# Structural and Biochemical Mechanisms of Notch Signal Activation and Inhibition

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Jarrett, Sanchez Mario. 2019. Structural and Biochemical Mechanisms of Notch Signal Activation and Inhibition. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013090">http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013090</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Structural and Biochemical Mechanisms of Notch Signal Activation and Inhibition

A dissertation presented

by

Sanchez Mario Jarrett

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biomedical and Biological Sciences

Harvard University

Cambridge, Massachusetts

August 2019
Abstract

Cell signaling encompasses the fundamental processes which enable an organism to perceive and respond to its environment. Productive cell signaling occurs in a series of tightly regulated steps that require stimulus detection, signal transmission, and a downstream response.

The Notch signaling pathway is an essential cell-cell communication system in metazoans that has critical roles in cell fate determination, growth and development. In mammals, there are two families of Notch-activating ligands, Jagged and Delta-like. The Delta-like paralogues, DLL1 and DLL4, have high structural similarity but exhibit context-dependent functional differences. Notch pathway activation induces transcription of Notch-responsive genes including the Notch regulated ankyrin repeat protein (NRARP) which acts as a feedback inhibitor of the Notch response and is among the most commonly Notch-induced genes in many cell types. In this thesis, I describe the molecular basis for functional divergence of the highly similar ligands DLL1 and DLL4 and a molecular mechanism for NRARP-mediated feedback inhibition of Notch signaling.

First, I analyze the functional divergence of DLL1 and DLL4 using cellular assays and biochemical studies. DLL1 and DLL4 activate NOTCH1 and NOTCH2 differently in cell-based assays and this discriminating potential lies in the region between the N-terminus and EGF repeat three. Our results reveal that DLL4 preferentially activates NOTCH1 over NOTCH2, whereas DLL1 exhibits no preference for either receptor. These studies establish that ligand
ectodomains dictate selective function and features outside the receptor-binding interface contribute to functional differences.

Next, I show that NRARP exerts its inhibitory effect by binding directly to the Notch transcriptional activation complex (NTC), requiring both the RBPJ transcription factor and Notch intracellular domain, but not Mastermind-like proteins or DNA. The X-ray structure of a NRARP/RBPJ/NOTCH1/DNA complex reveals that assembly of NRARP/RBPJ/NOTCH1 complexes relies on simultaneous engagement of RBPJ and NOTCH1 in a non-canonical binding mode involving the extension of the NOTCH1 ankyrin repeat stack by the ankyrin repeats of NRARP. Interface mutations of NRARP disrupt entry into NTCs and abrogate feedback inhibition of Notch signaling. These studies establish the structural basis for NTC engagement by NRARP and provide insights into a critical negative feedback mechanism that regulates Notch signaling.
Table of Contents

Abstract...........................................................................................................................................iii

Table of Contents..............................................................................................................................v

List of Figures.....................................................................................................................................vi

List of Tables.....................................................................................................................................vii

Acknowledgments............................................................................................................................viii

Chapter 1 – Introduction....................................................................................................................1

Chapter 2 – Refining Notch – Delta like ligands’ interaction and ligand in vivo functions........30

Chapter 3 – Extension of the NICD Ankyrin Repeat Stack by NRARP Promotes Feedback
Inhibition of Notch Signaling.............................................................................................................58

Chapter 4 – Conclusions and Future Directions.........................................................................102

Appendix 1 – Ligand ectodomains determine function in vivo and selectivity of DLL1 and
DLL4 toward NOTCH1 and NOTCH2 in vitro.................................................................................113

Appendix 2 – Supplemental figures for Chapter 3.........................................................................141

Appendix 3 – Bispecific Forkhead Transcription Factor FoxN3 Recognizes Two Distinct
Motifs with Different DNA Shapes.................................................................................................145
List of Figures

Figure 1.1 Notch mutant female.................................................................5
Figure 1.2 Notch signaling pathway..........................................................9
Figure 2.1. Contributions of DLL4 contact residues to NOTCH1 activation...........37
Figure 2.2. DLL4 binding to NOTCH1 on the cell surface..................................39
Figure 2.3. DLL1 and DLL4 differentially activate NOTCH1 and NOTCH2..............42
Figure 2.4. Schematic representation of DLL1 and DLL4 and variant proteins.............44
Figure 2.5. Contributions of the MNNL-EGF3 portion and contact amino acids to ligand
selectivity towards N1 and N2........................................................................47
Figure 3.1. NRARP is a negative feedback regulator of Notch signaling......................66
Figure 3.2. Requirements for complexation of NRARP with NOTCH1 and RBPJ...........70
Figure 3.3. Structure of an NRARP-NOTCH1-RBPJ complex on DNA, and comparison with the
human Notch1-RBPJ-MAML1.........................................................................74
Figure 3.4. Interface between NRARP and NICD1-RBPJ........................................76
Figure 3.5. A. Effect of NRARP interface mutations on Notch-dependent luciferase reporter
gene activity....................................................................................................79
Figure 3.6. Effect of NRARP on the degradation of activated NOTCH1 complexes.........82
List of Tables

Table 1. Tandem immunoprecipitation and mass spectrometry results from Jurkat cells using tagged NRARP as bait. .................................................................100

Table 2. Data collection, structure determination and refinement statistics........................101
Acknowledgements

Pursuing a PhD has been an arduous but satisfying endeavor and for those who have helped me along this journey you have my undying gratitude.

To my advisor Steve, I would like to express my utmost appreciation for your support over the years. Steve’s passion for science has always been inspiring and his guidance has shaped the way I think as a scientist. His leadership and unwavering efforts have created a haven for scientific pursuits in the form of the Blacklow laboratory. Thank you for your mentorship and the freedom and opportunity to explore.

To my lab members, the Blacklow lab has been an exceptional place to develop as a scientist and I am grateful to every member, past and present, for thoughtful discussions, experimental advice and for always being available for a laugh or a drink to celebrate our victories and to commiserate lessons learned. Thanks in particular to Tom and Brandon, for all your advice and teachings whether directly or by example – I am forever grateful.

Many thanks to Jon Aster, Joe Loparo and Mike Eck for serving on my Dissertation Advisory Committee over the years. Your advice throughout my dissertation work has truly been invaluable and I thank you for guiding me to good ideas and away from bad ones. And thanks to my Examination Committee for reading this dissertation.

To all my friends and BBS classmates, you have taught me so much and helped me grow as a scientist as well as a person. Thank you for enriching my life, I hope I have enriched yours.
Finally, I would like to thank my family for the love and support you have always given me.

Thanks to my parents who have always supported me in my pursuits. I will cherish the lessons learned from observing your persistence and sacrifice no matter where life takes me.

To my late grandmother, I am not a fraction of the person I am today without you. Thank you for your love and support when I needed it most.
CHAPTER 1

Introduction
1.1 Introduction to Intercellular Signaling Pathways

The ability of cells to perceive and respond to their environment is fundamental to the development of multicellular organisms. Cells are constantly exposed to signals that may arrive in the form of soluble factors generated locally or distantly, ligands expressed on other cells, or components of the extracellular matrix. The capacity to perceive these signals and coordinate a physiological response facilitates the formation and maintenance of the specialized tissues of multicellular organisms.

Specialized tissues that arise during embryogenesis are characterized by the careful regulation of cellular behaviors, so that cells proliferate, migrate, differentiate and form tissues in the proper place and at the proper time. These processes are genetically programmed and depend on cell history, lineage, and the action of the signaling pathways which coordinate the cellular interactions that lead to organogenesis.

A limited number of signaling pathways regulate many of the key events in normal development. Among the most important systems for intercellular communication are the Notch, Wnt, Hedgehog, FGF, and TGF β signaling systems. These pathways operate repeatedly in various contexts and elicit diverse signaling outcomes. Not only do they control cell fate decisions in development, they also function throughout the lifetime of an organism. These core developmental signaling pathways are among the most thoroughly investigated systems in biology, as they are evolutionarily conserved across species. At the molecular level the ligands, receptors, modulators, and effectors are largely preserved, as are the molecular interactions among these components.

Our knowledge of how cells integrate the information they receive through these pathways and produce diverse responses in a cell-specific manner while using the same signaling
molecule, however, is remarkably incomplete. That is, an “operational” understanding of these pathways, including how cellular context affects the information transmitted, remains elusive. Signal transduction in these conserved development pathways have varying signaling architectures – i.e. these pathways achieve signal propagation through various mechanisms which may include intermediates, post-translational modifications, signal amplification, relief of inhibition, etc. The presence of various signal transduction architectures provides an additional mode for transmitting information outside of the presence/absence of the initiating signal. A growing number of studies suggest that an additional layer of information is conveyed in the dynamics of signaling molecules which is determined by the molecular nature of the signaling architecture.

This work focuses on the Notch signaling pathway at points of signal initiation and termination. The relative simplicity of Notch signaling, in which a single ligand-binding event produces one nuclear effector protein, raises the question of how Notch signals are used iteratively and in different cell types to drive developmental fate decisions. This receptor property, wherein the processed receptor is itself the signaling effector molecule, combined with capacity of different ligands to produce distinct responses in signal-receiving cells, suggests complexity in the information flow represented by the amount or dynamics of the Notch intracellular domain (NICD). This thesis focuses on molecular factors that influence NICD dynamics and abundance in signal-receiving cells: first, on receptor activation by different ligands, and second, on the modulation of effector activity in the nucleus.

1.2 Notch Signaling Overview
The Notch signaling pathway is named in homage to the first observed phenotype for mutations in the receptor. The mutant phenotype was reported by Morgan and Bridges in 1916 following the first observation by John Dexter in 1913, who noticed that fruit fly *Drosophila melanogaster* haploinsufficient in a yet unidentified gene showed a phenotype of notches at the wing margins (1) (Figure 1.1). The *NOTCH* gene, encoding a transmembrane receptor, was identified several decades later (2-4) and since then a plethora of studies have shown that Notch signaling is a conserved signal transduction system that plays an essential and fundamental role in numerous developmental processes (5-7).
Figure 1.1. Notch mutant female. Credit: Thomas Hunt Morgan, Geneticist/Naturalist. 1919
Notch signaling defines one of the earliest cellular pathways controlling cellular processes necessary for metazoan development. The pathway is repeatedly used throughout development as well as postnatally in tissue maintenance/homeostasis. Evolutionary divergence of invertebrates and vertebrates has resulted in gene duplication events in both receptors and ligands. Flies possess two Notch ligands (Serrate and Delta), worms three (APX-1, DSL-1 and LAG-2), and mammals five (Jagged 1 & 2 and Delta-like 1, 3 & 4). Notch receptors have similarly been subjected to at least two rounds of gene duplication: flies possess a single Notch gene, worms two (GLP-1 and LIN-12), and mammals four (NOTCH1-4).

Not surprisingly, dysregulation of Notch signaling, either through mutations of receptors or ligands, or by changes in pathway modifiers, is linked to a number of developmental abnormalities and diseases, including human cancer. The mammalian Notch receptors and ligands exhibit varying expression patterns and accordingly different developmental relationships exist between specific receptors and/or ligands. NOTCH1 and NOTCH2 are broadly expressed during development and in adult organisms and accordingly, genetic knockout of Notch1 or Notch2 is embryonic lethal in mice, and conditional knockouts have revealed developmental defects in many organ systems. In contrast, NOTCH3 and NOTCH4 are more restricted in expression (primarily to vascular smooth muscle and endothelium, respectively), with Notch3 or Notch4 knockout mice viable with subtle phenotypes observed primarily in blood vessels (8-10).

On the ligand side, mutations and their associated developmental syndromes have also revealed tissue specific ligand functions and have suggested the existence of preferred ligand-receptor pairs. For example, Alagille syndrome, which primarily affects the liver, biliary tree and heart, can be caused by loss-of-function mutations in either JAGGED1 or NOTCH2 (11, 12);
Spondylocostal dysostosis and spondylothoracic dysostosis associated with rib fusions and deletions can be caused by loss-of-function mutations in DLL3 (reviewed in (13). Likewise, adult onset diseases such as CADASIL which affects arteries in the CNS, is caused by missense mutations in NOTCH3 (14). In addition, aberrant or dysregulated Notch signaling is associated with several different human cancers, most notably T cell acute lymphocytic leukemia (T-ALL), in which activating mutations of the NOTCH1 receptor are found in more than half of all cases (reviewed in (15)).

1.3 Molecular Mechanism of Notch Pathway Activation

Signals transmitted through the Notch signaling pathway depend on cell-cell contact between a ligand-expressing “sender” cell and a receptor-expressing “receiver” cell. The core events necessary for productive Notch signaling involve: 1) ligand binding 2) receptor cleavage, 3) NICD translocation and transcription activation (Figure 1.3 Top). There is also a strict requirement for endocytosis of ligands into the signal-sending cells.

Because both ligands and receptors are transmembrane proteins, initiation of the signal occurs at sites of cell-cell contact, where ligands bind to Notch. Receptor activation appears to require the application of a mechanical pulling force by the bound ligand, leading to relief of receptor autoinhibition and susceptibility to proteolysis (16, 17). The force required for this regulatory step depends on the endocytosis of the ligand in the signal sending cell.

After force-induced conformational changes overcome autoinhibition of the Notch ectodomain, the receptor undergoes a series of proteolytic cleavage events in a process referred to as regulated intramembrane proteolysis. The first cleavage event is carried out by a disintegrin and metalloprotease 10 (ADAM10) at a membrane-proximal extracellular site called S2 (18, 19).
ADAM10 cleavage renders the resulting Notch molecule a substrate for gamma secretase, which cleaves at an intramembrane site called S3 to liberate the intracellular region of Notch receptor (NICD) from the membrane. The intracellular portion of Notch then migrates to the nucleus to form a Notch transcription complex (NTC) (20-23).

The NTC is assembled in a stepwise fashion and is composed of NICD, the transcription factor RBPJ, and a coactivator protein of the Mastermind-like (MAML) family. Initially, NICD binds to RBPJ through a high affinity interaction of the “RBPJ-associated molecule” (RAM) and an Ankyrin repeat (ANK) domain (24). The MAML protein then recognizes and binds a composite Notch-RBPJ surface, which includes the ANK domain of NICD. The binding of MAML stabilizes the complex and induces the transcription of Notch target genes via a largely undefined mechanism (reviewed in (25)).
**Figure 1.2.** Notch signaling pathway

Model for the main events on the Notch signaling pathway as well as the mammalian Receptor – Ligand repertoire. *Top:* Engagement of Notch receptor by ligands induces receptor cleavage by ADAM metalloprotease and γ-secretase allowing the bioactive intracellular domain of the receptor to translocate to the nucleus and form the transcription activation complex. *Bottom:* Domain organization of human Notch receptor and ligand homologs. Adapted from (26).
1.4 Receptor – Ligand Interactions

The Notch signaling repertoire in mammals consists of four receptors and five canonical ligands. The four mammalian homologs of Notch receptors are large single-pass type I transmembrane glycoproteins. The extracellular domain of Notch receptors is largely composed of tandem epidermal growth factor (EGF)-like repeats ranging from 29-36 repeats. Many of the repeats bind calcium, which helps to introduce regions of localized rigidity in these receptors. The bound calcium ions are important both for structural integrity of the receptor as well as for ligand-binding and productive signaling (27, 28). Following the tandem EGF-like repeats is a juxtamembrane region termed the Negative Regulatory Region (NRR), which contains three Lin12-Notch repeats (LNR) and a heterodimerization domain. This module of the receptor is the key regulatory element protecting Notch from premature metalloprotease cleavage and aberrant signal initiation (29). Following its transmembrane segment, the Notch receptor has an intracellular region which begins with the RAM regions, a high affinity RBPJ-binding module of ~20 amino acids. A long unstructured linker region leads to an ankyrin repeat (ANK) domain containing seven ankyrin repeats, which is flanked by two nuclear localization signals (NLS). The C-terminus of the receptor contains a Proline/Glutamic-acid/Serine/Threonine-rich (PEST) region which is important for regulating protein degradation. Between the PEST and ANK domains lies a loosely defined and poorly conserved transactivation (TAD) domain (Figure 1.2: *Bottom Left*).

There are two families of the canonical Delta, Serrate, Lag2 (DSL) ligands, which have five total members in mammals (Delta-like 1, 3 and 4; Jagged 1 and 2). All canonical ligands are themselves type I transmembrane proteins and are largely comprised of three functional domains, namely, an MNNL domain (module at the N-terminus of Notch ligands), a DSL
domain and a series of 6-16 EGF-like repeats. Ligand families are classified based on the presence or absence of an additional cysteine rich (CR) domain next to the plasma membrane (Jagged or Delta-like, respectively) (Figure 1.2: Bottom Right). All ligands also contain a transmembrane segment and an unstructured C-terminal tail that serves as the recognition motif for ubiquitin-ligases to stimulate endocytosis of ligand (30).

Formation of Notch-ligand complexes is mediated through a region comprised of EGF repeats 8-12 of the receptors and the MNNL domain through EGF3 of the ligands (31). Structures of rat Notch1 in complex with a representative ligand from each ligand family have been solved, using yeast display to evolve high-affinity ligand variants. The Notch1-Dll4 complex was readily assembled with receptor and ligand fragments that contact each other using only EGF repeats 10-12 of Notch1, and short fragments of Dll4, with contacts limited to EGF11-12 of Notch1 and the MNNL and DSL domains of Dll4 as a result (32). The Notch1-Jagged1 (Jag1) complex was reconstituted with an extended Notch1 fragment containing EGF repeats 8-12 and a Jag1 fragment from MNNL-EGF3, and all domains were seen to participate in binding (33). The x-ray structures of both complexes revealed an antiparallel arrangement of receptor and ligand, confirming that both families of DSL ligands engage Notch1 in a similar fashion, using the identical intermolecular domain register and analogous contact surfaces within the overlapping region.

A distinctive feature of Notch signaling is its reliance on mechanical force and protein glycosylation for pathway activation. The structure of both Jag1 and Dll4-Notch1 complexes as well as the accompanying biochemistry gives new insights into how these two factors promote receptor activation. Biomembrane force-probe force-clamp spectroscopy studies (33) were used to suggest that both Jag1 and Dll4 ligand interactions with Notch1 have “catch bond” properties,
in which the lifetime of a bond/interaction is prolonged by the application of tensile force. Interestingly, Jag1-mediated activation appears to be more reliant on this property of the receptor-ligand interaction, as it appears to require a larger force for NOTCH1 activation. The basis for this catch bond property may lie in the conformational change ligands experience under tension. Comparison of the apo- and receptor-bound states of the Jag1 ligand revealed hinge-like movements in all the domains with a dramatic 32-degree pivot at the MNNL - DSL junction.

Another fascinating quality of Notch-ligand complexes partly explained by the structural data is the susceptibility to modulation of binding interactions by sugar modifications on Notch receptors. There are glycosylation sites along the Notch receptor extracellular domain, some of which occur within the ligand binding region. These modifications have profound effects on Notch activity, with O-fucosylation and O-glucosylation essential for optimal Notch signaling (34, 35). In both receptor-ligand structures, the ligands directly contact sugar residues present on Notch1. The O-fucose modification present on Thr466 of Notch1 EGF12 serves as the anchor for interactions with the MNNL domain of both ligands, which explains the deleterious effect of mutating Thr466 in Notch1 (36). Additional sugar modifications on Thr311 and Ser435 make direct contacts with Jag1 and Dll4 respectively. It is important to note that both sugar modifications on Thr311 and 466 are O-fucose and extension by Fringe proteins would be likely to change the binding interface, thus providing a rationale for how glycan modifications alter ligand affinity. The role of receptor glycosylation in signaling outcomes is further explored in later sections.

1.5 Signaling Context Dictates Response
The Notch signaling pathway is relatively simple mechanistically. Ligand engagement triggers receptor cleavage, which in turn liberates the intracellular domain for entry into the nucleus to initiate Notch dependent genetic programs. In canonical Notch signaling there are no well-established intermediates following Notch receptor activation and assembly of the transcription activation complex; that is, there are no identified attenuators or amplifiers of signaling prior to Notch target gene induction. Despite this apparent simplicity, Notch has vastly different functions in various developmental settings and in multiple disease contexts. This complexity of response is indicative of a much larger and sophisticated signaling network capable of many diverse biological outcomes.

1.5.1 Cell surface context

Given the lack of intermediates and the apparent straightforward nature of Notch signaling, how does the induction of a Notch response achieve such a wide array of biological outcomes? First, there are multiple receptors and activating ligands in the mammalian Notch repertoire, and by the spatial and temporal regulation of expression of these molecules certainly plays an important role. Additionally, there is regulation of the expression of Notch signaling components by other pathways thus illuminating how Notch may feature into larger signaling pathway networks. One example in which spatio-temporal regulation of Notch signals is prominent is during somitogenesis in the presomitic mesoderm. The Notch1-Dll1 ligand-receptor pair helps to drive this process, in which somites - blocks of mesoderm that give rise to axial muscles, bones and dermis in vertebrates - are formed. The Notch response is reinforced by Notch-dependent induction of the expression of NOTCH1 itself, whereas the expression of the ligand Dll1 is regulated by the Wnt signaling pathway. The Wnt-dependence of ligand
expression guarantees that the activity of Notch pathway is now linked with another core signaling component involved in the segmentation clock (37). This exchange between signaling pathways partially regulates the oscillatory gene expression program necessary for the periodic formation of somites.

In addition to the pathway interactions presented above, diverse biological outcomes can be achieved by the interplay of Notch ligands functioning in the same setting in conjunction with modifying enzymes such as Fringe glycosyltransferases. For example, it has been shown that Notch signaling has a critical function in vascular patterning (8, 38). During blood vessel formation the process of endothelial sprouting must be balanced for proper angiogenesis. Notch signaling acts in vessel formation to limit the development of endothelial tip cells which are the leading endothelial cells that sprout towards the angiogenic stimulus. Productive Notch signaling in this setting is mediated by Dll4 interaction with Notch1 in trailing cells (39). Endothelial cells, however, also express the Jag1 ligand. Interestingly, the loss of Jag1 decreases sprouting behavior and angiogenesis due to an increase in Notch signaling, indicating an antagonist function for Jag1 with respect to Dll4 in this setting. This interplay is facilitated through Fringe enzymes that modify Notch1 to potently increase Dll4 signaling capabilities while, conversely, decreasing Jag1-mediated signaling (40). Expression of these two ligands is differentially regulated with levels fluctuating between tip and stalk endothelial cells such that the equilibrium between these two ligands with conflicting functional roles in part regulates angiogenesis with the aid of Fringe enzymes (41).

The above detail on the interplay of ligands and the influence of modifying enzymes describes a scenario where competing ligands interacting with receptors might regulate biological outcome by differences in trans-signaling capacity. However, the Notch pathway also
incorporates a non-productive parallel interaction, in which ligands and receptors present in the same cell (in cis) have an inhibitory interaction, a phenomenon referred to as cis-inhibition (42, 43). Thus, ligands and receptors expressed in cis introduce an additional complexity. Where both receptors and ligands are expressed in the same cell, relative levels will not only determine whether there is productive signaling or not but by having both classes of molecules in the same cell, directionality of Notch signaling can be assigned and now be dynamically regulated. Modeling studies (44) suggest that cis-inhibition has a sharp ligand threshold while trans-activation shows a graded response to ligand levels. The mutual inactivation of ligand and receptor in cis-inhibition generates an ultrasensitive switch between two mutually exclusive signaling states, sending (high Delta/low Notch) and receiving (high Notch/low Delta). At the multicellular level the presence of this signaling switch can amplify small differences to facilitate the formation of sharp boundaries in cell type and in tissues. For example, the angiogenesis model described above can be further refined by cis-inhibitory interactions. Cis-inhibition likely helps to reinforce the distinction of endothelial tip and stalk cell fates ensuring that both maintain a signal sending and receiving state, respectively.

Interestingly, the mammalian Delta like ligand family member Dll3, does not seem to function in trans-activation and only exhibits cis-inhibitory activity (45). As noted, Notch signaling is a key regulator of somitogenesis, the process by which vertebrae are formed. Notch pathway genes have been associated with the congenital abnormalities spondylocostal dysostosis (SCD) and spondylothoracic dysostosis (STD), which are disorders characterized by congenital malformations in which the vertebrae are fused or altered in shape, position, or size. Mutations in Dll3 have been associated with SCD and STD, highlighting the importance of the cis-inhibitory interaction mode as a very important aspect of Notch signaling in development.
So far, we have looked at the interplay of ligands based on their expression in relation to each other as well as to receptors (cis or trans), the influence modifying enzymes can have on their interaction, and aspects of signaling that are mechanistically tied to spatiotemporal regulation of receptors, ligands and modifiers. But Notch signaling is far from an “all or nothing” pathway.

1.5.2 Nuclear context

Notch receptors are bioactive and function as transcription co-factors, in that they regulate transcription without binding DNA directly. All known Notch transcription responses are mediated through association with a single DNA binding partner, RBPJ (46, 47), and the identification of genes that are directly induced by Notch is an active area of research. RBPJ and RBPJ complexes have relatively modest DNA affinity (48, 49) and RBPJ is capable of binding both transcription activating (NICD) and repressing cofactors such as KYOT2 (50) and SHARP (51, 52).

RBPJ is thought to be incapable of binding target motifs when they are inaccessible such as when wrapped around nucleosomes. Thus, it is thought not be able to initiate changes in regulation at silent enhancers as a pioneer factor. This feature of the response mechanism renders Notch susceptible to the influence of extrinsic circumstances such as cell lineage and history, which influence the chromatin state and define the cohort of additional transcription factors that are present. Studies have shown that RBPJ/NICD complexes can act in cooperation with other transcription factors in both a predictable and defined manner (53-55). Binding motifs with specific configurations can also promote direct or indirect interactions among transcription
factors. The arrangement of response elements creates recognizable enhancer signatures, but these signatures do not need to be present at all Notch response elements.

One example of a well characterized enhancer signature is the SPS+A combinatorial code that functions in neural precursor specification in Drosophila. This signature combines a pair of specifically oriented (head-to-head) RBPJ motifs known as Su(H)-paired-sites (SPS) with a motif for pro-neural bHLH activator proteins (A). SPS+A sites cooperate to allow functional interaction of RBPJ/NICD complexes with bHLH activators to yield strong activation of gene expression, implying a Notch-bHLH synergy in the response (53, 56).

Cooperation of Notch with other transcription factors with no recognized distinctive signature is also evident. There is an association between RBPJ sites and RUNT-related transcription factors (RUNX) in several different cellular contexts, even though there is no discernable organization or precise arrangement relating the RBPJ and RUNX binding sites (57-59). In these cell types, binding of RUNX is required for the enhancers to be competent for Notch engagement, though the basis for this functional interaction remains poorly understood (57, 60).

In vivo, DNA binding by RBPJ is highly dynamic (61). RBPJ has a relatively modest DNA affinity, with $K_d$ for consensus binding sites of approximately 60 nM (49), and affinities for variant sites of as low as 1 µM. The modest affinity manifests itself as a low dwell time on canonical enhancers where the dwell time has been measured. Interestingly, NICD binding to RBPJ increased the observed abundance of RBPJ at enhancer binding sites, suggesting that complexation enhances RBPJ engagement with DNA in living cells (60, 62). In the specific case of paired sites (SPS), there are two specifically oriented (head-to-head) RBPJ motifs. These SPS can contribute to regulation of transcription by cooperative dimerization of NTC to tune
transcriptional responses. In such a cooperative scenario, SPS-regulated genes might be early or preferential responders to low doses of Notch signals. In agreement with this idea, previous studies have shown that E(spl) genes associated with promoter-localized paired sites have a very rapid response to Notch pathway activation (63). In addition, it has also been shown that SPS in many cases associate with early responding genes (64). Moreover, RBPJ/NICD ChIP-seq signals were stronger on average for SPS, suggesting that genes responding to paired-site binding may have larger and/or longer durations of response to Notch activation. However, the SPS element cannot completely account for the kinetics and size of Notch responses as the architecture of Notch response elements at many genes are complicated by a diversity in the number and arrangement of monomer binding sites, SPS elements, as well as their affinity and the presence of motifs for binding of other transcription factors. It is noteworthy that the SPS regulatory element is completely absent from worms, suggesting that SPS elements may have arisen to permit more precise and diverse responses to Notch signals in more complex metazoans.

Finally, RBPJ has a Notch-independent activity in transcription repression. It can bind to corepressors such as Kdm5a, KYOT2 and SHARP as well as the negative regulator RITA. Whereas the role of RBPJ in Notch mediated transcription activation is well established and characterized, its role as a repressor is less defined. What is clear from recent studies is that the different co-repressors recruit repression complexes that differ in their ability or degree to which they modify chromatin. For example, SHARP can mediate recruitment of two different repression complexes to regulate the transcriptional response at genes that are at other times stimulated by Notch (65). Based on the composition of the RBPJ-repressor complex (which is dependent on what co-repressors and modifiers are present), then, enhancers can either be completely suppressed or exist in a state that is relatively refractory to Notch activity.
1.6 Notch Feedback Modulation by NRARP

Another powerful mode of regulating the Notch signaling pathway is by modulating the activity or stability of the NICD following receptor activation. There is limited knowledge about how NICD activity is regulated after its entry into nuclear complexes. It is clear, however, that NICD is subjected to a number of post-translational modifications that have been linked to altered signaling (for a review see (66)). Interactions with other proteins also could modulate NICD nuclear levels and potentially its activity.

The Notch-regulated ankyrin repeat protein (NRARP) is one of the primary Notch target genes. NRARP is one of a small number of core Notch target genes that are induced in many different cell types. The NRARP gene, originally called 5D9, was first identified as a novel component of the Notch signaling pathway, following analysis of the Delta-Notch synexpression group of developmentally expressed genes in Xenopus embryos using in situ hybridization screening (67). The cDNA was later shown to encode a 114-amino acid protein consisting of at least two tandem C-terminal ankyrin repeats. NRARP was also shown to be regulated at the transcriptional level by the Notch signaling pathway in Xenopus and mice (68).

Previous studies have shown that NRARP is a negative regulator of Notch signaling. The forced expression of NRARP in mice results in a blockade of T-cell development in multiple stages, a process known to require Notch signaling (69). In addition, Nrarp knockout mice show defects in somitogenesis and in vascular trimming in the eye (70, 71). Both phenotypes are associated with excessive Notch activity and, together with studies investigating the influence of Nrar on cell fate decisions in the mouse retina (72), support the conclusion that Nrarp is a negative regulator of Notch signaling. A major focus of the work in Chapter 3 of this thesis
centers on elucidation of the molecular basis for the function of NRARP as an attenuator of Notch signal transduction.

