

# **Molecular Mechanisms of Teneurin Function in Transcriptional Regulation and Cell Adhesion**

**Inauguraldissertation**

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Jonas Schöler

aus Heidelberg, Deutschland

Basel, 2015

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät  
auf Antrag von:

Prof. Dr. Ruth Chiquet-Ehrismann

Richard P. Tucker, PhD

Basel, den 9.12.2014

Prof. Dr. Jörg Schibler  
Dekan

# Table of Contents

<b>1. Summary</b> .....	<b>1</b>
<b>2. Introduction</b> .....	<b>3</b>
2.1 ECM and cell-ECM adhesions.....	5
2.2 Cell-cell adhesions .....	15
2.3 Early neural development .....	24
2.4 Patterning of the brain.....	27
2.5 Axon guidance and synaptogenesis .....	31
2.6 Teneurin overview and structure.....	35
2.7 Teneurin expression and functions.....	39
2.8 Teneurins in disease .....	46
<b>3. Aim of the thesis</b> .....	<b>47</b>
<b>4. Publications</b> .....	<b>48</b>
4.1 Teneurins in evolution .....	48
4.2 Teneurins as transcriptional regulators .....	60
<b>5. Unpublished work</b> .....	<b>85</b>
5.1 Methods & Materials .....	85
5.1.1 Cloning and testing of the cten-1 and -2 NHL repeat domain constructs .....	85
5.1.2 Large-scale expression and purification of the cten-2 1262-1607 construct .....	88
5.1.3 Limited proteolysis .....	91
5.1.4 Recipes of buffers and media.....	93
5.2 Results.....	96
<b>6. Conclusion</b> .....	<b>104</b>
<b>7. References</b> .....	<b>108</b>
<b>8. Abbreviations</b> .....	<b>116</b>
<b>9. List of Figures</b> .....	<b>118</b>
<b>10. List of Tables</b> .....	<b>120</b>
<b>11. Acknowledgements</b> .....	<b>121</b>
<b>Appendix A - Microarray data of TEN1-ICD overexpression in BS149 cells</b> .....	<b>123</b>
<b>Appendix B - Microarray data of MITF overexpression in BS149 cells</b> .....	<b>133</b>
<b>Appendix C - General bacterial expression and purification protocols by PSF</b> .....	<b>145</b>
<b>Appendix D - General limited proteolysis protocol by PAF</b> .....	<b>150</b>
<b>Appendix E - Curriculum Vitae</b> .....	<b>151</b>

# 1. Summary

Teneurins are large transmembrane glycoproteins that are well-conserved across phyla and show their strongest expression in the developing central nervous system (CNS). The large extracellular domain (ECD) includes several structural features like an NHL repeat domain, a predicted beta-propeller, which is responsible for homophilic rather than heterophilic interactions. It has been shown that the homophilic interaction of the ECD leads to its release, and the subsequent regulated intramembrane proteolysis (RIP) of the intracellular domain (ICD). RIP cleaves the ICD at the membrane, after which it translocates to the nucleus, where it is known to affect transcriptional regulation.

The first part of my thesis discusses the evolution of teneurins. Teneurins are ancient proteins that are well-conserved across phyla from unicellular eukaryotic organisms like the choanoflagellate *Monosiga brevicollis* to higher multicellular organisms like vertebrates. The study suggests that teneurins may have evolved from a choanoflagellate via horizontal gene transfer from a prokaryote. It also describes the structural features of teneurins in detail, and identifies splice variants of chicken and human teneurin ICDs.

The second part of my thesis describes a novel molecular mechanism in transcriptional regulation for the intracellular domain of human teneurin-1 (TEN1-ICD). We identified several new interaction partners of the TEN1-ICD in a yeast-2 hybrid screen. Concurrently, we performed a whole transcriptome analysis of a glioblastoma cell line engineered for inducible overexpressing of the TEN1-ICD comparing induced to non-induced cells, to determine potential target genes. Results included several microphthalmia-associated transcription factor (MITF) target genes. Interestingly, MITF is directly inhibited at the promoter by transcriptional repressor

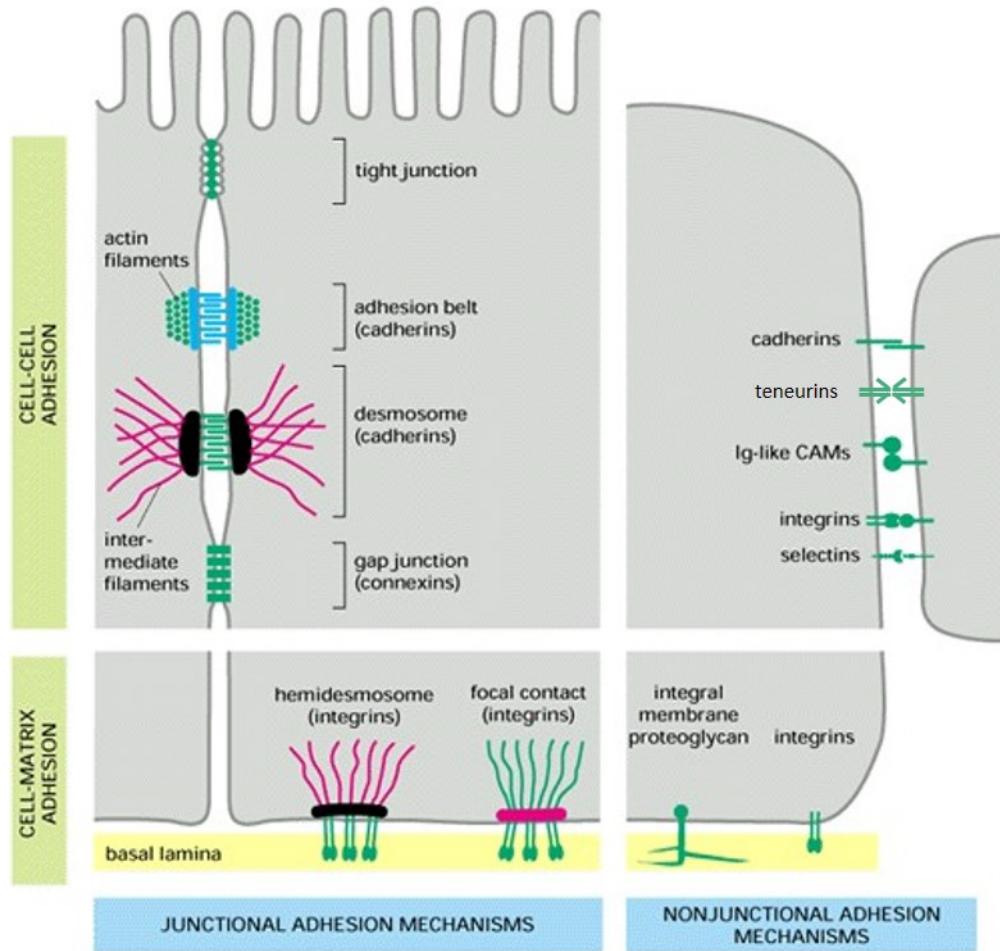
histidine-triad nucleotide binding protein 1 (HINT1), one of the novel TEN1-ICD interaction partners. Further experiments show that the TEN1-ICD competes for HINT1 binding to positively regulate MITF-dependent transcription of target gene *GPNMB*.

The third part of my thesis discusses the NHL repeat domain, located in the ECD of teneurins. Since this predicted beta-propeller is responsible for homophilic, but not heterophilic interactions in chicken teneurins-1 and -2, we were interested to learn more about the structure of the domain. For this, we started by purifying the NHL repeat domain of chicken teneurin-2 and set up drops for crystallization studies, to resolve the structure by X-ray crystallography. We have set up the purification protocol, but have not yet determined the 'right' conditions for crystallization of the protein.

## 2. Introduction

The interactions of cells with either, other cells or the extracellular matrix (ECM), are fundamentally important processes as they determine the structure of all multicellular organisms. In addition, the interactions are involved in crucial biological processes such as cell-cell communication, immune response, and embryonic development. Examples range from epithelial cells lining the gut to axonal pathfinding and synaptogenesis. All of these processes have one thing in common: they require transmembrane proteins. These proteins connect the cytoskeleton, usually through linker proteins, to other cells or the ECM (Fig. 2.1). This link is needed as the plasma membrane of cells is too weak to withstand strong forces.<sup>1</sup> One example of such transmembrane proteins is the teneurin family. Indeed, teneurins can connect the cytoskeleton of cells to other neighboring cells, while at the same time having signaling capabilities.<sup>2</sup>

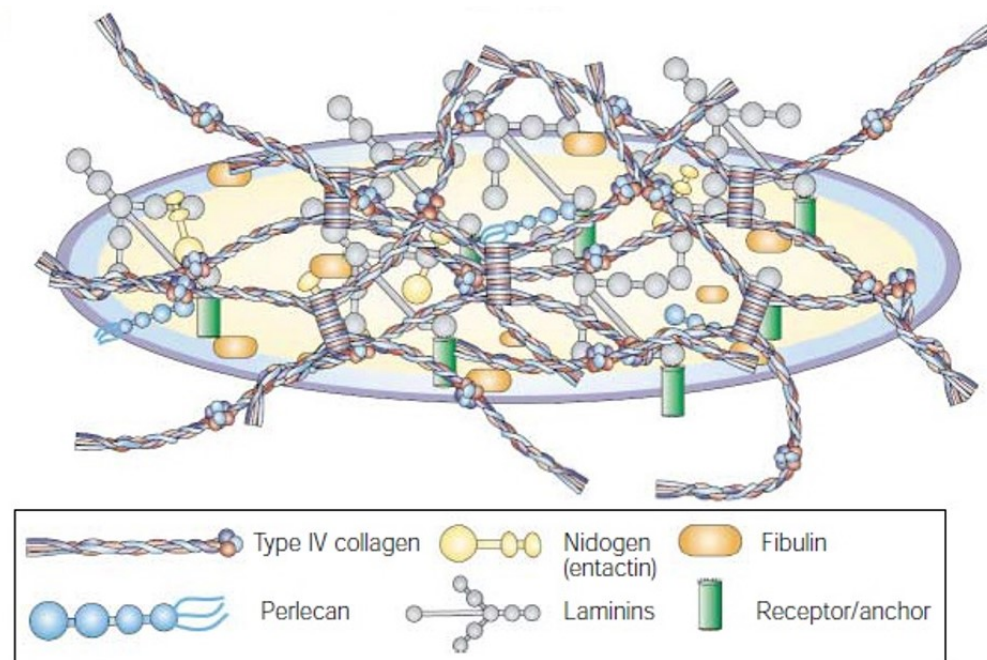
The introduction will start with the ECM, its significance, and how cells can interact with it, followed by an overview of cell-cell adhesion. Next, I will continue with early neural development and brain patterning, and get more into detail with axonal guidance and synaptogenesis, both of which are developmental processes that teneurins have been implicated in. I will conclude the introduction by giving an overview of the teneurin family, and what is known about it thus far.



**Figure 2.1 Cell-cell and cell-ECM adhesion molecules and junctions (adapted from <sup>1</sup>)**  
 Overview of the different types of cell adhesion molecules (CAMs) and their involvement in cell-cell and cell-ECM junctions.

## 2.1 ECM and Cell-ECM adhesions

The ECM is a crucial feature of multicellular organisms. For one, it gives organisms their structure and acts as a scaffold. Depending on the tissue, the quantity and the composition of ECM can vary significantly.<sup>1</sup> For example, in bone, ECM makes up more than 90% of the composition, and has a very high tensile strength, due to calcified collagen I.<sup>3</sup> Conversely, in the gut epithelium, ECM proteins are much less abundant, but no less important.<sup>1</sup> The basement membrane is a specific type of ECM (Figure 2.2) that acts as a barrier to other tissues and can help to selectively filter nutrients of certain sizes.<sup>4</sup> Thus, ECM does not only have a structural role but is also involved in a variety of cellular processes like cell proliferation, migration, and differentiation.<sup>1</sup>

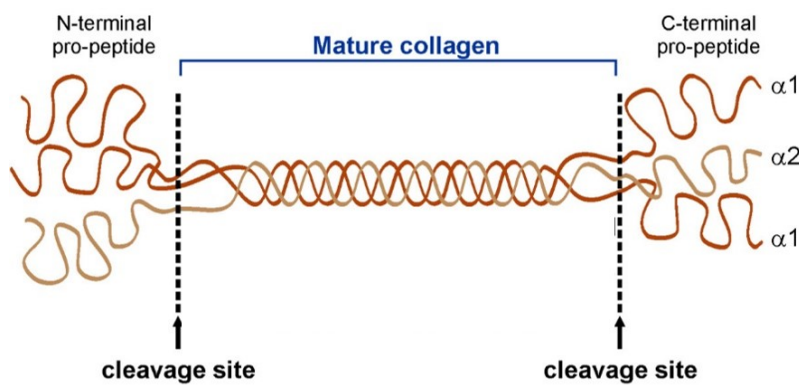


**Figure 2.2 Schematic drawing of the basement membrane<sup>4</sup>**

The basement membrane consists of several ECM proteins and is anchored to the receptors of neighboring cells. Laminin and Collagen IV form independent networks that interact with each other and the other ECM components.



The most abundant, and probably the best studied ECM proteins is the collagen family (reviewed in <sup>5</sup>). Collagens make up around one third of the total protein mass of most animals and mostly acts as a scaffold protein. Collagens are defined by the presence of one or several triple helices called collagenous domains. These domains are composed of three polypeptide  $\alpha$ -chains coiling around each other to form a triple-stranded helix (Figure 2.3). Even though there are at least 46 different  $\alpha$ -chains found in vertebrates, potentially leading to a large number of different combinations, only 28 collagen molecules have thus far been identified.<sup>6</sup> Collagen I, for example, which constitutes about 90% of the collagen of multicellular organisms, is made up of two  $\alpha 1$ [I] chains and one  $\alpha 2$ [I] chain, which is typically written as  $\alpha 1$ [I]<sub>2</sub> $\alpha 2$ [I]. Other forms of collagens also exist: network collagens, anchoring fibrils, fibril-associated collagens with interrupted triple helices, membrane-associated collagens with interrupted triple helices, and collagens with multiple triple helix domains and interruptions.<sup>7</sup> Since collagens are such important and abundant proteins, they are also involved in numerous diseases. Mutations in collagen genes can lead to chondrodysplasias, osteogenesis imperfecta, Alport syndrome, Ehler's Danlos Syndrome, and epidermolysis bullosa.<sup>7</sup>



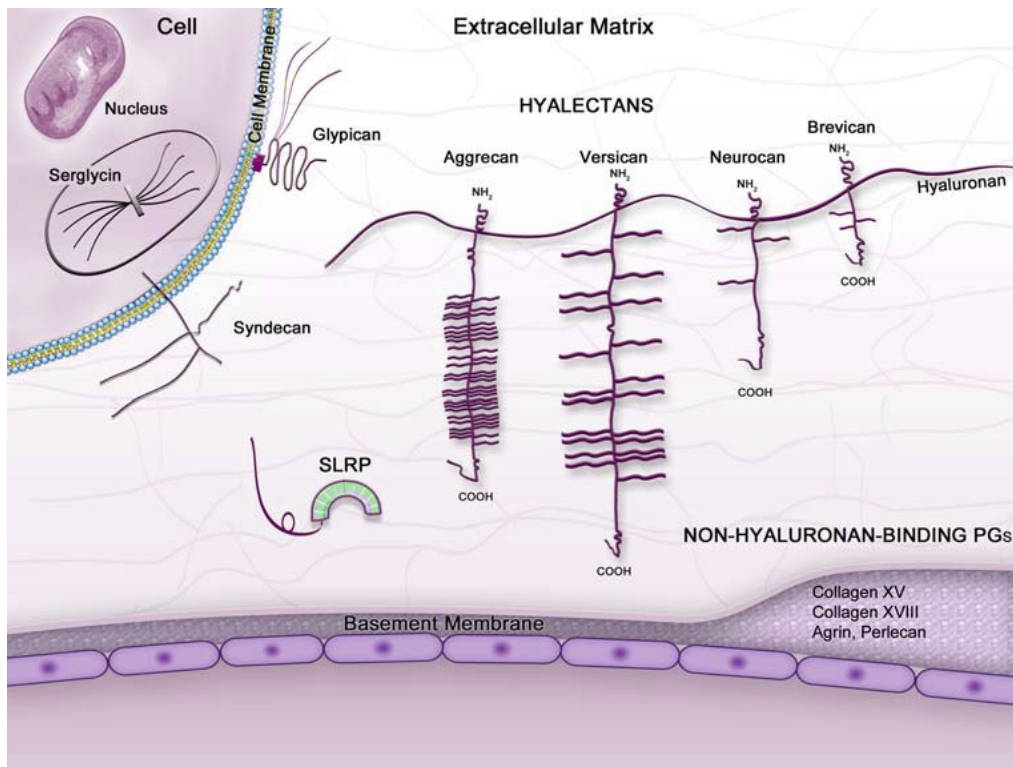
**Figure 2.3 Collagen type I structure (adapted from <sup>8</sup>)**

Example of a collagen triple helix. Pro-collagen forms in the cell into a triple helix, is secreted into the extracellular milieu, and is processed into mature collagen by collagen peptidases.

Another important family of ECM proteins are the proteoglycans (PGs) (reviewed in <sup>9</sup>). They consist of a core protein that is covalently bound to glycosaminoglycans (GAGs). The only exception is hyaluronan (HA), which is part of this family, but is made up only of GAGs and lacks the core protein. GAGs are long unbranched polysaccharides with a highly negative charge. There are two types of GAGs: sulfated GAGs like heparin, and unsulfated GAGs like HA. All GAGs consist of repeating subunits. For example, HA has a repeating glucuronic acid and N-acetylglucosamine disaccharide subunit, which can extend up to 25,000 sugars. The class of sulfated GAGs are generally bound to their core proteins via serine residues. One exception is keratin sulfate, which can be linked via N- and O-linked oligosaccharides. It is also interesting to note that the same core protein can differ in the number and types of sugars attached to it, while even the GAGs themselves can also be modified.<sup>1,9</sup>

The core proteins of PGs are a very diverse family of proteins, which can be divided into several subgroups depending on their location and binding: small leucine-rich PGs (SLRP), modular PGs (hyalectins and non-HA binding PGs), and cell-surface PGs (Figure 2.4)<sup>9</sup>. The heterogeneity of PGs also leads to a great variety of functions. Originally, like most other ECM proteins, PGs were thought to be mostly of structural importance. The GAG chains tend to be very hydrophilic and stiff, which can be helpful in forming hydrogels that can withstand very high compressive forces.<sup>1</sup> These properties are very important in the knee joint, where they act as natural lubricant.<sup>10</sup> However, PGs also contribute to other processes like cell adhesion, migration, and proliferation. One of the hyalectins, neurocan, is even involved in inhibiting neuronal attachment and neurite outgrowth. Another example is the SLRP decorin, which functions in signaling, as it can bind to multiple receptors. Decorin has been shown to inhibit transforming growth factor- $\beta$  receptor signaling or regulates fibrillin-1 synthesis. Thus, PGs not

only serve a variety of functions, but are also distributed in many different types of tissues, ranging from cartilage to the CNS. Mutations in PGs also cause several disease like Knobloch syndrome, intervertebral disc degeneration, and Schwartz-Jampel syndrome.<sup>1,9</sup>

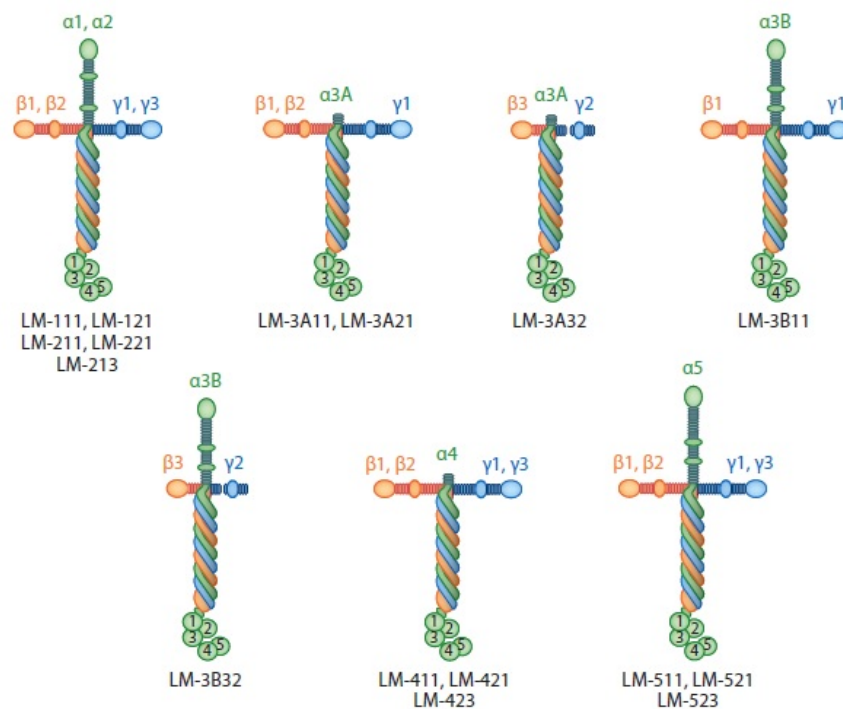


**Figure 2.4 Overview of the proteoglycan family<sup>9</sup>**

Classification of the heterogeneous PG family, based on their location and binding. (1) ECM PGs like SLRP, (2) Modular PGs that are divided into hyalectins and non-hyaluronan-binding PGs, and (3) cell-surface PGs like Syndecan.

Another family of ECM proteins with a very unique structure are the laminins (LMs) (reviewed in <sup>11</sup>). The LM structure is rather unique since the chains form a characteristic cross-like structure. Many Christians have thus dubbed it the ‘God protein’, as they believe that it proves the existence of god; a theory made popular by the American preacher Louie Giglio. LMs consist of an  $\alpha$ ,  $\beta$ , and  $\gamma$  chain, making them heterotrimeric and quite diverse (Figure 2.5). At least sixteen isoforms have been described thus far, which are named with a very simple nomenclature. For example, LM-521 consists of  $\alpha$ -chain 5,  $\beta$ -chain 2, and  $\gamma$ -chain 1. There are 5

genes encoding for  $\alpha$ -chains, 4 genes for  $\beta$ -chains, and 3 genes for  $\gamma$ -chains. LMs are vital in many tissues and processes. They are one of the major components of the basal lamina<sup>12</sup> and are involved in cell-specific processes like differentiation and adhesion<sup>11</sup>. For example, laminins are part of the basement membrane of the blood brain barrier.<sup>13</sup> When binding to certain integrins, they can also induce signaling pathways, like LM-521 binding to  $\alpha6\beta1$ -integrin (for integrin nomenclature, see below) activating the PI3K/Akt pathway, thus keeping hESCs pluripotent.<sup>14</sup> As one would expect, some laminins are also associated with diseases, like mutations in *LAMB2* causing Pierson's syndrome<sup>15</sup>, and mutations in *LAMA2* causing muscular dystrophy<sup>16,17</sup>.



**Figure 2.5 Structure of laminins<sup>11</sup>**

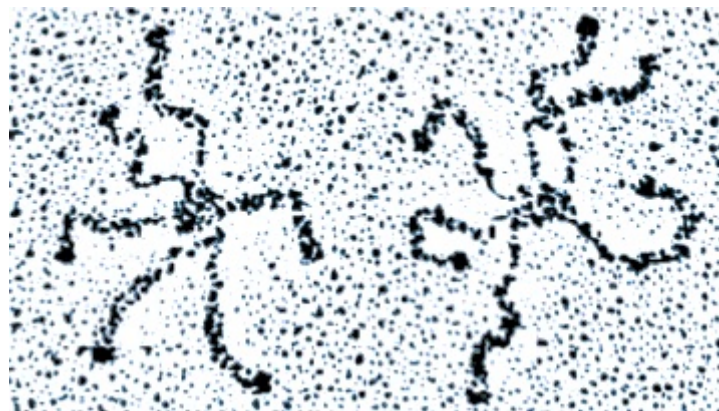
Overview of the structures of the laminin family of proteins, and the types of  $\alpha$ ,  $\beta$ , and  $\gamma$  chains that these ECM proteins consist of.

Finally, further common and important ECM proteins are the tenascins and fibronectin.

Their structures can be described as 'beads-on-a-string', since both consist of many small

domains in a row, connected by short linker regions. In addition, both are glycoproteins, and both consist to a large part of fibronectin type III domains.<sup>18</sup>

Fibronectins (FN) are large dimeric proteins, about 230-270 kDa in size (reviewed in<sup>18,19</sup>). They can be subdivided into two groups: plasma-fibronectin (p-FN) and cellular-fibronectin (c-FN). P-fibronectin is expressed and secreted by hepatocytes, whereas c-fibronectin can be expressed by many different cell types, like fibroblasts and chondrocytes. These FN variants are all derived from a single FN gene by alternative splicing of its primary transcript. FNs have functions as scaffold proteins, and influence cellular processes like migration and differentiation. During murine development, a homozygous interruption of the fibronectin gene (*FNI*) causes early embryonic lethality due to defects in the mesoderm, neural tube, and vascular system.<sup>20</sup> FNs play an important role in the tissue repair process, like wound healing, but also have a pathological role in fibrosis.<sup>21</sup> Mutations in *FNI* also lead to disease like glomerulopathy with fibronectin deposits.<sup>22</sup>



**Figure 2.6 Tenascin-C hexabrachion<sup>23</sup>**  
Rotary shadowing electron microscopy pictures of two mouse tenascin-C homo-hexamers.

There are four tenascins described thus far: tenascin-C, tenascin-W, tenascin-R, and tenascin-X (reviewed in<sup>18,24</sup>). The structure of all tenascins is very similar, starting with heptad

repeats at the N-terminus, EGF repeats, FN type III repeats, and finally a fibrinogen-like globular domain at the C-terminus. The heptad domain at the N-terminus is responsible for tenascins to assemble into homo-trimers which in the case of tenascin-C and tenascin.W can dimerize by disulfide bonds through their N-termini to homo-hexamers (Figure 2.6). The FN type III repeats and the fibrinogen globe seem to be responsible for most interactions with other proteins, like integrins, fibronectin, neurocan, and many more.<sup>18</sup> Opposite to FN, which is a very good cell adhesion protein, tenascins seem to be anti-adhesive and influence the adhesion of cells to other ECM proteins<sup>25</sup>, thereby having an effect on processes like cell adhesion and migration<sup>26</sup>. Tenascins play a significant role in embryonic development. Indeed, both tenascins-C and -R seem to be important in the development of the nervous system, though functions in adult tissues have also been described.<sup>24</sup> The importance of tenascins in human is highlighted by mutations, which in case of mutations in tenascin-X can lead to Ehlers Danlos syndrome.<sup>27</sup>

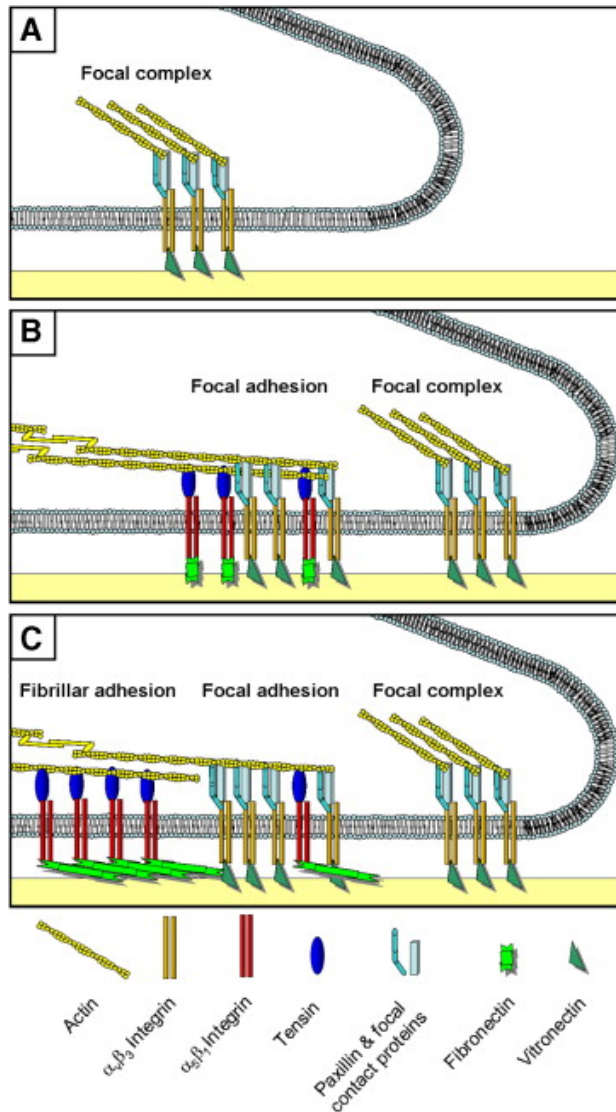
There are many other known ECM components which, together with the above mentioned proteins, make up the matrisome (reviewed in <sup>28</sup>). The mammalian core matrisome includes all common ECM proteins and consists of approximately 300 proteins. Other examples of proteins that are part of the core matrisome are thrombospondin, elastin, fibrillin, etc. There are also a large number of ECM-affiliated and ECM-modifying proteins that are not included in the core matrisome.<sup>28</sup> As discussed above the ECM plays a crucial role in many processes of a multicellular organism. Cells not only express and secrete the ECM proteins, but they also benefit from them in many vital processes. For this, transmembrane proteins of the cells need to be able to interact with the ECM proteins, as this is a way for cells to communicate with their environment, while the ECM influences the cells' behavior.<sup>1</sup> I will now outline how such an interaction between cells and the ECM can happen.

There are several transmembrane proteins that direct cell-ECM adhesion and signaling. The largest and most well-known family are the integrins (reviewed in <sup>29,30</sup>). Integrins are  $\text{Ca}^{2+}$ -dependent (or sometimes  $\text{Mg}^{2+}$ -dependent) cell adhesion molecules (CAMs). So far, twenty-four integrins have been described. Each consists of a heterodimer composed of an  $\alpha$ - and a  $\beta$ -subunit. The eighteen known  $\alpha$ -subunits and eight known  $\beta$ -subunits give integrins their name, like  $\alpha 5\beta 1$  consisting of  $\alpha$ -subunit 5 and a  $\beta$ -subunit 1. In order to drive adhesion and signaling events between cells and the ECM, integrins need to connect the ECM to the cytoskeleton. The extracellular domain of integrins can bind many ECM proteins, like collagens and fibronectin. The intracellular domain is connected to the actin cytoskeleton via adapter proteins like talin, kindlins, and vinculin.<sup>30</sup> Integrins do not simply connect the actin cytoskeleton to ECM proteins, but rather work in large protein complexes that are responsible for adhesion, outside-in, and inside-out signaling. This means that the ECM has an influence on the cell it interacts with, like causing a rearrangement of the cytoskeleton, and in turn the cell can also influence the ECM surrounding it. Some of these complexes are called focal adhesions (FAs), focal complexes and fibrillar adhesions (Figure 2.7). FAs are the best characterized of these complexes.<sup>30</sup> One example of FAs are costameres, which are important complexes in skeletal muscles, as they connect the sarcomere and the sarcolemma and are involved in transferring force in muscles.<sup>31</sup> Integrin-mediated adhesion in FAs involves a large variety of proteins. Up to 100 different proteins can make up FAs, while the entire adhesome of integrins consists of about 160 proteins.<sup>32</sup> However, integrins are thought to initiate the assembly of the FA complex. Due to the relatively low affinity of these receptors to their partners, the integrins start to cluster, also increasing the adhesion strength of the cell to the ECM. Other proteins of the complex are then subsequently recruited, such that the nascent complex can mature into a FA.<sup>33</sup> Since FAs have to

be dynamic, they can also be disassembled by ECM proteins like tenascins. Integrins do not strictly function in cell-ECM adhesion and signaling, but they are also involved in mediating cell-cell interactions.<sup>30,34</sup>

There are also non-integrin receptors for the ECM. One prominent example is dystroglycan (reviewed in <sup>35</sup>). Dystroglycan consists of two glycoproteins, but is encoded by a single gene (*DAG1*). After expression of *DAG1*, the protein gets cleaved post-translationally to give rise to the two components  $\alpha$ - and  $\beta$ -dystroglycan. The  $\beta$ -dystroglycan subunit contains the transmembrane domain of the receptor and connects to the actin cytoskeleton via linkers, while  $\alpha$ -dystroglycan is responsible for interacting with ECM components like laminin, agrin, and perlecan.<sup>35</sup> Another class is called discoidin domain receptors (DDR). DDRs are receptor tyrosine kinases that are activated by binding collagens and indirectly influence cell-ECM adhesion by influencing the activation of collagen-binding integrins.<sup>30</sup>





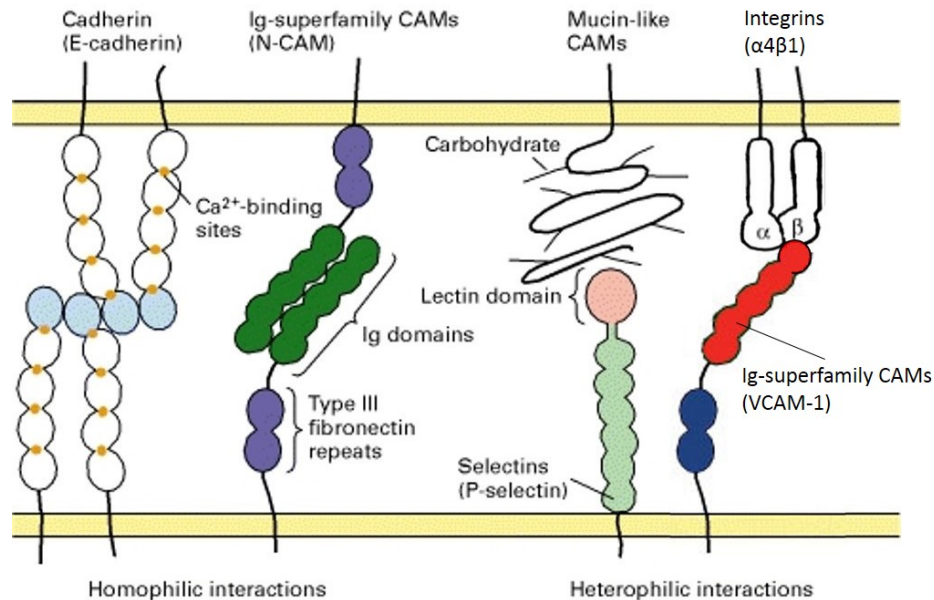
**Figure 2.7 Cell-ECM adhesion contacts<sup>34</sup>**

Summary of several cell-ECM adhesion contacts and the key players involved. (A) Initial cell-ECM contact is made by focal complexes, (B) which can mature into larger and more stable focal adhesions. (C) Fibrillar adhesions can then pull out certain components of the focal adhesion, which leads to a partial sorting.

## 2.2 Cell-cell adhesions

As mentioned above, cell-cell adhesion is another important process, especially in multicellular organisms. The adhesion can be temporary or permanent, depending on the tissue. In temporary adhesion, cells are typically not bound very tightly to each other. This process can be as short-term as leukocytes rolling along the endothelial cells of blood vessels to slow them down for leukocyte extravasation, where selectins on the surface of the endothelial cells bind their carbohydrate partners on the leukocytes.<sup>36</sup> On the other hand, the adhesion can be more permanent, with the cells bound tightly to each other, for example to act as a barrier, as is the case in the intestinal epithelium, requiring so-called occluding junctions.<sup>1,37</sup> These are two 'extreme' examples of cell adhesion, while many more examples fall somewhere in between, like synapses. While the adhesion of the pre- and post-synaptic partners is also permanent, they are less stable than the cell junctions of the intestinal epithelial cells. Especially during neuronal development, it is thought that many more connections are formed than are finally needed in the adult. Hence, synapse adhesion has to be reversible when some of the neuronal connections are pruned in children and early adulthood, also called synaptic plasticity.<sup>38</sup>

One thing that all cell adhesion events have in common is that CAMs are required. CAMs can be subdivided into the following subgroups: cadherins, selectins, immunoglobulin (Ig) superfamily of proteins, and integrins (Figure 2.8). Just like integrins in cell-ECM adhesion, CAMs during cell-cell adhesion have a relatively low affinity for each other. Stronger forces between cells are accomplished by the clustering of CAMs in some areas of the membrane, where cells will attach. Once the cells adhere to each other through CAMs, can other proteins be recruited into complexes, called anchoring junctions, leading to a stronger attachment of cells (for more details, see below after CAMs).<sup>1,39</sup>



**Figure 2.8 Major families of CAMs (adapted from 40,41)**

Overview of the four major families of CAMs. Cadherins and Ig-like CAMs undergo homophilic, selectins and integrins heterophilic interactions.

Cadherins (CDH) are a large and well-studied glycoprotein family of CAMs (reviewed in 42,43). They are  $\text{Ca}^{2+}$ -dependent adhesion proteins, giving cadherins their name. Calcium stabilizes adjacent cadherin repeats, where increasing amounts of bound  $\text{Ca}^{2+}$  ions also increases the overall stiffness of the protein. A lack of  $\text{Ca}^{2+}$  leads to proteolytic degradation of the extracellular domain.<sup>42</sup> Most cadherins are single-pass transmembrane domains, where the intracellular domain is linked to the actin cytoskeleton via catenins. The extracellular domains of cadherins mostly consist of extracellular cadherin (EC) repeats.<sup>42,43</sup> Classical cadherins have five EC repeats, while non-classical cadherins can have additional EC repeats or other additional domains like EGF-like repeats.<sup>42</sup> Cadherins are vital in embryonic development (and mature tissues), where different cadherins are responsible for the adhesion in specific tissues.<sup>43</sup> Cell types expressing the same cadherins, like N-cadherin, will only assemble with each other, rather than with cells expressing another cadherin. This led to the conclusion that cadherins can help to form specific tissues during development due to homophilic, rather than heterophilic

interactions.<sup>42</sup> Homophilic interactions take place when two or more of the same transmembrane proteins interact on apposing cells, while heterophilic interactions describe the interaction of different types of proteins. It has also been shown that cadherins can interact heterophilically or through a linker molecule, but this is less common.<sup>1</sup> The founding members of the cadherin family were N-cadherin, E-cadherin, and P-cadherin.<sup>1</sup> The letter indicates the tissue or cell type that the cadherin was originally identified in: N-cadherin (nerve cells), E-cadherin (epithelial cells), and P-cadherin (placenta).<sup>44</sup> These cadherins are not exclusively expressed in the above-mentioned tissues. N-cadherin is also expressed in the muscle, fibroblasts, etc.<sup>45</sup> Next to the classical cadherins, there are also many non-classical cadherins found in vertebrates. One such example are the protocadherins, which are mainly expressed in the brain. They seem to be particularly important in the formation and stabilization of synapses.<sup>46</sup>

Another family of  $\text{Ca}^{2+}$ -dependent glycoprotein CAMs are the selectins (reviewed in <sup>47,48</sup>). Selectins are type I transmembrane domain proteins with a small intracellular domain (ICD). The larger extracellular domain (ECD) consists of a lectin domain, EGF-like repeats, and between two and nine consensus repeats with homology to complement regulatory (CR) proteins, depending on the type of selectin.<sup>48</sup> Other than the number of CR domains, is the ECD highly conserved between paralogs and orthologs, where paralogs are two or more genes that evolved from the same ancestral genes in the same organism by duplications of parts of the genome, and orthologs are genes in different species which evolved from the same ancestral gene. The lectin domains of the different selectins, which are responsible for the carbohydrate binding, can thus bind very similar sugar structures. These sugars are usually oligosaccharides on glycoproteins or glycolipids on other cells, meaning that selectins exclusively interact heterophilically.<sup>47</sup> The transmembrane domain and ICDs are only conserved in orthologs, but not

in paralogs, which is how selectins gain their specific functions. There are three types of selectins, all of which are expressed on the surfaces of cells of the bloodstream. P-selectin is expressed in endothelial cells and platelets, L-selectin in monocytes, granulocytes, and most lymphocytes, and E-selectin is usually only expressed due to an induction by inflammatory cytokines.<sup>47,48</sup> It was mentioned previously, as an example of a temporary cell-cell adhesion, that selectins are responsible for leukocyte capture and rolling along the endothelial cells lining the blood vessels, prior to leukocyte extravasation. As the leukocytes roll along the blood vessel, the selectin-carbohydrate interactions are quickly made and broken again in order to slow down the cell, which can also be described as a Velcro-like mechanism.<sup>49</sup> This mechanism also shows how well different CAMs work perfectly in sync. Once selectins have slowed down the leukocyte, its integrins help stabilize the adhesion and allow it to extravasate from the blood stream into the tissue.<sup>36</sup>

Integrins have been mentioned extensively above, since they are mainly known as cell-ECM receptors. However, as can be seen during leukocyte extravasation, integrins are also involved in heterophilic cell-cell interactions.<sup>36</sup> A mutation in the  $\beta 2$ -integrin gene results in leukocyte adhesion deficiency-I, also supporting the significance of integrins interacting heterophilically.<sup>41</sup> These interactions include both, adhesion and signaling capabilities. Integrins are widely expressed in vertebrates and have many different functions, like aggregating keratinocytes or stabilizing synapses.<sup>30</sup>

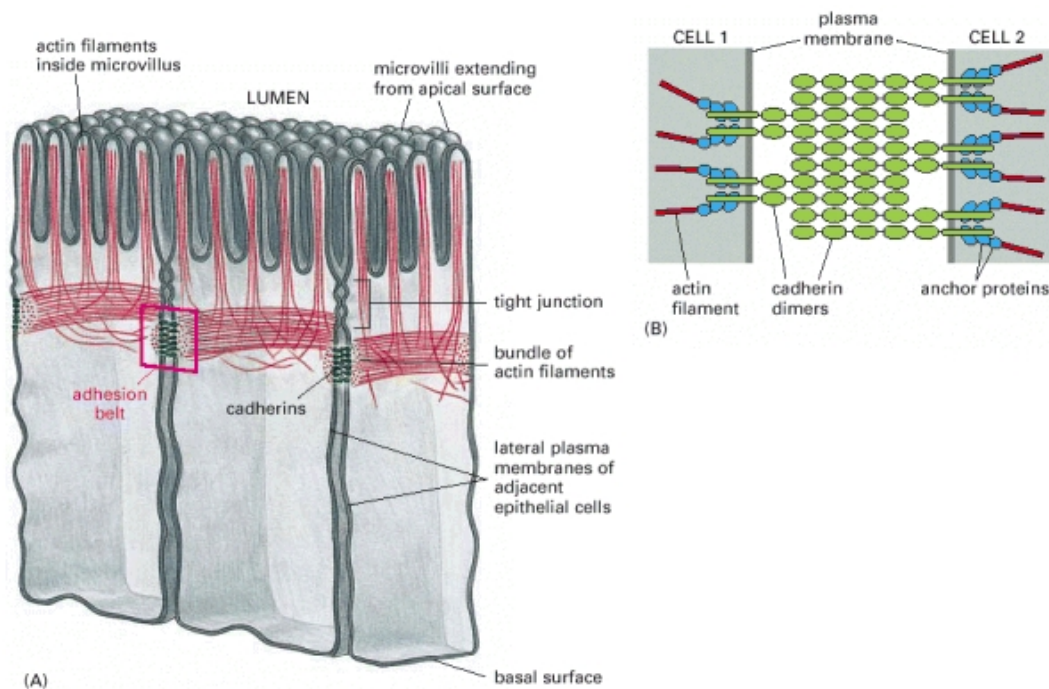
Finally, there is the major  $\text{Ca}^{2+}$ -independent family of CAMs, the Ig superfamily (IgSF) of proteins (reviewed in <sup>50</sup>). IgSF is a very diverse family of proteins, currently including 765 members, ranging from the neural cell adhesion molecule (N-CAM) to the major histocompatibility complex class I and II molecules. Members of the IgSF are typically single-

pass type I transmembrane proteins, where the ECD has to contain at least one Ig-like domain. These Ig-like domains are mainly responsible for homophilic interactions. Heterophilic interactions with integrins or carbohydrates have also been described.<sup>50</sup> Just like other CAMs, the ICD of IgSF members interacts with the cytoskeleton and can be involved in signaling. Probably the best-studied member of the IgSF is neural CAM (N-CAM). More than twenty isoforms of N-CAM have been identified thus far. N-CAM is expressed in a variety of cell types, but especially in subsets of neurons. Its most common function is to be involved in the development of the CNS, just like many other Ig-CAMs. These Ig-like proteins aid in axon guidance and synaptogenesis.<sup>39</sup> Cadherins and IgSF members are often involved in the adhesion of the same cells. Since the interactions of Ig-CAMs are generally much weaker than those of cadherins, and mutations are less dramatic in embryonic development, it is likely that the members of the IgSF family are generally involved in the fine-tuning of cell-cell interactions. IgSF proteins also play a role in pathological processes.<sup>1</sup> For example, several of the members like N-CAM, L1-type CAM, melanoma CAM, activated leukocyte CAM, etc. are involved in metastasis.<sup>50</sup>

There are several cell junctions that the CAMs help assemble: anchoring junctions, occluding junctions, and communicating junctions (Figure 2.1). Anchoring junctions mainly have a structural function by “anchoring” cells to each other or to the ECM, occluding junctions act as barriers across which water and small molecules can be selectively transported, and communicating junctions directly connect apposing cells to each other for mediating chemical or electrical signals. Not all CAMs are involved in assembling cell junctions, especially if they are not involved in selective adhesion. However, it is important that CAMs drive the adhesion of cells, before junction proteins are recruited, often becoming core proteins of the junction

themselves. Here, I will give a quick overview of cell junctions, by describing one example of each type of junction.<sup>1</sup>

The adherens junction is an example of an anchoring junction (reviewed in <sup>51,52</sup>). One of the central CAMs responsible for the formation of an adherens junction is E-cadherin. E-cadherin forms homo-dimers that interact homophilically with an E-cadherin homodimer of the apposing cell. Other core proteins include  $\beta$ -catenin and p120-catenin. Catenins connect cadherins to the actin cytoskeleton. Adherens junctions have several functions, like the adhesion of two neighboring cells by connecting their cytoskeletons, especially when higher forces act on the tissue. In epithelial cells, the adherens junctions recruit actin filaments to form belt-like structures lining the plasma membranes of attached cells (Figure 2.9). Other functions include intracellular signaling and transcriptional regulation. For example, one of the core adherens proteins  $\beta$ -catenin is a transcriptional cofactor in Wnt signaling.<sup>53,54</sup> Interestingly, while tight junction assembly requires adherens junctions, E-cadherin is not necessary for the stability of the tight junction.<sup>51</sup>

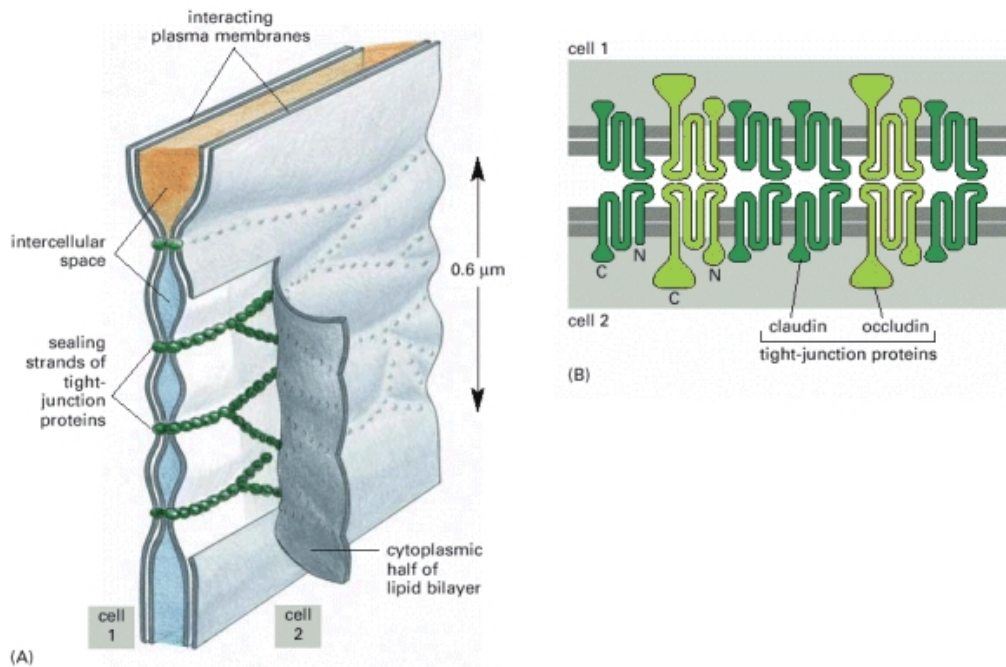


### Figure 2.9 Adherens junctions<sup>1</sup>

(A) Belt-like adherens junction between epithelial cells of the small intestine. (B) Basic structure of an adherens junction.

