Fritz, A; Ho, RK; Langeland, J; Prince, V; Wang, YL;. Zebrafish hox clusters and vertebrate genome evolution. *Science* (1998) 282: pp. 1711-1714.

[19] Duboule, D;. The rise and fall of hox gene clusters. *Development* (2007) 134: pp. 2549-2560.

[20] Duboule, D; Tarchini, B; Zàkàny, J; Kmita, M;. Tinkering with constraints in the evolution of the vertebrate limb anterior-posterior polarity. *Novartis Found Symp* (2007) 284: p. 130-7; discussion 138-41, 158-63.

[21] Lewis, EB;. A gene complex controlling segmentation in drosophila. *Nature* (1978) 276: pp. 565-570.

[22] Sanchez-Herrero, E; Vernos, I; Marco, R; Morata, G;. Genetic organization of drosophila bithorax complex. *Nature* (1985) 313: pp. 108-113.

[23] Kessel, M; Gruss, P;. Murine developmental control genes. *Science* (1990) 249: pp. 374-379.

[24] Lufkin, T; Mark, M; Hart, CP; Dolle, P; LeMeur, M; Chambon, P;. Homeotic transformation of the occipital bones of the skull by ectopic expression of a homeobox gene. *Nature* (1992) 359: pp. 835-841.

[25] Sayle, RA; Milner-White, EJ;. Rasmol: biomolecular graphics for all. *Trends Biochem Sci* (1995) 20: p. 374.

[26] Svingen, T; Tonissen, KF;. Hox transcription factors and their elusive mammalian gene targets. *Heredity* (2006) 97: pp. 88-96.

[27] Scott, MP; Tamkun, JW; Hartzell, G W, 3rd;. The structure and function of the homeodomain. *Biochim Biophys Acta* (1989) 989: pp. 25-48.

[28] Hayashi, S; Scott, MP;. What determines the specificity of action of drosophila homeodomain proteins?. *Cell* (1990) 63: pp. 883-894.

[29] Ekker, SC; Jackson, DG; von Kessler, DP; Sun, BI; Young, KE; Beachy, PA;. The degree of variation in dna sequence recognition among four drosophila homeotic proteins. *Embo J* (1994) 13: pp. 3551-3560.

[30] Gehring, WJ; Qian, YQ; Billeter, M; Furukubo-Tokunaga, K; Schier, AF; Resendez-Perez, D; Affolter, M; Otting, G; Wuthrich, K;. Homeodomain-dna recognition. *Cell* (1994) 78: pp. 211-223.

[31] Biggin, MD; McGinnis, W;. Regulation of segmentation and segmental identity by drosophila homeoproteins: the role of dna binding in functional activity and specificity. *Development* (1997) 124: pp. 4425-4433.

[32] Walter, J; Dever, CA; Biggin, MD;. Two homeo domain proteins bind with similar specificity to a wide range of dna sites in drosophila embryos. *Genes Dev* (1994) 8: pp. 1678-1692.

[33] Carr, A; Biggin, MD;. A comparison of in vivo and in vitro dna-binding specificities suggests a new model for homeoprotein dna binding in drosophila embryos. *Embo J* (1999) 18: pp. 1598-1608.

[34] Li, X; Murre, C; McGinnis, W;. Activity regulation of a hox protein and a role

for the homeodomain in inhibiting transcriptional activation. *Embo J* (1999) 18: pp. 198-211.

[35] Tour, E; Hittinger, CT; McGinnis, W;. Evolutionarily conserved domains required for activation and repression functions of the drosophila hox protein ultrabithorax. *Development* (2005) 132: pp. 5271-5281.

[36] Zhao, JJ; Lazzarini, RA; Pick, L;. Functional dissection of the mouse hox-a5 gene. *EMBO J* (1996) 15: pp. 1313-1322.

[37] Gehring, WJ; Affolter, M; Burglin, T;. Homeodomain proteins. *Annu Rev Biochem* (1994) 63: pp. 487-526.

[38] 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) 131: pp. 530-543.

[39] Qian, YQ; Otting, G; Furukubo-Tokunaga, K; Affolter, M; Gehring, WJ; Wüthrich, K;. Nmr structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the drosophila antennapedia protein. *Proc Natl Acad Sci U S A* (1992) 89: pp. 10738-10742.

[40] 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) 71: pp. 437-450.

[41] Percival-Smith, A; Teft, WA; Barta, JL;. Tarsus determination in drosophila melanogaster. *Genome* (2005) 48: pp. 712-721.

[42] Mannervik, M;. Target genes of homeodomain proteins. *Bioessays* (1999) 21: pp. 267-270.

[43] Galant, R; Walsh, CM; Carroll, SB;. Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* (2002) 129: pp. 3115-3126.

[44] Amoutzias, GD; Robertson, DL; Van de Peer, Y; Oliver, SG;. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem Sci* (2008) 33: pp. 220-229.

[45] Wilson, DS; Guenther, B; Desplan, C; Kuriyan, J;. High resolution crystal structure of a paired (pax) class cooperative homeodomain dimer on dna. *Cell* (1995) 82: pp. 709-719.

[46] Wolberger, C; Vershon, AK; Liu, B; Johnson, AD; Pabo, CO;. Crystal structure of a mat alpha 2 homeodomain-operator complex suggests a general model for homeodomain-dna interactions. *Cell* (1991) 67: pp. 517-528.

[47] Hirsch, JA; Aggarwal, AK;. Structure of the even-skipped homeodomain complexed to at-rich dna: new perspectives on homeodomain specificity. *Embo J* (1995) 14: pp. 6280-6291.

[48] Kissinger, CR; Liu, BS; Martin-Blanco, E; Kornberg, TB; Pabo, CO;. Crystal structure of an engrailed homeodomain-dna complex at 2.8 a resolution: a framework for understanding homeodomain-dna interactions. *Cell* (1990) 63: pp. 579-590.

[49] Neuteboom, ST; Murre, C;. Pbx raises the dna binding specificity but not the selectivity of antennapedia hox proteins. *Mol Cell Biol* (1997) 17: pp. 4696-4706.

[50] Phelan, ML; Featherstone, MS;. Distinct hox n-terminal arm residues are responsible for specificity of dna recognition by hox monomers and hox.pbx heterodimers. *J Biol Chem* (1997) 272: pp. 8635-8643.

[51] 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) 16: pp. 2032-2042.

[52] Beachy, PA; Varkey, J; Young, KE; von Kessler, DP; Sun, BI; Ekker, SC;. Cooperative binding of an ultrabithorax homeodomain protein to nearby and distant dna sites. *Mol Cell Biol* (1993) 13: pp. 6941-6956.

[53] Featherstone, M;. Hox proteins and their co-factors in transcriptional regulation. *Murine Homeobox Gene Control of Embryonic Patterning and Organogenesis* (2003) 13: pp. 1-42.

[54] Mann, RS; Chan, SK;. Extra specificity from extradenticle: the partnership between hox and pbx/exd homeodomain proteins. *Trends Genet* (1996) 12: pp. 258-262.

[55] Mann, RS; Affolter, M;. Hox proteins meet more partners. *Curr. Opin. Genet. Dev* (1998) 8: pp. 423-429.

[56] LaRonde-LeBlanc, NA; Wolberger, C;. Structure of hoxa9 and pbx1 bound to dna: hox hexapeptide and dna recognition anterior to posterior. *Genes Dev* (2003) 17: pp. 2060-2072.

[57] Ryoo, HD; Mann, RS;. The control of trunk hox specificity and activity by extradenticle. *Genes Dev* (1999) 13: pp. 1704-1716.

[58] 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) 126: pp. 5137-5148.

[59] Spirov, AV; Bowler, T; Reinitz, J;. Hox pro: a specialized database for clusters and networks of homeobox genes. *Nucleic Acids Res* (2000) 28: pp. 337-340.

[60] Kel, OV; Romaschenko, AG; Kel, AE; Wingender, E; Kolchanov, NA;. A compilation of composite regulatory elements affecting gene transcription in vertebrates. *Nucleic Acids Res* (1995) 23: pp. 4097-4103.

[61] Delano, WL;. Pymol. *Delano Scientific, San Carlos, CA* (2002) : .

[62] Sprules, T; Green, N; Featherstone, M; Gehring, K;. Lock and key binding of the hox ypwm peptide to the pbx homeodomain. *J. Biol. Chem* (2003) 278: pp. 1053-1058.

[63] Jacobs, Y; Schnabel, CA; Cleary, ML;. Trimeric association of hox and tale homeodomain proteins mediates hoxb2 hindbrain enhancer activity. *Mol Cell Biol* (1999) 19: pp. 5134-5142.

[64] 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) 126: pp. 5137-5148.

[65] Gebelein, B; Culi, J; Ryoo, HD; Zhang, W; Mann, RS;. Specificity of distalless repression and limb primordia development by abdominal hox proteins. *Dev Cell* (2002) 3: pp. 487-498.

[66] Gebelein, B; McKay, DJ; Mann, RS;. Direct integration of hox and segmentation gene inputs during drosophila development. *Nature* (2004) 431: pp. 653-659.

[67] Prince, F; Katsuyama, T; Oshima, Y; Plaza, S; Resendez-Perez, D; Berry, M; Kurata, S; Gehring, WJ;. The ypwm motif links antennapedia to the basal transcriptional machinery. *Development* (2008) 135: pp. 1669-1679.

[68] Taghli-Lamallem, O; Gallet, A; Leroy, F; Malapert, P; Vola, C; Kerridge, S; Fasano, L;. Direct interaction between teashirt and sex combs reduced proteins, via tsh's acidic domain, is essential for specifying the identity of the prothorax in drosophila. *Dev Biol* (2007) 307: pp. 142-151.

[69] Hueber, SD; Lohmann, I;. Shaping segments: hox gene function in the genomic age. *Bioessays* (2008) 30: pp. 965-979.

[70] Li, X; Veraksa, A; McGinnis, W;. A sequence motif distinct from hox binding sites controls the specificity of a hox response element. *Development* (1999) 126: pp. 5581-5589.

[71] Grienenberger, A; Merabet, S; Manak, J; Iltis, I; Fabre, A; Berenger, H; Scott, MP; Pradel, J; Graba, Y;. Tgfbeta signaling acts on a hox response element to confer specificity and diversity to hox protein function. *Development* (2003) 130: pp. 5445-5455.

[72] Walsh, CM; Carroll, SB;. Collaboration between smads and a hox protein in target gene repression. *Development* (2007) 134: pp. 3585-3592.

[73] Stoebe, P; Stein, MAS; Habring-Mueller, A; Bezdan, D; Fuchs, AL; Hueber, S; Wu, H; Lohmann, I;. Segment morphogenesis by multifactorial regulation of the drosophila hox target gene reaper.

[74] Reuter, R; Panganiban, GE; Hoffmann, FM; Scott, MP;. Homeotic genes regulate the spatial expression of putative growth factors in the visceral mesoderm of drosophila embryos. *Development* (1990) 110: pp. 1031-1040.

[75] Udagawa, Y; Hanai, J; Tada, K; Grieder, NC; Momoeda, M; Taketani, Y; Affolter, M; Kawabata, M; Miyazono, K;. Schnurri interacts with mad in a dpp-dependent manner. *Genes Cells* (2000) 5: pp. 359-369.

[76] Saleh, M; Rambaldi, I; Yang, XJ; Featherstone, MS;. Cell signaling switches hox-pbx complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases. *Mol Cell Biol* (2000) 20: pp. 8623-8633.

[77] Taghli-Lamallem, O; Hsia, C; Ronshaugen, M; McGinnis, W;. Context-dependent regulation of hox protein functions by ck2 phosphorylation sites. *Dev Genes Evol* (2008) 218: pp. 321-332.

[78] Perrin, L; Monier, B; Ponzielli, R; Astier, M; Semeriva, M;. Drosophila cardiac tube organogenesis requires multiple phases of hox activity. *Dev Biol* (2004) 272: pp. 419-431.

[79] Brodu, V; Elstob, PR; Gould, AP;. Abdominal a specifies one cell type in drosophila by regulating one principal target gene. *Development* (2002) 129: pp. 2957-2963.

[80] Bateson, WPOBAC;. Materials for the study of variation, treated with especial regard to discontinuity in the origin of species. pp. xv. 598. Macmillan & Co.: London, 1894.

[81] Postlethwait, JH; Schneiderman, HA;. Pattern formation and determination in the antenna of the homoeotic mutant antennapedia of drosophila melanogaster. *Dev Biol* (1971) 25: pp. 606-640.

[82] Stuart, JJ; Brown, SJ; Beeman, RW; Denell, RE;. A deficiency of the homeotic complex of the beetle tribolium. *Nature* (1991) 350: pp. 72-74.

[83] Casares, F; Mann, RS;. The ground state of the ventral appendage in drosophila. *Science* (2001) 293: pp. 1477-1480.

[84] Emerald, BS; Cohen, SM;. Spatial and temporal regulation of the homeotic selector gene antennapedia is required for the establishment of leg identity in drosophila. *Dev Biol* (2004) 267: pp. 462-472.

[85] Denell, RE; Brown, SJ; Beeman, RW;. Evolution of the organization and function of insect homeotic complexes. In . 1996.

[86] Panganiban, G; Irvine, SM; Lowe, C; Roehl, H; Corley, LS; Sherbon, B; Grenier, JK; Fallon, JF; Kimble, J; Walker, M;. The origin and evolution of animal appendages. *Proc Natl Acad Sci U S A* (1997) 94: pp. 5162-5166.

[87] Hughes, CL; Kaufman, TC;. Hox genes and the evolution of the arthropod body plan. *Evol Dev* (2002) 4: pp. 459-499.

[88] Carroll, SB; Weatherbee, SD; Langeland, JA;. Homeotic genes and the regulation and evolution of insect wing number. *Nature* (1995) 375: pp. 58-61.

[89] Struhl, G;. Genes controlling segmental specification in the drosophila thorax. *Proc Natl Acad Sci U S A* (1982) 79: pp. 7380-7384.

[90] Abbott, MK; Kaufman, TC;. The relationship between the functional complexity and the molecular organization of the antennapedia locus of drosophila melanogaster. *Genetics* (1986) 114: pp. 919-942.

[91] Kurata, S; Go, MJ; Artavanis-Tsakonas, S; Gehring, WJ;. Notch signaling and the determination of appendage identity. *Proc Natl Acad Sci U S A* (2000) 97: pp. 2117-2122.

[92] 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) 129: pp. 4509-4521.

[93] Lovato, TL; Nguyen, TP; Molina, MR; Cripps, RM;. The hox gene abdominal-a specifies heart cell fate in the drosophila dorsal vessel. *Development* (2002) 129: pp. 5019-5027.

[94] Lo, PCH; 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) 251: pp. 307-319.

[95] Lohmann, I; McGinnis, N; Bodmer, M; McGinnis, W;. The drosophila hox gene deformed sculpts head morphology via direct regulation of the apoptosis activator reaper. *Cell* (2002) 110: pp. 457-466.

[96] Lewis, RA; Kaufman, TC; Denell, RE; Tallerico, P;. Genetic analysis of the antennapedia gene complex (ant-c) and adjacent chromosomal regions of drosophila melanogaster. i. polytene chromosome segments 84b-d. *Genetics* (1980) 95: pp. 367-381.

[97] Panzer, S; Weigel, D; Beckendorf, SK;. Organogenesis in drosophila melanogaster: embryonic salivary gland determination is controlled by homeotic and dorsoventral patterning genes. *Development* (1992) 114: pp. 49-57.

[98] Reuter, R; Scott, MP;. Expression and function of the homoeotic genes antennapedia and sex combs reduced in the embryonic midgut of drosophila. *Development* (1990) 109: pp. 289-303.

[99] Chauvet, S; Merabet, S; Bilder, D; Scott, MP; Pradel, J; Graba, Y;. Distinct hox protein sequences determine specificity in different tissues. *Proc Natl Acad Sci U S A* (2000) 97: pp. 4064-4069.

[100] Heuer, JG;. Homeotic genes have specific functional roles in the establishment of the drosophila embryonic peripheral nervous system. (1992) 115: pp. 35-47.

[101] Hirth, F; Hartmann, B; Reichert, H;. Homeotic gene action in embryonic brain development of drosophila. *Development* (1998) 125: pp. 1579-1589.

[102] Hu, N; Castelli-Gair, J;. Study of the posterior spiracles of drosophila as a model to understand the genetic and cellular mechanisms controlling morphogenesis. *Dev Biol* (1999) 214: pp. 197-210.

[103] Myers, C; Charboneau, A; Boudreau, N;. Homeobox b3 promotes capillary morphogenesis and angiogenesis. *J Cell Biol* (2000) 148: pp. 343-351.

[104] Myers, C; Charboneau, A; Cheung, I; Hanks, D; Boudreau, N;. Sustained expression of homeobox d10 inhibits angiogenesis. *Am J Pathol* (2002) 161: pp. 2099-2109.

[105] Morata, G;. How drosophila appendages develop. *Nat Rev Mol Cell Biol* (2001) 2: pp. 89-97.

[106] Cohen, SM; Jurgens, G;. Proximal-distal pattern formation in drosophila: cell autonomous requirement for distal-less gene activity in limb development. *Embo J* (1989) 8: pp. 2045-2055.

[107] Cohen, SM; Bronner, G; Kuttner, F; Jurgens, G; Jackle, H;. Distal-less encodes a homoeodomain protein required for limb development in drosophila. *Nature* (1989) 338: pp. 432-434.

[108] Panganiban, G;. Distal-less function during drosophila appendage and sense organ development. *Dev Dyn* (2000) 218: pp. 554-562.

[109] Casares, F; Mann, RS;. Control of antennal versus leg development in drosophila. *Nature* (1998) 392: pp. 723-726.

[110] Dong, PD; Chu, J; Panganiban, G;. Coexpression of the homeobox genes distal-less and homothorax determines drosophila antennal identity.

Development (2000) 127: pp. 209-216.

[111] Rogers, BT; Peterson, MD; Kaufman, TC;. Evolution of the insect body plan as revealed by the sex combs reduced expression pattern. *Development* (1997) 124: pp. 149-157.

[112] 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) 12: pp. 1474-1482.

[113] Mohit, P; Bajpai, R; Shashidhara, LS;. Regulation of wingless and vestigial expression in wing and haltere discs of drosophila. *Development* (2003) 130: pp. 1537-1547.

[114] Akam, M;. Hox genes: from master genes to micromanagers. *Curr Biol* (1998) 8: p. R676-8.

[115] Garcia-Bellido, A;. Genetic control of wing disc development in drosophila. *Ciba Found Symp* (1975) 0: pp. 161-182.

[116] Lovegrove, B; Simoes, S; Rivas, ML; Sotillos, S; Johnson, K; Knust, E; Jacinto, A; Hombria, JC;. Coordinated control of cell adhesion, polarity, and cytoskeleton underlies hox-induced organogenesis in drosophila. *Curr Biol* (2006) 16: pp. 2206-2216.

[117] Merrill, VK; Turner, FR; Kaufman, TC;. A genetic and developmental analysis of mutations in the deformed locus in drosophila melanogaster. *Dev Biol* (1987) 122: pp. 379-395.

[118] Wagner-Bernholz, JT; Wilson, C; Gibson, G; Schuh, R; Gehring, WJ;. Identification of target genes of the homeotic gene antennapedia by enhancer detection. *Genes Dev* (1991) 5: pp. 2467-2480.

[119] Graba, Y; Aragnol, D; Laurenti, P; Garzino, V; Charmot, D; Berenger, H; Pradel, J;. Homeotic control in drosophila; the scabrous gene is an in vivo target of ultrabithorax proteins. *Embo J* (1992) 11: pp. 3375-3384.

[120] Feinstein, PG; Kornfeld, K; Hogness, DS; Mann, RS;. Identification of homeotic target genes in drosophila melanogaster including nervy, a proto-oncogene homologue. *Genetics* (1995) 140: pp. 573-586.

[121] Mastick, GS; McKay, R; Oligino, T; Donovan, K; Lopez, AJ;. Identification of target genes regulated by homeotic proteins in drosophila melanogaster through genetic selection of ultrabithorax protein-binding sites in yeast. *Genetics* (1995) 139: pp. 349-363.

[122] Botas, J; Auwers, L;. Chromosomal binding sites of ultrabithorax homeotic proteins. *Mech Dev* (1996) 56: pp. 129-138.

[123] Wu, Z; Irizarry, RA; Gentleman, R; Martinez-Murillo, F; Spencer, F;. A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American Statistical Association* (2004) 99: pp. 909-918.

[124] Leemans, R; Loop, T; Egger, B; He, H; Kammermeier, L; Hartmann, B; Certa, U; Reichert, H; Hirth, F;. Identification of candidate downstream genes for the homeodomain transcription factor labial in drosophila through oligonucleotide-array transcript imaging. *Genome Biol* (2001) 2: p.

RESEARCH0015.

[125] Mohit, P; Makhijani, K; Madhavi, MB; Bharathi, V; Lal, A; Sirdesai, G; Reddy, VR; Ramesh, P; Kannan, R; Dhawan, J;. Modulation of ap and dv signaling pathways by the homeotic gene ultrabithorax during haltere development in drosophila. *Dev Biol* (2006) 291: pp. 356-367.

[126] Hersh, BM; Nelson, CE; Stoll, SJ; Norton, JE; Albert, TJ; Carroll, SB;. The ubx-regulated network in the haltere imaginal disc of d. melanogaster. *Dev Biol* (2007) 302: pp. 717-727.

[127] Shen, J; Wu, H; Gudas, LJ;. Molecular cloning and analysis of a group of genes differentially expressed in cells which overexpress the hoxa-1 homeobox gene. *Exp Cell Res* (2000) 259: pp. 274-283.

[128] Zhao, Y; Potter, SS;. Functional specificity of the hoxa13 homeobox. *Development* (2001) 128: pp. 3197-3207.

[129] Valerius, MT; Patterson, LT; Witte, DP; Potter, SS;. Microarray analysis of novel cell lines representing two stages of metanephric mesenchyme differentiation. *Mech Dev* (2002) 112: pp. 219-232.

[130] Hedlund, E; Karsten, SL; Kudo, L; Geschwind, DH; Carpenter, EM;. Identification of a hoxd10-regulated transcriptional network and combinatorial interactions with hoxa10 during spinal cord development. *J Neurosci Res* (2004) 75: pp. 307-319.

[131] Martinez-Ceballos, E; Chambon, P; Gudas, LJ;. Differences in gene expression between wild type and hoxa1 knockout embryonic stem cells after retinoic acid treatment or leukemia inhibitory factor (lif) removal. *J Biol Chem* (2005) 280: pp. 16484-16498.

[132] Lei, H; Wang, H; Juan, AH; Ruddle, FH;. The identification of hoxc8 target genes. *Proc Natl Acad Sci U S A* (2005) 102: pp. 2420-2424.

[133] Cobb, J; Duboule, D;. Comparative analysis of genes downstream of the hoxd cluster in developing digits and external genitalia. *Development* (2005) 132: pp. 3055-3067.

[134] Williams, TM; Williams, ME; Kuick, R; Misek, D; McDonagh, K; Hanash, S; Innis, JW;. Candidate downstream regulated genes of hox group 13 transcription factors with and without monomeric dna binding capability. *Dev Biol* (2005) 279: pp. 462-480.

[135] Schwab, K; Hartman, HA; Liang, HC; Aronow, BJ; Patterson, LT; Potter, SS;. Comprehensive microarray analysis of hoxa11/hoxd11 mutant kidney development. *Dev Biol* (2006) 293: pp. 540-554.

[136] Rohrschneider, MR; Elsen, GE; Prince, VE;. Zebrafish hoxb1a regulates multiple downstream genes including prickle1b. *Dev Biol* (2007) 309: pp. 358-372.

[137] Ferrell, CM; Dorsam, ST; Ohta, H; Humphries, RK; Derynck, MK; Haqq, C; Largman, C; Lawrence, HJ;. Activation of stem-cell specific genes by hoxa9 and hoxa10 homeodomain proteins in cd34+ human cord blood cells. *Stem Cells* (2005) 23: pp. 644-655.

[138] Hueber, SD; Bezdan, D; Henz, SR; Blank, M; Wu, H; Lohmann, I;. Comparative

analysis of hox downstream genes in drosophila. *Development* (2007) 134: pp. 381-392.

[139]Wakimoto, BT; Turner, FR; Kaufman, TC;. Defects in embryogenesis in mutants associated with the antennapedia gene complex of drosophila melanogaster. *Dev Biol* (1984) 102: pp. 147-172.

[140]Brand, AH; Perrimon, N;. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* (1993) 118: pp. 401-415.

[141]Sanson, B; White, P; Vincent, JP;. Uncoupling cadherin-based adhesion from wingless signalling in drosophila. *Nature* (1996) 383: pp. 627-630.

[142]Gould, A; Morrison, A; Sproat, G; White, RA; Krumlauf, R;. Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping hox expression patterns. *Genes Dev* (1997) 11: pp. 900-913.

[143]Miller, DF; Rogers, BT; Kalkbrenner, A; Hamilton, B; Holtzman, SL; Kaufman, T;. Cross-regulation of hox genes in the drosophila melanogaster embryo. *Mech Dev* (2001) 102: pp. 3-16.

[144]Schneuwly, S; Klemenz, R; Gehring, WJ;. Redesigning the body plan of drosophila by ectopic expression of the homoeotic gene antennapedia. *Nature* (1987) 325: pp. 816-818.

[145]Zeng, W; Andrew, DJ; Mathies, LD; Horner, MA; Scott, MP;. Ectopic expression and function of the antp and scr homeotic genes: the n terminus of the homeodomain is critical to functional specificity. *Development* (1993) 118: pp. 339-352.

[146]Ashburner, M; Ball, CA; Blake, JA; Botstein, D; Butler, H; Cherry, JM; Davis, AP; Dolinski, K; Dwight, SS; Eppig, JT;. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat Genet* (2000) 25: pp. 25-29.

[147]Harden, N;. Signaling pathways directing the movement and fusion of epithelial sheets: lessons from dorsal closure in drosophila. *Differentiation* (2002) 70: pp. 181-203.

[148]Xia, Y; Karin, M;. The control of cell motility and epithelial morphogenesis by jun kinases. *Trends Cell Biol* (2004) 14: pp. 94-101.

[149]Weber, U; Paricio, N; Mlodzik, M;. Jun mediates frizzled-induced r3/r4 cell fate distinction and planar polarity determination in the drosophila eye. *Development* (2000) 127: pp. 3619-3629.

[150]Wu, J; Cohen, SM;. Proximodistal axis formation in the drosophila leg: subdivision into proximal and distal domains by homothorax and distal-less. *Development* (1999) 126: pp. 109-117.

[151]Carthew, RW;. Gene silencing by double-stranded rna. *Curr Opin Cell Biol* (2001) 13: pp. 244-248.

[152]Smolik-Utlaut, SM;. Dosage requirements of ultrabithorax and bithoraxoid in the determination of segment identity in drosophila melanogaster. *Genetics* (1990) 124: pp. 357-366.

[153]Scanga, S; Manoukian, A; Larsen, E;. Time- and concentration-dependent

response of the drosophila antenna imaginal disc to antennapedia. *Dev Biol* (1995) 169: pp. 673-682.

[154]Lemon, WJ; S, L; M, Y;. A high performance test of differential gene expression for oligonucleotide arrays. *Genome Biol* (2003) 4: p. R67.

[155]Jones, B; McGinnis, W;. The regulation of empty spiracles by abdominal-b mediates an abdominal segment identity function. *Genes Dev* (1993) 7: pp. 229-240.

[156]Castelli-Gair, J;. The lines gene of drosophila is required for specific functions of the abdominal-b hox protein. *Development* (1998) 125: pp. 1269-1274.

[157]Hinz, U; Wolk, A; Renkawitz-Pohl, R;. Ultrabithorax is a regulator of b3 tubulin expression in the drosophila visceral mesoderm. *Development* (1992) 116: pp. 543-554.

[158]Kremser, T; Hasenpusch-Theil, K; Wagner, E; Buttgereit, D; Renkawitz-Pohl, R;. Expression of the beta3 tubulin gene (beta tub60d) in the visceral mesoderm of drosophila is dependent on a complex enhancer that binds tinman and ubx. *Mol Gen Genet* (1999) 262: pp. 643-658.

[159]Vanario-Alonso, CE; O'Hara, E; McGinnis, W; Pick, L;. Targeted ribozymes reveal a conserved function of the drosophila paired gene in sensory organ development. *Mech Dev* (1995) 53: pp. 323-328.

[160]Mahaffey, JW; Jones, DF; Hickel, JA; Griswold, CM;. Identification and characterization of a gene activated by the deformed homeoprotein. *Development* (1993) 118: pp. 203-214.

[161]Pederson, JA; LaFollette, JW; Gross, C; Veraksa, A; McGinnis, W; Mahaffey, JW;. Regulation by homeoproteins: a comparison of deformed-responsive elements. *Genetics* (2000) 156: pp. 677-686.

[162]Capovilla, M; Kambris, Z; Botas, J;. Direct regulation of the muscle-identity gene apterous by a hox protein in the somatic mesoderm. *Development* (2001) 128: pp. 1221-1230.

[163]O'Hara, E; Cohen, B; Cohen, SM; McGinnis, W;. Distal-less is a downstream gene of deformed required for ventral maxillary identity. *Development* (1993) 117: pp. 847-856.

[164]Zhou, B; Bagri, A; Beckendorf, SK;. Salivary gland determination in drosophila: a salivary-specific, fork head enhancer integrates spatial pattern and allows fork head autoregulation. *Dev Biol* (2001) 237: pp. 54-67.

[165]Chauvet, S; Maurel-Zaffran, C; Miassod, R; Jullien, N; Pradel, J; Aragnol, D;. Dlarp, a new candidate hox target in drosophila whose orthologue in mouse is expressed at sites of epithelium/mesenchymal interactions. *Dev Dyn* (2000) 218: pp. 401-413.

[166]Graba, Y; Laurenti, P; Perrin, L; Aragnol, D; Pradel, J;. The modifier of variegation modulo gene acts downstream of dorsoventral and hom-c genes and is required for morphogenesis in drosophila. *Dev Biol* (1994) 166: pp. 704-715.

[167] Wiellette, EL; McGinnis, W;. Hox genes differentially regulate serrate to generate segment-specific structures. *Development* (1999) 126: pp. 1985-1995.

[168] McCormick, A; Core, N; Kerridge, S; Scott, MP;. Homeotic response elements are tightly linked to tissue-specific elements in a transcriptional enhancer of the teashirt gene. *Development* (1995) 121: pp. 2799-2812.

[169] Chiang, C; Young, KE; Beachy, PA;. Control of drosophila tracheal branching by the novel homeodomain gene unplugged, a regulatory target for genes of the bithorax complex. *Development* (1995) 121: pp. 3901-3912.

[170] Chandrasekaran, V; Beckendorf, SK;. Senseless is necessary for the survival of embryonic salivary glands in drosophila. *Development* (2003) 130: pp. 4719-4728.

[171] Chan, SK; Mann, RS;. A structural model for a homeotic protein-extradenticle-dna complex accounts for the choice of hox protein in the heterodimer. *Proc Natl Acad Sci U S A* (1996) 93: pp. 5223-5228.

[172] Liang, Z; Biggin, MD;. Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in drosophila. *Development* (1998) 125: pp. 4471-4482.

[173] Castelli-Gair, J; Akam, M;. How the hox gene ultrabithorax specifies two different segments: the significance of spatial and temporal regulation within metameres. *Development* (1995) 121: pp. 2973-2982.

[174] Cribbs, DL; Benassayag, C; Randazzo, FM; Kaufman, TC;. Levels of homeotic protein function can determine developmental identity: evidence from low-level expression of the drosophila homeotic gene proboscipedia under hsp70 control. *Embo J* (1995) 14: pp. 767-778.

[175] Segal, E; Raveh-Sadka, T; Schroeder, M; Unnerstall, U; Gaul, U;. Predicting expression patterns from regulatory sequence in drosophila segmentation. *Nature* (2008) 451: pp. 535-540.

[176] Merabet, S; Pradel, J; Graba, Y;. Getting a molecular grasp on hox contextual activity. *Trends Genet* (2005) 21: pp. 477-480.

[177] Li, XY; MacArthur, S; Bourgon, R; Nix, D; Pollard, DA; Iyer, VN; Hechmer, A; Simirenko, L; Stapleton, M; Luengo Hendriks, CL;. Transcription factors bind thousands of active and inactive regions in the drosophila blastoderm. *PLoS Biol* (2008) 6: p. e27.

[178] Zhu, X; Sen, J; Stevens, L; Goltz, JS; Stein, D;. Drosophila pipe protein activity in the ovary and the embryonic salivary gland does not require heparan sulfate glycosaminoglycans. *Development* (2005) 132: pp. 3813-3822.

[179] Abrams, EW; Andrew, DJ;. Prolyl 4-hydroxylase alpha-related proteins in drosophila melanogaster: tissue-specific embryonic expression of the 99f8-9 cluster. *Mech Dev* (2002) 112: pp. 165-171.

[180] Abrams, EW; Mihoulides, WK; Andrew, DJ;. Fork head and sage maintain a uniform and patent salivary gland lumen through regulation of two downstream target genes, ph4alphasg1 and ph4alphasg2. *Development*

(2006) 133: pp. 3517-3527.

[181]Kerman, BE; Cheshire, AM; Andrew, DJ;. From fate to function: the drosophila trachea and salivary gland as models for tubulogenesis. *Differentiation* (2006) 74: pp. 326-348.

[182]Kuziora, MA; McGinnis, W;. Autoregulation of a drosophila homeotic selector gene. *Cell* (1988) 55: pp. 477-485.

[183]Regulski, M; McGinnis, N; Chadwick, R; McGinnis, W;. Developmental and molecular analysis of deformed: a homeotic gene controlling drosophila head development. *EMBO J* (1987) 6: pp. 767-777.

[184]Wing, JP; Karres, JS; Ogdahl, JL; Zhou, L; Schwartz, LM; Nambu, JR;. Drosophila sickle is a novel grim-reaper cell death activator. *Curr Biol* (2002) 12: pp. 131-135.

[185]Srinivasula, SM; Datta, P; Kobayashi, M; Wu, JW; Fujioka, M; Hegde, R; Zhang, Z; Mukattash, R; Fernandes-Alnemri, T; Shi, Y;. Sickle, a novel drosophila death gene in the reaper/hid/grim region, encodes an iap-inhibitory protein. *Curr Biol* (2002) 12: pp. 125-130.

[186]Christich, A; Kauppila, S; Chen, P; Sogame, N; Ho, SI; Abrams, JM;. The damage-responsive drosophila gene sickle encodes a novel iap binding protein similar to but distinct from reaper, grim, and hid. *Curr Biol* (2002) 12: pp. 137-140.

[187]Stowers, RS; Garza, D; Rascle, A; Hogness, DS;. The l63 gene is necessary for the ecdysone-induced 63e late puff and encodes cdk proteins required for drosophila development. *Dev Biol* (2000) 221: pp. 23-40.

[188]Giraldez, AJ; Cohen, SM;. Wingless and notch signaling provide cell survival cues and control cell proliferation during wing development. *Development* (2003) 130: pp. 6533-6543.

[189]Rusch, DB; Kaufman, TC;. Regulation of proboscipedia in drosophila by homeotic selector genes. *Genetics* (2000) 156: pp. 183-194.

[190]Jasper, H; Benes, V; Schwager, C; Sauer, S; Clauder-Munster, S; Ansorge, W; Bohmann, D;. The genomic response of the drosophila embryo to jnk signaling. *Dev Cell* (2001) 1: pp. 579-586.

[191]Drysdale, RA; Crosby, MA;. Flybase: genes and gene models. *Nucleic Acids Res* (2005) 33: p. D390-5.

[192]Flyexpress: an image-matching web-tool for finding genes with overlapping patterns of expression in drosophila embryos. (www.flyexpress.net)(2006).

[193]Keller, PJ; Schmidt, AD; Wittbrodt, J; Stelzer, EH;. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* (2008) 322: pp. 1065-1069.

[194]Appel, B; Sakonju, S;. Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products ubx and abd-a in drosophila embryos. *Embo J* (1993) 12: pp. 1099-1109.

[195]Ebner, A; Cabernard, C; Affolter, M; Merabet, S;. Recognition of distinct target sites by a unique labial/extradenticle/homothorax complex. *Development* (2005) 132: pp. 1591-1600.