1.7 Goals of the Dissertation

This thesis work has focused on developing a more complete molecular understanding of Notch signaling events that directly influence the signals that cells detect and respond to. Specifically, I look at signaling at the steps of signal production and termination, using a combination of biochemical and biophysical techniques.

In Chapter 2, I present work probing molecular mechanisms underlying the observed differences in signaling activity of Delta like family activating ligands Dll1 and Dll4. This work revealed that Dll4 preferentially activates Notch1 while Dll1 exhibits equal efficiency in activating Notch1 and Notch2. Importantly, we show that the ligand ectodomains dictate their selective functions found in vivo and that features that lie outside of the structurally visualized binding interface may underlie these differences.

In Chapter 3 of this thesis, I describe my work studying the structural mechanisms underlying the observed Notch inhibitory activity of the Notch signaling modulator NRARP. This work aims to refine our understanding of an often-overlooked essential aspect of cell signaling, which is termination of activation and resetting of the system. By utilizing an unbiased proteomics approach to identify NRARP interacting proteins, I found that NRARP binds directly to the core Notch transcription complex. By determining a crystal structure of NRARP in complex with the core Notch transcription activation complex, I defined the molecular basis for
binding of NRARP to the NTC. Additional cell-based and biochemical studies identified a
potential mechanism for the activity of NRARP as a Notch pathway inhibitor.

In Chapter 4, I conclude by discussing the main findings of this work and presenting
interesting potential areas of future study.
References


37. R. A. Bone et al., Spatiotemporal oscillations of Notch1, Dll1 and NICD are coordinated across the mouse PSM. Development 141, 4806-4816 (2014).


40. R. Benedito et al., The Notch Ligands Dll4 and Jagged1 Have Opposing Effects on Angiogenesis. Cell 137, 1124-1135 (2009).

42. J. F. Celis de, S. Bray, Feed-back mechanisms affecting Notch activation at the

43. C. A. Micchelli, E. J. Rulifson, S. S. Blair, The function and regulation of cut expression
on the wing margin of Drosophila: Notch, Wingless and a dominant negative role for

44. D. Sprinzak *et al.*, Cis-interactions between Notch and Delta generate mutually exclusive

45. E. Ladi *et al.*, The divergent DSL ligand Dll3 does not activate Notch signaling but cell
autonomously attenuates signaling induced by other DSL ligands. *The Journal of Cell

46. M. E. Fortini, S. Artavanis-Tsakonas, The suppressor of hairless protein participates in

47. S. Christensen, V. Kodoyianni, M. Bosenberg, L. Friedman, J. Kimble, lag-1, a gene
required for lin-12 and glp-1 signaling in Caenorhabditis elegans, is homologous to


49. R. Torella *et al.*, A combination of computational and experimental approaches identifies
DNA sequence constraints associated with target site binding specificity of the

50. Y. Taniguchi, T. Furukawa, T. Tun, H. Han, T. Honjo, LIM Protein KyoT2 Negatively
Regulates Transcription by Association with the RBP-J DNA-Binding Protein. *Molecular
61. D. Castel et al., Dynamic binding of RBPJ is determined by Notch signaling status. 


CHAPTER 2

Refining Notch – Delta like ligands interaction and ligand *in vivo* functions.

Sections of this chapter are taken from:

Author Contributions:

I designed the experiments, purified proteins for biochemical and biophysical characterization, and performed analysis unless otherwise stated. Figures 2.3, 2.4, and 2.5 are adapted, wholly or in part, from the following study I assisted in authorship.

For the larger, published study (1):

I assisted in molecular design of all structure-based receptor interaction mutants (2) for assaying function both in vivo and in vitro. All experiments carried out with purified proteins were performed by me, including the data analysis as well as production and purification of all molecules. I was also involved in manuscript revision and editing.
Abstract

Notch signaling is an essential cell-cell communication system in multicellular organisms that has critical roles in cell fate determination, growth and development. In mammals, there are two Notch-activating Delta paralogues, DLL1 and DLL4, that have the same modular domain organization but context-dependent functional differences. Here, we analyze their functional divergence using cellular co-culture assays and biochemical studies. DLL1 and DLL4 activate Notch1 and Notch2 differently in cell-based assays and this discriminating potential lies in the region between the N-terminus and EGF repeat three. However, substitution of Notch1-binding interface residues in the MNNL and DSL domains of DLL1 with the analogous residues of DLL4 does not disrupt DLL1 function. Collectively, our results reveal that DLL4 preferentially activates Notch1 over Notch2, whereas DLL1 does not exhibit a strong preference for either Notch1 or Notch2. These studies establish that the ectodomains dictate selective ligand function and that features outside the structurally determined binding interface contribute to the observed activity differences.
2.1 Introduction

Notch proteins serve as the receptors for a signaling pathway that plays an essential role in a wide range of cell fate decisions in animals ranging from sea urchins to humans. This pathway is repeatedly used throughout development as well as in postnatal tissue maintenance and organ homeostasis.

Notch signal transduction relies on regulated intramembrane proteolysis to create a transcriptional effector in response to ligand stimulation. Ligand binding relieves autoinhibition, exposing a juxtamembrane processing site to ADAM metalloproteases. ADAM processing liberates the intracellular portion of Notch (NICD) from the membrane (3-5). Once freed from the membrane, NICD translocates into the nucleus where it forms the Notch transcriptional activation complex (NTC) consisting of NICD, the transcription factor RBPJ, and a coactivator protein of the Mastermind-like (MAML) family (6-8). The NTC then recruits additional transcriptional coregulators resulting in the initiation of transcription of Notch responsive genes (see (9) for a review).

The Notch signaling repertoire in mammals consists of four receptors and five canonical ligands. The four mammalian homologs of Notch receptors are large single-pass type I transmembrane glycoproteins. The extracellular domain of all four receptors contains a series of 29-36 N-terminal epidermal growth factor (EGF)-like repeats that include the ligand binding site. The EGF repeats are followed by three LIN-12/Notch repeat (LNR) modules, which are responsible for maintaining the autoinhibited state by masking the metalloprotease cleavage site.
There are two families of the canonical Delta, Serrate, Lag2 (DSL) ligands, which include five members in mammals (Delta-like 1, 3 and 4; Jagged 1 and 2). All canonical ligands are themselves type I transmembrane proteins and are largely comprised of three functional domains, namely, the MNNL (module at the N-terminus of Notch ligands), the DSL domain and 6-16 EGF-like repeats. Ligand families are classified based on the presence or absence of an additional cysteine rich (CR) domain next to the plasma membrane (Jagged or Delta-like, respectively).

Notch receptor-ligand interactions rely on a core interaction region comprised of EGF repeats 8-12 of the receptors and the segment including the MNNL domain through EGF3 of the ligands. Previous studies identified a requirement that truncated receptors begin with EGF8 in order to signal productively, and that the minimum DLL4 construct that can achieve near maximal signaling efficiency terminates at EGF3(10). An x-ray structure of an affinity matured DLL4-NOTCH1 complex, however, only includes receptor and ligand fragments that contact each other using EGF repeats 11 and 12 of NOTCH1, and the MNNL and DSL domains of DLL4 (2). The structure of the NOTCH1-JAGGED1(Jag1) complex also revealed an antiparallel arrangement of receptor and ligand, confirming that both families of DSL ligands engage NOTCH1 in a similar fashion (11). In this structure, a NOTCH1 fragment from EGF8-12 and a Jag1 fragment from MNNL-EGF3 were used to form complexes, and all domains participate in the binding interface. It is notable that binding studies comparing the DLL4 affinity of Notch1 EGF8-12 with that of Notch1 EGF10-12 showed only a 1.3-fold affinity difference, raising the question of what receptor property, other than affinity, accounts for the requirement that EGF8 and 9 be present in the receptor in order to respond to DLL4. When comparing the binding of the
MNNL-DSL domains of both ligands to their binding surface on Notch1 EGF11-12, Dll4 buries 200Å² more surface area, with Dll4 having twice the number of polar contacts with Notch1 in comparison to Jag1 (11).

Several studies have shown that in certain contexts DLL1 and DLL4 are functionally non-equivalent, but it remains unclear which portions of these similar DSL proteins account for their functional divergence (12-15). Previous work observed that DLL4 and DLL1 elicit Notch1 responses with different dynamics – DLL4 produces a sustained response while DLL1 produces pulsatile response – which were functionally attributed to the intracellular region of the ligands (16). However, it was previously shown that DLL4 has a higher intrinsic binding affinity than DLL1 for Notch1 (10), suggesting that the ectodomains of these ligands should play a role in their functional divergence. Here, we investigate the contributions of the extracellular and intracellular regions of DLL1 and DLL4 chimeric proteins on ligand function in cell-based assays and in biochemical binding assays for select chimeras. We observe that DLL1 and DLL4 can discriminate between Notch1 and Notch2 in vitro and that this discriminatory function is encoded by the ligand ectodomains. Interestingly, we also found that regions outside of the known DLL - Notch binding interface are major contributors to the discriminatory function of DLL1 and DLL4.

2.2 Results

Conservation of the DLL4 contact residues across species

Previous work using rat homologs of Notch1 and DLL4 was able to provide structural information about the Notch1-Dll4 complex via x-ray crystallography. The model revealed an
antiparallel, linear interaction between the receptor and ligand with interactions mediated through both protein-protein and protein-glycan contacts (Figure 2.1A). The interaction zones were localized to two regions, with “site 1” comprised of the MNNL and EGF12 domains and “site 2” the DSL and EGF11 domains of DLL4 and Notch1, respectively. To assess whether the ligand interaction surface revealed in the structure of the affinity matured, rat protein complex is conserved in human DLL4, we used a well-established luciferase reporter-gene assay (17) to test the relative signaling capabilities of DLL4 proteins with mutations at key interfacial residues (Figure 2.1B). When cells transiently transfected with full-length DLL4 mutants were co-cultured with cells stably expressing Notch1 and transiently transfected with a luciferase reporter, receptor cells showed defective signaling based on luciferase activity at approximately half of the key residues probed.
Figure 2.1. Contributions of DLL4 contact residues to NOTCH1 activation. A. Molecular surface representation highlighting interactions between DLL4 MNNL-DSL domains and EGF 11-12 of NOTCH1. B. Sequence conservation of human DLL1 and DLL4 as well as rat DLL4 sequence used for structural studies. Contact amino acids (boxed) are mutated in alanine scanning mutagenesis experiments. C. Effect of DLL4 interface mutations on Notch-dependent luciferase reporter gene activity.
The most detrimental mutations were in the DSL domain of DLL4, with a third of the mutants exhibiting no detectable signaling activity (Figure 1D, E). These results suggest that “site 2” of the Notch1-DLL4 interaction has a large contribution to signaling in vivo. Mutants of “site 1” residues - present in the MNNL domain – did not generally interfere with signaling, with the exception of F65A, which was partially defective in signaling. Interestingly, previous work (18) showed that further modification of the O-fucose modification of T466 (at the center of this interaction site) by fringe glycotransferases increases the binding affinity of Jagged1 and DLL1 ligands for Notch1, with no effect on DLL4 binding. This result may indicate that the “site 1” interaction plays a larger role in Jagged1 and DLL1 binding than in DLL4 binding, at least for the Notch1 interaction, in accordance with our signaling data (Figure 2.1C).

Direct visualization of Dll4 variants expression and quantitative analysis of Notch1 interactions

To exclude the possibility that mutagenesis of DLL4 disrupted protein production or trafficking to the cell surface, we assessed the localization of DLL4-mCherry fusion variants using fluorescence microscopy (Figure 2.2A). All DLL4-mCherry mutants exhibited similar localization patterns and expression levels, including the representative D193A mutant that is completely null in signaling assays. To further confirm that loss-of-function mutants were indeed defective in Notch1 binding, direct visualization of ligand binding to Notch1 expressing cells was done using flow cytometry.
Figure 2.2. DLL4 binding to Notch1 on the cell surface. A. Visualization of ligand-bearing cells with fluorescence microscopy. DLL4 mutants are C-terminal tagged with mCherry. B. Detection of purified ligands binding to Notch-bearing cells. Biotinylated ligands were bound to Alexa fluor 647-conjugated streptavidin, incubated with NOTCH1-gal4 U2OS or U2OS control cells and analyzed using flow cytometry. C. Quantification of ligand binding affinity using biolayer interferometry (BLI). Biotinylated Notch1(6–15) was immobilized on a streptavidin biosensor and bound to varying concentrations of DLL4(1–5) variants.
Recombinantly purified Dll4 molecules containing the N-terminal MNNL domains through EGF5 were purified, captured with streptavidin conjugated to Alexa fluor 647, and then subsequently incubated with cells stably expressing Notch1. A shift in fluorescence was observed for functionally active Dll4 constructs compared to Notch1-null control cells, but this shift was absent in the inactive DLL4 mutants (Figure 2.2B).

We next measured the affinity of a Notch1 fragment including EGF (6–15) for selected DLL4 constructs using biolayer interferometry to compare binding affinities directly. For F195A, we obtained a saturation-binding curve that could be fit readily by a single-site binding model (Figure 2.2C), whereas for the inactive D193A variant, binding was undetectable. Together, these data are consistent with the binding surface seen in the structure of the rDll4-Notch1 complex.

**DLL1 and DLL4 exhibit differential receptor selectivity in vitro**

DLL1 and DLL4 have discrete and overlapping expression patterns during embryonic development and in adult tissues. Interestingly, in regions where both ligands are expressed, they exhibit both redundant and different functional outcomes depending on the developmental context. For example, mouse Dll1 and Dll4 are both expressed in adult thymic epithelial cells where Dll4 is essential for Notch1-mediated T-lymphopoiesis. T cell development is unaffected in mice which lack Dll1 in the thymic epithelium suggesting that in this context Dll1 and Dll4 are functionally distinct. Biochemically, a previous study has shown that human Notch1 exhibits an intrinsic selectivity for DLL4 in comparison to DLL1, providing a potential explanation for the Dll4 preference seen in mouse thymic development (10).
We thus assessed whether other Notch receptors, particularly focusing on Notch2, also exhibits any selectivity for DLL1 or DLL4. To detect potential differences in ligand activity towards Notch1 or Notch2, our collaborators generated mouse E14TG2a ES (E14) cells stably expressing either Notch1 (N1rep) or Notch2 (N2rep). Co-cultures of cells expressing Dll1 or Dll4 with N1rep ES cells consistently revealed higher activation of Notch1 by Dll4 than by Dll1 (Figure 2.3A). In contrast, Dll4 activated Notch2 significantly less efficiently than Dll1 (Figure 2.3B).
Figure 2.3. DLL1 and DLL4 differentially activate NOTCH1 and NOTCH2. **A.** DLL4 activates Notch1 about 10-fold more strongly than DLL1 in co-culture assays. Left graph shows non-normalized Notch1 activation. Lines connect values measured in the same assay. Right graph shows values normalized to DLL1 activation and corrected for protein expression and cell surface presentation. **B.** DLL4 activates Notch2 about half as strongly as does DLL1. Left graph shows non-normalized Notch2 activation. Lines connect values measured in the same assay. Right graph shows values normalized to DLL1 activation and corrected for protein expression and cell surface presentation. Each dot represents a technical replicate. Adapted from (1).
Regions outside of the MNNL-DSL contact sites of DLLs contribute to the functional divergence of DLL1 and DLL4

To define the regions of DLL1 and DLL4 that confer the functional difference in activating NOTCH1 and NOTCH2, a series of domain swaps were made to generate a set of chimeric ligands (II-V, VII-X in Figure 2.4) for stimulation of NOTCH1 and NOTCH2-expressing cells in co-culture assays. To ensure that any observed differences in NOTCH1 and NOTCH2 activation was not due to differences in ligand expression or cell surface levels, receptor activation was analyzed as the response ratio of N1/N2 to stimulation by cells stably expressing chimeric ligands. Interestingly, chimeric ligands with ectodomains that encompassed the MNNL-EGF3 region of a ligand retain the full signaling capability of the corresponding full-length wild-type ligand. For example, the chimera having MNNL-EGF3 domains of DLL4 elicited a N1/N2 response ratio of ~20 which is identical to full-length DLL4. Likewise, the chimera containing MNNL-EGF3 of DLL1 showed a response ratio of ~1 or 2 like full-length DLL1. Strikingly, when chimeras include the MNNL-EGF2 or MNNL-DSL domains of one ligand and the remainder of the other, the N1/N2 response ratio of these corresponding chimeric pairs is equivalent (Figure 2.5A), suggesting that EGF3 makes an important contribution to receptor selectivity.
**Figure 2.4.** Schematic representation of DLL1 and DLL4 and variant proteins. I-X, full-length and chimeric ligands generated by domain swaps. XI and XII, ligands with exchanges of the known NOTCH1 contact amino acids in the MNNL and DSL domains. XIII, DLL4 variant with an N109G mutation that eliminates the N-glycosylation site in DLL4. XIV-XVIII, soluble proteins encoding the N-terminal region up to and including EGF5 carrying a C-terminal Avi-His-tag for protein purification. I-XIII were tested in cell-based Notch activation assays and XIV-XVIII used for measurements of binding affinities to NOTCH1. Adapted from (I).
To assess the extent to which the MNNL and DSL contact residues at the NOTCH1 binding interface influence NOTCH1 and NOTCH2 selectivity by DLL1 and DLL4, we reciprocally exchanged these amino acids (XI-XII in Figure 2.4; Figure 2.5B) based on alignments of the DLL4 (2) and DLL1 (19) structures. Replacing the contact residues of DLL1 with the analogous residues of DLL4 does not substantially affect the N1/N2 response ratios in comparison to WT DLL1 itself, indicating that the selectivity of N1 or N2 by DLL1 and DLL4 cannot simply be accounted for by interfacial residues in the MNNL-DSL region (Figure 2.5C). Similarly, swapping the DLL1 contact residues onto DLL4 does not collapse the response ratio to 1 (Figure 5C), however it slightly reduces the mean N1/N2 activation ratio (to ~13), again strongly suggesting that properties outside of the MNNL-DSL contact interface contribute to receptor selectivity. This is in accordance with the prior observation that affinity maturation of DLL4 accumulates mutations in the protein core but not in the binding interface (2). A DLL4 variant containing the analogous contact residues DLL1 did not show reduced binding affinity for Notch1 (D4contD1 $K_D = 0.327 \pm 0.036$ mM; Figure 2.5D panel c), fully consistent with the interpretation that the protein core of DLL4 contributes to Notch1 binding affinity, likely by influencing allosteric changes to induce a higher affinity conformation. Swapping the contact residues of DLL4 onto DLL1 did increase binding affinity for N1 (D1contD4 $K_D = 0.326 \pm 0.044$ mM; Figure 2.5D panel d), however the N1/N2 response ratio was not affected indicating that the contact amino acids and different binding affinities for N1 are not the only influence on the selectivity of the two ligands for N1 or N2.

Previous studies have shown that Notch signaling can be modified by a number of post-translational modifications including glycosylation. Like the Notch receptors, Notch ligands are glycoproteins and the ligand DLL4 contains three putative conserved glycosylation sites in its
MNNL domain, one of which - N109 - is not shared with the other activating member of the Delta-like family of ligands, DLL1. Additionally, glycosylation of N109 was confirmed in the NOTHC1-DLL4 complex structure where it lies adjacent to the interaction surface. To determine if glycosylation of N109 plays a role in the functional divergence of DLL1 and DLL4 we created a DLL4 N109G (D4\textsuperscript{N109G}) mutant which contains the analogous amino acid from DLL1 and assessed the effect on DLL4-mediated signaling and NOTCH1 binding. D4\textsuperscript{N109G} had no effect on the relative activation potential of DLL4 for Notch1 versus Notch2 (Figure 2.5C), and its affinity for Notch1 was not altered (K\textsubscript{D} = 0.341 ± 0.015 mM; Figure 2.5D panel e), indicating that N-glycosylation at this site does not significantly modulate N1 binding or contribute to the relative selectivity of DLL4 towards Notch1 and Notch2.
**Figure 2.5.** Contributions of the MNNL-EGF3 portion and contact amino acids to ligand selectivity towards N1 and N2. A. N1/N2 activation ratios by DLL1 and DLL4 chimeric proteins show that receptor selectivity of DLL1 and DLL4 is encoded by the extracellular domain and that EGF3 contributes to N1/N2 selectivity. DLL4, D4ECD_D1ICD, and D4N-E3_D1 show N1/N2 induction ratios of ~20. DLL1, D1ECD_D4ICD, and D1N-E3_D4 exhibit induction ratios of 1–3. Chimeric pairs with domain exchanges between EGF2 and EGF3 or between DSL domain and EGF1 show equivalent stimulation ratios. B. Segments of the MNNL and DSL sequences showing the contact amino acids (boxed), the divergent amino acids of DLL1 (red)
and DLL4 (blue), and the sequence of ligands with amino acid exchanges. The N-glycosylation site at residue N109 of DLL4 is indicated in green. D. N1/N2 activation ratios of ligands with exchanged N1 contact amino acids. D1contD4 does not show changes in receptor selectivity compared to DLL1. Replacing the contact residues of DLL4 with those of DLL1 only reduces N1/N2 activation ratio to ~13. Elimination of the N-glycosylation site of DLL4 with the N109G mutation (the corresponding amino acid of DLL1) does not change DLL4 receptor selectivity. Adapted from (1).
Discussion

In vertebrates the Delta-like family of proteins has two Notch activating ligands, DLL1 and DLL4. DLL1 and DLL4 are similar in sequence (human paralogues have an amino acid similarity of 63% with 49% identity), size and domain structure. In addition to high sequence similarity both proteins occasionally have overlapping expression during embryonic development and in adult tissues. In sites where both DLL1 and DLL4 are expressed, they appear to have both redundant and divergent functions, but the bases for these differences are unclear.

Here, we used alanine scanning mutagenesis of contact amino acids in the MNNL and DSL domains of DLL4 to first assess the relative contributions of the contact sites to Notch1 receptor activation in our cell-based assay. The integrity of the DSL domain is necessary for Notch receptor binding (20) and this domain is in contact with Notch1 in the Notch1-Dll4 crystal structure (2). Our mutagenesis results confirm that alanine substitution of select contact residues in the DSL domain of DLL4 is able to abolish Notch signaling. In addition, we have shown that mutants that disrupt Dll4-induced Notch signaling do so through abrogating ligand binding. Of the two DLL4 binding sites revealed in the x-ray model, the DSL domain is more sensitive to interface mutations, suggesting that this contact interface makes a dominant contribution to receptor binding.

The Notch1 contacting residues of DLL4 are largely conserved in DLL1 – 75% similar with 55% identity – and the most detrimental mutations are of residues that are identical in the two proteins. Nevertheless, DLL1 and DLL4 exhibit a receptor response “discrimination
potential” of ~20-fold between Notch1 and Notch2 in cell-based assays. Systematic domain swap analysis revealed that the region responsible for this receptor discrimination maps to the MNNL-EGF3 region (Figure 2.5A), extending beyond the MNNL and DSL interfaces visualized in the Notch1-Dll4 x-ray structure. This finding is consistent with previous work that points to the importance of the MNNL as well as the DSL domain and EGF repeats 1-3 for strong activation of Notch1 and to the importance of EGF repeats 8-10 of Notch1 for signal activation by DLL ligands.

These observations suggest that interactions of EGF repeats 1-3 of the DLL ligands with EGF repeats 8-10 of Notch also contribute to recognition and impart discriminatory potential. Together, the results also favor the conclusion that the contact amino acids in the MNNL and DSL domains do not make the dominant contributions to the functional divergence of DLL1 and DLL4 in vivo, suggesting instead, that differences in the domain cores, and/or contacts outside of the known DLL4-Notch1 interface, are the factors that contribute most to functional divergence.
Materials and Methods

Cloning of constructs

Constructs with exchanges of domains or individual amino acids in the extracellular domain of DLL1 and DLL4 were generated by standard cloning procedures using either synthesized gene fragments (II-IV, VII-IX, XI-XIII in Fig 1) or fragments obtained by restriction digests from Dll1 and Dll4 cDNA constructs (V, X in Fig 1). For production and purification of proteins for binding assays (XIV-XVIII in Fig 1) fragments of encompassing the N-terminus up to and including EGF5 were PCR amplified and cloned into pLexM-Avi-His vector (10). Hprt targeting constructs for expression of D1ECD_D4ICD and D4ECD_D1ICD were generated by cloning the respective cDNA into pMP8-CAG.Stop.

Culture and generation of cells stably expressing ligand proteins

Mouse E14TG2a ES cells were cultured in DMEM (Invitrogen) cell culture medium supplemented with 15% FCS (Biochrom AG), Glutamax, Pen/Strep, Sodium Pyruvate, MEM Non-Essential Amino Acid Solution, β-mercaptoethanol, and leukemia inhibitory factor (21). ES cells were electroporated with linearized pMP8 targeting vectors and selected with HAT (1:300; Gibco). Correct integration of the 5’ homology arm in HAT resistant clones was verified with long-range PCR using following primers: For/Rev:

GGGAACCTGTAGAAAAAAGAAAACATGAAAGAAC /
GGCTATGAACTATGACCCCG. Expression of proteins was verified using Western Blot analyses.
Trans-activation assay

To assess the activity of DLL4 ligand point mutations, a luciferase reporter assay was performed using U2OS cells transiently transfected with plasmid DNA encoding the different DLL4 proteins, co-cultured with U2OS Flp-in Notch1-Gal4 receiver cells transfected with Gal4 reporter plasmid and a Renilla luciferase internal control plasmid. On day 2, DLL4 cells were split onto 6-well plates containing Notch1-Gal4 cells which were induced with doxycycline (1µg/ml). On day 3, following 24 h of culture, firefly and Renilla luciferase activities were measured on a luminometer (Turner Systems) in whole cell extracts using the Dual Luciferase kit (Promega). Each measurement was normalized to non-ligand bearing U2OS cells. All data points within experiments were obtained in triplicate, and each experiment was repeated at least three times.

For in vitro cell co-culture assays (using stable ligand and receptor bearing cells) ES cells were counted in PBS using LUNA-II (logos biosystems) and 9.25x10^5 ligand and 0.75x10^5 receptor expressing cells were plated on gelatin coated 6 well plate dishes. After 24hrs fresh medium was added. 48-52hrs after co-cultivation cells were washed once with PBS, lysed in 250µl 1xCCLR (Luciferase Cell Culture Lysis Reagent, Promega), transferred into 1.5ml tubes, and frozen at -80°C. For measurements lysates were thawed, vortexed, and briefly centrifuged. 20µl aliquots of each lysate was measured with Luciferase Assay Reagent in duplicates or quadruplicates using GloMax-96 (Promega).
**Fluorescence Microscopy**

U2OS cells were transiently transfected with plasmid DNA encoding the different DLL4 proteins containing C-terminal mCherry fusions. The cells were plated onto Mat-tek glass bottom dishes and then imaged and photographed on a Nikon Eclipse Ti fluorescence microscope linked to a Photometrics CoolSNAP EZ camera.

**Flow Cytometry Binding Assay**

N1gal4 U2OS cells were induced with doxycycline (1 µg/ml) overnight, then trypsinized, washed twice with TBS, and blocked in TBS containing 3% BSA on ice for 30 min. Alexa Fluor 647-coupled streptavidin was incubated with 2 µM biotinylated ligand for 1 h in TBS containing 5 mM CaCl2 (TBS-Ca). Cells were spun down, resuspended in TBS-Ca, and incubated with the corresponding ligand-streptavidin complexes for 1 h. Cells were washed 2–3 times, resuspended in TBS-Ca, and analyzed using a BD Accuri C6 Plus personal flow cytometer.

**Protein Expression and Purification**

Notch1 fragment encoding Notch1 signal sequence followed by EGF repeats 6-15 (amino acids 216-604) was cloned into pLexM-Avi-His vector. Expi293F cells maintained in Expi293 expression media were grown to cell density of $10^6$ cells/ml and then transiently transfected with Dll1 ligand, Dll4 ligand or Notch1 DNA (1 mg/liter of cells) and FectoPro transfection reagent (Polyplus) at 1:1 DNA/FectoPro ratio. For biotinylation of Avi-tagged Notch1 protein, cells were
co-transfected with biotin ligase (BirA) DNA as well as DNA encoding Protein O-fucosyltransferase-1 (POFUT1) which enhances Notch folding and secretion.

The encoded proteins were expressed and secreted in cultured FreeStyle 293 expression media. Following expression for 3-4 days the cultured media was collected, separated by centrifugation and supplemented with 50mM Tris buffer, pH 8.0. The resulting supernatant was bound to Ni-NTA beads (Qiagen) over a 3-hour incubation at 4⁰ C. After a wash with ten column volumes of 50mM Tris buffer, pH 8.0, containing 150mM NaCl, 5mM CaCl₂, and 20mM Imidazole, bound protein was eluted with the same buffer supplemented with 250mM Imidazole. Following elution, fractions containing the partially purified proteins were concentrated and further purified by gel-filtration chromatography using a Superdex 200 Increase 10/300 GL column in 50 mM Tris, pH 8.0, 150mM NaCl, and 5mM CaCl₂. The quality of the resulting purified proteins was assessed via non-reducing SD-PAGE and pure fractions were pooled, flash frozen and stored at -80⁰ C. The efficiency of biotinylation was estimated by immunoprecipitation with streptavidin resin.

