Understanding the logic of TCF target location in *Drosophila* Wnt signaling

by

Hilary C. Archbold

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2014

Doctoral Committee:

Professor Kenneth M. Cadigan, Chair
Associate Professor Scott E. Barolo
Associate Professor Gyorgyi Csankovszki
Professor James Douglas Engel
Associate Professor Jorge A. Iñiguez-Lluhí

"Did you get your questions answered?"
"Yes. And every answer led to a bigger question."

—**Ultra Magnus** and **Optimus Prime**

--The Transformers

© Hilary Archbold
All rights reserved
2014

To my amazing family.

Acknowledgements

Science, like life, is a collaborative enterprise. This thesis represents the culmination of 6 years of effort on my part, but I could not have done it without the support, encouragement and freely shared knowledge of many, many people whom I hold dear.

My decision to attend grad school and work in *Drosophila* was in no small part due to the fantastic Victoria Meller at Wayne State University. My first research experience in her lab was addictive, and her guidance was invaluable. I am eternally grateful to have been given the opportunity to do my thesis research at an outstanding institution, the University of Michigan. I would like to thank my mentor, Ken Cadigan for giving me the opportunity to work on such an interesting project. As a mentor, Ken has consistently been generous with his time and his knowledge, and I am grateful for everything he has taught me in my time here.

I also would like to thank all the current and former members of the Cadigan Lab whom I have had the pleasure of working with over the years. I was lucky to get to share a bay first with Yunyun Ni and then Lisheng Chen (before I defected to the warmer climes of the annex). Aditi Ravindranath, Yaxuan Yang, Chen Zhang, and Chandan Bhambhani were always willing to discuss a problem, or offer advice, and Aditi and Yaxuan deserve special mention for proofreading this thesis. Aaron Buckingham and Pete Burby always made the lab run a little smoother. Mikyung Chang left for

Korea almost immediately after I joined the lab, but I would like to thank her for keeping the most immaculately organized research notebooks I have ever seen. The answer to many a quandary was found there, and my notebooks are probably far more legible now than they would have been without her shining example. Tim Blauwkamp and Jenni Kennell are lab members from before my time, who I had the pleasure of getting to know at conferences across the country. Their advice and scientific discussions have always been helpful. I would also like to thank Nicholas Tolwinski for advice, especially that which helped me out of the “Invert repeat cloning quagmire of oh-nine,” and Chris Broussard for his programming acumen.

I feel extremely lucky to have had the committee that I do. Their questions, comments and suggestions have helped to steer my project, as well as my growth as a scientist. Scott Barolo, Gyorgyi Csankovszki, James Douglas Engel, and Jorge A. Iñiguez-Lluhí each brought a different perspective to our meetings, and this project wouldn’t be what it is without their guidance. As Gyorgyi’s office is adjacent to Ken’s, I would also like to thank her for all the extra advice she has dispensed over the years, and thanks also to Laura Buttitta, who has done the same.

The community of Drosophila Labs at University of Michigan is fantastic. All of the labs are helpful and generous. I would specifically like to mention Andrea, Nikki and Lisa from the Barolo lab, Kerry, Yiqin, Kiki, and Dan from the Buttitta Lab, Gizem, Elizabeth and Lisa from the Wittkopp lab, Harper and Katie from the Shafer Lab, and Jill, Charlie and Derek from the Lee Labs, who over the years have been especially helpful to me personally. Not a fly lab, but also fantastic, are our neighbors in the Csankovszki Lab, especially Martha Snyder, who not only deserves thanks for

proofreading this thesis (all British spellings are my affectation, not hers), but also for being a sounding board, and a true friend. Work is a little more fun when Martha is around to talk to.

As a CMB student in an MCDB lab, I would like to thank both Cathy Mitchell from CMB and Mary Carr from MCDB for keeping all of my ducks in a row, and making sure I didn't starve. Both Jessica Schwartz and Bob Fuller have been excellent directors of CMB during my tenure, and I am glad I was a part of CMB.

I would also like to thank my family and friends for all of their love and support during this journey. I have a marvelous set of friends, but I'd like to thank Fran and Doran Steele specifically for going above and beyond, and for doing more extra-curricular ferrying of my child than I do. The ladies of Book club, and all the Gwaffle day regulars, especially my dear friends the Corvidaes, also deserve a mention for taking my mind off science occasionally. I couldn't have made it through the past six years without my incredible family. My daughter Octavia has been coming to lab and filling pipette tip boxes on snow days since before she could print. My amazing husband Zach has been my rock. He believes in me even when I don't believe in myself. He once drove 9 hours straight without complaining, so that I could finish a last minute grant application that was due right in the middle of our scheduled vacation on the Appalachian Trail. I don't know what I would do without them. My Mom and Dad, who are always proud of me, have spent every summer up here to help make life a little less crazy. It took me a long time to get to this place in life, and my parents have been supportive of me the whole way, for which I am forever grateful. I would also like to thank the rest of my loving and far-flung family, (all of the Archbolds, Hollers, Nagles,

and Taylors) and especially the Toledo contingent: May Towey, and Anne and Randy Ruch, who have been the source of excellent holiday dinners and great conversation. I love you all.

Table of Contents

List Of Figures

List of Tables

Chapter I:

General Introduction

I. Wnt signaling

Wnts are a family of secreted proteins that can exert profound influences on cell behaviour through activation of several signaling pathways. By interacting with multiple classes of cell surface molecules, Wnt ligands activate diverse downstream effectors. Wnt/β-catenin signaling, sometimes called 'canonical' Wnt signaling, is the best studied of these pathways. This particular Wnt pathway acts by increasing levels of nuclear β-catenin, which serves as a co-regulator for transcription factors (TFs) such as the TCF/LEF family of TFs. TCFs recruit β-catenin to specific regulatory elements (Cadigan, 2008; Cadigan and Peifer, 2009; MacDonald et al., 2009) to modulate target gene activity. "Non-canonical" signaling, in contrast, refers to pathways which act in a β-catenin independent fashion, and include the planar cell polarity, Ca++, and receptor tyrosine kinase pathways (De, 2011; Petrova et al., 2014; van Amerongen, 2012).

Wnt/β-catenin signaling is known to play many pivotal roles in animal development (Grigoryan et al., 2008; Logan and Nusse, 2004; Niehrs, 2010; Petersen and Reddien, 2009) and adult tissue maintenance (Haegebarth and Clevers, 2009; Nusse et al., 2008; Polakis, 2007; Wend et al., 2010). In addition, aberrant Wnt signaling has been linked to several human diseases, most notably (but not restricted

to) several cancers (Clevers, 2006; Polakis, 2012). The ability of this pathway to activate diverse transcriptional programs in different contexts is remarkable, and the mechanism by which transcriptional targets are identified and regulated is a primary focus of this thesis work.

Overview of Wnt/β-catenin signaling

The stability and nuclear localization of β-catenin protein is thought to play a central role in determining the level of Wnt/β catenin signaling. In the absence of Wnt stimulation, β-catenin is constitutively targeted for degradation by a complex (termed the β-catenin destruction complex, or destructosome) containing glycogen synthase kinase 3 (GSK3) and casein kinase I (CKI), as well as the scaffolding proteins Axin and adenomatous polyposis coli (APC) protein (Cadigan and Peifer, 2009; Kennell and Cadigan, 2009). Phosphorylation of specific residues in the N-terminus of β-catenin by CKI and GSK3 is followed by ubiquitination and proteosomal degradation (Cadigan and Peifer, 2009; MacDonald et al., 2009), (Figure 1.1A). In addition, Axin and APC are thought to sequester β-catenin in the cytosol, and/or to promote β-catenin efflux from the nucleus (Brocardo and Henderson, 2008). Without Wnt signaling, the β-catenin destruction complex keeps the level of β-catenin low, restricting β-catenin to its essential role in supporting cadherin-mediated cell adhesion (Stepniak et al., 2009).

When Wnt protein is recognized at the cell surface by members of the Frizzled (Visel et al.) family of proteins and low density lipoprotein receptor related protein 5 or 6 (LRP 5/6), a large complex termed the 'Wnt signalosome' is formed (Cadigan and Peifer, 2009; MacDonald et al., 2009). The signalosome interacts with the β-catenin

destruction complex, inhibiting its activity. Although the mechanism by which the complex is inhibited is still under debate (Li et al., 2012; Taelman et al., 2010), it is agreed that this event results in the accumulation of β-catenin, some of which enters the nucleus (Figure 1.1B).

Once in the nucleus, β-catenin can bind to several DNA-binding proteins, the best understood of which are the TCF/LEF family transcription factors. Some TCFs are thought to act as a transcriptional switch, repressing Wnt target gene expression in the absence of signaling and activating transcription upon forming a complex with β-catenin (Figure 1.1B). This switch activity requires interactions with additional co-activators and co-repressors, several of which alter the state of Wnt target gene chromatin (Arce et al., 2006; Cadigan, 2012; Mosimann et al., 2009; Willert and Jones, 2006).

Wnt signaling in *Drosophila*

The model organism *Drosophila melanogaster* provides the advantage of a simplified system to study Wnt signaling. In vertebrates, gene duplication and functional divergence of many components of the Wnt pathway results in a more complicated system in which partial redundancy between similar, but not identical, family members can obscure issues of necessity and sufficiency. While humans have 19 reported Wnt proteins, in *Drosophila*, only 7 have been identified (Chien et al., 2009). Wingless (Wg) was first identified in a EMS mutagenesis screen in which a hypomorphic allele exhibited wing and haltere defects with variable penetrance in adult flies (Sharma and Chopra, 1976). In addition, bristle, scutellar, hair, and thoracic defects were noted.

After homology to the mouse *int1* gene was recognized (Rijsewijk et al., 1987), the contraction Wnt (Wg + int) was given to the growing family of secreted proteins. The role of Wingless as a segment polarity gene during embryonic patterning was identified by Nusslein-Vollhard and Wiechaus (Nusslein-Volhard and Wieschaus, 1980). Wg is the best studied Wnt in *Drosophila* development. It is required for both embryonic and later development, and its loss results in lethality (Baker, 1988b). Wg is required for the segmental pattern of alternating denticle and naked cuticle in the larvae (Bejsovec and Martinez Arias, 1991), and Wg gradients are crucial for development of adult structures from larval imaginal discs such as the wing (Cadigan, 2002).

There is a single TCF/LEF family transcriptional effector in *Drosophila,* called dTCF, TCF or *pangolin* (TCF/Pan) (Clevers and van de Wetering, 1997; van de Wetering et al., 1997). The fly β-catenin, Armadillo (Arm), promotes activation by binding TCF and displacing co-repressors such as Groucho (a homolog of Transducin-like Enhancer of split 2) (Parker DS, 2007). Armadillo then forms a complex with co-activators such as Legless (Lgs) and Pygopus (Pygo) (Parker et al., 2002). Pygopus was shown to bind subunits of the *Drosophila* Mediator complex (Carrera et al., 2008), providing a mechanistic link between distant TCF bound Wingless Cis-Regulatory Modules (W-CRMs) and the Transcriptional Start Site (TSS) of Wg target genes.. Physical interactions have also been shown with a number of co-factors which positively or negatively affect transcription by altering local chromatin structure (Cadigan and Waterman, 2012) (Parker DS, 2007). Recruitment of the Histone Acetyltransferase (HAT) CBP by the Arm/TCF complex results in histone acetylation and chromatin

remodeling at target loci preceding transcriptional activation (Parker et al., 2008). In contrast, ISWI and ACF1 negatively regulate Wg targets by antagonizing histone H4 acetylation, and localizing TCF to negatively regulated genes (Liu et al., 2008). Chromatin remodelers *osa* and *brahma* have also been linked to target gene repression (Parker DS, 2007). However, even in the "simplified" *Drosophila* system complexities abound. In addition to its positive role in target gene regulation, CBP also has a demonstrated negative effect on TCF activity (Li et al., 2007; Waltzer and Bienz, 1998), and the co-factor CtBP has been shown to potentiate both activation and repression of Wg target genes (Fang et al., 2006) depending upon its oligomeric state (Bhambhani et al., 2011). In flies, there is also evidence of a reverse transcriptional switch in which TCF and Arm work together to directly repress some targets (Blauwkamp et al., 2008) (Zhang et al., 2014), although in this case the cofactors involved are unknown.

II. Wnt/β-catenin signaling acts in diverse and sometimes contradictory roles in development and disease

Wnt/β−catenin signaling plays an essential role in promoting cell proliferation during development, and inappropriately hyperactivated Wnt signaling is linked to tissue hyperplasia and a variety of cancers (Cadigan and Peifer, 2009; Polakis, 2012). However, the pathway is more than just a proliferative signal. Early in development, a gradient of Wnt/β-catenin activity is required for specification of the anterior/posterior axis (Niehrs, 2010). Later, Wnt/β-catenin signaling is important for a multitude of developmental decisions. These include limb formation, bone, hair and teeth

development, and the formation of every major organ (Cadigan and Nusse, 1997; Clevers, 2006; Gessert and Kuhl, 2010; Liu and Millar, 2010; Regard et al., 2012; Tanaka et al., 2011; Towers et al., 2012). In adult organisms, the pathway is required for tissue homeostasis, stem cell maintenance, and wound healing (Arwert et al., 2012; Lim and Nusse, 2013; Whyte et al., 2012). How Wnt/β-catenin signaling controls such a broad range of processes likely requires the ability to regulate multiple, distinct transcriptional programs in a time and tissue specific fashion. Some direct targets of the pathway have been identified in many of these processes. However, it is clear that a deeper understanding of how target genes are regulated in a coordinated fashion is needed to understand Wnt activity.

In the following section, only a small subset of Wnt roles in development and disease will be discussed. Heart formation provides a good example of how the Wnt/β-catenin pathway is used repeatedly to achieve different outcomes in different developmental contexts, even in the same tissue. Stem cell biology, tissue homeostasis and regeneration are interrelated processes, subject to Wnt regulation, in which the direct targets are only partially known. A third example, concerning the link between Wnt/ β-catenin signaling and oxidative stress in the context of diabetes and neurodegeneration will also be discussed.

Cardiogenesis: reiterative Wnt/β-catenin signaling required

Although the tube-like insect heart is morphologically different from the multichambered vertebrate heart, they appear to share a common ancestry. One

striking example supporting this view is provided by the tinman/Nkx2.5 gene, which is required for heart formation in *Drosophila* and several vertebrate systems (Bodmer and Venkatesh, 1998; Evans, 1999) In flies, Wg/Arm signaling is required for *tinman* expression and heart formation in general (Park et al., 1996; Wu et al., 1995). In contrast, the first studies in vertebrates found that Wnt/β-catenin signaling is required to restrict specification of cardiac mesoderm (Lickert et al., 2002; Marvin et al., 2001; Schneider and Mercola, 2001).

This paradox between invertebrate and vertebrates has been resolved by the realization that the Wnt/β-catenin pathway plays opposing roles in early vertebrate cardiogenesis. Temporal control of Wnt expression revealed that before gastrulation, the pathway activates Nkx2.5 expression in zebrafish embryos (Ueno et al., 2007). After gastrulation, the previously described inhibitory role was evident. This biphasic relationship was also observed in mouse embryonic stem (ES) cells, which spontaneously differentiate into cardiomyocytes. Activation of Wnt/β-catenin signaling in early cultures dramatically enhanced cardiomyocyte differentiation but pathway activation in later cultures reduced cardiogenesis (Ueno et al., 2007). These data suggest that the initial pro-cardiogenic effect of the Wnt/β-catenin pathway in vertebrates is analogous to the positive effect that Wg/Arm signaling has on heart development in the fly.

What are some of the direct targets in the initial regulation of cardiogenesis by the pathway? In flies, the presence of functional TCF binding sites in the enhancers that drive the expression of the transcription factors Sloppy paired 1 (Slp1) and Even-

skipped (Eve) in the cardiac precursors provides strong evidence that these genes are directly activated by Wg/Arm signaling (Halfon et al., 2000; Han et al., 2002; Knirr and Frasch, 2001; Lee and Frasch, 2000; Ueno et al., 2007). In mouse ES cells, gene profiling revealed many potential targets, including Brachyury, Mesp1 and Sox17 (Liu et al., 2007; Ueno et al., 2007). In regard to repression of cardiogenesis, the homeodomain protein Hex is an important Wnt target that is repressed in the presumptive cardiac mesoderm (Foley and Mercola, 2005). In Xenopus, GATA6 expression is repressed by Wnt/β-catenin signaling, and forced expression of GATA6 is sufficient to rescue many aspects of heart development that are disrupted by ectopic Wnt pathway activation (Afouda et al., 2008). The mechanism by which these genes are repressed by Wnt/β-catenin signaling is not known.

After the initial specification of the presumptive heart field, a population of cardiomyocytes known as the secondary heart field (SHF) will give rise to the future right ventricle and inflow and outflow tracts of the heart (Dyer and Kirby, 2009). Conditional knockout of β-catenin causes a significant reduction in these structures (Ai et al., 2007; Cohen et al., 2007; Klaus et al., 2007; Kwon et al., 2007; Tian et al., 2010). Removal of β-catenin before SHF specification resulted in a loss of Islet1 (a LIM homeodomain) expression, a marker of SHF cells (Cai et al., 2003). There is some evidence to indicate that the activation of Islet1 transcription by Wnt/β-catenin signaling is direct (Lin et al., 2007). Later removal of β-catenin in the developing SHF resulted in normal Islet1 expression, but the right ventricle and outflow tracts still failed to form (Ai et al., 2007; Cohen et al., 2007; Kwon et al., 2007). These results indicate multiple roles

for Wnt/β-catenin signaling in SHF development, both in establishing the SHF and subsequent differentiation of heart tissue (Figure 1.2).

The complexity of target gene regulation by the Wnt/β-catenin pathway in the development of the secondary heart field is evidenced by the findings that Wnt/β-catenin signaling represses Islet1 expression in differentiating cardiomyocytes (Kwon et al., 2009). Likewise, GATA6, which is repressed by Wnt/β-catenin signaling in early heart development in Xenopus (Afouda et al., 2008), is directly activated by Wnt/β-catenin signaling in the posterior SHF (Tian et al., 2010). The factors that enable one pathway to both activate and repress the same targets during the cardiomyocyte cell lineage are not known.

There are several additional events in heart development where Wnt/β-catenin signaling is also required. Mutations lowering pathway activity in cardiac NC cells result in defects to the cardiac outflow tract (Hamblet et al., 2002; Kioussi et al., 2002). This phenotype is similar to that observed when the gene encoding the bicoid homeodomain protein PitX2 is mutated (Kioussi et al., 2002). PitX2 is directly activated by LEF1 and β-catenin. PitX2 subsequently recruits β-catenin to the Cyclin D2 regulatory region, activating this cell cycle regulator and promoting proliferation of the cardiac NC cells (Kioussi et al., 2002). The Wnt/β-catenin pathway also promotes endocardial cell proliferation, which contributes to heart valve formation (Alfieri et al.; Gitler et al., 2003; Hurlstone et al., 2003; Liebner et al., 2004). Likewise, loss of β-catenin in the developing epicardium results in defects in coronary artery formation (Zamora et al., 2007).

Heart development highlights the multiple roles that the Wnt/β-catenin pathway plays in regulating cell fate and organogenesis (Figure. 1.2). How can one signaling pathway be utilized so many times to regulate different genes in the cardiac cell lineage? Understanding how Wnt transcriptional output diversity is generated will require a more detailed understanding of how TCF family members, and other transcription factors that mediate Wnt/β-catenin-dependent gene regulation, function on target gene chromatin.

Wnt/β-catenin signaling in stem cell biology and regeneration

The role of Wnt/β-catenin signaling in stem cell maintenance has been well documented in the intestine, hair follicles and in the skin (Barker and Clevers, 2010; Blanpain and Fuchs, 2009; Blanpain et al., 2007; Haegebarth and Clevers, 2009; Pinto and Clevers, 2005). In the mouse small intestine, TCF4 and β-catenin are required for maintenance of the crypt stem cells (Pinto and Clevers, 2005). Microarray profiling has identified many transcriptional targets of the Wnt/β-catenin pathway (van de Wetering et al., 2002; Van der Flier et al., 2007). One biologically important target is *c-myc*, which is required for normal intestinal crypt development (Muncan et al., 2006). C-myc appears to be a direct target of the pathway. Another gene that is activated by Wnt/β-catenin signaling encodes the orphan receptor Lgr5, which has received a great deal of attention, since it marks a population of multipotent cells in the crypt that give rise to all the specialized cells of the intestinal epithelia (Barker and Clevers). Another biologically important Wnt target is the bHLH transcription factor Achaete scute-like 2 (Ascl2), which

is required for the maintenance of the Lgr5-positive stem cells and can induce crypt hyperplasia upon forced expression (van der Flier et al., 2009). Both *lgr5* and *ascl2* were identified in a genome-wide survey of chromatin enriched for TCF4 binding (Hatzis et al., 2008), suggesting they may be direct targets of Wnt/β-catenin signaling. In addition to intestinal stem cells, Lgr5 and the related protein Lgr6 also appear to mark stem cell populations in several other organs (Haegebarth and Clevers, 2009; Snippert et al., 2010).

Wnt/β-catenin signaling has also been linked to the ability to replace damaged cells or to regenerate deleted tissues in a wide array of metazoans. In Hydra and Planaria, organisms renowned for their ability to regenerate large portions of their bodies after bisection, the pathway promotes the development of posterior cell fates, recapitulating its early developmental role in anterior/posterior (A/P) axis formation. In Hydra, decapitation of the head results in rapid induction of Wnt3 expression in the epithelial cells which is required for head regeneration (Chera et al., 2009a; Hobmayer et al., 2000; Lengfeld et al., 2009). It should be noted that the Hydra are unusual in that the head is a posterior, and not anterior, structure (Niehrs, 2010). When Hydra is bisected in the midgastric region, a wave of apoptosis among the interstitial cells is coupled to release of Wnt3 from these dying cells, which results in subsequent activation of Wnt3 transcription in epithelial and head regeneration (Chera et al., 2009b; Galliot and Chera, 2010; Niehrs, 2010).

The coupling of apoptosis to generation of a Wnt signal in regenerating Hydra is reminiscent of a similar phenomenon described in *Drosophila* imaginal discs, where

apoptotic cells express Wg. This expression is thought to promote compensatory proliferation of neighbouring cells to maintain the size of the tissue (Fan and Bergmann, 2008; Martin et al., 2009). While Wg is dispensable for disc repair in response to irradiation (Perez-Garijo et al., 2009), a functional role for Wg has been reported in disc regeneration following expression of a pro-apoptotic signal. In this case, Wg induces expression of *myc* and *cyclinE* to promote proliferation (Smith-Bolton et al., 2009).

In vertebrates, Wnt/β-catenin signaling has been shown to be required for tail fin regeneration in zebrafish (Stoick-Cooper et al., 2007) Wnt10a is induced after fin cutting, and blocking the pathway prevents induction of *fibroblast growth factor 20a* (*fgf20a*) expression, which is required for regeneration of this tissue (Wills et al., 2008). Interestingly, *fgf20a* is directly activated by TCF-β-catenin in cultured human cells (Chamorro et al., 2005), though it is not clear whether this is the case in regenerating zebrafish tails. The pathway is also thought to play a role in the initial step of limb regeneration in Xenopus (Yokoyama et al., 2007), though the molecular targets remain to be identified.

Wnts and oxidative stress: diabetes and neurodegenerative disorders

While the direct links between metabolic disorders or neurodegenerative diseases and Wnt/β-catenin signaling are not as well established as in cancer, there are some candidate targets where misregulation of the pathway could underlie the pathology. Emerging evidence has linked several players in Wnt/β-catenin signaling to metabolic disorders and type 2 diabetes mellitus. Genome-wide association studies

demonstrated a strong connection between diabetes type 2 risk and single nucleotide polymorphisms (SNPs) within the TCF7L2 (TCF4) gene, although the cellular basis of this association is still uncertain (Schinner et al., 2009). Mutations in the Wnt receptors LRP5 and 6 have also been implicated in obesity and type 1 diabetes and metabolic syndrome respectively (Jin, 2008). Wnt/β-catenin signaling appears to play roles both in the proliferation of pancreatic β-cells, and in the insulin release from islet cells. β-cell proliferation in cell culture and transgenic mouse models appear to be the result of Wnt activation of targets like cyclin D1 (Liu and Habener, 2008; Schinner et al., 2009), as well as cyclin D2 and PitX2 (Rulifson et al., 2007). Consistent with this, knockdown of TCF4 decreases β-cell proliferation and promotes apoptosis (Liu and Habener, 2008; Shu et al., 2008).

In addition to promoting β-cell proliferation, the Wnt/β-catenin pathway is also required for efficient insulin secretion (Fujino et al., 2003), in part through direct regulation of *proglucagon* (*pgc*) transcription (Ni et al., 2003; Shao et al., 2013; Yi et al., 2005). The pathway may play a role in insulin sensing by activating glucokinase transcription (Schinner et al., 2008). ChIP-seq (Chromatin immunoprecipitation and sequencing) of hepatocytes identified bound regions near a number of genes implicated in gluconeogenesis including *Fbp1, pck1* and *G6pc*, as well as a large set of genes implicated in glucose metabolism (Norton et al., 2011). It is interesting to note that antagonism of Wnt signaling by oxidative stress appears to play an important role in the pathology of insulin resistance and diabetes. The forkhead box DNA-binding protein FOXO, has been shown to compete with TCFs for β-catenin binding (Hoogeboom et al.,

2008). Upregulation of FOXOs in response to oxidative stress may contribute to insulin resistance by the promotion of gluconeogenesis and/or the promotion of apoptosis and downregulation of TCF-mediated gene expression (Manolagas and Almeida, 2007).

The effects of oxidative stress may also be a factor in the neurodegeneration of Alzheimer's disease (AD), where Wnt signaling has been ascribed a neuroprotective role (Manolagas and Almeida, 2007). Extracellular accumulation of amyloid-β (A-β) has been shown to bind Frizzled (Fz) receptors and downregulate Wnt signaling. It has been hypothesized that some of the cytotoxity caused by A-β may be the result of chronic suppression of Wnt/β-catenin signaling (Inestrosa and Toledo, 2008). The protective effects of β-catenin overexpression, but not of transcriptionally inactive β-catenin, indicate that transcriptional activation of target genes plays a role in ameliorating A-β toxicity (Chacon et al., 2008). One possibility is that oxidative stress may exacerbate cytotoxity in AD by increasing FOXO, which then competes with TCF for β-catenin, reducing the expression of anti-apoptotic factors.

A similar neuroprotective role for the Wnt/beta-catenin pathway has been described in the midbrain dopaminergic neurons affected in Parkinson's disease (Berwick and Harvey, 2012; L'Episcopo et al., 2014b). Here again, oxidative stress and inflammation have been implicated in decreased neuronal survival, and increased levels of β-catenin are thought to increase neuronal survival by defending against oxidative stress and increasing transcription of pro-survival genes (L'Episcopo et al., 2014a). The identity of these direct targets in adult neurons is still the subject of investigation, but a combination of bioinformatic analysis and ChIP in rat thalamic neurons has identified a

number of potential targets (Wisniewska et al., 2012) required for neuronal excitability and function. Many of these targets may be regulated in the neuronal populations affected in Parkinson's or Alzheimer's, as well as those in amyotrophic lateral sclerosis (ALS), where disrupted Wnt signaling has also been implicated in disease pathology (Pinto et al., 2013).

Clearly, much remains to be elucidated concerning the role Wnt/β-catenin signaling plays in neurodegeneration and other diseases. Information gained from the study of Wnt-mediated transcriptional regulation in model systems could facilitate the identification of the important targets in many pathological states where genetics suggests that aberrant Wnt signaling plays a causal role. Given the prominence of TCF family members in regulating numerous Wnt targets, the next few sections will review TCF function in detail. Understanding how direct targets are identified by TCF can shed light on this complexity.

III. TCFs are major mediators of Wnt signaling

The TCF family of transcription factors are the best characterized DNA-binding regulators of Wnt/β-catenin target gene expression. In addition to a β-catenin binding domain at the N-terminus, a hallmark of this family is the presence of a highly conserved HMG domain, followed by a stretch of basic residues (Figure 1.3, 1.5). The TCF subfamily of HMG domains is found throughout metazoans, but not in the sister group choanoflagellates (King et al., 2008). The HMG domain contacts DNA largely through minor groove contacts and results in a large bending of the double helix (Love et al.,

1995). High affinity binding sites for these HMGs have been determined for *Drosophila* TCF/Pangolin (TCF/Pan) and all four mammalian TCFs (Atcha et al., 2007; Giese and Grosschedl, 1993; Hallikas et al., 2006; van Beest et al., 2000; van de Wetering et al., 1997). These studies show that a site of CCTTTGATS (S = G/C) is bound with highest affinity *in vitro*. Many functional TCF binding sites in Wnt cis-regulatory modules (W-CRMs) fit this consensus, while others diverge markedly.

There is a single TCF gene in almost all invertebrate species that have been examined thus far. TCF/Pan from flies and POP-1 from *C. elegans* are the most thoroughly characterized. In contrast, amphibians and mammals have four TCF genes. These are most commonly referred to as TCF1 (TCF7), LEF1 (LEF1), TCF3 (TCF7L1) and TCF4 (TCF7L2). The names in parentheses are from the Human Genome Organization (HUGO). In zebrafish, the TCF7L1 gene is duplicated (TCF7L1a and TCF7L1b), giving a total of five TCF genes (Dorsky et al., 2003). A second DNA binding domain, the C-Clamp, is present in most invertebrates, but only a subset of vertebrate family members (Figure 1.3, 1.5). As will be discussed below, there is evidence that some of the vertebrate TCFs are more specialized in their function compared to their invertebrate counterparts.

The TCF family: a historical perspective

The members of the TCF family of transcription factors were first discovered by researchers interested in lymphocyte gene regulation. A protein originally called TCF-1α or lymphoid enhancer-binding factor 1 (LEF1) was highly expressed in pre-B and pre-T

cells and bound a specific DNA sequence in an enhancer controlling the T-cell receptor α (TCRα) gene (Travis et al., 1991; Waterman et al., 1991; Waterman and Jones, 1990). Another protein enriched in immature T-cells called TCF1 bound to a similar sequence in a CD3a enhancer (van de Wetering et al., 1991). Both LEF1 and TCF1 were found to contain a single HMG domain, which was sufficient for DNA-specific binding (Giese et al., 1991; Oosterwegel et al., 1991; Waterman et al., 1991). TCF1 and LEF1 belong to a subgroup within a larger family of HMG domain containing proteins, and are most closely related to the subgroups containing the SOX proteins (e.g. SRY) and fungal mating type proteins (e.g. STE11) (Laudet et al., 1993).

In addition to sequence specific DNA binding, the HMG domain of LEF1 has been shown to bend DNA up to 130°, (Giese et al., 1992) which was confirmed by NMR based structural analysis of a LEF1–DNA-binding site complex (Love et al., 1995). This bending was proposed to play an architectural role in coordinating the binding of several other factors to the TCRα enhancer (Carlsson et al., 1993; Giese and Grosschedl, 1993; Giese et al., 1995). This protein–enhancer complex is sometimes referred to as the 'T cell enhanceosome' (Balmelle et al., 2004). The high degree of conservation between the HMG domains of the TCF family (Figure 1.3) suggests that all members have the ability to bend DNA, though this remains to be tested directly.

Although interest in LEF1 and TCF1 was initially focused on lymphocytes, the finding that mice lacking a functional LEF1 gene displayed defects to several organ systems indicated a much broader role in developmental biology (van Genderen et al., 1994). This connection was further solidified by the findings that LEF1 and TCF3

(another member of the family) could bind to β-catenin (Behrens et al., 1996; Huber et al., 1996; Molenaar et al., 1996). It was known that mutants in the fly homolog of β-catenin, *armadillo* (*arm),* caused defects very similar to *wingless* (*wg*) mutants, a Wnt gene important in many aspects of fly development (Noordermeer et al., 1994; Peifer et al., 1991; Riggleman et al., 1989; Siegfried et al., 1994). In addition, mis-expression of β-catenin in ventral blastomeres of *Xenopus* embryos induced a secondary body axis (Heasman et al., 1994), reminiscent of mis-expression of several Wnt genes (McMahon and Moon, 1989; Smith and Harland, 1991; Sokol et al., 1991).

The N-terminus of TCFs are required for binding to β-catenin, and deletion of this portion of a TCF gene produces a protein that can dominantly inhibit Wnt signaling in several organisms (Behrens et al., 1996; Kratochwil et al., 2002; Molenaar et al., 1996; van de Wetering et al., 1997). In addition, placing multiple copies of high affinity HMG binding sites upstream of a minimal promoter-reporter gene cassette results in reporter gene expression that is highly activated by Wnt/β-catenin signaling in a number of contexts (Korinek et al., 1997; Molenaar et al., 1996; van de Wetering et al., 1997). These now classic observations contribute to the current working model of Wnt target gene activation depicted in Figure 1.1.

The TCF transcriptional switch

The current working model for TCF-mediated regulation of W-CRMs proposes the existence of a transcriptional switch. TCF and co-repressors bind relevant W-CRMs and inhibit target gene expression in the absence of signaling, and then act with β-

catenin and other co-activators to activate transcription of targets (Figure 1.1, 1.4) after signal is sensed at the cell surface. Evidence for this model was first obtained in the *Drosophila* embryo. *TCF/Pan* mutants display mis-patterning of the epidermis indicative of a reduction in Wg signaling (Brunner et al., 1997; Schweizer et al., 2003; van de Wetering et al., 1997). This defect was not as severe as that of null alleles of *wg*. However, a *wg*; *TCF/Pan* double mutant looked identical to *TCF/Pan* mutants (Cavallo et al., 1998). In *wg* mutants, there is no activation of targets and TCF repression is intact, resulting in a severe loss of the Wg signaling phenotype. But in *wg; TCF/Pan* double mutants, loss of repression of Wg targets allows some expression (i.e. derepression), resulting in a less severe phenotype (Cavallo et al., 1998). Genetic and physical interactions between TCFs and TLE family co-repressors (Cavallo et al., 1998; Roose et al., 1998) provided further support for the transcriptional model. The model has also been confirmed in fly cell culture using a combination of RNAi and ChIP of TCF/Pan in the absence and presence of Wg signaling (Fang et al., 2006).

In *C. elegans*, the POP-1 loss-of-function phenotypes indicate both positive and negative roles in regulating Wnt targets (Phillips and Kimble, 2009). In some contexts, repression of Wnt targets is the predominant mechanism of action, for example, when blocking mesoderm cell fate in the early embryo (Rocheleau et al., 1997; Thorpe et al., 1997). But in other stages, e.g., QL neuroblast migration and stem cell specification in the somatic gonad, loss of POP-1 has a similar phenotype as loss of other Wnt/β-catenin components (Herman, 2001; Lam et al., 2006), indicating an activating role. Clearly, POP-1 and TCF/Pan can both repress and activate Wnt targets.

In vertebrates, some TCF family members are more closely linked to either repression or activation. For example, *headless* (*hdl*) mutants in zebrafish are loss of function TCF3a alleles and display a lack of head structures (Kim et al., 2000), similar to Dkk1 knockouts in mice (Mukhopadhyay et al., 2001). While *hdl* mutants could be efficiently rescued with a TCF3a transgene, adding the VP16 domain, a potent transcriptional activation domain to TCF3a, abolished rescue activity of the transgene (Kim et al., 2000). This suggests that TCF3a is functioning primarily as a transcriptional repressor. Knockout of TCF3 in mice also resulted in phenotypes associated with an increase in Wnt/β-catenin signaling (Merrill et al., 2004) and siRNA inhibiton of TCF3 in mouse ES cells largely results in increased expression of target genes (Cole et al., 2008). In contrast, LEF1 knockouts have phenotypes best explained by a loss of Wnt/β-catenin signaling (Kratochwil et al., 2002; Reya et al., 2000; van Genderen et al., 1994).

In contrast to TCF3 and LEF1, TCF1 appears to be more versatile. Mouse embryos lacking both TCF1 and LEF1 have a loss of caudal somites that is reminiscent of Wnt3a mutants (Galceran et al., 1999). Furthermore, TCF1 and LEF1 also act redundantly to pattern the mesoderm in Xenopus embryos, and this activity is linked to transcriptional activation (Liu et al., 2005). However, TCF1 and TCF3 have also been shown to act redundantly in repressing Spemann organizer genes such as *siamois* in ventral blastomeres (Houston et al., 2002; Standley et al., 2006). Loss of the TCF1 gene in mice demonstrated that it was a tumour suppressor in the intestine and mammary gland (Roose et al., 1999), suggesting further repressive functions.

The situation for TCF4 also indicates both positive and negative roles in regulating Wnt targets. TCF4 knockouts display a loss of stem cells in the intestinal crypts (Korinek et al., 1998), consistent with a loss of Wnt/β-catenin signaling. Furthermore, TCF4 is required for activation of Wnt regulated Spemann organizer genes in Xenopus (Standley et al., 2006). Conversely, loss of TCF4 can result in elevated activation of a Wnt/β-catenin signaling in CRC cells, suggesting a possible role for TCF4 as a tumour suppressor (Tang et al., 2008). Further support for a bi-modal role for TCF4 comes from studies of TCF3; TCF4 double knockouts in the skin epithelia of mice (Nguyen et al., 2009). Loss of both TCFs in the skin epithelia resulted in a dramatic decline in epidermal survival, which was not observed when β-catenin was removed (Nguyen et al., 2009). Microarray profiling revealed that many genes were repressed by TCF3 and TCF4 in a redundant manner, which were either activated or not regulated by β-catenin (Nguyen et al., 2009). These data fit a model where the two TCFs are repressing gene expression in the absence of Wnt/β-catenin signaling. Like TCF1, TCF4 can activate or repress Wnt targets, depending on the context.

The loss-of-function studies summarized above suggest a model where, in some cases, the transcriptional switch in vertebrate W-CRMs is mediated by two distinct TCFs (Figure 1.4). For example, in two CRC cell lines, siRNA data fit a model where TCF4 represses Wnt targets in the absence of signaling while TCF1 works with β-catenin to activate targets in the presence of signaling (Tang et al., 2008). Given that these results are not consistent with the phenotype of TCF1 and TCF4 knockouts in mice (Korinek et al., 1998; Roose et al., 1999), follow-up experiments with ChIP and reporter genes are

required to confirm this model. In the presumptive Spemann organizer, TCF3 occupies the W-CRM upstream of the *siamois* TSS, and its binding to chromatin is reduced by Wnt/β-catenin signaling (Hikasa et al., 2010). It will be interesting to determine the occupancy of TCF4 on the *siamois* W-CRM in response to pathway activation, since this TCF is required for *siamois* regulation (Standley et al., 2006).

In *Drosophila,* testing the functionality of TCF binding sites in W-CRMs has suggested that the switch model for TCF transcriptional regulation is W-CRM-specific (Figure 1.4). In several cases, mutation of the TCF binding sites clearly demonstrated both a repressive and an activating role. In *C. elegans*, when a single TCF site is destroyed in the END-1 W-CRM, a reduction in activation by Wnt/β-catenin signaling was observed, but the reporter was also expressed in cells where it was normally repressed by POP-1 (Shetty et al., 2005). Similar phenotypes were observed when TCF sites were mutated in the *siamois* W-CRM in Xenopus (Brannon et al., 1997; Fan and Bergmann, 2008) as well as an *eve* W-CRM in flies (Knirr and Frasch, 2001). In contrast, mutation of TCF sites in a W-CRM controlling expression of *decapentapleigic* (*dpp*) in the fly visceral mesoderm resulted in a massive derepression of expression with no loss in maximal activation (Yang et al., 2000). It appears that for this module, Wg/Arm signaling activates expression by alleviating TCF/Pan repression. At the other extreme, mutation of TCF sites often results in loss of expression of the W-CRM reporter (Chang et al., 2008b; Lee and Frasch, 2000; Yamaguchi et al., 1999). For these W-CRMs, there appears to be little role for TCF repression, and regulation by the

pathway occurs through TCF-β-catenin mediated activation (see Figure 1.4 for further explanation).

IV: DNA Binding by TCF

DNA recognition by the TCF HMG domain

The DNA sequence motif (CCTTTGATS) that mediates high affinity binding of the HMG domain of TCF/LEFs *in vitro* (Hallikas et al., 2006; van Beest et al., 2000; van de Wetering et al., 1997) has been shown to be necessary and sufficient for activation of TCF-β-catenin-dependent transcription in many contexts. This sequence, called the HMG site, has been found in many W-CRMs, and mutation of these motifs abolishes activation by the pathway (Barolo, 2006; Chang et al., 2008a; He et al., 1998; Yamaguchi et al., 1999). In addition, TOPFLASH or TOPFLASH-style reporters, which contain multiple copies of an extended 11-bp motif (CCTTTGATCTT) placed upstream of a minimal promoter, confer Wnt/β-catenin responsiveness in cell culture (DasGupta et al., 2005; Korinek et al., 1997; Lum et al., 2003) and in transgenic mice and fish (DasGupta and Fuchs, 1999; Dorsky et al., 2002; Maretto et al., 2003; Nakaya et al., 2005).

However, in many cellular contexts in which Wnt/β-catenin signaling is presumed to operate, TOPFLASH style reporters fail to respond to the signal. For instance, the p-LEF_7-fos-GFP reporter fails to respond to endogenous Wnt ligands in the *Xenopus* eye, otic vesicle, spinal cord, MHB and tailbud (Geng et al., 2003), while the murine BAT-GAL reporter fails to faithfully recapitulate the TCF response in intestinal crypts

(Dessimoz et al., 2005) (for an in depth discussion, see (Barolo, 2006)). In *Drosophila* embryos and larval imaginal discs, which have well characterized Wg expression patterns, concatemerized TCF sites fail to respond to this signal entirely (Barolo, 2006).

Furthermore, like many transcription factors, the HMG consensus site is not inviolate. Many functional HMG binding sites in W-CRMs have one or more substitutions from the consensus (Barolo, 2006). Systematic analysis of TCF4-DNA binding *in vitro* also demonstrated that single substitutions from the CCTTTGATS consensus reduced, but did not abolish, recognition by TCF4 (Hallikas et al., 2006). A polymorphic HMG site that is in the far upstream c-myc W-CRM has either one (CCTTTCATG) or two (TCTTTCATG) changes from the consensus, with the first site having increased affinity for TCF4 (Tuupanen et al., 2009; Wright et al.). Likewise, binding sites that are highly functional in fly W-CRMs from the *naked cuticle* (*nkd*) locus can have two (GCTTTGTTC) or three (GCTTTGACA) differences from the consensus (Chang et al., 2008a). In addition to high affinity sites in the *eve* and *slp1* W-CRMs, more divergent sites (e.g. ACTTCACAG) were footprinted by TCF/Pan *in vitro* and contributed to activation in transgenic fly reporter assays (Knirr and Frasch, 2001; Lee and Frasch, 2000). The heterogeneity of what constitutes an HMG site makes locating biologically relevant W-CRMs by sequence analysis alone extremely difficult.

A simple analysis of randomly selected human intergenic DNA helps to illustrate how common predicted HMG binding sites are in the genome (Table 1.1). Perfect or near perfect sites (CCTTTGAWS) are rare (1 every approx. 22 kb). But if one allows a modest level of degeneracy, for example, SCTTTGAWS or CTTTGWWS, the frequency increases to 1/10,300 or 1/2500 respectively (Table 1.1). These sequences are well

within the range of known functional HMG sites (see preceding paragraph and (Chang et al., 2008a; Knirr and Frasch, 2001; Lee and Frasch, 2000). When the consensus is loosened to the level of the polymorphic TCF site found in the upstream c-myc W-CRM (GTTTGWWS; (Pomerantz et al., 2009; Tuupanen et al., 2009; Wright et al.)), the frequency in random DNA is 1 in 645 bp. This analysis suggests that there are upwards of 5 million potential TCF binding sites in the haploid human genome, and the challenge for researchers is to determine how TCF can identify functional sites among the sea of irrelevant ones.

C-clamp dependent bipartite binding

In some TCF family members, a second sequence specific DNA binding domain called the C-clamp is comprised of approximately 30 amino acids. This domain is located C-terminal to the HMG domain and basic tail, following a linker region which varies in size (between 6-44 amino acids) depending on species and isoform (Figure 1.3). This second DNA binding domain was originally identified in a subset of vertebrate TCFs, the E-box containing isoforms TCF-1E and TCF-4 (Atcha et al., 2003), and coined the C-clamp, as it contains four highly conserved cysteine residues (Figure 1.3). Recent work in our lab has indicated that these cysteines coordinate a Zinc ion, and that DNA binding by this motif can be disrupted by chelating agents (Ravindranath and Cadigan, 2014). A related C-clamp domain capable of sequence specific DNA binding has also been identified in the zinc finger transcription factor family which includes HDBP1/GEF (GlutEF), HDBP2/PBF, and Gig1 (Tanaka et al., 2004).

In cell culture, the presence of a C-clamp downstream of the HMG domain enabled TCF1 to recognize sequences containing the classic HMG binding site and an additional sequence of RCCG (Atcha et al., 2007). W-CRMs from the *LEF1*, *cdx1* and *Sp5* genes, which contain this bipartite motif, are only activated by TCFs containing a C-clamp (Atcha et al., 2007; Hecht and Stemmler, 2003; Hoverter et al., 2012). The RCCG motif from vertebrates is similar to a longer motif independently identified in *Drosophila*, coined the Helper site (Chang et al., 2008b).