In vertebrates, the occluding junction is called tight junction (reviewed in <sup>51</sup>). Tight junctions are assembled by two transmembrane proteins, occludins and claudins (Figure 2.10). Intracellular zona occludens proteins link the junction to the actin cytoskeleton. The main site of assembly is the epithelium, where it helps create a barrier by tightly binding epithelial cells to each other, thus keeping contents on either side of the epithelium separate, like in the intestine. Certain nutrients need to be able to pass the epithelium. Hence, another property of the tight junction is that it is selectively permeable to ions, depending on which tissue the junction is located in. Tight junctions also create a barrier that the apical and basolateral receptors cannot diffuse onto either side. Since macromolecules cannot pass through the junctions, they must be transported through the cells. Glucose gets transported into the epithelial cells by apical surface receptors, and leaves the cell via receptors on the basolateral side. The nutrients can then enter small blood vessels that are embedded in the tissues.<sup>51</sup>



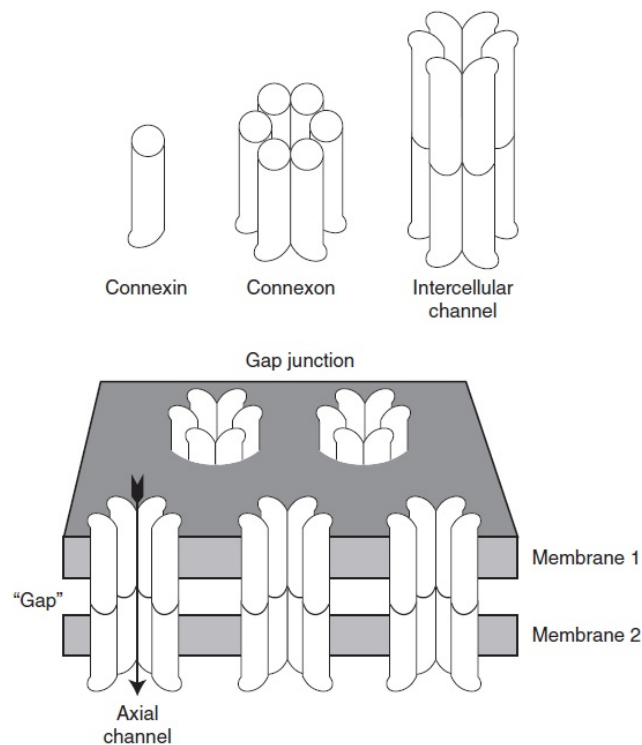


**Figure 2.10 Tight junctions<sup>1</sup>**

(A) Tight junctions sealing two cells together by (B) interacting claudins and occludins.

There are several communication junctions that exist in vertebrates. Compared to the other two types of junctions, this junction's main function is not structural, but rather in communication, by letting small molecules (less than 1 kDa)<sup>55</sup> pass directly between cells. An example of a communication junction is the gap junction (reviewed in <sup>56</sup>). One type of gap junction is the electrical synapse, while chemical synapses are a different type of non-gap communication junction. Synapses will be discussed later in more detail.<sup>57</sup> The channel of the gap junction, the connexon, is a hexamer and consists of connexin molecules (Figure 2.11). Connexins are a 21-member family of four-pass transmembrane proteins. Different combinations of connexins can form either homomeric or heteromeric connexons, but only certain combinations have been identified thus far. Connexons of two apposing cells need to line up to form a complete channel between the two cells. Both, homotypic and heterotypic intercellular channels exist. Gap junctions are assembled by neighboring cells in most tissues. The pore of the

channel is about 1.5 nm in size, which means that only ions and other small molecules, but no macromolecules can pass through, leading to an electrical and/or metabolic coupling. The close proximity of about 2 nm between two cells that is required for the gap junction to form, also keeps most other proteins from occupying that space.<sup>56</sup>



### Figure 2.11 Gap junctions<sup>56</sup>

Six connexins form homomeric or heteromeric connexons, which can interact homo- or heterophilically on apposing cells to form a gap junction.

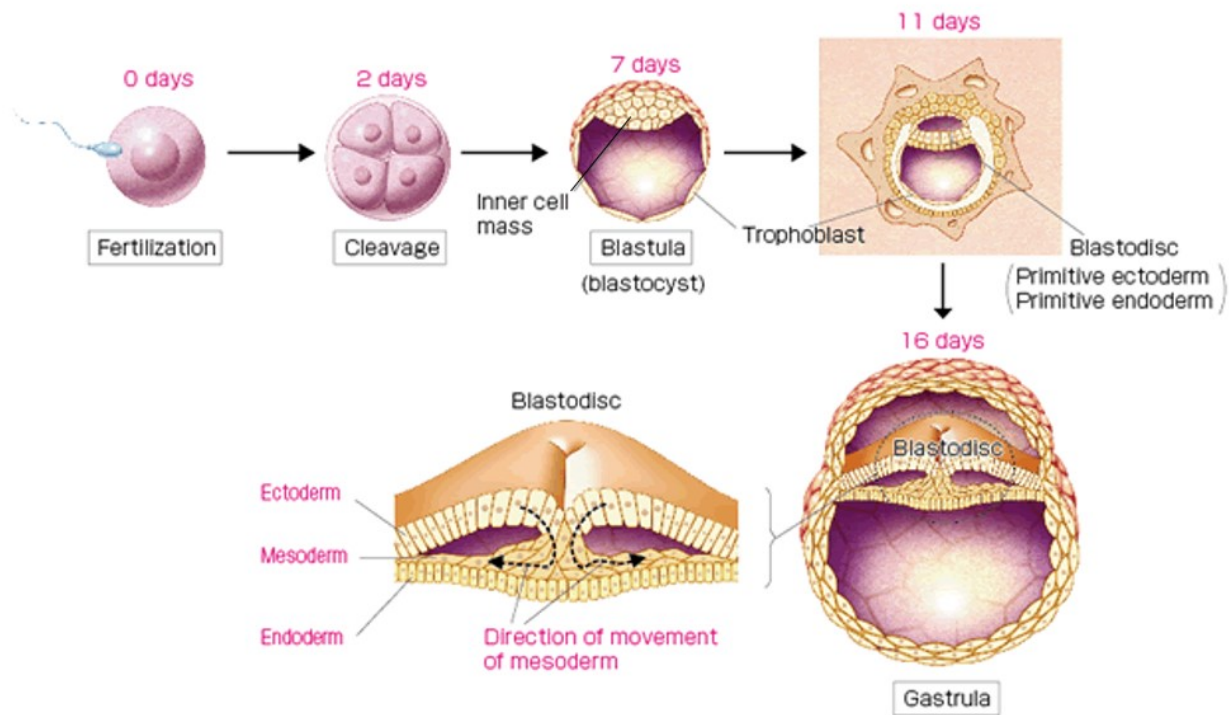
As can be seen in the previous two sections, both cell-ECM and cell-cell adhesion are vital in multicellular organisms. Vertebrate development would not be possible without these processes. In the developing CNS, cells migrate during early neural development to their appropriate locations, different parts of the brain develop during patterning and arealization, and neurons connect the different regions through axon guidance and synaptogenesis. One thing they have in common is that they all require cell-cell and/or cell-ECM interactions.

## 2.3 Early neural development

A lot is known about the mammalian CNS and its development. Since my projects were mainly set in a human context, I will focus on the human, or at least general mammalian, neural development when possible, starting with early embryonic development and how it gives rise to early neural progenitors.

After fertilization of the oocyte and once meiosis took place, it takes about another day until the first cleavage, which brings the embryo to the two-cell stage. Due to the asynchronous division of the blastomeres, the embryo reaches the ten-cell stage within four days, at which point it hatches from the zona pellucida. At the sixteen-cell stage, the morula has formed, with the outer part consisting of tightly bound cells, which will give rise to the trophoblast. This becomes more apparent at the blastocyst stage, when the trophoblast is formed around a fluid-filled cavity (blastocoel), also containing the inner cell mass (ICM) (Figure 2.12)<sup>35</sup>. The ICM now is in a pluripotent state (compared to the previously totipotent cell stages), i.e. its cells during gastrulation will give rise to the three germ layers that during further development will form the actual embryo, the embryo proper. The trophoblast is required for the blastocyst to implant, as it contacts the uterine epithelium, and it gives rise to the placenta that is required to nurture the developing embryo. This complete separation of the trophoblast and the ICM is the first of many differentiation events.<sup>58</sup> By the time the embryo reaches E13, gastrulation starts. The ICM has segregated into two cell layers, consisting of the lower layer, the hypoblast (a.k.a. primitive or embryonic endoderm) and the upper layer, the epiblast (a.k.a. primitive or embryonic ectoderm). The hypoblast cells are released from the ICM and become part of extraembryonic tissue by lining the blastocoel. At this time, the epiblast starts to form the three primordial germ layers: mesoderm, endoderm, and ectoderm. Gastrulation starts by primitive

streak formation and individual epiblast cells migrating through it, and towards the rostral end of the embryo, forming two new layers. The lower layer becomes the definitive endoderm, replacing the hypoblast, and the intermediate layer becomes the mesoderm. The endoderm gives rise to tissues like the gut and respiratory system, and the mesoderm forms tissues like muscle, bone and cartilage.<sup>59</sup>



**Figure 2.12 Early human embryonic development<sup>60</sup>**  
 Developing human embryo, from fertilization to the gastrula stage.

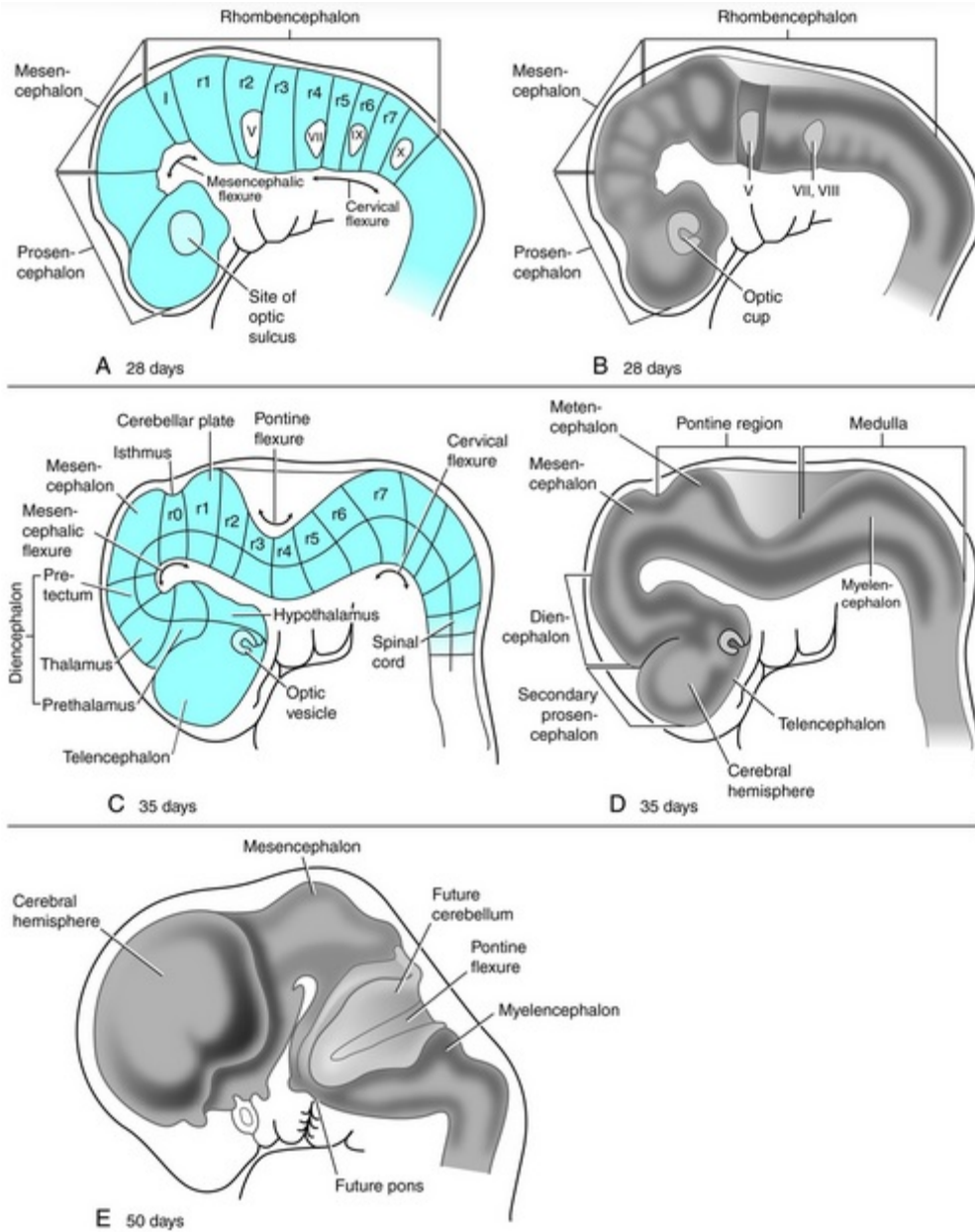
Most significant to neural development at this point is the ectodermal layer, which forms into epidermal-ectodermal and neurectodermal progenitors. The epidermal-ectodermal lineage will form tissues like skin, nails and sweat glands, while neurectodermal cells are the neural progenitors that give rise to the CNS. These neural progenitors are now lining the rostral-caudal midline (neural plate) and start forming their first neural structure around E20-E27, the neural

tube. Formation of the tube starts by two ridges emerging on the neural plate, with the neural progenitors located between them. The ridges then fold up and inward, and fuse to each other to form a hollow tube. Inside, the hollow tube is lined by a single-cell layer of neural progenitors. From now on, the CNS formation becomes more complex, with the rostral end of the neural tube forming the brain and the caudal end developing into the hindbrain and spinal column.<sup>59,61</sup>

## 2.4 Patterning of the brain

Once the primitive neural structures have been formed, they have to develop further into different parts of the CNS including the complex organization of the brain. While patterning of the CNS starts in the embryo, at this point still quite primitive, it goes on for several years. Brain patterning starts by three primary brain vesicles (neuromeres) developing at the anterior part of the neural tube, just before it closes. These primary neuromeres are called prosencephalon (which itself is divided into the telencephalon and diencephalon), mesencephalon, and rhombencephalon, and subsequently divide into secondary neuromeres during different stages of development (Figure 2.13). This division takes place along the rostral-caudal axis. The vesicles of the telencephalon start to take the form of the brain hemispheres by eight weeks of gestation, which is also the end of the embryonic period. Other parts of the primitive CNS, like the spinal column, hindbrain, and sensorimotor regions of the neocortex are starting to be specified as well at this point.<sup>61</sup>

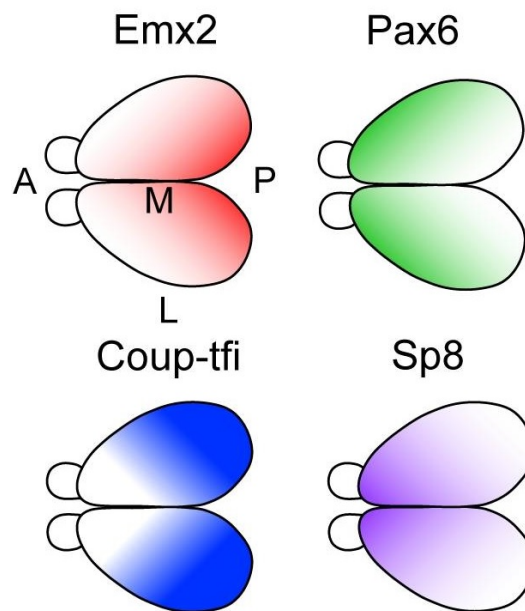
Organization and arealization of the brain is driven by a complicated network of transcription factors, CAMs and secreted proteins. Homeobox (HOX) proteins for example are often expressed in gradients during development, specifying the anteroposterior order of segments in the hindbrain and spinal cord.<sup>62</sup> However, the neocortex is the best-studied example in brain patterning. The cerebral cortex is the largest and most complex part of the brain and is formed from the dorsal telencephalon. The neocortex makes up the largest part of the cerebral cortex, and consists of six layers (I-VI, numbered from the outermost to the innermost layer). Functions of the neocortex include the processing of visual and auditory information, and the voluntary movement of body parts.<sup>63</sup>



**Figure 2.13 Lateral view of the developing brain<sup>64</sup>**

Several stages of early human brain development. (A) and (B) show the primary neuromeres in brain patterning, which (C) and (D) turn into secondary neuromeres. (E) At 50 days, the brain has started to develop from the vesicles of the telencephalon.

The signaling molecules that regulate transcription factors responsible for patterning of the neocortex, are secreted by so-called patterning centers. Some of the major patterning centers are called commissural plate, cortical hem, and antihem, and secrete factors such as members of the FGF, WNT, BMP, and NRG families of proteins. SHH is secreted by a patterning center of the ventral domain, but only indirectly affects arealization of the neocortex. Four transcription factors directly involved in the patterning of the neocortex have been identified in cortical progenitors thus far: COUP-TFI, EMX2, PAX6, and SP8.<sup>63</sup>



**Figure 2.14 Patterning of the neocortex<sup>65</sup>**

The four transcription factors expressed in gradients during the arealization of the neocortex, responsible for specifying the anterior (A) – posterior (P) and lateral (L) – medial (M) axes.

COUP-TFI, EMX2, PAX6, and SP8 are expressed as gradients and specify the arealization of the anterior-posterior and lateral-medial cortical axes (Figure 2.14). Specifically altering the expression levels of any of these transcription factors also changes the proportion and position of the different cortical areas. When overexpressing Emx2 in the neocortex, it



causes the size of the primary sensory and frontal/motor cortical areas to change, without increasing or decreasing the overall size of the cortex. Further, the transcription factors specifying the different regions of the neocortex are highly concentration dependent, also seen by the gradient patterns of expression. The “Cooperative Concentration Model” states that all four transcription factors are expressed across all cortical progenitors, but differing levels of expression are responsible for generating the regions of the neocortex.<sup>66</sup>

The development of the brain lasts throughout the embryonic and fetal periods, and into the post-natal period. During this time, the brain is still malleable and can be influenced by different inputs and experiences.<sup>59</sup> For example, the sizes of primary cortical areas can vary between two- and three-fold, across individuals of a normal population. The overall volume of the cortex only varies around 30%. Responsible cues or inputs have not been identified yet.<sup>66</sup>

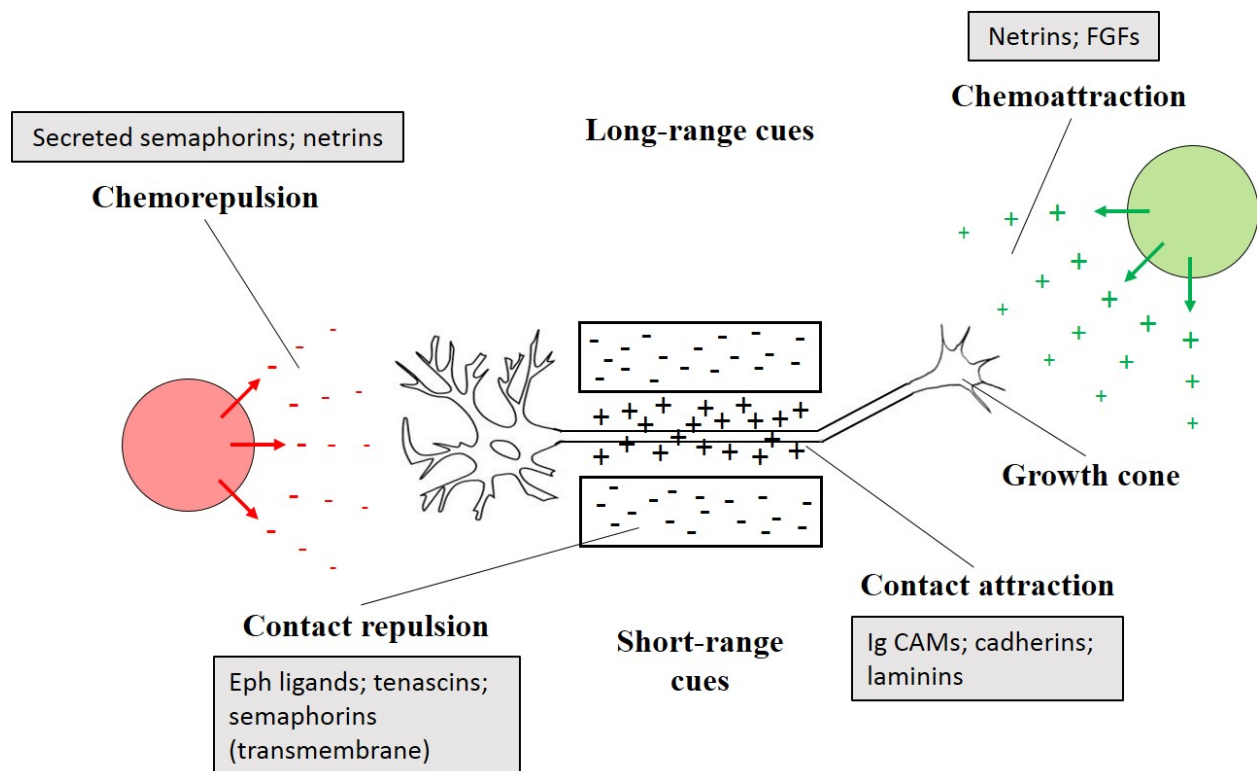
## 2.5 Axon guidance and synaptogenesis

Neurons and their processes connect the different areas of the brain and make up a significant part of the peripheral nervous system. They are responsible for sending electrochemical signals within the entire body in milliseconds. Neurons consist of a cell body, a single axon, and several branched dendrites. Axons can be as short as 1 mm and reach a length of up to one meter and may branch hundreds of times, making neurons very unique cell types. Different types of glial cells support neurons structurally, provide them with nutrients, and help form insulating myelin sheaths around their axons. In most cases, the dendrites and axons of different neurons are connected and communicate with each other via synapses.<sup>1</sup> The cell bodies of neurons start to appear around E42, by asymmetric division of the neural progenitors. Previous symmetric division of neural progenitors led to two subsequent progenitor cells, whereas asymmetric division now leads to one neuron and one neural progenitor. Regions of the brain that contain neuronal cell bodies give it a grey color, thus naming it grey matter. Fibrous parts of the brain that lack these cell bodies are called white matter. The axons then start to grow out to connect with their dendritic partners. It is estimated that more than 100 billion neurons exist in the human brain, each making an average of 1000 connections. Most of these connections are established by mid-gestation, but neuronal development is not completed until young adulthood, or arguably not completed for the entirety of the lifetime. This leads to a very complicated, but perfectly organized neuronal network. To make the connections between neurons, axons have to “find” the right path in a process called axon pathfinding or axon guidance. Once the axon has reached its proper target, a synapse is formed with the dendrite of another neuron, which is called synaptogenesis.<sup>59</sup>

Axon pathfinding was first mentioned more than a century ago by Ramón y Cajal, who proposed that the axons are guided to their proper targets by a set of long-range chemical attractants, secreted by the target cells. This was a remarkable observation, even though axon guidance is a bit more complicated as the past decades of research have shown. He also gave the club-like structure at the tip of the axon its name, the growth cone.<sup>67</sup> The growth cone is a highly motile structure that extends the axon and determines the direction of axon guidance, by receiving signals through its receptors. A significant part of the growth cone is the lamellipodium, which consists of an F-actin network and is considered to drive the extension of stable microtubules in the axon. Small protrusions of F-actin bundles, called filopodia, continuously extend and retract from the lamellipodium, “searching” for appropriate cues. Attractive or repulsive signals are recognized by the aforementioned receptors in the membranes of the filopodia, the growth cone undergoes rapid cytoskeletal rearrangement and then turns where the guidance cues direct it.<sup>68</sup> These signals involve direct interactions of receptors of the growth cone with receptors of other cells or ECM proteins. Many of the guidance cues can act as both, an attraction or a repulsion signal, depending on the neuron. A molecule may even attract and later repulse the same growth cone at different points of its path. There are four main classes of proteins and their receptors that are involved in guiding axons: netrins bind to deleted in colorectal cancer receptors, slit binds to roundabout, semaphorins bind to plexins, and ephrins (EFN) bind to ephrin receptors (EPH) (Figure 2.15). Other proteins that had previously been identified as potential candidates, like N-CAMs and integrins, are more involved in axon outgrowth than its guidance.<sup>67,69</sup>

Once the axon has reached its target dendrite, a synapse has to form, in a process called synaptogenesis. As mentioned previously, synapses are communicating junctions, which are

involved in sending electrochemical signals from a presynaptic neuron to its postsynaptic partner. Unlike other communication junctions, like gap junctions (e.g. electrical synapses), chemical synapses are very asymmetric, while its basic functions are the same. Synapses are also involved in physically coupling two cells to each other, as well as in intracellular membrane trafficking and communication between the two cells. The main communication between neurons is via chemical neurotransmitters that are released in vesicles from the membrane of the pre-synaptic neuron, move across the small gap and are recognized by receptors of the post-synaptic neuron.<sup>1,57</sup>



**Figure 2.15 Axon guidance (adapted from <sup>70</sup>)**

Several cues guide the axon towards its target, either in an attractive or a repulsive manner. Short-range cues are membrane-bound, and long-range cues are secreted. Guidance molecules are recognized by receptors on the growth cone.

Synaptogenesis is a very complex process, as there are over 100 billion neurons in the brain, forming an average of 1000 synapses each. It has been proposed that synapses do not form

between two specific neurons, but between two classes of neurons, putting an emphasis on the total number of formed synapses. This slightly reduces the complexity, and may also explain the significance of pruning during development, if an unspecified number of synapses are formed between classes of neurons.<sup>71</sup> Synaptic density is at its highest in humans around the age of five, and synaptic pruning is thought to continue until the age of thirty.<sup>38</sup> The specificity of synapse formation is determined by the axon finding its target dendrite and getting in close proximity of it, before synapse cell adhesion can start. Once the growth cone enters the neuropil, CAMs start forming the synapse. There are several proteins involved, like CDHs, N-CAMs, EFNs/EPHs, and neurexins/neuroligins. The specific roles of these CAMs in synapses have not yet been identified, however it is proposed that they are involved in the formation and mechanical stability of synapses, as well as trans-synaptic signaling. Their involvement in the specificity of the formed synapse is unlikely.<sup>71</sup> Other protein families involved in axon guidance and synaptogenesis, like the teneurins, are only starting to be investigated in more detail in recent years.<sup>72</sup>

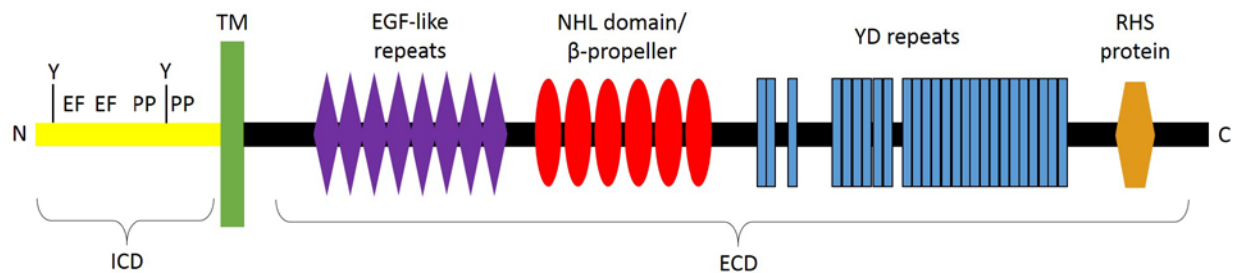
## 2.6 Teneurin overview and structure

Teneurins were originally discovered in *Drosophila* by two independent laboratories. The first approach identified two teneurin paralogs while screening a *Drosophila* expression library for tenascin-C homologs in arthropods, based on its EGF-like repeats.<sup>73,74</sup> The two newly discovered proteins received the names tenascin-like accessory (ten-a), and tenascin-like major (ten-m). Both names are still in use today, but only in a *Drosophila* context. Further studies showed that ten-a and ten-m differed structurally and functionally from tenascins, with only the EGF-like repeats in common. The other approach identified ten-m, when screening for phosphotyrosine-containing proteins in *Drosophila*. It was given the name odd Oz (odz), due to mutations in embryos showing an “oddless” pair-rule phenotype.<sup>75</sup> It was only discovered recently that the pair-rule phenotype was an artefact, due to a problem with the balancer chromosome in the mutant embryos.<sup>76</sup> The now commonly used name ‘teneurin’ was proposed by the original discoverers of the protein family, and stems from its historic name in *Drosophila*, and its neuronal context, which will be discussed later in more detail.<sup>77</sup>

Recent papers that focus on the evolution of teneurins, have identified orthologs across phyla, even in eukaryotic unicellular organisms, such as the choanoflagellate *Monosiga brevicollis* (Figure 2.15).<sup>78</sup> Teneurins have also been studied in *C. elegans*, in which a long (TEN-1L) and a short isoform (TEN-1S) of the same gene, but no paralogs, are expressed.<sup>79</sup> Vertebrate orthologs of teneurins were first identified in the mouse.<sup>80</sup> Most vertebrates express four paralogs, called teneurins 1-4 (Tenm1-4 in most vertebrates, TENM1-4 in humans).<sup>81,82</sup> Zebrafish are an exception with five paralogs, teneurins-1, -2A/B, -3, and -4.

Teneurins are a family of single-span type II (amino-terminus is in the cytosol and carboxy-terminus is extracellular) transmembrane glycoproteins that are well-conserved across phyla (basic structure of teneurins summarized in Figure 2.16). This family of proteins consists of a relatively small N-terminal ICD (ca. 45 kDa), followed by a single transmembrane domain, and a much larger C-terminal ECD (ca. 260 kDa). The sequence of the large ECD is very well-conserved between ortho- and paralogs across phyla, unlike the ICD.<sup>83</sup> The overall sequence identity in vertebrates is between 60% (orthologs) and 98% (paralogs), and between 33% and 41% between vertebrate and *Drosophila* or *C. elegans* teneurins. The vertebrate ICDs have a sequence similarity of 50-60% (paralogs) and 70% (orthologs). Vertebrate teneurin ICDs have several features: the polyproline-rich region can bind SH3-domain containing proteins (also found in invertebrate ICDs) such as CAP/Ponsin, several putative phosphorylation sites and an EF-hand like domain that can bind Ca<sup>2+</sup> ions.<sup>84,85</sup> Some ICDs also contain predicted nuclear localization signals (NLS). The presence of an NLS was confirmed in chicken teneurin-1, by site-directed mutagenesis.<sup>86</sup> The *C. elegans* isoforms of TEN-1 only differ in their ICDs, the longer isoform containing a predicted NLS, while the shorter isoform does not.<sup>79</sup> Adjacent to the ICD is a transmembrane domain, followed by a short linker region and EGF-like repeats of the ECD.<sup>77</sup> There are eight EGF-like repeats, which are responsible for dimerization of teneurins. Most repeats contain an even number of cysteines that form disulfide bridges with each other, except in repeats two and five. Instead of one of the cysteines in each of the two repeats, there is a tyrosine in its place. Hence, two neighboring teneurin molecules can form disulfide bonds in repeats two and five, and thus dimerize.<sup>81,87</sup> The following NHL-repeat domain is a predicted beta-propeller and is involved in protein-protein interactions.<sup>81,88</sup> Teneurins are known to interact with proteins on apposing cells and with ECM proteins.<sup>83,88,89</sup> In teneurins-1 and -2, the NHL-

repeat domain is responsible for homophilic, rather than heterophilic interactions.<sup>88,90</sup> Other studies have shown a heterophilic interaction between teneurins and latrophilins.<sup>89,91</sup> The subsequent 26 YD-repeats are unique to teneurins as non-bacterial proteins<sup>82</sup>, followed by a short peptide resembling the corticotrophin releasing factor (CRF), the teneurin c-terminal associated peptide (TCAP).<sup>92</sup> YD-repeats in bacteria are part of cell wall proteins in gram-negative bacteria and have a strong affinity for carbohydrate-binding; their functions in teneurins are mainly unknown.<sup>82</sup> Previous studies suggest that TCAP has neuromodulatory activity.<sup>93</sup>



**Figure 2.16 General teneurin structure (adapted from <sup>94</sup>)**

Overview of the basic domain organization. ICD = intracellular domain; ECD = extracellular domain; N = amino-terminus; C = carboxy-terminus; Y = tyrosine residue; EF = EF-hand-like motif; PP = polypoline-rich motif; TM = transmembrane domain.

There are several predicted furin-cleavage sites in the teneurin ECD. The teneurin molecule could thus be processed into several individual domains, i.e. shedding the entire ECD, and the TCAP into the extracellular milieu.<sup>94</sup> Cleavage of the ECD at the membrane is one of the requirements for releasing the ICD via regulated intramembrane proteolysis (RIP).<sup>95</sup> Proteases for teneurin RIP have not been identified yet. Signal peptide peptidase or site-2 protease are likely candidates, as they are involved in the processing of type II transmembrane proteins.<sup>94</sup> Several studies have shown the release and subsequent translocation of the ICD into the nucleus.<sup>84,86,96</sup>



**Table 2.1 Summary of teneurin nomenclature and expression during development**

Includes only proteins with known expression patterns; adapted from <sup>97</sup>

Species	Name	Synonyms	Expression pattern
<i>Caenorhabditis elegans</i>	Ten-1L <sup>79</sup>		Somatic gonad <sup>79,98</sup> , vulva, subset of neurons, gut, some hypodermal and muscle cells <sup>79</sup> , pharyngeal cells <sup>98,99</sup> , intestine cells <sup>99</sup>
	Ten-1S <sup>79</sup>		Subset of neurons, some hypodermal cells <sup>79,99</sup>
<i>Drosophila melanogaster</i>	Ten-m <sup>74,83</sup>	Odz <sup>75,100</sup>	Odd-numbered parasegments <sup>74,75</sup> , subset of neurons <sup>74,101,102</sup> , tracheal system, cardiac cells, lymph glands, muscle attachment sites <sup>74</sup> , morphogenetic furrow <sup>83,100</sup> , wing pouch, leg and antennal discs <sup>100</sup> , motor neuron <sup>76</sup>
	Ten-a <sup>73,103</sup>		CNS <sup>73,103,104</sup> , eye, muscle attachment sites <sup>73,103</sup>
<i>Danio rerio</i>	Teneurin-3	Ten-m3 <sup>105</sup>	Developing brain, somites, notochord, pharyngeal arches <sup>105</sup> , subset of neurons <sup>106</sup>
	Teneurin-4	Ten-m4 <sup>105</sup>	Developing brain, spinal cord <sup>105</sup>
<i>Gallus gallus domesticus</i>	Teneurin-1 <sup>77,83</sup>		Developing CNS and eye <sup>77,83,86,107</sup> , limb buds <sup>107</sup>
	Teneurin-2 <sup>77,108</sup>		Developing brain and eye <sup>77,86,107</sup> , AER of limb buds, tendon primordia, pharyngeal arches, heart, somites, neural tube, craniofacial mesenchyme <sup>108</sup>
	Teneurin-3 <sup>107</sup>		Developing CNS, central retina, limb bud <sup>107</sup>
	Teneurin-4 <sup>107,109</sup>		Developing CNS, ZPA of limb buds, pharyngeal arches <sup>109</sup> , heart, lung bud <sup>107</sup>
<i>Rattus rattus</i>	Teneurin-2	Neurestin <sup>110,111</sup> , lasso <sup>91</sup>	Developing and adult CNS <sup>110,111</sup> , somites <sup>110</sup>
<i>Mus musculus</i>	Teneurin-1	Ten-m1 <sup>81</sup> , odz1 <sup>112,113</sup> , ten-m/odz1 <sup>114</sup>	Developing and adult CNS <sup>81,113,114</sup> , eye, smooth muscle cells in lungs, kidney glomeruli, adult testes <sup>81</sup>
	Teneurin-2	Ten-m2 <sup>81</sup> , odz1 <sup>115</sup> , odz2 <sup>112</sup> , ten-m/odz2 <sup>114</sup>	Developing and adult CNS <sup>113,114</sup> , developing eye <sup>116</sup>
	Teneurin-3	Ten-m3 <sup>81</sup> , odz3 <sup>112,113</sup> , ten-m/odz3 <sup>114</sup>	Developing and adult brain <sup>113-115</sup> , developing eye <sup>112,117-119</sup> , spinal cord, notochord, craniofacial mesenchyme, tongue, dermis, saccule, developing limb, periosteum <sup>112</sup> , cartilage <sup>120</sup>
	Teneurin-4	Ten-m4 <sup>81</sup> , odz4 <sup>112,113,121</sup> , ten-m/odz4 <sup>114</sup> , Doc4 <sup>80,112,115</sup>	Developing and adult brain <sup>80,113-115,122</sup> , developing eye, somites, spinal chord, trachea, nasal epithelium, saccule, joints, adipose tissue, tail bud and limbs <sup>114</sup> , cartilage <sup>123</sup> , embryonic mesoderm <sup>124</sup>
<i>Macropus eugenii</i>	Teneurin-3	Ten-m3 <sup>125</sup>	Developing visual system <sup>125,126</sup>

## 2.7 Teneurin expression and functions

One of the key sites of expression of teneurins is the (developing) CNS, especially in well-defined subpopulations of neurons. Expression of teneurins is also found in many non-neural tissues and has been investigated in several animal models, ranging from *Drosophila* to the mouse (Table 2.1).

Several developmental functions involving ten-a and ten-m have been identified in *Drosophila*. Loss-of-function of ten-a and ten-m lead to eye defects, like photoreceptor loss.<sup>127</sup> Ten-m and filamin physically interact to affect the axon guidance of motor neurons in *Drosophila*.<sup>76</sup> Heterophilic interaction of ten-a in the pre-synaptic and ten-m in the post-synaptic neuron are required for synaptic organization in the neuromuscular junction.<sup>102</sup> Guidance and connectivity of olfactory receptor neurons with projection neurons in an olfactory map is driven by homophilic interactions of ten-a and ten-m.<sup>101</sup> Finally, ten-a is one of the genes involved in central complex formation, the part of the brain involved in several processes like visual and olfactory memory.<sup>104</sup>

Expression of the two ten-1 isoforms in *C. elegans* had very complex and distinct patterns, not only in developing larvae, but also in adult worms. Patterns of expression are not restricted to the developing nervous system, but also include the gut, gonads, and some hypodermal cells. Both, an RNAi knockdown and the knockout (KO) of the ten-1 gene in *C. elegans* had drastic effects on some larvae, which included protusion of the vulva, defects in the CNS and disintegration of the somatic gonad.<sup>79</sup> Further studies show the expression and involvement of ten-1 in somatic precursors of the gonad, which also seem important for the integrity of the pharyngeal basement membrane.<sup>98</sup> Another RNAi approach identified phy-1, the

catalytic domain of collagen prolyl 4-hydroxylase as an interaction partner of ten-1. Double-mutants embryos of ten-1 and phy-1 disrupted the integrity of the basement membrane and showed epidermal defects.<sup>128</sup>

In the zebrafish, only expression patterns of Ten-m3 and Ten-m4 have been described during development of the embryo. Ten-m4 has a faint expression pattern during gastrulation, while no Ten-m3 expression can be detected. During later stages of development, Ten-m3 and Ten-m4 expression was detected in the brain, while Ten-m3 was also expressed in somites, notochord and the pharyngeal arches. The two genes showed complimentary expression patterns during different stages of development of the forebrain and midbrain, which is often characteristic for teneurins.<sup>105</sup> Another study showed Ten-m3 expression in a subset of developing neurons, especially of the zebrafish visual system.<sup>106</sup>

Teneurins have also been studied in the avian model, *Gallus gallus domesticus*. As expected, teneurins have their strongest expression in the developing chick CNS. Teneurin-1<sup>77,83</sup>, as well as teneurin-2<sup>77</sup>, are expressed in subsets of neurons in the developing visual system. This distinct and largely non-overlapping expression of teneurins-1 and -2 in the visual system is especially found in the tectofugal and thalamofugal pathways, respectively.<sup>90</sup> Complementary patterns of teneurin-1 and -2 expression have also been described for other parts of the developing chick CNS, like the retina, olfactory bulb and cerebellum.<sup>86</sup> During different stages of development, teneurin-4 is first expressed in bundles of axons in the nasal retina, and later in the temporal ganglion cell layer.<sup>107</sup> Teneurin-2 and -4 are both expressed in non-neural tissues, like the developing limbs.<sup>107-109</sup> In vitro, teneurin-2 overexpression leads to an increased number of filopodia, and enlarged growth cones.<sup>77</sup>

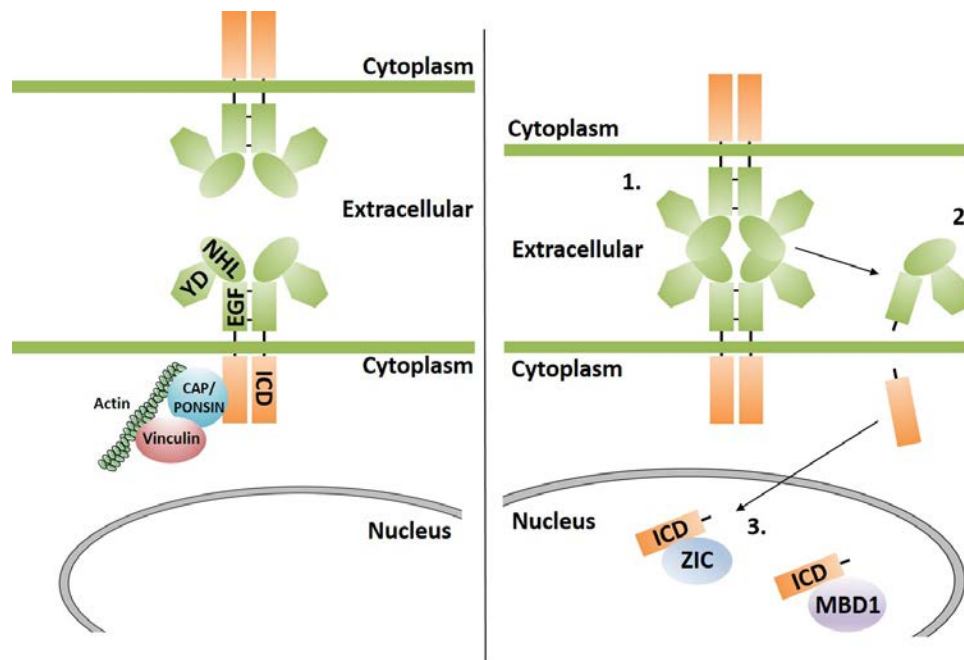
The first teneurin discovered in mice, and vertebrates in general, was teneurin-4, in a screen to identify genes up-regulated by CHOP. It was originally called DOC4, or “downstream of CHOP 4”.<sup>80</sup> Interestingly, the micro-RNA miR-708 is located in a teneurin-4 intron, and is co-regulated by CHOP, together with ten-m4, to control the expression of rhodopsin.<sup>129</sup> Expression studies in mice focused mostly on patterns in neural development, where teneurins are widely expressed.<sup>81,112</sup> Other sites of expression included the trachea and mesodermally-derived tissues of the developing embryo.<sup>112</sup> One very complete study of teneurins-1-4 utilized immunochemistry and in situ hybridization, and described the predominantly distinct yet overlapping expression patterns in the developing and adult nervous systems.<sup>114</sup> Two studies describe non-neural functions of teneurin-4 in the gastrulation of the mouse embryo, where it establishes the anterior-posterior axis and is important for mesoderm-derived tissues. The teneurin-4 mutation led to an arrest of the embryo prior to E6.5.<sup>121,124</sup> Teneurin-4 also plays a role in oligodendrocyte differentiation and myelination of small-diameter axons, by activating focal adhesion kinase (FAK) through phosphorylation.<sup>130</sup> The same group reported that the activation of FAK by teneurin-4 is also responsible for neurite outgrowth and cellular protrusion formation in vitro.<sup>122</sup> Other studies have also investigated teneurins in neurite outgrowth.<sup>83,90,131</sup> Only recently have knockout mouse models of teneurins been described. Ten-m2<sup>116</sup> and ten-m3<sup>117</sup> KO mice have given some important insights into the functions of teneurins during development. The effects were not as drastic, as may have been expected, especially in comparison to the *C. elegans* KOs. However, since four teneurins are expressed in mice, it is possible that redundant expression or up-regulation of another paralog rescues some of the phenotypes. Particularly the developing visual system of the mouse has been investigated in the KO mice. Most importantly, ten-m2 and 3 both regulate ipsilateral projections and are thus

critical for binocular mapping.<sup>116,117,119</sup> Another non-neural function in mice involves teneurins in chondrogenesis. Ten-m3 acts as a positive regulatory factor in the early stage of chondrogenic differentiation<sup>120</sup>, while ten-m4 suppresses the differentiation of chondrocytes.<sup>123</sup>

Teneurins have also been studied in two other vertebrate models, namely in rat and the marsupial wallaby. Neurestin, which corresponds to teneurin-2, is the only ortholog described in the rat thus far. It was cloned from the olfactory bulb, and its expression was described by Northern blot and in situ hybridization. As expected from expression studies in other vertebrates, neurestin/teneurin-2 is also mainly expressed in the developing and adult brain of rats, like the neocortex or the olfactory bulb.<sup>111</sup> In the other vertebrate model, the marsupial wallaby, *Macropus eugenii* the authors compared the role of ten-m3 in binocular vision in the mouse and the wallaby. Ten-m3 affects ipsilateral projections in both animal models, confirming a role in binocular mapping.<sup>117,119,125,126</sup>

Teneurins were the first transmembrane proteins, identified as pair-rule genes. This was very unusual, because all pair-rule genes described to date were transcription factors.<sup>74,75,81</sup> Later it was shown that teneurins can be cleaved at the membrane and translocate to the nucleus, explaining the potential link to transcriptional regulation.<sup>84,86,96</sup> While the evidence of teneurins being pair-rule genes was later refuted, it paved the way towards investigating the ICD in more detail.<sup>76</sup> Since none of the ICDs contain an intrinsic DNA binding sequence, interaction partners are required to regulate transcription. Studies with ICDs have mainly focused on teneurins-1 and -2. Interaction partners of the TEN1-ICD have been identified by several yeast-2 hybrid (Y2H) screens, initially in an avian context. The ICD of chicken teneurin-1 can bind to CAP/Ponsin, linking it to the cytoskeleton. It also translocates to the nucleus in vitro, where it interacts with MBD1, a transcriptional repressor with DNA-binding activity.<sup>84</sup> The chicken teneurin-2 ICD

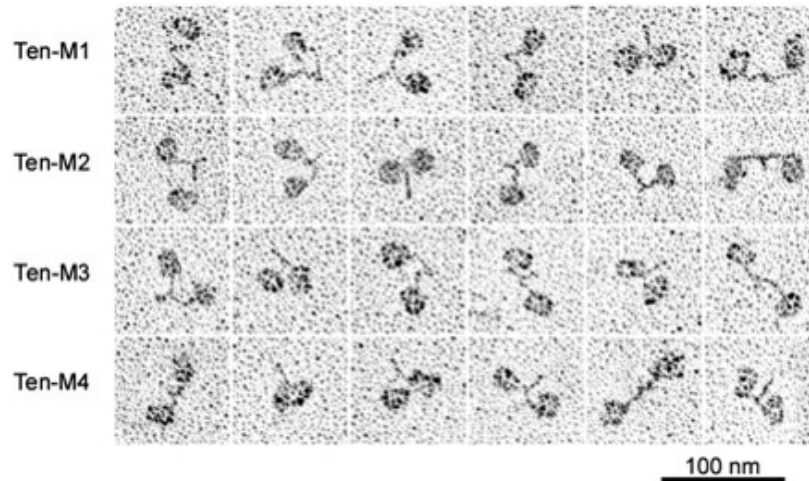
affects Zic-mediated repression and localizes to PML bodies in the nucleus, so-called transcriptional hotspots (Figure 2.17).<sup>96</sup>



**Figure 2.17 Current model of teneurin molecular function (adapted from <sup>94</sup>)**

Teneurins dimerize through the EGF repeats. The ICD can link to the cytoskeleton via CAP/Ponsin binding. 1. Once the ECDs interact homophilically, they get cleaved and released into the extracellular milieu. 2. The ICD gets processed by RIP and translocates to the nucleus. 3. In the nucleus, the ICD interacts with transcriptional regulators like Zic and MBD1.

