The Dissertation Committee for Julius Christopher Barsi Certifies that this is the approved version of the following dissertation:

A Core Signaling Component of the Notch Network + A Molecular Interaction Database Accessible through an Online VLSIC-like Interface

Committee:

Karen Artzt, Supervisor

Henry R. Bose, Jr.

Paul M. Macdonald

Theresa J. O'Halloran

Philip W. Tucker

A Core Signaling Component of the Notch Network + A Molecular Interaction Database Accessible through an Online VLSIC-like Interface

by

Julius Christopher Barsi, Lic. Biological Sciences

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2006

Acknowledgements

Karen Artzt, my mentor; simply and impossibly, for everything. Rashmi Rajendra, Gabriela Vasquez and Margaret Centilli; above all, for their friendship. I also care to distinguish two of the several undergraduates whom at one point or another worked in the laboratory; Zack Mahdavi and Maren Yngve.

A Core Signaling Component of the Notch Network + A Molecular Interaction Database Accessible through an Online VLSIC-like Interface

Publication No._____

Julius Christopher Barsi, Ph.D.

The University of Texas at Austin, 2006

Supervisor: Karen Artzt

Signal transduction among adjacent cells is central to all aspects of metazoan development. Distinct signaling pathways have evolved around this premise. Of these, the *Notch* pathway has perhaps been the most extensively studied. Much of our current understanding regarding *Notch* stems from *Drosophila* genetics, where components of the pathway were first identified. The discovery of a novel core *Notch* pathway component in a vertebrate model organism comes as a surprise, given the extensive genetic screens carried out in *Drosophila*. Herein I describe how, through the use of reverse genetics this novel gene was studied in *Mus musculus*. I then take a radically different approach towards understanding the *Notch* pathway as it pertains to mammals. This revolves around an *in silico* molecular interaction map whose purpose is two fold:

· Serve as a portal for a new type of database capable of retrieving all raw data supporting any given mapped interaction.

· Provide a theoretical framework with which to analyze, model and predict network behavior when natural or designed mutations are introduced.

Table of Contents

List of Tables

List of Figures

Introduction

Mammalian embryonic development is incredibly complex. The fact that a single cell ultimately gives rise to different tissues and organs all with the correct anatomical relationship surely must interest any biologist regardless of their field of expertise. At the base of this process lies cellular differentiation. To address this *in vivo*, one must understand how cells in the developing embryo interact with each other, and what modes of communication are used to control cell fate. Research in the latter has revealed at least seventeen distinct intercellular signal transduction pathways. Five of these are recurrent throughout early mammalian embryogenesis[1]. Each of these pathways is referred to by the name of one of its components. Hence, we have the wingless related mouse mammary tumor virus integration site (*Wnt*), transforming growth factor beta (*Tgfβ*), hedgehog (*Hh*), neurotrophic tyrosine kinase receptor (*Ntrk*) and notch (*Notch*) signal transduction pathways. This dissertation revolves around the latter.

While referred to as pathways, signal transduction differs from metabolic pathways in that only the signal is relayed as opposed to carbon atoms and energy. Although these signals may result in any one of numerous cellular responses, signal transduction throughout early development generally targets transcriptional activity. The fact that early development depends heavily on signal transduction would suggest that a null mutation for any given pathway component should result in embryonic lethality. Indeed, this is generally the case when dealing with invertebrates. However, this rationale does not necessarily hold for vertebrates.

Mus musculus is currently the premier mammalian model system for biological research mainly due to its close genetic and physiological similarities to *Homo sapiens*. Technological advances in this system allow for powerful genetic approaches. Nevertheless, dissecting these pathways through the use of a mammalian model organism requires the following considerations. At some point during early vertebrate evolution, two rounds of whole genome duplication took place followed by extensive gene loss and specialization[2]. The statistics of genomic features across genomes suggests that the proportion of genes in essential low-level functional categories such as protein biosynthesis and deoxyribonucleic acid (DNA) replication decreases with genome size, whereas the proportion of regulatory genes involved in signal transduction and transcriptional regulation increases approximately in a linear mode[3]. Should this

hold true for vertebrate evolution, it would explain the complexity of mammalian pathways as a result of genomic size. Paralogs have diversified not only their coding sequences, but their *cis*-regulatory elements as well. As a result, for any given pathway component it is not uncommon to find proteins of similar function with overlapping expression profiles. Such redundancy, adds to the robustness of the pathway[4]. Therefore, attempting to tease out the significance of individual pathway components through a loss-of-function approach is not necessarily straightforward. Interpretation is further complicated when a particular component serves as a negative regulator, where loss-of-function at the protein level may translate into a constitutively active pathway.

Apart from the dissection of the mechanistic aspects of the signaling pathway, a more difficult (but perhaps more interesting) challenge is to define the genetic circuitry of the pathway[5]. One step in this direction is to build comprehensive regulatory-circuitry maps. This is essential to understand biological networks, such as protein-protein interaction and gene regulatory networks. However, to achieve this it becomes necessary to first adopt a notation system tailored for molecular biology. Despite demand, only a few labs have pioneered work in this field[6-8]. Although no one notation system has yet been universally recognized, the architecture underlying a molecular interaction map is based on Very Large Scale Integrated Circuits (VLSIC), analogous to those used in microelectronics[6,7,9,10].

In order for such maps to be comprehensive, they must integrate a wide variety of data, collected at scales from the behavior of individual proteins to whole genome analyses. Data obtained from genomics is quantitative, thus easily manipulated and assigned confidence limits, whereas local data are often qualitative and report on a limited set of parameters. Furthermore, its correct interpretation is often restricted to a small number of experts. Ideally, each component within a map should expand so as to reveal the underlying data supporting its position within the map. This is necessary in order to address the nature and quality of the data itself. The construction of such databases would create a theoretical framework with which to analyze, model and predict network behavior when natural or engineered mutations are introduced.

If any predictive capabilities are to be expected, it is impossible to ignore cross-talk with other pathways, non-linear structures, and reactions that restore the pathway to its original state once input is removed. Hence, mapping the universe of all possible molecular interactions is important. However impressive this endeavor may appear, its biological relevance is minimal if analyzed out of context. A first approach to confine the

plethora of information to the space of biologically relevant interactions would require focusing on a particular cell type. A simple query as to *species*, *developmental stage*, *tissue*, *cell type* and perhaps *cell cycle stage*, could then highlight the subset of mapped components which are co-expressed in time and space.

The subsequent dissertation follows a pattern similar to that of its introduction. The first chapter reviews our current understanding of how the *Notch* pathway operates in mammals. The second chapter goes on to describe a case study regarding a murine gene presumed to influence *Notch* signal transduction, and how this was verified. The third and forth chapters address a larger scale, where they describe biological networks and a novel type of database, respectively. Finally, the last chapter encompasses avenues of research born out of all of the above.

Chapter 1: The *Notch* signaling pathway

GENE REGULATORY NETWORKS

Development of the animal *bauplan* is controlled by Gene Regulatory Networks (GRNs) which ultimately dictate cellular division, differentiation and apoptosis. GRNs are complex dynamic biological systems composed of modular subsystems[11,12]. In much the same way a cell within a multi-cellular system defines a discrete unit capable of certain autonomy, so too can functional and spatiotemporal modules be recognized within GRNs. Mounting evidence points towards a densely interconnected core of robust GRN modules, believed to serve as the kernel for developmental programs[13]. Additional classes of GRN modules may be distinguished, though all appear to be less robust than those that make up the kernel[3]. Signal transduction systems are one such class. This is consistent with the observation that the global topological structure of biological networks resembles that of a double-napped cone, with the more robust GRN modules closest to its vertex. It has been argued, that this in of itself could simultaneously account for short-term robustness and long-term evolvability inherent to biological systems[4]. It is within this framework that I wish to introduce the *Notch* signal transduction pathway.

MODUS OPERANDI (MO)

MO Preface

The *Notch* signal transduction pathway is an evolutionarily conserved GRN module necessary for numerous developmental functions. I prefer to use the term *GRN module* for two reasons. First, the term *pathway* does not accurately convey the degree to which non-linear interactions are involved, and second it is important to address modularity when operating at the systems level. For the sake of clarity, hereafter I will simply refer to it as the *Notch* network (Nn). Furthermore, I wish to clarify that the description I compile below pertains to the Nn as observed in mammalian systems. However, in those instances where mammalian research is lagging, I will integrate findings made in non-mammalian organisms if homologous components can be

identified. Finally, I will forego the historical events leading towards our current understanding and, as plainly as possible describe the Nn.

Nn nodes

The Nn is a mechanism through which adjacent cells communicate. Although the core components are universal, modulation occurs at almost every level in a context specific manner. Genetic aberration in any one of the modulatory genes generally results in tissue specific defects, whereas a deficient core gene disrupts the Nn in all its instances. This results in a range of embryological defects often too great to overcome, culminating with developmental arrest. Those regarded as core genes have been found to encode cell surface receptors and ligands, components of the endocytic machinery, key proteases, ubiquitin protein ligases, [1]deubiquitinating enzymes (DUBs), transcription factors, and many of their downstream targets. Having been grouped by function, these are discussed below.

Receptors

The four cell surface receptors from which the network takes its name (*Notch1* – *Notch4*) undergo a similar, albeit subtlety different maturation process within the trans-Golgi network. It is here that *Notch* precursor proteins are first cleaved by a furin-like convertase at a location termed site1 (S1)[14,15]. In all cases, the two resulting subunits remain noncovalently associated as a heterodimer. Epidermal Growth Factor (EGF) repeats constitute most of what will be the extracellular portion of this heterodimer. These EGF repeats are modified by the protein O-fucosyltransferase-1 (*Pofut1*)[16-18], several of which are further modified by the actions of three distinct β1,3N-acetylglucosaminyl (GlcNac) transferases[17-23], in particular, manic fringe (*Mfng*), lunatic fringe (*Lfng*), and radical fringe (*Rfng*). Modified heterodimers are expressed at the plasma membrane as single-pass transmembrane (TM) glycoproteins. In this form, *Notch* proteins act as bifunctional cell signaling components in that they serve as cell surface receptors as well as direct mediators of gene activation[24]. In their capacity as

[1] Of the DUBs studied to date, none has revealed an essential role within the Nn. However, given the complementary nature of this group with respect to ubiquitin protein ligases, I have decided to include them in this context.

receptors, each display a varying degree in affinity towards different cell surface ligands. Not only can this be attributed to differences in coding sequence among *Notch* homologs, but also to their cell specific glycosylation signature.

Ligands

As for the cell surface ligands capable of binding to, and activating these receptors, the following seven genes have been identified and surely more will follow. Three delta-like homologs (*Dll1*, *Dll3* and *Dll4*)[25-28], two jagged homologs (*Jag1* and *Jag2*)[26,29], the delta/notch-like EGF-related receptor (*Dner*)[30] and contactin (*Cntn1*)[31]. All encode single-pass TM proteins with the exception of *Cntn1*, the gene product of which is a glycosyl phosphatidylinositol (GPI) anchored molecule. Hereafter, I will refer to them collectively as Nn ligands. In addition, two microfibrillar associated protein homologs (*Mfap2* and *Mfap5*) have been shown to possess ligand potential[32]. However, their biological relevance is not yet understood, as they are secreted components of the extracellular matrix. Little is known about the signals and factors that regulate Nn ligand transcription. In principle, cell specific expression constitutes the first level of control within the Nn because regardless of whether or not a cell expresses *Notch*, signaling does not occur unless a neighboring cell expresses a ligand. At the cell surface, it is likely these arrange as clusters which, if heterogenic in nature might function synergistically.

Endocytic machinery

Endocytosis, the process by which cells internalize molecules at the cell surface into internal membrane compartments, and the intracellular traffic among such compartments, follows two distinct routes. One mediated by clathrin (*Clt*) and another, *Clt*-independent. The latter appears to be lipid-raft related[33], though both require the mechanochemical activity of the large GTPase dynamin (*Dnm*)[34,35] without which, the Nn is disrupted[36]. This reveals endocytosis of Nn ligands/receptors to be necessary for signal transduction. Whether this requirement is mediated by *Clt* or not, remains to be determined. The Nn as it pertains to endocytosis, may very well utilize both routes, each functionally distinct. Whereby, general ligand/receptor turnover most likely follows a separate route than that used for signaling. Though it is important to note that neither

function has yet been attributed a particular route. Many signal transduction receptors tend to avoid endocytic uptake until activated by ligand. Therefore, context will determine whether general turnover or signaling events prevail within a given cell. Epsins (*Epn1*, *Epn2*, and *Epn3*) are Ubiquitin Interaction Motif (UIM)-containing proteins which localize to the plasma membrane, where they are believed to function as Clathrin-Associated Sorting Proteins (CLASPs) in conjunction with AP2[37,38]. As with most CLASPs, *Epn* physically interacts with phosphatidylinositol 4,5-bisphosphate [PtdIns(4,5)P2], *Clt*, the alpha1 subunit of AP2 (*Ap2a1*) and particular cargo proteins. However, *Epn* is unique among CLASPs in that it is required for Nn signaling[39-42].

Although very little is known regarding the mechanisms by which Nn ligands/receptors get endocytosed, I believe that in order to design fruitful experiments, one must first attempt to understand the logic which drives similar GRN modules. An important distinction regarding the signal transduction apparatus can be made with respect to vesicular traffic. In some cases, distinct endocytic routes are utilized for signaling *versus* degradation; in others, signaling is intimately linked to receptor degradation. Whether ligand is tethered to the plasma membrane or secreted as a soluble molecule, may dictate which strategy is taken. For example, cells must be able to measure the concentration of extracellular factors accurately and continuously in order to respond appropriately to morphogen gradients within a developmental field. For this, the response should be proportional to the absolute number of activated receptors. Ligand independent endocytosis and uncoupling signaling from degradation clearly benefit this purpose, as it allows continuous monitoring of ligand concentration and dose sensitivity since prior exposure to ligand would not down regulate receptors. On the other hand, tethering a ligand to the cell surface restricts receptor activation exclusively to those cells that are in direct contact with the ligand expressing cell. In this case, signaling and degradative trafficking are often linked so as to ensure tight spatial control. The signal transduction apparatus of morphogen *versus* membrane tethered ligand trafficking may have therefore evolved around the very different biological functions that these signaling systems serve. Upon ligand dependent activation, Nn receptors undergo proteolytic events required for signaling, such that recycling to the plasma membrane would be futile. Furthermore, accumulating evidence suggests that similar proteolytic events occur within ligand as well[43-45]. This, in addition to the difference in strategy suggested above, would imply that signaling within the Nn is most likely coupled to degradative trafficking.

Key proteases

Already mentioned, is the paired basic amino acid cleaving enzyme furin (*Furin*), whose convertase activity is required in the trans-Golgi network in order to constitutively cleave *Notch* precursor proteins at S1. At the cell surface, following the binding of *Notch* receptors to their respective ligands and subsequent endocytosis, a disintegrin and metallopeptidase (*Adam*) domain proteins *Adam10* and *Adam17* cleave both *Notch* receptors as well as Nn ligands at an extracellular location termed site2 (S2)[45,46]. Membrane tethered portions of receptor and ligand, then serve as substrate for the γ-secratase complex, an aspartyl protease made up of at least four different proteins. Several variants of this complex are believed to exist, though all must contain anterior pharynx defective1 (*Aph1a*, *Aph1b* or *Aph1c*), nicastrin (*Ncstn*), presenilin (*Psen1* or *Psen2*) and the presenilin enhancer2 (*Psenen*). The intramembrane cleavage mediated by this complex, at a location termed site3 (S3), releases the intracellular domains of both receptor and ligand into the cytoplasm[24,47-51].

Ubiquitin protein ligases

Proteins are ubiquitinated in a variety of ways[52]. The most obvious distinction being the attachment of a single ubiquitin moiety (monoubiquitination) *versus* that of a polyubiquitin chain linked through the lysine residues of ubiquitin at position 48 (K48). Though not as common, polyubiquitin chains using alternative lysines on ubiquitin have been described as well. Multiple residues on a protein are often susceptible to ubiquitination, as indicated by experiments in which a key lysine is mutated, only to observe the ubiquitination machinery exploit alternative residues[53]. Finally, experiments utilizing mutant ubiquitin, incapable of forming polyubiquitin chains, have revealed the possibility of multiple monoubiquitination[54]. Traditionally, proteins tagged with a polyubiquitin chain were thought to be targeted for degradation *via* the 26S proteasome, whereas monoubiquitination was thought to regulate subcellular localization and recruitment of ubiquitin binding proteins. Though several examples support the above, current studies challenge this view and offer data where such roles are reversed, indicating that no such clear distinction exists[37,38]. In fact, in light of the various ways in which proteins can be ubiquitinated, one might suspect subtle differences in ubiquitination to dictate biological function.

Identification of ubiquitin protein ligases containing protein-protein interaction domains, which are expressed in temporally and spatially controlled dynamic circumstances, bring them to the forefront as regulators of the Nn[55]. Thus, by targeting many of the cell signaling components for proteosome mediated degradation, ubiquitination acts as an important physiological regulator. Moreover, recent evidence suggests that both ligand and receptor activity are regulated by ubiquitination triggered endocytosis[49,56]. Subsequent endosomal sorting most likely depends on similar ubiquitination signatures.

Before I proceed to outline the various ubiquitin protein ligases and their Nn related substrates, I would like to remind the reader of the critique on *in vitro* experiments by emphasizing its significance regarding ubiquitination. Clearly, an *in vitro* ubiquitination assay can prove to be quite informative. However, I feel it is necessary to highlight two points overlooked by several publications. First, I would like to point out that an appropriate negative control for an *in vitro* ubiquitination assay should not only contain the ubiquitin activating enzyme (E1), but the ubiquitin conjugating enzyme (E2) as well. This is important because if E2 is present in high enough concentrations, it may contribute toward substrate ubiquitination in the absence of ubiquitin protein ligase (E3) (R. Shiekhattar, personal communication). Second, the fact that a particular protein serves as substrate for an *in vitro* ubiquitination assay merely indicates a physiological possibility, rather than confirmation. With this in mind, I encourage the reader to explore the references used for the following section, as a few may be subject to interpretation.

Delta like proteins (*Dll1*, *Dll3* and *Dll4*) serve as substrate for at least two independent ubiquitination events. One mediated by the neuralized family (*Neurl* and *Neurl2*)[57,58] and the other by the mind bomb family (*Mib1* and *Mib2*)[56,59-63]. Whether neuralized is essential during mammalian embryogenesis remains to be seen, as the double knockout of both neuralized genes has not yet been generated. Individually, knockouts of either member are viable[64-66]. The second chapter of this dissertation revolves around the founding member of the mind bomb family; therein I argue that *Mib1* dependent ubiquitination of delta is central to the Nn. In fact, results which I have published supporting this view have been independently verified by an unrelated laboratory using a similar approach[62,63]. These and other studies support the notion that *Mib1* activity is required for the S3 cleavage of Nn receptors. *Mib2* on the other hand, does not appear to be an essential component of the Nn as the phenotype associated with its knockout is not fully penetrant (unpublished data). A few studies have been

published regarding *Mib2* function; however the supporting data is unconvincing and conclusions remain contradictory[60,67-69].

The *Notch* proteins (*Notch1* – *Notch4*) are themselves recognized by several distinct E3's. Members of the casitas B-lineage lymphoma (*Cbl*, *Cblb* and *Cblc*)[70], deltex (*Dtx1*, *Dtx2*, *Dtx3*, *Dtx3l* and *Dtx4*)[71,72], numb (*Numb* and *Numbl*)[73,74] and neural precursor cell expressed developmentally down-regulated gene4 (*Nedd4* and *Nedd4l*)[75,76] families, as well as itchy (*Itch*)[77] and [2]F-box and WD-40 domain protein7 (*Fbxw7*)[78-83], have all been proposed to promote *Notch* ubiquitination. In most cases, supporting data seems to be accompanied by the suggestion that such interactions target *Notch* towards lysosomal compartments. For additional information regarding interactions amongst Nn related E3's, refer to Table 1[74,84-98].

Deubiquitinating enzymes (DUBs)

Enzymes that remove ubiquitin from proteins are collectively referred to as DUBs. The vast majority of DUBs are cysteine proteases, which can further be classified depending upon the ubiquitin protease domains they contain. Those which possess the Ubiquitin Specific Protease (USP) domain represent the majority of DUBs encoded by mammalian organisms.

Throughout vertebrate evolution, there has been a significant increase in the number of genes encoding E3's. Likewise, so has the number of USPs increased[99]. The fact that both these groups represent functionally distinct aspects of a common system, in addition to this apparent coevolution, suggests that their activity is intimately related. However in mammals, E3 genes outnumber DUBs by what appears to be a biologically relevant margin. Three possibilities may explain why. The fact that many DUBs and/or cofactors remain to be identified surely accounts for a portion of the observed difference, though it is impossible to say to what extent. One could entertain the possibility that DUBs are promiscuous in their interactions, however current data suggests otherwise.

[2] On occasion, the activity of a particular E3 in regard to Notch has been further defined. For example *Fbxw7* has been shown to bind the nuclear form of activated notch in a phosphorylation dependent manner so as to promote its proteasome mediated degradation. However, few attempts have been made to explore the interactions amongst these E3's, not to mention the less evident connections known to exist (Table 1).

Finally, I favor the notion that only a fraction of the targets that are ubiquitinated require the additional regulatory step of deubiquitination. DUBs are sure to play key regulatory roles in multiple processes, though little is known as to their mode of regulation and substrate specificity. They have been implicated in the endocytic pathway at multiple levels and are known to play an important role in other forms of vesicular traffic. Though no DUB has yet been recognized as a core component of the Nn, research in *Drosophila melanogaster* suggests that ubiquitin specific peptidase9 X-chromosome (*Usp9x*) may serve to modulate *Notch* signaling in a tissue specific manner by influencing the cellular levels of *Epn*[42,100,101]. In light of all of the above, I would be surprised if DUBs are not in some way essential components of the Nn.