In flies, the major isoform of TCF/Pan contains a C-clamp, and this domain is required for activation of the W-CRMs containing Helper sites (Chang et al., 2008b). The Helper motif, with a consensus of GCCGCCR (R = A/G), was shown to be critical for the activation of six W-CRMs in fly cells (Chang et al., 2008b). However, unlike classic HMG sites, multiple copies of this element are not sufficient for activation of a promoter/reporter cassette. Instead, these elements greatly enhance the ability of HMG domain binding sites to respond to pathway activation (Chang et al., 2008b), hence the moniker “Helper site.” Furthermore, recombinant TCF/Pan had a dramatic increase in affinity for an HMG site if a Helper site was present, and this enhanced binding was C-clamp dependent (Chang et al., 2008b). These results suggest a model where TCF/Pan, TCF1E and TCF4E recognize DNA through a bipartite mechanism involving HMG domain-HMG-site and C-clamp-Helper site interactions. Interestingly, although it appears that the presence of a Helper site near an HMG site facilitates TCF/Pan activation of W-CRMs, the location of flanking Helper sites seems quite variable in the tested enhancers (Chang et al., 2008b). The location of vertebrate

Helper sequences also seems surprisingly flexible, and CASTing analysis of bipartite binding sites for TCF1E indicates that Helper motifs are enriched both up and downstream of The HMG motif at variable spacing (1-8 and 0-11, respectively) (Hoverter et al., 2012).

In contrast to vertebrates, where only some TCF1 and TCF4 isoforms contain a C-clamp, almost all invertebrate genomes examined contain only one TCF gene with a C-clamp. The fly *TCF/Pan* locus is subject to alternative splicing but the RNA-seq profiling indicates that the two most abundant isoforms expressed throughout development (van de Wetering et al., 1997), possess both HMG and a C-clamp domains (see modencode website: http://modencode.oicr.on.ca/fgb2/gbrowse/fly/?name=4:87956..131430). While such detailed analysis of TCF isoforms in other invertebrates has not yet been performed, it appears that the ancestral TCF gene contained both domains, and that after amplification during the vertebrate lineage, the C-clamp was lost (LEF1 and TCF3) or partially retained through alternative splicing (TCF1 and TCF4) (Figure 1.5). This model makes the prediction that Helper sites will play a central role in specifying invertebrate W-CRMs, as has been found in *Drosophila* (Chang et al., 2008b) and *C.elegans*, (Bhambhani et al., 2014), while additional mechanisms exist for target location of vertebrate TCFs lacking a C-clamp. It should be noted that despite the high degree of similarity among C-clamps, some of the invertebrate domains have non-conservative changes at some positions. For example, the *M. leidyi* C-clamp has an arginine in place of the third cysteine (Figure 1.3). Direct analysis of these C-clamps will be required to

determine whether they enhance TCF binding, as is the case for TCF/Pan, POP-1, TCF1E and TCF4E.

V. TCF target gene identification by genome-wide analysis methods

Microarray based screens and RNA-seq

In an attempt to define the Wnt/β-catenin transcriptome, microarray based screens have been performed in a variety of cell types. The number of genes whose expression is altered in the cells varies from hundreds to thousands (Jackson et al., 2005; Klapholz-Brown et al., 2007; van de Wetering et al., 2002; Van der Flier et al., 2007). A list of these microarray studies can be found on the Wnt homepage curated by the Nusse Lab at: http://www.stanford.edu/group/nusselab/cgi-bin/wnt/ and also at the Stanford Microarray Database website: http://genome-www5.stanford.edu/. One simple message from these data sets is the limited amount of overlap between Wnt targets in different cell types. It has been estimated that as few as 5% of targets are identified in all studies (Vlad et al., 2008). While variations in experimental protocols and signal detection may contribute to this low number, most of the cell-specific differences likely reflect the existence of discrete transcriptional programs. In one study, microarray analysis of PC12 and NIH3T3 cells identified 129 and 355 genes with alteration of expression in response to Wnt3a treatment respectively (Railo et al., 2009). Only two genes were commonly activated in both cell lines, one of which was *axin2*, often considered a universally induced feedback antagonist of Wnt/β-catenin signaling (Jho et al., 2002).

While some classes of Wnt targets, such as Wnt pathway components, proliferative genes or anti-apoptotic genes are found in multiple studies (Chen et al., 2007; Klapholz-Brown et al., 2007; Longo et al., 2002; Railo et al., 2009; Van der Flier et al., 2007), other classes may be more restricted in their expression domains. For instance, angiogenic (Masckauchan et al., 2005) or osteoblastic and adipogenic targets (Jackson et al., 2005) are most likely Wnt responsive only in permissive tissue specific contexts. It is important to note that there are several important shortfalls in this experimental method, and a newer technique, RNA-seq, is becoming a more popular method to assay global transcript changes (Malone and Oliver, 2011). One pitfall of microarrays is the fact that the entire genomic sequence is not represented, so changes in regions not present on the array will be missed. Secondly, the hybridization method has a narrow window of detection, and genes with large fold changes but extremely low total transcript levels may be missed. Furthermore, genes at the other end of the spectrum may also be missed, where high basal levels of transcription may saturate the assay, and fold changes would be undectectable. In RNA-seq, changes in high copy transcripts are quite easy to detect, and low copy number transcripts can be detected by increasing sequencing depth (Malone and Oliver, 2011). However, increased depth comes at a literal cost, and the need to amplify starting material can lead to expression level artifacts. RNA-seq was recently used to identify 84 genes differentially expressed in diaphyseal bone tissue null for the Wnt receptor LRP5 (Ayturk et al., 2013). In this case, surgical contamination from surrounding tissue was a serious concern, and therefore, concomitant assays of blood and muscle tissue were performed to filter and

identify bone specific transcriptional changes. As RNA-seq becomes more widespread, it will be interesting to see how many identified targets are tissue specific.

Chromatin Immunoprecipitation based screens

Another limitation of microarrays is that they do not distinguish between direct and indirect targets. ChIP-based genomic surveys offer the potential to identify regions of the genome that are enriched for bound β-catenin or a particular TCF. However, this method is also not without caveats. For example, how many of these bound regions actually correspond to a functional W-CRMs? In one study of TCF4 binding sites in LS174T cells (a colorectal cancer cell line), over 6,800 high quality binding peaks were identified using a ChIP-CHIP approach (Hatzis et al., 2008). More than 70% of the identified peaks were over 10 kb from the nearest TSS, highlighting the tremendous amount of genomic real estate that must be examined when searching for W-CRMs. In many cases, several TCF4-bound regions were found near a single gene, such as the 11 peaks surrounding the Axin2 gene. Four of these regions had W-CRM activity in a reporter assay, while 10 out of 22 other TCF4 bound regions from other locations tested positive in this assay (Hatzis et al., 2008). It remains to be seen whether the regions that tested negative in the reporter assay are simply non-functional binding sites for TCF4 or represent W-CRMs that are not active in a simple reporter assay. The relatively small degree of overlap (12.5–20.5% depending on how the comparison is made) between whether a TCF4-bound region was found within 100 kb of a TSS from a gene

upregulated in adenomas suggests that many of these Wnt targets may be indirectly regulated (Hatzis et al., 2008).

Another study using ChIP followed by high throughput sequencing (ChIP-seq) identified over 20,000 TCF4 bound regions in the human CRC cell line HCT116 (Blahnik et al., 2010). Over 6000 of these motifs map to putative enhancer regions within 10–100 kb of a TSS, while over 9000 mapped within 2 kb of a TSS. Using the same cell line, over 2100 regions enriched for β-catenin binding were also reported (Bottomly et al., 2010). In this study, only 47% of the peaks contained at least one consensus TCF motif in the vicinity of the peak. The remaining peaks may represent TCF-independent β-catenin targets, peaks containing TCF binding motifs which diverge from the consensus, peaks in which TCF is recruited not by direct DNA contact, but by protein-protein interactions with other TFs, or false positives. Currently, the false positive rate for these experiments is unclear. In *Arabidopsis*, the use of null mutants indicates many peaks are caused by non-specific binding of the antisera (Wierzbicki et al., 2012; Zheng et al., 2012), and the adoption of this important control would greatly benefit studies in metazoans. In cases where lethality precludes this approach, comparing data from multiple antibodies with different epitope specificities may be an alternative.

In Freitze, et al., the binding profile of TCF4 was studied in 6 distinct human call lines. Between 24,000 to 53,000 peaks were identified in each line, and the overlap varied from 18% to 46% in pairwise comparisons, indicating that many targets of Wnt signaling are likely to be tissue specific (Frietze et al., 2012). Interestingly, enrichment of other transcription factor binding sites seemed highly cell line specific, and GATA3

co-occupancy was identified in MCF7 cells, while HNF4alpha and FOXA2 co-occupancy was found in HepG2 cells. Furthermore, in MCF7 cells, the enrichment of consensus TCF4 binding sites was only found in peaks without flanking GATA motifs, and TCF localization was lost in this subset of peaks upon GATA3 depletion. These data argue that either tethering by, or cooperativity with, other TFs may be an important mechanism for cell-type specific recruitment of TCFs to W-CRMS. This may partially explain the generally low percentage of high quality HMG binding sites found in ChIP–seq peaks. The idea of transcription factor cooperativity is supported by a ChIP-seq analysis of TCF/Pan and four other cardiogenic TFs (Pannier, Dorsocross, pMad and Tinman) in 4-8 hour *Drosophila* embryos (Junion et al., 2012). Enrichment of HMG motifs was lower in peaks shared by all five TFs than in peaks shared by only TCF and one other TF. It should be noted that the Helper motif has been ignored in these analyses, and the presence or absence of Helper site motif in bound DNA is likely to be an important piece of information in identifying *bona fide* W-CRMs for C-Clamp containing TCFs. It should also be mentioned that in order to decrease the number of false positives in these screens, peak calling algorithms are set to high stringency, and weakly bound but functional sites are likely to be overlooked. This is of particular concern in studies where the whole organism or multiple tissue types are assayed together (Junion et al.). In these cases, higher peaks may disproportionately represent ubiquitously regulated targets at the expense of tissue specific targets, which may be bound only in a subset of the total cells fractionated.

Bioinformatic Identification of W-CRMs

The above studies indicate that TCFs can bind to regions far removed from the TSS of target genes. This suggests that the most common way of determining whether a developmental gene is directly regulated by Wnt/β-catenin signaling, that is, scanning the region immediately upstream of a candidate gene's TSS for conserved TCF binding sites, followed by site-directed mutagenesis, may miss many W-CRMs. While continued genome-wide surveys of TCF-bound chromatin in interesting developmental contexts is desirable, an alternative is to use computational approaches to detect W-CRMs within entire genomes. For example, an algorithm called the Enhancer Element Locator (EEL) utilized binding site affinity matrixes and motif clustering conservation between two or more species to identify potential W-CRMs (Hallikas et al., 2006). Several putative elements were tested in a transgenic mouse assay. Their expression patterns were consistent with positive regulation by Wnt/β-catenin signaling, though this was not directly confirmed by site-directed mutation of the conserved TCF sites. While this method is likely to identify some W-CRMs, the challenge of sorting through the entire genome requires stringent screening parameters which likely miss many biologically relevant elements, and may bias findings towards elements regulating ubiquitously expressed genes, at the expense of tissue specific modules. One of the benefits of this algorithm is the reliance on conservation of motif clusters rather than strict sequence conservation. In some instances, though, enhancer elements in divergent species have been shown to retain functional conservation while rapidly altering motif organization (Swanson et al., 2011) or losing motif clustering or locational conservation (Kalay and Wittkopp, 2010), and thus may elude this type of analysis.

Given the likelihood that many targets of the Wnt/β-catenin pathway are cell specific, it seems that a combination of transcriptome analysis, physical localization assays (e.g. ChIP-seq) and further refinement of computational approaches, including a better understanding of the sequence requirements for functional TCF binding, will be needed to efficiently identify novel and important W-CRMs. A panel of 162 genes, called "a Wnt target signature", were identified in a colon cancer cell line by combining ChIP-seq binding data for β-catenin with micro-array identified gene expression changes (Watanabe et al., 2014). Since many Wnt targets are controlled by TCFs in combination with other transcription factors, often in direct contact with each other, integrating binding site information for other TFs, as was done by the Furlong and Jin groups (Frietze et al., 2012; Junion et al., 2012), should also aid these efforts.

Combinatorial interactions of TCFs and other transcription factors

Combinatorial interactions between TCFs and other transcription factors may take several forms. The formation of complexes able to recruit transcriptional machinery may require the presence of multiple TFs bound to the DNA in close proximity. Alternately, protein-protein interactions may allosterically change the affinity of TCF for a DNA binding site, or tether TCF to a W-CRM in the absence of a functional binding site. Negative interactions are also possible, as in the case of TF competition for overlapping DNA binding sites. Interactions between TCFs and GATA factors may serve as examples for several of these mechanisms, depending on the family member and context. A direct protein-protein interaction has been demonstrated for LEF1 and

GATA3, which requires the HMG domain (Hossain et al., 2008). In MCF7 cells, GATA3 occupancy is required for TCF localization to some loci, arguing for a tethering mechanism resulting in transcriptional repression (Frietze et al., 2012). In contrast, TCF4 and GATA1 (and 2) colocalize at W-CRMs which contain binding motifs for both factors in hematopoietic cells, resulting in transcriptional activation (Trompouki et al., 2011). In *Drosophila*, TCF/Arm dependent direct repression occurs at the *Ugt36B* and *tiggrin* loci through WGAWA motifs (Blauwkamp et al., 2008; Zhang et al., 2014), whose similarity to GATA binding sites suggests that displacement of GATA factors may occur. In this section, several other transcription factors are discussed that interact with TCFs and/or β-catenin and appear to act cooperatively with TCFs to bind to regulatory elements, primarily for positive combinatorial effect.

The Smad Family Transcription Factors

One family of transcription factors that interact with TCFs on cis-regulatory elements are the Smads, which mediate many aspects of TGF-β signaling (Moustakas and Heldin, 2009). Although the Wnt/β-catenin and TGF-β pathways cross-talk at several levels (Eivers et al., 2009; Itasaki and Hoppler), here the discussion will focus on reports of cis-regulatory cooperation, in which Smad-TCF-β-catenin complexes assemble on composite DNA regulatory elements. This was first shown for the regulatory region of the *twin* gene in Xenopus embryos (Nishita et al., 2000) and mammalian cells (Labbe et al., 2000). In both contexts, both Smad and HMG binding sites were required for maximal activation of reporter constructs by Wnt/β-catenin and

TGF-β signaling. Additionally, the HMG domain of LEF1 can directly interact with Smad 3 or Smad 4 (Labbe et al., 2000; Nishita et al., 2000). These results suggest a model where a combination of protein-DNA interactions and protein-protein interactions can promote the formation of a Smad-TCF-β-catenin complex in a signaling-dependent manner (Figure 1.6).

Since these initial reports, other genes have been identified that are co-regulated by TCFs and Smads. Most of these studies are in the context of cell culture, such as regulatory elements from the *Msx2* (Hussein et al., 2003), *c-myc* (Hu and Rosenblum, 2005), *gastrin* (Lei et al., 2004) (Chakladar et al., 2005), *Sm22a* (Shafer and Towler, 2009), *TMEPA1 (Nakano et al., 2010)* and several osteogenic genes (Rodriguez-Carballo et al., 2010). But the existence of functional Smad and TCF binding sites in close proximity to each other has also been found in *Emx2* elements active in the developing CNS of the mouse (Suda et al.; Theil et al., 2002). While these studies mostly relied on reporter constructs, some ChIP data suggests that Wnt/β-catenin signaling can increase Smad recruitment to regulatory chromatin (Hussein et al., 2003) (Shafer and Towler, 2009). Conversely, TGF-β signaling can recruit LEF-1 or TCF4 to chromatin as well (Hussein et al., 2003; Nakano et al., 2010). The presence of both Smad and β-catenin on the chromatin has been proposed to increase binding for the histone acetyltransferases CBP/p300, leading to increased histone acetylation and transcription (Rodriguez-Carballo et al., 2010) (Lei et al., 2004) (Figure 1.6B).

Interestingly, a direct protein-protein interaction between TCF/Pan and the *Drosophila* Smad, Mothers against Dpp (Mad), has been reported. In this case, it

appears that some Wnt targets, such as the gene *Senseless*, require Mad, and activity of a TOPFLASH reporter is increased with overexpression of Mad in cell culture (Eivers et al., 2011) . In this case, arm/TCF/Mad ternary complexes are thought to bind DNA through TCF binding motifs, which are most likely bipartite.

AP-1

Another transcription factor linked with TCF-β-catenin transcriptional activation is c-Jun, a basic leucine zipper domain protein that can bind DNA specifically as a homodimer or as a heterodimer with c-fos (constituting AP-1) (Shaulian and Karin, 2002). In CRC cells, c-Jun, TCF4 and β-catenin cooperated in activating c-Jun expression (Nateri et al., 2005). Phosphorylated c-Jun was found to associate with TCF4 and both transcription factors occupy the c-Jun regulatory region (Nateri et al., 2005). These data complement genetic interaction studies in the mouse intestine to support a model, where Wnt/β-catenin signaling acts with c-Jun in a positive feedback loop to promote carcinogenesis (Nateri et al., 2005) (Sancho et al., 2009). In contrast to most of the elements co-regulated by TCF and Smad, the distance of the functional TCF and AP-1 site suggests the existence of a DNA loop stabilized by protein-protein and protein-DNA interactions (Figure 1.6C).

While the Wnt/β-catenin-c-Jun autoregulatory loop may be crucial for intestinal cancer in mice and CRC in humans, additional evidence suggests that many Wnt transcriptional targets in CRC cells are co-regulated by TCF4 and c-Jun. A *c-myc* W-CRM located downstream of the *c-myc* gene contains a functional AP-1 site that is

required for synergistic activation between the Wnt/β-catenin pathway and serum-derived mitogens (Yochum et al., 2008). A genome-wide survey of chromatin sites with β-catenin enrichment revealed that 40% of the β-catenin bound regions contain both TCF and AP-1 binding sites (Bottomly et al., 2010). More than a dozen sites were bound by TCF4, β-catenin and c-Jun. As previously shown for *c-myc*, the activation of several Wnt targets were enhanced by serum in CRC cells arrested in G0/G1 (Bottomly et al., 2010). The connection between Wnt/β-catenin signaling and cell cycle progression has also been noted further upstream in the pathway (Davidson and Niehrs).

The Wnt/β-catenin pathway-c-Jun connection has also been observed outside the context of intestinal cells and CRC. Regulatory elements controlling either the *matrilysin* gene in kidney or the *versican* gene in melanoma require both TCF and AP-1 sites (Domenzain-Reyna et al., 2009; Rivat et al., 2003). In addition, interactions between TCF4 and c-Jun (Gan et al., 2008) or β-catenin and the AP-1 complex (Toualbi et al., 2007) can regulate Wnt targets independent of AP-1 binding sites, though whether this type of regulation occurs under physiological conditions is not clear.

Other factors

The GATA, Smad and c-Jun/AP-1 studies described above are examples where distinct signaling pathways and Wnt/β-catenin signaling converge on regulatory elements to activate transcription. Such combinatorial control of Wnt targets can also occur through interactions with transcription factors not directly controlled by cell–cell

signaling. One candidate for such factors is the Cdx family of homeodomain proteins. C*dx1* and *cdx4* are known to be direct transcriptional targets of the Wnt/β-catenin pathway (Ikeya and Takada, 2001; Lickert et al., 2000; Pilon et al., 2006). In addition, there is some developmental genetic data suggesting that the TCF-β-catenin complex may functionally interact with Cdx proteins (Young et al., 2009). Cdx1 autoregulation has been shown to require a Cdx1-LEF1-β-catenin complex, involving direct interactions between the homeodomain and HMG domains (Beland et al., 2004). More recently, a genome-wide survey of Cdx2 binding in intestinal cell lines revealed a significant overlap between Cdx2 and TCF4 chromatin bound regions (Verzi et al., 2010). TCF4 binding to chromatin was found to be partially dependent on Cdx2 at several locations (Verzi et al., 2010). Interestingly, nested TCF-Cdx binding sites have been shown to be required for an intronic *raldh2* enhancer that is active in the dorsal spinal cord of the chick (Castillo et al., 2010).

There are other examples of TCFs interacting with other transcription factors to regulate gene expression. LEF1 and microphthalmia-associated transcription factor are thought to physically interact to regulate gene expression in melanocyte differentiation (Yasumoto et al., 2002). Likewise, LEF1 and the homeodomain protein Pitx2 may interact in the developing dental epithelium and other tissues (Amen et al., 2007). The short list of transcription factors that interact with TCFs discussed here is likely only the tip of the iceberg. Genome-wide studies of TCF4 binding patterns have found an enrichment for many other transcription factor binding sites besides AP-1, including NF1, PPARγ, HNF4, Elk-1, GATA1, -2, and -3, c-Ets-1, Bach-1 and FoxD-1 (Blahnik et

al., 2010; Frietze et al., 2012; Hatzis et al., 2008; Trompouki et al., 2011). For TCF/Pan, the cardiogenic factors Tin, Pannier, Doc, and Mad, have been identified (Junion et al., 2012). TCF-protein interactions are likely a common mechanism to locate W-CRMs in the information rich nucleus, though the binding partner is likely to be different for different targets.

C-Clamp dependent interactions

While C-Clamp dependent interactions have not been studied in detail, several reports have indicated that the C-Clamp may direct some protein-protein interactions. An interaction between the transcriptional Co-activator p300 and the C-clamp containing isoform TCF4E is lost in a deletion mutant missing the C-Clamp domain (Hecht and Stemmler, 2003), although it is unknown which specific residues in the 46 aa region are required for the interaction. In another example, the Mucin-1C protein has been shown to bind to both TCF4E and β-catenin in a human breast cancer cell line, and in an *in vitro* binding assay (Rajabi et al., 2012). This interaction was shown to require the two central cysteines in the C-Clamp, as well as two cysteines in a CQC motif in Muc-1C. This interaction is thought to promote activation of targets such as Cyclin-D1, both by stabilizing the β-catenin-TCF4 interaction, and by blocking C-terminal binding sites for the co-repressor CtBP (Rajabi et al., 2012). How this interaction affects the ability of the C-clamp to bind DNA is unknown.

VI. Spacing and Orientation Requirements for other bipartite DNA binding proteins

The use of two DNA binding sites is a common strategy to increase the affinity and specificity of a single transcription factor or a transcription factor complex. The previous section covered a number of combinatorial interactions which improve the targeting or activity of TCFs (with or without the C-clamp) at specific cis-regulatory regions by other transcription factors. As the mechanism by which HMG/Helper pairs improve TCF/Pan specificity in Drosophila is a major focus of this thesis, the manner in which other transcription factors utilize bipartite binding sites is of great interest. This section will look at binding site spacing and orientation requirements for several classes of transcription factors. Monomeric transcription factors which contain multiple DNA binding domains may be the most relevant comparison to HMG and C-clamp containing TCFs. The Paired-homeodomain proteins are an example of TFs which contain multiple DNA binding domains. The factor Pax-3, which contains both the paired domain and the homeodomain, has been shown to flexibly bind bipartite elements with spacing between 1-13 bp (Phelan and Loeken, 1998). For Pax-3, the highest affinity binding is seen at 5-6 bp intersite spacing. Increased or decreased spacing correlates with decreased binding and activity, indicating a limit to Pax-3's flexibility. TFs like ZEB1/δEF1 and ZEB2/SIP-1, which bind DNA through two zinc finger clusters separated by a flexible linker (Remacle et al., 1999), are also relevant. The POU domain containing proteins are an interesting case, as they can multimerize, and the monomers each contain two DNA binding domains (Herr and Cleary, 1995; Phillips and Luisi,

2000). This creates a situation in which DNA motif architecture can potentially direct assemblage of a plethora of unique transcriptional complexes leading to either activation or repression. Other classes of multimeric transcription factors will also be discussed, as they often display very specific half-site spacing and orientation requirement for correct assemblage on the DNA. These fall into two general categories: those which bind DNA as homomers, such as the p53 tetramer (Tokino et al., 1994), and those that can bind as heteromers. For example, the spacing and orientation preferences for class II nuclear receptors are determined by the subunit composition of the receptor dimers (Mangelsdorf and Evans, 1995; Tata, 2002). This correlation has been dubbed the "1 to 5 rule'.

Nuclear hormone receptors

Nuclear hormone receptors (NHR) are transcription factors with the ability to complex with hormone ligands. They are divided into 5 classes based on phylogenetic similarity and the mechanism of DNA binding (Olefsky, 2001). As several of the classes bind DNA as homo- or heterodimers, bipartite binding site spacing requirements have been studied for many of the hormone receptors. The steroid hormone (class I) receptors, (androgen, estrogen, progesterone, mineralocorticoid and glucocorticoid) (Olefsky, 2001), bind DNA as homodimers which recognize an inverted repeat of two hexamer motifs with a 3bp intersite spacing (IR3) (Geserick et al., 2005; Shaffer et al., 2004). However, many functional motifs deviate from this pattern, and allosteric regulation of these factors is thought to be a common mechanism (Geserick et al., 2005; Lefstin and Yamamoto, 1998; Watson et al., 2013). The binding site architecture

of the glucocorticoid receptor (GR) may be the best studied example in this class. Classic Glucocorticoid Response Elements (GREs) are IR3 elements which drive transcriptional activation when bound by GR homodimers (Geserick et al., 2005; Meijsing et al., 2009). In contrast, elements which contain a shorter spacer of 0-2 bp act as repressive elements, downregulating target genes upon GR binding (Hudson et al., 2013; Surjit et al., 2011) (Figure 1.7A). The 3 bp spacer is required for self-association of two molecules on the same DNA face, and steric hindrance prevents dimer formation at elements with shorter spacer elements (Hudson et al., 2013). At these repressive GREs, (Zanke et al.) GR monomers instead bind to opposite faces of the DNA, and these monomers preferentially recruit co-repressor molecules such as NCoR (Nuclear receptor CoRepressor 1) and SMRT (Silencing Mediator of Retinoic acid and Thyroid hormone receptor) (Hudson et al., 2013; Surjit et al., 2011).

The class II hormone receptor family can bind DNA as either homo- or heterodimers. For instance, the retinoid X receptor (RXR) forms heterodimers with other subfamily members such as the peroxisome proliferator-activated, thyroid hormone, vitamin D and retinoic acid receptors (PPAR, TXR, VDR and RAR, respectively), and the “1 to 5” rule has implicated half site spacing in the choice of heterodimer partners (Mangelsdorf and Evans, 1995; Tata, 2002). This class of NHRs binds to direct repeats, and heterodimers containing the receptors PPAR, RAR, VDR, TXR and RAR preferentially recognize half sites pairs separated by 1, 2, 3, 4, and 5 bp respectively (Figure 1.7B) (Katz and Koenig, 1994; Mangelsdorf and Evans, 1995; Naar et al., 1991; Perlmann et al., 1993; Tata, 2002; Umesono et al., 1991). RXR generally binds the 5’ half site while the 3’ half site is bound by the other receptor (Tata, 2002).

However, there is evidence that these rules are not cut and dry. For instance, retinoic acid/retinoid x (RAR/RXR) heterodimers have also been shown to bind direct repeats at 0, 1 and 8 bp spacing, and 0 space invert repeats (Moutier et al.). Furthermore, Thyroid Hormone homo- and heterodimers appear to bind invert repeats at a variety of spacings *in vitro* (Katz and Koenig, 1994), although the functional relevance of this observation has not been well studied. In addition to half site spacing, the sequence composition of half-sites is also important to direct specificity of binding partners, as are sequences flanking the bi-partite motif (Katz and Koenig, 1994; Phan et al., 2010). Interestingly, heterodimers of RXR and the Class II receptor, COUP-TF, also bind direct repeats with 1 bp spacing, however, COUP-TF directs transcriptional repression (Kliewer et al., 1992). Binding site architecture likely works in combination with receptor availability to dictate tissue specific patterns of gene expression. Small differences in binding affinity seen *in vitro* may have large functional effects in directing which homo- or hetero-dimers are formed at CRMs.

p53 family

The p53 tumor suppressor gene has been widely studied, and exerts many of its effects through both transcriptional activation and repression. p53 binds DNA as a tetramer, and canonical p53 response elements (p53REs) are defined as two repeating sites (which are themselves inverted repeats of a heptamer) with the consensus "RRRCWWGYYY" separated by a 0–13 bp spacer (Wang et al., 1995). While some repressive activity of p53 is indirect, there are many reported cases of direct transcriptional repression by p53, and potential mechanisms have long been a topic of

study. In 1994, Tokino et al. reported that intersite spacing was critical to p53 mediated transcriptional activation. A synthetic reporter was highly activated at 4 bp, but not at 0 or 14 bp between sites (Tokino et al., 1994). The idea that half site arrangement may dictate activation versus repression of p53 targets was bolstered by several following reports. Removing a 3 bp spacer from the Survivin p53RE switched the element from a repressive to an activating element (Hoffman et al., 2002). Similarly, a non-canonical repressive p53RE in the MDR1 promoter was changed into an activator by replacing the direct repeat half sites with canonical invert repeat half sites (Johnson et al., 2001). p53 has been shown to interact with both co-activator and co-repressor molecules (Bansal et al., 2011; Barlev et al., 2001), and the possibility that the specific DNA sequence bound may alter the affinity of p53 for particular co-factors is an attractive model. However, with the identification of more p53REs, a clear regulatory code has not emerged. Rather, spacing and orientation constraints seem to be context-specific, and nucleotide choice, especially in the CWWG core motif, within the half site also affects whether the transcriptional outcome is activation or repression (Wang et al., 2010).

Interestingly, the distantly related, and possibly ancestral LSF/CP2/GRH family of proteins (Kokoszynska et al., 2008), has members reported to bind DNA half-sites as tetramers, with half-site spacing sensitivity. For the human LSF/CP2 oligomer, half site spacing is strongly correlated with binding affinity, with a preferred center to center spacing of 10 bp (which correlates with the 4bp intersite spacing reported for p53) for direct repeat half-sites (Figure 1.7C) (Huang et al., 1990; Shirra et al., 1994). This argues that the tetramer is formed on a single face of the DNA. While LSF appears primarily dimeric in solution, it is thought to form stable tetramers at the DNA. This

differs from a related factor in *Drosophila*, Grainyhead (GRH), which binds DNA as a dimer (Shirra and Hansen, 1998). Consequently, GRH is able to bind a single LSF half-site, while LSF requires both sites to bind.

Zinc finger clusters

The zinc-finger E-box-binding (ZEB) transcriptional repressors, ZEB-1(δEF1/Zfhx1a) and ZEB-2 (SIP1/Zfhx1b) bind DNA as a monomer via a bipartite binding mechanism (Remacle et al., 1999). Both family members contain two zinc-finger clusters separated by a relatively long and presumably flexible linker region (Comijn et al., 2001; Postigo and Dean, 2000). Both the C-terminal 4 zinc finger cluster and the N-terminal cluster of 3 fingers, which are highly conserved between the two proteins, recognize the motif CACCTN *in vitro* (Postigo and Dean, 2000; Remacle et al., 1999). However, high affinity binding requires one CACCTN and one CACCTG motif, and functional bipartite motifs have been reported with spacing that varies from 24-44 bp, in both direct and invert repeat orientations (Figure 1.7D) (Remacle et al., 1999). It is likely that this mode of bipartite DNA binding is utilized in other classes of zinc-finger TFs, such as in the MBP/PRDII-BF1 and NZF families.

For the mammalian MBP/PRDII-BF1 protein (a homolog of *Drosophila* Schnurri (Arora, 1995 #1414;Grieder, 1995 #1415), both zinc finger clusters have been shown to bind palindromic motifs conforming to the consensus GGGAYTTYCCC with high affinity (van 't Veer et al., 1992). In this case, it is unclear if this palindrome represents a bipartite motif that can be separated into two widely spaced half-sites, or if two palindromic element are required for high affinity binding. The NZF family members

MyT1 and NZF-3 bind tandem direct repeats of the motif AAASTTT, with spacing of 1-11 bp between elements (Yee and Yu, 1998). A final example, Evi-1, binds to a bipartite motif containing a 1-28 bp spacer. In this case, the N-terminal 3 finger cluster binds to an invert repeat of the core sequence recognized by the C-terminal 7 zinc finger cluster ((GACAAGATAAGATAA-N1–28-CTCATCTTC) (Morishita et al., 1995).

As the HMG and C-Clamp domains in many TCF/LEFs are separated by a relatively short linker (9, 10 and 11 amino acids for TCF/Pan, POP-1 and TCF1E respectively) (Figure 1.3), the amount of motif flexibility seen in zinc finger cluster TFs may not be attainable by TCFs. TF families with smaller linker regions, such as the POU protein families may be more informative.

POU domain TFs

The POU domain family of transcription factors is named for the shared DNA binding domain identified in the Pit-1, Oct-1 and-2, and Unc-86 proteins. This DNA binding domain is actually comprised of two individual DNA binding domains, POU_H and POU_S, which are joined by a flexible linker that varies in size depending on the family member (Herr and Cleary, 1995). POU_S, the POU specific domain, binds to the consensus sequence ATGC, while POU_H binds the homeodomain-like tretramer AAAT (Klemm et al., 1994). The flexible linker allows the two domains to interact with a wide array of half-site sequences either on the same or opposite faces of the DNA molecule (Jacobson et al., 1997; Tomilin et al., 2000). The ability to homo- and hetero-dimerize further increases the number of possible functional binding sites for these TFs (Jacobson et al., 1997; Kemler et al., 1989; LeBowitz et al., 1989; Poellinger et al.,

1989; Voss et al., 1991). Allosteric regulation of these factors, therefore, has been widely studied. There is ample evidence that the half-site architecture can affect transcriptional output by changing the conformation of the bound TF, thereby altering affinity for co-activator or co-repressor molecules (reviewed in Phillips, 2000 #1428).

For example, at the growth hormone promoter, the ability of Pit-1 to repress this target is dependent upon a 6 bp intersite spacer (Scully et al., 2000). The conformation of Pit-1 on this motif, determined by crystal structure, allows for interaction with the co-repressor molecule NCoR, and this interaction is lost when the spacer is decreased to 4 bp (Figure 1.7E) (Scully et al., 2000). A 4bp bipartite motif, also found at the prolactin promoter, instead confers activation, presumably as the conformation at this binding site preferentially recruits activating cofactors (Scully et al., 2000). The ability of Oct-1 and -2 dimers to bind the transcriptional co-activator OBF-1 is dependent upon the half-site-directed dimer conformation. In contrast, the co-activator E1A appears to be able to bind Oct4 in multiple dimer conformations (Tomilin et al., 2000). Furthermore, for the palindromic MORE binding element, inserting a 4 bp spacer between half-sites was shown to abolish dimeric, but not monomeric Oct4 binding, while a 2 bp spacer was tolerated (Tomilin et al., 2000). These studies argue that the tremendous diversity seen in POU binding site composition may be a primary mechanism in creating tissue and context-specific transcriptional outputs.

The size of the linker region in specific family members may be a major determinant of the bound protein conformation. For instance Pit1, which has the shortest linker at 15 amino acids, binds to the prolactin element (4 bp intersite spacing) as a homodimer, with each subunit binding to a single perpendicular face of the DNA.

Oct1, with a 24 amino acid linker, binds as a monomer to a similar element with 2 bp spacing by wrapping around the DNA strand and binding opposite faces of the DNA (Jacobson et al.)(Figure 1.7F). Brn-2, which has a linker of 17 amino acids, can interact with motifs containing half-site spacing of 0, 2 and 3 bp (Li et al., 1993). Interestingly, the 2 bp motif is bound with the highest affinity, and displays a bound conformation with the shortest distance between DNA binding domains (Jacobson et al., 1997; Li et al., 1993). This suggests that increasing or decreasing motif spacing may increase stress on the linker region, and this stress results in decreased binding affinity. Strangely, Brn3, which also has a 17 bp linker, displays much less tolerance, and binds exclusively to motifs with 2bp intersite spacing. Altering amino acids in the Pou_H domain, but not in the linker, relaxes this constraint (Li et al., 1993), indicating that the DNA binding domains may effect linker conformation, and the linker may not always be quite as flexible as it is thought.

While these studies indicate that many protein families exhibit flexibility in what constitutes an allowable bipartite binding site, the binding site architecture can be of primary importance in determining the conformation and composition of the transcription factor complexes formed at the DNA. Sequence directed allosteric regulation can thereby affect the type of transcriptional regulation (activation versus repression), the strength of the transcriptional response, and also the location where a response is seen. Known and putative HMG and Helper sites are found in close proximity to each other in a number of *Drosophila* W-CRMs, but their orientation and spacing appears highly flexible (Chang et al., 2008b). Therefore, we sought to investigate the limits of this flexibility to better delineate what makes a functional bipartite binding site for TCF/Pan.

Rationale

In *Drosophila*, TCF/Pan acts as the primary transcriptional regulator of Wingless (Wg, a fly Wnt). Wg has multiple stage and tissue specific roles during development, but how TCF/Pan mediates the pleiotropic effects of Wg signaling is poorly understood. Tissue specific transcriptional programs likely integrate multiple signaling inputs, but to dissect these combinatorial interactions requires knowledge of the location and sequence composition of these cis-regulatory DNA elements. Understanding how TCF identifies and bind these elements will improve our ability to identify novel W-CRMs. Analysis of these W-CRMS, in turn, will allow us to dissect these complicated combinatorial inputs.

Our data support a model where TCF binds two distinct DNA motifs, the HMG and Helper sites, via two closely spaced domains, the HMG domain and the C-Clamp. Both DNA sequences have been shown to be indispensable for activation of multiple Wg targets, but surprisingly, spacing and orientation of these two motifs varies both within and between W-CRMs (Wingless responsive Cis-Regulatory Elements). The focus of this work was twofold: one) to investigate the effects of HMG/Helper motif configuration on TCF binding affinity and transcriptional activation potential, and two) use the knowledge of bipartite motif usage to improve *in silico* searches for novel W-CRMs.

Experiments *in vitro* and in cell culture were designed to survey a variety of bipartite motif conformations for binding and transcriptional activity. Analysis of a

subset of these conformations in synthetic W-CRMs was carried out in transgenic *Drosophila*. These experiments were designed to investigate whether spacing/orientation constraints identified *in vitro* were universal or tissue specific. In addition, the genome-wide distribution of bipartite motif configurations was analyzed bioinformatically. A publicly available ChIP-seq data set for germband extended embryos was used to compare motif enrichment in TCF bound regions versus random DNA.

To identify novel W-CRMs, we tailored our search for high affinity bipartite motifs which were responsive to Wg signaling in multiple tissues that regulate Wg target genes. A candidate approach was used to search the genomic locus of a putative Wg target, *ladybird early*, and regions were chosen for analysis based on the clustering of multiple bipartite motifs. A broader search of the entire Chromosome arm 3R used criterion focusing on a single high affinity motif, and two regions, which also contained other nearby motifs of lower quality, were chosen for analysis. The expression patterns of these putative W-CRMS were evaluated in transgenic *Drosophila,* and the requirement for Wg- dependent activation was tested.

This introductory chapter was adapted from the previously published review article:

Archbold, H. C., Yang, Y. X., Chen, L., and Cadigan, K. M. (2012). **How do they do Wnt they do?: regulation of transcription by the Wnt/beta-catenin pathway**. Acta physiologica *204*, 74-109.

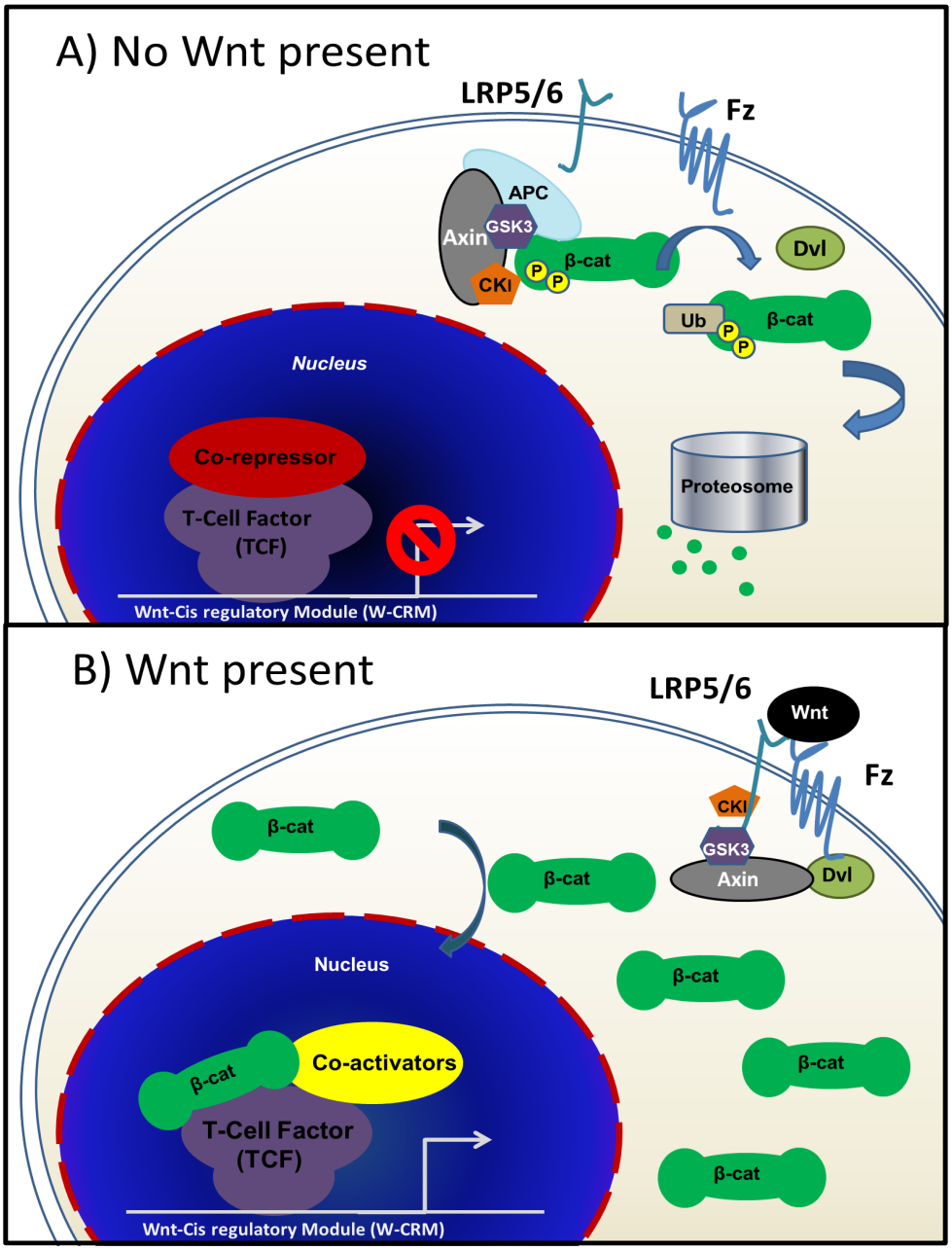


Figure 1.1 Basic outline of the Wnt/b-catenin signaling pathway.
(A) In the absence of Wnt ligand, a "destruction complex" containing Axin, APC1, and kinases GSK3 and CKI targets b-catenin for ubiquitination and proteosomal degradation. In the nucleus, TCF preferentially binds co-repressors, keeping many target genes off. (B) Upon Wnt ligand binding, disruption of the destruction complex results in the stabilization and increased nuclear translocation of b-catenin, where it binds TCF and recruits transcriptional co-activators to the W-CRM. See text for more details. APC, adenomatous polyposis coli; CKI, casein kinase I; GSK3, glycogen synthase kinase 3; TCF, T-cell factor 1; Fz, Frizzled; LRP, lipoprotein receptor related protein; Dvl, Dishevelled. Adapted from Archbold et al. 2012

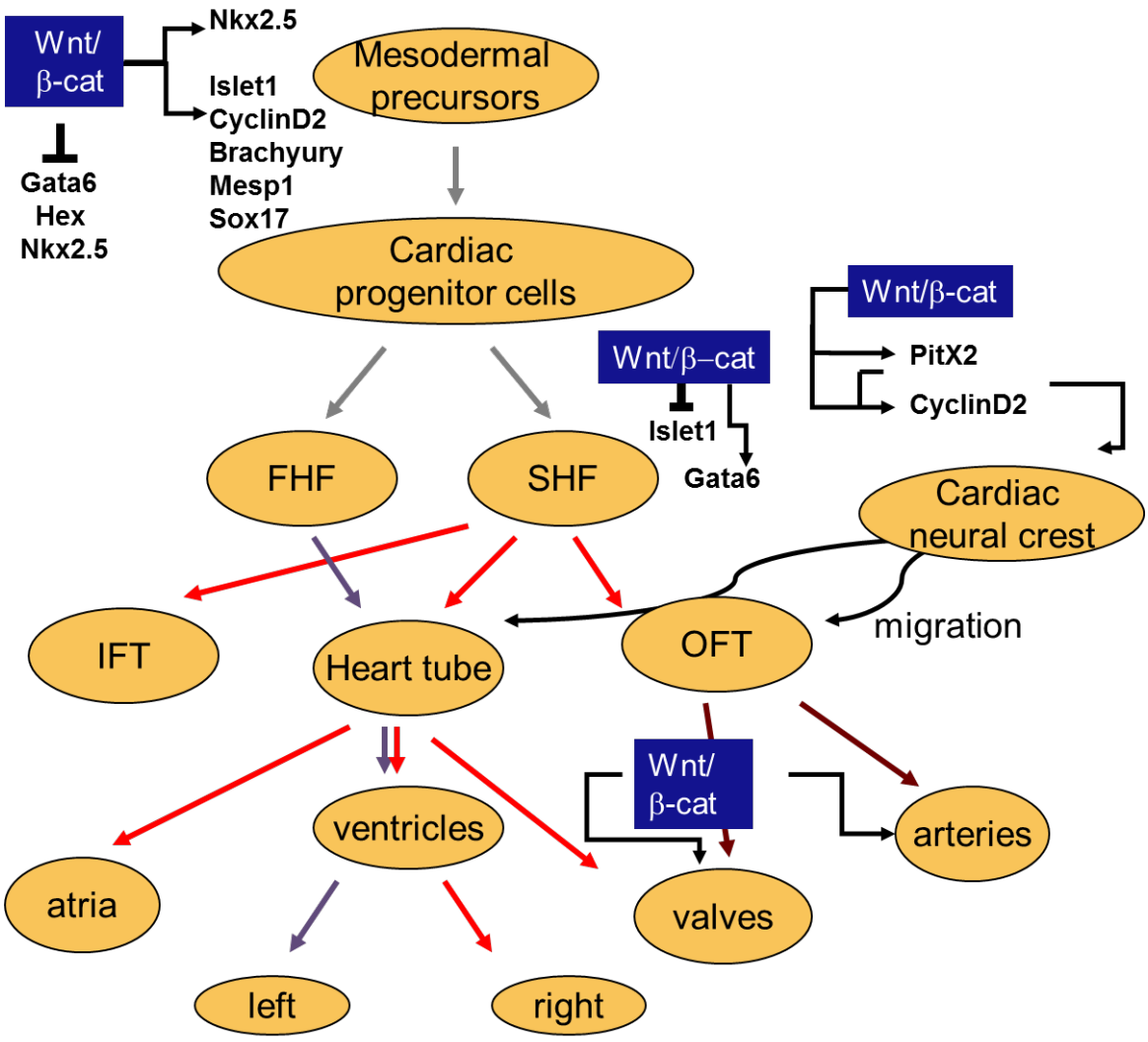


Figure 1.2 Wnt/β-catenin signaling (WBS) has multiple roles in heart development.
This schematic depicts a few of the roles WBS plays in mammalian heart development, both as an activator and as a repressor of gene activity. Prior to gastrulation, *nkx2.5* is positively activated by WBS, while after gastrulation it is repressed, as are cardiogenic factors GATA6 and Hex. Activation of Wnt target genes such as Islet1 early in development, has been linked to the proliferation of Cardiac Progenitor Cells (CPC). Differentiation of these CPCs has been linked to downregulation of WBS. These Islet1+ CPC contribute to the Second Heart Field (SHF), inflow tract (IFT) and outflow tract (OFT), and the SHF is required for proper formation of the atria and right ventricle. In the posterior SHF, Wnt activates GATA6, in contrast to its earlier inhibitory role. At this stage, another example of this contradictory activity is seen in Islet1+ cells where WBS may have a repressive or no effect on continued Islet1 expression. WBS is also instrumental in driving proliferation of cardiac neural crest cells, which migrate to the heart tube and OFT, and are important in septation and valve and artery formation along with WBS driven endocardial proliferation. See text for details. Adapted from Archbold et al. 2012.