Biolayer Interferometry

Ligand binding affinities were quantified by biolayer interferometry using BLItz instrument (ForteBio). Streptavidin biosensors were loaded with biotinylated Notch1 fragment, equilibrated in buffer for 30 seconds, then dipped into samples of serial dilutions of ligand concentrations until equilibrium was observed. All ligand-receptor binding experiments were done in HBS-P buffer containing 0.005% surfactant P20 and supplemented with 5Mm CaCl₂. Equilibrium binding curves were fitted with one site- specific binding models using GraphPad Prism.
References


CHAPTER 3

Extension of the NICD Ankyrin Repeat Stack by NRARP Promotes Feedback Inhibition of Notch Signaling

This Chapter is adapted from:

Sanchez M. Jarrett, Tom C. M. Seegar, Mark Andrews, Guillaume Adelmant, Jarrod A. Marto, Jon C. Aster, and Stephen C. Blacklow. In Review

http://biorxiv.org/cgi/content/short/645937v1
Author Contributions:

I designed the experiments and analysis, purified protein for crystallization, and determined the crystal structure. Tom Seegar assisted with structure determination. I cloned all the proteins, as well as the crystallization constructs except for NOTCH1 ANK and RAMANK domains. Mark Andrews assisted with protein production and retroviral experiments. Guillaume Adelmant performed the mass spectrometry analysis used to determine NRARP interactors. Stephen C. Blacklow supervised the research.
Abstract

Canonical Notch signaling relies on regulated proteolysis of the receptor Notch to generate a nuclear effector that induces the transcription of Notch-responsive genes. In higher organisms, one Notch-responsive gene that is activated in many different cell types encodes the Notch-regulated ankyrin repeat protein (NRARP), which acts as a negative feedback regulator of the Notch response. Here, we show that NRARP exerted its inhibitory effect by binding directly to the core Notch transcriptional activation complex (NTC), requiring both the transcription factor RBPJ and the Notch intracellular domain (NICD), but not Mastermind-like proteins or DNA. The X-ray structure of a NRARP-RBPJ-NICD1-DNA complex, determined to 3.75 Å resolution, revealed that assembly of NRARP-RBPJ-NICD1 complexes relies on simultaneous engagement of RBPJ and NICD1 in a non-canonical binding mode in which the three ankyrin repeats of NRARP extend the Notch1 ankyrin repeat stack. NRARP-NICD1 interface mutations disrupt entry into NTCs and abrogate feedback inhibition in Notch signaling assays in cultured cells. These studies establish the structural basis for NTC engagement by NRARP and provide insights into a critical negative feedback mechanism that regulates Notch signaling.
Introduction

Cell signaling enables an organism to perceive and respond to its local environment. This fundamental process occurs in a series of tightly regulated steps that require stimulus detection, signal transmission, and a downstream response. The amplitude and duration of the response can be tuned by various signaling modulators, which can vary widely based on the cellular context. One common mechanism of signal modulation is feedback inhibition, in which the downstream response to the signal produces an output that suppresses the initiating signal. Feedback regulation is particularly important in developmental signaling, wherein control of the timing and strength of the signal is critical to ensure proper cellular proliferation and differentiation.

Notch signaling is a major primary juxtacrine developmental signaling pathway controlling cell fate decisions in multicellular organisms (1). Mutations of the core components of this pathway give rise to various developmental disorders, including Alagille syndrome (2), left ventricular non-compaction (3), spondylcostal dysostosis (4), and Hajdu-Cheney syndrome (5-7) as well as adult-onset diseases such as cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) (8). In addition, aberrant or dysregulated Notch signaling is associated with many different human cancers, including T cell acute lymphocytic leukemia (T-ALL), in which activating mutations of Notch1 are found in more than half of all cases (9).

Notch signaling activation depends on cell-cell contact between a ligand-expressing “sender” cell and a receptor-expressing “receiver” cell (10, 11). Ligand binding results in regulated
intramembrane proteolysis of the receptor, liberating the intracellular portion of Notch (NICD) from the membrane (12-14). NICD then migrates into the nucleus and enters into a Notch transcriptional activation complex (NTC), which also includes the transcription factor RBPJ and a coactivator protein of the Mastermind-like (MAML) family (15-17), resulting in induced transcription of Notch-responsive genes (see (I) for a review).

The gene encoding Notch-regulated ankyrin repeat protein (NRARP) is one of a small number of core Notch target genes. NRARP cDNA was first identified in an in situ hybridization screen following analysis of the Delta1 synexpression group of developmentally expressed genes in *Xenopus laevis* embryos (18). The cDNA was later shown to encode a 114 amino acid protein consisting of at least two tandem C-terminal ankyrin repeats, and to be regulated at the transcriptional level by the Notch signaling pathway in *Xenopus* and mice (19, 20).

Several studies in different organisms and developmental contexts have reported that NRARP is a negative regulator of Notch signaling. Enforced expression of *Nrar* in mice leads to a block in T cell development, which requires Notch signaling at multiple stages (21). In addition, *Nrar* knockout mice exhibit defects in somitogenesis and vascular pruning in the eye (22, 23). Both of these phenotypes are associated with excess Notch activity, and, together with studies investigating the influence of *Nrar* on cell fate decisions in the mouse retina (24), further support the conclusion that NRARP counteracts Notch signals in vivo.

The molecular basis for the action of NRARP as a negative regulator of Notch signaling is incompletely understood. Experiments in *Xenopus* embryos using overexpressed tagged Notch
signaling components and NRARP detected association of NRARP with NICD and RBPJ, suggesting that the NRARP protein enters into a complex with the NTC (19). However, proteomic studies to uncover NRARP-interacting, endogenously expressed proteins have not been performed, nor has an NRARP-NTC complex been reconstituted using purified proteins. Moreover, there are no structural data available for NRARP or NRARP-containing complexes. Thus, the molecular basis for the function of NRARP as an attenuator of Notch signal transduction has remained elusive.

Using mass spectrometry of tandem-affinity–purified NRARP complexes, we show here that human NRARP associated with endogenous NTCs containing NICD1, RBPJ, and a MAML coactivator. Using purified proteins, we found that NRARP bound directly to RBPJ-NICD1 complexes, requiring both RBPJ and NICD1, but not MAML or DNA, for entry into NTCs. The X-ray structure of a NRARP-RBPJ-NICD1-DNA complex, determined to 3.75 Å resolution, revealed that assembly of NRARP-RBPJ-NICD1 complexes relies on NRARP simultaneously engaging RBPJ and Notch1 in a non-canonical binding mode involving the extension of the Notch1 ankyrin repeat stack by the three ankyrin repeats of NRARP. NICD1-NRARP interface mutations disrupted entry into NTCs and abrogated feedback inhibition in Notch signaling assays. These studies establish the structural basis for NTC engagement by NRARP and provide insights into a critical negative feedback mechanism that regulates Notch signaling.

**Results**

*NRARP inhibits Notch signaling and suppresses growth of Notch-dependent T-ALL cells*
Previous studies in *Xenopus* (19) and in mice (20, 23, 24) have implicated NRARP as a negative regulator of Notch signaling (Figure 3.1A). To test whether human NRARP inhibits Notch activity in cells, we used a well-established luciferase reporter-gene assay (25). Enforced expression of NRARP suppressed reporter activity induced by intracellular Notch1 (NICD1) in a dose-dependent manner (Figure 3.1B), consistent with prior studies. To test whether the negative regulatory activity of NRARP is selective for NICD1, we also tested the effect of NRARP on reporter gene induction by NICD2, NICD3, and NICD4. The data show that NRARP inhibited reporter gene induction by all four human Notch proteins, although the inhibitory effect on NICD4-dependent reporter activity was not as strong as the inhibitory effect on the other NICDs (supplementary figure S1, A to D).

Previous studies have shown that genetic and chemical inhibition of Notch in Notch-mutated T-ALL cell lines results in suppression of cell growth. We tested whether enforced expression of NRARP in two Notch-dependent T-ALL cell lines, DND-41 and HPB-ALL, resulted in growth suppression, as predicted for a negative modulator of signaling. Cells were infected with green fluorescent protein (GFP)-expressing retroviruses that were either empty or carried the cDNA for NRARP or a dominant-negative form of MAML1 (dnMAML1) (25). The fraction of GFP-positive cells in the population was then monitored over time to assess the effect of NRARP or dnMAML on cell growth (in this assay, a reduction in the GFP positive population over time is indicative of growth suppression). Transduction with NRARP suppressed growth of both T-ALL lines to a similar extent as did dnMAML, whereas empty virus had no effect (Figure 3.1, C and D). Analysis of the Notch-responsive genes *HES1, HES4*, and *DTX1* in cells expressing NRARP revealed decreased amounts of mRNA (Figure 3.1E), again consistent with the conclusion that NRARP is a negative regulator of Notch signaling. The amount of NOTCH1 mRNA, on the
other hand, was unchanged in NRARP-expressing cells, suggesting that NRARP regulates NOTCH1 not by influencing its expression, but by modulating its activity at the protein level (Figure 3.1E).
Figure 3.1. NRARP is a negative feedback regulator of Notch signaling. (A) NRARP expression induced by Notch transcriptional activation complexes results in feedback repression of Notch signaling activity mediated by the NRARP protein. (B) Luciferase reporter assay probing the effect of NRARP on NICD1 transcriptional induction activity. Quantification of
luciferase activity in NIH 3T3 cells that were transiently transfected with the indicated amounts of pcDNA3-based plasmids for expression of NICD1 and NRARP, a firefly luciferase reporter plasmid under control of the TP1 Notch response element, and a plasmid encoding Renilla luciferase. Firefly luciferase activity is reported relative to Renilla control, with the ratio for cells transfected with empty pcDNA vector set to a value of 1. n=3 (independent experiments). Statistical analysis was performed using ANOVA, and a Dunnett’s multiple comparison post hoc test was performed comparing test samples to the control. ****p < 0.0001. (C, D). Effect of NRARP or dnMAML on the growth of DND-41 (C) and HPB-ALL (D) T-ALL cells. Each cell line was transduced with retrovirus expressing dnMAML1-GFP, NRARP with GFP behind an IRES, or GFP alone. The relative fraction of GFP-positive cells is plotted as a function of time. The asterisk indicates an impurity band (likely streptavidin lost from the streptavidin-coated beads). n=2 (independent experiments). (E) Quantification of the abundance of NOTCH1 and sentinel Notch target gene (HES1, HES4, and DTX) transcripts in NRARP-transduced Jurkat cells, measured using qPCR. N=2 (independent experiments). Statistical analysis was performed with Prism 8 software (Graphpad) using a two-tailed, unpaired t test. **p < 0.01; ***p < 0.001; ****p < 0.0001.
Direct binding of NRARP to NOTCH1-RBPJ complexes requires both RBPJ and NOTCH1

Previous work carried out in Xenopus embryos using forced expression of tagged proteins showed that NRARP co-immunoprecipitates with Xenopus RBPJ and Xenopus NICD1, suggesting that NRARP enters a complex with Notch and RBPJ (19). To identify the complete spectrum of proteins that associate with human NRARP, we used tandem-affinity purification of HA-FLAG–tagged NRARP from Jurkat cells followed by mass spectrometry of the recovered endogenous proteins. We consistently recovered peptides for the three core components of the NTC: RBPJ, NOTCH1, and MAML1, in four independent experiments (Table 1; see also table S1).

RBPJ contains three structured domains that encompass most of the coding sequence, followed by a region that is not required for assembly of RBPJ-NICD1-MAML1 complexes on DNA. NICD1 has a RAM (RBP-associated molecule) region, a series of ankyrin repeats (ANK), a transcriptional activation region (TAD), and a C-terminal PEST sequence. NRARP is predicted to have three ankyrin repeats, and MAML1 is predicted to be unstructured C-terminal to the region required for formation of the NTC (Figure 3.2A).

To map the domain requirements for formation of complexes between NTC components and NRARP, we purified the RAM-ANK region (hereafter NICD1) of NOTCH1 and the structured portion of RBPJ. Neither of these proteins, alone or in combination, formed stable complexes with NRARP (Figure 3.2B). However, extension of the C-terminus of RBPJ to residue 452 enabled purification of stable complexes, but only when both RBPJ and NICD1 were present.
(Figure 3.2B). Further domain-mapping studies established that the only region of NICD1 that was required for complex formation was the ANK domain (Figure 3.2C), and that the association of NRARP with NICD1 and RBPJ did not compete with binding to either MAML1 or DNA (Figure 3.2D).
Figure 3.2. Requirements for complexation of NRARP with Notch1 and RBPJ. A. Domain organization of the protein components found in NRARP complexes. NRARP has three predicted ankyrin-repeats (ANK). RBPJ contains an N-terminal domain (NTD), β-trefoil domain (BTD), and a C-terminal domain (CTD). Intracellular Notch1 (NICD1) contains an RBPJ associated molecule (RAM), an ankyrin-repeat domain (ANK), a transcriptional activation domain (TAD), and a PEST (proline, glutamate, serine and threonine) sequence. MAML1 contains an N-terminal Notch/RBPJ binding region (red) followed by a long C-terminal portion predicted to be natively disordered (white). B. Binding of NRARP to RBPJ, NICD1, and RBPJ-
NICD1 complexes. Biotinylated-NRARP, RAMANK, and the indicated forms of RBPJ were combined, complexes were recovered using streptavidin-sepharose beads, and the recovered proteins were analyzed on a Coomassie-stained gel. Data are representative of n=2 (independent experiments). C. NICD1 domain requirements for formation of complexes with RBPJ and NRARP. Biotinylated-NRARP, RBPJ and either ANK or RAMANK were combined, complexes were recovered using avidin-sepharose beads, and the recovered proteins were analyzed on a Coomassie-stained gel. Data are representative of n=2 (independent experiments). D. Association of NRARP with NICD1-RBPJ in the presence of MAML1 and DNA. Biotinylated-DNA (containing a single RBPJ binding site), MAML1(13-74), NRARP, RBPJ and Notch1 RAMANK proteins were combined, complexes were recovered using streptavidin-sepharose beads, and the recovered proteins were analyzed on a Coomassie-stained gel. Data are representative of n=3 (independent experiments).
Crystal structure of an NRARP-NOTCH1-RBPJ-DNA complex reveals a composite binding interface

To determine the structural basis for recognition of NICD1-RBPJ complexes by NRARP, we determined a 3.75 Å X-ray structure of an NRARP-NICD1-RBPJ complex bound to DNA, phased using molecular replacement with the human NICD1-RBPJ-MAML1-DNA complex (PDB ID code 2F8X; (26)) as a search model (Table 2). The structural features described here are drawn from the better ordered of the two assemblies seen in the asymmetric unit (figure 3.S2A, B).

The most striking feature of the complex is the assembly of the ankyrin repeats from NICD1 and NRARP into a pseudo-continuous stack that wraps around the RBPJ-DNA complex in a crescent-shaped arc (Figure 3.3A). The extended ankyrin repeat stack results in an elongated assembly overall, with dimensions of approximately 120 x 70 x 60 Å. The arrangement of RBPJ, NICD1, and the DNA within the complex are minimally affected by the binding of NRARP, as the NICD1 and RBPJ subunits of the NRARP complex superimpose with a backbone root-mean-square deviation (RMSD) of 1.06 Å when compared with the transcriptional activation complex that contains NICD1, RBPJ and MAML on DNA (Figure 3.3B, C).

NRARP itself is a single protein domain with three predicted ankyrin repeats. In the structure of the complex, however, the first ankyrin repeat is less ordered that the other two, and its first helix is modeled only as polyalanine even in the better-defined copy of the asymmetric unit. The third ankyrin repeat of NRARP engages the NICD1-RBPJ interface at a composite surface that
includes the first ankyrin repeat of NICD1 and the C-terminal domain of RBPJ. At this interface, NRARP is oriented with its C-terminal ankyrin repeat abutting the N-terminal ankyrin repeat of NICD1, thereby creating the pseudo-continuous stack of repeats that travels along the C-terminal Rel-homology domain of RBPJ (Figure 3.3A). The NRARP-NICD1 interface results in improved electron density for the first ankyrin repeat of NICD1 when compared with the structure of the human NTC, suggesting that NRARP binding stabilizes the structure of this repeat. The interaction between NRARP and RBPJ relies on the canonical concave binding surface of the third ankyrin repeat of NRARP, which appears to approach within contact distance of the C-terminal extension of RBPJ. The interface between NRARP and the NICD1-RBPJ complex does not overlap the NICD1-RBPJ interface with MAML1 and is completely compatible with the observed simultaneous binding of NRARP and MAML1 by RBPJ-NICD1 complexes on DNA (Figure 3.3C). Moreover, the NRARP binding site is also non-overlapping with the NTC dimerization interface (27) on the convex face of the NICD1 ankyrin domain (figure 3.S2C). Key NRARP residues at the contact interface with NICD1-RBPJ include Trp^85 and Ala^92 of the third ankyrin repeat (Figure 3.4A). Trp^85 makes contacts in a cleft created primarily by residues on RBPJ, with an additional potential contact with Pro^1880 of NICD1, whereas Ala^92 (stick) of NRARP approaches the first helix of the NICD1 ANK domain (Figure 3.4B).
Figure 3.3. Structure of an NRARP-Notch1-RBPJ complex on DNA, and comparison with the human Notch1-RBPJ-MAML1 (PDB ID code 3V79; (28)) transcriptional activation complex. A.
Ribbon representation of the NRARP-Notch1-RBPJ complex on DNA. The complex contains NRARP (pink), RBPJ (green), Notch1 RAMANK (blue) and a 16-mer DNA (orange) containing a single RBPJ binding site. B. X-ray structure of RBPJ in complex with the ankyrin domain of Notch1 (blue) and MAML1 (red) on DNA (orange). C. Overlay of the NRARP-Notch1-RBPJ-DNA complex (colors) with the RBPJ-Notch1-MAML1-DNA complex (gray).
Figure 3.4. Interface between NRARP and NICD1-RBPJ. A. Molecular surface representation with open book views. Center panel: NRARP is pink, RBPJ is green and the ANK domain of Notch1 is blue. Left panel: open book view of NRARP. NRARP is rotated 60 degrees clockwise and residues that approach within 4 Å of the NICD1-RBPJ surface are colored in a darker shade. NRARP interface residues Trp\textsuperscript{85} and Ala\textsuperscript{92} are indicated, as are Pro\textsuperscript{1880} and helix one of NICD1.
Right panel: open book view of the NICD1-RBPJ surface. NICD1-RBPJ is rotated 60 degrees counterclockwise and residues that approach within 4 Å of NRARP are colored in darker shades. B. Close-up view of the binding interface. Trp$^{85}$ of and Ala$^{92}$ of NRARP, and Pro$^{1880}$ of NICD1, are shown in ball and stick form. Helix one of NICD1 and the C-terminal extension of RBPJ are shown in a darker shade.
Inhibition of Notch signaling by NRARP depends on the binding interface

To determine whether inhibition of Notch activity by NRARP relies on the binding interface seen in the X-ray structure, we tested the effect of mutating conserved NRARP residues at this interface in the reporter gene assay. The first mutation, W85E, significantly attenuated the inhibitory effect seen with wild-type NRARP. When combined with an additional A92W mutation (W85E/A92W), the attenuation was even greater (Figure 3.5A). Neither the single nor the double mutation disrupted the overall structural integrity of purified NRARP protein (Figure 3.5B), as judged by the near equivalence of their circular dichroism (CD) spectra (Figure 3.5C).

To determine whether the reduced inhibitory activity of the NRARP mutants was indeed due to a decrease in NRARP binding as predicted, we directly tested binding of purified recombinant NRARP polypeptides to RBPJ-NICD1 complexes on DNA. Whereas wild-type NRARP was efficiently pulled down by streptavidin in complexes with NICD1-RBPJ on biotinylated DNA, the W85E and W85E/A92W NRARP variants were not, indicating that both mutants are defective in forming complexes (Figure 3.5D).
Figure 3.5. A. Effect of NRARP interface mutations on Notch-dependent luciferase reporter gene activity. Firefly luciferase activity is reported relative to that of the Renilla luciferase, setting the firefly:Renilla ratio in cells transfected with empty pcDNA vector control to a value of one. n=3 (data are representative of 3 independent experiments). Statistical analysis was performed using ANOVA, and a Dunnett’s multiple comparison post hoc test was performed.
comparing test samples to the control. *p < 0.05; ****p < 0.0001. B. Coomassie-stained gel of purified wild-type (WT), W85E, and W85E/A92W forms of NRARP. n=1. C. CD spectra of purified WT, W85E, and W85E/A92W NRARP proteins. n=1. D. Effect of NRARP interface mutations on binding to preassembled NICD1-RBPJ complexes, captured on biotinylated DNA. Proteins were analyzed on a Coomassie-stained gel. The asterisk indicates an impurity band (likely streptavidin lost from the streptavidin-coated beads). n=2 (data representative of 2 independent experiments).
NRARP promotes NOTCH turnover

Prior studies have reported a decrease in detectable levels of NICD when both NOTCH1 and NRARP are transiently co-expressed in *Xenopus* embryonic extracts (19). To determine whether NRARP affected the abundance of endogenous Notch1 in human cells, we infected Jurkat cells with control retrovirus expressing GFP only, virus expressing dnMAML1, or virus expressing NRARP, and probed cell lysates for both total Notch1 and activated Notch1 (NICD1). Whereas expression of dnMAML led to accumulation of activated Notch1 (Figure 3.6A) compared to vector control, expression of NRARP reduced the level of NICD1 without depleting total Notch1 (Figure 3.6A), suggesting that NRARP selectively promoted degradation of the active intracellular form of Notch1 (NICD1).
Figure 3.6. Effect of NRARP on the degradation of activated Notch1 complexes. A. Jurkat cells were transduced with retrovirus expressing GFP alone, a GFP-DnMAML fusion protein, or
NRARP, and the amounts of activated Notch1 (top panel), total Notch1 (middle panel), or GAPDH (bottom panel) were determined by Western blot with the indicated antibodies. \( n = 3 \) (independent experiments). B. Proposed model for NRARP function in Notch signaling. Notch pathway activation results in transcription of NRARP, a Notch target gene. NRARP binds to Notch transcriptional activation complexes, accelerating degradation of intracellular Notch and downregulating target gene expression.
Discussion

The induced expression of Notch target genes relies on the formation of an NTC containing the transcription factor RBPJ, NICD, and a MAML coactivator on DNA. The stepwise assembly of an NTC begins with NICD binding to RBPJ, which enables the engagement of MAML followed by the recruitment of additional co-factors for target gene transcription. RBPJ also mediates transcriptional repression in the absence of Notch by interacting with an alternative set of cofactors, including several co-repressors, such as SHARP \((29, 30)\) and KyoT2 \((31)\).

We report here the structure of an NRARP-NICD1-RBPJ complex bound to DNA, revealing the molecular basis for interaction of the Notch feedback inhibitor NRARP with the core NTC. Previous work anticipated the direct interaction of NRARP with Notch-RBPJ complexes from studies with overexpressed proteins in Xenopus extracts \((19)\). Consistent with those observations, we detected a robust interaction between human NRARP and endogenous components of the core human NTC in Jurkat cells using an unbiased proteomic approach. When we reconstituted a complex with purified proteins, we found that NRARP associated directly with NICD1-RBPJ complexes and required the presence of both NICD1 and RBPJ for complex formation, much as MAML proteins require both NICD1 and RBPJ for NTC assembly. Our work with purified proteins also established that a C-terminal extension of RBPJ beyond the region used in previous structural studies is required for stable entry of NRARP into RBPJ-Notch complexes, but that neither MAML nor DNA was needed for NRARP recruitment.
NRARP binding did not induce any major conformational changes in the NICD1-RBPJ complex, nor did it interfere with binding of MAML1 or DNA. Instead, NRARP extended the ankyrin repeat stack of NICD1 by three repeats, engaging the first of the NICD1 ankyrin repeats with its C-terminal repeat. NOTCH4, which is the human homolog least sensitive to inhibition in reporter gene assays (figure 3.S1), also shows the greatest divergence in its first ankyrin repeat, explaining its reduced sensitivity to NRARP inhibition. In contrast, NRARP relies on the concave surface of its ankyrin repeat stack to contact RBPJ, using a binding mode seen frequently in other ankyrin repeat protein complexes. The induced ordering of the C-terminal extension of RBPJ forces the serine/threonine-rich C-terminal segment of RBPJ away from the core of the NTC, potentially exposing it to post-translational modifications that may regulate NTC turnover.

Because NRARP binds to a composite RBPJ-Notch surface, its action is restricted to effector signaling complexes that are engaged in inducing a transcriptional response. Prior work (32) pointed to a link between the assembly of Notch transcription complexes to their timed destruction, with estimates for the half-life of activated (gamma-secretase-cleaved) Notch of roughly 2-4 h. Transcriptional induction of canonical target genes like NRARP in response to activated Notch can occur within 1 h, indicating that negative feedback regulation by direct binding of NRARP to promote Notch turnover may be one of the early steps in the molecular mechanism underlying this “timed destruction” program.

A number of mechanistic explanations may account for why binding of NRARP to the NTC inhibits Notch target gene activation. One possibility is that NRARP binding directly alters the
ability of the NTC to recruit co-factors to stimulate transcription, but this explanation seems unlikely because NRARP does not appreciably affect either DNA or MAML binding. Another possibility is that bound NRARP recruits enzymes that directly modify the NTC to suppress transcriptional induction. Along these lines, there are phosphorylation sites near the NRARP binding site on NOTCH1 (33). Though these sites (T1898 and S1901) are distant from the DNA contact interface on RBPJ and unlikely to directly affect DNA binding when modified, their phosphorylation could indirectly affect the ability of NTCs to stimulate transcription. The most likely model for NRARP action, however, supported both by data presented here (Figure 3.6A) and by complementary studies using Xenopus extracts (19), is that NRARP accelerates NTC turnover (Figure 3.6B), likely by promoting such posttranslational modifications of Notch, as well as of RBPJ and/or MAML. The molecular pathway for NRARP-mediated NTC turnover, which appears to be present in both physiological and pathological contexts, could be a future avenue for development of therapeutics designed to modulate the dynamic response of cells to Notch pathway activation.
**Materials and Methods**

*Protein expression and purification*

Sequences encoding wild type and mutant full length NRARP proteins were inserted into a pETHSUL vector at BamHI and XhoI sites to produce NRARP as a cleavable His-SUMO fusion protein. To make a biotin-tagged form of NRARP for pull-down assays, a cassette encoding a biotinylation (avi) tag was inserted between the His-SUMO affinity tag and the full-length NRARP sequence, and this expression cassette was inserted into a pETDUET vector encoding biotin ligase (BirA) for simultaneous NRARP expression and in-cell biotinylation. Point mutants were produced by site-directed mutagenesis.

For protein production, expression constructs were transformed into Rosetta(DE3)pLysS cells and cells were grown in rich media at 18°C. Protein expression was induced with 0.5 mM isopropyl β-D-1 thiogalactopyranoside (IPTG), and cells were grown for an additional 16 h at 18 degrees C. Cells were harvested by centrifugation, subjected to a freeze-thaw cycle, and lysed by sonication in 50 mM Tris buffer, pH 8.0, containing 500 mM NaCl, 5% glycerol, and 2 mM tris(2-carboxyethyl)phosphine (TCEP), supplemented with EDTA-free protease inhibitor tablets (Roche).

NRARP protein was captured from cleared lysates by affinity chromatography using HisPur Ni-NTA resin (Thermo Scientific). The His-SUMO tag was cleaved using Ulp1 protease, releasing NRARP from the beads. NRARP was then purified by anion exchange chromatography on Mono-Q resin using a linear gradient of NaCl (0.05-1 M) in 20 mM Tris buffer, pH 8.0,
containing 1 mM EDTA and 5 mM DTT. Fractions containing NRARP were combined, concentrated and further purified by size-exclusion chromatography using a Superdex 200 column equilibrated in 20 mM Tris buffer, pH 8.5, containing 500 mM NaCl, 5% glycerol and 2 mM TCEP. Fractions that were >95% pure as assessed by SDS-PAGE were pooled, flash frozen, and stored at -80°C.

RBPJ molecules (9-452 and 9-435) were expressed and purified using the same series of chromatographic steps. The only modification was the buffer used for anion exchange on Mono-Q resin. Because RBPJ is not stable in low-salt buffer, a linear gradient of NaCl from 0.1-1 M was used in 20 mM Tris buffer, pH 8.3, containing 5 mM DTT. MAML1, Notch1 ANK and Notch1 RAMANK proteins were expressed and purified as previously reported (34).

**Crystallization and Data collection**

The components used to generate protein complexes for crystallography were full-length human NRARP, residues 1760-2126 of human Notch1 (comprising the RAM and ANK domains) and residues 9-452 of human RBPJ. Complexes were purified by size exclusion chromatography using a Superdex 200 column equilibrated in 10 mM Hepes buffer, pH 7.8, containing 150 mM NaCl and 2 mM TCEP. A 16-mer DNA duplex containing an RBPJ binding site with two nucleotide overhangs was generated by annealing oligonucleotides (5’-TTGACTGTGGGAAAGA-3’ and 5’-AATCTTTCCCACAGTC-3’) at 95 degrees C for 5 min and slowly cooling to room temperature. The protein complex was combined with the DNA duplex in a stoichiometric ratio of 1:1.1. Crystals of protein-DNA complex (4 mg/ml) grew in sitting drops at 16 degrees C after 24-36 hours in 50 mM Hepes pH 6.8, 200 mM Sodium
Fluoride, 18% PEG3350, Crystals were cryoprotected by supplementing the mother liquor with 20% ethylene glycol (v/v). Data were collected at the Advanced Photon Source, beamline 23-ID-B (GM/CA).

Structure Determination

The structure was solved by molecular replacement in Phenix with Phaser (35), using RBPJ and the NOTCH1 ANK domain from the human NTC structure, PDB 2F8X (26). The presence of density for DNA and for the RAM portion of NOTCH1 confirmed that the molecular replacement solution was correct. Two NRARP-NICD1-RBPJ-DNA complexes were identified in the asymmetric unit. Iterative rounds of manual model building and refinement were performed with COOT (36) and phenix.refine (37), respectively. All crystallographic data processing, refinement, and analysis software was compiled and supported by the SBGrid Consortium (38). PyMOL (Schrodinger) was used to prepare all Figures as well as perform structural superpositions. Coordinates have been deposited in the protein data bank with PDB ID code 6PY8.

Reporter assay

NIH 3T3 cells cultured in 24-well plate format were transfected with pcDNA3-based plasmids expressing NICD1 (100 ng), a varying amount of NRARP (as indicated in Figures 1 (0.5-2 µg) and 5 (500 ng)), a firefly luciferase reporter plasmid under control of the TP1 Notch response element, and a Renilla luciferase control plasmid. Cells were harvested 24 hours after transfection. Firefly luciferase activity, relative to Renilla control, was measured using a Dual
Luciferase assay kit (Promega). Data were normalized to the signal from 3T3 cells transfected with empty pcDNA3 vector, which was set to a value of 1.

Cell growth assay
MigR1 retroviruses encoding NRARP followed by GFP under control of an internal ribosome entry site, a MAML1(13-74)-GFP fusion protein, or GFP alone were used at titers where only a subpopulation of cells was infected. The fraction of GFP-positive cells was monitored daily by flow cytometry using a BD Accuri C6 Plus flow cytometer. The fraction of GFP positive cells was plotted as a function of time, normalized to the maximum fraction of GFP positive cells (typically observed at 72 - 96 hours after infection).

Quantitative PCR
Total RNA was recovered from Jurkat cells using an RNAeasy Mini Kit (Qiagen), and cDNA was prepared using an iScript cDNA synthesis kit (BioRad). qPCR for NOTCH1, HES1, HES4 and DTX1 was carried out using a modified Quant Studio 6 instrument (Life Technologies). Gene expression was normalized to 28s rRNA as an endogenous control.