Several ligands and molecular functions have also been identified for the teneurin ECDs. As previously mentioned, the ECDs of teneurins-1 and -2 can interact homophilically.<sup>88,90</sup> Lasso, a splice-variant of teneurin-2, was identified in the rat brain. Pre-synaptic latrophilin 1 can specifically bind post-synaptic lasso, to form a trans-synaptic complex.<sup>91</sup> Going more in-depth into this interaction than the previous study, another group discusses the combinations of teneurins and latrophilins in trans-synaptic binding, and suggests a function in synapse formation and maintenance.<sup>89</sup>



**Figure 2.18 Ten-m1-4 homodimers<sup>87</sup>**

Rotary shadowing electron microscopic images of purified recombinant ten-m1-4 extracellular domains as homodimers.

It has recently been proposed that the NHL domain, YD-repeats and the adjacent C-terminal peptide of the teneurin ECDs may have a novel function in eukaryotes that would also put previously identified processes of TCAP function into question. Sequence similarity of this region with ABC toxins of the gram-negative bacterium *Yersinia entomophaga* is striking. ABC toxins are generally secreted to target other bacterial or eukaryotic cells. Three proteins (A, B, C) need to assemble for the complex to be toxic. The complex functions either by the A protein binding directly to the target cell, or by forming a pore. The C-terminal region of the C protein is the cytotoxic component. Thus it is possible that the YD-repeats also form an encapsulation device that contains the C-terminal toxic peptide, which is subsequently delivered to the target cell. Potential initiators, delivery systems, and functions would have to be investigated in future studies.<sup>132,133</sup> Transmission electron microscope pictures of murine teneurin dimers show ‘lollipop-like’ structures, which resemble BC toxin dimers (Figure 2.18).<sup>81</sup> The fact that teneurins are ancient proteins and most likely evolved from a bacterial protein through horizontal gene transfer may further support this hypothesis.<sup>78</sup>

Little is known about the regulation of teneurin genes. There is evidence that FGF8 regulates teneurin-2 expression in avian limb buds.<sup>108</sup> Another study suggested teneurins to be potential target genes of Emx2 and Pax6 in the mouse cortex.<sup>113</sup> Human TEN-M1 was later confirmed as a direct target gene of EMX2.<sup>134</sup> In *Drosophila*, ten-m may be regulated by transcription factors Abrupt and Knot.<sup>135</sup>

In summary, teneurins are a family of proteins, involved in the development of the CNS, and some other (adult) tissues. Evidence from the studies above strongly implicate teneurins to have important functions in axon guidance and synaptogenesis. Other functions may include an involvement in the integrity of the basement membrane and chondrogenesis. However, a lot is still left to be learned about teneurins. For example, most molecular mechanisms of function are still unknown.



## 2.8 Teneurins in disease

The teneurin family of proteins has been implicated in several diseases. One of these diseases has a confirmed link, a null mutation in teneurin-3 causing microphthalmia in humans.<sup>136</sup> Teneurin-1 may be one of the proteins involved in X-linked mental retardation (XLMR). TEN-M1 is located on chromosome Xq25, an area of the genome to which several XLMR-related genes have been mapped. Interestingly, some of the symptoms of XLMR include visual impairment and motor sensory neuropathy, which would fit into the context of teneurin function in CNS development.<sup>83</sup>

Some of the binding partners of teneurin ICDs discussed above, may also imply teneurins in disease. The ICD of teneurin-1 directly interacts with MBD1, a protein that is associated with autism.<sup>84</sup> Mutations in *zic1*, an interaction partner of the teneurin-2 ICD, are linked with Dandy-Walker syndrome.<sup>96</sup>

Teneurins have also been implicated in bipolar disorder. A large-scale study of individuals with bipolar disease compared to a control group revealed a new susceptibility locus of bipolar disease near the TEN-M4 gene.<sup>137</sup>

A recent review discusses expression levels of teneurins in different types of tumors and infers a function in human malignancy. The authors used a bioinformatics approach, only focusing on teneurins-2 and -4, which may also be due to a lack of the data that is currently available. Both show altered expression levels in a variety of tumors, like TEN-M2 in hepatocellular carcinomas, and TEN-M4 in brain tumors. It remains to be seen, which mechanisms teneurins are involved in and whether they are oncogenes, tumor suppressors or neither.<sup>138</sup>

### 3. Aim of the thesis

Teneurins are a well-conserved family of proteins that are expressed in many tissues, but especially in the developing CNS of *C. elegans*, *Drosophila*, and vertebrates. An interesting feature of teneurins are the YD repeats located near the C-terminal end of the ECD. YD repeats are unique to eukaryotic proteins, since they are usually expressed in bacterial cell wall proteins of gram-negative bacteria. It not only raises the question of how teneurins acquired this particular domain, but also how this protein family evolved (Chapter 4.1).

The teneurin ICDs are cleaved at the membrane and released to translocate to the nucleus. The transcriptional regulation capabilities of the teneurin ICDs are of particular interest. Previous studies have shown that the teneurin-1 ICD interacts with transcriptional repressor MBD1, and the teneurin-2 ICD affects *Zic*-mediated repression. However, little is known about the molecular mechanisms of teneurin function. We performed two types of unbiased screens, (1) yeast-2 hybrid screen and (2) whole transcriptome analyses, to gain new insight into the functions of teneurin ICDs (Chapter 4.2).

An elegant study utilized the cutting-edge technique AFM-SCFS to determine that chicken teneurins-1 and -2 interact homophilically rather than heterophilically. Domain-swapping experiments between both teneurins showed that the NHL repeat domain is responsible for the specificity of this interaction. Interestingly, NHL repeat domains are predicted beta-propellers. Beta-propellers have previously been implicated in protein-protein interactions, but not in determining the specificity of homophilic versus heterophilic interactions. Hence, we were interested in resolving the structure of the NHL repeat domains of teneurins (Chapter 5).

## 4. Publications

### 4.1 Phylogenetic Analysis of the Teneurins: Conserved Features and Premetazoan Ancestry

Richard P. Tucker, Jan Beckmann, Nathaniel T. Leachman, Jonas Schöler,  
and Ruth Chiquet-Ehrismann

Mol Biol Evol (2012) 29 (3): 1019-1029

First published online: October 31, 2011

My contributions to this paper:

For this study, I cloned and described the splice variants of the human teneurin intracellular domains and performed some initial sequence alignments. Further, I gave input to the manuscript.

# Phylogenetic Analysis of the Teneurins: Conserved Features and Premetazoan Ancestry

Richard P. Tucker,<sup>\*1</sup> Jan Beckmann,<sup>2,3</sup> Nathaniel T. Leachman,<sup>1</sup> Jonas Schöler,<sup>2,3</sup> and Ruth Chiquet-Ehrismann<sup>2,3</sup>

<sup>1</sup>Department of Cell Biology and Human Anatomy, University of California at Davis

<sup>2</sup>Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Basel, Switzerland

<sup>3</sup>Faculty of Science, University of Basel, Basel, Switzerland

\*Corresponding author: E-mail: rptucker@ucdavis.edu.

Associate editor: Billie Swalla

## Abstract

Teneurins are type II transmembrane proteins expressed during pattern formation and neurogenesis with an intracellular domain that can be transported to the nucleus and an extracellular domain that can be shed into the extracellular milieu. In *Drosophila melanogaster*, *Caenorhabditis elegans*, and mouse the knockdown or knockout of teneurin expression can lead to abnormal patterning, defasciculation, and abnormal pathfinding of neurites, and the disruption of basement membranes. Here, we have identified and analyzed teneurins from a broad range of metazoan genomes for nuclear localization sequences, protein interaction domains, and furin cleavage sites and have cloned and sequenced the intracellular domains of human and avian teneurins to analyze alternative splicing. The basic organization of teneurins is highly conserved in Bilateria: all teneurins have epidermal growth factor (EGF) repeats, a cysteine-rich domain, and a large region identical in organization to the carboxy-half of prokaryotic YD-repeat proteins. Teneurins were not found in the genomes of sponges, cnidarians, or placozoa, but the choanoflagellate *Monosiga brevicollis* has a gene encoding a predicted teneurin with a transmembrane domain, EGF repeats, a cysteine-rich domain, and a region homologous to YD-repeat proteins. Further examination revealed that most of the extracellular domain of the *M. brevicollis* teneurin is encoded on a single huge 6,829-bp exon and that the cysteine-rich domain is similar to sequences found in an enzyme expressed by the diatom *Phaeodactylum tricorutum*. This leads us to suggest that teneurins are complex hybrid fusion proteins that evolved in a choanoflagellate via horizontal gene transfer from both a prokaryotic gene and a diatom or algal gene, perhaps to improve the capacity of the choanoflagellate to bind to its prokaryotic prey. As choanoflagellates are considered to be the closest living relatives of animals, the expression of a primitive teneurin by an ancestral choanoflagellate may have facilitated the evolution of multicellularity and complex histogenesis in metazoa.

**Key words:** Teneurin, Odz, Ten-m, evolution, horizontal gene transfer, choanoflagellate, *Monosiga brevicollis*.

## Introduction

Teneurins are phylogenetically conserved transmembrane proteins (see reviews by Tucker and Chiquet-Ehrismann 2006; Young and Leamey 2009). The name “teneurin” honors their discovery in *Drosophila melanogaster* by combining the names of the two dipteran teneurin homologs, Ten-a (Baumgartner and Chiquet-Ehrismann 1993) and Ten-m (Baumgartner et al. 1994, also referred to as Odz [Levine et al. 1994]), with neurons, which are common sites of expression (e.g., Minet et al. 1999). In *D. melanogaster*, chicken, and mouse the teneurin homologs have the following conserved features: 1) teneurins are type II transmembrane proteins with an N-terminal intracellular domain (ICD) and a large extracellular domain (ECD); 2) teneurins have eight epidermal growth factor (EGF) repeats; 3) the third cysteine residue in the second and fifth EGF repeat is replaced with a tyrosine or phenylalanine residue, which results in the potential for teneurins to dimerize side-by-side through disulfide bonds (Oohashi et al. 1999); and 4) the C-terminal two-thirds

of teneurins is similar to the YD-repeat proteins of prokaryotes, with characteristic NHL (from NCL-1, HT2A, and Lin-41) repeats, tyrosine and aspartate-rich YD repeats, and a region similar to the core-associated domain of retrotransposon hot spot (RHS) proteins. In addition, many teneurins can be proteolytically processed, freeing the ICD for transport to the nucleus (Bagutti et al. 2003; Nunes et al. 2005; Kenzelmann et al. 2008) and/or releasing the ECD for incorporation in the extracellular matrix (ECM; Rubin et al. 1999; Trzebiatowska et al. 2008). An additional cleavage site near the C-terminus can lead to the creation of a neuropeptide (reviewed by Lovejoy et al. 2009). Proline-rich Src homology 3 (SH3)-binding domains have been identified in the ICD of teneurins cloned from chordates and ecdysozoans, and ICD-interacting partners have been characterized that may mediate associations between teneurins and the cytoskeleton and methylated DNA (Nunes et al. 2005). Mutation analysis and RNAi-mediated knockdown of teneurin expression in *D. melanogaster* and *Caenorhabditis elegans* reveal fundamental roles for teneurins in

© The Author(s) 2011. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Mol. Biol. Evol. 29(3):1019–1029. 2012 doi:10.1093/molbev/msr271 Advance Access publication October 31, 2011 1019

pattern formation (Baumgartner et al. 1994; Rakovitsky et al. 2007), axonal fasciculation (Drabikowski et al. 2005), and the integrity of basement membranes (Trzebiatowska et al. 2008). In chordates, teneurins are best studied in mouse and chicken, where they are predominantly expressed in the developing nervous system in area-specific patterns mediated in part by *EMX2* (Li et al. 2006; Beckmann et al. 2011). Knockout of the gene encoding mouse teneurin-3 by homologous recombination results in abnormal pathfinding in the visual system and a loss of binocular vision (Leamey et al. 2007).

In order to identify novel features and learn more about the potential evolutionary origin of teneurins, we searched for and compared sequences encoding teneurin-like proteins in opisthokont genomes and collections of expressed sequence tags (ESTs). We also cloned and sequenced cDNAs encoding the ICD of teneurins from human and chicken to study alternative splicing. By aligning and analyzing proteins for predicted nuclear localization sequences (NLSs), SH3-binding domains, and furin-type proteolytic cleavage sites, we have refined our knowledge of conserved teneurin structure and function. In addition, we identified a gene encoding a teneurin in the choanoflagellate *Monosiga brevicollis*, which suggests that teneurins may have played a role in the early evolution of metazoan tissues.

## Materials and Methods

### Sequence Analysis

Novel teneurin sequences were identified by sequence homology using tBLASTn (<http://blast.ncbi.nlm.nih.gov/>) and by domain architecture using Pfam (<http://pfam.sanger.ac.uk/>) with “view a family,” SMART (<http://smart.embl-heidelberg.de/smart/>) with “architecture queries,” and Superfamily (<http://supfam.org/SUPERFAMILY/>) with “domain combinations.” Boundaries of regions, domains, and repeats were determined using Pfam for EGF repeats, NHL repeats, RHS repeats (related to YD repeats), RHS protein, Ten\_N domains, and PfamB PB025792 (the region between the transmembrane domain and the EGF repeats, which was identified as a phylogenetically conserved region by Pfam), and SMART for transmembrane domains and EGF repeats. Alignments and phylogenetic relationships were determined using ClustalW (<http://www.genome.jp/tools/clustalw/>) and the settings “pair alignment slow/accurate” (gap open penalty 10, gap extension penalty 0.1). Importin  $\alpha/\beta$  pathway NLSs were identified using NLS Mapper (<http://nls-mapper.iab.keio.ac.jp/>), and furin cleavage sites were predicted with ProP (<http://www.cbs.dtu.dk/services/ProP/>). SH3-binding domains were identified by hand from consensus sequences described by others (Kay et al. 2000; Mayer 2001; Kowanetz et al. 2003).

### Reverse Transcriptase PCR and Sequencing

Human adult brain cDNA was generated out of total human brain RNA (AMS Biotechnology, Oxon, UK) using Superscript III (Invitrogen, Carlsbad, CA) polymerase and random hexamer primers (Invitrogen). Sequences corresponding to the ICDs of teneurins-1 through -4 were

amplified with PfuTurbo polymerase (Stratagene/Agilent Technologies, Santa Clara, CA) using specific primers (teneurin-1: 5'-ACTAGCGGCCGCACCATGGAGCAAAGTACTGACTGC-3'/5'-ACTACTCGAG GCAGCACCTGTAAGGTTTG-3'; teneurin-2: 5'-ACTAGCGGCCGCACCATGGATGTAAGAGACCCG-3'/5'-ACTACTCGAGGCAGTATTTGGAGGGCTTC-3'; teneurin-3: 5'-ACTAGCGGCCGCACCATGGATGTGAAAGAACGC-3'/5'-ACTACTCGAGACAGTACTTTGAAGACTTC-3'; teneurin-4: 5'-ACTAGCGGCCGCACCATGGAGGTGAAGGAGAGG-3'/5'-ACTACTCGAGACAGTACTTGGAGGGCTTC-3') including restriction sites NotI and XhoI. Amplified products were separated on a 0.8% agarose gel, and fragments were excised, gel extracted, and cloned into pcDNA3. Positive clones were sequenced using forward primer T7 and reverse primer Sp6.

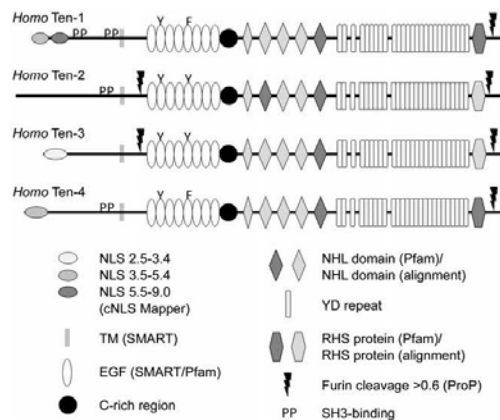
The sequence of avian teneurin-3 including alternatively spliced variants was assembled from overlapping fragments cloned by polymerase chain reaction (PCR). cDNA was prepared from total RNA extracted from embryonic day 16 chicken cerebellum using the RNeasy Mini kit (Qiagen, Germantown, MD). PCR was performed with the Platinum Pfx DNA Polymerase System (Invitrogen). Five sets of primers were used to divide avian teneurin-3 into five segments. Segment 1 used primer pair 5'-ATGGATGTGAAAGAAGTCCG-3'/5'-CACGTGGAGGGTAAACGATAA-3'; segment 2 used primer pair 5'-ACTGTGAAAGACGGATTGC-3'/5'-GACCCCAAAAGTCACTAGA-3'; segment 3 used primer pair 5'-TGATGGGACCATGATCAGAA-3'/5'-ACCAGACCGCAGACATGAAC-3'; segment 4 used primer pair 5'-AGCGAGGGACGACTAGTAA-3'/5'-GGAGAAAGGATAGAGTGAAA-3'; and segment 5 used primer pair 5'-AGGCTGTGACAGAAAGGAGA-3'/5'-GGTCCTACTTGGATGACT-3'. Each segment was TOPO cloned into the pCR-II vector (Invitrogen) for sequencing.

## Results

### Overview: Identification of Teneurins and Analysis of Teneurins from *Homo sapiens*

Genes encoding teneurin-like proteins and predicted proteins with the characteristic domain organization of known teneurins were identified by sequence similarity (e.g., tBLASTn) and by the presence of combinations of domains (e.g., predicted proteins with both EGF repeats and NHL repeats using Pfam or SMART; for details, see Materials and Methods). Teneurins identified in this way from chordates are summarized in [supplementary table S1, Supplementary Material online](#); teneurins from nonchordates are summarized in [supplementary table S2, Supplementary Material online](#).

To illustrate the features of these teneurins identified through proteomic analysis, the four teneurins from *H. sapiens* are shown in [figure 1](#). The variant of teneurin-1 shown in [figure 1](#) is a type II transmembrane protein with a 317aa N-terminal ICD, a 23aa transmembrane domain, and a 2385aa C-terminal ECD. Within the ICD, there are two predicted importin  $\alpha/\beta$  pathway NLSs. The first NLS (aa11-40) has an NLS Mapper score of 4.8, and the second NLS (aa60-69) has an NLS Mapper score of 6.0. Higher



**FIG. 1.** The human teneurins. Human teneurins share a basic domain organization but they differ in the presence of NLSs, SH3-binding motifs, and furin cleavage sites. Following processing the ICD of teneurin-1, for example, is predicted to be located in the nucleus but not the ICD of teneurin-2. Similarly, the ECD of teneurins-2 and -3 are predicted to be shed into the ECM but not the ECD of teneurins-1 and -4. Accession numbers, the positions of domains as well as the NLS and ProP scores of each teneurin are found in [supplementary table S3, Supplementary Material online](#).

scores represent a greater likelihood of nuclear localization (Kosugi et al. 2009) and are indicated on the figure with progressively darker ovals. Two proline-rich SH3-binding domains (indicated by “PP” on [fig. 1](#)) are also found in the ICD of human teneurin-1. The first (aa193-199; RPLPPP) is consistent with the consensus sequence for Class I SH3 ligands (+xφPxφP); the second (aa292-297; PRPLPR) is consistent with the atypical SH3-binding motif (PxxxPR) of cbl proteins (Kowanetz et al. 2003). In the ECD of human teneurin-1, there are eight EGF repeats (aa531-796). The third cysteine residue of the second EGF repeat has been replaced with a tyrosine residue, and the third cysteine residue of the fifth EGF repeat has been replaced with a phenylalanine (indicated by the “Y” and “F” in [fig. 1](#)). This substitution results in the potential for dimerization of teneurin monomers through disulfide bonds between cysteines that lack an intramodular partner (Oohashi et al. 1999). A cysteine-rich domain is found adjacent to the eighth EGF repeat (aa797-836). The carboxy two-thirds of human teneurin-1 shares the same domain organization as the YD-repeat proteins of some prokaryotes (e.g., *Myxococcus xanthus*, where it is required for gliding motility [Youderian and Hartzell 2007]); five NHL domains, YD-repeats (similar to “RHS repeats”), and a region near the C-terminal tail identified by Pfam as RHS protein (similar to “RHS-associated core domain”). The NHL domains of human teneurin-1 were identified by Pfam (dark gray, [fig. 1](#)) or by alignment using ClustalW (light gray, [fig. 1](#)). Similarly, the RHS protein domain identified by Pfam in human teneurin-1 is indicated in dark gray, and those identified in other teneurins by alignment are indicated by lighter shades. Finally,

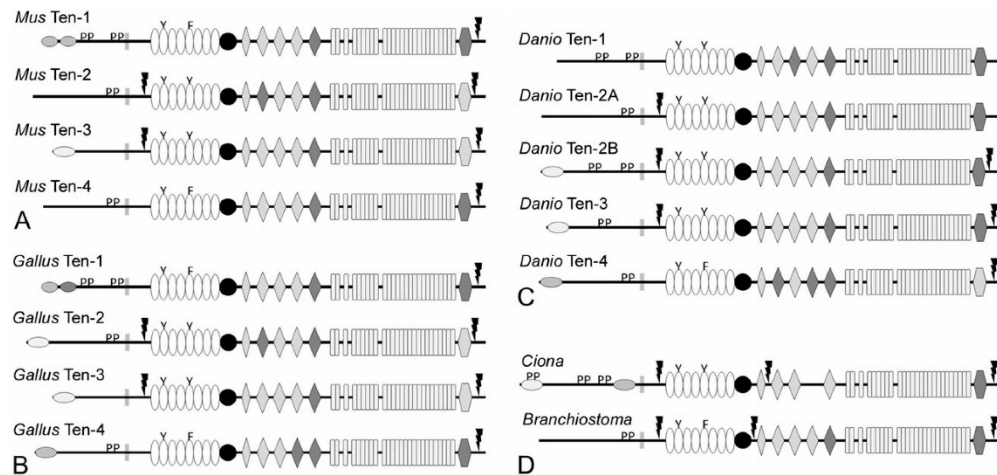
human teneurin-1 has a single predicted furin cleavage site with a ProP score at or above 0.55 (threshold = 0.50) at aa2618 (LNGRTRR/FA). This would create a 107aa C-terminal peptide similar to teneurin C-terminal-associated peptide-1 (Trubiani et al. 2007).

There are four teneurin genes in *H. sapiens*. The basic organization of teneurins-1, -2, -3, and -4 is the same, most notably in the ECD: each teneurin has eight EGF repeats with aromatic residues substituting for cysteines in the second and fifth repeat, each has a cysteine-rich domain and a C-terminal region similar to the YD-repeat proteins of prokaryotes, and each has a predicted furin cleavage site near the C-terminus ([fig. 1](#); details can be found in [supplementary table S3, Supplementary Material online](#)). One difference is the presence of a second predicted cleavage site between the transmembrane domain and EGF repeats of teneurins-2 and -3 that would permit shedding of the ECD into the ECM; these cleavage sites are not found in teneurins-1 and -4. Additional differences are seen in the ICD. Teneurins-1, -2, and -4 have proline-rich motifs that match consensus SH3-binding domains but teneurin-3 does not. However, the proline-rich sequence PPTRPLPR is found in the ICD of human teneurin-3, which resembles, but does not exactly match, known SH3-binding motifs. Teneurin-2 does not have a predicted NLS, and the NLS of teneurin-3 has a lower NLS Mapper score than the NLSs of teneurin-1 and teneurin-4.

Representative teneurins from major taxonomic groups were analyzed in this manner and are described below.

#### Identification and Analysis of Chordate Teneurins

The teneurins of mouse (*Mus musculus*), chicken (*Gallus gallus*), zebrafish (*Danio rerio*), and the protochordates *Ciona intestinalis* and *Branchiostoma floridae* were identified and analyzed. There are four teneurins in *M. musculus* and *G. gallus* and they share many of the features described above for human teneurins. The few differences include the absence of a predicted NLS in murine teneurin-4, and the observation that chicken teneurin-2 has a predicted NLS, albeit a weak one ([fig. 2A and B](#); [supplementary table S3, Supplementary Material online](#)). In *D. rerio*, there are five teneurins. ClustalW alignment and basic phylogenomic analysis identify two teneurin-2 paralogs that we have named teneurin-2A and teneurin-2B ([figs. 2C and 3](#); [supplementary table S3, Supplementary Material online](#)). The predicted sequences of teneurin-1, teneurin-2B, and teneurin-4 appear to be complete, but the predicted N-termini of teneurin-2A and teneurin-3 were completed by translating potential open reading frames and aligning them with known teneurin sequences and by piecing together ESTs ([supplementary table S3, Supplementary Material online](#); FASTA files can be found in [supplementary table S4, Supplementary Material online](#)). As with the chicken teneurins, the basic features of the zebrafish teneurins are conserved. Differences include the absence of a potential furin cleavage site near the C-terminus of teneurin-1, an additional potential furin cleavage site between the NHL repeats and YD repeats of teneurin-2A and the absence of predicted NLSs



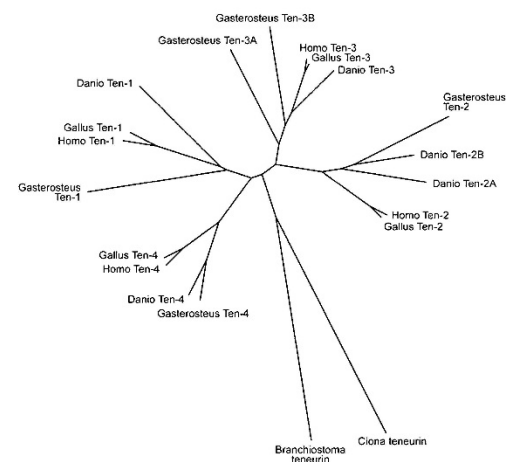
**FIG. 2.** The domain organization and possible relationships of chordate teneurins. There are four teneurins in mouse (*Mus musculus*) and chicken (*Gallus gallus*). Basic features are highly conserved between mouse (A), chicken (B), and human teneurins (fig. 1). There are five teneurin genes in the zebrafish, *Danio rerio*, including two teneurin-2 paralogs. The ICD of the predicted teneurin-1 does not have an NLS, and an additional furin cleavage site is found between the NHL domains and the YD repeats (C). The genomes of the tunicate *Ciona intestinalis* and the cephalochordate *Branchiostoma floridae* each contain a single teneurin gene (D).

in the ICDs of teneurin-1 and teneurin-2A. In addition, the ICD of *D. rerio* teneurin-3 has a predicted proline-rich SH3-binding domain, unlike the ICDs of teneurin-3 in chicken and man.

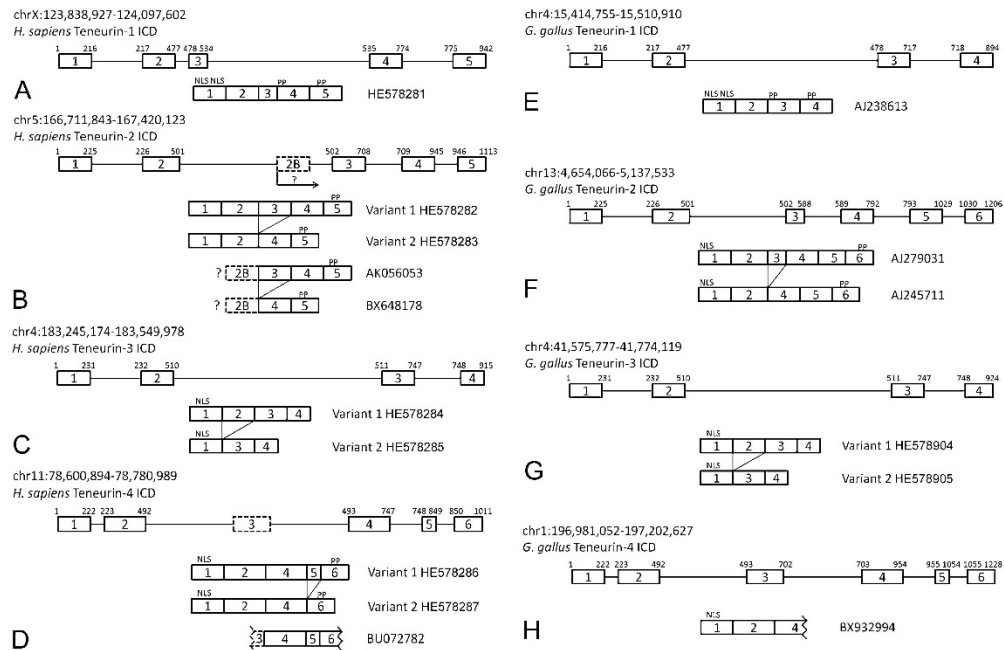
To determine if the duplication of teneurin-2 is a common feature in bony fish, the teneurins of the stickleback *Gasterosteus aculeatus* were also identified (supplementary table S1, Supplementary Material online) and aligned with the teneurins of other chordates. Like *D. rerio*, *G. aculeatus* has five teneurin genes. However, there are two teneurin-3 paralogs (teneurin-3A and teneurin-3B) and only one teneurin-2 (fig. 3). The teneurin-1 of *G. aculeatus* has a potential furin cleavage site near the C-terminus (aa2642; ProP score = 0.74), indicating that the absence of this site in *D. rerio* may not be typical of teneurin-1 in actinopterygians. The *G. aculeatus* teneurin-1 also has a predicted NLS in the ICD (aa11-41, NLS Mapper score = 4.0).

The genomes of *C. intestinalis* and *B. floridae* each encode a single teneurin (fig. 2D, supplementary table S3, Supplementary Material online). The basic organization of these teneurins resembles those of the craniate teneurins. The ICD of the predicted teneurin from *C. intestinalis* has two NLSs: one near the N-terminus and the other near the transmembrane domain. A predicted NLS near the transmembrane domain is commonly observed in the ICD of teneurins from protostomes (see below). The ICDs from both of the protochordate teneurins have predicted SH3-binding domains, but the ICD from *B. floridae* does not have an NLS that is recognized by NLS Mapper. Both of the predicted protochordate teneurins have potential furin cleavage sites that would shed the ECD and process the C-terminus like those of teneurin-2 and teneurin-3 in fish,

birds, and man, but like *D. rerio* teneurin-2A (and unlike other craniate teneurins examined) they also have a third predicted furin cleavage site near the center of the ECD. A second teneurin-like sequence is found in the *B. floridae* genome when two adjacent predicted proteins (XP\_002592160 and XP\_002592161) are combined. However, the C-terminal two-thirds of the second predicted protein also



**FIG. 3.** An unrooted phylogenetic tree based on ClustalW alignment of real and predicted teneurin amino acid sequences suggests that teneurins-1 and -4 share a common ancestor, as do teneurins-2 and -3. The stickleback *Gasterosteus aculeatus* has five teneurins, but unlike *Danio rerio* it has retained two teneurin-3 paralogs.



**FIG. 4.** Comparison of alternative splicing in the ICD of human and chicken teneurins. The human teneurin-1 ICD is encoded on five exons. No splice variants were found using RT-PCR and primers corresponding to sequences in the first and fifth exon of human teneurin-1 (A). In contrast, RT-PCR reveals two variants amplified with primers based on sequences in the first and fifth exon of human teneurin-2. Variant 1 is encoded on 5 exons, whereas Variant 2 is composed of four. ESTs suggest an additional way to generate diversity in the ICD of human teneurin-2: An exon found between the second and third exon (exon 2B) is associated with long and short variants as well. This may represent an alternative start site for teneurin-2 (B). Human teneurin-3 has an ICD encoded on four exons, and exon 2 can be spliced out to generate a second variant (C). RT-PCR with primers based on sequences in exon 1 and exon 6 reveal long and short variants of human teneurin-4. Variant 1 is encoded on five exons, whereas Variant 2 is encoded on four. Use of an additional exon (exon 3) that does not contain a start codon is found in an EST (D). The ICDs of teneurins-1 and -2 from the chicken were described previously. As in human, there is a single teneurin-1 isoform, and there are long and short variants of teneurin-2. There is no evidence from chicken of an alternative start site in teneurin-2 (E, F). PCR using cDNA from embryonic chicken cerebellum and primers based on sequences in exons 1 and 4 reveal identical variants in chicken and human teneurin-3 (G). Sequence from a single EST from the chicken is consistent with the organization of the ICD of human teneurin-4 (H).

has lysozyme and keratin-related sequences, and some teneurin and lysozyme-like sequences are encoded on the same large predicted exon, which leads us to suggest that this is a pseudogene. This is supported by the total absence of ESTs.

#### Alternatively Spliced Variants

Previously we showed that there are a number of isoforms of avian teneurin-2 and that these variants are derived from alternative splicing of regions of transcripts encoding both the ICD and the ECD (Tucker et al. 2001). Here, we examined ESTs and used RT-PCR to determine if the ICD variants are specific for teneurin-2 and if they are conserved in mammals and birds. A single PCR product is amplified from adult human brain-derived cDNA using primers corresponding to the 3' end of the first exon of the human teneurin-1 gene and the 5' end of the fifth exon, which encodes the transmembrane domain (fig. 4A). When the same strategy is applied using primers based on human teneurin-2 sequences, two variants are observed: Variant

1 is encoded by all five previously identified exons and Variant 2 lacks the third exon (fig. 4B). These variants are supported by EST data, which also reveal the use of a sixth exon found between exon 2 and exon 3. EST sequences containing this alternative exon, which we have named exon 2B, do not contain sequences corresponding to either exon 1 or exon 2. As a putative start codon is found in exon 2B, this exon may be used as an alternative start site for teneurin-2 transcripts (and therefore would not have been amplified using our flanking primer pairs). The ICD of teneurin-3 is encoded on four exons and like teneurin-2 there are two ICD splicing variants: Variant 1 uses all four exons, whereas Variant 2 is encoded on exons 1, 3, and 4 (fig. 4C). Finally, RT-PCR reveals two alternatively spliced variants of the human teneurin-4 ICD. The larger is encoded by five exons, and a smaller by four exons. Interestingly, an EST (BU72782) shows the potential use of an additional exon that was not amplified by our primer pair (fig. 4D).

ESTs demonstrate that some of the ICD variants we observed in human teneurin-1 and teneurin-4 are conserved



in *G. gallus*, just as our previous work with avian teneurin-2 showed the presence of two ICD variants (fig. 4E, F, and H). To study the alternative splicing of teneurin-3, we used RT-PCR to amplify products corresponding to the ICD and cDNA derived from embryonic chicken brain. Just as in human, the avian teneurin-3 ICD is encoded on four exons. A large variant contains sequences corresponding to all four exons, and exon 2 is spliced from a smaller variant (fig. 4G). Note that we could not identify an exon homologous to exon 2B of human teneurin-2 in the chicken genomic sequence, but there is a homologous potential exon 3 in chicken teneurin-4 DNA.

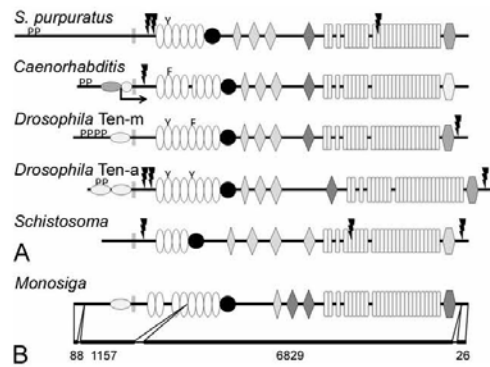
There is also evidence of teneurin variants derived from alternative splicing in the region encoding the ECD. For example, a short (8aa) stretch of amino acids can be present between the seventh and eighth EGF repeats of teneurin-2 from the chicken, and at least one variant of avian teneurin-2 is truncated between the seventh and eighth EGF repeats, resulting in an isoform lacking the cysteine-rich domain and the region homologous to the YD repeat proteins of prokaryotes (Tucker et al. 2001). Alternative splicing that results in additional sequence between the seventh and eighth EGF repeats may be common in teneurins, as similar variants are found in mRNA sequences in mouse teneurin-3 (NP\_035987) and teneurin-4 (BAE28005). However, there is no evidence supporting grossly truncated isoforms of teneurins in other species.

#### Identification and Analysis of Teneurins from an Echinoderm and Protostomes

The same methods used to identify teneurins in chordates were applied to other metazoan sequences. In addition to the known teneurins of *D. melanogaster* and *C. elegans*, predicted complete or partial teneurins were found in the purple sea urchin (*Strongylocentrotus purpuratus*), a mollusk (*Lottia gigantea*), an annelid (*Capitella teleta*), a trematode (*Schistosoma mansoni*), and a wide variety of nematodes and arthropods (supplementary table S2, Supplementary Material online). Interestingly, no teneurin-like sequences were identified in the genomes or ESTs from cnidarians, ctenophores, sponges, or *Trichoplax adhaerens*.

The single teneurin from *S. purpuratus* is remarkable for a few features not seen in teneurins from chordates: 1) it has only six EGF repeats and only the second EGF repeat has an aromatic residue substituting for a cysteine residue; 2) it lacks a predicted furin cleavage site near the C-terminus; and 3) it has two predicted furin cleavage sites between the transmembrane domain and the EGF repeats (fig. 5A). Like the teneurins from protochordates, it has a predicted furin cleavage site near the center of the ECD.

The teneurins from *C. elegans* and *D. melanogaster* are well known, and the sequences that were analyzed here came from cDNAs. There are two teneurins from *C. elegans*, Ten-1L and Ten-1S; they are encoded by the same gene, but two different promoters regulate the expression of "long" and "short" transcripts (Drabikowski et al. 2005). Ten-1L is illustrated in figure 4A; Ten-1S would be identical except the ICD is much smaller (the approximate location of



**FIG. 5.** The domain organization of nonchordate teneurins. The ECDs of teneurins from the purple sea urchin (*Strongylocentrotus purpuratus*), *Caenorhabditis elegans*, *Drosophila melanogaster*, and the trematode *Schistosoma mansoni* have the same basic organization as chordate teneurins, but there is variation in the number of EGF repeats. Aromatic residues substitute for cysteines in at least one EGF repeats in all the species examined except the fluke, which suggests that the teneurin from *S. mansoni* does not dimerize (A). The genome of the choanoflagellate *Monosiga brevicollis* encodes a protein with a domain organization that is identical to metazoan teneurins. The predicted protein does not have furin cleavage sites or SH3-binding domains, and its EGF repeats contain a full complement of cysteines. The predicted *M. brevicollis* teneurin is encoded on just four exons, and most of the ECD is encoded on a single mega-exon of 6829 bp (B).

the N-terminus of Ten-1S is indicated by the crooked arrow in fig. 5A). Ten-1L and the two *D. melanogaster* teneurins, Ten-m and Ten-a, have putative SH3-binding domains and one or more NLS. Unlike the NLSs of most chordate teneurins, the NLSs from the ecdysozoans tend to be found near the transmembrane domain and not at the N-terminus. The ECDs of these teneurins are similar to those found in chordate teneurins: note that both Ten-m and Ten-a (but not Ten-1L) have potential furin cleavage sites near the C-terminus, and both Ten-1L and Ten-a have potential furin cleavage sites that could shed the ECD into the ECM. The fifth EGF repeat of *C. elegans* Ten-1L is incomplete; the part of this repeat encoding both the second and third cysteine residues in other teneurins is missing. Also, the part of the ECD near the C-terminus that is predicted by Pfam to be homologous to the RHS core-associated protein domain is more unlike this domain in other teneurins, though some identity could be found by alignment.

Two teneurin genes that encode predicted proteins that align with either *D. melanogaster*'s Ten-a or Ten-m were identified in the genomes of a number of insect species, including *Apis mellifera* (honey bee), *Tribolium castaneum* (flour beetle), and the mosquitoes *Aedes aegypti* and *Culex quinquefasciatus* (supplementary table S2, Supplementary Material online). However, single teneurin genes were found in the genomes of the branchiopod crustacean *Daphnia pulex* and the arachnid *Ixodes scapularis* (deer tick). This suggests that the duplication of teneurins in

**Table 1.** Alignment of Representative Sequences with the Cysteine-Rich Domain Core Sequence of *Monosiga brevicollis* Teneurin.

Species (common name)	Core Sequence	% Id	% Sim
<i>Monosiga brevicollis</i> (choanoflagellate)	CNDGIDNDNDRVTDCNDADCCSS		
Teneurin			
<i>Phaeodactylum tricornutum</i> (diatom)	CNDGIDNDNDGLFDCEDPDCAND		
Endo-1,3-beta-glucosidase	***** : **:*,**...	65%	91%
<i>Schistosoma mansoni</i> (flake)	CDDGIDNDHDDLVDCLDPDCCTS		
Teneurin	*:*****:* :_*_*_*_*_*_*_*_*	65%	91%
<i>Caenorhabditis elegans</i> (roundworm)	CDDGLDNDSDGLIDCDDPECCSS		
Ten-a	*:*_*_*_*_*_*_*_*_*_*_*_*_*_*_*	61%	91%
<i>Homo sapiens</i> (human)	CGDNLNDNDGGLTDCVDPDCQQQ		
Teneurin-1	*_*_*_*_*_*_*_*_*_*_*_*_*_*_*_*	57%	91%
<i>Volvox carteri</i> (volvox)	CDDGIDNDNDGLVDMDDPDCNTS		
Gametolysin	*:***** * :_*_*_*_*_*_*_*_*	57%	83%
<i>Strongylocentrotus purpuratus</i> (purple sea urchin)	CTDEVNDGDSLIDCEDPDCCLS		
Teneurin	*_*_*_*_*_*_*_*_*_*_*_*_*_*_*_*	57%	83%
Pfam: Cu-binding_MopE PF11617*	C.DGVNNDGGLTDCVDPDCQQQ		
	*_*_*_*_*_*_*_*_*_*_*_*_*_*_*_*	54%	77%

NOTE.—\*The Cu<sup>++</sup>-binding consensus domain of MopE and related proteins.

the protostome lineage is limited to insects and is not a feature of all arthropods.

The trematode *S. mansoni* has a single predicted teneurin (fig. 5A; supplementary table S3, Supplementary Material online) that has a number of distinctive features. Its ICD is relatively short, and it does not contain predicted SH3-binding motifs or an NLS. Like many teneurins it has a putative furin cleavage site between the transmembrane domain and the EGF repeats, but it also has a second predicted furin cleavage site amidst the YD repeats. Finally, the fluke teneurin has only four EGF repeats, and all four EGF repeats have a full complement of cysteine residues, so unlike other teneurins studied to date it probably fails to dimerize.

#### A Teneurin Is Encoded in the Genome of the Choanoflagellate *M. brevicollis*

The absence of teneurin genes in the complete and assembled genomes of the cnidarian *Nematostella vectensis* and placozoan *T. adhaerens* (that could be identified using the search methods employed to find teneurins in other metazoans) initially suggested that teneurins may have evolved about the time of the Cambrian radiation. However, during a routine search of predicted protein domain architectures that included RHS core-associated protein domains using the Pfam program, a sequence encoding EGF repeats, NHL domains, and YD repeats (in addition to the RHS core-associated protein domain) was identified in the genome of the choanoflagellate *Monosiga brevicollis*. Further analysis revealed that this predicted protein has the basic features of a metazoan teneurin: it is a type II transmembrane protein with a putative NLS in the ICD, eight EGF repeats, a cysteine-rich domain, and a C-terminal two-thirds with the same domain architecture as a prokaryotic YD-repeat protein (i.e., NHL domains, YD repeats, and an RHS core-associated protein

domain; fig. 5B; supplementary table S3, Supplementary Material online). The predicted sequence (XP\_001749414) is shown in its entirety in supplementary figure S1, Supplementary Material online, together with relevant alignments generated with ClustalW. The expression of the *M. brevicollis* teneurin is supported by two nonoverlapping ESTs (FE890769 and FE895158), both of which correspond to regions encoding the YD repeats.