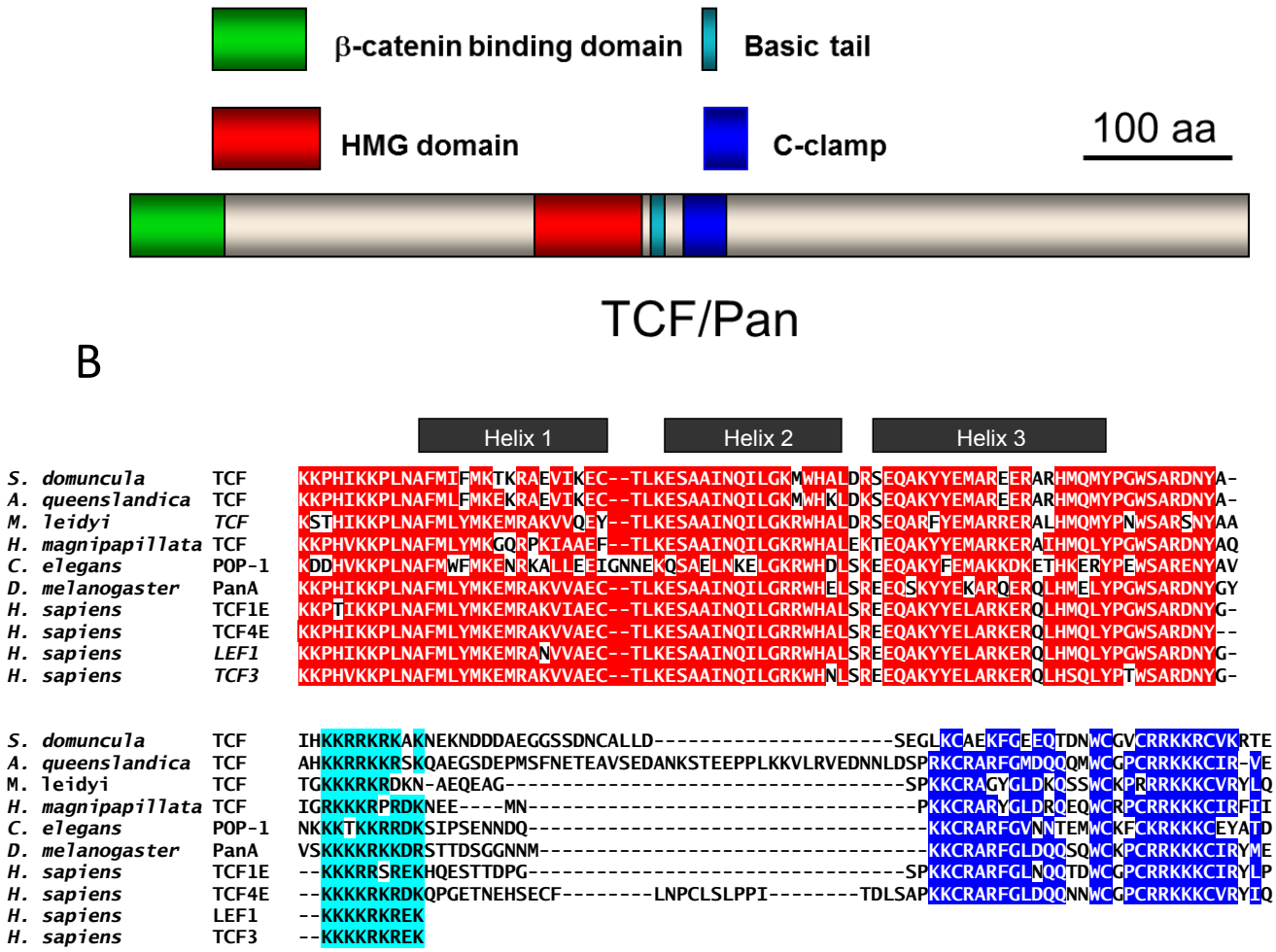


Figure 1.3 Conservation of the HMG and C-Clamp domains among metazoans. (A) Cartoon depicting Drosophila TCF/Pan (the PanA isoform; 751 aa) showing the location of the b-catenin binding domain (green), the HMG domain (red), the basic tail (aqua) and the C-clamp (blue). (B) Alignment of the HMG domains, basic tails and C-clamps among metazoan TCFs. Non-conserved residues are not coloured in the alignment. The positions of the three a-helices of the HMG domain, based on the structure of LEF1 are indicated at the top of the figure. The six invertebrate TCFs possess all three domains, while only the E box isoforms of vertebrate TCF1 and TCF4 possess C-clamps. The degree of conservation in the HMG domain is quite high, e.g. the TCF of Suberities domuncula and human TCF4E are 79.5% identical, 85.9% conserved. The C-clamp is less conserved (55.2% identity; 58.6% for the S. domuncula-human TCF4E comparison). The number of non-conserved residues between the basic tail and C-clamps are highly variable. The GenBank accession number of each protein sequence is in parentheses: S. domuncula (CAH04889.1); Amphimedon queenslandica (ADO16566.1); Mnemiopsis leidyi (ADO34164.1); Hydra magnipapillata (XP_002159974.1); Caenorhabditis elegans (NP_491053.3); Drosophila melanogaster (isoform A; NP_726522); human TCF1E (EAW62279.1); TCF4E (CAB97213.1); LEF1 (NP_001124185) and TCF3 (NP_112573.1). TCF, T-cell factor 1; HMG, high mobility group; LEF1, lymphoid enhancer-binding factor 1. Adapted from Archbold et al. 2012.

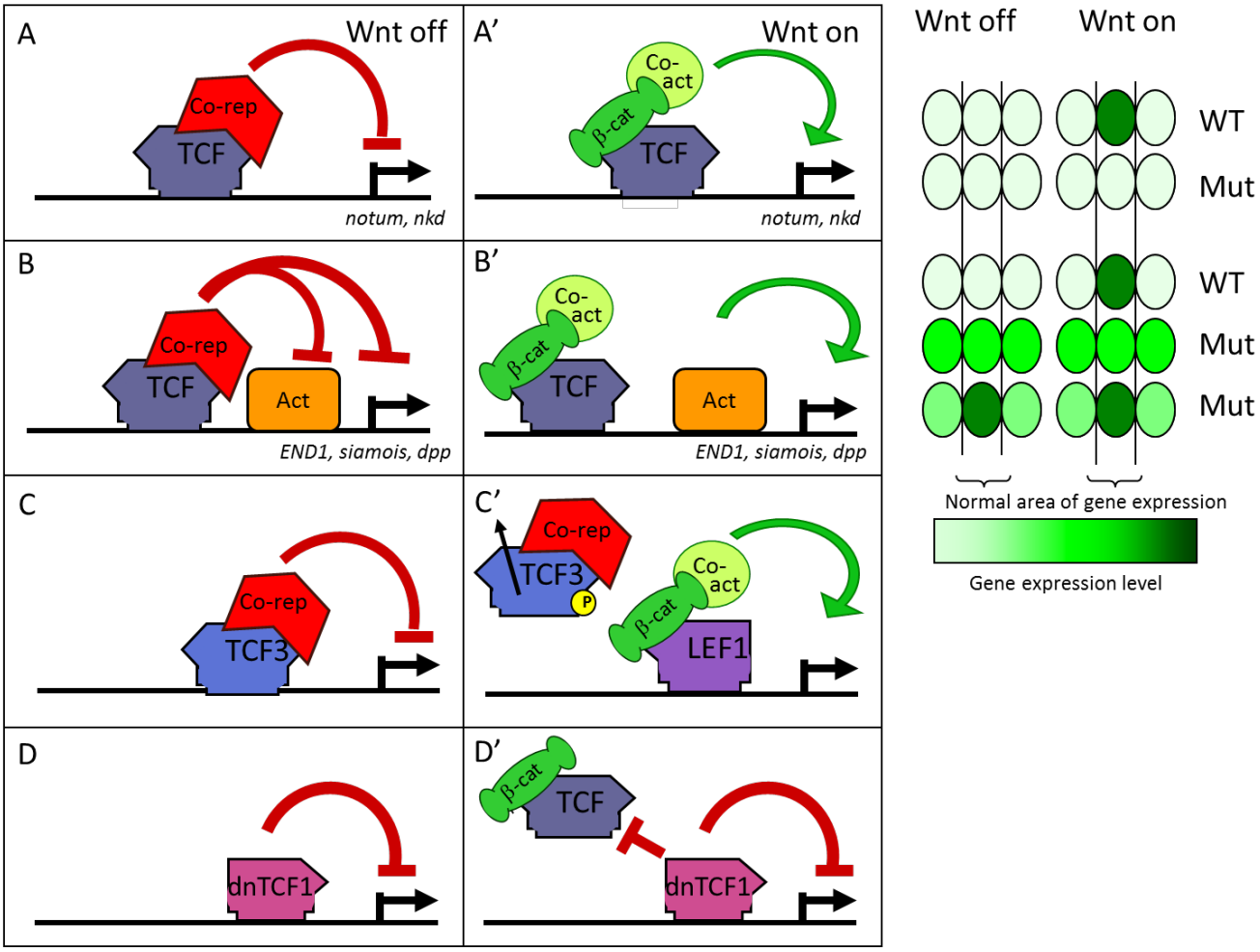


Figure 1.4 TCF acts as a transcriptional switch.
(A) TCF interacts with co-repressors in the absence of signaling. When b-catenin is stabilized by Wnt signaling, it is able to displace repressors and recruit co-activators to turn on target gene transcription. In W-CRMs such as nkdIntE or notumUPEB, which lack other basal activating elements, mutation of TCF binding sites results in a loss of activation, but no change in repression (panel E). (B) In W-CRMs which contain other activating elements, such as END-1, siamois or eve, mutation of TCF sites results in decreased activation in areas with active Wnt signaling, while de-repression is seen in other areas (panel F, mutant A). In some W-CRMs with potent activator elements, mutation of the TCF site results primarily in loss of repression, with minimal effect on activation, for instance, in the visceral mesodermal W-CRMs for dpp (panel F, mutant B). (C) In vertebrates, some instances of the transcriptional switch may be accomplished by the exchange of a repressive TCF, ie. TCF3, for an activator such as TCF1 or TCF4. Phosphorylation events may be responsible for altering DNA binding affinity in this model. (D) The transcriptional switch can be blocked by dominant negative isoforms of TCF which lack b-catenin binding domains. Adapted from Archbold et al. 2012

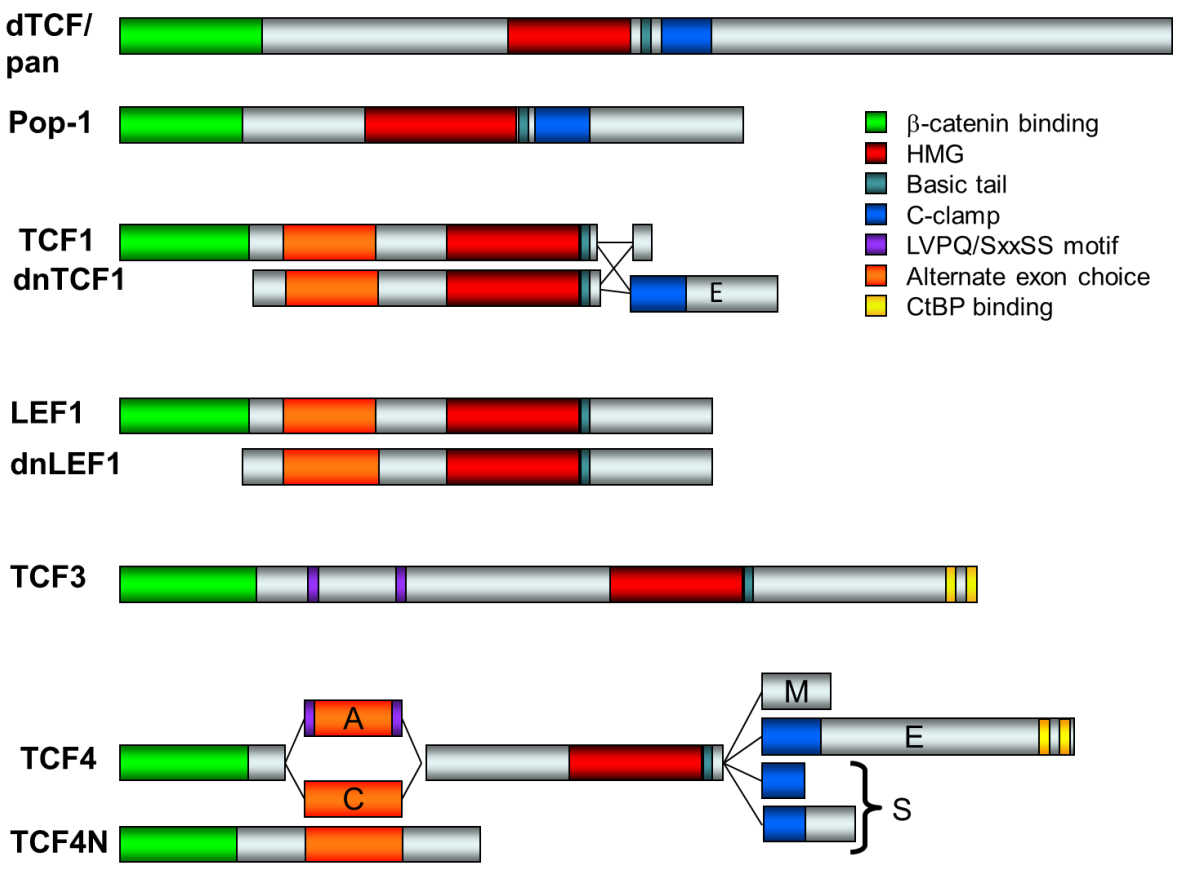


Figure 1.5 Diversity of TCF/LEFs.
Invertebrates contain a single TCF member containing the β -catenin binding (green), HMG (red), basic tail (turquoise), and C-clamp (blue) domains. Pictured here is the most abundant isoform in Drosophila (Pan A) and the C. elegans POP-1. In vertebrates, alternate promoter usage and alternative splicing result in a myriad of TCF isoforms with diverse functional properties. Alternate usage of downstream promoters can result in isoforms which lack the β -catenin binding domain, and function as natural dominant negatives, such as dnTCF1 and dnLEF1 (Roose et al. 1999, Hovanes et al. 2001). Alternate exon usage (orange) occurs in all family members except TCF3, and the LVPQ/SXXSS motif (purple) which is invariant in TCF3 confers repressive activity on TCF4 isoforms which contain it (as in TCF4A) (Liu et al. 2005). Inclusion of the C-clamp motif is seen in E-tail containing isoforms TCF1E and TCF4E. M isoforms lack the C-clamp, while S isoforms contain truncated C-clamp domains (Weise et al. 2010). Some TCF3 and TCF4 isoforms also contain CtBP binding sites. TCF, T-cell factor; LEF, lymphoid enhancer-binding factor; HMG, high mobility group. Adapted from Archbold et al. 2012

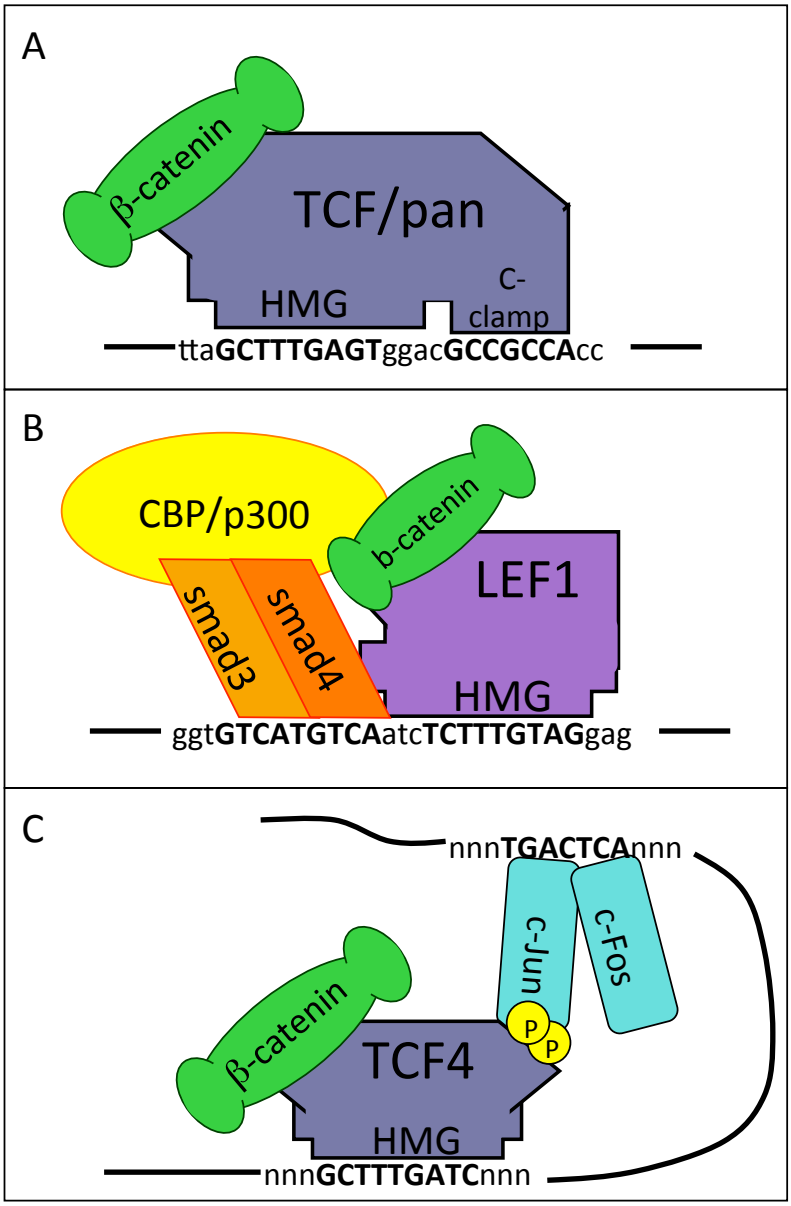


Figure 1.6 Three different mechanisms that contribute to TCF target selection in the nucleus.
(A) Bipartite binding of TCF/Pan with HMG domain–HMG site and C-clamp–Helper site interactions at a binding site in the intronic W-CRM from *nkd*. This strategy increases the TCF recognition site to approx. 16 bp. (B) Combinatorial binding between LEF1 and a Smad heterodimer on the twin WRE in Xenopus. The adjacent location of the Smad and TCF binding site again increases the amount of basepairs required for binding. Smads and β-catenin are also thought to cooperate in recruiting p300/CBP to TGFβ regulated W-CRMs. (C) In the case of the *c-jun* and *c-myc* regulatory regions, the TCF and AP-1 sites are not near each other, suggesting a model where DNA looping is stabilized by interactions between c-Jun and TCF. Adapted from Archbold et al. 2012

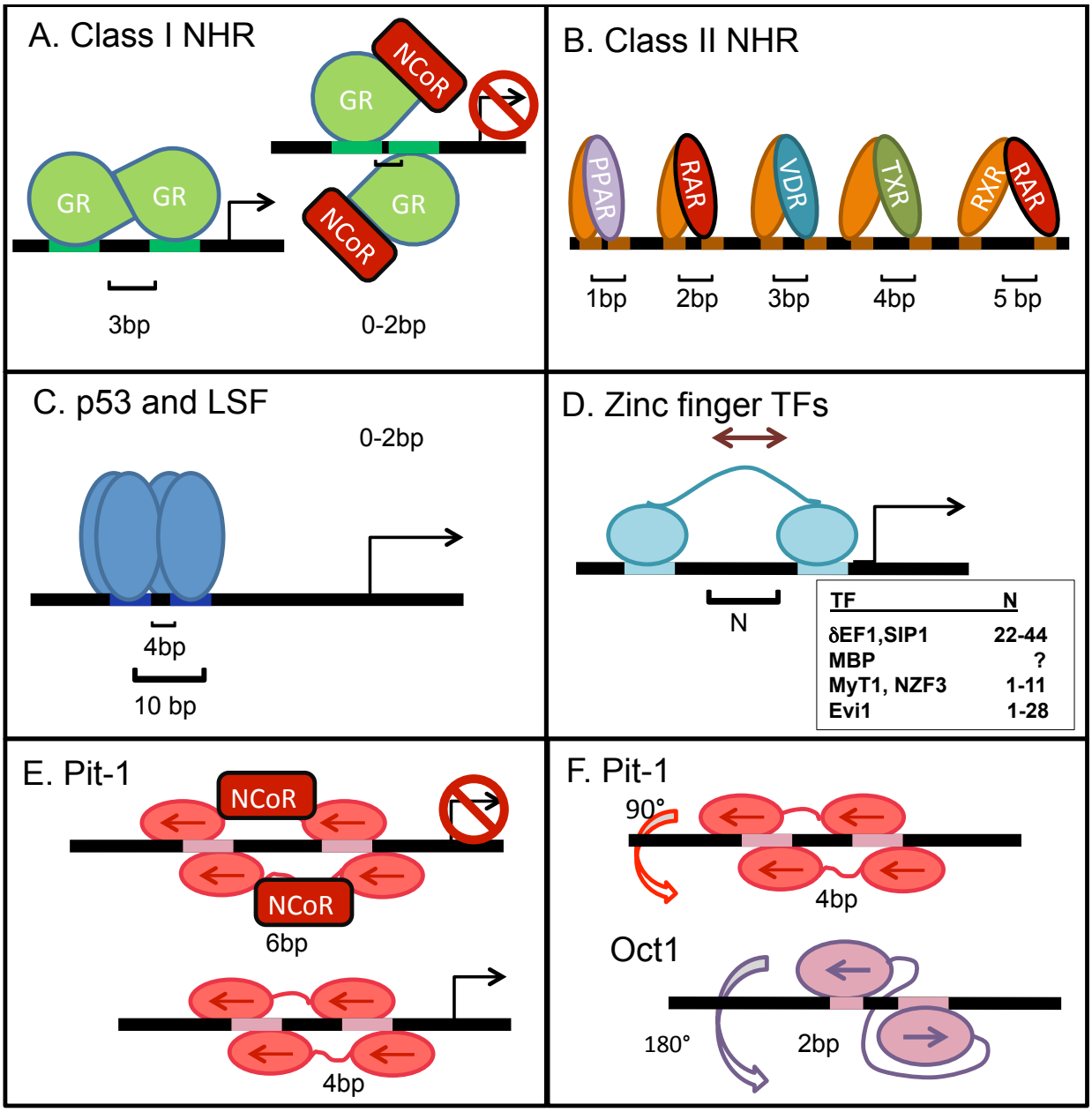


Figure 1.7 Bipartite binding proteins.

Figure 1.7 Bipartite binding proteins.
(A) Glucocorticoid receptor dimerizes at Invert Repeats (IR) separated by 3 bp. IRs with shorter spacers (0-2 bp) are bound by monomer which recruit co-repressors like NCoR. (B) The "1-5 rule": class II NHR heterodimers composed of RXR and PPAR, RAR,VDR,TXR and RAR preferentially bind bipartite motifs with 1,2,3,4 and 5 bp intersite spacing. This rule is not invariant, see text for details. (C) p53 tetramers bound to direct repeats with 4 bp intersite spacing activate transcription. Inverting the motifs, or increasing or decreasing this spacer results in repression. The related factor LSF has a similar spacing requirement. (D) Zinc finger cluster TFs contain two DNA binding domains separated by a flexible linker. These factors can bind invert or direct repeats with a wide range of spacers (N). Range of functional spacer length for each TF indicated in box. (E) Pit-1 bound to the GH element with a 6 bp spacer binds NCoR and represses GH. Decreasing spacing to 4bp activates the target instead. (F) Pit1, with a 15 aa linker, binds to the prolactin element (4 bp spacing) as a homodimer. Each subunit binds a single perpendicular face of the DNA. Oct1, with a 24 aa linker, binds as a monomer to a 2 bp element by wrapping around the DNA strand and binding opposite faces of the DNA.

TCF Site	Frequency in random intergenic DNA (bp)
CCTTTGAWS	~1/22,300
SCTTTGAWS	~1/10,300
CTTTGWWS	~1/2500
SCTTTGWW, SGTTTGWWS, or SCTTTCWWS	~1/1175
CTTTGWW, GTTTGWW, Or CTTTCWW	~1/645

Table 1.1 The frequency of TCF sites increases with sequence degeneracy. 134 kb of Human intergenic DNA was searched for potential TCF sites. Target Explorer was used to create a weighted matrix of high to low quality TCF sites. While TCF sites matching the high affinity consensus (CCTTTGAWS) are relatively rare, increasing the level of degeneracy causes a rapid amplification in the number of potential sites. Adapted from Archbold et al. 2012.

Chapter II:

Examining TCF bipartite motif architecture *in vitro* and *in vivo*

Abstract

The T-cell factor (TCF) family of transcription factors are major mediators of Wnt/β-catenin signaling in metazoans. All TCFs contain a High Mobility Group (HMG) domain that possesses specific DNA binding activity. In addition, many TCFs contain a second DNA binding domain, the C-clamp, which binds to DNA motifs referred to as Helper sites. While HMG and Helper sites are both important for the activation of several Wnt dependent cis-regulatory modules (W-CRMs), the rules of what constitutes a functional HMG-Helper site pair are unknown. In this report, we employ a combination of *in vitro* binding, reporter gene analysis and bioinformatics to address this question, using the *Drosophila* family member TCF/Pangolin (TCF/Pan) as a model. We find that while there are constraints for the orientation and spacing of HMG-Helper pairs, the presence of a Helper site near an HMG site in any orientation can increase binding and transcriptional response, and some orientations display tissue-specific patterns. We find that altering an HMG-Helper site pair from a low affinity to a high affinity orientation/spacing dramatically increases the responsiveness of a W-CRM in several fly tissues. In sum, this work extends the importance of Helper sites in fly W-CRMs and suggests that the quality of an HMG-Helper pair is a major factor in setting the threshold for Wnt activation and tissue-responsiveness.

Introduction

During metazoan development, Wnt/β-catenin signaling, often called "canonical" Wnt signaling, and hereafter referred to as "Wnt signaling," is required to drive multiple stage and tissue specific events (Archbold et al., 2012; Cadigan and Nusse, 1997; Clevers, 2006; Grigoryan et al., 2008). Wnt signaling is essential in such diverse events as specification of the anterior/posterior body axis, as well as limb, heart, intestinal and craniofacial development (Archbold et al., 2012; Gessert and Kuhl, 2010; Liu and Millar, 2010; Niehrs, 2010; Towers et al., 2012). In several cases, Wnts have been shown to act as morphogens, regulating different targets in a concentration dependent manner (Perrimon et al., 2012; Sanchez-Camacho and Bovolenta, 2009; Swarup and Verheyen, 2012). The pathway is also needed in adult tissues for stem cell maintenance and wound healing (Arce et al., 2006; Barker et al., 2013; Holland et al., 2013; Lim and Nusse, 2013; Sato and Clevers, 2013). Disregulated Wnt signaling has been implicated in a host of cancers and other human pathologies (Joiner et al., 2013; Polakis, 2012; Regard et al., 2012). How a single signaling pathway accomplishes such a wide range of outcomes remains a major question in developmental biology and tissue homeostasis.

Variations in Wnt-dependent *cis*-regulatory modules (W-CRMs) likely contribute to the diversity of Wnt transcriptional responses, though the mechanisms are poorly understood. Members of the T-cell factor (TCF) family of transcription factors (TFs) are principal mediators of Wnt signaling (Brantjes et al., 2002; Cadigan, 2012). In many contexts, TCFs act as a transcriptional switch, binding with co-repressors on W-CRM

chromatin in the absence of signal, and then recruiting β-catenin and other co-activators in response to Wnt signaling (Cadigan, 2012; Valenta et al., 2012). ChIP-seq studies have found that TCFs co-localize with several other TFs in specific cell types (Blahnik et al., 2010; Bottomly et al., 2010; Cole et al., 2008; Frietze et al., 2012; Junion et al., 2012; Trompouki et al., 2011; Verzi et al., 2010; Wu et al., 2012), and combinatorial control may be one method to achieve tissue or temporal specificity. While not as well appreciated, the sequence composition of the TCF binding sites in W-CRMs can also have a major influence on its transcriptional output (Blauwkamp et al., 2008). A better understanding of the *cis*-regulatory logic of W-CRMs will shed more light on how they differ in their responsiveness to Wnt signaling, and how TCFs regulate this process.

All TCFs share a highly conserved High Mobility Group (HMG) domain, which binds DNA with sequence specificity (Giese et al., 1991; Hallikas et al., 2006; van Beest et al., 2000; van de Wetering et al., 1991). The HMG recognition motif is a 9-11 bp sequence with the consensus SCTTTGWWSWW. Sequences roughly conforming to this consensus have been shown to be required for activation of numerous W-CRMs (Archbold et al., 2012; Barolo, 2006). Reporter genes with 3 to16 copies of high affinity HMG binding sites behind a basal promoter, such as TOPFLASH, have been used successfully as an experimental readout for Wnt signaling in numerous contexts (Barolo, 2006; DasGupta et al., 2005; Korinek et al., 1997; Lum et al., 2003). However, such high-density clusters of perfect HMG sites are not found in naturally occurring W-CRMs (Archbold et al., 2012; Barolo, 2006). Furthermore, there are several instances where synthetic HMG site reporters do not respond to endogenous Wnt signaling in vertebrate tissues (Dessimoz et al., 2005; Geng et al., 2003). In *Drosophila* embryos

and larval imaginal discs, where Wingless (Wg, a fly Wnt) signaling is highly active, synthetic HMG site reporters have little or no expression (Barolo, 2006; Chang et al., 2008b). These results strongly suggest that under certain physiological conditions, HMG sites are not sufficient for Wnt activation of W-CRMs.

We have previously reported that several fly W-CRMs contain a GC-rich motif, found near HMG sites, that is critical for Wnt activation (Chang et al., 2008b). This motif, termed the Helper site, is bound by a second DNA-binding domain in TCF/Pangolin (TCF/Pan, the fly TCF) known as the C-clamp (Chang et al., 2008b). The C-clamp was originally discovered in "E-tail" isoforms of mammalian TCF1 and TCF4 genes (Atcha et al., 2007). These TCF isoforms also bind Helper sites, which are essential for the activation of specific mammalian W-CRMs (Atcha et al., 2007; Hoverter et al., 2012; Wallmen et al., 2012). Reporters containing only multimerized copies of Helper sites cannot respond to Wnt signaling, but these motifs synergize with HMG sites to greatly enhance the Wnt activation of reporter constructs (Chang et al., 2008b). The presence of an intact C-clamp domain imparts increased affinity for DNA containing both HMG and Helper sites and a functional C-clamp is required for TCF/Pan activation of fly W-CRMs (Chang et al., 2008b; Ravindranath and Cadigan, 2014). These data support a bipartite binding model for C-clamp containing TCF family members, where HMG domain-HMG site and C-clamp-Helper site interactions allow TCF to properly locate W-CRMs and regulate Wnt target genes.

Surprisingly, our initial characterization of Helper site sequences in *Drosophila* W-CRMs identified numerous putative Helper elements with variable spacing and orientation with respect to HMG sites (Figure 2.1A). This was interesting because

bipartite binding by TFs is often sensitive to the spacing and orientation of the two sites. Examples of this spacing/orientation constraint include several type II nuclear receptor/RXR heterodimers (Katz and Koenig, 1994; Naar et al., 1991; Phan et al., 2010) and Smad heterodimers (Pyrowolakis et al., 2004; Weiss et al., 2010). Spacing and orientation is also important for the POU family member Pit-1 (Dawson et al., 1996; Scully et al., 2000), where the spacing of half-sites has been shown to determine whether target genes are activated or repressed. In contrast, the related zinc finger DNA binding proteins SIP1 and δEF1 have a high tolerance for half-site spacing and orientation variability, perhaps because the two DNA-binding zinc finger clusters are separated by a large and presumably flexible linker region (Remacle et al., 1999). Given the short (10 aa) spacer between the HMG and C-clamp domains in TCF/Pan, it was unclear whether all the variable HMG-Helper site pairs found in W-CRMs were *bona fide* TCF binding sites. As no consistent organizational preference was seen among the functional HMG and Helper sites (asterisks, Figure 2.1A), a systematic approach was needed to determine the constraints of HMG-Helper pair flexibility.

In this work, we examine the rules of TCF/Pan binding to HMG-Helper site pairs using several experimental approaches. We identified two HMG-Helper site configurations that were bound by TCF/Pan with highest affinity *in vitro*. In one case the Helper site is located 6 bp upstream of the HMG site and in the other it is immediately downstream. These two HMG-Helper site configurations also have the greatest transcriptional activation in many tissues, and are most enriched in genomic regions bound by TCF/Pan. We suggest a model where the DNA-bending activity of the HMG domain enables TCF/Pan to recognize both these HMG-Helper site

configurations. However, our data also makes it clear that the presence of a Helper site near an HMG site in any orientation and with variable spacing can enhance TCF/Pan binding, and many of these “non-optimal” arrangements have transcriptional activity, some with striking tissue-specificity. In addition, we have shown that altering the orientation and spacing of an HMG-Helper site pair in a W-CRM has a dramatic effect on its sensitivity to the Wg morphogen in imaginal tissues.

Results

HMG and Helper Sites Work in Pairs

The *Drosophila* Helper site was previously defined by sequence alignment of several functional motifs as having the consensus GCCGCCR (R=A/G) (Chang et al., 2008b). However, a shorter consensus has been reported for vertebrate E-tail TCFs (RCCG) (Atcha et al., 2007). To test whether all seven nucleotides of the longer consensus were required for maximal activation, we performed serial mutagenesis on the second Helper motif in the *nkd*IntE W-CRM luciferase reporter (Figure 2.1B). This reporter is highly activated by expression of a constitutively active form of Armadillo (Arm, the fly β-catenin), which contains a point mutation rendering it resistant to degradation (Arm*) (Chang et al., 2008a; Chang et al., 2008b). Substitution of any of the first four positions had as dramatic a reduction in reporter activation as mutating the entire 7 bp motif. Mutation of the last three positions had a slightly less severe reduction (Figure 2.1B). Thus, at least in this context, all seven bp in the GCCGCCR motif are important for maximal activation by Wnt signaling.

Several previous pieces of evidence support the idea that HMG and Helper sites work in closely spaced pairs. The contribution of individual HMG sites to W-CRM activation varies widely, with HMG sites proximal to Helpers sites more likely to contribute to activation (Chang et al., 2008a; Chang et al., 2008b). In addition, mutation of a Helper site typically has a comparable effect on W-CRM activation as the mutation of the nearest HMG site (Chang, Chang, Archbold and Cadigan, unpublished observations). To further test this hypothesis, we again used the *nkd*IntE W-CRM, previously found to contain three functional HMG binding sites, and two functional Helper sites (Chang et al., 2008b). The arrangement of these functional sites suggests that there are two closely spaced HMG-Helper site pairs, separated by 101 bp (Figure 2.1A), but there remained a formal possibility of longer-range interactions between HMG and Helper sites.

As previously reported (Chang et al., 2008b), activation by Arm* is nearly abolished by mutation of the three HMG binding sites (Figure 2.1C). Four additional *nkd*IntE mutants were created, leaving one HMG and one Helper site intact. The two constructs retaining an HMG site and Helper site in close proximity were able to activate target gene transcription at levels higher than the HMG mutant control. The first pair had a small but reproducible activation, while the activation of the second intact pair was more pronounced (Figure 2.1C). In contrast, the reporters where the intact HMG and Helper sites were separated (P1Dist & P2Dist) were not activated. These data support the idea that HMG and Helper sites must be in close proximity to respond to Wnt signaling.

TCF/Pan Prefers Specific HMG-Helper Configurations *in Vitro*

There are four possible orientations for HMG-Helper site pairs, which we have termed Akimbo (AK), Rewind (RW), Fast Forward (FF), and Knock Knee (KK) (Figure 2.1D). Helper sites are defined by the aforementioned seven bp GCCGCCR consensus (Figure 2.1B). We used the eleven bp consensus of SCTTTGWSWW determined for TCF/Pan (van de Wetering et al., 1997) to define HMG sites. It should be noted that the four orientations indicate the relationship between the HMG and Helper sites, and not the relationship of these bipartite motifs to the nearest Transcriptional Start Site (TSS). Therefore, it is possible to have either the Helper or HMG site first in all four orientations, depending on which strand contains the consensus (Figure 2.1D). The spacing of each pair is defined by the number of bp between the two motifs, e.g., the examples in Figure 2.1D have a spacing of 6 bp and will hereafter be referred to as AK6, FF6, etc.

We previously reported that the presence of a Helper site increases the ability of TCF/Pan to bind to DNA *in vitro* (Chang et al., 2008b). These experiments utilized an AK5 HMG-Helper site configuration. To determine the relative binding affinities of different HMG-Helper site pairs, we performed EMSAs with a recombinant His-tagged protein containing both the HMG and C-Clamp domains of TCF/Pan, a labeled AK6 probe (see Figure 2.1D for sequence) and unlabeled competitor oligonucleotides containing the 0 and 6 bp versions of each orientation. The AK6 probe was labeled with an infrared (IR)-dye, allowing quantification of the gel shift with the LiCor Odyssey IR platform (see Materials and Methods for further details). Representative blots are presented (Figure 2.2A) and the data from multiple experiments were summarized by

showing the half maximal inhibitory concentrations (IC_{50}) for each competitor (Figure 2.2B) and the dose-response curves on semi-log line graphs (Figure 2.2C).

The competition assays clearly show that TCF/Pan has a preference for oligonucleotides containing an AK6 or FF0 motif. The IC_{50} for AK6 and FF0 are 5.1 and 9.4 nM, respectively (Figure 2.2B). RW6, KK0, FF6 and AK0 are in the next group, with IC_{50} values between 38.7-66.6 nM. KK6 and RW0 have the lowest relative affinity (IC_{50} of 99.7 and 189 nM, respectively). The data indicate that AK6 and FF0 are bound with the greatest affinity by TCF/Pan, but also demonstrate that the presence of a nearby Helper site in any orientation enhances its recognition by TCF/Pan. It should be noted that this experiment was done under ligand depleted conditions. While the experimentally determined IC50 values give us an estimate of the relative affinities, calculating K_i values would require a mathematical correction (Munson and Rodbard, 1988) or, preferably, experimental conditions in which the free and total ligand concentrations are approximately equal. The IC50 values for RW0 and KK6 are lower than for the two HMG site only oligonucleotides, indicating they may be bound with higher affinity. However, the large range in values for the 95% confidence intervals (Figure 2.2B) indicates additional measurements would be required to conclude this with certainty.

How can the HMG-C-clamp of TCF/Pan bind to HMG-Helper pairs in all four orientations? One possibility takes into account the ability of the HMG domain to bend DNA. Structural data indicates that the HMG domain of human LEF-1 imparts an approximate 117° ± 10° bend in the DNA, with three alpha helixes in the domain making contact in the minor groove of the DNA while the basic tail region makes contact with

the major groove on the inside of the bend (Love et al., 1995)(Figure 2.3). The HMG domain and basic tail of TCF/Pan share 96% conservation with LEF-1 (Archbold et al., 2012), and TCF/Pan can significantly bend DNA in solution (Zhang et al., 2014). The C-Clamp is located 10 amino acids C-terminal to the basic tail in TCF/Pan (Archbold et al., 2012), which may place the C-clamp in the interior of the DNA bend, allowing it to "swing", and interact with Helper sites located either "upstream" of the HMG binding site (AK) or "downstream" (FF) (Figure 2.3).

The bend in DNA by the HMG domain is centered between positions 3 and 4 in the 11 bp consensus binding site (Love et al., 1995)(Figure 2.3). This asymmetry may provide mechanistic insight into the logic behind the optimal placement of Helper sites. The preferred configurations for TCF/Pan binding are FF0 and AK6, where the Helper sites are roughly equidistant from the HMG imposed bend, within reach of the C-clamp (Figure 2.3).

The ability of the HMG-C-clamp of TCF/Pan to recognize KK and RW configurations is likely due to the semi-palindromic nature of the Helper site. For example, the KK0 sequence (HMG site-TGGCGGCG) can also be viewed as a degenerate FF1, with a C to G substitution at positions 2 and 5 of the Helper site (Figure 2.1D). The same is true for the RW configuration (e.g., RW0 is a degenerate AK1). Viewed in this way, the IC_{50} data becomes more coherent, with the FF and KK configurations ranked FF0 > KK0 ≈ FF6 > KK6 in terms of affinity for TCF/Pan and the AK and RW ones ranked AK6 > RW6 ≥ AK0 > RW0 (Figure 2.2B). These data support a model where the HMG and Helper sites in the FF/KK orientations are preferentially

bound when adjacent, while the preferred spacing for AK/RW is further removed (e.g., 6 bp).

HMG-Helper Site Configuration Preferences in Cell Culture Assays

To explore the functional orientation/spacing constraints between various HMG-Helper site configurations, we created a series of synthetic W-CRMs containing two HMG-Helper site pairs upstream of a minimal promoter. All four orientations were tested for the ability to activate a luciferase reporter gene at 0, 3, 6, 9 and 12 bp spacing in transfected Kc cells. Three out of the four orientations (AK, FF & KK) exhibited levels of activation by Arm* higher than reporters containing HMG sites alone (Figure 2.4A). Spacing of HMG-Helper pairs affected the level of activation in an orientation-dependent manner. The AK reporters were significantly different from the HMG site only reporters at most spacings tested, but peak activity occurred with AK6 (Figure 2.4A). In contrast, for FF, activation was greatest at 0 bp spacing, with much weaker activation at greater distances. The KK orientation constructs showed weak activation at several spacings, though activation was slightly greater when the HMG and Helper sites were closer together. Unlike the other three orientations, activity of the RW reporters were not significantly different than the relevant HMG control at any of the spacings tested (Figure 2.4A).