Transcription factors

The recombining binding protein suppressor of hairless (*Rbpsuh* and *Rbpsuhl*) family encodes two DNA binding transcription factors, the first of which is required for both the repression and activation of *Notch* target genes. In the absence of *Notch* signal, DNA target sites are occupied by *Rbpsuh* in association with corepressor proteins. These are thought to include CBF1 interacting corepressor (*CIR*), one member of the C-terminal binding protein (*Ctbp1* and *Ctbp2*) family, one member of the nuclear receptor co-repressor (*Ncor1* and *Ncor2*) family, retinoblastoma binding protein8 (*Rbbp8*), split ends transcriptional regulator (*Spen*) and one member of the transducin-like enhancer of split (*Tle1*, *Tle2*, *Tle3*, *Tle4*, *Tle4l* and *Tle6*) family[102,103]. Whether the displacement of these corepressors is necessary in order to convert *Rbpsuh* from transcriptional repressor to activator, remains to be determined. Nevertheless, once in the nucleus *Notch* binds to *Rbpsuh* *via* its RAM (*Rbpsuh*-associated molecule) domain. This interaction, although not entirely necessary, further enables *Notch* to dock by way of its ankyrin repeats to the reticuloendotheliosis oncogene (*Rel*)-homology region of *Rbpsuh*. This interaction is required in order to create a composite surface recognized by members of the mastermind like (*Maml1*, *Maml2* and *Maml3*) family[102,103]. Transcriptional activation by *Notch* associated *Rbpsuh*, ultimately relies upon E1A binding protein p300 (*Ep300*) to recruit the transcriptional apparatus. The extent to which its components, such as *Ep300*, cyclinC (*Ccnc*), cyclin dependent kinase8 (*Cdk8*) and ribonucleic acid (RNA) polymerase II are preassociated, is not known.

Downstream targets

Until recently, the hairy and enhancer of split (*Hes1*, *Hes2*, *Hes3*, *Hes5*, *Hes6* and *Hes7*) family has been the only known effector of *Notch* signaling. However, *Hes* was not always found in areas where Nn ligand and receptor expression overlap, suggesting that *Notch* signaling may target additional genes. The identification of another basic helix-loop-helix (bHLH) family known as hairy and enhancer of split related with YRPW motif (*Hey1*, *Hey2* and *Heyl*) has fulfilled this prediction. Data suggests that *Hes* and *Hey* not only may function as homodimers, but that *Hes*-*Hey* heterodimers are also possible in those instances where they are coexpressed[104]. Although both represent transcriptional repressors, it appears that each functions through distinct targets. In as far as transcriptional activity is concerned; *Hes1* has been shown to negatively regulate its own promoter activity as well as that of CD4 antigen (*Cd4*)[105,106], more importantly *in vivo* it functions to down regulate neurogenesis by directly repressing the proneural gene, achaete-scute complex like1 (*Ascl1*)[107]. The same can be said for *Hes5*. Members of the *Hey* family likely contribute towards similar roles; however, an ever increasing amount of data suggests that they are also capable of repressing transcription of the GATA binding protein (*Gata1*, *Gata2*, *Gata3*, *Gata4*, *Gata5* and *Gata6*) family[108]. Finally, the question remains as to whether these bHLH transcriptional repressors are directly and exclusively activated through the Nn. Several lines of evidence strongly suggest that *Hes1*, *Hes5* and *Hey1* are all primary targets of *Notch*, very likely more will soon be revealed.

MO *per se*

Let us imagine a field of cells, all of which express at least one of the various types of *Notch* receptors. Let us now assume a subset of these express at least one of the various Nn ligands as well. In this way, intercellular signal transduction *via* the Nn is restricted exclusively to those cells that are in direct contact with a ligand expressing cell. The manner in which both ligands and receptors are presented is not known. Clustering, either as homo or hetero multimers may be a prerequisite for active cell surface components. Aside from the latter, in this scenario we will assume that contact with ligand expressing cells is sufficient to permit receptor activation. If we now focus our attention to any two such cells, the current model holds that endocytosis is required in

both the signal sending cell (defined as that which expresses ligand), as well as the signal receiving cell (defined as that which expresses receptor). Once the ligand has engaged with its receptor, endocytosis within the signaling cell is thought to be required in order to mechanically facilitate exposure of the receptors S2 cleavage site. This is supported in part by the trans-endocytosis of the *Notch* extracellular domain (NECD) into the signaling cell[109]. Endocytosis of what remains of the *Notch* receptor into the receiving cell, is believed to be necessary for the γ-secratase complex to process *Notch* at the S3 cleavage site and in so doing, release the *Notch* intracellular domain (NICD)[49]. The latter translocates to the nucleus where it influences gene transcription.

Anyone who has read up to this point can immediately see this model falls short on many levels. First, once all said is done, cells may clearly be distinguished as either signal sending or signal receiving. However, to define these *a priori* simply as those which express ligand or receptor, respectively, is an oversimplification knowing full well that most cells express both ligands *and* receptors of one sort or another. Many however, suggest that random fluctuations in gene expression establish a population of cells with a higher ligand:receptor ratio, and that such a difference ultimately dictates the degree to which *Notch* is activated. Second, there is evidence indicating that components necessary for ligand endocytosis function in a non cell autonomous fashion, in other words, that ligand endocytosis is required for *Notch* to be activated in *neighboring* cells[56]. Although this has been interpreted as described in the model above, I do not believe that we have any idea as to how ligands and receptors interact in *cis* at the cell surface. If any such interactions exist, this may well offer an alternate explanation. Finally, whether NICD is generated at the plasma membrane or within endocytic vesicles remains an open question[49,110-112]. If the latter is true as the model predicts, then so should the same rationale hold true with respect to ligands. For instance, increasing amounts of data suggest that both delta and jagged are processed in a manner analogous to that of *Notch*[44,113]. Furthermore, no transcriptional activity has been attributed to the delta intracellular domain (DICD) despite it being found in the nucleus[44]. In short, new insights with regards to ligand/receptor *cis* interaction and endocytosis, in addition to DICD transcriptional capacity, must be obtained before expecting a more parsimonious model.

Chapter 2 The murine gene mind bomb

MIND BOMB PREFACE

It is my intention to introduce this chapter in light of the circumstance under which my first project took form. My reason for doing so is simple; such events determined, at least in part, the course I was to follow while at the University of Texas (UT). During the late nineties, in what some have argued to be the last large scale alliance between academia and the private sector[114], the Massachusetts Institute of Technology (MIT) secured $1.2 million *per* year over a three year period from the pharmaceutical firm *Amgen* in order to conduct a large scale insertional mutagenesis screen in *Danio rerio*.

In this type of screen, not only can the integration of an exogenous DNA sequence into a genome be mutagenic, but the inserted DNA may also serve as a tag with which to clone the mutated gene. While on sabbatical Dr. Karen Artzt, whose laboratory I would later join, collaborated on this effort[115]. By the time I joined the laboratory, she had brought back to UT one mutant strain created in this manner, known as *hi904*. This mutant had a typically neurogenic phenotype. As it turns out, this was the third large scale mutagenesis screen in *Danio rerio* to generate such a phenotype. One laboratory out of Tübingen previously named their mutant white tail[116,117], whereas another laboratory out of Boston had named theirs mind bomb[118]. After it was confirmed that both mutants represented different alleles of the same gene, the latter was accepted as the official name. The similarity of our mutant's phenotype to those described above, drove us to carry out complementation tests, and in so doing we found *hi904* to be allelic to mind bomb. However, due to the method by which our mutant was generated, we had some knowledge with regard to the gene involved. Hence, we knew that the mind bomb phenotype resulted from the disruption of a novel gene. But it soon became apparent that another laboratory was already working at characterizing it in *Danio rerio*. We then turned to genomic sequence analysis which revealed two highly conserved homologs in *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens*. This prompted us to shift our focus towards *Mus musculus*, which in any case had always been our laboratory's preferred model organism. There are two highly conserved mind bomb genes in mammals: mind bomb1 (*Mib1*) and mind bomb2 (*Mib2*) (Figure 1).

Therefore, in order to address their function in a mammalian system, we genetically engineered mice carrying a germ line mutation in either one of these genes.

However, I will restrict myself to discussing *Mib1*, as the phenotype associated with knocking out *Mib2* was not fully penetrant. Finally, for the sake of clarity to the reader which is unfamiliar with mouse embryo anatomy, Figure 2 provides a sketch which illustrates the different levels of magnification at which morphological features and expression pattern data were collected.

GENE AND PROTEIN CHARACTERISTICS

In *Mus musculus*, the gene referred to as *Mib1* maps to chromosome 18 at location 10,770,970-10,863,191 base pairs (bps) according to Ensembl Mouse version 37 (v37), based on the National Center for Biotechnology Information (NCBI) mouse build 34 (m34) mouse assembly (freeze May 17, 2005, strain C57BL/6J). *Mib1* spans just over 92 kilobases (Kb) coding for a primary transcript of 10 Kb which is then translated into a 110 kilo Dalton (kDa) protein. At its N-terminus, it contains a ZZ-type Zinc Finger region flanked on either side by what has been called the Mib/Herc2 domain. The middle portion of the protein contains nine Ankyrin repeats, and towards the C-terminus we find three RING-type Zinc Fingers; the first two of which are atypical and separated from the third by a Coiled-coil region. The structure of this protein strongly resembles that of a RING-type ubiquitin protein ligase (Figure 3). At the amino acid level, *Mus musculus* and *Homo sapiens Mib1* are 99.3% identical, whereas *Mus musculus* and *Danio rerio Mib1* are 91.7% identical (Figure 4). Further analysis reveals the possibility that the mammalian *Mib1* gene may also encode two distinct microRNAs (miRNAs) within its twelfth intron. Specifically, *mmu-miR-1-2* at location 10,830,826-10,830,897 bps and *mmu-miR-133a-1* at location 10828254-10828321 bps according to the same genomic coordinate system specified above (Figure 5).

GENE EXPRESSION

Mib1 protein coding transcript

The expression of *Mib1* protein coding transcript was analyzed by northern blot and RNA *in situ* hybridization to embryos. Its expression peaked during two periods of embryonic development [embryonic day (E) 6.5 and E10.5-11.5], after which point expression appears to decline (Figure 6). Whole mount RNA *in situ* hybridization to E9.5

embryos revealed a ubiquitous expression pattern (data not shown). In adult tissues, *Mib1* expression is highest in brain, lung, kidney and testis. Two alternatively spliced messages were detected; one of 10 kb and another of 4 kb particularly highly expressed in testis, though the functional relevance of the latter remains unknown (Figure 6).

Mib1 non-coding transcripts

The expression of *Mib1* non-coding transcripts were analyzed by RNase protection assay and locked nucleic acid (LNA) *in situ* hybridization to embryos. Specifically, sequences corresponding to the mature form of *mmu-miR-1-2* and *mmu-miR-133a-1* were assessed (Figure 5). Because the mature forms of both *mmu-miR-1-2* and *mmu-miR-133a-1* appeared to be co-expressed, their expression profile will be treated as one. However, in all instances analyzed, the expression levels of *mmu-miR-1-2* were much higher than those of *mmu-miR-133a-1*. Embryonic expression was detected as early as E8.5 up to E10.5, though it is likely to continue beyond this developmental window. During which period, whole mount LNA *in situ* hybridization revealed expression to be restricted to the developing heart and somites (Figure 7). This not only confirms, but also expands results published by other groups[119-121]. In adult tissues, expression was absent in brain, but as in the embryo remained prominent within the heart (Figure 8).

GENETICALLY ENGINEERED MUTANT MICE

The *Mus musculus Mib1* complementary DNA (cDNA) sequence was identified using Basic Local Alignment Search Tool (BLAST) analyses to the *Danio rerio hi904* partial sequence. To address *Mib1* protein function, we generated a targeted mutation that deleted a portion of its first exon including the start codon (Figure 9). Correctly targeted embryonic stem (ES) cell clones were used to generate chimeras, which were then bred in order to obtain germ line transmission. As of yet, no attempts were made on our behalf to generate a mouse model which would exclusively address *Mib1* miRNA loss of function. Hence, I will further restrict my results to *Mib1* protein. Nevertheless, I cannot exclude the possibility that our targeted mutation may have in some way disturbed *Mib1* miRNA expression, and by extension contributed towards the observed phenotype.

RESULTS

Mib1 null mutants are embryonic lethal

Mice capable of propagating the *Mib1* targeted mutation (*Mib1*^{tm1Art}) were backcrossed to an in-house strain. First filial generation (F1) animals which were determined to be heterozygous for this mutation were selfed. No homozygous mutants were observed when newborns were genotyped, suggesting a critical role in post gastrulation development. As suspected, mutants were found to be recessive embryonic lethal by E10.5, which was confirmed by yolk sac Polymerase Chain Reaction (PCR) genotyping. Reverse Transcriptase-PCR (RT-PCR) analysis of homozygous mutants showed complete absence of *Mib1* transcript (Figure 10). *Mib1* immunoblot detects a single specific protein of a 110 kDa (Figure 11). Although several cross reacting proteins are present, the 110 kDa band is absent in *Mib1* homozygous mutants, indicating that we have generated a null mutation and I will hereafter refer to them as *Mib1*$^{-/-}$ embryos.

The most visible defects in *Mib1*$^{-/-}$ embryos are neural tube kinks apparent by E8.5. Somites if at all present, are fewer than in normal littermates and irregular with indiscrete edges (Figure 12a-b). To further highlight the defect in somitogenesis, we took advantage of *mmu-miR-1-2*'s well defined embryonic expression pattern. By utilizing *mmu-miR-1-2* as a marker for the developing heart and somites, it becomes clearly visible that most *Mib1*$^{-/-}$ embryos fail to develop somites by E9.5 (Figure 7a-c). At this point in development, the defects are more evident; resulting in a shortened body axis and the second branchial arch is never made. At this stage pericardial edema and a poorly folded heart also become visible (Figure 12c-d). Upon dissection, vascularization defects were observed in the embryo as well as the yolk sac (data not shown). By E10.5, although the heart frequently continues to beat weakly, mutant embryos display a histologically necrotic neuroepithelium and are moribund (data not shown). They die around this time presumably from lack of a placental connection, as indicated by a rounded allantois (Figure 12d).

Mib1 loss of function results in neuronal differentiation defects

In contrast to the somite and vascular defects, the head of E9.5 mutant embryos appeared grossly normal. However, histopathology revealed that *Mib1* loss of function

results in dramatic neuroepithelial defects (Figure 13). Within this tissue there is a dearth of cells (Figure 13c) and structural integrity is frequently compromised (Figure 13b). To test for differentiation in *Mib1*$^{-/-}$ embryos, transverse sections were taken from embryonic brain. Tubulin beta3 (*Tubb3*) immunohistochemistry (IHC) was assessed on neuroepithelium, as it is one of the earliest known neuronal markers[122]. Wild type (WT) embryos at E9.5 revealed a row of nascent neurons along the pial surface (Figure 14a). Immediately neighboring these *Tubb3* positive cells, there are many negative, as yet uncommitted cells consistent with *Notch* mediated lateral inhibition. In contrast, *Mib1*$^{-/-}$ neuroepithelial cells prematurely differentiate *en masse*. Not only is there no evidence for lateral inhibition, but differentiation within the tissue was no longer polarized towards the pial surface (Figure 14b). Despite the presence of non-specific proteins in *Mib1* immunoblots (Figure 11), cross-reactivity in mutants is restricted to the mesenchyme and not present in neuroepithelium (Figure 15a). This makes it possible to follow the *in vivo* expression pattern of *Mib1* within WT neuroepithelium.

Using confocal immunofluorescence, it appears *Mib1* protein expression is modified during neurogenesis. In WT E9.5 embryos, *Mib1* is up-regulated within the neuroepithelium specifically in the region undergoing differentiation (Figure 15d-f)[123]. This is supported by an increase in *Mib1* immunoreactivity towards the pial surface where nascent neurons are found, as shown by *Mib1*/*Tubb3* co-localization (Figure 16c). This implies that *Mib1* is functionally relevant during neurogenesis. Furthermore, unrestricted *Tubb3* expression in mutants coincides with activated caspase3 (*Casp3*) (Figure 17a-c). Suggesting that unlike the zebrafish *Mib1* mutant, premature neurons undergo apoptosis soon after differentiation[56,59]. This may explain the dearth of neuroepithelial cells often seen in hematoxylin and eosin (H&E) stained mutant sections (Figure 13c).

Mib1$^{-/-}$ embryos misexpress Nn transcripts

The phenotypic similarity of *Mib1*$^{-/-}$ embryos to the *Notch1* and *Rbpsuh* knockouts led us to analyze the RNA *in situ* expression pattern of several Nn nodes E9.5[124,125]. *Notch1* transcript in WT localizes to the neural tube, paraxial mesenchyme, developing vasculature, somites and the presomitic mesoderm[124]. Interestingly, in the *Mib1*$^{-/-}$ embryos, *Notch1* transcript was found to be misexpressed predominantly in the mesencephalon and neural tube, with lowered expression in the presomitic mesoderm.

In WT, *Dll1* transcript is expressed in the neural tube, presomitic mesoderm and also defines the posterior half of all somites[126]. In contrast, *Mib1*$^{-/-}$ embryos showed definite spatial misexpression of *Dll1* transcript exclusively in the neural tube in addition to a significant reduction within the presomitic mesoderm. Similarly, mutant embryos showed a lower level of *Jag1* in all the relevant structures with prominent misexpression to the mesencephalon, whereas in WT it localizes to the developing brain, ear, branchial arches, trunk mesenchyme, and tail bud[127]. Of the relevant GlcNac transferases, *Lfng* in WT has a characteristic pattern in the future nervous system. Particularly striking at E9.5, is its absence in the forebrain-midbrain boundary and the isthmus that defines the mesencephalon. Interestingly, no such patterning of the midbrain was observed in mutants even though a strong neural expression was evident. Furthermore, upon examining *Lfng* expression at E8.5, *Mib1*$^{-/-}$ embryos do not reveal newly formed somites as defined by this transcript. Finally, the transcriptional repressors *Hes1* and *Hes5* showed greatly diminished expression in null embryos. Taken together, this data indicates that absence of *Mib1* disrupts *Notch1* signaling and neural patterning.

Since death associated protein kinase1 (*Dapk1*) may also represent a *Mib1* substrate *in vivo*[84], the expression pattern of its transcript was analyzed in the embryo. In WT at E9.5, *Dapk1* transcript was observed in the developing ear, eye, branchial arches, paraxial mesoderm and in some vascular tissue, excluding the heart. Interestingly, *Mib1*$^{-/-}$ embryos seem to also misexpress *Dapk1* in the brain and neural tube. The implications of this result regarding a possible role for *Dapk1* along with *Mib1* during embryonic development, is as yet unclear.

Data described within this section, as well as the materials and methods employed, may be found in reference[63].

Mib1 deficiency disrupts the Nn

To obtain a broader profile of Nn genes deregulated in the *Mib1*$^{-/-}$, we chose to perform RT-PCR analyses on individual E9.5 WT and mutant embryos (Figure 10). Primers to *Notch1* and *Jag1* showed lower levels within the mutant, whereas *Dll1* remained unchanged. However, it is important to note that *Dll1* expression within the embryo is mislocalized as seen by RNA *in situ* hybridization. Also analyzed were the *Notch1* transcriptional co-activators *Rbpsuh* and *Maml1*, in addition to genes they activate[128]. The latter consisted of *Hes1*, *Hes5*, *Hey1* and *Hey2*. With the exception of

Rbpsuh and *Hey2*, all of these genes showed lowered expression in the mutant as compared to WT. Noteworthy is the down regulation of *Maml1*, which may further contribute towards the disruption of the Nn. *Rbpsuh* serves as a co-activator for *Notch* signaling but is regulated in a *Notch* independent manner[124]. *Hey2* transcript levels in the mutant can be attributed to the fact that only *Hey1* and not *Hey2*, strictly reflects *Jag1* and *Dll1* stimulated *Notch1* signaling[104,129]. Lastly, analyses of proneural transcription factors found *Ascl1* and atonal homolog1 (*Atoh1*) to be down regulated in the mutant. This might appear counterintuitive given the neurogenic nature of our phenotype. However, *Ascl1* and *Atoh1* are pivotal during later stages of differentiation where they specify subsets of differentiating neurons[130-132].

Mib1$^{-/-}$ embryos fail to activate *Notch1 in vivo*

Mib1$^{-/-}$ embryos express full length *Notch1* protein (Figure 18c). Therefore, with the intention to better understand how the Nn is disrupted, we asked whether *Mib1* acts up- or down-stream of *Notch1* activation. To answer this question, efforts were focused on the *Mib1*$^{-/-}$ embryo's capacity to generate *Notch1* intracellular domain (NICD1). For this, we resorted to an antibody which specifically recognizes a neo-epitope exposed only after the *Notch1* receptor has been cleaved between glycine 1743 and valine 1744 (S3 cleavage). With a WT littermate as control, NICD1 immunoblot detects a single specific protein of 110 kDa. This band is absent in the *Mib1*$^{-/-}$ embryo (Figure 18a). Together with the RT-PCR data, this proves that *Mib1*$^{-/-}$ embryos fail to activate the *Notch1* receptor. In order to prove that this underlies the *Mib1*$^{-/-}$ neurogenic phenotype, we had to corroborate *Hes5* down regulation (Figure 10) at the protein level. Therefore, we analyzed NICD1 and *Hes5* co-localization within embryonic neuroepithelium at E9.5. In WT, NICD1 immunoreactivity was weak (Figure 19b). Higher magnification revealed a speckled-like distribution (Figure 19f). NICD1 was observed to co-localize with DAPI, corroborating its nuclear presence. However its subcellular localization was by no means restricted to the nucleus. This is in disagreement with an earlier study[133]. In contrast, no immunoreactivity was observed in sections derived from *Mib1*$^{-/-}$ embryos (Figure 19n). The absence of immunoreactivity within the negative control (*Notch1*$^{-/-}$) confirms the specificity of the antibody (Figure 19j). These results directly correlate with *Hes5* immunoreactivity, which was prominent throughout WT neuroepithelium, but not in that

of the mutants (Figure 19g *versus* k, o). Finally, *Tubb3* was assessed so as to correlate these results with aberrant neurogenesis (Figure 19p).