The ICD of the *M. brevicollis* teneurin does not align significantly with the ICDs from other teneurins, and it lacks SH3-binding motifs. In addition, ProP fails to identify any potential furin cleavage sites in this teneurin. There are eight EGF repeats, but there are no free cysteines to support dimerization. Adjacent to the EGF repeats is a cysteine-rich region that is highly conserved: the exact 23aa consensus sequence ExxCx(D/N)xxDx(D/E)xDxxxDCxxx(D/E)CCxxxCxxxxC is found in all the teneurins analyzed except for *S. purpuratus* teneurin (which has one additional "x" between the fourth and fifth cysteine) and *S. mansoni* teneurin (which is missing the sixth cysteine). In fact, using the *M. brevicollis* cysteine-rich domain sequence in a tBLASTn search of all nucleotide sequences uncovers all the teneurins identified above that are listed on GenBank. However, neither this method nor the other search methods we employed to identify teneurin sequences revealed teneurins in sequences from sponges, placozoans, ctenophores, cnidarians, fungi, ichthyospores or nucleariids. Interestingly, a similar cysteine-rich sequence is found in an endo-1,3-beta-glucosidase from the diatom *Phaeodactylum tricornutum* (XP\_002181321). This sequence is 46% identical and 71% similar to the 35aa cysteine-rich domain of *M. brevicollis* and it includes a core stretch of 23aa that is 65% identical and 91% similar (table 1). This 23aa core domain aligns well with the Cu<sup>++</sup>-binding motif of MopE (Helland et al. 2008), and similar sequences are found in

**Table 2.** Alignment of Most Similar Sequences with YD Repeats from *Monosiga brevicollis* Teneurin<sup>a</sup>

<i>M. brevicollis</i> <i>Syntrophobacter fumaroxidans</i> <sup>b</sup>	YDVDGQLTQVLEDEGAEVE SYSYDVNCRVAVNWVRG---AAHSATYGADDA YDSLGRLLAVRLDGVPAEEYRYDVNGNRVEETNTPRGIIGRTSTYSEEDH ** *:* * * * . * . * * * * * * * * . . . : : * * . : *
<i>M. brevicollis</i> <i>S. fumaroxidans</i>	VFTVDGQSYAVDVGFLTSVRG---MSLAYSGRGELLSATLPSGAGTVR LLTSGGTVYRYDADGFLTRTEGSAVTRYVYSSRGELLSVALPDGK-RIE : : * . * * * * . * * * * * : . * * . * * * * * . : * * . : .
<i>M. brevicollis</i> <i>S. fumaroxidans</i>	YRYDGFGRRI YVNDPLGRRI 46% identical/69% similar * * : * * * *
<i>M. brevicollis</i> <i>Desulfococcus oleovorans</i> <sup>c</sup>	YDVDGQLTQVLEDEGAEVE SYSYDVNGN----RVAWNVRGAAHSATYGAD YDEMGRLTIVTKDGLTVE SYSYDSTPYGTCTYQMNTLRGIAGRVLDYDAE ** *:* * * : * : * * * * * * * . . : : * . : * . * :
<i>M. brevicollis</i> <i>D. oleovorans</i>	DAVFTVDGQSYAVDVGFLTSVRG---MSLAYSGRGELLSATLPSGAGTVR DHLISAGGTDYQYDLGFLTSKTSGAETTYDYSSRGELLSVDLPDGT-D * : : . . * . * * : * * * * * . * * . * * * * * . * * . * :
<i>M. brevicollis</i> <i>D. oleovorans</i>	VRYRYDGFGRRI ITYVHDPLGRRI 43% identical/67% similar : * * : * * * *
<i>M. brevicollis</i> <i>Homo sapiens Ten-4</i>	YDVDGQLTQVLEDEGAEVE SYSYDVNCRVAVNWVRGAAHS--ATYGADDAV YDADGQLQTVSINDKPLWRYSYDLNGLHLLSPGNSARLTPLRYDIRDRI ** . * * * * * : . : * * * * * * * . : * : * . * :
<i>M. brevicollis</i> <i>H. sapiens Ten-4</i>	FTVDGQSYAVDVGFLTSVRGMSLAYSGRGELLSATLPSGAGTVRYRYD TRLGDVQYKMDDEGFLRQGGDIFBYNSAGLLIKAYNRAGSWSVRYRYD : . . * * : * * * * * . * : * . . * * * : * : * * * * * :
<i>M. brevicollis</i>	FGRRRI LGRRV 38% identical/63% similar : * * :

<sup>a</sup> As determined by tBLASTn with aa2276–aa2379 of *M. brevicollis* teneurin against the NCBI Nucleotide Collection database and aligned with ClustalW 2.1.

<sup>b</sup> YD protein from genomic sequence CP000478 (4941818–4942148).

<sup>c</sup> YD protein from genomic sequence CP000859 (1374170–1373806).

the metal-binding region of a predicted gametolysin from *Volvox carteri* (table 1 [XP\_002958497]) and *Chlamydomonas reinhardtii* (XP\_001695639). Alignment of the 23aa motif reveals that the core of the cysteine-rich region of *M. brevicollis* teneurin is most similar to the diatom sequence and the core of the cysteine-rich region of trematode teneurin; the same regions from other metazoan teneurins are conserved but not to the same extent (table 1).

The NHL repeats and YD repeats of *M. brevicollis* align better with the YD-repeat proteins of some prokaryotes than with the YD repeats found in metazoan teneurins. The NHL domains are most similar (31% identical) to the YD-repeat protein of *Herpetosiphon aurantiacus* (ABX04679), a predatory filamentous chloroflex bacterium that lives in soil and freshwater. A stretch of 103aa corresponding to YD re-

peats 17–21 of *M. brevicollis* teneurin was analyzed further using tBLASTn and the entire NCBI nucleotide collection. This stretch is most similar to the YD-repeat proteins of *Syntrophobacter fumaroxidans* (a freshwater bacterium) and *Desulfococcus oleovorans* (which lives in coastal waters) and then to the YD repeats found in human teneurin-4 (table 2).

The *M. brevicollis* teneurin is predicted from sequences encoded on just four exons that are separated by introns of 129, 206 and 105 bp (in contrast, human teneurin-1 is encoded on 29 exons and the average intron is 8 kb). Remarkably, the region of the predicted protein corresponding to the four C-terminal EGF repeats, the cysteine-rich domain, the NHL repeats, the YD repeats, and the RHS core-associated protein domain is encoded on a single giant exon of 6829 bp (fig. 5B). For comparison, the

corresponding regions of human teneurin-1, *S. mansoni* teneurin, and *C. elegans* Ten-1L are encoded on 20, 21, and 7 exons, respectively (see also Minet and Chiquet-Ehrismann 2000).

## Discussion

Here, we have used predictions based on proteomics to determine which teneurins may be processed such that the ECD becomes incorporated into the ECM and which teneurins may be processed such that the ICD is transported to the nucleus. Our predictions are validated by our previous experimental studies with avian teneurins. For example, we showed that a recombinant fusion protein with the ECD of chicken teneurin-2 was cleaved *in vitro* at a furin site between the transmembrane domain and the EGF repeats (Rubin et al. 1999). Consistent with this observation, we also showed that antibodies to the ECD of chicken teneurin-2 not only labeled the cell surface but also the ECM of chicken embryos (Tucker et al. 2001). When tagged chicken teneurin-2 ICD is overexpressed in HT1080 cells the recombinant ICD localizes to the nucleus (Bagutti et al. 2003), but there is no evidence published to date that the teneurin-2 ICD is processed and transported to the cell nucleus *in vivo*. In contrast, antibodies to the ECD of chicken teneurin-1 failed to stain the ECM, but antibodies to the ICD of teneurin-1 routinely stained the nuclei of cells *in vitro* and in histological sections of embryos (Kenzelmann et al. 2008). Moreover, when the sequence RKRK in the avian teneurin-1 ICD is mutated to AAAA it no longer localizes to the nucleus *in vitro* (Kenzelmann et al. 2008). Here, we show that the ICD of chicken teneurin-1 (specifically, the RKRK and flanking sequences) is predicted with a high likelihood to be located in the nucleus (NLS Mapper score = 9.0) and that the ICD of chicken teneurin-2 is much less likely to be nuclear (NLS Mapper score = 2.7). Similarly, the chicken teneurin-2 furin-cleavage site that we previously demonstrated to be functional is predicted by ProP to be active (score = 0.65), but no such site is found in chicken teneurin-1. Thus, teneurin-1 and/or teneurin-4 are most likely to be processed (by a yet unknown mechanism) so that the ICD can move to the nucleus, and teneurin-2 and/or teneurin-3 are more likely to have the ECD shed into the ECM. The shared features of these pairs of teneurins are also reflected in their predicted origins: teneurins-1 and -4 appear to have evolved from a gene duplication, as do teneurins-2 and -3 (fig. 3; see also Minet and Chiquet-Ehrismann 2000). All the chordate teneurins examined here except *D. rerio* teneurin-1 are likely to be cleaved near the C-terminus. This may be a step that precedes the formation of the teneurin-derived C-terminal neuropeptides characterized by others (reviewed by Rotzinger et al. 2010).

Using a yeast two-hybrid screen, Nunes et al. (2005) found that the SH3 domains of CAP/ponsin interact with the second proline-rich SH3-binding motif of chicken teneurin-1; the identical motif is present in teneurin-4. CAP/ponsin in turn binds to vinculin, which could anchor the ICD of

teneurins to the actin cytoskeleton. A predicted SH3-binding motif at the same location in teneurin-2 does not bind CAP/ponsin even though it varies from the motif in teneurins-1 and -4 by only a single amino acid (Nunes et al. 2005). This led us to analyze teneurins from a broad range of taxa for SH3-binding motifs. The teneurin ICDs from each species examined, except for *S. mansoni* and *M. brevicollis*, contained one or more consensus SH3-binding motif. Interestingly, *S. mansoni* and *M. brevicollis* are the only species examined with teneurins lacking the capacity to dimerize. Perhaps dimerization is necessary for the ICD-interacting proteins to link teneurins to the cytoskeleton or to regulate the processes necessary for ICD nuclear localization.

Databases (e.g., GenBank, Ensembl, JGI, UniProt) contain listings for numerous teneurin variants. Most of these variants are based on predicted sequences, but some are based on cDNAs and ESTs. Here, we chose to study the range of alternative splicing in the ICD of human and chicken teneurin by PCR. The ICDs of human and chicken teneurins tend to be encoded on two pairs of neighboring exons separated by a large intron. Additional exons, which often are not conserved between birds and man and which frequently are subjected to alternative splicing, are sometimes found between the two pairs of exons. Alternatively spliced exons do not contain recognizable SH3-binding domains or NLSs, so the significance of these variations is not clear. Interestingly, an alternatively spliced exon in human teneurin-2 may represent an alternative start site, as ESTs with this sequence do not contain sequence from exons 1 or 2, and sequences encoded on this exon are not found in the PCR products amplified using a primer based on sequences found in exon 1. A similar method for generating teneurin splice variants was shown previously for *C. elegans* (Drabikowski et al. 2005).

The extraordinary diversity of teleost fish is commonly attributed to the duplication of their genome followed by the selective retention of certain duplicated genes (see Jozefowicz et al. 2003; Postlethwait et al. 2004; Volff 2005). This has been supported by studies of *Hox* genes (Kurosawa et al. 2006; Zou et al. 2007). In contrast, comparisons of *Sox* genes in the zebrafish *D. rerio* and the stickleback *G. aculeatus* (Cresko et al. 2003) show the mutual retention of *Sox9a* and *Sox9b*, albeit with subtle differences in their patterns of expression. Here, we show that the zebrafish has retained genes encoding teneurin-2A and teneurin-2B, whereas the stickleback has retained genes encoding teneurin-3A and teneurin-3B. Current models of selective gene retention in teleosts predict that genes are preserved following degeneration of regulatory elements and the partitioning of function between the duplicated gene products (Force et al. 1999). It is likely that a large, multifunctional protein like a teneurin would be selected in this way, and differential retention and expression could contribute to speciation.

Previously we speculated that the RHS proteins of bacteria, which share significant sequence homologies with the C-terminal portion of teneurins, may have evolved from horizontal gene transfer from a metazoan teneurin to a symbiotic or pathogenic prokaryote (Minet and Chiquet-

Ehrismann 2000). However, the presence of a teneurin gene in *M. brevicollis*, but not in the other nonmetazoan opisthokonts (e.g., fungi), suggests that the gene evolved in a choanoflagellate, and the horizontal gene transfer was from a prokaryote to a eukaryote instead of the other way around. Horizontal gene transfer between predatory *M. brevicollis* and their prokaryotic prey has been described previously. For example, Foerster et al. (2008) reported that a nitrile hydratase is encoded in the *M. brevicollis* genome that is most closely related to enzymes from proteobacteria; the absence of this enzyme from other eukaryotic genomes strongly implies horizontal gene transfer from prokaryotic prey to eukaryotic predator. Over a hundred genes originating from haptophytes and diatoms have also been found in *M. brevicollis* (Nedelcu et al. 2008; Sun et al. 2010), indicating that gene transfer may be a relatively common occurrence in these organisms. In fact, this may explain the origin of the highly conserved cysteine-rich domain, which is nearly identical to part of an enzyme from the diatom *P. tricornutum* (and is also similar to an enzyme in *V. carteri*) but is only found in teneurins in metazoa. If this is the case, *M. brevicollis* teneurin originated as a fusion protein acquired by horizontal gene transfer from both a prokaryote and a diatom or algae.

Choanoflagellates are believed to be the closest living relatives of metazoans (King and Carroll 2001; Philippe et al. 2004; King et al. 2008). The presence of teneurins (which have been shown to play roles in cell–cell and cell–ECM interactions in a variety of tissues) on the surface of an ancestral choanoflagellate may have facilitated the evolution of metazoan multicellularity and the development of complex tissues. Similar roles have been proposed for cadherins, which appear to have evolved in a choanoflagellate as well (Abedin and King 2008). The two cadherins of *M. brevicollis* are found in the microvilli that form the feeding collar that surrounds the base of the flagellum, which has led to the hypothesis that this family of proteins, which is indispensable in the formation of meaningful cell–cell contacts in animal tissues, evolved as a means of catching prey. Teneurins may have evolved to do something similar. YD-repeat proteins are found on the surface of aquatic bacteria, and in vitro studies with eukaryotic cells show that teneurin expression leads to increased cell–cell adhesion (Rubin et al. 2002). The acquisition of the carbohydrate-rich YD-repeat proteins from a prokaryote by a choanoflagellate may have improved “fishing” for bacterial prey in the feeding collar. It will be interesting to determine where *M. brevicollis* teneurin is expressed to test this hypothesis.

The lowest branches of the metazoan tree of life include the ancestors of sponges, ctenophores, and cnidarians. Therefore, it is puzzling that we were able to identify teneurins in a choanoflagellate and in all the available genomes of Bilateria (i.e., deuterostomes and protostomes) but not in modern sponges or cnidarians. It is possible that our search methods were insufficient to find them. More likely they are present in some of these organisms but not in the organisms with complete and well-assembled genomes like *N. vectensis*. It will be important to scrutinize newly sequenced

and assembled sponge and cnidarian genomes for teneurin genes as they become available. Another possibility is that teneurins evolved in a relatively advanced common ancestor of protostomes and deuterostomes, after the evolution of sponges and cnidarians. In this scenario, the teneurin in *M. brevicollis* would have been acquired by horizontal gene transfer from metazoan-derived detritus and not a prokaryote. Evidence against this hypothesis includes the relative similarity of the core region of the cysteine-rich domain and a diatom enzyme and the YD repeats to YD-repeat proteins from bacteria, as well as the organization of the *M. brevicollis* teneurin gene: most of the ECD is encoded on a single huge exon, not unlike the YD-repeat proteins of prokaryotes, and unlike the ECD of metazoan teneurins. Others (King et al. 2008) have reported that the number of introns per gene in *M. brevicollis* is similar (6.6) to the number found in human genes (7.7), so the unusually large exon encoding the ECD of *M. brevicollis* teneurin argues for origins from a prokaryotic and not metazoan, horizontal gene transfer, and the subsequent loss of teneurins from the genomes of modern sponges and cnidarians.

Teneurins are phylogenetically conserved among Bilateria, where they have been demonstrated to play critical roles in pattern formation, the organization of the ECM, and the development of the nervous system. Genomic analysis reveals an ancient origin of teneurins in single-celled choanoflagellates that may have assembled teneurins via horizontal gene transfer from two of its prey: diatoms and prokaryotes. Thus, the talent for gene acquisition by an ancestral choanoflagellate, perhaps to diversify its metabolic pathways and improve its ability to capture prey, may have contributed to the development of multicellularity in metazoans.

### Supplementary Material

Supplementary tables S1–S4 and figure S1 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

### Acknowledgments

We would like to thank John Hess for his comments on the manuscript and for assistance with the cloning and sequencing of avian teneurin-3.

### References

- Abedin M, King N. 2008. The premetazoan ancestry of cadherins. *Science* 319:946–948.
- Bagutti C, Forro G, Ferralli J, Rubin B, Chiquet-Ehrismann R. 2003. The intracellular domain of teneurin-2 has a nuclear function and represses zic-1-mediated transcription. *J Cell Sci* 116:2957–2966.
- Baumgartner S, Chiquet-Ehrismann R. 1993. Tena, a *Drosophila* gene related to tenascin, shows selective transcript localization. *Mech Dev* 40:165–176.
- Baumgartner S, Martin D, Hagios C, Chiquet-Ehrismann R. 1994. Tenna, a *Drosophila* gene related to tenascin, is a new pair-rule gene. *EMBO J* 13:3728–3740.
- Beckmann J, Vitobello A, Ferralli J, Kenzelmann Broz D, Rijli FM, Chiquet-Ehrismann R. 2011. Human teneurin-1 is a direct target

- of the homeobox transcription factor EMX2 at a novel alternate promoter. *BMC Dev Biol.* 11:35.
- Cresko WA, Yan YL, Baltrus DA, Amores A, Singer A, Rodriguez-Mari A, Postlethwait JH. 2003. Genome duplication, subfunction partitioning and lineage divergence: *sox9* in stickleback and zebrafish. *Dev Dyn.* 228(3):480–489.
- Drabikowski K, Trzebiatowska A, Chiquet-Ehrismann R. 2005. *ten-1*, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in *Caenorhabditis elegans*. *Dev Biol.* 282:27–38.
- Foerster KU, Doerks T, Muller J, Raes J, Bork P. 2008. A nitrile hydratase in the eukaryote *Monosiga brevicollis*. *PLoS One* 3:e3976.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- Helland R, Fjellbirkeland A, Karlsen OA, Ve T, Lillehaug JR, Jensen HB. 2008. An oxidized tryptophan facilitates copper binding in *Methylococcus capsulatus*-secreted protein MopE. *J Biol Chem.* 283:13897–13904.
- Jozefowicz C, McClintock J, Prince V. 2003. The fates of zebrafish *Hox* gene duplicates. *J Struct Funct Genomics.* 3:185–194.
- Kay BK, Williamson MP, Sudol M. 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* 14:231–241.
- Kenzelmann D, Chiquet-Ehrismann R, Leachman NT, Tucker RP. 2008. Teneurin-1 is expressed in interconnected regions of the developing brain and is processed in vivo. *BMC Dev Biol.* 8:30.
- King N, Carroll SB. 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. *Proc Natl Acad Sci U S A.* 98:15032–15037.
- King N, Westbrook MJ, Young SL, Kuo A, et al. (36 co-authors). 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788.
- Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A.* 106:10171–10176.
- Kowanetz K, Szymkiewicz I, Haglund K, Kowanetz M, Husnjak K, Taylor JD, Soubeyran P, Engstrom U, Ladbury JE, Dikic I. 2003. Identification of a novel proline-arginine motif involved in CIN85-dependent clustering of Cbl and down-regulation of epidermal growth factor receptors. *J Biol Chem.* 278:39735–39746.
- Kurosawa G, Takamatsu N, Takahashi M, et al. (15 co-authors). 2006. Organization and structure of *hox* gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. *Gene* 370:75–82.
- Leamey CA, Merlin S, Lattouf P, Sawatari A, Zhou X, Demel N, Glendinning KA, Oohashi T, Sur M, Fässler R. 2007. *Ten\_m3* regulates eye-specific patterning in the mammalian visual pathway and is required for binocular vision. *PLoS Biol.* 5:e241.
- Levine A, Bashan-Ahrend A, Budai-Hadrian O, Gartenberg D, Menasherov S, Wides R. 1994. *Odd Oz*: a novel *Drosophila* pair rule gene. *Cell* 77:587–598.
- Li H, Bishop KM, O'Leary DD. 2006. Potential target genes of EMX2 include *Odz/Ten-M* and other gene families with implications for cortical patterning. *Mol Cell Neurosci.* 33:136–149.
- Lovejoy DA, Rotzinger S, Barsyte-Lovejoy D. 2009. Evolution of complementary peptide systems: teneurin C-terminal-associated peptides and corticotropin-releasing factor superfamilies. *Ann N Y Acad Sci.* 1163:215–220.
- Mayer BJ. 2001. SH3 domains: complexity in moderation. *J Cell Sci.* 114:1253–1263.
- Minet AD, Rubin BP, Tucker RP, Baumgartner S, Chiquet-Ehrismann R. 1999. Teneurin-1, a vertebrate homologue of the *Drosophila* pair-rule gene *ten-m*, is a neuronal protein with a novel type of heparin-binding domain. *J Cell Sci.* 112:2019–2032.
- Minet AD, Chiquet-Ehrismann R. 2000. Phylogenetic analysis of teneurin genes and comparison to the rearrangement hot spot elements of *E. coli*. *Gene* 257:87–97.
- Nedelcu AM, Miles IH, Fagir AM, Karol K. 2008. Adaptive eukaryote-to-eukaryote lateral gene transfer: stress-related genes of algal origin in the closest unicellular relatives of animals. *J Evol Biol.* 21:1852–1860.
- Nunes SM, Ferralli J, Choi K, Brown-Luedi M, Minet AD, Chiquet-Ehrismann R. 2005. The intracellular domain of teneurin-1 interacts with MBD1 and CAP/ponsin resulting in subcellular codistribution and translocation to the nuclear matrix. *Exp Cell Res.* 305:122–132.
- Oohashi T, Zhou XH, Feng K, Richter B, Morgelin M, Perez MT, Su WD, Chiquet-Ehrismann R, Rauch U, Fässler R. 1999. Mouse *ten-m/Odz* is a new family of dimeric type II transmembrane proteins expressed in many tissues. *J Cell Biol.* 145:563–577.
- Philippe H, Snell EA, Baptiste E, Lopez P, Holland PW, Casane D. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol Biol Evol.* 21:1740–1752.
- Postlethwait J, Amores A, Cresko W, Singer A, Yan YL. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* 20:481–490.
- Rakovitsky N, Buganim Y, Swissa T, Kinel-Tahan Y, Brenner S, Cohen MA, Levine A, Wides R. 2007. *Drosophila Ten-a* is a maternal pair-rule and patterning gene. *Mech Dev.* 124:911–924.
- Rotzinger S, Lovejoy DA, Tan LA. 2010. Behavioral effects of neuropeptides in rodent models of depression and anxiety. *Peptides* 31:736–756.
- Rubin BP, Tucker RP, Martin D, Chiquet-Ehrismann R. 1999. Teneurins: a novel family of neuronal cell surface proteins in vertebrates, homologous to the *Drosophila* pair-rule gene product *Ten-m*. *Dev Biol.* 216:195–209.
- Rubin BP, Tucker RP, Brown-Luedi M, Martin D, Chiquet-Ehrismann R. 2002. Teneurin 2 is expressed by the neurons of the thalamofugal visual system in situ and promotes homophilic cell-cell adhesion in vitro. *Development* 129:4697–4705.
- Sun G, Yang Z, Ishwar A, Huang J. 2010. Algal genes in the closest relatives of animals. *Mol Biol Evol.* 27:2879–2889.
- Trubiani G, Al Chawaf A, Belsham DD, Barsyte-Lovejoy D, Lovejoy DA. 2007. Teneurin carboxy (C)-terminal associated peptide-1 inhibits alkalosis-associated necrotic neuronal death by stimulating superoxide dismutase and catalase activity in immortalized mouse hypothalamic cells. *Brain Res.* 1176:27–36.
- Trzebiatowska A, Topf U, Sauder U, Drabikowski K, Chiquet-Ehrismann R. 2008. *Caenorhabditis elegans* teneurin, *ten-1*, is required for gonadal and pharyngeal basement membrane integrity and acts redundantly with integrin *ina-1* and dystroglycan *dgn-1*. *Mol Biol Cell.* 19:3898–3908.
- Tucker RP, Chiquet-Ehrismann R. 2006. Teneurins: a conserved family of transmembrane proteins involved in intercellular signaling during development. *Dev Biol.* 290:237–245.
- Tucker RP, Chiquet-Ehrismann R, Chevron MP, Martin D, Hall RJ, Rubin BP. 2001. Teneurin-2 is expressed in tissues that regulate limb and somite pattern formation and is induced in vitro and in situ by FGF8. *Dev Dyn.* 220:27–39.
- Volff JN. 2005. Genome evolution and biodiversity in teleost fish. *Heredity* 94:280–294.
- Youderian P, Hartzell PL. 2007. Triple mutants uncover three new genes required for social motility in *Myxococcus xanthus*. *Genetics* 177:557–566.
- Young TR, Leamey CA. 2009. Teneurins: important regulators of neural circuitry. *Int J Biochem Cell Biol.* 41:990–993.
- Zou SM, Jiang XY, He ZZ, Yuan J, Yuan XN, Li SF. 2007. *Hox* gene clusters in blunt snout bream, *Megalobrama amblycephala* and comparison with those of zebrafish, fugu and medaka genomes. *Gene* 400:60–70.

## 4.2 The intracellular domain of Teneurin-1 influences MITF-dependent transcription by HINT1 binding

Jonas Schöler, Jacqueline Ferralli, Stéphane Thiry,  
and Ruth Chiquet-Ehrismann

MS ID#: JBC/2014/615922

In Revision

My contribution to this paper:

For this study, I cloned the constructs and established the stable cell lines. I performed the Western Blot, all real-time Q-PCR experiments and prepared the total RNA for the whole transcriptome analyses. Further, I was involved in the planning and analysis of all other experiments. I wrote the first and final versions of the manuscript, with the input of the co-authors. Finally, I revised the manuscript and performed the necessary revision experiments.

**The intracellular domain of Teneurin-1 influences MITF-dependent transcription by HINT1 binding**

Jonas Schöler<sup>1,2</sup>, Jacqueline Ferralli<sup>1</sup>, Stéphane Thiry<sup>1</sup>, Ruth Chiquet-Ehrismann<sup>1,2\*</sup>

<sup>1</sup>Novartis Research Foundation, Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

<sup>2</sup>Faculty of Science, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

\*Corresponding author: ruth.chiquet@fmi.ch

Running title: *Teneurin 1 ICD as a transcriptional regulator*

Keywords: Synapse, glioblastoma, microarray, transcription repressor, Notch pathway, ODZ, TENM1, SH3 domain, MACF1, BEX1.

**CAPSULE**

**Background:** The intracellular domain of teneurins translocates to the nucleus and is thought to influence transcription.

**Results:** The teneurin-1 intracellular domain (TEN1-ICD) binds transcriptional repressor HINT1 and induction of TEN1-ICD expression induces MITF target genes including *GPNMB*.

**Conclusion:** TEN1-ICD prevents HINT1 to repress MITF at its target gene promoters, inducing MITF-mediated transcription.

**Significance:** This is a novel mechanism for a teneurin ICD influencing transcription.

**ABSTRACT**

Teneurins are large type II transmembrane proteins that are necessary for the normal development of the central nervous system (CNS). While many studies highlight the significance of teneurins, especially during development, there is only limited information known about the molecular mechanisms of function. Previous studies have shown that the N-terminal intracellular domain (ICD) of teneurins can be cleaved at the membrane and subsequently translocates to the nucleus where it can influence gene transcription. Since teneurin ICDs do not contain any intrinsic DNA binding sequences, interaction partners are required to affect transcription. Here, we identified histidine triad nucleotide binding

protein 1 (HINT1) as a human teneurin-1 ICD interaction partner in a yeast-2 hybrid screen. This interaction was confirmed in human cells, where HINT1 is known to inhibit the transcription of target genes by directly binding to transcription factors at the promoter. In a whole transcriptome analysis of BS149 glioblastoma cells overexpressing the teneurin-1 ICD, several microphthalmia-associated transcription factor (MITF) target genes were found to be up-regulated. Directly comparing the transcriptomes of MITF versus TEN1-ICD overexpressing BS149 cells revealed 42 co-regulated genes, including glycoprotein non-metastatic b (GPNMB). Using real-time Q-PCR to detect endogenous GPNMB expression upon overexpression of MITF and HINT1 as well as promoter reporter assays using GPNMB promoter constructs, we could demonstrate that the teneurin-1 ICD binds HINT1, thus switching on MITF-dependent transcription of GPNMB.

**INTRODUCTION**

Teneurins are a well-conserved protein family that were originally identified in *Drosophila* (1-3). While two paralogs have been described in the fruit fly, and one in *C. elegans*, four members, namely teneurins 1-4, have been identified in vertebrates (4,5). Expression patterns during development taken together with functional studies strongly suggest a role for teneurins during



development of the CNS in *Drosophila* (6-8), *C. elegans* (9,10), chicken (6,11-15), and mouse (5,16-19), though teneurins are likely to play roles in other regions, like the developing limbs, as well (12,13).

Members of this protein family are type II transmembrane proteins containing a single-spanning transmembrane domain. Teneurins are about 300 kDa in size, the largest part being the C-terminal extracellular domain (ECD), which contains several conserved domains (11). Following the transmembrane domain, many teneurins have furin cleavage sites (Tucker et al. 2012) preceding eight EGF repeats, through which teneurins can dimerize due to free cysteines in repeats 2 and 5 forming inter-chain disulfide bonds (5,20). The following NHL repeat domain is a predicted beta-propeller, and is responsible for homophilic rather than a heterophilic interactions in some teneurins (5,21), though heterophilic interactions have also been observed (22). Interestingly, these interactions can lead to the release of the ECD (11,23). Further, there are 26 YD repeats, which are unique for a eukaryotic protein, as they are generally found in bacterial cell wall proteins (4). Finally, the most C-terminal domain or teneurin C-terminal associated peptide (TCAP) has been described as having sequence similarity to corticotropin-releasing factor (24) and has neuromodulatory activity (25). More recently striking homology of the C-terminal half of the teneurin ECDs, including the NHL, YD and TCAP regions, was discovered with bacterial ABC toxins (26,27).

The N-terminal intracellular domain is about 45 kDa in size and is not as conserved across all phyla as the ECD (6,28). Nevertheless, vertebrate ICDs share a sequence similarity of up to 70% between orthologs and 50-60% between paralogs and the ICDs contain several common features and domains. There is an EF-hand-like Ca<sup>2+</sup> binding site, predicted phosphorylation sites, and poly-proline rich regions shown to bind SH3 domain-containing proteins like CAP/Ponsin (28,29).

Teneurins were first identified as pair-rule genes (2,3), which was curious as pair-rule genes were known to be transcription factors rather than transmembrane proteins (5). While this evidence was recently refuted (30), it still paved the way to

implicate the teneurin ICDs in transcriptional regulation. Regulated intramembrane proteolysis by either Site-2 protease, signal peptide peptidase (SPP) or SPP-like proteases is predicted to release the ICD (23) before it translocates to the nucleus, as shown by several in vitro and in vivo studies (14,28,31). In the nucleus, the teneurin-2 ICD affects zic-mediated transcription, and is localized to PML bodies, which are so-called transcriptional hotspots (31). The teneurin-1 ICD can bind to MBD1, a transcriptional repressor (28). Interaction partners are crucial for the ICDs to influence transcriptional regulation since they do not contain any intrinsic DNA binding sequences.

While the teneurin ICDs are strongly implicated in transcriptional regulation, target genes as well as mechanisms have yet to be elucidated. In this study, we are investigating the transcriptional activity of the human teneurin-1 ICD, which will be referred to as simply TEN1-ICD from this point on. We performed two unbiased screens to identify: 1) interacting proteins, and 2) target genes involved in the transcriptional control of these target genes. We will show how the ICD can influence transcription via HINT1, a transcriptional repressor that directly binds to transcription factors at the promoters of their target genes (32,33).

#### EXPERIMENTAL PROCEDURES

*Cloning* – All constructs were prepared by classical cloning procedures and verified by sequencing. TEN1-ICD was cloned from human adult brain cDNA (34), HINT1 from pPR3-N-HINT1 (Dualsystems), eGFP-His from pcDNA3-EGFP (Addgene), RFP-HA from pQCXIX-RFP (Clontech), CFP-MYC from pECFP1-C1 (Addgene) and MITF from pCMV6-MITF variant 4 (Origene). The following constructs were prepared in pcDNA3 (Invitrogen): HINT1-MYC, TEN1-ICD-HA, HINT1-CFP-MYC, RFP-HA, MITF-RFP-HA (Note: all tags are C-terminal). For inducible overexpression studies with the TEN1-ICD we used the highly predictable and tightly regulated gene expression system described by Anastassiadis et Al. (35) using the following plasmids: pirtetR-GBD as tet-activator plasmid and eGFP-His and TEN1-ICD-eGFP-His in the tet-promoter plasmid ptetO as described in (35). For the yeast-2 hybrid screen

TEN1-ICD was cloned into pDHB1 (Dualsystems). For promoter reporter studies the GPNMB and GPNMB  $\Delta$ M-box promoters were cloned into pSEAP2basic (Clontech). Promoter sequences GPNMB (1096 bps) and GPNMB  $\Delta$ M-box (144 bps) as described in (36) were cloned from HT1080 cell whole genomic DNA, extracted with DNeasy Tissue kit (Qiagen) using the following primers:

(fw: ACTAGCTAGCGCCACATTTTCTGCATACTCTG; rev: ACTACTCGAGCATCTGTGGTGCCTCCCTCT) and (fw: ATGCTAGCGAAC TTGAGAGACCAGATCAGGC; rev: ATCTCGAGCAGTGTTCCTCTGGCATCTGTGGTGCCTCCCTCTC), respectively.

**Cell Culture** - BS149, SH-SY5Y, U373, LN229, MO95S, LN408, HS643, U343MG, T98G, LN319 cells (Kind gift of the Hemmings lab at the Friedrich Miescher Institute for Biomedical Research) and COS-7 (ATCC) cells were cultured in DMEM medium containing 10% Fetal Bovine Serum (FBS), 100 mg/L penicillin and 100 mg/L streptomycin. Genes were either overexpressed by transient transfection or by a modified tet-system kindly provided by Gerrit Fischedick of the Schöler lab at the Max Planck Institute for Molecular Biomedicine (35). Cells were transfected with jetPEI (Polyplus). BS149 cells were first transfected with pirtetR-GBD and made stable by 0.5  $\mu$ g/ $\mu$ l Puromycin selection, and then after further transfection with either ptetO-eGFP-His or ptetO-TEN1-ICD-eGFP-His by 150  $\mu$ g/ $\mu$ l Hygromycin selection. The tet-system was induced by adding  $10^{-7}$  M dexamethasone (Dex) 100% ethanol and 1  $\mu$ g/ml doxycycline (Dox) in Milli-Q water to the medium.

**Yeast-2 Hybrid Screen** – The Yeast-2 Hybrid screen was performed using the DUALhunter starter kit (Dualsystems) and following its manual. Prey proteins were fished with the bait protein Ten1 ICD, out of the Normalized Human Fetal Brain cDNA (NubG-X) library (Dualsystems).

**Proximity Ligation Assay and Immunocytochemistry** – The Proximity Ligation Assay was performed using the DUOLink system (Sigma-Aldrich) and following the appropriate protocol provided by the manufacturer. COS-7 cells were transfected with either HINT1-MYC as negative control or both, HINT1-MYC and TEN1-ICD-HA,

to identify an interaction. For detection, mouse monoclonal  $\alpha$ -HA (42F13, used as hybridoma supernatant) and rabbit polyclonal  $\alpha$ -MYC (Abcam) antibodies were used. For immunocytochemistry parallel cultures were stained with the same primary antibodies and Alexa Fluor 488 Goat Anti-Rabbit and Alexa Fluor 568 Goat Anti-Mouse (both life technologies) as secondary antibodies. Cells were viewed with an Axioskop 50 Microscope (Zeiss) and pictures were taken with an ORCA-ER digital camera (Hamamatsu).

**Real-time Q-PCR** – For the glioblastoma cell line screen of endogenous teneurin-1 expression,  $1 \times 10^6$  cells were used for RNA extraction. All overexpressing cells were FACS-sorted directly into RLT lysis buffer (Qiagen) at a 3:1 volume ratio of lysis buffer to cells in PBS, 24 h post-induction or -transfection. Total RNA was extracted with an RNeasy Mini kit and QIAshredder columns (both Qiagen), and reverse-transcribed into cDNA, using random hexamer primers with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems).

For Real-time Q-PCR, gene-specific primers (Supplemental Table 1) were designed and all data were normalized to TBP, using Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen) on a StepOnePlus Real-Time PCR System (Applied Biosystems).

**Western Blot** – BS149 cells were induced, or kept uninduced as a negative control, at 70% confluency on a 3.5 cm plate and harvested after 24 h directly in sample buffer, containing  $\beta$ -mercaptoethanol. Samples were run on a 10% polyacrylamide gel and blotted on a PVDF membrane. Protein bands were detected by a mouse monoclonal  $\alpha$ -GFP antibody (Roche) in 5% skim milk in 1xTBS-Tween-20 (0.05%), a mouse monoclonal  $\alpha$ -vinculin antibody (Sigma-Aldrich) as a loading control, and a goat  $\alpha$ -mouse secondary horseradish peroxidase-conjugated antibody (Invitrogen) with SuperSignal West Dura Chemiluminescent Substrate (Thermo Scientific).

**Microarray** – For the whole transcriptome analysis of TEN1-ICD overexpression, the stable BS149 cell lines with pirtetR-GBD/ptetO-eGFP-His or pirtetR-GBD/ptetO-TEN1-ICD-eGFP-His were split into three 10 cm Petri dishes each. The

triplicate cell lines were cultured once before induction with Dex and Dox. For the whole transcriptome analysis of MITF, cells were transiently transfected in triplicates with either, pcDNA3-RFP-HA or pcDNA3-MITF-RFP-HA. The overexpressing cells were then FACS-sorted directly into RLT lysis buffer (Qiagen) at a 3:1 volume ratio of lysis buffer to cells in PBS, 24 h post-induction or -transfection. Total RNA was extracted with an RNeasy Mini kit and QIAshredder columns (both Qiagen), and reverse-transcribed into cDNA, using random hexamer primers with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). 100 ng of extracted total RNA was amplified using the Ambion WT Expression kit (Ambion) and the resulting sense-strand cDNA was fragmented and labeled using the Affymetrix GeneChip WT Terminal Labeling kit (Affymetrix). Affymetrix Gene Chip Arrays were hybridized following the "GeneChip Whole Transcript (WT) Sense Target Labeling Assay Manual" (Affymetrix) with a hybridization time of 16 h. The Affymetrix Fluidics protocol FS450\_0007 was used for washing. Scanning was performed with Affymetrix GCC Scan Control v. 3.0.0.1214 on a GeneChip Scanner 3000 with autoloader. Arrays were normalized and probeset-level expression values were calculated with R/Bioconductor's (v2.14) 'affy' package using the rma() function. Differential gene expression was calculated with the Bioconductor package limma using a linear fold-change cutoff = 1.5 and a Benjamini-Hochberg corrected p-value = 0.05. Normalized unscaled expression values for the genes passing these filters were plotted using the heatmap.2() function of the gplots package. Microarray data files and detailed protocols for procedures and statistical analysis are available from Gene Expression Omnibus (GEO), accession numbers 61704, 61705 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61704>; <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61705>).

*SEAP Reporter Assays* – Promoter reporter assays were performed by cotransfecting BS149 cells with 1 µg of pSEAP2basic-promoter constructs and either 0.5 µg empty pcDNA3 vector, or 0.5 µg MITF, or 0.05 µg MITF / 0.45 µg empty pcDNA3 vector or 0.05 µg MITF / 0.45 µg Ten1 ICD, and 150 ng of the *Metridia* luciferase pMetLuc

reporter vector (Clontech) for the normalization of the transfection efficiency. SEAP activity measurements were taken 24 h post-transfection by using the chemiluminescent SEAP Reporter Gene Assay (Roche). The *Metridia* luciferase assay for normalization was performed with the Ready-to-Glow Secreted Luciferase Reporter Assay (Clontech). All values were measured by the Mithras LB940 Luminometer (Berthold Technologies).

*Statistical Analysis* - All grouped data are means ± SD. Statistical analysis was completed using GraphPad InStat v3.05. Differences between two groups were evaluated using a two-tailed Student's t test for parametric data or a Mann-Whitney U test for nonparametric data. Multiple comparisons were performed using one-way analysis of variance (ANOVA). Values of P < 0.05 were considered statistically significant.

## RESULTS

### *Novel interaction partners of TEN1-ICD*

To identify novel interaction partners, we performed a Yeast-2 Hybrid screen using the TEN1-ICD as a membrane-anchored bait protein fishing for prey encoded by a human fetal brain cDNA library. The benefit of this method is that baits which might activate transcription independently of a prey are prevented from entering the nucleus, making them nevertheless usable for a screen. Furthermore, the readout is based on the split-ubiquitin method, in which an interaction of the bait and prey proteins also joins the N- and C-terminal parts of split-ubiquitin. A ubiquitin-specific protease then releases the artificial transcription factor LexA-VP16 to switch on the two reporters HIS3 and LacZ (Figure 1A). The yeast strain NMY51 can thus grow on a defined minimal medium lacking histidine only when an interaction occurs, and turns the medium blue due to β-galactosidase cleaving added X-gal. Our screen resulted in the isolation of more than one hundred positive colonies and their bait plasmids were isolated and sequenced. The isolated plasmids were re-transformed into NMY51 together with the TEN1-ICD or the empty prey plasmid and tested at different dilutions for growth on medium lacking histidine (Figure 1B). The positive controls show that the yeast do contain all plasmids and grow when the defined

minimal medium contains histidine and no interaction is required between bait and prey proteins for their growth (Figure 1B, left panel). When the medium lacks histidine there is a robust growth of the NYM51 colonies only when the bait and prey proteins interact (Figure 1B, right panel). There is also a blue coloration for all candidates in the  $\beta$ -galactosidase assay (Figure 1B). We thus identified a total of eleven novel interaction partners, each being potentially interesting in the context of teneurin function. The main features of each interacting protein is summarized in Table 2.

*Several MITF target genes are up-regulated in TEN1-ICD overexpressed cells*

In parallel we wanted to identify a system in which we could study how the overexpression of the TEN1-ICD influences transcription assuming that cells with endogenous teneurin-1 expression would also have the machinery to respond to teneurin-1 signaling. To get started, we performed a Q-PCR screen identifying glioblastoma cell lines that express endogenous teneurin-1 (Figure 2A). We picked the BS149 cell line for our further studies as it has a relatively high endogenous expression of teneurin-1 and excellent transfection efficiency.

To study the influence of teneurin-1 on transcriptional regulation in BS149 cells, we needed an inducible method to overexpress the TEN1-ICD transiently. Thus, we used a modified and improved tet-system, which eliminates the leakiness of the promoter (35). Two separate vectors are required to make the cell line stable (Figure 2B). One contains the tet-activating domain fused to a glucocorticoid binding domain (GBD). The other contains the tetO operator sequences directly upstream of a CMV promoter and the gene to be overexpressed. The tet-activating domain is continuously expressed and bound to HSP90 in the cytosol via the GBD. Adding dexamethasone (Dex) releases the fusion protein from HSP90 and subsequently binds the tetO-CMV promoter, switching on gene expression due to the addition of doxycycline (Dox). First, we were able to show the functionality of the system by western blot of cell extracts of the stable BS149 cells induced by Dex/Dox addition. TEN1-ICD-eGFP-His and eGFP-His fusion proteins were detected only when

the modified tet-system was induced (Figure 2C). Q-PCR confirmed the specific overexpression of the TEN1-ICD, resulting in a 791.6 fold increase on a transcript level. In comparison, there was no change of endogenous teneurin-1 levels (Figure 2D).

To show the effects on transcriptional regulation in BS149 cells, we overexpressed the TEN1-ICD-eGFP-His construct by induction of the tet-system and compared it to the overexpression of the eGFP-His construct in a whole transcriptome microarray analysis. For more meaningful results, we compared biological triplicates in which we separately induced and FACS-sorted the cells for GFP expression before extraction of their total RNA. With a fold change cutoff excluding values below 1.5 and an adjusted p-value cutoff excluding values above 0.05, we came up with a list of 430 differentially regulated genes. Interestingly, this list of genes contains seven known MITF target genes (Figure 3A) (37). All seven target genes are up-regulated in the microarray (Figure 3B), which we could also confirm by Q-PCR (Figure 3C). To exclude that these target genes are a result of the particular stable cell line used for the transcript profiling, we created a second set of stable BS149 cell lines using the same vectors of the modified tet-system. Again, we confirmed all target genes by Q-PCR (data not shown). MITF is a well-described basic helix-loop-helix leucine zipper transcription factor that has previously been shown to be directly inhibited by the histidine triad nucleotide-binding protein HINT1 (32,33). Interestingly, HINT1 was one of the interaction partners found in the Yeast-2 Hybrid screen described above and we considered that this interaction may be involved in the induction of MITF target genes by the TEN1-ICD.

*PLA confirms interaction of TEN1-ICD and HINT1*

To confirm the interaction between HINT1 and TEN1-ICD not only in yeast but also in human cells, we recloned HINT1 with a MYC-tag into a mammalian expression vector. We used the Proximity Ligation Assay (PLA) to test whether the proteins interact when expressed in COS-7 cells. We co-transfected vectors encoding HINT1-MYC and TEN1-ICD-HA and if the two proteins are located in close proximity to each other, the

oligonucleotides present on the secondary antibodies used to detect the two proteins can hybridize, be ligated, and amplified using fluorescent-labeled nucleotides. Immunocytochemistry images of co-transfected cells show that many COS-7 cells overexpress both proteins in the cytoplasm as well as in the nucleus (Figure 4A). Using PLA, the HINT1-MYC and TEN1-ICD-HA co-transfection confirms an interaction, as seen by the fluorescent red dots in the cells that are overexpressing both proteins (Figure 4B; left and center image). As a negative control, the same primary and secondary antibodies were used, but only HINT1-MYC was overexpressed in COS-7 cells (Figure 4B; right image). Therefore, we concluded that the TEN1-ICD indeed binds to HINT1 and we hypothesized that TEN-1 ICD might act by sequestering HINT1 thereby compromising its repression of MITF-mediated transcription.

*GPNMB is co-regulated by the TEN1-ICD, MITF, and HINT1*

If our hypothesis is correct and there is a connection between the TEN1-ICD and MITF in regulating transcription in BS149 cells, MITF overexpression should result in the induction of genes in common with TEN1-ICD induction. To determine the overlap of differentially regulated genes, the MITF-RFP-HA construct was transiently overexpressed in BS149 cells and compared to RFP-HA in a whole transcriptome microarray analysis. Again, for more meaningful results, we compared biological triplicates by separately transfecting and sorting the cells and extracting their total RNA. With a fold change cutoff (excluding values below 1.5) and an adjusted p-value cutoff (excluding values above 0.05) we came up with a list of 497 differentially regulated genes. Between the 430 genes differentially regulated by the TEN1-ICD and the 497 genes affected by MITF, there is an overlap of 42 genes that were either up- or down-regulated in both microarrays (Figure 5A).

One of the overlapping genes is *GPNMB*, a target gene of MITF with a well-described promoter that is directly bound by MITF (36). Therefore, we further investigated the effect of the TEN1-ICD on the expression of this particular gene. First, we could confirm by Q-PCR that MITF also regulates

*GPNMB* in BS149 cells, showing a 4.14 fold increase in expression (Figure 5B). To show a link of the TEN1-ICD and MITF via HINT1, HINT1 would have to influence MITF-regulated transcription. We therefore compared the effect on *GPNMB* transcripts after co-transfection of cells with the MITF-RFP-HA, and HINT1-CFP-MYC constructs. To make sure to analyze cells that overexpress both proteins, cells were FACS-sorted for expression of RFP and CFP before RNA isolation. Indeed, we saw a significant upregulation of *GPNMB* transcripts when we transiently overexpressed the MITF-RFP-HA together with the empty control plasmid CFP-MYC, and this induction was attenuated by co-transfection with HINT1-CFP (Figure 5C).

Finally, we wanted to investigate whether MITF directly influences *GPNMB* transcription in BS149 cells using promoter reporter experiments and whether this was affected by TEN1-ICD. To do this we used two different promoter constructs of *GPNMB* in SEAP reporter gene assays: the full promoter, and one missing the crucial MITF binding site M-box (*GPNMB*  $\Delta$ M-box) as described in (36) (Figure 6A). Overexpressed MITF in BS149 cells strongly induced the full *GPNMB* promoter, with a 24.15 fold increase compared to the control, while the *GPNMB*  $\Delta$ M-box promoter could not be induced by MITF (Figure 6B). Thus, MITF directly binds to the *GPNMB* promoter and influences its transcription in BS149 cells. Next we tested the influence of the TEN1-ICD on MITF-induced transcriptional activation of the *GPNMB* promoter reporter. This experiment showed that the TEN1-ICD further increases MITF-dependent transcription by 1.59 fold.