[196] Heuer, JG; Li, K; Kaufman, TC;. The drosophila homeotic target gene centrosomin (cnn) encodes a novel centrosomal protein with leucine zippers and maps to a genomic region required for midgut morphogenesis. *Development* (1995) 121: pp. 3861-3876.

[197] Gould, AP; White, RA;. Connectin, a target of homeotic gene control in drosophila. *Development* (1992) 116: pp. 1163-1174.

[198] Capovilla, M; Brandt, M; Botas, J;. Direct regulation of decapentaplegic by ultrabithorax and its role in drosophila midgut morphogenesis. *Cell* (1994) 76: pp. 461-475.

[199] Sun, B; Hursh, DA; Jackson, D; Beachy, PA;. Ultrabithorax protein is necessary but not sufficient for full activation of decapentaplegic expression in the visceral mesoderm. *Embo J* (1995) 14: pp. 520-535.

[200] Capovilla, M; Botas, J;. Functional dominance among hox genes: repression dominates activation in the regulation of dpp. *Development* (1998) 125: pp. 4949-4957.

[201] Zeng, C; Pinsonneault, J; Gellon, G; McGinnis, N; McGinnis, W;. Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in drosophila embryos. *Embo J* (1994) 13: pp. 2362-2377.

[202] Hersh, BM; Carroll, SB;. Direct regulation of knot gene expression by ultrabithorax and the evolution of cis-regulatory elements in drosophila. *Development* (2005) 132: pp. 1567-1577.

[203] Grieder, NC; Marty, T; Ryoo, HD; Mann, RS; Affolter, M;. Synergistic activation of a drosophila enhancer by hom/exd and dpp signaling. *Embo J* (1997) 16: pp. 7402-7410.

[204] Gould, AP; Brookman, JJ; Strutt, DI; White, RA;. Targets of homeotic gene control in drosophila. *Nature* (1990) 348: pp. 308-312.

[205] Strutt, DI; White, RA;. Characterisation of t48, a target of homeotic gene regulation in drosophila embryogenesis. *Mech Dev* (1994) 46: pp. 27-39.

[206] Graba, Y; Gieseler, K; Aragnol, D; Laurenti, P; Mariol, MC; Berenger, H; Sagnier, T; Pradel, J;. Dwnt-4, a novel drosophila wnt gene acts downstream of homeotic complex genes in the visceral mesoderm. *Development* (1995) 121: pp. 209-218.

Publications

Stefanie D. Hueber, Daniela Bezdan, Stefan R. Henz, Martina Blank, Haijia Wu and Ingrid Lohmann

Comparative analysis of Hox downstream genes in *Drosophila*

Development 134, 381-392 (2007)

Stefanie D. Hueber and Ingrid Lohmann

Shaping segments: Hox gene function in the genomic age

BioEssays 30, 965-979 (2008)

Petra Stöbe, Matthias A. S. Stein, Anette Habring-Müller, Daniela Bezdan, Aurelia Fuchs, Stefanie D. Hueber, Haijia Wu and Ingrid Lohmann

Multifactorial Regulation of a Hox Target Gene

PLOS Genetics 5 (3), e1000412 (2009)

Development 134, 381-392 (2007) doi:10.1242/dev.02746

Comparative analysis of Hox downstream genes in *Drosophila*

Stefanie D. Hueber[1,*], Daniela Bezdan[1,*], Stefan R. Henz[2], Martina Blank[1], Haijia Wu[1] and Ingrid Lohmann[1,†]

Functional diversification of body parts is dependent on the formation of specialized structures along the various body axes. In animals, region-specific morphogenesis along the anteroposterior axis is controlled by a group of conserved transcription factors encoded by the Hox genes. Although it has long been assumed that Hox proteins carry out their function by regulating distinct sets of downstream genes, only a small number of such genes have been found, with very few having direct roles in controlling cellular behavior. We have quantitatively identified hundreds of Hox downstream genes in *Drosophila* by microarray analysis, and validated many of them by in situ hybridizations on loss- and gain-of-function mutants. One important finding is that Hox proteins, despite their similar DNA-binding properties in vitro, have highly specific effects on the transcriptome in vivo, because expression of many downstream genes respond primarily to a single Hox protein. In addition, a large fraction of downstream genes encodes realizator functions, which directly affect morphogenetic processes, such as orientation and rate of cell divisions, cell-cell adhesion and communication, cell shape and migration, or cell death. Focusing on these realizators, we provide a framework for the morphogenesis of the maxillary segment. As the genomic organization of Hox genes and the interaction of Hox proteins with specific co-factors are conserved in vertebrates and invertebrates, and similar classes of downstream genes are regulated by Hox proteins across the metazoan phylogeny, our findings represent a first step toward a mechanistic understanding of morphological diversification within a species as well as between species.

KEY WORDS: *Drosophila*, Deformed, Morphogenesis, Realizators, Microarray, Hox downstream target genes

INTRODUCTION

In animals, morphology along the various body axes is very diverse, requiring both a system that confers positional identity and a means to respond to these positional cues. It has long been known that the system specifying positional identity along the anteroposterior (AP) axis is based on an evolutionarily conserved set of regulators, the Hox genes (Carroll, 1995; Mann and Morata, 2000; McGinnis and Krumlauf, 1992). In *Drosophila*, Hox genes are expressed in defined domains along the AP axis, and their activity assigns distinct morphologies to the various body segments (McGinnis and Krumlauf, 1992). In addition, Hox genes are very often expressed in overlapping domains and crossregulate each other (McGinnis and Krumlauf, 1992; Miller et al., 2001). Consequently, loss of function of one Hox gene frequently leads to the ectopic expression of neighboring Hox genes, which is one of the reasons for the drastic homeotic transformations of body segments initially identified by Ed Lewis (Lewis, 1978). Therefore, only a subset of Hox functions can be identified in loss-of-function mutants.

Hox genes code for transcription factors with a highly conserved DNA-binding domain, the homeodomain (McGinnis et al., 1984; Scott and Weiner, 1984), and it has been postulated that Hox proteins direct morphogenesis by regulating appropriate sets of downstream genes in a segment-specific manner (Graba et al., 1997; Hombria and Lovegrove, 2003). Although a wide range of strategies has been used to identify Hox downstream genes (Graba et al., 1997; Hombria and Lovegrove, 2003; Pradel and White, 1998), our knowledge of their nature is still limited. Initial attempts have focused on in vitro studies or on heterologous systems; however, Hox proteins acquire DNA-binding specificity mostly through interactions with various co-factors in vivo (Ebner et al., 2005; Gebelein et al., 2004; Mahaffey, 2005; Mann, 1995; Mann and Affolter, 1998). Therefore, most known Hox downstream genes have been identified by candidate gene approaches based on expression patterns or similar mutant phenotypes (Pearson et al., 2005), highlighting the power of in vivo strategies to identify Hox target genes. This notion is further supported by recent successful approaches combining loss- or gain-of-function alleles of Hox genes and microarray experiments to identify Hox downstream genes on a larger scale (Cobb and Duboule, 2005; Hedlund et al., 2004; Lei et al., 2005; Williams et al., 2005). Still, most previous efforts were biased toward the identification of direct Hox target genes, and, while knowledge of direct Hox targets is a prerequisite to understanding how Hox proteins acquire DNA-binding specificity in vivo, we need to know the entire Hox-dependent regulatory network with all its tiers of regulatory interactions to understand how Hox proteins control morphogenesis on a cellular level.

Most of the known Hox downstream genes code either for transcription factors or for signaling molecules (Hombria and Lovegrove, 2003; Pearson et al., 2005). These two classes represent the top tiers of regulatory cascades and are able to coordinate many downstream events. Hence, they are not informative for elucidating the role of Hox proteins in the specification of morphological properties on a cellular level per se. To this end, the functional analysis of the so-called realizators, which directly influence the morphology by regulating cytodifferentiation processes (Garcia-Bellido, 1975; Pradel and White, 1998), is required. Unfortunately, even though the concept of realizators was postulated more than 30 years ago, so far very few Hox realizator genes have been identified and studied mechanistically (Bello et al., 2003; Lohmann et al., 2002). One

[1]Max Planck Institute for Developmental Biology, [2]Department of Molecular Biology, Spemanstrasse 37-39, D-72076 Tübingen, Germany.

*These authors contributed equally to this work
†Author for correspondence (e-mail: ingrid.lohmann@tuebingen.mpg.de)

Accepted 1 November 2006

well-studied example of a realizator gene in *Drosophila* is the apoptosis-inducing gene *reaper* (*rpr*), which is expressed in the maxillary segment in *Drosophila* embryos and is directly controlled by the Hox protein Deformed (Dfd) (Lohmann et al., 2002). In addition, the Dfd-dependent expression of *rpr* and, consequently, the activation of apoptosis was shown to be necessary and sufficient for the maintenance of the boundary between the maxillary and mandibular segments of the embryonic head (Lohmann et al., 2002). This is one of the few examples demonstrating how a Hox protein can execute, via a single realizator gene, one specific aspect of segmental morphology on the cellular level. To understand and mechanistically link the many remaining Hox functions with morphogenetic outputs, we need to quantitatively identify Hox downstream genes. Functional analysis of this set will then allow us to elucidate all tiers of interactions within the Hox-regulatory network, and to establish links between Hox genes and realizator genes. This seems fundamental for a complete understanding of the role of Hox genes in development and evolution.

MATERIALS AND METHODS

Drosophila genetics

The wild-type strain used was Oregon-R. UAS-*lacZ*, UAS-*Dfd*, UAS-*Antp* and *arm*-GAL4 strains were from W. McGinnis, UAS-*Scr*, UAS-*Ubx* and UAS-*Abd-B* from F. Hirth (Hirth et al., 2001), UAS-*Abd-A* from A. M. Michelson (Michelson, 1994) and UAS-*hep*$^{act.}$ from M. Mlodzik (Weber et al., 2000). For trans-heterozygous mutants the following alleles were used: *Dfd*^{r11} and *Dfd*^{w21} from W. McGinnis; *Scr*^{1}, *Scr*^{4}, *Abd-B*^{M2}, *wg*$^{l-12}$ and *wg*$^{l-17}$ from the Bloomington Stock Center; and *Abd-B*^{M5} from C. Nüsslein-Volhard (Tübingen *Drosophila* Stock Collection). *Dfd* mutant embryos for BrdU staining were *Dfd*^{w21}/TM3*Sb*[*twi*::GFP] crossed to *Dfd*^{r11}/TM3*Sb*[*twi*::GFP] and homozygous *Dfd* mutants (*Dfd*^{w21}/*Dfd*^{r11}) were identified by absence of GFP signal.

Plasmids

cDNAs were from the *Drosophila* Genomics Resource Center: *CG5080* (LD34147), *CG7447* (LD16414), *disco* (GH27656), *Dll* (LP01770), *ImpL2* (SD07266), *gt* (RE29225), *sage* (RE59356), *skl* (RE14076), *spz* (SD07354), *LysE* (LP07339), *CG8193* (GH07976), *CG3097* (RE43153), *Mp20* (RE55741), *CG17052* (LD43683), *Ance* (LD11258), *Hsp23* (LD06759), *sn* (RH62992), *mas* (LP06006), *pav* (RE22456), *wrapper* (GH03113), *wg* (RE02607) and *W* (AT13267). *prd* cDNA was from W. McGinnis, *Eip63E* cDNA and predicted Dfd response elements tested by EMSA were PCR amplified, cloned and sequenced. Expression plasmids for Dfd and Ubx were obtained from W. McGinnis and S. Carroll, respectively.

Histology and scanning electron microscopy

In situ hybridization and immunochemistry were performed as described (Bergson and McGinnis, 1990; Tautz and Pfeifle, 1989), and BrdU labeling and scanning electron microscopy were done as described (Dolbeare and Selden, 1994; Lohmann et al., 2002). Hox protein expression was measured by the fluorescent intensity of a standardized area of individual nuclei using the Zeiss LSM 510 META confocal microscope. Twenty nuclei of four independent embryos were analyzed for each expression domain and genotype. Antibodies were: anti-Dfd, W. McGinnis; anti-Scr, anti-Antp, anti-Abd-B and anti-wg, Developmental Studies Hybridoma Bank (Iowa, University), anti-Ubx, R. White (Cambridge); anti-Abd-A, I. W. Duncan (Washington, University); anti-GFP, Torrey Pines Biolabs (Houston); anti-BrdU, Roche; anti-mouse AlexaFluor 488, anti-guinea pig AlexaFluor 488 and anti-rabbit AlexaFluor 488, Molecular Probes.

Microarray experiments

Microarray hybridizations were carried out as described (Schmid et al., 2003) in biological triplicates with RNA from pools of stage 11 or stage 12 embryos. Raw data were quantile normalized and expression estimates were calculated using gcRMA (Wu et al., 2004) implemented in R. Statistical testing for differential expression was carried out using LogitT (Lemon et al., 2003). Microarray data discussed here have been deposited with ArrayExpress database at the EBI (http://www.ebi.ac.uk/arrayexpress-old/; Accession number E-MEXP-879). For analysis of gene ontology categories, GO lists from FlyBase were used. Genes were sorted using a combination of molecular and biological GO terms. Genes containing the following description in their GO annotations were classified as realizators: apoptosis, cell death, cell adhesion, cell shape, cell cycle, mitosis, cell proliferation, cytoskeleton, proteolysis, peptidolysis, cytoskeleton, structural constituent of larval cuticle or peritrophic membrane.

Quantitative real-time PCR

Quantitative real-time PCR (qRT-PCR) was carried out in triplicates from RNA of pooled tissue as described (Schmid et al., 2003) using SYBR-green QPCR Master Mix (Invitrogen). Expression of β-*Tubulin* was used for cross-experiment normalization. Primer and probe sequences are available upon request.

Bioinformatics

For cluster identification the Cis-Analyst algorithm (Berman et al., 2004) was used with a Position-Frequency-Matrix (PFM) based on DNaseI footprint data (Bergman et al., 2005) and consensus sequences from the literature. The PFM shown in Fig. S1B in the supplementary material was generated by PATSER. To define enhancer parameters, such as length of enhancer, number of Dfd-binding sites per enhancer, distance between binding sites, known Dfd-dependent enhancers were analyzed. The parameters identified, as shown in Fig. S1B in the supplementary material, were used to predict clusters of Dfd-binding sites in the regulatory regions of selected genes in *Drosophila melanogaster*. To this end, intergenic and intronic sequences of *D. melanogaster* were aligned to a multiple sequence file, sorted and separated to segment-files, which included annotation information. The PATSER program used these segment-files as a template to generate *P*-values for each Dfd-binding site identified according to the PFM. Using this binding site information, clusters of Dfd-binding sites were predicted using the standalone version of cis-Analyst-helper. To validate this approach statistically and to optimize the parameters chosen, Dfd downstream genes identified in the microarray experiment were used. The logic of this approach is based on the assumption that direct Dfd target genes should be enriched among the Dfd downstream genes identified in the microarray experiment when compared with randomly selected genes. To identify Dfd clusters in other *Drosophila* species (*D. simulans*, *D. yakuba*, *D. erecta*, *D. pseudoobscura*), a NCBI BLAST search was performed. To consider clusters as being conserved, the following conservation criteria had to be fulfilled: (1) conservation of the enhancer in at least two other *Drosophila* species; (2) the length of the homologues enhancer had to be ⩾50% of the enhancer length identified in *D. melanogaster*; (3) conservation of at least two Dfd-binding sites within the conserved enhancer elements; and (4) conserved enhancers with less then 50% of length conservation but more then two Dfd-binding sites conserved were treated as minor hits. Conserved enhancers were ranked according to the following parameters: (1) evolutionary distance of *Drosophila* species; (2) overall sequence similarity of conserved enhancers; (3) numbers of binding sites present in conserved clusters; and (4) degree of variation in enhancer length. Binding site matches for other transcription factors located in the Hox response elements were identified by using rVISTA, Transfac and Jasper databases.

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSAs) were performed as described (Lohmann et al., 2002).

RESULTS

Identification of Hox responsive genes by microarray analysis

In order to systematically elucidate Hox-regulatory networks, we performed a comparative microarray screen using stage 11 and stage 12 embryos ubiquitously overexpressing six out of eight Hox genes – *Dfd*, *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal A* (*abd-A*) and *Abdominal B* (*Abd-B*) – by means of the UAS-GAL4 system (Brand and Perrimon, 1993) (Fig. 1A; for

endogenous expression of Hox proteins see Fig. 1E). The Hox genes *labial* (*lab*) and *proboscipedia* (*pb*) were not included, as a similar study has been performed for *lab* (Leemans et al., 2001), and as *pb* mutant embryos do not exhibit any obvious defects (Wakimoto et al., 1984). We opted to use overexpression instead of a loss-of-function experiments for several reasons. First, the analysis of Hox mutants is complicated by the extensive crossregulatory interactions of Hox genes (Gould et al., 1997; Miller et al., 2001); loss of expression of one Hox gene often results in ectopic expression of other Hox genes, thereby obscuring the effects on downstream genes. Another important limitation of a loss-of-function approach in conjunction with microarray analysis is sensitivity; due to the small expression domains of many Hox genes (McGinnis and Krumlauf, 1992; Pearson et al., 2005), locally restricted differences in gene expression caused by Hox mutations will be diluted in RNA isolated from whole embryos, and therefore many downstream genes might not be detected. Isolation of cells expressing individual Hox genes by cell sorting (Wang et al., 2006) could provide a means to circumvent this problem; however, the required reporter genes that are expressed in specific Hox domains, although independent from Hox gene activity, currently do not exist. Conversely, although ectopic expression of one Hox gene affects the expression of other Hox genes, their ubiquitous overexpression should allow us to even detect genes whose expression is only weakly, or locally, affected, because the Hox expression domains are expanded manyfold and consequently their transcriptional output is amplified. To achieve ubiquitous Hox overexpression in the desired stages of development, we used an *armadillo* (*arm*)-GAL4 driver line (Sanson et al., 1996), which confers ubiquitous expression starting at stage 10, as judged from analyzing GFP activity in embryos carrying an additional UAS-*2xEGFP* transgene (Fig. 1A). Previous studies have shown that ubiquitous overexpression of Hox genes in UAS fly strains, which were also used in our study, is sufficient to induce ectopic differentiation of Hox-dependent structures without affecting the development of early embryonic stages in an unspecific manner (Li et al., 1999). Thus, a substantial part of Hox downstream genes seem to be responsive to Hox signaling even at ectopic locations and should be detectable by microarray analysis.

One concern with overexpression experiments, however, is that they might result in varying or even unphysiological concentrations of transcription factor proteins in the nucleus, which might unspecifically affect gene expression. Therefore, we confirmed similar RNA and protein expression levels in our overexpression lines by qRT-PCR and antibody stainings (Fig. 1B,E). In addition,

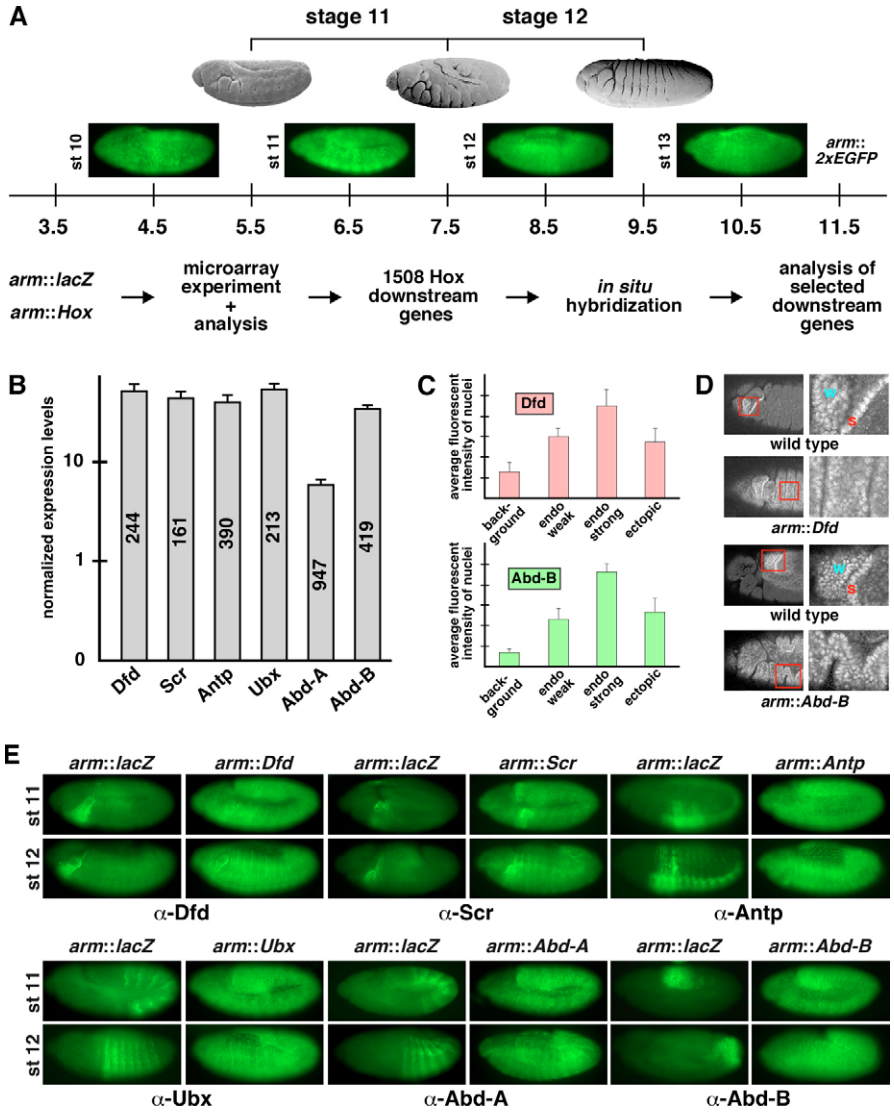

Fig. 1. Identification of Hox downstream genes during early embryogenesis. (**A**) Outline of microarray analysis. Scanning electron micrographs of embryos at late stage 10-early stage 11, late stage 11-early stage 12, and late stage 12-early stage 13. *arm*::*2xEGFP* embryos demonstrate ubiquitous transgene expression beginning at stage 10. Time scale shows hours of embryogenesis. (**B**) Quantitative real-time PCR shows similar levels of overexpression of Dfd, Scr, Antp, Ubx, Abd-A and Abd-B transgenes. Numbers of genes regulated by the different Hox proteins are indicated. (**C**) Average fluorescent intensity (in arbitrary units) of 20 independent nuclei at different locations in wild-type, *arm*::*Dfd* and *arm*::*Abd-B* embryos stained either with α-Dfd or α-Abd-B antibody. Due to variable expression levels of endogenous Hox proteins, fluorescence from nuclei in different expression domains was measured (marked as 'endo weak' and 'endo strong'). (**D**) Representative embryo used for measuring fluorescent intensity of nuclei is shown. The upper two rows show wild-type and *arm*::*Dfd* embryos stained with α-Dfd antibody, the lower two rows wild-type and *arm*::*Abd-B* embryos stained with α-Abd-B antibody. Red boxes mark the areas used for fluorescence analysis. s, strong endogenous expression domain; w, weak endogenous expression domain. (**E**) Ubiquitous overexpression of Hox proteins in stage 11 and 12 was confirmed by antibody staining on *arm*::*lacZ* and *arm*::*Hox* embryos.

we quantified Dfd and Abd-B protein levels in the nuclei of wild-type and the respective overexpression embryos and found that protein levels in transgenic embryos were on the lower end of the wild-type expression spectrum (Fig. 1C,D). Thus, non-physiological effects of Hox misexpression due to increased protein levels should not interfere with our experiment.

For the microarray experiment, total RNA from *arm*::*lacZ* (control), *arm*::*Dfd*, *arm*::*Scr*, *arm*::*Antp*, *arm*::*Ubx*, *arm*::*abd-A* and *arm*::*Abd-B* embryos collected separately at the different stages was prepared and hybridized in biological triplicates to Affymetrix *Drosophila* Genome 1 arrays, which contains probe sets interrogating more than 13,500 genes. A combination of per-gene (Lemon et al., 2003) and common variance (>1.5-fold change) filtering was used to identify 1508 Hox responsive genes (Table S1, Table S2, Table S3, Table S4, Table S5 and Table S6 in the supplementary material). This list contained six of the 18 genes previously shown to be under Hox control during stages 11 and 12 (Fig. 2A), and in situ hybridizations for three of these transcripts confirmed the microarray results (Fig. 2A). The fact that we were only able to recover one-third of the known Hox targets can be explained by a number of differences between our experimental setup and the ones used before, such as detection method (Capovilla et al., 2001; Mahaffey et al., 1993), timing and level of overexpression (Feinstein et al., 1995), and use of mutants instead of overexpression (Mahaffey et al., 1993; Ryoo and Mann, 1999). Based on this observation, it is conceivable that the actual number of all Hox downstream genes is two- to threefold the number we have discovered in our study, which is still significantly lower than previously suggested (Liang and Biggin, 1998). The microarray data also showed that anterior Hox genes were repressed by those normally expressed more posteriorly (data not shown), a crossregulatory interaction known as posterior suppression (Miller et al., 2001). Again, we could confirm the microarray data by performing antibody stainings for all Hox proteins on embryos ubiquitously misexpressing either Dfd or Abd-B (Fig. 2B).

Verification of Hox downstream genes identified in the microarray analysis

To verify differential expression of the newly identified genes at the cellular level, we carried out in situ hybridization on embryos misexpressing the various Hox genes (Fig. 3; see Fig. S1 in the supplementary material). Twenty-four of the 25 randomly selected genes that showed a specific in situ signal behaved as observed in the microarray experiment. In addition, for a selected subset of seven genes Hox-dependent regulation could also be shown in Hox mutants (Fig. 3), demonstrating the power of the initial microarray experiment. For example, three transcripts found to be induced by Dfd in the microarray experiment were *sickle* (*skl*), a known apoptosis activator (Wing et al., 2002), *CG5080*, a gene putatively involved in cytoskeletal regulation (Jasper et al., 2001), and CG7447, a gene of unknown function. In situ analysis confirmed strong and ectopic induction of all three genes in response to Dfd misexpression (Fig. 3B,F,J), and showed that their expression in the maxillary segment was lost in *Dfd* mutants (Fig. 3D,H,L). Similarly, mRNA levels of *salivary gland-expressed bHLH* (*sage*), a transcription factor gene exclusively expressed in the salivary gland primordium (Chandrasekaran and Beckendorf, 2003), were increased in response to ectopic Scr activity (Fig. 3N). By contrast, *sage* expression was abolished in *Scr* mutants (Fig. 3P), consistent with Scr being a master regulator of salivary gland morphogenesis (Panzer et al., 1992). Among the genes that were induced by Abd-B were *Ecdysone-inducible gene L2* (*ImpL2*), putatively involved in cell adhesion (Garbe et al., 1993), and *spätzle* (*spz*), which encodes a Toll receptor ligand involved in embryonic axis specification (DeLotto et al., 2001). Again, we observed strong ectopic expression of *ImpL2* and *spz* in *arm*::*Abd-B* embryos (Fig. 3R,V), whereas expression in the posterior end was lost in *Abd-B* mutants (Fig. 3T,X). In summary, in situ hybridization with probes for 24 randomly selected genes (Fig. 3; see Fig. S1 in the supplementary material) not only confirmed the microarray results, but also demonstrated that many of the identified Hox downstream genes responded in a converse manner in the respective Hox mutants.

Direct versus indirect Hox downstream genes

To understand the logic of Hox-dependent morphogenesis, it is important to place the newly identified downstream genes within the underlying regulatory hierarchy. To this end, we developed a bioinformatics tool to detect direct Hox target genes, based on the identification of evolutionarily conserved clusters of Hox consensus binding sites in the genome (see Materials and methods and Fig. S2 in the supplementary material for detail). Using this approach, we were able to identify a large number of putative direct targets of Dfd. From the 240 genes found to be significantly regulated by Dfd, 75 had clusters of Dfd-binding sites (31% of all identified Dfd responsive genes), which was significantly more than expected by chance ($P<0.001$). In addition, 46 of these clusters were well conserved in at least two other *Drosophila* species (19% of all identified Dfd responsive genes) (see Tables S7 and S8 in the supplementary material). Most of the predicted Dfd response elements also contained binding sites for other transcription factors (data not shown), a known prerequisite for functional enhancer elements (Berman et al., 2004). We randomly selected six of the 75 predicted Dfd response elements and performed EMSA to test whether Dfd protein could bind to these elements. All enhancer elements tested were bound by Dfd in vitro (Fig. 4), whereas Ubx, a Hox protein specifying trunk identity, was not able to interact with these enhancers (see Fig. S2C in the supplementary material). In addition, competition experiments showed that Dfd specifically bound some, but not all, of the predicted Dfd-binding sites in these enhancers (Fig. 4A-D), demonstrating that the simple presence of a consensus binding site is not sufficient for Dfd binding in the context of these enhancers in vitro and/or that some of the predicted sites are not functional in vivo. Based on our results with Dfd, it seems likely that about 20 to 30% of the identified downstream genes are direct Hox targets. In sum, the combination of microarray analysis with bioinformatics approaches will allow us in the future to not only identify direct Hox target genes, but also to construct complete Hox-regulatory networks.

Specificity of Hox-dependent regulation

To assess the specificity of Hox gene regulation, the 1508 responsive genes were classified according to the number of Hox proteins that influenced their expression and the influence of the developmental stage. Remarkably, most downstream genes (1039, 68.9%) were affected by only a single Hox protein, with Abd-A having a very high proportion of unique response genes (two-thirds of its downstream genes were unique), whereas the fraction of unique response genes was smaller (18 to 36%) for the other Hox genes (Fig. 5A). The use of various statistical cut-offs showed that this result is not an artifact of arbitrary thresholding (data not shown). In addition, we were able to confirm the specificity of the Hox response by analyzing the expression of some of the unique downstream genes by in situ hybridizations in embryos misexpressing any of the

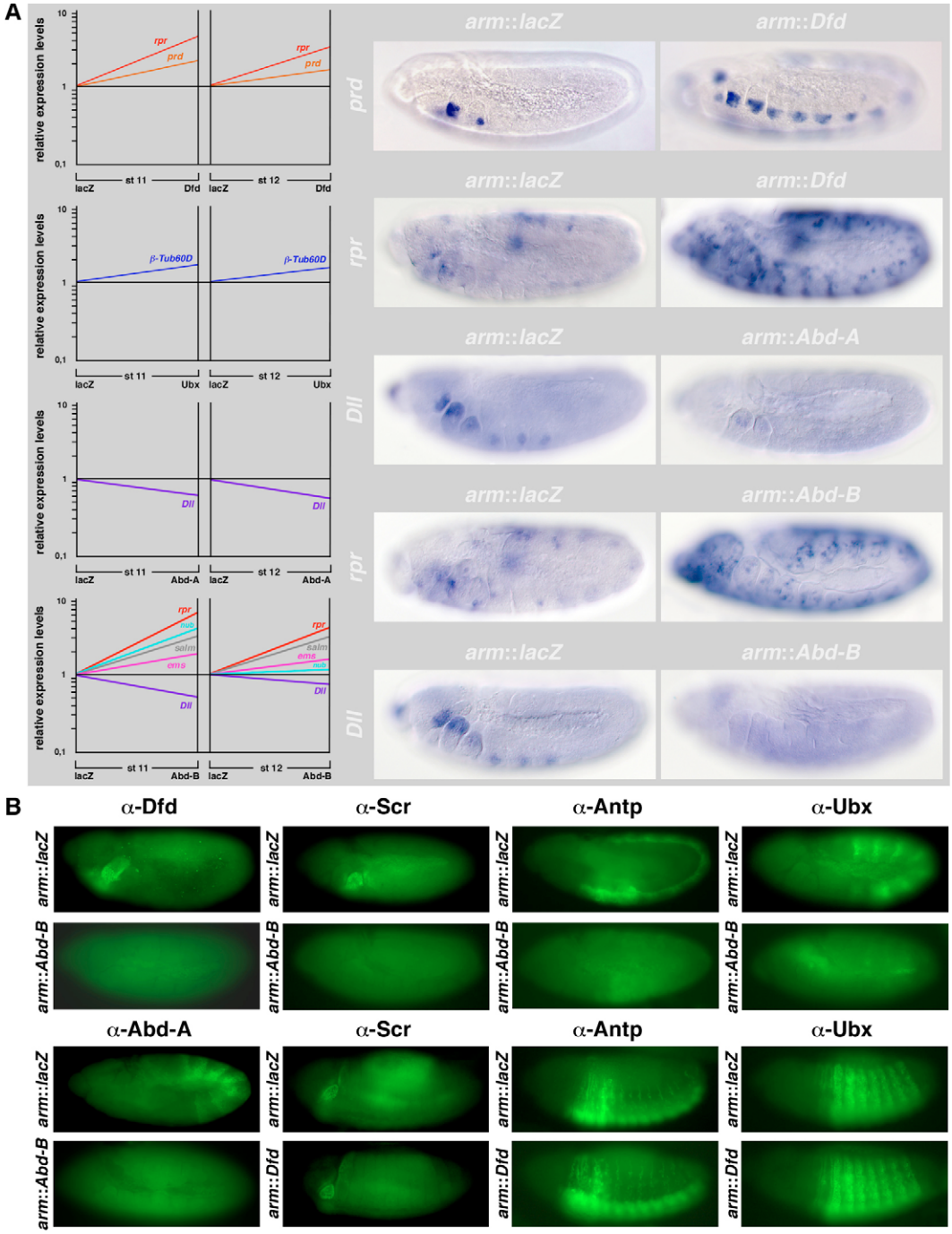

Fig. 2. Verification of known Hox downstream genes identified by microarray analysis. (**A**) Relative expression levels of seven known Hox downstream genes identified in a microarray screen are shown. For three of the seven genes, in situ hybridizations were performed. Genes shown are: *reaper* (*rpr*), *paired* (*prd*), *β-Tubulin at 60D* (*β-Tub60D*), *Distal-less* (*Dll*), *spalt major* (*salm*), *empty spiracles* (*ems*) and *nubbin* (*nub*). (**B**) α-Dfd and α-Abd-B antibody stainings on embryos misexpressing different *Hox* genes confirmed posterior suppression as seen in the microarray experiment. Antibody stainings for all (Abd-B misexpression) or for some (Dfd misexpression) Hox proteins are shown.

Hox genes (see Fig. S3 in the supplementary material). About one-third of the identified downstream genes (449, 29.8%) were affected by several Hox proteins, and only 20 genes (1, 3%) responded to all Hox proteins, representing the classes of regional and common downstream genes, respectively (Fig. 5A). Even when we excluded the Abd-A experiment, which was performed slightly differently from the rest of the set and therefore could interfere with this type of analysis, the result did not change: 63% of the genes were uniquely regulated by only one Hox protein, 34.5% of the genes by some and 2.5% by all Hox proteins (Fig. 5A). Remarkably, among the predicted direct Dfd target genes the distribution of unique and regional Hox downstream genes was similar to their distribution among all identified Hox downstream genes (Fig. 5B). Taken together, these results indicate that the specific effects of Hox proteins on morphology are largely mediated by regulatory interactions with uniquely regulated downstream genes, and that

despite the very similar DNA-binding sequences for all Hox proteins observed in vitro (Ekker et al., 1994), the overlap of commonly regulated genes in vivo is relatively small.

To evaluate the influence of co-factors on the regulation of Hox downstream genes, we analyzed the Hox response at the two developmental stages. The idea was that overexpression of Hox genes remained the same, while the regulatory environment changed during the progression of embryogenesis. We found that most of the downstream genes were Hox responsive at either embryonic stage 11 or stage 12 (Fig. 5C,D), consistent with the hypothesis that Hox proteins strongly change their transcriptional output through the interaction with differentially expressed co-factors (Gebelein et al., 2004; Mahaffey, 2005; Merabet et al., 2005). To confirm stage-specific regulation of Hox downstream genes on a cellular level, we performed in situ hybridizations for some of the differentially expressed genes (Fig. 5D), and indeed found that most of these genes were Hox responsive primarily at one of the two stages (Fig. 5D).