DISCUSSION

Mib1 encoded protein

Mib1 was first characterized as a ubiquitin protein ligase through its capacity to interact with and ubiquitinate *Dapk1*[84]. However, the latter has been most extensively studied for its relevance in apoptosis and oncogenesis, with little known about its role during embryonic development[134,135]. Previous studies have established the *Notch* ligand *Dll1* as another substrate of *Mib1*. This interaction has been shown to regulate *Dll1* cell-surface levels, possibly through ubiquitin-triggered endocytosis[56,59,61,136]. Our experiments with the *Mib1* knockout have revealed that its developmental relevance is maintained in higher vertebrates. Even though *Mib1* is known to act upon two different proteins, it would appear that the main phenotype observed in our mutant pertains to *Dll1*. Since the aim of this study was to provide an initial characterization of the murine *Mib1* knockout, we did not directly assess how this might affect the Nn's signaling mechanism *per se*. However, our data does allow us to infer conclusions in this respect. We show that abolishing *Mib1* activity in *Mus musculus* severely reduces *Notch* signaling. Four lines of evidence are consistent with earlier studies. The first comes as a direct result from our mutant phenotype, where developmental defects can be seen in all instances where *Notch* signaling is required. Similar to the *Notch1* and *Rbpsuh* knockouts, *Mib1*$^{-/-}$ embryos undergo mid-gestation lethality. Second, when assessing several of the Nn nodes at E9.5, mutant embryos clearly misexpress many downstream targets. Third, our IHC data reveals patterns of expression which reflect the loss of *Notch* mediated lateral inhibition and directly implicates *Mib1* in neurogenesis. We demonstrate that active cell death is deregulated in this process. Finally, confirmed through immunoblot was the mutant's failure to generate NICD1. Hence, establishing that *Mib1* activity is required for S3 cleavage of the *Notch1* receptor. As a result, the near absence of *Hes1* and *Hes5* would account for the neurogenic phenotype, as these transcriptional repressors are responsible for maintaining neuroepithelial cells in an undifferentiated state[137]. *Mib1* activity in as far as the Nn is concerned, may possibly extend beyond its interaction with *Dll1*[61,136]. Recent work in *Drosophila melanogaster*

suggests that *Mib1* modulates the cell surface levels of yet another *Notch* ligand, two homologs of which are found in mammals, *Jag1* and *Jag2*[138,139]. Whether *Notch* ligands and their respective receptors get mono-, multiple mono-, or poly-ubiquitinated at either the cell surface or within endocytic compartments, is important in order to understand the mechanism by which adjacent cells signal[140,141]. Whether *Mib1* functions in a cell-autonomous fashion has yet to be clearly determined as well. Finally, how failure to down regulate Nn ligand cell-surface levels could ultimately abolish *Notch* signaling rather than enhance it, remains a central issue[55,142].

Mib1 encoded miRNAs

The majority of miRNA genes map to locations on the genome distant from previously annotated genes, though several variations exist. In *Drosophila melanogaster*, many appear to be clustered in such a manner which would suggest they are first transcribed as a multi cistronic transcript and latter processed into the biologically relevant mature miRNAs. In mammals however, this genomic arrangement is much less frequent. Another interesting arrangement can be found in a quarter of all mammalian miRNA genes, specifically when these are encoded within the introns of protein coding genes. In such instances, miRNAs are sometimes found in the same orientation as the host gene, suggesting that their transcription is under the control of the host gene promoter, analogous to what is seen for many small nucleolar RNA (snoRNA) genes. Other times, these intronic miRNA genes are found on the strand opposite from which the mRNA is transcribed and hence, will most likely be under independent cis-regulatory control. The expression data I have shown regarding *Mib1* encoded miRNAs *mmu-miR-1-2* and *mmu-miR-133a-1*, pertain to this latter class.

The distinct expression pattern between these miRNAs and their host gene clearly discredits the notion that intronic miRNA spatio-temporal expression can be inferred from the host gene mRNA[143]. Because intronic miRNA expression profiles highly correlate with those of their host genes[144], discrepancies in this correlation have previously been attributed to the existence of miRNA paralogs whose potentially distinct expression cannot be distinguished by microarray technology. However, if what has been observed in *Drosophila melanogaster* holds true for mammals, my observation would not be restricted to intronic miRNAs residing on the opposite strand, as *mmu-mir-*

7-1 and heterogeneous nuclear ribonucleoproteinK (*Hnrpk*) would provide yet another example[145].

Finally, miRNA expression patterns strongly correlate with what little is known about their function. In this case, the fact that during development *mmu-miR-1-2* is exclusively expressed within the heart and somites, reflects the location of two of its known targets, heart and neural crest derivatives expressed transcript2 (*Hand2*)[146] and *Dll1*[147]. A similar rationale may apply to *mmu-miR-133a-1*, where in conjunction with *mmu-miR-1-2*, has been shown to modulate skeletal muscle proliferation and differentiation[119,148]. This obviously is not surprising; however, I believe that this correlation could be exploited to further refine miRNA *in silico* target predictions[149-152]. Currently, several routes are being taken towards the development of new technology capable of physically identifying miRNA targets in a high throughput manner. Unfortunately, to this day all have encountered significant obstacles (as an example see reference[153]). Therefore, it is important to incorporate variables such as spatio-temporal co-expression, into the prediction algorithm.

MATERIALS AND METHODS

Analysis of *Mib1*$^{-/-}$ embryos

Noon of the first day in which vaginal plugs were observed in female mice, was considered as E0.5. Mutant embryos were identified by their phenotype and confirmed by yolk sac PCR typing. This multiplex PCR reaction was done with three primers, which gave a 600 bps band for the WT allele and a 450 bps band for the mutant allele. Primers used were New1p-5' CGAGTGATGGTGGAGGG; Neo66-5' ATGCTCCAGACTGCCTTG and 9233-3' GCTCAACAAAGCAAAGGCACCAAGAGACGGATCC.

Antibodies

Primary antibodies used were, rabbit anti-*Casp3* (which specifically recognizes the p17 fragment of active caspase3) (Abcam Inc. Cambridge, Massachusetts, USA), rabbit anti-NICD1 (which recognizes the cytosolic domain of *Notch1* only when cleaved between glycine 1743 and valine 1744) (Cell Signaling Technology Inc. Danvers, Massachusetts, USA), hamster anti-*Notch1* (Merck Inc. Darmstadt, Hesse, Germany),

rabbit anti-*Mib1* was a gift from Patricia Gallagher (raised against *Mib1* peptide residues 493-1006), mouse monoclonal anti-*Tubb3* (reacts with neuron specific Tubulin beta3) (Abcam Inc. Cambridge, Massachusetts, USA) and goat anti-*Hes5* (Santa Cruz Biotechnology Inc. Santa Cruz, California, USA).

Secondary antibodies used were, horseradish-peroxidase (HRP) conjugated goat anti- hamster, rabbit IgG (Santa Cruz Biotechnology Inc. Santa Cruz, California, USA), Cy2 conjugated donkey anti-goat IgG (Abcam Inc. Cambridge, Massachusetts, USA), Cy3 conjugated goat anti-mouse IgG, anti-rabbit IgG and donkey anti-rabbit IgG (GE Healthcare Inc. Piscataway, New Jersey, USA). Donkey anti- goat, rabbit and mouse IgG are all affinity purified sera conjugated to Alexa Fluor® 488, 594 and 647, respectively (Invitrogen Inc. Carlsbad, California, USA).

Generation of mutant mice

The knockout construct was made by inserting a 1.7 kb 129SvEv (129S6) genomic fragment upstream of *Mib1*'s first exon into the pCH110 vector. Together with the *β-galactosidase* gene, this was cloned into the pPNTeI vector, which contains a PGK-neo cassette. A 4 kb fragment downstream of the start codon was then inserted into the Xho1/NotI restriction site to make the complete construct. Correctly targeted ES cell clones were identified by Southern blotting using probes external to both the 5' and 3' regions of the construct. Chimeras were mated to BTNTTF/Art in order to obtain germ line transmission. The mutation was then backcrossed to BTNTTF/Art. Heterozygous F1 animals were tested by Southern blotting and subsequent PCR genotyping. All mice were maintained under UT at Austin Institutional Animal Care and Use Committee (IACUC) protocol #03030701.

Histology

For histological analysis, embryos were fixed in 4% paraformaldehyde (PFA) during 4 hours and embedded in Paraplast® (Oxford Labware Inc. St. Louis, Missouri, USA). 7 µm serial sections were obtained and stained with H&E for routine analysis or further processed for immunohistochemistry (IHC).

Image acquisition and processing

Light microscopy was carried out on a SZH10 microscope (Olympus Inc. Center Valley, Pennsylvania, USA) and standard fluorescence microscopy on an Axioplan 2 MOT epifluorescence microscope (Carl Zeiss Inc. Thornwood, New York, USA). Images were captured *via* charge-coupled device (CCD) camera (Optronics Inc. Goleta, California, USA). Confocal imaging was achieved with a TCS 4D confocal microscope (Leica Inc. Bannockburn, Illinois, USA). Confocal stacks were processed using *ImageJ* software <http://rsb.info.nih.gov/ij>. Final image panels were collated using Photoshop® cs (Adobe Systems Inc. San Jose, California, USA).

Immunoblot

Embryos were homogenized in 2X sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) sample buffer. Total protein from individual embryos was loaded *per* lane. Protein lysate was then fractionated by electrophoresis through a 7.5% SDS-polyacrylamide gel and transferred to Optitran® nitrocellulose (Whatman Inc. Florham Park, New Jersey, USA). Protein transfer was verified by *Ponceau S* stain. Blots were blocked overnight in 5% nonfat milk in 1X phosphate buffered saline (PBS), incubated with primary antibody (1:1000) for 1 hour, washed and then incubated with HRP conjugated secondary antibody (1:5000) for 1 hour. Immunoreactivity was detected on BioMax® film (Kodak Inc. New Haven, Connecticut, USA) using Western Lightning® chemiluminescent enhanced luminol (PerkinElmer Inc. Boston, Massachusetts, USA). Optimal film exposure was timed at 45 seconds.

Immunohistochemistry (IHC)

Histological samples were deparaffinized. Antigen retrieval was achieved by 30 minutes incubation with 10 mM citric acid buffer (pH 3) at 37^{o}C. Sections were blocked with 3% bovine serum albumin (BSA) in 1X PBS then incubated with primary antibodies (1:500) at 4^{o}C overnight. Samples were drained and washed before incubating with secondary antibodies (1:200) for 5 minutes. Following thorough washes, samples were briefly dehydrated, air dried and mounted with Permount™ (Fisher Scientific Inc. Hampton, New Hampshire, USA).

LNA *in situ* hybridization

Whole mount embryo LNA *in situ* hybridizations were done using digoxigenin (DIG) labeled LNA probes as described in reference[121]. BM purple (Roche Diagnostics Inc. Indianapolis, Indiana, USA) was used to detect the alkaline phosphatase (AP) conjugated anti-DIG antibody (Roche Diagnostics Inc. Indianapolis, Indiana, USA). The antisense probes used were: *mmu-miR-1-2* and *mmu-miR-133a-1* LNA oligonucleotides (Exiqon Inc. Vedbaek, Capitol Region, Denmark).

Northern blot

Mouse embryonic and adult tissue total RNA blots were purchased (Seegene Inc. Rockville, Maryland, USA). They were probed with *Mib1* cDNA (1-1170 bp), stripped, and then re-probed with glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) as a loading control.

RNase protection assay

Radiolabeled probes were generated by *in vitro* transcription, gel purified and their specific activity determined. All samples consisted in FirstChoice® Total RNA (Ambion Inc. Austin, Texas, USA). Probe was hybridized to sample, RNase digested/inactivated and precipitated. End products were separated on a 15% urea-polyacrylamide gel and protected fragments detected by autoradiography. The antisense probes used were: *mmu-miR-1-2* GGGAGACAGGUACAUACUUCUUUACAUUCCA; *mmu-miR-133a-1* GGGAGACAGGACAGCUGGUUGAAGGGGACCAA. *Note*. Due to the short length of the probes, *all* U represent ^{32}P labeled uracil (PerkinElmer Inc. Boston, Massachusetts, USA).

RT-PCR

RT-PCR was carried out on individual embryos. 3 µg of total RNA was used for each RT reaction using SuperScript II™ (Invitrogen Inc. Carlsbad, California, USA). 2 µl of the RT reaction was used for every PCR reaction. PCR products were resolved on a

1% agarose gel in 1X Tris-borate-ethylenediaminetetraacetic acid (TBE). *Gapdh* primers were used as control.

Chapter 3 Modeling biological networks

PREFACE

Despite my interest in how studying the Nn first originated and later evolved, I have chosen to omit the historical perspective from chapter one so as not to digress. In contrast, I will open this chapter by addressing what I believe to be pivotal events that drove the modeling of biological large-scale networks, and conclude this preface with the current status of the field.

The latter, which only recently can be referred to as 'a field of study' in and of itself, is comprised of *graph theory* and *network analysis*, which in turn stem from mathematical and social studies, respectively. However powerful a tool it may represent for biologists, it remains an engineering problem rather than a biological one. In the late nineteen nineties, the formation of the Joint Genome Institute (JGI) represented a shift toward high-throughput large-scale analysis. This in turn, precipitated the topic of *knowledge representation* (a rather thorough paper discussing the latter can be found online <http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html> and two straightforward examples where this topic has been addressed in an unrelated field can be found in references[154,155]). The latter, commonly takes the form of rectangular data tables, where each row represents a unit and the columns contain data (or conclusions thereof) that describe the attributes of each unit. This has become a popular method of *knowledge representation* because of its compatibility with algorithmic processing, and hence statistical analysis. However, the very nature of a table implies that all units posses an identical set of attributes.

Ontology represents an alternative approach, whereby the concepts of a given field of study are used to create a formal model. The units (whatever they may represent) are assigned properties and the relationships among units are explicitly defined. However, in order to perform statistical analysis such properties must first be coded as discrete variables.

In anticipation to what the post genomic era would bring, several biologists tackled this issue. Among these, Dr. Minoru Kanehisa spearheaded the Kyoto Encyclopedia of Genes and Genomes (KEGG)[10,156] project and Dr. Kurt Kohn was one of the first to conceptualize an interaction map that is capable of unambiguously representing a network composed of biological entities resolved at the molecular

level[7,157]. Some may argue that the two examples above are distinct in nature; the point I am trying to make is that when faced with how best to represent biological knowledge at the molecular level, both resorted to what ultimately could be viewed as *graph theory*. More recently, an international effort lead by Dr. Hiroaki Kitano has taken this a step further so as to develop an environment in which an all inclusive (by this, I mean that metabolic networks[158-161], cell-signaling pathways[162], protein interaction networks[163,164], gene regulatory networks[13], etc. would all be represented simultaneously) biological super network can be modeled. However complex this may appear, it still boils down to *graph theory*.

The past few years have seen an increased effort to standardize the modeling of biological networks, both in terms of computer compatibility and human interpretability. For the most part, this can be attributed to the Systems Biology Markup Language (SBML)[165,166] and the proposal of process diagrams[6] as defined by the Systems Biology Graphical Notation (SBGN), respectively.

SBML can best be described as a vehicle by which researchers can not only exchange models, but also enables them to run simulations of such models utilizing a wide array of software. In fact SBML is a file format, much like a word document or a portable document format (PDF). At present, a database termed BioModels has been created at the European Bioinformatics Institute (EBI) with which to house SBML compatible models. However, the converse is not true. That is to say, no network model utilizing the SBGN system has been created, which in itself serves as a database. In chapter four of this dissertation I describe one such example.

Finally, all of the efforts described above without a doubt serve the intended purpose of their creators. In some instances, such tools overlap with the interest of others and may be further exploited. Nevertheless, none are modular. Interestingly enough, modularity is not only an important property of complex systems, but as it turns out, has become a necessary feature of the very tools utilized for their analysis. The degree, to which previous efforts have been wasted, is tremendous. In the past, researchers resorted to different notation systems and simulation environments, such that their models were only rarely explored by others, but never reused. To some extent, this has been corrected by SBML and has been further addressed by Dr. Nicolas Le Novère and colleagues[167]. However, many of the models found within the BioModels Database are redundant. The value of such models is limited without modularity, since there is no way with which to incorporate them into a larger model.

At present, I am aware of only one group attempting to resolve this issue, that of Dr. Jeremy Gunawardena. Their project, termed *Little b*, consists of a modular modeling language based in turn, on the programmable programming language known as LISt Processing (LISP).

Ultimately, modeling biological networks aims to integrate both topological and kinetic properties with the intention of resolving a complex scenario. If we are to achieve this aim, future research should address the method of computation *per se*. In order to execute a command, it is not uncommon for computer programs to employ symbolic mathematics, which in turn are enabled by a computer algebra system. For simulation purposes, such a program could operate under *control theory* (CT), *metabolic control analysis*[159] (MCA) or *soft computing*[168] (SC). An engineer might favor CT and compile a mathematically elegant, though completely impractical program. The reason being is evident; it would be impossible to obtain all the necessary values for the many differential equations. For instance, the dissociation constant (Kd) for every reaction modeled under different environments. Hence the beauty of SC (which synergistically resorts to fuzzy logic, neural network theory, and probabilistic reasoning), which (unlike other methods of computation) is tolerant of imprecision, uncertainty and partial truth.

The level of success of such projects, will also hinge on the level of sophistication of their graphical interface.

I will close by paraphrasing Dr. Eörs Szathmáry; *…its main intellectual 'deliverable'* [in this case referring to the modeling of biological networks] *will be to show that we have understood some basic biology* [as evidenced through a model's predictive capacity]*; just as the total synthesis of a molecule proves that the chemist knew what he was doing*.

GRAPH THEORY

What follows, is a brief outline of the most basic concepts in graph theory, in particular the elemental components of a graphical model and a few properties of the network resulting thereof. For a more thorough introduction, I encourage the reader to enjoy reference[169].

Graphical model

In its simplest form a graphical model, as it is understood by graph theory, consists of a set of ‘nodes’ interconnected by ‘edges’. Several layers of complexity may be added to this scheme. For instance, ‘edges’ which by definition are undirected, may be given direction and in so doing are referred to as ‘arcs’, these give rise to ‘directed graphical models’. Edges and/or arcs may also be ‘weighted’. Nodes themselves need not be of the same type, this gives rise to multipartite graphical models. In fact the latter is particularly useful when attempting to represent different networks (such as genetic, protein and metabolic information) within a single super network. The complexity which emerges from large-scale networks is surprising, given how simple this form of *knowledge representation* is at its core.

Network properties

If at least one path exists between any two nodes of a graphical model, the model is considered to be ‘connected’. The number of nodes to which a given node is directly connected, is referred to as its ‘degree’. In the case of a directed multipartite graphical model, for any given node there will be an ‘in-degree’ and an ‘out-degree’ for however many different types of arcs exist within the model. By ‘neighborhood’, it is understood that one is referring to a space defined by a given node and those connected directly to it.

All of the statistical measurements applicable to the other forms of *knowledge representation* mentioned above (rectangular data tables and ontology) may also be readily applied to the data contained within these types of models. What distinguishes this approach from the others is the topological structure that emerges from the graphical model of a large-scale network. Because this particular form of *knowledge representation* focuses on relationships, one may now in addition to the above, perform statistical measurements on the relationships themselves and in so doing, uncover the properties of a particular topological structure.

In the case where a directed multipartite graphical model is used to represent biological molecular interactions, it is not entirely obvious that separate statistics should be compiled for each type of node and/or arc. Whether such components are treated differently or in unison will ultimately depend on what is being thought after.

A few of the more simple statistical measurements include 'average node degree', 'average path length' and the 'clustering coefficient'. The average node degree reflects to a certain extent the density of the graphical model. It is defined as the average number of edges *per* node. To illustrate the case of a directed multipartite graphical model with respect to statistical analysis, a generic average node degree could be calculated across all types of nodes and/or arcs, or a specific average node degree could be calculated for each type of node-arc pair. Which analytical route should be taken deserves special attention, as each will lead to different conclusions. By and large, the same level of caution should be taken for all statistical measurements carried out on any graphical model other than the most basic in nature. The average path length is the average distance between any two given nodes. The clustering coefficient can be thought of in terms of the extent to which the neighborhoods of two neighboring nodes overlap. This carries a value between zero and one, the latter representing a fully connected graphical model.

These and many more properties help define the topological structure of a particular graphical model. Historically, the properties of random graphical models were amongst the first to be explored, and hence such topological structures are well defined. Graphical models based on real world data however, share common global properties, in particular the ways in which they differ from random graphical models. For instance, if one were to analyze the three statistical measurements defined above, one would find that the global topological structure of a graphical model pertaining to real world data is defined by:

- A distribution in node degree characterized by a 'power law'; that is to say that only a small number of nodes have many neighbors, whereas the majority of nodes have but a few.
- A relatively short average path length.
- A clustering coefficient close to one.

  For more information see references[170,171].

Kitano and colleagues in several related studies, point out the presence of 'highly connected hubs' within biological networks[162,172]. This could very well suggest a power law distribution in node degree, however these 'hubs' have been empirically defined rather than through statistical analysis. Similarly, the 'robustness' of a network is often

discussed without any formal measure. In such a case, the average path length and clustering coefficient could indicate the statistical likelihood that by removing any node at random, it would constitute a 'hub', and by extension the likelihood of changing the overall topology of a network. In other words, such statistical measurements could help better define robustness.

In part, these characteristics define the underlying structure of a network. However, to further differentiate amongst networks with similar global properties (for instance, two distinct signal transduction networks); one's attention must focus on the local topological structure. Currently, such efforts are directed towards identifying 'network motifs'[173-175]. In a nutshell, this consists in identifying topologically distinct subsets within graphical models, and then exploiting their frequency of occurrence within each model in its entirety; to characterize the topological structure at a local level[176].

With the closing of this chapter, I would like to clarify that in it, I present a completely biased view on the modeling of biological networks and furthermore, I do not pretend to have covered any of the mentioned topics in an exhaustive manner. In short, this chapter highlights those aspects related to the modeling of biological networks, which I personally view as interesting and worthy of further study.

Chapter 4 Murine Notch Signal Transduction Database

PREFACE

Most molecular, cellular and developmental biologists, as well as geneticists and surely numerous other specialists, work in a given area of research because they are attempting to answer a particular question pertaining to biological function. Although this might seem obvious at first, given that they are all biologists, this is not necessarily true for an important sector from each of the guilds mentioned above. In these latter cases, research can primarily be viewed as goal oriented. That is to say, it is not a question *per se* that is the motivating factor, as much as an obstacle to be solved. Traditionally, this distinction used to be perceived as the 'basic science' carried out in academia *versus* the 'applied science' of the private sector. However, both motivating factors are currently prominent in either sector.

Regardless of which sector and/or motivation drives research, it is safe to say that only a minority of the researchers are clever enough to design the correct experiments which can directly answer a given question or solve a given problem. Nevertheless, the bulk of all experiments, though they may have failed to resolve the original question or problem, by and large still generate quality data. Furthermore, a number of researchers altogether skip the hypothesis stage, and the gathering of quality data in and of itself becomes the motivating factor (for instance, the identification of all the possible protein-protein interactions on a genome wide scale[163,164] or the creation of a mouse atlas of gene expression[177-179]). If in addition to the above, we factor in the fact that an ever increasing number of researchers are resorting to high-throughput analysis, either to complement more traditional methods or in some instances as a replacement of such methods, it becomes clear that data complexity as well as the overall volume of data, is on the rise.

In order to gain knowledge from the plethora of available data, the use of one or more databases has become an integral part of the scientific process. Nonetheless, there are few tools which enable the visualization of data across multiple scales and datasets. Ideally, one could envision modeling biological networks in which the model itself serves as a portal to the underlying database. This database in turn, would serve as a reservoir for only those datasets which support interactions mapped out in the model. If, as I argued earlier, this were engineered in a modular fashion, then models

and data alike could be integrated into a common framework. More importantly, certain datasets would surely overlap, and in so doing new relationships would emerge which could then be mapped back into the model. Only in this way will we commence to understand biological processes in their entirety.