In summary, our results show that the transcriptional repressor HINT1 is a novel interaction partner of teneurin-1, through which the TEN-1 ICD can influence MITF-dependent transcription as depicted in the model presented in Figure 7.

## DISCUSSION

We propose a function for the ICD of teneurin-1 that is very similar to the Notch signaling pathway. Notch is a well-studied type I transmembrane protein with a single-spanning

transmembrane domain and an ICD that is much smaller in size than the ECD. The ICD can regulate transcription of several target genes by replacing repressors with other positive regulators in a transcription complex once it translocates to the nucleus (38).

In this study, we show that the teneurin-1 ICD can influence MITF-dependent transcription of *GPNUMB* by binding to repressor HINT1. While the ICDs of teneurins have been implicated in transcriptional activity in the past (28,31), here we are elucidating a new molecular mechanism to explain how a teneurin ICD can influence transcriptional activity. In our model, the ECD domain is released following homo- or heterophilic interaction, which is required for subsequently releasing the ICD.

Following its release, the ICD can translocate to the nucleus due to its predicted nuclear localization signal, as has previously been demonstrated in experimental studies (14,28,31). Here we have shown that the ICD can bind HINT1, thus switching on MITF-dependent transcription (Figure 7). We showed this by using the MITF target gene *GPNUMB* as an example, though there were 41 other genes differentially regulated by MITF and the TEN1-ICD. Most importantly, this highlights the potential of how teneurins can influence transcriptional regulation.

The mechanism of action of the TEN-1 ICD may be slightly different from Notch ICDs. Rather than replacing transcriptional repressors, recruiting positive regulators, and taking part in a transcriptional complex, the teneurin-1 ICD might regulate transcription by binding the transcriptional repressor and either releasing it from the transcription factor MITF or by competing for binding to MITF. We are not ruling out the possibility that the TEN-1 ICD or the ICDs of the other teneurins have other modes of action, more like the Notch ICD. As we will discuss below, there are other interesting genes that might be regulated by the teneurin-1 ICD, but not MITF.

While MITF-dependent transcription has mostly been studied in a melanocyte-specific context, other sites of expression like the retinal pigment epithelium have also been identified (39). Additionally, according to the Allen Mouse Brain Atlas (<http://mouse.brain-map.org/>) (40) there is a

weak MITF expression in the mouse olfactory bulb. However, MITF expression seems to have been rarely studied in the context of the central nervous system.

The expression of teneurin-1, HINT1, and *GPNUMB* in the CNS has been studied more extensively. Unrelated studies suggest at least partially overlapping patterns of the three genes in the olfactory bulb, hippocampus, and cerebral cortex (14,17,41,42). It will be very interesting to further explore the overlapping expression patterns of these genes in the CNS, including MITF, and determine the function of the teneurin-1 ICD regulating MITF target genes like *GPNUMB*. In the real-time Q-PCR screen, all tested glioblastoma cell lines expressed endogenous teneurin-1. In a recent review discussing teneurins in human tumorigenesis and malignancy, teneurins-2 and -4 were described as potential tumor suppressors or oncogenes (43). Teneurin-1 has not been mentioned in the review, which may be due to a lack of data about the expression of this gene in tumors. *GPNUMB* not only promotes invasiveness of glioma cells, but is also elevated in malignant glioblastomas. Taken together, this could be a first link of teneurin-1 to cancer, and would make it a potential target in glioblastomas (44).

*GPNUMB* is another type I transmembrane protein originally identified in human melanoma cells (44). Since then, its expression has also been identified in chondrogenesis (45), in differentiating osteoclasts and osteoblasts (45), and in normal and diseased brain tissue (42,46). The function in the CNS is mostly unknown, though it is suggested to play a role in the stability of neurons and the immune/inflammatory response in the CNS (42). Interestingly, *GPNUMB* also contains an RGD sequence in its ECD, which is likely to bind integrins (44). Integrins have also been studied in the context of axon guidance and neuronal connectivity (47). This leaves us with several potential functions of the teneurin-1 ICD regulating *GPNUMB* expression. For one, *GPNUMB* could regulate the stability of neurons once they have found their post-synaptic partner due to teneurin-1 homophilic binding. On the other hand, the capability of *GPNUMB* to bind integrins could further specify or stabilize an interaction of the presynaptic neuron with its postsynaptic partner.

Finally, we cannot exclude the possibility that our identified mechanism takes place in a non-neuronal context. Recent papers have identified teneurins-3 and -4 as important players in chondrogenic differentiation (48,49). While teneurin-1 has not been studied in chondrogenesis, GPNMB's likely role as a key regulator in chondrogenesis could implicate a function of both proteins in this event.

Beyond the main aim of the study, we have also made several additional observations and findings. In addition to HINT1, we identified a total of ten other novel interaction partners of the ICD by a Yeast-2 hybrid screen. While we only focused on one of the interaction partners in the present work, each of the other proteins could be interesting to study in the context of teneurins. The three candidates brain-expressed X-linked 1 (BEX1), microtubule-actin cross-linking factor 1 (MACF1), and amyloid beta A4 precursor protein-binding family B member 1 (APBB1) particularly stand out, as they have all been implicated in neurite outgrowth. BEX1, for example, has also been shown to affect neuronal differentiation and cell cycle arrest, depending on its subcellular localization. One of the potential target genes of the TEN-1 ICD, nerve growth factor receptor (NGFR), seems to be essential in determining the localization of BEX1. In the nucleus, BEX1 inhibits cell-cycle arrest, whereas in the cytosol it competes with receptor-interacting protein 2 (RIP2) binding of NGFR, thus blocking the NF- $\kappa$ B pathway (50,51). The teneurin-1 ICD could have a dual mode of action, by not only regulating transcription of NGFR, but also aiding in the localization of BEX1, or working in a regulatory complex with BEX1. Another potential interaction partner, MACF1, contains an SH3 domain and could be another link of teneurins to the cytoskeleton by binding to its poly-proline rich region (52). This interaction has previously been identified between teneurin-1 and CAP/Ponsin (28). It is also interesting that MACF1 has an important function in filopodia formation and axon elongation (53). Finally, APBB1 has been shown to bind the cleaved ICD of amyloid precursor protein (APP) and together translocate to the nucleus to regulate transcriptional activity in a complex (54). One possibility would be that the

TEN1-ICD interacts with APBB1 in a similar mechanism and may regulate a different set of genes. APBB1 and APP are also strongly expressed in the growth cones of neurons, where they are thought to regulate actin dynamics (55). However, the mechanism has not been resolved thus far. Since there is evidence that the TEN1-ICD can be linked to the cytoskeleton (28), it may also be involved with APBB1 and APP in the growth cone.

In our whole transcriptome analysis of overexpressing the TEN-1 ICD in BS149 cells we identified several hundred genes that were affected, many of which merit further analysis. For example, two of the potential target genes, close homolog of L1 (CHL1) and Ng-CAM related cell adhesion molecule (Nr-CAM), are paralogs of the *Drosophila* gene neuroglian (Nrg). While teneurins are responsible for proper connectivity and Nrg for synaptic stability in the neuromuscular junction (NMJ), both have a function in synaptic organization (8,56). The ICDs could thus be involved in the regulation of Nrg in the NMJ of *Drosophila* or its orthologs in vertebrates. Interestingly, we also identified other target genes known to be inhibited by HINT1. For example, TGFB2 is regulated by USF2 and CCND1 by  $\beta$ -catenin/TCF4, and both transcription factors have been shown to be inhibited by HINT1 (57). Hence, it is possible that the teneurin-1 ICD not only regulates MTF-dependent transcription via HINT1, but all HINT1-mediated inhibition of transcription.

Here we elucidated a novel mechanism of how the teneurin ICD can influence transcription. It is likely that additional mechanisms will be identified in the future, as this is a very diverse protein family with a wide range of important functions.

#### ACKNOWLEDGMENTS

We would like to thank Richard Tucker (UC Davis, U.S.A.) for critically reviewing the manuscript, Tim Roloff for helping with the analysis of the microarray data, and Hubertus Kohler for performing the FACS experiments (both Friedrich Miescher Institute, Basel, Switzerland).

1. Baumgartner, S., and Chiquet-Ehrismann, R. (1993) Tena, a Drosophila gene related to tenascin, shows selective transcript localization. *Mech. Dev.* **40**, 165-176
2. Baumgartner, S., Martin, D., Hagios, C., and Chiquet-Ehrismann, R. (1994) Tenm, a Drosophila gene related to tenascin, is a new pair-rule gene. *EMBO J.* **13**, 3728-3740
3. Levine, A., Bashan-Ahrend, A., Budai-Hadrian, O., Gartenberg, D., Menasherow, S., and Wides, R. (1994) Odd Oz: a novel Drosophila pair rule gene. *Cell* **77**, 587-598
4. Minet, A. D., and Chiquet-Ehrismann, R. (2000) Phylogenetic analysis of teneurin genes and comparison to the rearrangement hot spot elements of E. coli. *Gene* **257**, 87-97
5. Oohashi, T., Zhou, X. H., Feng, K., Richter, B., Morgelin, M., Perez, M. T., Su, W. D., Chiquet-Ehrismann, R., Rauch, U., and Fassler, R. (1999) Mouse ten-m/Odz is a new family of dimeric type II transmembrane proteins expressed in many tissues. *J. Cell Biol.* **145**, 563-577
6. Minet, A. D., Rubin, B. P., Tucker, R. P., Baumgartner, S., and Chiquet-Ehrismann, R. (1999) Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene ten-m, is a neuronal protein with a novel type of heparin-binding domain. *J. Cell Sci.* **112 ( Pt 12)**, 2019-2032
7. Hong, W., Mosca, T. J., and Luo, L. (2012) Teneurins instruct synaptic partner matching in an olfactory map. *Nature* **484**, 201-207
8. Mosca, T. J., Hong, W., Dani, V. S., Favaloro, V., and Luo, L. (2012) Trans-synaptic Teneurin signalling in neuromuscular synapse organization and target choice. *Nature* **484**, 237-241
9. Drabikowski, K., Trzebiatowska, A., and Chiquet-Ehrismann, R. (2005) ten-1, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in *Caenorhabditis elegans*. *Dev. Biol.* **282**, 27-38
10. Morck, C., Vivekanand, V., Jafari, G., and Pilon, M. (2010) *C. elegans* ten-1 is synthetic lethal with mutations in cytoskeleton regulators, and enhances many axon guidance defective mutants. *BMC Dev. Biol.* **10**, 55
11. Rubin, B. P., Tucker, R. P., Martin, D., and Chiquet-Ehrismann, R. (1999) Teneurins: a novel family of neuronal cell surface proteins in vertebrates, homologous to the Drosophila pair-rule gene product Ten-m. *Dev. Biol.* **216**, 195-209
12. Tucker, R. P., Martin, D., Kos, R., and Chiquet-Ehrismann, R. (2000) The expression of teneurin-4 in the avian embryo. *Mech. Dev.* **98**, 187-191
13. Tucker, R. P., Chiquet-Ehrismann, R., Chevron, M. P., Martin, D., Hall, R. J., and Rubin, B. P. (2001) Teneurin-2 is expressed in tissues that regulate limb and somite pattern formation and is induced in vitro and in situ by FGF8. *Dev. Dyn.* **220**, 27-39
14. Kenzelmann, D., Chiquet-Ehrismann, R., Leachman, N. T., and Tucker, R. P. (2008) Teneurin-1 is expressed in interconnected regions of the developing brain and is processed in vivo. *BMC Dev. Biol.* **8**, 30
15. Kenzelmann-Broz, D., Tucker, R. P., Leachman, N. T., and Chiquet-Ehrismann, R. (2010) The expression of teneurin-4 in the avian embryo: potential roles in patterning of the limb and nervous system. *Int. J. Dev. Biol.* **54**, 1509-1516
16. Ben-Zur, T., Feige, E., Motro, B., and Wides, R. (2000) The mammalian Odz gene family: homologs of a Drosophila pair-rule gene with expression implying distinct yet overlapping developmental roles. *Dev. Biol.* **217**, 107-120
17. Zhou, X. H., Brandau, O., Feng, K., Oohashi, T., Ninomiya, Y., Rauch, U., and Fassler, R. (2003) The murine Ten-m/Odz genes show distinct but overlapping expression patterns during development and in adult brain. *Gene Expr Patterns* **3**, 397-405
18. Leamey, C. A., Merlin, S., Lattouf, P., Sawatari, A., Zhou, X., Demel, N., Glendining, K. A., Oohashi, T., Sur, M., and Fassler, R. (2007) Ten\_m3 regulates eye-specific patterning in the mammalian visual pathway and is required for binocular vision. *PLoS Biol.* **5**, e241



19. Young, T. R., Bourke, M., Zhou, X., Oohashi, T., Sawatari, A., Fassler, R., and Leamey, C. A. (2013) Ten-m2 is required for the generation of binocular visual circuits. *J. Neurosci.* **33**, 12490-12509
20. Feng, K., Zhou, X. H., Oohashi, T., Morgelin, M., Lustig, A., Hirakawa, S., Ninomiya, Y., Engel, J., Rauch, U., and Fassler, R. (2002) All four members of the Ten-m/Odz family of transmembrane proteins form dimers. *J. Biol. Chem.* **277**, 26128-26135
21. Beckmann, J., Schubert, R., Chiquet-Ehrismann, R., and Muller, D. J. (2013) Deciphering teneurin domains that facilitate cellular recognition, cell-cell adhesion, and neurite outgrowth using atomic force microscopy-based single-cell force spectroscopy. *Nano Lett* **13**, 2937-2946
22. Boucard, A. A., Maxeiner, S., and Sudhof, T. C. (2014) Latrophilins function as heterophilic cell-adhesion molecules by binding to teneurins: regulation by alternative splicing. *J. Biol. Chem.* **289**, 387-402
23. Kenzelmann, D., Chiquet-Ehrismann, R., and Tucker, R. P. (2007) Teneurins, a transmembrane protein family involved in cell communication during neuronal development. *Cell. Mol. Life Sci.* **64**, 1452-1456
24. Wang, L., Rotzinger, S., Al Chawaf, A., Elias, C. F., Baryte-Lovejoy, D., Qian, X., Wang, N. C., De Cristofaro, A., Belsham, D., Bittencourt, J. C., Vaccarino, F., and Lovejoy, D. A. (2005) Teneurin proteins possess a carboxy terminal sequence with neuromodulatory activity. *Brain Res. Mol. Brain Res.* **133**, 253-265
25. Lovejoy, D. A., Al Chawaf, A., and Cadinouche, M. Z. (2006) Teneurin C-terminal associated peptides: an enigmatic family of neuropeptides with structural similarity to the corticotropin-releasing factor and calcitonin families of peptides. *Gen. Comp. Endocrinol.* **148**, 299-305
26. Busby, J. N., Panjkar, S., Landsberg, M. J., Hurst, M. R., and Lott, J. S. (2013) The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. *Nature* **501**, 547-550
27. Zhang, D., de Souza, R. F., Anantharaman, V., Iyer, L. M., and Aravind, L. (2012) Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* **7**, 18
28. Nunes, S. M., Ferralli, J., Choi, K., Brown-Luedi, M., Minet, A. D., and Chiquet-Ehrismann, R. (2005) The intracellular domain of teneurin-1 interacts with MBD1 and CAP/ponsin resulting in subcellular codistribution and translocation to the nuclear matrix. *Exp. Cell Res.* **305**, 122-132
29. Tucker, R. P., and Chiquet-Ehrismann, R. (2006) Teneurins: a conserved family of transmembrane proteins involved in intercellular signaling during development. *Dev. Biol.* **290**, 237-245
30. Zheng, L., Michelson, Y., Freger, V., Avraham, Z., Venken, K. J., Bellen, H. J., Justice, M. J., and Wides, R. (2011) Drosophila Ten-m and filamin affect motor neuron growth cone guidance. *PLoS One* **6**, e22956
31. Bagutti, C., Forro, G., Ferralli, J., Rubin, B., and Chiquet-Ehrismann, R. (2003) The intracellular domain of teneurin-2 has a nuclear function and represses zic-1-mediated transcription. *J. Cell Sci.* **116**, 2957-2966
32. Razin, E., Zhang, Z. C., Nechushtan, H., Frenkel, S., Lee, Y. N., Arudchandran, R., and Rivera, J. (1999) Suppression of microphthalmia transcriptional activity by its association with protein kinase C-interacting protein 1 in mast cells. *J. Biol. Chem.* **274**, 34272-34276
33. Genovese, G., Ghosh, P., Li, H., Rettino, A., Sioletic, S., Cittadini, A., and Sgambato, A. (2012) The tumor suppressor HINT1 regulates MITF and beta-catenin transcriptional activity in melanoma cells. *Cell Cycle* **11**, 2206-2215
34. Tucker, R. P., Beckmann, J., Leachman, N. T., Scholer, J., and Chiquet-Ehrismann, R. (2012) Phylogenetic analysis of the teneurins: conserved features and premetazoan ancestry. *Mol. Biol. Evol.* **29**, 1019-1029

35. Anastassiadis, K., Kim, J., Daigle, N., Sprengel, R., Scholer, H. R., and Stewart, A. F. (2002) A predictable ligand regulated expression strategy for stably integrated transgenes in mammalian cells in culture. *Gene* **298**, 159-172
36. Ripoll, V. M., Meadows, N. A., Raggatt, L. J., Chang, M. K., Pettit, A. R., Cassady, A. I., and Hume, D. A. (2008) Microphthalmia transcription factor regulates the expression of the novel osteoclast factor GPNMB. *Gene* **413**, 32-41
37. Hoek, K. S., Schlegel, N. C., Eichhoff, O. M., Widmer, D. S., Praetorius, C., Einarsson, S. O., Valgeirsdottir, S., Bergsteinsdottir, K., Schepsky, A., Dummer, R., and Steingrimsson, E. (2008) Novel MITF targets identified using a two-step DNA microarray strategy. *Pigment Cell Melanoma Res* **21**, 665-676
38. Guruharsha, K. G., Kankel, M. W., and Artavanis-Tsakonas, S. (2012) The Notch signalling system: recent insights into the complexity of a conserved pathway. *Nat Rev Genet* **13**, 654-666
39. Adjianto, J., Castorino, J. J., Wang, Z. X., Maminishkis, A., Grunwald, G. B., and Philp, N. J. (2012) Microphthalmia-associated transcription factor (MITF) promotes differentiation of human retinal pigment epithelium (RPE) by regulating microRNAs-204/211 expression. *J. Biol. Chem.* **287**, 20491-20503
40. Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., van de Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L., Abajian, C., Beckmann, C. F., Bernard, A., Bertagnolli, D., Boe, A. F., Cartagena, P. M., Chakravarty, M. M., Chapin, M., Chong, J., Dalley, R. A., Daly, B. D., Dang, C., Datta, S., Dee, N., Dolbeare, T. A., Faber, V., Feng, D., Fowler, D. R., Goldy, J., Gregor, B. W., Haradon, Z., Haynor, D. R., Hohmann, J. G., Horvath, S., Howard, R. E., Jeromin, A., Jochim, J. M., Kinnunen, M., Lau, C., Lazarz, E. T., Lee, C., Lemon, T. A., Li, L., Li, Y., Morris, J. A., Overly, C. C., Parker, P. D., Parry, S. E., Reding, M., Royall, J. J., Schulkin, J., Sequeira, P. A., Slaughterbeck, C. R., Smith, S. C., Sodt, A. J., Sunkin, S. M., Swanson, B. E., Vawter, M. P., Williams, D., Wohnoutka, P., Zielke, H. R., Geschwind, D. H., Hof, P. R., Smith, S. M., Koch, C., Grant, S. G., and Jones, A. R. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399
41. Liu, Q., Puche, A. C., and Wang, J. B. (2008) Distribution and expression of protein kinase C interactive protein (PKCI/HINT1) in mouse central nervous system (CNS). *Neurochem. Res.* **33**, 1263-1276
42. Huang, J. J., Ma, W. J., and Yokoyama, S. (2012) Expression and immunolocalization of Gpnmb, a glioma-associated glycoprotein, in normal and inflamed central nervous systems of adult rats. *Brain Behav* **2**, 85-96
43. Ziegler, A., Corvalan, A., Roa, I., Branes, J. A., and Wollscheid, B. (2012) Teneurin protein family: an emerging role in human tumorigenesis and drug resistance. *Cancer Lett.* **326**, 1-7
44. Maric, G., Rose, A. A., Annis, M. G., and Siegel, P. M. (2013) Glycoprotein non-metastatic b (GPNMB): A metastatic mediator and emerging therapeutic target in cancer. *Onco Targets Ther.* **6**, 839-852
45. Abdelmagid, S. M., Barbe, M. F., Hadjiargyrou, M., Owen, T. A., Razmpour, R., Rehman, S., Popoff, S. N., and Safadi, F. F. (2010) Temporal and spatial expression of osteoactivin during fracture repair. *J. Cell. Biochem.* **111**, 295-309
46. Tanaka, H., Shimazawa, M., Kimura, M., Takata, M., Tsuruma, K., Yamada, M., Takahashi, H., Hozumi, I., Niwa, J., Iguchi, Y., Nikawa, T., Sobue, G., Inuzuka, T., and Hara, H. (2012) The potential of GPNMB as novel neuroprotective factor in amyotrophic lateral sclerosis. *Sci. Rep.* **2**, 573
47. Myers, J. P., Santiago-Medina, M., and Gomez, T. M. (2011) Regulation of axonal outgrowth and pathfinding by integrin-ECM interactions. *Dev. Neurobiol.* **71**, 901-923

48. Murakami, T., Fukunaga, T., Takeshita, N., Hiratsuka, K., Abiko, Y., Yamashiro, T., and Takano-Yamamoto, T. (2010) Expression of Ten-m/Odz3 in the fibrous layer of mandibular condylar cartilage during postnatal growth in mice. *J. Anat.* **217**, 236-244
49. Suzuki, N., Mizuniwa, C., Ishii, K., Nakagawa, Y., Tsuji, K., Muneta, T., Sekiya, I., and Akazawa, C. (2014) Teneurin-4, a transmembrane protein, is a novel regulator that suppresses chondrogenic differentiation. *J. Orthop. Res.* **32**, 915-922
50. Carter, B. D. (2006) A Bex-cycle built for two. *EMBO Rep* **7**, 382-384
51. Vilar, M., Murillo-Carretero, M., Mira, H., Magnusson, K., Besset, V., and Ibanez, C. F. (2006) Bex1, a novel interactor of the p75 neurotrophin receptor, links neurotrophin signaling to the cell cycle. *EMBO J.* **25**, 1219-1230
52. Margaron, Y., Fradet, N., and Cote, J. F. (2013) ELMO recruits actin cross-linking family 7 (ACF7) at the cell membrane for microtubule capture and stabilization of cellular protrusions. *J. Biol. Chem.* **288**, 1184-1199
53. Sanchez-Soriano, N., Travis, M., Dajas-Bailador, F., Goncalves-Pimentel, C., Whitmarsh, A. J., and Prokop, A. (2009) Mouse ACF7 and drosophila short stop modulate filopodia formation and microtubule organisation during neuronal growth. *J. Cell Sci.* **122**, 2534-2542
54. Cao, X., and Sudhof, T. C. (2001) A transcriptionally [correction of transcriptively] active complex of APP with Fe65 and histone acetyltransferase Tip60. *Science* **293**, 115-120
55. Sabo, S. L., Ikin, A. F., Buxbaum, J. D., and Greengard, P. (2003) The amyloid precursor protein and its regulatory protein, FE65, in growth cones and synapses in vitro and in vivo. *J. Neurosci.* **23**, 5407-5415
56. Enneking, E. M., Kudumala, S. R., Moreno, E., Stephan, R., Boerner, J., Godenschwege, T. A., and Pielage, J. (2013) Transsynaptic coordination of synaptic growth, function, and stability by the L1-type CAM Neuroglian. *PLoS Biol.* **11**, e1001537
57. Wang, L., Li, H., Zhang, Y., Santella, R. M., and Weinstein, I. B. (2009) HINT1 inhibits beta-catenin/TCF4, USF2 and NFkappaB activity in human hepatoma cells. *Int. J. Cancer* **124**, 1526-1534
58. Cao, J. Y., Shire, K., Landry, C., Gish, G. D., Pawson, T., and Frappier, L. (2014) Identification of a novel protein interaction motif in the regulatory subunit of casein kinase 2. *Mol. Cell. Biol.* **34**, 246-258
59. Proshkin, S. A., Shematorova, E. K., Souslova, E. A., Proshkina, G. M., and Shpakovski, G. V. (2011) A minor isoform of the human RNA polymerase II subunit hRPB11 (POLR2J) interacts with several components of the translation initiation factor eIF3. *Biochemistry (Mosc.)* **76**, 976-980
60. Durrin, L. K., and Krontiris, T. G. (2002) The thymocyte-specific MAR binding protein, SATB1, interacts in vitro with a novel variant of DNA-directed RNA polymerase II, subunit 11. *Genomics* **79**, 809-817
61. Fanciulli, M., Bruno, T., Di Padova, M., De Angelis, R., Iezzi, S., Iacobini, C., Floridi, A., and Passananti, C. (2000) Identification of a novel partner of RNA polymerase II subunit 11, Che-1, which interacts with and affects the growth suppression function of Rb. *FASEB J.* **14**, 904-912
62. Cheung, H. N., Dunbar, C., Morotz, G. M., Cheng, W. H., Chan, H. Y., Miller, C. C., and Lau, K. F. (2014) FE65 interacts with ADP-ribosylation factor 6 to promote neurite outgrowth. *FASEB J.* **28**, 337-349
63. Kypri, E., Christodoulou, A., Maimaris, G., Lethan, M., Markaki, M., Lysandrou, C., Lederer, C. W., Tavernarakis, N., Geimer, S., Pedersen, L. B., and Santama, N. (2014) The nucleotide-binding proteins Nubp1 and Nubp2 are negative regulators of ciliogenesis. *Cell. Mol. Life Sci.* **71**, 517-538
64. Zheng, Y., and Lu, Z. (2013) Regulation of tumor cell migration by protein tyrosine phosphatase (PTP)-proline-, glutamate-, serine-, and threonine-rich sequence (PEST). *Chin. J. Cancer* **32**, 75-83

65. Li, W., Wang, C., Juhn, S. K., Ondrey, F. G., and Lin, J. (2009) Expression of fibroblast growth factor binding protein in head and neck cancer. *Arch. Otolaryngol. Head Neck Surg.* **135**, 896-901
66. Kolpakova, E., Wiedlocha, A., Stenmark, H., Klingenberg, O., Falnes, P. O., and Olsnes, S. (1998) Cloning of an intracellular protein that binds selectively to mitogenic acidic fibroblast growth factor. *Biochem. J.* **336 ( Pt 1)**, 213-222
67. Tian, Y., Chang, J. C., Fan, E. Y., Flajolet, M., and Greengard, P. (2013) Adaptor complex AP2/PICALM, through interaction with LC3, targets Alzheimer's APP-CTF for terminal degradation via autophagy. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17071-17076
68. Gharesouran, J., Rezazadeh, M., Khorrami, A., Ghojzadeh, M., and Talebi, M. (2014) Genetic Evidence for the Involvement of Variants at APOE, BIN1, CR1, and PICALM Loci in Risk of Late-Onset Alzheimer's Disease and Evaluation for Interactions with APOE Genotypes. *J. Mol. Neurosci.*
69. Du, Y., Meng, J., Huang, Y., Wu, J., Wang, B., Ibrahim, M. M., and Tang, J. (2014) Guanine nucleotide-binding protein subunit beta-2-like 1, a new Annexin A7 interacting protein. *Biochem. Biophys. Res. Commun.* **445**, 58-63

**Table 1. Real-time Q-PCR primers.**

Human primers for all Real-time Q-PCR experiments, using Platinum SYBR Green qPCR SuperMix-UDG with ROX.

**Table 2. Novel teneurin-1 ICD interaction partners.**

Information about all novel interaction partners of the TEN1-ICD, as identified in the Yeast-2 Hybrid screen.

**Figure 1. Novel interaction partners of teneurin-1 ICD identified by Yeast-2 Hybrid screen.**

A. Dual Hunter Yeast-2 Hybrid screen. 1. When the teneurin-1 ICD (bait protein) and the prey protein do not interact, the reporter genes are not expressed. 2. Once an interaction takes place, the C- and N-terminal parts of split ubiquitin are united. 3. An ubiquitin-specific protease releases the attached transcription factor. 4. The transcription factor switches on expression of the HIS3/LacZ reporters. B. Interaction partners identified by the screen. Left panel: Serial dilutions of yeast strain NMY51 transformed with the prey-NubG plasmid or an empty prey-NubG plasmid (Control), together with either the membrane-anchored bait plasmid TEN1-ICD-Cub-LexA-VP16 (+) or the membrane-bound empty bait plasmid Cub-LexA-VP16 (-) grown on -trp/-leu medium containing histidine. Center panel: The same dilutions of the transformed NMY51 yeast as shown in the left panel, grown on medium lacking histidine, thus showing growth only if an interaction between bait and prey proteins took place. Right panel:  $\beta$ -galactosidase assay of yeast transformed with the bait TEN1-ICD-Cub-LexA-VP16 together with the prey plasmids indicated or with an empty prey plasmid (Control).

**Figure 2. Overexpression of the teneurin-1 ICD through a modified Tet-system in BS149 cells, a cell line expressing endogenous teneurin-1.**

A. Quantitative RT-PCR screen of teneurin-1 expression in glioblastoma cell lines. Values are normalized to TBP, fold change values are compared to the lowest expressing neuroblastoma cell line SH-SY5Y. B. Scheme of the modified Tet-system. Through the addition of dexamethasone (Dex) and doxycycline (Dox), the Tet-activator construct (irtTA/VP16/GBD) is released from HSP90 in the cytosol and induces expression of the ICD through the Tet-CMV promoter. C. Western blot with anti-GFP showing the expression of the control construct GFP-His, and the teneurin-1 ICD-GFP-His 24 h after the addition of Dex and Dox. Anti-vinculin is the internal control for equal loading. D. Quantitative RT-PCR results show overexpression of the teneurin-1 ICD, but not the Teneurin 1 ECD (endogenous control). Values were normalized to TBP, fold changes values compared to the negative control construct GFP-His.

**Figure 3. Microarray identifies seven up-regulated MITF target genes due to teneurin-1 ICD overexpression.**

A. A total of 430 genes are differentially regulated when the teneurin-1 ICD is overexpressed in BS149 cells compared to BS149 cells with overexpression of GFP only. Seven of these genes are MITF target genes indicated to the right. Overexpression of teneurin-1 ICD-GFP-His is compared to GFP-His. Biological triplicates were used for the microarray. B. Detailed view of the expression levels of the seven MITF target genes, where bright green means little or no expression and bright red means high expression. All seven genes are up-regulated with the fold change indicated. C. Quantitative RT-PCR confirmed the up-regulation of the MITF target genes. Values are normalized to TBP, fold change values compare overexpression of teneurin-1 ICD-GFP-His to GFP-His.

**Figure 4. Proximity Ligation Assay confirms the interaction between HINT1 and the TEN1-ICD.**A. Immunocytochemistry (ICC) images of COS-7 cells co-transfected with TEN1ICD-HA and HINT1-MYC

stained with anti-HA (red) and anti-myc (green) and the merged channels showing co-expression in yellow.

B. Proximity Ligation Assay (PLA) images of COS-7 cells co-transfected with TENE1ICD-HA and HINT1-MYC (left and center image) and COS-7 cells transfected only with HINT1-MYC as a negative control (right image); nuclei are stained with DAPI; white bar = 50  $\mu$ m.

**Figure 5. MITF-induced overexpression of *GPNMB* is attenuated by co-transfection of HINT1.**

A. A total of 497 genes are differentially regulated when MITF is overexpressed in BS149 cells. Of these genes, 42 are either up- or down-regulated due to both, MITF and Teneurin 1 ICD overexpression. One gene that is up-regulated in both microarrays is MITF target gene *GPNMB* (fold change after MITF induction in brackets). Overexpression of MITF-RFP-HA is compared to RFP-HA. Biological triplicates were used for the microarray.

B. Quantitative RT-PCR confirmed the up-regulation of *GPNMB* due to MITF overexpression. Values are normalized to TBP, fold change value compares overexpression of MITF-RFP-HA to RFP-HA.

C. Quantitative RT-PCR results show a reduction of MITF-dependent transcription of *GPNMB* in the presence of HINT1. Values are normalized to TBP, fold change values are compared to co-transfection of RFP-HA and CFP-C1 constructs (Neg. ctrl); graph is representative of four separate experiments. The inhibition of *GPNMB* due to HINT1 compares the co-transfection of MITF-RFP-HA and HINT1-CFP-MYC to the co-transfection of MITF-RFP-HA and CFP-C1. Co-transfected cells were FACS sorted before RNA isolation.

**Figure 6. The *GPNMB* promoter is induced by MITF in BS149 cells and further increased by overexpression of the teneurin-1 ICD.**

A. Diagram of the full length *GPNMB* promoter and the *GPNMB*  $\Delta$ M-box promoter, missing the M-box required for MITF-dependent regulation.

B. SEAP assays show that MITF can induce the full length *GPNMB* promoter activity while the *GPNMB*  $\Delta$ M-box promoter is not inducible. The transfection efficiency was normalized to luciferase assay values, fold change values indicated by FC above the bars are compared to the full length *GPNMB* promoter, co-transfected with empty pcDNA3.1 vector.

C. SEAP assays show an increase of full length *GPNMB* promoter activity when the teneurin-1 ICD is co-transfected with MITF, compared to MITF co-transfected with an empty pcDNA3.1 vector. The transfection efficiency was normalized to luciferase assay values and fold change values between the different conditions are indicated above the bars.

**Figure 7. The ICD of Teneurin 1 regulates MITF-dependent *GPNMB* expression by competing for HINT1.**

A. Teneurin-1 ECD is not bound to an interaction partner and the intracellular domain is not released from the plasma membrane. HINT1 in the nucleus inhibits MITF at the promoter of *GPNMB*.

B. 1. Teneurin-1 interacts with a transmembrane protein of another cell. 2. Both, the extracellular and intracellular domains are released from the membrane. The ICD translocates to the nucleus. 3. The ICD competes for HINT1 binding, thus switching on MITF-dependent transcription of *GPNMB*.

*Teneurin 1 ICD as a transcriptional regulator*

Gene Name	Forward primer	Reverse primer
Teneurin 1 ICD	CTACCTGGATTCCCAAACC	CCAGGATAGTCTTCCAGGA
Teneurin 1 ECD	GCATAGTTCCTGTTTGCCA	TCTGCACATCTTGAGTAGAC
GPNMB	AAGTGAAAGATGTGTACGTGGTAACAG	TCGGATGAATTCGATCGTTCT
SCARB1	AATAAGCCCATGACCCTGAAGC	GCCCCACATGATCTCACCC
EDNRB	CTGCTGCACATCGTCATTGAC	GCTCCAAATGGCCAGTCCT
SLC1A4	CAGCGACCCTCCCTCTATGA	GCCCCGATGGGGAGAATAAAC
SEMA6A	ACATTGCTGCTAGGGACCATA	TCTGCATGTGTCTACATCGGC
ERBB3	GTCTGTGTGACCCACTGCAACT	GGGTGGCAGGAGAAGCATT
CHL1	ACCAACATTTTCGTGGACTAAGG	TCGATGGAATTATCCGATGGTCA
TBP	TGCACAGGAGCCAAGAGTGAA	CACATCACAGCTCCCCACCA

Table 1  
Schöler et al.

Gene	Protein Name	Information
CSNK2B	Casein Kinase II subunit beta	Regulatory subunit of Casein Kinase II (CK2); not required for catalytic activity; recruits substrates/ regulators; CK2 is a pleiotropic kinase involved in e.g. cell proliferation, transcription, etc. (58)
POLR2J	DNA-directed RNA polymerase II subunit RPB11-a	Subunit of the RNA polymerase II complex; forms heterodimer with POLR2C, which is part of the core enzyme; involved in the termination of transcription; binds to SATB1 and Che-1 (59-61)
APBB1	Amyloid beta A4 precursor protein-binding family B member 1	Adapter protein; aids in processing of amyloid precursor protein (APP) by binding to it; this interaction also influences production of amyloid- $\beta$ peptides that are found in Alzheimer's patients; promotes neurite outgrowth (62)
BEX1	Brain-expressed X-linked protein 1	Small adapter protein; involved in NGFR signaling and transcriptional regulation of cell-cycle arrest genes; implicated in promoting survival of neurons and neurite outgrowth (50,51)
NUBP1	Nucleotide binding protein 1	MRP/MinD-type P-loop NTPase; interacts with motor protein KIFC5A; involved in regulation of centriole duplication; implicated in assembly of cytosolic iron-sulfur proteins (63)
HINT1	Histidine triad nucleotide-binding protein 1	Member of the evolutionary conserved HIT superfamily; tumor suppressor gene; inhibits transcription factors like MITF and $\beta$ -catenin by directly binding them at the promoters of their target genes (33)
MACF1	Microtubule-actin cross-linking factor 1	Spectraplakin and +TIP protein; conserved SH3 domain; role in crosslinking microtubules to actin cytoskeleton; regulates growth of neuronal microtubules, and subsequently filopodia formation and axon extension (52,53)
PTPN12	Tyrosine-protein phosphatase non-receptor type 12	Part of non-receptor PTP subfamily; ubiquitously expressed phosphatase; important functions in early embryogenesis, like development of mesenchyme; involved in cell spreading and migration (64)
FIBP	Acidic fibroblast growth factor intracellular-binding protein	Selectively binds aFGF; implicated in mitogenic action of aFGF; up-regulated in some cancers; involved in angiogenesis of tumors (65,66)
PICALM	Phosphatidylinositol binding clathrin assembly protein	Ubiquitously expressed adapter protein; strongest expression in neurons, especially in pre- and post-synaptic structures; risk factor for Alzheimer's disease; plays a role in clathrin-mediated endocytosis (67,68)
GNB2L1	Guanine nucleotide-binding protein	Scaffold protein; originally identified as anchoring protein for protein kinase C; also recruits and binds other proteins like integrins; altered expression levels in many cancers (69)

Table 2  
Schöler et al.



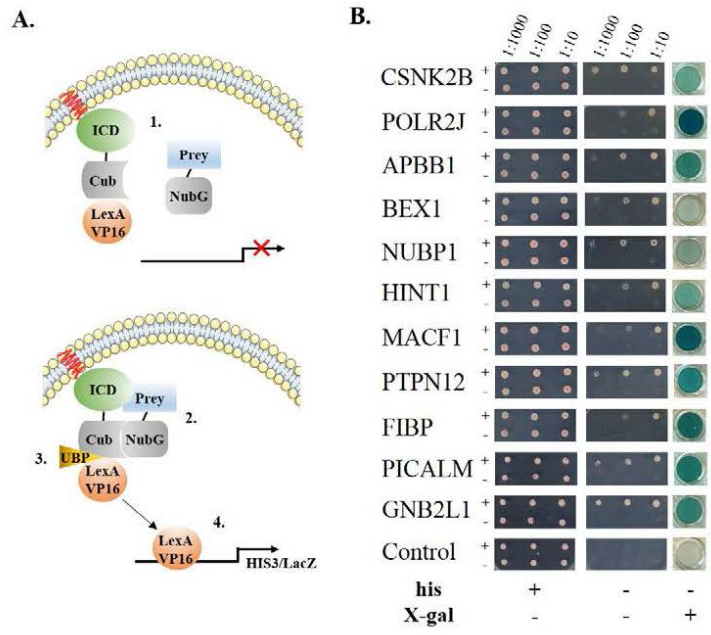


Figure 1  
Schöler et al.

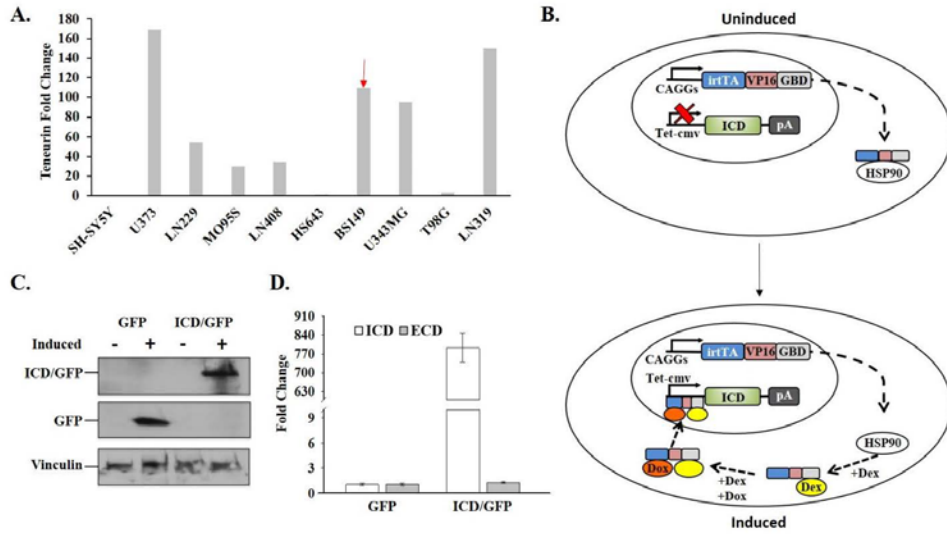


Figure 2  
Schöler et al.

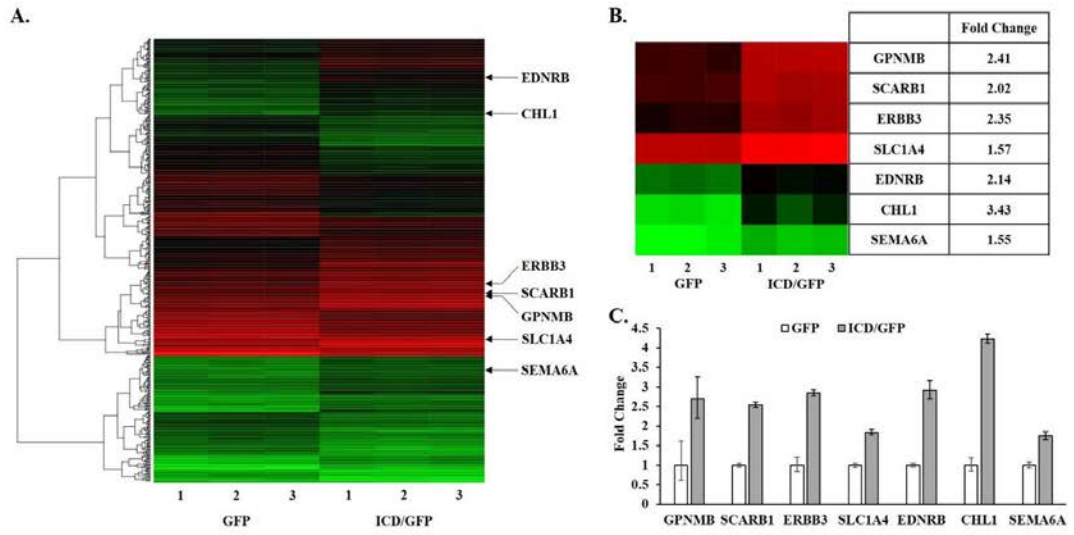


Figure 3  
Schöler et al.

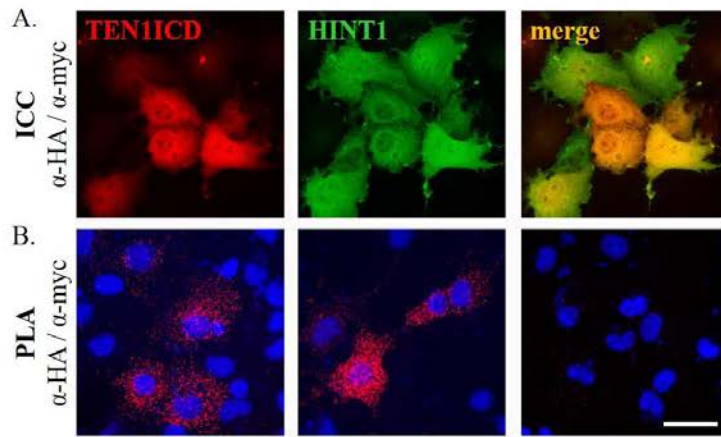


Figure 4  
Schöler et al.

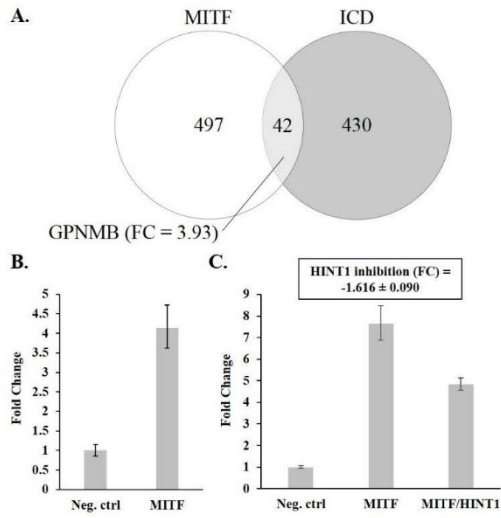


Figure 5  
Schöler et al.

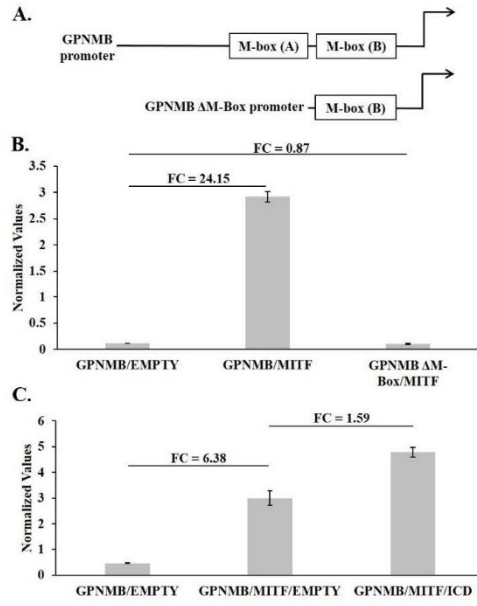


Figure 6  
Schöler et al.

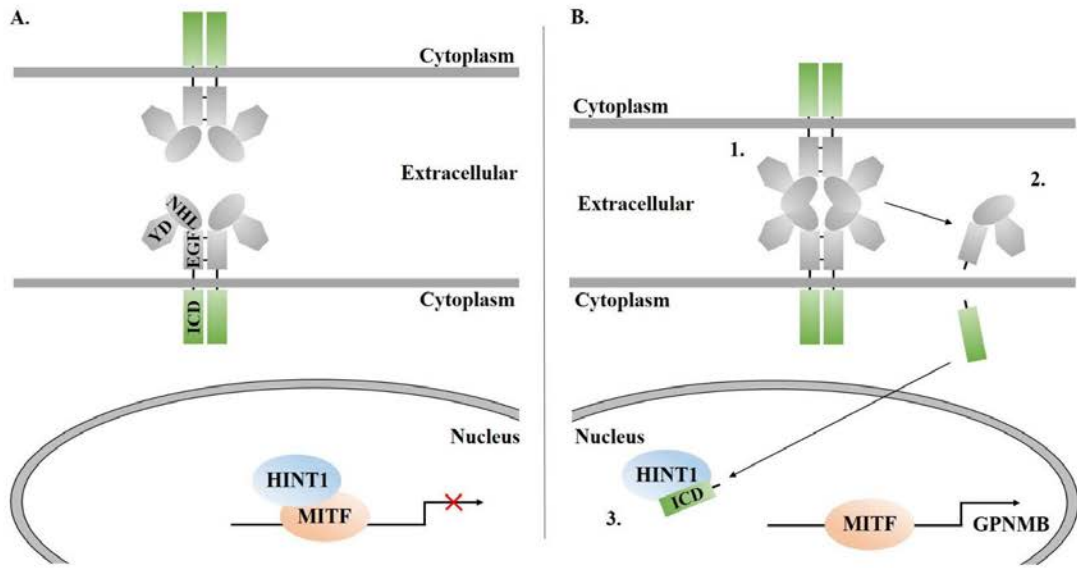


Figure 7  
Schöler et al.