Functional classification and comparative analysis of Hox downstream genes

To group the newly identified Hox downstream genes functionally, we used Gene Ontology (GO) annotations (Fig. 6A) and analyzed the distribution of GO categories within the Hox-responsive genes. We found that the two major groups of Hox downstream genes encoded metabolic and realizator functions, followed by the transcription or translation, signaling, transport, stress or defense response and DNA repair or replication classes (Fig. 6A). Supporting the significance of the realizator genes for the Hox response, we found that this group was the only class statistically over-represented among the Hox downstream genes by Fisher's exact test after Bonferoni correction ($P<0.001$). By contrast, the transcription or translation and transport classes were under-represented ($P<0.001$). Focusing on Dfd downstream genes, we also found that realizators were over-represented, whereas the transcription or translation functions were under-represented, albeit at weaker P-values. Intriguingly, even within the group of direct Dfd targets, the transcription or translation class was not over-represented. As sequence-specific transcription factors were never over-represented among any category tested, this suggests that the activation of regulatory proteins is not the preferred mode of Hox action. The fact that realizator processes at the same time are under direct as well as indirect Hox control demonstrates that Hox-dependent morphogenesis is indeed achieved by regulatory networks, rather than linear pathways. Furthermore, these results also argue against the idea that Hox genes mainly act on other transcription factors, a hypothesis that was based on the limited information on known direct Hox targets.

Since the realizator gene concept was postulated almost 30 years ago (Garcia-Bellido, 1975), but only a few such genes have previously been identified as Hox downstream genes in *Drosophila*, we decided to study this class of Hox response genes in more detail.

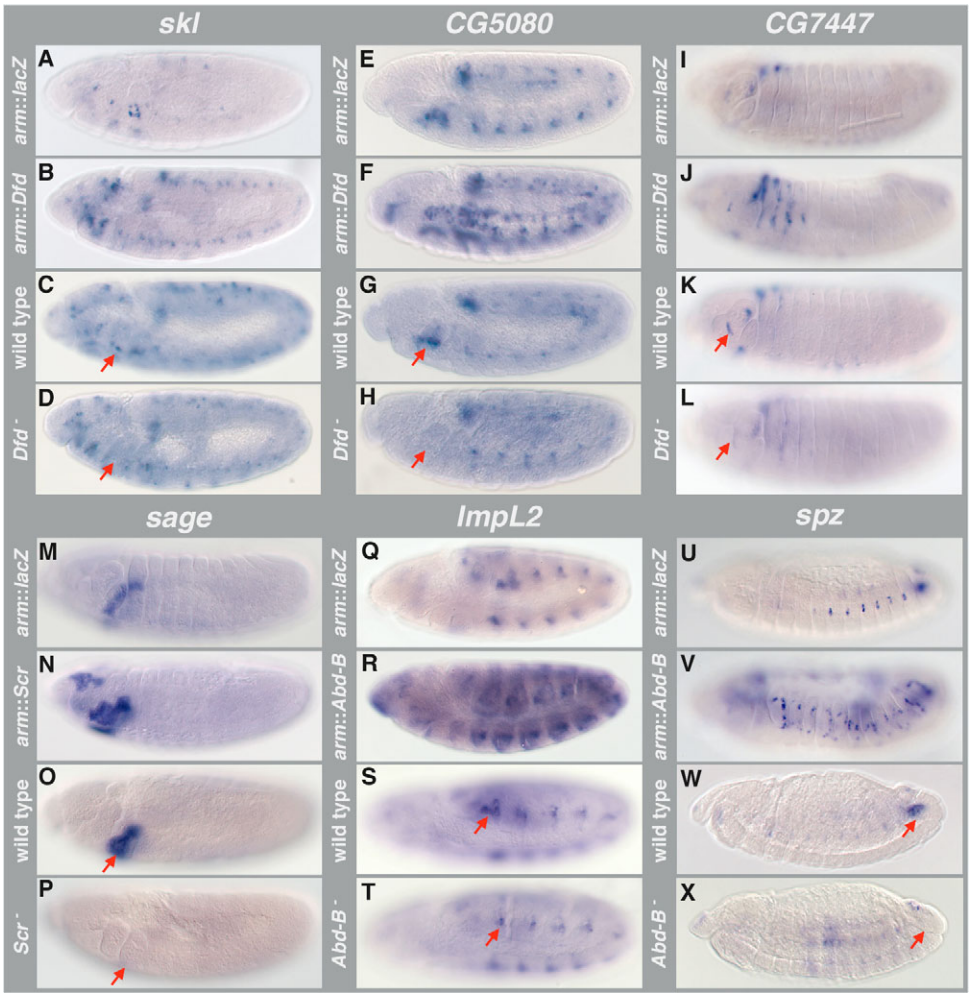

Fig. 3. Verification of newly identified Hox downstream genes by in situ hybridization. In situ hybridizations of the indicated genes in stage 11 and/or stage 12 *arm::lacZ*, *arm::Hox*, wild-type and *Hox* mutant embryos. Genes shown are: *skl* (**A-D**), *CG5080* (**E-H**), *CG7447* (**I-L**), *sage* (**M-P**), *lmpL2* (**Q-T**) and *spz* (**U-X**). Hybridizations on embryos misexpressing *Hox* genes and on *Hox* mutant embryos were performed independently (with the respective *arm::lacZ* and wild-type controls). Differences in staining intensities are due to differences in the in situ hybridization procedures. Pictures of *arm::lacZ* and wild-type embryos were taken at different focal planes and thus expression patterns in these embryos appear slightly different. Red arrows mark expression domains changed in wild-type and *Hox* mutant embryos.

Among the realizators, the largest subgroup comprises genes involved in proteolytic processes, followed by genes with cytoskeleton functions, a diverse group containing cuticle, chorion and peritrophic membrane genes, cell cycle or cell proliferation genes, apoptosis or cell death genes and cell adhesion genes (Fig. 6B). Genes within realizator subclasses are often coordinately regulated: most apoptotic (7/8) and cell cycle or proliferation genes (18/21) were activated, whereas almost all cell adhesion genes (12/14) and the majority of genes involved in proteolytic processes (56/75) were repressed by Hox proteins (Fig. 6C). Re-analyzing data from a more restricted microarray study, a similar trend can be identified for *lab*, another Hox gene (Leemans et al., 2001): one apoptotic gene and six cell cycle or cell proliferation genes were activated, whereas three cell adhesion genes were all repressed by *lab*. This suggests that a variety of cellular processes need to be regulated in a coordinated fashion in every segment in order to realize common aspects of segmental morphology. Support for this notion also comes from a previous analysis in *Drosophila*, showing that two Hox proteins, Dfd and Abd-B, locally activate the apoptosis gene *rpr* and thus the apoptotic machinery at segment boundaries for their maintenance (Lohmann et al., 2002).

Finally, we also wanted to analyze which role the identified Hox downstream genes play during the morphogenesis of segments. To this end, we first asked whether differences in morphology along the AP axis are reflected in the differential regulation of Hox responsive genes. One line of evidence supporting this idea is that many downstream genes responded only to a single Hox protein (Fig. 5A). In addition, we found that coordinated regulation of shared downstream genes was more frequent among Hox proteins specifying segments with similar morphologies than among those that specify diverse segmental morphologies (Fig. 6D). For example, both Dfd and Scr specify small lobe-shaped gnathal segments and only 4% (n=2) of their shared downstream genes are regulated in an opposite manner (Fig. 6D). One of the differentially regulated Dfd or Scr downstream genes is *PAPS synthetase*, which is repressed by Dfd and activated by Scr in our dataset (see Tables S1 and S2 in the supplementary material). Consistent with this observation, it had been shown that *PAPS synthetase* is specifically expressed in the salivary gland placodes in a Scr-dependent manner, while it is absent from Dfd-expressing cells in stage 11 and 12 embryos (Jullien et al., 1997). Another example for a strong correlation of differences in morphology and the differential regulation of shared downstream genes is the Scr-Abd-A pair. In this case, opposite regulation was found for 58% (n=55) of the targets shared by Scr and Abd-A, in line with the very different morphologies specified by these two Hox proteins (Fig. 6D). Again, two selected examples, *pipe* and *PH4αSG2*, both activated by Scr and repressed by Abd-A (see Table S2 and Table S5 in the supplementary material), are known to be expressed exclusively in the Scr-specified salivary glands at stages 11 and 12 (Abrams and Andrew, 2002; Zhu et al., 2005). Interestingly, it has been shown only recently that *pipe*, differentially regulated by Dfd and Abd-A, and *PAPS synthetase*, differentially regulated by Scr and Dfd in our microarray analysis, are both necessary for the production of sulfated macromolecules in the salivary glands of *Drosophila* embryos (Zhu et al., 2005). Thus, it seems that the diversification of segments is achieved, on the one hand, through the regulation of unique downstream genes, and, on the other hand, through the differential regulation of shared downstream genes.

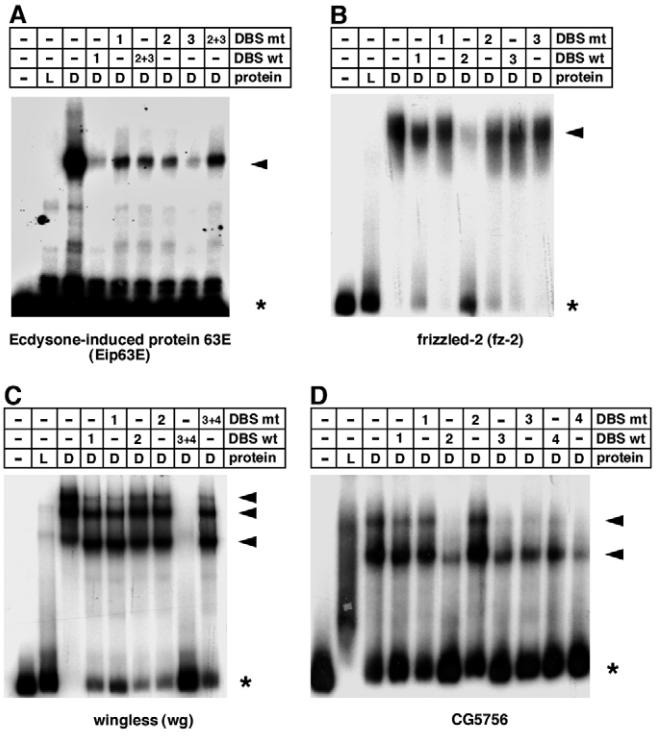

Fig. 4. Confirmation of predicted direct Dfd downstream genes by EMSA. (**A-D**) EMSA for four predicted direct Dfd downstream genes tested using no protein, translation lysate only (L) and lysate with Dfd protein (D). To test the specificity of binding of Dfd protein to the DNA fragments, competitor oligonucleotides for the individual Dfd-binding sites (DBS) were used either in their wild-type (wt) or mutant (mt) sequence versions. The black arrowhead indicates the specific DNA-protein complex containing Dfd protein. Asterisks indicate the unbound labeled probe. Predicted Dfd response enhancers of the following genes were used: *Eip63E* (A), *frizzled 2* (*fz2*) (B), *wg* (C) and *CG5756* (D).

A framework for the morphogenesis of the maxillary segment

To analyze the morphogenetic function of Hox responsive genes in more detail, we focused on the potential role of several newly identified Dfd downstream genes during the development of the maxillary segment. It has long been known that Dfd is expressed in the maxillary and mandibular segments, and is necessary for the morphological specializations (mouth hooks, cirri, ventral organ) of these head segments (McGinnis et al., 1990). However, only a single cellular event necessary for the morphogenesis of the maxillary segment and under the control of Dfd has been explained mechanistically so far: the maintenance of the boundary between the maxillary and mandibular segments, which is dependent on Dfd-mediated activation of *rpr* expression in the anterior part of the maxillary segment (Lohmann et al., 2002). Consistently, *rpr* was found among the activated Dfd downstream genes in our microarray analysis (see Table S1 in the supplementary material). Another prominent feature of *Dfd* mutants is the displacement of maxillary and mandibular segments to a more dorsal position, caused by the accumulation of supernumerary cells at the ventral side of both segments, which had been observed almost 20 years ago (Fig. 7B) (Regulski et al., 1987). There are two alternative explanations for this defect: loss of cell death and/or overactivation of cell proliferation. Consistent with the former explanation, we observed reduced local expression of the apoptosis activator *skl* (Fig. 7K,P), one of the newly identified genes activated by Dfd, and a concurrent

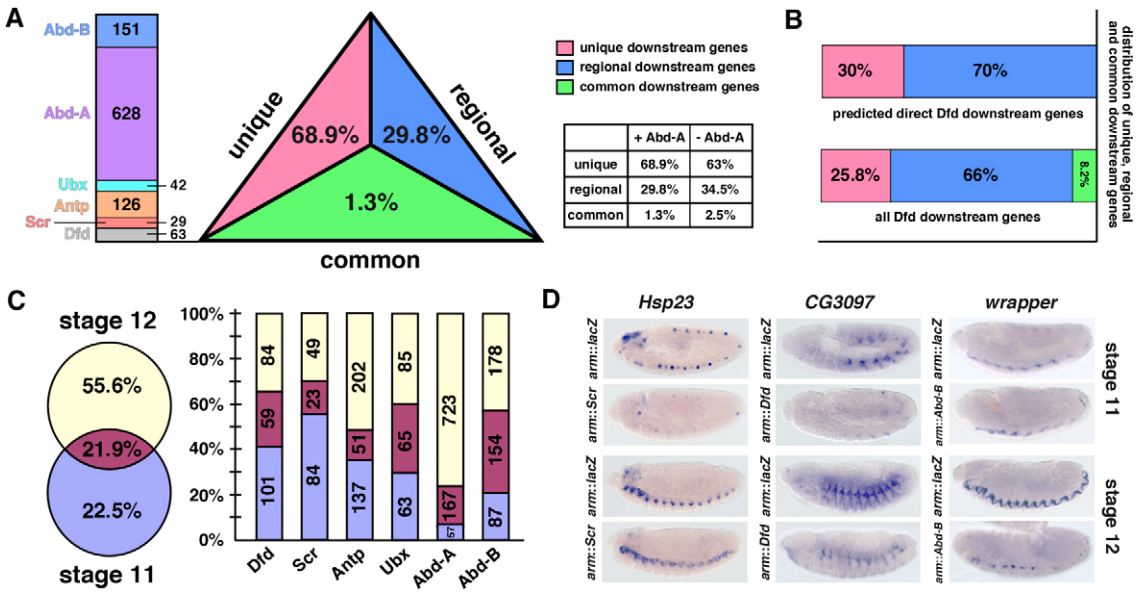

Fig. 5. Specificity of Hox downstream gene regulation. (**A**) Classification of Hox downstream genes according to their regulation by one (unique), several (regional) and all (common) Hox proteins. Numbers of unique downstream genes for each Hox protein are shown on the left. The distribution of classes does not change when the Abd-A experiment, which was performed independently, is excluded from the analysis (shown in the table). (**B**) Distribution of the regulatory classes among all identified Dfd downstream genes and predicted direct Dfd target genes is very similar. No commonly regulated downstream genes are found among the predicted direct Dfd target genes. (**C**) Distribution of Hox downstream genes regulated at the two different stages analyzed. On the left side, the percentage of all Hox downstream genes regulated at the two stages are shown; on the right side, the distribution for each individual Hox protein is indicated. Numbers of genes are shown within bars. (**D**) In situ hybridizations of selected examples of genes regulated at specific stages (early-specifc, early and late, late-specific). Expression patterns of the following genes are displayed: *Heat shock protein 23 (Hsp23), CG3097* and *wrapper*.

reduction in the number of apoptotic cells at the ventral side of the maxillary segment in *Dfd* mutants (data not shown) (Lohmann et al., 2002). Additionally, we were able to show that cell proliferation at the ventral region of the maxillary segment is increased in *Dfd* mutant embryos by performing BrdU labeling experiments (Fig. 7E,J). This might be attributed to the de-repression of two genes identified as repressed by Dfd in this study (see Table S1 in the supplementary material): *Ecdysone-induced protein 63E* (*Eip63E*), encoding a cyclin-dependent protein kinase (Stowers et al., 2000), and *wingless* (*wg*), encoding a signaling molecule known to play a role in cell proliferation (Giraldez and Cohen, 2003). In line with a role of these genes in shaping the maxillary segment, we found that both genes are ectopically expressed in the dorsal part of the maxillary segment in *Dfd* mutants (Fig. 7Q,R). Although the function of *Eip63E* during the morphogenesis of the maxillary segment could not be analyzed due to the lack of mutants, we could confirm that *wg* mutants have reduced gnathal lobes (Fig. 7D) (Rusch and Kaufman, 2000), suggesting an important role of *wg* in the regulation of cell proliferation in the maxillary segment. The third notable defect of *Dfd* mutants is the loss of the maxillary cirri primordium (Regulski et al., 1987). *paired* (*prd*), one of the transcription factor genes identified in our screen (see Table S1 in the supplementary material), is known to be important for development of cirri and the maxillary ventral organ (Vanario-Alonso et al., 1995). Because late *prd* expression is completely under the control of Dfd (Fig. 7N,S), we conclude that some aspects of ventral maxillary identity are specified by Dfd via *prd* regulation. Finally, we analyzed Dfd-dependent regulation of cell shape changes, because cells at ventral positions of wild-type maxillary segments are round (Fig. 7G), whereas in *Dfd* mutants many appeared elongated (Fig. 7H). The JNK pathway has been implicated in cell shape changes in *Drosophila*, for example during embryonic dorsal closure and adult thorax closure (Harden, 2002; Xia and Karin, 2004) and because we had identified several genes responsive to the JNK pathway (Jasper et al., 2001) [*Ras-related protein* (*Rala*), *Angiotensin converting enzyme* (*Ance*) and *CG5080*] (see Table S1 in the supplementary material) as Dfd downstream genes, we tested the contribution of the JNK pathway to the cell shape phenotype of *Dfd* mutants. After ubiquitous activation of the JNK pathway by overexpressing a constitutively active form of Hemipterous (Weber et al., 2000) using the *arm*-GAL4 driver, we observed elongated cells in the maxillary segment (Fig. 7I), as well as in other parts of the embryo (data not shown). As we could confirm for one of the JNK-responsive Dfd downstream genes, *CG5080*, implicated in the regulation of cytoskeletal dynamics (Jasper et al., 2001), strong upregulation by Dfd (Fig. 7T), we conclude that the JNK pathway plays a major role in organizing cell shapes in the maxillary segment.

DISCUSSION

More than 30 years ago Antonio Garcia-Bellido proposed that a hierarchy of three classes of genes, activators, selectors and realizators, accounts for cell differentiation during development, thereby providing a functional scheme for the control of morphogenetic processes. The key proposal was that, once activated in their appropriate territories by so-called activator genes, selector genes (he applied this name to homeotic genes) would not directly specify morphological differences between different segments, but would rather select a battery of subordinate downstream genes, the realizator genes, encoding cellular proteins directly required in cell

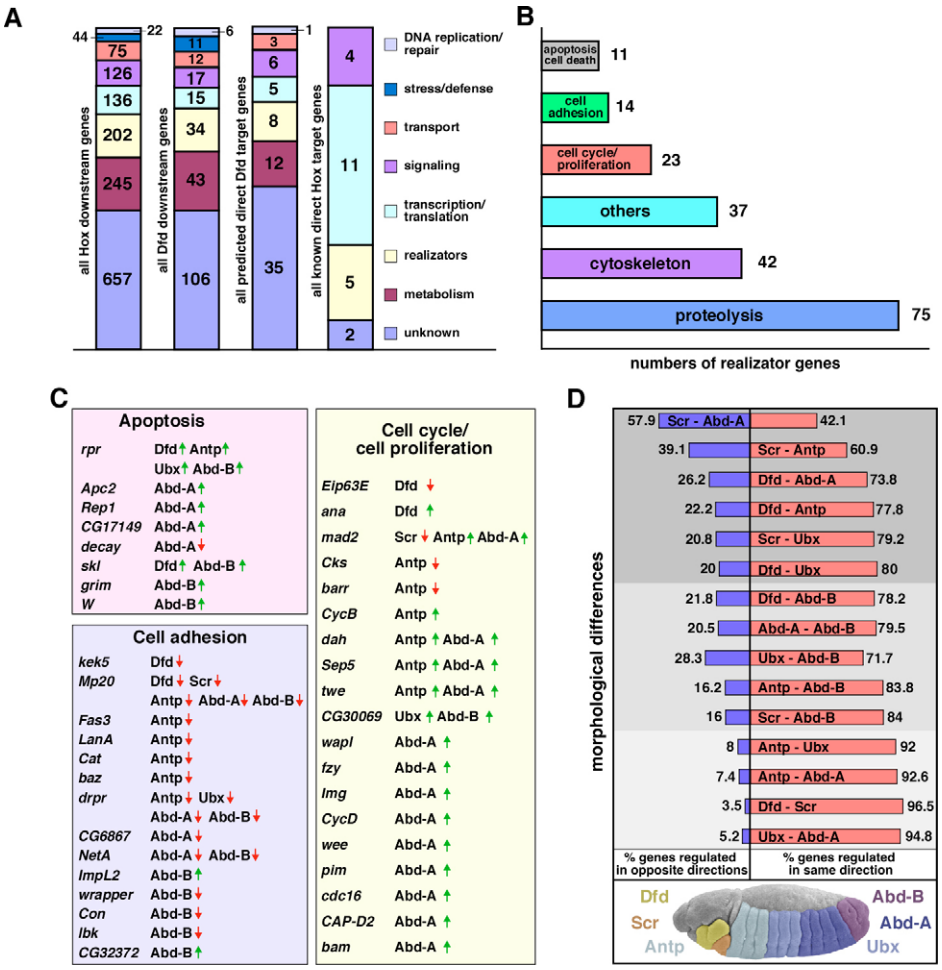

Fig. 6. Functional classification of Hox downstream genes using GO categories. (**A**) Functional categories of downstream genes are shown for all newly identified Hox downstream genes (first column), for all identified Dfd downstream genes (second column), for all predicted direct Dfd target genes (third column) and for all known direct Hox target genes (fourth column). Numbers of genes for each category are indicated within bars. (**B**) Diagram showing subclasses of realizator genes, with numbers of genes for each class indicated. (**C**) Subclasses of realizators are often coordinately regulated, as shown here by three examples (apoptosis, cell adhesion, cell cycle or cell proliferation). Green arrows, increased expression; red arrows, reduced expression. (**D**) Morphological differences along the AP axis are reflected in the percentage of shared downstream genes regulated opposingly by two different Hox proteins. Light gray indicates similar morphologies; medium gray and dark gray indicate increasing differences in morphologies directed by the *Hox* genes compared. The scanning electron micrograph shows the morphology of a stage 13 embryo, with the expression domains of the different Hox proteins highlighted.

differentiation processes (Garcia-Bellido, 1975). Until now much effort has gone into elucidating the nature and function of all three hierarchical levels, with a substantial amount of knowledge having accumulated at the activator and selector level. It is now well established that a genetic cascade comprising maternal and various classes of segmentation genes regulate the temporal and spatial expression of a unique combination of Hox genes in different segments, which subsequently specifies the identities of individual segments (McGinnis and Krumlauf, 1992; St Johnston and Nusslein-Volhard, 1992). Additionally, the discovery that Hox proteins act as transcriptional regulators established the general view that each segment will enter a specific morphogenetic program and develop unique shape and function depending on Hox downstream genes, in particular the realizator genes. Although the question of Hox downstream gene identity and function is not a novel problem, and although a considerable amount of progress has been made in recent years, our knowledge of their nature and function is still far from complete, especially with regard to the realizator genes in the sense of Garcia-Bellido.

Many of the known Hox downstream genes coded either for transcription factors or signaling molecules, and only very few of them were realizators (Hombria and Lovegrove, 2003; Pearson et al., 2005; Pradel and White, 1998). This was puzzling, as the primary function of Hox proteins is to specify the morphology of different segments, thus one would have expected to find a bias toward realizators. Moreover, this finding established the view that most of the cellular responses mediated by Hox proteins, including realizator functions, are indirectly influenced through the action of intermediate regulatory molecules. Our analysis of Hox downstream genes in *Drosophila*, which was designed to allow for a quantitative identification of Hox-regulatory networks (including most realizator genes), revealed that a major group of genes responsive to Hox input did indeed code for realizators. Therefore, our results constitute the first experimental support of the concept postulated by Garcia-Bellido more than three decades ago. We could furthermore show that a substantial part of the Hox output is directly transferred to the realizator level, suggesting that intermediate regulators might play a smaller role than previously thought. One possible explanation why so few Hox realizators had been identified before is that most realizators will be required for general functions in many cells. Consequently, mutations in realizator genes are likely to result either in early embryonic lethality or in pleiotropic effects, making it difficult to correlate their phenotypes to those found in Hox mutants. In addition, it seems likely that realizators act redundantly or have very subtle effects, making their identification in forward genetic screens extremely difficult. Similarly, individual mutations in all known guidance factors for border cell migration in *Drosophila* produce either no, or only mild, defects and thus they could be identified only by expression profiling studies (Wang et al., 2006). In this study we have quantitatively identified Hox realizator genes by a comparative microarray analysis, which now can serve as a resource to study the mechanisms of segmental morphogenesis. Focusing on the differentiation of the maxillary segment, we were able to functionally correlate all major morphological defects

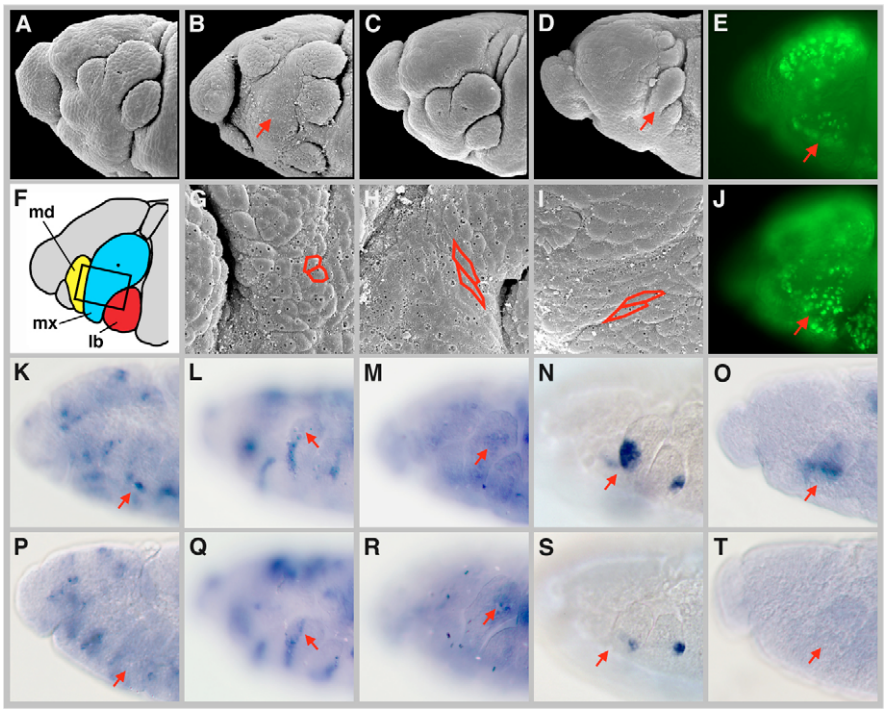

Fig. 7. A framework for the morphogenesis of the maxillary segment in *Drosophila*. (**A**,**C**) Scanning electron micrographs of heads of stage 12 wild-type embryos. (**B**) Head of a stage 12 *Dfd* mutant embryo. The red arrow marks additional cells at the ventral side of the maxillary segment. (**D**) Head of a *wg* mutant embryo. The red arrow marks the size-reduced maxillary segment. (**E**,**J**) BrdU labeling of stage 12 wild-type and *Dfd* mutant embryos, respectively. Red arrows mark proliferating BrdU-positive cells at the ventral side of the maxillary segment. (**F**) Diagram of a stage 12 wild-type embryo. Mandibular (md), maxillary (mx) and labial (lb) segments are indicated; the box marks the area shown in G,H,I. (**G**,**H**,**I**) In stage 12 wild-type embryos, cells in the ventral part of the maxillary segment are round (G), whereas in *Dfd* mutant (H) and *arm*::*hep*$^{act.}$ (I) embryos, cells are elongated (marked in red). (**K-T**) *skl* (K,P), *wg* (L,Q), *Eip63E* (M,R), *prd* (N,S) and *CG5080* (O,T) RNA expression in wild-type and *Dfd* mutant embryos, respectively. The red arrows indicate the expression of the respective genes that differs in wild-type and *Dfd* mutant embryos.

observed in *Dfd* mutants with newly identified Dfd downstream genes, many of which code for realizators, demonstrating the validity of this approach.

Another important discovery of our analysis is the enormous specificity of Hox protein action on the transcriptome in vivo, which sharply contrasts with the low DNA-binding specificity in vitro. Hox proteins have been shown to bind to very similar, relatively simple, DNA sequences containing a TAAT core sequence in vitro (Biggin and McGinnis, 1997; Carr and Biggin, 1999; Ekker et al., 1994; Walter and Biggin, 1996), whereas many of the identified Hox downstream genes are uniquely regulated by only a single Hox protein. This contrast may be explained by our observation that the majority of genes are primarily regulated at only one of the two stages, implicating that Hox proteins excessively interact with the regulatory environment in which they are embedded. Support for the notion that co-factor interactions have a major impact on Hox output also comes from a recent study, which has provided direct evidence that Hox proteins gain the ability to regulate their target genes in a context-specific manner by interaction with known cell- and/or tissue-specific transcription factors in vivo (Gebelein et al., 2004). In addition, this study also suggests that a large number of transcription factors might function as Hox co-factors, which could dictate the outcome of Hox gene action. Along these lines, we found that ubiquitous overexpression of Hox proteins never caused ubiquitous activation of downstream genes, but that ectopic expression was always locally restricted, suggesting that regional transcription factors are essential for Hox output. This is also reflected in our finding that conserved clusters of Hox binding sites in the regulatory regions of direct targets frequently contain binding sites for unrelated transcription factors. Taken together, these results support the hypothesis put forward by Michael Akam in 1998, that "we should think of the Hox genes with their short and relatively non-specific target sequences, as cofactors that modify the actions of other more specific transcription factors, rather than proteins in need of cofactors themselves" (Akam, 1998).

We thank Jan U. Lohmann for help with the microarray analysis and quantitative real-time PCR, Benjamin Berman for receiving the stand-alone binaries and support for Cis-analyst, Tancred Frickey for help with gene annotations, Jürgen Berger for advice with the SEMs, Stephan Ossowski for help with databases, Petra Stöbe for assistance with gelshifts and Jan U. Lohmann and Detlef Weigel for discussion and for critically reading the manuscript. Supported by the Deutsche Forschungsgemeinschaft (I.L.; LO 844/3-1) and by the Max Planck Society.

Supplementary material

Supplementary material for this article is available at
http://dev.biologists.org/cgi/content/full/134/2/381/DC1

References

Abouelhoda, M. I., Kurtz, S. and Ohlebusch, E. (2002). The enhanced suffix array and its applications to genome analysis. In *Algorithms in Bioinformatics: Second International Workshop, WABI, Rome, Italy 17-21, 2002, Proceedings* (*Lecture Notes in Computer Science*). Vol. 2452 (ed. G. Goos, J. Hartmanis and J. van Leeuwen), pp. 449-463. Berlin, Heidelberg: Springer Verlag.

Abrams, E. W. and Andrew, D. J. (2002). Prolyl 4-hydroxylase alpha-related proteins in *Drosophila melanogaster*: tissue-specific embryonic expression of the 99F8-9 cluster. *Mech. Dev.* **112**, 165-171.

Akam, M. (1998). *Hox* genes: from master genes to micromanagers. *Curr. Biol.* **8**, R676-R678.

Bello, B. C., Hirth, F. and Gould, A. P. (2003). A pulse of the *Drosophila* Hox protein Abdominal-A schedules the end of neural proliferation via neuroblast apoptosis. *Neuron* **37**, 209-219.

Bergman, C. M., Carlson, J. W. and Celniker, S. E. (2005). *Drosophila* DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, *Drosophila melanogaster*. *Bioinformatics* **21**, 1747-1749.

Bergson, C. and McGinnis, W. (1990). An autoregulatory enhancer element of the *Drosophila* homeotic gene *Deformed*. *EMBO J.* **9**, 4287-4297.

Berman, B. P., Pfeiffer, B. D., Laverty, T. R., Salzberg, S. L., Rubin, G. M., Eisen, M. B. and Celniker, S. E. (2004). Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol.* **5**, R61.

Biggin, M. D. and McGinnis, W. (1997). Regulation of segmentation and segmental identity by *Drosophila* homeoproteins: the role of DNA binding in functional activity and specificity. *Development* **124**, 4425-4433.