AIM

The main purpose behind this database was to create a proof of concept for part of a more general theme. Given the nature of my previous research topic, I chose to model *Notch* signal transduction as observed in *Mus musculus*. In particular, this proof of concept was to demonstrate the usefulness of generating a graphical model (in this case, referred to more specifically as a molecular interaction map) using a clearly defined notation system tailored for molecular biology, which in itself would serve as a portal for an underlying database containing all information pertinent to the model.

Moreover, this was organized in such a manner so that every mapped interaction would contain a hyperlink directed towards all data from which the interaction itself was inferred. In this way, the user could judge the nature and quality of the data without having to wade through the entire list of papers, though these are made available should one care to delve deeper.

In addition, I wanted to generate a theoretical framework with which to analyze, model and to a limited extent predict, network behavior when natural or targeted mutations are introduced into the Nn.

The entire model/database was coded using strictly a web-based platform such that all components were generated in a hyper text markup language (HTML) compatible format and are freely accessible through the world wide web (WWW) to anybody with an internet connection.

RESULTS

Uniform resource locator (URL)

The Murine Notch Signal Transduction Database is currently hosted by the UT biological sciences computer support facility (BSCSF) and may be accessed at the following URL <http://web.biosci.utexas.edu/artzt_lab/notch>.

Interface

The Murine Notch Signal Transduction Database portal consists of two main components (Figure 20).

The first allows the user to access the database through a query box, either by explicitly searching for a component or by simply leaving the query box empty and activating the search button. The latter action directs the user to an alphabetically organized list of all the database components. In addition, a help toggle in the form of a question mark has been added adjacent to the query box such that when activated, instructions appear as how to operate the database (Figure 21).

The second component requires *Macromedia Flash Player* (Adobe Systems Inc. San Jose, California, USA) version 6 or later to be installed on the user's web browser. *Macromedia Flash Player* is a multi-platform client that enables rich content interface compatible with most web applications. It has been chosen in part for its popularity. The fact that it has been installed on 98% of internet enabled desktops worldwide assures that very few users will experience compatibility issues when attempting to access the database. It is this second component which offers a unique approach with respect to how biological data is organized and accessed.

The particular configuration employed consists of two windows, a 'preview' window in which the graphical model in its entirety is displayed and a 'diagram' window below it, which reveals the molecular interaction map in full detail. A diagram 'key' (Figure 21) explaining the SBGN system employed, can be accessed through a second toggle, labeled 'click for help' (Figure 20). The interface is equipped with a tool that enables the user to zoom in at will anywhere within the graphical model (Figure 22).

For this particular database, the graphical model was drawn so as to represent two adjacent cells. For clarity, Nn receptors were mapped to one cell, whereas Nn ligands were mapped to the other.

Database

All components visible through the diagram window, hyperlink out to the underlying data which justifies their position within the molecular interaction map. This information is displayed under four main categories (underlined below) and their respective subcategories (Figure 23). These are organized as follows:

Component Info
Component Details
- Name
- Synonyms
- Component Type
- Dimer Number
- Compartment
- Molecular Weight
- Chromosomal Location
- Amino Acids
- Motifs/Domains

Linked Components
- Directly Connected
- (n) Level Indirect

Associated Interactions & Components
- Interaction (n^o)
  - Interaction Type
    - Interaction Table (Component Name, Type, Compartment and Connection)
    - Experiments Associated With This Interaction
      - Experiment Details
        - Related Text
        - Image
      - Paper Details
        - Author
        - Date Published
        - Title
        - Journal
        - PubMedID
        - Abstract

For the most part, each subcategory is self-explanatory. However, I will briefly explain subcategories with variable parameters or those instances where a subcategory's content cannot be directly inferred from its title.

Both, 'Component Type' (i.e. receptor) and 'Compartment' (i.e. cytoplasm) hyperlink out so as to reveal all other components of the same type or those that reside within the same compartment, respectively.

'Chromosomal location', displays the genomic location for any given component in physical terms (chromosome number and base pair range) in *Mus musculus* and hyperlinks out so as to view the location with Ensembl Mouse ContigView.

'Amino acids', in addition to displaying the number of residues for any given component, hyperlinks out to the Geneatlas: Gene Database.

In order to facilitate the analysis of network topology and/or predict repercussions throughout the network, I created a subcategory labeled 'Linked Components'. This in turn, is further subdivided in two. The first, labeled 'Directly Connected', identifies all immediate neighbors of a given component. Whereas the second, labeled 'n Level Indirect', enables the user to choose the level of indirectly connected components displayed.

In as far as data *per se* is concerned; each 'Image' expands so as to reveal a full resolution image of the relevant data.

Finally, the 'PubMedID' subcategory hyperlinks out so as to reveal the corresponding paper as indexed by PubMed. However, if the user is logged on to the Artzt laboratory web site, an additional hyperlink becomes visible, by which the user may directly download the paper as a PDF.

Administration

One of the main factors behind this endeavor was to create a format in which a model can be amended as new data becomes available. The nature of the database is such, that only minimal effort is required for its upkeep. Though most issues have been addressed so as to facilitate this process, the database is not, nor was it ever intended to be automated and hence, requires a curator.

The latter, whether a single individual or a team, is granted administrative privileges *via* a secure login <http://web.biosci.utexas.edu/artzt_lab/admin/login/login.aspx> accessible from any

computer with an active internet connection (Figure 24a). The fact that most modifications to the database can be made online simultaneously by different administrators, further streamlines this process. Upon successful login, the curator is directed towards the administrator portal (Figure 24b); from which four distinct web pages can be accessed. These pages allow all aspects of the database *per se* to be modified at will (Figure 25).

There is one feature whose modification remains unavailable by hyper text transfer protocol (http), that of the Macromedia Flash Player enabled interface. Once modified however, this can easily be updated through a file transfer protocol (ftp) connection.

Analytics

In an effort to assess the response generated by the Murine Notch Signal Transduction Database, statistics regarding internet traffic were gathered during the spring semester of 2006 (specific range: 01/01/2006 00:00:00 – 05/08/2006 11:40:28).

During this time, the database had been accessed a total of 23,578 times. For comparative purposes, the Artzt laboratory home page was visited only 588 times within the same time frame. The average time spent on the database was 00:07:27; however interpretation of this value is limited. 71.64% of the users were confirmed to be accessing the database from within the United States. Perhaps of more significance is the fact that 391 distinct users [as defined by their internet protocol (IP) address and further authenticated through cookies] accessed the database recurrently with session lengths lasting a minimum of 00:30:00.

I am fully aware that from these statistics one cannot accurately judge the level of success (or lack thereof) of the database. Nonetheless, the numbers are encouraging.

DISCUSSION

Although the Murine Notch Signal Transduction Database was created as a proof of concept, I must admit I had anticipated submitting it for publication. Early on however, it was brought to my attention that in its original form, I would be infringing copyright laws.

As described elsewhere, one main focus was to serve up all relevant data surrounding a given model. To this end, numerous papers spanning a great number of subscription-only journals were 'harvested'. That is to say, the key experiment (in as far as a given model is concerned) within each paper was identified, stored, and cataloged along with a full version of the paper itself. Therefore an important part of all information within the database is currently under copyright protection. Hence, I was cautioned against making all of this freely available. Furthermore, I felt that restricting the type of information available would strip the database of its most important feature, and for this reason I have not pursued its publication.

As it currently stands, the Murine Notch Signal Transduction Database is fully operational to all members of the UT community, thanks in part to the UT library system subscription umbrella. As for those outside the UT community, unfortunately direct access to all experimental data has been restricted, as have all papers. In lieu of the latter, a hyperlink to PubMed has been provided for each of the papers cited within the database.

As of today, there is one key feature that remains to be integrated to the model/database. Briefly, it consists of a simple query as to the *developmental stage* of *Mus musculus* and *tissue* of interest. In response to this query, only the subset of mapped components which are co-expressed in time and space would be highlighted within the graphical model.

MATERIALS AND METHODS

Website development

The Murine Notch Signal Transduction Database was created in collaboration with Zack Mahdavi as a series of extensible markup language (xml) and html *Flash* enabled files utilizing *Macromedia Studio* version 8 (Adobe Systems Inc. San Jose, California, USA).

Nn molecular interaction map

Molecular interactions were inferred from the literature. Relevant articles were identified by searching the MEDLINE/PubMed database. Journals were accessed

through the UT library system or purchased whenever necessary. All molecular interactions found in this manner considered important to the Nn as observed in mammalian systems, were first mapped out utilizing *CellDesigner* version 3.0 (http://celldesigner.org/) so as to create a template from which a *Macromedia Flash Player* file was generated.

Analytics

Statistics regarding internet traffic through the Artzt laboratory web space was collected with *Analytics* version 8 (WebTrends Inc. Portland, Oregon, USA).

Chapter 5 Conclusion and future research

In this dissertation I describe, through *in vivo* experimentation, a core signaling component of the Nn previously uncharacterized in a mammalian model organism. I conclude that the gene officially known as mind bomb1, is essential for *Mus musculus* embryogenesis due to its requirement for *Notch* signal transduction. In addition I present an *in silico* model of the Nn, which serves as an online portal to a mammalian Nn molecular interaction database. The latter I have named, the Murine Notch Signal Transduction Database and it may be accessed online at <http://web.biosci.utexas.edu/artzt_lab/notch>.

Throughout my studies here at UT, I have experienced a change in how I desire to approach biology. As I believe is common, at first I strove to obtain high quality data. However, I am at present attracted to the prospect of uncovering previously unrecognized principles which govern all aspects of biology. I firmly believe that the key to the above lies in utilizing established methods or perhaps developing new ones, to extract biological knowledge from the plethora of data available. By no means should this be confused with data mining, rather it should be viewed as defining a given topic *a priori* based purely on personal interest, then approaching the topic from all angles (Top-down, Bottom-up, *in vivo*, *in vitro* and *in silico*).

At present, I cannot help but recognize two general topics which promise to be particularly fertile territory. The *Flow of Information* through biological systems, and the study of *microRNA*s during mammalian development. These topics are clearly distinct in nature, whereas the first will benefit from the vast volume of data already available, the second because in its infancy will require an emphasis on generating *in vivo* information. In the immediate future, I intend to address both these topics precisely as described.

Let me finish by stating that although I strongly commend translational research, I have always supported science for science's sake. That is to say, I never felt the need to justify basic science by invoking its potential application towards healthcare. It is embarrassing to the extent to which many biologists go in order to justify their research, knowing full well that their results, though valuable to biological understanding, will unlikely be translated to a clinical setting. I would argue that attempting to predict which projects most likely constitute translational research, in fact diminishes the chances of success (as defined by a blockbuster drug; though I do not agree with this particular

definition)[180]. It is the nature of basic science which ultimately allows for a greater understanding of biology and with that, comes the potential to improve healthcare.

Substrate / E3	*Cbl*	*Ccne1*	*Dapk1*	*Eps15*	*Itch*	*Jun*	*Myc*	*Numb*	*Psen1*
Fbxw7		●[85-87]				●[88]	●[87]		●[89]
Itch	●[90]								
Lnx1								●[91-93]	
Mdm2								●[94, 95]	
Mib1			●[84]						
Nedd4	●[90]								
Numb				●[96, 97]	●[74]				
Siah1a								●[98]	

Table 1. Nn interactions that warrant further research. A black dot denotes the interaction results in a ubiquitination event; in which case, the rows represent the active E3 and the columns represent their substrates (some of which are themselves E3's whose substrates are not included within this table). A red dot denotes that the nature of the interaction remains unknown. Casitas B-lineage lymphoma (*Cbl*, *Cblb* and *Cblc*) family, cyclin E1 (*Ccne1*), death associated protein kinase1 (*Dapk1*), epidermal growth factor receptor pathway substrate15 (*Eps15*), F-box and WD-40 domain protein7 (*Fbxw7*), itchy (*Itch*), jun oncogene (*Jun*), ligand of numb-protein X1 (*Lnx1*), mind bomb1 (*Mib1*), myelocytomatosis oncogene (*Myc*), neural precursor cell expressed developmentally down-regulated gene 4 (*Nedd4*), numb (*Numb*), presenilin1 (*Psen1*), seven in absentia1A (*Siah1a*), transformed mouse 3T3 cell double minute2 (*Mdm2*). In all instances, the references from which the interactions were inferred, accompany each dot.

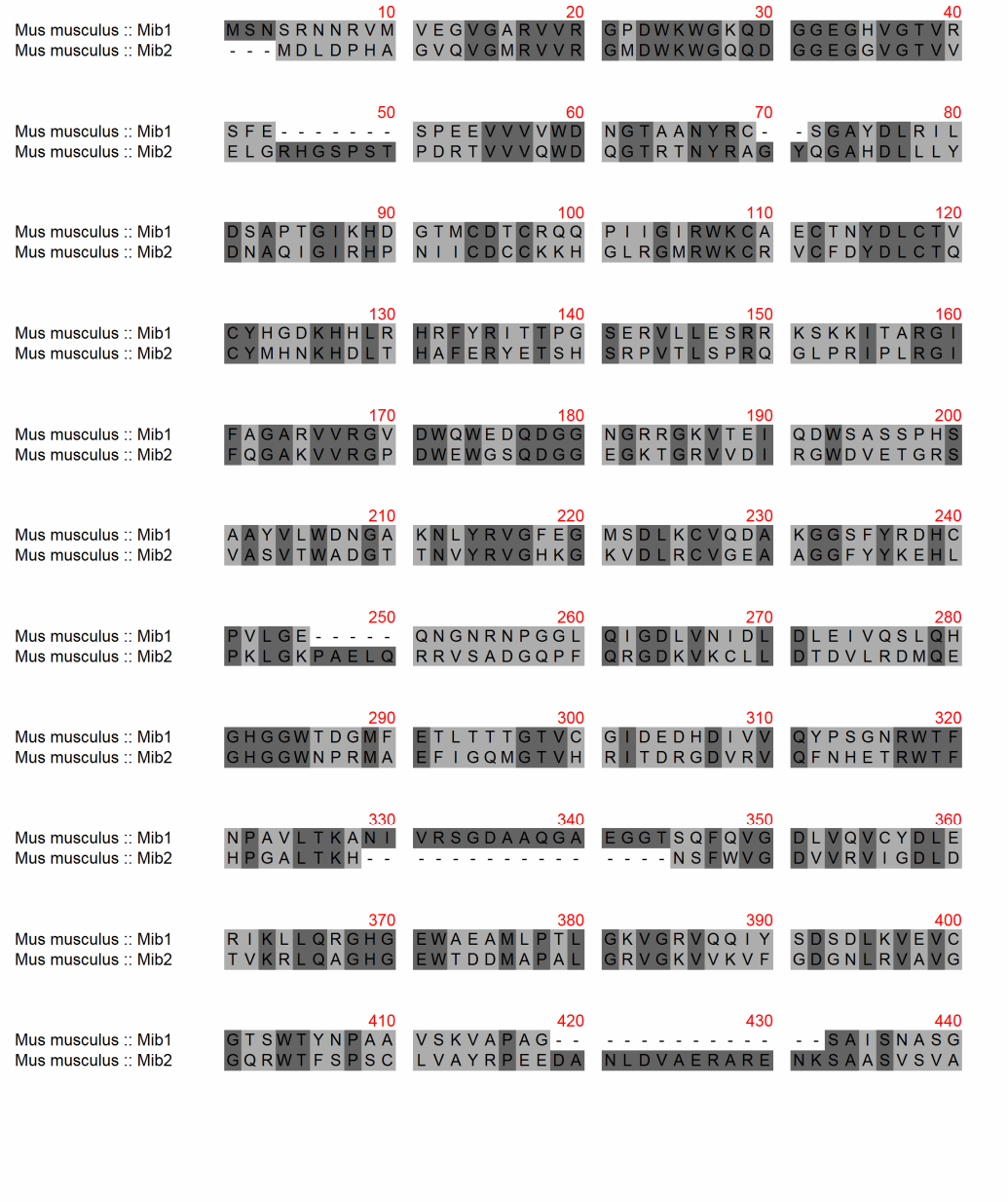


Figure 1. Protein sequence alignment of *Mus musculus Mib1* and *Mib2*. Different shades of grey reflect sequence similarity for any given position. Identical residues have been shaded dark grey, whereas similarities are a lighter tone of grey and mismatches were left uncolored. Every tenth residue has been numbered in red.

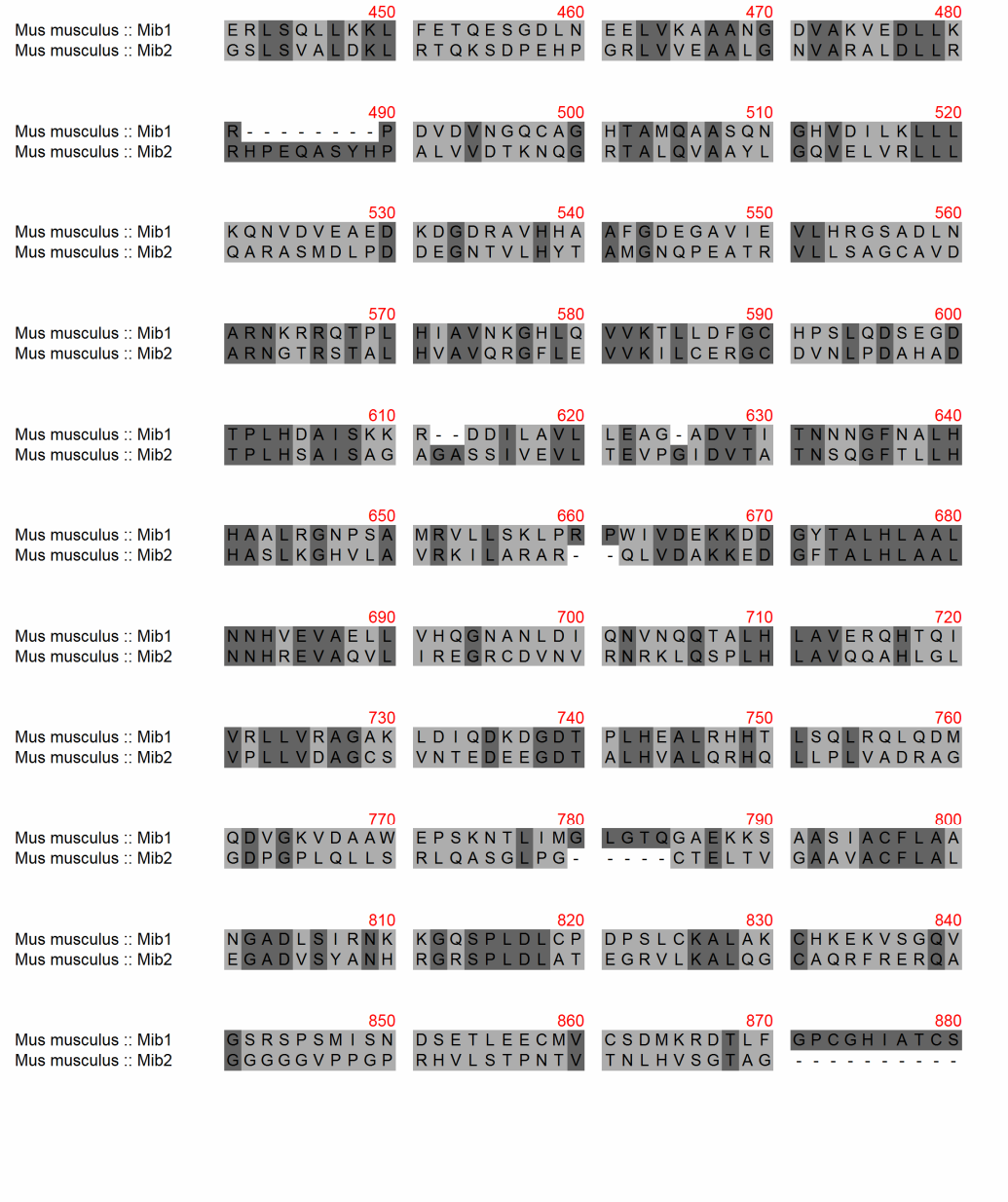


Figure 1 (*continued*). Protein sequence alignment of *Mus musculus Mib1* and *Mib2*. Different shades of grey reflect sequence similarity for any given position. Identical residues have been shaded dark grey, whereas similarities are a lighter tone of grey and mismatches were left uncolored. Every tenth residue has been numbered in red.

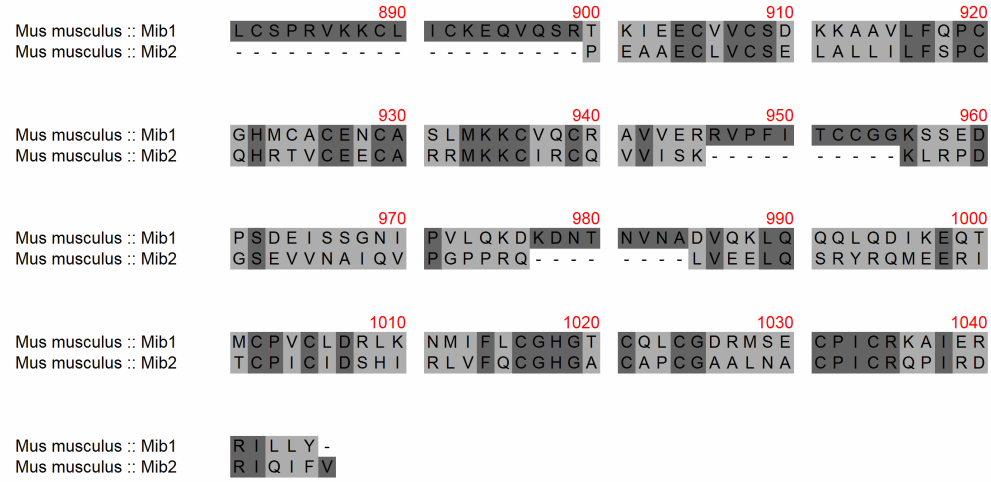


Figure 1 (*continued*). Protein sequence alignment of *Mus musculus Mib1* and *Mib2*. Different shades of grey reflect sequence similarity for any given position. Identical residues have been shaded dark grey, whereas similarities are a lighter tone of grey and mismatches were left uncolored. Every tenth residue has been numbered in red.