To explore the spacing requirements of the AK and FF HMG-Helper site pairs in the context of endogenous enhancers, we chose two previously characterized W-CRMs from the *nkd* locus. First, we used a modified *nkd*IntE, termed *nkd*IntEP2P, where the first two HMG sites and Helper are mutated, leaving only the endogenous AK6 motif

(Figure 2.4B). We replaced this motif with either AK or FF motifs containing 0, 6, or 12 bp spacers. In this context, AK6 promoted the most robust activation, while the AK0 and AK12 constructs had lower levels of activation, consistent with the behavior of the synthetic constructs. Also consistent with the synthetic data, FF0 was the only spacing of the FF *nkd*IntEP2P constructs to activate at levels significantly different than the HMG only control (Figure 2.4B).

We then examined a second W-CRM, *nkd*UPE2, previously shown to have a specific HMG and Helper site that were major contributors to Wg activation (Chang et al., 2008a)(Chang, JL and Cadigan KM, unpublished observations). This HMG-Helper pair (green box in Figure 2.4C cartoon) has a degenerate FF1/KK0 configuration. Mutation of the HMG site results in a dramatic decrease in activation by Arm* (Figure 2.4C). We altered this HMG-Helper site pair to an AK1, AK6, or FF6 configuration. The AK motifs were more flexible in its range of functional spacing, as both AK1 and AK6 containing W-CRMs activated transcription as robustly as the WT FF1 element (Figure 2.4C). The FF motif displayed a strong preference for the 1bp spacer configuration, with strongly decreased activation from the FF6 element (Figure 2.4C). However, the FF6 motif retained some activation by Arm*, as compared to an HMG site mutant (Figure 2.4C). The data with the *nkd*IntE and *nkd*UpE reporters indicate that the configurations that work well (*e.g.,* AK6, FF0) in the synthetic reporters in cell culture (Figure 2.4A) also are optimal for the *nkd* W-CRMs in cell culture reporter assays.

HMG-Helper Site Synthetic Reporters Reveal Tissue-Specific Expression in *Drosophila* Tissues

To test whether the functional constraints for HMG-Helper site configurations observed in cell culture assays also held true in the context of an intact organism, transgenic reporter lines with different HMG-Helper pairs were generated in *Drosophila*. ΦC31 site directed integration of reporter constructs was utilized to eliminate position effects (Bischof et al., 2007). All four orientations at 0 and 6 spaces were tested, as these HMG-Helper pairs displayed distinct outputs in cell culture (Figure 2.4A). In imaginal discs from wandering 3rd instar larva, the HMG site only reporters had no detectable expression over the basal *Hsp70* promoter (Figure 2.5A, 2.5B). In contrast, several HMG-Helper site reporters had expression patterns consistent with activation by Wg signaling (Archbold et al., 2012; Baker, 1988a; Couso et al., 1993; Phillips and Whittle, 1993). Consistent with our cell culture data, the most potent activity was seen with both the AK6 and FF0 HMG-Helper pairs (Figure 2.5A, 2.5B). Other configurations (AK0, FF6, KK0) displayed weaker expression. Interestingly, with the exception of RW0, the presence of Helper sites in all other conformations tested displayed more activity than the HMG site only controls in the imaginal discs (Figure 2.5A, 2.5B, 2.6G). These results indicate that Helper sites have a surprising degree of flexibility in potentiating the ability of HMG site to respond to Wg signaling.

While the AK6 and FF0 synthetic reporters displayed the most activity in imaginal discs, there were tissue-specific differences in their expression. AK6 was the most robust responder to Wg signaling in wing imaginal discs (Figure 2.5A), while FF0 had the highest expression of all reporters in eye/antennal discs (Figure 2.5B). In some

non-imaginal tissues, the other two orientations displayed the highest level of activation. For example, RW0 drove robust expression in the larval epidermis, in the cells underlying the naked cuticle located between denticle belts (Figure 2.6C), while other favorable configurations, like FF0, had less expression (Figure 2.6B). In addition, the AK6 reporter had extremely weak expression in the Corpora Allata (CA, also known as the medial secretory cells) of the ring gland (Figure 2.6E), while KK6 was expressed at much higher levels (Figure 2.6F).

A summary of all the collected expression data from the eight HMG-Helper site reporters is shown in Figure 2.6G. FF0 and AK6 are clearly the strongest reporters in imaginal discs and have intermediate expression in the body wall. However, they are weakly expressed in the CA. Strikingly, RW0, which has no detectable expression in the imaginal discs, has high expression in the body wall and CA. KK0 & KK6 have weak expression in the discs, no activity in the body wall and the highest expression in the CA (Figure 2.6G). These data suggest the possibility that altering HMG-Helper site architecture may be a way to create a repertoire of tissue-specific responses to Wg signaling.

TCF/Pan-bound Embryonic Chromatin is Enriched for Optimal HMG-Helper Site Configurations

The *in vitro* DNA binding assays described earlier (Figure 2.2) are a reductionist approach to understanding HMG-Helper site recognition by TCF/Pan. An alternative is to determine whether HMG-Helper site pairs are enriched in genomic sequences bound by TCF/Pan. A genome-wide survey of TCF/Pan localization has been performed in

germband extended *Drosophila* embryos (Junion et al., 2012). At this developmental stage, Wg signaling is patterning the embryonic epidermis and mesoderm (Bejsovec and Martinez Arias, 1991; Klingensmith and Nusse, 1994; Lawrence et al., 1995; Park et al., 1996). For one time point (6-8 hr after fertilization), 2079 high confidence TCF/Pan peaks were identified (Junion et al., 2012). The DNA covered by these TCF/Pan peaks (~2.9 x 10^6 bp) was analyzed for HMG-Helper pairs and compared to randomly selected intronic and intergenic DNA.

To analyze these genomic sequences, we created a program to identify HMG and Helper site pairs, which could then be sorted for orientation and distance (see Materials and Methods). Position Weight Matrices (PWMs) of each motif were created from the collection of functional HMG and Helper sites we have identified (Chang et al., 2008a; Chang et al., 2008b)(Figure 2.7). This allowed us to analyze DNA sequences using different stringencies for calling HMG and Helper sites. We considered PWM values of 4.5 for HMG sites and 6.5 for Helper sites to be a fairly stringent criteria for these motifs, while 3.5 and 5.0 (for HMG and Helper sites respectively) was a more relaxed calling criteria.

Regardless of the criteria used, HMG-Helper pairs are enriched in the TCF/Pan bound regions. With the stringent criteria, pairs with 0-15 bp spacers were 3.48 times more likely to occur in bound peaks than in random DNA (Figure 2.8A). This enrichment level is considerably higher than that obtained for HMG sites only (1.46 times enriched in bound DNA) or for the Helper sites, which were underrepresented in bound DNA (0.76 times) compared to random DNA. Using the relaxed criteria for

calling motifs, many more HMG-Helper sites are identified (2139 versus 448), and they are 2.4 fold enriched in TCF/Pan bound versus random DNA (Figure 2.8B).

A closer look at the spacing between HMG-Helper pairs in all four orientations reveals two general messages. First, the enrichment over random DNA is most pronounced in configurations that were favorable for *in vitro* binding and/or transcriptional activity in cell culture and imaginal discs. For example, at the stringent calling criteria, FF0-2 and AK0-6 pairs were 6.1 times as likely to be found in TCF/Pan bound compared to random DNA (Figure 2.8A). Second, despite the first point, it is also true that HMG-Helper sites in every orientation at almost every spacing are enriched in TCF/Pan bound DNA (Figure 2.8A), and this is also true at the more relaxed criteria for calling motifs (Figure 2.8B). It should also be noted that there were a number of palindromic motifs (e.g. YGCCGGCR) that were double called, either as both AK and RW or as both FF and KK. These pairs are represented as the overlapping area in the Venn diagram (Figure 2.8A).

In addition to examining TCF/Pan localization in the *Drosophila* genome, Junion and co-workers surveyed four other TFs involved in cardiogenesis: the GATA factor Pannier, phosphorylated Mad (pMAD), Tinman (Tin) and Dorsocross (Doc) (Domenzain-Reyna et al.). They found that many genomic locations contained several of these TFs, which often contained functional W-CRMs that were active in cardiac or mesodermal cells (Junion et al., 2012). To determine if the frequency of HMG-Helper site pairs was different at sites where TCF/Pan co-localized with these TFs, we partitioned the TCF/Pan bound peaks into those in which the peak center was within 150 bp of another TF's peak, and those in which the center was not within 150 bp of any

of the tested TFs. We called this latter class of peaks "TCF unique," though this is only known for the TFs included in the analysis. This caveat aside, it is still interesting to note that FF0-2 and AK0-6 pairs are 16.25 times more likely to be found in the TCF unique peaks compared to random DNA, while these motifs are less enriched in the peaks shared with Pannier (4.42 fold) and pMad (3.78 fold) (Figure 2.9B). Even less enrichment was observed in the peaks TCF/Pan shared with Tinman and Dorsocross (3.07 & 1.80 fold, respectively) (Figure 2.9B). These data suggest that the mechanism(s) for recruitment of TCF/Pan to chromatin differs depending on the prevalence of co-localizing TFs.

Altering HMG-Helper Site Architecture Increases W-CRM Sensitivity to Wg Signaling

We next wanted to test if we could alter the activity of an endogenous W-CRM *in vivo* by replacing a suboptimal HMG-Helper site pair with an "optimal" configuration. *nkd*UPE2 was a good candidate, since this W-CRM is active in the imaginal discs (Chang et al., 2008a; Chang et al., 2008b), and contains an endogenous RW4 HMG-Helper site pair (Figure 2.10A, green box) which contributes weakly to activation by Wg signaling in cell culture (Chang et al., 2008a)(Chang JL and Cadigan KM, unpublished data). The RW4 motif was reconfigured to an AK6 pair through site-directed mutagenesis (Figure 2.10A). Strikingly, this "optimized" W-CRM reporter displayed increased expression in the wing, haltere and eye/antennal imaginal discs (Figure 2.10B-D'). The domain of reporter gene expression is also increased in the wing discs

(arrows in Figure 2.10B, B'), suggesting that the optimized W-CRM has greater sensitivity to the secreted Wg morphogen.

Discussion

The Rules of TCF/Pan Binding to HMG-Helper Site Pairs

Previous work has shown that TCFs containing C-clamp domains recognize two distinct DNA sequence motifs; HMG sites (via the HMG domain) and Helper sites (via the C-clamp) (Atcha et al., 2007; Bhambhani et al., 2014; Chang et al., 2008b; Hoverter et al., 2012; Ravindranath and Cadigan, 2014). The close proximity of these motifs suggested that they act as HMG-Helper site pairs, which we confirmed through site-directed mutagenesis (Figure 2.1C). Since HMG and Helper sites are often clustered in W-CRMs (Figure 2.1A), it was not readily apparent what orientation and spacing constraints exist for these sites to form a functional bipartite TCF binding site. In this report, we employed a variety of approaches to determine which HMG-Helper site configurations enhanced TCF/Pan binding *in vitro* and *in vivo*, and which ones allowed transcriptional activation by Wnt/β-catenin signaling.

Our analysis revealed that HMG-Helper pairs in the FF0 and AK6 arrangement are preferred in a number of situations. These configurations were bound by TCF/Pan with the highest affinity *in vitro* (Figure 2.2) and were highly enriched in chromatin bound by TCF/Pan in embryos (Figure 2.8). In cell culture, synthetic reporters with FF0 and AK6 pairs were the most highly activated by Wnt signaling (Figure 2.4A). Similar results were also obtained in transgenic reporter assays in several imaginal discs (Figure 2.5, 2.6G). These results demonstrate a strong correlation between DNA binding affinity of

HMG-Helper pairs for TCF/Pan and their ability to mediate Wnt-dependent activation of transcription in several contexts.

While the aforementioned data support the view that some HMG-Helper site configurations are better than others, additional analyses paint a more complex picture. In the context of endogenous W-CRMs, FF1 and AK6 are also the most active in promoting transcriptional activation, but AK1 is just as good in some contexts (Figure 2.4B-C). This dovetails well with the computational analysis of TCF/Pan ChIP-Seq data, where AK0-6 showed the highest enrichment for this orientation (Figure 2.8). However, AK0 shows only moderate affinity *in vitro* (Figure 2.2), similar to other configurations (KK0, FF6, RW6) which have reduced or no functional activity in synthetic reporters in cultured cells (Figure 2.4A) and imaginal discs (Figure 2.5, 2.6G). The correlation between DNA binding affinity and transcriptional activation is poorest in the larval epidermis and CA, e.g., RW0 and KK6. A disconnect between *in vitro* binding affinity and transcriptional activation in cells has also been observed for glucocorticoid receptor (Meijsing et al., 2009). This work and our data demonstrate that some caution is needed when inferring functional significance from *in vitro* binding studies.

Another general lesson from our work is that the presence of a Helper site near a HMG site, no matter the orientation, increases TCF/Pan binding affinity and its ability to mediate Wnt activation of transcription in at least some tissues. This is evident in the EMSA data, where all eight HMG-Helper pairs are bound with greater affinity than HMG sites alone (Figure 2.2), and in TCF/Pan bound chromatin, where enrichment of HMG-Helper pairs is observed over a surprisingly wide array of orientations and spacings (Figure 2.8). This flexibility is also observed functionally in the synthetic reporters,

where HMG site alone had no detectable expression but all eight configurations tested have detectable reporter activity in some tissues (Figure 2.6G).

How can the HMG and C-clamp domains, which are separated by only ten amino acid residues, bind to HMG-Helper pairs with such diversity? We think it likely that DNA bending by TCF/Pan is a major contributor to this flexibility of DNA recognition. Murine LEF1 has been shown to bend DNA more than 110˚ (Love et al., 1995) and TCF/Pan possesses a similar ability (Zhang et al., 2014). Such a bend could allow the C-clamp to reach a Helper site either 5' or 3' of the HMG site (Figure 2.3). The bend is centered between the third and fourth position in the eleven bp HMG site, placing Helpers in the FF orientation further away from the C-terminus of the basic tail (BT) of TCF/Pan (Figure 2.3). This could explain why FF0 is bound preferentially over FFs with larger spacing between the HMG and Helper sites. Conversely, AK6 may be bound with highest affinity (at least *in vitro*) compared to AK0 due to less steric hindrance from the amino acids connecting the BT and the C-clamp (Figure 2.3).

In addition to DNA bending, the semi-palindromic nature of the Helper site likely explains why KK and RW configurations also enhance TCF/Pan binding (Figure 2.2 & 2.8) and have transcriptional activity (Figure 2.4A, 2.5 & 2.6). Defining KK and RW as degenerate FF and AK orientations, respectively, could explain why these motifs mirror the spacing constraints of their reverse configuration partners. It would also explain why they are bound with weaker affinity and typically display less transcriptional activity.

Biological Relevance of HMG-Helper Site Configurations for Wnt Gene Regulation

In the wing imaginal disc, Wg has been proposed to act as a morphogen, forming

a concentration gradient emanating from the dorsal/ventral boundary and regulating target gene expression in a concentration-dependent manner (Baeg et al., 2001; Cadigan et al., 1998; Strigini and Cohen, 2000; Zecca et al., 1996). How W-CRMs differently respond to this Wg morphogen gradient has not been previously investigated. To address this important question, we utilized the *nkd*UPE2 reporter, which is activated in areas of high Wg ligand concentration in the wing disc (Chang et al., 2008a). Replacing a low affinity RW4 motif in this W-CRM with a high affinity AK6 motif elevated the level of reporter gene expression, and broadened the expression domain (Figure 2.10B, B'). These results argue that increasing the affinity of TCF/Pan for the W-CRM increases the sensitivity of the W-CRM to respond to the Wg morphogen.

Our data are reminiscent of classic studies of CRMs that are controlled by gradients of TFs in the syncytial blastoderm stage of *Drosophila* embryogenesis. The affinity of the binding sites for the c-rel homolog, Dorsal, has been shown to set threshold responsiveness in dorsal/ventral patterning, with higher affinity sites being more sensitive to the Dorsal gradient (Jiang and Levine, 1993). Such a correlation was also observed with the Bicoid morphogen gradient using synthetic reporters (Driever et al., 1989; Struhl et al., 1989), though this correlation is not found in endogenous CRMs regulated by Bicoid (Chen et al., 2012; Ochoa-Espinosa et al., 2005; Segal et al., 2008). In contrast, higher affinity sites have been shown to restrict the domain of expression of CRM reporters for the transcription factor Cubitus Interruptus (Ci), an effector of Hedgehog signaling (Parker et al., 2011; Ramos and Barolo, 2013), possibly due to homo-cooperative interactions with the repressive form of Ci (Parker et al., 2011; White et al., 2012). Although Ci and TCF/Pan both act as transcriptional switches, repressing

CRMs until the signaling pathway converts these TFs to activators (Barolo and Posakony, 2002), the relationship between binding site affinity and interpretation of the signaling gradient are diametrically opposed.

Another interesting feature of our work is the tissue-specific responses of our synthetic HMG-Helper site reporters in transgenic fly tissues. In imaginal discs, the strength of expression of these reporters was largely correlated with binding affinity (Figure 2. 5, 2.6G). However, low affinity RW and KK motifs, which had little or no activity in imaginal tissues, drove robust expression in the larval epidermis and the CA cells of the ring gland (Figure 2.6C,F). Given that these simple reporters presumably only contain TCF/Pan sites plus a minimal promoter, the data suggest that TCF/Pan is allosterically regulated by DNA in a tissue-specific manner. Allosteric regulation of TFs by their cognate binding sites is known to occur (Johnson et al., 2001; Meijsing et al., 2009; Pyrowolakis et al., 2004; Scully et al., 2000; Surjit et al., 2011), and has been proposed previously for TCF/Pan (Blauwkamp et al., 2008). In these cases, the type of DNA binding site controlled whether the TF activated or repressed transcription. Our data suggest another possible aspect of TF allosteric regulation. TCF/Pan bound to different HMG-Helper pairs may allow interactions with distinct co-regulators, which enable it to activate transcription in a tissue-specific manner.

The aforementioned data demonstrates that different HMG-Helper pairs can profoundly influence the strength and tissue-responsiveness of promoters to Wnt signaling. While this was only examined in detail for a handful of reporters, our computational analysis supports the view that HMG-Helper pairs of all four orientations and various spacings contribute to TCF/Pan binding to chromatin (Figure 2.8, 2.9).

Therefore, we speculate that there are many other such examples in the genome, and that the flexibility of TCF/Pan to HMG-Helper pairs provides a versatile evolutionary mechanism for CRMs to modulate their response to Wnt signaling.

C-clamp Containing TCFs in Other Systems

The genome sequences of many metazoans indicates that almost all invertebrates have a single TCF containing a C-Clamp, while vertebrates have four or more TCFs, with E-tail isoforms of the TCF1 and TCF4 genes containing a C-clamp (Archbold et al., 2012; Cadigan, 2012). While the HMG and C-clamp domains are highly conserved in most metazoans, POP-1, the *C. elegans* TCF, is somewhat divergent (Archbold et al., 2012). Perhaps more importantly, the linker sequence between the HMG and C-clamp domains is variable. For example it is 23 aa in human TCF1E, compared with 10 aa in TCF/Pangolin and 9 in POP-1 (Archbold et al., 2012). These differences could influence the rules for preferred HMG-Helper site configurations in different organisms.

Despite these concerns, the available data suggests that other metazoans have a bias for HMG-Helper pair configurations similar to what we have found in *Drosophila*. We have recently characterized four W-CRMs in *C. elegans*, identifying a functionally important HMG-Helper pair in each one. Three of these were FF orientations of 0, 1 & 2 spaces, while the fourth is an AK7 (Bhambhani et al., 2014). Furthermore, in a search for new *C. elegans* W-CRMs, 3 putative modules containing HMG and Helper clusters were chosen based on sequence conservation and individual site quality. However, only the module containing an optimal motif (AK7) was bound by POP-1 in an *in vitro* binding assay (Bhambhani et al., 2014). In humans, an *in vitro* protocol for enriching

preferred sequences flanking an HMG site for TCF1E reveals Helper-like motifs (RCCG) in either an AK2-9 or FF0-11 configurations (Hoverter et al., 2012). Functional Helper sites were found in several W-CRMs in a colon cancer cell line, where the C-clamp of TCF-1E is required for maximal cell growth (Atcha et al., 2007; Hoverter et al., 2012). In the case of the *Sp5* W-CRM, systematic mutagenesis of Helper sites revealed three motifs that contribute to Wnt-dependent activation, with configurations of AK7, RW3 and FF1 (Hoverter et al., 2012). Other W-CRMs that were Helper site-dependent also show predominantly FF and AK configurations (Hoverter et al., 2012). While analysis of additional W-CRMs in flies, worms, humans and other systems is required, the general rules for TCF-DNA recognition outlined in this report clearly provide a strong foundation for further studies.

Materials and Methods

Plasmids

Synthetic HMG-Helper pairs were synthesized by Integrate DNA Technologies (IDT; Coralville, IA) and cloned into a modified pGL3-Basic vector (Promega) containing an hsp70 minimal promoter (Blauwkamp et al., 2008) for cell culture assays, or the pLacZattB vector (Bischof et al., 2007) for transgenic fly generation, using BglII and XhoI restriction sites. The *nkd*IntE and *nkd*UPE2 reporter gene vectors were described previously (Chang et al., 2008a; Chang et al., 2008b), and mutagenesis was carried out using the Stratagene QuickChange kit (Agilent). pAcArm* and p*arm*LacZ have been described previously (Blauwkamp et al., 2008; Chang et al., 2008a; Chang et al., 2008b). The protein expression vector for EMSA was generated by cloning the region

encoding the HMG domain and the C-clamp into the XmaI and SacI restriction sites of the pET52b(+) vector (Merck Millipore).

Cell culture

Drosophila Kc167 cells were cultured in Schneider's *Drosophila* Medium (Gibco) supplemented with 10% Fetal Bovine Serum (Gemini Bioscience). 250 ul of cells were seeded in 48 well plates, at a density of 1million cells/ml, and transient transfections were performed using Fugene transfection agent (Roche). Each well received 20ng luciferase reporter vector and 2 ng pArmLacZ. Wnt signaling was activated by transfection with 10 ng pAcArm*, (a constitutively active Arm protein), and pAC5.1 EV was used as filler DNA to 100 ng total for each well. Cells were lysed and treated three days later using the Tropix Luc-screen kit (Applied Biosciences) and Luciferase and LacZ activity assayed using the Promega Glomax system. pArmLacZ was used to normalize for transfection efficiency.

EMSA

A His-tagged fragment of TCF/Pan containing both the HMG and C-Clamp domains was purified from *E.coli* strain BL21 following IPTG induction for 4 hours @ 37° using column purification on Nickel beads (Invitrogen) with Immidazole elution. LB growth media supplemented with 10uM ZnCl. dsDNA probes were purchased from IDT and labeled probe was tagged with a 5' 700 IR moiety on both strands. Competition assays were performed using the LI-COR Odyssey Infrared platform, and infrared intensity of the IR dye-labeled probe/protein complexes were calculated using Image Studio 2.0.

The IC_{50} values were calculated using Prism 6 for Mac OS X (Graphpad Software, La Jolla California), as were the saturation binding curves. Three independent experiments were used to perform a least-squares non-linear fit. Binding reactions were performed as described in (Chang et al., 2008b), briefly, with 50 ug/ml poly(dIdC). 0.05% NP40, 50 mM MgCl2 and 3.5% glycerol in binding buffer (10 mM Tris-HCl, pH7.5, 50 mM KCl, 1mM DTT). Each reaction, containing 6 pmol recombinant protein and 0-2.4 pmol competitor dsDNA (dose indicated in figure 2.2A) was incubated for 5 min on ice, 25 minutes at RT before 20 fmol IR-dye labeled probe was added and reactions were incubated for an additional 30 minutes. A complete list of the probes used can be found in Table 2.1.

Drosophila genetics

Synthetic and endogenous W-CRMs were cloned into the pLacZattB vector (Bischof et al., 2007) and injected by Rainbow Transgenics (Camarillo, CA) using a ϕ-C31 site directed integration strategy. All constructs were injected into line 24749, integration site 86Fb. 1-3 individual lines were analyzed for each construct, and as expected, no variation in expression level or pattern was seen between lines.

Immunohistochemistry and Microscopy

To detect β-galactosidase activity, third-instar larval discs were fixed in 1% gluteraldehyde (in PBS), and incubated in staining solution (10 mM $NaPO_4$, 150 mM NaCl, 1 mM $MgCl_2$, 6 mM $K_4[Fe^{II}(CN)_6]$, 6 mM $K_3[Fe^{III}(CN)_6]$, and 0.3% Triton X-100, plus 2 mg/ml X-gal) for 25 minutes at room temperature. After the reaction was stopped, discs were mounted in 70%

glycerol. Images were taken on a Nikon Eclipse E600 upright microscope with Spot basic software and processed using Adobe Photoshop CS5.1.

Bioinformatic analysis of ChIP seq data

Training sequences for PWMs were taken from previously defined functional sites in W-CRMs depicted in Figure 2.1A. PWM scores were calculated using the formula: $weight_{i,j} = \ln\{[(n_{i,j} + p_i)/(N+1)]/p_i\} \sim \ln(f_{i,j}/p_i)$. The high confidence TCF/Pan bound regions (Junion et al., 2012) were searched for bipartite motifs and binned according to orientation and spacing using the dm3 genomic assembly in Matlab. To generate a random set of DNA sequences to analyze, an aggregate list of all sequences found in the 5', intergenic, intronic, and 3' data sets was created. Each sequence from the set was assigned an index 1 through N, where N was index of the last sequence in the aggregate set. A random ordering of all indices was then created and used to iterate over the data set, thus guaranteeing the same sequence could not be selected more than once. For each iteration, if a sequence contained a minimum size of 50 base pairs it was analyzed using the same processes as was used on the target data set. When the number of random sequence base pairs equaled or exceeded the number of base pairs in the target data set, the random data analysis was concluded. For each run of the random sequence analysis, the random number generator was seeded such that successive runs did not analyze the same fragments.

Acknowledgements

Thanks to Martha Snyder, Yaxuan Yang, Aditi Ravindranath and Chen Zhang for careful reading of the manuscript. HA was supported by NIH training grant T-32-GM007315 and a predoctoral fellowship from the AHA.

The data presented in this chapter is a major portion of the manuscript**: Bipartite Recognition of DNA by TCF/Pangolin is Remarkably Flexible and Contributes to Transcriptional Responsiveness and Tissue Specificity of Wingless Signaling**
Hilary C. Archbold, Chris Broussard, Mikyung V. Chang, and Ken M. Cadigan,
This manuscript has been submitted to PLoS Genetics, and is currently in revision.

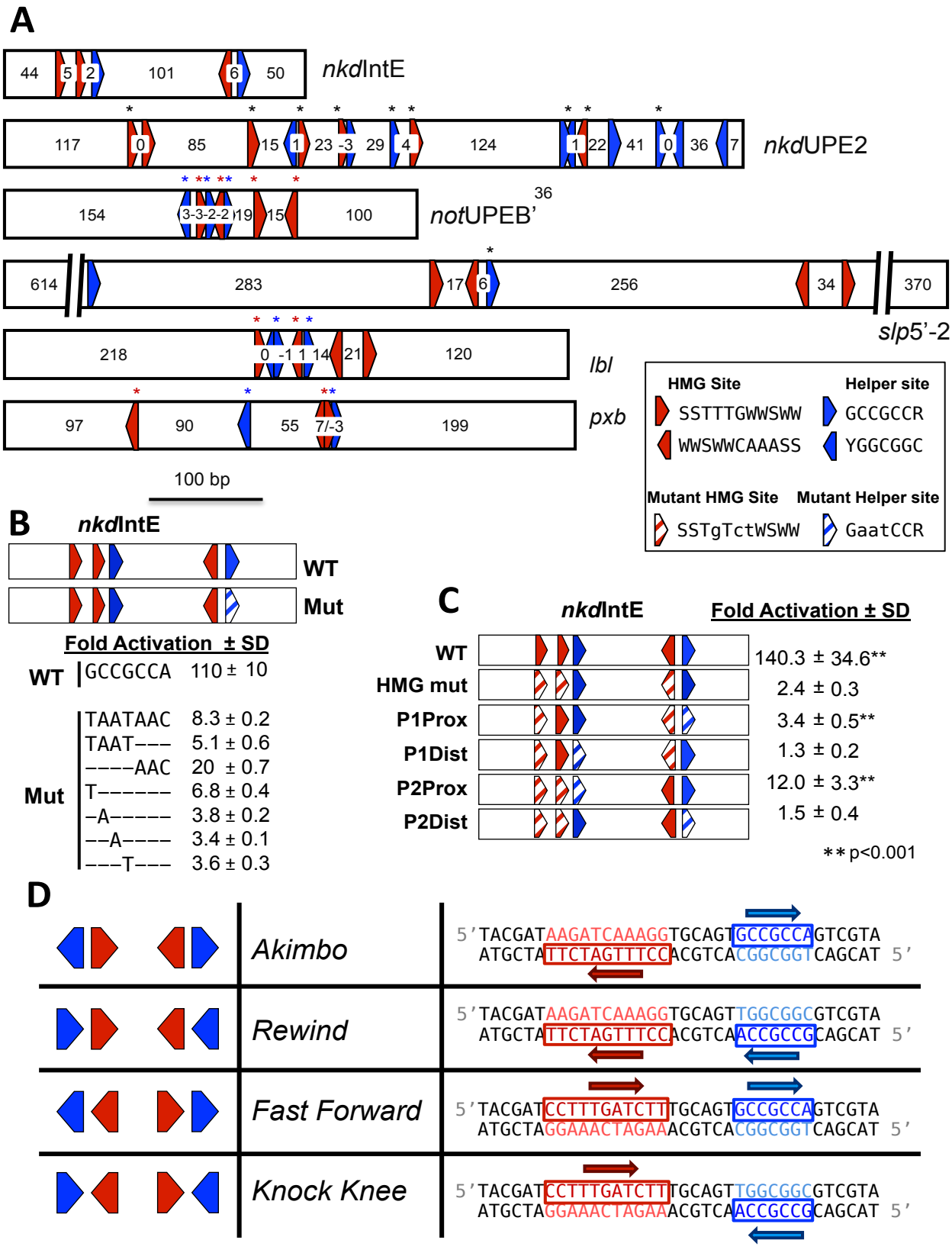
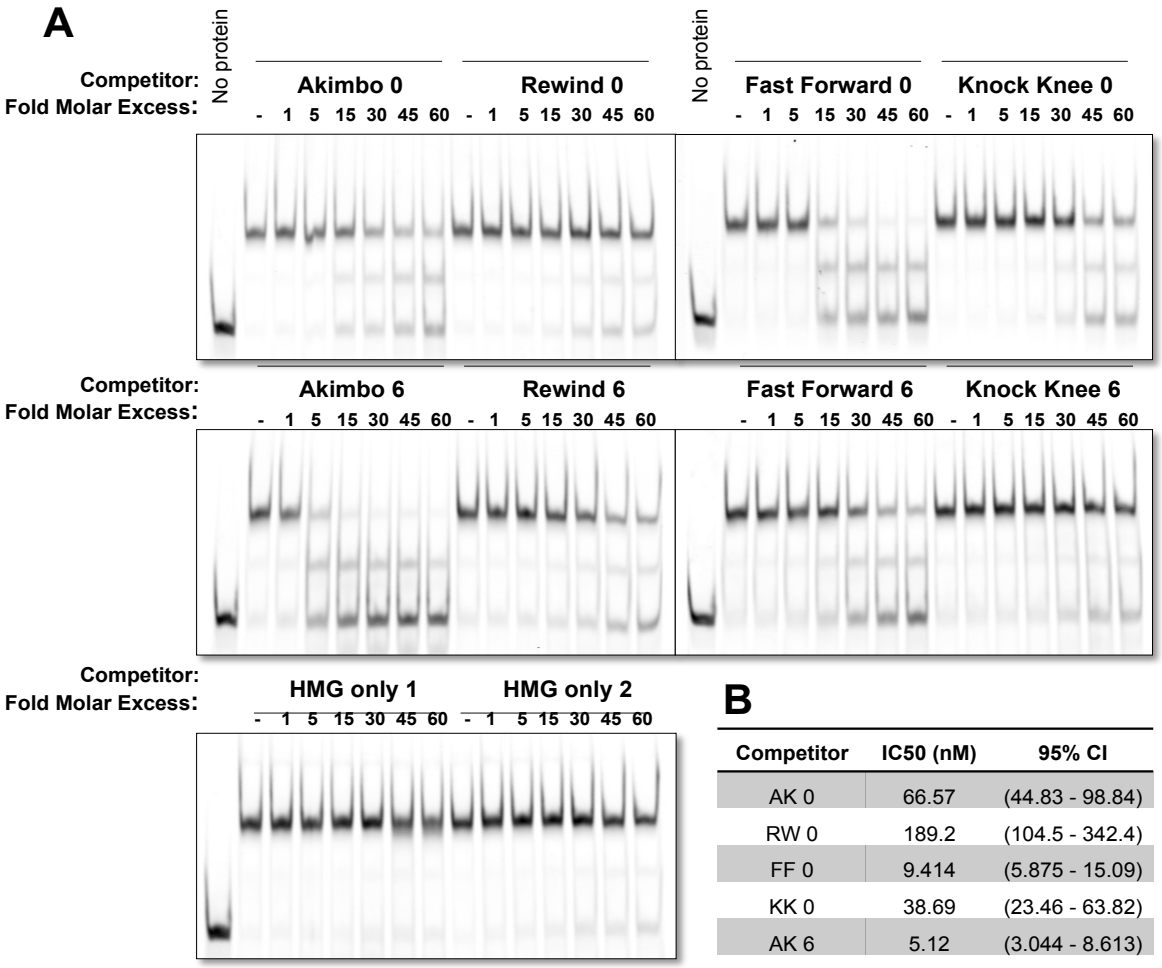


Figure 2.1. HMG and Helper site configurations in W-CRMs.

Figure 2.1. HMG and Helper site configurations in W-CRMs. (A) Previously characterized W-CRMs [43,56] with location of predicted HMG (red arrows) and Helper (blue arrows) indicated. Position Weight Matrices for site prediction indicated in Figure S1. Cutoffs for HMG and Helper sites were 5.35 and 6.5, respectively. Direction of arrow indicates orientation of motif (see inset for consensus sequences). Number of nucleotides between each motif is indicated. Black asterisks indicate sites that contributed to W-CRM activation by Wnt signaling in cell culture when mutated individually ([56], Chang, Chang and Cadigan, unpublished results). Red and blue asterisks denote function when all indicated HMG or Helper sites were mutated simultaneously. (B) Mutagenesis of second Helper site in nkdIntE W-CRM reporter indicates all seven positions contribute to W-CRM activation. Mutations in the Helper site are indicated, while "-" denotes the wild-type nucleotide is retained. Reporter constructs were transfected into Drosophila Kc cells with or without a plasmid expressing Arm*. Fold activation represents the ratio of Arm*/control, SD equals standard deviation of three biological replicates. (C) HMG and Helper sites work in closely spaced pairs. The nkdIntE reporter contains three functional HMG and two functional Helper sites [43], which were mutated (striped arrows) in different combinations, and tested for Arm* activation in Kc cells. Values represent the mean of three biological replicates ± SD. A Student's T-test was employed to determine significance. (D) Nomenclature and symbology for the four possible HMG-Helper site pair orientations. A right pointing arrow indicates the consensus sequence is read 5' to 3' on the "top" strand, a left arrow indicates the consensus is read 5' to 3' on the "bottom" strand. Sequences shown for HMG and Helper sites are identical to those used for DNA binding experiments (Figure 2) and synthetic Wnt reporter constructs (Figure 4A, 5 and 6). Supplemental Table 1 contains the complete list.



B

Competitor	IC50 (nM)	95% CI
AK 0	66.57	(44.83 - 98.84)
RW 0	189.2	(104.5 - 342.4)
FF 0	9.414	(5.875 - 15.09)
KK 0	38.69	(23.46 - 63.82)
AK 6	5.12	(3.044 - 8.613)
RW 6	44.52	(26.21 - 75.63)
FF 6	47.67	(23.28 - 97.60)
KK 6	99.72	(44.29 - 224.5)
HMG 1	969.1	(112.1 - 8378)
HMG 2	292	(124.8 - 683.4)

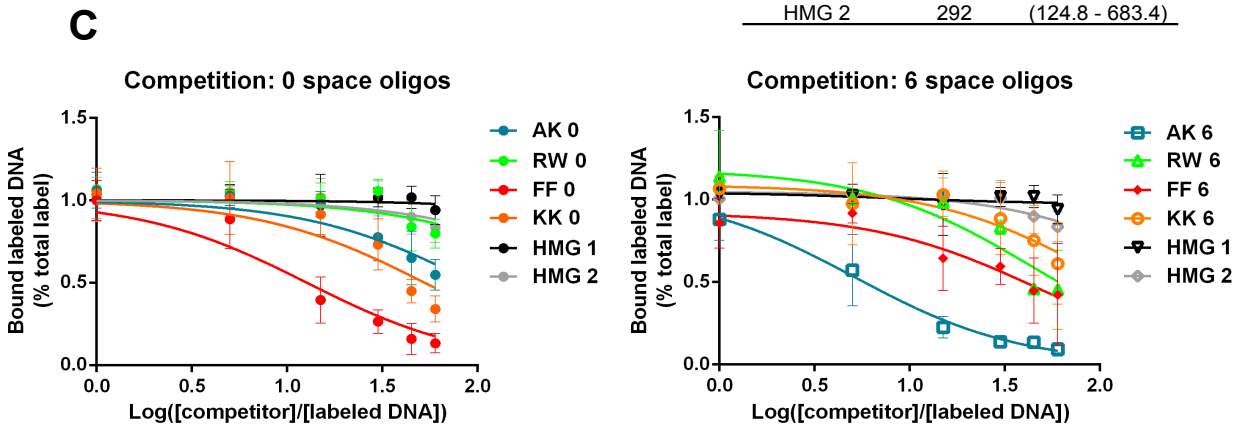


Figure 2.2. Binding preferences of TCF/Pan for various HMG-Helper site configurations *in vitro*.

Figure 2.2. Binding preferences of TCF/Pan for various HMG-Helper site configurations *in vitro*. (A) Competition EMSA experiments performed with a recombinant TCF fragment containing the HMG and C-Clamp domains, a AK6 IR-labeled probe, and competitor oligonucleotides containing one of the four orientations at 0 or 6 spaces. Images were taken on the Licor Odyssey, and binding intensity quantified with Image Studio 2.0. AK6 and FF0 were the strongest competitors, but Helper sites in all positions improve binding affinity when compared to binding of the HMG sites alone (HMG only 1 and 2). (B) The IC_{50} value (the measure of DNA concentration required to reduce binding of the labeled probe to 50% of uncompeted levels) for each competitor was calculated using Prism 6.1 (Graphpad). Asterisks indicate values with large confidence intervals, indicative of an imprecise measurement. (C) Semilog graphs depicting competition results from three independent experiments. Error bars indicate SD.

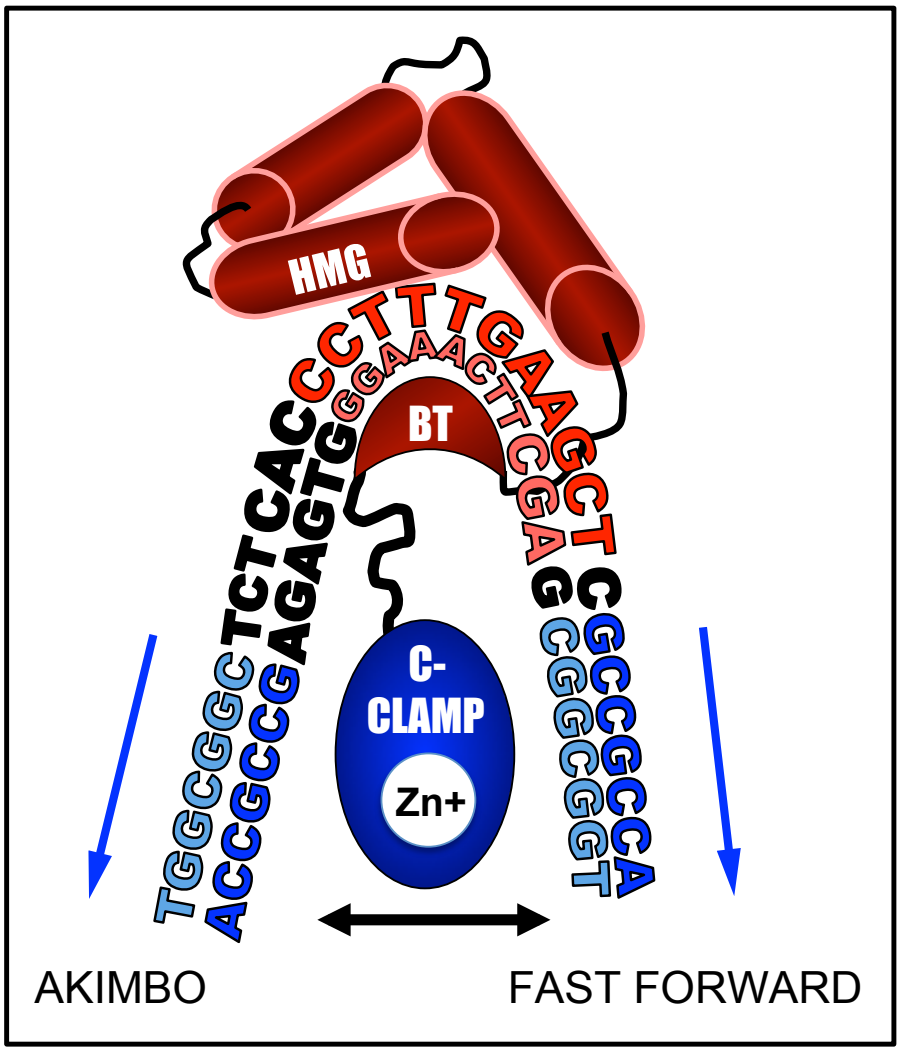


Figure 2.3. DNA bending by the HMG domain of TCF/Pan could explain preferential binding for AK6 and FF0 HMG-Helper site configurations.
A cartoon based on the NMR-deduced structure of the HMG domain and basic tail (BT) of human LEF-1 (Love,1995) bending the HMG site. The HMG domain is composed of three alpha-helices (red barrels), the first of which binds the minor groove of the HMG site (shown in red), while the BT (red crescent) wraps around to make contact with the major groove. This binding induces a sharp bend in the DNA, most pronounced between T3 and T4 (CCTTTGATCTT). In TCF/Pan, which shares 92% identity in the HMG domain of LEF1, the BT is followed by a 10 residue linker, and then by the C-Clamp (blue oval), recently shown to coordinate a zinc ion (Ravindranath, 2014). The C-clamp can bind to Helper sites (blue sequences) either upstream (AK) or downstream (FF) of the HMG site. We postulate that the linker places a constraint on the optimal spacing for the FF and AK orientation of 0 and 6 bp, respectively.