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

Capovilla, M., Kambris, Z. and Botas, J. (2001). Direct regulation of the muscle-

identity gene apterous by a Hox protein in the somatic mesoderm. *Development* **128**, 1221-1230.
Carr, A. and Biggin, M. D. (1999). A comparison of *in vivo* and *in vitro* DNA-binding specificities suggests a new model for homeoprotein DNA binding in *Drosophila* embryos. *EMBO J.* **18**, 1598-1608.
Carroll, S. B. (1995). Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479-485.
Chandrasekaran, V. and Beckendorf, S. K. (2003). senseless is necessary for the survival of embryonic salivary glands in *Drosophila*. *Development* **130**, 4719-4728.
Cobb, J. and Duboule, D. (2005). Comparative analysis of genes downstream of the *Hoxd* cluster in developing digits and external genitalia. *Development* **132**, 3055-3067.
DeLotto, Y., Smith, C. and DeLotto, R. (2001). Multiple isoforms of the *Drosophila* Spatzle protein are encoded by alternatively spliced maternal mRNAs in the precellular blastoderm embryo. *Mol. Gen. Genet.* **264**, 643-652.
Dolbeare, F. and Selden, J. R. (1994). Immunochemical quantitation of bromodeoxyuridine application to cell-cycle kinetics. *Methods Cell Biol.* **41**, 297-316.
Ebner, A., Cabernard, C., Affolter, M. and Merabet, S. (2005). Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex. *Development* **132**, 1591-1600.
Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E. and Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *EMBO J.* **13**, 3551-3560.
Feinstein, P. G., Kornfeld, K., Hogness, D. S. and Mann, R. S. (1995). Identification of homeotic target genes in *Drosophila melanogaster* including *nervy*, a proto-oncogene homologue. *Genetics* **140**, 573-586.
Garbe, J. C., Yang, E. and Fristrom, J. W. (1993). IMP-L2: an essential secreted immunoglobulin family member implicated in neural and ectodermal development in *Drosophila*. *Development* **119**, 1237-1250.
Garcia-Bellido, A. (1975). Genetic control of wing disc development in *Drosophila*. *Ciba Found. Symp.* **29**, 161-182.
Gebelein, B., McKay, D. J. and Mann, R. S. (2004). Direct integration of Hox and segmentation gene inputs during *Drosophila* development. *Nature* **431**, 653-659.
Giraldez, A. J. and Cohen, S. M. (2003). Wingless and Notch signaling provide cell survival cues and control cell proliferation during wing development. *Development* **130**, 6533-6543.
Gould, A., Morrison, A., Sproat, G., White, R. A. and Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. *Genes Dev.* **11**, 900-913.
Graba, Y., Aragnol, D. and Pradel, J. (1997). *Drosophila* Hox complex downstream targets and the function of homeotic genes. *BioEssays* **19**, 379-388.
Harden, N. (2002). Signaling pathways directing the movement and fusion of epithelial sheets: lessons from dorsal closure in *Drosophila*. *Differentiation* **70**, 181-203.
Hedlund, E., Karsten, S. L., Kudo, L., Geschwind, D. H. and Carpenter, E. M. (2004). Identification of a Hoxd10-regulated transcriptional network and combinatorial interactions with Hoxa10 during spinal cord development. *J. Neurosci. Res.* **75**, 307-319.
Hirth, F., Loop, T., Egger, B., Miller, D. F., Kaufman, T. C. and Reichert, H. (2001). Functional equivalence of *Hox* gene products in the specification of the tritocerebrum during embryonic brain development of *Drosophila*. *Development* **128**, 4781-4788.
Hombria, J. C. and Lovegrove, B. (2003). Beyond homeosis – HOX function in morphogenesis and organogenesis. *Differentiation* **71**, 461-476.
Jasper, H., Benes, V., Schwager, C., Sauer, S., Clauder-Munster, S., Ansorge, W. and Bohmann, D. (2001). The genomic response of the *Drosophila* embryo to JNK signaling. *Dev. Cell* **1**, 579-586.
Jullien, D., Crozatier, M. and Kas, E. (1997). cDNA sequence and expression pattern of the *Drosophila melanogaster* PAPS synthetase gene: a new salivary gland marker. *Mech. Dev.* **68**, 179-186.
Leemans, R., Loop, T., Egger, B., He, H., Kammermeier, L., Hartmann, B., Certa, U., Reichert, H. and Hirth, F. (2001). Identification of candidate downstream genes for the homeodomain transcription factor Labial in *Drosophila* through oligonucleotide-array transcript imaging. *Genome Biol.* **2**, RESEARCH0015.
Lei, H., Wang, H., Juan, A. H. and Ruddle, F. H. (2005). The identification of Hoxc8 target genes. *Proc. Natl. Acad. Sci. USA* **102**, 2420-2424.
Lemon, W. J., Liyanarachchi, S. and You, M. (2003). A high performance test of differential gene expression for oligonucleotide arrays. *Genome Biol.* **4**, R67.
Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565-570.
Li, X., Murre, C. and McGinnis, W. (1999). Activity regulation of a Hox protein and a role for the homeodomain in inhibiting transcriptional activation. *EMBO J.* **18**, 198-211.
Liang, Z. and Biggin, M. D. (1998). Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in *Drosophila*. *Development* **125**, 4471-4482.
Lohmann, I., McGinnis, N., Bodmer, M. and McGinnis, W. (2002). The *Drosophila Hox* gene *Deformed* sculpts head morphology via direct regulation of the apoptosis activator *reaper*. *Cell* **110**, 457-466.
Mahaffey, J. W. (2005). Assisting Hox proteins in controlling body form: are there new lessons from flies (and mammals)? *Curr. Opin. Genet. Dev.* **15**, 422-429.
Mahaffey, J. W., Jones, D. F., Hickel, J. A. and Griswold, C. M. (1993). Identification and characterization of a gene activated by the Deformed homeoprotein. *Development* **118**, 203-214.
Mann, R. S. (1995). The specificity of homeotic gene function. *BioEssays* **17**, 855-863.
Mann, R. S. and Affolter, M. (1998). Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* **8**, 423-429.
Mann, R. S. and Morata, G. (2000). The developmental and molecular biology of genes that subdivide the body of *Drosophila*. *Annu. Rev. Cell Dev. Biol.* **16**, 243-271.
McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68**, 283-302.
McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. and Gehring, W. J. (1984). A conserved DNA sequence in homoeotic genes of the *Drosophila Antennapedia* and *bithorax* complexes. *Nature* **308**, 428-433.
McGinnis, W., Jack, T., Chadwick, R., Regulski, M., Bergson, C., McGinnis, N. and Kuziora, M. A. (1990). Establishment and maintenance of position-specific expression of the *Drosophila* homeotic selector gene *Deformed*. *Adv. Genet.* **27**, 363-402.
Merabet, S., Pradel, J. and Graba, Y. (2005). Getting a molecular grasp on Hox contextual activity. *Trends Genet.* **21**, 477-480.
Michelson, A. M. (1994). Muscle pattern diversification in *Drosophila* is determined by the autonomous function of homeotic genes in the embryonic mesoderm. *Development* **120**, 755-768.
Miller, D. F., Rogers, B. T., Kalkbrenner, A., Hamilton, B., Holtzman, S. L. and Kaufman, T. (2001). Cross-regulation of *Hox* genes in the *Drosophila melanogaster* embryo. *Mech. Dev.* **102**, 3-16.
Panzer, S., Weigel, D. and Beckendorf, S. K. (1992). Organogenesis in *Drosophila melanogaster*: embryonic salivary gland determination is controlled by homeotic and dorsoventral patterning genes. *Development* **114**, 49-57.
Pearson, J. C., Lemons, D. and McGinnis, W. (2005). Modulating *Hox* gene functions during animal body patterning. *Nat. Rev. Genet.* **6**, 893-904.
Pradel, J. and White, R. A. (1998). From selectors to realizators. *Int. J. Dev. Biol.* **42**, 417-421.
Regulski, M., McGinnis, N., Chadwick, R. and McGinnis, W. (1987). Developmental and molecular analysis of *Deformed*: a homeotic gene controlling *Drosophila* head development. *EMBO J.* **6**, 767-777.
Rusch, D. B. and Kaufman, T. C. (2000). Regulation of *proboscipedia* in *Drosophila* by homeotic selector genes. *Genetics* **156**, 183-194.
Ryoo, H. D. and Mann, R. S. (1999). The control of trunk Hox specificity and activity by Extradenticle. *Genes Dev.* **13**, 1704-1716.
Sanson, B., White, P. and Vincent, J. P. (1996). Uncoupling cadherin-based adhesion from wingless signalling in *Drosophila*. *Nature* **383**, 627-630.
Schmid, M., Uhlenhaut, N. H., Godard, F., Demar, M., Bressan, R., Weigel, D. and Lohmann, J. U. (2003). Dissection of floral induction pathways using global expression analysis. *Development* **130**, 6001-6012.
Scott, M. P. and Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **81**, 4115-4119.
St Johnston, D. and Nusslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201-219.
Stowers, R. S., Garza, D., Rascle, A. and Hogness, D. S. (2000). The L63 gene is necessary for the ecdysone-induced 63E late puff and encodes CDK proteins required for *Drosophila* development. *Dev. Biol.* **221**, 23-40.
Tautz, D. and Pfeifle, C. (1989). A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. *Chromosoma* **98**, 81-85.
Vanario-Alonso, C. E., O'Hara, E., McGinnis, W. and Pick, L. (1995). Targeted ribozymes reveal a conserved function of the *Drosophila paired* gene in sensory organ development. *Mech. Dev.* **53**, 323-328.
Wakimoto, B. T., Turner, F. R. and Kaufman, T. C. (1984). Defects in embryogenesis in mutants associated with the antennapedia gene complex of *Drosophila melanogaster*. *Dev. Biol.* **102**, 147-172.
Walter, J. and Biggin, M. D. (1996). DNA binding specificity of two homeodomain proteins *in vitro* and in *Drosophila* embryos. *Proc. Natl. Acad. Sci. USA* **93**, 2680-2685.
Wang, X., Bo, J., Bridges, T., Dugan, K. D., Pan, T. C., Chodosh, L. A. and Montell, D. J. (2006). Analysis of cell migration using whole-genome expression profiling of migratory cells in the *Drosophila* ovary. *Dev. Cell* **10**, 483-495.
Weber, U., Paricio, N. and Mlodzik, M. (2000). Jun mediates Frizzled-induced R3/R4 cell fate distinction and planar polarity determination in the *Drosophila* eye. *Development* **127**, 3619-3629.
Williams, T. M., Williams, M. E., Kuick, R., Misek, D., McDonagh, K., Hanash,

S. and Innis, J. W. (2005). Candidate downstream regulated genes of HOX group 13 transcription factors with and without monomeric DNA binding capability. *Dev. Biol.* **279**, 462-480.
Wing, J. P., Karres, J. S., Ogdahl, J. L., Zhou, L., Schwartz, L. M. and Nambu, J. R. (2002). *Drosophila sickle* is a novel grim-reaper cell death activator. *Curr. Biol.* **12**, 131-135.
Wu, Z., Irizarry, R. A., Gentleman, R., Murillo, F. M. and Spencer, F. A. A. (2004). A model based background adjustment for oligonucleotide expression arrays. Working paper 1. http://www.bepress.com/jhubiostat/paper1.
Xia, Y. and Karin, M. (2004). The control of cell motility and epithelial morphogenesis by Jun kinases. *Trends Cell Biol.* **14**, 94-101.
Zhu, X., Sen, J., Stevens, L., Goltz, J. S. and Stein, D. (2005). *Drosophila* pipe protein activity in the ovary and the embryonic salivary gland does not require heparan sulfate glycosaminoglycans. *Development* **132**, 3813-3822.

Shaping segments: *Hox* gene function in the genomic age

Stefanie D. Hueber[1] and Ingrid Lohmann[1,2]*

Summary

Despite decades of research, morphogenesis along the various body axes remains one of the major mysteries in developmental biology. A milestone in the field was the realisation that a set of closely related regulators, called *Hox* genes, specifies the identity of body segments along the anterior–posterior (AP) axis in most animals. *Hox* genes have been highly conserved throughout metazoan evolution and code for homeodomain-containing transcription factors. Thus, they exert their function mainly through activation or repression of downstream genes. However, while much is known about *Hox* gene structure and molecular function, only a few target genes have been identified and studied in detail. Our knowledge of *Hox* downstream genes is therefore far from complete and consequently Hox-controlled morphogenesis is still poorly understood. Genome-wide approaches have facilitated the identification of large numbers of *Hox* downstream genes both in *Drosophila* and vertebrates, and represent a crucial step towards a comprehensive understanding of how *Hox* proteins drive morphological diversification. In this review, we focus on the role of *Hox* genes in shaping segmental morphologies along the AP axis in *Drosophila*, discuss some of the conclusions drawn from analyses of large target gene sets and highlight methods that could be used to gain a more thorough understanding of *Hox* molecular function. In addition, the mechanisms of *Hox* target gene regulation are considered with special emphasis on recent findings and their implications for *Hox* protein specificity in the context of the whole organism. *BioEssays* 30:965–979, 2008. © 2008 Wiley Periodicals, Inc.

Introduction

All bilateral animals possess a common genetic mechanism regulating development along the AP axis,[1,2] and *Hox* proteins are among the key regulators in specifying morphological diversity along this axis[3–6] (Fig. 1). In all animals studied, *Hox* genes are expressed in defined and often overlapping domains along the AP axis, and it is their activity that assigns distinct morphologies to the various body segments.[3,5] This becomes most evident when *Hox* gene function is disrupted, which frequently results in "homeotic transformations".[6,7] The term "homeotic transformation", defined by Bateson in 1894,[8] is used to describe the transformation of one structure to resemble, in form and shape, a homologous structure present in the body. For example, in *Drosophila* mutations in the *Hox* gene, *Ultrabithorax* (*Ubx*) result in the development of an additional pair of wings instead of halteres, two small balancing organs, giving rise to the famous four-winged fly, discovered by Ed Lewis.[6] Although first observed in *Drosophila*, homeotic transformations are found in many other organisms,[9,10] which led to the assumption that *Hox* proteins act as master regulators of morphogenesis. However, mutations in *Hox* genes do not always result in such dramatic phenotypes—they can also cause very subtle defects, as frequently observed in organisms with multiple *Hox* clusters (e.g. vertebrates) due to the overlapping expression and functional redundancy of paralogous *Hox* genes of different clusters.[11] In these cases, major morphological changes are only observed when paralogous *Hox* genes are simultaneously mutated. But even in organisms with a single *Hox* cluster, as in *Drosophila*, homeotic transformations are primarily observed after mutations in those *Hox* genes that either have overlapping expression domains or are engaged in a negative cross-regulation with other *Hox* genes.[5,12] Loss of function of one *Hox* gene allows the overlapping or ectopically expressed *Hox* gene to exert its function, which results in the transformation of one segment identity towards the identity of neighbouring segments.[5,13] This implies that homeotic transformations are actually not very informative with regards to the function of the mutant *Hox* gene, but rather provide insights about the function of nearby or overlapping *Hox* genes.[5] Since it is mostly the more posterior *Hox* protein repressing the expression of a more anterior one, this phenomenon was termed posterior suppression.[14,15] When no "backup" *Hox* gene is present, the functional elimination of a *Hox* gene does not result in homeotic transformation, but in structural deficiencies,[5,16] as observed for many other mutations.

On the molecular level, *Hox* genes encode proteins with a highly conserved 60-amino-acid DNA-binding motif, the homeodomain,[17–19] and function as transcription factors by directly binding to DNA sequences in *Hox* response elements (HREs)[20,21] (Fig. 2). Thus, it seems obvious that *Hox* proteins

[1]Department of Molecular Biology, AG I. Lohmann, MPI for Developmental Biology, Tübingen, Germany.
[2]BIOQUANT Center, NWG Developmental Biology, University of Heidelberg, Germany.
*Correspondence to: Ingrid Lohmann, BIOQUANT Zentrum, BQ 0032 NWG Developmental Biology, Im Neuenheimer Feld 267, 69120 Heidelberg. E-mail: ilohmann@flydev.org
DOI 10.1002/bies.20823
Published online in Wiley InterScience (www.interscience.wiley.com).

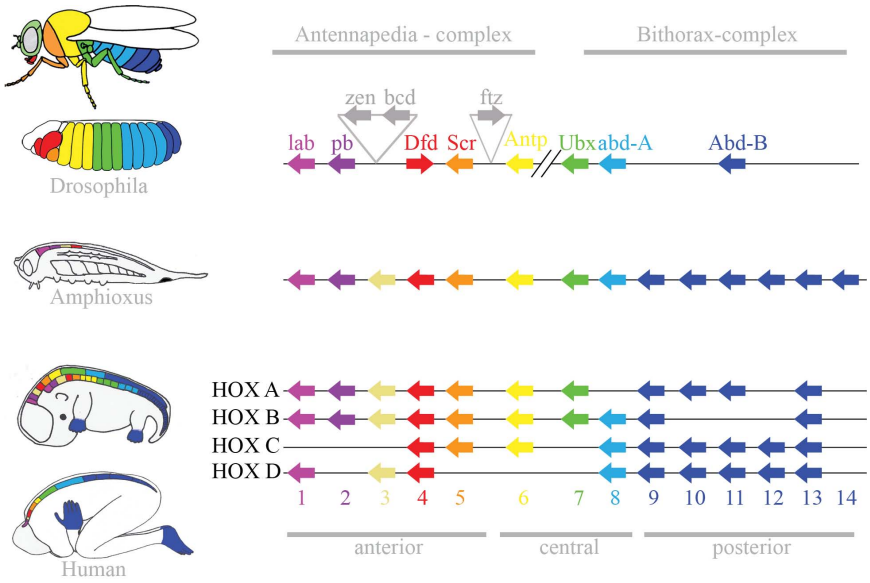

Figure 1. *Hox* expression and genomic organization in different organisms. **Left panel:** Schematic representations of different organisms, the fruit fly (*Drosophila melanogaster*, adult fly and a stage 13 embryo), amphioxus (*Branchiostoma belcheri*), mouse (*Mus musculus*) and human (*Homo sapiens*; adult and embryonic stage) are shown, with approximate RNA expression domains of most *Hox* genes. Colour-code as in the *Hox* cluster diagrams. **Right panel:** A scheme of the *Hox* gene clusters (not to scale) in the genomes of the species shown on the left panel. Genes are coloured to differentiate between *Hox* family members, orthologous genes are labelled in the same colour. Genes are shown in the order in which they are found on the chromosomes, the positions of three non-*Hox* homeodomain genes, *zen, bcd* and *ftz*, within the cluster of the fly genome are shown in grey. Gene abbreviations are: *lab, labial; pb, proboscipedia; Dfd, Deformed; Scr, Sex combs reduced; Antp, Antennapedia; Ubx, Ultrabithorax; abd-A, abdominal-A; Abd-B, Abdominal-B; zen, zerknüllt; bcd, bicoid; ftz, fushi-tarazu*.

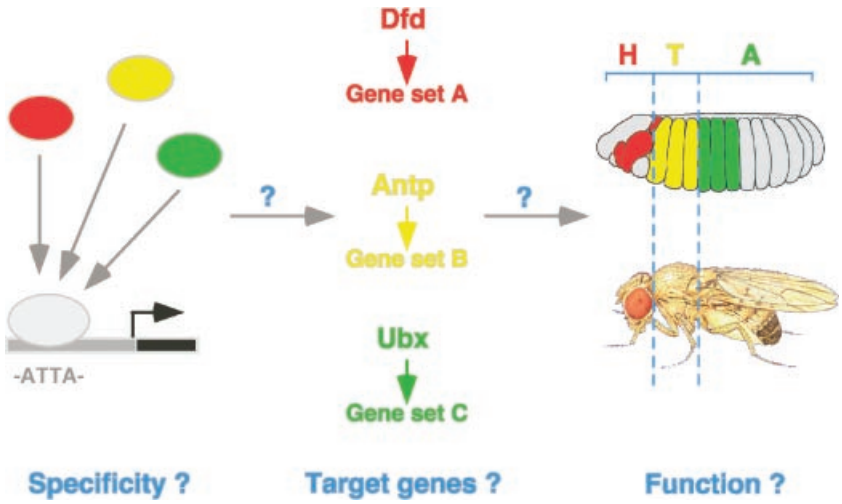

Figure 2. Schematic illustration of the "*Hox* paradox". Since *Hox* genes encode transcription factors, it seems obvious that they control different morphological outputs along the AP axis by controlling distinct sets of downstream genes. Thus, the key to understanding *Hox* function is to identify these genes and analyze their role during development. However, there are many problems: in vitro *Hox* proteins have rather unspecific DNA-binding specificities, thus they all bind to very similar and frequently occurring DNA core sequences. Therefore, it has been unclear for a long time which specific sets of downstream genes they regulate and how they can do so. The unspecific DNA-binding preference is also one of the major reasons why so few *Hox* downstream genes have been identified and functionally analyzed in the pre-genomic age. Consequently, many aspects of the morphogenetic function of *Hox* proteins are still not understood.

drive the morphological diversification of body segments by differentially controlling the expression of downstream genes (Fig. 2). While this is a straightforward assumption, there is one major problem: all *Hox* proteins display a poor in vitro DNA-binding specificity and recognize highly similar nucleotide sequences containing an -ATTA- core[22–26] (Fig. 2). In contrast, *Hox* proteins have very specific effects in vivo, and different *Hox* transcription factors target diverse sets of downstream genes.[27] Furthermore, even a single *Hox* protein is able to regulate different sets of downstream genes depending on tissue context or developmental stage, and some downstream genes are activated in one context and repressed in another.[27–29] The ambiguous nature of DNA binding by *Hox* proteins, along with the complexity of the biological processes controlled by them has hampered the identification of *Hox* target genes, despite the use of a wide range of strategies.[30–32] Only recently it has become feasible to quantitatively identify novel *Hox* downstream by genome-wide approaches.[27,33–35]

In this article, we will first summarize what was known about *Hox* downstream genes and mechanisms of *Hox* target gene regulation in the pre-genomic age, next describe recent progress in these two fields using genome-wide approaches and, finally, discuss how these recent findings have influenced our views of how *Hox* proteins exert their fundamental role in the morphological diversification of segments along the AP axis in vivo. Given the enormous amount of published data on these topics, this review cannot be exhaustive. Therefore, we focus on what is known in *Drosophila* and only include selected data from other organisms.

Hox genes—the pre-genomic era

Although *Hox* proteins have highly complex roles in specifying segment identities along the AP axis in animals, one can simplify their activities by focussing on their molecular nature. As transcription factors, *Hox* proteins control morphogenesis by regulation of distinct sets of target genes (Fig. 2). Thus, the key to understanding *Hox* function is to identify these genes and analyze their role during development.

Identification of Hox downstream genes

Before the advent of large-scale techniques, a diverse repertoire of approaches was used to identify *Hox* downstream genes[30–32] (Table 1). Initial attempts focused on screens in heterologous systems, like yeast one-hybrid assays performed to identify regulatory elements mediating *Hox* responses.[36] However, these approaches showed limited success, since only very few *Hox* target genes could be identified. We now know that the most likely reason for this limitation lies in the fact that *Hox* proteins acquire DNA-binding specificity and thus specificity in target gene selection through interactions with additional DNA-binding proteins in vivo.[37–40] Thus, it is not surprising that most *Hox* downstream genes were initially identified by candidate gene approaches based on homeotic responses of transcript or enhancer trap patterns[3]. These findings clearly highlight the power of in vivo strategies for the identification of *Hox* target genes. For example, two well-characterized *Hox* targets identified in enhancer trap screens are *decapentaplegic* (*dpp*),[43] a gene member of the TGF-β family of signaling proteins, and the homeobox transcription factor gene *Distal-less* (*Dll*).[44] Although very powerful in identifying *Hox* downstream genes, these approaches did not allow a discrimination of direct and indirect targets without additional and tedious experimentation. This knowledge, however, was regarded as essential, since direct targets could be used not only to explore *Hox* function, but also to elucidate the mechanisms of *Hox* target gene selection and regulation. In this context, chromatin immunoprecipitation (ChIP) was developed, which has become one of the most powerful tools for the study of protein–DNA interactions in vivo. Before genomic arrays or massively parallel sequencing technologies became available, immunoprecipitation of genomic DNA fragments associated with *Hox* proteins in vivo was used to clone transcription units in their vicinity. In addition to isolating targets, this procedure had the added advantage that HREs in the regulatory regions of these target genes were identified. Some of the direct *Hox* targets identified using this technique are *scabrous* (*sca*), *Transcript 48* (*T48*) and *centrosomin* (*cnn*), which are under direct control of Ubx.[45–47] Taken together, all these methods resulted in the identification of 24 direct *Hox* downstream genes in *Drosophila* (Table 1), which then served as models to study *Hox* gene output.

Nature and function of Hox downstream genes: transcription factors—realisators—regulatory networks

In 1975, Antonio Garcia-Bellido established a concept in which he proposed a hierarchy of three classes of genes, activators, selectors and realisators to be responsible for cell differentiation in development. One key element of his proposal was that, once activated in their appropriate territories by the activator genes, the homeotic (*Hox*) selector genes would select a large number of subordinate targets, the realisator genes, that would directly influence the morphology of segments by regulating cytodifferentiation processes.[48] Based on this idea, it had been expected that realisator genes would constitute a large fraction of *Hox* target genes. However, examination of the few *Hox* downstream genes known at that time showed that many of them encoded regulatory molecules, mostly transcription factors, like Distal-less (Dll), Forkhead (Fkh) and Teashirt (Tsh), and some signalling proteins, like Decapentaplegic (Dpp), Wingless (Wg) and Scabrous (Sca)[3] (Table 1). These molecules act as high-level executives and very often function themselves to select the activity of a large number of downstream genes.

Table 1. Direct Hox target genes identified in the pre-genomic age in *Drosophila* (adapted from Pearson et al., 2005)

Target gene	Regulated by	Function	Target class	Validation	References
1.28	Dfd	unknown	unknown	reporter lines	Pederson et al., 2000[118]
Antennapedia	Antp, Ubx, Abd-A	homeobox TF; thorax development	transcription factor	reporter lines	Appel and Sakonju,1993[51]
apterous	Antp	homeobox TF; muscle identity	transcription factor	reporter lines	Capovilla et al., 2001[111]
CG11339	Lab	actin binding protein	realisator	reporter lines	Ebner et al., 2005[41]
CG13222	Ubx	cuticle protein	realisator	EMSA	Hersh et al., 2007[93]
centrosomin	Antp	centrosomal protein; PNS and CNS development	realisator	ChIP	Heuer et al., 1995[47]
connectin	Ubx, Abd-A	GPI linked cell surface protein; neuromuscular connection	realisator	ChIP	Gould and White, 1992[54]
decapentaplegic	Ubx, Abd-A	Tgf-ß protein; D/V polarity, midgut morphogenesis	signaling molecule	reporter lines	Capovilla et al., 1994[86]
Deformed	Dfd	homeobox TF; head development	transcription factor	reporter lines	Zeng et. al 1994[50]
Distal-less	Ubx, Abd-A	homeobox TF; limb development	transcription factor	reporter lines	Vachon et el., 1992[40]
empty spiracles	Abd-B	homeobox TF; head development, filzkörper specification	transcription factor	reporter lines	Jones and McGinnis, 1993[115]
forkhead	Scr	forkhead domain TF; specification of the terminal region	transcription factor	reporter lines	Ryoo and Mann, 1999[78]; Zhou et al., 2001[121]
knot	Ubx	EBF/Olf1 TF; development of wing imaginal disc	transcription factor	reporter lines	Hersh and Carroll, 2005[122]
labial	Lab	homeobox TF; head development	transcription factor	reporter lines	Grieder et al., 1997[114]
La-related protein	Scr, Ubx	autophagic cell death	realisator	ChIP	Chauvet et al., 2000[112]
reaper	Dfd	apoptosis activator	realisator	reporter lines	Lohmann et al., 2002[52]
serpent	Ubx	Zn finger TF	transcription factor	One-hybrid assay	Mastick, 1995[36]
scabrous	Ubx	secreted signal transducer; eye morphogenesis; CNS and PNS development	signaling molecule	ChIP	Graba et al., 1992[45]
spalt major	Ubx	Zn finger TF; development of wing disc	transcription factor	reporter lines	Galant et al., 2002[83]
Transcript 48	Ubx	transmembrane protein	unknown	ChIP	Strutt and White, 1994[46]
teashirt	Antp, Ubx	Zn finger TF; specification of trunk identity	transcription factor	reporter lines	McCormick et al., 1995[90]
ß-3-tubulin	Ubx	cytoskeletal protein; visceral mesoderm differentiation	realisator	reporter lines	Hinz et al., 1992[53]; Kremser et al., 1999[116]
wingless	Abd-A	Wnt signal transducer; midgut morphogenesis	signaling molecule	reporter lines	Grienenberger et al., 2003[28]
Wnt-4	Ubx	Wnt protein	signaling molecule	ChIP	Graba et al., 1995[113]

Consequently, mutations in these genes resulted in major morphological and patterning defects, and sometimes even in homeotic transformations similar to the ones observed in *Hox* mutants. This is surely one of the reasons why initially transcription factors were preferentially identified as *Hox* targets (in genetic screens). One well-studied example of *Hox* target gene coding for a transcription factor is *Dll*, which is required for appendage formation in ventral regions of *Drosophila* embryos.[44] The *Hox* proteins Ubx, Abdominal-A (Abd-A) and Abdominal-B (Abd-B) repress *Dll* expression, resulting in the absence of limbs in the abdomen[44] (Fig. 3A), whereas *Hox* transcription factors expressed in the head and thorax preferentially do not repress *Dll* transcription (Fig. 3A). This allows the formation of appendages, while the precise spatial context dictates which kind of appendage will develop[49] (Fig. 3A). When *Dll* is co-expressed with *homothorax* (*hth*), another homeodomain containing transcription factor gene under the control of *Hox* proteins,[49] antennae are formed, whereas cell-specific expression of *Dll* (in the absence of *hth* expression) results in the formation of legs[49] (Fig. 3A). Other interesting examples of *Hox* target genes coding for transcription factors are the *Hox* genes themselves. Although there are many examples, we would like to focus on two that have been understood at the molecular level. First, Deformed (Dfd), a head-specific *Hox* protein, is known to maintain its own expression in the maxillary and mandibular segments by interacting with specific binding sites in Dfd autoregulatory enhancer elements.[50] Second, *Antennapedia* (*Antp*), a *Hox* gene primarily expressed in thoracic segments, has been shown to be directly regulated by three different *Hox* proteins, Antp, Ubx and Abd-A:[51] Antp positively autoregulates its own expression in neuronal cells of the thorax by binding to specific DNA sites in a P2-specific enhancer, whereas Ubx and Abd-A prevent this autoregulation in abdominal neuronal cells by competitively interacting with the same sites. If this cross-regulation of *Hox* genes fails in *Hox* mutants, *Hox* proteins are

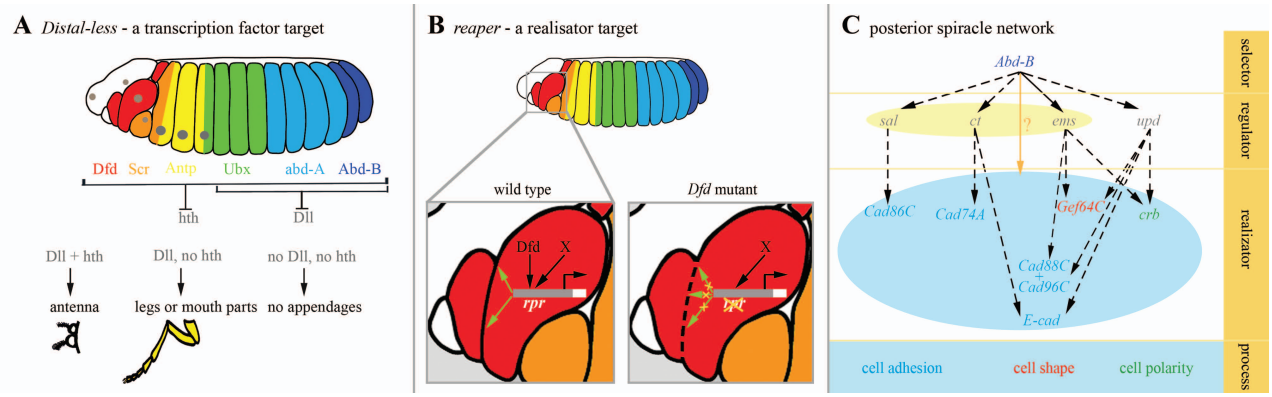

Figure 3. Examples of *Hox* target gene regulation on the executive, realisator and network level. **A:** The *Hox* target gene *Distal-less* (*Dll*), required for appendage formation, encodes a homeobox transcription factor and is one of the high-level executives regulated by *Hox* proteins. The area of *Dll* expression in the embryo (shown as grey patches) corresponds to sites where in the course of development imaginal discs and, subsequently, appendages are formed. This is due to the inability of *Hox* transcription factors expressed in the head and thorax to repress *Dll* expression. In the abdomen, where no appendages are formed, the *Hox* transcription factors Ubx, Abd-A and Abd-B repress *Dll* transcription. **B:** The apoptosis gene *reaper* (*rpr*) is an example of a *Hox* realisator gene target. In wild-type embryos, Deformed (Dfd) directly activates the apoptotic gene *rpr*, leading to localized cell death (marked by green arrows), which sculpts a 'groove' between two head structures, the mandibular and maxillary segments (marked in red). In *Dfd* mutants, *rpr* is not activated, thus localized cell death is not induced, resulting in the fusion of the mandibular and maxillary head segments (indicated by dashed line). **C:** The selector protein Abd-B, supposedly through the interaction with cofactors, activates the spatially and temporarily restricted expression of four "intermediate" regulator genes, *spalt* (*sal*), *cut* (*ct*), *empty spiracles* (*ems*) and *unpaired* (*upd*). Subsequently, their gene products (potentially in combination with Abd-B) control (directly or indirectly) the local expression of a battery of realisator genes, which confer unique properties to cells by influencing morphogenetic processes like cell adhesion, cell polarity or organization of the cytoskeleton. This ultimately leads to the proper morphogenesis of the posterior spiracle. Figure adapted from Lovegrove B, Somoes S, Rivas ML, Sotillos S, Johnson K et al. 2006 Curr Biol 16:2206–2216.

expressed outside their normal expression domains, which in turn results in homeotic transformations.[5]

Only very few *Hox* target genes initially identified coded for realisator genes, which was rather unexpected and suggested that *Hox* proteins exert their function primarily through the regulation of other high-executive genes. However, analysis of the few known realisator genes has been instrumental for understanding the morphogenetic function of *Hox* proteins.[52–56] For example, one of the best-studied *Hox* realisator genes in *Drosophila* is the apoptosis inducing gene *reaper* (*rpr*). *rpr* is expressed in a small number of cells in the anterior part of the maxillary segment in *Drosophila* embryos, and is directly controlled by the *Hox* protein Dfd[52] (Fig. 3B). In *Dfd* mutant embryos, *rpr* expression in the maxillary segment is abolished, which results in a loss of the boundary between the maxillary and mandibular segments (Fig. 3B). When *rpr* expression is restored in *Dfd* mutants, the segment boundary is maintained, showing that the Dfd-dependent activation of *rpr* and, consequently, the local activation of apoptosis is necessary and sufficient for the maintenance of the maxillary–mandibular segment boundary.[52] Looking at this example, it becomes clearer why the identification of realisator genes among the *Hox* targets has been so difficult. Realisator proteins are required for general functions (cell adhesion, cell proliferation, cell death etc.) in many cells at many different developmental stages. Consequently, mutations in realisator genes either result in early embryonic lethality or in pleiotropic effects. Alternatively, realisators very often act redundantly or have very subtle and context-dependent effects, like in the case of *rpr*. These complications make it difficult to correlate their mutant phenotypes to those found in *Hox* mutants. And here lies another problem: although mutations in *Hox* genes have been analyzed for decades, their phenotypic analysis is far from complete and many of the subtle morphological changes in *Hox* mutants may have gone unnoticed. Knowledge of these phenotypes, however, is a prerequisite to correlate *Hox*-dependent morphological output with the activity of downstream gene(s). Taken together, several lessons can be learned. First, although it was easier to identify and study transcription factors and signalling molecules as *Hox* target genes, per se these two classes of *Hox* target genes are not so informative in elucidating the role of *Hox* proteins in the specification of morphological properties on a cellular level (but more so in understanding the patterning properties of *Hox* proteins). Second, in order to gain an in-depth understanding of Hox-dependent morphogenesis, it is essential to study the function of *Hox* realisator genes irrespective of whether they are direct or indirect *Hox* targets. Third, since many realisator genes will be indirectly regulated by several executive *Hox* target genes, we need to elucidate

the nature of these Hox-modulated regulatory networks to understand how *Hox* genes control morphogenesis.

One such regulatory network, which is fairly well understood in *Drosophila*, is the posterior spiracle network (Fig. 3C). This network is activated in the abdominal segment 8 (A8) by *Abdominal-B* (*Abd-B*), the *Hox* gene specifying the morphology of the posterior region in *Drosophila* embryos.[57] Activation of the posterior spiracle network results in the formation of an ectodermal structure composed of the spiracular chamber, a tube connecting the trachea to the exterior, and the stigmatophore representing the external protrusion, in which the spiracular chamber is located.[57] The formation of these structures is dependent on the activation of four primary Abd-B target genes, the transcription factors *cut* (*ct*), *empty spiracles* (*ems*) and *spalt* (*sal*), and the signal transduction ligand of the JAK/STAT pathway *unpaired* (*upd*). The partially overlapping activity of the four primary Abd-B targets subsequently activates (potentially in combination with Abd-B) different sets of realisator genes in particular subsets of spiracle cells[58] (Fig. 3C). The targets include cell adhesion molecules, like E-cadherin or non-classical cadherins, like Cad86C or Cad74A, the cell-polarity protein Crumbs (Crb), and two regulators of the actin-cytoskeleton organization, the Rho GTPases RhoGAP88C and Gef64C[58] (Fig. 3C). Although not yet understood at the cellular level, it is thought that the region-specific expression of these and probably numerous other realisators confers unique morphogenetic properties to the cells, which ultimately lead to the formation of a segment-specific organ, the posterior spiracle.

The development of the posterior spiracle is one example for the complexity of Hox-modulated regulatory networks and illustrates that *Hox* proteins are able to regulate, directly and indirectly, many levels of such a network. Thus, in order to fully understand all functions of *Hox* proteins, it is necessary to elucidate these regulatory networks, which includes a complete knowledge of all direct and indirect *Hox* downstream genes.

Specificity in Hox target gene regulation: the Hox paradox

The functional analysis of direct *Hox* target genes, which inevitably includes the identification and characterization of regulatory sequences directly mediating the homeotic response, the so-called *Hox* response elements (HREs), has been extremely difficult. The limited success in identifying in vivo relevant HREs can be primarily attributed to an important intrinsic property of *Hox* proteins: a poor specificity in sequence recognition and binding exhibited by *Hox* proteins in vitro. Why is that? Why do *Hox* proteins (at least when present as monomers) recognize very similar and rather unspecific DNA sequences? The answer lies in the DNA-binding domain of *Hox* proteins, the homeodomain, which is, including its functionality, greatly conserved over large evolutionary distances.[59–61] This high conservation is reflected in an almost identical three-dimensional structure of the homeodomain in all *Hox* proteins studied so far.[60] As a consequence of this almost invariant molecular structure, the majority of Hox proteins, including the paralogues within one species, preferentially recognize a conserved, but fairly unspecific, -ATTA- core motif.[26] This low DNA-binding specificity, however, sharply contrasts with the highly specific effects *Hox* transcription factors exert on distinct and different sets of target genes in vivo.[27] Another dimension of *Hox* transcriptional specificity is reflected in their ability to act both as transcriptional repressors and activators and to regulate their target genes in highly specific spatial and temporal patterns in the animal, despite their rather large domains of expression.[27–29] This paradox of high in vivo (functional) but low in vitro (binding) specificity has raised the fundamental question: how do *Hox* proteins achieve regulation of selected target genes and what are the molecular mechanisms allowing *Hox* proteins to achieve their high developmental specificity in vivo?