Proteomics Studies: Tandem Affinity Purification
Jurkat cells were transduced with a retrovirus encoding NRARP with an N-terminal tandem HA-FLAG tag and a cassette encoding the interleukin-2 receptor (IL2R) after an internal ribosomal entry site (IRES) (pOZ-FH-N) (39). Stable cell lines were generated by magnetic sorting for IL2R-positive cells. For proteomic studies, the resulting cell line was grown in RPMI with 10% bovine growth serum to a density of 3x10^6 cells/ml. Cells were harvested by centrifugation and
lysed with a Dounce homogenizer. The resulting lysate was immunoprecipitated using anti-FLAG-conjugated agarose beads (Sigma) in 50 mM Tris buffer, pH 7.5, containing 150 mM NaCl, 1 mM EDTA, 0.5% NP40, 2 mM TCEP and 10% glycerol, supplemented with EDTA-free protease inhibitor tablets (Roche). The beads were washed three times and the immunoprecipitated protein was eluted with Flag peptide. The eluate was then re-immunoprecipitated with anti-HA-conjugated agarose beads and eluted from the beads with HA peptide for mass spectrometry analysis.

**Mass spectrometry**

Protein complexes isolated by tandem affinity purification (40) were directly processed in solution: Cysteine residues were first reduced with 10 mM dithiothreitol for 30 minutes at 56°C in the presence of 0.1% RapiGest SF (Waters, Milford, MA) and then alkylated with 22.5 mM iodoacetamide for 20 minutes at room temperature in the dark. Proteins were digested overnight at 37°C using 2.5 micrograms of trypsin after adjusting the pH to 8.0 with Tris.

RapiGest SF was cleaved for 30 minutes at 37°C and its by-products were removed by centrifugation. Tryptic peptides were desalted by batch-mode reverse phase solid phase extraction (Poros 10R2) and concentrated in a vacuum concentrator. Peptides were solubilized in 25% acetonitrile containing 0.1% formic acid and further purified by strong cation exchange (Poros 10HS). Peptides were eluted with 25% acetonitrile containing 0.1% formic acid and 300 mM potassium chloride. Acetonitrile was removed using a vacuum concentrator and peptides were reconstituted with 20 µL of 0.1% TFA.
Peptides were loaded onto a precolumn (4 cm POROS 10R2, Applied Biosystems) and eluted with an HPLC gradient (NanoAcquity UPLC system, Waters; 5%–40% B in 90 min; A = 0.2 M acetic acid in water, B = 0.2 M acetic acid in acetonitrile). Peptides were resolved on a self-packed analytical column (50 cm Monitor C18, Column Engineering) and introduced in the mass spectrometer (QExactive HF mass spectrometer, Thermo, Waltham, MA) equipped with a Digital PicoView electrospray source platform (New Objective, Woburn, MA) (41).

The mass spectrometer was operated in data dependent mode where the top 10 most abundant ions in each MS scan were subjected to high energy collision induced dissociation (HCD, 30% normalized collision energy) and subjected to MS2 scans (isolation width = 1.6 Da, intensity threshold = 2e5). Dynamic exclusion was enabled with an exclusion duration of 15 seconds. ESI voltage was set to 3.8 kV.

MS spectra were recalibrated using the background ion (Si(CH3)2O)6 at m/z 445.12 +/- 0.03 and converted into a Mascot generic file format (.mgf) using multiplierz scripts (PMID: 19333238; PMID: 19874609). Spectra were searched using Mascot (version 2.6) against three appended databases consisting of: i) human protein sequences (downloaded from RefSeq on 11/19/2010); ii) common lab contaminants and iii) a decoy database generated by reversing the sequences from these two databases. For Mascot searches, precursor tolerance was set to 15 ppm and product ion tolerance to 25 mmu. Search parameters included trypsin specificity, up to 2 missed cleavages, fixed carbamidomethylation (C, +57 Da) and variable oxidation (M, +16 Da). Spectra matching to peptides from the reverse database were used to calculate a global false discovery rate and were discarded. Data were further processed to remove peptide spectral matches (PSMs)
to the forward database with an FDR greater than 1.0%. Peptides shared by two or more genes were excluded from consideration when constructing the final protein list. Any protein identified in more than 1% of 108 negative TAP controls or any of the negative control TAP experiments (PMID: 22810586) was removed from the sets of interactors.

**In vitro biotin pull-down assays**

Biotin pull-down assays were performed using streptavidin-conjugated agarose (Thermo Fisher) in 20 mM HEPES buffer, pH 7.6, containing 150 mM NaCl, 2 mM TCEP and 0.2% Tween-20. Purified recombinant biotinylated-NRARP, RBPJ and NICD1 (ANK or RAM-ANK) proteins were combined with streptavidin beads at 2 µM and incubated for 30 minutes at room temperature. The beads were washed three times, transferred into gel loading buffer, and the recovered molecules were analyzed by SDS-PAGE using a 4-20% gradient gel followed by staining with Coomassie blue.

**Circular Dichroism Spectroscopy**

Circular dichroism (CD) spectra of wild-type, W85E, and W85E/A92W variants of NRARP were acquired at 20 ºC at a protein concentration of 8 µM on a Jasco J-815 instrument in 10 mM phosphate buffer, pH 7.6, containing 150 mM potassium fluoride and 1 mM DTT. Data were acquired in a 0.1 cm pathlength cell and represent the average of 5 scans taken at a 50 nm/min scan rate with a 0.5-nm step size.

**Quantification and Statistical Analysis**
Bar graphs display mean ± SD. P values were calculated by one-way ANOVA followed by post hoc Dunnett’s multiple comparison tests where applicable using GraphPad Prism (version 8.0).

**NICD1 stability analysis**

Jurkat cells were infected with control retrovirus expressing GFP only, virus expressing dnMAML1-GFP, or virus expressing NRARP followed by GFP under control of an IRES and grown for 96 hours at 37 C. GFP positive cell populations were then sorted to isolate the GFP positive cells from each virally infected population. Sorted cells were grown for 72 hours, lysed on ice, and probed for total Notch1, activated Notch1 (NICD1), and GAPDH with anti-Notch1 (anti-TAD; Weng et al., 2003), anti-V1744 (Cell Signaling antibody D3B8), and anti-GAPDH (Cell Signaling antibody D16H11) antibodies, respectively.

**Supplementary Materials**

figure 3.S1. NRARP inhibition of reporter gene induction by all four human Notch receptors.

figure 3.S2. Electron density and unit cell features.

table S1. List of NRARP-associated proteins detected using tandem-affinity purification followed by mass spectrometry.
References


Table 1. Tandem immunoprecipitation and mass spectrometry results from Jurkat cells using tagged NRARP as bait. The number of unique endogenous peptides recovered for components of the core Notch transcription activation complex is listed for 4 independent experiments.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Recovered peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOTCH1</td>
<td>12; 5; 5; 9</td>
</tr>
<tr>
<td>RBPJ</td>
<td>10; 8; 5; 6</td>
</tr>
<tr>
<td>MAML1</td>
<td>10; 3; 2; 5</td>
</tr>
</tbody>
</table>
Table 2. Data collection, structure determination and refinement statistics.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>NRARP-NTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.979</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>48.5 - 3.75 (3.89 - 3.75)</td>
</tr>
<tr>
<td>Space group</td>
<td>P2_{1}2_{1}2_{1}</td>
</tr>
<tr>
<td>Unit cell (Å, degrees)</td>
<td>79.88, 103.65, 301.42, 90, 90, 90</td>
</tr>
<tr>
<td>Total reflections</td>
<td>114,867 (9,192)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>26,271 (2,504)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>4.4 (3.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.17 (97.13)</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>7.39 (1.16)</td>
</tr>
<tr>
<td>Wilson B-factor</td>
<td>114.5</td>
</tr>
<tr>
<td>R-merge</td>
<td>0.174 (1.155)</td>
</tr>
<tr>
<td>R-meas</td>
<td>0.1973 (1.343)</td>
</tr>
<tr>
<td>CC_{1/2}</td>
<td>0.998 (0.689)</td>
</tr>
<tr>
<td>Reflections used in refinement</td>
<td>26,221 (2,502)</td>
</tr>
<tr>
<td>Reflections used for R-free</td>
<td>1,997 (190)</td>
</tr>
<tr>
<td>R-work</td>
<td>0.27 (0.36)</td>
</tr>
<tr>
<td>R-free</td>
<td>0.32 (0.40)</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms</td>
<td>13,310</td>
</tr>
<tr>
<td>Protein residues</td>
<td>1,551</td>
</tr>
<tr>
<td>RMS (bonds, Å)</td>
<td>0.002</td>
</tr>
<tr>
<td>RMS (angles, degrees)</td>
<td>0.58</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
<td>92.4</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
<td>7.3</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
<td>0.3</td>
</tr>
<tr>
<td>Rotamer outliers (%)</td>
<td>0</td>
</tr>
<tr>
<td>Clash score</td>
<td>10.1</td>
</tr>
<tr>
<td>Average B-factor (Å²)</td>
<td>119.4</td>
</tr>
</tbody>
</table>

*Highest shell statistics are reported in parentheses.*
CHAPTER 4

Conclusions and Future Directions
4.1 Introduction

In this dissertation, I have presented studies of the Notch signaling pathway, in which the interactions between different receptors-ligand pairs achieve a wide array of biological outcomes. Work presented in this dissertation has shown that NOTCH1 and NOTCH2 can discriminate between ligands with a high degree of structural similarity, and that this discrimination relies on the ectodomains of these proteins, not on their intracellular regions. Additionally, I have revealed the molecular basis for the function of a NOTCH pathway regulator that modulates the activity and/or stability of the NICD following receptor activation. This new understanding of the pathway at the levels of signal initiation and termination will allow us to develop better tools – such as antibodies and/or small molecules – allowing for improved manipulation of specific in vivo ligand functions, and for manipulating the overall Notch response in biotechnological and pharmaceutical applications.

In my work on the mammalian Delta paralogues, DLL1 and DLL4, I have shown that in vitro, these molecules are discriminatory in their activation of NOTCH1 and NOTCH2. The functional divergence of these ligands was not simply explained by the intracellular portion of the molecules as proposed by recent work in the Notch signaling field (1). The discriminatory activity of the receptors in responding to these ligands is tied to the ectodomains of the proteins.

This finding is consistent with structural studies of the Jagged family ligand Jagged1 in complex with NOTCH1 which showed a more extensive ligand-receptor interface than is reported in the model of the DLL4-NOTCH1 complex, which was determined using smaller fragments of both proteins. A feature of the more extensive Jag1-NOTCH1 structure is an additional contact site at EGF3 of the ligand with EGF8 of the receptor. In NOTCH1 signaling assays, stimulation by DLL4 needed the MNNL domain as well as the DSL domain and EGF
repeats 1-3 to induce strong signaling, indicating that interactions outside of the MNNL-DSL region reported in the NOTCH1-DLL4 complex structure are important for function.

The mechanisms by which these additional interactions enable receptor discrimination will be an interesting area of future study. Additionally, the interdomain arrangement is different when the structures of the Jagged1 apoprotein and receptor-bound molecules are compared, revealing potentially interesting roles for differences in the domain cores, and/or contacts outside of the known DLL4-NOTCH1 interface.

In Chapter 3, I described a set of biochemical and structural studies that conclusively show that NRARP is an attenuator of Notch signaling and reveal the molecular basis for this inhibitory function. NRARP is a known downstream target of Notch signaling and has been shown to inhibit Notch signaling, identifying NRARP as a feedback regulator. As the molecular determinants for NRARP inhibition of Notch signaling had not been deeply explored, we utilized an unbiased proteomics study to identify molecules that interact with NRARP when present at their natural abundance in cells.

We found that among other proteins, NRARP interacted with components of the core Notch transcription activation complex: NOTCH1, RBPJ and MAML1. Biochemical studies with purified recombinant molecules revealed that this recruitment requires the presence of both NOTCH1 and the transcription factor RBPJ. The requirement for this minimal complex indicated that NRARP selectively targets activated NOTCH1 complexes. Excitingly, we observed that enforced expression of NRARP in NOTCH1-mutated T-ALL cell lines resulted in suppression of cell growth, confirming that the mechanism of action for NRARP is still present in this subset of cancers. Additionally, enforced NRARP expression in a NOTCH1 independent T-ALL cell line
resulted in decreased amounts of mRNA of NOTCH responsive genes which was tied to a reduction of activated NOTCH protein levels.

Structural studies of a minimal NRARP-NOTCH1-RBPJ complex bound to DNA showed that NRARP bound to a defined composite NOTCH1-RBPJ surface. The NRARP binding site is proximal to regions of both NOTCH1 and RBPJ that have potential as well as verified post-translational modification (PTM) sites that have been linked to Notch inhibition (2). NRARP binding may be responsible for direct or indirect recruitment of modifying enzymes that target Notch for degradation. Structurally, binding of NRARP introduced local stability into a region of Notch that has some of the verified PTM sites as well as protrusion of the serine/threonine rich C-terminus of RBPJ. These structural features were not previously observed in NTC structures lacking NRARP. The mechanism by which NRARP binding results in NOTCH inhibition will be an interesting area of future study and the above observations may hint at a role for NRARP in substrate presentation for PTM.

The results presented here suggest the need for further study into: 1) the mechanism of NOTCH inhibition by the pathway feedback inhibitor NRARP as well as 2) receptor recognition by ligands in order better explain distinct biological functions of different ligand-receptor pairs, especially the context dependence non-equivalency of closely related ligands. In this chapter, I will describe potential future experimental studies that could extend the scientific goals of this thesis.
4.2 Future Directions

4.2.1 Further study of functional divergence of DLLs

The analysis of functional divergence for Delta-like ligands in binding to NOTCH receptors raises questions of how these effects arise, and if there are ways to better engineer tools to modulate ligand specific functions. In the work presented in Chapter 2, we examined the contribution of individual domains to functional differences of the two canonical DLLs, DLL1 and DLL4. DLLs have a core interaction mediated by the MNNL and DSL domains with likely secondary interactions mediated by the EGF repeats. Analysis of the contributions of individual domains using domain swapping experiments between the ligands show that EGF repeats are the most critical determining factors in functional divergence. For example, swapping the EGF repeats of DLL4 with those from DLL1 removes the intrinsic preferential activation of NOTCH1 versus NOTCH2 by this ligand despite the MNNL-DSL portion being unaltered. Interestingly, others reported (3) only a modest fold-change (1.3) in affinity of an extended Dll4 construct for the NOTCH receptor fragment EGF8-12 vs EGF10-12, suggesting that the signaling efficiency of more extensive molecules may be linked to biochemical or biophysical properties outside of affinity changes.

Structural elucidation of a much larger DLL1/4-NOTCH complex would give key insights into the nature of the anticipated interactions of EGF3 with NOTCH receptors. The importance of any additional atomic interactions identified would be analyzed using a similar approach to the study presented in Chapter 2. It is important that the implications of any additional interactions identified from structures of more extensive receptor-ligand complexes be
assessed in larger biological contexts - such as a well characterized tissue or organistic setting - to fully characterize how they affect signaling.

In a more recent structural study of the Jag1-NOTCH1 complex, conformational changes in Jag1 domains and force-probe studies revealed a catch bond in NOTCH1 engagement (3). This catch bond property was also observed in the DLL4-NOTCH1 interaction. It will be interesting to see if this property is preserved in all canonical ligands and to what extent catch bond behavior influences the interaction of different receptor-ligand pairs and the resulting NOTCH response.

4.2.2 Mechanism of action of NRARP-mediated NOTCH inhibition

The direct interaction of NRARP with NTCs indicates that the mechanism of action for NOTCH inhibition likely involves the recruitment of post-translational modifiers and/or disruption of recruitment of the transcription machinery, given the lack of any enzymatic property of NRARP. The structure we elucidated of the NRARP-NOTCH1-RBPJ complex revealed a distinct composite binding surface containing the N-terminus of the NOTCH1 ankyrin repeat domain and the C-terminus of RBPJ. The location of the binding site could indicate allosteric modulation of protein-protein interactions given the presence of multiple PTM sites – both predicted and verified – in this region of both NTC components (4).

In order to test this hypothesis about the recruitment of protein modifiers that alter the stability and/or function of the NTC, one would need to determine the in vivo interacting molecules. In chapter 3 of this thesis, the unbiased proteomics approach I used to identify stable
interactors of NRARP was fundamentally restricted in its ability to capture transient or weak interactions such as those one may find in protein modifiers. In our initial proteomics study, I found that the enriched NRARP interactors not only included the core NTC components but also several protein modifiers including kinases and ubiquitin ligases. To confirm this finding and expand this work to identify any transient interactor, one approach likely to be productive would be to perform peroxidase-catalyzed proximity labeling combined with mass tag labelling and mass spectrometry to enable quantitative and temporal insight into NRARP interactions. These studies would be performed in Jurkat cells as previously done in chapter 3. Treatment with a gamma secretase inhibitor (GSI) followed by washout of the inhibitor would effectively control NOTCH activation and thus enable temporal resolution. The biotin ligase APEX2 would be fused to the N-terminus of NRARP: based on structural and biochemical data this region of the molecule should be amenable to protein modification.

Proximity labelling results would be analyzed to look for changes in the biotinylated proteome upon NOTCH pathway inhibition following GSI treatment. NRARP-mediated inhibition of NOTCH appears to be linked to accelerated degradation of the NICD. Transient molecular interactors – specifically protein modifiers such as kinases - that function in the inhibition of NOTCH should therefore exhibit peak biotinylation prior to a decrease in the abundance of NOTCH while stable interactors that are directly mediating degradation should have a labelling pattern that parallels NOTCH labelling. Chemical and/or genetic inhibition of identified interactors that are of interest – based on the scenarios described above - could then be used to assess the necessity for NRARP-mediated inhibition in cellular contexts where NRARP function has been verified.
NRARP partiality for binding NOTCH transcription complexes – as is evident from the requirement of RBPJ for binding - provides an additional level of potential complexity in NOTCH signaling. NTCs could use NRARP binding to direct interactions with different co-factors at RBPJ binding motifs, potentially adding complexity and nuance to gene regulatory networks under extended NOTCH signaling control. Additionally, the affinity of NRARP for the different paralogs of NOTCH is likely different – we have observed weaker interaction with NOTCH4 complexes - meaning whatever the functional outcome of NRARP is in NOTCH signaling there could be intrinsic bias for different receptor ICDs. In such a scenario, the NRARP-mediated effect in cellular contexts where multiple receptors are present is interesting to consider especially if different ICDs are generating distinct responses. While the idea that different NOTCH ICDs are endowed with distinct functions is still a hypothesis, there is some indication that NOTCH1 and NOTCH2 ICDs can yield distinct outputs depending on the cellular context. For example, NOTCH1 and NOTCH2 have different effects in medulablastoma development and similarly in bladder cancer, where in both cases NOTCH2 acts as an oncogene whereas NOTCH1 shows a tumor suppressor phenotype (5, 6).

Unfortunately, the specifics of co-factor interactions in transcription machinery recruitment for canonical NOTCH signaling is still an area where there is a surprising paucity of knowledge. As a result, examining the effect of NRARP on NOTCH nuclear activity will require an extensive study. The work presented in this thesis showed conclusively that NRARP interacts directly with NTCs. The effect of NRARP binding, however, on productive normal NOTCH nuclear interactions in cells remains unclear. Does NRARP binding interrupt canonical NOTCH
nuclear interactions? Additionally, do the molecules identified in our NRARP interactome studies represent components of larger alternative NOTCH nuclear complexes? Studies using proximity labeling combined with mass spectrometry can be used to address these questions. In such studies, endogenous NOTCH ICD would be fused to a suitable biotin ligase and using temporal control afforded by pathway inhibition – GSI treatment and washouts is a proven method to control NOTCH pathway activation – we can analyze difference in protein interactors as assessed by presence of biotin modification via mass spectrometry. In this approach the choice of cell line will be important, I propose we use Jurkat cell lines that are parental, NRARP−/− and stably express NRARP, which we’ve already created. As a reference, this approach has been used for multidimensional tracking of GPCR signaling (7).

4.3 Concluding Remarks

Overall, the findings presented in this thesis document my efforts to understand the molecular basis for 1) functional divergence of the closely related canonical NOTCH activating ligands DLL1 and DLL4 and 2) negative feedback regulation of NOTCH signaling in higher organisms by the NOTCH Regulated Ankyrin Repeat Protein (NRARP). The use of domain swapping mutagenesis and more precise binding site amino acid substitutions in the ectodomain of DLLs highlight the importance of detailed biochemical studies to uncover the true contributions of individual functional domains.

The results presented here show the importance of studies into the mechanism by which protein-protein interactions can elicit distinct responses and which features of proteins can affect functional divergence especially in molecules that have a high degree of similarity. Our
observations highlight the fact that functional outcome is not simply determined by the direct amino acid contacts in protein-protein interactions but can be modulated through differences in domain cores as well as interdomain properties. The canonical NOTCH ligands DLL1 and DLL4 exhibit receptor selectivity, yet our findings suggest that this property is not determined exclusively by differences in receptor binding affinity. Physical properties afforded by intra- and interdomain properties – such as catch bond behavior – appear to be major factors that contribute to functional divergence.

In chapter 3, the use of unbiased mass spectrometry-based proteomic studies enabled the discovery of endogenous NRARP interactors, including both direct and potential indirect interacting molecules. This depth of proteomics data has allowed us to develop a more comprehensive model of the NRARP interactome. The accompanying combination of structural studies with functional assays clarifies how NRARP binds to RBPJ-NOTCH complexes and opens new avenues for acquiring a deeper understanding of the molecular mechanism of action of a critical negative feedback mechanism that regulates Notch signaling.
References


Appendix 1

The ectodomains determine ligand function in vivo and selectivity of DLL1 and DLL4 toward NOTCH1 and NOTCH2 in vitro

This appendix has been published as:

The ectodomains determine ligand function in vivo and selectivity of DLL1 and DLL4 toward NOTCH1 and NOTCH2 in vitro

Lena Tveriakhina1, Karin Schuster-Gossler1, Sanchez M Jarrett2, Marie B Andrawes2†, Meike Rohrbach1, Stephen C Blacklow2,3*, Achim Gossler1*

1Institute for Molecular Biology, Medizinische Hochschule Hannover, Hannover, Germany; 2Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts; 3Department of Cancer Biology, Dana Farber Cancer Institute, Boston, Massachusetts

Abstract DLL1 and DLL4 are Notch ligands with high structural similarity but context-dependent functional differences. Here, we analyze their functional divergence using cellular co-culture assays, biochemical studies, and in vivo experiments. DLL1 and DLL4 activate NOTCH1 and NOTCH2 differently in cell-based assays and this discriminating potential lies in the region between the N-terminus and EGF repeat three. Mice expressing chimeric ligands indicate that the ectodomains dictate ligand function during somitogenesis, and that during myogenesis even regions C-terminal to EGF3 are interchangeable. Substitution of NOTCH1-interface residues in the MNNL and DSL domains of DLL1 with the corresponding amino acids of DLL4, however, does not disrupt DLL1 function in vivo. Collectively, our data show that DLL4 preferentially activates NOTCH1 over NOTCH2, whereas DLL1 is equally effective in activating NOTCH1 and NOTCH2, establishing that the ectodomains dictate selective ligand function in vivo, and that features outside the known binding interface contribute to their differences.

DOI: https://doi.org/10.7554/eLife.40045.001

Introduction

The Notch signaling pathway mediates communication between neighboring cells in metazoans and thereby regulates a multitude of developmental processes in various tissues (Artavanis-Tsakonas et al., 1995; Yoon and Gaiano, 2005; Bolós et al., 2007; Gridley, 2007; Radtke et al., 2010; Koch and Radtke, 2011; reviewed in Louvi and Artavanis-Tsakonas, 2012; Kopan, 2012). This communication depends on the interaction of Notch receptors on the surface of the signal receiving cells with transmembrane ligands on the surface of adjacent cells. Ligand binding then leads to a sequence of proteolytic cleavages of the receptor releasing the Notch intracellular domain (NICD) from the membrane. NICD translocates into the nucleus where it enters into a complex with a CSL protein (CBF-1/RBPJ in mammals, Suppressor of Hairless in flies, and Lag-1 in worms) and a protein of the Mastermind family (Petcherski and Kimble, 2000; Wu et al., 2000; Nam et al., 2003; Nam et al., 2006; Wilson and Kovall, 2006; Choi et al., 2012) to regulate transcription of target genes (reviewed in Bray, 2016).

Mammals have four Notch receptors (N1-N4) and four activating ligands of the DSL (Delta, Serrate, Lag-2) family: DLL1 and DLL4, orthologs of Drosophila Delta, and JAG1 and JAG2, orthologs of Drosophila Serrate. DLL1 and DLL4 are similar in domain structure, size and sequence (Shutter et al., 2000). Both proteins contain an N-terminal MNNL (also referred to as C2) domain...
eLife digest A small number of signaling systems control how an animal develops from a single cell into a complex organism made up of many different cell types. Signals pass back and forth between cells, switching genes on and off to direct the development of tissues and organs. One of these signaling systems, called Notch, is so ancient that it appears in nearly all multicellular organisms.

A cell sends a Notch signal using proteins called Delta or Jagged ligands that span membrane of the cell, so that part of the protein sits inside the cell and part remains outside. To change the behavior of another cell, the ligands bind to proteins called Notch receptors that span the membrane of the receiving cell.

Mammals have two types of Delta ligand, two types of Jagged ligand and four types of Notch receptor. Cells in different tissues display different combinations of these eight proteins. Two Delta ligands called DLL1 and DLL4 often appear together in developing organisms. Some tissues need both and some only the one or the other. In some cases one ligand can compensate if the other is missing, but in others not. It was not clear why this is, or which parts of the proteins are responsible.

Tveriakhina et al. used mouse cells to investigate how DLL1 and DLL4 interact with two Notch receptors, called NOTCH1 and NOTCH2. The results of these experiments show that while DLL1 can bind and activate both Notch receptors equally, DLL4 prefers to partner with NOTCH1. To find out which parts of the ligands are responsible for this selectivity, Tveriakhina et al. created hybrid ligands that contained a mixture of regions from DLL1 and DLL4. These suggest that the different binding preferences depend on parts of the ligands that sit outside cells and that lie outside the known sites of binding contact with the Notch receptors.

Further experiments studied mice that had been engineered to produce hybrid ligands as replacements for DLL1. A hybrid ligand consisting of the part of DLL1 that sits outside cells and the part of DLL4 found inside cells generated Notch signals in the tissue that depended on the activity of DLL1. However, a hybrid consisting of the part of DLL4 that sits outside cells and the part of DLL1 found inside cells did not, showing that in developing mice the parts that sit outside the cells contribute to the different functions of DLL1 and DLL4.

Overall, the results presented by Tveriakhina et al. show that interactions between specific ligands and receptors play important roles in how mammals develop. Further efforts to understand which parts of the ligands affect selectivity could ultimately allow researchers to develop ways to modify how ligands and receptors interact. Such “molecular engineering” strategies could enable cell responses to be precisely controlled by pairing designer ligand-receptor pairs to develop cell-based therapies.

DOI: https://doi.org/10.7554/eLife.40045.002

(Chillakuri et al., 2013; Suckling et al., 2017), followed by a DSL domain and eight EGF-like repeats in their extracellular portion, and a less well conserved intracellular domain. The MNNL and DSL domains, required for high-affinity binding of Delta-like ligands to Notch receptors (Rebay et al., 1991; Cordle et al., 2008), contact EGF repeats 12 and 11 of Notch, respectively (Luca et al., 2015). Contributions from adjacent EGF-like repeats, however, are required for signal transduction by Delta-like ligands (Andrawes et al., 2013) as well as for optimal interaction with Serrate (Yamamoto et al., 2012) and Jagged (JAG)-family ligands (Luca et al., 2017). Although the biological activities of DLL1 and DLL4 are partially overlapping, the two proteins are not equivalent in vitro or in vivo. In cell culture studies, DLL4 is more effective than DLL1 in activating N1 signaling during T cell development (Besseyrias et al., 2007), consistent with its ten-fold higher binding affinity in binding studies using purified fragments of N1, DLL1, and DLL4 (Andrawes et al., 2013). In vivo, studies of adult intestinal epithelium in mice have shown that DLL1 and DLL4 are co-expressed in crypts and act redundantly to maintain the intestinal stem cell pool (Pellegrinet et al., 2011). In contrast, however, mouse DLL1 cannot fully replace DLL4 in its ability to trigger T lineage commitment (Besseyrias et al., 2007; Mohtashami et al., 2010). Conversely, endogenous DLL4 does not substitute for DLL1 in its ability to promote development of the arterial vascular epithelium (Sørensen et al., 2009), nor does it compensate for the function of DLL1 in the paraxial mesoderm,
Figure 1. Schematic representation of DLL1 and DLL4 and variant proteins. I-X, full-length and chimeric ligands generated by domain swaps. XI and XII, ligands with exchanges of the known NOTCH1 contact amino acids in the MNNL and DSL domains. XIII, DLL4 variant with an N109G mutation that eliminates the N-glycosylation site in DLL4. XIV-XVIII, soluble proteins encoding the N-terminal region up to and including EGF5 carrying a C-terminal Avi-His-tag for protein purification. I-XIII were tested in cell-based Notch activation assays, II, III, VII and XI in transgenic mice, XIV-XVIII used for measurements of binding affinities to N1. Proteins analyzed in cell-based assays were C-terminally Flag-tagged, proteins analyzed in mice were untagged. Break points and surrounding amino acid sequences and point substitutions are illustrated in Figure 1—figure supplement 1. Red domains/spikes: DLL1; blue domains/spikes: DLL4; white asterisks: N109G mutation.

ECD, extracellular domain; N, N-terminus; D, DSL domain; E, EGF repeat, TM, transmembrane domain; ICD, intracellular domain; D, DLL; cont, N1 contact amino acids.