## 5. Unpublished work

### 5.1 Methods and Materials

Together with the Protein Structure Facility (PSF) at the FMI, we have set up a protocol for the expression and purification of the chicken teneurin-1 (cten-1) and -2 (cten-2) NHL repeat domains in a bacterial system. For the general bacterial expression and purification protocols established by the PSF, please refer to Appendix C.

#### 5.1.1 Cloning and testing of the cten-1 and -2 NHL repeat domain constructs

All NHL repeat domain constructs were cloned from full length cten-1 and -2 constructs. Full length constructs were provided by Jan Beckmann, a former PhD student in the lab; preparation of the constructs is described in <sup>88</sup>. NHL repeat domain constructs were prepared with specific primers by in-fusion cloning (Clontech) into the pOPINF bacterial expression vector (Addgene) (Table 5.1).

**Table 5.1 Primers used for in-fusion cloning of constructs**

Primer Name	Primer Sequence
pOPINF cten1 for1112	AAGTTCTGTTTCAGGGCCC GAACAAACACCATGTA TTAAATCCACAAAGTGGAATTG
pOPINF cten1 for1164	AAGTTCTGTTTCAGGGCCC GAACAGCAAACCTCTTT GCCCCAGTTGCTC
pOPINF cten1 rev1510	ATGGTCTAGAAAGCTTTTAGTCGCTGAGGTGAGCCT TGTTTCGGCT
pOPINF cten1 rev1450	ATGGTCTAGAAAGCTTTAATCTGACGGGGCTCCGG CGATTATGGA
pOPINF cten2 for1210	AAGTTCTGTTTCAGGGCCC GGATAAACATCATGTA CTGAATGTCAAGAGTGGTATATTGCAC
pOPINF cten2 for1262	AAGTTCTGTTTCAGGGCCC GGGAAATAAGCTTTTG GCCCTGTAGCACTG
pOPINF cten2 rev1607	ATGGTCTAGAAAGCTTTAAGAATTAAGAATGGGCC TGTTTTACTGACAGCCCTAATG
pOPINF cten2 rev1547	ATGGTCTAGAAAGCTTTAGTCTGAAGCTGCCCTG CAAGAAGGCATATTT



Bioinformatics software (HHpred, I-TASSER, Phyre-2) predicted the different size constructs, predicted to contain the beta-propeller, (Table 5.2).

**Table 5.2 Construct boundaries and sizes**

Construct Name (includes boundaries)	Length of construct	Approximate MW (kDa)
pOPINF cten1 1112_1510	418	45.4
pOPINF cten1 1164_1510	366	39.8
pOPINF cten1 1112_1450	358	39.0
pOPINF cten1 1164_1450	306	33.5
pOPINF cten2 1210_1607	417	45.1
pOPINF cten2 1262_1607	365	39.5
pOPINF cten2 1210_1547	357	38.7
pOPINF cten2 1262_1547	305	33.1

Transform the pOPINF vectors containing the different size constructs into TAM1 cells (Active Motif) and grow overnight (o/n) at 37°C. Isolate the plasmids with the QIAprep Spin Miniprep kit (Qiagen). Vectors can be verified by sequencing with T7 forward and pOPIN reverse primers.

Small-scale expression of the constructs was tested in Origami2 DE3 cells (Merck Millipore). Origami2 DE3 cells are *E. coli* K-12 derivatives that enhance disulfide bond formation. Note that the recipes of all buffers and media are listed at the end of Materials and Methods, in section 5.4. Transform 1 µl of the minipreps into the bacteria on LB-agar plates containing 50 µg/ml carbenicillin and 12.5 µg/ml tetracycline. Grow o/n at 37°C and set up starter cultures the next day in 96-DWB plates (Abgene). Pick single colonies and inoculate 1.5 ml GS96 medium (Recipe I), containing glycerol, 1% glucose, 50 µg/ml carbenicillin, and 12.5 µg/ml tetracycline, and seal the DWB with a gas-permeable adhesive seal (Abgene). Shake the plate o/n at 225 rpm and 37°C in a Kühner DWB holder. The next morning, dilute the starter culture at 1:50 in 2.5 ml ZYP-5052 autoinduction medium (Recipe II), supplemented with 50 µg/ml carbenicillin and 12.5 µg/ml tetracycline, in a 24-DWB plate (Abgene). Shake the plate 4-

5 hours at 225 rpm and 37°C until cultures are cloudy, reduce temperature to 20°C and grow another 20 hours. Transfer 1.2 ml of each well into a 96-DWB plate, and spin down for 30 minutes at 6000 x g and 4°C, using a Beckman Coulter JS5.3 rotor. Carefully remove supernatant and freeze pellets on dry ice for at least 30 minutes. Cell pellets can be directly used for small-scale purification or stored at -80°C.

Re-suspend cell pellets in 280 µl NPI-10-Tween buffer (Recipe III), supplemented with 1 mg/ml lysozyme and 3 units/ml Benzonase (both Sigma). Shake DWB plate in an orbital shaker for 30 minutes at 1400 rpm and 4°C. Spin down for 30 minutes at 6000 x g and 4°C. During the centrifugation step, dispense 20 µl of Ni-NTA bead suspension (Qiagen) into a flat-bottom 96-well plate (MTP, Qiagen). Transfer 180 µl of the cleared lysate into the 96-well plate containing the Ni-NTA bead suspension. Mix for 30 minutes at room temperature (RT) and 800 rpm on a vortex. Place plate on a 96-well type A magnet (Qiagen) for 2 minutes and carefully remove supernatant. To run the different fractions of the small-scale purification on a protein gel, take a sample of the supernatant, called flow-through (FT). Add 200 µl NPI-20-Tween wash buffer (Recipe IV) to the beads and mix on the vortex shaker for 5 minutes. Place plate on the 96-well type A magnet for 2 minutes and carefully remove wash buffer. Repeat wash step once more. Take samples of each wash step (W1, W2). Add 25 µl of NPI-250-Tween elution buffer (Recipe V) to each of the wells and mix on the vortex shaker for 2 minutes. Place plate on the 96-well type A magnet for 2 minutes and transfer the eluate to clean Eppendorf tubes (E). Mix 15 µl of the eluate with 5 µl of sample buffer containing DTT, and load 15 µl onto a 17-well NuPAGE Novex Bis-Tris protein gel and run for 35 minutes at 200 V in NuPAGE MES SDS running buffer, the inner running chamber also containing NuPAGE LDS sample buffer. Take gel out of running chamber, remove from the plastic plates it is cast in, and stain in an appropriate plastic

dish with enough InstantBlue (Expedeon via Lucerna) to coat the gel. Depending on the amount of protein, bands will start to appear in a matter of minutes. Viewing of the bands is improved on a bright light source.

### 5.1.2 Large-scale expression and purification of the cten-2 1262-1607 construct

Optimally, pick just one construct that was well-expressed in the small-scale expression test (in our case cten-2 1262-1607) and focus on its purification before moving on to the next construct. The large-scale expression and purification process requires time, resources and enough space, especially in incubators.

Take a 20  $\mu$ l aliquot of Origami 2 competent bacteria and leave it on ice for 5 minutes. Gently add 1  $\mu$ l of 10 ng/ $\mu$ l pOPINF + cten-2 (1262-1607) construct to the middle of the bacteria, flick the tube once, and incubate for 5min on ice. Heatshock for 30 seconds at 42°C in a waterbath. Put back on ice for 2 minutes. Add 80  $\mu$ l of GS96 + 1% Glucose medium to the bacteria and flick gently. Incubate at 37°C for 1 hour without shaking. Carefully spread on a pre-warmed/dry LB-Agar plate, supplemented with 12.5  $\mu$ g/ml Tetracycline and 50  $\mu$ g/ml Carbenicillin. Incubate o/n at 37°C. Keep at 4°C during the day, but inoculate starter culture the same day. Inoculate an 80ml starter culture of GS96 + 1% Glucose medium supplemented with 12.5  $\mu$ g/ml Tetracycline and 50  $\mu$ g/ml Carbenicillin with 6 colonies. Incubate overnight at 37°C at 225 rpm. Continue with the large-scale culture the next morning. Add supplements (20x NPS, 50x 5052, 1M MgSO<sub>4</sub>) and antibiotics (50  $\mu$ g/ml Carbenicillin, 12.5  $\mu$ g/ml Tetracycline) to the ZY medium to make a total of 8x 500ml culture (Recipe VI). Dilute 5 ml of overnight starter culture (i.e. 1:100) into each of the 2 L flasks, containing the 500 ml medium. Grow for 5 hours at 37°C at 225rpm, then turn down the temperature to 20°C for 20 hours. Equally distribute the 4

L of culture into 2x 2 L centrifugation containers. Take 1ml sample from each container for a small-scale expression test; i.e. to determine whether the construct was expressed. Harvest cells by spinning down both, the 2 L and the 1 ml samples for 30 minutes at 6500 x g and 4°C.

Remove the medium from the 1 ml samples, directly freeze on dry ice, and store at -80°C until needed. Also decant medium from the 2 L cultures. Resuspend each cell pellet in 100 ml Lysis Buffer (Recipe VII) until it is a homogenous mixture, freeze on dry ice in 50 ml Falcon tubes and store at -80°C until needed. For the small-scale expression test, quickly thaw the cell pellets at RT and follow the protocol of Chapter 5.1.1, starting at re-suspending the cells in NPI-10-Tween buffer. If there are few samples (in this case two), the small-scale expression test should be done in 1.5 ml Eppendorf tubes, rather than 24- and 96-DWB plates.

After determining whether the construct is expressed on a small scale, continue with the large-scale purification. Slowly thaw re-suspended pellets in beaker filled with water at room temperature; exchange water regularly. Meanwhile prepare more lysis buffer, add 4 small tablets of protease inhibitor cocktail (Roche) to 20 ml of lysis buffer, and 4 µl of Benzonase (Sigma). Incubate lysis buffer on a roller until the tablets have been dissolved. Once the re-suspended bacterial pellets are thawed, filter out any debris by pipetting the suspension into a funnel containing a piece of cloth, and collecting the flow-through in a clean beaker. Use a cooled down EmulsiFlex-C3 cell disruptor (AVESTIN) to lyse the cells. Repeat this step once. Centrifuge the lysate in the Beckman Coulter JA-17 rotor at 30,000 x g for 30 minutes at 4°C. Collect the soluble fraction by decanting the supernatant into a clean beaker, keeping it on ice. Filter the soluble fraction through a 0.45 µm minisart plus filter (Sartorius Stediim) into 50 ml Falcon tubes. Take an aliquot of the soluble cleared lysate (CL) to run as a control on a protein gel. Use 2 ml of 50% Ni-NTA agarose beads (Qiagen) for 200 ml of cleared lysate, i.e. 1 ml of pure

beads. Incubate the cleared lysate – beads mixture on a roller for 1 hour at 4°C. Spin down the beads at 500 x g for 5 minutes at 4°C. Carefully transfer the beads with a 5ml pipette into a 25ml polyprep column (Bio-Rad). Collect flow-through (FT), and take an aliquot to run as a control on a protein gel. The beads should at least be covered by a minimal amount of liquid at all times. Wash the beads twice with 10:1 volumes of Nickel wash buffer (recipe VIII) to Ni-NTA agarose beads. Collect each wash step separately (W1, W2) and take an aliquot to run as a control on a protein gel. Transfer the washed beads into a 15 ml Falcon tube and fill with Nickel wash buffer to 14.5 ml. Add 0.5 ml of 3C protease. The final volume should be close to the maximum volume of the tube, so the beads cannot stick to the plastic. Mix well, then incubate o/n at 4°C. Keep a small aliquot of the 3C protease (3C) to run as a control on a protein gel. Transfer the beads into a 25 ml polyprep column by pouring. Collect the flow-through (3C FT) and take an aliquot to run as a control on a protein gel. Pour 5 ml Nickel wash buffer into the Falcon tube to wash out the rest of the beads and pour into the column. Take an aliquot from this wash step (3C W1), and a second wash step (3C W2), to run both as controls on a protein gel. Elute the impurities (3C E) with 10 ml of elution buffer (recipe IX). Mix 15 µl of the samples with 5 µl of sample buffer containing DTT, and load 15 µl onto a 12-well NuPAGE Novex Bis-Tris protein gel and run for 35 minutes at 200 V in NuPAGE MES SDS running buffer, the inner running chamber also containing NuPAGE LDS sample buffer. Take gel out of running chamber, remove from the plastic plates it is cast in, and stain in an appropriate plastic dish with enough InstantBlue to coat the gel.

Samples 3C FT, 3C W1, and 3C W2 should contain the protein to be purified. Depending on the amount of protein in each sample, combine all, or at least some of the samples for continuation of the large-scale purification. Concentrate combined samples using Amicon tubes

with 10 kDa cutoff (Merck Millipore), to a volume of less than 2 ml. Load the sample onto an Äkta Purifier, using the Hi-load 16/60 S200 column, for size-exclusion chromatography. Collect at least 95 fractions, also containing the protein to be purified. Mix 15  $\mu$ l of the fractions around the predicted protein peak with 5  $\mu$ l of sample buffer containing DTT, and load 15  $\mu$ l onto a 12-well NuPAGE Novex Bis-Tris protein gel and run for 35 minutes at 200 V in NuPAGE MES SDS running buffer, the inner running chamber also containing NuPAGE LDS sample buffer. Take gel out of running chamber, remove from the plastic plates it is cast in, and stain in an appropriate plastic dish with enough InstantBlue to coat the gel. Decide which fractions to pool based on the protein, depending on the amount of protein and impurities in the samples. Concentrate the combined fractions to a relatively high protein concentration, first using Amicon tubes with a 10 kDa cutoff and then a 30 kDa cutoff. Relatively high protein concentration means at least 4 mg/ml, but as high as possible before too much of the protein precipitates out of solution, depending on the properties of the protein. Protein drops can now be set up in hanging drop crystal growth chambers with a Phoenix nano-liter crystallization robot (ARI). Several crystallizing conditions, like differing buffers, pH, temperature, or adding proteases can be tested. Buffer screens were tested in this study at RT as follows: IndexHT, JBS Solubility, JCSGI, JCSGII, JCSGIII, JCSGIV, PEGSI, PEGSII (all Qiagen), MIDAS, Morpheus, PGA (all Molecular Dimensions). IndexHT, JCSGI, JCSGII, JCSGIII, JCSGIV screens were also tested at RT with trypsin and chymotrypsin, and at 4°C.

### 5.1.3 Limited proteolysis

Limited proteolysis determines the boundaries of the protein domain to be purified, and whether to add any proteases to the crystallization conditions. All limited proteolysis

experiments were set up together with the Protein Analysis Facility (PAF) at the FMI, using previously purified protein of the cten2 NHL repeat domain (1262-1607). For the general limited proteolysis protocol established by the PAF, please refer to Appendix D. Trypsin was used as the protease to determine the boundaries.

Dissolve enough trypsin powder in 20 mM Tris pH 8.0 + 0.2M NaCl to prepare a 1 mg/ml stock solution of trypsin. Determine the exact protein concentration by Bradford assay. Mix 10  $\mu$ l of the purified protein solution, optimally containing 10  $\mu$ g of protein, with differing amounts of trypsin in 10  $\mu$ l of solution (Table 5.3). Incubate 1.5 hours at 37°C and then stop reaction with phenylmethanesulfonylfluoride (PMSF). Run samples on a 10% polyacrylamide gel, but switch off before dye front has run out of the gel. Stain gel with InstantBlue to view the protein fragment bands.

**Table 5.3 Starting concentrations of trypsin in limited proteolysis**

Total amount of trypsin	Concentration in 20 $\mu$ l volume
0.1 ng	10 $\mu$ l at 0.01 ng/ $\mu$ l
0.3 ng	10 $\mu$ l at 0.03 ng/ $\mu$ l
1 ng	10 $\mu$ l at 0.1 ng/ $\mu$ l
3 ng	10 $\mu$ l at 0.3 ng/ $\mu$ l
10 ng	10 $\mu$ l at 1 ng/ $\mu$ l
30 ng	10 $\mu$ l at 3 ng/ $\mu$ l
100 ng	10 $\mu$ l at 10 ng/ $\mu$ l
300 ng	10 $\mu$ l at 30 ng/ $\mu$ l

Once you have decided, which trypsin concentration gives you several clear and distinct fragments, expand on this concentration to optimize this step of the process. Incubation time is another variable. Then, up-scale the trypsin concentration to 40-50  $\mu$ g of purified protein and incubate at 37°C for the determined amount time. Stop reaction with PMSF and analyze samples by liquid chromatography-mass spectrometry; LTQ Orbitrap Velos (Thermo Scientific) equipped

with a Thermo EASY-Spray source or the New Objective Digital Pico View source, coupled to an EASY-nLC 1000 Liquid Chromatograph and using Scaffold software (Proteome Software).

#### 5.1.4 Recipes of buffers and media

##### Recipe I – GS96 medium (per 1 L)

49.1 g GS96 powder (MP biomedical)

0.5 ml glycerol (Sigma)

Fill up with Milli-Q water, autoclave, and store at 4°C.

##### Recipe II – ZYP-5052 medium

ZY medium 463 ml (see recipe VI)

1 M MgSO<sub>4</sub> 1 ml

50 x 5052 10 ml

20 x NPS 25 ml

1 M MgSO<sub>4</sub>:

24.65 g MgSO<sub>4</sub>·7H<sub>2</sub>O (Merck)

Fill up to 500 ml with milli-Q H<sub>2</sub>O.

50x 5052 (500 ml):

125 g glycerol (Merck)

365 ml milli-Q H<sub>2</sub>O

12.5 g glucose (Sigma)

50 g α-lactose (Sigma)

Fill up to 500 ml with milli-Q H<sub>2</sub>O.



20x NPS:

450 ml        milli-q H<sub>2</sub>O  
33 g         (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (Merck)  
68 g         KH<sub>2</sub>PO<sub>4</sub> (Merck)  
71 g         Na<sub>2</sub>HPO<sub>4</sub> (Merck)

Fill up to 500 ml with milli-Q H<sub>2</sub>O.

Recipe III – NPI-10-Tween buffer

50 mM NaH<sub>2</sub>PO<sub>4</sub> (Merck)  
300 mM NaCl (Sigma)  
10 mM imidazole (Sigma)  
1% v/v Tween 20 (Sigma)

Adjust pH to 8.0 using NaOH (Merck); sterile-filter and store at 4°C.

Recipe IV – NPI-20-Tween wash buffer

50 mM NaH<sub>2</sub>PO<sub>4</sub> (Merck)  
300 mM NaCl (Sigma)  
20 mM imidazole (Sigma)  
0.05% v/v Tween 20 (Sigma)

Adjust pH to 8.0 using NaOH (Merck); sterile-filter and store at 4°C.

Recipe V – NPI-250-Tween elution buffer

50 mM NaH<sub>2</sub>PO<sub>4</sub> (Merck)  
300 mM NaCl (Sigma)  
250 mM imidazole (Sigma)  
0.05% v/v Tween 20 (Sigma)

Adjust pH to 8.0 using NaOH (Merck); sterile-filter and store at 4°C.

Recipe VI – ZY medium (per 500 ml)

5 g tryptone (BD)

2.5 g yeast extract (BD)

463 ml milli-Q water

Autoclave, and store at 4°C.

Recipe VII – Lysis buffer

50 mM Tris pH 7.5 (VWR)

500 mM NaCl (Sigma)

20 mM imidazole (Sigma)

0.2% Tween (Add 2 ml 10% Tween 20 to 100 ml of Wash buffer) (Sigma)

0.2 µM sterile-filtered and degassed (just before use).

Recipe VIII – Nickel wash buffer

50 mM Tris pH 7.5 (VWR)

500 mM NaCl (Sigma)

20 mM imidazole (Sigma)

0.2 µM sterile-filtered and degassed (just before use).

Recipe IX – Elution buffer

50 mM Tris pH 7.5 (VWR)

500 mM NaCl (Sigma)

500 mM imidazole (Sigma)

0.2 µM sterile-filtered and degassed (just before use).

## 5.2 Results

In the recent paper by Beckmann et al., the authors determined through atomic force microscopy/single-molecule force spectroscopy experiments which domain is responsible for the homophilic rather than heterophilic interaction of the teneurin extracellular domains.<sup>88</sup> By swapping the different full length chicken teneurin-1 and -2 domains they showed that the NHL repeat domain determines the specificity of the homophilic interaction. Interestingly, NHL repeat domains are classic beta-propellers, suggesting that a beta-propeller is responsible for determining this specificity. Thus, we were interested in purifying the teneurin-1 and -2 NHL repeat domains, set up crystallization studies to determine the structure and elucidate the binding mechanism of these domains.

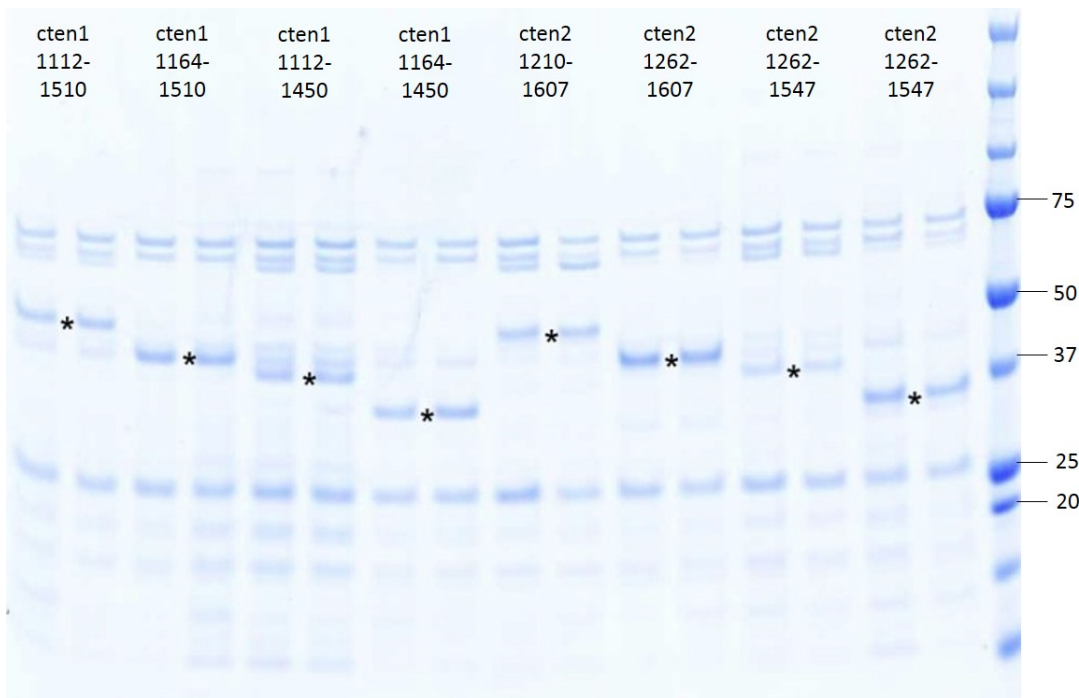
After insertion into the pOPINF expression vector, the constructs were verified by sequencing. There were several point mutations in the constructs, which were already present in the original full length chicken teneurin-1 and -2 constructs (Table 5.4). However, these point mutations did not appear to be detrimental to the structure or function of the NHL repeat domain and were thus not corrected by site-directed mutagenesis.

**Table 5.4 Point mutations identified in chicken teneurin-1 and -2 NHL repeat domain constructs**

Teneurin 1 Mutations	Teneurin 2 Mutation
M 1144 V	D 1451 N
P 1419 L	
R 1422 A	

Next, we tested the expression of the different constructs on a small scale. Experiments were performed in duplicates by auto-induction in Origami2 DE3 cells (Figure 5.1). Since the predicted beta-propellers in cten-1 and -2 also contain several disulfide bonds, we chose to work

with Origami2 DE3 cells. These particular competent bacteria have a mutation in their thioredoxin reductase (trxB) and glutathione reductase (gor) genes, which according to Merck Millipore significantly enhances the formation of disulfide bonds.

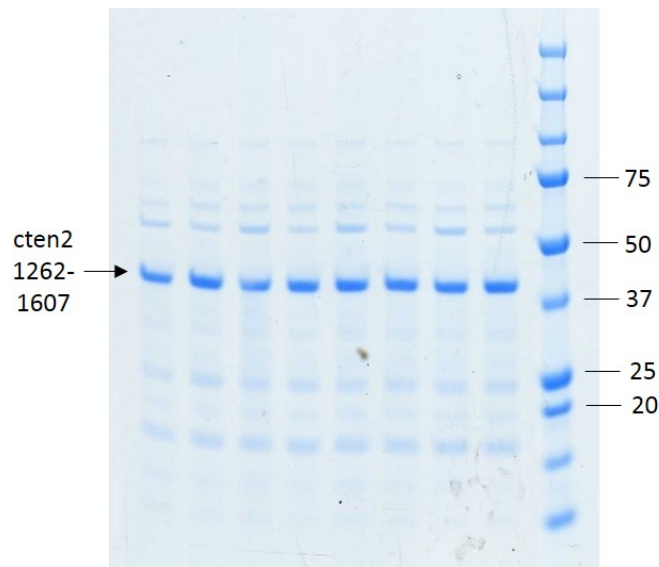


**Figure 5.1 Chicken teneurin-1 and -2 NHL repeat domain constructs**

Small-scale expression test of cten1 and cten2 constructs in duplicates predicted to contain the beta-propeller domain. Test expression in Origami2 DE3 cells by auto-induction. Star (\*) denotes the correct size of the constructs.

Teneurin-2 (1262-1607) was the larger of the two well-expressing constructs and we decided to focus on this particular construct first. The other construct, teneurin-2 (1262-1547) is missing 60 amino acids at the C-terminal part of the domain, but is identical otherwise. Using the larger construct increases the chance that the complete predicted beta-propeller will be expressed. Missing part of the domain could be detrimental to the crystallization of the protein, since it could be misfolded and unstable. However, the construct should not be too large either. Flexible and unfolded domains on either side of the well-structured domain to be crystallized can

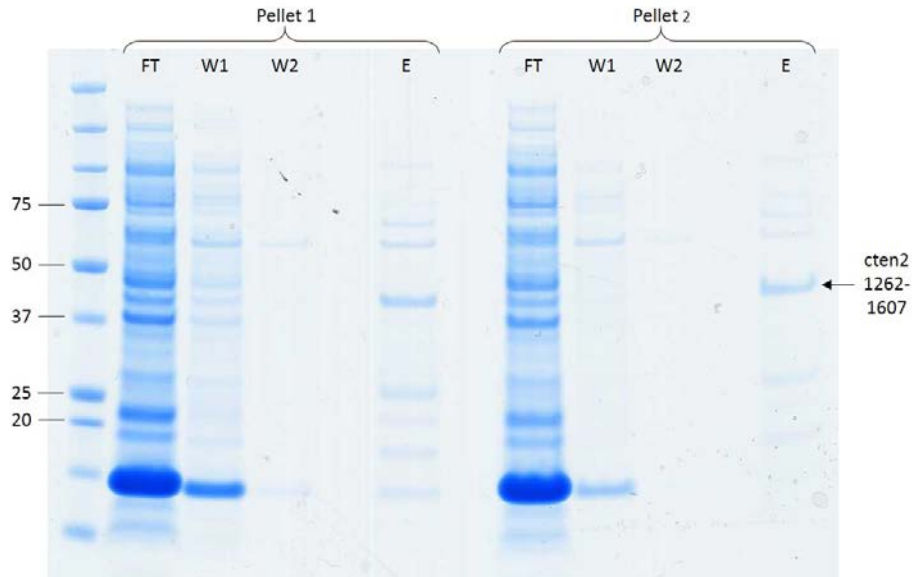
prevent it from tightly packing in an organized manner. While the predicted beta-propeller would be intact, this can also keep the protein from crystallizing. After choosing the cten-2 (1262-1607) construct, we up-scaled the expression system to purify larger quantities of the protein. Proteins have to be highly concentrated (ideally  $>4.5 \mu\text{g/ml}$ ) for setting up crystallization studies. We performed four purifications of the chosen construct, the results below showing a representative round of purification.



**Figure 5.2 Retest of chosen construct**

Confirmation of the expression of the cten2 (1262-1607) construct on a small-scale in eight randomly picked colonies.

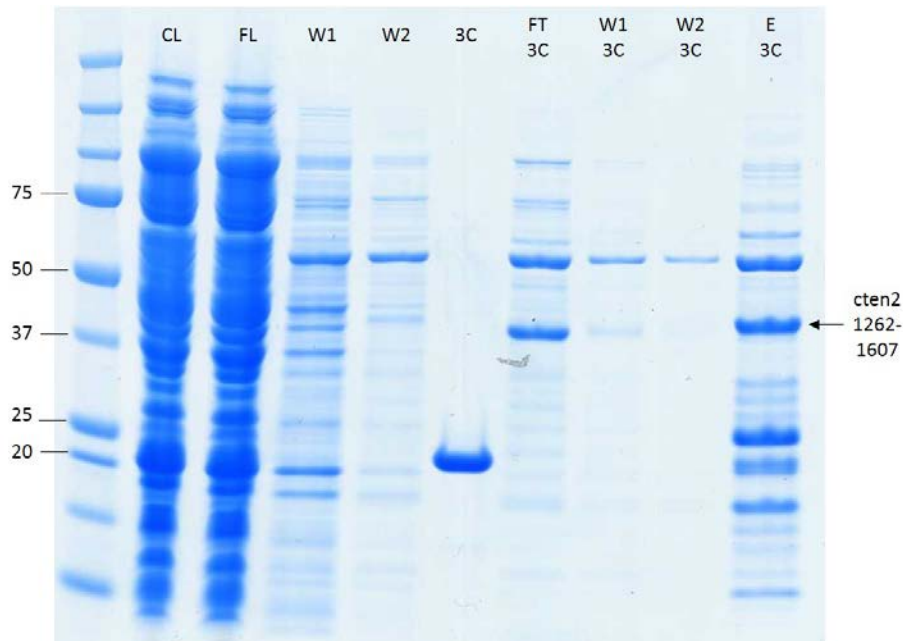
Initially, we retested the expression of the construct on a small-scale before continuing with the large-scale culture. All eight randomly picked colonies expressed the construct at about 39.5 kDa (Figure 5.2). After upscaling the bacterial culture to 4 L, we again tested the expression by taking a 1 ml sample from each of the two 2 L cultures, before moving on to the large-scale purification (Figure 5.3). To save time in future experiments using this construct, we only test expression after up-scaling the bacterial culture.



**Figure 5.3 Pretesting expression in large-scale culture**

Small-scale expression test of the two 2 L bacterial cultures, before continuing with large-scale purification. FT = flow through; W1 = wash 1; W2 = wash 2; E = elution; fractions on the gel are explained in more detail in Chapter 5.1.

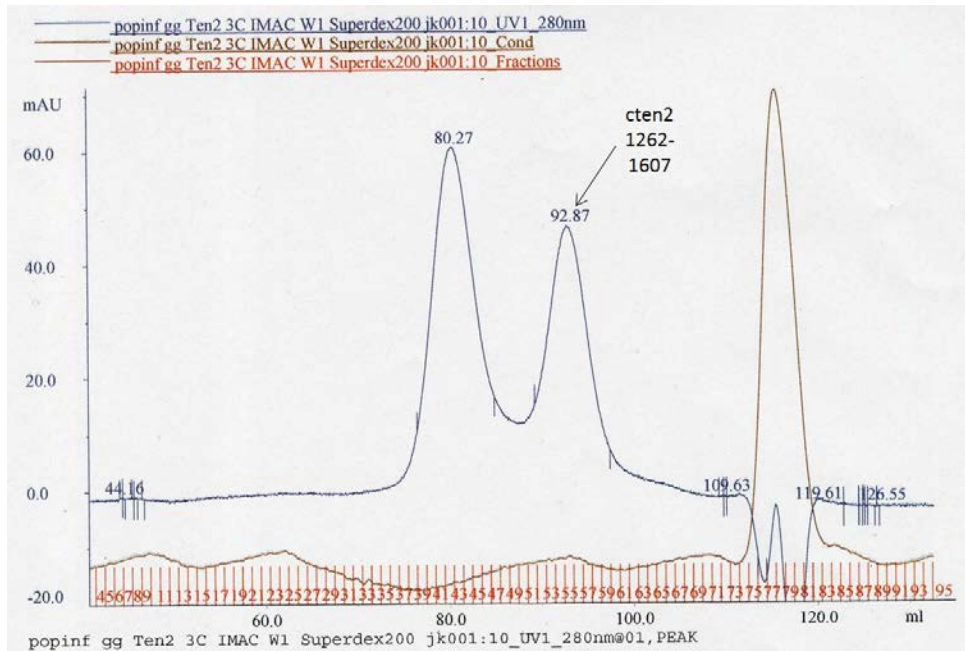
Once we lysed the cells of the large scale bacterial pellet and removed the insoluble fraction and other cell debris, we started with purification by Ni-NTA beads. All constructs in the pOPINF vectors are also fused to a His-tag. Thus, we first removed impurities that do not bind to the beads by several washing steps. Subsequently we removed impurities that do stick to beads, by cleaving the His-tag off the cten-2 construct with 3C-protease and washing it out of the column. The protease also contains a His-tag, thus binding to the beads and not contaminating the sample. Most of the purified protein is found in the 3C FT and 3C W1 fractions. However, at this point the sample contains too many impurities (Figure 5.4).



**Figure 5.4 Large-scale purification by Ni-NTA agarose beads**

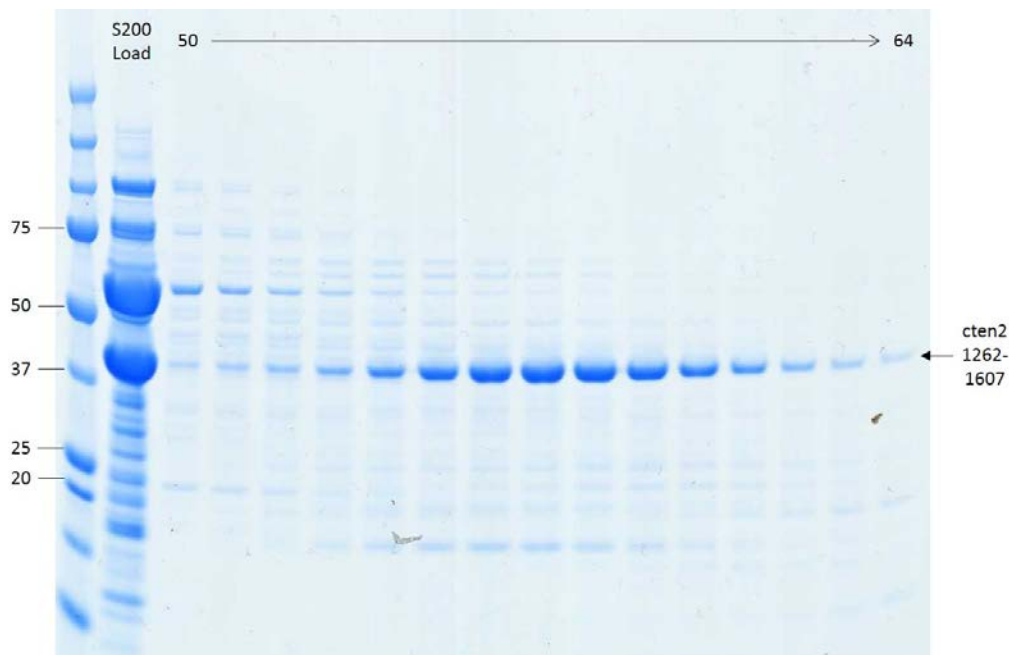
Fractions taken throughout the large-scale purification process, using Ni-NTA agarose beads. The cten2 (1262-1607) construct sticks to the beads via its His-tag, which is cleaved off by 3C-protease, and the construct is subsequently washed out. CL = cleared lysate; FL = flow through; W1 = wash 1; W2 = wash 2; 3C = 3C-protease; FT 3C = flow through post-3C cleavage; W1 3C = wash 1 post-3C cleavage; W2 3C = wash 2 post-3C cleavage; E 3C = elution post-3C cleavage; fractions on the gel are explained in more detail in Chapter 5.2.

After concentrating the combined fractions to below 2 ml, we also ran them on a size-exclusion chromatography column. Here, we collected a total of 95 fractions, also containing the cten-2 construct (Figure 5.5). The protein peak around fraction 55 corresponds 39.5 kDa, the calculated molecular weight of cten-2 (1262-1607), while the peak around fraction 41 contains an impurity. After running the samples on a polyacrylamide gel, we decided to pool fractions 55-64. The ratio of purified protein to impurities in the other fractions was too low (Figure 5.6).



**Figure 5.5 Size-exclusion chromatography**

UV1 spectrum (blue line) shows two distinct protein peaks. Peak 1 around fraction 41 corresponds to a larger molecular weight than cten2 (1262-1607), while peak 2 around fraction 55 is approximately the correct size. Fraction numbers are shown in red.

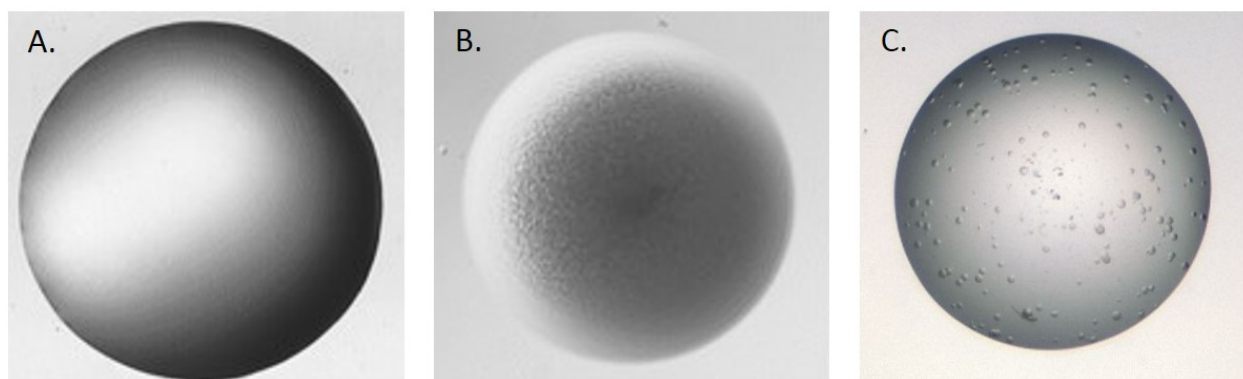


**Figure 5.6 Fractions including purified protein**

Fifteen fractions collected surrounding the protein peak at fraction 55, including the cten2 (1262-1607) construct expressed at 39.5 kDa. 'S200 Load' is the sample injected into the Äkta purifier.



After concentrating the pooled fractions down to 6.5 mg/ml, we set up drops for crystallization in chambers containing different crystallization solutions. The following conditions were tested thus far: different sets of crystallization solutions, at 4°C and at RT, and adding the proteases trypsin and chymotrypsin. Drops in the different conditions have shown no precipitation, complete precipitation, or microspherulite structures (incomplete crystallization), but no crystals (Figure 5.7).

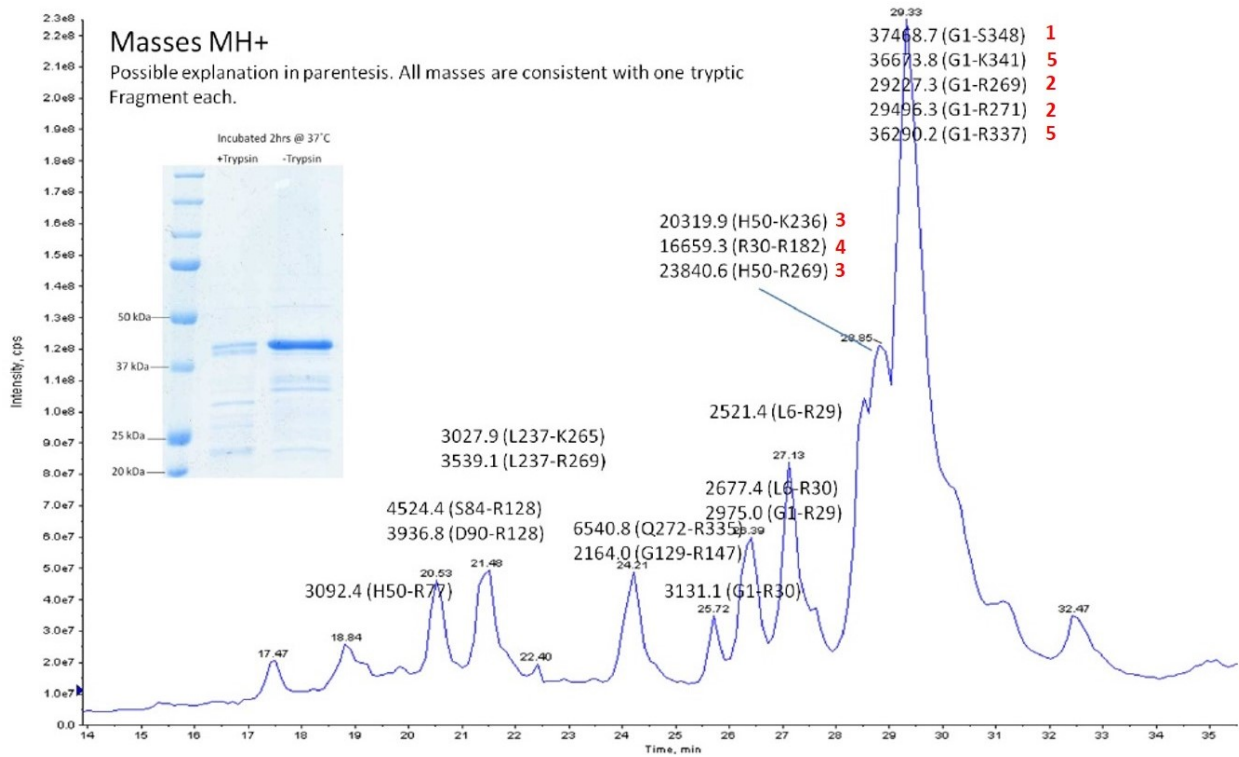


**Figure 5.7 Crystallization drops**

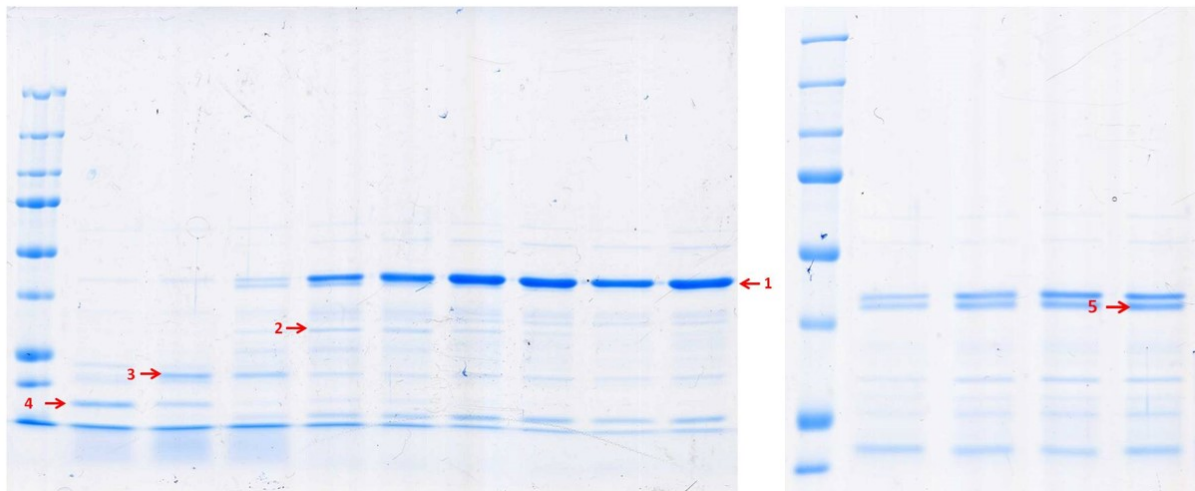
Examples of the different types of precipitates that appeared in the crystallization drops: A. no precipitate B. complete precipitate C. microspherulite structures.

Limited proteolysis can be utilized to determine the boundaries or stability of a certain purified domain. We used it to determine the boundaries of the cten2 (1262-1607) construct, or whether there are any unfolded regions on the N- or C-terminal end of the domain. Here, we able to show that the construct seems to be stable, with a short seven amino acid C-terminal region lysed by trypsin (Figure 5.8). The smaller fragments were all predicted to end within one of the beta-propellers, and were thus not potential constructs. Since according to the limited proteolysis experiments cten2 (1262-1607) seemed to be almost perfect with only a short C-terminal exposed tail, we concluded that our crystallization attempts failed for other reasons than the selection of the wrong domain boundaries.

A.



B.



**Figure 5.8 Limited Proteolysis**

A. Results from the liquid chromatography-mass spectrometer after limited proteolysis with trypsin. Peaks correspond to different size fragments, with molecular weights (Da) and possible explanations in parenthesis. Red numbers 1-5 denote the largest fragments, potentially containing the full NHL repeat domain.

B. Red numbers 1-5 show the different fragments from A. in protein gels that ran the limited proteolysis samples.

## 6. Conclusion

Results of my work are discussed in detail in Chapter 4 of the thesis. Here I will include a short conclusion for the unpublished results, and discuss how the different studies are related to each other.

Teneurins are an ancient family of proteins with several known functions and features, some of which are unique pertaining to eukaryotic proteins. Numerous studies have shown teneurins to be involved in the developing CNS, particularly in axon guidance and synapse formation. Several ligands for the teneurin ECDs have been described and functions for the ICDs include linking teneurins to the cytoskeleton, and influencing transcriptional regulation.<sup>84,96</sup> However, most molecular mechanisms of function have yet to be elucidated. The study presented in Chapter 4.1 investigates the evolution of teneurins. Sequence analyses suggest that teneurins evolved from *Monosiga brevicollis* via horizontal gene transfer from an ancient prokaryotic protein. YD repeats are one indicator, as they are generally found in bacterial cell wall proteins.<sup>82</sup> In-depth analysis of the structure identifies additional properties and features of teneurins, like additional furin cleavage sites where the ECD could be processed, and the probabilities of the ICDs to translocate to the nucleus. Due to its importance in teneurin function, there is a focus on the ICD. We described splice variants of human and chicken ICDs. Interestingly, no NLS or putative binding sites were missing in any of the variants, making a prediction of their functions purely speculative. While the full length ICDs have been implicated in influencing transcriptional regulation, the splice variants may be involved in other regulatory processes. Our study presented in Chapter 4.2 describes a novel mechanism of teneurin ICD function. Here, we showed that the TEN1-ICD can influence MITF-dependent transcription of target gene *GPNMB*, by competing for HINT1 binding. In addition to elucidating this molecular mechanism, we also

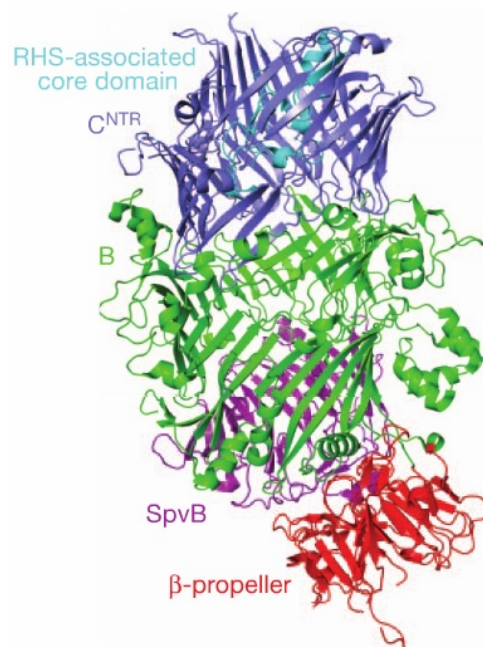
produced a lot of further interesting data when performing the unbiased screens: 1. Yeast-2 hybrid screen and 2. Whole-transcriptome analyses. The discussion in Chapter 4.2 also contains an outlook about determining new functions of the ICDs, especially in the context of our newly discovered ICD interacting proteins, as well as genes regulated by the expression of the teneurin-1 ICD. Examining the splice variants in more detail would be another interesting future direction.