The most-likely explanation is that other factors influence the functional specificity of *Hox* proteins. Initial support for this hypothesis came from studies that tested the effects of chimeric *Hox* proteins in vivo in order to identify functional domains within the *Hox* proteins.[15,62–65] All these studies suggested that multiple domains within any given *Hox* protein are essential for in vivo specificity. Based on these findings, the idea emerged that *Hox* proteins would heterodimerize with many other factors, so-called cofactors, which would subsequently enhance their sequence selectivity and binding specificity. This "cooperative/co-selective binding model" seemed the most-plausible mechanism used by *Hox* proteins (Fig. 4), since the cooperative binding of the yeast homeodomain transcription factors a1 and α2 had only been discovered a couple of years before.[66] Both transcription factors, a1 and α2, bind poorly to DNA alone, but specificity of DNA binding is greatly enhanced when both factors form a complex via protein–protein interactions. Therefore, it is not surprising that a hunt for factors began that were able to influence the binding behaviour of *Hox* proteins in a similar manner.[4,49,40,42,67] The resolution to the *Hox* pardox seemed close, when it was demonstrated that one candidate, the homeodomain protein Extradenticle (Exd), which was shown to directly interact with *Hox* proteins via a conserved YPWM hexapeptide motif,[68] was able to cooperatively bind DNA with many *Hox* proteins, thereby selectively increasing Hox-binding affinity on specific DNA sites.[69–71] In the following years, much research focused on Exd, not only because of its important role in *Hox* target gene regulation, but also because of major difficulties (despite large efforts) in identifying additional *Hox* cofactors of the Exd-type. The results showed that *Hox* target gene regulation is amazingly diverse and complex even when only a single *Hox* cofactor is considered. For example, the activity of Exd is regulated at the level of its

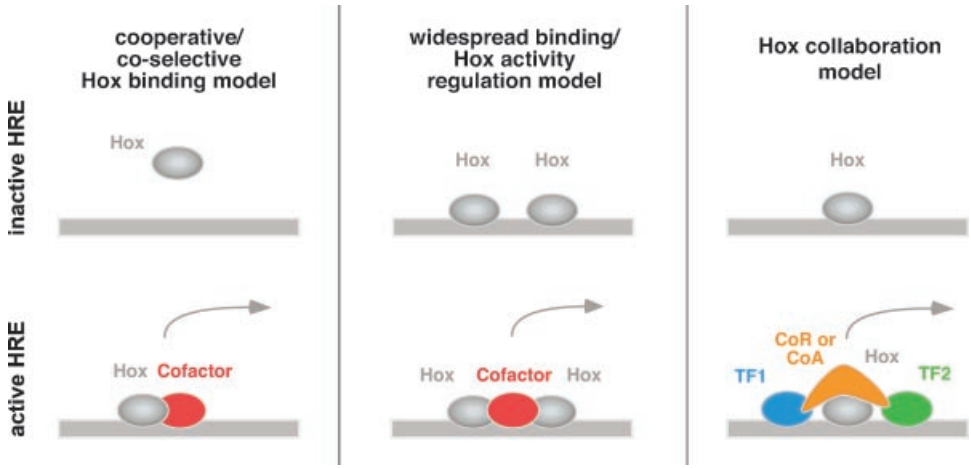

Figure 4. Three models of *Hox* target gene regulation in vivo. In the cooperative/co-selective binding model, *Hox* proteins cannot significantly occupy any HRE without the aid of cooperative interactions with cofactors (like Exd). Only in those cells co-expressing the *Hox* protein and the cofactor specific interactions occur, allowing *Hox* proteins to recognize and occupy individual HREs with increased binding specificity, and thus enabling them to activate or repress target genes accordingly. In the widespread binding/activity regulation model, *Hox* proteins bind many HREs in the absence of cofactors, but are unable to change target gene expression under these conditions. In the presence of cofactors, the activity of *Hox* proteins is changed, thus allowing them to activate or repress specific target genes. In the collaboration model, *Hox* proteins and many other context-specific transcription factors independently bind to HREs. The sign of the transcriptional output is determined by the combination of bound factors and the recruitment of specific co-repressors and/or co-activators. TF, transcription factor; CoA, co-activator; CoR, co-repressor.

subcellular localization, since Exd requires direct interaction with another homeodomain protein encoded by the gene *homothorax* (*hth*) for its nuclear translocation.[72] Once in the nucleus, Exd acts as a *Hox* cofactor by several means.

Detailed analysis of several *Hox* response elements led to the conclusion that Exd, as predicted by the "cooperative/co-selective binding model"[25] (Fig. 4), helps *Hox* proteins achieve DNA-binding specificity.[69,73–75] A couple of studies have meanwhile underlined the significance of this model in vivo.[76–78] For example, it has been shown that a 37 bp HRE from the *forkhead* (*fkh*) gene, a direct target of the *Hox* protein Scr, is not only cooperatively bound by Exd and Scr in vitro, but that activation of this element in vivo requires both genes, *Scr* and *exd*, and that other *Hox*–Exd heterodimers do not exert these specific in vitro and in vivo effects.[78] When two base pairs within this element were mutated, the element was bound by different *Hox*/Exd heterodimers with almost the same affinities in vitro and was specifically regulated by different *Hox* proteins in vivo. Structural analysis of both *Hox*–Exd–DNA ternary complexes, containing either the natural occurring *Hox*–Exd consensus sites or the mutated versions, now elegantly revealed a potential mechanism used by *Hox* proteins to select specific binding sequences in vivo: it was well established that *Hox* proteins recognize generic Hox-binding sites through major groove-recognition helix interactions,[60,61] but the structural analysis of both Scr–Exd–DNA complexes showed that selection among sites is critically dependent on minor groove interactions determined by two positively charged amino acid residues located in the N-terminal arm and linker region of the *Hox* protein Scr,[79] These residues, which are only correctly positioned through interaction with Exd, recognize the structure of the minor groove in a sequence-specific fashion.[79] Interestingly, the Scr-dependent regulation of *fkh* highlights another level of complexity in *Hox* target gene regulation: Scr and Exd are able to regulate the *fkh* HRE only during early stages of embryogenesis, since Scr negatively regulates *hth* expression and thus nuclear translocation of Exd later in development.[78] Thus, *Hox* proteins are also able to regulate their own activity (on specific enhancers) by regulating the availability of their cofactors. Finally, Hth does not only promote Exd's nuclear translocation, but it also acts as a *Hox* cofactor itself by increasing the DNA-binding specificities and/or affinities of *Hox*–Exd complexes through direct protein–protein interaction.[40,41,70,77]

Alternative to the assembly of large protein complexes on HREs, recent findings on *Hox*–Exd interactions indicate that only two protein partners, such as Exd and *Hox* proteins, are sufficient to generate specificity in target sequence recognition simply by interacting via different protein domains. Although *Hox*–Exd interactions for a long time were thought to be mediated primarily by the short hexapeptide N-terminal to the homeodomain.[68,80,81] newer evidence suggests that the physical interactions of both proteins are more complex than anticipated. For example, it has been shown for the *Hox* protein Ubx that the unrelated UbdA motif, located C-terminal

to the homeodoamin, mediates Exd recruitment, and that this interaction is essential for repression of the Ubx target gene *Dll in vivo*.[82] In contrast, it is well established that Ubx–Exd interaction via the hexapeptide motif is important for the specification of other Ubx-dependent segmental features.[14,83] Thus, specificity in *Hox* target gene recognition is (at least in some cases) achieved by at least two distinct interactions of *Hox* proteins with a single cofactor, Exd. This presumably allows the Hox–Exd complex to adopt different conformations, which can in turn recognize different target sequences. An open question in this context remains how the different Hox–Exd interactions are regulated in vivo. In addition, other *Hox* cofactors of the Exd type have been identified in recent years, including Teashirt and Disconnected.[67,84] However, the cooperative interaction of these factors with *Hox* proteins and the selective enhancement of Hox-binding specificities have remained unclear.

While work on the ''cooperative/co-selective binding model'' focused on the regulation of Hox–DNA interactions, much less progress has been made to elucidate how the transcriptional activity of *Hox* proteins bound to DNA is modulated in vivo. According to the ''widespread binding/activity regulation model'' cofactors, such as Exd, function to convert *Hox* proteins, which are bound to a very large number of Hox-binding sites in vivo, from a neutral to an active state capable of transcriptional activation or repression[25] (Fig. 4). This ''widespread binding/activity regulation model'' has three major implications: first, *Hox* proteins should be able to bind many genes in the genome, which has been confirmed for homeodomain-containing transcription factors by in vivo cross-linking experiments.[85] Second, *Hox* proteins should be able to bind to DNA independently of Exd. And consistent with this assumption, some naturally occurring Hox-dependent enhancers contain functionally important high-affinity Hox-binding sites that are not closely juxtaposed to high-affinity Exd sites.[50,86,87] In addition, it has only been shown recently that, for some enhancers, even binding of a Hox–Exd complex alone is not sufficient for target gene regulation.[40] Third, the ''widespread binding/activity regulation model'' implies that Exd should be able to switch *Hox* proteins into both transcriptional activators and repressors. Although Exd is able to change a *Hox* protein from a transcriptional repressor into a transcriptional activator (at least in one case), probably by masking a repressor domain contained in some *Hox* proteins,[88] a switch from activator into repressor has, to our knowledge, never been shown. Thus it has been postulated that the sign of transcriptional effect is not primarily determined by the Hox–Exd interaction, but due to the recruitment of additional factors into the Hox–cofactor complex. Interestingly, two such factors have been identified recently: in addition to the binding of a Hox–Exd–Hth complex, which is itself not sufficient for target gene regulation, two segmentation proteins, Engrailed (En) and Sloppy paired 1 (Slp1), and their sequence-specific recruitment have been shown to be required for the repression of the *Hox* target gene *Dll*.[40] Since En and Slp1 harbour motifs for interaction with the co-repressor Groucho, it is assumed that both proteins (through the recruitment of Groucho) act as intermediate regulatory molecules determining the sign of the transcriptional *Hox* output (in this case, repression of *Dll* transcription). In this scenario, the Hox–Exd–Hth complex serves to select the correct Hox-binding site(s) without regulating target gene expression, whereas the regulatory activity of the *Hox* protein is dependent on the surrounding binding sites and the activity of the factors interacting with these sites.

The finding that transcription factors with very-well-known functions in development (in the case of En and Slp1, the generation of anterior and posterior compartments in segments) work directly with *Hox* proteins in regulating their target genes was not completely unanticipated, since it had been realised before that other transcription factors and conserved sequences surrounding Hox-binding sites are important for Hox-dependent transcriptional control.[28,50,89,90] However, it seems that, in these days, the time was not right for the idea that *Hox* protein function and activity does not only dependent on specific Hox-cofactor interactions, but also (and perhaps more so) on the combinatorial interaction with other transcription factors, allowing *Hox* proteins to mediate context-specific activation or repression of target genes (''*Hox* collaboration model'') (Fig. 4). Evidence for the general importance of this new concept was provided only recently, when it was shown that the *Hox* protein Ubx collaborates with two transcription factors downstream of the Dpp/TGF-β pathway, Mothers against Dpp (Mad) and Medea (Med), to repress the *Hox* target gene spalt major (*sal*) in the haltere.[42] In addition, this study showed that cooperative interaction of *Hox* proteins with other regulatory factors is not required to modulate *Hox* target gene selection. On the contrary, the repression of *sal*, which does not require Exd and Hth activity, is mediated by the independent binding of Ubx and the collaborating transcription factors Mad and Med to the *sal* HRE.[42] And again, as in the case of En and Slp1, Mad and Med do not themselves function as transcriptional repressors, but they determine the regulatory activity of the *Hox* protein through recruitment of the co-repressor Schnurri (Shn), which was shown to be necessary for *sal* repression.[42] Taken together, these findings have revolutionized our picture of *Hox* target gene regulation: previously, much attention has focused on cofactors of the Exd- and Hth-type and their control of *Hox* DNA-binding selectivity via cooperative interactions with *Hox* proteins. However, *Hox* proteins (positively and negatively) regulate in a context-dependent manner a large diversity of target genes that are also regulated by other transcription factors. Thus, it has been argued that it would be too great a constraint to require that *Hox* proteins physically interact with

the large and diverse repertoire of transcription factors with which they act. In the light of recent findings, it seems more plausible that, even in the absence of any direct physical interaction, *Hox* proteins work together with many other transcription factors in a combinatorial fashion through a selective recruitment of all regulatory factors to target-specific and nearby binding sites in HREs, a transcriptional control mechanism meanwhile termed collaboration.[42] Since all of the collaborating transcription factors identified so far (En, Slp1, Mad and Med) harbour motifs for interactions with co-repressors/co-activators, it seems quite attractive to assume that the different transcriptional inputs from *Hox* proteins and collaborators are integrated and mediated to the transcription machinery via the recruitment and assembly of different co-repressor and/or co-activator complexes. In addition, the collaboration model offers a very simple, but nonetheless elegant explanation to the mystery how *Hox* proteins can act as repressors in one context and as activators in another: it seems very likely that *Hox* proteins primarily function as placeholders in HREs, but that the sign of *Hox* action (and thus the transcriptional output) is mainly dictated by the regulatory activity of all collaborating transcription factors assembled on these HREs. Finally, we would like to take the collaboration model to the next level: we propose that, in principle, every transcription factor could act as *Hox* collaborator. Since every cell has a unique combination of transcription factors, the combinatorial interactions for the broadly expressed *Hox* proteins would be almost limitless in such a scenario. This would allow for a very precise modulation and fine-tuning of *Hox* target gene regulation, even on the level of the individual cell, eventually leading to the amazing functional diversity that *Hox* proteins achieve in development and evolution. Thus, previously identified tissue-specific transcription factors or sequence elements shown to be necessary for the regulation of Hox target genes[28,67,84,89–91] could represent, in our view, additional collaborators of *Hox* proteins or their respective binding sites in HREs.

Taken together, many aspects of *Hox* function and *Hox* target gene regulation were understood in much detail in the pre-genomic age. However, the fact that only few *Hox* target genes were known severely limited our ability to assess the relative contribution of the various modes of *Hox* action during development of the entire organism. The focus on a few selected target genes and HREs implicates that we could be dealing with the exceptions rather than the most widely used mechanisms for *Hox* target gene regulation. In addition, the small number of known target genes made it impossible to infer how *Hox* proteins carry out their morphogenetic function in vivo. Therefore, to draw more general conclusions about *Hox* target recognition and regulation, as well as Hox target function, a genome-wide inventory of *Hox* targets and HREs is required.

Large-scale analysis of *Hox* downstream genes and regulation: the genomic era

With the advent of genome-wide approaches in the last decade, we are now in a position to overcome some of the limitations outlined above and characterize *Hox* downstream genes and HREs on a large-scale to more fully understand all aspects and mechanisms of *Hox* function. Again, we will focus our review on findings made in *Drosophila*.

Large-scale identification of Hox downstream genes: microarray expression profiling

With the introduction of DNA microarray technology, transcript profiling was used to systematically detect genes that showed differential expression in response to *Hox* proteins. Though this method is extremely powerful, additional methods are required to distinguish between direct and indirect targets. DNA microarray technology has been used quite extensively in vertebrates (summarized in Table 2), while only few groups have used it for the identification of *Hox* downstream genes in *Drosophila* (summarized in Table 2). Recent papers report genes regulated by the *Hox* genes *Dfd*, *Scr*, *Antp*, *Ubx*, *abd-A* and *Abd-B* in the embryo,[27] as well as downstream genes of Ubx in developing wing and haltere imaginal discs.[92,93] What can be learned from these studies? Which concepts established previously were confirmed, which ones have to be newly defined?

First of all, both embryo and imaginal studies show that *Hox* genes regulate a large number of downstream genes despite the fact that only very tightly defined developmental stages were analyzed (Table 2). Since *Hox* genes are required throughout development, it follows that they very likely regulate thousands of genes during a fly's life. Although at first glance this might be a surprising finding, it is not a new one. Already in 1998, Liang and colleagues[85] have characterized the expression of randomly selected genes at different stages of *Drosophila* embryogenesis, and their results suggested that selector homeoproteins, including *Hox* proteins, regulate the expression of most genes throughout development,[85] a view now supported by genomic data. Is this also true for other organisms? In vertebrates, various groups have used microarray expression profiling to identify downstream genes of different *Hox* proteins (Table 2). One major problem in higher organisms is the functional redundancy of *Hox* genes, thus it is almost impossible to study the effects of a single *Hox* gene in vivo. To circumvent this problem many groups have expressed individual *Hox* genes in cell cultures and assessed gene expression changes with high-density gene arrays.[35,94–97] However, as outlined above, *Hox* genes fulfil only a subset of their functions if taken out of context due to the lack of assisting transcription factors. This is also reflected in the outcome of the studies mentioned above, since they very often report low numbers of putative *Hox* downstream genes (Table 2). However, when such studies were performed in vivo, the

Table 2. Large-scale identification of *Hox* downstream genes and *Hox* response elements in the genomic age

References	*Hox* genes	Organism	Tissue	Stage	#Targets
Leemanns et al., 2001[117]	Lab	*Drosophila*	whole embryo	embryonic stage 10-17	96
Mohit et al., 2006[92]	Ubx	*Drosophila*	haltere and wing disc	3rd instar larvae	541
Hersh et al., 2007[93]	Ubx	*Drosophila*	haltere and wing disc	3rd instar larvae	447
Hueber et al., 2007[27]	Dfd, Scr, Antp, Ubx, Abd-A, Abd-B	*Drosophila*	whole embryo	embryonic stage 11+ 12	1508
Shen et al., 2000[97]	*Hox*A1	mouse	cell culture – teratocarcinoma		28
Zhao and Potter, 2001[120]	*Hox*A13	mouse	uterus and cervix tissue	4.5 weeks old	unclear
Valerius et al., 2002[119]	*Hox*A11	mouse	kidney tissue	embryonic stage 18.5	10
Hedlund et al., 2004[34]	*Hox*D10	mouse	spinal cord tissue	embryonic stage 12.5	69
Martinez-Ceballos et al., 2005[96]	*Hox*A1	mouse	cell culture – embryonic blastocysts		145
Lei et al., 2005[95]	*Hox*C8	mouse	cell culture – embryonic fribroblasts		34
Cobb et al., 2005[33]	*Hox*D cluster genes	mouse	mouse tissue of limbs and genitalia	embryonic stage 12.5	16
Williams et al., 2005[35]	*Hox*A13	mouse	cell culture – embryonic fribroblasts		68
Schwab et al., 2006[99]	*Hox*A11 + HoxD11	mouse	whole embryonic kidneys and urogenital tissue	embryonic stage 11.5, 12.5 13.5, 16.5 + adult	1518
Rohrschneider et al., 2007[98]	*Hox*B1a	zebrafish	whole embryo	19-20 hours post fertilization	471
Ferrell et al., 2005[94]	*Hox*A10	human	cell culture – umbilical cord cells		115

Large-scale identification of *Hox* response elements

References	*Hox* genes	Organism	Approach
Ebner et al., 2005[41]	Lab	*Drosophila*	*in silico* prediction
Hueber et al., 2007[27]	Dfd	*Drosophila*	*in silico* prediction
McCabe et al., 2005[101]	*Hox*A13 + HoxD13	mouse	ChIP

results were often similar to those obtained in *Drosophila*,[98,99] again highlighting the importance of in vivo studies.

Another important observation of the large-scale analyses in *Drosophila* is that *Hox* downstream genes can be found across diverse functional classes, ranging from regulatory molecules, like transcription factors and signalling components, to realisators. This notion is also mirrored by studies performed in vertebrates, which resulted in the identification of similar classes of *Hox* downstream genes. These findings contrast with the view, prevalent in the pre-genomic age, that *Hox* proteins primarily affect regulatory genes, especially transcription factors. While this concept was based on the knowledge of 24 *Hox* targets (Table 1), a recent study now identified thousands of *Hox* response genes and showed that 13% code for realisator proteins.[27] A similar result was also obtained by a study performed in vertebrates.[98] This result lends support for the concept postulated by Garcia-Bellido more than 30 years ago[48] and showed, for the first time, that *Hox* proteins regulate morphogenesis at least in part through the regulation of terminal differentiation genes.[27] Most of these realisator genes likely act redundantly in general cellular processes required in many cells, which probably has precluded their discovery by genetic approaches. This highlights one of the advantages of genomic approaches, namely the ability to identify targets irrespective of their molecular nature. Conversely, in many cases, it will be very difficult to elucidate the in vivo function of the identified realisators by reverse genetics.

Each *Hox* protein specifies distinct morphological features within segments, and understanding how this specificity is achieved has been one of the major goals since the discovery of *Hox* genes. So far, there has been only a single study in which the effects of different *Hox* proteins under the same experimental conditions were tested.[27] One of the major findings is that many of the identified *Hox* downstream genes are primarily affected by a single *Hox* protein, implying that there is tremendous specificity in target gene regulation, despite the similarities in in vitro DNA binding. Moreover, there was a clear trend for distinct regulatory interactions in those cases where downstream genes were regulated by more than one *Hox* protein. These genes were likely to be affected in a similar manner when targeted by *Hox* proteins specifying segments with similar morphologies, whereas they were more often regulated in the opposite direction when targeted by *Hox* proteins functioning in different body parts.[27] The authors therefore concluded that the diversification of segments is achieved through the regulation of unique downstream genes on the one hand and through the differential regulation of shared downstream genes on the other hand.

The issue of *Hox* protein specificity on the transcriptome was not only reflected in the identification of a large portion of unique *Hox* downstream genes, but also in the observation that the majority of genes are primarily regulated at only one of the two developmental stages analyzed. The authors suggested that this might be achieved by an extensive interaction of *Hox* proteins with the regulatory environment that they are embedded in. Support for this notion not only comes from previous studies showing that regulation of *Hox* target genes is dependent on the context,[28,29,91,100] but more so from recent studies, which have provided direct evidence that *Hox* proteins gain the ability to regulate their target genes in a context-specific manner in vivo by interaction with known cell- and/or tissue-specific transcription factors, so-called collaborators.[40,42]

Taken together, transcriptomic approaches so far have been very informative about the number and nature of *Hox* response genes, but many questions about the mechanisms of regulatory interactions as well as the function of downstream genes remain open.

Large-scale identification of Hox response elements: in silico and in vivo approaches

An essential aspect for our mechanistic understanding of *Hox*-dependent processes is the identification of direct targets versus downstream genes that are controlled via intermediate factors. The transcriptome datasets described thus far include both direct and indirect targets and so far there are only three published studies, which aim to identify *Hox* response elements on a genome-wide scale.[27,41,101] In principle, two complementing strategies have been used: in silico by searching for genomic regions that are enriched in transcription-factor-binding sites using computational tools,[27,41] and in vivo by identifying DNA fragments associated with transcription factors using chromatin immunoprecipitation (ChIP).[101] The bioinformatics detection of *Hox* response elements is hampered by the fact that individual *Hox* proteins have rather poorly defined binding sequences, which in addition occur very frequently in the genome. However, two distinct computational approaches have been applied to the identification of HREs in *Drosophila*. To enhance the stringency of the search criteria, the first study was based the observation that *Hox* proteins can bind to their target sequences in association with the cofactors Exd and Hth.[102] In this scenario, distinct *Hox*–Exd–Hth complexes recognize and select specific DNA sequences depending on the *Hox* protein included in the complex. Based on this model, Ebner and colleagues[41] searched the *D. melanogaster* genome for Lab–Exd heterodimer-binding sequences within 40 base pairs of an Hth consensus site.[41] Although they identified 40 putative target sequences for the Lab–Exd–Hth complex, only a single gene (*CG11339)* in the vicinity of the identified binding sites showed a Lab-like expression pattern. However, when the predicted Lab response element was tested in vivo, it did not show the expected enhancer activity. Interestingly, another DNA fragment nearby the *CG11339* transcription unit, which was not predicted to be bound by Lab, was able to drive reporter gene expression in Lab-expressing cells, despite the fact that a putative Lab-binding site within this enhancer was highly divergent from the consensus binding sequence. While the results of Ebner and colleagues[41] might not seem very encouraging with regards to the reliability of computational identification of HREs, they highlight two important issues for in silico strategies: First, in vivo *Hox* binding sites might be more divergent than anticipated and therefore, the stringency of the initial motif search might have been to high. And second, the cooperative binding of *Hox* proteins with dedicated cofactors of the Exd-type might not be of such a general importance for *Hox* target gene regulation as previously anticipated. Thus, the findings of Ebner and colleagues[41] should, in our view, not be considered as a failure of in silico approaches to predict HREs. There are meanwhile many examples of the successful in silico prediction of regulatory elements.[103–107] On the contrary, it could be very well argued that the findings of Ebner and colleagues[41] underline the peculiarity of the cooperative/co-selective binding model to explain the *Hox* paradox and support the possibility that collaboration might be indeed the more general mode of *Hox* proteins to select and regulate their target genes. And to be even a little provocative: Exd might assist *Hox* proteins in target gene regulation not only by increasing the DNA-binding specificity of *Hox* proteins through cooperate binding, but perhaps more so by acting, like other *Hox* collaborators, in a combinatorial fashion with *Hox* proteins. In this context, the outcome of an improved in silico search using less stringent Lab- and Exd-binding sequences and a more relaxed spacing of these sites would be very interesting.

In recent years, it has been realised that a combination of in silico prediction and in vivo approaches has a higher success rate in identifying in vivo functional regulatory elements than approaches only based on computational calculations.[108–110] This has been successfully integrated in the strategy used by Hueber and colleagues.[27] After a detailed analysis of all known HREs, which also included enhancers regulated independently of the Exd/Hth input, the main requirement for their *in silico* search was an accumulation of Hox-binding sites within a limited stretch of DNA sequence. In addition, for a *Hox* response element to be considered, it had to pass a sequence conservation filter across four *Drosophila* species. And finally, in contrast to the approach of Ebner and colleagues,[41] the parameters of the search were optimized and validated using in vivo results from transcript profiling experiments. By applying this approach, Hueber and colleagues[27] were able to identify a large number of putative response elements for the *Hox* protein Dfd. Two findings indicate that these may constitute true target sequences for

Dfd: first, all of the predicted Dfd response elements contained several conserved binding sites for other transcription factors, a known prerequisite for functional enhancer elements.[107] In the light of the collaboration model, these sites could represent interaction sites for collaborating transcription factors, which needs further experimental proof. Second, many elements have been tested experimentally by in vitro analysis and in all cases showed that they were specifically bound by Dfd, whereas Ubx, a *Hox* protein specifying trunk identity, was not able to interact with these enhancers. Meanwhile, some of the enhancers were also tested in the embryo, showing that all of them function in a Hox-dependent manner in vivo (Bezdan D, Schäder N, Piediotta M, Hent S, Lohmonn I, unpublished data. Thus, less-stringent sequence requirements and the incorporation of in vivo data seem to increase the power of computational methods for predicting HREs.

The only study using an in vivo approach for the large-scale identification of *Hox* response elements was performed by McCabe and Innis[101] in 2005. Here, genomic fragments bound by the *Hox* protein *HOX*A13, misexpressed in embryonic fibroblast cells, where isolated using ChIP. DNA fragments were eluted, cloned and 5% of the clones were sequenced.[101] To verify the identified fragments, the authors analyzed expression changes of nearby genes in response to *HOXA*13 activity and studied putative enhancers in reporter gene assays. Only seven new high-confidence HREs passed all requirements, which might be explained by the fact that this analysis was performed in tissue culture and thus in an environment deprived of other factors assisting *Hox* proteins in target gene regulation.

Taken together, it is obvious that further experimentation is required to identify HREs and thus the genes directly regulated by *Hox* proteins on a genome-wide scale. Due to the limited amount of data and in particular the lack of in vivo analysis of the identified enhancers, it is so far impossible to draw general conclusions on the mechanisms of how *Hox* proteins achieve specificity in target gene regulation. One of the major challenges on our way to deducing more general rules for *Hox*–DNA interactions will be the integration of data derived from diverse experiments, including more traditional gene by gene methods, to cover all aspects of *Hox* protein activity.

Conclusions and future directions

Although large-scale approaches are extremely important for a comprehensive understanding of all aspects of *Hox* gene function, only few such studies exist despite the widespread availability of many useful tools, such as expression profiling, or ChIP-on-chip experiments. Thus, the field so far suffers from a limitation of available data, which together with the enormous complexity of *Hox* protein activity makes it difficult to reconstruct the regulatory networks orchestrated by them. In addition, *Hox* output largely depends on the regulatory context within every cell, in the sense that other transcription factors dictate the "When", "Where" and "How" of *Hox* target gene regulation. Consequently, *Hox*-modulated regulatory networks will change dramatically during development. Thus, in order to understand those networks and their dynamic behaviour, we suggest a more-refined large-scale identification of direct and indirect *Hox* target genes and of active HREs in consecutive developmental stages using genome-wide approaches, like microarray expression profiling and ChIP-on-chip or ChIP-Seq experiments. In addition, data from other resources (like ArrayExpress, modENCODE, BDGP in situ database) need to be incorporated, since it is now realised that many other transcription factors will assist *Hox* proteins in target gene regulation. With new large-scale datasets and innovative mechanistic studies becoming available, the new millenium certainly is an exciting time for the *Hox* field. If we succeed in integrating results generated by large-scale in vivo, in vitro and in silico strategies, we might be able to decipher the mysteries of *Hox* activity that have caught the imagination of developmental biologists ever since the discovery of homeotic transformation by Bateson in 1894.

Acknowledgments

We thank Jan U. Lohmann, Daniela Bezdan, Aurelia L. Fuchs, Petra Stöbe and Zongzhao Zhai for critically reading the manuscript and helpful comments.

References

1. Duboule D, Dolle P. 1989. The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. EMBO J 8:1497–1505.
2. Graham A, Papalopulu N, Krumlauf R. 1989. The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell 57:367–378.
3. Pearson JC, Lemons D, McGinnis W. 2005. Modulating *Hox* gene functions during animal body patterning. Nature Genetics Reviews 6:893–904.
4. Mann RS, Morata G. 2000. The developmental and molecular biology of genes that subdivide the body of Drosophila. Annu Rev Cell Dev Biol 16:243–271.
5. McGinnis W, Krumlauf R. 1992. Homeobox genes and axial patterning. Cell 68:283–302.
6. Lewis EB. 1978. A gene complex controlling segmentation in Drosophila. Nature 276:565–570.
7. Kaufman TC, Seeger MA, Olsen G. 1990. Molecular and genetic organization of the Antennapedia gene complex of Drosophila melanogaster. Adv Genet 27:309–362.
8. Bateson W. 1894. Materials for the study of variation: treated with special regard to discontinuity in the origin of species. Cambridge: Cambridge Press.
9. Beeman RW, Stuart JJ, Haas MS, Denell RE. 1989. Genetic analysis of the homeotic gene complex (HOM-C) in the beetle Tribolium castaneum. Dev Biol 133:196–209.
10. Krumlauf R. 1994. *Hox* genes in vertebrate development. Cell 78:191–201.
11. Kmita M, Tarchini B, Zakany J, Logan M, Tabin CJ, et al. 2005. Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. Nature 435:1113–1116.
12. Hafen E, Levine M, Gehring WJ. 1984. Regulation of Antennapedia transcript distribution by the Bithorax complex in Drosophila. Nature 307:287–289.

13. Struhl G. 1982. Genes controlling segmental specification in the Drosophila thorax. Proc Natl Acad Sci USA 79:7380–7384.
14. Gonzalez-Reyes A, Morata G. 1990. The developmental effect of overexpressing a Ubx product in Drosophila embryos is dependent on its interactions with other homeotic products. Cell 61: 515–522.
15. Mann RS, Hogness DS. 1990. Functional dissection of Ultrabithorax proteins in D. melanogaster. Cell 60:597–610.
16. Regulski M, McGinnis N, Chadwick R, McGinnis W. 1987. Developmental and molecular analysis of Deformed: a homeotic gene controlling Drosophila head development. EMBO J 6:767–777.
17. Scott MP, Weiner AJ. 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proc Natl Acad Sci USA 81:4115–4119.
18. McGinnis W, Levine MS, Hafen E, Kuroiwa A, Gehring WJ. 1984. A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature 308:428–433.
19. Scott MP, Tamkun JW, Hartzell GW 3rd. 1989. The structure and function of the homeodomain. Biochim Biophys Acta 989:25–48.
20. Krasnow MA, Saffman EE, Kornfeld K, Hogness DS. 1989. Transcriptional activation and repression by Ultrabithorax proteins in cultured Drosophila cells. Cell 57:1031–1043.
21. Desplan C, Theis J, O'Farrell PH. 1988. The sequence specificity of homeodomain-DNA interaction. Cell 54:1081–190.
22. Carr A, Biggin MD. 1999. A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos. EMBO J 18:1598–1608.
23. Walter J, Dever CA, Biggin MD. 1994. Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in Drosophila embryos. Genes Dev 8:1678–1692.
24. Li X, Murre C, McGinnis W. 1999. Activity regulation of a *Hox* protein and a role for the homeodomain in inhibiting transcriptional activation. EMBO J 18:198–211.
25. Biggin MD, McGinnis W. 1997. Regulation of segmentation and segmental identity by Drosophila homeoproteins: the role of DNA binding in functional activity and specificity. Development 124:4425–4433.
26. Ekker SC, Jackson DG, von Kessler DP, Sun BI, Young KE, et al. 1994. The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. EMBO J 13:3551–3560.
27. Hueber SD, Bezdan D, Henz SR, Blank M, Wu H, et al. 2007. Comparative analysis of *Hox* downstream genes in Drosophila. Development 134:381–392.
28. Grienenberger A, Merabet S, Manak J, Iltis I, Fabre A, et al. 2003. TGF-? signaling acts on a *Hox* response element to confer specificity and diversity to *Hox* protein function. Development 130:5445–5455.
29. Bondos SE, Tan XX. 2001. Combinatorial transcriptional regulation: the interaction of transcription factors and cell signaling molecules with homeodomain proteins in Drosophila development. Crit Rev Eukaryot Gene Expr 11:145–171.
30. Graba Y, Aragnol D, Pradel J. 1997. Drosophila *Hox* complex downstream targets and the function of homeotic genes. Bioessays 19:379–388.
31. Hombria JC, Lovegrove B. 2003. Beyond homeosis–HOX function in morphogenesis and organogenesis. Differentiation 71:461–476.
32. Pradel J, White RA. 1998. From selectors to realizators. Int J Dev Biol 42:417–421.
33. Cobb J, Duboule D. 2005. Comparative analysis of genes downstream of the Hoxd cluster in developing digits and external genitalia. Development 132:3055–3067.
34. Hedlund E, Karsten SL, Kudo L, Geschwind DH, Carpenter EM. 2004. Identification of a Hoxd10-regulated transcriptional network and combinatorial interactions with Hoxa10 during spinal cord development. J Neurosci Res 75:307–319.
35. Williams TM, Williams ME, Kuick R, Misek D, McDonagh K, et al. 2005. Candidate downstream regulated genes of HOX group 13 transcription factors with and without monomeric DNA binding capability. Dev Biol 279:462–480.
36. Mastick GS, McKay R, Oligino T, Donovan K, Lopez AJ. 1995. Identification of target genes regulated by homeotic proteins in Drosophila melanogaster through genetic selection of Ultrabithorax protein-binding sites in yeast. Genetics 139:349–363.
37. Mann RS. 1995. The specificity of homeotic gene function. Bioessays 17:855–863.
38. Mann RS, Affolter M. 1998. *Hox* proteins meet more partners. Curr Opin Genet Dev 8:423–429.
39. Mahaffey JW. 2005. Assisting *Hox* proteins in controlling body form: are there new lessons from flies (and mammals)? Curr Opin Genet Dev 15:422–429.
40. Gebelein B, McKay DJ, Mann RS. 2004. Direct integration of *Hox* and segmentation gene inputs during Drosophila development. Nature 431:653–659.
41. Ebner A, Cabernard C, Affolter M, Merabet S. 2005. Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex. Development 132:1591–1600.
42. Walsh CM, Carroll SB. 2007. Collaboration between Smads and a *Hox* protein in target gene repression. Development 134:3585–3592.
43. Manak JR, Mathies LD, Scott MP. 1994. Regulation of a decapentaplegic midgut enhancer by homeotic proteins. Development 120:3605–3619.
44. Vachon G, Cohen B, Pfeifle C, McGuffin ME, Botas J, et al. 1992. Homeotic genes of the Bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distalless. Cell 71:437–450.
45. Graba Y, Aragnol D, Laurenti P, Garzino V, Charmot D, et al. 1992. Homeotic control in Drosophila; the scabrous gene is an in vivo target of Ultrabithorax proteins. EMBO J 11:3375–3384.
46. Strutt DI, White RA. 1994. Characterisation of T48, a target of homeotic gene regulation in Drosophila embryogenesis. Mech Dev 46:27–39.
47. Heuer JG, Li K, Kaufman TC. 1995. The Drosophila homeotic target gene centrosomin (cnn) encodes a novel centrosomal protein with leucine zippers and maps to a genomic region required for midgut morphogenesis. Development 121:3861–3876.
48. Garcia-Bellido A. 1975. Genetic control of wing disc development in Drosophila. Ciba Found Symp 0:161–182.
49. Percival-Smith A, Teft WA, Barta JL. 2005. Tarsus determination in Drosophila melanogaster. Genome 48:712–721.
50. Zeng C, Pinsonneault J, Gellon G, McGinnis N, McGinnis W. 1994. Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in Drosophila embryos. EMBO J 13:2362–2377.
51. Appel B, Sakonju S. 1993. Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in Drosophila embryos. EMBO J 12:1099–1109.
52. Lohmann I, McGinnis N, Bodmer M, McGinnis W. 2002. The Drosophila *Hox* gene Deformed sculpts head morphology via direct regulation of the apoptosis activator reaper. Cell 110:457–466.
53. Hinz U, Wolk A, Renkawitz-Pohl R. 1992. Ultrabithorax is a regulator of b3 tubulin expression in the Drosophila visceral mesoderm. Development 116:543–554.
54. Gould AP, White RA. 1992. Connectin, a target of homeotic gene control in Drosophila. Development 116:1163–1174.
55. Li K, Kaufman TC. 1996. The homeotic target gene centrosomin encodes an essential centrosomal component. Cell 85:585–596.
56. Bello BC, Hirth F, Gould AP. 2003. A pulse of the Drosophila *Hox* protein Abdominal-A schedules the end of neural proliferation via neuroblast apoptosis. Neuron 37:209–219.
57. Hu N, Castelli-Gair J. 1999. Study of the posterior spiracles of Drosophila as a model to understand the genetic and cellular mechanisms controlling morphogenesis. Dev Biol 214:197–210.
58. Lovegrove B, Somoes S, Rivas ML, Sotillos S, Johnson K, et al. 2006. Co-ordinated control of cell adhesion, cell polarity and cytoskeleton underlies *Hox* induced organogenesis in Drosophila. Curr Biol 16:2206–2216.
59. Malicki J, Schughart K, McGinnis W. 1990. Mouse Hox-2.2 specifies thoracic segmental identity in Drosophila embryos and larvae. Cell 63:961–967.

60. Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, et al. 1994. Homeodomain-DNA recognition. Cell 78:211–223.
61. McGinnis N, Kuziora MA, McGinnis W. 1990. Human Hox-4.2 and Drosophila Deformed encode similar regulatory specificities in Drosophila embryos and larvae. Cell 63:969–976.
62. Kuziora MA, McGinnis W. 1989. A homeodomain substitution changes the regulatory specificity of the Deformed protein in Drosophila embryos. Cell 59:563–571.
63. Lin L, McGinnis W. 1992. Mapping functional specificity in the Dfd and Ubx homeo domains. Genes Dev 6:1071–1081.
64. Gibson G, Schier A, LeMotte P, Gehring WJ. 1990. The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. Cell 62:1087–1103.
65. Zeng W, Andrew DJ, Mathies LD, Horner MA, Scott MP. 1993. Ectopic expression and function of the Antp and Scr homeotic genes: the N terminus of the homeodomain is critical to functional specificity. Development 118:339–352.
66. Keleher CA, Goutte C, Johnson AD. 1988. The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. Cell 53:927–936.
67. Taghli-Lamallem O, Gallet A, Leroy F, Malapert P, Vola C, et al. 2007. Direct interaction between Teashirt and Sex combs reduced proteins, via Tsh's acidic domain, is essential for specifying the identity of the prothorax in Drosophila. Dev Biol 307:142–151.
68. Johnson FB, Parker E, Krasnow MA. 1995. Extradenticle protein is a selective cofactor for the Drosophila homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. Proc Natl Acad Sci USA 92:739–743.
69. Chan SK, Jaffe L, Capovilla M, Botas J, Mann RS. 1994. The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with Extradenticle, another homeoprotein. Cell 78:603–615.
70. Popperl H, Bienz M, Studer M, Chan SK, Aparicio S, et al. 1995. Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon Exd/Pbx. Cell 81:1031–1042.
71. van Dijk MA, Murre C. 1994. Extradenticle raises the DNA binding specificity of homeotic selector gene products. Cell 78:617–624.
72. Rieckhof GE, Casares F, Ryoo HD, Abu-Shaar M, Mann RS. 1997. Nuclear translocation of Extradenticle requires Homothorax, which encodes an Extradenticle-related homeodomain protein. Cell 91:171–183.
73. Chan SK, Mann RS. 1996. A structural model for a homeotic protein-extradenticle-DNA complex accounts for the choice of HOX protein in the heterodimer. Proc Natl Acad Sci USA 93:5223–5228.
74. Phelan ML, Featherstone MS. 1997. Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX.PBX heterodimers. J Biol Chem 272:8635–8643.
75. Shen WF, Chang CP, Rozenfeld S, Sauvageau G, Humphries RK, et al. 1996. *Hox* homeodomain proteins exhibit selective complex stabilities with Pbx and DNA. Nucleic Acids Res 24:898–906.
76. Chan SK, Ryoo HD, Gould A, Krumlauf R, Mann RS. 1997. Switching the in vivo specificity of a minimal Hox-responsive element. Development 124:2007–2014.
77. Gebelein B, Culi J, Ryoo HD, Zhang W, Mann RS. 2002. Specificity of Distalless repression and limb primordia development by abdominal *Hox* proteins. Dev Cell 3:487–498.
78. Ryoo HD, Mann RS. 1999. The control of trunk *Hox* specificity and activity by Extradenticle. Genes Dev 13:1704–1716.
79. Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, et al. 2007. Functional specificity of a *Hox* protein mediated by the recognition of minor groove structure. Cell 131:530–543.
80. Knoepfler PS, Kamps MP. 1995. The pentapeptide motif of *Hox* proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. Mol Cell Biol 15:5811–5819.
81. Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK. 1999. Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature 397:714–719.
82. Merabet S, Saadaoui M, Sambrani N, Hudry B, Pradel J, et al. 2007. A unique Extradenticle recruitment mode in the Drosophila *Hox* protein Ultrabithorax. Proc Natl Acad Sci USA 104:16946–16951.
83. Galant R, Walsh CM, Carroll SB. 2002. *Hox* repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. Development 129:3115–3126.
84. Robertson LK, Bowling DB, Mahaffey JP, Imiolczyk B, Mahaffey JW. 2004. An interactive network of zinc-finger proteins contributes to regionalization of the Drosophila embryo and establishes the domains of HOM-C protein function. Development 131:2781–2789.
85. Liang Z, Biggin MD. 1998. Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in Drosophila. Development 125:4471–4482.
86. Capovilla M, Brandt M, Botas J. 1994. Direct regulation of decapentaplegic by Ultrabithorax and its role in Drosophila midgut morphogenesis. Cell 76:461–475.
87. Regulski M, Dessain S, McGinnis N, McGinnis W. 1991. High-affinity binding sites for the Deformed protein are required for the function of an autoregulatory enhancer of the Deformed gene. Genes Dev 5:278–286.
88. Pinsonneault J, Florence B, Vaessin H, McGinnis W. 1997. A model for extradenticle function as a switch that changes HOX proteins from repressors to activators. EMBO J 16:2032–2042.
89. Marty T, Vigano MA, Ribeiro C, Nussbaumer U, Grieder NC, et al. 2001. A HOX complex, a repressor element and a 50 bp sequence confer regional specificity to a DPP-responsive enhancer. Development 128:2833–2845.
90. McCormick A, Core N, Kerridge S, Scott MP. 1995. Homeotic response elements are tightly linked to tissue-specific elements in a transcriptional enhancer of the teashirt gene. Development 121:2799–2812.
91. Zaffran S, Kuchler A, Lee HH, Frasch M. 2001. Biniou (FoxF), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in Drosophila. Genes Dev 15:2900–2915.
92. Mohit P, Makhijani K, Madhavi MB, Bharathi V, Lal A, et al. 2006. Modulation of AP and DV signaling pathways by the homeotic gene Ultrabithorax during haltere development in Drosophila. Dev Biol 291:356–367.
93. Hersh BM, Nelson CE, Stoll SJ, Norton JE, Albert TJ, et al. 2007. The UBX-regulated network in the haltere imaginal disc of D. melanogaster. Dev Biol 302:717–727.
94. Ferrell CM, Dorsam ST, Ohta H, Humphries RK, Derynck MK, et al. 2005. Activation of stem-cell specific genes by HOXA9 and HOXA10 homeodomain proteins in CD34+ human cord blood cells. Stem Cells 23:644–655.
95. Lei H, Wang H, Juan AH, Ruddle FH. 2005. The identification of Hoxc8 target genes. Proc Natl Acad Sci USA 102:2420–2424.
96. Martinez-Ceballos E, Chambon P, Gudas LJ. 2005. Differences in gene expression between wild type and Hoxa1 knockout embryonic stem cells after retinoic acid treatment or leukemia inhibitory factor (LIF) removal. J Biol Chem 280:16484–16498.
97. Shen J, Wu H, Gudas LJ. 2000. Molecular cloning and analysis of a group of genes differentially expressed in cells which overexpress the Hoxa-1 homeobox gene. Exp Cell Res 259:274–283.
98. Rohrschneider MR, Elsen GE, Prince VE. 2007. Zebrafish Hoxb1a regulates multiple downstream genes including prickle1b. Dev Biol 309:358–372.
99. Schwab K, Hartman HA, Liang HC, Aronow BJ, Patterson LT, et al. 2006. Comprehensive microarray analysis of Hoxa11/Hoxd11 mutant kidney development. Dev Biol 293:540–554.
100. Bondos SE, Tan XX, Matthews KS. 2006. Physical and genetic interactions link *Hox* function with diverse transcription factors and cell signaling proteins. Mol Cell Proteomics 5:824–834.
101. McCabe CD, Innis JW. 2005. A genomic approach to the identification and characterization of HOXA13 functional binding elements. Nucleic Acids Res 33:6782–6794.
102. Ryoo HD, Marty T, Casares F, Affolter M, Mann RS. 1999. Regulation of *Hox* target genes by a DNA bound Homothorax/Hox/Extradenticle complex. Development 126:5137–5148.
103. Rajewsky N, Vergassola M, Gaul U, Siggia ED. 2002. Computational detection of genomic cis-regulatory modules applied to body patterning in the early Drosophila embryo. BMC Bioinformatics 3:30.

104. Schroeder MD, Pearce M, Fak J, Fan H, Unnerstall U, et al. 2004. Transcriptional control in the segmentation gene network of Drosophila. PLoS Biol 2:E271.
105. Kel A, Konovalova T, Waleev T, Cheremushkin E, Kel-Margoulis O, et al. 2006. Composite Module Analyst: a fitness-based tool for identification of transcription factor binding site combinations. Bioinformatics 22: 1190–1197.
106. Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, et al. 2002. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. Proc Natl Acad Sci USA 99:757–762.
107. Berman BP, Pfeiffer BD, Laverty TR, Salzberg SL, Rubin GM, et al. 2004. Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in Drosophila melanogaster and Drosophila pseudoobscura. Genome Biol 5:R61.
108. Kim SY, Kim Y. 2006. Genome-wide prediction of transcriptional regulatory elements of human promoters using gene expression and promoter analysis data. BMC Bioinformatics 7:330.
109. Nilsson R, Bajic VB, Suzuki H, di Bernardo D, Bjorkegren J, et al. 2006. Transcriptional network dynamics in macrophage activation. Genomics 88:133–142.
110. Ramsey SA, Klemm SL, Zak DE, Kennedy KA, Thorsson V, et al. 2008. Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. PLoS Comput Biol 4:e1000021.
111. Capovilla M, Kambris Z, Botas J. 2001. Direct regulation of the muscle-identity gene apterous by a *Hox* protein in the somatic mesoderm. Development 128:1221–1230.
112. Chauvet S, Maurel-Zaffran C, Miassod R, Jullien N, Pradel J, et al. 2000. dlarp, a new candidate *Hox* target in Drosophila whose orthologue in mouse is expressed at sites of epithelium/mesenchymal interactions. Dev Dyn 218:401–413.
113. Graba Y, Gieseler K, Aragnol D, Laurenti P, Mariol MC, et al. 1995. DWnt-4, a novel Drosophila Wnt gene acts downstream of homeotic complex genes in the visceral mesoderm. Development 121:209–218.
114. Grieder NC, Marty T, Ryoo HD, Mann RS, Affolter M. 1997. Synergistic activation of a Drosophila enhancer by HOM/EXD and DPP signaling. EMBO J 16:7402–7410.
115. Jones B, McGinnis W. 1993. The regulation of empty spiracles by Abdominal-B mediates an abdominal segment identity function. Genes Dev 7:229–240.
116. Kremser T, Hasenpusch-Theil K, Wagner E, Buttgereit D, Renkawitz-Pohl R. 1999. Expression of the ? tubulin gene (? Tub60D) in the visceral mesoderm of Drosophila is dependent on a complex enhancer that binds Tinman and UBX. Mol Gen Genet 262:643–658.
117. Leemans R, Loop T, Egger B, He H, Kammermeier L, et al. 2001. Identification of candidate downstream genes for the homeodomain transcription factor Labial in Drosophila through oligonucleotide-array transcript imaging. Genome Biol 2: RESEARCH0015.
118. Pederson JA, LaFollette JW, Gross C, Veraksa A, McGinnis W, et al. 2000. Regulation by homeoproteins: a comparison of Deformed-responsive elements. Genetics 156:677–686.
119. Valerius MT, Patterson LT, Witte DP, Potter SS. 2002. Microarray analysis of novel cell lines representing two stages of metanephric mesenchyme differentiation. Mech Dev 110:151–164.
120. Zhao Y, Potter SS. 2001. Functional specificity of the Hoxa13 homeobox. Development 128:3197–3207.
121. Zhou B, Bagri A, Beckendorf SK. 2001. Salivary gland determination in Drosophila: a salivary-specific, fork head enhancer integrates spatial pattern and allows fork head autoregulation. Dev Biol 237:54–67.
122. Hersh BM, Carroll SB. 2005. Direct regulation of *knot* gene expression by Ultrabithorax and the evolution of cis-regulatory elements in *Drosophila*. Development 132:1567–1577.

OPEN ACCESS Freely available online

PLoS GENETICS

Multifactorial Regulation of a Hox Target Gene

Petra Stöbe[1], Sokrates M. A. Stein[1], Anette Habring-Müller[1], Daniela Bezdan[1], Aurelia L. Fuchs[1,2], Stefanie D. Hueber[1], Haijia Wu[1], Ingrid Lohmann[1,2]*

1 Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany, **2** BIOQUANT Center, Heidelberg, Germany

Abstract

Hox proteins play fundamental roles in controlling morphogenetic diversity along the anterior–posterior body axis of animals by regulating distinct sets of target genes. Within their rather broad expression domains, individual Hox proteins control cell diversification and pattern formation and consequently target gene expression in a highly localized manner, sometimes even only in a single cell. To achieve this high-regulatory specificity, it has been postulated that Hox proteins co-operate with other transcription factors to activate or repress their target genes in a highly context-specific manner *in vivo*. However, only a few of these factors have been identified. Here, we analyze the regulation of the cell death gene *reaper* (*rpr*) by the Hox protein Deformed (Dfd) and suggest that local activation of *rpr* expression in the anterior part of the maxillary segment is achieved through a combinatorial interaction of Dfd with at least eight functionally diverse transcriptional regulators on a minimal enhancer. It follows that context-dependent combinations of Hox proteins and other transcription factors on small, modular Hox response elements (HREs) could be responsible for the proper spatio-temporal expression of Hox targets. Thus, a large number of transcription factors are likely to be directly involved in Hox target gene regulation *in vivo*.

Citation: Stöbe P, Stein SMA, Habring-Müller A, Bezdan D, Fuchs AL, et al. (2009) Multifactorial Regulation of a Hox Target Gene. PLoS Genet 5(3): e1000412. doi:10.1371/journal.pgen.1000412

Editor: Mary C. Mullins, University of Pennsylvania School of Medicine, United States of America

Received October 8, 2008; **Accepted** February 9, 2009; **Published** March 13, 2009

Funding: Supported by the Deutsche Forschungsgemeinschaft (IL; LO 844/3-2) and the Max Planck Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ilohmann@flydev.org

Introduction

Distinct morphological structures exist along the anterior-posterior (A/P) axes of animals, and the *Hox* genes represent the major regulators for patterning of this body axis in organisms as diverse as fruit flies, fish and humans [1,2,3]. Since *Hox* genes code for transcription factors, Hox-dependent morphogenesis is driven by the differential regulation of downstream genes [4,5,6]. In line with the very diverse and many-fold effects of Hox proteins on morphogenesis, Hox transcription factors are known to regulate a large number of Hox downstream genes [7,8], including genes that themselves have broad effects on morphology, as well as genes involved in terminal differentiation [reviewed in 3].

Hox genes are expressed in broad and partially overlapping domains along the A/P axis [1,2,3], and their constant and simultaneous activity within hundreds of cells is required throughout development [1]. Despite being active in a very large number of cells, Hox proteins affect target gene expression in precisely defined sub-domains in the animal [for example 9,10,11,12,13,14,15]. In the most extreme case the regulation of a Hox target gene can be limited to a single cell [16]. In addition, some downstream genes can be activated and repressed by the same Hox protein depending on the tissue or developmental stage. Finally, this context dependency also allows a single Hox protein to affect distinct sets of target genes in the same cells during the course of development [8,11,17]. While more and more of these complex regulatory interactions are being described, the molecular mechanisms underlying the spatio-temporal precision of Hox target gene regulation is only poorly understood. This is in large part due to our limited knowledge of the design and function of Hox-dependent enhancers and promoters and their interaction with the regulatory environment.

Only a few Hox regulated enhancers have been analyzed in some detail in *Drosophila* [3]. For example, the activation of *wingless* (*wg*) expression in the visceral mesoderm of *Drosophila* embryos had been shown to be dependent on the Hox protein Abdominal-A (Abd-A) and the Dpp/TGF-β signalling pathway, with both activities functioning on a small *wg* enhancer [11]. Here, two transcriptional effectors of the Dpp/TGF-β pathway, Mother against dpp (Mad) and Creb, had been shown to mediate the Dpp response on the *wg* enhancer, and were thus assumed to work in concert with the Hox protein Abd-A [11]. Only recently, Mad and another effector of the Dpp/TGF-β pathway, Medea (Med), have been found to collaborate with the Hox protein Ultrabithorax (Ubx) to repress transcription of the Hox target gene *spalt major* (*sal*) in the haltere by independently interacting with adjacent Mad/Med and Ubx binding sites in a small *sal* enhancer [18]. And finally, two transcription factors very well known for their function in the *Drosophila* segmentation cascade, Engrailed (En) and Sloppy paired 1 (Slp1), were shown to assist the Hox proteins Ubx and Abd-A in repressing *Distal-less* (*Dll*) expression in the abdomen of *Drosophila* embryos by occupying their identified binding sites in a minimal *Dll* enhancer [10].

Another well-studied direct Hox target gene in *Drosophila* is the apoptosis gene *rpr*, *which* is activated by the Hox protein Dfd in the anterior part of the maxillary segment through four binding sites located in the *rpr*-4S3 regulatory fragment [13]. Since Dfd is active throughout the maxillary segment, whereas *rpr* RNA is found only locally [19], it seemed likely that additional factors contribute to region specific Dfd-dependent *rpr* expression. Here, we find that

Author Summary

Bilateral animals share a common genetic mechanism to control development along the anterior–posterior body axis, and transcription factors of the Hox class are key regulators of this conserved process. It is thought that Hox proteins drive morphological diversification of body segments by differentially controlling the expression of downstream genes. However, due to their highly conserved DNA binding domain, the homeodomain, Hox proteins alone bind to very similar and frequently occurring sequences in the genome. This implies that Hox proteins alone are likely insufficient to activate or repress their target genes in a proper spatio-temporal fashion. In contrast to this observation, Hox proteins have very specific effects during development and execute their function with high precision. A solution to this paradox could lie in the context-specific interaction of Hox proteins with other transcriptional regulators; however, only a few examples are known. By analysing the mechanism underlying the regulation of the Hox target gene *reaper*, we identified a set of eight transcription factors to be important for the precise spatio-temporal regulation of this gene. Based on our findings, we suggest that Hox proteins functionally interact with a plethora of unrelated transcription factors on small, yet complex enhancer elements to execute their specific functions throughout development of diverse organisms.

eight transcriptional regulators, with diverse roles in patterning or differentiation processes, co-operate with Dfd in the regulation of *rpr*. Within their spatially restricted expression domains, these regulators are recruited to a minimal *rpr* enhancer through specific cis-regulatory DNA sequences and act together with Dfd to regulate *rpr* expression in the appropriate spatio-temporal pattern. Thus, our data support the idea that the combinatorial activity of Hox proteins and diverse transcriptional regulators on small regulatory elements is responsible for the spatially and temporarily restricted expression of Hox target genes *in vivo*. In addition, our data show that even small Hox-dependent enhancers are complex and integrate diverse regulatory inputs, which result in precise spatio-temporal expression of Hox target genes.

Results

Dissection of the Dfd-Dependent *rpr* Enhancer

To isolate a minimal *rpr* regulatory element able to recapitulate endogenous expression in the maxillary segment, we divided the known 674 bp long *rpr*-4S3 enhancer element [13], and analyzed *lacZ* expression driven by the resulting sub-fragments (Figure 1). We found that the 3′ part of the enhancer termed *rpr*-4S3/3′, which contained all previously defined Dfd binding sites [13], was sufficient to drive *lacZ* expression in a few cells located in the anterior part of the maxillary segment (Figure 1C′). Double-labelling experiments demonstrated co-localization of *rpr* and *lacZ* transcripts in the *rpr*-4S3/3′ line in a subset of *rpr* expressing cells (Figure 1C″).

We next tested the functional relevance of the Dfd binding sites in the *rpr*-4S3/3′ enhancer by mutational analysis. Surprisingly, we observed an increase in *lacZ* expression in the anterior part of the maxillary segment after mutating all Dfd binding sites (Figure 1D′), rather than a reduction, as observed with the same mutations in the context of the larger *rpr*-4S3 reporter [13]. Additionally, weak *lacZ* expression was observed in the anterior part of all other segments in the *rpr*-4S3/3′-Dfdmt line (Figure 1D). These findings showed that the Dfd binding sites are not exclusively used for activation, but also for repression of *rpr* transcription. Additionally, these results indicated that the overall binding site composition of the *rpr*-4S3 enhancer determines its regulatory output and that most information for repression is located in the 3′ part of the enhancer. Consistently, we found ectopic reporter gene expression in the posterior part of the maxillary segment when using the remaining 5′ part of the *rpr*-4S3 enhancer (*rpr*-4S3/5′) (Figures 1E and 1E′). While the *rpr*-4S3/3′ enhancer harbours most of the binding sites for transcriptional repression, it still has the capacity to direct region-specific activation of *rpr*, since *lacZ* expression is maintained in a few cells in the anterior part of the maxillary segment in the *rpr*-4S3/3′ line (Figure 1C′). Taken together, these results showed that the Dfd-dependent regulation of *rpr* is highly complex and that Dfd has activating and repressing activity even when acting on a small regulatory element. Thus, we decided to study the *rpr*-4S3/3′ enhancer in detail, because its reduced complexity provided a sensitive background to uncover the mechanisms of Hox target regulation *in vivo*.

Identification of Transcription Factors Necessary for Proper *rpr* Expression

To test the effect of Dfd on the minimal *rpr*-4S3/3′ enhancer fragment, we ubiquitously expressed Dfd in the *rpr*-4S3/3′ reporter strain using the *armadillo* (*arm*)-GAL4 driver [20]. We observed specific reporter gene activation in the anterior part of every segment (Figure 2F), a result we had observed before when using the *rpr*-4S3 reporter line (data not shown). *lacZ* RNA never extended into the anterior-dorsal or anterior-ventral zone (Figures 2A and 2F). This led us to hypothesize that essential factors for the Dfd-dependent *rpr* expression are locally expressed in sub-domains of every segment, either in the anterior part if they act as activators or in the posterior, dorsal or ventral part if they act as repressors on the *rpr*-4S3/3′ enhancer.

To test this hypothesis mechanistically, we assayed 16 transcription factors, which meet the expression criteria outlined above, along with two known Dfd interactors, Apontic (Apt) and Disconnected (Disco) [21,22] for their ability to modulate Dfd-dependent *rpr* expression (Table 1). We ubiquitously mis-expressed all factors in embryos harbouring the *rpr*-4S3/3′ reporter, and categorized them dependent on their capacity to affect reporter gene expression (Figure 2). Seven transcription factors were able to elicit the response predicted by their expression patterns (Figure S1): Apt and Glial cells missing (Gcm) activated reporter gene expression (Figures 2C and 2D), whereas Brinker (Brk), Disco, Dorsocross 1 (Doc1), En and Slp1 repressed *lacZ* expression (Figures 2K to 2O). Two of the factors identified in our screen, En and Slp1, have recently been shown to assist Hox proteins in target gene regulation [10], supporting the validity of our approach. Additionally, Disco and Apt were known to genetically and/or biochemically interact with Dfd [21,22]. While over-expression of the activating transcription factors alone had a modest effect (Figures 2B to 2D), simultaneous over-expression with Dfd strongly enhanced reporter gene expression (Figures 2G to 2I). Similar effects were observed when we analyzed endogenous *rpr* RNA expression in these embryos (Figures S2E to S2H), suggesting that these factors are likely to function in concert with Dfd in the induction of *rpr*. Using this co-expression strategy, we identified another factor modulating *rpr* expression: Empty spiracles (Ems) enhanced the ability of Dfd to activate reporter gene expression (Figure 2G), although Ems mis-expression alone had no effect (Figure 2B). Since other candidates tested had no effect on reporter gene expression either alone or in combination with Dfd

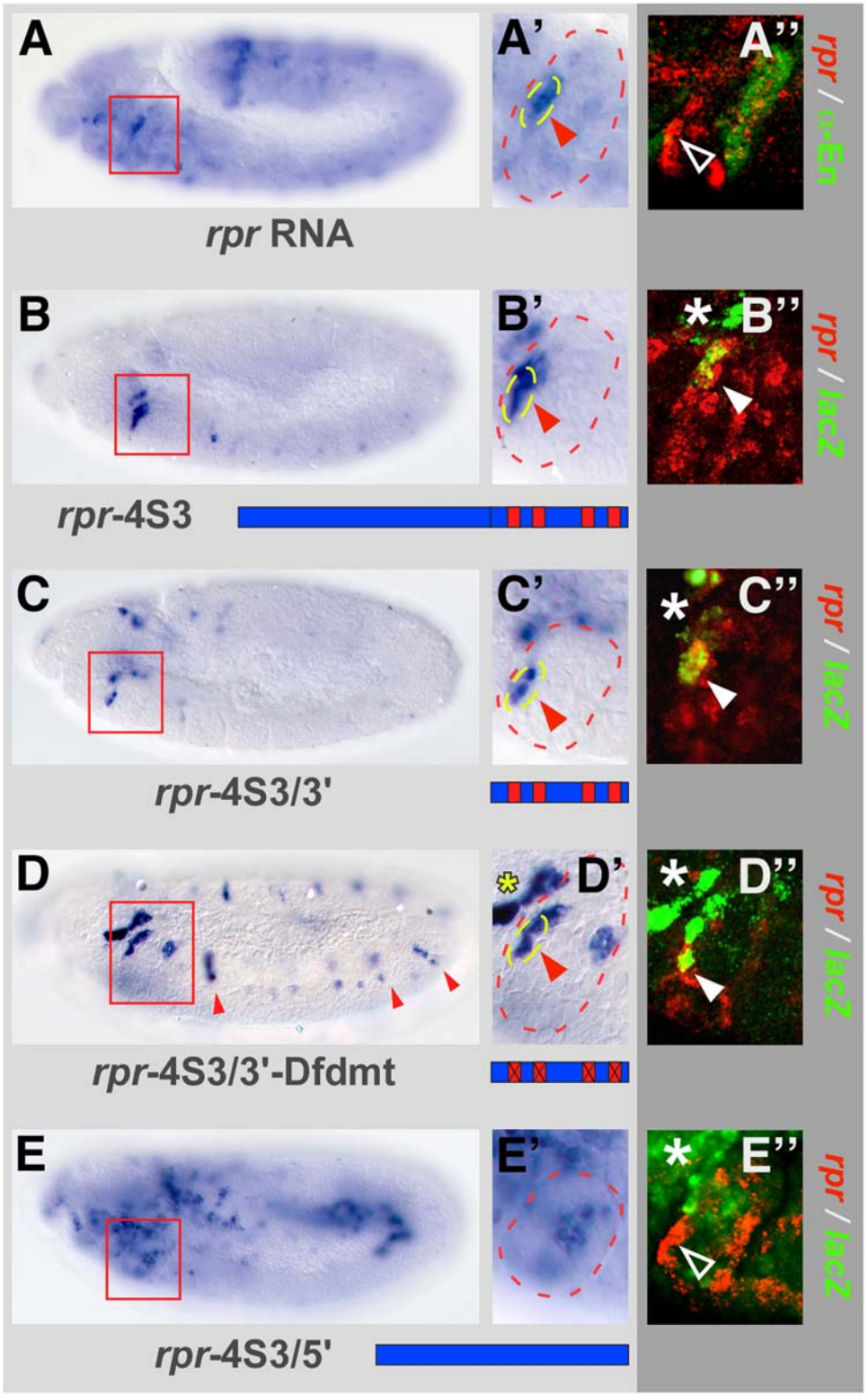

Figure 1. Identification of minimal Dfd response element in the *rpr* enhancer using stage 11 wild-type embryos. (A and A′) *rpr* RNA is strongly expressed in the anterior part of the maxillary segment. (A″) Double-labelling of *rpr* RNA and Engrailed (En) protein. Arrowhead marks *rpr* transcripts, mostly excluded from the posterior part of the maxillary segment (highlighted by En expression). (B and B′) *lacZ* RNA expression in the *rpr*-4S3 reporter line. (C and C′) In the *rpr*-4S3/3′ reporter line, *lacZ* expression recapitulates endogenous *rpr* transcription in the maxillary segment. (D and D′) In the *rpr*-4S3/3′-Dfdmt reporter line all four Dfd binding sites are mutated, resulting in strong *lacZ* activation in the anterior part of the maxillary segment. Small, red arrowheads in (D) indicate ectopic *lacZ* expression in trunk segments. (E and E′) In the *rpr*-4S3/5′ reporter line, *lacZ* is expressed in a broad stripe close to the posterior end. (B″ to E″) Double-labelling of *rpr* and *lacZ* RNA in the *rpr*-4S3 (B″), *rpr*-4S3/3′ (C″), *rpr*-4S3/3′-Dfdmt (D″) and *rpr*-4S3/5′ (E″) transgenic lines. The closed arrowheads in (B″ to D″) mark areas of co-localization of *rpr* and *lacZ* transcripts, the open arrowhead in (E″) marks area of *rpr* expression in the anterior part of the maxillary segment without any *lacZ* transcripts. Red boxes in (A to E) mark the maxillary segment, close-ups of which are shown in (A′ to E′). Asterisks in (B″, C″, D″ and E″) indicate area of *lacZ* expression in procephalic lobe. Blue bars in (B to E) represent different parts of *rpr* enhancer, Dfd binding sites are indicated as small red boxes.
doi:10.1371/journal.pgen.1000412.g001

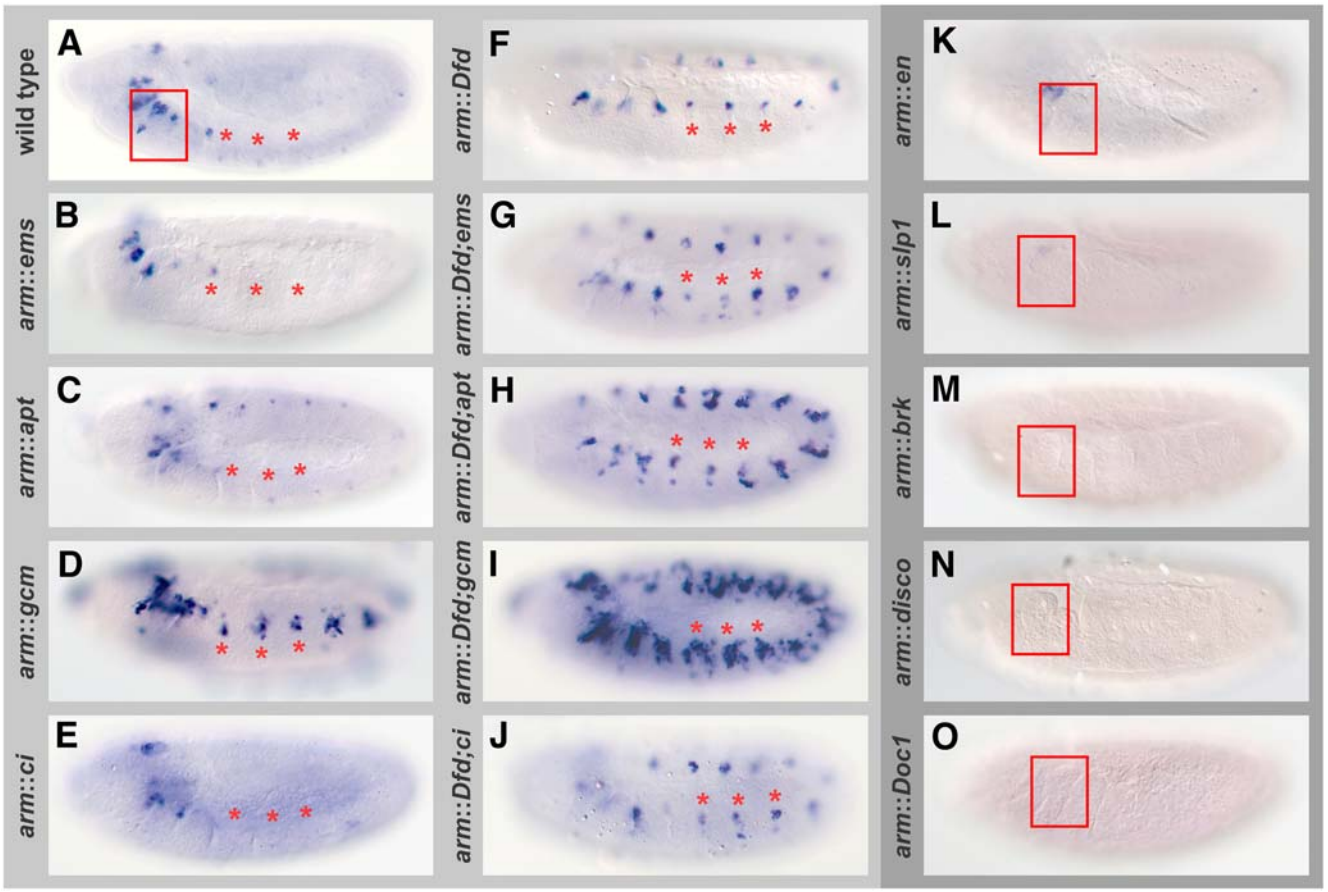

Figure 2. Approach to identify factors for Dfd-dependent *rpr* expression. *lacZ* RNA *in situ* hybridizations in stage 11 embryos ubiquitously mis-expressing different genes in *rpr*-4S3/3′ reporter line using the *arm*-GAL4 driver are shown: (A) *rpr*-4S3/3′ control, (B) *arm::ems*, (C) *arm::apt*, (D) *arm::gcm*, (E) *arm::ci*, (F) *arm::Dfd*, (G) *arm::Dfd;ems*, (H) *arm::Dfd;apt*, (I) *arm::Dfd;gcm*, (J) *arm::Dfd;ci*, (K) *arm::en*, (L) *arm::slp1*, (M) *arm::brk*, (N) *arm::disco*, (O) *arm::Doc1*. The screen is based on the observation that ubiquitous mis-expression of Dfd in the *rpr*-4S3/3′ line leads to ectopic *lacZ* expression in anterior part of every segment (shown in F). In (A to J) asterisks mark three spots of *lacZ* expression in trunk, box in (A, K to O) highlights the maxillary segment.
doi:10.1371/journal.pgen.1000412.g002

(Figures 2E and 2J), we concluded that the effects of Ems on the *rpr*-4S3/3′ enhancer are specific.