DOI: https://doi.org/10.7554/eLife.40045.003

The following figure supplements are available for figure 1:

Figure supplement 1. Amino acid exchanges of DLL variant proteins.
DOI: https://doi.org/10.7554/eLife.40045.004

Figure supplement 2. Analysis of ligand receptor binding.
DOI: https://doi.org/10.7554/eLife.40045.005

Figure supplement 3. N109 is highly conserved and N-glycosylated in DLL4.
DOI: https://doi.org/10.7554/eLife.40045.006

a tissue where these ligands are normally not co-expressed: mice in which DLL1 was replaced by DLL4 had severe somite patterning defects and showed premature myogenic differentiation leading to reduced skeletal muscles. However, the function of DLL1 during early retina development was rescued by DLL4 in these mice (Preuß et al., 2015).
Figure 2. The extracellular domains of DLL1 and DLL4 determine ligand behavior during somitogenesis. (A) Scheme of the targeting vector pMP8. CAG-Stop used to introduce inducible chimeric ligands into the Hprt locus, and of Cre-mediated activation of transgene (D1ECD_D4ICD or D4ECD_D1ICD) expression driven by the CAG promotor (CAG prom), 5' hom and 3' hom, Hprt 5' and 3' homology regions; ex1-3 (grey boxes), Hprt exons; neo', neomycin phosphotransferase; pA, polyadenylation signal; hHPRT prom, human Hprt promoter; DLL1/4iresdsRed, chimeric ORF–linked to dsRed tag by an internal ribosomal entry site (IRES). (B) Uncx expression in E9.5 wild type embryos (a, a'; n = 28), embryos lacking DLL1 in the mesoderm (b, b'; n = 12) and male embryos lacking DLL1 in the mesoderm that express either D1ECD_D4ICD (c, c'; n = 9) or D4ECD_D1ICD (d, d'; n = 8) showing that the extracellular domain of DLL1 but not of DLL4 can restore Uncx expression. (C) Whole mount immunofluorescent staining of wild type (a–c) and D4ECD_D1ICD/Y;T(s):Cre (d–f) PSMs using antibodies recognizing the extracellular domain of DLL4 showing co-localization of the exogenous chimeric ligand with pan-Cadherin (panCad) at the cell surface. Additional intracellular staining most likely reflects the presence of the ligand in the ER and trans Golgi as observed previously for DLL1 in cultured cells (Geffers et al., 2007; Müller et al., 2014) and for endogenous DLL1 and transgenic DLL4 in the PSM (Preuße et al., 2015). n = 3 for wild type, n = 4 for D4ECD_D1ICD/Y;T(s):Cre; Scale bar = 10 mm.

DOI: https://doi.org/10.7554/eLife.40045.007

Collectively these studies indicate that the functionality of DLL1 and DLL4 strongly depends on context, but it remains unclear which portions of these similar DSL proteins account for their functional non-equivalence. A recent study in cell culture observed that DLL1 and DLL4 stimulate NOTCH1 receptors to produce responses with different dynamics, attributing differences between pulsatile signaling of DLL1 and sustained signaling by DLL4 to the intracellular, rather than the extracellular, regions of the proteins (Nandagopal et al., 2018). Here, we investigate the influence of the extracellular and intracellular regions of DLL1 and DLL4 chimeric proteins on ligand function in cell culture assays, and for selected chimeras, in biochemical binding assays and in vivo in mice.
observe that in vivo differences of DLL1 and DLL4 function during somite patterning and myogenesis are encoded by the ligands ectodomains, that DLL1 and DLL4 are able to discriminate between NOTCH1 and NOTCH2 in vitro, and that ligand residues outside of the known binding interface are important contributing factors for ligand function in vivo.

**Results**

**The extracellular domain dominates ligand function during somitogenesis**

Previous in vivo analyses indicated that DLL4 cannot substitute for DLL1 function during embryonic development (Preuße et al., 2015). To test whether the inability of DLL4 to rescue the loss of DLL1 in the paraxial mesoderm in vivo resides in its extra- or intracellular domain we generated single copy transgenic mice allowing for the conditional expression of chimeric DLL molecules consisting of the extracellular domain of one ligand and transmembrane and intracellular domain of the other (D1ECD_D4ICD and D4ECD_D1ICD, II and VII in Figure 1). Transgenes were introduced into the Hprt-deficient E14TG2a ES cells by homologous recombination using the strategy already employed for the initial analysis of full length DLL1 and DLL4 during somitogenesis (Preuße et al., 2015). Briefly, cDNAs encoding chimeric ligands were cloned into the targeting vector pMP8 in reverse orientation downstream of neomycin phosphotransferase (neo') driven by the CAG promoter. Cre-mediated recombination of two loxP sites and two mutant loxP2272 (loxM) sites removes the neo' cassette and flips the gene of interest and results in its expression from the CAG promoter (Figure 2A).

To test whether the extracellular or intracellular domain determines the inability of DLL4 to rescue the loss of DLL1 in mesodermal tissues of early embryos, we induced expression of either chimeric ligand and simultaneously removed endogenous DLL1 using a floxed Dll1 allele and a Cre transgene expressed in the primitive streak driven by a promoter derived from brachyury (T(s):Cre) (Feller et al., 2008). Because the Hprt locus is located on the X-chromosome, we used hemizygous male embryos for the analysis. As previously described, inactivation of DLL1 in the mesoderm resulted in loss of Uncx (formerly called Uncx4.1) expression in caudal somite compartments (n = 12; Figure 2Bb, a) indicating severe somite patterning defects compared to wild type embryos (n = 28; Figure 2Ba, a). Expression of D1ECD_D4ICD in Dll1-deficient embryos (n = 9) restored robust expression of Uncx similar to full length DLL1 (Preuße et al., 2015). Uncx expression expanded into cranial somite compartments (Figure 2Bc, c) reminiscent of ectopic Notch activity (Feller et al., 2008), probably reflecting non-restricted D1ECD_D4ICD expression throughout the PSM and somites. In contrast, expression of D4ECD_D1ICD barely restored Uncx expression in the majority (n = 8/12) of Dll1-deficient embryos (Figure 2Bd, d), a phenotype similar to that seen with full-length DLL4 (Preuße et al., 2015), even though the chimeric ligand was expressed and detected on the cell surface of PSM cells (Figure 2Cd-f). As observed previously for full-length DLL4 (Preuße et al., 2015) some embryos (n = 4) displayed essentially normal Uncx expression (not shown), which might result from some perdurance of DLL1 activity or delayed or inefficient excision of endogenous Dll1. Overall, this analysis strongly suggests that the functional difference between DLL1 and DLL4 observed in vivo during somitogenesis resides in the extracellular domains.

**Regions outside the known receptor binding domain are essential for full DLL1 function in vivo**

The N-terminal MNNL and DSL domains and adjacent EGF repeats 1–3 constitute the major interface for interaction between DSL ligands and Notch receptors, and are essential for (full) activation of Notch signaling (Cordle et al., 2008; Andrawes et al., 2013; Luca et al., 2015; Schuster-Gossler et al., 2016; Luca et al., 2017). To analyze whether this region accounts for the observed differences between DLL1 and DLL4 in vivo we generated a chimeric ligand that contained the N-terminal region up to and including EGF3 of DLL1 fused to EGF4 and the remaining C-terminal portion of DLL4 (D1N-E3_D4, III in Figure 1; the amino acid sequence around the fusion is shown in Figure 1—figure supplement 1Aa). We then tested whether this chimeric ligand is sufficient for normal DLL1 function during development. We generated mice (Dll1D1N-E3_D4_h) expressing D1N-E3_D4 instead of DLL1 using the "mini-gene" knock-in strategy (Figure 3A) that disrupts endogenous Dll1,
Figure 3. D1N-E3_D4 is not able to compensate for DLL1 function during somitogenesis. (A) "Mini-gene" targeting strategy to express DLL1 or DLL4 variants from the Dll1 locus (a) and alleles generated in this study (d and e). The Dll1(Dom) (b) and Dll1(Dom) (c) control alleles were described previously (Preuße et al., 2015; Schuster-Gossler et al., 2016). Dll1(D1N-E3_D4D) (d) encodes a fusion protein between the N-terminal part of DLL1 including EGFR3 fused to EGFR4 and the remaining C-terminal portion of DLL4 (III in Figure 1 and Figure 1—figure supplement 1Ab). Dll1(D1N-E3_D4D) (e) encodes a

Figure 3 continued on next page
variant whose predicted amino acids of the MNNL and DSL domains that contact N1 are replaced by the corresponding amino acids of DLL4 (XI in Figure 1, Figure 5C, and Figure 1—figure supplement 1B). All alleles have an identical structure and intron 9 and 10 sequences ofDll1. (B) External phenotypes of wild type (a; n = 19), homozygous Diff10^{DIN,E3}_D4ki (b; n = 11), Diff10^{DIN} (c; n = 3) and Diff10^{DOM} (d; n = 3) control E15.5 fetuses. Arrow in f points to the short tail. Arrowhead in (c) points to edemas present in homozygous Diff10^{DOM} fetuses. (C) Indirect immunofluorescence staining of wild type (a–c, j–l), homozygous Diff10^{DIN,E3}_D4ki (d–f, m–o), and homozygous Diff10^{DOM} (g–i, p–r) E9.5 PSMs using antibodies recognizing the extracellular domain of DLL4 (a, d, g) and DLL1 (j, m, p) and pan-Cadherin (panCad; b, e, h, k, n, q) showing expression of D1N-E3_D4 and co-localization with the cell surface marker pan-Cadherin. Staining of D1N-E3_D4 appears weaker than DLL4 most likely because much of the epitope recognized by the polyclonal anti-DLL4 antibody is missing in this chimeric protein. n ≤ 3; Scale bar = 5 mm. (D) Wish of E9.5 embryos showing that D1N-E3_D4 does not restore normal Uncx expression (e; n = 10) resembling the Diff10^{DOM} phenotype (d; n = 7). (E) Skeletal preparations of wild type (a; n = 11), homozygous Diff10^{DIN} (b; n = 6), heterozygous (c; n = 14/16) and homozygous (d; n = 3) Diff10^{DOM}, and heterozygous (e; n = 14) and homozygous (f; n = 10) Diff10^{DIN,E3}_D4ki/E15.5 fetuses. Arrow and arrowheads in (c) point to axial skeleton defects that were not detected in Diff10^{DIN,E3}_D4ki heterozygotes (e). (F) Cross-sections of hind limbs of wild type (a), homozygous Diff10^{DEN} (b), homozygous Diff10^{DOM} (c), and homozygous (d–f; n = 3) Diff10^{DIN,E3}_D4ki/E18.5 fetuses stained for expression of Myosin Heavy Chain (MHC) indicating that D1N-E3_D4 rescues the skeletal muscle phenotype in contrast to DLL4. Arrows in (c) point to skeletal muscle remnants.

DOI: https://doi.org/10.7554/eLife.40045.008

successfully employed previously to express either a Dll4 orDll1 (control) mini-gene (Schuster-Gossler et al., 2007; Preuße et al., 2015; Schuster-Gossler et al., 2016). Heterozygous mice obtained from two independent targeting events carrying the Diff10^{DIN,E3,D4}_E3 allele were viable although no apparent phenotype. Homozygous Diff10^{DIN,E3,D4}_E3 mice were stillborn (n = 3 and 4 respectively), indicating that D1N-E3_D4 cannot fully replace DLL1 during development although it is present on the cell surface of PSM cells (Figure 3Cd–f). At E15.5 homozygous Diff10^{DIN,E3,D4}_E3 fetuses showed a stumpy tail (n = 5 and 6, respectively; arrow in Figure 3Bb) similar to Diff10^{DOM} mutants (Figure 3Bd); however, they lacked the edema observed in Diff10^{DOM} homozygotes (arrow head in Figure 3Bc). D1N-E3_D4 was also not able to restore normal Uncx expression (Figure 3De). Axial skeletons of homozygous Diff10^{DIN,E3,D4}_E3 fetuses were severely disorganized (n = 10; Figure 3Ef), a phenotype consistent with abnormal Uncx expression and similar to Diff10^{DOM} heterozygote axial skeletons (Figure 3Eo), although the rib cage appeared less compressed. In contrast to Diff10^{DOM}, which often displayed axial skeleton defects (n = 14/16) such as hemivertebrae (arrow in Figure 3Ec) and fused ribs (arrowheads in Figure 3Ec) heterozygous Diff10^{DIN,E3,D4}_E3 fetuses showed no defects of the axial skeleton (n = 0/14; Figure 3Ee) indicating that D1N-E3_D4 lacks the dominant interfering activity of DLL4.

Deletion of DLL1 during myogenesis leads to premature differentiation of myogenic progenitor cells resulting in severe skeletal muscle hypotrophy at fetal stages (Schuster-Gossler et al., 2007). This phenotype cannot be suppressed by DLL4 expression (Figure 3Fc; Preuße et al., 2015). In contrast, skeletal muscles of Diff10^{DIN,D4,E3,D4,E3}= Diff10^{DIN,E3}_D4 homozygotes (Figure 3Ff; n = 3) were indistinguishable from Diff10^{DEN,DIN} (Figure 3Fb) and wild type fetuses (Figure 3Fa). These in vivo analyses indicate that, unlike the D1EC_D4CD chimera, D1N-E3_D4 is not a fully functional DLL1 ligand during somite patterning. However, D1N-E3_D4 remains functional during myogenesis and restricts muscle progenitor differentiation despite the presence of the DLL4 ICD, consistent with the conclusion that in vivo the functional difference between DLL1 and DLL4 is encoded in the ECDs.

DLL1 and DLL4 exhibit differential receptor selectivity in vitro

In cell-based trans-activation assays using HeLa cells stably expressing murine N1 (HeLaN1) co-cultured with CHO cells expressing mouse DLL1 (mDLL1) or DLL4 from the same locus both ligands activated a transiently expressed Notch reporter similarly (Preuße et al., 2015). However, a purified fragment of the extracellular domain of human DLL4 (N-terminus up to and including EGF5: hD4N-E5) bound to hN1 with an approximately ten-fold higher affinity than the corresponding hDLL1 fragment (Andrawes et al., 2013). Like hD4N-E5, mD4N-E5 has a higher affinity for hN1 (Kd = 0.43 ± 0.046 mM; Figure 1—figure supplement 2Aa) than the corresponding mDLL1 fragment (Kd = 1.56 ± 0.207 mM; Figure 1—figure supplement 2Ab), as judged by bilayer interferometry measurements. To find a potential explanation for the discrepancy between binding affinities and Notch activation in HeLaN1 cells we analyzed these cells for expression of other Notch receptors and found that in addition to exogenous mouse Notch1 HeLaN1 cells express endogenous NOTCH2.
Figure 4. DLL1 and DLL4 differentially activate NOTCH1 and NOTCH2 in cell-based co-culture assays. (A) RT-PCR analysis using RNA of HeLaN1 cells shows the expression of endogenous human NOTCH2 and NOTCH3 in addition to the exogenous murine Notch1. (B) ES cell-based trans-activation assays demonstrate that E14TG2a ES cells express negligible amounts of endogenous Notch receptors and ligands. Co-cultivation of ES cells with DLL1 expressing cells with N1rep ES cells showed a 6–10-fold increase in luciferase activity (VI), whereas co-culture of DLL1 expressing cells with N1rep ES cells showed a 6–10-fold increase in luciferase activity (VI). n ≤ 3 co-cultures with 2–4 replicate measurements per n (Figure 4—source data 1). Mean ± SD, ns = p ≤ 0.05, ****=p < 0.0001, one-way ANOVA followed by Tukey’s multiple comparison test. (C) Protein expression analysis indicating similar expression levels of DLL1 and DLL4 in the ES cell clones used. Each DLL4 value represents a technical replicate, which was referenced to its paired DLL1 value, which was arbitrarily set to one for each measurement. The non-normalized values (DLL/b-Tub ratios) are depicted in a graph in Figure 4—figure supplement 1A (Figure 4—source data 2). (D) Cell-surface biotinylation demonstrating that a slightly higher fraction of DLL4 is present at the cell surface compared to DLL1 (n ≤ 6; Figure 4—source data 3). (E) DLL4 activates N1 about 10-fold more strongly than DLL1 in co-culture assays. Left graph shows non-normalized N1 activation. Lines connect values measured in the same assay. Right graph shows values normalized to DLL1 activation, and corrected for protein expression and cell surface presentation. (F) DLL4 activates N2 about half as strongly as does DLL1. Left graph shows non-normalized N2 activation. Lines connect values measured in the same assay. Right graph shows values normalized to DLL1 activation, and corrected for protein expression and cell surface presentation. Each dot represents a technical replicate. Raw data are shown in Figure 4—source data 4 and Figure 4—source data 5. Co-cultures (n = 39) with two replicate measurements per n. Mean ± SD, ns = p ≤ 0.05, ****=p < 0.0001, Student’s paired t-test.

DOI: https://doi.org/10.7554/eLife.40045.009

The following source data and figure supplements are available for figure 4:

Source data 1. Raw data used to generate the graph in Figure 4B.
DOI: https://doi.org/10.7554/eLife.40045.013

Source data 2. Data used to generate the graphs in Figure 4C and Figure 4—figure supplement 1A.

Figure 4 continued on next page
and NOTCH3 (Figure 4A), which might have masked underlying differences in the intrinsic N1 response to the DLL1 and DLL4 ligands.

To detect potential differences in ligand activity towards N1 or N2, the two Notch receptors present during somitogenesis, and to reduce variability due to transient reporter expression we stably integrated a Notch luciferase reporter in the Hprt locus (E14rep) of mouse E14TG2a ES (E14) cells, and generated stable cell lines expressing either Notch1 (N1rep) or Notch2 (N2rep) in these cells (Schuster-Gossler et al., 2016). When co-cultured with E14 cells or DLL1 or DLL4 expressing cells, E14rep cells show luciferase activity similar to wild type E14 levels (compare I to II, III, and IV in Figure 4B; numerical values in Figure 4—source data 1), indicating that negligible amounts of functional endogenous NOTCH receptors are present in E14 cells. Likewise, N1rep cells show essentially no activation above background when co-cultured with wild type E14 ES cells (compare V to I in Figure 4B; numerical values in Figure 4—source data 1), indicating insignificant amounts of functional endogenous Notch ligands in these cells. ES cells expressing exogenous DLL1 activate the luciferase reporter approximately ten-fold above the basal signal in E14 ES cells when co-cultured with N1rep cells (compare VI to V in Figure 4B; numerical values in Figure 4—source data 1) indicating that our co-culture system reliably measures specific Notch signaling activity.

To create ligand presenting cells for a comparison between mDLL1 and mDLL4, we generated ES cells expressing either mDLL1 or mDLL4 from single copy integrations into the Hprt locus. Co-cultures (n = 39) of cells expressing DLL1 or DLL4 with N1rep ES cells consistently revealed higher activation of N1 by DLL4 than by DLL1 (mean 12.454 ± 3.961 SD fold of non-normalized luciferase activity, 9.42 ± 2.997 SD fold, when normalized to DLL1 activation and corrected for protein expression and cell surface levels (Figure 4C–E; numerical values Figure 4—source data 2, Figure 4—source data 3, Figure 4—source data 4). In contrast, DLL4 activated N2 significantly less efficiently than did DLL1 (n = 39; mean 0.468 ± 0.161 SD fold of non-normalized luciferase activity, 0.35 ± 0.12 SD fold, when normalized to DLL1 activation and corrected for protein expression and cell surface levels (Figure 4F; numerical values in Figure 4—source data 5). To confirm that the observed differences between DLL1 and DLL4 in activating N1 and N2 were not a secondary consequence of clonal selection (however unlikely), we also analyzed additional DLL1 (n = 3) and DLL4 (n = 9) expressing ES cell clones for protein expression and N1 or N2 activation. Despite some variability of protein expression (Figure 4—figure supplement 1B and Figure 4—figure supplement 1—source data 1) and Notch activation levels between individual clones and co-cultures, all DLL4 clones consistently activated N1 significantly better than all DLL1 clones, and all DLL4 clones stimulated N2 significantly less efficiently than DLL1 (Figure 4—figure supplement 1C,D; numerical values in Figure 4—figure supplement 1—Source Data 2), indicating that both ligands differ significantly in their ability to activate different Notch receptors in our cell-based assay. Consistent with the differences in N2 stimulation by mDLL1- and mDLL4-expressing cells, the highly homologous human hD1N-E5 exhibits a higher affinity (Kₐ = 0.36 ± 0.11 mM; Figure 1—figure supplement 2Ba) for human NOTCH2 than D4N-E5 (Kₐ = 1.28 ± 0.2 mM; Figure 1—figure supplement 2Bb).
Figure 5. Contributions of the MNNL-EGF3 portion and contact amino acids to ligand selectivity towards N1 and N2. (A) N1/N2 activation ratios by DLL1 and DLL4 chimeric proteins show that receptor selectivity of DLL1 and DLL4 is encoded by the extracellular domain and that EGF3 contributes to N1/N2 selectivity. DLL4, DLL4_ECD_D4_ICD, and DLL4_N-E3_D4 show N1/N2 induction ratios of ~20. DLL1, D1_ECD_D4_ICD, and D1N-E3_D4 show N1/N2 induction ratios of ~20. DLL1, D1_ECD_D4_ICD, and D1N-E3_D4 show N1/N2 induction ratios of ~20.
Figure 5 continued

exhibit induction ratios of 1–3. Chimeric pairs with domain exchanges between EGF2 and EGF3 or between DSL domain and EGF1 show equivalent stimulation ratios. Each dot represents the mean of N1 (relative luciferase units; Figure 5—source data 1)/N2 (relative luciferase units; Figure 5—source data 2) of n ≤ 3 measurements per clone of a given ligand construct. Bars represent the Mean ± SD of n ≤ 3 clones per construct (Figure 5—source data 3). (B) Structure-based superposition of DLL1 and DLL4 (PDB ID codes 4XBM and 4XLW, respectively; (Kershaw et al., 2015; Luca et al., 2015). Top panel: NOTCH1 is rendered as a molecular surface (wheat), and DLL4 is rendered in ribbon representation (cyan). N1 contact residues on DLL4 were rendered as sticks, and were used to predict N1 contact amino acids of the MNNL and DSL domains of DLL1 (red). Domains are labeled above and below the structures, respectively, and individual domains are identified by different degrees of color shading/intensity. (C) Parts of the MNNL and DSL sequences showing the contact amino acids (boxed), the divergent amino acids of DLL1 (red) and DLL4 (blue), and the sequence of ligands with amino acid exchanges (complete sequences of the changed MNNL and DSL domains are shown in Figure 1—figure supplement 1B). The N-glycosylation site at residue N109 of DLL4 is indicated in green. (D) N1/N2 activation ratios of ligands with exchanged N1 contact amino acids. D1contD4 does not show changes in receptor selectivity compared to DLL1. Replacing the contact residues of DLL4 with those of DLL1 only reduces N1/N2 activation ratio to ~13. Elimination of the N-glycosylation site of DLL4 with the N109G mutation (the corresponding amino acid of DLL1) does not change DLL4 receptor selectivity. Each dot represents the mean of N1 (relative luciferase units; Figure 5—source data 1)/N2 (relative luciferase units; Figure 5—source data 2) of n ≤ 3 measurements per clone of a given ligand construct. Bars represent the Mean ± SD of n ≤ 3 clones per construct (Figure 5—source data 3). DOI: https://doi.org/10.7554/eLife.40045.018

The following source data and figure supplement are available for figure 5:

Source data 1. Raw data (RLUs) of luciferase activity in co-cultures with N1rep cells used to generate the graph in Figure 5—figure supplement 1A.

DOI: https://doi.org/10.7554/eLife.40045.020

Source data 2. Raw data (RLUs) of luciferase activity in co-cultures with N1rep cells used to generate the graph in Figure 5—figure supplement 1B.

DOI: https://doi.org/10.7554/eLife.40045.021

Source data 3. N1/N2 activation ratios.

DOI: https://doi.org/10.7554/eLife.40045.022

Figure supplement 1. N1 and N2 activation by different ligand proteins.

DOI: https://doi.org/10.7554/eLife.40045.019

The region encompassing the MNNL up to and including EGF3 encodes the differential receptor selectivity of DLL1 and DLL4

In an attempt to identify the domains of DLL1 and DLL4 that contribute to differences in activating N1 and N2, we carried out a series of domain swaps to generate a set of chimeric ligands (II-V, VII-X in Figure 1) for stimulation of N1 and N2-expressing cells in our co-culture assay. Like wild-type ligands, chimeric ligands were expressed from single copy transgene integrations in the Hprt locus of murine ES cells. All chimeric proteins were expressed and present on the cell surface (Supplementary file 1), but expression levels varied among the chimeras (Supplementary file 2) despite integration into the Hprt locus by homologous recombination. We thus analyzed receptor selectivity of the chimeras in stimulating N1 and N2 responses using the co-culture assay by determining the N1/N2 response ratio. Stimulation with DLL1 gives a N1/N2 response ratio of approximately 1, DLL4 of ~20 (Figure 5A; numerical values used for calculations in Figure 5—source data 1, Figure 5—source data 2, Figure 5—source data 3; graphical representations of the relative luciferase activities of the ligands are shown in Figure 5—figure supplement 1). Strikingly, chimeras which retain the full ectodomain or the MNNL-EGF3 region of DLL4 have a N1/N2 stimulation ratio of approximately 20 similar to DLL4, whereas chimeras that retain the ectodomain, or MNNL-EGF3 region of DLL1 have a stimulation ratio of between one and two, resembling DLL1 (Figure 5A). These results indicate that the differences in activation potential of DLL4 and DLL1 toward N1 and N2 are encoded in the N-terminal part of the protein, encompassed by MNNL-EGF3. When chimeras include the MNNL-EGF2 or MNNL-DSL region of one ligand and the remainder of the other, the N1/N2 stimulation ratios of the chimeric pairs are equivalent (Figure 5A), indicating that the third EGF-like repeat makes an important contribution to receptor selectivity.
Regions outside of the MNNL-DSL contact interface contribute to the functional difference of DLL1 and DLL4 in vitro and in vivo

To analyze to what extent the amino acids that contact N1 in the binding interfaces of the MNNL and DSL domains might contribute to the different activity of DLL1 and DLL4 toward N1 and N2 we reciprocally exchanged these amino acids (XI-XII in Figure 1; Figure 5C and Figure 1—figure supplement 1B) based on alignments of the DLL4 (Luca et al., 2015) and DLL1 (Kershaw et al., 2015) structures (Figure 5B). Western blot analyses of cell lysates and cell surface biotinylation and immunoprecipitation showed that all variants were present on the cell surface (Supplementary File 1). The N1/N2 response ratios show that swapping the contact residues of DLL4 onto DLL1 do not substantially affect the activation ratio when compared to DLL1 itself, indicating that the differences between DLL1 and DLL4 in N1/N2 selectivity cannot simply be accounted for by interfacial residues in the MNNL-DSL region (Figure 5D; numerical values used for calculations are in Figure 5—source data 1, Figure 5—source data 2, Figure 5—source data 3). Similarly, replacement of the DLL4 contact residues by the analogous residues of DLL1 slightly reduces the mean N1/N2 activation ratio (to ~13), but does not collapse the ratio to 1 (Figure 5D), again strongly suggesting that residues outside of the MNNL-DSL contact interface contribute to the relative N1 selectivity of DLL4. These results are consistent with 1) the domain swap data, which argue that discrimination between DLL1 and DLL4 depends on the EGF repeats as well as on the MNNL-DSL region, and 2) the prior observation that variants of DLL4 selected for high N1 affinity accumulate mutations in the protein core, but not in the binding interface (Luca et al., 2015). Swapping the contact residues of DLL1 onto DLL4 did not reduce the binding affinity of DLL4 for N1 (D4contD1 Kd = 0.327 ± 0.036 mM; Figure 1—figure supplement 2Ac), fully consistent with the interpretation that the protein core of DLL4 contributes to N1 binding affinity, likely by influencing the fraction of molecules in a binding-active conformation. Although swapping the contact residues of DLL4 onto DLL1 increased binding affinity for N1 (D1contD4 Kd = 0.326 ± 0.044 mM; Figure 1—figure supplement 2Ad), the substitution did not substantially change the N1/N2 activation ratio, indicating that binding affinity for N1 is not the only influence on the selectivity of the two ligands for N1 or N2.

The DLL4 MNNL domain contains three N-glycosylation sites, one of which (N109) is conserved from amphibian to mammalian DLL4 ligands but absent in DLL1. This residue resides adjacent to the contact amino acid F110 (Figure 1—figure supplement 3A). We confirmed that DLL4 can actually be N-glycosylated at this site (Figure 1—figure supplement 3C) and tested whether N109-glycosylation contributes to DLL4 activity and selectivity by mutating N109 to G (XIII in Figure 1), which is the amino acid present in DLL1 in the equivalent position (G112). D4N109G had no effect on the relative activation potential of DLL4 for N1 versus N2 (Figure 5D), and its affinity for N1 was not altered (Kd = 0.341 ± 0.015 mM; Figure 1—figure supplement 2Ae), indicating that N-glycosylation at this site does not significantly modulate N1 binding or contribute to the relative selectivity of DLL4 towards N1 and N2.

To test whether the contact amino acids of DLL1 and DLL4 contribute to their functional divergence in vivo we generated a mouse line expressing D1contD4 (XI in Figure 1) instead of wild type DLL1 using our "mini-gene" knock-in strategy (Figure 3A). Heterozygous mice carrying this allele (DLL1D1contD4K) are indistinguishable from wild type. Homozygous mutants obtained from heterozygous matings at the expected Mendelian ratio (6/27) were viable and fertile, and indistinguishable from wild type and DLL1D1contD4K controls (Figure 6A). Uncx was expressed in regular pattern in the caudal halves of the somites of homozygous embryos (Figure 6Cd,d), consistent with only subtle abnormalities of individual vertebral bodies in the lower thoracic region of DLL1D1contD4K/D1contD4K fetuses (Figure 6D; n = 3/4) indicating that the contact amino acids and different binding affinities are not a major discriminating feature of the two ligands in vivo.

Discussion

DLL1 and DLL4 have context-dependent redundant and divergent functions, but the bases for these differences are unclear. Here, using systematic domain exchanges and mutation of contact amino acids in the MNNL and DSL domains of DLL1 and DLL4, cell-based and biochemical assays, and transgenic mice we show that 1) DLL1 and DLL4 differ significantly in their potential to activate N1 and N2 and this difference is encoded in the ligand ectodomains, 2) regions outside the known
Figure 6. DLL1 carrying the DLL4 contact amino acids in the MNNL and DSL domains is a functional DLL1 ligand in vivo. (A) E15.5 $D_{ll1}^{1contD4/1contD4}$ (c; n = 12) fetuses are indistinguishable from wild type (a; n = 19) and $D_{ll1}^{1contD4/1contD4}$ (b; n = 3) controls. (B) D1contD4 co-localizes with pan-Cadherin (panCad) at the cell surface of $D_{ll1}^{1contD4/1contD4}$ PSM cells (e-h; n ≤ 3); Scale bars: a, e = 500 mm; b-d, f-h = 5 mm. (C) Whole mount in situ hybridization showing that D1contD4 induces normal Uncx expression during somitogenesis (d,d'; n ≤ 5). (D) Skeletal preparations of $D_{ll1}^{1contD4/1contD4}$ E15.5 fetuses showing minor defects of single vertebrae in the lower thoracic region (c,c'; n = 3/4).