The NHL repeat domain is a predicted beta-propeller and is part of all teneurin ECDs. While it is not directly involved in transcriptional regulation, it is no less important to the function of teneurins. As Beckmann et al. discovered recently, the NHL repeat domain of teneurins is responsible for homophilic rather than heterophilic interaction with other teneurins.<sup>88</sup> This interaction likely drives the cleavage of the ECD and subsequently the ICD, which is vital for the ICD to affect transcriptional regulation. But how does such a conserved beta-propeller distinguish between homophilic over heterophilic interactions? The purification and crystallization studies discussed in Chapter 5 were supposed to shed some light into this question. While we are able to purify constructs of the NHL repeat domain, they did not crystallize in the different conditions we had set up. This is not a trivial process, as several issues have to be overcome, such as determining the correct boundaries of the domain, purifying the protein with as little impurities as possible at a high concentration yet quick enough for the protein not to break down, the domain has to be stable and not flexible, and the purified protein has to reproducibly pack into an ordered structure that leads to crystals under certain conditions.<sup>139</sup> Future studies addressing these issues could finally lead to the crystallization of the predicted beta propellers in the teneurin ECDs. We tried to resolve the crystal structure of one domain of a larger multi-domain protein. While prediction software was able to estimate boundaries for the beta-propeller domain, it is vital for the boundaries to be as close to the

domain as possible, while keeping the whole domain intact. Flexible linkers that are not part of the folded domain can keep it from tightly packing and thus crystallizing. If the domain is not fully intact, it cannot take a stable conformation, and thus will not crystallize either. Since the software cannot precisely predict the exact boundaries, it gave us different size constructs that are most likely to include the full beta-propeller. Since different protein drop conditions did not lead to crystallization of the protein, we tested the boundaries of the protein by limited proteolysis, in case there would be unfolded linker regions on either side of the domain. The C-terminal side did have a short peptide that could be cleaved during limited proteolysis, nevertheless proteases did not improve the crystallization process. There are several other options that can be tested in future studies. We purified the NHL repeat domain from bacteria. While the purification process is quicker and simpler, it may not be optimal for proper folding of the domain. Purifying the domain from mammalian cells may be a good option to optimize this process. Since the NHL repeat domain is well-conserved between different species and the other teneurins, it would be also be a viable option to purify and set up crystallization conditions with another teneurin paralog or ortholog.

Recently, the crystal structure of the BC component of ABC toxins was published.<sup>132</sup> Interestingly, the domain structure is similar to that of the teneurin ECD, particularly the NHL repeat domain, YD repeats/RHS protein and the TCAP. Further, the peptide at the very C-terminal end of the teneurin ECD, TCAP, has a striking sequence similarity to the Tox-GHH domain, the toxic peptide of ABC toxins. While the structure of the teneurin ECDs has not been solved to date, if modeling the ECD structure after the BC component of ABC toxins, the YD repeats would form a barrel, protecting the RHS-protein and Tox-GHH domain (Figure 6.1). There are several potential mechanisms of how teneurin ECDs could work in an ABC toxin-like

manner. ABC toxins work in two mechanisms of how the toxic peptide gets internalized by the target cell, either by forming a pore to inject the peptide or by endocytosis of the entire complex and subsequent release inside the cell. Since the ECDs only contain a structural similarity to the BC component of ABC toxins, the  $\beta$ -propeller may be responsible for heterophilic binding to a protein on another cell, acting as the A protein. Further, homo- or heterophilic interactions could lead to the release of the toxic peptide into the extracellular milieu, though this would differ from the typical ABC toxin mechanisms. It would be very interesting to study this new aspect of teneurin ECDs. A crystal structure of the ECD, at least ranging from the NHL domain to the TCAP could prove the structural similarity to ABC toxins. Further, the type of release of the teneurin Tox-GHH domain and its function inside the cell or the extracellular milieu could be studied. This could be an entirely novel mechanism of action for a eukaryotic protein, and open up a new way of thinking about how teneurins function in cell-cell communication.



**Figure 6.1 BC component structure**<sup>132</sup>

Structure of the BC component of the *Yersinia entomophaga* ABC toxin, with the  $\beta$ -sheets of the SpvB, the rest of the B-protein and the C<sup>NTR</sup> domain forming a central cavity.

## 7. References

- 1 Alberts, B. *Molecular biology of the cell*. 4th edn, (Garland Science, 2002).
- 2 Leamey, C. A. & Sawatari, A. The teneurins: New players in the generation of visual topography. *Semin. Cell Dev. Biol.*, doi:10.1016/j.semcdb.2014.08.007 (2014).
- 3 Viguet-Carrin, S., Garnero, P. & Delmas, P. D. The role of collagen in bone strength. *Osteoporos. Int.* **17**, 319-336, doi:10.1007/s00198-005-2035-9 (2006).
- 4 Kalluri, R. Basement membranes: structure, assembly and role in tumour angiogenesis. *Nat. Rev. Cancer* **3**, 422-433, doi:10.1038/nrc1094 (2003).
- 5 Ricard-Blum, S. The collagen family. *Cold Spring Harb. Perspect. Biol.* **3**, a004978, doi:10.1101/cshperspect.a004978 (2011).
- 6 Shoulders, M. D. & Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929-958, doi:10.1146/annurev.biochem.77.032207.120833 (2009).
- 7 Gelse, K., Poschl, E. & Aigner, T. Collagens--structure, function, and biosynthesis. *Advanced drug delivery reviews* **55**, 1531-1546 (2003).
- 8 Fan, D., Takawale, A., Lee, J. & Kassiri, Z. Cardiac fibroblasts, fibrosis and extracellular matrix remodeling in heart disease. *Fibrogenesis & tissue repair* **5**, 15, doi:10.1186/1755-1536-5-15 (2012).
- 9 Schaefer, L. & Schaefer, R. M. Proteoglycans: from structural compounds to signaling molecules. *Cell Tissue Res.* **339**, 237-246, doi:10.1007/s00441-009-0821-y (2010).
- 10 Schmidt, T. A., Gastelum, N. S., Nguyen, Q. T., Schumacher, B. L. & Sah, R. L. Boundary lubrication of articular cartilage: role of synovial fluid constituents. *Arthritis Rheum.* **56**, 882-891, doi:10.1002/art.22446 (2007).
- 11 Domogatskaya, A., Rodin, S. & Tryggvason, K. Functional diversity of laminins. *Annu. Rev. Cell Dev. Biol.* **28**, 523-553, doi:10.1146/annurev-cellbio-101011-155750 (2012).
- 12 Yurchenco, P. D. Basement membranes: cell scaffoldings and signaling platforms. *Cold Spring Harb. Perspect. Biol.* **3**, doi:10.1101/cshperspect.a004911 (2011).
- 13 Cardoso, F. L., Brites, D. & Brito, M. A. Looking at the blood-brain barrier: molecular anatomy and possible investigation approaches. *Brain research reviews* **64**, 328-363, doi:10.1016/j.brainresrev.2010.05.003 (2010).
- 14 Rodin, S. *et al.* Clonal culturing of human embryonic stem cells on laminin-521/E-cadherin matrix in defined and xeno-free environment. *Nature communications* **5**, 3195, doi:10.1038/ncomms4195 (2014).
- 15 Matejas, V. *et al.* Mutations in the human laminin beta2 (LAMB2) gene and the associated phenotypic spectrum. *Hum. Mutat.* **31**, 992-1002, doi:10.1002/humu.21304 (2010).
- 16 Gupta, V. A. *et al.* A splice site mutation in laminin-alpha2 results in a severe muscular dystrophy and growth abnormalities in zebrafish. *PLoS One* **7**, e43794, doi:10.1371/journal.pone.0043794 (2012).
- 17 Patton, B. L., Wang, B., Tarumi, Y. S., Seburn, K. L. & Burgess, R. W. A single point mutation in the LN domain of LAMA2 causes muscular dystrophy and peripheral amyelination. *J. Cell Sci.* **121**, 1593-1604, doi:10.1242/jcs.015354 (2008).
- 18 Van Obberghen-Schilling, E. *et al.* Fibronectin and tenascin-C: accomplices in vascular morphogenesis during development and tumor growth. *Int. J. Dev. Biol.* **55**, 511-525, doi:10.1387/ijdb.103243eo (2011).
- 19 Schwarzbauer, J. E. & DeSimone, D. W. Fibronectins, their fibrillogenesis, and in vivo functions. *Cold Spring Harb. Perspect. Biol.* **3**, doi:10.1101/cshperspect.a005041 (2011).

- 20 George, E. L., Georges-Labouesse, E. N., Patel-King, R. S., Rayburn, H. & Hynes, R. O. Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin. *Development* **119**, 1079-1091 (1993).
- 21 To, W. S. & Midwood, K. S. Plasma and cellular fibronectin: distinct and independent functions during tissue repair. *Fibrogenesis & tissue repair* **4**, 21, doi:10.1186/1755-1536-4-21 (2011).
- 22 Castelletti, F. *et al.* Mutations in FN1 cause glomerulopathy with fibronectin deposits. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2538-2543, doi:10.1073/pnas.0707730105 (2008).
- 23 Jones, F. S. & Jones, P. L. The tenascin family of ECM glycoproteins: structure, function, and regulation during embryonic development and tissue remodeling. *Dev. Dyn.* **218**, 235-259, doi:10.1002/(SICI)1097-0177(200006)218:2<235::AID-DVDY2>3.0.CO;2-G (2000).
- 24 Chiquet-Ehrismann, R. & Tucker, R. P. Tenascins and the importance of adhesion modulation. *Cold Spring Harb. Perspect. Biol.* **3**, doi:10.1101/cshperspect.a004960 (2011).
- 25 Chiquet-Ehrismann, R. Anti-adhesive molecules of the extracellular matrix. *Curr. Opin. Cell Biol.* **3**, 800-804 (1991).
- 26 Murphy-Ullrich, J. E. The de-adhesive activity of matricellular proteins: is intermediate cell adhesion an adaptive state? *J. Clin. Invest.* **107**, 785-790, doi:10.1172/JCI12609 (2001).
- 27 Hsia, H. C. & Schwarzbauer, J. E. Meet the tenascins: multifunctional and mysterious. *J. Biol. Chem.* **280**, 26641-26644, doi:10.1074/jbc.R500005200 (2005).
- 28 Hynes, R. O. & Naba, A. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb. Perspect. Biol.* **4**, a004903, doi:10.1101/cshperspect.a004903 (2012).
- 29 Campbell, I. D. & Humphries, M. J. Integrin structure, activation, and interactions. *Cold Spring Harb. Perspect. Biol.* **3**, doi:10.1101/cshperspect.a004994 (2011).
- 30 Hohenester, E. Signalling complexes at the cell-matrix interface. *Curr. Opin. Struct. Biol.* **29C**, 10-16, doi:10.1016/j.sbi.2014.08.009 (2014).
- 31 Fluck, M., Ziemiecki, A., Billeter, R. & Muntener, M. Fibre-type specific concentration of focal adhesion kinase at the sarcolemma: influence of fibre innervation and regeneration. *J. Exp. Biol.* **205**, 2337-2348 (2002).
- 32 Geiger, B., Spatz, J. P. & Bershadsky, A. D. Environmental sensing through focal adhesions. *Nat. Rev. Mol. Cell Biol.* **10**, 21-33, doi:10.1038/nrm2593 (2009).
- 33 Wu, C. Focal adhesion: a focal point in current cell biology and molecular medicine. *Cell adhesion & migration* **1**, 13-18 (2007).
- 34 Chiquet, M., Gelman, L., Lutz, R. & Maier, S. From mechanotransduction to extracellular matrix gene expression in fibroblasts. *Biochim. Biophys. Acta* **1793**, 911-920, doi:10.1016/j.bbamcr.2009.01.012 (2009).
- 35 Moore, C. J. & Winder, S. J. Dystroglycan versatility in cell adhesion: a tale of multiple motifs. *Cell communication and signaling : CCS* **8**, 3, doi:10.1186/1478-811X-8-3 (2010).
- 36 Middleton, J., Patterson, A. M., Gardner, L., Schmutz, C. & Ashton, B. A. Leukocyte extravasation: chemokine transport and presentation by the endothelium. *Blood* **100**, 3853-3860, doi:10.1182/blood.V100.12.3853 (2002).
- 37 Peterson, L. W. & Artis, D. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat. Rev. Immunol.* **14**, 141-153, doi:10.1038/nri3608 (2014).
- 38 Bianchi, S. *et al.* Synaptogenesis and development of pyramidal neuron dendritic morphology in the chimpanzee neocortex resembles humans. *Proc. Natl. Acad. Sci. U. S. A.* **110 Suppl 2**, 10395-10401, doi:10.1073/pnas.1301224110 (2013).
- 39 Juliano, R. L. Signal transduction by cell adhesion receptors and the cytoskeleton: functions of integrins, cadherins, selectins, and immunoglobulin-superfamily members. *Annu. Rev. Pharmacol. Toxicol.* **42**, 283-323, doi:10.1146/annurev.pharmtox.42.090401.151133 (2002).



- 40 Lodish, H. F. *Molecular cell biology*. 4th edn, (W.H. Freeman, 2000).
- 41 Schmidt, S., Moser, M. & Sperandio, M. The molecular basis of leukocyte recruitment and its deficiencies. *Mol. Immunol.* **55**, 49-58, doi:10.1016/j.molimm.2012.11.006 (2013).
- 42 Shapiro, L. & Weis, W. I. Structure and biochemistry of cadherins and catenins. *Cold Spring Harb. Perspect. Biol.* **1**, a003053, doi:10.1101/cshperspect.a003053 (2009).
- 43 Halbleib, J. M. & Nelson, W. J. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes Dev.* **20**, 3199-3214, doi:10.1101/gad.1486806 (2006).
- 44 Stepniak, E., Radice, G. L. & Vasioukhin, V. Adhesive and signaling functions of cadherins and catenins in vertebrate development. *Cold Spring Harb. Perspect. Biol.* **1**, a002949, doi:10.1101/cshperspect.a002949 (2009).
- 45 Derycke, L. D. & Bracke, M. E. N-cadherin in the spotlight of cell-cell adhesion, differentiation, embryogenesis, invasion and signalling. *Int. J. Dev. Biol.* **48**, 463-476, doi:10.1387/ijdb.041793ld (2004).
- 46 Morishita, H. & Yagi, T. Protocadherin family: diversity, structure, and function. *Curr. Opin. Cell Biol.* **19**, 584-592, doi:10.1016/j.ceb.2007.09.006 (2007).
- 47 Ley, K. The role of selectins in inflammation and disease. *Trends Mol. Med.* **9**, 263-268 (2003).
- 48 Vestweber, D. & Blanks, J. E. Mechanisms that regulate the function of the selectins and their ligands. *Physiol. Rev.* **79**, 181-213 (1999).
- 49 Carlow, D. A. *et al.* PSGL-1 function in immunity and steady state homeostasis. *Immunol. Rev.* **230**, 75-96, doi:10.1111/j.1600-065X.2009.00797.x (2009).
- 50 Wai Wong, C., Dye, D. E. & Coombe, D. R. The role of immunoglobulin superfamily cell adhesion molecules in cancer metastasis. *Int. J. Cell Biol.* **2012**, 340296, doi:10.1155/2012/340296 (2012).
- 51 Hartsock, A. & Nelson, W. J. Adherens and tight junctions: structure, function and connections to the actin cytoskeleton. *Biochim. Biophys. Acta* **1778**, 660-669, doi:10.1016/j.bbamem.2007.07.012 (2008).
- 52 Meng, W. & Takeichi, M. Adherens junction: molecular architecture and regulation. *Cold Spring Harb. Perspect. Biol.* **1**, a002899, doi:10.1101/cshperspect.a002899 (2009).
- 53 Perez-Moreno, M. & Fuchs, E. Catenins: keeping cells from getting their signals crossed. *Dev. Cell* **11**, 601-612, doi:10.1016/j.devcel.2006.10.010 (2006).
- 54 Brembeck, F. H., Rosario, M. & Birchmeier, W. Balancing cell adhesion and Wnt signaling, the key role of beta-catenin. *Curr. Opin. Genet. Dev.* **16**, 51-59, doi:10.1016/j.gde.2005.12.007 (2006).
- 55 Hu, X. & Dahl, G. Exchange of conductance and gating properties between gap junction hemichannels. *FEBS Lett.* **451**, 113-117 (1999).
- 56 Goodenough, D. A. & Paul, D. L. Gap junctions. *Cold Spring Harb. Perspect. Biol.* **1**, a002576, doi:10.1101/cshperspect.a002576 (2009).
- 57 Pereda, A. E. Electrical synapses and their functional interactions with chemical synapses. *Nat. Rev. Neurosci.* **15**, 250-263, doi:10.1038/nrn3708 (2014).
- 58 Gilbert, S. F. *Developmental biology*. 6th edn, (Sinauer Associates, 2000).
- 59 Stiles, J. & Jernigan, T. L. The basics of brain development. *Neuropsychol. Rev.* **20**, 327-348, doi:10.1007/s11065-010-9148-4 (2010).
- 60 U-Tokyo. *Introduction to Life Science*, <<http://csls-text2.c.u-tokyo.ac.jp/>> (2010).
- 61 Levene, M. I. & Chervenak, F. A. *Fetal and neonatal neurology and neurosurgery*. 4th edn, (Elsevier Churchill Livingstone, 2009).
- 62 Reichert, H. Conserved genetic mechanisms for embryonic brain patterning. *Int. J. Dev. Biol.* **46**, 81-87 (2002).
- 63 O'Leary, D. D., Chou, S. J. & Sahara, S. Area patterning of the mammalian cortex. *Neuron* **56**, 252-269, doi:10.1016/j.neuron.2007.10.010 (2007).

- 64 Schoenwolf, G. C. & Larsen, W. J. *Larsen's human embryology*. 4th edn, (Churchill Livingstone/Elsevier, 2009).
- 65 Giacomantonio, C. E. & Goodhill, G. J. A Boolean model of the gene regulatory network underlying Mammalian cortical area development. *PLoS Comput. Biol.* **6**, doi:10.1371/journal.pcbi.1000936 (2010).
- 66 O'Leary, D. D. & Sahara, S. Genetic regulation of arealization of the neocortex. *Curr. Opin. Neurobiol.* **18**, 90-100, doi:10.1016/j.conb.2008.05.011 (2008).
- 67 Chilton, J. K. Molecular mechanisms of axon guidance. *Dev. Biol.* **292**, 13-24, doi:10.1016/j.ydbio.2005.12.048 (2006).
- 68 Purves, D. & Williams, S. M. *Neuroscience*. 2nd edn, (Sinauer Associates, 2001).
- 69 Dickson, B. J. Molecular mechanisms of axon guidance. *Science* **298**, 1959-1964, doi:10.1126/science.1072165 (2002).
- 70 Tessier-Lavigne, M. & Goodman, C. S. The molecular biology of axon guidance. *Science* **274**, 1123-1133 (1996).
- 71 Dityatev, A. & El-Husseini, A. *Molecular mechanisms of synaptogenesis*. (Springer, 2006).
- 72 Chia, P. H., Li, P. & Shen, K. Cell biology in neuroscience: cellular and molecular mechanisms underlying presynapse formation. *J. Cell Biol.* **203**, 11-22, doi:10.1083/jcb.201307020 (2013).
- 73 Baumgartner, S. & Chiquet-Ehrismann, R. Tena, a Drosophila gene related to tenascin, shows selective transcript localization. *Mech. Dev.* **40**, 165-176 (1993).
- 74 Baumgartner, S., Martin, D., Hagios, C. & Chiquet-Ehrismann, R. Tenm, a Drosophila gene related to tenascin, is a new pair-rule gene. *EMBO J.* **13**, 3728-3740 (1994).
- 75 Levine, A. *et al.* Odd Oz: a novel Drosophila pair rule gene. *Cell* **77**, 587-598 (1994).
- 76 Zheng, L. *et al.* Drosophila Ten-m and filamin affect motor neuron growth cone guidance. *PLoS One* **6**, e22956, doi:10.1371/journal.pone.0022956 (2011).
- 77 Rubin, B. P., Tucker, R. P., Martin, D. & Chiquet-Ehrismann, R. Teneurins: a novel family of neuronal cell surface proteins in vertebrates, homologous to the Drosophila pair-rule gene product Ten-m. *Dev. Biol.* **216**, 195-209, doi:10.1006/dbio.1999.9503 (1999).
- 78 Tucker, R. P. Horizontal gene transfer in choanoflagellates. *Journal of experimental zoology. Part B, Molecular and developmental evolution* **320**, 1-9, doi:10.1002/jez.b.22480 (2013).
- 79 Drabikowski, K., Trzebiatowska, A. & Chiquet-Ehrismann, R. ten-1, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in *Caenorhabditis elegans*. *Dev. Biol.* **282**, 27-38, doi:10.1016/j.ydbio.2005.02.017 (2005).
- 80 Wang, X. Z. *et al.* Identification of novel stress-induced genes downstream of chop. *EMBO J.* **17**, 3619-3630, doi:10.1093/emboj/17.13.3619 (1998).
- 81 Oohashi, T. *et al.* Mouse ten-m/Odz is a new family of dimeric type II transmembrane proteins expressed in many tissues. *J. Cell Biol.* **145**, 563-577 (1999).
- 82 Minet, A. D. & Chiquet-Ehrismann, R. Phylogenetic analysis of teneurin genes and comparison to the rearrangement hot spot elements of *E. coli*. *Gene* **257**, 87-97 (2000).
- 83 Minet, A. D., Rubin, B. P., Tucker, R. P., Baumgartner, S. & Chiquet-Ehrismann, R. Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene ten-m, is a neuronal protein with a novel type of heparin-binding domain. *J. Cell Sci.* **112 ( Pt 12)**, 2019-2032 (1999).
- 84 Nunes, S. M. *et al.* The intracellular domain of teneurin-1 interacts with MBD1 and CAP/ponsin resulting in subcellular codistribution and translocation to the nuclear matrix. *Exp. Cell Res.* **305**, 122-132, doi:10.1016/j.yexcr.2004.12.020 (2005).
- 85 Tucker, R. P. & Chiquet-Ehrismann, R. Teneurins: a conserved family of transmembrane proteins involved in intercellular signaling during development. *Dev. Biol.* **290**, 237-245, doi:10.1016/j.ydbio.2005.11.038 (2006).

- 86 Kenzelmann, D., Chiquet-Ehrismann, R., Leachman, N. T. & Tucker, R. P. Teneurin-1 is expressed in interconnected regions of the developing brain and is processed in vivo. *BMC Dev. Biol.* **8**, 30, doi:10.1186/1471-213X-8-30 (2008).
- 87 Feng, K. *et al.* All four members of the Ten-m/Odz family of transmembrane proteins form dimers. *J. Biol. Chem.* **277**, 26128-26135, doi:10.1074/jbc.M203722200 (2002).
- 88 Beckmann, J., Schubert, R., Chiquet-Ehrismann, R. & Muller, D. J. Deciphering teneurin domains that facilitate cellular recognition, cell-cell adhesion, and neurite outgrowth using atomic force microscopy-based single-cell force spectroscopy. *Nano Lett* **13**, 2937-2946, doi:10.1021/nl4013248 (2013).
- 89 Boucard, A. A., Maxeiner, S. & Sudhof, T. C. Latrophilins function as heterophilic cell-adhesion molecules by binding to teneurins: regulation by alternative splicing. *J. Biol. Chem.* **289**, 387-402, doi:10.1074/jbc.M113.504779 (2014).
- 90 Rubin, B. P., Tucker, R. P., Brown-Luedi, M., Martin, D. & Chiquet-Ehrismann, R. Teneurin 2 is expressed by the neurons of the thalamofugal visual system in situ and promotes homophilic cell-cell adhesion in vitro. *Development* **129**, 4697-4705 (2002).
- 91 Silva, J. P. *et al.* Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12113-12118, doi:10.1073/pnas.1019434108 (2011).
- 92 Wang, L. *et al.* Teneurin proteins possess a carboxy terminal sequence with neuromodulatory activity. *Brain Res. Mol. Brain Res.* **133**, 253-265, doi:10.1016/j.molbrainres.2004.10.019 (2005).
- 93 Lovejoy, D. A., Al Chawaf, A. & Cadinouche, M. Z. Teneurin C-terminal associated peptides: an enigmatic family of neuropeptides with structural similarity to the corticotropin-releasing factor and calcitonin families of peptides. *Gen. Comp. Endocrinol.* **148**, 299-305, doi:10.1016/j.ygcen.2006.01.012 (2006).
- 94 Kenzelmann, D., Chiquet-Ehrismann, R. & Tucker, R. P. Teneurins, a transmembrane protein family involved in cell communication during neuronal development. *Cell. Mol. Life Sci.* **64**, 1452-1456, doi:10.1007/s00018-007-7108-9 (2007).
- 95 Ebinu, J. O. & Yankner, B. A. A RIP tide in neuronal signal transduction. *Neuron* **34**, 499-502 (2002).
- 96 Bagutti, C., Forro, G., Ferralli, J., Rubin, B. & Chiquet-Ehrismann, R. The intracellular domain of teneurin-2 has a nuclear function and represses zic-1-mediated transcription. *J. Cell Sci.* **116**, 2957-2966, doi:10.1242/jcs.00603 (2003).
- 97 Tucker, R. P., Kenzelmann, D., Trzebiatowska, A. & Chiquet-Ehrismann, R. Teneurins: transmembrane proteins with fundamental roles in development. *Int. J. Biochem. Cell Biol.* **39**, 292-297, doi:10.1016/j.biocel.2006.09.012 (2007).
- 98 Trzebiatowska, A., Topf, U., Sauder, U., Drabikowski, K. & Chiquet-Ehrismann, R. *Caenorhabditis elegans* teneurin, ten-1, is required for gonadal and pharyngeal basement membrane integrity and acts redundantly with integrin ina-1 and dystroglycan dgn-1. *Mol. Biol. Cell* **19**, 3898-3908, doi:10.1091/mbc.E08-01-0028 (2008).
- 99 Morck, C., Vivekanand, V., Jafari, G. & Pilon, M. C. *elegans* ten-1 is synthetic lethal with mutations in cytoskeleton regulators, and enhances many axon guidance defective mutants. *BMC Dev. Biol.* **10**, 55, doi:10.1186/1471-213X-10-55 (2010).
- 100 Levine, A., Weiss, C. & Wides, R. Expression of the pair-rule gene odd Oz (odz) in imaginal tissues. *Dev. Dyn.* **209**, 1-14, doi:10.1002/(SICI)1097-0177(199705)209:1<1::AID-AJA1>3.0.CO;2-M (1997).
- 101 Hong, W., Mosca, T. J. & Luo, L. Teneurins instruct synaptic partner matching in an olfactory map. *Nature* **484**, 201-207, doi:10.1038/nature10926 (2012).

- 102 Mosca, T. J., Hong, W., Dani, V. S., Favaloro, V. & Luo, L. Trans-synaptic Teneurin signalling in neuromuscular synapse organization and target choice. *Nature* **484**, 237-241, doi:10.1038/nature10923 (2012).
- 103 Fascetti, N. & Baumgartner, S. Expression of Drosophila Ten-a, a dimeric receptor during embryonic development. *Mech. Dev.* **114**, 197-200 (2002).
- 104 Cheng, X. *et al.* Ten-a affects the fusion of central complex primordia in Drosophila. *PLoS One* **8**, e57129, doi:10.1371/journal.pone.0057129 (2013).
- 105 Mieda, M., Kikuchi, Y., Hirate, Y., Aoki, M. & Okamoto, H. Compartmentalized expression of zebrafish ten-m3 and ten-m4, homologues of the Drosophila ten(m)/odd Oz gene, in the central nervous system. *Mech. Dev.* **87**, 223-227 (1999).
- 106 Antinucci, P., Nikolaou, N., Meyer, M. P. & Hindges, R. Teneurin-3 specifies morphological and functional connectivity of retinal ganglion cells in the vertebrate visual system. *Cell reports* **5**, 582-592, doi:10.1016/j.celrep.2013.09.045 (2013).
- 107 Kenzelmann-Broz, D., Tucker, R. P., Leachman, N. T. & Chiquet-Ehrismann, R. The expression of teneurin-4 in the avian embryo: potential roles in patterning of the limb and nervous system. *Int. J. Dev. Biol.* **54**, 1509-1516, doi:10.1387/ijdb.103139dk (2010).
- 108 Tucker, R. P. *et al.* Teneurin-2 is expressed in tissues that regulate limb and somite pattern formation and is induced in vitro and in situ by FGF8. *Dev. Dyn.* **220**, 27-39, doi:10.1002/1097-0177(2000)9999:9999<::AID-DVDY1084>3.0.CO;2-B (2001).
- 109 Tucker, R. P., Martin, D., Kos, R. & Chiquet-Ehrismann, R. The expression of teneurin-4 in the avian embryo. *Mech. Dev.* **98**, 187-191 (2000).
- 110 Otaki, J. M. & Firestein, S. Segregated expression of neurestin in the developing olfactory bulb. *Neuroreport* **10**, 2677-2680 (1999).
- 111 Otaki, J. M. & Firestein, S. Neurestin: putative transmembrane molecule implicated in neuronal development. *Dev. Biol.* **212**, 165-181, doi:10.1006/dbio.1999.9310 (1999).
- 112 Ben-Zur, T., Feige, E., Motro, B. & Wides, R. The mammalian Odz gene family: homologs of a Drosophila pair-rule gene with expression implying distinct yet overlapping developmental roles. *Dev. Biol.* **217**, 107-120, doi:10.1006/dbio.1999.9532 (2000).
- 113 Li, H., Bishop, K. M. & O'Leary, D. D. Potential target genes of EMX2 include Odz/Ten-M and other gene families with implications for cortical patterning. *Mol. Cell. Neurosci.* **33**, 136-149, doi:10.1016/j.mcn.2006.06.012 (2006).
- 114 Zhou, X. H. *et al.* The murine Ten-m/Odz genes show distinct but overlapping expression patterns during development and in adult brain. *Gene Expr Patterns* **3**, 397-405 (2003).
- 115 Ben-Zur, T. & Wides, R. Mapping homologs of Drosophila odd Oz (odz): Doc4/Odz4 to mouse chromosome 7, Odz1 to mouse chromosome 11; and ODZ3 to human chromosome Xq25. *Genomics* **58**, 102-103, doi:10.1006/geno.1999.5798 (1999).
- 116 Young, T. R. *et al.* Ten-m2 is required for the generation of binocular visual circuits. *J. Neurosci.* **33**, 12490-12509, doi:10.1523/JNEUROSCI.4708-12.2013 (2013).
- 117 Leamey, C. A. *et al.* Ten\_m3 regulates eye-specific patterning in the mammalian visual pathway and is required for binocular vision. *PLoS Biol.* **5**, e241, doi:10.1371/journal.pbio.0050241 (2007).
- 118 Merlin, S. *et al.* Deletion of Ten-m3 induces the formation of eye dominance domains in mouse visual cortex. *Cereb. Cortex* **23**, 763-774, doi:10.1093/cercor/bhs030 (2013).
- 119 Dharmaratne, N. *et al.* Ten-m3 is required for the development of topography in the ipsilateral retinocollicular pathway. *PLoS One* **7**, e43083, doi:10.1371/journal.pone.0043083 (2012).
- 120 Murakami, T. *et al.* Expression of Ten-m/Odz3 in the fibrous layer of mandibular condylar cartilage during postnatal growth in mice. *J. Anat.* **217**, 236-244, doi:10.1111/j.1469-7580.2010.01267.x (2010).

- 121 Lossie, A. C., Nakamura, H., Thomas, S. E. & Justice, M. J. Mutation of I7Rn3 shows that Odz4 is required for mouse gastrulation. *Genetics* **169**, 285-299, doi:10.1534/genetics.104.034967 (2005).
- 122 Suzuki, N. *et al.* Teneurin-4 promotes cellular protrusion formation and neurite outgrowth through focal adhesion kinase signaling. *FASEB J.* **28**, 1386-1397, doi:10.1096/fj.13-241034 (2014).
- 123 Suzuki, N. *et al.* Teneurin-4, a transmembrane protein, is a novel regulator that suppresses chondrogenic differentiation. *J. Orthop. Res.* **32**, 915-922, doi:10.1002/jor.22616 (2014).
- 124 Nakamura, H., Cook, R. N. & Justice, M. J. Mouse Tenm4 is required for mesoderm induction. *BMC Dev. Biol.* **13**, 9, doi:10.1186/1471-213X-13-9 (2013).
- 125 Carr, O. P., Glendining, K. A., Leamey, C. A. & Marotte, L. R. Retinal overexpression of Ten-m3 alters ipsilateral retinogeniculate projections in the wallaby (*Macropus eugenii*). *Neurosci. Lett.* **566**, 167-171, doi:10.1016/j.neulet.2014.02.048 (2014).
- 126 Carr, O. P., Glendining, K. A., Leamey, C. A. & Marotte, L. R. Overexpression of Ten-m3 in the retina alters ipsilateral retinocollicular projections in the wallaby (*Macropus eugenii*). *Int. J. Dev. Neurosci.* **31**, 496-504, doi:10.1016/j.ijdevneu.2013.05.011 (2013).
- 127 Kinel-Tahan, Y., Weiss, H., Dgany, O., Levine, A. & Wides, R. Drosophila odz gene is required for multiple cell types in the compound retina. *Dev. Dyn.* **236**, 2541-2554, doi:10.1002/dvdy.21284 (2007).
- 128 Topf, U. & Chiquet-Ehrismann, R. Genetic interaction between *Caenorhabditis elegans* teneurin ten-1 and prolyl 4-hydroxylase phy-1 and their function in collagen IV-mediated basement membrane integrity during late elongation of the embryo. *Mol. Biol. Cell* **22**, 3331-3343, doi:10.1091/mbc.E10-10-0853 (2011).
- 129 Behrman, S., Acosta-Alvear, D. & Walter, P. A CHOP-regulated microRNA controls rhodopsin expression. *J. Cell Biol.* **192**, 919-927, doi:10.1083/jcb.201010055 (2011).
- 130 Suzuki, N. *et al.* Teneurin-4 is a novel regulator of oligodendrocyte differentiation and myelination of small-diameter axons in the CNS. *J. Neurosci.* **32**, 11586-11599, doi:10.1523/JNEUROSCI.2045-11.2012 (2012).
- 131 Leamey, C. A. *et al.* Differential gene expression between sensory neocortical areas: potential roles for Ten\_m3 and Bcl6 in patterning visual and somatosensory pathways. *Cereb. Cortex* **18**, 53-66, doi:10.1093/cercor/bhm031 (2008).
- 132 Busby, J. N., Panjikar, S., Landsberg, M. J., Hurst, M. R. & Lott, J. S. The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. *Nature* **501**, 547-550, doi:10.1038/nature12465 (2013).
- 133 Zhang, D., de Souza, R. F., Anantharaman, V., Iyer, L. M. & Aravind, L. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* **7**, 18, doi:10.1186/1745-6150-7-18 (2012).
- 134 Beckmann, J. *et al.* Human teneurin-1 is a direct target of the homeobox transcription factor EMX2 at a novel alternate promoter. *BMC Dev. Biol.* **11**, 35, doi:10.1186/1471-213X-11-35 (2011).
- 135 Hattori, Y. *et al.* Sensory-neuron subtype-specific transcriptional programs controlling dendrite morphogenesis: genome-wide analysis of Abrupt and Knot/Collier. *Dev. Cell* **27**, 530-544, doi:10.1016/j.devcel.2013.10.024 (2013).
- 136 Aldahmesh, M. A., Mohammed, J. Y., Al-Hazaa, S. & Alkuraya, F. S. Homozygous null mutation in ODZ3 causes microphthalmia in humans. *Genet. Med.* **14**, 900-904, doi:10.1038/gim.2012.71 (2012).

- 137 Psychiatric, G. C. B. D. W. G. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43**, 977-983, doi:10.1038/ng.943 (2011).
- 138 Ziegler, A., Corvalan, A., Roa, I., Branes, J. A. & Wollscheid, B. Teneurin protein family: an emerging role in human tumorigenesis and drug resistance. *Cancer Lett.* **326**, 1-7, doi:10.1016/j.canlet.2012.07.021 (2012).
- 139 Smyth, M. S. & Martin, J. H. x ray crystallography. *Mol. Pathol.* **53**, 8-14 (2000).

## 8. Abbreviations

- ECM – extracellular matrix
- CAM – cell adhesion molecule
- PG – proteoglycan
- GAG – glycosaminoglycan
- HA – hyaluronan
- SLRP – small leucine-rich PGs
- CNS – central nervous system
- LM – laminin
- kDa – kilo-dalton
- FN – fibronectin
- p-FN – plasma fibronectin
- c-FN – cellular fibronectin
- EGF – epidermal growth factor
- FA – focal adhesion
- DDR – discoidin domain receptors
- Ig – immunoglobulin
- CDH - cadherin
- EC – extracellular cadherin
- ICD – intracellular domain
- ECD – extracellular domain
- CR – complement regulatory
- IgSF – immunoglobulin superfamily
- N-CAM – neural cell adhesion molecules
- Nm – nanometer
- ICM – inner cell mass
- E13 – embryonic day 13
- HOX protein – homeobox protein
- FGF – fibroblast growth factor
- SHH – sonic hedgehog

- WNT – *Drosophila* wingless
- BMP – bone morphogenic protein
- NRG – neuregulin
- COUP-TFI – chicken ovalbumin upstream promoter transcription factor 1
- EMX2 – empty spiracles homeobox 2
- PAX6 – paired box 6
- SP8 – specificity protein 8
- F-actin – filamentous actin
- NTN - netrin
- DCC – deleted in colorectal carcinoma
- ROBO - roundabout
- SEMA - semaphorin
- PLXN - plexin
- EFN - ephrin
- EPH – ephrin receptor
- NRXN – neurexin
- NLGN – neuroligin
- Ten-a – tenascin-like accessory
- Ten-m – tenascin-like major
- Odz – odd Oz
- NLS – nuclear localization signal
- NHL domain – NCL-1, HT2A and Lin-41 domain
- SH3 domain – Src-homology 3 domain
- CRF – corticotrophin releasing factor
- TCAP – teneurin C-terminal associated peptide
- RIP – regulated intramembrane proteolysis
- KO – knockout
- XLMR – X-linked mental retardation



## 9. List of Figures

Figure 2.1 Cell-cell and cell-ECM adhesion molecules and junctions

Figure 2.2 Schematic drawing of the basement membrane

Figure 2.3 Collagen type I structure

Figure 2.4 Overview of the proteoglycan family

Figure 2.5 Structure of laminins

Figure 2.6 Tenascin-C hexabrachion

Figure 2.7 Cell-ECM adhesion contacts

Figure 2.8 Major families of CAMs

Figure 2.9 Adherens junctions

Figure 2.10 Tight junctions

Figure 2.11 Gap junctions

Figure 2.12 Early human embryonic development

Figure 2.13 Lateral view of the developing brain

Figure 2.14 Patterning of the neocortex

Figure 2.15 Axon guidance

Figure 2.16 General teneurin structure

Figure 2.17 Current model of teneurin molecular function

Figure 2.18 Ten-m1-4 homodimers

Figure 5.1 Chicken teneurin-1 and -2 NHL repeat domain constructs

Figure 5.2 Retest of chosen construct

Figure 5.3 Pretesting expression in large-scale culture

Figure 5.4 Large-scale purification by Ni-NTA agarose beads

Figure 5.5 Size-exclusion chromatography

Figure 5.6 Fractions including purified protein

Figure 5.7 Crystallization drops

Figure 5.8 Limited Proteolysis

Figure 6.1 BC component structure

## **10. List of Tables**

Table 2.1 Summary of teneurin nomenclature and expression during development

Table 5.1 Primers used for in-fusion cloning of constructs

Table 5.2 Construct boundaries and sizes

Table 5.3 Starting concentrations of trypsin in limited proteolysis

Table 5.4 Point mutations identified in chicken teneurin-1 and -2 NHL repeat domain constructs

## 11. Acknowledgements

First of all, I would like to thank my boss, Ruth Chiquet-Ehrismann, for giving me the chance to do my PhD work in her lab and for her continuous support and guidance throughout the different projects I was working on. The open door policy always made it easy to discuss with her anytime. I am very grateful for everything she did for me!

Further, I am very thankful to my thesis committee members Richard Tucker, and Patrick Matthias, who have also guided me throughout my PhD years. They have always been there to talk to, and have given great input for the projects.

I am also happy that I got the chance to work in such a great lab, which has become like a second family to me. We always had fun together, and you all made it a very enjoyable work environment. I miss the people who have already left, and I will miss everyone else, once I leave.

The ‘good soul’ of the lab is of course Jacqueline Ferralli. She keeps the whole lab running smoothly, and makes it a lot easier for us to do our work. In my case (as for many others), she has given so much support with experiments, discussions, and otherwise that she has helped me substantially with finishing my PhD. I would also like to think that she has become a friend over the past 5 years, with whom I have always enjoyed talking to.

Many thanks also to Maria Asparuhova, Jan Beckmann, Ismaïl Hendaoui, Enrico Martina, and Matthias Scharenberg for being kind and patient when helping me with experiments and giving me scientific input when I had questions.

Matze, Jan, Dominik, Ismaïl, and Rahel, I want to thank you all for just being my friends, and being there for me. Coffee breaks, soccer games, lunches, discussions and so many other

good times would not have been the same without all of you! I have also made other friendships and met so many great people at the FMI, but you guys are the most important.

I want to truly thank my parents Imke Trautwein-Schöler and Hans Schöler for letting me choose my own path, supporting my decisions, and for everything they have done for me. However, I am certain that my dad's passion for science had some influence on my educational direction. I can still remember when I was about four years old, and my dad and I went on several bike tours around Lake Starnberg (German: Starnberger See) in Bavaria and we kept stopping to look at many different insects along the way. This may already have piqued my interest in science at an early age. But, no matter how I found my path to this point, I know my parents were there every step of the way, and I wouldn't be where I am today without them.

Last but definitely not least, I want to thank my partner Sarah Bosshard for her continuous support throughout my PhD years and especially during the last months of writing my thesis. You gave me our wonderful daughter Hanna Sophia, and took care of her so many nights and weekends while I had to write. I want to end in a quote from my dad's doctoral thesis that he was writing when I was a baby and my mom took care of me nights and weekends, showing how history repeats itself.

Original: "...und nicht zu vergessen meinem Sohn Jonas, der oft genug schon so einsichtig war während der Niederschrift dieser Arbeit zu schlafen, statt mit mir weiter spielen zu wollen."  
( "...and not to forget my son Jonas, who was already understanding often enough to sleep during the writing of my thesis, rather than to keep playing with me.")

So, I want to thank my daughter by slightly adapting that quote because it fits perfectly: "...und nicht zu vergessen meine Tochter Hanna, die oft genug schon so einsichtig war während der Niederschrift dieser Arbeit zu schlafen, statt mit mir weiter spielen zu wollen."