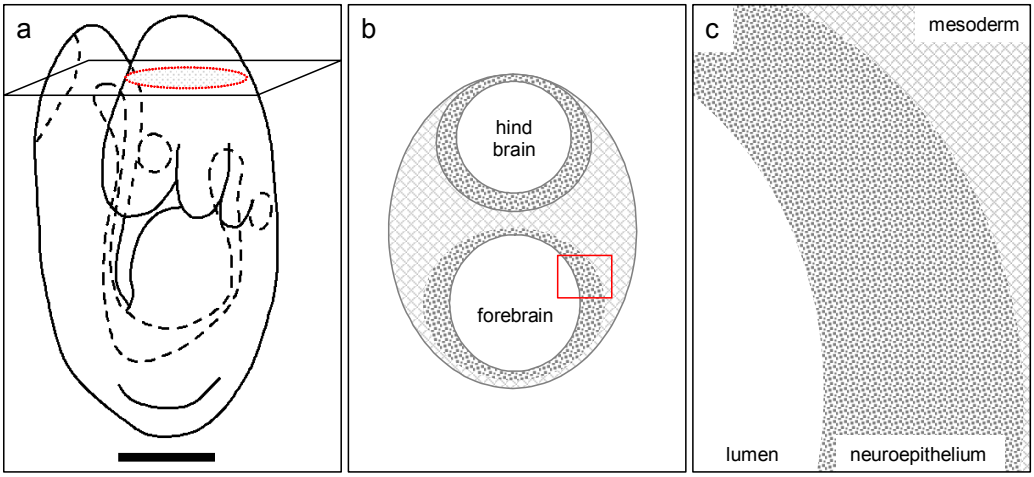


Figure 2. Sketch illustrating the different levels of magnification at which *in vivo* data was collected. (a) Whole mount. As an example, an E9.5 embryo is depicted showing the plane through which sections were taken. Area outlined in red corresponds to what is shown in 'b'. Scale bar represents 500 µm. (b) Histological transverse section of embryonic brain. Area outlined in red corresponds to what is shown in 'c'. (c) High magnification of the neuroepithelium.

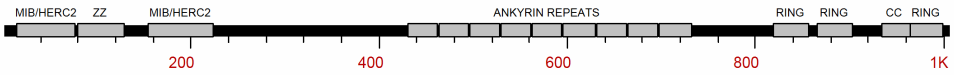


Figure 3. Graphical representation of *Mus musculus Mib1* protein domains, regions and repeat locations with regard to the amino acid sequence numbered in red. (MIB/HERC2) Mib/Herc2 domain. (ZZ) ZZ-type Zinc Finger region. (ANKYRIN REPEATS) Ankyrin repeats. (RING) RING-type Zinc Finger region. (CC) Coiled-coil region. N-terminus at left.

Figure 4. *Homo sapiens*, *Mus musculus* and *Danio rerio Mib1* protein sequence aligned across species. Different shades of grey reflect sequence similarity for any given position. Identical residues have been shaded dark grey, whereas similarities are a lighter tone of grey and mismatches were left uncolored. Every tenth residue has been numbered in red.

Figure 4 (*continued*). *Homo sapiens*, *Mus musculus* and *Danio rerio Mib1* protein sequence aligned across species. Different shades of grey reflect sequence similarity for any given position. Identical residues have been shaded dark grey, whereas similarities are a lighter tone of grey and mismatches were left uncolored. Every tenth residue has been numbered in red.

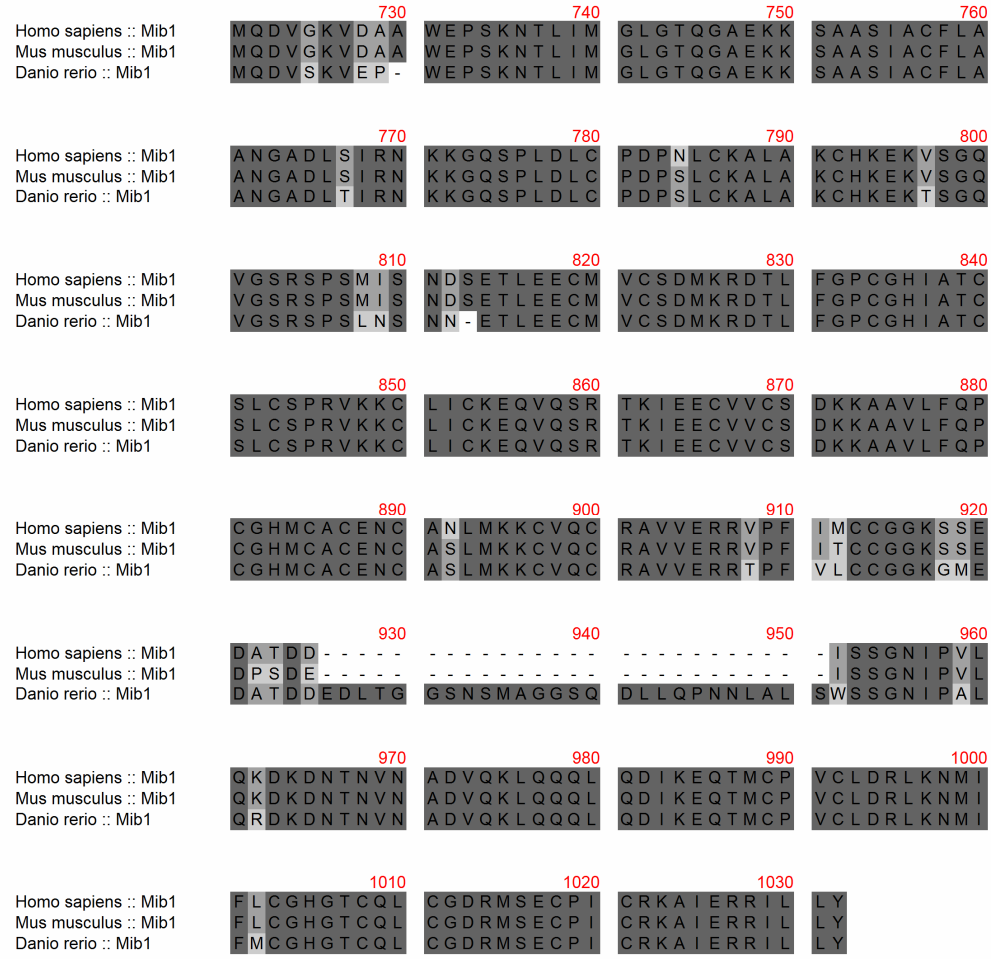
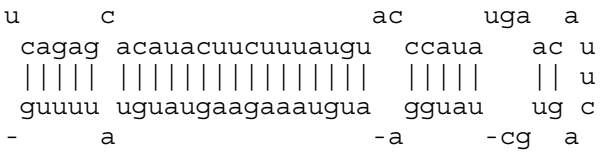


Figure 4 (*continued*). *Homo sapiens*, *Mus musculus* and *Danio rerio Mib1* protein sequence aligned across species. Different shades of grey reflect sequence similarity for any given position. Identical residues have been shaded dark grey, whereas similarities are a lighter tone of grey and mismatches were left uncolored. Every tenth residue has been numbered in red.

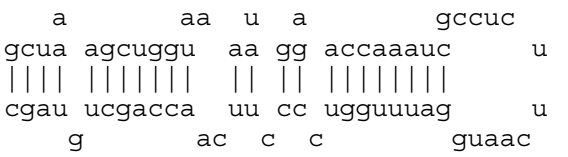
(a) *Mus musculus miR-1-2* stem-loop



(b) *Mus musculus miR-1-2* mature sequence

uggaauguaaagaaguaugua

(c) *Mus musculus miR-133a-1* stem-loop



(d) *Mus musculus miR-133a-1* mature sequence

uuggucccuucaaccagcugu

Figure 5. Two distinct miRNAs encoded within *Mib1*'s twelfth intron. (a, c) Predicted secondary structure. (b, d) Biologically active mature sequence.

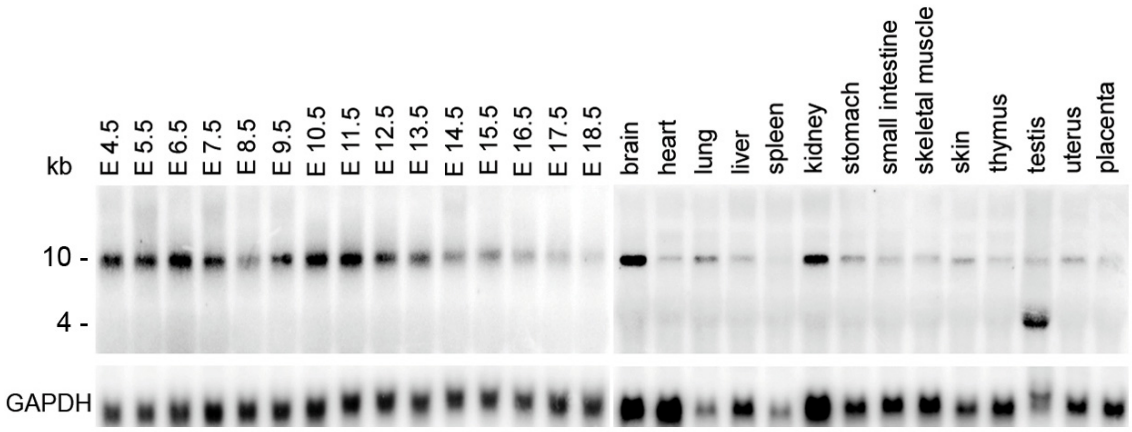


Figure 6. Northern blot. *Mib1* protein coding transcript expression throughout embryonic development and in several adult tissues. Data collected by Jiang Wu.

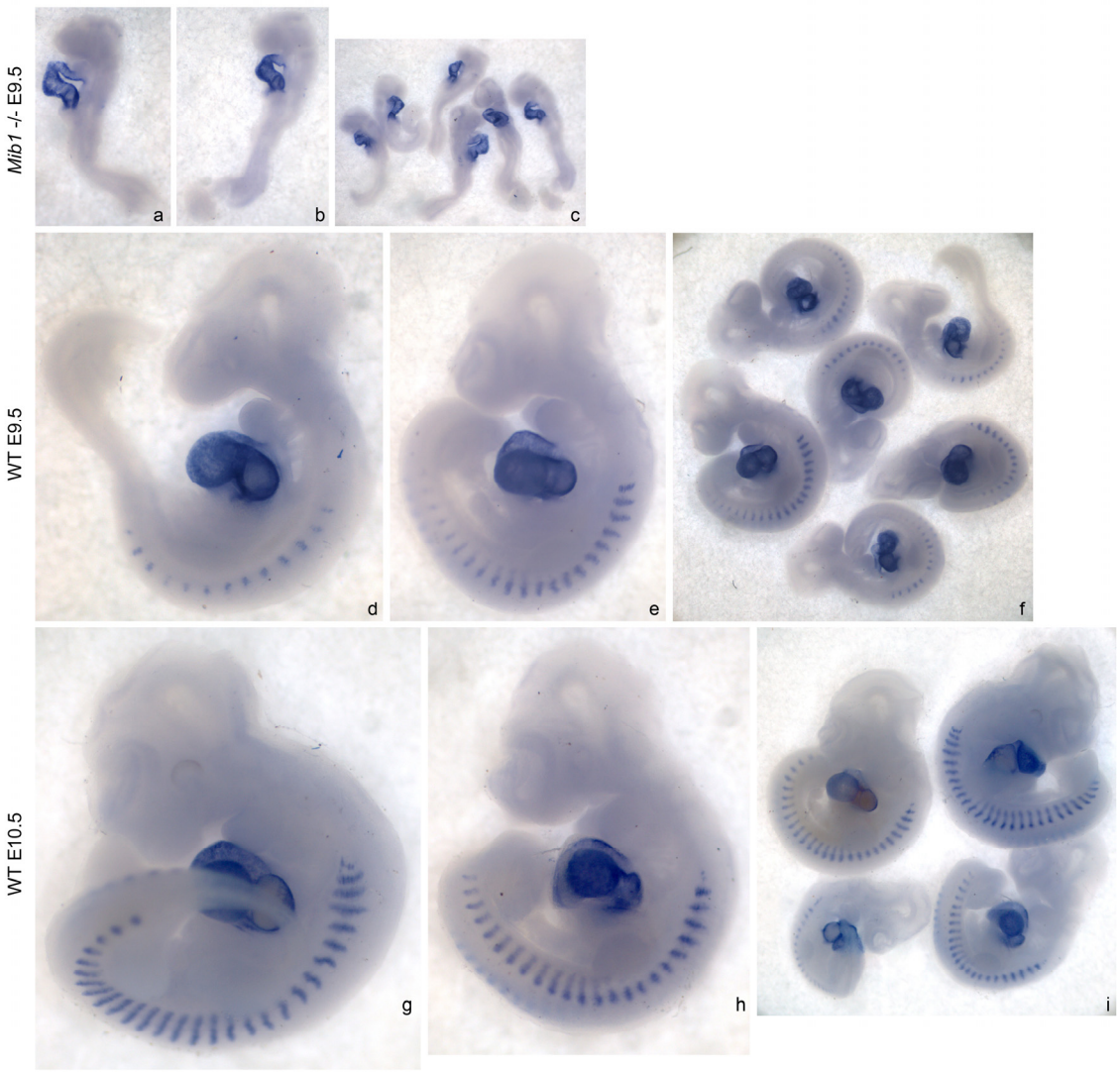


Figure 7. Whole mount LNA *in situ* hybridization. Embryonic expression pattern of the mature form of *mmu-miR-1-2*. (a-c) *Mib1*$^{-/-}$ E9.5. (d-f) WT E9.5 littermates. (g-i) WT E10.5. All images captured with the same magnification, except for (c, f, i), which correspond to a lower magnification. Data collected by Maren Yngve.

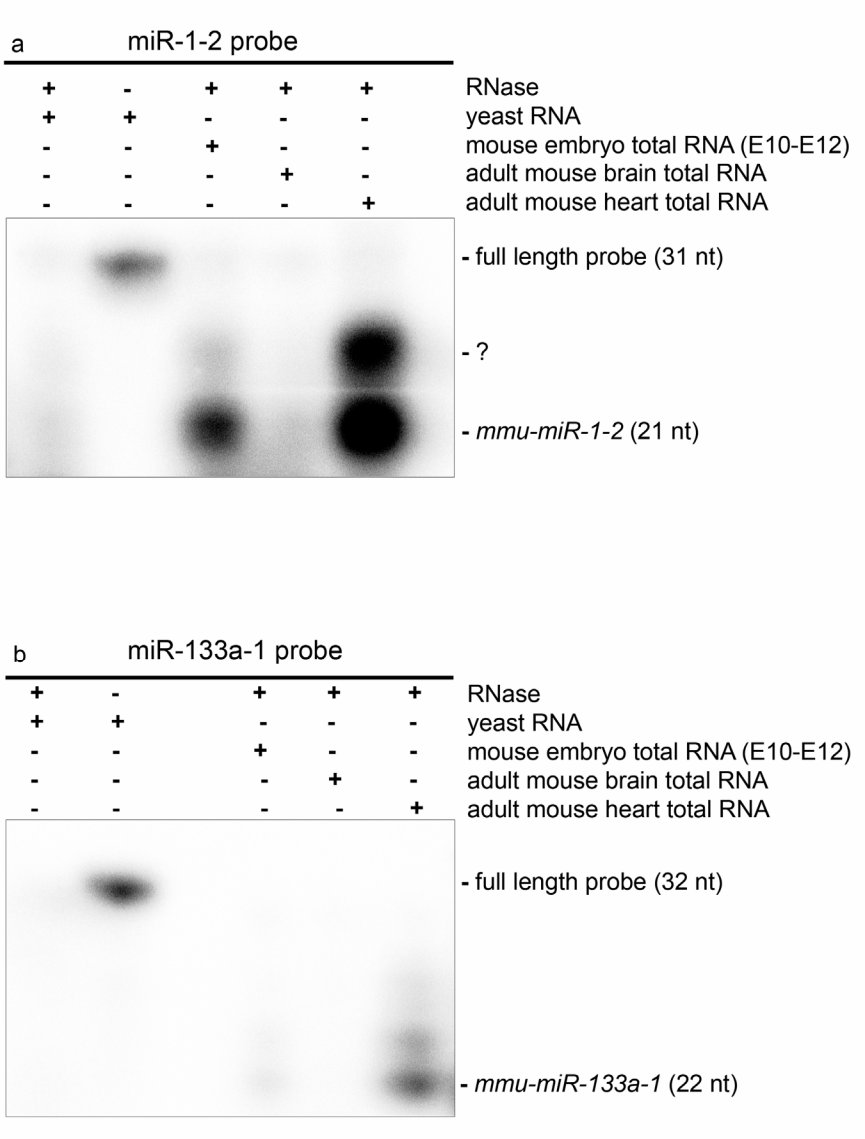


Figure 8. *Mib1*-encoded miRNA RNase protection assay. *Mib1* non-coding transcript expression during embryonic development (E10-E12) and in adult brain and heart. (a) *mmu-miR-1-2* expression. Question mark may represent a slightly larger mature miRNA with similar sequence. (b) *mmu-miR-133a-1* expression.

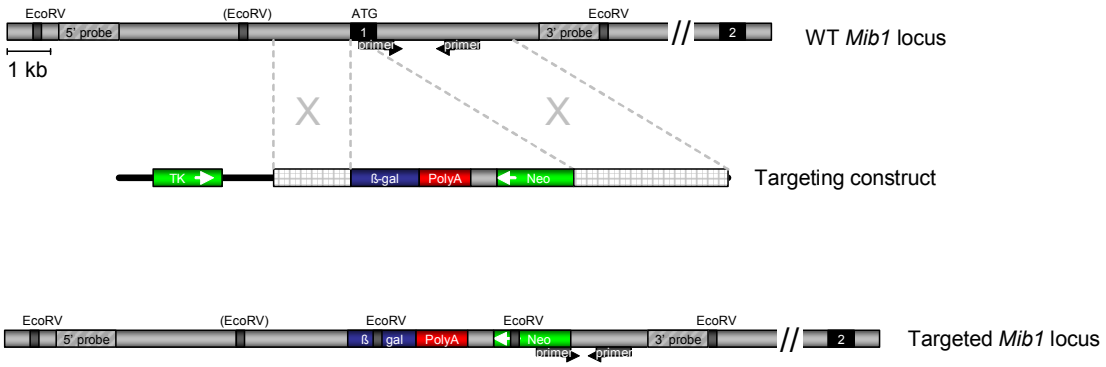


Figure 9. *Mib1* knockout strategy. Organization of the *Mib1* WT allele 5' region and targeting construct. The targeted allele resulting from homologous recombination has most of its first coding exon replaced. Exons (1) and (2) are depicted in black, with (ATG) indicating the start codon. (EcoRV) was used to digest genomic DNA, with dark grey bands indicating cleavage sites. The second cleavage site is polymorphic. It exists in B6 mice, but not 129. The two pairs of arrows which read (primer) below WT and targeted alleles, indicate the positions of the primers used for genotyping. (TK) Thymidine kinase. Homologous region shown as textured pattern. (β-gal) *β-galactosidase*. Note: *β-galactosidase* showed no activity *in vivo*. (5' probe) and (3' probe) were used for Southern hybridization. (PolyA) Simian Virus 40 poly adenylation signal. (Neo) Phosphoglycerate kinase-neomycin cassette.

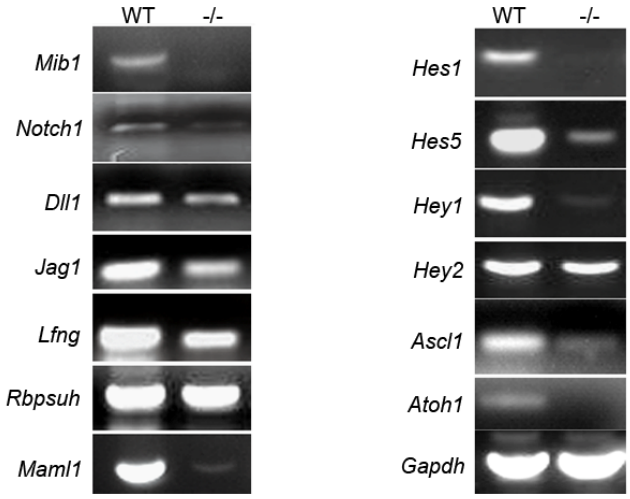


Figure 10. RT-PCR analyses. Expression levels of various genes analyzed at E9.5. Genes analyzed are listed. Individual *Mib1*$^{-/-}$ embryo and WT littermate used. Experiment done in duplicate. Data collected by Rashmi Rajendra.

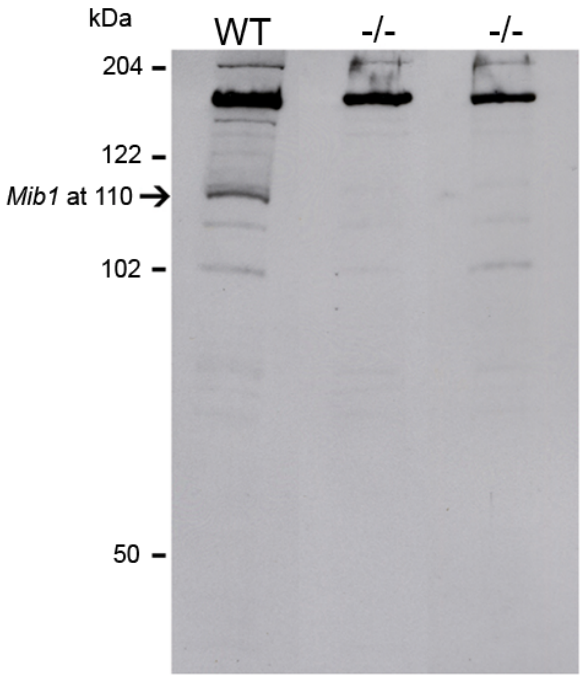


Figure 11. *Mib1* immunoblot. Left lane corresponds to WT, right two lanes *Mib1* homozygous null. *Mib1* migrates as a 110 kDa protein. Heavy band at around 200 kDa is non-specific cross-reactivity. Single E9.5 embryo lysate loaded *per* lane.