DOI: https://doi.org/10.7554/eLife.40045.023

Contact interface contribute to context-dependent ligand function, and (3) the contact amino acids are not the sole or primary determinant of this discrimination between the two receptors.

Analysis of our transgenic mice expressing D1ECD_D4ICD or D4ECD_D1ICD indicate a critical role of the ECD for the function of DLL1 during somite patterning in vivo. This resembles intrinsic
functional differences that reside in the extracellular domains of mN1 and mN2 during kidney development (Liu et al., 2013), whereas the N1 and N2 ICDs appear to be functionally equivalent in various developmental contexts (Liu et al., 2015). Functional equivalence of DLL1’s and DLL4’s ICDs in vivo is further supported by the rescue of the skeletal muscle phenotype in our D1N-E3_D4 knock-in mice, which harbor the DLL4 ICD. In this developmental context even domains C-terminal to EGF3 of DLL1 appear to be interchangeable. Analyses of the ECD/ICD domain swaps in the cell-based assay also suggest that the discriminatory potential of the ligands tracks with the ECD, and not with the ICD, even though the ICD appears to affect the strength and/or dynamics of the signal in co-culture assays where ligand and receptor expression is enforced in vitro (Nandagopal et al., 2018). Additional sources of complexity in vivo, like the stronger cis-inhibitory potential of the DLL4 ECD on Notch signaling (Preuße et al., 2015), or cyclic modulation of Notch by LFNG in the paraxial mesoderm, or different interactions with lipids (Suckling et al., 2017) might account for the resistance to loss of function phenotypes from ligand ICD swaps in vivo.

EGF-like repeats 11 and 12 of mouse N1 and N2 are highly similar (56/83 residues identical, 14 similar amino acids), and 13 of the 17 amino acid residues at the DLL4-binding interface are identical. Moreover, the x-ray structures of the EGF11-13 fragments of human N1 and N2 adopt a very similar arrangement (Suckling et al., 2017). Nevertheless, DLL1 and DLL4 exhibit a "discrimination potential" of ~20 fold in terms of receptor response in culture assays, suggesting that either the few different contact amino acids in EGF 11 and 12 of N1 and N2 have a significant impact or interactions of DLL1 and DLL4 with N1 and N2 are not limited to the MNNL and DSL interfaces with receptor EGF repeats 11 and 12. Domain swaps carried out here show that the region responsible for this receptor discrimination maps to the MNNL-EGF3 region (Figure 5). These findings are consistent with previous work uncovering the requirement of EGF repeats 1–3 of the DLL ligands for NOTCH1 activation, the importance of this region in the binding of Serrate family ligands to Notch receptors and in Serrate/Jagged-induced signaling, and the importance of EGF repeats 8–10 of NOTCH1 for signal activation by DLL ligands (Shimizu et al., 1999; Cordle et al., 2008; Yamamoto et al., 2012; Andrawes et al., 2013; Schuster-Gossler et al., 2016; Luca et al., 2017; Liu et al., 2017). Together, this body of work suggests that interactions of the N-terminal EGF repeats of the DLL ligands with EGF repeats 8–10 of Notch also contribute to recognition and impart discriminatory potential. The D1N-E3_D4 knock-in mice also point a functional role for domains outside the known binding interface, since this chimeric ligand does not substitute fully for DLL1 in vivo during somite patterning despite harboring the MNNL and DSL domains and EGF1-3 of DLL1, supporting context-dependent contributions of additional C-terminal EGF repeats observed previously in mice (Schuster-Gossler et al., 2016).

Remarkably, the exchange of the contact amino acids in DLL1 with those of DLL4 in the D1contD4 protein does not alter receptor selectivity in cultured cells even though these changes increase N1 binding affinity. This result suggests that receptor selectivity of DLL1 and DLL4 is not determined exclusively by the differences in binding strength. The D1contD4 chimera even substitutes almost completely for DLL1 function in mice during somite patterning, which is highly sensitive to altered Notch signaling (Schuster-Gossler et al., 2009) and therefore a suitable in vivo read out to detect even minor differences of Notch ligand function. Together, our results also favor the conclusion that the contact amino acids in the MNNL and DSL domains do not make the dominant contributions to the functional divergence of DLL1 and DLL4 in vivo, suggesting instead that differences in the domain cores, and/or contacts outside of the known DLL4-NOTCH1 interface, are the factors that most contribute to this functional divergence.

Materials and methods

<table>
<thead>
<tr>
<th>Key resources table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reagent type (species) or resource</td>
</tr>
<tr>
<td>Gene (Mus musculus)</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Reagent type (species) or resource</th>
<th>Designation</th>
<th>Source or reference</th>
<th>Identifiers</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene (Mus musculus)</td>
<td>DLL4</td>
<td>MGI:1859388; NCBI Gene: 54485</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain, strain background (Mus musculus)</td>
<td>CD1</td>
<td>Charles River Laboratories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain, strain background (Mus musculus)</td>
<td>129Sv/CD1 hybrids</td>
<td>own colony</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>Dll1lacZ</td>
<td>PMID: 9109488; DOI: 10.1038/386717a0</td>
<td>RRID:MGI:5780046</td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>Dll1loxP</td>
<td>PMID: 15146182; DOI: 10.1038/n1075</td>
<td>RRID:MGI:5431505</td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>T(s):Cre</td>
<td>PMID: 18708576; PMCID: PMC2518812; DOI: 10.1101/gad.480408</td>
<td>MGI:3811072</td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>ZP3:Cre</td>
<td>PMID: 10686600; MGI:2176187</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>Dll1loxN</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534 genetics.115.184515</td>
<td>MGI:5790945</td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>Dll1loxM</td>
<td>PMID: 26114479; PMCID: PMC4482573; DOI: 10.1371/journal.pgen.1005328</td>
<td>MGI:5779556</td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>Dll1loxE3_D4ki</td>
<td>this paper</td>
<td>mini gene insertion in the Dll1 locus</td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>Dll1loxD4ki</td>
<td>this paper</td>
<td>mini gene insertion in the Dll1 locus</td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>HprtloxECD_Dll4ICD</td>
<td>this paper</td>
<td>inducible insertion into Hprt locus</td>
<td></td>
</tr>
<tr>
<td>Genetic reagent (Mus musculus)</td>
<td>HprtloxECD_Dll1ICD</td>
<td>this paper</td>
<td>inducible insertion into Hprt locus</td>
<td></td>
</tr>
<tr>
<td>Cell line (Mus musculus)</td>
<td>E14TG2a</td>
<td>PMID: 26114479; PMCID: PMC4482573; DOI: 10.1371/journal.pgen.1005328</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell line (Mus musculus)</td>
<td>129Sv/cast</td>
<td>PMID: 26114479; PMCID: PMC4482573; DOI: 10.1371/journal.pgen.1005328</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell line (Homo sapiens)</td>
<td>HeLaN1</td>
<td>PMID: 9653148; DOI: 10.1073/pnas.95.14.8108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell line (Mus musculus)</td>
<td>pMP8.CAG-DLL1</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534 genetics.115.184515</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell line (Mus musculus)</td>
<td>E14rep</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534 genetics.115.184515</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Reagent type (species) or resource</th>
<th>Designation</th>
<th>Source or reference</th>
<th>Identifiers</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell line (<strong>Mus musculus</strong>)</td>
<td>N1rep</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell line (<strong>Mus musculus</strong>)</td>
<td>N2rep</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial strain (<strong>E. coli</strong>)</td>
<td>SW106</td>
<td>PMID: 15731329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-DLL4</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D1ECD-D4ICD</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D1N-E3_D4</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D1N-E2_D4</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D1N-D_D4</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D4ECD-D1ICD</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D4N-E3_D1</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D4N-E2_D1</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D4N-D_D1</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D1contD4</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D4contD1</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-D4N109G</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-Stop-D1ECD-D4ICD</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pMP8.CAG-Stop-D4ECD-D1ICD</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pMP8.CAG</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>D1N-E3_D4-targeting</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>based on Dil1Dll1ki targeting</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>D1contD4-targeting</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>based on Dil1Dll1ki targeting</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pLexM-Avi-His</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pLexM</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pLexM-D1N-E5-Avi-His</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pLexM</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pLexM-D4N-E5-Avi-His</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pLexM</td>
</tr>
<tr>
<td>Transfected construct (<strong>Mus musculus</strong>)</td>
<td>pLexM-D1N-E3_D4-E5-Avi-His</td>
<td>this paper</td>
<td>PMID: 26801181; PMCID: PMC4788113; DOI: 10.1534/genetics.115.184515</td>
<td>progenitor: pLexM</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Reagent type (species) or resource</th>
<th>Designation</th>
<th>Source or reference</th>
<th>Identifiers</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfected construct (Mus musculus)</td>
<td>pLexM-D4N-E3_D1-E5-Avi-His</td>
<td>this paper</td>
<td></td>
<td>progenitor: pLexM</td>
</tr>
<tr>
<td>Transfected construct (Mus musculus)</td>
<td>pLexM-D1contD4-E5-Avi-His</td>
<td>this paper</td>
<td></td>
<td>progenitor: pLexM</td>
</tr>
<tr>
<td>Transfected construct (Mus musculus)</td>
<td>pLexM-D4contD1-E5-Avi-His</td>
<td>this paper</td>
<td></td>
<td>progenitor: pLexM</td>
</tr>
<tr>
<td>Transfected construct (Mus musculus)</td>
<td>pLexM-D4N109G-E5-Avi-His</td>
<td>this paper</td>
<td></td>
<td>progenitor: pLexM</td>
</tr>
<tr>
<td>Antibody</td>
<td>Rat anti-DLL1</td>
<td>PMID: 17664336;</td>
<td>(1F9, rat monoclonal)</td>
<td>1:50 (IF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMCID: PMC2064846;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DOI: 10.1083/jcb.200702009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Goat anti-DLL4</td>
<td>R and D Systems</td>
<td>Cat. #AF1389</td>
<td>1:50 (IF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_354770</td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Mouse anti-panCadherin</td>
<td>Sigma-Aldrich</td>
<td>Cat. #C1821</td>
<td>1:250 (IF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_476826</td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Donkey anti-mouse Alexa 555</td>
<td>Invitrogen</td>
<td>Cat. #A-31570</td>
<td>1:100 (IF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_2536180</td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Donkey anti-goat Alexa 488</td>
<td>Invitrogen</td>
<td>Cat. #A-11055</td>
<td>1:100 (IF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_2534102</td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Donkey anti-rat Alexa 488</td>
<td>Invitrogen</td>
<td>Cat. #A-21208</td>
<td>1:100 (IF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_2535794</td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Anti-FLAG-Peroxidase (HRP)</td>
<td>Sigma-Aldrich</td>
<td>(M2 mouse, monoclonal purified)</td>
<td>1:10 000 (WB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cat. #A8592</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_439702</td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Mouse anti-b-Tubulin</td>
<td>Sigma-Aldrich</td>
<td>Cat. #T7816</td>
<td>1:500 000; 1:1 000 000 (WB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_261770</td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Anti-mouse HRP</td>
<td>Amersham</td>
<td>Cat. #NA931</td>
<td>1:10 000 (WB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_772210</td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>MHC (Myosin Heavy Chain)</td>
<td>Sigma-Aldrich</td>
<td>Cat. #M4276</td>
<td>1:250 (IHC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_477190</td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Anti-DIG AP fab fragment</td>
<td>Roche</td>
<td>Cat. #1093274</td>
<td>1: 5 000 (ISH)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antibody</td>
<td>Anti-mouse biotinylated (BA9200/goat)</td>
<td>Vector Laboratories</td>
<td>Cat. #BA-9200</td>
<td>1:200 (IHC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RRID:AB_2336171</td>
<td></td>
</tr>
<tr>
<td>Commercial assay or kit</td>
<td>Luciferase Cell Culture Lysis 5X Reagent</td>
<td>Promega</td>
<td>Cat. #E1531</td>
<td></td>
</tr>
<tr>
<td>Commercial assay or kit</td>
<td>Luciferase Assay Reagent</td>
<td>Promega</td>
<td>Cat. #E1483</td>
<td></td>
</tr>
<tr>
<td>Commercial assay or kit</td>
<td>SuperScript IV Reverse Transcriptase</td>
<td>Invitrogen</td>
<td>Cat. #18090050</td>
<td></td>
</tr>
<tr>
<td>Commercial assay or kit</td>
<td>Expand High-Fidelity PCR system</td>
<td>Roche</td>
<td>Cat. #04743733001</td>
<td></td>
</tr>
<tr>
<td>Commercial assay or kit</td>
<td>Tri-Reagent</td>
<td>Sigma-Aldrich</td>
<td>Cat. #T9424</td>
<td></td>
</tr>
<tr>
<td>Chemical compound, drug</td>
<td>Sulfo-NHS-LC-Biotin</td>
<td>Thermo</td>
<td>Cat. #21335</td>
<td></td>
</tr>
<tr>
<td>Chemical compound, drug</td>
<td>Pierce NeutrAvidin Agarose</td>
<td>Thermo</td>
<td>Cat. #29200</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Reagent type (species) or resource</th>
<th>Designation</th>
<th>Source or reference</th>
<th>Identifiers</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical compound, drug</td>
<td>cOmplete, Mini, EDTA-free Proteinase Inhibitor Cocktail</td>
<td>Roche</td>
<td>Cat. #04693159001</td>
<td></td>
</tr>
<tr>
<td>Chemical compound, drug</td>
<td>BM-Purple AP substrate</td>
<td>Sigma-Aldrich</td>
<td>Cat. #11442074001</td>
<td></td>
</tr>
<tr>
<td>Chemical compound, drug</td>
<td>G418</td>
<td>Biochrom</td>
<td>Cat. #291–25</td>
<td>125 mg/ml</td>
</tr>
<tr>
<td>Chemical compound, drug</td>
<td>HAT</td>
<td>Gibco</td>
<td>Cat. #31062–037</td>
<td></td>
</tr>
<tr>
<td>Chemical compound, drug</td>
<td>HT</td>
<td>Gibco</td>
<td>Cat. #11067030</td>
<td>1:100</td>
</tr>
<tr>
<td>Chemical compound, drug</td>
<td>Tunicamycin</td>
<td>Sigma-Aldrich</td>
<td>Cat. #7765</td>
<td>1 mg/ml</td>
</tr>
<tr>
<td>Chemical compound, drug</td>
<td>Alcian blue</td>
<td>Sigma-Aldrich</td>
<td>Cat. #A5268</td>
<td>5% working solution</td>
</tr>
<tr>
<td>Chemical compound, drug</td>
<td>Alizarin red</td>
<td>Sigma-Aldrich</td>
<td>Cat. #A5533</td>
<td>5% working solution</td>
</tr>
<tr>
<td>Other</td>
<td>WesternBright Quantum HRP substrate</td>
<td>Advansta</td>
<td>Cat. #12042-D20</td>
<td>as recommended by the manufacturer</td>
</tr>
<tr>
<td>Other</td>
<td>Amersham ECL Detection Reagent</td>
<td>GE Healthcare Life Sciences</td>
<td>#RPN2106</td>
<td>as recommended by the manufacturer</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>DLL1 wt For</td>
<td>other</td>
<td>NA</td>
<td>5'-CTGAAGCCGACCTGCCCTGATAGCAC-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>DLL1 wt Rev</td>
<td>other</td>
<td>NA</td>
<td>5'-GGAGCTCCAGA CCTGGCGGGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Dil1lox2 For</td>
<td>other</td>
<td>NA</td>
<td>5'-ATCCCTTGGGT TTGAAGAAG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Dil1lox2 Rev</td>
<td>other</td>
<td>NA</td>
<td>5'-TGTTGAGCGGTA ACAAACCGTGAGATT-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Dil1Dw1 For</td>
<td>other</td>
<td>NA</td>
<td>5'-GTACAGCTACTTC CAGTAAGAGAAG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Dil1Dw1 Rev</td>
<td>other</td>
<td>NA</td>
<td>5'-AAAGACAACC TATCCCTGCGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Dil1Dw4 For</td>
<td>other</td>
<td>NA</td>
<td>5'-TCCATTCATCG CTCATCTTAC-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Dil1Dw4 Rev</td>
<td>other</td>
<td>NA</td>
<td>5'-GCACATCCTAC ACACTC-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>T(s):Cre For</td>
<td>other</td>
<td>NA</td>
<td>5'-AATCTTTGG GCTGCCAGAG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>T(s):Cre Rev</td>
<td>other</td>
<td>NA</td>
<td>5'-AGCTTCAGCAGG ATCAAG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>ZP3:Cre For</td>
<td>other</td>
<td>NA</td>
<td>5'-GCTTCAGTACC GGTCGATGCAACG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>ZP3:Cre Rev</td>
<td>other</td>
<td>NA</td>
<td>5'-GCTGAGTACTT GATGGAGCAAG-3'</td>
</tr>
</tbody>
</table>
| Sequence-based reagent           | Hprt-CAGD1ECD_ D4ICD + neo For   | this paper          | NA                           | 5'-CCTAGCCCTGCA AGAACGGAGC-3'
<table>
<thead>
<tr>
<th>Reagent type (species) or resource</th>
<th>Designation</th>
<th>Source or reference</th>
<th>Identifiers</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence-based reagent</td>
<td>Hprt-CAGD1ECD_ D4ICD + neo Rev</td>
<td>this paper</td>
<td>NA</td>
<td>5'-TTGGCCACAAATTG GACCTTGTCC-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Hprt-CAGD4ECD_ D1ICD + neo For</td>
<td>this paper</td>
<td>NA</td>
<td>5'-CAGTTGAGCAT AGTACC TTGAC-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Hprt-CAGD4ECD_ D1ICD + neo Rev</td>
<td>this paper</td>
<td>NA</td>
<td>5'-CATGGTTTCTGTCT CTCCCCACAGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Hprt- D1ECD_D4ICDrec and Hprt- D4ECD_D1ICDrec For (activated allele)</td>
<td>this paper</td>
<td>NA</td>
<td>5'-ACATGGCCGTCATC AAAGAG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Hprt- D1ECD_D4ICDrec and Hprt- D4ECD_D1ICDrec Rev (activated allele)</td>
<td>this paper</td>
<td>NA</td>
<td>5'-GGGCAACAGAGA AATATCTGTCATC-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Dll1- D1N- E3_D4ki For</td>
<td>this paper</td>
<td>NA</td>
<td>5'-CTGGAGCTCT ACAGTGATGA-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Dll1- D1N- E3_D4ki Rev</td>
<td>this paper</td>
<td>NA</td>
<td>5'-AGCTCATCA GGGGATCTG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>mNotch1 For</td>
<td>other</td>
<td>NA</td>
<td>5'-CAACCGCCA GTGTGTTCAAG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>mNotch1 Rev</td>
<td>other</td>
<td>NA</td>
<td>5'-GAGGATCTCA CGTGGACAGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>hNOTCH1 For</td>
<td>PMID: 22002304; DOI: 10.1038/onc.2011.467</td>
<td>NA</td>
<td>5'-GCCGACAGA AAAGAGAATCAGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>hNOTCH1 Rev</td>
<td>PMID: 22002304; DOI: 10.1038/onc.2011.467</td>
<td>NA</td>
<td>5'-GCCGACAGA AAAGAGAATCAGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>hNOTCH2 For</td>
<td>this paper</td>
<td>NA</td>
<td>5'-CAACCGCCA GTGTGTTCAAG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>hNOTCH2 Rev</td>
<td>this paper</td>
<td>NA</td>
<td>5'-GAGGATCTCA CGTGGACAGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>hNOTCH3 For</td>
<td>PMID: 16327489; PMCID: PMC1409885</td>
<td>NA</td>
<td>5'-GAGGATCTCA CGTGGACAGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>hNOTCH3 Rev</td>
<td>PMID: 16327489; PMCID: PMC1409885</td>
<td>NA</td>
<td>5'-GAGGATCTCA CGTGGACAGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>hGAPDH For</td>
<td>PMID: 22002304; DOI: 10.1038/onc.2011.467</td>
<td>NA</td>
<td>5'-GAGGATCTCA CGTGGACAGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>hGAPDH Rev</td>
<td>PMID: 22002304; DOI: 10.1038/onc.2011.467</td>
<td>NA</td>
<td>5'-GAGGATCTCA CGTGGACAGG-3'</td>
</tr>
<tr>
<td>Sequence-based reagent</td>
<td>Forward primer - correct integration into Hprt locus</td>
<td>other</td>
<td>NA</td>
<td>5'-GGGAACTGTGTT AGAAAAAAAGA AACTATGAAGA-3'</td>
</tr>
</tbody>
</table>
### Generation and husbandry of transgenic mice

**Ethics statement**

All animal experiments were performed according to the German rules and regulations (Tierschutzgesetz) and approved by the ethics committee of Lower Saxony for care and use of laboratory animals (LAVES Niederschlesisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit; refs.: 33.12-42505-04-13/1314 and 33.14-42505-04-13/1293). Mice were housed in the central animal facility of Hannover Medical School (ZTL) and were maintained as approved by the responsible Veterinary Officer of the City of Hannover. Animal welfare was supervised and approved by the Institutional Animal Welfare Officer (Tierschutzbeauftragter).

**Mouse strains**

Wild type mice were CD1 and 129Sv/CD1 hybrids; *Dll1lacZ* (Hrabec de Angelis et al., 1997), *Dll1loxP* (Hozumi et al., 2004), T(s):Cre (Feller et al., 2008) and ZP3:Cre (de Vries et al., 2000), *Dll1ki* (Schuster-Gossler et al., 2016), and *Dll1D10A* (Preuße et al., 2015) were described previously.

**Generation of transgenic mice**

Mice allowing for inducible expression of chimeric ligands were generated by morula injection of E14TG2a ES cells carrying the expression construct in the *Hprt* locus. E14TG2a cells were
electroporated with linearized targeting constructs and correct integrations were identified by HAT selection and validated by long-range PCR using primers: For/Rev: GGGAACCTGTTAGAAAAAAA-GAAGCTTGAAGAAC/GGCTATGAACTAATGACCCCG.

DII1(D1-E3, D4)ki and DII1(D1contD4)ki mice were generated with 129Sv/cast ES cells. ES cells were electroporated with linearized targeting constructs, Cas9 D10A nickase (Addgene #42335; Cong et al., 2013) expression vector and guide RNAs targeting the first intron of DII1 to increase the frequency of homologous recombination (guide-A:FOR: GGCGAGGCGGACGCTCGGAT; guide-B:REV: GTCCTCGGTCGTGAC, according to http://crispr.mit.edu/) the pair score for A and B = 79, 0 off target pairs, and 0 genic OT pairs). G418 resistant clones were screened for targeted integrations by Southern blot analysis (5' probe: a 316 bp BamHI/EcoRI fragments containing the Hprt intron five digestion from either synthesized gene fragments (II, IV, VII, IX, XI-XIII in Figure 1) or fragments obtained by restriction digests from DII1 and DII4 cDNA constructs (V, X in Figure 1). Tagged cDNAs were cloned into pMP8-CAG.Stop shuttle as EcoRI/BamHI or EcoRI/NcoI fragments. The stop cassette was excised by Cre mediated recombination of the loxp sites in bacterial SW106 cells.

Genotyping of mice and embryos

Genomic DNA was isolated from ear or tail biopsies, yolk sacs or umbilical cords and used as template in PCRs with the following primer pairs:

- Hprt(FOR: GGCAGCGGGCAGCTCCGGAT; guide-B:REV: GCTC TCGGTCGTGAC, according to http://crispr.mit.edu/) the pair score for A and B = 79, 0 off target pairs, and 0 genic OT pairs). G418 resistant clones were screened for targeted integrations by Southern blot analysis (5' probe: a 316 bp BamHI/Avall fragment 3.8 kb upstream of DII1 exon 1; 3' probe: a 528 bp PCR fragment in DII1 intron five obtained with primers: CCTGTGAGACTTTCTACGTTGCTC/CACAACCATGTCACCTTCTAGATTC).

The neo cassette was excised in the female germ line using ZP3:Cre mice.

Cloning of constructs

Hprt targeting constructs for expression from single copy integrations in ES cells
cDNAs encoding Flag-tagged ligand proteins with exchanges of domains or individual amino acids in the extracellular domain of DII1 and DII4 were generated by standard cloning procedures using either synthesized gene fragments (II-V, VII-X, XI-XIII in Figure 1) or fragments obtained by restriction digests from DII1 and DII4 cDNA constructs (V, X in Figure 1). Tagged cDNAs were cloned into pMP8-CAG.Stop shuttle as EcoRI/BamHI or EcoRI/NcoI fragments. The stop cassette was excised by Cre mediated recombination of the loxp sites in bacterial SW106 cells.

Hprt targeting constructs for inducible expression in transgenic mice
D1ECD_D4/ICD and D4ECD_D1/ICD were generated by PCR amplification of the respective untagged cDNAs and subcloned into shuttle vector pSLL1800tomato containing the wild type and mutant loxp sites and irsRED. Subsequently, the fragments encoding the chimeric ligands fused to irsRED were cloned into pMP8-CAG.Stop (Preuße et al., 2015) using MluI and Sva1 restriction sites.

Mini gene constructs for targeting the DII1 locus
DII1(D1-E3, D4)ki and DII1(D1contD4)ki targeting constructs were generated by standard cloning procedures based on the DII1(D1-E3, D4)ki or DII1(D1contD4)ki targeting vectors (Preuße et al., 2015; Schuster-Gossler et al., 2016). First, the 3' DT cassette was removed by Pmel and AatI digest and relegation of the blunt ended plasmid. EcoRI fragments containing the DII1 or DII4 coding sequences in the targeting vector lacking the 3' DT cassette were excised by EcoRI and cloned.
into pCR-TOPO-XL. The wild type Dll1 sequence was replaced in pCR-TOPO-XL by a D1contD4 cDNA, the Dll4 sequence by D1N-E3_D4 cDNA. Fragments were ligated back into the targeting vectors as EcoRI fragments.

Avi-His-tagged ligand fragments for protein expression and purification

For production and purification of proteins for binding assays (XIV-XVIII in Figure 1) fragments encompassing the N-terminus up to and including EGF5 were PCR amplified and cloned into pLexM-Avi-His vector (Andrewes et al., 2013) as EcoRI/BamHI fragments by standard procedures.

Analysis of gene expression patterns and phenotypes

Whole mount in situ hybridization

E9.5 embryos were collected in ice cold PBS and fixed in 4% formaldehyde/PBS overnight at 4°C and dehydrated in methanol. In situ hybridization was performed according to standard procedures with digoxigenin labelled cDNA probes for Uncx (Neidhardt et al., 1997).

Antibody staining

E18.5 embryos were collected in ice cold PBS, fixed in 4% formaldehyde/PBS overnight at 4°C, dehydrated in methanol, ethanol, and 2-propanol. Hind limbs were paraffin embedded and 10 mm transverse sections stained for Myosin Heavy Chain (MHC).

Whole mount immunofluorescence

E9.5 embryos were collected in ice cold PBS, fixed in 4% formaldehyde in PBS and immunofluorescence staining was performed as described in Bone et al. (2014). Used primary antibodies: anti-DLL1 (1F9; 1:50) (Geffers et al., 2007), anti-DLL4 (AF1389, R and D; 1:50), and anti-pan-Cadherin (C1821, Sigma; 1:250). Used secondary antibodies: Alexa488/555 conjugated antibodies (Invitrogen; 1:100). Images were taken using OLYMPUS FV1000.

Skeletal preparations

E15.5 fetuses were collected in ice cold PBS and dehydrated in EtOH. Alcian blue and Alizarin red staining was performed using standard procedures (Cordes et al., 2004).

Western blot analyses

Cells were lysed in 2x sample buffer (0.125M Tris pH 6.8; 4% SDS; 20% glycin; 5% b-mercaptoethanol; 0.025% bromphenol blue). Proteins were separated by SDS-PAGE and transferred onto Immobilon-P Transfer membranes (Millipore) by wet tank or SemiDry blotting. Membranes were blocked in 5% nonfat dried milk powder (AppliChem) in PBS/0.1% Tween20 and subsequently incubated in 5% nonfat dried milk powder containing primary antibodies. Used primary antibodies: anti-Flag HRP (mouse monoclonal; clone M2; Sigma; 1:10 000), anti-b-Tubulin I (Sigma; 1:500 000/1:1 000 000). Used secondary antibodies: anti-mouse HRP (Amersham; 1:10 000). For HRP detection ECL Western Blotting Detection Reagent (Amersham) and WesternBright Quantum (advansta) were used with Luminisent Image Analyser LAS4000 (Fujifilm). ImageJ was used to quantify signals.

RT-PCR

HeLaN1 cells were lysed in Tri-Reagent (Sigma) and RNA was isolated according to the manufacturer’s instructions. Reverse transcription was performed using SuperScript IV (Invitrogen) according to the manufacturer’s instructions. Primers used for RT-PCR analysis were: mNotch1 For/Rev TAGG TGCTCTTGCGTCAGGCTGCTCGAGGTGTTCTCAGTTGATGTTC; hNOTCH1 For/Rev TCCACCAG TTTGAATGGTGCAAGCTCATCTGGGACAGG (Ding et al., 2012); hNOTCH2 For/Rev CAACCCCGATGTTGCTGAGACCTAGGCTGTTGCTTCAGG; hNOTCH3 For/Rev AGATTCTCA TCCGAACCGCTTCA/GGGGTCTCTTCCTTGCTATCTCG (Büchler et al., 2005); hGAPDH For/Rev GAGTCACCGATTTGGCTGTTGATTTGGGAGGGATCTCG (Ding et al., 2012).
Southern blot analyses
Genomic DNA was isolated from ES cells, digested with BamHI overnight and separated on an 0.7% agarose gel. Blotting, crosslinking, hybridization, and signal detection were performed using Immobilon-Ny+ membrane (Millipore) according to the manufacturer’s instructions.

Cell culture experiments
Culture of cells
Mouse E14TG2a and 129Sv/cast ES cells were cultured in DMEM (Invitrogen) cell culture medium supplemented with 15% FCS (Biochrom AG), Glutamax, Pen/Strep. Sodium Pyruvate, MEM Non-Essential Amino Acid Solution, b-mercaptoethanol, and leukemia inhibitory factor (LIF). HeLaN1 cells were cultured in DMEM (Invitrogen) cell culture medium supplemented with 10% FCS (Biochrom AG), Glutamax and Pen/Strep. All cell lines were tested negative for mycoplasma. No authentication of cell lines was performed.