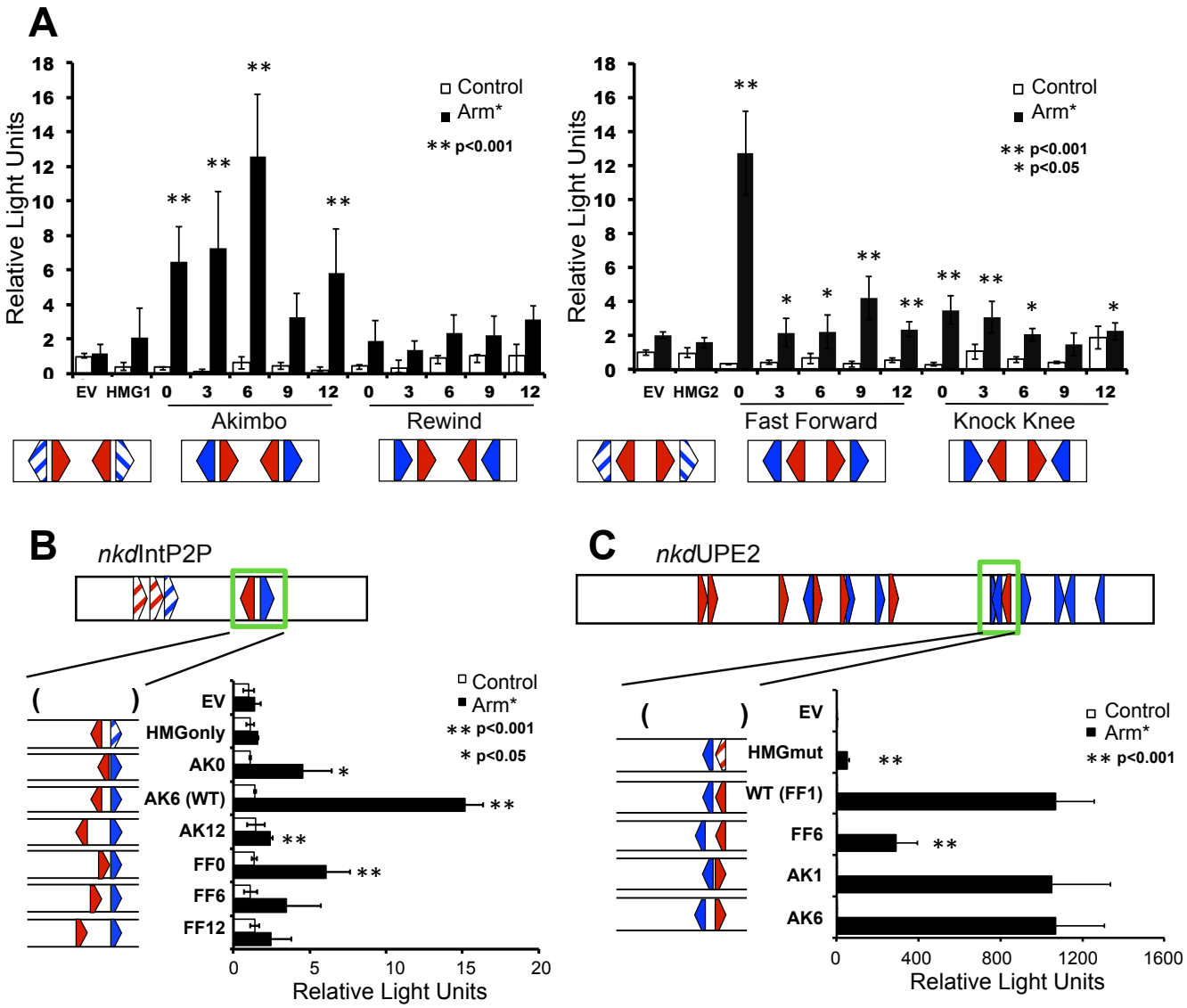


Figure 2.4. HMG-Helper pair configuration preferences in cell culture reporter assays.

Figure 2.4. HMG-Helper pair configuration preferences in cell culture reporter assays. Kc cells were transfected with the indicated reporters with or without a plasmid expressing Arm* and assayed for luciferase expression. (A) Synthetic W-CRMs containing two HMG-Helper site pairs in all four orientations with 0, 3, 6, 9 or 12 bp spacing between the HMG and Helper site. In every construct, the two pairs were separated by 6 bp. Two constructs containing only HMG sites (HMG1 and HMG2) were included in the analysis. The data shown represents the average of three independent experiments with three biological replicates each. Error bars represent SD and asterisks indicate a significant difference in activation compared to the HMG site constructs. (B) The cartoon at the top of the panel depicts the nkdIntE2P2 W-CRM, where all HMG and Helpers sites were mutated (striped arrowheads) except for one AK6 pair (green box). The spacing and orientation of this pair were altered as indicated. The endogenous AK6 configuration displayed the highest level of activation by Arm*, while the AK0, AK12 and FF0 constructs also exhibited higher activation than the HMG site only control. The data are the means of two independent experiments with three biological replicates each, ± SD. Asterisks indicate a significant increase in activation compared to the HMG site only construct. (C) In the nkdUPE2 W-CRM, the HMG and Helper site pair (green box) which contributes most potently to activation (Chang et al., 2008a) was altered as indicated. AK1 and AK6 configurations responded as well as the endogenous FF1 pair, while the FF6 configuration was less active. The data are the means of three biological replicates ± SD, and are representative of several independent experiments. Asterisks indicate reduced activation by Arm* compared to the wild-type construct. In all cases, Student's T-tests were employed to determine significance.

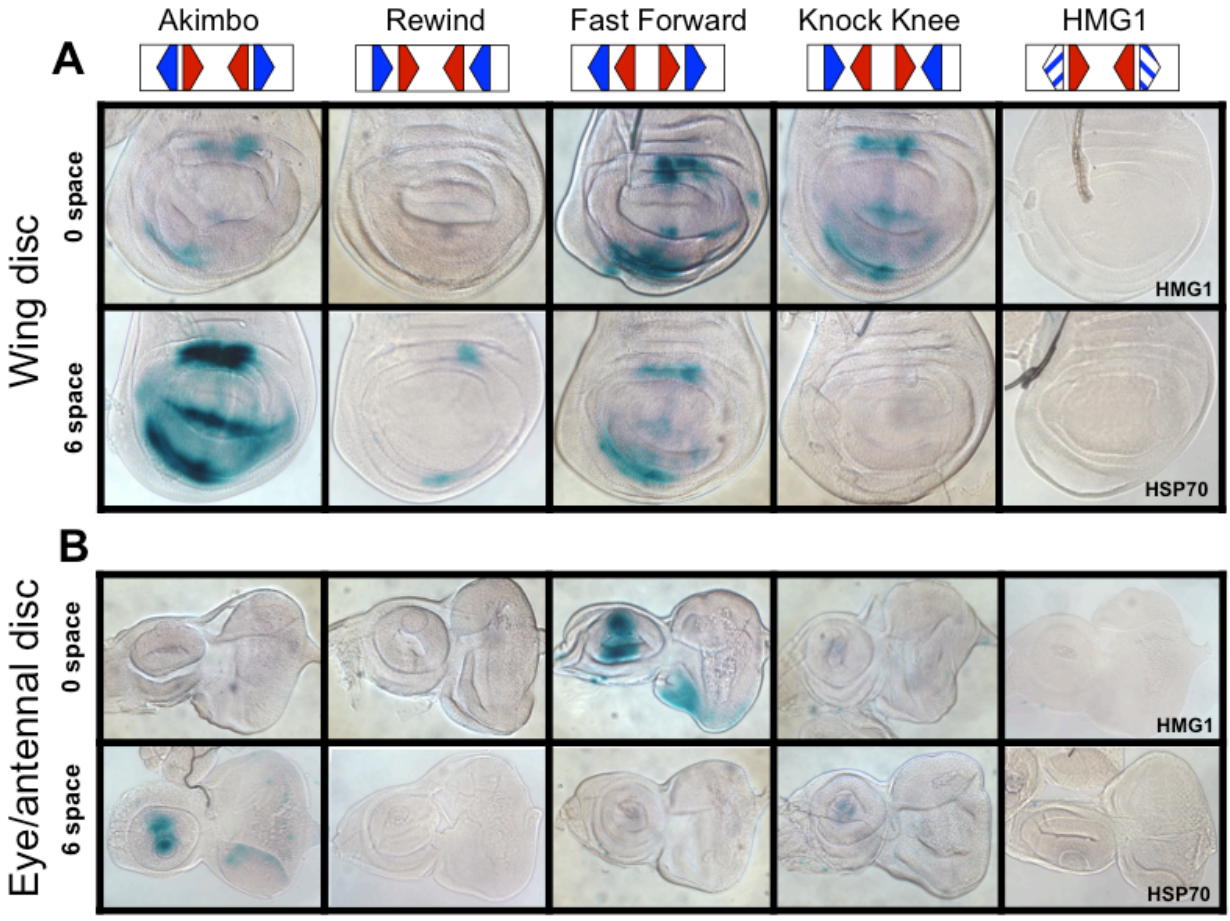
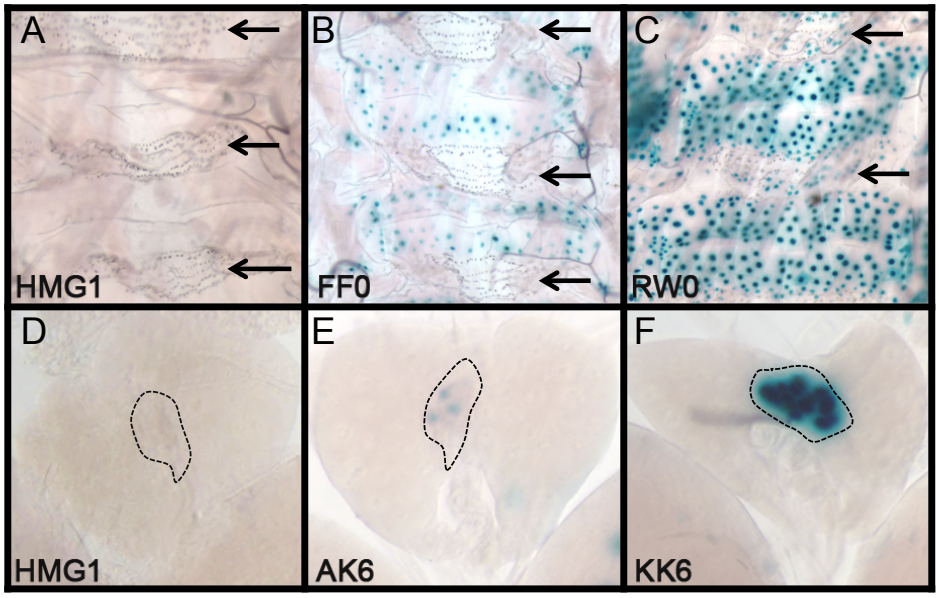


Figure 2.5. HMG-Helper pair configuration preferences in imaginal discs. Brightfield images of imaginal discs from 3rd instar larva containing the indicated lacZ reporter constructs stained for lacZ activity. (A) Wing imaginal discs. (B) Eye/antennal discs. The FF0 and AK6 reporters display the highest expression in these tissues. Neither the promoter alone (HSP70) nor the HMG site only (HMG1 shown) constructs have detectable expression. At least 20 discs for each reporter were analyzed, with representative images shown.



G	Wing	Eye/ Antennal	Leg	Haltere	Body Wall	Corpora Allata
AK0	+	-	-	+	+	-
RW0	-	-	-	-	++++	+++
FF0	+++	++++	++++	++++	++	+
KK0	++	-	-	++	-	++++
AK6	++++	+++	++++	++++	+++	+
RW6	+	-	-	-	+	+++
FF6	++	-	-	++	-	+++
KK6	-	+	-	-	-	++++
HMG1	-	-	-	-	-	-
HMG2	-	-	-	-	-	-
HSP70	-	-	-	-	-	-

Figure 2.6. Tissue-specific activity of HMG-Helper pair reporters.
Brightfield images of tissues from 3rd instar larva containing the indicated lacZ reporter constructs stained for lacZ activity. (Kawamoto et al.) (A-C) LacZ expression in larval body wall. Expression is seen in cells underlying the naked cuticle, located between denticle belts (arrows). The HMG1 reporter has no detectable expression; FF0 has weak expression and RW0 drives robust expression. (D-F) Ring glands with expression in the Corpora Allata (dotted lines). HMG1 has no detectable expression with AK6 and KK6 displaying weak and strong expression, respectively. (G) Summary of expression data from all tissues examined, with number of plus signs and blue hue indicative of the relative level of reporter gene expression. At least 12 samples of each reporter line

were analyzed for each tissue, with similar results observed between individual samples.

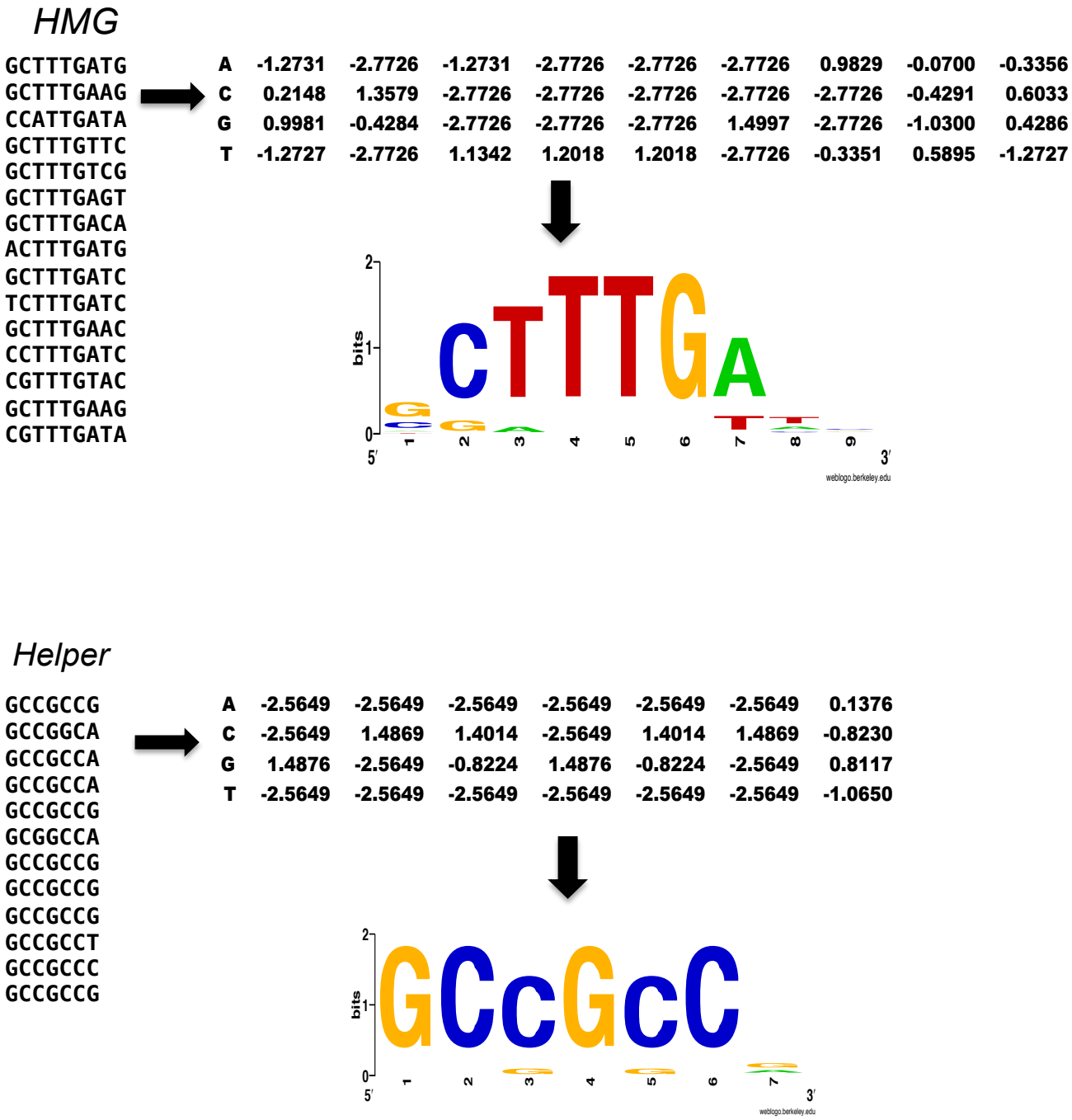


Figure 2.7. Position Weight Matrixes for HMG and Helper Sites.
Training sequences for matrixes shown to the left. Weighted scores (in bold on right) were calculated using the formula $weight_{i,j} = \ln\{[(n_{i,j} + p_i)/(N+1)]/p_i\} \sim \ln(f_{i,j}/p_i)$. Sequence logos shown below the position weight matrixes were designed using Weblogo (http://weblogo.berkeley.edu).

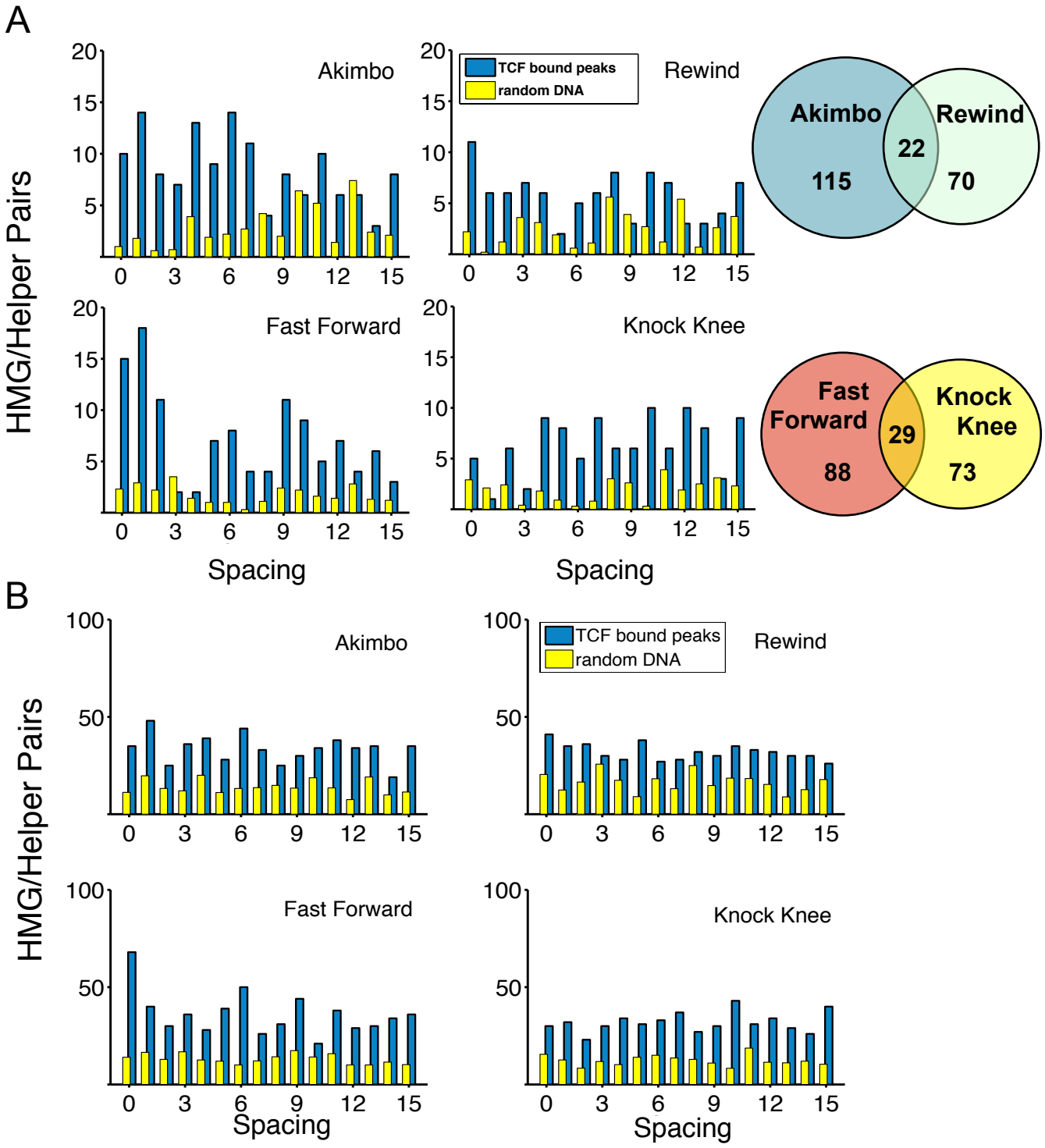


Figure 2.8. TCF/Pan-bound chromatin is enriched for HMG-Helper site pairs.

Figure 2.8. TCF/Pan-bound chromatin is enriched for HMG-Helper site pairs. (A) Distribution of HMG-Helper pairs in genomic sequences bound by TCF/Pan (blue bars) versus random DNA (yellow bars). TCF/Pan-bound sequences obtained from a ChIP-seq data set from germband extended *Drosophila* embryos (Junion et al., 2012). Equivalent amounts of random DNA from intergenic, intronic and 5'/3'UTR regions were analyzed, the average of ten runs is displayed. Nearly all HMG-Helper site configurations were enriched in TCF/Pan bound regions, with the highest degree of enrichment in FF0-2 and AK0-6. Due to the semi-palindromic nature of the Helper site, many sites with nucleotide mismatches were called as both AK/RW or FF/KK. Overlap is indicated by a Venn diagram to the right of each pair of graphs. (B) Enrichment using a lower stringency calling criteria. A cutoff of 3.5 for HMG and 5.0 for Helper sites (based on PWM in Fig. 2.7) was used to identify HMG-Helper pairs in TCF/Pan bound and random DNA . HMG-Helper pairs are ~ 2.4 times more likely to occur in TCF/Pan bound regions than in random DNA (2139 hits vs. 893.1).

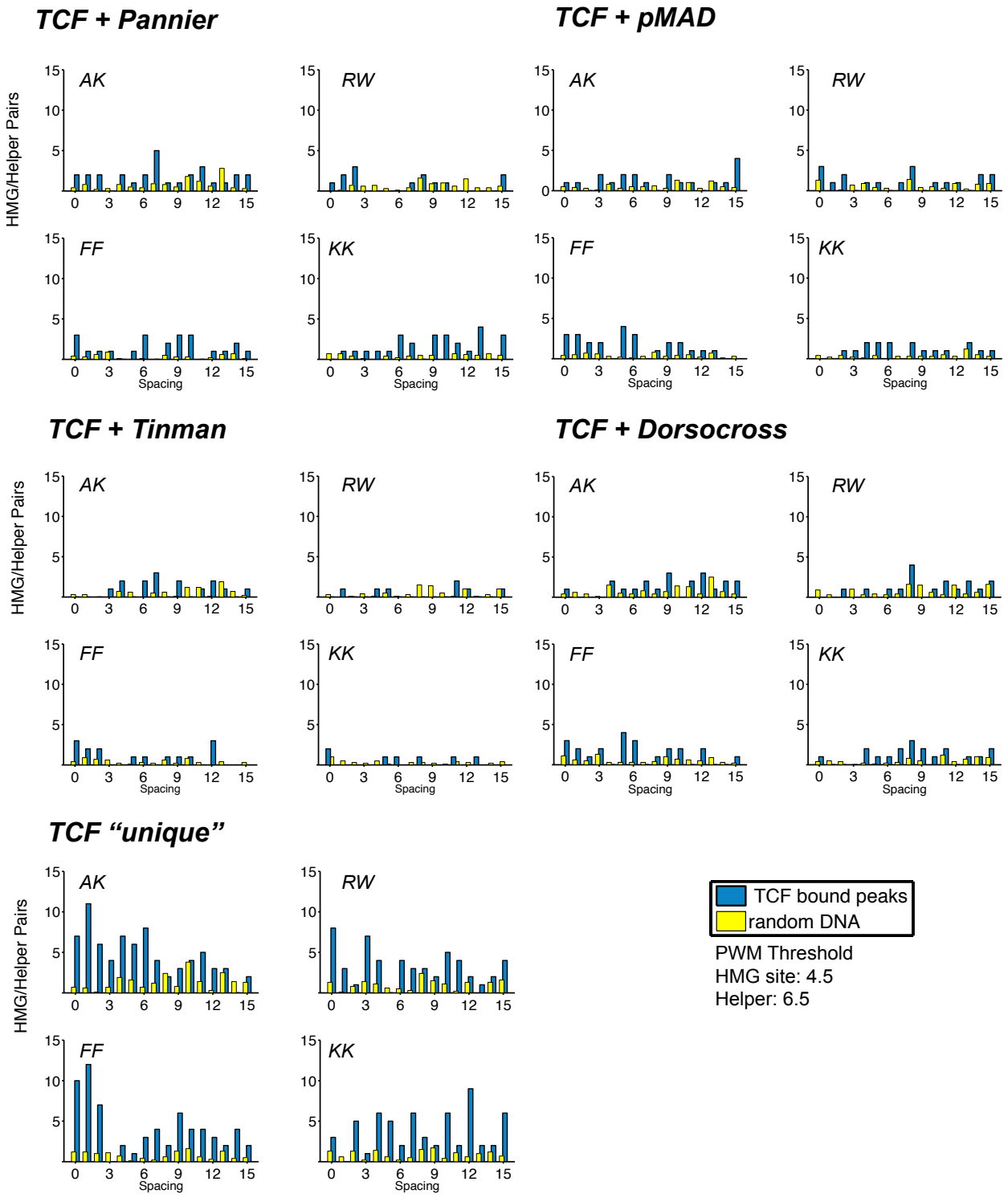


Figure 2.9. HMG-Helper pairs are less enriched in TCF/Pan bound regions shared by other cardiogenic TFs.
TCF/Pan bound peaks were portioned into groups based on whether the center of the peak was located within 150 bp of the peak for another TF. HMG-Helper pair enrichment is much greater in "unique peaks" than in shared peaks. The difference is especially evident in the FF0-2 and AK0-6 range.

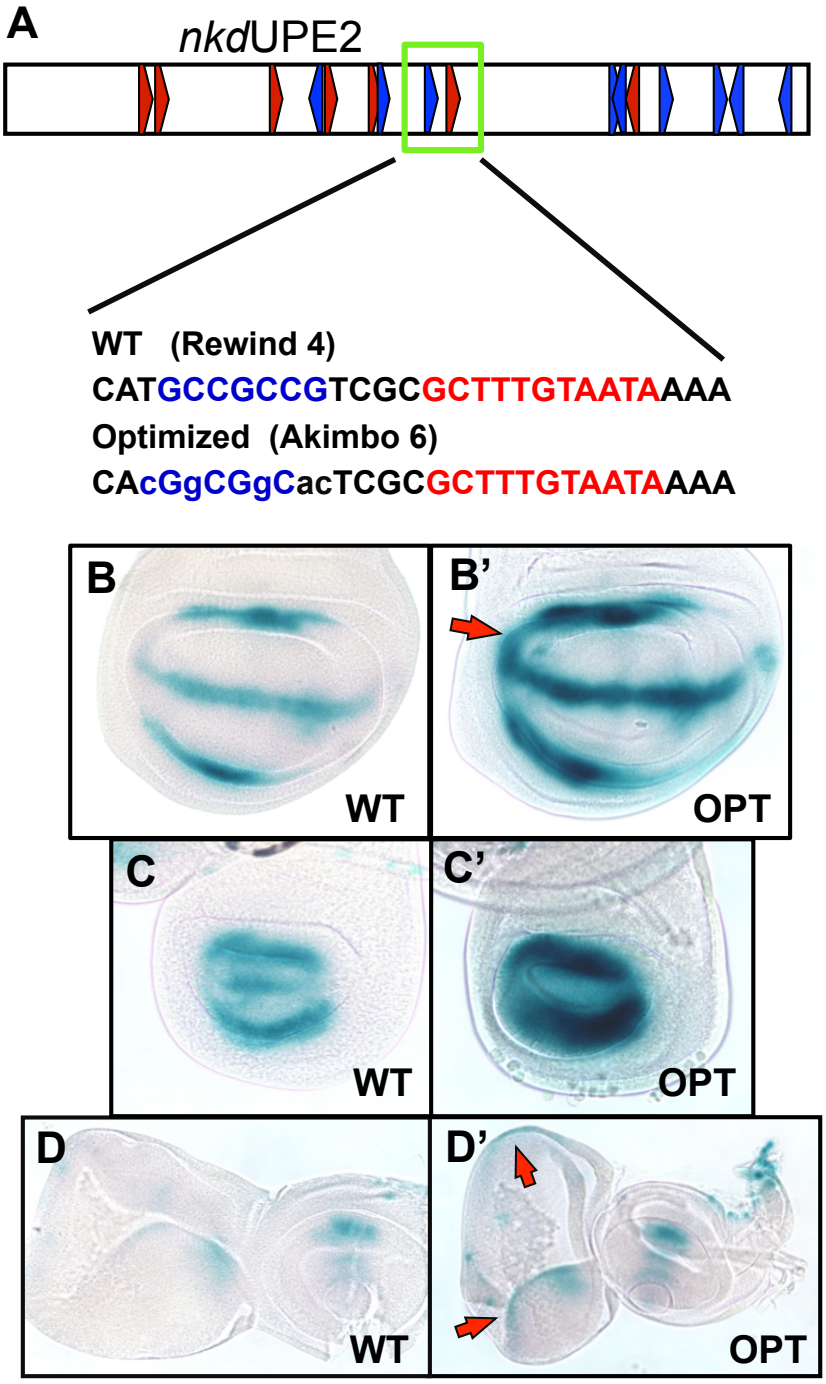


Figure 2.10. Optimizing motif architecture increases transcriptional output. (A) Cartoon of the nkdUPE2 W-CRM. A RW4 motif (green box) was altered to an AK6, point mutations indicated in lowercase. (B-D) Brightfield images of 3rd instar imaginal discs stained with X-gal. The "optimized" AK6 containing (OPT) W-CRM (B', C', D') drives higher expression levels of the reporter transgene than the wildtype (WT) W-CRM (B,C,D), and expands the region which responds to Wg signaling (red arrows). Representative wing (B), haltere (C), and eye/antennal (D) imaginal discs shown.

Table 2.1 Experimental Probes and Primers

Labeled probe for EMSA

IRSynAK6spFwd	/5IRD700/tacgatAAGATCAAAGGTGCAGTGCCGCCAgtcgta
IRSynAK6spRev	/5IRD700/tacgacTGGCGGCactgcaCCTTTGATCTTatcgac

Competitor probes for EMSA

SynAK0spFwd	tggtacgatAAGATCAAAGGGCCGCCAgtcgtacac
SynAK0spRev	gtgtacgacTGGCGGCCCTTTGATCTTatcgtacca
SynRW0spFwd	tggtacgatAAGATCAAAGGTGGCGGCgtcgtacac
SynRW0spRev	gtgtacgacGCCGCCACCTTTGATCTTatcgtacca
SynFF0spFwd	tggtacgatCCTTTGATCTTGCCGCCAgtcgtacac
SynFF0spRev	gtgtacgacTGGCGGCAAGATCAAAGGatcgtacca
SynKK0spFwd	tggtacgatCCTTTGATCTTTGGCGGCgtcgtacac
SynKK0spRev	gtgtacgacGCCGCCAAAGATCAAAGGatcgtacca
SynAK6spFwd	tacgatAAGATCAAAGGTGCAGTGCCGCCAgtcgta
SynAK6spRev	tacgacTGGCGGCactgcaCCTTTGATCTTatcgac
SynRW6spFwd	tacgatAAGATCAAAGGTGCAGTTGGCGGCgtcgta
SynRW6spRev	tacgacGCCGCCAactgcaCCTTTGATCTTatcgta
SynFF6spFwd	tacgatCCTTTGATCTTTGCAGTGCCGCCAgtcgta
SynFF6spRev	tacgacTGGCGGCactgcaAAGATCAAAGGatcgta
SynKK6spFwd	tacgatCCTTTGATCTTTGCAGTTGGCGGCgtcgta
SynKK6spRev	tacgacGCCGCCAactgcaAAGATCAAAGGatcgta
SynHMGonly1Fwd	tacgatAAGATCAAAGGTGCAGTGAATCCAgtcgta
SynHMGonly1Rev	tacgacTGGATTCactgcaCCTTTGATCTTatcgta
SynHMGonly2Fwd	tacgatCCTTTGATCTTTGCAGTGaatCCAgtcgta
SynHMGonly2Rev	tacgacTGGattCactgcaAAGATCAAAGGatcgta

Primers for cloning Candidate W-CRMs

fd96cMID

FWD	aaaagcggCCGCGTTGCACACTTCTGACT
REV	aaaaggtaccGGAGCCGCAGGGGCCAGATA

fkh3'UTR

FWD	aaagcggccgcTTGGGTTTCTTGGGGAGGGGG
REV	aaaaggtaccGCCGGCTACAACGCCCTCAG

Site Directed Mutagenesis Primers

*nkd*IntE mutants

HMG1/2mutFWD	CGCTGTGGAGGTCATTAGCTGTGGATGGACGCCGCCACCGCCGC
HMG1/2mutREV	GCGGCGGTGGCGGCGTCCATCCACAGCTAATGACCTCCACAGCG
HMG1/2Help1mutFWD	GCTGTGGAGGTCATTAGCTGTGCTTGGACGAATCCACCGCACTTTGCCGC
HMG1/2Help1mutREV	GCGGCAAAGTGCGGCGGTGGATTCGTCCAAGCACAGCTAATGACCTCCACAGC

HMG1mut/Help1mutFWD	GCTGTGGAGGTCATTAGCTTTGAGTGGACGAATCCACCGCACTTTGCCG
HMG1/Help1mutREV	CGGCAAAGTGCGGCGGTGGATTCGTCCACTCAAAGCTAATGACCTCCACAGC
Help2mutFWD	GATTATGTCAAAGCGAAAAATAATCCAAACATATGTGCCTCCAACC
Help2mutREV	GGTTGGAGGCACATATGTTTGGATTATTTTTCGCTTTGACATAATC
P2mutFWD	CGGTTTGGTTTTGATTATAGCACAGCGAAAAAGAATCCAAACATATGTGCCTCCAACC
P2mutREV	GGTTGGAGGCACATATGTTTGGATTCTTTTTCGCTGTGCTATAATCAAAACCAAACC
P2AK0spFWD	GGTTTTGATTATGTCAAAGCGCCGCCAAACATATGTGCCTCC
P2AK0spREV	GGAGGCACATATGTTTGGCGGCGCTTTGACATAATCAAAACC
P2AK12spFWD	GGTTTTGATTATGTCAAAGCACTAGGACTAGTGCCGCCAAACATATGTGCCTCC
P2AK12spREV	GGAGGCACATATGTTTGGCGGCACTAGTCCTAGTGCTTTGACATAATCAAAACC
P2FF0spFWD	GGTTTTGATGCTTTGACATAGCCGCCAAACATATGTGCCTCC
P2FF0spREV	GGAGGCACATATGTTTGGCGGCTATGTCAAAGCATCAAAACC
P2FF6spFWD	GGTTCGGTTTGGTTTTGATGCTTTGACATAGAAAAAGCCGCCAAACATATGTGC
P2FF6spREV	GCACATATGTTTGGCGGCTTTTTCTATGTCAAAGCATCAAAACCAAACCGAACC
P2FF12spFWD	GGTTTTGATGCTTTGACATAACTAGGACTAGTGCCGCCAAACATATGTGCCTCC
P2FF12spREV	GGAGGCACATATGTTTGGCGGCACTAGTCCTAGTTATGTCAAAGCATCAAAACC
*nkd*UPE2	
FF1(PERF)FWD	CCATTAAACGCTGGCGGCAAGAACAAAGCCGAACCAACCGACCGGC
FF1(PERF)REV	GCCGGTCGGTTGGTTCGGCTTTGTTCTTGCCGCCAGCGTTTAATGG
FF6FWD	CCATTAAACGCTGCCGGCAGAAAAAAAGAACAAAGCCGAACCAACCGACCGGC
FF6REV	GCCGGTCGGTTGGTTCGGCTTTGTTCTTTTTTTCTGCCGGCAGCGTTTAATGG
AK1FWD	CCATTAAACGCTGCCGGCAGCTTTGTTCTTCGAACCAACCGACCGGC
AK1REV	CCGGTCGGTTGGTTCGAAGAACAAAGCTGCCGGCAGCGTTTAATG
AK6FWD	CCATTAAACGCTGCCGGCATTTTTCGCTTTGTTCTTCGAACCAACCGACCGGC
AK6REV	GCCGGTCGGTTGGTTCGAAGAACAAAGCGAAAAATGCCGGCAGCGTTTAATGG
*nkd*UPE2RW>AK	
FWD	CGACGCAAATTGAAATATCGCATCAATTTGCTCGGCTGTCGCTTGTAATAAAAGTAAATAATA
REV	TATTATTTACTTTTATTACAAAGCGCGACAGCCGCCGAGCAAATTGATGCGATATTTCAATTTGCGTCG
SYNTHETIC W-CRMS	
HMGonly1FWD	GATCTAAAAAACTCGAGTGGATTCACTGCACCTTTGATCTTATCGACAAGATCAAAGGACGTCAGAATCCA
HMGonly1REV	TCGACTGGATTATGACGTCCTTTGATCTTGTCGATAAGATCAAAGGTGCAGTGAATCCACTCGAGTTTTTTA
HMGonly2FWD	GATCTAAAAAACTCGAGTGGATTCACTGCAAAGATCAAAGGATCGACCCTTTGATCTTACGTCAGAATCCA
HMGonly2REV	TCGACTGGATTCTGACGTAAGATCAAAGGGTCGATCCTTTGATCTTTGCAGTGAATCCACTCGAGTTTTTTA
AK0FWD	GATCTAAAAAACTCGAGTGGCGGCCCTTTGATCTTATCGACAAGATCAAAGGGCCGCCA
AK0REV	TCGACTGGCGGCCCTTTGATCTTGTCGATAAGATCAAAGGGCCGCCACTCGAGTTTTTTA
RW0FWD	GATCTAAAAAACTCGAGGCCGCCACCTTTGATCTTATCGACAAGATCAAAGGTGGCGGC
RW0REV	TCGACGCCGCCACCTTTGATCTTGTCGATAAGATCAAAGGTGGCGGCCTCGAGTTTTTA

FF0FWD	GATCTAAAAAACTCGAGTGGCGGCAAGATCAAAGGATCGACCCTTTGATCTTGCCGCCA
FF0REV	TCGACTGGCGGCAAGATCAAAGGGTCGATCCTTTGATCTTGCCGCCACTCGAGTTTTTTA
KK0FWD	GATCTAAAAAACTCGAGGCCGCCAAAGATCAAAGGATCGACCCTTTGATCTTTGGCGGC
KK0REV	TCGACGCCGCCAAAGATCAAAGGGTCGATCCTTTGATCTTTGGCGGCCTCGAGTTTTTTA
AK6FWD	GATCTAAAAAACTCGAGTGGCGGCACTGCACCTTTGATCTTATCGACAAGATCAAAGGACGTCAGCCG CCA
AK6REV	TCGACTGGCGGCTGACGTCCTTTGATCTTGTCGATAAGATCAAAGGTGCAGTGCCGCCACTCGAGTTTT TTA
RW6FWD	GATCTAAAAAACTCGAGGCCGCCAACTGCACCTTTGATCTTATCGACAAGATCAAAGGACGTCATGGC GGC
RW6REV	TCGACGCCGCCATGACGTCCTTTGATCTTGTCGATAAGATCAAAGGTGCAGTTGGCGGCCTCGAGTTTT TTA
FF6FWD	GATCTAAAAAACTCGAGTGGCGGCACTGCAAAGATCAAAGGATCGACCCTTTGATCTTACGTCAGCCG CCA
FF6REV	TCGACTGGCGGCTGACGTAAGATCAAAGGGTCGATCCTTTGATCTTTGCAGTGCCGCCACTCGAGTTTT TTA
KK6FWD	GATCTAAAAAAACTCGAGGCCGCCAACTGCAAAGATCAAAGGATCGACCCTTTGATCTTACGTCATGG CGGC
KK6REV	TCGACGCCGCCATGACGTAAGATCAAAGGGTCGATCCTTTGATCTTTGCAGTTGGCGGCCTCGAGTTTT TTA

Chapter III:

In Silico Searches for Novel W-CRMs Using Optimized HMG-Helper Site Architecture

Abstract

Wnt/β-catenin signaling is presumed to exert its many and varied effects on development by activating distinct transcriptional programs at different developmental time points, and in different tissues. How this feat is accomplished is only partially understood, although it is generally assumed that TCFs act together with tissue specific cofactors to coordinate regulation of the appropriate targets. Studying these combinatorial interactions requires knowledge of the modules where these inputs are coordinated. However, identifying Wnt-Responsive *Cis*-Regulatory Modules (W-CRMs) is not a trivial matter. *Cis*-regulatory modules can be located almost anywhere in the genome. They are generally thought to be located in intergenic or intronic regions near the genes they regulate, but they have also been found at great distances, sometimes several genes away, and even in coding regions of other genes (Harmston and Lenhard, 2013). In this section, we used our knowledge of the *cis*-regulatory code for TCF/Pan binding to computationally identify new W-CRMs in *Drosophila*. Four candidate W-CRMs were chosen for analysis, and all tested elements drove reporter gene expression in patterns which overlap Wg protein expression domains in transgenic *Drosophila*. Two novel elements on Chromosome 3R were shown to require Arm/TCF

for activation, indicating they are *bona-fide* W-CRMs. One of these elements drives expression in the larval prothoracic gland, a tissue not previously linked to Wg signaling. This finding highlights the importance of W-CRM discovery in contributing to the understanding of the many facets of Wnt/β-catenin signaling.

Introduction

In metazoans, Wnt/β-catenin signaling is important for a multitude of cell fate decisions, beginning with specifying the A/P axis early in development (Niehrs, 2010), and continuing with important roles in the formation of every major organ system (Grigoryan et al.) (Cadigan and Peifer, 2009) (Tanaka et al., 2011). In the adult organism, Wnt/β-catenin signaling continues to be important in regulating tissue homeostasis, stem cell maintenance, and wound healing (Arwert et al., 2012; Lim and Nusse, 2013; Whyte et al., 2012). How Wnt signaling controls such a broad range of processes likely requires the ability to regulate multiple, distinct transcriptional programs in a time and tissue specific fashion.

The TCF/LEF family of transcription factors are the primary mediators of Wnt/β-catenin signaling (Brantjes et al., 2002; Cadigan and Waterman, 2012). Understanding how TCFs identify target genes in the information rich nucleus is a critical step in dissecting the complex regulatory programs which create such diversity in cell type and function. In the hopes that a better understanding of the interaction between TCF and DNA will improve our ability to bioinformatically identify Wnt-responsive *Cis*-Regulatory Modules (W-CRMs),we have used the single *Drosophila* TCF/Pan to ask questions about the DNA sequence requirements for TCF binding and activity.

TCF/Pan contains two DNA binding domains: the HMG domain, and the C-Clamp. In chapter II, we showed that although TCF/Pan binds a bipartite HMG/Helper motif with remarkable flexibility, some orientations (such as Akimbo or Fast Forward) show higher affinity binding at specific spacings, and these motifs direct robust transcriptional activation in multiple Wg-dependent contexts. In this chapter, we use this knowledge of TCF/Pan binding to computationally identify new W-CRMs.

Previously, we used *in silico* searches for clusters of HMG and Helper sites to identify novel W-CRMs, without factoring in the orientation and spacing of potential HMG and Helper site pairs (Chang et al., 2008b). In this work, we chose two methods to test our ability to identify novel W-CRMs using searches biased for "optimal" bipartite motif configuration. First, using a candidate gene approach, we scanned regions flanking the gene, *ladybird early,* known to be downstream of Wg signaling, for clusters of optimal bipartite motifs. Second, we scanned the entire Chromosome arm 3R for optimal Fast Forward sites. We chose two candidate W-CRMS to test, one located between the related genes fd96Ca and fd96Cb, and one found in the 3' UTR of the *forkhead* (*fkh*) gene. These candidates are expressed in patterns consistent with Wg activation, and are negatively affected by disruption of Wg signal transduction.

Interestingly, the *fkh*3'UTR W-CRM drives expression in the larval ring gland (also called the prothoracic gland, PG), a tissue not previously linked to Wg signaling. We found that Wg is expressed in the ring gland, and that *fkh* W-CRM reporter expression and ring gland morphology require Wg signaling. These findings highlight how a better understanding of DNA recognition by TCF/Pan can enhance our ability to identify novel W-CRMs and discover new aspects of Wnt biology.

Results

The candidate gene approach: W-CRMs at the *ladybird early (lbe)* locus

As mentioned, we previously used an *in silico* search for clusters of HMG and Helper sites to identify novel W-CRMs (Chang et al., 2008b). One cluster occurred in the first intron of the *ladybird late* (*lbl*) locus, and a 500 bp stretch containing this cluster was shown to have W-CRM activity in cell culture (Chang et al., 2008b). While this W-CRM has four HMG and three Helper sites in close proximity, only one pair is in an optimal configuration (AK1; data not shown). *lbl* and its sister gene, *ladybird early* (*lbe*) are homeobox transcription factors positively regulated by Wg signaling in the embryonic epidermis, some neuroblasts, the heart primordium, and the anal pads (Jagla et al., 1997a; Jagla et al., 1997b). To see if we could use HMG/Helper pair conformation to identify additional W-CRMs, we searched a 30 kb region flanking the *lbe* TSS for HMG/Helper pairs using the online algorithm Target Explorer (Sosinsky et al., 2003). After examining the regions surrounding the top hits, we chose two regions to test based on the relative density and orientation of HMG/Helper pairs, rather than individual PWM scores. These regions, termed *lbe*UPE1 and *lbe*UPE2, were upstream of the *lbe* transcription unit (Figure 3.1A) and were tested for W-CRM activity using a lacZ reporter in transgenic flies.

The *lbe*UPE1 element drives reporter gene expression in the presumptive anterior sensory anlage and the proventriculus from embryonic stage 12-17 in an overlapping pattern with secreted Wg protein (Figure 3.1B, white arrowheads). *lbe*UPE2, on the other hand, drives lacZ expression in the anal pads from stage 13 (Figure 3.1C, white

arrow heads) to stage 17 (3.1D, blue arrowhead). This overlaps the Wg expression domain, consistent with previous reports that *lbe* is downstream of Wg signaling in the anal pads (Jagla et al., 1997b). *lbe*UPE2 also drives mesodermal expression (Figure 3.1C, D) during these stages.