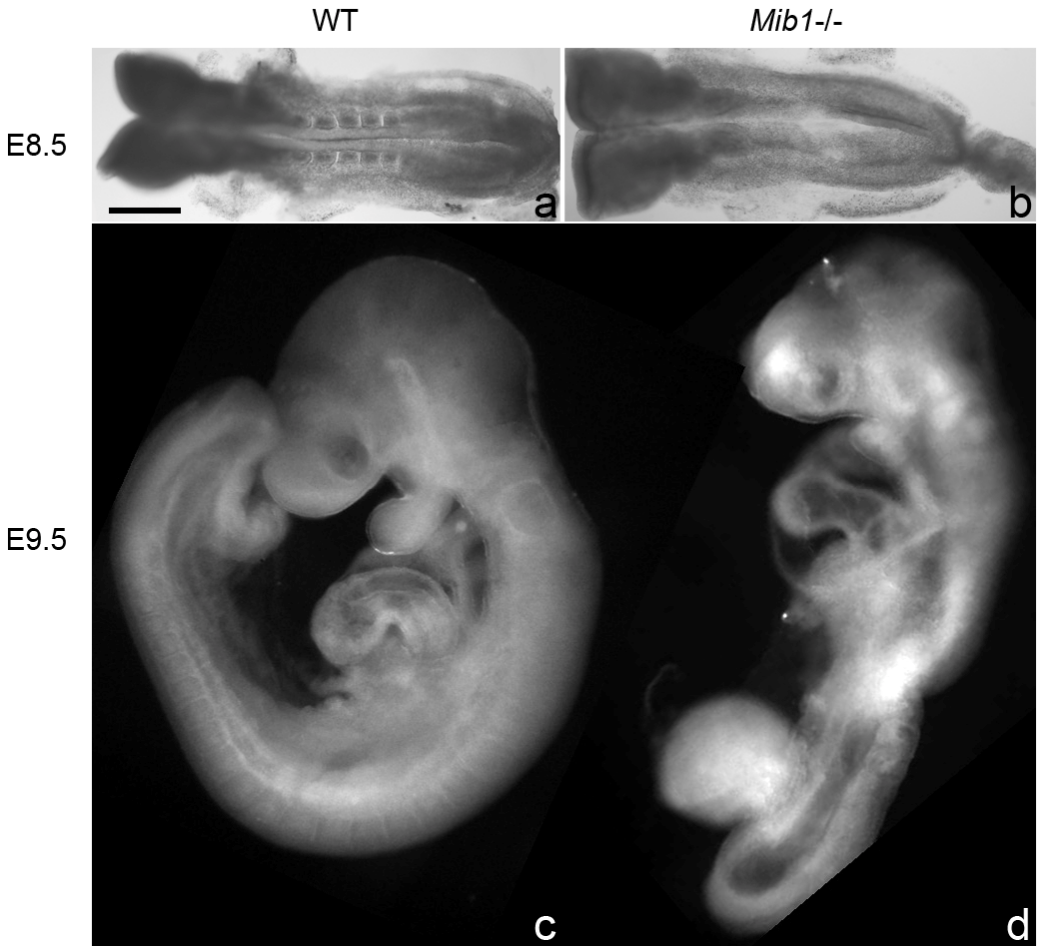


Figure 12. *Mib1*$^{-/-}$ phenotype. Unstained whole mount embryos at E8.5 and E9.5 (a, b and c, d, respectively). WT littermates (a, c) alongside *Mib1*$^{-/-}$ (b, d). Scale bar represents 500 μm. Photograph by Rashmi Rajendra.

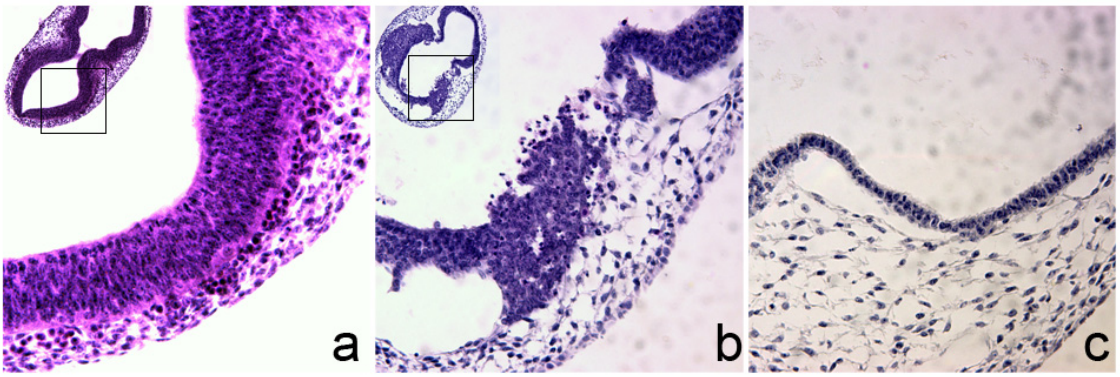


Figure 13. Transverse sections through embryonic brain at E9.5 stained with H&E. Entire field of view inset upper left hand corner in (a) and (b), area outlined delimits region of higher magnification. (a) WT littermate control reveals a normal pseudostratified neuroepithelium. (b) *Mib1*$^{-/-}$ in which structural integrity has been compromised. (c) *Mib1*$^{-/-}$ depicting a dearth of neuroepithelial cells. All sections 7 µm.

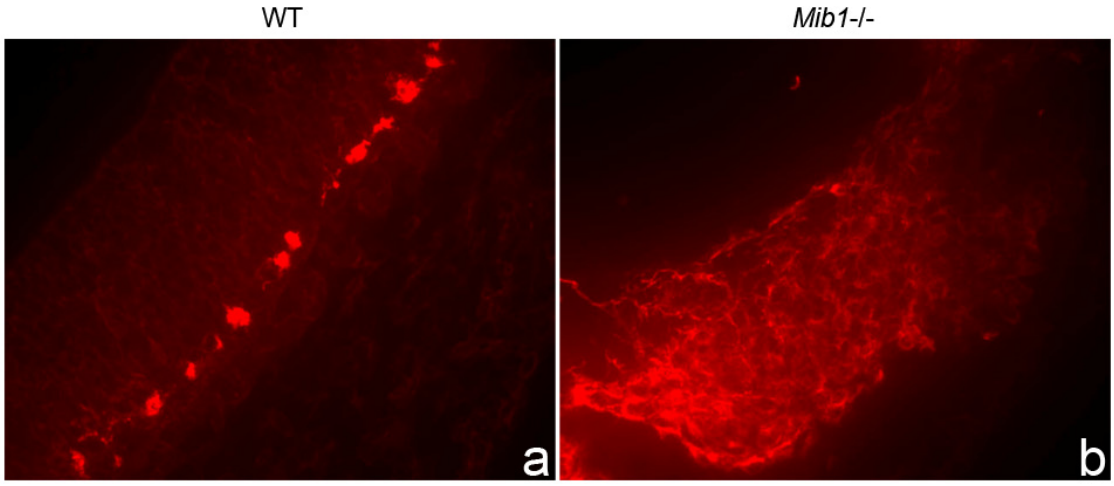


Figure 14. E9.5 *Tubb3* immunoreactivity. (a) WT littermate control neuroepithelium depicting the first wave of neuronal differentiation. (b) *Mib1*$^{-/-}$ neuroepithelium prematurely differentiates *en masse*. All sections 7 µm.

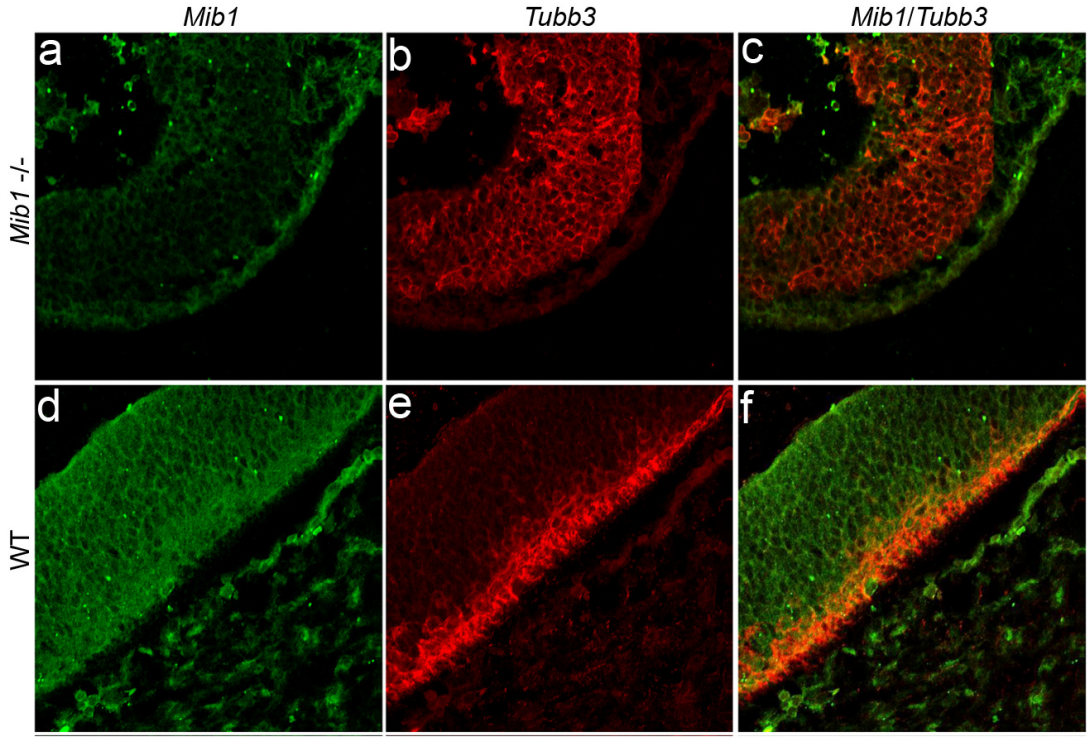


Figure 15. *Mib1 in vivo* expression in neurogenesis. (a–c) *Mib1*$^{-/-}$. (d–f) WT. (a, d) *Mib1* immunoreactivity. (b, e) *Tubb3* immunoreactivity. (c, f) *Mib1* and *Tubb3* co-localization. All sections 7 µm. All embryos analyzed at E9.5. Images obtained through confocal microscopy.

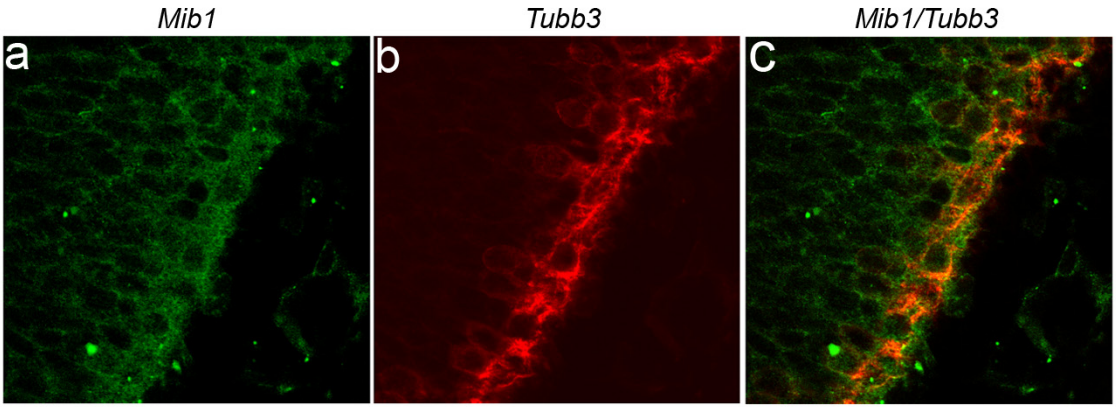


Figure 16. High magnification of nascent WT neurons. (a) *Mib1* immunoreactivity (b) *Tubb3* immunoreactivity. (c) *Mib1* and *Tubb3* co-localization. All sections 7 μm. All embryos analyzed at E9.5. Images obtained through confocal microscopy.

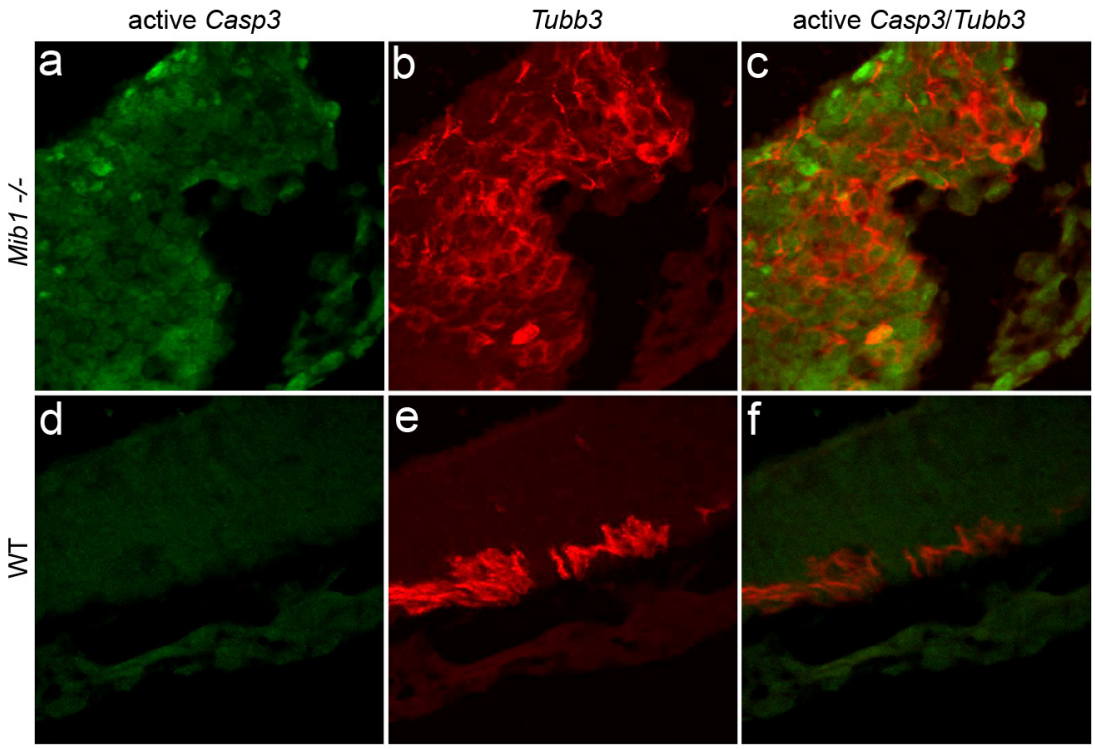


Figure 17. Premature neurons in *Mib1*$^{-/-}$ undergo apoptosis soon after differentiation. (a, d) Active *Casp3* immunoreactivity. (b, e) *Tubb3* immunoreactivity. (c, f) *Casp3* and *Tubb3* co-localization. All sections 7 µm. All embryos analyzed at E9.5. Images obtained through confocal microscopy.

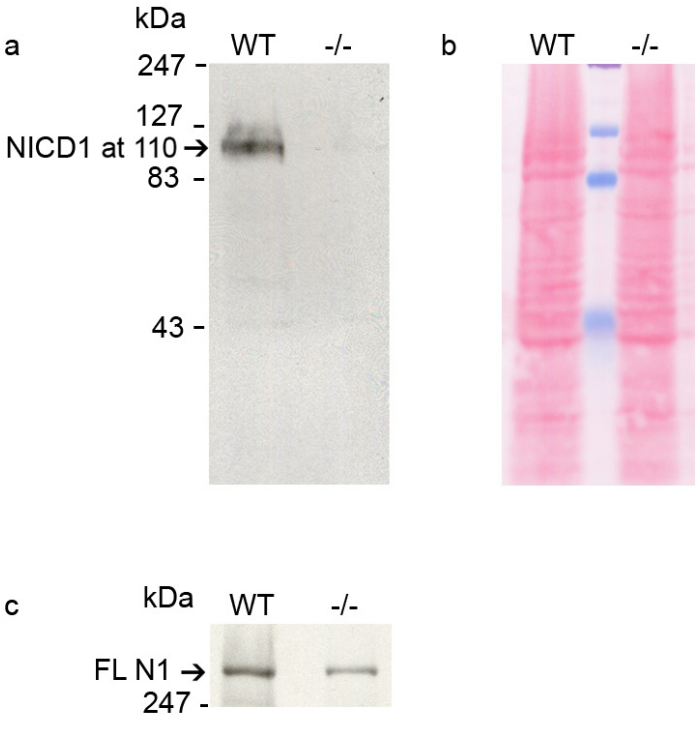


Figure 18. NICD1 immunoblot. (a) Left lane corresponds to WT, right lane to *Mib1* homozygous null. NICD1 migrates as a 110 kDa protein. (b) The same blot as 'a' stained with *Ponceau S* so as to reveal total protein loaded. (c) *Notch1* immunoblot. Full length (FL) *Notch1* (N1) detected in both WT and *Mib1* homozygous null embryos. Single E9.5 embryo lysate loaded *per* lane.

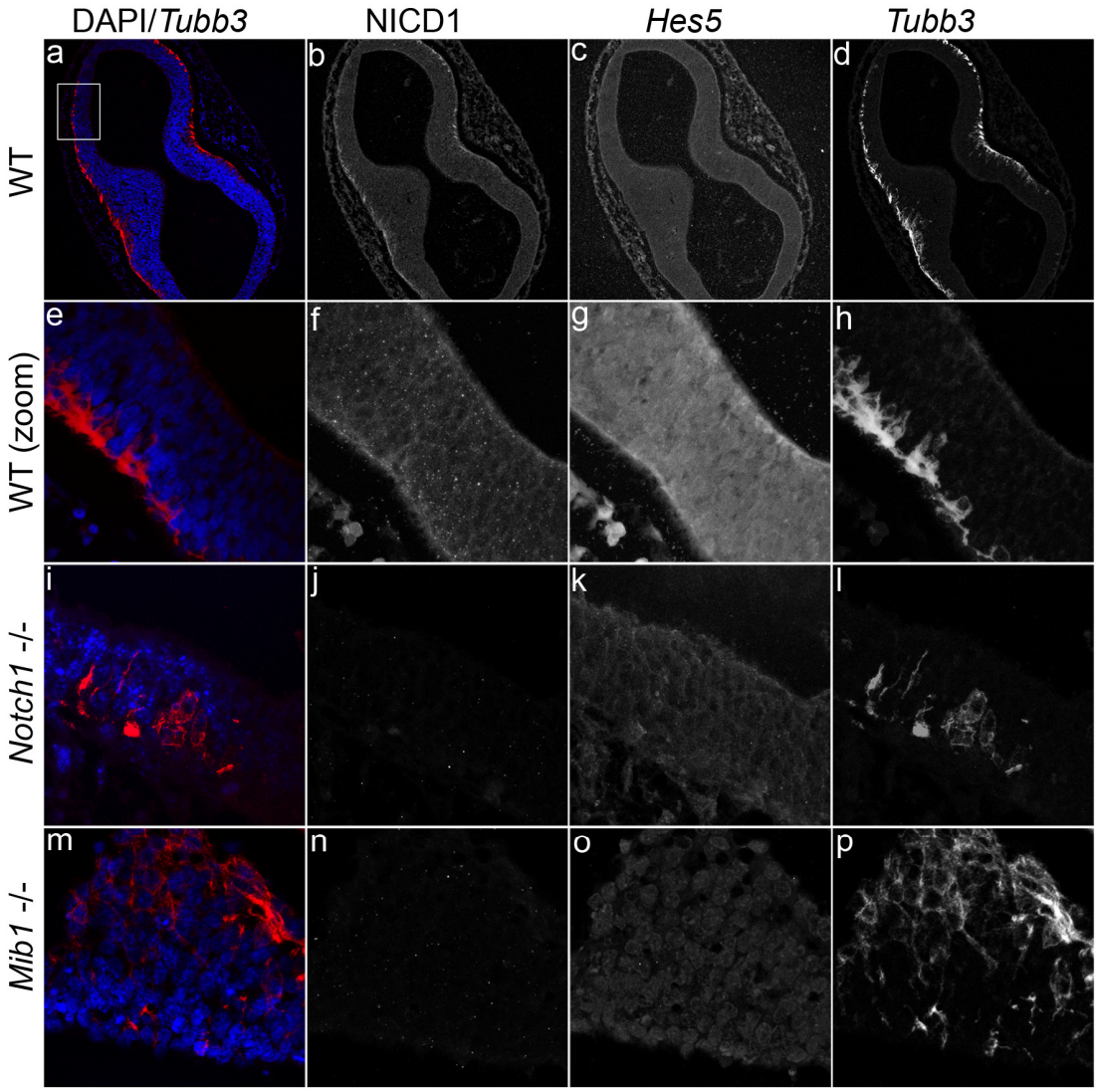


Figure 19. *Notch* signaling within the embryonic neuroepithelium. (a–d) Entire field of view of a transverse histological section through embryonic brain. (e-p) High magnification of embryonic neuroepithelium which approximately corresponds to an area similar to that outlined in 'a'. (a-h) WT. (i-l) *Notch1*$^{-/-}$. (m-p) *Mib1*$^{-/-}$. (a, e, i, m) 4',6-diamidino-2-phenylindole (DAPI) and *Tubb3* co-localization. (b, f, j, n) NICD1 immunoreactivity. (c, g, k, o) *Hes5* immunoreactivity. (d, h, l, p) *Tubb3* immunoreactivity. All sections 7 µm. All embryos analyzed at E9.5 with the exception of *Notch1*$^{-/-}$ being analyzed at E10.5 due to availability. Images obtained through confocal microscopy.

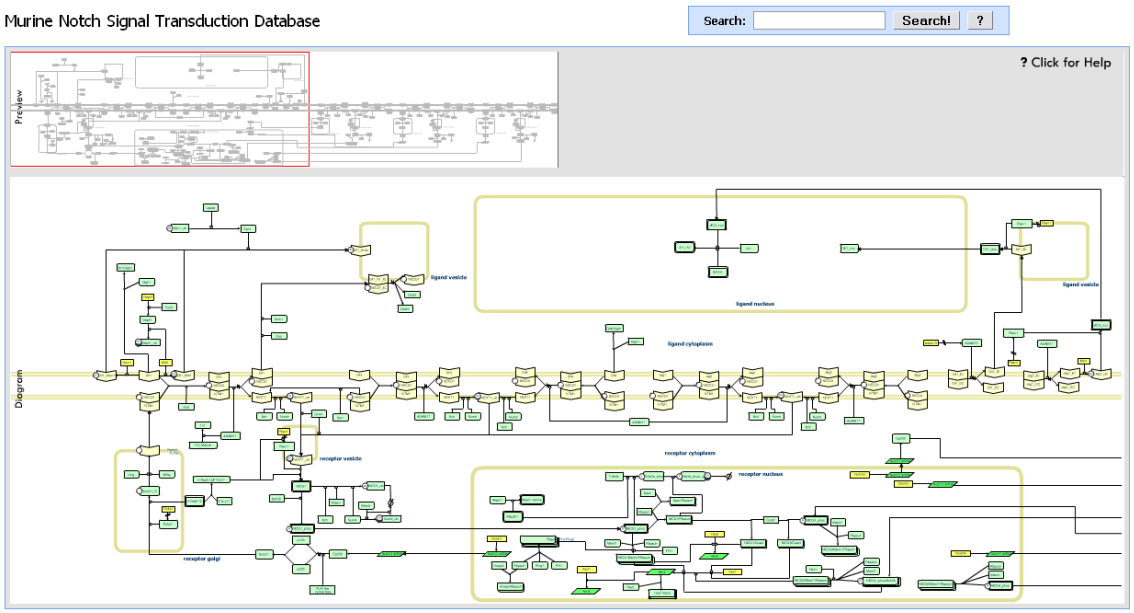


Figure 20. Screen capture. Murine Notch Signal Transduction Database portal. Macromedia Flash Player operated user interface.