Generation of cells stably expressing ligand proteins
ES cells were electroporated with linearized pMP8 targeting vectors and selected with HAT (1:300; Gibco). Correct integration of the 5’ homolog arm in HAT resistant clones was verified with long-range PCR using following primers: For/Rev: GGGAACTGTGAAAAAGAAAATGAAGAAC/GGCTATGAACTAATGACCCG. Expression of proteins was verified using Western Blot analyses.

Trans-activation assay
For in vitro cell co-culture assays ES cells were counted in PBS using LUNA-II (logos biosystems) and 9.25 × 10⁵ ligand and 0.75 × 10⁵ receptor expressing cells were plated on gelatin coated six well plate dishes. After 24 hr fresh medium was added. 48–52 hr after co-cultivation cells were washed once with PBS, lysed in 250 ml 1xCCLR (Luciferase Cell Culture Lysis Reagent, Promega), transferred into 1.5 ml tubes, and frozen at −80°C. For measurements lysates were thawed, vortexed, and briefly centrifuged. 20 ml aliquots of each lysate was measured with Luciferase Assay Reagent in duplicates or quadruplicates using GloMax-96 (Promega).

Biotinylation assay
For determination of relative cell surface protein levels, cells were treated with Sulfo-NHS-LC-Biotin (Pierce; 0.25 mg/ml PBS supplemented with 1 mM MgCl₂ and 0.1 mM CaCl₂), quenched with 100 mM glycine in DMEM and lysed in lysis buffer supplemented with Complete Proteinase Inhibitor Cocktail Tablets (Roche). Biotinylated proteins were immunoprecipitated using NeutrAvidin beads (Thermo Scientific) and analyzed by Western blotting. For detailed information see (Braune et al., 2014; Preuße et al., 2015).

Protein expression and purification
The cDNA for expression of the N1 fragment using the pLexM vector was described previously (Andrawes et al., 2013) and encodes the N1 signal sequence followed by EGF repeats 6-15 (amino acids 216–604), a biotinylation (avi) tag, and a His₆ tag. The cDNAs for expression of DLL1, DLL4, and all chimeric proteins extend from the N-terminus through EGF5. These proteins were also subcloned into pLexM as described (Andrawes et al., 2013). The N2(1–15)-Fc protein was purchased from R and D systems and used without further purification.

Exp293F cells maintained in Expi293 expression media were grown to cell density of 10⁶ cells/ml and then transiently transfected withDll1 ligand, Dll4 ligand or N1 DNA (1 mg/liter of cells) and FectoPro transfection reagent (Polysynt) at 1:1 DNA/FectoPro ratio. For biotinylation of Avi-tagged NOTCH1 protein, cells were co-transfected with biotin ligase (BirA) DNA as well as with DNA encoding Protein O-fucosyltransferase-1 (POFUT1), which enhances Notch folding and secretion.

Transfected cells were then cultured in FreeStyle293 media for 3–4 days to produce protein. The media was collected, separated from the cells by centrifugation and supplemented with 50 mM Tris buffer, pH 8.0. The resulting supernatant was bound to Ni-NTA beads (Qiagen) over a 3 hr incubation at 4°C. After a wash with ten column volumes of 50 mM Tris buffer, pH 8.0, containing 150 mM NaCl, 5 mM CaCl₂, and 20 mM Imidazole, bound protein was eluted with the same buffer supplemented with 250 mM Imidazole. Following elution, fractions containing the partially purified proteins were pooled and dialyzed against 250 mM Imidazole.
were concentrated and further purified by gel-filtration chromatography using a Superdex 200 column in 50 mM Tris, pH 8.0, containing 150 mM NaCl, and 5 mM CaCl₂. The quality of the resulting purified proteins was assessed using non-reducing SDS-PAGE. Pure fractions were pooled, flash frozen and stored at —80°C. The efficiency of biotinylation was estimated by immunoprecipitation with streptavidin resin.

Biolayer interferometry
Ligand binding affinities were quantified by biolayer interferometry using a BLItz instrument (ForteBio). For N1 binding, streptavidin biosensors were loaded with the biotinylated Notch1 fragment, equilibrated in buffer for 30 s, then dipped into ligand samples of varying concentration until equilibrium was observed. For N2 binding, protein A biosensors were used for the capture step. All ligand-receptor binding experiments were done in HBS-P buffer containing 0.005% surfactant P20, supplemented with 5 mM CaCl₂. Equilibrium binding curves were fitted with a one site - specific binding model using GraphPad Prism.

Statistical analysis
Statistical analyses were done using Prism7 (GraphPad) as indicated in Figure legends.

Acknowledgements
We thank Kristina Preusse for providing the cDNAs for D1N-E2_D4, D1N-D_D4, D4N-E2_D1, and D4N-D_D1 chimeric ligands and Patricia Delany-Heiken for excellent technical assistance.

Additional information

<table>
<thead>
<tr>
<th>Funder</th>
<th>Grant reference number</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harvard Medical School</td>
<td>van Maanen Graduate fellowship</td>
<td>Sanchez M Jarrett</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>R35-CA220340</td>
<td>Stephen C Blacklow</td>
</tr>
<tr>
<td>Deutsche Forschungsgemeinschaft</td>
<td>GO 449/13-1</td>
<td>Achim Gossler</td>
</tr>
<tr>
<td>Deutsche Forschungsgemeinschaft</td>
<td>REBIRTH</td>
<td>Achim Gossler</td>
</tr>
</tbody>
</table>

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions
Lena Tveriakhina, Formal analysis, Investigation, Visualization, Writing—original draft, Writing—review and editing; Karin Schuster-Gossler, Resources, Investigation, Writing—review and editing; Sanchez M Jarrett, Marie B Andrawes, Meike Rohrbach, Investigation, Writing—review and editing; Stephen C Blacklow, Achim Gossler, Conceptualization, Supervision, Funding acquisition, Visualization, Writing—original draft, Writing—review and editing

Author ORCIDs
Achim Gossler http://orcid.org/0000-0002-9103-9116

Ethics
Animal experimentation: All animal experiments were performed according to the German rules and regulations (Tierschutzgesetz) and approved by the ethics committee of Lower Saxony for care and use of laboratory animals LAVES (Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit; refs.: 33.12-42502-04-13/1314 and 33.14-42502-04-13/1293). Mice were housed in the central animal facility of Hannover Medical School (ZTL) and were maintained as approved by
the responsible Veterinary Officer of the City of Hannover. Animal welfare was supervised and approved by the Institutional Animal Welfare Officer (Tierschutzbeauftragter).

Decision letter and Author response
Decision letter https://doi.org/10.7554/eLife.40045.029
Author response https://doi.org/10.7554/eLife.40045.030

Additional files
Supplementary files
- Supplementary file 1. Relative cell surface expression levels of the ligand proteins used co-culture studies. Levels of one representative clone for each ligand protein were determined by cell surface biotinylation and quantitative analysis of Western blots after immunoprecipitation. Values for DLL1 and DLL4 see Figure 4—source data 3. ND: due to closely co-migrating background band protein levels could not be quantified. Surface expression validated by biotinylation of ES cells and antibody staining of PSMs.
  DOI: https://doi.org/10.7554/eLife.40045.024
- Supplementary file 2. Relative ligand protein expression level in ES cell clones. The protein level of three independent clones used for co-culture studies was determined by quantitative analysis of Western blots and normalized to DLL1 clone #1 protein level measured in the same assay. Values for DLL1 and DLL4 see Figure 4—source data 2. ND: due to closely co-migrating background band protein levels could not be quantified.
  DOI: https://doi.org/10.7554/eLife.40045.025
- Transparent reporting form
  DOI: https://doi.org/10.7554/eLife.40045.026

Data availability
All data generated or analysed during this study are included in the manuscript and supporting files and source data files.

References

138
The DOI: [10.1083/jcb.200702009], PMID: [17664336]


Appendix 2

Supplemental figures for CHAPTER 3
Fig. S1. NRARP inhibits reporter gene induction by all four human Notch receptors. NIH 3T3 cells were transiently transfected with pcDNA3-based plasmids for expression of NRARP and NICD1 (A), NICD2 (B), NICD3 (C), or NICD4 (D), a firefly luciferase reporter plasmid under control of the TP1 Notch-response element, and a plasmid expressing Renilla luciferase. Firefly luciferase activity is reported relative to that of the Renilla luciferase, setting the firefly:Renilla ratio in cells transfected with empty pcDNA vector control to a value of one. n=2 (independent experiments).
**Fig. S2.** Electron density and unit cell features. A. Electron density map of the region at the Notch1-NRARP interface, contoured at 1σ after refinement. B. Ribbon representation of the asymmetric unit of the NRARP-Notch1-RBPJ complex on DNA. The protein assemblies in the two copies of the complex in the asymmetric unit are green and blue, and the DNA is orange. C. Overlay of the NRARP-NICD1-RBPJ complex (NRARP pink, NICD1 blue, RBPJ green, and DNA orange) onto the dimeric RBPJ-NICD1-MAML1-DNA complex (gray; PDB ID code 3NBN).
Table S1. Mass spectrometry results from tandem affinity purification of NRARP complexes.

The full list of interactors from multiple experimental repeats and control experiments can be accessed following the instructions provided below.

Protein accession IDs, names, gene symbols, number of times the protein has been identified in control immunoprecipitates (tandem detections) and number of gene-unique peptides are given in columns A-E. In the NRARP-tagged samples, only proteins that have had fewer than two tandem detections are listed. Data are from four independent experiments and two control immunoprecipitations in parental cells lacking NRARP. Native mass spectrometry data files are freely available for download from the MassIVE data archive hosted at the University of California, San Diego (ftp://massive.ucsd.edu) with the accession code/identifier MSV000084131.
Appendix 3

Bispecific Forkhead Transcription Factor FoxN3 Recognizes Two Distinct Motifs with Different DNA Shapes

This appendix has been published as:

Author Contributions:

I assisted in the development of the standard operating procedure for production and purification of the proteins used for biochemical and biophysical characterization. I was also involved in manuscript revision and editing.
Bispecific Forkhead Transcription Factor FoxN3 Recognizes Two Distinct Motifs with Different DNA Shapes

Julia M. Rogers, 1,2 Colin T. Waters, 1,3 Tom C.M. Seegar, 4 Sanchez M. Jarrett, 3,4 Amelia N. Hallworth, 1 Stephen C. Blacklow, 2,3,4,5,6,* and Martha L. Bulyk, 1,2,3,5,7,*

1Division of Genetics, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA 02115, USA
2Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA 02138, USA
3Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA 02138, USA
4Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
5Department of Pathology, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA 02115, USA
6Department of Cancer Biology, Dana Farber Cancer Institute, Boston, MA 02215, USA
7Lead Contact
*Correspondence: stephen_blacklow@hms.harvard.edu (S.C.B.), mlbulyk@genetics.med.harvard.edu (M.L.B.)
https://doi.org/10.1016/j.molcel.2019.01.019

SUMMARY

Transcription factors (TFs) control gene expression by binding DNA recognition sites in genomic regulatory regions. Although most forkhead TFs recognize a canonical forkhead (FKH) motif, RYAAAYA, some forkheads recognize a completely different (FHL) motif, GACGC. Bispecific forkhead proteins recognize both motifs, but the molecular basis for bispecific DNA recognition is not understood. We present co-crystal structures of the FoxN3 DNA binding domain bound to the FKH and FHL sites, respectively. FoxN3 adopts a similar conformation to recognize both motifs, making contacts with different DNA bases using the same amino acids. However, the DNA structure is different in the two complexes. These structures reveal how a single TF binds two unrelated DNA sequences and the importance of DNA shape in the mechanism of bispecific recognition.

INTRODUCTION

Transcription factors (TFs) must accurately distinguish target sites from the rest of the genome in order to properly regulate gene expression. Structural studies have revealed mechanisms used by many TFs to contact specific DNA sequences, largely by hydrogen bonding interactions between amino acid side chains and DNA bases in the major groove (Garvie and Wolberger, 2001; Harrison, 1991). Global structural features of DNA can also be recognized through a shape readout mechanism, involving contacts to the DNA backbone instead of directly to the bases (Otwinski et al., 1988). TFs can also recognize particular structural features of DNA, such as a narrow minor groove (Rohs et al., 2009).

Forkhead TFs are one of the major TF families in eukaryotes and play prominent roles in development, immunity, metabolism, and cell cycle control (Lam et al., 2013). Within this family, structural studies have revealed how members of this family bind sequences closely matching the canonical forkhead motif, RYAAAYA, and a related lower affinity motif, AHAACA (Boura et al., 2010; Brent et al., 2008; Clark et al., 1993; Li et al., 2017; Littler et al., 2010; Stroud et al., 2006; Tsai et al., 2006, 2007). Forkhead factors share a canonical forkhead DNA binding domain (DBD), which adopts a winged-helix fold, a modification of the helix-turn-helix DBD, to interact with DNA. In this motif, a three-helix bundle presents the third helix, the recognition helix, into the major groove of DNA. Additionally, two loop structures, known as the wings, make additional DNA contacts, typically to the DNA backbone.

We previously showed that some forkhead proteins can recognize an alternate DNA motif, GACGC, with some individual proteins able to bind both the canonical forkhead (FKH) and alternate forkhead-like (FHL) motifs (Nakagawa et al., 2013). These two motifs are different lengths and GC content, and their sequences are divergent enough that there is no clear sequence-based alignment between them; therefore, it is not clear how the documented binding mechanism to the FKH motif could also enable binding to the FHL motif. Previously determined mechanisms of recognition of multiple motifs by a single TF involve binding sequences of the same length that can be aligned (Badis et al., 2009; Gorda’n et al., 2011; Morgunova et al., 2018). In other cases, dimeric TFs, such as bZIP TFs, can bind motifs comprising very similar or identical DNA half-sites separated by different spacer lengths. Recognition of different DNA sequence motifs can also be achieved by the binding of different sets of fingers, linkers, or flanking regions within a multi-fingered C2H2 zinc finger protein (Siggers et al., 2014). However, the bispecificity observed for forkheads is achieved by a single, monomeric DBD, binding two sites of different lengths and very different sequences, and thus cannot be explained by these mechanisms.

Human FoxN3 was first identified as a suppressor of checkpoint mutations in S. cerevisiae (Pati et al., 1997). It has been shown to act as a transcriptional repressor and interacts with
histone deacetylase complexes involved in the DNA damage response (Busygina et al., 2006; Scott and Plon, 2003, 2005). FoxN3 has also been implicated in craniofacial and eye development and in regulation of metabolism and the cell cycle (Chang et al., 2005; Huot et al., 2014; Karanth et al., 2016; Markowski et al., 2009; Nagel et al., 2017; Samaan et al., 2010; Schmidt et al., 2011; Schuff et al., 2007; Sun et al., 2016). The molecular mechanisms by which FoxN3 carries out these diverse functions remain unclear.

Here, we show that FoxN3 is a bispecific TF that binds both the FKH and FHL sites in cells. We report the co-crystal structures of the bispecific human protein FoxN3 in complex with both FKH and FHL consensus sequences. The structures reveal that the forkhead DBD adopts remarkably similar structures to contact both motifs, using the same residues to specifically recognize two distinct DNA motifs. However, the shape of the DNA, particularly the bend of the DNA helix, throughout the recognition motif is strikingly different between the structures.

RESULTS

FoxN3 Is a Bispecific Transcription Factor

Human FoxN3 is a member of the FoxN forkhead subfamily, which contains bispecific and FHL monospecific TFs (Nakagawa et al., 2013). We assayed the binding specificity of FoxN3 by universal protein binding microarray (PBM) and found that the FoxN3 DBD recognizes both the FKH and FHL motifs (Figures 1A and 1B). We also measured the binding affinity of FoxN3 to DNA oligonucleotides containing the FKH or FHL sequence and showed that FoxN3 binds both sequences with mid-nanomolar affinity (Figure 1C). The $K_d$ to the FKH site is 60 ± 20 nM and to the FHL site is 238 ± 69 nM.

The ability of a forkhead factor to recognize both the FKH and FHL sites in the same cells in vivo has not been reported in prior studies. Therefore, we performed chromatin immunoprecipitation sequencing (ChIP-seq) experiments on FoxN3 and found that FoxN3 also binds both motifs in HepG2 cells (Figures 2A, 2B, and S1; Table S1). The FKH ($p = 1.3 \times 10^{-21}$) and FHL ($p = 3.3 \times 10^{-8}$) motifs are both enriched among the top 2,000 ChIP-seq peaks (Figure 2C), with centralization of both motifs (Figures 2D and 2E), supporting direct DNA binding of FoxN3 to these motifs within the peaks. 78 of these peaks contained an FHL match, and 365 contained an FKH match (1 peak had matches to both motifs).

FoxN3 Adopts the Canonical Winged-Helix Fold to Bind Both the FKH and FHL DNA Sequences

In order to understand how the forkhead domain contacts these two sites, we determined co-crystal structures of the DBD of FoxN3 in complex with the FKH and, separately, the FHL consensus sequence (Figures 3A and 3B; Table 1). In both structures, FoxN3 adopts the same overall forkhead winged-helix fold that has been observed for other forkhead proteins (Boura et al., 2010; Brent et al., 2008; Clark et al., 1993; Li et al., 2017; Littler et al., 2010; Stroud et al., 2006; Tsai et al., 2006, 2007). In the FoxN3:FHL structure, two DBDs bind the same DNA in the crystallographic asymmetric unit (Figure S2A). Molecule A directly contacts the FHL motif within the DNA, and molecule B contacts the end of the crystallized DNA sequence, making contacts with a weak match to an FKH site created by formation of a pseudo-continuous DNA helix in the crystal (Figure S2B). Therefore, for the rest of the analysis of the FHL site presented here, we will discuss molecule A of this structure. The root-mean-square deviation (RMSD) between the forkhead domain in the FKH-bound structure and that in the FHL-bound structure is 0.533 Å. In both structures, there is no observed electron density for wing 1 for amino acid (aa) positions 178–185, indicating that this wing may be flexible when FoxN3 binds DNA. Wing 2 is partially a-helical, with electron density extending to aa position 207 in the FKH structure and position 210 in the FHL structure.

FKH and FHL DNA Adopt Different Shapes in Complex with FoxN3

Although the overall protein conformation is similar between the FKH-contacting and FHL-contacting structures, the DNA molecules in the two structures adopt different conformations (Figures 3A and 3B). The DNA is bent away from the FoxN3 DBD by 22.2° in the FHL structure but toward the DBD in the FKH
structure by 13.5°. This bend toward the protein in the FKH structure is consistent with the DNA conformation in other co-crystal structures of forkhead domains with the FKH motif, but the FHL DNA has a very different conformation from that observed in any other published forkhead co-crystal structure (Boura et al., 2010; Brent et al., 2008; Clark et al., 1993; Li et al., 2017; Littler et al., 2010; Stroud et al., 2006; Tsai et al., 2006, 2007; Figures 3C and 3D).

This difference in the DNA shape explains how the same DBD can recognize two motifs of different lengths: the bend in the FHL DNA reduces the number of DNA base pairs that contact the DBD (Figure 4A). The guanines at the 5’ end of both the FKH (G7) and FHL (G10) motifs are in the same position with respect to the protein, as are G12’ in the FKH motif and G14’ in the FHL motif. These guanines can be thought of as “registration positions” that orient the DNA with respect to the protein. Between these two positions, the FKH DNA is bent 6.5° toward the protein, and FHL is bent 9.5° away. In the FKH motif, there are four base pairs between these two positions, and there are three in the FHL motif. In order to accommodate one fewer base in the FoxN3-FHL structure, many aspects of DNA shape are different between this structure and the FoxN3-FKH structure. The minor groove width in the FHL motif is much narrower than that in the FKH motif, averaging 5.90 Å over the FHL motif compared to 7.05 Å over the FKH motif (Figures 3C and 3E). Additionally, the helical rise (the distance between adjacent base pairs in the DNA helix) is much larger in the FHL structure than in the FKH structure (average of 3.38 Å over the FHL motif

Figure 2. FoxN3 Recognizes Both Motifs in Cells
(A and B) Browser shot for an ~40-kb region of chromosome 3 flanking a peak with an FKH binding site (A) or a ~40-kb region of chromosome 15 flanking a peak with an FHL binding site (B). An additional track for ChromHMM genome segmentations is displayed (red regions are annotated as promoters, yellow regions are annotated as enhancers, and green regions are annotated as regions of transcription; Ernst and Kellis, 2012).
(C) A table summarizing enrichment p values for FKH or FHL motifs in individual and pooled replicate samples, as well as the percent of the top 2,000 peaks from the pooled sample that contain matches to each motif.
(D and E) Composite profile of binding site location within the top 2,000 peaks after scanning with either the FKH (D) or FHL (E) motif. LOESS fits of the data are displayed in black.

See also Figure S1 and Table S1.
Comparison to 3.17 Å over the FKH motif; Figures 3D and 3F). The helical twist is lower in the FKH structure (average 31.4°) than in the FHL structure (average 33.8°), although roll is higher in FKH (average 2.7°) than in FHL (average —0.5°; Figures S3A and S3B).

In the two DNA conformations, the sequence of the intervening bases between the recognition positions is very different; these sequences may be partially specified by their propensity to form the required DNA structure, as well as by specific interactions with base-contacting protein side chains. The predicted DNA shape is different from that observed in the crystal structure for both the FKH and FHL sequences, so some deformation from inherent shape must occur to accommodate protein binding (Figures S3C–S3H). Specific DNA base contacts are largely made through the recognition helix, helix 3, in both structures (Figures 4B and 4C). In both structures, Arg163 makes bidentate hydrogen bonds to one of the guanine registration positions: G12 in the FKH structure and G14 in the FHL structure. The other conserved base-contacting positions in the recognition helix (His164 and Asn160) contact different bases in the two complexes (Figures 4B, 4C, 4A, and S4B). In both structures, contacts to the DNA backbone are made by the bases of wing 1 and helix 1 (Figures S4C–S4F).

In contrast, wing 2 adopts different positions in the two structures (Figure 4D). In the FKH structure, wing 2 is angled closer to helix 1, not making any direct ordered contacts to the DNA backbone. In the FHL structure, wing 2 extends toward the DNA backbone and contacts the backbone of G10 through His209 (Figure 4E).

Chimeric Proteins Reveal the Importance of the Wings in FHL Binding
These observed differences between the structures suggest that the wings may contribute to DNA binding specificity. Therefore, we created chimeric proteins in which we swapped segments of the forkhead DBD between FoxN3 and the FKH monospecific FoxJ3 (Table S2). Replacing the wings of FoxN3 with the wings of FoxJ3 significantly reduced binding to the FHL site while maintaining binding to the FKH site (Figures 5 and S5). Performing these swaps in the context of a different bispecific forkhead protein, FoxN2, also showed the requirement of the wings for FHL binding, indicating that this effect is not limited to FoxN3.
as the wing swap was not sufficient on its own. However, none of the tested combinations of swaps could increase FHL binding, indicating that a more distributed set of amino acids throughout the forkhead domain is required for FHL binding.

**DISCUSSION**

The structures presented in this work demonstrate the ability of the forkhead DBD to recognize vastly different DNA sequences using the same overall fold and DNA-contacting residues. The FKH and FHL sites are much more dissimilar than other examples of sequences bound by the same DBD (Badis et al., 2009; Gordań et al., 2011). Structures of FoxO1 in complex with the FKH primary site and a related lower affinity motif (AHAAACA) showed that this protein also rearranges the interactions of the amino acids in the recognition helix to bind different sequences (Brent et al., 2006). The FoxN3 structures presented here reveal even more flexibility in the interactions of these amino acids, as they can not only tolerate substitutions leading to DNA sequences still related to the FKH motif but also interact with the highly divergent FHL motif of different sequence length. The interactions made by FoxN3 with the registration positions in both the FKH and FHL motifs are enabled by the difference in DNA shape, orienting the FHL DNA in a strikingly different conformation than the FKH DNA. The structures presented here highlight that DNA motif recognition cannot be cleanly partitioned into sequence versus shape recognition, as TFs recognize the totality of the DNA structure. The role of DNA shape in motif recognition represents a different mechanism of binding to distinct DNA sequences than others previously reported. For example, dimeric TFs can bind DNA half-sites with different spacer lengths (Badis et al., 2009; Gordań et al., 2011). Alternatively, the monomers of a dimeric TF can adopt different protein conformations to recognize distinct half-site sequences (Kalodimos et al., 2002). Monomeric HoxB13 can recognize two sequences (CCAATAAA and CTCGTAATA) using different contributions of enthalpy and entropy to binding (Morgunova et al., 2018). FoxN3 uses a different strategy, recognizing two DNA sequences without major structural differences within the DBD via differences in DNA shape.

Co-crystal structures of other forkhead TFs in complex with the FKH site show that these wings can adopt vastly different positions and conformations and so may be a wider source of functional divergence within this family (Boura et al., 2010; Brent et al., 2008; Clark et al., 1993; Littler et al., 2010; Stroud et al., 2006; Tsai et al., 2006, 2007). Previous studies of FKH-binding forkhead proteins showed that swapping wings between two forkhead proteins was sufficient to switch their preferences for the flanking sequences surrounding the core FKH motif (Pierrou et al., 1994). Our study reveals that the protein sequence of the wings can exert a much more dramatic effect in determining specificity for the core recognition sequence.

Lastly, the differences in DNA recognition between the two complexes might have gene regulatory consequences. For example, different co-factors might interact with FoxN3 when bound to an FKH versus FHL site; this mode of sequence readout by TFs affecting co-factor recruitment has been observed for single-nucleotide differences in nuclear factor kb (NF-kb) binding sites, suggesting that the much more dramatic

(Figure S5). A shorter stretch of 6 amino acids in wing 2, L199-K204, recapitulated the effect of swapping the entire wing 2 sequence. This stretch of amino acids is conserved between the bispecific proteins FoxN2 and FoxN3 (Figure S6) and does not directly contact DNA in either structure. Interestingly, swapping the sequence of the FoxN3 wings into FoxJ3 did not increase FHL binding. These findings reveal that features in the wings of the proteins that bind the FHL motif are necessary, but not sufficient, for binding this motif, indicating allosteric interactions with other regions of the protein.

Given that the repositioned portion of wing 2 is important for DNA binding specificity, we also tested other portions of the domain that adopted different positions in the two structures: helix 4, the short helix preceding the recognition helix, and the N-terminal loop. Neither of these regions affected binding to either the FKH or FHL sequence (Figure S5). The N-terminal loop contacts wing 2 and helix 4, suggesting that combining these swaps may be sufficient to confer FHL binding to FoxJ3.
Figure 4. Contacts between FoxN3 and the FKH and FHL Sites
(A) The guanine bases at the end of each DNA strand in both motifs are in the same position with respect to the protein. The helical axis calculated by Curves+ (Lavery et al., 2009) is shown for both structures.
(B) Specific contacts between FoxN3 (magenta) and the FKH DNA (yellow) are shown.
(C) Contacts between these same amino acids in FoxN3 (blue) and the FHL DNA (green) are shown.
(D) The FKH structure (magenta protein and yellow DNA) and the FHL structure (blue protein and green DNA) are aligned to highlight the differences in positions of wing 2.
(E) In the FHL structure, wing 2 contacts the DNA backbone through His209.
See also Figure S4.

Differences over the lengths of the FKH versus FHL motifs might also contribute to different regulatory output (Leung et al., 2004). Overall, the work presented here reveals the ability of the FoxN3 DBD to recognize two DNA bind sites of vastly different DNA sequences and structures, highlighting the surprising plasticity of DNA recognition within the forkhead family. Given the prevalence of diverse DNA binding motif preferences within numerous TF families, such flexibility in TF-DNA recognition may be a more universal principle and might play an important role in the evolution of gene regulatory networks (Badis et al., 2009; Gordon et al., 2011; Nakagawa et al., 2013).

STAR+METHODS

Detailed methods are provided in the online version of this paper and include the following:

KEY RESOURCES TABLE
CONTACT FOR REAGENT AND RESOURCE SHARING

EXPERIMENT MODEL AND SUBJECT DETAILS

Cell culture
Protein Production

METHOD DETAILS

Preparation of whole cell lysates
Western blot
Cross-linking and harvest of cells for ChIP
Sonication of cross-linked material for ChIP
Immunoprecipitation
Reversal of crosslinks
Preparation of sequencing libraries and sequencing
ChIP-seq data processing and analysis
Cloning
Protein Purification
Crystallization and Data Collection
Structure Determination
DNA Shape Analysis
Protein Binding Microarrays
MicroScale Thermophoresis
Figure 5. Subdomain Swap Experiments Show the Involvement of the Wings in FHL Recognition

(A) Chimeric proteins were designed to test the importance of parts of the forkhead DBD for binding specificity. Positions are numbered with respect to the full-length FoxN3 protein.

(B) The location of the swaps on the forkhead structure is shown.

(C) Boxplots show PBM E-scores for 8 mers containing the FKH site (GTAAACA; top panel) or the FHL site (GACGC; bottom panel) for each chimeric protein. FoxN3 and chimeras of it are shown in blue, and FoxJ3 is shown in red. The FoxN3 boxplot represents the average of three replicate PBM experiments, and N2(J3-wing) and N3(J3-6aa) represent the average of tworeplicates. Boxplots for individual replicates are shown in Figure S5. * indicates p < 9.3 × 10^{-16}; one-sided Mann-Whitney test. See also Figures S5 and S6 and Table S2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Protein Binding Microarray Analysis
MST Affinity Analysis

DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and two tables and can be found with this article online at https://doi.org/10.1016/j.molcel.2019.01.019.

ACKNOWLEDGMENTS

We thank Jesse V. Kurland for performing a PBM experiment and Katy Weinand for assistance in ChIP-seq analysis. We thank Steve Gisselbrecht, Kelly Arnett, Karen Adelman, Remo Rohs, and Kevin Struhl for helpful discussion. We thank The Bauer Core Facility at Harvard University for high-throughput sequencing services. We thank the Snyder lab and the ENCODE Consortium for generating and hosting FOXA1 ChIP-seq data. Microscale thermophoresis (MST) experiments were performed at the Center for Macromolecular Interactions at Harvard Medical School. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the NIH (P41 GM103403). The Pilatus 6M detector on 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). The Eiger 16M detector on 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. A.N.H. was supported by NSF Dbl 1263215 and NIH R25HL121029. This work was supported by NIH grant R01 HG003985 to M.L.B., R35 CA220340 to S.C.B., and by National Science Foundation Graduate Research Fellowships to J.M.R. and C.T.W.