The overlap in expression patterns is consistent with the idea that *lbe* is a direct target of Wg signaling. To further explore this question, we disrupted signal transduction by expressing *arm*RNAi in developing embryos using the ubiquitous Daughterless- (Da-) Gal4 driver (Wodarz et al., 1995). Surprisingly, in Arm depleted embryos, the expression of *lbe*UPE1 in the anlage and proventriculus is unchanged, while there is an increase in expression of epidermal stripes (Figure 3.1B'). *lbe*UPE2 expression in the anal pads is slightly weaker at stage 13 (Figure 3.1C'), but by stage 17, both anal pad and mesodermal expression is much stronger (Figure 3.1D'). This puzzling result may indicate that these W-CRMs do not mediate direct positive regulation by Arm/TCF. Another possibility is that multiple Wnt dependent factors may co-regulate this W-CRM.

The genomic approach: Identifying novel W-CRMs on Chromosome 3R

As our data indicated certain conformations, such as FF0-1, are overrepresented in TCF/Pan-bound DNA and drive robust activation by Wg signaling in multiple contexts, we tailored a computational search for FF1 motifs. A stringent calling criterion was used, to keep the number of hits at a manageable level. The search was performed on the right arm of chromosome 3, containing more than 20 Mb of sequence, using Target Explorer, an on-line search algorithm (Sosinsky et al., 2003). The stringent criteria

resulted in a short list of 23 hits (Table 3.1). We chose two putative W-CRMs for further analysis because they contained additional lower stringency HMG-Helper pairs near the initial FF1 hit. One cluster of pairs was located in the intergenic region between the related genes *forkhead domain containing 96c a* and *b* (*fd96ca* and *fd96cb*) (Figure 3.2A), while the other is from the 3' UTR of the transcription factor *forkhead* (*fkh*) (Figure 3.3A). These were interesting targets because *fd96ca* and *b* were previously reported to be expressed in 14 pairs of neuroblasts after germband extension in a pattern reminiscent of Wg expression (Hacker et al., 1992). *fkh* has been previously shown to be downstream of Wg signaling in the salivary placode (Zhou et al., 2001) and has also been shown to be required for the maintenance of Wg expression in the developing hindgut (Hoch and Pankratz, 1996).

To test the functionality of these putative W-CRMs, lacZ reporter constructs were inserted into the fly genome and their expression was examined. The *fd96C* W-CRM reporter is weakly active in a small number of cells at stage 11-12 (data not shown) and by stage 13 drives robust expression of stripes in the ventral mesoderm, overlapping the Wg pattern (Figure 3.2B). To determine if this reporter was dependent on Wg signaling, we examined its expression in embryos where Arm was depleted by driving an ArmRNAi transgene via the ubiquitous Daughterless (Da)-Gal4 driver (Wodarz et al., 1995). Arm depletion resulted in a nearly complete loss of *fd96C* reporter expression (Figure 3.2B'), indicating that it is a *bona fide* W-CRM.

In 3rd instar larvae, the *fkh* 3'UTR reporter drives strong expression in the PG, part of the ring gland (Figure 3.3B, middle panel). Although the PG had not been previously linked to Wg signaling, Wg protein was clearly detectable in this tissue by

immunostaining (Figure 3.3B, left panel). To confirm that the 3rd instar expression pattern was dependent on Wg signaling, the PG-specific *phantom* (*phm*)-Gal4 driver (Gibbens et al., 2011) was used to drive dominant negative TCF. A tub-Gal80^{ts} transgene was included (McGuire et al., 2004), so that signaling disruption was limited to 24 hr prior to dissection and staining. This treatment resulted in reduced lacZ expression compared to controls in late 3rd larval instars (Figure 3.3C'). The reduction is quantified in 3.3D. These results indicate that the 3' UTR of *fkh* contains a PG-specific W-CRM.

Reduction of Wg signaling in the PG had a significant effect on PG morphology and the development of the animal. Distortion of the PG was observed for *phm*Gal4::UAS-DNTCF in the absence of the Gal80 transgene and when Gal80 repression was relieved for more than 24 hours (data not shown). In addition, *phm*Gal4::UAS-DNTCF and phm::*arm*RNAi animals die during pupation and the larvae were, in general, smaller and slightly delayed in development compared to controls (data not shown). These results argue that Wg signaling has a previously unappreciated role in the development of the PG.

Discussion

In silico Identification of Novel W-CRMs

The high level of degeneracy in TCF binding sites (Badis et al., 2009) makes *in silico* detection of W-CRMs difficult. The use of evolutionary conservation can facilitate such searches, e.g., with the EEL algorithm (Hallikas et al., 2006). We previously demonstrated that searching for clusters of HMG and Helper sites in the fly genome

could identify W-CRMs that are directly activated by Wnt signaling in cell culture (Chang et al., 2008b). In this work, we incorporated the knowledge gained from analyzing the functional architecture of HMG-Helper site pairs to refine our computational searching. Our basic strategy employed searching for high quality "optimal conformation" HMG-Helper pairs, followed by secondary searches for nearby lower quality pairs, which resulted in the identification of several novel W-CRMs.

Searching the *lbe* locus identified multiple clusters of bipartite motifs. The two regions we analyzed drive expression in patterns which partially overlap Wg expression. Since *lbe* participates in a feed forward regulatory loop and maintains late Wg expression in the anal plate and dorsal epidermis (Jagla et al., 1997b), this overlap could be the result of *lbe* activity and not vice versa. Depletion of Arm had ambiguous effects, and could indicate these CRMS are not directly regulated by Arm/TCF. However, the increased late expression could be the result of compensatory activation of ladybird, or by loss of Wg-dependant repressors, such as *even-skipped* (*eve*) (Han et al., 2002). *lbe* and *eve,* which are both targets of Wg in the cardiac mesoderm, participate in mutual repression in this tissue (Han et al., 2002), and a similar mechanism could be at work here. To distinguish between these possibilities, mutation of the putative TCF sites would allow for the direct contribution of TCF to be assayed, without altering the endogenous expression of *lbe* or other Wnt dependent factors which may co-regulate this W-CRM. In addition, sequence analysis of the W-CRMs could identify candidate binding sites for possible co-factors involved in combinatorial regulation at these CRMs.

Our screen of chromosome 3R for high quality FF1 pairs revealed stretches containing multiple HMG-Helper pairs in the *fkh* and *fd96C* loci (Figure 3.2A, 3.3A). These regions were demonstrated to possess W-CRM activity in embryos and the ring gland (Figure 3.2B, 3.3B,C). Our results indicate that searches biased for those HMG-Helper site configurations that are bound by TCF/Pan with highest affinity *in vitro* can successfully identify novel W-CRMs.

Given our functional data that other "non-optimal" HMG-Helper pairs can also recruit TCF/Pan and promote Wnt-dependent transcription, often in tissue-specific ways (Figures 2.5-2.6), additional searches for these configurations should be a useful approach for W-CRM identification. For example, the *mab-5* gene in *C. elegans* is a known target of Wnt signaling (Maloof et al., 1999), but a W-CRM in its regulatory DNA had not been identified (Ji et al., 2013). Using our search protocol, we identified a FF7 pair 9.4 kB upstream of the *mab-5* ATG, which was demonstrated by others to have W-CRM activity in mab-5 expressing cells (Ji et al., 2013). Expression of this reporter was significantly reduced by mutation of the HMG site identified by our search (Ji et al., 2013). These HMG and Helper sites are fairly divergent (i.e., TCTTTTGCCTC & GCCATAA) which highlights another application of the results in our report: functional TCF sites that diverge from the consensus can still be identified if HMG-Helper site pairing is considered, as long as the amount of DNA to be searched is not too extensive (e.g., <12 kb).

In silico searching for HMG-Helper pairs offers a complimentary approach to genome-wide surveying of TCF/Pan binding using ChIP-seq. While the region containing the *fd96c* W-CRM was identified as bound by TCF/Pan in fly embryos

(Junion et al., 2012), the *fkh* W-CRM was not. This is perhaps because it is active in only a extremely small number of cells during embryogenesis (data not shown). While computational analysis of HMG-Helper pairs may help to prioritize which TCF/Pan ChIP-seq peaks might be functionally relevant, given that HMG-Helper pair enrichment is markedly reduced in TCF/Pan-bound regions that are also occupied by other TFs, it is also likely that TCF/Pan is recruited to many W-CRMs by protein-protein interactions (Figure 2.9).

Another benefit of *in silico* based discovery of W-CRMs is highlighted by our identification of the *fkh* W-CRM, which is expressed in the PG cells of the ring gland (Figure 3.3). This endocrine organ is a master regulator of *Drosophila* molting behavior (Di Cara and King-Jones, 2013; Rewitz et al., 2013), but had not been previously linked to Wnt signaling. Wg protein was detected on PG cells, and disruption of Wg signaling in the PG results in reduced expression of the *fkh* W-CRM reporter, as well as morphological abnormality of the ring gland and organismal lethality (Figure 3.3). We are currently exploring the role of Wg signaling in ring gland biology and think it likely that computational searches for W-CRMs will uncover additional roles for the Wg pathway in other tissues.

Materials and Methods

Plasmids

For the *fd96C*Mid and *fkh*3'UTR W-CRMs, the fragments were amplified using Roche High Fidelity enzyme, using w^{118} genomic DNA as the template, and cloned into

TOPO TA (Invitrogen) as an intermediate before being moved into the pLacZattB vector, using the Acc65I and NotI sites.

Drosophila genetics

Synthetic and endogenous W-CRMs were cloned into the pLacZattB vector (Bischof et al., 2007) and injected by Rainbow Transgenics (Camarillo, CA) using a ϕ-C31 site directed integration strategy. All constructs were injected into line 24749, integration site 86Fb. 1-3 individual lines were analyzed for each construct, and as expected, no variation in expression level or pattern was seen between lines. Candidate W-CRM constructs were recombined with UAS lines expressing a dominant negative TCF/Pan (van de Wetering et al., 1997) or an ArmRNAi hairpin (Dietzl et al., 2007) and crossed to the appropriate GAL4 driver line using standard techniques. *Da*Gal4 (Wodarz et al., 1995) was used to drive expression in the embryonic epidermis, while the ring gland-specific driver *phm*GAL4 (created by M. B. O'Connor) was obtained from Michael Stern.

Immunohistochemistry, Immunostaining, and Microscopy

Immunostaining was performed as described in (Cadigan and Nusse, 1996), using rabbit anti-LacZ (MP biomedicals) and mouse anti-Wg concentrate (Developmental Studies Hybridoma Bank, University of Iowa). Embryos were collected for 24 hours before processing, and both antibodies used at a dilution factor of 1:1200. For the PG, larvae were collected at the third instar larval phase, and a 1:600 dilution of each antibody was used. For all samples, CY3 (Jackson Immunoresearch) and Alexa

488 (Molecular Probes) conjugated secondary antibodies were used at a 1:300 dilution. Images were taken on a Leica DM6000B confocal microscope and processed using Adobe Photoshop CS5.1. Normalized pixel intensity was calculated using Leica LAS software to measure pixel intensity in bounded nuclei. Mean LacZ fluorescent intensity for each nucleus was normalized to mean DAPI fluorescent intensity. A Tukey box plot was created at http://boxplot.tyerslab.com/. n=120 per condition. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots, notches indicate +/-1.58*IQR/sqrt(n) and non-overlapping notches gives 95% confidence that two medians differ.

Acknowledgements

We thank the Vienna *Drosophila* RNAi center and Bloomington Stock Center for providing fly strains. Special thanks to Michael Stern, Ulrike Heberlein, Laura Buttitta, and Tom Kerppola for additional fly lines. Thanks also to Martha Snyder, Yaxuan Yang, Aditi Ravindranath and Chen Zhang for careful reading of the manuscript. HA was supported by NIH training grant T-32-GM007315 and a pre-doctoral fellowship from the AHA.

The data presented in figures 3.2 and 3.3 of this chapter are included in the manuscript**:**
Bipartite Recognition of DNA by TCF/Pangolin is Remarkably Flexible and Contributes to Transcriptional Responsiveness and Tissue Specificity of Wingless Signaling
Hilary C. Archbold, Chris Broussard, Mikyung V. Chang, and Ken M. Cadigan,

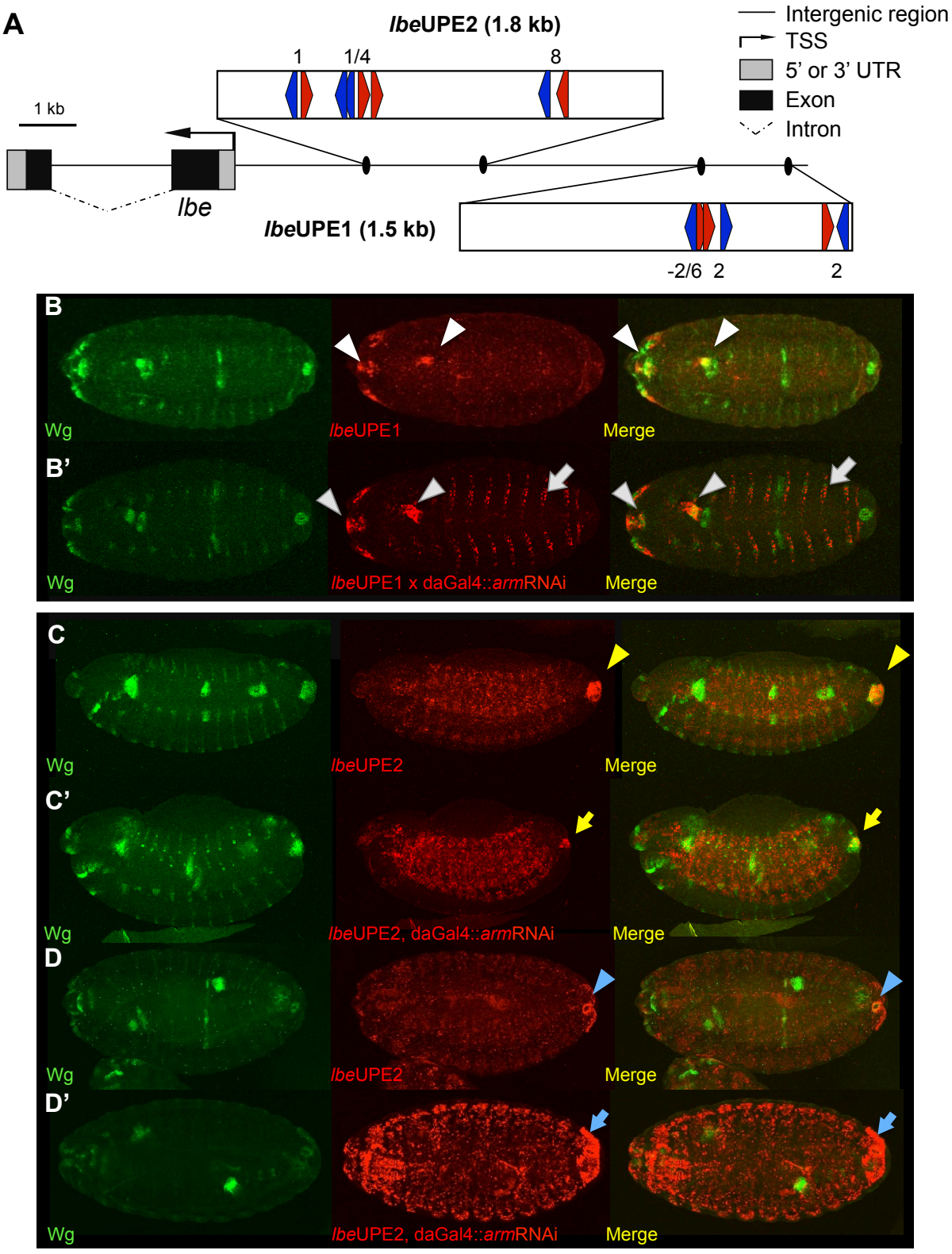


Figure 3.1 Candidate W-CRMs upstream of the *ladybird early* locus.

Figure 3.1 Candidate W-CRMs upstream of the *ladybird early* locus. (A) Cartoon of candidate W-CRM regions identified by Target Explorer. Red and blue arrows indicate HMG and Helper sites, respectively. Motif spacing indicated by black numbers above or below each pair. (B-D') Confocal images of embryos immunostained for Wg protein and a LacZ reporter. *lbe*UPE1 (B) expresses in the anterior sensory anlage and the proventriculus (white arrowheads), which overlaps Wg expression. (B') *arm*RNAi driven by the da-Gal4 transgene has no effect on this expression, but epidermal expression is increased (grey arrow). *lbe*UPE2 (C,D) expresses in the anal pads (yellow, blue arrowheads), also a Wg expressing domain. (C')Early expression (~stage 13) is only slightly reduced by *arm*RNAi (yellow arrow) while (D') later expression (stage15-17) is stronger (blue arrow), as is the mesodermal expression. B,D are top views, C is side view. In this, and all pictures, anterior is to the left.

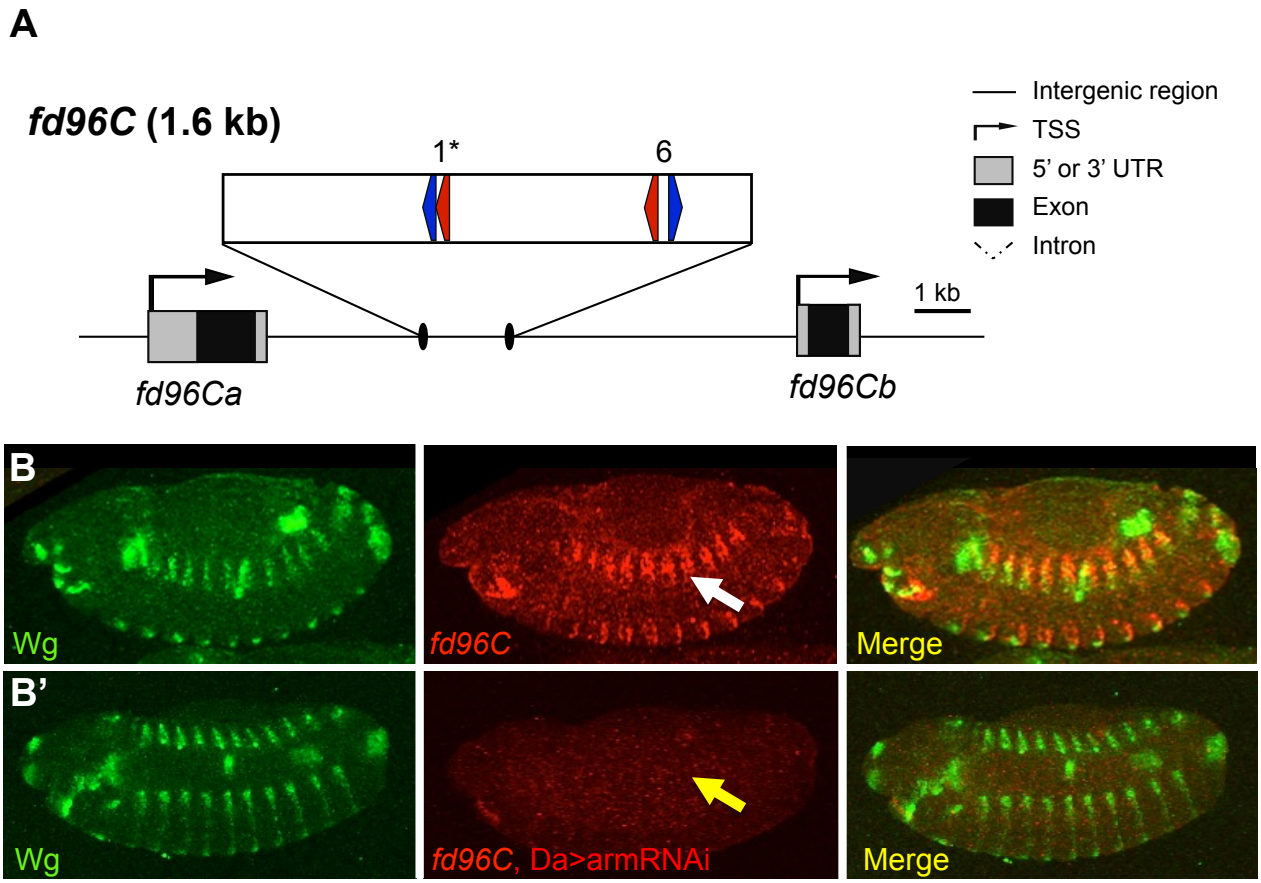


Figure 3.2 W-CRM located in the *fd96C* locus.
(A) Cartoon of candidate W-CRM region located between the related genes *fd96ca* and *fd96cb*. Bipartite motifs depicted by red and blue arrow pairs. Asterisks indicate the high scoring FF1 motif identified in the initial computational search. (B) The *fd96c* W-CRM drives stripes of expression in both the dorsal and ventral epidermis (white arrow), overlapping the Wg expression domain. (B') RNAi depletion of Arm results in loss of reporter gene activation (yellow arrow).

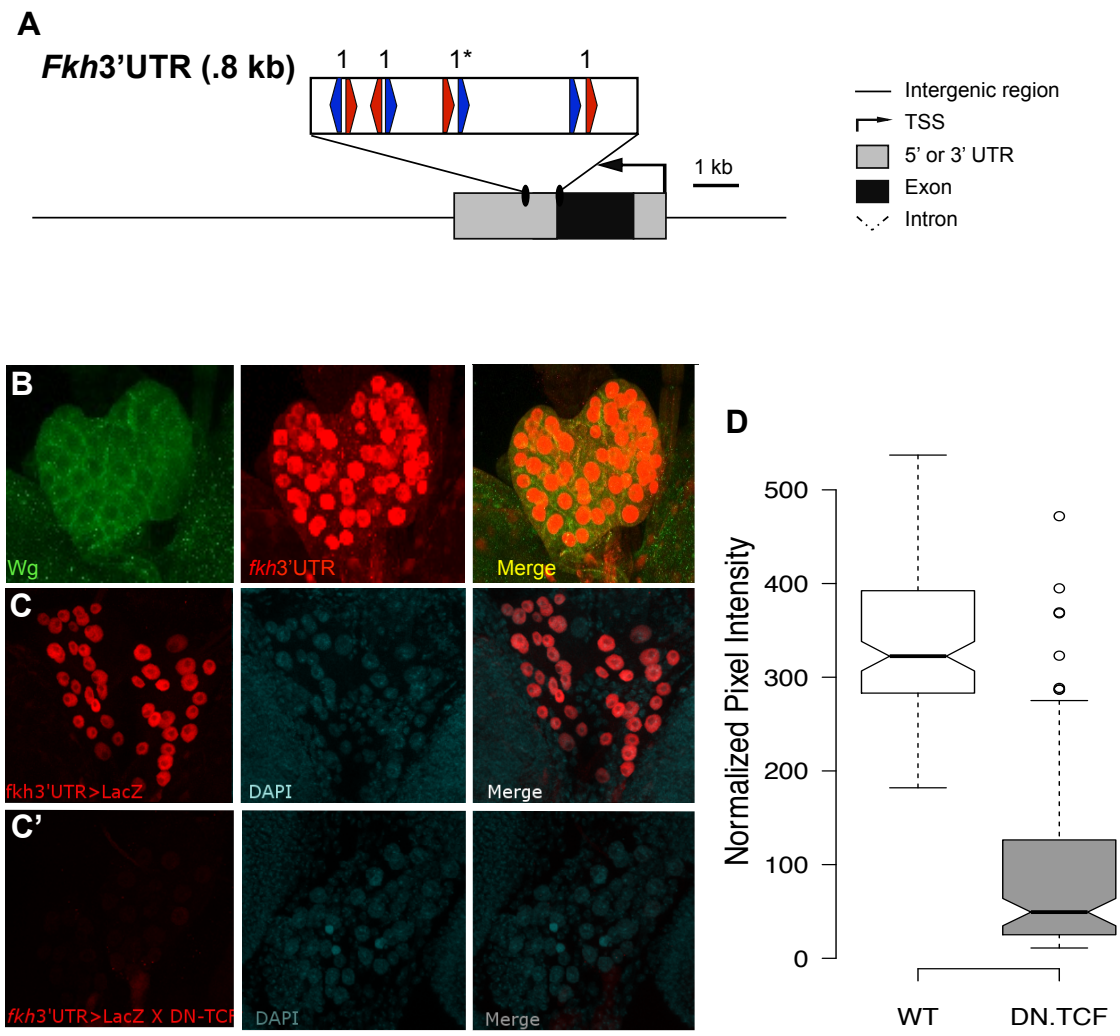


Figure 3.3 W-CRM in the 3' UTR of the *fkh* gene.
(A) Cartoon of region cloned for reporter analysis. Blue and red arrow pairs indicate bipartite motifs. Asterisk indicates the high scoring FF1 motif identified in the initial computational search. (B) Confocal images of a 3rd instar larval ring gland immunostained for Wg and fkh3'UTR>LacZ. The large lateral cells of the prothoracic gland (PG) display strong nuclear LacZ staining. Wg protein can be seen at the cortical surface of these cells. (C-D) Reporter activity is significantly reduced by overexpression of DN-TCF. Fluorescent images of ring glands from control (C) or phmGal4::UAS-DNTCF (C') ring glands. (D) Normalized pixel intensity of the LacZ reporter was calculated for 120 nuclei representing each condition, and plotted using a Tukey box plot. Horizontal lines at the notch mark the median, the notches indicate the 95% confidence interval. See materials and methods for further details.

Chr.	Location	Score	Sequence	Environment	Nearest gene
3R	20238070	17.03	GCTTTGATTTATGCCGCCG	intron	nAcRa96a
3R	25819986	16.87	GCTTGGACCTGGGCCGCCG	intron	CG7903
3R	14966083	15.76	TGCCGGCCACACACAAAGC	3' UTR	Ppcs
3R	1964090	15.51	GCTTTGACTTGTGCCGTCG	intergenic (3')	CG31559
3R	22306630	15.34	GCTTTGGCCTCGGCCGCCA	intron	plum
3R	4783106	15.23	CGGCGGCAAAGGTCAAAAC	intron	CG11966
3R	14763530	14.91	GCTTTCAATTTGGCCGCCA	intron	CG42613
3R	18892883	14.91	GCTTTGACTTTTGCCACCA	5'UTR	CG17111
3R	22195296	14.91	GCTTTGAACGTTGCCCCCA	intergenic (3')	CG33970
3R	**24408217**	**14.87**	**CGGCGGCTCCATTGAAAGC**	**3'UTR**	**fkh**
3R	11995504	14.86	ACTTTGATCTGTGCCGGCG	coding	CG6006
3R	11919815	14.71	TGACGGCCCAGCTCCAAGC	intron	Mhcl
3R	5345999	14.62	TGGCGGCCACGCAGAAAGC	intron	JHDM2/CG8176
3R	**20914519**	**14.54**	**TGCCGGCACATATCAAAGG**	**intergenic (3'/5')**	**fd96Ca/fd96Cb**
3R	22952109	14.54	GCTTTGGTTGGAGCCGGCG	coding	CG6051
3R	1937901	14.43	GCTCTGACCTGGGCCACCA	intron	Obp83g
3R	2312629	14.40	TGGCGGCTGCGTTCAATGC	intron	CG15186
3R	13009598	14.37	TGGCGGCCAAAAACAATGC	intron	Mur89F
3R	21786096	14.29	TCTTTGACCTGATCCGCCA	coding	CG14540
3R	22361482	14.29	GCTTTGAAGAGAGCCGGCA	intergenic (5')	plum/scrib
3R	7724385	14.26	TGCCGGCAAATTCCAAAGC	coding	Cad87A
3R	5968268	14.15	GCCTTGCCCTTGGCCGCCA	coding	CG6241
3R	9502929	14.11	TGGAGGCACAGGTCAAGGC	coding	lkb1

Table 3.1. *In silico* search for FF1 HMG-Helper site pairs.
List of the top 23 hits on Chromosome arm 3R. The nearest gene and genomic environment were determined using FlyBase GeneBrowse. Highlighted in bold are two candidates chosen for further analysis.

Chapter IV:

Conclusions and Future directions

Summary of Contributions

More than 30 years since the discovery of the first Wnt ligands, the number of direct transcriptional targets of Wnt/β-catenin signaling that have been identified is still surprisingly modest. The TCF/LEF family of transcription factors are the primary mediators of Wnt activated transcription. Understanding how TCFs identify and regulate target genes is of primary importance in identifying and dissecting tissue specific Wnt transcriptional programs. Although the presence of multiple HMG binding sites is sufficient for a Wnt response in some contexts, it is not sufficient in many others (Barolo, 2006). Previous work in our lab has shown that TCF/Pan recognizes two different DNA sequence motifs, HMG sites (via the HMG domain) and Helper sites (via the C-clamp domain) (Chang et al., 2008b). The C-clamp is present in almost all invertebrate TCF proteins, and in some vertebrate isoforms (Archbold et al., 2012), indicating that bipartite motifs are important for directing many aspects of Wnt signaling. The main focus of this thesis was to investigate the functional constraints of HMG-Helper pair architecture.

The clustering of HMG and Helper sites in known W-CRMs indicated that they likely worked in pairs with some functional flexibility, which I confirmed (Figure 1.1C). However, it was unclear what the limits of this flexibility were. To investigate this question I assayed multiple conformations of all four possible orientations for *in vitro* binding affinity, and for activation efficiency, both in cell culture and *in vivo* (Chapter 2). I found that TCF/Pan bound all tested conformations of HMG/Helper pairs more efficiently than probes containing only HMG sites, but relative affinity was orientation specific (Figure 2.2). The two strongest interactions were with the Akimbo 6 (AK6) and Fast Forward 0 (FF0) probes. These conformations place the Helper site roughly equal distances from the HMG induced DNA bend, offering a model which explains why TCF can bind Helper sites both 5' and 3' of the HMG site. The orientations bound with the highest relative affinity were the motifs which drove the most robust reporter gene activity, both in Drosophila Kc culture and in larval imaginal discs in transgenic *Drosophila* (Figure 2.3). However, intriguingly, in other larval tissues, this correlation between binding affinity and activation did not hold true. In the Corpus Allatum and the larval epidermis, several synthetic W-CRMs, which displayed weak to no activation in the imaginal discs, displayed robust gene activation (Figure 2.4). These data argue that motif architecture is a mechanism by which tissue–specific response patterns can be modulated.

How this tissue specific response is mediated is still unknown, but the "enhanceosome" model of transcriptional regulation provides a plausible explanation. This model posits that very specific binding site architecture is required for proper formation of enhancer complexes at the DNA (Arnosti and Kulkarni, 2005). The

competing “flexible billboard” model, argues that arrangement of transcription factor binding sites is flexible, and activation requires surpassing some threshold for TF density. There is evidence for CRMs which function using each of these mechanisms, or a combination of the two (Arnosti and Kulkarni, 2005). The enhanceosome can refer to multiple co-operative binding events at the DNA, or to multi-protein complexes. Our data is consistent with a model in which the sequence mediating TCF/Pan binding affects the tissue displaying a response by altering the identity of the “enhanceosome” which is formed. The allosteric regulation of the TF Pit1 is an example of “enhanceosome” style regulation. In this case, the distance between half sites determines co-factor binding activity, and the identity of the complex that is assembled (Scully et al., 2000). Another example is seen in the allosteric regulation of NF-kappaB heterodimers. A change of a single nucleotide in the binding site alters the recruitment of co-factors, without altering the ability of dimers to bind to the DNA (Leung et al., 2004). This model may be applicable to TCF/Pan, in that the conformation of TCF/Pan bound to RW0 or KK6 conformations may preferentially bind co-factors that are expressed in the larval CA or epidermis.

Using a publicly available data set, I performed bioinformatics analysis of TCF bound DNA regions in 4-6 and 6-8 hour *Drosophila* embryos. In TCF bound DNA regions, all four bipartite motifs were enriched at a variety of intersite spacings, but the enrichment was most pronounced for high affinity Akimbo and Fast Forward motifs (Figure 2.8). Taken together with our functional data, this supports a model in which TCF recognizes and binds to many permutations of the bipartite motif, but the motif has an instructive role in determining the strength or location of the response (Figures 2.4,

2.5, 2.6). Using binding data for other cardiogenic TFs, I also found that high affinity sites are less enriched in regions where other TFs localize when compared to “unique “ peaks (Figure 2.9). This data is consistent with the idea that TCF may be localized to the DNA by multiple mechanisms, and that TCF binding in the absence of high affinity binding sites may be mediated by protein-protein interactions either at the DNA, or through tethering mechanisms.

In this work, I also tested our ability to identify novel W-CRMs by searching for high affinity binding sites. I identified four modules containing bipartite binding motifs on Chromosome 3R which respond in regions which overlap with areas of secreted Wg protein. Activity of two of these modules, the *fd96c* and *fkh3'*UTR W-CRMS, is lost when Wg signal transduction is disrupted (Figures 3.3,3.4), indicating that they are *bona fide* W-CRMs. Surprisingly, the modules upstream of the *lbe* locus had a variable response to depletion of arm, ranging from a slight reduction to a strong increase in expression. The overlap in expression could be due to *lbe* activation of late Wg expression, or these results could indicate that multiple Wnt responsive factors co-regulate these W-CRMs. The information I have gleaned by studying the functional consequences of bipartite motif architecture will be useful in future informatics searches for W-CRMs. The tools we developed to analyze TCF ChIP data can easily be used to query the entire genome for clusters of binding sites which respond in the same tissues. The ability to correlate specific bipartite motif cluster “signatures” with tissue specific transcriptional programs is a long-term goal.

The data gleaned in this study support a model in which motif configuration plays an instructive role in setting the threshold, magnitude, and tissue-specificity of W-CRM

driven transcriptional responses. Whether this type of instructive flexibility is the rule or the exception for bipartite DNA binding proteins remains to be seen. Systematic analysis of synthetic motifs for other transcription factors has not been performed, and the analysis that has been done on these factors has been primarily in the context of cell culture. My work suggests that cell culture assays are capable of revealing only part of the story, and that transgenic analysis of transcriptional responses for other TFs may reveal equally complicated stories. How TCF/Pan regulates tissue–specific responses via distinct HMG/Helper configurations is a major question raised by this work. Other TFs, such as the POU domain family proteins, have been shown to be allosterically regulated by bipartite motif configuration in some contexts. Conformational changes in the bound TF affect co-factor recruitment, thus altering transcriptional responses. Whether this is the case for TCF/Pan is unknown. In the following section, I will discuss this issue, as well as several other questions raised by this work, and possible experiments to address these questions.

Questions and Future Directions

How much sequence degeneracy is allowable in the Helper site?

Mutation scanning analysis of the *nkd*IntE W-CRM initially identified two helper sequences (GCCGCCG, GCCGCCA), and searches for Helper sites in other W-CRMs studied in our lab identified several more 7 bp sequences which were highly similar (Chang et al., 2008b). However, while there are several indications that some degeneracy may be allowed in this motif, we have not thoroughly answered the

question: "How much degeneracy *is* tolerated, and which positions are more sensitive to substitutions?" In Chapter 2, we showed that for the second motif in *nkd*IntE, substitutions in the last three nucleotides had a less severe effect than mutations in the first four (Figure 2.1B). In *C.elegans*, several functional helper sites were identified, all of which displayed divergence from the 7 bp consensus, while still maintaining at least one GCCR core element (e.g. GCCGAAA, GCCGACA, GCCGCTT, GCCAAGT, TCCGCCA) (Bhambhani, 2014 #1333). In vertebrates, a shorter motif, RCCG has been identified through *in vitro* binding analysis, (Atcha, 2007 #228;Hoverter, 2012 #371). These data, taken together, could indicate that degeneracy is allowable (possibly at some cost) as long as a four bp core is retained. To test this supposition, SELEX-seq could be used to assay the permutations of Helper site that TCF/Pan can recognize and bind to. For this assay, a library of oligonucleotides is created, in which an HMG consensus binding site is fixed at one end of the oligomer, with random nucleotide sequences either upstream or downstream (20-25 bp should be sufficient). EMSA is used to select and then PCR amplify oligonucleotides which are bound by TCF. With traditional SELEX, only the highest affinity binding sites are identified after multiple rounds of selection and PCR amplification. However, in SELEX-seq, Illumina sequencing of bound oligomers occurs at each selection and PCR amplification step (Slattery, 2011 #1468). Therefore, more possible sites are identified, and relative affinity for these sites can be determined by comparing the number of reads in successive cycles. An alternative experimental method to address this question is protein binding microarray technology (Badis et al., 2009). This technique measures the intensity of a fluorophore-conjugated antibody bound to a tagged TF, and also

allows for computational rank ordering of the binding sites identified. A His-tagged C-clamp fragment which directly bind a consensus Helper site (Ravindranath and Cadigan, 2014) could be used in conjunction with a 10bp oligonucleotide array. Alternately, longer oligonucleotides encompassing both a fixed HMG and random 10 mers could be used with the HMG/C-Clamp fragment. The benefit of using the longer protein and oligonucleotides is that information about sequence preference can be gleaned in the context of the Helper site orientation and spacing. By better understanding Helper site sequence composition, we can improve our ability to correctly predict which HMG/Helper pairs identified bioinformatically are likely to represent functional elements *in vivo*.

How does the C-Clamp interact with the Helper site?

I have shown the position of the Helper site with respect to the HMG domain affects both binding affinity and transcriptional activation (chapter 2). Our model posits that *in vitro* binding affinity is correlated to the position of the Helper site in relation to the HMG imposed bend in the DNA. Structural analysis of HMG and C-Clamp fragments bound to DNA through either crystallization or nuclear magnetic resonance could provide useful information about how the C-Clamp interacts with different bipartite binding sites. Both AK6 and FF0 motifs were highly active in cell culture (Figure (2.4A), and strongly bound in the *in vitro* binding assays (Figure 2.2), however, their activity in larval imaginal discs was not entirely equivalent (Figure 2.5, 2.6) and their activity in the body wall and the CA was quite divergent (Figure 2.6). These two motifs represent HMG-Helper pairs where the Helper site is located on opposite sides of the DNA bend,

so comparing the conformation of recombinant TCF on these two motifs may provide interesting information about how specific bipartite motifs are bound by TCF. Analyzing TCF bound to KK6 or RW0 motifs might also prove informative in understanding how bipartite motif conformation allosterically regulates TCF, as these motifs are inactive in several classically Wg dependent contexts, but activate highly in the body wall and/or CA (Figure 2.6). Structural analysis may indicate surfaces which are more or less exposed in these conformations than in AK6 or FF0, and these may represent regulated binding interfaces for specific co-factors.

The use of a TCF fragment containing only the HMG and C-Clamp will, by definition, miss allosteric changes which might occur in either the N-terminal or C-terminal portions of the protein. However, expressing full length recombinant TCF (~81 Kda) is difficult. Furthermore, in NMR analysis, backbone assignments are possible for proteins up to 80 kDA, but challenging above 20-30 kDa (Skrisovska et al., 2010). No protein size limits exist for X-ray crystallography, but this technique depends on the ability of the protein to form regular crystals, which is not always a simple matter (Giege, 2013). The murine LEF-1 HMG fragment has previously been solved using NMR (Love et al., 1995), indicating that this technique is the logical place to begin, using a slightly larger HMG-C-Clamp containing minimal fragment of TCF/Pan.

Partial protease digestion is a technique which could be used to screen bipartite motifs for possible conformational changes before performing structural analysis. TCF/Pan has been shown to bind to the repressive WGAWAW motif in a different conformation than it binds to the classic HMG consensus (CCTTTGAT) using this technique (Zhang et al., 2014). If the recombinant TCF/Pan is bound to different

bipartite motifs with different conformations, the regions accessible to protease digestion could change, resulting in unique digestion patterns. Motifs causing unique patterns would be high priority probes for use in the NMR analysis.

Are the synthetic W-CRM driven expression patterns seen in the Epidermis and Corpus Allatum Wnt-dependant?

In the embryo and imaginal discs, expression patterns driven by the synthetic W-CRMs I created mirror well-characterized Wg expression patterns, and the strength of the transcriptional response roughly correlates to *in vitro* binding affinity. However, in the larval CA and epidermis, this correlation does not hold true. In the larval epidermis, robust activation from the RW0 W-CRM is quite different from its inactivity in other Wnt responsive contexts. Several of the other W-CRMs also respond in this tissue, which is suggestive. However, demonstrating that this response is indeed Wnt dependent is important. To address this question, the level of LacZ staining seen with or without disruption of Wg signaling will be compared for several of the W-CRMs. To disrupt Wg signaling in the larval epidermis, the A58-Gal4 transgenes, an epidermis-specific driver which turns on early in larval development (Galko, 2004 #1463), will be used to drive expression of either UAS-Arm RNAi or a dominant-negative TCF (UAS-ΔN4). Should removing Wg signaling too early in larval development make collecting samples difficult, the addition of a temperature sensitive Gal80 transgene can be used to keep Gal4 activity low until larvae are moved to a non-permissive temperature, relieving the Gal80 repression of Gal4.

In the Corpora Allata, most of the Helper-site containing W-CRMs display robust activity, including the KK6 motif, which was inactive in many of the other tissues assayed. Wg activity in this tissue has not been studied, and interestingly, the *fkh3'UTR* enhancer which is active in the neighboring PG does not display activity in the CA. Preliminary evidence indicates that this expression pattern is indeed Wnt dependent. Driving dominant negative TCF with the CA specific Aug21-Gal4 driver (Adam et al., 2003) completely abolished the expression of the RW0 W-CRM (Figure 4.1). The FF0, KK6, and AK 6 W-CRMs will also be tested to determine if there is a general requirement for TCF at these W-CRMs.

Although the expression pattern seen in the CA is dependent upon both the Helper site sequence, and (at least for RW0) on WT TCF, we have not demonstrated that these synthetic W-CRMs are being bound by TCF *in vivo*. Tissue-specific Chromatin Immunoprecipitation (ChIP) (Slattery et al., 2011) could be used to assay TCF occupancy at both robustly activated (e.g. KK6) and weakly activated (e.g. AK6) motifs. TCF binding at these motifs would be compared to the HSP70 only construct. It would also be interesting to compare binding between the CA and the PG to see if the difference in activation is due to differences in TCF localization at the W-CRM, or to difference in co-factor recruitment. However, the CA is an incredibly small tissue, and dissection of a large number of ring glands would be required. To make the CA/PG comparison, FACS sorting of the cells in the ring gland could possibly be done if the W-CRMS to be tested were placed upstream of GFP instead of LacZ, or if a UAS-GFP transgene were expressed in either tissue with a tissue-specific Gal4 driver (either Aug21- or phm-Gal4). FACS sorting of GFP expressing cells in larval imaginal discs

(de la Cruz, 2008 #1465) and larval neuroblasts (Harzer, 2013 #1464) has been described, and the protocols could presumably be modified. Our lab has previously used TCF antibody to perform ChIP in cell culture (Fang, 2006 #1357) (Parker, 2008 #282). However, as the CA is an extremely small tissue, overexpression of a tagged TCF protein in the CA may be required, although a result with endogenous TCF would be preferable.

Are tissue specific co-factors required for the expression pattern seen in the Corpus Allatum?

It seems likely that the presence or absence of tissue-specific co-factors may account for the unique pattern of W-CRM response seen in the Corpus Allatum. Due to the relatively small size and informational content of the synthetic W-CRMs, it also seems likely that these putative co-factors may interact directly with TCF, altering its binding affinity or transcriptional activity, rather than interacting with flanking DNA sequences. Unfortunately, there is not currently a robust method to directly assay protein-protein interactions in this tissue. A technique like ChIP-Mass Spec (ChIP-MS) would be theoretically ideal for directly assaying protein-protein interactions at the DNA, but unfortunately, is technically unfeasible to answer this particular question at this time. ChIP–MS was recently used to successfully identify components of the MSL complex, which regulates gene dosage compensation in male *Drosophila* (Wang et al., 2013). The MSL complex, though, coats the entire X chromosome, providing an input several orders of magnitude higher than TCF bound to two copies of a transgenic W-CRM. Furthermore, this assay was done in cell culture, with 5 X 10^9 cells per sample. Primary

cell culture in Drosophila, while possible, it not a simple matter, and the Corpus Allatum is a fully differentiated tissue. Disassociated CA from the cockroach, *Diploptera punctata*, have been induced to undergo a second cell division when hemolymph is added to the FBS growth media (Tsai et al., 1995), but a single division is far short of what would be required to derive enough CA cells to attempt an experiment of this sort.

A candidate based RNAi screen for possible CA specific co-factors is a more feasible method to investigate this question. As X-gal staining gives a reasonably quick visual method to assay W-CRM activity, a panel of possible cofactors could be tested by driving UAS-RNAi against each co-factor with the Aug21-Gal4 driver (Adam et al., 2003), and looking for any which result in a reduction or loss of LacZ activity. Logical candidates would be factors previously found to bind TCF, such as *Mad, HIP-R, ebd1, or CG7737 (Benchabane et al., 2011; Eivers et al., 2011; Guruharsha et al., 2011)*, or factors which have reported CA expression, and are thought to have effects on transcription, like *retained* (*retn*), or *CG11723* (Curators. et al., 2004; Valentine et al., 1998). Transcriptional co-factors previously unlinked to the CA would also be good candidates.

Does the activity seen in the Corpus Allatum and Larval Epidermis require different TCF Isoforms?