Role of Co-Regulatory Factors in Dfd-Dependent *rpr* Expression

To test whether the factors identified are necessary for proper *rpr* transcription in the maxillary segment, we analyzed *rpr* transcripts in embryos mutant for the individual transcription factor genes (Figure 3). Additionally, we studied the morphology of the gnathal lobes, in particular the boundary between the maxillary and mandibular segments, since it is known that maintenance of this boundary depends on proper *rpr* activity [13]. In mutant embryos of two activators, *gcm* and *apt*, we observed a reduction of *rpr* expression in the anterior part of the maxillary segment (Figures 3C′ and 3E′) and a slight fusion of the maxillary and mandibular segments (Figures 3C″ and 3E″). The latter phenotype was not as pronounced as in *Dfd* mutants (Figure 3B″), which exhibit a strongly reduced *rpr* expression in the anterior part of the maxillary segment (Figure 3B′; [13]). In *ems* mutants, the maxillary-mandibular boundary developed normally (Figure 3F″). Here, *rpr* transcript levels were only reduced in the middle part of the anterior *rpr* expression domain, whereas dorsally and ventrally to this area *rpr* transcript levels were elevated (Figure 3F′). As reported previously [23], a loss of the mandibular segment was observed in *ems* mutant embryos (Figure 3F″). To test the interactions between Dfd and the activating transcription factors genetically, we extended our studies to *Dfd gcm* double mutants. In these embryos, *rpr* expression and the formation of the maxillary-mandibular boundary were completely lost (Figures 3D′ and 3D″). This result not only confirmed an important role of the activating factors for *rpr* expression and the maintenance of the segment boundary, but also suggested that Dfd and Gcm act independently.

In embryos mutant for the repressing transcription factor genes, we observed ectopic *rpr* expression in the maxillary segment, primarily in the central or posterior part (Figures 3G′ to 3K′), showing that the factors are involved in repressing *rpr* transcription. The de-repression of *rpr* transcription in only a few cells suggested that repression of *rpr* transcription is redundant. Thus, we aimed to analyze *rpr* expression in embryos mutant for multiple transcription factor genes with repressive function. However, due to lethality, we were not able to generate any double mutant combination. To circumvent this problem, we made use of the fact that the activity of Slp1, Brk and En is dependent on the transcriptional co-repressor Groucho (Gro) [24,25,26,27]. We hypothesized that *gro* mutant embryos should behave similarly to a

Table 1. Transcription factors tested for effect on *rpr*-4S3/3′ reporter gene expression.

Expression	Gene	Effect
anterior	*cubitus interruptus* (*ci*)	none
	cubitus interruptus 75 (*ci75*)	none
	glial cells missing (*gcm*)	activating
	empty spiracles (*ems*)	activating
	gooseberry (*gsb*)	none
	stripe (*sr*)	none
posterior	*engrailed* (*en*)	repressive
	sloppy paired 1 (*slp1*)	repressive
	sloppy paired 2 (*slp2*)	none
dorsal	*Dorsocross 1* (*Doc1*)	repressive
	Dorsocross 2 (*Doc2*)	none
	Dorsocross 3 (*Doc3*)	none
ventral	*brinker* (*brk*)	repressive
	runt (*r*)	none
other	*disconnected* (*disco*)	repressive
	apontic (*apt*)	activating

doi:10.1371/journal.pgen.1000412.t001

triple knock-out of the repressor genes with regards to *rpr* regulation. Consistently, the number of cells ectopically expressing *rpr* was further increased in *gro* mutants when compared to the single mutants. The effect was most pronounced in the posterior part of the maxillary segment (Figures 3L and 3L′). Phenotypic analysis of the repressor mutants revealed that the maxillary-mandibular boundary was not affected (Figures 3F″ to 3K″). This was consistent with largely unchanged *rpr* transcription in the anterior part of the maxillary segment in all mutants for repressive transcription factors (Figures 3F′ to 3K′). Nevertheless, the overall morphology of the gnathal lobes in these mutants was abnormal (Figure 3).

To exclude the possibility that changes in Dfd expression cause the modifications in *rpr* activity and boundary formation observed in the transcription factor mutants, we analyzed Dfd protein localization in these embryos. Since Dfd expression was always unaffected (Figures 4B″ to 4E″), we concluded that the factors identified do not act upstream, but in parallel to Dfd in the regulation of *rpr*. In addition, we could rule out cross-regulatory effects between *gcm* and Slp [28] in the maxillary segment (data not shown).

After having shown a functional relevance for the identified transcription factors, we studied their contribution to *rpr* expression in the context of the *rpr*-4S3/3′ enhancer element, since gene expression is often resistant to the modulation of individual trans-acting factors acting on large and redundant enhancers. We observed a strong reduction of *lacZ* RNA in embryos mutant for the activating transcription factors (Figures 4B′ to 4D′). This result suggests that all three factors play important roles in *rpr* activation and that they act on regulatory elements contained within the *rpr*-4S3/3′ enhancer. In embryos mutant for repressing transcription factors, ectopic *lacZ* activation was observed only in some maxillary cells, as shown for the *Doc* mutant (Figure 4E′), suggesting that transcriptional repression is redundant even at the level of the *rpr*-4S3/3′ enhancer. Consistent with the binding site mutations (Figure 1D′), *lacZ* was strongly activated in the anterior part (but also in other parts) of the maxillary segment in *Dfd* mutants (Figures 4F and 4F′). These results confirmed that in the *rpr*-4S3/3′ context Dfd acts primarily as a repressor, and suggested that full repression is achieved by the combined action of Dfd and additional transcription factors.

Direct Interaction of Co-Regulatory Factors with Minimal *rpr* Enhancer

We next addressed whether the identified factors act directly in the regulation of *rpr* in the maxillary segment. Thus, we studied the expression of *rpr* and all factors with cellular resolution using double-labelling experiments. *rpr* transcripts always co-localized with the activating transcription factors, whereas they were mostly excluded from cells positive for the repressing transcription factors (Figures 5F′ to 5M′). We obtained the same result when *lacZ* transcript distribution and expression of the co-regulatory factors in the *rpr*-4S3 reporter line were analyzed (Figures 5N to 5U). *rpr* transcripts and the activators co-localized in distinct sub-domains of the *rpr* expression zone: Ems in the dorsal most, Gcm in the middle and Apt in the ventral most part (Figures 5F′ to 5H′). This suggests that individual activating factors are responsible for *rpr* transcription in distinct cells in the anterior part of the maxillary segment and that their combined activity is required for the expression of *rpr* in its complete domain.

To further test whether the identified transcription factors are directly involved in the expression of *rpr* on the mechanistic level, we mapped transcription factor binding sites in the *rpr*-4S3/3′ enhancer using phylogenetic footprinting [29,30,31]. Using species-specific *rpr* RNA probes, we could show that *rpr* was expressed specifically in the anterior part of the maxillary segment in all seven *Drosophila* species chosen (Figure S3). We then isolated the *rpr*-4S3/3′ enhancer fragment from all species and after aligning the sequences using the TCoffee algorithm [32,33], we identified three highly conserved boxes, which contained all four Dfd binding sites previously characterized (Figure S3). Furthermore, we found known consensus binding motifs for three of the eight factors within the conserved regions (Figure 5A, Figure S3) [10,34,35,36].

To molecularly test direct binding of all factors identified in the *rpr* enhancer, we performed electrophoretic mobility shift assays (EMSA). We found that all eight transcription factors interact with conserved regions in the *rpr*-4S3/3′ enhancer *in vitro* (Figure 5, Figure S4). To define the DNA sequences necessary for this interaction, systematic competition experiments using overlapping and mutated oligonucleotides for each conserved box were performed. This analysis allowed us to confirm the published Gcm consensus sequence -ACCCGCAT- [37] (Figure 5D), which is located directly adjacent to Dfd binding site 1 in the *rpr*-4S3/3′ fragment (Figure 5A, Table 2). Similarly, En and Slp1 binding sites are found in close proximity to Dfd binding sites 2 and 3 (Figure 5A). Our EMSA analysis uncovered that the assisting factors Apt, En, Slp1 and Brk interact with binding sites slightly divergent from published consensus sequences (Figure 5, Figure S4, Table 2) [10,38,39,40]. Finally, we identified unknown DNA binding sequences for the assisting factors Doc1 and Disco (Figure 5B, C, Figure S4, Table 2). Our competition experiments revealed that Disco protein interacts with two binding sites in the *rpr*-4S3/3′ enhancer (Figure 5B, Figure S4), which share an invariant five nucleotide core motif, -TGACA- (Figure 5A, Table 2).

To test if the identified target sequences are directly bound by the factors *in vivo*, we analyzed the ability of the transcription factors to bind to the *rpr*-4S3/3′ enhancer in the context of chromatin. To this end, we performed chromatin immunoprecipitation assays (ChIPs) for all factors for which functional

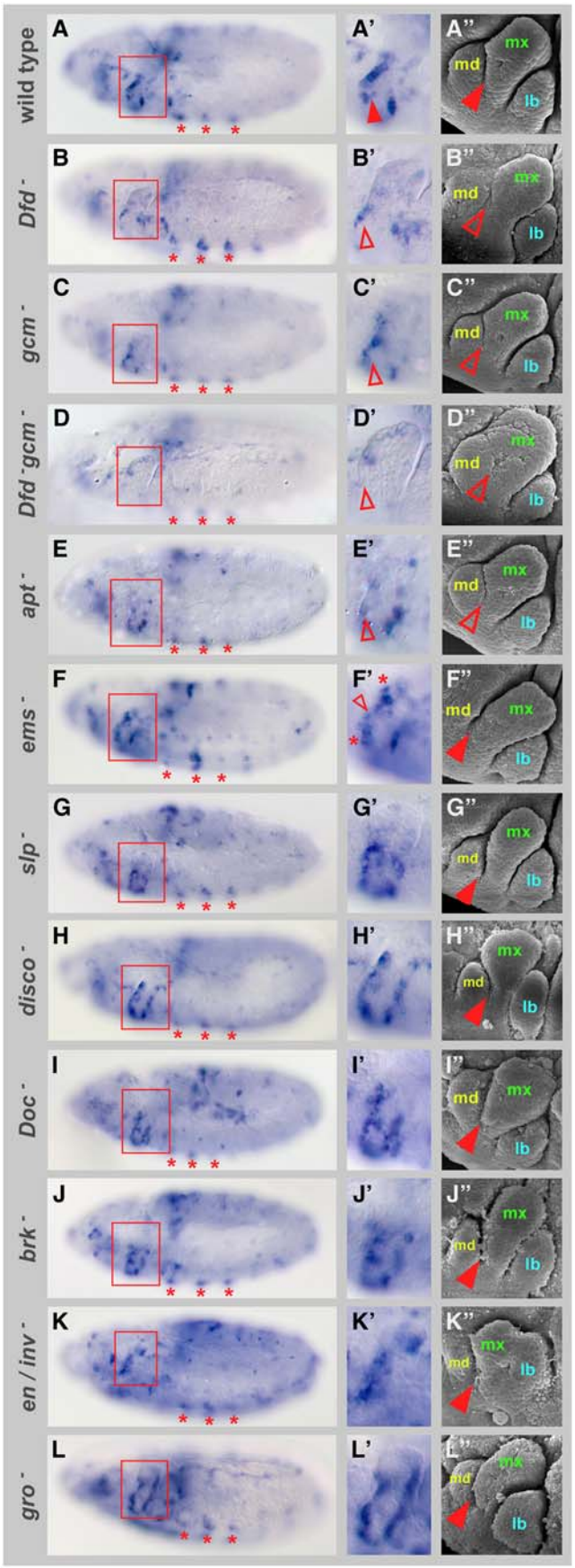

Figure 3. Requirement of transcription factors for *rpr* expression and development of the maxillary segment. (A–L) *rpr* RNA expression in stage 11 wild-type (A), Dfd^{w21} (B), $gcm^{N7\text{-}4}$ (C), Dfd^{w21}; $gcm^{N7\text{-}4}$ (D), apt^{03041} (E), ems^{9G}/ems^{7D99} (F), *Df(2L)slp2-Δd66C* (G), *Df(1)XR14* (H), *Df(3L)DocA* (I), brk^{M68} (J), $Df(2R)en^{E}$ (K) and gro^{B48} (L) mutant embryos. To select identical stages, two criteria were used: 1) overall morphology of embryos; 2) three spots of *rpr* expression in thoracic segments characteristic for stage 11 wild-type embryos (marked by three asterisks). Red boxes in (A to L) highlight maxillary segments. (A′ to L′) Close-up of maxillary segments in respective mutants. In $gcm^{N7\text{-}4}$ and apt^{03041} mutants, *rpr* expression is reduced (C′ and E′), in Dfd^{w21}; $gcm^{N7\text{-}4}$ double mutants expression is lost (D′) (open arrowhead). In ems^{9G}/ems^{7D99} mutants, levels of *rpr* transcripts are reduced in middle part of anterior *rpr* expression area (small open arrowhead), in ventral-anterior and dorsal-anterior part *rpr* transcript levels are increased (highlighted by asterisks). In embryos mutant for repressing transcription factor genes, cells ectopically expressing *rpr* are observed in various parts of the maxillary segment (G′ to K′). In gro^{B48} mutants, *rpr* expression in anterior and posterior parts is increased (L and L′). (A″ to L″) Scanning electron micrographs of gnathal segments of late stage 12 embryos of respective mutants. Mandibular (md), maxillary (mx) and labial (lb) segments are indicated in this panel. In mutants for the activating transcription factor genes, the boundary between the maxillary and mandibular segments is reduced or abolished (C″ to E″) (open arrowhead), reminiscent to the effects seen in *Dfd* mutants (B″), in mutants for the repressing transcription factor genes this boundary is unaffected (G″ to K″) (closed arrowhead).
doi:10.1371/journal.pgen.1000412.g003

antibodies were available. A significant enrichment of the *rpr*-4S3/3′ locus was observed using Dfd, Gcm and En antibodies (Figure 6). These results demonstrated that Dfd, Gcm and En directly interact with specific target sequences in the *rpr*-4S3/3′ enhancer *in vivo*.

Our EMSA experiments also revealed that all factors are able to bind independently of Dfd to the *rpr*-4S3/3′ enhancer, since formation of protein complexes between Dfd and the co-regulatory transcription factors was not observed (Figures 5B to 5E, Figure S4). Consistently, GST pulldown and yeast-two hybrid assays did not provide any evidence for direct interactions of the identified transcription factors with Dfd (data not shown). To exclude the possibility that more than two factors are required for complex formation on DNA, which has been shown before for other HREs [41], we performed EMSA experiments using a mixture of three transcription factors and conserved box 1 as probe. In this region the binding sites for Disco, Dfd and Gcm lie in close proximity, which is considered a requirement for cooperative binding [41]. However, we did not observe a higher-order complex when incubating conserved box 1 with extracts containing all three transcription factor proteins (data not show). Thus, we conclude that these regulators do not bind the *rpr*-4S3/3′ enhancer in a cooperative manner.

Contribution of Co-Regulatory Factors to Dfd-Dependent *rpr* Expression

We next tested the importance of the identified DNA binding sites for Dfd-dependent *rpr* expression in the embryo. To this end, we mutated all sites for the activating or repressing transcription factors in the *rpr*-4S3/3′ element and analyzed reporter gene expression. We found that *lacZ* expression was abolished, when binding sites for all three activating factors, Gcm, Apt and Ems, either alone or in combination with the Dfd binding sites, were mutated (Figures 7C and 7D). These results show that these factors, independently of Dfd, are responsible for activation of the *rpr*-4S3/3′ enhancer element in the anterior part of the maxillary segment. A reduction of *lacZ* expression was even observed when a single binding site for an activating factor was mutated (Figures 7G

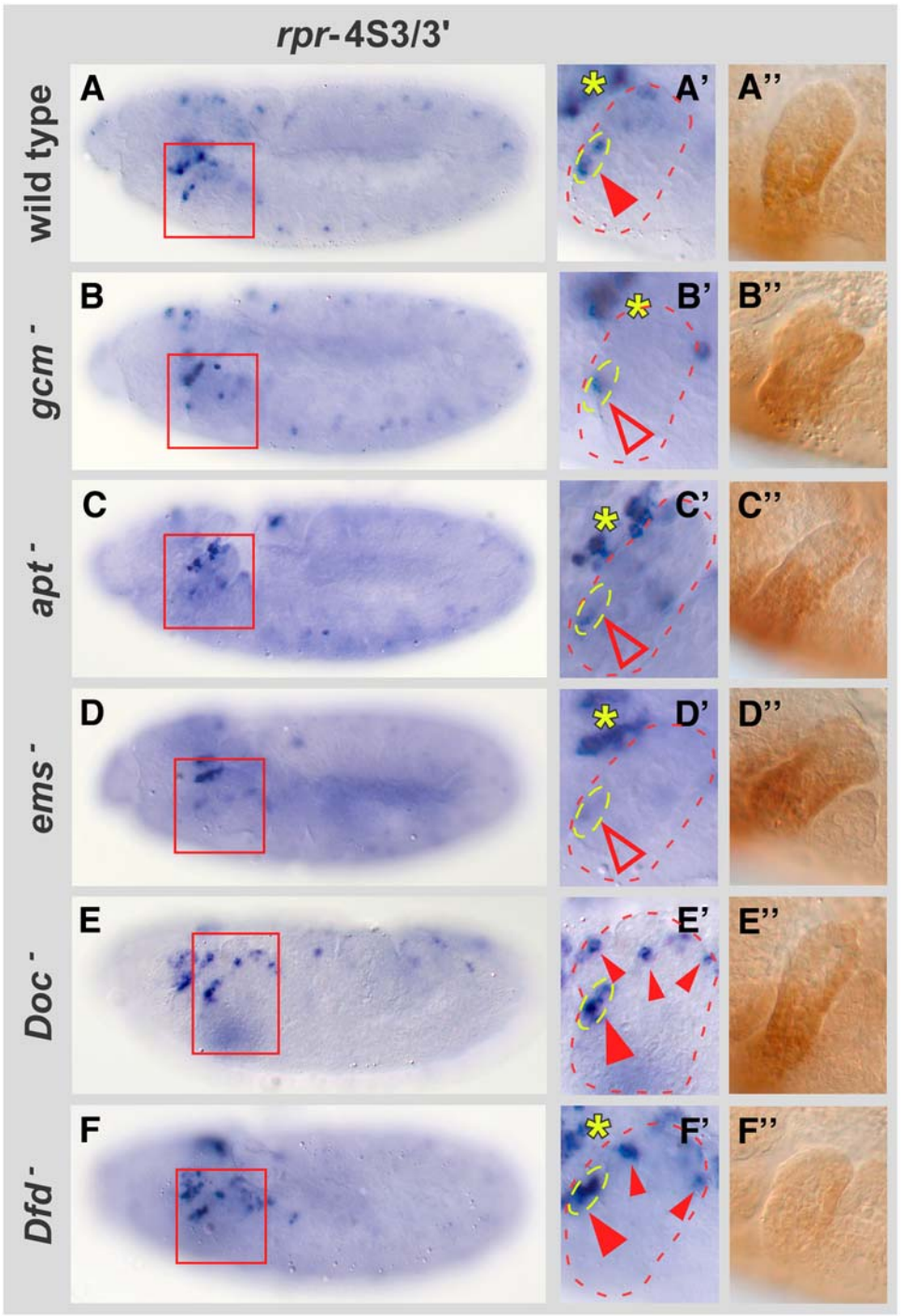

Figure 4. Co-regulatory transcription factors are required for proper *lacZ* expression in stage 11 *rpr*-4S3/3′ reporter line. (A to F) In *rpr*-4S3/3′ reporter line, β-galactosidase expression in the following genetic backgrounds is shown: (A) *rpr*-4S3/3′ control, (B) $gcm^{N7\text{-}4}$, (C) apt^{03041}, (D) ems^{9G}/ems^{7D99}, (E) *Df(3L)DocA*, (F) Dfd^{w21} mutant embryos. Red boxes in (A to F) highlight maxillary segments. (A′ to F′) Close-up of maxillary segments in respective mutants. The yellow asterisks in (A′, B′, C′, D′ and F′) mark expression of *lacZ* in procephalic lobes. (A″ to F″) Dfd protein expression in the respective genotypes. Note that although the morphology of the maxillary segment is changed, the expression domain and intensity of Dfd protein in the respective mutants (B″ to E″) is very similar to wild-type Dfd protein expression (A″).
doi:10.1371/journal.pgen.1000412.g004

and 7G′), further suggesting that the combined action of all three factors is required for activation of the *rpr*-4S3/3′ element. In embryos carrying the *rpr*-4S3/3′ enhancer fragment with all sites for the repressing transcription factors mutated, we observed ectopic reporter gene expression in some, but not all maxillary cells (Figure 7E′). Additionally, *lacZ* was expressed throughout the embryo (Figure 7E). Finally, when all Dfd and repressor sites were mutated, *lacZ* transcription was activated in even more cells of the maxillary segment (Figure 7F and 7F′), confirming that Dfd acts in parallel to the repressing factors. Here, reporter gene expression

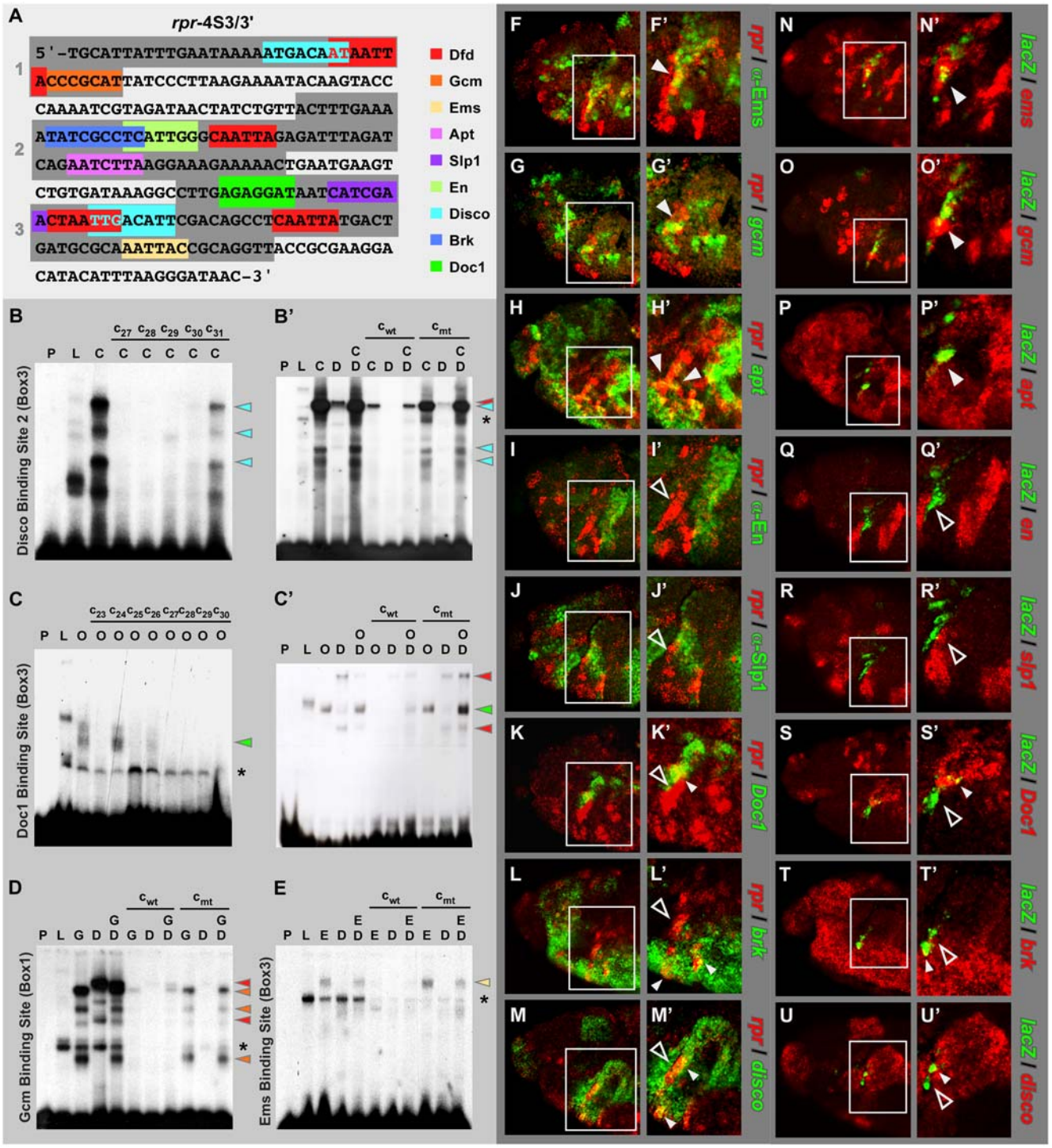

Figure 5. Co-regulatory transcription factors directly interact with *rpr*-4S3/3′ enhancer. (A) Sequence of *rpr*-4S3/3′ enhancer fragment with binding sites for Dfd (shown in red) and all identified co-regulatory transcription factors (highlighted in different colours) is shown. Conserved regions 1 to 3 within the *rpr*-4S3/3′ enhancer are highlighted as dark grey boxes. (B and C) EMSAs for mapping of Disco binding site 2 (B) and Doc1 binding site (C) using box 3 as shift probe. EMSA was performed using no protein (P), translation lysate only (L), lysate with Dfd protein (D), lysate with Disco protein (C) and lysate with Doc1 protein (O). c_{27} to c_{31} in (B) and c_{23} to c_{30} in (C) represent consecutive competitor oligonucleotides with their middle base-pairs mutated. Competition experiments revealed that sequences mutated in the oligonucleotide c_{31} include binding site for the Disco protein (B), whereas oligonucleotides c_{24} include binding site for the Doc1 protein (C). The turquoise and green arrowheads indicate specific DNA-protein complexes containing either Disco or Doc1 protein, respectively. (B′, C′, D, E) EMSAs using no protein (P), translation lysate (L), lysate with Dfd protein (D), Doc1 protein (O), Gcm protein (G), Ems protein (E) and lysate with Dfd protein (D). To test specificity of binding of the proteins to the DNA fragments, competitor oligonucleotides for the mapped binding sites were used either in their wild-type (c_{wt}) or mutant (c_{mt}) sequence versions. Red arrowheads indicate specific DNA-protein complexes containing Dfd protein, turquoise, green, orange or light-yellow arrowheads indicate specific DNA-protein complexes containing Disco, Doc1, Gcm or Ems proteins, respectively. Note that in all competitor oligonucleotides only binding site sequences for co-regulatory transcription factors are mutated, but not for Dfd binding sites. (F to U′) Protein or RNA co-localization of co-regulatory

transcription factors and *rpr* (F to M′) or *lacZ* RNA (N to U′) in head of stage 11 wild-type (F to M′) or *rpr*-4S3 reporter line (N to U′) embryos. Boxes mark maxillary segment with *rpr* or *lacZ* RNAs present in anterior part. In (F′ to M′ and N′ to U′) close-ups of maxillary segments are shown. Co-localization of *rpr* or *lacZ* RNAs and co-regulator RNA is observed in individual cells for Doc1, Brk, Disco (K′ to M′ and S′ to U′; small, closed arrowheads). Closed arrowheads mark cells co-expressing *rpr* and *lacZ* RNAs and RNA or protein of activating co-regulators, open arrowheads highlight areas of *rpr* or *lacZ* transcription and missing expression of repressing co-regulators in anterior part of maxillary segments. Asterisks in (B′, C, D and E) indicate complexes with lysate protein seen also in the controls.
doi:10.1371/journal.pgen.1000412.g005

Table 2. Published and mapped binding sites for all transcription factors within the *rpr*-4S3/3′ Hox response element.

Transcription factor	Mapped DNA binding sites	Published DNA binding sites	Reference
Gcm	**ACCCGCAT**	(**A**/G)**CCCGCAT**	Akiyama et al., 1996
Apt	AA**TCTTA**	(A/G)T**TC**(C/**T**)(A/**T**)**A**T(T/A)(G/A)GA(A/T)(T/C**)**	Liu et al., 2003
Ems	A**ATTA**C	AAXTX**TAAT**GACA	Taylor, 1998
En	**TCATT**GG	**TCATT**C	Gebelein et al., 2004
Slp1	**CATCGAA**	GGTGTGTTGA**CATCGAA**GA	Yu et al., 1999
Brk	TAT**CGCCT**C	(**C**/T)**GCC**A(G/**C)**	Sivasankaran et al., 2000
Doc1	AGAGGAT	-	-
Disco	A**TGACA**AT	-	-
	T**TGACA**TT		

Bold letters highlight identical nucleotides within the *rpr*-4S3/3′ fragment and published consensus sequences.
doi:10.1371/journal.pgen.1000412.t002

was strongly increased in the anterior and posterior part of the maxillary segment (Figure 7F′). However, we never observed *lacZ* expression throughout the whole maxillary segment, suggesting that there are additional, unidentified repressors of *rpr* expression. When single binding sites for repressing transcription factors were mutated, as shown for the Doc1 binding site (Figure 7H′), ectopic reporter gene expression in the maxillary segment was observed, confirming the importance for the direct interaction of repressors with *rpr*-4S3/3′ element.

Discussion

We have shown that eight transcriptional regulators, Apt, Gcm, Ems, En, Slp1, Brk, Doc1 and Disco, are required in addition to

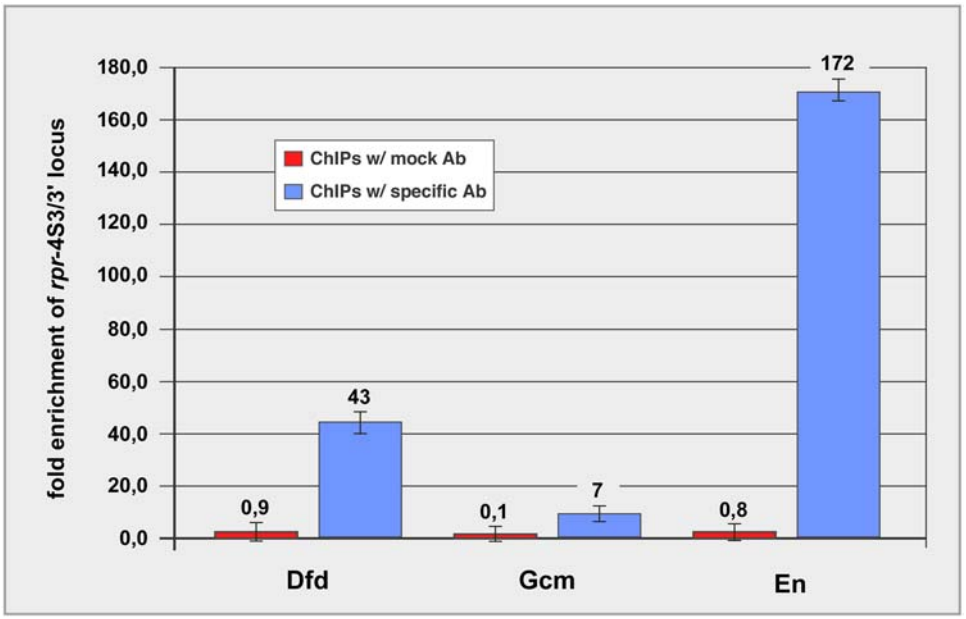

Figure 6. Chromatin immuno-precipitation (ChIP) for Dfd, En and Gcm confirms interaction with *rpr*-4S3/3′ enhancer *in vivo*. Specific enrichment of binding sites within the *rpr*-4S3/3′ enhancer was assayed by quantitative real-time PCR and compared to negative control locus. All ChIPs performed with specific antibodies (blue) yield at least 7-fold enrichment over the negative control, precipitations with mock antibodies (red) yield no enrichment (ratios below 1). Fold enrichment were normalized against input chromatin sample and to negative control region for primer normalization (for details: see Materials and Methods).
doi:10.1371/journal.pgen.1000412.g006

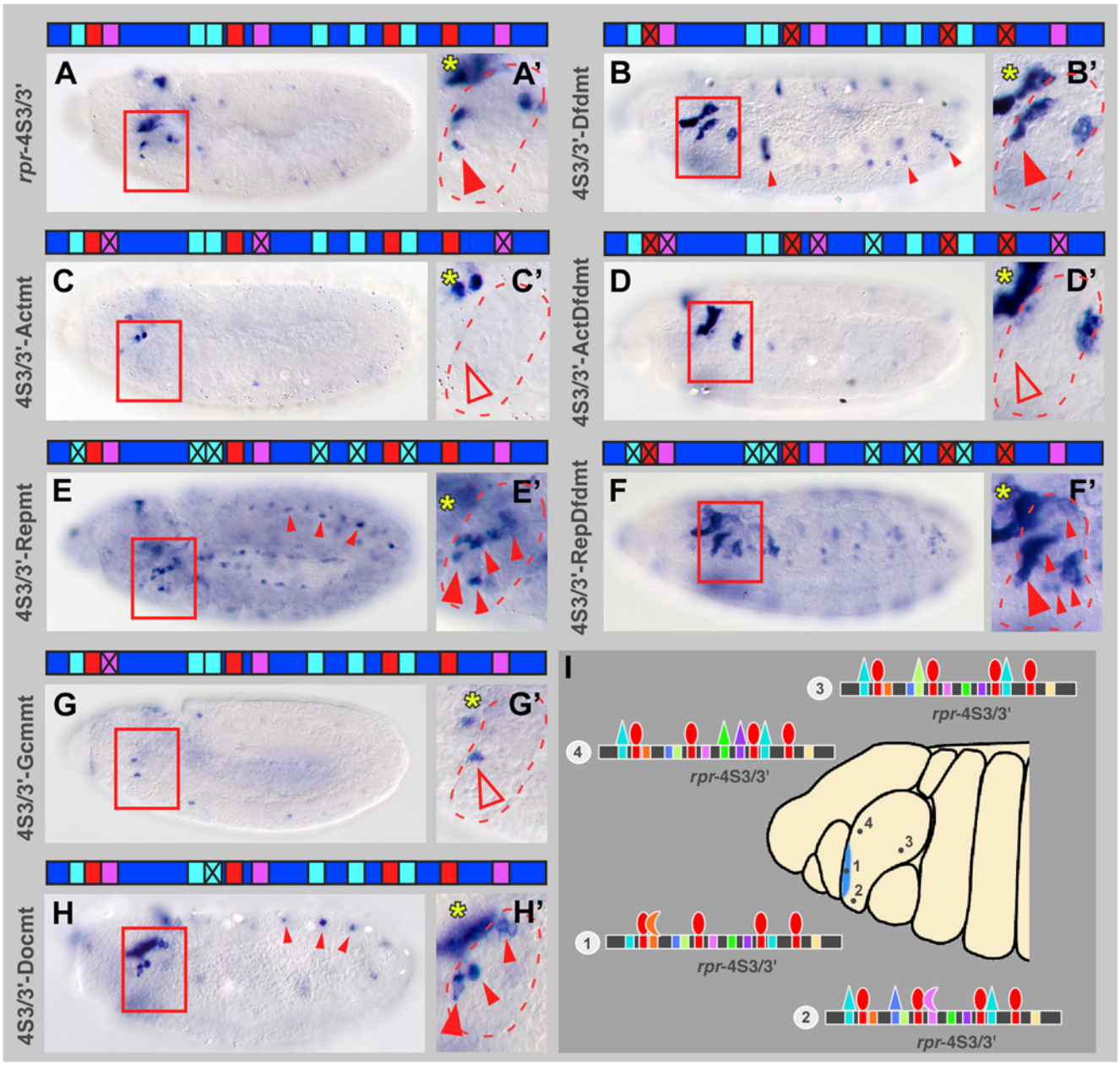

Figure 7. Binding sites for co-regulatory transcription factors are required for *rpr* enhancer activity in stage 11 embryos. (A and A′) β-galactosidase is expressed in anterior part of maxillary segment in *rpr*-4S3/3′ reporter line. Closed, red arrowhead marks anterior part of maxillary segment. (B and B′) In the *rpr*-4S3/3′-Dfdmt reporter, with all Dfd binding sites mutated, *lacZ* expression is increased in anterior part of maxillary segment (closed, red arrowhead). *lacZ* expression is ectopically induced in anterior part of every segment (small, closed arrowheads). (C and C′) In the *rpr*-4S3/3′-Actmt line, with all sites for activating co-regulators mutated, reporter gene expression in maxillary segment is lost (open arrowhead). (D and D′) In the *rpr*-4S3/3′-ActDfdmt reporter, with all Dfd binding sites and sites for activating co-regulators mutated, *lacZ* expression in the anterior part is lost (open arrowhead). (E and E′) In the *rpr*-4S3/3′-Repmt line, with all sites for repressing co-regulators mutated, additional cells in maxillary segment express reporter gene. In rest of embryo, *lacZ* expression is ectopically induced (small, closed arrowheads). (F and F′) In the *rpr*-4S3/3′-RepDfdmt reporter, with all Dfd binding sites and sites for repressing co-regulators mutated, *lacZ* expression in anterior and posterior parts is strongly induced (closed arrowheads). (G and G′) In stage 11 embryos of *rpr*-4S3/3′-Gcmmt line, with the Gcm binding site mutated, reporter gene expression in anterior part of maxillary segment is reduced (open arrowhead). (H and H′) In the *rpr*-4S3/3′-Docmt line, with the Doc1 binding site mutated, reporter gene expression is observed in additional cells in maxillary segment. In the rest of the embryo, *lacZ* expression is ectopically induced (small, closed arrowheads). (I) Model of *rpr* regulation through the *rpr*-4S3/3′ enhancer. Expression of *rpr* in the anterior part of the maxillary segment (highlighted in blue) is achieved through a combinatorial interaction of the Hox protein Dfd and co-regulatory transcription factors (represented as different-coloured triangles) to specific binding sites in the *rpr*-4S3/3′ enhancer. Each cell of the maxillary segment expresses different combinations of Dfd and the co-regulatory transcription factors, which is reflected in a cell type-specific occupancy of the *rpr*-4S3/3′ enhancer, as shown exemplarily for four different cells (marked 1 to 4). According to the model, the decision whether *rpr* transcription is activated or repressed in individual maxillary cells depends on the nature and combination of regulatory factors interacting with the *rpr*-4S3/3′ enhancer. Boxes in (A to F) highlight maxillary segments, yellow asterisks in (A′ to H′) mark *lacZ* expressing cells in procephalic lobes. *rpr*-4S3/3′ enhancer in (A to H) is represented as blue bar, Dfd binding sites as red, sites for activating co-regulators as pink and sites for repressing co-regulators as turquoise boxes.
doi:10.1371/journal.pgen.1000412.g007

the Hox protein Dfd to properly regulate the expression of the apoptosis gene *reaper* in a specific subset of cells of the maxillary segment. The finding that such a large number of structurally unrelated transcription factors with important and diverse functions during differentiation and cell-type specification processes assist Dfd in the regulation of the *rpr*4S3/3′ enhancer element was surprising. For example, Gcm is one of the major regulators of glial cell differentiation, consistent with its expression in glial precursor cells during embryogenesis [42,43]. Two other factors, Brk and Doc1, are both known to play important roles in the Dpp/TGF-β signalling pathway: Brk, a genuine transcriptional repressor [24], acts as a negative regulator of Dpp-dependent genes [44,45,46], whereas Doc1, one of the three genetically redundant Dorsocross transcription factors required for amnioserosa development in *Drosophila* [47], is a direct target of the Dpp pathway [48]. Furthermore, Apt and Disco have previously been found in genetic screens designed to identify modifiers/interactors of Dfd [21,22]. Mutations in these genes, which are expressed in the gnathal segments [21,22], result in severe malformations of structures derived from these segments, similarly to the defects observed in *Dfd* mutants. And finally, we have identified three factors known to be critically involved in patterning the A/P axis as important regulators of *rpr* expression: the gap-like segmentation gene product Ems, which is important for patterning embryonic head structures [49,50], and two segment polarity factors, En and Slp1. Taken together, our findings suggest that proper spatio-temporal Hox target gene regulation is achieved by the combined action of multiple transcriptional regulators: the Hox proteins themselves and a large number of structurally diverse transcription factors. Although the interaction with additional transcription factors has been reported before [11,14,51], the finding that a multitude of diverse factors is required to regulate the activity of a small HRE adds a new layer of complexity to the mechanisms of Hox target gene regulation.