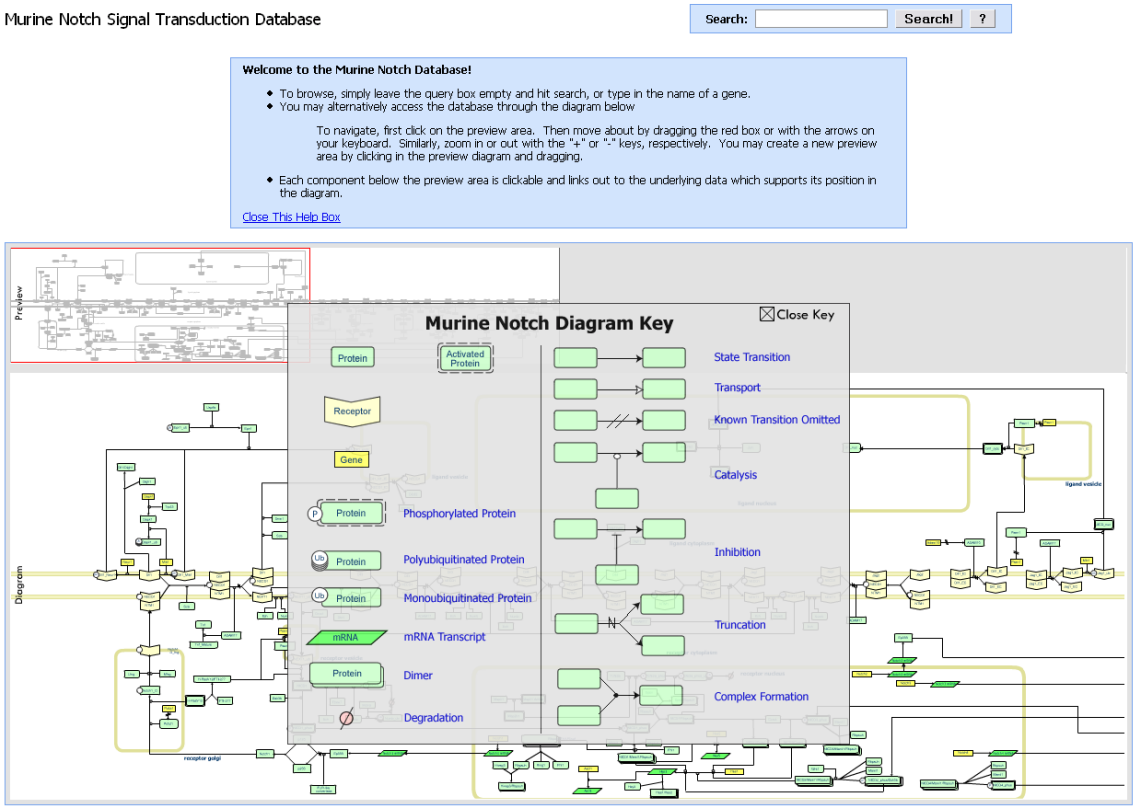


Figure 21. Screen capture. Murine Notch Signal Transduction Database as viewed with both help toggles activated.

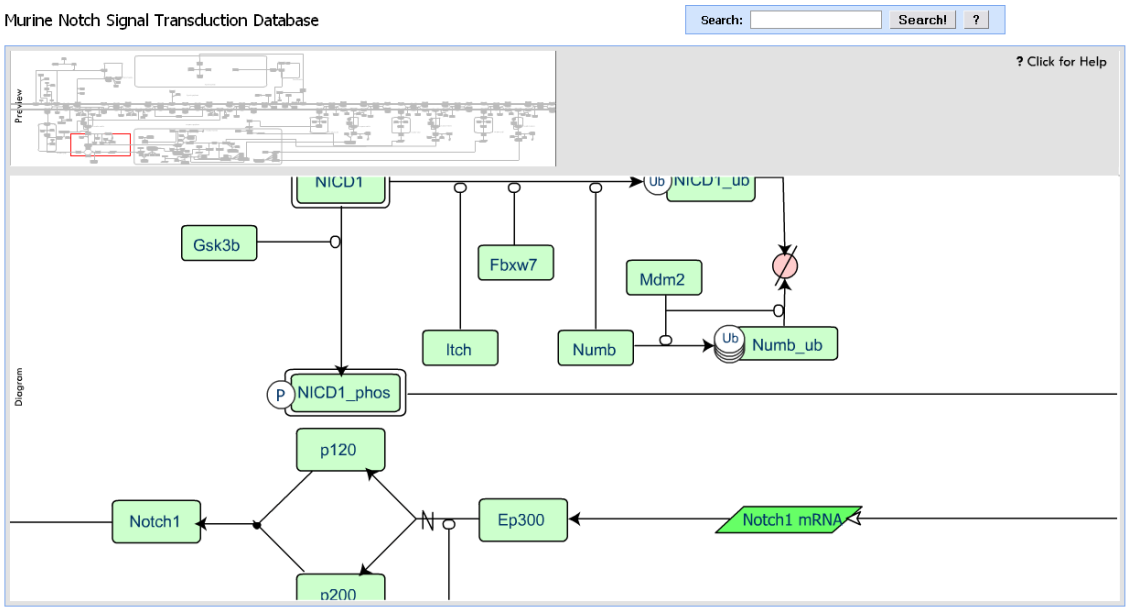


Figure 22. Screen capture. Murine Notch Signal Transduction Database portal. The Macromedia Flash Player operated interface is equipped with a tool that enables the user to zoom in on any region of the molecular interaction map. The area outlined in red within the 'preview' window is magnified below within the 'diagram' window.

Murine Notch Signal Transduction Database

Search: Search!

Component Info - Notch1

There are 2 connections associated with Notch1

Component Details

Name:	Notch1
Synonyms:	Notch-1; Notch 1
Component Type:	Protein generic
Dimer Number:	1
Compartment:	receptor cytoplasm
Molecular Weight (kDa):	270.83414
Mouse Location:	Chromosome 2: 26.4 Mb (26418058 - 26463820) (More Info)
Amino Acids:	2444 aa
Motifs/Domains:	- a signal peptide - thirty six extracellular EGF repeats - three lin/notch repeats - HDn and HDc domain - five cdc10/ankyrin repeats - a RAM23 domain - a transcriptional activation domain - C terminal PEST domain

Linked Components

Directly Connected:	1 Level Indirect:
Notch1	Ep300
Notch1_O	FUR-like convertase
p120	Lfng
p200	Mfng
Pofut1	Notch1
	Notch1_O
	Notch1_O_fng
	p120
	p200
	Pofut1
	Pofut1

Change Indirect Levels: 1

Associated Interactions & Components

Interaction 1

Interaction Type: Heterodimer formation

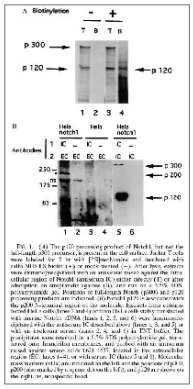
Component Name	Component Type	Compartment	Connection
p120	Protein generic	receptor cytoplasm	From
p200	Protein generic	receptor cytoplasm	From
Notch1	Protein generic	receptor cytoplasm	To

Experiments Associated With This Interaction

Experiment Details

Related Text: Notch1 present on surface as a heterodimeric molecule (p120/p200)whereas the p300 molecule does not reach the surface.

Image



Paper Details

Author:	Logeat
Date Published:	1998
Title:	The Notch1 receptor is cleaved constitutively by a furin-like convertase
Journal:	Developmental Biology
PubMedID:	9653148 (View PDF)
Abstract:	The Notch receptor, which is involved in numerous cell fate decisions in invertebrates and vertebrates, is synthesized as a 300-kDa precursor molecule (p300). We show here that proteolytic processing of p300 is an essential step in the formation of the biologically active receptor because only the cleaved fragments are present at the cell surface. Our results confirm and extend recent reports indicating that the Notch receptor exists at the plasma membrane as a heterodimeric molecule, but disagree as to the nature of the protease that is responsible for the cleavage that takes place in the extracellular region. We report here that constitutive processing of murine Notch1 involves a furin-like convertase. We show that the calcium ionophore A23187 and the alpha1-antitrypsin variant, alpha 1-PDX, a known inhibitor of furin-like convertases, inhibit p300 processing. When expressed in the furin-deficient Lovo cell line, p300 is not processed. In vitro digestion of a recombinant Notch-derived substrate with purified furin allowed mapping of the processing site to the carboxyl side of the sequence RQRR (amino acids 1651-1654). Mutation of these four amino acids (and of two secondary dibasic furin sites located nearby) completely abolished processing of the Notch1 receptor.

Figure 23. Screen capture. Murine Notch Signal Transduction Database component information. For illustrative purposes, let us assume we care to probe the 'Notch1' component shown in Figure 22. All general information pertaining to this component, as well as the data justifying its position within the map is displayed as shown. For clarity, the content has been cropped so as to only reveal the first experiment annotated. On line however, additional content may be accessed by scrolling down the web page.

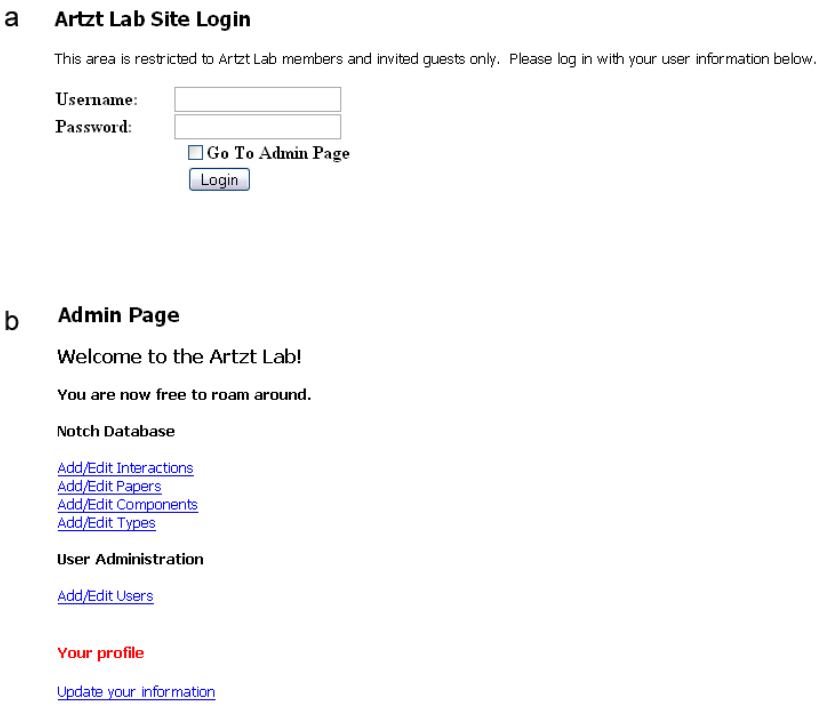


Figure 24. Screen capture. Murine Notch Signal Transduction Database. (a) Administrator secure login. (b) Administrator portal.

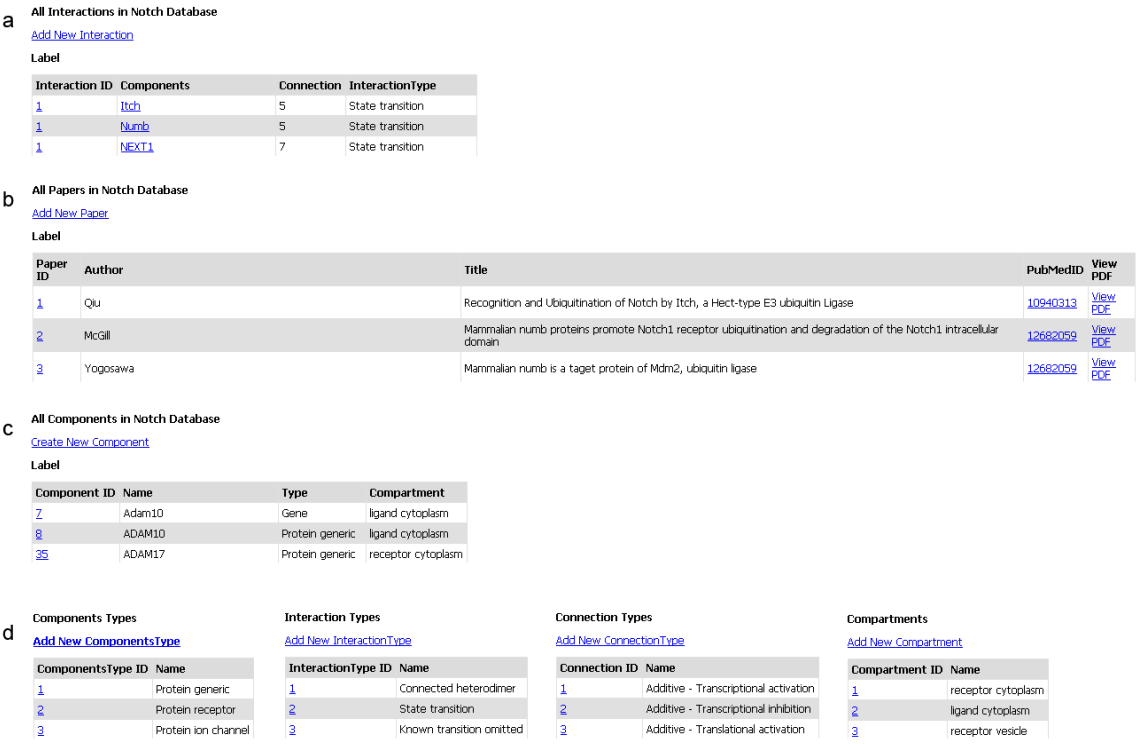


Figure 25. Screen capture. Murine Notch Signal Transduction Database. The following database properties may be edited. (a) Model interactions. (b) Papers. (c) Model components. (d) Model types. Each of the above items is displayed as an individual web page. For clarity, the content of all items has been cropped so as to only reveal the first three entries. On line however, all entries are visible and may be edited.

References

1. Gerhart, J. 1998 Warkany lecture: signaling pathways in development. *Teratology* **60**, 226-39 (1999).
2. Dehal, P. & Boore, J. L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* **3**, e314 (2005).
3. van Nimwegen, E. Scaling laws in the functional content of genomes. *Trends Genet* **19**, 479-84 (2003).
4. Kitano, H. Biological robustness. *Nat Rev Genet* **5**, 826-37 (2004).
5. Louvi, A. & Artavanis-Tsakonas, S. Notch signalling in vertebrate neural development. *Nat Rev Neurosci* **7**, 93-102 (2006).
6. Kitano, H., Funahashi, A., Matsuoka, Y. & Oda, K. Using process diagrams for the graphical representation of biological networks. *Nat Biotechnol* **23**, 961-6 (2005).
7. Kohn, K. W., Aladjem, M. I., Weinstein, J. N. & Pommier, Y. Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. *Molecular biology of the cell* **17**, 1-13 (2006).
8. Longabaugh, W. J., Davidson, E. H. & Bolouri, H. Computational representation of developmental genetic regulatory networks. *Developmental biology* **283**, 1-16 (2005).
9. Kurata, H., Masaki, K., Sumida, Y. & Iwasaki, R. CADLIVE dynamic simulator: direct link of biochemical networks to dynamic models. *Genome Res* **15**, 590-600 (2005).
10. Kanehisa, M. et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research* **34**, D354-7 (2006).
11. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101-13 (2004).
12. Wuchty, S., Barabasi, A. L. & Ferdig, M. T. Stable evolutionary signal in a Yeast protein interaction network. *BMC Evol Biol* **6**, 8 (2006).
13. Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796-800 (2006).
14. Logeat, F. et al. The Notch1 receptor is cleaved constitutively by a furin-like convertase. *Proc Natl Acad Sci U S A* **95**, 8108-12 (1998).
15. Blaumueller, C. M., Qi, H., Zagouras, P. & Artavanis-Tsakonas, S. Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane. *Cell* **90**, 281-91 (1997).
16. Wang, Y. et al. Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase. *J Biol Chem* **276**, 40338-45 (2001).

17. Okajima, T., Xu, A. & Irvine, K. D. Modulation of notch-ligand binding by protein O-fucosyltransferase 1 and fringe. *J Biol Chem* **278**, 42340-5 (2003).
18. Panin, V. M. et al. Notch ligands are substrates for protein O-fucosyltransferase-1 and Fringe. *J Biol Chem* **277**, 29945-52 (2002).
19. Blair, S. S. Notch signaling: Fringe really is a glycosyltransferase. *Curr Biol* **10**, R608-12 (2000).
20. Bruckner, K., Perez, L., Clausen, H. & Cohen, S. Glycosyltransferase activity of Fringe modulates Notch-Delta interactions. *Nature* **406**, 411-5 (2000).
21. Moloney, D. J. et al. Fringe is a glycosyltransferase that modifies Notch. *Nature* **406**, 369-75 (2000).
22. Munro, S. & Freeman, M. The notch signalling regulator fringe acts in the Golgi apparatus and requires the glycosyltransferase signature motif DXD. *Curr Biol* **10**, 813-20 (2000).
23. Bayer, P. & Fanghanel, J. Fringe gives a saccharine to notch. *Trends Biochem Sci* **25**, 485-6 (2000).
24. Schroeter, E. H., Kisslinger, J. A. & Kopan, R. Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain. *Nature* **393**, 382-6 (1998).
25. Bettenhausen, B., Hrabe de Angelis, M., Simon, D., Guenet, J. L. & Gossler, A. Transient and restricted expression during mouse embryogenesis of Dll1, a murine gene closely related to Drosophila Delta. *Development* **121**, 2407-18 (1995).
26. Kiernan, A. E., Cordes, R., Kopan, R., Gossler, A. & Gridley, T. The Notch ligands DLL1 and JAG2 act synergistically to regulate hair cell development in the mammalian inner ear. *Development* **132**, 4353-62 (2005).
27. Dunwoodie, S. L., Henrique, D., Harrison, S. M. & Beddington, R. S. Mouse Dll3: a novel divergent Delta gene which may complement the function of other Delta homologues during early pattern formation in the mouse embryo. *Development* **124**, 3065-76 (1997).
28. Shutter, J. R. et al. Dll4, a novel Notch ligand expressed in arterial endothelium. *Genes Dev* **14**, 1313-8 (2000).
29. Varnum-Finney, B. et al. The Notch ligand, Jagged-1, influences the development of primitive hematopoietic precursor cells. *Blood* **91**, 4084-91 (1998).
30. Eiraku, M. et al. DNER acts as a neuron-specific Notch ligand during Bergmann glial development. *Nat Neurosci* **8**, 873-80 (2005).
31. Hu, Q. D. et al. F3/contactin acts as a functional ligand for Notch during oligodendrocyte maturation. *Cell* **115**, 163-75 (2003).
32. Miyamoto, A., Lau, R., Hein, P. W., Shipley, J. M. & Weinmaster, G. Microfibrillar proteins MAGP-1 and MAGP-2 induce notch1 extracellular domain dissociation and receptor activation. *J Biol Chem* (2006).

33. Le Roy, C. & Wrana, J. L. Clathrin- and non-clathrin-mediated endocytic regulation of cell signalling. *Nat Rev Mol Cell Biol* **6**, 112-26 (2005).
34. Praefcke, G. J. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat Rev Mol Cell Biol* **5**, 133-47 (2004).
35. Takei, K., Yoshida, Y. & Yamada, H. Regulatory mechanisms of dynamin-dependent endocytosis. *J Biochem (Tokyo)* **137**, 243-7 (2005).
36. Seugnet, L., Simpson, P. & Haenlin, M. Requirement for dynamin during Notch signaling in Drosophila neurogenesis. *Dev Biol* **192**, 585-98 (1997).
37. Barriere, H. et al. Molecular basis of oligoubiquitin-dependent internalization of membrane proteins in Mammalian cells. *Traffic* **7**, 282-97 (2006).
38. Hawryluk, M. J. et al. Epsin 1 is a polyubiquitin-selective clathrin-associated sorting protein. *Traffic* **7**, 262-81 (2006).
39. Wang, W. & Struhl, G. Distinct roles for Mind bomb, Neuralized and Epsin in mediating DSL endocytosis and signaling in Drosophila. *Development* **132**, 2883-94 (2005).
40. Wang, W. & Struhl, G. Drosophila Epsin mediates a select endocytic pathway that DSL ligands must enter to activate Notch. *Development* **131**, 5367-80 (2004).
41. Tian, X., Hansen, D., Schedl, T. & Skeath, J. B. Epsin potentiates Notch pathway activity in Drosophila and C. elegans. *Development* **131**, 5807-15 (2004).
42. Overstreet, E., Fitch, E. & Fischer, J. A. Fat facets and Liquid facets promote Delta endocytosis and Delta signaling in the signaling cells. *Development* **131**, 5355-66 (2004).
43. Kolev, V. et al. The intracellular domain of Notch ligand Delta1 induces cell growth arrest. *FEBS Lett* **579**, 5798-5802 (2005).
44. Bland, C. E., Kimberly, P. & Rand, M. D. Notch-induced proteolysis and nuclear localization of the Delta ligand. *J Biol Chem* **278**, 13607-10 (2003).
45. Six, E. et al. The Notch ligand Delta1 is sequentially cleaved by an ADAM protease and gamma-secretase. *Proc Natl Acad Sci U S A* **100**, 7638-43 (2003).
46. Hartmann, D. et al. The disintegrin/metalloprotease ADAM 10 is essential for Notch signalling but not for alpha-secretase activity in fibroblasts. *Hum Mol Genet* **11**, 2615-24 (2002).
47. De Strooper, B. et al. A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. *Nature* **398**, 518-22 (1999).
48. Kopan, R. & Goate, A. Aph-2/Nicastrin: an essential component of gamma-secretase and regulator of Notch signaling and Presenilin localization. *Neuron* **33**, 321-4 (2002).
49. Gupta-Rossi, N. et al. Monoubiquitination and endocytosis direct gamma-secretase cleavage of activated Notch receptor. *J Cell Biol* **166**, 73-83 (2004).