It is thought that TCF/Pan exists primarily as a 751 amino acid protein. However, 9 CDNAs have been annotated, all of which contain an HMG and a C-Clamp (Hoskins et al., 2007), and alternate isoform usage in *Drosophila* has not been characterized. This presents a situation in which unique tissue specific transcriptional responses could

be mediated by very specific isoforms. The variants A, B, D and R are identical, and represent the 751 amino acid isoform which is the most abundantly expressed both during embryogenesis and adulthood (van de Wetering et al., 1997). While variant S only differs by an 18 amino acid deletion at the C-terminus, variant H has 12 amino acid substitutions in the HMG domain and a 4 amino acid deletion in the basic tail. The much shorter variants Q and I (494 and 410 AA respectively) are identical to the H variant for AA 1-384 but have unique C-clamp domains and highly divergent C termini through alternate exon usage. In these isoforms, the C-clamp diverges at the residue after the third cysteine, and both are missing the fourth cysteine, which is presumed to help coordinate a zinc ion (Ravindranath and Cadigan, 2014). Variant J, on the other hand, which encodes an 1192 amino acid isoform, is identical to A/B/D/R through the HMG and C-clamp region, but contains a large and unique N-terminus.

To test if the CA and/or body wall express isoforms other than the primary A/B/D/R variant, Q-PCR can be used to assay the abundance of each possible isoform in each of the two tissues. Primers directed against regions unique to the H, Q, I, or J isoforms should be relatively simple to design, and the abundance of A isoforms can be inferred mathematically by comparing the unique primer products to a pan TCF primer product. S abundance can be inferred by comparisons between the pan TCF primer set, and a primer amplifying the region which is present in the A, H and J variants, but not in S.

If isoform expression differences are identified in these tissues, the functional relevance of these isoforms can be tested by isoform specific RNAi knockdown and rescue experiments, using cDNAs with heterologous 3'UTRs. Further analysis of the

isoform specific domains may yield clues as to possible protein-protein binding regions for specific cofactors. A yeast-two hybrid screen using relevant isoform specific domains fused to a Gal4-DBD as bait may identify possible co-factors.

Are the *lbe*UPE1 and 2 W-CRMS directly regulated by TCF?

Depletion of Armadillo in *Drosophila* embryos containing the lbeUPE1 and lbeUPE2 W-CRMS resulted in increased expression in late stage embryos. It is possible that these W-CRMs do not mediate direct positive regulation by Arm/TCF. Direct repression is also possible and has been reported for Arm/TCF. But, in this case, TCF acts through divergent WGAWA motifs (Blauwkamp et al., 2008) (Zhang et al., 2014) and possibly Helper like sequences, not classic HMG/Helper pairs, making this scenario unlikely. *lbe* and Wg participate in a feed-forward regulatory loop, and *lbe* is required to maintain Wg expression in later embryonic stages (Jagla et al., 1997b), which could indicate that coexpression of the W-CRM and Wg protein is the result of late *lbe* activation of Wg, and not vice versa. Alternately, this confusing result may indicate that multiple Wnt dependent factors may co-regulate this W-CRM. *lbe* and *even-skipped* (*eve*) are both targets of Wg in the cardiac mesoderm, and in this tissue they participate in mutual repression, creating unique zones of expression (Han et al., 2002). Complex combinatorial regulation of this W-CRM by TCF and other Wnt dependent factors is therefore also a possibility. To determine if this is the case, site-directed mutagenesis of the W-CRM can be used to disrupt the putative bipartite TCF binding motifs. Comparing the WT and mutant W-CRMS will allow us to measure TCF

direct effects on the reporter gene, without altering the endogenous expression of *lbe* or other Wnt dependent factors which may co-regulate this W-CRM.

Which genes do the novel W-CRMs identified on Chromosome 3R regulate?

Although it is often presumed that *cis*-regulatory elements (CREs) regulate the closest TSS, this is not always the case (Harmston and Lenhard, 2013). Chromatin looping is thought to bring distal enhancers in proximity to the TSS, and in a set of human cell cells, only ~7% of looping interactions identified were with the closest gene (Sanyal et al., 2012). Functional examples of this long range regulation have been found in multiple organisms. For example, a crucial murine regulatory element for the *sonic hedgehog* (*shh*) gene was identified in an intron of the Lmbr1 gene, over 1MB away (Lettice et al., 2003), and this regulatory organization is conserved in zebrafish and humans.

To investigate whether either of the forkhead domain containing genes (*fd96Ca* and *b*) or *forkhead* are regulated by the respective adjoining W-CRMs, the first step is determining whether endogenous transcripts are present in the tissues which express the transgene, and whether this expression is Wg dependent. To answer this question, I would perform *in situ* hybridization in embryos, (and larval PGs for the *forkhead* transcript). For any genes which display mRNA expression in appropriate tissues, expression levels in embryos where da-Gal4 is driving dominant negative TCF or *arm*RNAi will be compared to a da-Gal4 self-cross. Wg^{TS} trans-heterozygotes can also be used to assay if mRNA levels are decreased. Flies homozygous for the mutant Wg can be compared to siblings who retain one functional copy of Wg on the balancer

chromosome, or siblings left at the permissive temperature. While demonstrating that *fd96Ca* and/or *b*, or *forkhead* are expressed in a Wg dependent manner in the same (or partially overlapping) pattern would indicate these are likely the regulatory targets for the W-CRMs, it is not conclusive proof.

We have not been able to identify Forkhead protein in the PG using an anti-Forkhead antibody which has been successfully used in the larval fatbodies (Bulow, 2010 #1467). While it is possible that *forkhead* is transcribed, and protein levels are kept extremely low by post-transcriptional regulation, it is also quite possible that the *fkh*3'UTR W-CRM is regulating a different gene. If the former case is true, we should be able to detect transcript through the *in situ* assays described above, or through qPCR. However, if *fkh* is not detected in the relevant tissues, the question becomes "Which gene is this W-CRM regulating?" There are 9 annotated genes in the 100kb region surrounding *fkh*3'UTR W-CRM (*CG12413, PPN, MRE23, CR43440, Noa36, snoRNA:Psi28s, Hrb98DE and CG10011*). None stand out in particular as a more likely target than others based on expression profiling data or suspected molecular functions. qPCR could be performed to ascertain if any of these candidates are expressed in the PG.

It is possible that this enhancer participates in even longer-range interactions, and 4C (Circularized Chromosome Conformation Capture) or ChIP-Loop could be used to identify a region (or regions) that physically interacts with the W-CRM. Like 3C, 4C utilizes crosslinking followed by digestion and intramolecular ligation to identify long-range DNA interactions. However, 3C can only be used to identify previously suspected connections between two loci, as PCR primers from both regions need to be designed

to amplify the chimeric fragment. In contrast, for 4C only the "bait" loci need to be known, as high throughput sequencing is used to identify all captured regions after a second digestion and circularization step (Simonis et al., 2006; van de Werken et al., 2012).

How can we improve *in silico* searches for novel W-CRMs and Target genes?

In Chapter 3, we demonstrated that novel W-CRMs could be identified by targeted searches for high affinity motifs. In our search of Chromosome arm 3R, we used the online program Target Explorer (Sosinsky et al., 2003) which allows searching for multiple motifs, but has no sophisticated method to filter by spacing or orientation of bipartite motif "hits." Therefore we limited our search to FF1 motifs, as it was a configuration we could fix as a single motif. The program we wrote to search the TCF ChIP-seq data set (Figure 2.8) allows us to quickly sort identified bipartite motifs by spacing and orientation. By changing the search region, we can quickly scan whole chromosome arms or the entire genome and identify clusters of bipartite motifs which act robustly in the same tissue (in synthetic W-CRMs). Systematic analysis of Helper site sequence degeneracy, which I described earlier in this section, will also improve *in silico* searches by providing information to improve the Helper site position weight matrix. With a better PWM, more divergent, but possibly important bipartite motifs will be identified. Our transgenic data supports a model in which the bipartite motif plays an instructive role in transcription expression patterns. It would be interesting to see if we could identify W-CRMs which display clusters of the same types of bipartite motifs, and if these motif cluster "signatures" could be correlated with tissue specific transcriptional

programs. This may be a complicated proposition, but as more individual W-CRMs are identified and their cognate genes confirmed, patterns of this sort may become apparent, and these patterns may, in turn, make identifying W-CRMS easier. Hopefully, the work described in this dissertation has provided information which makes this goal easier to attain.

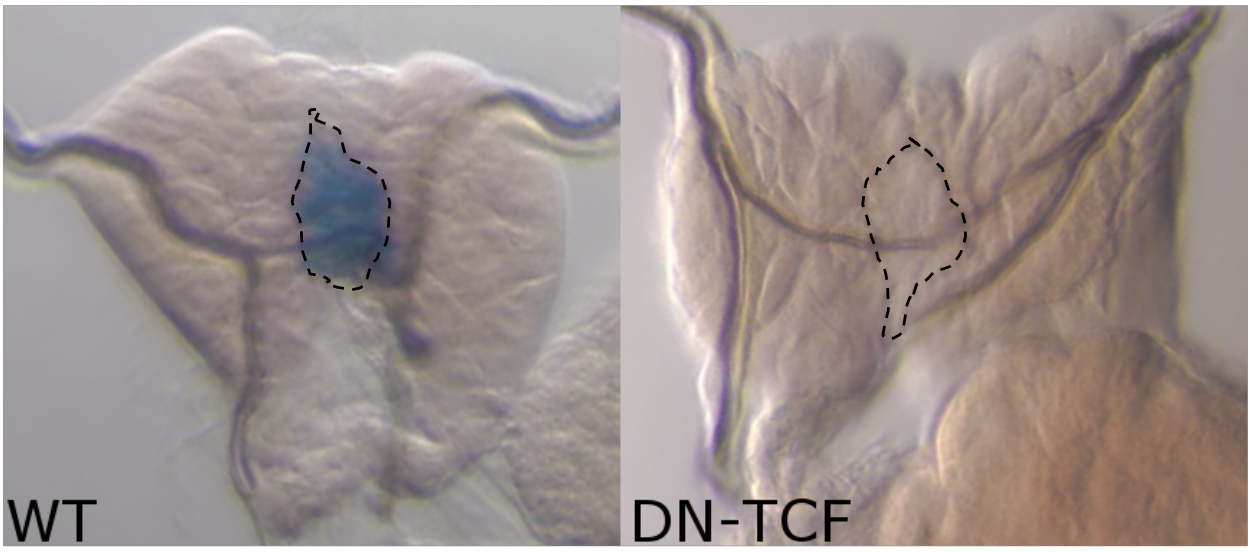


Figure 4.1 Expression of the RW0 W-CRM in the Corpus Allatum is blocked by dominant negative TCF.
Transcriptional activity of the RW0>LacZ synthetic W-CRM was assayed using X-gal staining (30 minutes). CA is indicated by dashed lines. Flies with the RW0 W-CRM and UAS-DNTCF were crossed to WT (w118, right) or Aug21-Gal4 (left) flies, and the F1 progeny assayed.

References

Adam, G., Perrimon, N., and Noselli, S. (2003). The retinoic-like juvenile hormone controls the looping of left-right asymmetric organs in Drosophila. Development *130*, 2397-2406.

Afouda, B. A., Martin, J., Liu, F., Ciau-Uitz, A., Patient, R., and Hoppler, S. (2008). GATA transcription factors integrate Wnt signalling during heart development. Development *135*, 3185-3190.

Ai, D., Fu, X., Wang, J., Lu, M. F., Chen, L., Baldini, A., Klein, W. H., and Martin, J. F. (2007). Canonical Wnt signaling functions in second heart field to promote right ventricular growth. Proceedings of the National Academy of Sciences of the United States of America *104*, 9319-9324.

Alfieri, C. M., Cheek, J., Chakraborty, S., and Yutzey, K. E. Wnt signaling in heart valve development and osteogenic gene induction. Developmental biology *338*, 127-135.

Amen, M., Liu, X., Vadlamudi, U., Elizondo, G., Diamond, E., Engelhardt, J. F., and Amendt, B. A. (2007). PITX2 and beta-catenin interactions regulate Lef-1 isoform expression. Molecular and cellular biology *27*, 7560-7573.

Arce, L., Yokoyama, N. N., and Waterman, M. L. (2006). Diversity of LEF/TCF action in development and disease. Oncogene *25*, 7492-7504.

Archbold, H. C., Yang, Y. X., Chen, L., and Cadigan, K. M. (2012). How do they do Wnt they do?: regulation of transcription by the Wnt/beta-catenin pathway. Acta physiologica *204*, 74-109.

Arnosti, D. N., and Kulkarni, M. M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? J Cell Biochem *94*, 890-898.

Arwert, E. N., Hoste, E., and Watt, F. M. (2012). Epithelial stem cells, wound healing and cancer. Nat Rev Cancer *12*, 170-180.

Atcha, F. A., Munguia, J. E., Li, T. W., Hovanes, K., and Waterman, M. L. (2003). A new beta-catenin-dependent activation domain in T cell factor. The Journal of biological chemistry *278*, 16169-16175.

Atcha, F. A., Syed, A., Wu, B., Hoverter, N. P., Yokoyama, N. N., Ting, J. H., Munguia, J. E., Mangalam, H. J., Marsh, J. L., and Waterman, M. L. (2007). A unique DNA binding domain

converts T-cell factors into strong Wnt effectors. Molecular and cellular biology *27*, 8352-8363.

Ayturk, U. M., Jacobsen, C. M., Christodoulou, D. C., Gorham, J., Seidman, J. G., Seidman, C. E., Robling, A. G., and Warman, M. L. (2013). An RNA-seq protocol to identify mRNA expression changes in mouse diaphyseal bone: applications in mice with bone property altering Lrp5 mutations. J Bone Miner Res *28*, 2081-2093.

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X.*, et al.* (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720-1723.

Baeg, G. H., Lin, X., Khare, N., Baumgartner, S., and Perrimon, N. (2001). Heparan sulfate proteoglycans are critical for the organization of the extracellular distribution of Wingless. Development *128*, 87-94.

Baker, N. E. (1988a). Localization of transcripts from the wingless gene in whole Drosophila embryos. Development *103*, 289-298.

Baker, N. E. (1988b). Transcription of the segment-polarity gene wingless in the imaginal discs of Drosophila, and the phenotype of a pupal-lethal wg mutation. Development *102*, 489-497.

Balmelle, N., Zamarreno, N., Krangel, M. S., and Hernandez-Munain, C. (2004). Developmental activation of the TCR alpha enhancer requires functional collaboration among proteins bound inside and outside the core enhancer. J Immunol *173*, 5054-5063.

Bansal, N., Kadamb, R., Mittal, S., Vig, L., Sharma, R., Dwarakanath, B. S., and Saluja, D. (2011). Tumor suppressor protein p53 recruits human Sin3B/HDAC1 complex for down-regulation of its target promoters in response to genotoxic stress. PLoS One *6*, e26156.

Barker, N., and Clevers, H. (2010). Leucine-rich repeat-containing G-protein-coupled receptors as markers of adult stem cells. Gastroenterology *138*, 1681-1696.

Barker, N., Tan, S., and Clevers, H. (2013). Lgr proteins in epithelial stem cell biology. Development *140*, 2484-2494.

Barlev, N. A., Liu, L., Chehab, N. H., Mansfield, K., Harris, K. G., Halazonetis, T. D., and Berger, S. L. (2001). Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. Mol Cell *8*, 1243-1254.

Barolo, S. (2006). Transgenic Wnt/TCF pathway reporters: all you need is Lef? Oncogene *25*, 7505-7511.

Barolo, S., and Posakony, J. W. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. Genes & development *16*, 1167-1181.

Behrens, J., von Kries, J. P., Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R., and Birchmeier, W. (1996). Functional interaction of beta-catenin with the transcription factor LEF-1. Nature *382*, 638-642.

Bejsovec, A., and Martinez Arias, A. (1991). Roles of wingless in patterning the larval epidermis of Drosophila. Development *113*, 471-485.

Beland, M., Pilon, N., Houle, M., Oh, K., Sylvestre, J. R., Prinos, P., and Lohnes, D. (2004). Cdx1 autoregulation is governed by a novel Cdx1-LEF1 transcription complex. Molecular and cellular biology *24*, 5028-5038.

Benchabane, H., Xin, N., Tian, A., Hafler, B. P., Nguyen, K., Ahmed, A., and Ahmed, Y. (2011). Jerky/Earthbound facilitates cell-specific Wnt/Wingless signalling by modulating beta-catenin-TCF activity. The EMBO journal *30*, 1444-1458.

Berwick, D. C., and Harvey, K. (2012). The importance of Wnt signalling for neurodegeneration in Parkinson's disease. Biochem Soc Trans *40*, 1123-1128.

Bhambhani, C., Chang, J. L., Akey, D. L., and Cadigan, K. M. (2011). The oligomeric state of CtBP determines its role as a transcriptional co-activator and co-repressor of Wingless targets. The EMBO journal *30*, 2031-2043.

Bhambhani, C., Ravindranath, A. J., Mentink, R. A., Chang, M. V., Betist, M. C., Yang, Y. X., Koushika, S. P., Korswagen, H. C., and Cadigan, K. M. (2014). Distinct DNA binding sites contribute to the TCF transcriptional switch in C. elegans and Drosophila. PLoS genetics *10*, e1004133.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proceedings of the National Academy of Sciences of the United States of America *104*, 3312-3317.

Blahnik, K. R., Dou, L., O'Geen, H., McPhillips, T., Xu, X., Cao, A. R., Iyengar, S., Nicolet, C. M., Ludascher, B., Korf, I., and Farnham, P. J. (2010). Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data. Nucleic acids research *38*, e13.

Blanpain, C., and Fuchs, E. (2009). Epidermal homeostasis: a balancing act of stem cells in the skin. Nat Rev Mol Cell Biol *10*, 207-217.

Blanpain, C., Horsley, V., and Fuchs, E. (2007). Epithelial stem cells: turning over new leaves. Cell *128*, 445-458.

Blauwkamp, T. A., Chang, M. V., and Cadigan, K. M. (2008). Novel TCF-binding sites specify transcriptional repression by Wnt signalling. The EMBO journal *27*, 1436-1446.

Bodmer, R., and Venkatesh, T. V. (1998). Heart development in Drosophila and vertebrates: conservation of molecular mechanisms. Developmental genetics *22*, 181-186.

Bottomly, D., Kyler, S. L., McWeeney, S. K., and Yochum, G. S. (2010). Identification of {beta}-catenin binding regions in colon cancer cells using ChIP-Seq. Nucleic acids research *38*, 5735-5745.

Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T., and Kimelman, D. (1997). A beta-catenin/XTcf-3 complex binds to the siamois promoter to regulate dorsal axis specification in Xenopus. Genes & development *11*, 2359-2370.

Brantjes, H., Barker, N., van Es, J., and Clevers, H. (2002). TCF: Lady Justice casting the final verdict on the outcome of Wnt signalling. Biological chemistry *383*, 255-261.

Brocardo, M., and Henderson, B. R. (2008). APC shuttling to the membrane, nucleus and beyond. Trends Cell Biol *18*, 587-596.

Brunner, E., Peter, O., Schweizer, L., and Basler, K. (1997). pangolin encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in Drosophila. Nature *385*, 829-833.

Cadigan, K. M. (2002). Regulating morphogen gradients in the Drosophila wing. Seminars in cell & developmental biology *13*, 83-90.

Cadigan, K. M. (2008). Wnt/beta-catenin signaling: turning the switch. Developmental cell *14*, 322-323.

Cadigan, K. M. (2012). TCFs and Wnt/beta-catenin signaling: more than one way to throw the switch. Current topics in developmental biology *98*, 1-34.

Cadigan, K. M., Fish, M. P., Rulifson, E. J., and Nusse, R. (1998). Wingless repression of Drosophila frizzled 2 expression shapes the Wingless morphogen gradient in the wing. Cell *93*, 767-777.

Cadigan, K. M., and Nusse, R. (1996). wingless signaling in the Drosophila eye and embryonic epidermis. Development *122*, 2801-2812.

Cadigan, K. M., and Nusse, R. (1997). Wnt signaling: a common theme in animal development. Genes & development *11*, 3286-3305.

Cadigan, K. M., and Peifer, M. (2009). Wnt signaling from development to disease: insights from model systems. Cold Spring Harbor perspectives in biology *1*, a002881.

Cadigan, K. M., and Waterman, M. L. (2012). TCF/LEFs and Wnt signaling in the nucleus. Cold Spring Harbor perspectives in biology *4*.

Cai, C. L., Liang, X., Shi, Y., Chu, P. H., Pfaff, S. L., Chen, J., and Evans, S. (2003). Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Developmental cell *5*, 877-889.

Carlsson, P., Waterman, M. L., and Jones, K. A. (1993). The hLEF/TCF-1 alpha HMG protein contains a context-dependent transcriptional activation domain that induces the TCR alpha enhancer in T cells. Genes & development *7*, 2418-2430.

Carrera, I., Janody, F., Leeds, N., Duveau, F., and Treisman, J. E. (2008). Pygopus activates Wingless target gene transcription through the mediator complex subunits Med12 and Med13. Proceedings of the National Academy of Sciences of the United States of America *105*, 6644-6649.

Castillo, H. A., Cravo, R. M., Azambuja, A. P., Simoes-Costa, M. S., Sura-Trueba, S., Gonzalez, J., Slonimsky, E., Almeida, K., Abreu, J. G., de Almeida, M. A*., et al.* (2010). Insights into the organization of dorsal spinal cord pathways from an evolutionarily conserved raldh2 intronic enhancer. Development *137*, 507-518.

Cavallo, R. A., Cox, R. T., Moline, M. M., Roose, J., Polevoy, G. A., Clevers, H., Peifer, M., and Bejsovec, A. (1998). Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature *395*, 604-608.

Chacon, M. A., Varela-Nallar, L., and Inestrosa, N. C. (2008). Frizzled-1 is involved in the neuroprotective effect of Wnt3a against Abeta oligomers. J Cell Physiol *217*, 215-227.

Chakladar, A., Dubeykovskiy, A., Wojtukiewicz, L. J., Pratap, J., Lei, S., and Wang, T. C. (2005). Synergistic activation of the murine gastrin promoter by oncogenic Ras and beta-catenin involves SMAD recruitment. Biochem Biophys Res Commun *336*, 190-196.

Chamorro, M. N., Schwartz, D. R., Vonica, A., Brivanlou, A. H., Cho, K. R., and Varmus, H. E. (2005). FGF-20 and DKK1 are transcriptional targets of beta-catenin and FGF-20 is implicated in cancer and development. The EMBO journal *24*, 73-84.

Chang, J. L., Chang, M. V., Barolo, S., and Cadigan, K. M. (2008a). Regulation of the feedback antagonist naked cuticle by Wingless signaling. Developmental biology *321*, 446-454.

Chang, M. V., Chang, J. L., Gangopadhyay, A., Shearer, A., and Cadigan, K. M. (2008b). Activation of wingless targets requires bipartite recognition of DNA by TCF. Current biology : CB *18*, 1877-1881.

Chen, H., Xu, Z., Mei, C., Yu, D., and Small, S. (2012). A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes. Cell *149*, 618-629.

Chen, S., McLean, S., Carter, D. E., and Leask, A. (2007). The gene expression profile induced by Wnt 3a in NIH 3T3 fibroblasts. Journal of cell communication and signaling *1*, 175-183.

Chera, S., Buzgariu, W., Ghila, L., and Galliot, B. (2009a). Autophagy in Hydra: a response to starvation and stress in early animal evolution. Biochimica et biophysica acta *1793*, 1432-1443.

Chera, S., Ghila, L., Dobretz, K., Wenger, Y., Bauer, C., Buzgariu, W., Martinou, J. C., and Galliot, B. (2009b). Apoptotic cells provide an unexpected source of Wnt3 signaling to drive hydra head regeneration. Developmental cell *17*, 279-289.

Chien, A. J., Conrad, W. H., and Moon, R. T. (2009). A Wnt survival guide: from flies to human disease. J Invest Dermatol *129*, 1614-1627.

Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. Cell *127*, 469-480.

Clevers, H., and van de Wetering, M. (1997). TCF/LEF factor earn their wings. Trends in genetics : TIG *13*, 485-489.

Cohen, E. D., Wang, Z., Lepore, J. J., Lu, M. M., Taketo, M. M., Epstein, D. J., and Morrisey, E. E. (2007). Wnt/beta-catenin signaling promotes expansion of Isl-1-positive cardiac progenitor cells through regulation of FGF signaling. J Clin Invest *117*, 1794-1804.

Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H., and Young, R. A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes & development *22*, 746-755.

Comijn, J., Berx, G., Vermassen, P., Verschueren, K., van Grunsven, L., Bruyneel, E., Mareel, M., Huylebroeck, D., and van Roy, F. (2001). The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. Mol Cell *7*, 1267-1278.

Couso, J. P., Bate, M., and Martinez-Arias, A. (1993). A wingless-dependent polar coordinate system in Drosophila imaginal discs. Science *259*, 484-489.

Curators., F., Members., S.-P. P., and Members., I. P. (2004). Gene Ontology annotation in FlyBase through association of InterPro records with GO terms. .

DasGupta, R., and Fuchs, E. (1999). Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. Development *126*, 4557-4568.

DasGupta, R., Kaykas, A., Moon, R. T., and Perrimon, N. (2005). Functional genomic analysis of the Wnt-wingless signaling pathway. Science *308*, 826-833.

Davidson, G., and Niehrs, C. (2010). Emerging links between CDK cell cycle regulators and Wnt signaling. Trends Cell Biol *20*, 453-460.

Dawson, S. J., Liu, Y. Z., Rodel, B., Moroy, T., and Latchman, D. S. (1996). The ability of POU family transcription factors to activate or repress gene expression is dependent on the

spacing and context of their specific response elements. The Biochemical journal *314 (Pt 2)*, 439-443.

De, A. (2011). Wnt/Ca2+ signaling pathway: a brief overview. Acta Biochim Biophys Sin (Shanghai) *43*, 745-756.

Dessimoz, J., Bonnard, C., Huelsken, J., and Grapin-Botton, A. (2005). Pancreas-specific deletion of beta-catenin reveals Wnt-dependent and Wnt-independent functions during development. Current biology : CB *15*, 1677-1683.

Di Cara, F., and King-Jones, K. (2013). How clocks and hormones act in concert to control the timing of insect development. Current topics in developmental biology *105*, 1-36.

Dietzl, G., Chen, D., Schnorrer, F., Su, K. C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S.*, et al.* (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature *448*, 151-156.

Domenzain-Reyna, C., Hernandez, D., Miquel-Serra, L., Docampo, M. J., Badenas, C., Fabra, A., and Bassols, A. (2009). Structure and regulation of the versican promoter: the versican promoter is regulated by AP-1 and TCF transcription factors in invasive human melanoma cells. The Journal of biological chemistry *284*, 12306-12317.

Dorsky, R. I., Itoh, M., Moon, R. T., and Chitnis, A. (2003). Two tcf3 genes cooperate to pattern the zebrafish brain. Development *130*, 1937-1947.

Dorsky, R. I., Sheldahl, L. C., and Moon, R. T. (2002). A transgenic Lef1/beta-catenin-dependent reporter is expressed in spatially restricted domains throughout zebrafish development. Developmental biology *241*, 229-237.

Driever, W., Thoma, G., and Nusslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen. Nature *340*, 363-367.

Dyer, L. A., and Kirby, M. L. (2009). The role of secondary heart field in cardiac development. Developmental biology *336*, 137-144.

Eivers, E., Demagny, H., Choi, R. H., and De Robertis, E. M. (2011). Phosphorylation of Mad controls competition between wingless and BMP signaling. Science signaling *4*, ra68.

Eivers, E., Fuentealba, L. C., Sander, V., Clemens, J. C., Hartnett, L., and De Robertis, E. M. (2009). Mad is required for wingless signaling in wing development and segment patterning in Drosophila. PLoS One *4*, e6543.

Evans, S. M. (1999). Vertebrate tinman homologues and cardiac differentiation. Seminars in cell & developmental biology *10*, 73-83.

Fan, Y., and Bergmann, A. (2008). Apoptosis-induced compensatory proliferation. The Cell is dead. Long live the Cell! Trends Cell Biol *18*, 467-473.

Fang, M., Li, J., Blauwkamp, T., Bhambhani, C., Campbell, N., and Cadigan, K. M. (2006). C-terminal-binding protein directly activates and represses Wnt transcriptional targets in Drosophila. The EMBO journal *25*, 2735-2745.

Foley, A. C., and Mercola, M. (2005). Heart induction by Wnt antagonists depends on the homeodomain transcription factor Hex. Genes & development *19*, 387-396.

Frietze, S., Wang, R., Yao, L., Tak, Y. G., Ye, Z., Gaddis, M., Witt, H., Farnham, P. J., and Jin, V. X. (2012). Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. Genome biology *13*, R52.

Fujino, T., Asaba, H., Kang, M. J., Ikeda, Y., Sone, H., Takada, S., Kim, D. H., Ioka, R. X., Ono, M., Tomoyori, H.*, et al.* (2003). Low-density lipoprotein receptor-related protein 5 (LRP5) is essential for normal cholesterol metabolism and glucose-induced insulin secretion. Proceedings of the National Academy of Sciences of the United States of America *100*, 229-234.

Galceran, J., Farinas, I., Depew, M. J., Clevers, H., and Grosschedl, R. (1999). Wnt3a-/--like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. Genes & development *13*, 709-717.

Galliot, B., and Chera, S. (2010). The Hydra model: disclosing an apoptosis-driven generator of Wnt-based regeneration. Trends Cell Biol *20*, 514-523.

Gan, X. Q., Wang, J. Y., Xi, Y., Wu, Z. L., Li, Y. P., and Li, L. (2008). Nuclear Dvl, c-Jun, beta-catenin, and TCF form a complex leading to stabilization of beta-catenin-TCF interaction. The Journal of cell biology *180*, 1087-1100.

Geng, X., Xiao, L., Lin, G. F., Hu, R., Wang, J. H., Rupp, R. A., and Ding, X. (2003). Lef/Tcf-dependent Wnt/beta-catenin signaling during Xenopus axis specification. FEBS letters *547*, 1-6.

Geserick, C., Meyer, H. A., and Haendler, B. (2005). The role of DNA response elements as allosteric modulators of steroid receptor function. Molecular and cellular endocrinology *236*, 1-7.

Gessert, S., and Kuhl, M. (2010). The multiple phases and faces of wnt signaling during cardiac differentiation and development. Circulation research *107*, 186-199.

Gibbens, Y. Y., Warren, J. T., Gilbert, L. I., and O'Connor, M. B. (2011). Neuroendocrine regulation of Drosophila metamorphosis requires TGFbeta/Activin signaling. Development *138*, 2693-2703.

Giege, R. (2013). A historical perspective on protein crystallization from 1840 to the present day. FEBS J *280*, 6456-6497.

Giese, K., Amsterdam, A., and Grosschedl, R. (1991). DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. Genes & development *5*, 2567-2578.

Giese, K., Cox, J., and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell *69*, 185-195.

Giese, K., and Grosschedl, R. (1993). LEF-1 contains an activation domain that stimulates transcription only in a specific context of factor-binding sites. The EMBO journal *12*, 4667-4676.

Giese, K., Kingsley, C., Kirshner, J. R., and Grosschedl, R. (1995). Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. Genes & development *9*, 995-1008.

Gitler, A. D., Lu, M. M., Jiang, Y. Q., Epstein, J. A., and Gruber, P. J. (2003). Molecular markers of cardiac endocardial cushion development. Developmental dynamics : an official publication of the American Association of Anatomists *228*, 643-650.

Grigoryan, T., Wend, P., Klaus, A., and Birchmeier, W. (2008). Deciphering the function of canonical Wnt signals in development and disease: conditional loss- and gain-of-function mutations of beta-catenin in mice. Genes & development *22*, 2308-2341.

Guruharsha, K. G., Rual, J. F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D. Y., Cenaj, O*., et al.* (2011). A protein complex network of Drosophila melanogaster. Cell *147*, 690-703.

Hacker, U., Grossniklaus, U., Gehring, W. J., and Jackle, H. (1992). Developmentally regulated Drosophila gene family encoding the fork head domain. Proceedings of the National Academy of Sciences of the United States of America *89*, 8754-8758.

Haegebarth, A., and Clevers, H. (2009). Wnt signaling, lgr5, and stem cells in the intestine and skin. Am J Pathol *174*, 715-721.

Halfon, M. S., Carmena, A., Gisselbrecht, S., Sackerson, C. M., Jimenez, F., Baylies, M. K., and Michelson, A. M. (2000). Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. Cell *103*, 63-74.

Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E., and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell *124*, 47-59.

Hamblet, N. S., Lijam, N., Ruiz-Lozano, P., Wang, J., Yang, Y., Luo, Z., Mei, L., Chien, K. R., Sussman, D. J., and Wynshaw-Boris, A. (2002). Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure. Development *129*, 5827-5838.

Han, Z., Fujioka, M., Su, M., Liu, M., Jaynes, J. B., and Bodmer, R. (2002). Transcriptional integration of competence modulated by mutual repression generates cell-type specificity within the cardiogenic mesoderm. Developmental biology *252*, 225-240.

Harmston, N., and Lenhard, B. (2013). Chromatin and epigenetic features of long-range gene regulation. Nucleic acids research *41*, 7185-7199.

Hatzis, P., van der Flier, L. G., van Driel, M. A., Guryev, V., Nielsen, F., Denissov, S., Nijman, I. J., Koster, J., Santo, E. E., Welboren, W.*, et al.* (2008). Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. Mol Cell Biol *28*, 2732-2744.

He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B., and Kinzler, K. W. (1998). Identification of c-MYC as a target of the APC pathway. Science *281*, 1509-1512.

Heasman, J., Crawford, A., Goldstone, K., Garner-Hamrick, P., Gumbiner, B., McCrea, P., Kintner, C., Noro, C. Y., and Wylie, C. (1994). Overexpression of cadherins and underexpression of beta-catenin inhibit dorsal mesoderm induction in early Xenopus embryos. Cell *79*, 791-803.

Hecht, A., and Stemmler, M. P. (2003). Identification of a promoter-specific transcriptional activation domain at the C terminus of the Wnt effector protein T-cell factor 4. The Journal of biological chemistry *278*, 3776-3785.

Herman, M. (2001). C. elegans POP-1/TCF functions in a canonical Wnt pathway that controls cell migration and in a noncanonical Wnt pathway that controls cell polarity. Development *128*, 581-590.

Herr, W., and Cleary, M. A. (1995). The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. Genes & development *9*, 1679-1693.

Hikasa, H., Ezan, J., Itoh, K., Li, X., Klymkowsky, M. W., and Sokol, S. Y. (2010). Regulation of TCF3 by Wnt-dependent phosphorylation during vertebrate axis specification. Developmental cell *19*, 521-532.

Hobmayer, B., Rentzsch, F., Kuhn, K., Happel, C. M., von Laue, C. C., Snyder, P., Rothbacher, U., and Holstein, T. W. (2000). WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra. Nature *407*, 186-189.

Hoch, M., and Pankratz, M. J. (1996). Control of gut development by fork head and cell signaling molecules in Drosophila. Mechanisms of development *58*, 3-14.

Hoffman, W. H., Biade, S., Zilfou, J. T., Chen, J., and Murphy, M. (2002). Transcriptional repression of the anti-apoptotic survivin gene by wild type p53. The Journal of biological chemistry *277*, 3247-3257.

Holland, J. D., Klaus, A., Garratt, A. N., and Birchmeier, W. (2013). Wnt signaling in stem and cancer stem cells. Current opinion in cell biology *25*, 254-264.

Hoogeboom, D., Essers, M. A., Polderman, P. E., Voets, E., Smits, L. M., and Burgering, B. M. (2008). Interaction of FOXO with beta-catenin inhibits beta-catenin/T cell factor activity. The Journal of biological chemistry *283*, 9224-9230.

Hoskins, R. A., Carlson, J. W., Kennedy, C., Acevedo, D., Evans-Holm, M., Frise, E., Wan, K. H., Park, S., Mendez-Lago, M., Rossi, F.*, et al.* (2007). Sequence finishing and mapping of Drosophila melanogaster heterochromatin. Science *316*, 1625-1628.

Hossain, M. B., Hosokawa, H., Hasegawa, A., Watarai, H., Taniguchi, M., Yamashita, M., and Nakayama, T. (2008). Lymphoid enhancer factor interacts with GATA-3 and controls its function in T helper type 2 cells. Immunology *125*, 377-386.

Houston, D. W., Kofron, M., Resnik, E., Langland, R., Destree, O., Wylie, C., and Heasman, J. (2002). Repression of organizer genes in dorsal and ventral Xenopus cells mediated by maternal XTcf3. Development *129*, 4015-4025.

Hoverter, N. P., Ting, J. H., Sundaresh, S., Baldi, P., and Waterman, M. L. (2012). A WNT/p21 circuit directed by the C-clamp, a sequence-specific DNA binding domain in TCFs. Molecular and cellular biology *32*, 3648-3662.

Hu, M. C., and Rosenblum, N. D. (2005). Smad1, beta-catenin and Tcf4 associate in a molecular complex with the Myc promoter in dysplastic renal tissue and cooperate to control Myc transcription. Development *132*, 215-225.

Huang, H. C., Sundseth, R., and Hansen, U. (1990). Transcription factor LSF binds two variant bipartite sites within the SV40 late promoter. Genes & development *4*, 287-298.

Huber, O., Korn, R., McLaughlin, J., Ohsugi, M., Herrmann, B. G., and Kemler, R. (1996). Nuclear localization of beta-catenin by interaction with transcription factor LEF-1. Mechanisms of development *59*, 3-10.

Hudson, W. H., Youn, C., and Ortlund, E. A. (2013). The structural basis of direct glucocorticoid-mediated transrepression. Nature structural & molecular biology *20*, 53-58.

Hurlstone, A. F., Haramis, A. P., Wienholds, E., Begthel, H., Korving, J., Van Eeden, F., Cuppen, E., Zivkovic, D., Plasterk, R. H., and Clevers, H. (2003). The Wnt/beta-catenin pathway regulates cardiac valve formation. Nature *425*, 633-637.

Hussein, S. M., Duff, E. K., and Sirard, C. (2003). Smad4 and beta-catenin co-activators functionally interact with lymphoid-enhancing factor to regulate graded expression of Msx2. The Journal of biological chemistry *278*, 48805-48814.

Ikeya, M., and Takada, S. (2001). Wnt-3a is required for somite specification along the anteroposterior axis of the mouse embryo and for regulation of cdx-1 expression. Mechanisms of development *103*, 27-33.

Inestrosa, N. C., and Toledo, E. M. (2008). The role of Wnt signaling in neuronal dysfunction in Alzheimer's Disease. Mol Neurodegener *3*, 9.

Itasaki, N., and Hoppler, S. Crosstalk between Wnt and bone morphogenic protein signaling: a turbulent relationship. Developmental dynamics : an official publication of the American Association of Anatomists *239*, 16-33.

Jackson, A., Vayssiere, B., Garcia, T., Newell, W., Baron, R., Roman-Roman, S., and Rawadi, G. (2005). Gene array analysis of Wnt-regulated genes in C3H10T1/2 cells. Bone *36*, 585-598.

Jacobson, E. M., Li, P., Leon-del-Rio, A., Rosenfeld, M. G., and Aggarwal, A. K. (1997). Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. Genes & development *11*, 198-212.

Jagla, K., Frasch, M., Jagla, T., Dretzen, G., Bellard, F., and Bellard, M. (1997a). ladybird, a new component of the cardiogenic pathway in Drosophila required for diversification of heart precursors. Development *124*, 3471-3479.

Jagla, K., Jagla, T., Heitzler, P., Dretzen, G., Bellard, F., and Bellard, M. (1997b). ladybird, a tandem of homeobox genes that maintain late wingless expression in terminal and dorsal epidermis of the Drosophila embryo. Development *124*, 91-100.

Jho, E. H., Zhang, T., Domon, C., Joo, C. K., Freund, J. N., and Costantini, F. (2002). Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. Molecular and cellular biology *22*, 1172-1183.

Ji, N., Middelkoop, T. C., Mentink, R. A., Betist, M. C., Tonegawa, S., Mooijman, D., Korswagen, H. C., and van Oudenaarden, A. (2013). Feedback Control of Gene Expression Variability in the Caenorhabditis elegans Wnt Pathway. Cell *155*, 869-880.

Jiang, J., and Levine, M. (1993). Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. Cell *72*, 741-752.

Jin, T. (2008). The WNT signalling pathway and diabetes mellitus. Diabetologia *51*, 1771-1780.

Johnson, R. A., Ince, T. A., and Scotto, K. W. (2001). Transcriptional repression by p53 through direct binding to a novel DNA element. The Journal of biological chemistry *276*, 27716-27720.

Joiner, D. M., Ke, J., Zhong, Z., Xu, H. E., and Williams, B. O. (2013). LRP5 and LRP6 in development and disease. Trends in endocrinology and metabolism: TEM *24*, 31-39.

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E. H., Birney, E., and Furlong, E. E. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. Cell *148*, 473-486.

Kalay, G., and Wittkopp, P. J. (2010). Nomadic enhancers: tissue-specific cis-regulatory elements of yellow have divergent genomic positions among Drosophila species. PLoS genetics *6*, e1001222.

Katz, R. W., and Koenig, R. J. (1994). Nucleotide substitutions differentially affect direct repeat and palindromic thyroid hormone response elements. The Journal of biological chemistry *269*, 9500-9505.

Kemler, I., Schreiber, E., Muller, M. M., Matthias, P., and Schaffner, W. (1989). Octamer transcription factors bind to two different sequence motifs of the immunoglobulin heavy chain promoter. The EMBO journal *8*, 2001-2008.

Kennell, J., and Cadigan, K. M. (2009). APC and beta-catenin degradation. Advances in experimental medicine and biology *656*, 1-12.

Kim, C. H., Oda, T., Itoh, M., Jiang, D., Artinger, K. B., Chandrasekharappa, S. C., Driever, W., and Chitnis, A. B. (2000). Repressor activity of Headless/Tcf3 is essential for vertebrate head formation. Nature *407*, 913-916.

King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I.*, et al.* (2008). The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature *451*, 783-788.

Kioussi, C., Briata, P., Baek, S. H., Rose, D. W., Hamblet, N. S., Herman, T., Ohgi, K. A., Lin, C., Gleiberman, A., Wang, J.*, et al.* (2002). Identification of a Wnt/Dvl/beta-Catenin --> Pitx2 pathway mediating cell-type-specific proliferation during development. Cell *111*, 673-685.

Klapholz-Brown, Z., Walmsley, G. G., Nusse, Y. M., Nusse, R., and Brown, P. O. (2007). Transcriptional program induced by Wnt protein in human fibroblasts suggests mechanisms for cell cooperativity in defining tissue microenvironments. PLoS One *2*, e945.

Klaus, A., Saga, Y., Taketo, M. M., Tzahor, E., and Birchmeier, W. (2007). Distinct roles of Wnt/beta-catenin and Bmp signaling during early cardiogenesis. Proceedings of the National Academy of Sciences of the United States of America *104*, 18531-18536.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell *77*, 21-32.

Kliewer, S. A., Umesono, K., Heyman, R. A., Mangelsdorf, D. J., Dyck, J. A., and Evans, R. M. (1992). Retinoid X receptor-COUP-TF interactions modulate retinoic acid signaling. Proceedings of the National Academy of Sciences of the United States of America *89*, 1448-1452.

Klingensmith, J., and Nusse, R. (1994). Signaling by wingless in Drosophila. Developmental biology *166*, 396-414.

Knirr, S., and Frasch, M. (2001). Molecular integration of inductive and mesoderm-intrinsic inputs governs even-skipped enhancer activity in a subset of pericardial and dorsal muscle progenitors. Developmental biology *238*, 13-26.

Kokoszynska, K., Ostrowski, J., Rychlewski, L., and Wyrwicz, L. S. (2008). The fold recognition of CP2 transcription factors gives new insights into the function and evolution of tumor suppressor protein p53. Cell Cycle *7*, 2907-2915.

Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P. J., and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet *19*, 379-383.

Korinek, V., Barker, N., Morin, P. J., van Wichen, D., de Weger, R., Kinzler, K. W., Vogelstein, B., and Clevers, H. (1997). Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science *275*, 1784-1787.

Kratochwil, K., Galceran, J., Tontsch, S., Roth, W., and Grosschedl, R. (2002). FGF4, a direct target of LEF1 and Wnt signaling, can rescue the arrest of tooth organogenesis in Lef1(-/-) mice. Genes & development *16*, 3173-3185.

Kwon, C., Arnold, J., Hsiao, E. C., Taketo, M. M., Conklin, B. R., and Srivastava, D. (2007). Canonical Wnt signaling is a positive regulator of mammalian cardiac progenitors. Proceedings of the National Academy of Sciences of the United States of America *104*, 10894-10899.

Kwon, C., Qian, L., Cheng, P., Nigam, V., Arnold, J., and Srivastava, D. (2009). A regulatory pathway involving Notch1/beta-catenin/Isl1 determines cardiac progenitor cell fate. Nature cell biology *11*, 951-957.

L'Episcopo, F., Tirolo, C., Caniglia, S., Testa, N., Morale, M. C., Serapide, M. F., Pluchino, S., and Marchetti, B. (2014a). Targeting Wnt signaling at the neuroimmune interface for dopaminergic neuroprotection/repair in Parkinson's disease. Journal of molecular cell biology *6*, 13-26.