AUTHOR CONTRIBUTIONS


### STAR+METHODS

#### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibodies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-FoxN3</td>
<td>Abgent</td>
<td>AP19255B; RRID: AB_2783721</td>
</tr>
<tr>
<td>Anti-GST</td>
<td>Sigma</td>
<td>G7781; RRID: AB_259965</td>
</tr>
<tr>
<td>HRP-conjugated goat anti-rabbit</td>
<td>Thermo Fisher</td>
<td>31460; RRID: AB_228341</td>
</tr>
<tr>
<td>Alexa488-conjugated anti-GST</td>
<td>Thermo Fisher</td>
<td>A-11131; RRID: AB_2534137</td>
</tr>
<tr>
<td>Bacterial and Virus Strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Shot® BL21(DE3) Chemically Competent E. coli</td>
<td>Thermo Fisher Scientific</td>
<td>C600003</td>
</tr>
<tr>
<td>Chemicals, Peptides, and Recombinant Proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMEM</td>
<td>HyClone</td>
<td>SH30022.01</td>
</tr>
<tr>
<td>Fetal Bovine Serum</td>
<td>GIBCO</td>
<td>10082139</td>
</tr>
<tr>
<td>Formaldehyde</td>
<td>Sigma</td>
<td>F87750</td>
</tr>
<tr>
<td>Protein G Sepharose beads</td>
<td>Sigma</td>
<td>P3296</td>
</tr>
<tr>
<td>Salmon Sperm DNA</td>
<td>Sigma</td>
<td>D9156</td>
</tr>
<tr>
<td>BSA</td>
<td>New England Biolabs</td>
<td>B9000S</td>
</tr>
<tr>
<td>Recombinant GST</td>
<td>Sigma</td>
<td>G5663</td>
</tr>
<tr>
<td>Critical Commercial Assays</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QIAquick PCR purification kit</td>
<td>QIAGEN</td>
<td>28106</td>
</tr>
<tr>
<td>NEBNext Ultra II DNA Library Prep Kit</td>
<td>NEB</td>
<td>E7645</td>
</tr>
<tr>
<td>PURExpress in vitro protein synthesis kit</td>
<td>NEB</td>
<td>E6800S</td>
</tr>
<tr>
<td>8x60K universal protein binding microarray</td>
<td>Agilent Technologies</td>
<td>AMADID #030236</td>
</tr>
<tr>
<td>Monolith NT.115 Standard Treated Capillaries</td>
<td>NanoTemper</td>
<td>MO-K002</td>
</tr>
<tr>
<td>Tapestation 2200 D1000 Screen Tape and Reagents</td>
<td>Agilent</td>
<td>5067-5582 5067-5583</td>
</tr>
<tr>
<td>Deposited Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FoxA1 ChIP-seq, HepG2 cells</td>
<td>Dunham et al., 2012</td>
<td>ENCFF396NXZ, ENCFF988UCQ, ENCFF190EPQ</td>
</tr>
<tr>
<td>FoxN3 ChIP-seq</td>
<td>This study</td>
<td>GEO, GSE112672</td>
</tr>
<tr>
<td>PBM data</td>
<td>This study</td>
<td>Uniprobe, <a href="http://the_brain.bwh.harvard.edu/uniprobe/">http://the_brain.bwh.harvard.edu/uniprobe/</a>, ROG18A</td>
</tr>
<tr>
<td>FoxN3:FKH structure</td>
<td>This study</td>
<td>PDB 6NCe</td>
</tr>
<tr>
<td>FoxN3:FHL structure</td>
<td>This study</td>
<td>PDB 6NCM</td>
</tr>
<tr>
<td>Experimental Models: Cell Lines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HepG2</td>
<td>ATCC</td>
<td>HB-8065</td>
</tr>
<tr>
<td>Oligonucleotides</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEBNext Singleplex Oligos for Illumina Kit</td>
<td>NEB</td>
<td>E7350</td>
</tr>
<tr>
<td>Crystallization oligo FKH-F TCTTAAAGTAAACAATG</td>
<td>This paper, IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>Crystallization oligo FKH-R ACATTGTCTACTTAAG</td>
<td>This paper, IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>Crystallization oligo FHL-F TCACTGCTAAGCAGCTA</td>
<td>This paper, IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>Crystallization oligo FHL-R ATAGCGCCTTACAGCAG</td>
<td>This paper, IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>MST oligo FKH_affinity_F F56-FAMTTAAGTAAACAATG CAAATGAG</td>
<td>This paper, IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>MST oligo FKH_affinity_R CTCATTGTTTTACTTAA</td>
<td>This paper, IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>MST oligo FHL_affinity_F F56-FAMGTGAGGACGCTA GCTATTATC</td>
<td>This paper, IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>MST oligo FHL_affinity_R GATAATAGCCTCCTCA</td>
<td>This paper, IDT</td>
<td>N/A</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recombinant DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pDEST17</td>
<td>Invitrogen</td>
<td>11803012</td>
</tr>
<tr>
<td>pDEST15</td>
<td>Invitrogen</td>
<td>11802014</td>
</tr>
<tr>
<td>pDEST17-FoxN3</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-FoxN3</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-FoxN2</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-FoxJ3</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-N3(J3-wing)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-N3(J3-6aa)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-J3(N3-wing)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-J3(N3-6aa)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-N3(J3-Nterm)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-N3(J3-loop)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-J3(N3-6aa+Nterm)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-J3(N3-loop)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-J3(N3-6aa+loop)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-J3(N3-6aa+Nterm+loop)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-N2(J3-wing)</td>
<td>This study</td>
<td>n/a</td>
</tr>
<tr>
<td>pDEST15-J3(N2-wing)</td>
<td>This study</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Software and Algorithms

<table>
<thead>
<tr>
<th>Software and Algorithms</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FastQC v0.11.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fastx-trimmer v0.0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowtie v1.1.1</td>
<td>Langmead et al., 2009</td>
<td></td>
</tr>
<tr>
<td>MACS v2.1.1.20160309</td>
<td>Feng et al., 2012; Zhang et al., 2008</td>
<td></td>
</tr>
<tr>
<td>HOMER v4.9</td>
<td>Heinz et al., 2010</td>
<td></td>
</tr>
<tr>
<td>GENRE</td>
<td>Mariani et al., 2017</td>
<td></td>
</tr>
<tr>
<td>Integrative Genomics Viewer</td>
<td>Robinson et al., 2011</td>
<td></td>
</tr>
<tr>
<td>ChromHMM</td>
<td>Ernst and Kellis, 2012</td>
<td></td>
</tr>
<tr>
<td>UCSC LiftOver</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XDS</td>
<td>Kabsch, 2010</td>
<td></td>
</tr>
<tr>
<td>Phenix</td>
<td>Adams et al., 2010</td>
<td></td>
</tr>
<tr>
<td>COOT</td>
<td>Emsley and Cowtan, 2004</td>
<td></td>
</tr>
<tr>
<td>PyMOL</td>
<td>The PyMOL Molecular Graphics System, Schrödinger, LLC.</td>
<td></td>
</tr>
<tr>
<td>Curves+</td>
<td>Lavery et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Universal PBM Analysis Suite</td>
<td>Berger and Bulyk, 2009</td>
<td></td>
</tr>
<tr>
<td>MO.Control version 1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MO.Affinity Analysis version 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GraphPad Prism 7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
Continued

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monolith NT.115pico</td>
<td>NanoTemper Technologies</td>
<td></td>
</tr>
<tr>
<td>Biorupter Twin circulating bath sonicator</td>
<td>Diagenode</td>
<td>UCD-400</td>
</tr>
<tr>
<td>Illumina NextSeq sequencer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mono-S 10/100 GL column</td>
<td>GE</td>
<td>17516901</td>
</tr>
<tr>
<td>Superdex 75 10/300 GL</td>
<td>GE</td>
<td>17517401</td>
</tr>
</tbody>
</table>

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Martha L. Bulyk (mlbulyk@genetics.med.harvard.edu).

EXPERIMENT MODEL AND SUBJECT DETAILS

Cell culture
HepG2 cells were purchased from ATCC (HB-8065). Cells were cultured in DMEM with High Glucose and 4.0 mM L-Glutamine, without Sodium Pyruvate (HyClone SH30022.01), and supplemented with 10% heat inactivated fetal bovine serum (GIBCO 10082139) and 1% penicillin/streptomycin.

Protein Production
One Shot® BL21(DE3) Chemically Competent E. coli cells were purchased from Thermo Fisher Scientific.

METHOD DETAILS

Preparation of whole cell lysates
Whole cell lysates were prepared by placing a 15-cm culture dish on ice, aspirating culture media, and washing once in 15 mL cold PBS. Two mL of ice cold RIPA buffer (150 mM NaCl, 1% NP-40 substitute, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0) was then added. Cells were scraped in RIPA buffer and transferred to a cold microcentrifuge tube. The tube was then placed on a shaker platform at 100 rpm for 30 minutes at 4°C. After lysis, cell debris was pelleted by spinning at 14,000 rcf. for 20 minutes at 4°C.

The supernatant was removed, aliquoted into 300 mL aliquots, flash frozen in liquid nitrogen and stored at —80°C. One cOmplete ULTRA mini protease-inhibitor tablet was used per 10 mL of buffer (RIPA or PBS).

Western blot
Anti-FoxN3 antibody (Abgent AP19255B) was first evaluated for specificity via western blot against HepG2 whole cell lysate (Figure S1A). Ten to fifteen mL of whole cell lysate was run on a 4%–12% Criterion Bis-Tris acrylamide gel (Bio-Rad 3450125), and was blotted with a 1:100 dilution of primary antibody, followed by 1:2,000 dilution of an HRP-conjugated goat anti-rabbit secondary (Thermo Fisher #31460).

Cross-linking and harvest of cells for ChIP
To prepare material for ChIP-seq, cells were grown on 15-cm culture dishes. Two independent passages were maintained simultaneously, harvested independently, and processed in parallel, yielding two biological replicate datasets. Plates were removed from the incubator and placed at room temperature. Formaldehyde (Sigma F87750) was added directly to the culture medium to a final concentration of 1%. Plates were incubated at room temperature for 10 minutes, with swirling every 2.5 minutes. Crosslinking was quenched by adding 2.5 M stock glycine to a final concentration of 0.125 M and plates were swirled to mix. Media was then aspirated and cells were washed once in 1X PBS. Eight milliliters of cold lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40 substitute, prepared fresh and filter sterilized) was then added directly to cells on the dish. Cells were then scraped, transferred to a 15-cm conical tube, and spun at 2000 rpm for 5 minutes at 4°C. The supernatant was then removed and flash frozen in liquid nitrogen at a concentration of 2x10⁷ cells/mL for storage at —80°C.

Sonication of cross-linked material for ChIP
A 1-mL aliquot of cells was thawed and gently resuspended before being passed through a 20-gauge needle 20 times. Crude nuclear prep was then collected by spinning the lysate at 2000 rpm for 5 minutes at 4°C. The resulting pellet was resuspended in 300 mL RIPA
buffer and then processed in a Biorupter Twin circulating bath sonicator for 50 cycles of 30 s on and 30 s off at the high setting in a 4°C environmental chamber. Following sonication, samples were spun at 16,000 rcf for 15 minutes at 4°C and the supernatant was either snap frozen and stored at —80°C or used as input for immunoprecipitation. Prior to immunoprecipitation, the total protein concentration of each sample was measured by Bradford assay, and all sample concentrations were normalized to 1.75 mg/mL with 1X RIPA buffer and split into 1-mL aliquots.

Immunoprecipitation
Sixty mL of Protein G Sepharose beads (Sigma P3296) were used per sample. Beads were washed twice in 1X PBS and resuspended in 1X PBS to 60 mL. Thirty mL of washed beads were blocked by adding 9 mL of 0.3 mg/mL salmon sperm DNA and 12 mL 1 mg/mL BSA and incubating at 4°C for 1 hour on a rotisserie. The remaining 30 mL was added directly to a 1-mL sonicated cell aliquot and incubated for 1 hour at 4°C on a rotisserie to pre-clear the lysate. Following clearing, the sample was spun at 2500 rcf. for 1 minute and 950 mL of supernatant was transferred to a new tube. Ten micrograms of the anti-FoxN3 antibody (Abgent AP19255B) and 30 mL of blocked bead slurry was then added and samples were incubated at 4°C overnight with rotation on a rotisserie. A ‘no antibody’ negative control sample was also prepared from a paired aliquot by incubating overnight with 30 mL of blocked bead slurry, without addition of the primary antibody. Approximately 18 hours later, samples were washed at 4°C by spinning at 2,500 rcf. for 1 minute, adding 1 mL of the following buffers, incubating for 1 minute, and repeating. Washes began with 2 1X PBS washes, 4 washes with IP wash buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1% NP-40 substitute, and 1% sodium deoxycholate), and 1 wash with PBS-RIPA buffer (1X PBS, 1% NP-40 substitute, 0.5% sodium deoxycholate, and 0.1% SDS). Immunoprecipitated material was eluted off the beads by adding 200 mL elution buffer (70 mM Tris-Cl pH 8.0, 1 mM EDTA, 1.5% SDS), and incubating at 65°C for 10 minutes with vortexing every 2 minutes. Samples were then spun at 2500 rcf. for 2 minutes and the supernatants were transferred to fresh tubes.

Reversal of crosslinks
To each sample (antibody or no antibody), 13 mL of 4 M NaCl (200 mM final) was added, and samples were incubated at 65°C for 18 hours. Subsequently, 20 mg of Proteinase K was added to each sample and incubated at 45°C for 60 minutes. Five volumes of QIAGEN Buffer PB (QIAquick PCR purification kit) was then added to one volume of sample. This material was then processed according to the manufacturer’s instructions for the QIAquick PCR purification kit, eluting twice with 30 mL of pre-warmed QIAGEN Buffer EB.

Preparation of sequencing libraries and sequencing
Libraries for sequencing with Illumina high-throughput sequencing chemistry were prepared using the NEBNext Ultra II DNA Library Prep Kit (NEB E7645) with adaptors from the NEBNext Singleplex Oligos for Illumina Kit (NEB E7350) according to manufacturer’s instructions. A final cycle number of 12 cycles of amplification was required during the PCR step. Libraries were multiplexed at equal concentrations based on integration under TapeStation D1000 tape traces and sequenced on an Illumina NextSeq instrument at the Bauer Core Facility at Harvard University.

ChIP-seq data processing and analysis
Read quality was evaluated using FastQC v0.11.5. Reads were trimmed to 36 nt to remove lower quality 3′ end base calls using fastx-trimmer v0.0.13. Reads were then aligned to hg38 using bowtie v1.1.1 (bowtie -n 2 -m 1) (Langmead et al., 2009). Aligned reads from replicate sequencing runs were pooled either within biological replicates or across all replicates. Peaks were then called on pooled aligned reads using MACS2 v2.1.1.20160309 (macs2 callpeak -g hs -bw 400) (Feng et al., 2012; Zhang et al., 2008). Following peak calls, enriched motifs in the top 2,000 peaks, as ranked by the -log10(q-value) for each peak, were detected using HOMER v4.9 (findMotifsGenome.pl -size 50 -len 6,8,10 with hg38) (Heinz et al., 2010). Motif enrichment p values were calculated by the HOMER findMotifsGenome.pl script against a binomial null distribution.

The top 2,000 peaks were partitioned according to the presence of a match to the FoxN3 PBM-derived FKH or FHL motifs using an empirically-derived log odds detection threshold in HOMER (annotatePeaks.pl -size 50 or -size 200 with hg38). Composite profiles were generated using the HOMER annotatePeaks.pl function (-m -hist 10 -size 2000 with hg38). Plots were produced using R v3.2.4. The background dataset used for motif centralization was generated using the GENRE utility, matching on dinucleotide content and promoter enrichment (Mariani et al., 2017). Genome browser shots were depicted using Integrative Genomics Viewer (IGV), and chromatin state was analyzed using ChromHMM (Ernst and Kellis, 2012; Robinson et al., 2011). The top 2,000 peaks, annotated for FKH and FHL motif occurrence, are provide in Table S1.

FoxA1 ChIP-seq data from HepG2 cells were downloaded from the ENCODE repository (ENCFHF396NXZ and ENCFH988UQCQ) with corresponding control data (ENCFHF910EPO) (Dunham et al., 2012). Peaks were called using MACS2, as above, and the top 2,000 peaks as ranked by -log10(q-value) were analyzed for the presence of matches to either the FKH or the FHL motif, as above. Composite profiles were generated using the HOMER annotatePeaks.pl function (-m -hist 5 -size 2000 with hg38).

Cloning
Forkhead DBDs were generated through gene synthesis, flanked by Gateway attB recombination sites (GenScript USA, Inc.; IDT). For protein expression, the FoxN3 construct was transferred into the pDEST17 vector, which confers an N-terminal 6xHis tag, using
the Gateway cloning system (Invitrogen). The sequence encoding the TEV protease cleavage site (ENLYFQG) was inserted between the pDEST17 recombination site and the beginning of the FoxN3 sequence, in order to enable tag cleavage. For use in PBM experiments, constructs were transferred into pDEST15, which confers an N-terminal Glutathione S-transferase (GST) tag, using the Gateway system (Invitrogen). All cloned protein sequences are provided in Table S2.

**Protein Purification**

To produce protein for crystallization, BL21(DE3) cells were transformed with pDEST17 containing the FoxN3 DBD, grown at 37°C to OD∞00 of 0.6, and protein production was induced with 0.5 mM IPTG overnight at 16°C. Cells were harvested, resuspended in lysis buffer (20 mM Tris HCl pH7.7, 300 mM NaCl, 0.03% Triton X-100, 5 mM beta-mercaptoethanol, and protease inhibitor tablets (Roche)), lysed by sonication and clarified. Protein was bound to Ni-NTA beads (Qiagen), washed in wash buffer (20 mM Tris HCl pH 7.7, 300 mM NaCl, 0.03% Triton X-100, 10 mM imidazole, 5 mM 1,4-dithiothreitol (DTT)), and eluted in elution buffer (20 mM Tris HCl pH 7.7, 300 mM NaCl, 250 mM imidazole (500 mM in final elution), 5 mM DTT). The 6xHis tag was removed by overnight digestion with Tobacco Etch Virus (TEV) at 4°C. The cleaved protein solution was re-bound to Ni-NTA beads to capture the cleaved tag, and the flow-through was collected and loaded on to a Mono-S 10/100 GL column (GE) and eluted in 20 mM Tris HCl pH 7.7, 5 mM DTT with a 0 to 1 M NaCl gradient. Peak fractions were pooled and further purified by size exclusion chromatography on a Superdex 75 10/300GL column (GE) in 20 mM Tris HCl pH 7.7, 150 mM NaCl, 2 mM tris(2-carboxyethyl)phosphine (TCEP). The protein was either used fresh, or concentrated to 1 mg/mL and flash frozen in 20% glycerol for later use.

Protein for PBM experiments was produced through in vitro transcription and translation using the PURExpress in vitro protein synthesis kit (New England Biolabs), according to the manufacturer’s instructions. Protein quality and concentration was assessed by anti-GST western blot with a dilution series of a recombinant GST standard (Sigma G5663). The western blot was performed with 20 ng/mL anti-GST primary antibody (Sigma G7781) and a 1:2000 dilution of HRP-conjugated goat anti-rabbit secondary antibody (Thermo Fisher #31460).

**Crystallization and Data Collection**

DNA oligonucleotides (IDT) for crystallization were resuspended to 100 mM in DNA hybridization buffer (10 mM Tris HCl pH 8, 50 mM NaCl), and annealed by mixing the forward and reverse sequences at a 1:1 ratio, heating to 95°C for 5 minutes, and slowly cooling to room temperature overnight. The DNA sequences are: FKH-F (5’TCTTAAGTAAACAATG-3’), FKH-R (5’ACATTTTACCTTAGA-3’), FHL-F (5’TCATGTACAGCGTACT-3’), and FHL-R (5’ATAGCGTTCTAGCAT-3’). These sequences were selected from among PBM probes that were highly bound by bispecific forkhead proteins in our previous study (Nakagawa et al., 2013).

FoxN3 was mixed with the annealed DNA at a molar ratio of 1:1.2, and the mixture was incubated on ice for 5 min. The complexes were concentrated to a final protein concentration of 4 mg/mL for the FKH complex, and 5 mg/mL for the FHL complex. Initial crystallization tests with the Natrix, Peg-Ion, and Peg-Rx screens (Hampton) were performed using the NT8 liquid handling robot, in sitting drop format. Crystals were optimized in hanging drops in 24-well format at room temperature. The FoxN3:FKH crystals formed in 0.1 M BisTris pH 5.4, 0.2 M MgCl2, 22% PEG 3350. The FoxN3:FHL crystals formed in 0.1 M BisTris pH 5.5, 0.2 M NaCl, 22% PEG 3350.

Crystals were harvested, moved to cryoprotectant solution (for FKH crystal: 0.1 M BisTris pH 5.4, 0.2 M MgCl2, 25% PEG 3500, 5% glycerol; for FHL crystal: 0.1 M BisTris pH 5.5, 0.2 M NaCl, 22% PEG 3350, 10% glycerol), and flash frozen with liquid nitrogen. Diffraction data was collected at the Advanced Photon Source, beamline 24-ID-C for the FoxN3-FKH structure, and 24-ID-E for the FoxN3-FHL structure (NE-CAT).

**Structure Determination**

Diffraction images were indexed using XDS (Kabsch, 2010). Phases were produced by molecular replacement in Phenix (Adams et al., 2010). For the FoxN3:FKH structure, the structure of FoxK1a (PDB 2C6Y, chains A,C,D) was used as a search model (Tsai et al., 2006). The DNA and protein structures from the FoxN3:FKH structure were used as search models for the FoxN3:FHL structure. Model building was done in COOT (Emsley and Cowtan, 2004), and structures were refined with Phenix (Adams et al., 2010) using reciprocal space optimization of xyz coordinates, individual atomic B factors, optimization of X-ray/stereochemistry weights and optimization of ADP weights for both structures. TLS was used in the refinement of the FoxN3:FKH structure. The crystallographic data table (Table 1) was generated using Phenix. Figures were created using PyMOL (Schrodinger, 2018).

**DNA Shape Analysis**

Protein-DNA contacts were identified using the DNAproDB tool (Sagendorf et al., 2017). Superpositions of the FKH and FHL structures were performed using the align function in PyMOL, aligning the FoxN3 protein in the two structures. DNA shape parameters were determined from the structures using Curves+ (Lavery et al., 2009). For MGW for C14 at the end of the FHL DNA, which Curves+ could not compute, the measurement tool in PyMOL was used to determine the distance between the phosphate group of the +2 and −2 nucleotides. 5.8 A was subtracted from the measurement, for concordance with the Curves+ results. Shape parameters for unbound DNA were predicted using the DNAshape webserver (Zhou et al., 2013). Average shape parameter values were calculated over the GTAAAC or GACGC bases for the FKH and FHL motifs, respectively.
Protein Binding Microarrays

PBM assays were carried out essentially as described, using our 8 3 60K “all 10-mer” universal array design (Agilent Technologies, AMADID #030236) (Berger and Bulyk, 2009; Berger et al., 2006). Briefly, GST-tagged forhead DBDs, synthesized by IVT, were applied to the double-stranded DNA array, and detected with a fluorescently conjugated anti-GST antibody (Thermo Fisher #A-11131). Full experimental conditions (protein concentration and buffers used for each protein) are provided in Table S2. Microarray data were quantified as described previously (Berger and Bulyk, 2009; Berger et al., 2006).

MicroScale Thermophoresis

Fluorescein-labeled oligonucleotides (IDT) were resuspended to 100 mM in DNA hybridization buffer (10 mM Tris HCl pH 8.0, 50 mM NaCl), and annealed by mixing the forward and reverse sequences at a 1:1 ratio, heating to 95°C for 5 minutes, and slowly cooling to room temperature overnight. Concentrations of the labeled, annealed oligonucleotidess were determined by measuring A260 and A495, and using the following formula, with ε260, fluorescein = 20960, ε495, fluorescein = 75000, ε260, FKH = 252,998, and ε260, FHL = 264,337.

\[
M = \frac{\frac{A_{260} - A_{495}}{\varepsilon_{260\text{-fluorescein}}} \cdot \varepsilon_{495\text{-fluorescein}}}{A_{260\text{-oligo}}} 
\]

MST reaction buffer was 12mM Tris pH7.5, 200mM KCl, 1mM DTT, 5mM MgCl2, 0.05% Tween-20. A 1:2 dilution series of FoxN3 from 7.5 mM to 228pM was made in MST reaction buffer, and incubated with 20nM DNA for at least 30 minutes before measuring binding. Binding was measured by MicroScale Thermophoresis using the NanoTemper Monolith NT.115pico. MST measurements were made at 22°C, at 30% excitation power for the FKH sequence, 50% excitation power for the FHL sequence, and medium MST power (40%) for both. 3 dilution series of FoxN3 were performed for each DNA, and each dilution series was read twice by MST, resulting in six replicates.

QUANTIFICATION AND STATISTICAL ANALYSIS

Protein Binding Microarray Analysis

Boxplots were generated by filtering all 8-bp sequences for those that match the indicated sequence (FKH: GTAAACA, FHL: GACGC). The PBM E-scores of those 8-mers for each protein were plotted using the boxplot function in R. The negative control GST PBM data are from Barrera et al. (Barrera et al., 2016). P-values were computed using the one-sided Mann-Whitney test function in R, and were corrected for multiple hypothesis testing using the Bonferroni correction.

MST Affinity Analysis

Fnorm was calculated from the MST traces, with the ‘cold region’ defined as —1 s to 0 s, and the ‘hot region’ defined as 4 s to 5 s. Individual capillary traces were visualized to remove fluorescence outliers outside 20% of the average fluorescence. Data were fit to a one-to-one binding model accounting for ligand depletion in Prism to determine Kd: Y = Unbound+ (Bound-Unbound)*((Ro + X + Kd) -sqrt((Ro + X + Kd)^2 - 4*Ro*X))/(2*Ro), where Ro = the total fluorescent oligo concentration, fixed at 20nM, X is the protein concentration, Bound and Unbound are the Maximum and Minimum binding response measurements, and Y is the binding measurement (here, Fnorm). For plotting, data were normalized to represent fraction bound: fraction = (unbound-Fnorm)/(unbound-bound).

DATA AND SOFTWARE AVAILABILITY

The accession number for the ChIP data reported in this paper is GEO: GSE112672. The accession number for the PBM data reported in this paper is UniPROBE: ROG18A, at http://the_brain.bwh.harvard.edu/uniprobe/. The accession number for the crystal structures reported in this paper are PDB: 6NCE and 6NCM.
Supplemental Information

Bispecific Forkhead Transcription Factor FoxN3
Recognizes Two Distinct Motifs
with Different DNA Shapes

Julia M. Rogers, Colin T. Waters, Tom C.M. Seegar, Sanchez M. Jarrett, Amelia N. Hallworth, Stephen C. Blacklow, and Martha L. Bulyk
Figure S1. Related to Figure 2

FoxN3 ChIP.
(a) Western blot of HepG2 whole-cell lysate, probed with anti-FoxN3 antibody (Abgent #AP19255B). Sizes from a molecular weight standard are indicated on the left side of the film, and the predominant band (indicated with the arrow) at ~70 kDa is the expected FoxN3 band. FoxN3 ChIP-seq read pile-up at FKH-containing peaks (b), FHL-containing peaks (c), or peaks without either motif (d). Composite profile of binding site location at the top 2,000 FOXA1 ChIP-seq peaks after scanning with either the FKH (e) or FHL (f) motif. LOESS fits of the data are displayed in black. Motif centrality analysis of background DNA regions matched for dinucleotide content and promoter enrichment, scanned for FKH (g) or FHL (h) motif matches.
Figure S2. Related to Figure 3

Molecule B in the FHL structure binds a weak match to the FKH motif created by the pseudocontinuous DNA helix.

(a) Two molecules of FoxN3 (chain A, blue, chain B, cyan), in complex with DNA containing the FHL motif (green): GACGC. (b) Molecule B is shown in cyan, DNA from the same crystallographic unit is in dark green, and the DNA from the next unit is shown in light green. The bases comprising the weak match to the FKH sequence, GCTATCA, are annotated.
Figure S3. Related to Figure 3

Observed and Predicted DNA Shape parameters
Helical twist (a) and roll (b) are shown for both sequences (FKH in red, FHL in blue). DNA shape parameters were measured from the crystal structure using Curves+ (Lavery et al., 2009). Predicted and measured minor groove width (c,f), helical twist (d,g) and roll (e,h) are shown for both sequences. FKH predicted is shown in dark red, FKH measured in red, FHL predicted is in dark blue, and FHL measured in blue. DNA shape parameters were measured from the crystal structure using Curves+ (Lavery et al., 2009), and DNA shape predictions were made using the DNAshape webserver (Zhou et al., 2013).
Figure S4. Related to Figure 4

A

FKH

B

FHL

C

D

E

F
All protein-DNA contacts for the FKH (a) and FHL (b) co-crystal structures are shown. (c,d) Contacts between wing 1 and the DNA backbone are shown for (ca) the FKH structure (magenta protein, yellow DNA) and (d) the FHL structure (blue protein, green DNA). (e,f) Contacts between helix 1 at the N-terminal end of the DBD and the protein are shown for (e) the FKH and (f) the FHL structure.
Figure S5. Related to Figure 5.

The diagram shows the interaction of various proteins and motifs, with specific amino acid positions highlighted and scored for binding affinity. The sequences 'GACCC' and 'GTAAAACA' are shown with their respective PBM scores. The figure includes detailed images of protein interactions, with emphasis on key amino acid positions such as '6aa', 'Nterm', 'Nloop', and 'loop'.
FoxN3 subdomain swap boxplots.
(a) Schema of designed chimeric proteins with positions numbered according to the FoxN3 (left) and FoxN2 (right) full protein sequences. (b) Position of the substituted sequences are shown on the FoxN3 structure. (c) As in Figure 5, PBM E-scores for the FKH (top panel) and FHL (bottom panel) sites are shown.
Protein sequence alignment of the forkhead DBD. The DBD sequences of all human forkhead proteins were collected from UniProt and aligned using ClustalΩ. Alpha helices are boxed in blue, beta strands in green, and the two wings of the forkhead fold in orange. The 6 a.a. in FoxN3 substituted in Figure 5 are shown in red.