50. Chandu, D., Huppert, S. S. & Kopan, R. Analysis of transmembrane domain mutants is consistent with sequential cleavage of Notch by gamma-secretase. *J Neurochem* **96**, 228-35 (2006).
51. Saxena, M. T., Schroeter, E. H., Mumm, J. S. & Kopan, R. Murine notch homologs (N1-4) undergo presenilin-dependent proteolysis. *J Biol Chem* **276**, 40268-73 (2001).
52. Hochstrasser, M. Lingering mysteries of ubiquitin-chain assembly. *Cell* **124**, 27-34 (2006).
53. Rodriguez, M. S., Desterro, J. M., Lain, S., Lane, D. P. & Hay, R. T. Multiple C-terminal lysine residues target p53 for ubiquitin-proteasome-mediated degradation. *Mol Cell Biol* **20**, 8458-67 (2000).
54. Haglund, K. et al. Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nat Cell Biol* **5**, 461-6 (2003).
55. Le Borgne, R. Regulation of Notch signalling by endocytosis and endosomal sorting. *Curr Opin Cell Biol* **18**, 213-22 (2006).
56. Itoh, M. et al. Mind bomb is a ubiquitin ligase that is essential for efficient activation of Notch signaling by Delta. *Dev Cell* **4**, 67-82 (2003).
57. Lai, E. C., Deblandre, G. A., Kintner, C. & Rubin, G. M. Drosophila neuralized is a ubiquitin ligase that promotes the internalization and degradation of delta. *Dev Cell* **1**, 783-94 (2001).
58. Pavlopoulos, E. et al. neuralized Encodes a peripheral membrane protein involved in delta signaling and endocytosis. *Dev Cell* **1**, 807-16 (2001).
59. Chen, W. & Casey Corliss, D. Three modules of zebrafish Mind bomb work cooperatively to promote Delta ubiquitination and endocytosis. *Dev Biol* **267**, 361-73 (2004).
60. Koo, B. K. et al. Mind bomb-2 is an E3 ligase for Notch ligand. *J Biol Chem* **280**, 22335-42 (2005).
61. Lai, E. C., Roegiers, F., Qin, X., Jan, Y. N. & Rubin, G. M. The ubiquitin ligase Drosophila Mind bomb promotes Notch signaling by regulating the localization and activity of Serrate and Delta. *Development* **132**, 2319-32 (2005).
62. Koo, B. K. et al. Mind bomb 1 is essential for generating functional Notch ligands to activate Notch. *Development* **132**, 3459-70 (2005).
63. Barsi, J. C., Rajendra, R., Wu, J. I. & Artzt, K. Mind bomb1 is a ubiquitin ligase essential for mouse embryonic development and Notch signaling. *Mech Dev* **122**, 1106-17 (2005).
64. Nastasi, T. et al. Ozz-E3, a muscle-specific ubiquitin ligase, regulates beta-catenin degradation during myogenesis. *Dev Cell* **6**, 269-82 (2004).
65. Ruan, Y., Tecott, L., Jiang, M. M., Jan, L. Y. & Jan, Y. N. Ethanol hypersensitivity and olfactory discrimination defect in mice lacking a homolog of Drosophila neuralized. *Proc Natl Acad Sci U S A* **98**, 9907-12 (2001).

66. Vollrath, B., Pudney, J., Asa, S., Leder, P. & Fitzgerald, K. Isolation of a murine homologue of the Drosophila neuralized gene, a gene required for axonemal integrity in spermatozoa and terminal maturation of the mammary gland. *Mol Cell Biol* **21**, 7481-94 (2001).
67. Takeuchi, T., Adachi, Y. & Ohtsuki, Y. Skeletrophin, a novel ubiquitin ligase to the intracellular region of Jagged-2, is aberrantly expressed in multiple myeloma. *Am J Pathol* **166**, 1817-26 (2005).
68. Takeuchi, T., Adachi, Y. & Ohtsuki, Y. Skeletrophin, a novel RING molecule controlled by the chromatin remodeling complex, is downregulated in malignant melanoma. *DNA Cell Biol* **24**, 339-44 (2005).
69. Takeuchi, T. et al. Down-regulation of a novel actin-binding molecule, skeletrophin, in malignant melanoma. *Am J Pathol* **163**, 1395-404 (2003).
70. Jehn, B. M., Dittert, I., Beyer, S., von der Mark, K. & Bielke, W. c-Cbl binding and ubiquitin-dependent lysosomal degradation of membrane-associated Notch1. *J Biol Chem* **277**, 8033-40 (2002).
71. Hori, K., Fuwa, T. J., Seki, T. & Matsuno, K. Genetic regions that interact with loss- and gain-of-function phenotypes of deltex implicate novel genes in Drosophila Notch signaling. *Mol Genet Genomics* **272**, 627-38 (2005).
72. Kishi, N. et al. Murine homologs of deltex define a novel gene family involved in vertebrate Notch signaling and neurogenesis. *Int J Dev Neurosci* **19**, 21-35 (2001).
73. Petersen, P. H., Tang, H., Zou, K. & Zhong, W. The enigma of the numb-Notch relationship during mammalian embryogenesis. *Dev Neurosci* **28**, 156-68 (2006).
74. McGill, M. A. & McGlade, C. J. Mammalian numb proteins promote Notch1 receptor ubiquitination and degradation of the Notch1 intracellular domain. *J Biol Chem* **278**, 23196-203 (2003).
75. Wilkin, M. B. et al. Regulation of notch endosomal sorting and signaling by Drosophila Nedd4 family proteins. *Curr Biol* **14**, 2237-44 (2004).
76. Sakata, T. et al. Drosophila Nedd4 regulates endocytosis of notch and suppresses its ligand-independent activation. *Curr Biol* **14**, 2228-36 (2004).
77. Qiu, L. et al. Recognition and ubiquitination of Notch by Itch, a hect-type E3 ubiquitin ligase. *J Biol Chem* **275**, 35734-7 (2000).
78. Wu, G. et al. SEL-10 is an inhibitor of notch signaling that targets notch for ubiquitin-mediated protein degradation. *Mol Cell Biol* **21**, 7403-15 (2001).
79. Oberg, C. et al. The Notch intracellular domain is ubiquitinated and negatively regulated by the mammalian Sel-10 homolog. *J Biol Chem* **276**, 35847-53 (2001).
80. Tsunematsu, R. et al. Mouse Fbw7/Sel-10/Cdc4 is required for notch degradation during vascular development. *J Biol Chem* **279**, 9417-23 (2004).

81. Maruyama, S., Hatakeyama, S., Nakayama, K., Ishida, N. & Kawakami, K. Characterization of a mouse gene (Fbxw6) that encodes a homologue of Caenorhabditis elegans SEL-10. *Genomics* **78**, 214-22 (2001).
82. Foltz, D. R. & Nye, J. S. Hyperphosphorylation and association with RBP of the intracellular domain of Notch1. *Biochem Biophys Res Commun* **286**, 484-92 (2001).
83. Foltz, D. R., Santiago, M. C., Berechid, B. E. & Nye, J. S. Glycogen synthase kinase-3beta modulates notch signaling and stability. *Curr Biol* **12**, 1006-11 (2002).
84. Jin, Y., Blue, E. K., Dixon, S., Shao, Z. & Gallagher, P. J. A death-associated protein kinase (DAPK)-interacting protein, DIP-1, is an E3 ubiquitin ligase that promotes tumor necrosis factor-induced apoptosis and regulates the cellular levels of DAPK. *J Biol Chem* **277**, 46980-6 (2002).
85. Tetzlaff, M. T. et al. Defective cardiovascular development and elevated cyclin E and Notch proteins in mice lacking the Fbw7 F-box protein. *Proc Natl Acad Sci U S A* **101**, 3338-45 (2004).
86. Ye, X. et al. Recognition of phosphodegron motifs in human cyclin E by the SCF(Fbw7) ubiquitin ligase. *J Biol Chem* **279**, 50110-9 (2004).
87. Kwak, E. L. et al. Infrequent mutations of Archipelago (hAGO, hCDC4, Fbw7) in primary ovarian cancer. *Gynecol Oncol* **98**, 124-8 (2005).
88. Mao, J. H. et al. Fbxw7/Cdc4 is a p53-dependent, haploinsufficient tumour suppressor gene. *Nature* **432**, 775-9 (2004).
89. Li, J. et al. SEL-10 interacts with presenilin 1, facilitates its ubiquitination, and alters A-beta peptide production. *J Neurochem* **82**, 1540-8 (2002).
90. Magnifico, A. et al. WW domain HECT E3s target Cbl RING finger E3s for proteasomal degradation. *J Biol Chem* **278**, 43169-77 (2003).
91. Nie, J. et al. LNX functions as a RING type E3 ubiquitin ligase that targets the cell fate determinant Numb for ubiquitin-dependent degradation. *Embo J* **21**, 93-102 (2002).
92. Rice, D. S., Northcutt, G. M. & Kurschner, C. The Lnx family proteins function as molecular scaffolds for Numb family proteins. *Mol Cell Neurosci* **18**, 525-40 (2001).
93. Chen, J. et al. Characterization of human LNX, a novel ligand of Numb protein X that is downregulated in human gliomas. *Int J Biochem Cell Biol* **37**, 2273-83 (2005).
94. Yogosawa, S., Miyauchi, Y., Honda, R., Tanaka, H. & Yasuda, H. Mammalian Numb is a target protein of Mdm2, ubiquitin ligase. *Biochem Biophys Res Commun* **302**, 869-72 (2003).
95. Juven-Gershon, T. et al. The Mdm2 oncoprotein interacts with the cell fate regulator Numb. *Mol Cell Biol* **18**, 3974-82 (1998).
96. Salcini, A. E. et al. Binding specificity and in vivo targets of the EH domain, a novel protein-protein interaction module. *Genes Dev* **11**, 2239-49 (1997).

97. Naslavsky, N. & Caplan, S. C-terminal EH-domain-containing proteins: consensus for a role in endocytic trafficking, EH? *J Cell Sci* **118**, 4093-101 (2005).
98. Susini, L. et al. Siah-1 binds and regulates the function of Numb. *Proc Natl Acad Sci U S A* **98**, 15067-72 (2001).
99. Nijman, S. M. et al. A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123**, 773-86 (2005).
100. Chen, X., Overstreet, E., Wood, S. A. & Fischer, J. A. On the conservation of function of the Drosophila fat facets deubiquitinating enzyme and Fam, its mouse homolog. *Dev Genes Evol* **210**, 603-10 (2000).
101. Chen, X., Zhang, B. & Fischer, J. A. A specific protein substrate for a deubiquitinating enzyme: Liquid facets is the substrate of Fat facets. *Genes Dev* **16**, 289-94 (2002).
102. Wilson, J. J. & Kovall, R. A. Crystal structure of the CSL-Notch-Mastermind ternary complex bound to DNA. *Cell* **124**, 985-96 (2006).
103. Nam, Y., Sliz, P., Song, L., Aster, J. C. & Blacklow, S. C. Structural basis for cooperativity in recruitment of MAML coactivators to Notch transcription complexes. *Cell* **124**, 973-83 (2006).
104. Iso, T., Kedes, L. & Hamamori, Y. HES and HERP families: multiple effectors of the Notch signaling pathway. *J Cell Physiol* **194**, 237-55 (2003).
105. Allen, R. D., 3rd, Kim, H. K., Sarafova, S. D. & Siu, G. Negative regulation of CD4 gene expression by a HES-1-c-Myb complex. *Mol Cell Biol* **21**, 3071-82 (2001).
106. Kim, H. K. & Siu, G. The notch pathway intermediate HES-1 silences CD4 gene expression. *Mol Cell Biol* **18**, 7166-75 (1998).
107. Castella, P., Wagner, J. A. & Caudy, M. Regulation of hippocampal neuronal differentiation by the basic helix-loop-helix transcription factors HES-1 and MASH-1. *J Neurosci Res* **56**, 229-40 (1999).
108. Elagib, K. E. et al. Jun blockade of erythropoiesis: role for repression of GATA-1 by HERP2. *Mol Cell Biol* **24**, 7779-94 (2004).
109. Parks, A. L., Klueg, K. M., Stout, J. R. & Muskavitch, M. A. Ligand endocytosis drives receptor dissociation and activation in the Notch pathway. *Development* **127**, 1373-85 (2000).
110. Chyung, J. H., Raper, D. M. & Selkoe, D. J. Gamma-secretase exists on the plasma membrane as an intact complex that accepts substrates and effects intramembrane cleavage. *J Biol Chem* **280**, 4383-92 (2005).
111. Tarassishin, L., Yin, Y. I., Bassit, B. & Li, Y. M. Processing of Notch and amyloid precursor protein by gamma-secretase is spatially distinct. *Proc Natl Acad Sci U S A* **101**, 17050-5 (2004).

112. Kaether, C., Schmitt, S., Willem, M. & Haass, C. Amyloid Precursor Protein and Notch Intracellular Domains are Generated after Transport of their Precursors to the Cell Surface. *Traffic* **7**, 408-15 (2006).
113. LaVoie, M. J. & Selkoe, D. J. The Notch ligands, Jagged and Delta, are sequentially processed by alpha-secretase and presenilin/gamma-secretase and release signaling fragments. *J Biol Chem* **278**, 34427-37 (2003).
114. Lawler, A. University-industry collaboration. Last of the big-time spenders? *Science* **299**, 330-3 (2003).
115. Golling, G. et al. Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat Genet* **31**, 135-40 (2002).
116. Jiang, Y. J. et al. Mutations affecting neurogenesis and brain morphology in the zebrafish, Danio rerio. *Development* **123**, 205-16 (1996).
117. van Eeden, F. J. et al. Mutations affecting somite formation and patterning in the zebrafish, Danio rerio. *Development* **123**, 153-64 (1996).
118. Schier, A. F. et al. Mutations affecting the development of the embryonic zebrafish brain. *Development* **123**, 165-78 (1996).
119. Chen, J. F. et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature genetics* **38**, 228-33 (2006).
120. Mansfield, J. H. et al. MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nature genetics* **36**, 1079-83 (2004).
121. Kloosterman, W. P., Wienholds, E., de Bruijn, E., Kauppinen, S. & Plasterk, R. H. In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nature methods* **3**, 27-9 (2006).
122. Fanarraga, M. L., Avila, J. & Zabala, J. C. Expression of unphosphorylated class III beta-tubulin isotype in neuroepithelial cells demonstrates neuroblast commitment and differentiation. *The European journal of neuroscience* **11**, 517-27 (1999).
123. Haubensak, W., Attardo, A., Denk, W. & Huttner, W. B. Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3196-201 (2004).
124. de la Pompa, J. L. et al. Conservation of the Notch signalling pathway in mammalian neurogenesis. *Development (Cambridge, England)* **124**, 1139-48 (1997).
125. Oka, C. et al. Disruption of the mouse RBP-J kappa gene results in early embryonic death. *Development (Cambridge, England)* **121**, 3291-301 (1995).
126. Hrabe de Angelis, M., McIntyre, J., 2nd & Gossler, A. Maintenance of somite borders in mice requires the Delta homologue DII1. *Nature* **386**, 717-21 (1997).

127. Xue, Y. et al. Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. *Human molecular genetics* **8**, 723-30 (1999).
128. Wu, L. et al. Cloning and functional characterization of the murine mastermind-like 1 (Maml1) gene. *Gene* **328**, 153-65 (2004).
129. Iso, T., Chung, G., Hamamori, Y. & Kedes, L. HERP1 is a cell type-specific primary target of Notch. *The Journal of biological chemistry* **277**, 6598-607 (2002).
130. Ben-Arie, N. et al. Math1 is essential for genesis of cerebellar granule neurons. *Nature* **390**, 169-72 (1997).
131. Guillemot, F. & Joyner, A. L. Dynamic expression of the murine Achaete-Scute homologue Mash-1 in the developing nervous system. *Mechanisms of development* **42**, 171-85 (1993).
132. Guillemot, F. et al. Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell* **75**, 463-76 (1993).
133. Tokunaga, A. et al. Mapping spatio-temporal activation of Notch signaling during neurogenesis and gliogenesis in the developing mouse brain. *Journal of neurochemistry* **90**, 142-54 (2004).
134. Bialik, S. & Kimchi, A. The Death-Associated Protein Kinases: Structure, Function, and Beyond. (2006).
135. Tada, Y. et al. The association of death-associated protein kinase hypermethylation with early recurrence in superficial bladder cancers. *Cancer research* **62**, 4048-53 (2002).
136. Le Borgne, R., Remaud, S., Hamel, S. & Schweisguth, F. Two distinct E3 ubiquitin ligases have complementary functions in the regulation of delta and serrate signaling in Drosophila. *PLoS biology* **3**, e96 (2005).
137. Hitoshi, S. et al. Notch pathway molecules are essential for the maintenance, but not the generation, of mammalian neural stem cells. *Genes & development* **16**, 846-58 (2002).
138. Sidow, A. et al. Serrate2 is disrupted in the mouse limb-development mutant syndactylism. *Nature* **389**, 722-5 (1997).
139. Mitsiadis, T. A., Henrique, D., Thesleff, I. & Lendahl, U. Mouse Serrate-1 (Jagged-1): expression in the developing tooth is regulated by epithelial-mesenchymal interactions and fibroblast growth factor-4. *Development (Cambridge, England)* **124**, 1473-83 (1997).
140. Haglund, K., Di Fiore, P. P. & Dikic, I. Distinct monoubiquitin signals in receptor endocytosis. *Trends in biochemical sciences* **28**, 598-603 (2003).
141. Hicke, L. & Dunn, R. Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. *Annual review of cell and developmental biology* **19**, 141-72 (2003).
142. Chitnis, A. Why is delta endocytosis required for effective activation of notch? *Developmental dynamics* **235**, 886-94 (2006).

143. Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L. & Bradley, A. Identification of mammalian microRNA host genes and transcription units. *Genome research* **14**, 1902-10 (2004).
144. Baskerville, S. & Bartel, D. P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA (New York, N.Y* **11**, 241-7 (2005).
145. Aboobaker, A. A., Tomancak, P., Patel, N., Rubin, G. M. & Lai, E. C. Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18017-22 (2005).
146. Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* **436**, 214-20 (2005).
147. Kwon, C., Han, Z., Olson, E. N. & Srivastava, D. MicroRNA1 influences cardiac differentiation in Drosophila and regulates Notch signaling. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18986-91 (2005).
148. Sokol, N. S. & Ambros, V. Mesodermally expressed Drosophila microRNA-1 is regulated by Twist and is required in muscles during larval growth. *Genes & development* **19**, 2343-54 (2005).
149. Enright, A. J. et al. MicroRNA targets in Drosophila. *Genome biology* **5**, R1 (2003).
150. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20 (2005).
151. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787-98 (2003).
152. John, B. et al. Human MicroRNA targets. *PLoS biology* **2**, e363 (2004).
153. Vatolin, S., Navaratne, K. & Weil, R. J. A Novel Method to Detect Functional MicroRNA Targets. (2006).
154. Elman, J. L. Finding Structure in Time. *Cognitive Science* **14**, 179-211 (1990).
155. Larkin, J. & Herbert, S. A. Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* **11**, 65-99 (1987).
156. Ogata, H., Goto, S., Fujibuchi, W. & Kanehisa, M. Computation with the KEGG pathway database. *Bio Systems* **47**, 119-28 (1998).
157. Kohn, K. W. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molecular biology of the cell* **10**, 2703-34 (1999).
158. Rosen, R. The representation of biological systems from standpoint of the theory of categories. *Bull. Math. Biophys.* **20**, 317–341 (1958).
159. Kacser, H. & Burns, J. A. The control of flux. *Symposia of the Society for Experimental Biology* **27**, 65-104 (1973).

160. Rosen, R. A relational theory of biological systems II. *Bull. Math. Biophys.* **21**, 109–128 (1959).
161. Letelier, J. C., Soto-Andrade, J., Guinez Abarzua, F., Cornish-Bowden, A. & Luz Cardenas, M. Organizational invariance and metabolic closure: analysis in terms of (M,R) systems. *Journal of theoretical biology* **238**, 949-61 (2006).
162. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology* (2005).
163. Gavin, A. C. et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631-6 (2006).
164. Krogan, N. J. et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440**, 637-43 (2006).
165. Finney, A. & Hucka, M. Systems biology markup language: Level 2 and beyond. *Biochemical Society transactions* **31**, 1472-3 (2003).
166. Hucka, M. et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics (Oxford, England)* **19**, 524-31 (2003).
167. Le Novere, N. et al. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature biotechnology* **23**, 1509-15 (2005).
168. Zadeh, L. A. Toward a generalized theory of uncertainty (GTU) –an outline. *Information Sciences* **172**, 1-40 (2005).
169. Newman, M., Barabási, A. & Watts, D. J. *The Structure and Dynamics of Networks* (eds. Levin, S. A. & Strogatz, S. H.) (Princeton University Press, 2006).
170. Watts, D. J. The "New" Science of Networks. *Annual Review of Sociology* **30**, 243-270 (2004).
171. Newman, M. The Structure and Function of Complex Networks. *SIAM Review* **45**, 167-256 (2003).
172. Oda, K. & Kitano, H. A comprehensive map of the toll-like receptor signaling network. *Molecular Systems Biology* (2006).
173. Kashtan, N., Itzkovitz, S., Milo, R. & Alon, U. Topological generalizations of network motifs. *Physical review* **70**, 031909 (2004).
174. Yeger-Lotem, E. et al. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 5934-9 (2004).
175. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics* **31**, 64-8 (2002).
176. Milo, R. et al. Network Motifs: Simple Building Blocks of Complex Networks. *Science* **298**, 824-827 (2002).

177. Christiansen, J. H. et al. EMAGE: a spatial database of gene expression patterns during mouse embryo development. *Nucleic acids research* **34**, D637-41 (2006).

178. Siddiqui, A. S. et al. A mouse atlas of gene expression: large-scale digital gene-expression profiles from precisely defined developing C57BL/6J mouse tissues and cells. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18485-90 (2005).

179. Su, A. I. et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6062-7 (2004).

180. Couzin, J. The brains behind blockbusters. *Science* **309**, 728 (2005).

Vita

Julius Christopher Barsi was born in Los Angeles, California on January 15th, 1977, son to Julius Bela Barsi and Guillermina Julia Barsi. After growing up in California, he moved to San Carlos de Bariloche, Río Negro, Argentina; an epic ski town in Patagonia. It was there that between the years of 1995 and 2000, he attended undergraduate studies at the *Universidad Nacional del Comahue, Centro Regional Universitario Bariloche*. Upon earning the degree of *Licenciado en Ciencias Biológicas*, he returned to the United States to pursue graduate studies in molecular biology. In 2001, he entered the Cell and Molecular Biology graduate program of the University of Texas at Austin. Three years after having joined the laboratory of Dr. Karen Artzt, he published his first paper on mammalian development (Barsi *et al.* 2005).

Permanent address: 3820 Duval Street, Austin, Texas 78751

This dissertation was typed by the author.