Our results not only demonstrate that the newly identified factors are functionally involved in *rpr* regulation, but also show mechanistically that they contribute to localized *rpr* activation by direct interactions with specific DNA sequences located in the *rpr*4S3/3′ enhancer element both *in vitro* and *in vivo*. Additionally, all factors very likely bind to their target sequences independently of Dfd. Thus, our findings support and significantly extend recent observations: the repression of the Hox target *sal* in the haltere requires the direct interaction of the Hox protein Ubx and two Dpp downstream effectors, Mad and Med, with adjacent binding sites in the *sal*1.1 CRE [18]. As in our case, no evidence for a direct cooperative interaction of the assisting transcription factors with the Hox protein was detected. Thus, we postulate that the Hox-dependent regulation of *rpr* expression is, as in the case of *sal* repression by Ubx, achieved through combinatorial regulation, in which two or more regulatory proteins bind to nearby sites, but not necessarily to each other [18].

While combinatorial regulation of gene expression has been extensively studied for diverse transcription factors [52,53], our results shed new light on the mechanisms of Hox target gene regulation. Previously, much attention focused on the Hox cofactors Extradenticle (Exd) and Homothorax (Hth), which allow Hox proteins to differentially recognize and select some of their target genes through cooperative complex formation [54,55,56,57]. Although Hox cofactors like Exd can explain why different Hox proteins have different DNA binding specificities, the interaction with these factors was not able to explain how broadly expressed Hox proteins are able to affect target gene expression in only a subset of cells. One of the major reasons for that is that Exd and Hth, the only well-known Hox cofactors, are expressed throughout the embryo and interact promiscuously with most Hox proteins. In addition, studies on Exd and Hth revealed that Hox target gene regulation more or less inevitably includes complex formation between Hox proteins and assisting co-regulatory factors. We and others have now shown that the ability of a broadly expressed Hox protein to regulate a target gene in a proper spatial and temporal context is achieved by the Hox-independent recruitment of context-specific transcription factors to cis-regulatory sequences present in compact HREs [18]. Based on our findings, we now suggest that a large number of transcription factors could dictate the transcriptional output in combination with the respective Hox protein by binding selectively and independently to cis-regulatory sequences within HREs of target genes (Figure 7I). Since every cell has a unique combination of transcription factors, the combinatorial interactions for the broadly expressed Hox proteins are almost limitless in such a scenario, accounting for the precise modulation and fine-tuning of Hox target gene regulation, even on the level of the individual cell. Additionally, this model can explain how Hox proteins can act as repressors in one context and as activators in another, because the combined transcriptional output is dependent on the regulatory activity of all transcription factors assembled on a HRE (Figure 7I). There are several lines of evidence that support this model: first, the invariable ectopic activation of Hox downstream genes in spatially restricted domains of every segment when upstream Hox proteins are ubiquitously mis-expressed [8], and second, the accumulation of binding sites for additional transcription factors in enhancers predicted to be controlled by Hox proteins [8]. Alternatively, it seems also possible that Dfd regulates the expression of its target gene *rpr* only in some maxillary cells, while the novel co-regulatory factors mediate regulation in other maxillary cells. However, since Dfd protein is present in all cells of the segment, the first model seems more likely. Irrespective of the mechanism used by Dfd, it will be essential to study the architecture of HREs, with a special focus on the binding site composition of these enhancers and the diverse factors binding to them to further advance our understanding of Hox target gene regulation *in vivo*.

It has been argued before that context-specific transcription factors assisting Hox proteins in target gene regulation are not likely to act as transcriptional repressors or activators themselves, but rather recruit co-repressors and/or co-activators, and thereby dictate the transcriptional output imprinted in HREs [18,58]. Our finding of the co-repressor Groucho playing a role in the Dfd-dependent repression of *rpr* transcription now substantiates this hypothesis, since three of the factors identified in our work, En, Slp1 and Brk, are known to require interactions with the Groucho co-repressor for the transcriptional repression of some of their downstream genes [24,25,59]. Interestingly, at least two other co-regulatory transcription factors identified in this work are also known to interact with co-activators/co-repressors: Apt is able to recruit the transcriptional co-activator Multiprotein bridging factor 1 (MBF1), thereby mediating Apt-dependent transcriptional activation [35]. Disco has been found in a yeast-two-hybrid screen to interact with the well-known co-repressor C-terminal Binding Protein (CtBP) [60], which is also recruited by Brk to repress some Dpp-responsive genes [24]. Since there is accumulating evidence that co-repressors, like CtBP, execute their function on transcriptional regulation through chromatin modification [61,62], it is tempting to speculate that Hox proteins regulate their target genes also by epigenetic control mechanisms. In summary, the multitude of potential regulatory mechanisms used by Hox proteins might be the reason why it has been impossible to fully elucidate how Hox proteins mediate their function with high specificity and precision

in vivo. In the light of recent advances, it now seems likely that the mechanisms leading to functional specificity of Hox proteins are dependent on the cellular context, the composition of the target enhancer element and the identity of the individual Hox protein. Thus, one could argue that during evolution Hox proteins have undergone individualization in trans, as well as sequence diversification in cis.

Materials and Methods

Drosophila Genetics

D. simulans, *D. yakuba*, *D. erecta* and *D. mojavensis* were obtained from the Tucson *Drosophila* Stock Center. The *D. melanogaster* strain used was Oregon-R. apt^{03041}, ems^{7D99}, $gcm^{N7\text{-}4}$, $Df(2R)en^{E}$, UAS-*disco*, UAS-*gcm* strains were obtained from the Bloomington Stock Center; gro^{B48} line, P. Heitzler [63]; ems^{9G} flies, W. McGinnis [38]; brk^{M68}, UAS-*brk* lines, S. Roth [45]; *Df(1)XR14* flies, H. Saumweber [64]; *Df(3L)DocA*, UAS-*Doc1* strains, M. Frasch [48]; *Df(2L)slp2-Δd66C* strain, W. Gehring [65]; UAS-*en* flies, I. Guerrero [66]; UAS-*ems* line, H. Jäckle [67]; UAS-*slp1* flies, M. Leptin [68]; UAS-*apt* strain, R. Schuh [69]; UAS-*ci* flies, T. Kornberg [70]. The following lines are described in Hueber et al. (2007): Dfd^{ev21}, Dfd^{11}, *arm*-GAL4, *prd*-GAL4, UAS-*Dfd*, UAS-*lacZ*. The following green balancer lines were used: Dr^{Mio}/TM3*Sb*[*twi*::2xEGFP], *In(2LR)Gla wg*-Gla/Cyo[*twi*::2xEGFP] , *N*/FM7c[*twi*::2xEGFP].

Plasmids

cDNAs were obtained from: *disco*, *Drosophila* Genomics Resource Center (GH27656), *gcm* cDNA, G. Technau [71], *brk* cDNA, C. Rushlow [45,72], *Doc1* cDNA, M. Frasch [48], *apt* cDNA, R. Schuh [69]. Mutations in the *rpr*-4S3/3′ fragment were created by site-directed mutagenesis via two-step PCR or the QuickChange Multi Site-directed Mutagenesis Kit (Stratagene). Primer sequences are available upon request. All products were cloned, sequenced, and shuttled into pH-Pelican plasmid [73]. All transgenic fly lines were generated by the BestGene *Drosophila* Embryo Injection Service. At least three independent lines were analyzed for expression levels. *rpr* coding regions from different *Drosophila* species were PCR amplified with specific primers, cloned and sequenced.

Histology and Scanning Electron Microscopy

In situ hybridization and immunochemistry were performed as described [74,75]. Fluorescent RNA / protein double labelling and fluorescent duplex *in situ* hybridizations were done as described previously [13,76]. Probe detection was done using the TMR and Fluorescein TSA Amplification kits from PerkinElmer (Waltham, MA). Antibodies were: rat anti-Ems (1:200), U. Walldorf; mouse anti-En (mAb4D9) (1:200), Developmental Studies Hybridoma Bank (Iowa, University); guinea pig anti-Slp (1:200), J. Jäckle; anti-DIG POD, Roche (Penzberg, Germany); anti-mouse AlexaFluor 488, anti-guinea pig AlexaFluor 488 and anti-rat AlexaFluor 488, Molecular Probes. All fluorescent images were taken at Zeiss LSM510 META confocal microscope. SEM analysis was performed as described in Lohmann et al. (2002).

Electrophoretic Mobility Shift Assays

EMSA was performed as described previously (Lohmann et al., 2002). For the zinc finger transcription factor Disco, 1 mM $ZnSO_4$ was included in the binding reaction. For the mapping of binding sites, the ability of all eight transcription factors to interact with conserved boxes 1 to 3 was tested, to define the interaction domains on the *rpr*-4S3/3′ fragment. Subsequently, if known binding sites were present within the conserved boxes, competition experiments were performed to test if these sites are necessary for binding. For all factors with unknown binding sites, systematic competition experiments using overlapping and mutated oligonucleotides covering the binding region were performed. All oligonucleotide sequences used for these experiments can be obtained upon request.

Chromatin Immuno-Precipitation (ChIP)

ChIP experiments were performed as described previously at www.flychip.org. Four independently staged wild-type embryo populations were collected, chromatin samples were prepared from 5 to 9.5 hr embryo collections. Antibodies used were the following: guinea pig anti-Dfd, guinea pig anti-IgGs (gift from H. Schwarz, MPI Tuebingen), mouse anti-En, mouse anti-LacZ (Invitrogen), rat anti-Gcm (gift from M. Wegner, University Erlangen) and rat anti-GFP (Invitrogen). A dilution of 1:500 was used for the anti-Dfd, anti-En and anti-Gcm antibodies, the mock antibodies were used at equivalent protein concentrations. Amplification of the *rpr*-4S3/3′ and an unrelated, non-coding control locus were analyzed by quantitative real-time PCR in technical triplicates using at least two biological replicates. Precipitates were normalized to input DNA (i.e., sonicated, pre-ChIP DNA) and compared to the non-coding negative control region. A PCR efficiency of 1.8-fold amplifications per cycle was assumed. PCR primer sequences can be obtained upon request.

Supporting Information

Figure S1 Expression patterns of the identified transcription factors in stage 11 wild-type embryos. For the following genes, antibody stainings are shown: *ems* (B), *en* (C) and *slp1* (E). Due to the unavailability or inactivity of antibodies, *in situ* hybridizations for the following genes are shown: *gcm* (A), *apt* (D), *Doc1* (F), *brk* (G) and *disco* (H). Boxes in (A to H) highlight the maxillary segment.
Found at: doi:10.1371/journal.pgen.1000412.s001 (2.53 MB TIF)

Figure S2 Identified transcription factors modulate *rpr* expression when mis-expressed. *rpr* RNA in situ hybridizations in stage 11 embryos with the following genotypes are shown: (A) wild type, (B) *prd*::*Dfd*, (C) *prd*::*ems*, (D) *prd*::*Dfd;ems*, (E) *prd*::*apt*, (F) *prd*::*Dfd;apt*, (G) *prd*::*gcm*, (H), *prd*::*Dfd;gcm*. Co-expression of Ems, Apt and Gcm with Dfd enhances Dfd-dependent ectopic *rpr* induction (D, F and H), whereas Gcm is able to ectopically induce *rpr* expression alone. To select identical stages, three characteristic spots of *rpr* expression in the thoracic segments normally seen in stage 11 wild-type embryos (marked by three asterisks) were used. In (B to H) one spot of ectopic *rpr* expression at the very posterior end in the *prd*-GAL4 over-expression embryos is marked by a red arrow, the blue box outlines an additional stripe of *rpr* RNA expression in the T3 primordium.
Found at: doi:10.1371/journal.pgen.1000412.s002 (6.46 MB TIF)

Figure S3 Identification of conserved regulatory elements in the *rpr*-4S3/3′ enhancer by phylogenetic footprint analysis. Upper: *rpr* RNA expression in stage 11 embryos of different *Drosophila* species used for the phylogenetic footprint analysis. *In situ* hybridization experiments with species-specific probes show that *rpr* is expressed in the anterior part of all five *Drosophila* species (*D. melanogaster*, *D. simulans*, *D. yakuba*, *D. erecta*, *D. mojavensis*). The red boxes highlight the maxillary segment. Bottom: Alignment of the *rpr*-4S3/3′ enhancer from seven different *Drosophila* species revealed three highly conserved boxes (I to III). Identified and verified binding sites for all eight co-regulatory transcription factors are highlighted in different colours.
Found at: doi:10.1371/journal.pgen.1000412.s003 (3.37 MB TIF)

Figure S4 Identification and verification of binding sites for co-regulatory transcription factors in the *rpr*-4S3/3′ enhancer. (A) EMSA for mapping of Slp1 binding site in the *rpr*-4S3/3′ enhancer using box 3 (as shown in Figure S3) as shift probe. EMSA was performed using no protein (P), translation lysate only (L) and lysate with Slp1 protein (S). c_{30} to c_{36} represent competitor oligonucleotides with consecutive base-pairs mutated. Competition experiments revealed that sequences mutated in the oligonucleotides c_{33} include binding site for the Slp1 protein. The purple arrowheads indicate specific DNA-protein complexes containing Slp1 protein. The asterisk indicates a complex with lysate protein seen also in the control. (A′) EMSA using box 3 (as shown in Figure S3) and no protein (P), translation lysate (L), lysate with Slp1 protein (S) and lysate with Dfd protein (D). To test specificity of binding of Slp1 protein to the DNA fragment, competitor oligonucleotides for the mapped Slp1 binding site were used either in their wild-type (c_{wt}) or mutant (c_{mt}) sequence versions. The purple arrowheads indicate specific DNA-protein complexes containing Slp1 protein. Note that in the competitor oligonucleotides only the binding site sequence for the Slp1 protein is mutated, but not for the Dfd binding site sequence. (B) EMSA for mapping of Disco binding site 1 in the *rpr*-4S3/3′ enhancer using box 1 (as shown in Figure S3) as shift probe. EMSA was performed using no protein (P), translation lysate only (L) and lysate with Disco protein (C). c_1 to c_7 represent competitor oligonucleotides with consecutive base-pairs mutated. Competition experiments revealed that sequences mutated in the oligonucleotides c_3 and c_4 include binding site for the Disco protein. The turquoise arrowheads indicate the DNA-protein complexes containing Disco protein. The asterisk indicates a complex with lysate protein seen also in the control. (B′) EMSA using box 1 (as shown in Figure S3) and no protein (P), translation lysate (L), lysate with Disco protein (C) and lysate with Dfd protein (D). To test specificity of binding of Disco protein to the DNA fragment, competitor oligonucleotides for the mapped Disco binding site were used either in their wild-type (c_{wt}) or mutant (c_{mt}) sequence versions. The red and turquoise arrowheads indicate the specific DNA-protein complexes containing either Dfd or Disco protein, respectively. The asterisk indicates a complex with lysate protein seen also in the control. Note that in the competitor oligonucleotides only the binding site sequence for the Disco protein is mutated, but not for the Dfd binding site sequence. (C) EMSA for mapping of Brk binding site in the *rpr*-4S3/3′ enhancer using box 2 (as shown in Figure S3) as shift probe. EMSA was performed using no protein (P), translation lysate only (L) and lysate with Brk protein (B). c_8 to c_{13} represent competitor oligonucleotides with consecutive base-pairs mutated. Competition experiments revealed that sequences mutated in the oligonucleotide c_{10} and c_{11} include binding site for the Brk protein. The blue arrowheads indicate specific DNA-protein complexes containing Brk protein. The asterisks indicate complexes with lysate protein seen also in the control. (C′) EMSA using box 2 (as shown in Figure S3) and no protein (P), translation lysate (L), lysate with Brk protein (B) and lysate with Dfd protein (D). To test specificity of binding of Brk protein to the DNA fragment, competitor oligonucleotides for the mapped Brk binding site were used either in their wild-type (c_{wt}) or mutant (c_{mt}) sequence versions. The red and blue arrowheads indicate the specific DNA-protein complexes containing either Dfd or Brk protein, respectively. The asterisks indicate complexes with lysate protein seen also in the control. Note that in the competitor oligonucleotides only the binding site sequence for the Brk protein is mutated, but not for the Dfd binding site sequence. (D) EMSA using box 2 (as shown in Figure S3) and no protein (P), translation lysate (L), lysate with En protein (N) and lysate with Dfd protein (D). To test specificity of binding of En protein to the DNA fragment, competitor oligonucleotides for En binding site were used either in their wild-type (c_{wt}) or mutant (c_{mt}) sequence versions. The red and light green arrowheads indicate the specific DNA-protein complexes containing either Dfd or En protein, respectively. Note that in the competitor oligonucleotides only the binding site sequence for the En protein is mutated, but not for the Dfd binding site sequence. (E) EMSA using box 2 (as shown in Figure S3) and no protein (P), translation lysate (L), lysate with Apt protein (A) and lysate with Dfd protein (D). To test specificity of binding of Apt protein to the DNA fragment, competitor oligonucleotides for Apt binding site were used either in their wild-type (c_{wt}) or mutant (c_{mt}) sequence versions. The red and light purple arrowheads indicate the specific DNA-protein complexes containing either Dfd or Apt protein, respectively. Note that in the competitor oligonucleotides only the binding site sequence for the Apt protein is mutated, but not for the Dfd binding site sequence. Found at: doi:10.1371/journal.pgen.1000412.s004 (3.72 MB TIF)

Acknowledgments

We thank Jürgen Berger for advice with the SEMs, Jan U. Lohmann for discussion and critically reading the manuscript.

Author Contributions

Conceived and designed the experiments: PS IL. Performed the experiments: PS SMAS AHM DB ALF SDH HW. Analyzed the data: PS IL. Wrote the paper: IL.

References

1. McGinnis W, Krumlauf R (1992) Homeobox genes and axial patterning. Cell 68: 283–302.
2. Mann RS, Morata G (2000) The developmental and molecular biology of genes that subdivide the body of *Drosophila*. Annu Rev Cell Dev Biol 16: 243–271.
3. Pearson JC, Lemons D, McGinnis W (2005) Modulating *Hox* gene functions during animal body patterning. Nature Genetics Reviews 6: 893–904.
4. Botas J (1993) Control of morphogenesis and differentiation by *HOM/Hox* genes. Curr Opin Cell Biol 5: 1015–1022.
5. Graba Y, Aragnol D, Pradel J (1997) *Drosophila* Hox complex downstream targets and the function of homeotic genes. Bioessays 19: 379–388.
6. Pradel J, White RA (1998) From selectors to realizators. Int J Dev Biol 42: 417–421.
7. Hersh BM, Nelson CE, Stoll SJ, Norton JE, Albert TJ, et al. (2007) The UBX-regulated network in the haltere imaginal disc of *D. melanogaster*. Dev Biol 302: 717–727.
8. Hueber SD, Bezdan D, Henz SR, Blank M, Wu H, et al. (2007) Comparative analysis of Hox downstream genes in *Drosophila*. Development 134: 381–392.
9. Capovilla M, Brandt M, Botas J (1994) Direct regulation of *decapentaplegic* by Ultrabithorax and its role in *Drosophila* midgut morphogenesis. Cell 76: 461–475.
10. Gebelein B, McKay DJ, Mann RS (2004) Direct integration of Hox and segmentation gene inputs during *Drosophila* development. Nature 431: 653–659.
11. Grienenberger A, Merabet S, Manak J, Iltis I, Fabre A, et al. (2003) TGFb signaling acts on a Hox response element to confer specificity and diversity to Hox protein function. Development 130: 5445–5455.
12. Hersh BM, Carroll SB (2005) Direct regulation of *knot* gene expression by Ultrabithorax and the evolution of cis-regulatory elements in *Drosophila*. Development 132: 1567–1577.
13. Lohmann I, McGinnis N, Bodmer M, McGinnis W (2002) The *Drosophila Hox* gene *Deformed* sculpts head morphology via direct regulation of the apoptosis activator *reaper*. Cell 110: 457–466.
14. Marty T, Vigano MA, Ribeiro C, Nussbaumer U, Grieder NC, et al. (2001) A HOX complex, a repressor element and a 50 bp sequence confer regional specificity to a DPP-responsive enhancer. Development 128: 2833–2845.
15. Vachon G, Cohen B, Pfeifle C, McGuffin ME, Botas J, et al. (1992) Homeotic genes of the *Bithorax* complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *Distal-less*. Cell 71: 437–450.
16. Brodu V, Elstob PR, Gould AP (2002) Abdominal-A specifies one cell type in *Drosophila* by regulating one principal target gene. Development 129: 2957–2963.

17. Bondos SE, Tan XX (2001) Combinatorial transcriptional regulation: the interaction of transcription factors and cell signaling molecules with homeodomain proteins in *Drosophila* development. Crit Rev Eukaryot Gene Expr 11: 145–171.
18. Walsh CM, Carroll SB (2007) Collaboration between Smads and a Hox protein in target gene repression. Development 134: 3585–3592.
19. Lohmann I (2003) Dissecting the regulation of the *Drosophila* cell death activator *reaper*. Gene Expr Patterns 3: 159–163.
20. Sanson B, White P, Vincent JP (1996) Uncoupling cadherin-based adhesion from wingless signalling in *Drosophila*. Nature 383: 627–630.
21. Gellon G, Harding KW, McGinnis N, Martin MM, McGinnis W (1997) A genetic screen for modifiers of Deformed homeotic function identifies novel genes required for head development. Development 124: 3321–3331.
22. Mahaffey JW, Griswold CM, Cao QM (2001) The *Drosophila* genes *disconnected* and *disco-related* are redundant with respect to larval head development and accumulation of mRNAs from Deformed target genes. Genetics 157: 225–236.
23. Schmidt-Ott U, Gonzalez-Gaitan M, Jackle H, Technau GM (1994) Number, identity, and sequence of the *Drosophila* head segments as revealed by neural elements and their deletion patterns in mutants. Proc Natl Acad Sci U S A 91: 8363–8367.
24. Hasson P, Muller B, Basler K, Paroush Z (2001) Brinker requires two corepressors for maximal and versatile repression in Dpp signalling. EMBO J 20: 5725–5736.
25. Andrioli LP, Oberstein AL, Corado MS, Yu D, Small S (2004) Groucho-dependent repression by Sloppy-paired 1 differentially positions anterior pair-rule stripes in the *Drosophila* embryo. Dev Biol 276: 541–551.
26. Jimenez G, Paroush Z, Ish-Horowicz D (1997) Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. Genes Dev 11: 3072–3082.
27. Tolkunova EN, Fujioka M, Kobayashi M, Deka D, Jaynes JB (1998) Two distinct types of repression domain in Engrailed: one interacts with the Groucho corepressor and is preferentially active on integrated target genes. Mol Cell Biol 18: 2804–2814.
28. Mondal S, Ivanchuk SM, Rutka JT, Boulianne GL (2007) Sloppy paired 1/2 regulate glial cell fates by inhibiting Gcm function. Glia 55: 282–293.
29. Markstein M, Markstein P, Markstein V, Levine MS (2002) Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. Proc Natl Acad Sci U S A 99: 763–768.
30. Dermitzakis ET, Clark AG (2002) Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. Mol Biol Evol 19: 1114–1121.
31. Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, et al. (2002) Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. Proc Natl Acad Sci U S A 99: 757–762.
32. Poirot O, O'Toole E, Notredame C (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res 31: 3503–3506.
33. Moretti S, Armougom F, Wallace IM, Higgins DG, Jongeneel CV, et al. (2007) The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. Nucleic Acids Res 35: W645–648.
34. Granderath S, Bunse I, Klambt C (2000) Gcm and Pointed synergistically control glial transcription of the *Drosophila* gene *loco*. Mech Dev 91: 197–208.
35. Liu QX, Jindra M, Ueda H, Hiromi Y, Hirose S (2003) *Drosophila* MBF1 is a co-activator for *tracheae Defective* and contributes to the formation of tracheal and nervous systems. Development 130: 719–728.
36. Kirkpatrick H, Johnson K, Laughon A (2001) Repression of Dpp targets by binding of Brinker to Mad sites. J Biol Chem 276: 18216–18222.
37. Akiyama Y, Hosoya T, Poole AM, Hotta Y (1996) The Gcm-motif: a novel DNA-binding motif conserved in *Drosophila* and mammals. Proc Natl Acad Sci U S A 93: 14912–14916.
38. Dalton D, Chadwick R, McGinnis W (1989) Expression and embryonic function of *empty spiracles*: a *Drosophila* homeo box gene with two patterning functions on the anterior-posterior axis of the embryo. Genes Dev 3: 1940–1956.
39. Sivasankaran R, Vigano MA, Muller B, Affolter M, Basler K (2000) Direct transcriptional control of the Dpp target *omb* by the DNA binding protein Brinker. EMBO J 19: 6162–6172.
40. Yu Y, Yussa M, Song J, Hirsch J, Pick L (1999) A double interaction screen identifies positive and negative *ftz* gene regulators and Ftz-interacting proteins. Mech Dev 83: 95–105.
41. Ryoo HD, Marty T, Casares F, Affolter M, Mann RS (1999) Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. Development 126: 5137–5148.
42. Jones BW, Fetter RD, Tear G, Goodman CS (1995) *glial cells missing*: a genetic switch that controls glial versus neuronal fate. Cell 82: 1013–1023.
43. Hosoya T, Takizawa K, Nitta K, Hotta Y (1995) *glial cells missing*: a binary switch between neuronal and glial determination in *Drosophila*. Cell 82: 1025–1036.
44. Minami M, Kinoshita N, Kamoshida Y, Tanimoto H, Tabata T (1999) *brinker* is a target of Dpp in *Drosophila* that negatively regulates Dpp-dependent genes. Nature 398: 242–246.
45. Jazwinska A, Kirov N, Wieschaus E, Roth S, Rushlow C (1999) The *Drosophila* gene *brinker* reveals a novel mechanism of Dpp target gene regulation. Cell 96: 563–573.
46. Campbell G, Tomlinson A (1999) Transducing the Dpp morphogen gradient in the wing of *Drosophila*: regulation of Dpp targets by Brinker. Cell 96: 553–562.
47. Hamaguchi T, Yabe S, Uchiyama H, Murakami R (2004) *Drosophila Tbx6*-related gene, *Dorsocross*, mediates high levels of Dpp and Scw signal required for the development of amnioserosa and wing disc primordium. Dev Biol 265: 355–368.
48. Reim I, Lee HH, Frasch M (2003) The T-box-encoding *Dorsocross* genes function in amnioserosa development and the patterning of the dorsolateral germ band downstream of Dpp. Development 130: 3187–3204.
49. Cohen SM, Jurgens G (1990) Mediation of *Drosophila* head development by gap-like segmentation genes. Nature 346: 482–485.
50. Walldorf U, Gehring WJ (1992) *empty spiracles*, a gap gene containing a homeobox involved in *Drosophila* head development. EMBO J 11: 2247–2259.
51. Grieder NC, Marty T, Ryoo HD, Mann RS, Affolter M (1997) Synergistic activation of a *Drosophila* enhancer by HOM/EXD and DPP signaling. EMBO J 16: 7402–7410.
52. Oliveri P, Davidson EH (2004) Gene regulatory network controlling embryonic specification in the sea urchin. Curr Opin Genet Dev 14: 351–360.
53. Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, et al. (2007) Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. Genes Dev 21: 385–390.
54. Pederson JA, LaFollette JW, Gross C, Veraksa A, McGinnis W, et al. (2000) Regulation by homeoproteins: a comparison of Deformed-responsive elements. Genetics 156: 677–686.
55. Mann RS, Carroll SB (2002) Molecular mechanisms of selector gene function and evolution. Curr Opin Genet Dev 12: 592–600.
56. Chan SK, Ryoo HD, Gould A, Krumlauf R, Mann RS (1997) Switching the *in vivo* specificity of a minimal Hox-responsive element. Development 124: 2007–2014.
57. Chan SK, Jaffe L, Capovilla M, Botas J, Mann RS (1994) The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with Extradenticle, another homeoprotein. Cell 78: 603–615.
58. Merabet S, Pradel J, Graba Y (2005) Getting a molecular grasp on Hox contextual activity. Trends Genet 21: 477–480.
59. Alexandre C, Vincent JP (2003) Requirements for transcriptional repression and activation by Engrailed in *Drosophila* embryos. Development 130: 729–739.
60. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al. (2003) A protein interaction map of *Drosophila melanogaster*. Science 302: 1727–1736.
61. Qi D, Bergman M, Aihara H, Nibu Y, Mannervik M (2008) Drosophila Ebi mediates Snail-dependent transcriptional repression through HDAC3-induced histone deacetylation. EMBO J 27: 898–909.
62. Subramanian T, Chinnadurai G (2003) Association of class I histone deacetylases with transcriptional corepressor CtBP. FEBS Lett 540: 255–258.
63. Heitzler P, Bourouis M, Ruel L, Carteret C, Simpson P (1996) Genes of the *Enhancer of split* and *achaete-scute* complexes are required for a regulatory loop between Notch and Delta during lateral signalling in *Drosophila*. Development 122: 161–171.
64. Stanewsky R, Rendahl KG, Dill M, Saumweber H (1993) Genetic and molecular analysis of the X chromosomal region 14B17-14C4 in *Drosophila melanogaster*: loss of function in NONA, a nuclear protein common to many cell types, results in specific physiological and behavioural defects. Genetics 135: 419–442.
65. Grossniklaus U, Pearson RK, Gehring WJ (1992) The *Drosophila sloppy paired* locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. Genes Dev 6: 1030–1051.
66. Guillen I, Mullor JL, Capdevila J, Sanchez-Herrero E, Morata G, et al. (1995) The function of *engrailed* and the specification of *Drosophila* wing pattern. Development 121: 3447–3456.
67. Schock F, Reischl J, Wimmer E, Taubert H, Purnell BA, et al. (2000) Phenotypic suppression of *empty spiracles* is prevented by *buttonhead*. Nature 405: 351–354.
68. Riechmann V, Irion U, Wilson R, Grosskortenhaus R, Leptin M (1997) Control of cell fates and segmentation in the *Drosophila* mesoderm. Development 124: 2915–2922.
69. Eulenberg KG, Schuh R (1997) The *tracheae defective* gene encodes a bZIP protein that controls tracheal cell movement during *Drosophila* embryogenesis. EMBO J 16: 7156–7165.
70. Aza-Blanc P, Ramirez-Weber FA, Laget MP, Schwartz C, Kornberg TB (1997) Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. Cell 89: 1043–1053.
71. Soustelle L, Jacques C, Altenhein B, Technau GM, Volk T, et al. (2004) Terminal tendon cell differentiation requires the Glide/Gcm complex. Development 131: 4521–4532.
72. Rushlow C, Colosimo PF, Lin MC, Xu M, Kirov N (2001) Transcriptional regulation of the *Drosophila* gene *zen* by competing Smad and Brinker inputs. Genes Dev 15: 340–351.
73. Barolo S, Carver LA, Posakony JW (2000) GFP and β-galactosidase transformation vectors for promoter/enhancer analysis in *Drosophila*. Biotechniques 29: 726–732.
74. Bergson C, McGinnis W (1990) An autoregulatory enhancer element of the *Drosophila* homeotic gene *Deformed*. EMBO J 9: 4287–4297.
75. Tautz D, Pfeifle C (1989) A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. Chromosoma 98: 81–85.
76. Kosman D, Mizutani CM, Lemons D, Cox WG, McGinnis W, et al. (2004) Multiplex detection of RNA expression in *Drosophila* embryos. Science 305: 846.