L'Episcopo, F., Tirolo, C., Testa, N., Caniglia, S., Morale, M. C., Serapide, M. F., Pluchino, S., and Marchetti, B. (2014b). Wnt/beta-catenin signaling is required to rescue midbrain dopaminergic progenitors and promote neurorepair in ageing mouse model of Parkinson's disease. Stem Cells.

Labbe, E., Letamendia, A., and Attisano, L. (2000). Association of Smads with lymphoid enhancer binding factor 1/T cell-specific factor mediates cooperative signaling by the transforming growth factor-beta and wnt pathways. Proceedings of the National Academy of Sciences of the United States of America *97*, 8358-8363.

Lam, N., Chesney, M. A., and Kimble, J. (2006). Wnt signaling and CEH-22/tinman/Nkx2.5 specify a stem cell niche in C. elegans. Current biology : CB *16*, 287-295.

Laudet, V., Stehelin, D., and Clevers, H. (1993). Ancestry and diversity of the HMG box superfamily. Nucleic acids research *21*, 2493-2501.

Lawrence, P. A., Bodmer, R., and Vincent, J. P. (1995). Segmental patterning of heart precursors in Drosophila. Development *121*, 4303-4308.

LeBowitz, J. H., Clerc, R. G., Brenowitz, M., and Sharp, P. A. (1989). The Oct-2 protein binds cooperatively to adjacent octamer sites. Genes & development *3*, 1625-1638.

Lee, H. H., and Frasch, M. (2000). Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target sloppy paired. Development *127*, 5497-5508.

Lefstin, J. A., and Yamamoto, K. R. (1998). Allosteric effects of DNA on transcriptional regulators. Nature *392*, 885-888.

Lei, S., Dubeykovskiy, A., Chakladar, A., Wojtukiewicz, L., and Wang, T. C. (2004). The murine gastrin promoter is synergistically activated by transforming growth factor-beta/Smad and Wnt signaling pathways. The Journal of biological chemistry *279*, 42492-42502.

Lengfeld, T., Watanabe, H., Simakov, O., Lindgens, D., Gee, L., Law, L., Schmidt, H. A., Ozbek, S., Bode, H., and Holstein, T. W. (2009). Multiple Wnts are involved in Hydra organizer formation and regeneration. Developmental biology *330*, 186-199.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Human molecular genetics *12*, 1725-1735.

Leung, T. H., Hoffmann, A., and Baltimore, D. (2004). One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. Cell *118*, 453-464.

Li, J., Sutter, C., Parker, D. S., Blauwkamp, T., Fang, M., and Cadigan, K. M. (2007). CBP/p300 are bimodal regulators of Wnt signaling. The EMBO journal *26*, 2284-2294.

Li, P., He, X., Gerrero, M. R., Mok, M., Aggarwal, A., and Rosenfeld, M. G. (1993). Spacing and orientation of bipartite DNA-binding motifs as potential functional determinants for POU domain factors. Genes & development *7*, 2483-2496.

Li, V. S., Ng, S. S., Boersema, P. J., Low, T. Y., Karthaus, W. R., Gerlach, J. P., Mohammed, S., Heck, A. J., Maurice, M. M., Mahmoudi, T., and Clevers, H. (2012). Wnt signaling through inhibition of beta-catenin degradation in an intact Axin1 complex. Cell *149*, 1245-1256.

Lickert, H., Domon, C., Huls, G., Wehrle, C., Duluc, I., Clevers, H., Meyer, B. I., Freund, J. N., and Kemler, R. (2000). Wnt/(beta)-catenin signaling regulates the expression of the homeobox gene Cdx1 in embryonic intestine. Development *127*, 3805-3813.

Lickert, H., Kutsch, S., Kanzler, B., Tamai, Y., Taketo, M. M., and Kemler, R. (2002). Formation of multiple hearts in mice following deletion of beta-catenin in the embryonic endoderm. Developmental cell *3*, 171-181.

Liebner, S., Cattelino, A., Gallini, R., Rudini, N., Iurlaro, M., Piccolo, S., and Dejana, E. (2004). Beta-catenin is required for endothelial-mesenchymal transformation during heart cushion development in the mouse. The Journal of cell biology *166*, 359-367.

Lim, X., and Nusse, R. (2013). Wnt signaling in skin development, homeostasis, and disease. Cold Spring Harbor perspectives in biology *5*.

Lin, L., Cui, L., Zhou, W., Dufort, D., Zhang, X., Cai, C. L., Bu, L., Yang, L., Martin, J., Kemler, R.*, et al.* (2007). Beta-catenin directly regulates Islet1 expression in cardiovascular progenitors and is required for multiple aspects of cardiogenesis. Proceedings of the National Academy of Sciences of the United States of America *104*, 9313-9318.

Liu, F., and Millar, S. E. (2010). Wnt/beta-catenin signaling in oral tissue development and disease. Journal of dental research *89*, 318-330.

Liu, F., van den Broek, O., Destree, O., and Hoppler, S. (2005). Distinct roles for Xenopus Tcf/Lef genes in mediating specific responses to Wnt/beta-catenin signalling in mesoderm development. Development *132*, 5375-5385.

Liu, Y., Asakura, M., Inoue, H., Nakamura, T., Sano, M., Niu, Z., Chen, M., Schwartz, R. J., and Schneider, M. D. (2007). Sox17 is essential for the specification of cardiac mesoderm in embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *104*, 3859-3864.

Liu, Y. I., Chang, M. V., Li, H. E., Barolo, S., Chang, J. L., Blauwkamp, T. A., and Cadigan, K. M. (2008). The chromatin remodelers ISWI and ACF1 directly repress Wingless transcriptional targets. Developmental biology *323*, 41-52.

Liu, Z., and Habener, J. F. (2008). Glucagon-like peptide-1 activation of TCF7L2-dependent Wnt signaling enhances pancreatic beta cell proliferation. The Journal of biological chemistry *283*, 8723-8735.

Logan, C. Y., and Nusse, R. (2004). The Wnt signaling pathway in development and disease. Annu Rev Cell Dev Biol *20*, 781-810.

Longo, K. A., Kennell, J. A., Ochocinska, M. J., Ross, S. E., Wright, W. S., and MacDougald, O. A. (2002). Wnt signaling protects 3T3-L1 preadipocytes from apoptosis through induction of insulin-like growth factors. The Journal of biological chemistry *277*, 38239-38244.

Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature *376*, 791-795.

Lum, L., Yao, S., Mozer, B., Rovescalli, A., Von Kessler, D., Nirenberg, M., and Beachy, P. A. (2003). Identification of Hedgehog pathway components by RNAi in Drosophila cultured cells. Science *299*, 2039-2045.

MacDonald, B. T., Tamai, K., and He, X. (2009). Wnt/beta-catenin signaling: components, mechanisms, and diseases. Developmental cell *17*, 9-26.

Malone, J. H., and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. BMC biology *9*, 34.

Maloof, J. N., Whangbo, J., Harris, J. M., Jongeward, G. D., and Kenyon, C. (1999). A Wnt signaling pathway controls hox gene expression and neuroblast migration in C. elegans. Development *126*, 37-49.

Mangelsdorf, D. J., and Evans, R. M. (1995). The RXR heterodimers and orphan receptors. Cell *83*, 841-850.

Manolagas, S. C., and Almeida, M. (2007). Gone with the Wnts: beta-catenin, T-cell factor, forkhead box O, and oxidative stress in age-dependent diseases of bone, lipid, and glucose metabolism. Mol Endocrinol *21*, 2605-2614.

Maretto, S., Cordenonsi, M., Dupont, S., Braghetta, P., Broccoli, V., Hassan, A. B., Volpin, D., Bressan, G. M., and Piccolo, S. (2003). Mapping Wnt/beta-catenin signaling during mouse development and in colorectal tumors. Proceedings of the National Academy of Sciences of the United States of America *100*, 3299-3304.

Martin, F. A., Perez-Garijo, A., and Morata, G. (2009). Apoptosis in Drosophila: compensatory proliferation and undead cells. Int J Dev Biol *53*, 1341-1347.

Marvin, M. J., Di Rocco, G., Gardiner, A., Bush, S. M., and Lassar, A. B. (2001). Inhibition of Wnt activity induces heart formation from posterior mesoderm. Genes & development *15*, 316-327.

Masckauchan, T. N., Shawber, C. J., Funahashi, Y., Li, C. M., and Kitajewski, J. (2005). Wnt/beta-catenin signaling induces proliferation, survival and interleukin-8 in human endothelial cells. Angiogenesis *8*, 43-51.

McGuire, S. E., Mao, Z., and Davis, R. L. (2004). Spatiotemporal gene expression targeting with the TARGET and gene-switch systems in Drosophila. Science's STKE : signal transduction knowledge environment *2004*, pl6.

McMahon, A. P., and Moon, R. T. (1989). Ectopic expression of the proto-oncogene int-1 in Xenopus embryos leads to duplication of the embryonic axis. Cell *58*, 1075-1084.

Meijsing, S. H., Pufall, M. A., So, A. Y., Bates, D. L., Chen, L., and Yamamoto, K. R. (2009). DNA binding site sequence directs glucocorticoid receptor structure and activity. Science *324*, 407-410.

Merrill, B. J., Pasolli, H. A., Polak, L., Rendl, M., Garcia-Garcia, M. J., Anderson, K. V., and Fuchs, E. (2004). Tcf3: a transcriptional regulator of axis induction in the early embryo. Development *131*, 263-274.

Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O., and Clevers, H. (1996). XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. Cell *86*, 391-399.

Morishita, K., Suzukawa, K., Taki, T., Ihle, J. N., and Yokota, J. (1995). EVI-1 zinc finger protein works as a transcriptional activator via binding to a consensus sequence of GACAAGATAAGATAAN1-28 CTCATCTTC. Oncogene *10*, 1961-1967.

Mosimann, C., Hausmann, G., and Basler, K. (2009). Beta-catenin hits chromatin: regulation of Wnt target gene activation. Nat Rev Mol Cell Biol *10*, 276-286.

Moustakas, A., and Heldin, C. H. (2009). The regulation of TGFbeta signal transduction. Development *136*, 3699-3714.

Moutier, E., Ye, T., Choukrallah, M. A., Urban, S., Osz, J., Chatagnon, A., Delacroix, L., Langer, D., Rochel, N., Moras, D.*, et al.* (2012). Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. The Journal of biological chemistry *287*, 26328-26341.

Mukhopadhyay, M., Shtrom, S., Rodriguez-Esteban, C., Chen, L., Tsukui, T., Gomer, L., Dorward, D. W., Glinka, A., Grinberg, A., Huang, S. P.*, et al.* (2001). Dickkopf1 is required for embryonic head induction and limb morphogenesis in the mouse. Developmental cell *1*, 423-434.

Muncan, V., Sansom, O. J., Tertoolen, L., Phesse, T. J., Begthel, H., Sancho, E., Cole, A. M., Gregorieff, A., de Alboran, I. M., Clevers, H., and Clarke, A. R. (2006). Rapid loss of intestinal crypts upon conditional deletion of the Wnt/Tcf-4 target gene c-Myc. Molecular and cellular biology *26*, 8418-8426.

Munson, P. J., and Rodbard, D. (1988). An exact correction to the "Cheng-Prusoff" correction. Journal of receptor research *8*, 533-546.

Naar, A. M., Boutin, J. M., Lipkin, S. M., Yu, V. C., Holloway, J. M., Glass, C. K., and Rosenfeld, M. G. (1991). The orientation and spacing of core DNA-binding motifs dictate selective transcriptional responses to three nuclear receptors. Cell *65*, 1267-1279.

Nakano, N., Itoh, S., Watanabe, Y., Maeyama, K., Itoh, F., and Kato, M. (2010). Requirement of TCF7L2 for TGF-{beta}-dependent transcriptional activation of the TMEPAI gene. The Journal of biological chemistry.

Nakaya, M. A., Biris, K., Tsukiyama, T., Jaime, S., Rawls, J. A., and Yamaguchi, T. P. (2005). Wnt3a links left-right determination with segmentation and anteroposterior axis elongation. Development *132*, 5425-5436.

Nateri, A. S., Spencer-Dene, B., and Behrens, A. (2005). Interaction of phosphorylated c-Jun with TCF4 regulates intestinal cancer development. Nature *437*, 281-285.

Nguyen, H., Merrill, B. J., Polak, L., Nikolova, M., Rendl, M., Shaver, T. M., Pasolli, H. A., and Fuchs, E. (2009). Tcf3 and Tcf4 are essential for long-term homeostasis of skin epithelia. Nat Genet *41*, 1068-1075.

Ni, Z., Anini, Y., Fang, X., Mills, G., Brubaker, P. L., and Jin, T. (2003). Transcriptional activation of the proglucagon gene by lithium and beta-catenin in intestinal endocrine L cells. The Journal of biological chemistry *278*, 1380-1387.

Niehrs, C. (2010). On growth and form: a Cartesian coordinate system of Wnt and BMP signaling specifies bilaterian body axes. Development *137*, 845-857.

Nishita, M., Hashimoto, M. K., Ogata, S., Laurent, M. N., Ueno, N., Shibuya, H., and Cho, K. W. (2000). Interaction between Wnt and TGF-beta signalling pathways during formation of Spemann's organizer. Nature *403*, 781-785.

Noordermeer, J., Klingensmith, J., Perrimon, N., and Nusse, R. (1994). dishevelled and armadillo act in the wingless signalling pathway in Drosophila. Nature *367*, 80-83.

Norton, L., Fourcaudot, M., Abdul-Ghani, M. A., Winnier, D., Mehta, F. F., Jenkinson, C. P., and Defronzo, R. A. (2011). Chromatin occupancy of transcription factor 7-like 2 (TCF7L2) and its role in hepatic glucose metabolism. Diabetologia *54*, 3132-3142.

Nusse, R., Fuerer, C., Ching, W., Harnish, K., Logan, C., Zeng, A., ten Berge, D., and Kalani, Y. (2008). Wnt signaling and stem cell control. Cold Spring Harbor symposia on quantitative biology *73*, 59-66.

Nusslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795-801.

Ochoa-Espinosa, A., Yucel, G., Kaplan, L., Pare, A., Pura, N., Oberstein, A., Papatsenko, D., and Small, S. (2005). The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *102*, 4960-4965.

Olefsky, J. M. (2001). Nuclear receptor minireview series. The Journal of biological chemistry *276*, 36863-36864.

Oosterwegel, M. A., van de Wetering, M. L., Holstege, F. C., Prosser, H. M., Owen, M. J., and Clevers, H. C. (1991). TCF-1, a T cell-specific transcription factor of the HMG box family, interacts with sequence motifs in the TCR beta and TCR delta enhancers. International immunology *3*, 1189-1192.

Park, M., Wu, X., Golden, K., Axelrod, J. D., and Bodmer, R. (1996). The wingless signaling pathway is directly involved in Drosophila heart development. Developmental biology *177*, 104-116.

Parker DS, B. T., Cadigan KM (2007). Wnt mediated transcriptional regulation. Advances in Developmental Biology 17 : 1-60. Advances in Developmental Biology *17*, 1-60.

Parker, D. S., Jemison, J., and Cadigan, K. M. (2002). Pygopus, a nuclear PHD-finger protein required for Wingless signaling in Drosophila. Development *129*, 2565-2576.

Parker, D. S., Ni, Y. Y., Chang, J. L., Li, J., and Cadigan, K. M. (2008). Wingless signaling induces widespread chromatin remodeling of target loci. Molecular and cellular biology *28*, 1815-1828.

Parker, D. S., White, M. A., Ramos, A. I., Cohen, B. A., and Barolo, S. (2011). The cis-regulatory logic of Hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity. Science signaling *4*, ra38.

Peifer, M., Rauskolb, C., Williams, M., Riggleman, B., and Wieschaus, E. (1991). The segment polarity gene armadillo interacts with the wingless signaling pathway in both embryonic and adult pattern formation. Development *111*, 1029-1043.

Perez-Garijo, A., Shlevkov, E., and Morata, G. (2009). The role of Dpp and Wg in compensatory proliferation and in the formation of hyperplastic overgrowths caused by apoptotic cells in the Drosophila wing disc. Development *136*, 1169-1177.

Perlmann, T., Rangarajan, P. N., Umesono, K., and Evans, R. M. (1993). Determinants for selective RAR and TR recognition of direct repeat HREs. Genes & development *7*, 1411-1422.

Perrimon, N., Pitsouli, C., and Shilo, B. Z. (2012). Signaling mechanisms controlling cell fate and embryonic patterning. Cold Spring Harbor perspectives in biology *4*, a005975.

Petersen, C. P., and Reddien, P. W. (2009). Wnt signaling and the polarity of the primary body axis. Cell *139*, 1056-1068.

Petrova, I. M., Malessy, M. J., Verhaagen, J., Fradkin, L. G., and Noordermeer, J. N. (2014). Wnt signaling through the Ror receptor in the nervous system. Mol Neurobiol *49*, 303-315.

Phan, T. Q., Jow, M. M., and Privalsky, M. L. (2010). DNA recognition by thyroid hormone and retinoic acid receptors: 3,4,5 rule modified. Molecular and cellular endocrinology *319*, 88-98.

Phelan, S. A., and Loeken, M. R. (1998). Identification of a new binding motif for the paired domain of Pax-3 and unusual characteristics of spacing of bipartite recognition elements on binding and transcription activation. The Journal of biological chemistry *273*, 19153-19159.

Phillips, B. T., and Kimble, J. (2009). A new look at TCF and beta-catenin through the lens of a divergent C. elegans Wnt pathway. Developmental cell *17*, 27-34.

Phillips, K., and Luisi, B. (2000). The virtuoso of versatility: POU proteins that flex to fit. Journal of molecular biology *302*, 1023-1039.

Phillips, R. G., and Whittle, J. R. (1993). wingless expression mediates determination of peripheral nervous system elements in late stages of Drosophila wing disc development. Development *118*, 427-438.

Pilon, N., Oh, K., Sylvestre, J. R., Bouchard, N., Savory, J., and Lohnes, D. (2006). Cdx4 is a direct target of the canonical Wnt pathway. Developmental biology *289*, 55-63.

Pinto, C., Cardenas, P., Osses, N., and Henriquez, J. P. (2013). Characterization of Wnt/beta-catenin and BMP/Smad signaling pathways in an in vitro model of amyotrophic lateral sclerosis. Frontiers in cellular neuroscience *7*, 239.

Pinto, D., and Clevers, H. (2005). Wnt control of stem cells and differentiation in the intestinal epithelium. Exp Cell Res *306*, 357-363.

Poellinger, L., Yoza, B. K., and Roeder, R. G. (1989). Functional cooperativity between protein molecules bound at two distinct sequence elements of the immunoglobulin heavy-chain promoter. Nature *337*, 573-576.

Polakis, P. (2007). The many ways of Wnt in cancer. Current opinion in genetics & development *17*, 45-51.

Polakis, P. (2012). Wnt signaling in cancer. Cold Spring Harbor perspectives in biology *4*.

Pomerantz, M. M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M. P., Doddapaneni, H., Beckwith, C. A., Chan, J. A., Hills, A., Davis, M.*, et al.* (2009). The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet *41*, 882-884.

Postigo, A. A., and Dean, D. C. (2000). Differential expression and function of members of the zfh-1 family of zinc finger/homeodomain repressors. Proceedings of the National Academy of Sciences of the United States of America *97*, 6391-6396.

Pyrowolakis, G., Hartmann, B., Muller, B., Basler, K., and Affolter, M. (2004). A simple molecular complex mediates widespread BMP-induced repression during Drosophila development. Developmental cell *7*, 229-240.

Railo, A., Pajunen, A., Itaranta, P., Naillat, F., Vuoristo, J., Kilpelainen, P., and Vainio, S. (2009). Genomic response to Wnt signalling is highly context-dependent--evidence from

DNA microarray and chromatin immunoprecipitation screens of Wnt/TCF targets. Exp Cell Res *315*, 2690-2704.

Rajabi, H., Ahmad, R., Jin, C., Kosugi, M., Alam, M., Joshi, M. D., and Kufe, D. (2012). MUC1-C oncoprotein induces TCF7L2 transcription factor activation and promotes cyclin D1 expression in human breast cancer cells. The Journal of biological chemistry *287*, 10703-10713.

Ramos, A. I., and Barolo, S. (2013). Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. Philosophical transactions of the Royal Society of London Series B, Biological sciences *368*, 20130018.

Ravindranath, A., and Cadigan, K. M. (2014). Structure-function analysis of the C-clamp of TCF/Pangolin in Wnt/ss-catenin signaling. PLoS One *9*, e86180.

Regard, J. B., Zhong, Z., Williams, B. O., and Yang, Y. (2012). Wnt signaling in bone development and disease: making stronger bone with Wnts. Cold Spring Harbor perspectives in biology *4*.

Remacle, J. E., Kraft, H., Lerchner, W., Wuytens, G., Collart, C., Verschueren, K., Smith, J. C., and Huylebroeck, D. (1999). New mode of DNA binding of multi-zinc finger transcription factors: deltaEF1 family members bind with two hands to two target sites. The EMBO journal *18*, 5073-5084.

Rewitz, K. F., Yamanaka, N., and O'Connor, M. B. (2013). Developmental checkpoints and feedback circuits time insect maturation. Current topics in developmental biology *103*, 1-33.

Reya, T., O'Riordan, M., Okamura, R., Devaney, E., Willert, K., Nusse, R., and Grosschedl, R. (2000). Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. Immunity *13*, 15-24.

Riggleman, B., Wieschaus, E., and Schedl, P. (1989). Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene. Genes & development *3*, 96-113.

Rijsewijk, F., Schuermann, M., Wagenaar, E., Parren, P., Weigel, D., and Nusse, R. (1987). The Drosophila homolog of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless. Cell *50*, 649-657.

Rivat, C., Le Floch, N., Sabbah, M., Teyrol, I., Redeuilh, G., Bruyneel, E., Mareel, M., Matrisian, L. M., Crawford, H. C., Gespach, C., and Attoub, S. (2003). Synergistic cooperation between the AP-1 and LEF-1 transcription factors in activation of the matrilysin promoter by the src oncogene: implications in cellular invasion. FASEB J *17*, 1721-1723.

Rocheleau, C. E., Downs, W. D., Lin, R., Wittmann, C., Bei, Y., Cha, Y. H., Ali, M., Priess, J. R., and Mello, C. C. (1997). Wnt signaling and an APC-related gene specify endoderm in early C. elegans embryos. Cell *90*, 707-716.

Rodriguez-Carballo, E., Ulsamer, A., Susperregui, A. R., Manzanares-Cespedes, C., Sanchez-Garcia, E., Bartrons, R., Rosa, J. L., and Ventura, F. (2010). Conserved regulatory motifs in osteogenic gene promoters integrate cooperative effects of canonical Wnt and BMP pathways. J Bone Miner Res.

Roose, J., Huls, G., van Beest, M., Moerer, P., van der Horn, K., Goldschmeding, R., Logtenberg, T., and Clevers, H. (1999). Synergy between tumor suppressor APC and the beta-catenin-Tcf4 target Tcf1. Science *285*, 1923-1926.

Roose, J., Molenaar, M., Peterson, J., Hurenkamp, J., Brantjes, H., Moerer, P., van de Wetering, M., Destree, O., and Clevers, H. (1998). The Xenopus Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. Nature *395*, 608-612.

Rulifson, I. C., Karnik, S. K., Heiser, P. W., ten Berge, D., Chen, H., Gu, X., Taketo, M. M., Nusse, R., Hebrok, M., and Kim, S. K. (2007). Wnt signaling regulates pancreatic beta cell proliferation. Proceedings of the National Academy of Sciences of the United States of America *104*, 6247-6252.

Sanchez-Camacho, C., and Bovolenta, P. (2009). Emerging mechanisms in morphogen-mediated axon guidance. BioEssays : news and reviews in molecular, cellular and developmental biology *31*, 1013-1025.

Sancho, R., Nateri, A. S., de Vinuesa, A. G., Aguilera, C., Nye, E., Spencer-Dene, B., and Behrens, A. (2009). JNK signalling modulates intestinal homeostasis and tumourigenesis in mice. The EMBO journal *28*, 1843-1854.

Sanyal, A., Lajoie, B. R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109-113.

Sato, T., and Clevers, H. (2013). Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. Science *340*, 1190-1194.

Schinner, S., Ulgen, F., Papewalis, C., Schott, M., Woelk, A., Vidal-Puig, A., and Scherbaum, W. A. (2008). Regulation of insulin secretion, glucokinase gene transcription and beta cell proliferation by adipocyte-derived Wnt signalling molecules. Diabetologia *51*, 147-154.

Schinner, S., Willenberg, H. S., Schott, M., and Scherbaum, W. A. (2009). Pathophysiological aspects of Wnt-signaling in endocrine disease. Eur J Endocrinol *160*, 731-737.

Schneider, V. A., and Mercola, M. (2001). Wnt antagonism initiates cardiogenesis in Xenopus laevis. Genes & development *15*, 304-315.

Schweizer, L., Nellen, D., and Basler, K. (2003). Requirement for Pangolin/dTCF in Drosophila Wingless signaling. Proceedings of the National Academy of Sciences of the United States of America *100*, 5846-5851.

Scully, K. M., Jacobson, E. M., Jepsen, K., Lunyak, V., Viadiu, H., Carriere, C., Rose, D. W., Hooshmand, F., Aggarwal, A. K., and Rosenfeld, M. G. (2000). Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. Science *290*, 1127-1131.

Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature *451*, 535-540.

Shafer, S. L., and Towler, D. A. (2009). Transcriptional regulation of SM22alpha by Wnt3a: convergence with TGFbeta(1)/Smad signaling at a novel regulatory element. J Mol Cell Cardiol *46*, 621-635.

Shaffer, P. L., Jivan, A., Dollins, D. E., Claessens, F., and Gewirth, D. T. (2004). Structural basis of androgen receptor binding to selective androgen response elements. Proceedings of the National Academy of Sciences of the United States of America *101*, 4758-4763.

Shao, W., Wang, D., Chiang, Y. T., Ip, W., Zhu, L., Xu, F., Columbus, J., Belsham, D. D., Irwin, D. M., Zhang, H.*, et al.* (2013). The Wnt signaling pathway effector TCF7L2 controls gut and brain proglucagon gene expression and glucose homeostasis. Diabetes *62*, 789-800.

Sharma, R. P., and Chopra, V. L. (1976). Effect of the Wingless (wg1) mutation on wing and haltere development in Drosophila melanogaster. Developmental biology *48*, 461-465.

Shaulian, E., and Karin, M. (2002). AP-1 as a regulator of cell life and death. Nature cell biology *4*, E131-136.

Shetty, P., Lo, M. C., Robertson, S. M., and Lin, R. (2005). C. elegans TCF protein, POP-1, converts from repressor to activator as a result of Wnt-induced lowering of nuclear levels. Developmental biology *285*, 584-592.

Shirra, M. K., and Hansen, U. (1998). LSF and NTF-1 share a conserved DNA recognition motif yet require different oligomerization states to form a stable protein-DNA complex. The Journal of biological chemistry *273*, 19260-19268.

Shirra, M. K., Zhu, Q., Huang, H. C., Pallas, D., and Hansen, U. (1994). One exon of the human LSF gene includes conserved regions involved in novel DNA-binding and dimerization motifs. Molecular and cellular biology *14*, 5076-5087.

Shu, L., Sauter, N. S., Schulthess, F. T., Matveyenko, A. V., Oberholzer, J., and Maedler, K. (2008). Transcription factor 7-like 2 regulates beta-cell survival and function in human pancreatic islets. Diabetes *57*, 645-653.

Siegfried, E., Wilder, E. L., and Perrimon, N. (1994). Components of wingless signalling in Drosophila. Nature *367*, 76-80.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet *38*, 1348-1354.

Skrisovska, L., Schubert, M., and Allain, F. H. (2010). Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. Journal of biomolecular NMR *46*, 51-65.

Slattery, M., Ma, L., Negre, N., White, K. P., and Mann, R. S. (2011). Genome-wide tissue-specific occupancy of the Hox protein Ultrabithorax and Hox cofactor Homothorax in Drosophila. PLoS One *6*, e14686.

Smith, W. C., and Harland, R. M. (1991). Injected Xwnt-8 RNA acts early in Xenopus embryos to promote formation of a vegetal dorsalizing center. Cell *67*, 753-765.

Smith-Bolton, R. K., Worley, M. I., Kanda, H., and Hariharan, I. K. (2009). Regenerative growth in Drosophila imaginal discs is regulated by Wingless and Myc. Developmental cell *16*, 797-809.

Snippert, H. J., van der Flier, L. G., Sato, T., van Es, J. H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A. M., van Rheenen, J., Simons, B. D., and Clevers, H. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134-144.

Sokol, S., Christian, J. L., Moon, R. T., and Melton, D. A. (1991). Injected Wnt RNA induces a complete body axis in Xenopus embryos. Cell *67*, 741-752.

Sosinsky, A., Bonin, C. P., Mann, R. S., and Honig, B. (2003). Target Explorer: An automated tool for the identification of new target genes for a specified set of transcription factors. Nucleic acids research *31*, 3589-3592.

Standley, H. J., Destree, O., Kofron, M., Wylie, C., and Heasman, J. (2006). Maternal XTcf1 and XTcf4 have distinct roles in regulating Wnt target genes. Developmental biology *289*, 318-328.

Stepniak, E., Radice, G. L., and Vasioukhin, V. (2009). Adhesive and signaling functions of cadherins and catenins in vertebrate development. Cold Spring Harbor perspectives in biology *1*, a002949.

Stoick-Cooper, C. L., Weidinger, G., Riehle, K. J., Hubbert, C., Major, M. B., Fausto, N., and Moon, R. T. (2007). Distinct Wnt signaling pathways have opposing roles in appendage regeneration. Development *134*, 479-489.

Strigini, M., and Cohen, S. M. (2000). Wingless gradient formation in the Drosophila wing. Current biology : CB *10*, 293-300.

Struhl, G., Struhl, K., and Macdonald, P. M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. Cell *57*, 1259-1273.

Suda, Y., Kokura, K., Kimura, J., Kajikawa, E., Inoue, F., and Aizawa, S. The same enhancer regulates the earliest Emx2 expression in caudal forebrain primordium, subsequent expression in dorsal telencephalon and later expression in the cortical ventricular zone. Development *137*, 2939-2949.

Surjit, M., Ganti, K. P., Mukherji, A., Ye, T., Hua, G., Metzger, D., Li, M., and Chambon, P. (2011). Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. Cell *145*, 224-241.

Swanson, C. I., Schwimmer, D. B., and Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. Current biology : CB *21*, 1186-1196.

Swarup, S., and Verheyen, E. M. (2012). Wnt/Wingless signaling in Drosophila. Cold Spring Harbor perspectives in biology *4*.

Taelman, V. F., Dobrowolski, R., Plouhinec, J. L., Fuentealba, L. C., Vorwald, P. P., Gumper, I., Sabatini, D. D., and De Robertis, E. M. (2010). Wnt signaling requires sequestration of glycogen synthase kinase 3 inside multivesicular endosomes. Cell *143*, 1136-1148.

Tanaka, K., Shouguchi-Miyata, J., Miyamoto, N., and Ikeda, J. E. (2004). Novel nuclear shuttle proteins, HDBP1 and HDBP2, bind to neuronal cell-specific cis-regulatory element in the promoter for the human Huntington's disease gene. The Journal of biological chemistry *279*, 7275-7286.

Tanaka, S. S., Kojima, Y., Yamaguchi, Y. L., Nishinakamura, R., and Tam, P. P. (2011). Impact of WNT signaling on tissue lineage differentiation in the early mouse embryo. Development, growth & differentiation *53*, 843-856.

Tang, W., Dodge, M., Gundapaneni, D., Michnoff, C., Roth, M., and Lum, L. (2008). A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer. Proceedings of the National Academy of Sciences of the United States of America *105*, 9697-9702.

Tata, J. R. (2002). Signalling through nuclear receptors. Nat Rev Mol Cell Biol *3*, 702-710.

Theil, T., Aydin, S., Koch, S., Grotewold, L., and Ruther, U. (2002). Wnt and Bmp signalling cooperatively regulate graded Emx2 expression in the dorsal telencephalon. Development *129*, 3045-3054.

Thorpe, C. J., Schlesinger, A., Carter, J. C., and Bowerman, B. (1997). Wnt signaling polarizes an early C. elegans blastomere to distinguish endoderm from mesoderm. Cell *90*, 695-705.

Tian, Y., Yuan, L., Goss, A. M., Wang, T., Yang, J., Lepore, J. J., Zhou, D., Schwartz, R. J., Patel, V., Cohen, E. D., and Morrisey, E. E. (2010). Characterization and in vivo pharmacological rescue of a Wnt2-Gata6 pathway required for cardiac inflow tract development. Developmental cell *18*, 275-287.

Tokino, T., Thiagalingam, S., el-Deiry, W. S., Waldman, T., Kinzler, K. W., and Vogelstein, B. (1994). p53 tagged sites from human genomic DNA. Human molecular genetics *3*, 1537-1542.

Tomilin, A., Remenyi, A., Lins, K., Bak, H., Leidel, S., Vriend, G., Wilmanns, M., and Scholer, H. R. (2000). Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. Cell *103*, 853-864.

Toualbi, K., Guller, M. C., Mauriz, J. L., Labalette, C., Buendia, M. A., Mauviel, A., and Bernuau, D. (2007). Physical and functional cooperation between AP-1 and beta-catenin for the regulation of TCF-dependent genes. Oncogene *26*, 3492-3502.

Towers, M., Wolpert, L., and Tickle, C. (2012). Gradients of signalling in the developing limb. Current opinion in cell biology *24*, 181-187.

Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. Genes & development *5*, 880-894.

Trompouki, E., Bowman, T. V., Lawton, L. N., Fan, Z. P., Wu, D. C., DiBiase, A., Martin, C. S., Cech, J. N., Sessa, A. K., Leblanc, J. L.*, et al.* (2011). Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell *147*, 577-589.

Tsai, W. H., Holbrook, G. L., Schal, C., and Chiang, A. S. (1995). In vitro growth of corpora allata from Diploptera punctata. In vitro cellular & developmental biology Animal *31*, 542-546.

Tuupanen, S., Turunen, M., Lehtonen, R., Hallikas, O., Vanharanta, S., Kivioja, T., Bjorklund, M., Wei, G., Yan, J., Niittymaki, I.*, et al.* (2009). The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat Genet *41*, 885-890.

Ueno, S., Weidinger, G., Osugi, T., Kohn, A. D., Golob, J. L., Pabon, L., Reinecke, H., Moon, R. T., and Murry, C. E. (2007). Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *104*, 9685-9690.

Umesono, K., Murakami, K. K., Thompson, C. C., and Evans, R. M. (1991). Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. Cell *65*, 1255-1266.

Valenta, T., Hausmann, G., and Basler, K. (2012). The many faces and functions of beta-catenin. The EMBO journal *31*, 2714-2736.

Valentine, S. A., Chen, G., Shandala, T., Fernandez, J., Mische, S., Saint, R., and Courey, A. J. (1998). Dorsal-mediated repression requires the formation of a multiprotein repression complex at the ventral silencer. Molecular and cellular biology *18*, 6584-6594.

van 't Veer, L. J., Lutz, P. M., Isselbacher, K. J., and Bernards, R. (1992). Structure and expression of major histocompatibility complex-binding protein 2, a 275-kDa zinc finger protein that binds to an enhancer of major histocompatibility complex class I genes. Proceedings of the National Academy of Sciences of the United States of America *89*, 8971-8975.

van Amerongen, R. (2012). Alternative Wnt pathways and receptors. Cold Spring Harbor perspectives in biology *4*.

van Beest, M., Dooijes, D., van De Wetering, M., Kjaerulff, S., Bonvin, A., Nielsen, O., and Clevers, H. (2000). Sequence-specific high mobility group box factors recognize 10-12-base pair minor groove motifs. The Journal of biological chemistry *275*, 27266-27273.

van de Werken, H. J., de Vree, P. J., Splinter, E., Holwerda, S. J., Klous, P., de Wit, E., and de Laat, W. (2012). 4C technology: protocols and data analysis. Methods in enzymology *513*, 89-112.

van de Wetering, M., Cavallo, R., Dooijes, D., van Beest, M., van Es, J., Loureiro, J., Ypma, A., Hursh, D., Jones, T., Bejsovec, A.*, et al.* (1997). Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. Cell *88*, 789-799.

van de Wetering, M., Oosterwegel, M., Dooijes, D., and Clevers, H. (1991). Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. The EMBO journal *10*, 123-132.

van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A. P.*, et al.* (2002). The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell *111*, 241-250.

Van der Flier, L. G., Sabates-Bellver, J., Oving, I., Haegebarth, A., De Palo, M., Anti, M., Van Gijn, M. E., Suijkerbuijk, S., Van de Wetering, M., Marra, G., and Clevers, H. (2007). The Intestinal Wnt/TCF Signature. Gastroenterology *132*, 628-632.

van der Flier, L. G., van Gijn, M. E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D. E., Begthel, H., van den Born, M., Guryev, V., Oving, I.*, et al.* (2009). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell *136*, 903-912.

van Genderen, C., Okamura, R. M., Farinas, I., Quo, R. G., Parslow, T. G., Bruhn, L., and Grosschedl, R. (1994). Development of several organs that require inductive epithelial-

mesenchymal interactions is impaired in LEF-1-deficient mice. Genes & development *8*, 2691-2703.

Verzi, M. P., Hatzis, P., Sulahian, R., Philips, J., Schuijers, J., Shin, H., Freed, E., Lynch, J. P., Dang, D. T., Brown, M.*, et al.* (2010). TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. Proc Natl Acad Sci U S A *107*, 15157-15162.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F.*, et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Vlad, A., Rohrs, S., Klein-Hitpass, L., and Muller, O. (2008). The first five years of the Wnt targetome. Cell Signal *20*, 795-802.

Voss, J. W., Wilson, L., and Rosenfeld, M. G. (1991). POU-domain proteins Pit-1 and Oct-1 interact to form a heteromeric complex and can cooperate to induce expression of the prolactin promoter. Genes & development *5*, 1309-1320.

Wallmen, B., Schrempp, M., and Hecht, A. (2012). Intrinsic properties of Tcf1 and Tcf4 splice variants determine cell-type-specific Wnt/beta-catenin target gene expression. Nucleic acids research *40*, 9455-9469.

Waltzer, L., and Bienz, M. (1998). Drosophila CBP represses the transcription factor TCF to antagonize Wingless signalling. Nature *395*, 521-525.

Wang, B., Xiao, Z., Ko, H. L., and Ren, E. C. (2010). The p53 response element and transcriptional repression. Cell Cycle *9*, 870-879.

Wang, C. I., Alekseyenko, A. A., LeRoy, G., Elia, A. E., Gorchakov, A. A., Britton, L. M., Elledge, S. J., Kharchenko, P. V., Garcia, B. A., and Kuroda, M. I. (2013). Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. Nature structural & molecular biology *20*, 202-209.

Wang, Y., Schwedes, J. F., Parks, D., Mann, K., and Tegtmeyer, P. (1995). Interaction of p53 with its consensus DNA-binding site. Molecular and cellular biology *15*, 2157-2165.

Watanabe, K., Biesinger, J., Salmans, M. L., Roberts, B. S., Arthur, W. T., Cleary, M., Andersen, B., Xie, X., and Dai, X. (2014). Integrative ChIP-seq/Microarray Analysis Identifies a CTNNB1 Target Signature Enriched in Intestinal Stem Cells and Colon Cancer. PLoS One *9*, e92317.

Waterman, M. L., Fischer, W. H., and Jones, K. A. (1991). A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. Genes & development *5*, 656-669.

Waterman, M. L., and Jones, K. A. (1990). Purification of TCF-1 alpha, a T-cell-specific transcription factor that activates the T-cell receptor C alpha gene enhancer in a context-dependent manner. The New biologist *2*, 621-636.

Watson, L. C., Kuchenbecker, K. M., Schiller, B. J., Gross, J. D., Pufall, M. A., and Yamamoto, K. R. (2013). The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. Nature structural & molecular biology *20*, 876-883.

Weiss, A., Charbonnier, E., Ellertsdottir, E., Tsirigos, A., Wolf, C., Schuh, R., Pyrowolakis, G., and Affolter, M. (2010). A conserved activation element in BMP signaling during Drosophila development. Nature structural & molecular biology *17*, 69-76.

Wend, P., Holland, J. D., Ziebold, U., and Birchmeier, W. (2010). Wnt signaling in stem and cancer stem cells. Seminars in cell & developmental biology *21*, 855-863.

White, M. A., Parker, D. S., Barolo, S., and Cohen, B. A. (2012). A model of spatially restricted transcription in opposing gradients of activators and repressors. Molecular systems biology *8*, 614.

Whyte, J. L., Smith, A. A., and Helms, J. A. (2012). Wnt signaling and injury repair. Cold Spring Harbor perspectives in biology *4*, a008078.

Wierzbicki, A. T., Cocklin, R., Mayampurath, A., Lister, R., Rowley, M. J., Gregory, B. D., Ecker, J. R., Tang, H., and Pikaard, C. S. (2012). Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes & development *26*, 1825-1836.

Willert, K., and Jones, K. A. (2006). Wnt signaling: is the party in the nucleus? Genes & development *20*, 1394-1404.

Wills, A. A., Kidd, A. R., 3rd, Lepilina, A., and Poss, K. D. (2008). Fgfs control homeostatic regeneration in adult zebrafish fins. Development *135*, 3063-3070.

Wisniewska, M. B., Nagalski, A., Dabrowski, M., Misztal, K., and Kuznicki, J. (2012). Novel beta-catenin target genes identified in thalamic neurons encode modulators of neuronal excitability. BMC Genomics *13*, 635.

Wodarz, A., Hinz, U., Engelbert, M., and Knust, E. (1995). Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of Drosophila. Cell *82*, 67-76.

Wright, J. B., Brown, S. J., and Cole, M. D. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. Molecular and cellular biology *30*, 1411-1420.

Wu, J. Q., Seay, M., Schulz, V. P., Hariharan, M., Tuck, D., Lian, J., Du, J., Shi, M., Ye, Z., Gerstein, M.*, et al.* (2012). Tcf7 is an important regulator of the switch of self-renewal and differentiation in a multipotential hematopoietic cell line. PLoS genetics *8*, e1002565.

Wu, X., Golden, K., and Bodmer, R. (1995). Heart development in Drosophila requires the segment polarity gene wingless. Developmental biology *169*, 619-628.

Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N., and McMahon, A. P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. Genes & development *13*, 3185-3190.

Yang, X., van Beest, M., Clevers, H., Jones, T., Hursh, D. A., and Mortin, M. A. (2000). decapentaplegic is a direct target of dTcf repression in the Drosophila visceral mesoderm. Development *127*, 3695-3702.

Yasumoto, K., Takeda, K., Saito, H., Watanabe, K., Takahashi, K., and Shibahara, S. (2002). Microphthalmia-associated transcription factor interacts with LEF-1, a mediator of Wnt signaling. The EMBO journal *21*, 2703-2714.

Yee, K. S., and Yu, V. C. (1998). Isolation and characterization of a novel member of the neural zinc finger factor/myelin transcription factor family with transcriptional repression activity. The Journal of biological chemistry *273*, 5366-5374.

Yi, F., Brubaker, P. L., and Jin, T. (2005). TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta. The Journal of biological chemistry *280*, 1457-1464.

Yochum, G. S., Cleland, R., and Goodman, R. H. (2008). A genome-wide screen for beta-catenin binding sites identifies a downstream enhancer element that controls c-Myc gene expression. Molecular and cellular biology *28*, 7368-7379.

Yokoyama, H., Ogino, H., Stoick-Cooper, C. L., Grainger, R. M., and Moon, R. T. (2007). Wnt/beta-catenin signaling has an essential role in the initiation of limb regeneration. Developmental biology *306*, 170-178.

Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J. N.*, et al.* (2009). Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. Developmental cell *17*, 516-526.

Zamora, M., Manner, J., and Ruiz-Lozano, P. (2007). Epicardium-derived progenitor cells require beta-catenin for coronary artery formation. Proceedings of the National Academy of Sciences of the United States of America *104*, 18109-18114.

Zanke, B. W., Greenwood, C. M., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S. M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E.*, et al.* (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet *39*, 989-994.

Zecca, M., Basler, K., and Struhl, G. (1996). Direct and long-range action of a wingless morphogen gradient. Cell *87*, 833-844.

Zhang, C., Blauwkamp, T., Burby, P., and Cadigan, K. (2014). Wnt-mediated repression via bipartite DNA recognition by TCF in the Drosophila hematopoietic system. . PloS Genetics, In revision.

Zheng, Q., Rowley, M. J., Bohmdorfer, G., Sandhu, D., Gregory, B. D., and Wierzbicki, A. T. (2012). RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes. The Plant journal : for cell and molecular biology.

Zhou, B., Bagri, A., and Beckendorf, S. K. (2001). Salivary gland determination in Drosophila: a salivary-specific, fork head enhancer integrates spatial pattern and allows fork head autoregulation. Developmental biology *237*, 54-67.