## Appendix A - Microarray data of TEN1-ICD overexpression in BS149 cells

Up-regulated genes (p-value < 0.01; Fold Change > 1.5)

Gene Name	Gene Description	Fold Change
A2M	alpha-2-macroglobulin	1.69
ABI3BP	ABI family, member 3 (NESH) binding protein	1.56
ACAT2	acetyl-CoA acetyltransferase 2	1.50
ACSS2	acyl-CoA synthetase short-chain family member 2	1.59
ADAM22	ADAM metallopeptidase domain 22	1.54
AFAP1L2	actin filament associated protein 1-like 2	1.77
AGPHD1	aminoglycoside phosphotransferase domain containing 1	1.57
AIF1L	allograft inflammatory factor 1-like	1.52
ALDOC	aldolase C, fructose-bisphosphate	2.35
ALX1	ALX homeobox 1	3.29
ANGPTL1	angiopoietin-like 1	1.58
ANKRD30A	ankyrin repeat domain 30A	2.05
APOD	apolipoprotein D	1.65
ARHGEF6	Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6	1.62
ARHGEF9	Cdc42 guanine nucleotide exchange factor (GEF) 9	1.93
AZGP1	alpha-2-glycoprotein 1, zinc-binding	2.42
B3GALT1	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 1	1.82
BGN	biglycan	3.34
BST2	bone marrow stromal cell antigen 2	1.56
BTG2	BTG family, member 2	1.55
C1R	complement component 1, r subcomponent	1.65
C3	complement component 3	3.69
C5orf4	chromosome 5 open reading frame 4	2.12
C6orf192	chromosome 6 open reading frame 192	1.95
C7orf58	chromosome 7 open reading frame 58	1.75
CA12	carbonic anhydrase XII	1.95
CA9	carbonic anhydrase IX	1.78
CACNA2D1	calcium channel, voltage-dependent, alpha 2/delta subunit 1	2.18
CADM3	cell adhesion molecule 3	1.63
CCL20	chemokine (C-C motif) ligand 20	1.56
CCL5	chemokine (C-C motif) ligand 5	2.93
CDH19	cadherin 19, type 2	1.66
CELF2	CUGBP, Elav-like family member 2	1.58
CFI	complement factor I	1.59
CHL1	cell adhesion molecule with homology to L1CAM (close homolog of L1)	3.43
CHRDL1	chordin-like 1	2.99
CHRNA6	cholinergic receptor, nicotinic, alpha 6	2.69
CKB	creatine kinase, brain	1.69
CMTM5	CKLF-like MARVEL transmembrane domain containing 5	1.90
CNTN1	contactin 1	2.20
CP	ceruloplasmin (ferroxidase)	1.56
CPE	carboxypeptidase E	1.63
CRELD1	cysteine-rich with EGF-like domains 1	1.57
CSGALNACT1	chondroitin sulfate N-acetylgalactosaminyltransferase 1	1.72
CXADR	coxsackie virus and adenovirus receptor	2.42
DAAM2	dishevelled associated activator of morphogenesis 2	1.55
DDX58	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	1.59
DEPDC6	DEP domain containing 6	1.91

DEPDC7	DEP domain containing 7	1.90
DHCR7	7-dehydrocholesterol reductase	2.22
DKFZp434F142	hypothetical DKFZp434F142	1.52
DOCK4	dedicator of cytokinesis 4	1.83
EDIL3	EGF-like repeats and discoidin I-like domains 3	1.52
EDNRB	endothelin receptor type B	2.14
ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2	1.72
ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	2.35
EXTL1	exostoses (multiple)-like 1	1.84
FABP7	fatty acid binding protein 7, brain	1.91
FADS2	fatty acid desaturase 2	2.79
FAIM2	Fas apoptotic inhibitory molecule 2	1.61
FAM134B	family with sequence similarity 134, member B	3.06
FAM198B	family with sequence similarity 198, member B	1.60
FAM70A	family with sequence similarity 70, member A	2.10
FAM99A	family with sequence similarity 99, member A	1.56
FCRLA	Fc receptor-like A	3.61
FDFT1	farnesyl-diphosphate farnesyltransferase 1	1.54
FHL1	four and a half LIM domains 1	2.15
FLJ35776	hypothetical LOC649446	1.56
FOXP2	forkhead box P2	1.77
FXYD3	FXYD domain containing ion transport regulator 3	3.21
GAS7	growth arrest-specific 7	1.71
GBP2	guanylate binding protein 2, interferon-inducible	1.56
GLB1L	galactosidase, beta 1-like	1.60
GMFG	glia maturation factor, gamma	1.59
GPCPD1	glycerophosphocholine phosphodiesterase GDE1 homolog ( <i>S. cerevisiae</i> )	1.71
GPNMB	glycoprotein (transmembrane) nmb	2.41
GPR155	G protein-coupled receptor 155	1.79
GPR162	G protein-coupled receptor 162	1.68
GPR37	G protein-coupled receptor 37 (endothelin receptor type B-like)	2.13
GRAMD3	GRAM domain containing 3	1.64
GSTA4	glutathione S-transferase alpha 4	1.55
HCG27	HLA complex group 27	1.76
HEXIM1	hexamethylene bis-acetamide inducible 1	1.54
HEY2	hairly/enhancer-of-split related with YRPW motif 2	2.39
HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	4.80
HMCN1	hemicentin 1	2.16
HMGCR	3-hydroxy-3-methylglutaryl-CoA reductase	2.00
HMGCS1	3-hydroxy-3-methylglutaryl-CoA synthase 1 (soluble)	2.11
HOXA5	homeobox A5	1.94
HSD17B7	hydroxysteroid (17-beta) dehydrogenase 7	2.07
HSD17B7P2	hydroxysteroid (17-beta) dehydrogenase 7 pseudogene 2	1.72
ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein	1.53
IDI1	isopentenyl-diphosphate delta isomerase 1	1.86
IFI35	interferon-induced protein 35	1.53
IFIH1	interferon induced with helicase C domain 1	1.62
IFIT1	interferon-induced protein with tetratricopeptide repeats 1	1.80
IFIT2	interferon-induced protein with tetratricopeptide repeats 2	1.85
IFITM1	interferon induced transmembrane protein 1 (9-27)	2.78
IFITM2	interferon induced transmembrane protein 2 (1-8D)	1.94
IGFBP5	insulin-like growth factor binding protein 5	1.88
IGLJ3	immunoglobulin lambda joining 3	1.92

INSIG1	insulin induced gene 1	2.96
ITGA9	integrin, alpha 9	1.60
ITIH5L	inter-alpha (globulin) inhibitor H5-like	2.08
ITPR2	inositol 1,4,5-triphosphate receptor, type 2	1.63
KCND1	potassium voltage-gated channel, Shal-related subfamily, member 1	1.53
KCNK5	potassium channel, subfamily K, member 5	1.62
KIAA1161	KIAA1161	1.59
KLF9	Kruppel-like factor 9	2.00
KLHL13	kelch-like 13 (Drosophila)	1.53
KLHL24	kelch-like 24 (Drosophila)	1.60
LAMA4	laminin, alpha 4	2.45
LAPTM5	lysosomal protein transmembrane 5	1.50
LASS4	LAG1 homolog, ceramide synthase 4	1.60
LDLR	low density lipoprotein receptor	1.57
LGMN	legumain	1.53
LIFR	leukemia inhibitory factor receptor alpha	1.54
LIPG	lipase, endothelial	1.56
LOC400804	hypothetical LOC400804	1.59
LRRTM4	leucine rich repeat transmembrane neuronal 4	1.73
LSAMP	limbic system-associated membrane protein	1.93
LUZP4	leucine zipper protein 4	1.92
LY96	lymphocyte antigen 96	1.60
MANSC1	MANSC domain containing 1	1.55
MAOB	monoamine oxidase B	1.70
MAP3K12	mitogen-activated protein kinase kinase kinase 12	1.53
METTL7A	methyltransferase like 7A	1.55
MGP	matrix Gla protein	1.90
MIR186	microRNA 186	1.51
MKNK2	MAP kinase interacting serine/threonine kinase 2	1.51
MMP16	matrix metalloproteinase 16 (membrane-inserted)	1.64
MMP16	matrix metalloproteinase 16 (membrane-inserted)	1.50
MMP19	matrix metalloproteinase 19	1.82
MRGPRX4	MAS-related GPR, member X4	1.76
MYEF2	myelin expression factor 2	1.54
NCALD	neurocalcin delta	2.03
NCAM2	neural cell adhesion molecule 2	1.75
NFIA	nuclear factor I/A	2.63
NR4A1	nuclear receptor subfamily 4, group A, member 1	2.26
NR4A2	nuclear receptor subfamily 4, group A, member 2	2.45
NRCAM	neuronal cell adhesion molecule	2.65
OASL	2'-5'-oligoadenylate synthetase-like	1.55
ODZ1	odz, odd Oz/ten-m homolog 1(Drosophila)	2.79
ORAI3	ORAI calcium release-activated calcium modulator 3	1.64
PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	1.76
PCDH10	protocadherin 10	2.13
PDE3A	phosphodiesterase 3A, cGMP-inhibited	2.23
PDE4B	phosphodiesterase 4B, cAMP-specific (E4 dunce homolog, Drosophila)	1.68
PDE7B	phosphodiesterase 7B	1.72
PDGFRL	platelet-derived growth factor receptor-like	1.56
PKD4	pyruvate dehydrogenase kinase, isozyme 4	1.59
PELI1	pellino homolog 1 (Drosophila)	1.54
PHLPP1	PH domain and leucine rich repeat protein phosphatase 1	1.51
PLAC1	placenta-specific 1	1.57



PLCD4	phospholipase C, delta 4	1.66
PLD1	phospholipase D1, phosphatidylcholine-specific	1.52
PLEKHH1	pleckstrin homology domain containing, family H (with MyTH4 domain) 1	1.53
PLP1	proteolipid protein 1	1.87
PMP2	peripheral myelin protein 2	2.58
PNPLA3	patatin-like phospholipase domain containing 3	2.40
PODXL	podocalyxin-like	1.69
PRELP	proline/arginine-rich end leucine-rich repeat protein	3.41
PTGS1	prostaglandin-endoperoxide synthase 1	1.93
PTPN13	protein tyrosine phosphatase, non-receptor type 13	1.63
PTPRZ1	protein tyrosine phosphatase, receptor-type, Z polypeptide 1	2.54
PTTG1	pituitary tumor-transforming 1	1.50
RAB3A	RAB3A, member RAS oncogene family	1.69
RASGRP3	RAS guanyl releasing protein 3 (calcium and DAG-regulated)	1.78
RASSF4	Ras association (RalGDS/AF-6) domain family member 4	1.62
RDH11	retinol dehydrogenase 11 (all-trans/9-cis/11-cis)	1.52
RFTN2	raftlin family member 2	1.97
RNF150	ring finger protein 150	1.84
ROBO1	roundabout, axon guidance receptor, homolog 1 (Drosophila)	1.55
ROBO2	roundabout, axon guidance receptor, homolog 2 (Drosophila)	1.96
RPE65	retinal pigment epithelium-specific protein 65kDa	1.90
RXRG	retinoid X receptor, gamma	2.59
S100A3	S100 calcium binding protein A3	1.72
S100B	S100 calcium binding protein B	1.79
SAA1	serum amyloid A1	4.48
SAT1	spermidine/spermine N1-acetyltransferase 1	1.93
SC4MOL	sterol-C4-methyl oxidase-like	1.83
SCARB1	scavenger receptor class B, member 1	2.02
SCD	stearoyl-CoA desaturase (delta-9-desaturase)	2.44
SCRG1	stimulator of chondrogenesis 1	2.92
SDF2L1	stromal cell-derived factor 2-like 1	1.68
SEMA3A	semaphorin 3A	1.55
SEMA3D	semaphorin 3D	1.59
SEMA4B	semaphorin 4B	1.68
SEMA4G	semaphorin 4G	1.61
SEMA6A	semaphorin 6A	1.55
SEPT4	septin 4	1.76
SERPINA5	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), 5	1.63
SKAP2	src kinase associated phosphoprotein 2	1.78
SLAIN1	SLAIN motif family, member 1	2.10
SLC16A4	solute carrier family 16, member 4 (monocarboxylic acid transporter 5)	1.58
SLC1A1	solute carrier family 1 (neuronal/epithelial glutamate transporter) 1	1.53
SLC1A4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	1.57
SLC25A45	solute carrier family 25, member 45	1.67
SLC29A2	solute carrier family 29 (nucleoside transporters), member 2	1.51
SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3	1.57
SLC35F1	solute carrier family 35, member F1	1.59
SLC38A3	solute carrier family 38, member 3	1.81
SLC44A2	solute carrier family 44, member 2	1.53
SLCO4A1	solute carrier organic anion transporter family, member 4A1	1.85
SOBP	sine oculis binding protein homolog (Drosophila)	1.66
SOD2	superoxide dismutase 2, mitochondrial	2.08
SOD3	superoxide dismutase 3, extracellular	1.74

SOX5	SRY (sex determining region Y)-box 5	1.89
SPRY1	sprouty homolog 1, antagonist of FGF signaling (Drosophila)	1.50
SQLE	squalene epoxidase	1.93
SREBF1	sterol regulatory element binding transcription factor 1	1.69
SREBF2	sterol regulatory element binding transcription factor 2	2.00
ST3GAL5	ST3 beta-galactoside alpha-2,3-sialyltransferase 5	1.54
ST8SIA1	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 1	2.28
ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4	1.89
STARD4	StAR-related lipid transfer (START) domain containing 4	2.66
STC1	stanniocalcin 1	1.65
STMN2	stathmin-like 2	1.57
TDRKH	tudor and KH domain containing	1.74
TF	transferrin	1.64
TGFBR3	transforming growth factor, beta receptor III	1.54
THBS2	thrombospondin 2	1.71
THBS3	thrombospondin 3	1.52
TMEM135	transmembrane protein 135	1.51
TMEM195	transmembrane protein 195	1.65
TMEM229B	transmembrane protein 229B	2.53
TMPRSS15	transmembrane protease, serine 15	1.51
TNFAIP6	tumor necrosis factor, alpha-induced protein 6	2.26
TNFRSF19	tumor necrosis factor receptor superfamily, member 19	2.58
TP53INP1	tumor protein p53 inducible nuclear protein 1	1.59
TPCN1	two pore segment channel 1	1.50
TRIM9	tripartite motif-containing 9	1.58
TRPM8	transient receptor potential cation channel, subfamily M, member 8	2.26
TTC28	tetratricopeptide repeat domain 28	1.77
TUBB4	tubulin, beta 4	2.13
USP2	ubiquitin specific peptidase 2	1.88
WIP1	WD repeat domain, phosphoinositide interacting 1	1.57
YPEL2	yippee-like 2 (Drosophila)	1.66
ZKSCAN1	zinc finger with KRAB and SCAN domains 1	1.51

Down-regulated genes (p-value < 0.01; Fold Change < -1.5)

Gene Name	Gene Description	Fold Change
ABCC3	ATP-binding cassette, sub-family C (CFTR/MRP), member 3	-1.66
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	-1.62
ABLIM3	actin binding LIM protein family, member 3	-1.71
ACPL2	acid phosphatase-like 2	-1.56
ACTBL2	actin, beta-like 2	-2.31
ADD2	adducin 2 (beta)	-2.10
ADRA1B	adrenergic, alpha-1B-, receptor	-2.23
AGPAT9	1-acylglycerol-3-phosphate O-acyltransferase 9	-1.55
AK3L1	adenylate kinase 3-like 1	-1.56
AKR1C2	aldo-keto reductase family 1, member C2	-2.64
AKR1C3	aldo-keto reductase family 1, member C3	-1.93
ALDH1A3	aldehyde dehydrogenase 1 family, member A3	-1.84
AMIGO2	adhesion molecule with Ig-like domain 2	-3.65
ANGPTL4	angiopoietin-like 4	-3.90
ANKRD1	ankyrin repeat domain 1 (cardiac muscle)	-6.63
ANO1	anoctamin 1, calcium activated chloride channel	-1.93
AOC2	amine oxidase, copper containing 2 (retina-specific)	-1.56
APOBEC3G	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G	-1.51
ARRDC3	arrestin domain containing 3	-1.52
AVEN	apoptosis, caspase activation inhibitor	-1.51
B3GALNT1	beta-1,3-N-acetylgalactosaminyltransferase 1 (globoside blood group)	-1.91
BCAR3	breast cancer anti-estrogen resistance 3	-1.65
BCL2L1	BCL2-like 1	-1.51
BDNF	brain-derived neurotrophic factor	-1.62
BEND6	BEN domain containing 6	-1.66
BHLHE41	basic helix-loop-helix family, member e41	-1.81
BLID	BH3-like motif containing, cell death inducer	-1.50
C3orf59	chromosome 3 open reading frame 59	-1.66
C6orf138	chromosome 6 open reading frame 138	-1.81
C6orf145	chromosome 6 open reading frame 145	-1.64
C6orf168	chromosome 6 open reading frame 168	-1.57
C6orf191	chromosome 6 open reading frame 191	-1.56
CA8	carbonic anhydrase VIII	-1.92
CALB2	calbindin 2	-2.35
CAMK2N1	calcium/calmodulin-dependent protein kinase II inhibitor 1	-2.03
CAV2	caveolin 2	-1.63
CCND1	cyclin D1	-1.67
CCNE2	cyclin E2	-1.67
CDC25A	cell division cycle 25 homolog A (S. pombe)	-1.56
CDRT1	CMT1A duplicated region transcript 1	-2.40
CDRT1	CMT1A duplicated region transcript 1	-2.26
CDRT1	CMT1A duplicated region transcript 1	-2.15
CNGA2	cyclic nucleotide gated channel alpha 2	-2.46
CPA4	carboxypeptidase A4	-1.61
CPEB1	cytoplasmic polyadenylation element binding protein 1	-1.50
CYTSB	cytospin B	-1.51
DCBLD1	discoidin, CUB and LCCL domain containing 1	-1.82
DCLK2	doublecortin-like kinase 2	-1.55
DDAH1	dimethylarginine dimethylaminohydrolase 1	-1.81
DKK1	dickkopf homolog 1 (Xenopus laevis)	-2.83

DLX2	distal-less homeobox 2	-1.75
DPYSL3	dihydropyrimidinase-like 3	-1.57
DUSP10	dual specificity phosphatase 10	-1.67
EDA2R	ectodysplasin A2 receptor	-1.69
EDN1	endothelin 1	-1.62
EFNB2	ephrin-B2	-1.99
EGFR	epidermal growth factor receptor	-1.72
EGR2	early growth response 2	-1.90
EID3	EP300 interacting inhibitor of differentiation 3	-1.59
EIF4E3	eukaryotic translation initiation factor 4E family member 3	-1.68
EMB	embigin homolog (mouse)	-1.76
EPB41L4A	erythrocyte membrane protein band 4.1 like 4A	-2.43
EPHA2	EPH receptor A2	-1.54
EPHB2	EPH receptor B2	-2.11
ERRFI1	ERBB receptor feedback inhibitor 1	-1.63
ETS2	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)	-1.93
F3	coagulation factor III (thromboplastin, tissue factor)	-1.57
FAM105A	family with sequence similarity 105, member A	-1.81
FAM171A1	family with sequence similarity 171, member A1	-1.82
FAT4	FAT tumor suppressor homolog 4 (Drosophila)	-1.62
FERMT1	fermitin family homolog 1 (Drosophila)	-1.54
FGF12	fibroblast growth factor 12	-1.61
FJX1	four jointed box 1 (Drosophila)	-1.65
GADD45A	growth arrest and DNA-damage-inducible, alpha	-1.71
GAP43	growth associated protein 43	-1.57
GCLC	glutamate-cysteine ligase, catalytic subunit	-1.71
GCLM	glutamate-cysteine ligase, modifier subunit	-1.52
GDF15	growth differentiation factor 15	-2.06
GFPT2	glutamine-fructose-6-phosphate transaminase 2	-1.67
GHR	growth hormone receptor	-1.69
GRAMD1C	GRAM domain containing 1C	-1.68
GTF2IRD2B	GTF2I repeat domain containing 2B	-2.42
HEG1	HEG homolog 1 (zebrafish)	-1.66
HERC3	hect domain and RLD 3	-2.13
HMGA2	high mobility group AT-hook 2	-1.61
HOXC11	homeobox C11	-1.54
HOXC13	homeobox C13	-1.71
ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein	-1.76
IL7R	interleukin 7 receptor	-2.39
INHBA	inhibin, beta A	-1.70
ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	-1.63
JAG1	jagged 1 (Alagille syndrome)	-1.64
JUB	jub, ajuba homolog (Xenopus laevis)	-1.59
KCNH1	potassium voltage-gated channel, subfamily H (eag-related), member 1	-2.05
KIAA0748	KIAA0748	-1.58
KIAA1644	KIAA1644	-1.68
KLHL21	kelch-like 21 (Drosophila)	-1.54
KLHL4	kelch-like 4 (Drosophila)	-2.26
KLK6	kallikrein-related peptidase 6	-1.58
KLKB1	kallikrein B, plasma (Fletcher factor) 1	-1.53
KRT81	keratin 81	-1.61
LIF	leukemia inhibitory factor (cholinergic differentiation factor)	-1.51
LIMCH1	LIM and calponin homology domains 1	-1.70

LOC100133299	GALI1870	-1.51
LOC151009	hypothetical LOC151009	-1.52
LOC402778	CD225 family protein FLJ76511	-3.65
LPHN3	latrophilin 3	-2.27
LRRFIP1	leucine rich repeat (in FLII) interacting protein 1	-1.76
LRRFIP1	leucine rich repeat (in FLII) interacting protein 1	-1.61
LTBP1	latent transforming growth factor beta binding protein 1	-1.70
LYPD6	LY6/PLAUR domain containing 6	-1.60
MAGEC2	melanoma antigen family C, 2	-1.53
MAP1B	microtubule-associated protein 1B	-1.58
MGLL	monoglyceride lipase	-1.84
MICAL2	microtubule associated monooxygenase, calponin and LIM domain cont. 2	-1.95
MIR221	microRNA 221	-1.75
MLPH	melanophilin	-2.11
MT1E	metallothionein 1E	-1.62
MT1F	metallothionein 1F	-1.90
MTTP	microsomal triglyceride transfer protein	-1.56
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	-2.14
MYPN	myopalladin	-3.31
NAV3	neuron navigator 3	-2.02
NEO1	neogenin homolog 1 (chicken)	-2.46
NEXN	nexilin (F actin binding protein)	-2.25
NGFR	nerve growth factor receptor	-2.37
NIPSNAP3A	nipsnap homolog 3A (C. elegans)	-1.55
NPIPL3	nuclear pore complex interacting protein-like 3	-1.86
NPTX2	neuronal pentraxin II	-1.74
NRG1	neuregulin 1	-2.49
NRP1	neuropilin 1	-1.70
OBFC2A	oligonucleotide/oligosaccharide-binding fold containing 2A	-1.52
ODZ2	odz, odd Oz/ten-m homolog 2 (Drosophila)	-1.54
PAQR5	progesterin and adipoQ receptor family member V	-1.53
PAWR	PRKC, apoptosis, WTI, regulator	-1.75
PCBP3	poly(rC) binding protein 3	-1.70
PDE1C	phosphodiesterase 1C, calmodulin-dependent 70kDa	-2.05
PER2	period homolog 2 (Drosophila)	-1.53
PGM5P2	phosphoglucomutase 5 pseudogene 2	-1.60
PHF15	PHD finger protein 15	-1.55
PHLDB2	pleckstrin homology-like domain, family B, member 2	-1.58
PLXNA1	plexin A1	-1.61
PLXNA2	plexin A2	-1.80
PORCN	porcupine homolog (Drosophila)	-1.79
PRKAG2	protein kinase, AMP-activated, gamma 2 non-catalytic subunit	-1.64
PRR16	proline rich 16	-1.57
PRR4	proline rich 4 (lacrima)	-1.71
PTGER2	prostaglandin E receptor 2 (subtype EP2), 53kDa	-2.20
PTH1H	parathyroid hormone-like hormone	-2.79
PTPN22	protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	-2.05
PTPRB	protein tyrosine phosphatase, receptor type, B	-1.56
PTX3	pentraxin 3, long	-1.73
QRFPR	pyroglutamylated RFamide peptide receptor	-1.92
RAB27B	RAB27B, member RAS oncogene family	-1.74
RAC2	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP BP Rac2)	-1.67
RAD51	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)	-1.56

RALGPS2	Ral GEF with PH domain and SH3 binding motif 2	-1.79
RBPMS	RNA binding protein with multiple splicing	-1.61
RGAG1	retrotransposon gag domain containing 1	-1.66
RGNEF	190 kDa guanine nucleotide exchange factor	-1.64
RGS20	regulator of G-protein signaling 20	-1.52
RGS7	regulator of G-protein signaling 7	-1.86
RNU5D	RNA, U5D small nuclear	-1.66
RPSAP52	ribosomal protein SA pseudogene 52	-3.26
SCG2	secretogranin II	-1.54
SCNN1A	sodium channel, nonvoltage-gated 1 alpha	-1.67
SDC1	syndecan 1	-1.77
SEMA4A	Semaphoring 4A	-1.84
SERINC5	serine incorporator 5	-2.18
SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	-1.92
SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5	-1.53
SERPINE1	serpin peptidase inhibitor, clade E (nexin, PAI type 1), member 1	-3.38
SH3RF2	SH3 domain containing ring finger 2	-2.20
SLC12A8	solute carrier family 12 (potassium/chloride transporters), member 8	-1.69
SLC16A6	solute carrier family 16, member 6 (monocarboxylic acid transporter 7)	-3.38
SLC35F2	solute carrier family 35, member F2	-1.61
SLC37A1	solute carrier family 37 (glycerol-3-phosphate transporter) 1	-1.63
SLC37A2	solute carrier family 37 (glycerol-3-phosphate transporter) 2	-1.65
SLC7A11	solute carrier family 7, (cationic amino acid transporter, y+ system) 11	-2.75
SMAD7	SMAD family member 7	-1.82
SMAGP	small cell adhesion glycoprotein	-1.68
SNAI2	snail homolog 2 (Drosophila)	-2.02
SPAG11B	sperm associated antigen 11B	-1.69
SPANXA2	SPANX family, member A2	-1.60
SPANXE	SPANX family, member E	-1.84
SPRR2D	small proline-rich protein 2D	-1.62
SRXN1	sulfiredoxin 1 homolog (S. cerevisiae)	-1.71
SSX4	synovial sarcoma, X breakpoint 4	-1.62
SSX6	synovial sarcoma, X breakpoint 6 (pseudogene)	-1.51
STAC	SH3 and cysteine rich domain	-1.78
STC2	stanniocalcin 2	-2.40
STRA6	stimulated by retinoic acid gene 6 homolog (mouse)	-1.51
STX1A	syntaxin 1A (brain)	-1.53
SULT1B1	sulfotransferase family, cytosolic, 1B, member 1	-1.56
SYNC	syncoilin, intermediate filament protein	-1.59
TBX3	T-box 3	-1.88
TBXAS1	thromboxane A synthase 1 (platelet)	-1.69
TCF19	transcription factor 19	-1.70
TGFB2	transforming growth factor, beta 2	-2.03
TGFBR2	transforming growth factor, beta receptor II (70/80kDa)	-1.50
TGM2	transglutaminase 2	-1.52
THSD4	thrombospondin, type I, domain containing 4	-1.86
TINAGL1	tubulointerstitial nephritis antigen-like 1	-1.76
TLR4	toll-like receptor 4	-1.83
TM4SF18	transmembrane 4 L six family member 18	-1.96
TMEM171	transmembrane protein 171	-1.77
TMEM183A	transmembrane protein 183A	-1.80
TMTC1	transmembrane and tetratricopeptide repeat containing 1	-1.51
TRHDE	thyrotropin-releasing hormone degrading enzyme	-3.16

TRIM16L	tripartite motif-containing 16-like	-1.75
TRIM22	tripartite motif-containing 22	-1.66
TRIM73	tripartite motif-containing 73	-1.86
TRIML2	tripartite motif family-like 2	-2.02
TRPA1	transient receptor potential cation channel, subfamily A, member 1	-2.10
TRPC1	transient receptor potential cation channel, subfamily C, member 1	-1.53
TSKU	tsukushi small leucine rich proteoglycan homolog ( <i>Xenopus laevis</i> )	-1.51
TXNRD1	thioredoxin reductase 1	-1.94
UNG	uracil-DNA glycosylase	-1.55
VLDLR	very low density lipoprotein receptor	-1.68
WNT5A	wingless-type MMTV integration site family, member 5A	-1.90
XAGE1A	X antigen family, member 1A	-1.58
ZNF365	zinc finger protein 365	-1.51
ZNF643	zinc finger protein 643	-1.56

## Appendix B - Microarray data of MITF overexpression in BS149 cells

Up-regulated genes (p-value < 0.01; Fold Change > 1.5)

Gene Name	Gene Description	Fold Change
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	1.79
ABHD6	abhydrolase domain containing 6	1.96
ACTN2	actinin, alpha 2	3.54
ADM	adrenomedullin	2.00
AGXT2L2	alanine-glyoxylate aminotransferase 2-like 2	1.51
AHNAK2	AHNAK nucleoprotein 2	3.93
AIFM1	apoptosis-inducing factor, mitochondrion-associated, 1	1.90
ALDH1L2	aldehyde dehydrogenase 1 family, member L2	1.56
ALOXE3	arachidonate lipoxygenase 3	1.67
ALS2CR12	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 12	2.55
ALS2CR8	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 8	1.55
AMDHD2	amidohydrolase domain containing 2	2.07
AMZ2P1	archaelysin family metallopeptidase 2 pseudogene 1	1.97
ANGPTL2	angiopoietin-like 2	3.59
ANKMY2	ankyrin repeat and MYND domain containing 2	1.73
ANKRD1	ankyrin repeat domain 1 (cardiac muscle)	3.41
AP1AR	adaptor-related protein complex 1 associated regulatory protein	1.57
APBB3	amyloid beta (A4) precursor protein-binding, family B, member 3	1.64
ARFGAP1	ADP-ribosylation factor GTPase activating protein 1	1.67
ARMCX1	armadillo repeat containing, X-linked 1	1.66
ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase) 1	2.37
ATG2B	ATG2 autophagy related 2 homolog B ( <i>S. cerevisiae</i> )	1.61
ATP6V0D1	ATPase, H <sup>+</sup> transporting, lysosomal 38kDa, V0 subunit d1	1.57
ATP6V0D2	ATPase, H <sup>+</sup> transporting, lysosomal 38kDa, V0 subunit d2	10.18
ATP6V1B2	ATPase, H <sup>+</sup> transporting, lysosomal 56/58kDa, V1 subunit B2	1.71
ATP6V1C1	ATPase, H <sup>+</sup> transporting, lysosomal 42kDa, V1 subunit C1	1.76
ATP6V1F	ATPase, H <sup>+</sup> transporting, lysosomal 14kDa, V1 subunit F	1.71
ATP6V1H	ATPase, H <sup>+</sup> transporting, lysosomal 50/57kDa, V1 subunit H	1.72
ATP7A	ATPase, Cu <sup>++</sup> transporting, alpha polypeptide	2.45
ATXN7L1	ataxin 7-like 1	1.56
BAIAP2	BAI1-associated protein 2	1.96
BBS7	Bardet-Biedl syndrome 7	1.83
BEST1	bestrophin 1	6.95
BHLHE41	basic helix-loop-helix family, member e41	2.07
BLVRB	biliverdin reductase B (flavin reductase (NADPH))	1.64
BNC2	basonuclin 2	2.11
C12orf49	chromosome 12 open reading frame 49	2.80
C12orf50	chromosome 12 open reading frame 50	1.86
C12orf66	chromosome 12 open reading frame 66	1.72
C14orf128	chromosome 14 open reading frame 128	1.63
C14orf129	chromosome 14 open reading frame 129	1.55
C14orf79	chromosome 14 open reading frame 79	1.80
C15orf34	chromosome 15 open reading frame 34	2.48
C18orf19	chromosome 18 open reading frame 19	1.85
C1orf189	chromosome 1 open reading frame 189	1.78
C1orf51	chromosome 1 open reading frame 51	1.53
C1orf85	chromosome 1 open reading frame 85	1.57



C1orf97	chromosome 1 open reading frame 97	1.51
C20orf165	chromosome 20 open reading frame 165	1.53
C21orf90	chromosome 21 open reading frame 90	2.17
C21orf91	chromosome 21 open reading frame 91	1.74
C22orf25	chromosome 22 open reading frame 25	3.08
C2orf86	chromosome 2 open reading frame 86	1.54
C6orf115	chromosome 6 open reading frame 115	1.54
C6orf192	chromosome 6 open reading frame 192	1.66
C9orf91	chromosome 9 open reading frame 91	1.90
CAB39L	calcium binding protein 39-like	1.51
CABLES1	Cdk5 and Abl enzyme substrate 1	3.67
CAMKK1	calcium/calmodulin-dependent protein kinase kinase 1, alpha	1.72
CCBE1	collagen and calcium binding EGF domains 1	1.71
CCDC110	coiled-coil domain containing 110	1.68
CCDC113	coiled-coil domain containing 113	1.58
CCDC43	coiled-coil domain containing 43	1.63
CDK20	cyclin-dependent kinase 20	1.81
CDSN	corneodesmosin	11.45
CEP250	centrosomal protein 250kDa	1.76
CGB2	chorionic gonadotropin, beta polypeptide 2	1.58
CHKA	choline kinase alpha	2.00
CHRM4	cholinergic receptor, muscarinic 4	1.79
CHRN2	cholinergic receptor, nicotinic, beta 2 (neuronal)	2.04
CLCN6	chloride channel 6	2.28
CLCN7	chloride channel 7	2.23
CLDN12	claudin 12	1.62
CLN3	ceroid-lipofuscinosis, neuronal 3	1.68
CLN6	ceroid-lipofuscinosis, neuronal 6, late infantile, variant	1.54
COPZ2	coatamer protein complex, subunit zeta 2	1.51
CPNE7	copine VII	1.85
CPSF4	cleavage and polyadenylation specific factor 4, 30kDa	2.01
CRY1	cryptochrome 1 (photolyase-like)	1.96
CTNS	cystinosis, nephropathic	2.49
CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1	2.09
CYP2U1	cytochrome P450, family 2, subfamily U, polypeptide 1	2.23
DAB2	disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)	1.71
DCT	dopachrome tautomerase	1.58
DDI2	DNA-damage inducible 1 homolog 2 (S. cerevisiae)	1.58
DDIT3	DNA-damage-inducible transcript 3	2.15
DET1	de-etiolated homolog 1 (Arabidopsis)	1.51
DESI	Dexi homolog (mouse)	4.40
DFNB31	deafness, autosomal recessive 31	1.82
DNAJC12	DnaJ (Hsp40) homolog, subfamily C, member 12	1.59
DNAL4	dynein, axonemal, light chain 4	1.53
DNER	delta/notch-like EGF repeat containing	2.20
DOCK10	dedicator of cytokinesis 10	1.56
DPP4	dipeptidyl-peptidase 4	2.27
DPY19L2	dpy-19-like 2 (C. elegans)	1.62
DPY19L2P1	dpy-19-like 2 pseudogene 1 (C. elegans)	1.68
DPY19L2P2	dpy-19-like 2 pseudogene 2 (C. elegans)	1.73
DUSP10	dual specificity phosphatase 10	1.95
DUSP3	dual specificity phosphatase 3	1.54
DYRK3	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 3	3.08

EEA1	early endosome antigen 1	1.61
EEF1A2	eukaryotic translation elongation factor 1 alpha 2	1.55
EEPD1	endonuclease/exonuclease/phosphatase family domain containing 1	1.82
EFTUD1	elongation factor Tu GTP binding domain containing 1	1.50
EML5	echinoderm microtubule associated protein like 5	2.02
ENO2	enolase 2 (gamma, neuronal)	2.76
ENO3	enolase 3 (beta, muscle)	1.60
ENOSF1	enolase superfamily member 1	1.63
EPAS1	endothelial PAS domain protein 1	2.72
EPHA5	EPH receptor A5	2.72
EYA4	eyes absent homolog 4 (Drosophila)	3.18
FAM102A	family with sequence similarity 102, member A	1.78
FAM135A	family with sequence similarity 135, member A	1.56
FAM13C	family with sequence similarity 13, member C	1.57
FAM21A	family with sequence similarity 21, member A	1.87
FAM21B	family with sequence similarity 21, member B	2.24
FAM21C	family with sequence similarity 21, member C	2.02
FAM27A	family with sequence similarity 27, member A	1.57
FAM40B	family with sequence similarity 40, member B	2.88
FAM53B	family with sequence similarity 53, member B	1.57
FAM70A	family with sequence similarity 70, member A	1.93
FAM86B1	family with sequence similarity 86, member B1	1.67
FBXO25	F-box protein 25	2.32
FBXO32	F-box protein 32	1.77
FICD	FIC domain containing	2.70
FLJ10661	family with sequence similarity 86, member A pseudogene	2.02
FMN1	formin 1	2.21
FNIP2	folliculin interacting protein 2	2.58
FYB	FYN binding protein	2.48
GAB2	GRB2-associated binding protein 2	1.85
GBAP1	glucosidase, beta, acid pseudogene 1	2.32
GDF15	growth differentiation factor 15	2.81
GEM	GTP binding protein overexpressed in skeletal muscle	2.49
GK	glycerol kinase	1.82
GNPDA1	glucosamine-6-phosphate deaminase 1	2.18
GNPTAB	N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits	1.65
GOLGA4	golgin A4	1.92
GPNMB	glycoprotein (transmembrane) nmb	3.93
GPR56	G protein-coupled receptor 56	1.73
GREB1	growth regulation by estrogen in breast cancer 1	1.77
GYPC	glycophorin C (Gerbich blood group)	2.11
HCG4	HLA complex group 4	3.20
HEATR7A	HEAT repeat containing 7A	1.79
HEXA	hexosaminidase A (alpha polypeptide)	1.81
HEY1	hairy/enhancer-of-split related with YRPW motif 1	2.44
HEY2	hairy/enhancer-of-split related with YRPW motif 2	1.98
HIST2H2BF	histone cluster 2, H2bf	1.73
HK2	hexokinase 2	1.96
HLA-DOB	major histocompatibility complex, class II, DO beta	1.63
HLCS	holocarboxylase synthetase	1.68
HMGCL	3-hydroxymethyl-3-methylglutaryl-CoA lyase	1.60
HOXB9	homeobox B9	1.74
HSPB8	heat shock 22kDa protein 8	1.73

HSPBAP1	HSPB (heat shock 27kDa) associated protein 1	2.31
ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	1.63
IFI30	interferon, gamma-inducible protein 30	1.71
IL6R	interleukin 6 receptor	1.77
ILVBL	ilvB (bacterial acetolactate synthase)-like	1.54
INPP4B	inositol polyphosphate-4-phosphatase, type II, 105kDa	1.57
INSIG1	insulin induced gene 1	1.74
IVNS1ABP	influenza virus NS1A binding protein	1.60
KANK1	KN motif and ankyrin repeat domains 1	1.87
KCNAB1	potassium voltage-gated channel, shaker-related subfamily, beta member 1	1.57
KCNH1	potassium voltage-gated channel, subfamily H (eag-related), member 1	1.97
KCTD21	potassium channel tetramerisation domain containing 21	1.94
KCTD7	potassium channel tetramerisation domain containing 7	2.31
KIAA0415	KIAA0415	1.90
KIAA1009	KIAA1009	1.67
KIAA1632	KIAA1632	1.69
KIAA1671	KIAA1671	1.52
KIAA1737	KIAA1737	1.82
KIAA1919	KIAA1919	1.85
KIF3C	kinesin family member 3C	1.90
KLF15	Kruppel-like factor 15	2.67
KLF3	Kruppel-like factor 3 (basic)	1.53
KLHDC1	kelch domain containing 1	1.51
KLHL24	kelch-like 24 (Drosophila)	1.79
KMO	kynurenine 3-monooxygenase (kynurenine 3-hydroxylase)	1.54
KRTAP4-7	keratin associated protein 4-7	5.40
KU-MEL-3	KU-MEL-3	2.40
LACTB2	lactamase, beta 2	1.75
LAT	linker for activation of T cells	1.63
LCORL	ligand dependent nuclear receptor corepressor-like	1.77
LGALS3	lectin, galactoside-binding, soluble, 3	1.74
LN2	ligand of numb-protein X 2	1.69
LOC25845	hypothetical LOC25845	1.60
LOC285033	hypothetical protein LOC285033	1.59
LOC402778	CD225 family protein FLJ76511	1.63
LOC440957	similar to CG32736-PA	2.07
LOC654433	hypothetical LOC654433	1.86
LONRF3	LON peptidase N-terminal domain and ring finger 3	2.65
LRCH2	leucine-rich repeats and calponin homology (CH) domain containing 2	1.93
LRRC37B	leucine rich repeat containing 37B	1.53
LRRC48	leucine rich repeat containing 48	1.54
LY96	lymphocyte antigen 96	3.29
LYPLAL1	lysophospholipase-like 1	2.03
LYST	lysosomal trafficking regulator	1.77
MAGI2	membrane associated guanylate kinase, WW and PDZ domain containing 2	3.05
MARK1	MAP/microtubule affinity-regulating kinase 1	1.76
MBP	myelin basic protein	1.72
MCOLN1	mucolipin 1	1.56
MCOLN3	mucolipin 3	2.29
MCTP1	multiple C2 domains, transmembrane 1	1.78
MFSD1	major facilitator superfamily domain containing 1	1.63
MITD1	MIT, microtubule interacting and transport, domain containing 1	1.59
MLPH	melanophilin	1.84

MOSPD1	motile sperm domain containing 1	1.91
MRAS	muscle RAS oncogene homolog	2.35
MREG	melanoregulin	2.13
MRGPRX4	MAS-related GPR, member X4	3.18
MTHFR	methylenetetrahydrofolate reductase (NAD(P)H)	1.89
MXI1	MAX interactor 1	1.51
NAV2	neuron navigator 2	2.25
NCRNA00169	non-protein coding RNA 169	1.83
NEURL2	neuralized homolog 2 (Drosophila)	1.59
NEURL3	neuralized homolog 3 (Drosophila) pseudogene	2.01
NOTCH2NL	Notch homolog 2 (Drosophila) N-terminal like	1.51
NOV	nephroblastoma overexpressed gene	5.75
NPC1	Niemann-Pick disease, type C1	2.19
NRIP3	nuclear receptor interacting protein 3	1.55
NT5DC3	5'-nucleotidase domain containing 3	1.53
NUPR1	nuclear protein, transcriptional regulator, 1	2.18
OSBPL1A	oxysterol binding protein-like 1A	2.62
OXCT1	3-oxoacid CoA transferase 1	1.56
P2RX4	purinergic receptor P2X, ligand-gated ion channel, 4	2.13
P2RX7	purinergic receptor P2X, ligand-gated ion channel, 7	2.77
PBLD	phenazine biosynthesis-like protein domain containing	1.50
PDCD4	programmed cell death 4 (neoplastic transformation inhibitor)	1.78
PFKFB2	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	2.36
PFKFB4	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	3.64
PIK3C2B	phosphoinositide-3-kinase, class 2, beta polypeptide	1.52
PLA2G16	phospholipase A2, group XVI	1.69
PLA2G4C	phospholipase A2, group IVC (cytosolic, calcium-independent)	2.56
PLA2G6	phospholipase A2, group VI (cytosolic, calcium-independent)	1.61
PLEKHF1	pleckstrin homology domain containing, family F (with FYVE domain) 1	2.21
PLEKHF2	pleckstrin homology domain containing, family F (with FYVE domain) 2	1.57
PLEKHM1	pleckstrin homology domain containing, family M (with RUN domain) 1	2.42
PLEKHM1P	pleckstrin homology domain containing, family M (with RUN domain) 1 PsG	1.50
PLK3	polo-like kinase 3 (Drosophila)	2.46
PLLP	plasma membrane proteolipid (plasmolipin)	1.89
PLSCR1	phospholipid scramblase 1	1.58
PNLIPRP3	pancreatic lipase-related protein 3	11.79
PPARGC1A	peroxisome proliferator-activated receptor gamma, coactivator 1 alpha	9.04
PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	2.26
PPP1R3B	protein phosphatase 1, regulatory (inhibitor) subunit 3B	1.58
PPP1R3C	protein phosphatase 1, regulatory (inhibitor) subunit 3C	1.71
PQLC2	PQ loop repeat containing 2	1.68
PRDM4	PR domain containing 4	1.90
PRICKLE2	prickle homolog 2 (Drosophila)	1.52
PSEN2	presenilin 2 (Alzheimer disease 4)	1.80
PSG4	pregnancy specific beta-1-glycoprotein 4	2.97
PTGES	prostaglandin E synthase	2.11
QDPR	quinoid dihydropteridine reductase	1.70
RAB3A	RAB3A, member RAS oncogene family	1.72
RAB9A	RAB9A, member RAS oncogene family	1.53
RAD51L1	RAD51-like 1 ( <i>S. cerevisiae</i> )	1.51
RAD9A	RAD9 homolog A ( <i>S. pombe</i> )	1.85
RANBP10	RAN binding protein 10	2.07
RASGRP3	RAS guanyl releasing protein 3 (calcium and DAG-regulated)	2.93

RASSF3	Ras association (RalGDS/AF-6) domain family member 3	1.50
RCBTB1	RCC1 and BTB (POZ) domain containing protein 1	1.64
RCL1	RNA terminal phosphate cyclase-like 1	1.71
RENBP	renin binding protein	2.19
RGPD5	RANBP2-like and GRIP domain containing 5	1.79
RIMKLB	ribosomal modification protein rimK-like family member B	1.62
RNF144A	ring finger protein 144A	1.61
RPS6KA5	ribosomal protein S6 kinase, 90kDa, polypeptide 5	1.98
RRAGD	Ras-related GTP binding D	4.39
RTN4RL2	reticulon 4 receptor-like 2	2.21
RUFY3	RUN and FYVE domain containing 3	2.09
RUNDC2C	RUN domain containing 2C	1.99
SCD	stearoyl-CoA desaturase (delta-9-desaturase)	1.66
SDSL	serine dehydratase-like	1.56
SEMA6A	semaphorin 6A	1.85
SEPHS2	selenophosphate synthetase 2	1.64
SEPT3	septin 3	1.55
SERINC5	serine incorporator 5	2.53
SETD4	SET domain containing 4	1.76
SETDB2	SET domain, bifurcated 2	6.95
SGSH	N-sulfoglucosamine sulfohydrolase	1.57
SH3BP5	SH3-domain binding protein 5 (BTK-associated)	1.64
SIK1	salt-inducible kinase 1	1.54
SIRPA	signal-regulatory protein alpha	2.22
SLAMF7	SLAM family member 7	5.82
SLC16A6	solute carrier family 16, member 6 (monocarboxylic acid transporter 7)	4.10
SLC19A2	solute carrier family 19 (thiamine transporter), member 2	2.54
SLC25A25	solute carrier family 25 (mitochondrial carrier; phosphate carrier), 25	2.26
SLC25A26	solute carrier family 25, member 26	1.68
SLC26A11	solute carrier family 26, member 11	1.51
SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3	1.50
SLC36A1	solute carrier family 36 (proton/amino acid symporter), member 1	1.84
SLC38A6	solute carrier family 38, member 6	1.94
SLC38A7	solute carrier family 38, member 7	2.01
SLC43A1	solute carrier family 43, member 1	1.67
SLC9A9	solute carrier family 9 (sodium/hydrogen exchanger), member 9	2.20
SLCO4A1	solute carrier organic anion transporter family, member 4A1	2.38
SLFN5	schlafen family member 5	1.79
SNX10	sorting nexin 10	1.51
SNX25	sorting nexin 25	1.61
SNX8	sorting nexin 8	1.81
SOAT1	sterol O-acyltransferase 1	1.74
SORD	sorbitol dehydrogenase	1.57
SPHK1	sphingosine kinase 1	2.21
SPINK1	serine peptidase inhibitor, Kazal type 1	3.20
SPIRE1	spire homolog 1 (Drosophila)	1.76
SPSB1	splA/ryanodine receptor domain and SOCS box containing 1	2.49
ST3GAL6	ST3 beta-galactoside alpha-2,3-sialyltransferase 6	2.08
STAG3L4	stromal antigen 3-like 4	1.55
STARD10	StAR-related lipid transfer (START) domain containing 10	1.59
STK16	serine/threonine kinase 16	1.52
STK17B	serine/threonine kinase 17b	1.80
STX7	syntaxin 7	2.30

SULT1C2	sulfotransferase family, cytosolic, 1C, member 2	2.20
SYNC	syncoilin, intermediate filament protein	1.98
SYT14	synaptotagmin XIV	1.80
TAF4B	TAF4b RNA polymerase II, TBP-associated factor, 105kDa	2.26
TAP2	transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)	1.72
TBC1D14	TBC1 domain family, member 14	2.34
TBC1D16	TBC1 domain family, member 16	1.61
TBC1D7	TBC1 domain family, member 7	5.48
TM4SF19	transmembrane 4 L six family member 19	9.91
TMEM117	transmembrane protein 117	2.58
TMEM144	transmembrane protein 144	1.59
TMEM150A	transmembrane protein 150A	1.63
TMEM229B	transmembrane protein 229B	1.54
TMEM232	transmembrane protein 232	2.40
TMEM38B	transmembrane protein 38B	2.71
TMEM55A	transmembrane protein 55A	1.53
TMEM86A	transmembrane protein 86A	1.67
TNFRSF14	tumor necrosis factor receptor superfamily, member 14	4.79
TOM1	target of myb1 (chicken)	1.62
TOP3B	topoisomerase (DNA) III beta	1.91
TP53INP1	tumor protein p53 inducible nuclear protein 1	1.76
TPCN1	two pore segment channel 1	2.23
TPI1P2	triosephosphate isomerase 1 pseudogene 2	1.68
TPP1	tripeptidyl peptidase I	1.66
TPRA1	transmembrane protein, adipocyte associated 1	1.76
TRAF5	TNF receptor-associated factor 5	2.39
TRAK1	trafficking protein, kinesin binding 1	1.62
TRAK2	trafficking protein, kinesin binding 2	1.69
TRAPPC2L	trafficking protein particle complex 2-like	2.07
TRIB3	tribbles homolog 3 (Drosophila)	1.96
TRIM63	tripartite motif-containing 63	1.72
TSPAN10	tetraspanin 10	1.76
TTYH2	tweety homolog 2 (Drosophila)	1.63
UAP1L1	UDP-N-acetylglucosamine pyrophosphorylase 1-like 1	2.48
UCK1	uridine-cytidine kinase 1	1.56
UCN2	urocortin 2	5.25
UPP1	uridine phosphorylase 1	2.32
USP2	ubiquitin specific peptidase 2	1.65
VAC14	Vac14 homolog (S. cerevisiae)	1.85
VAT1	vesicle amine transport protein 1 homolog (T. californica)	1.66
VGf	VGf nerve growth factor inducible	2.68
VPS11	vacuolar protein sorting 11 homolog (S. cerevisiae)	1.61
VPS18	vacuolar protein sorting 18 homolog (S. cerevisiae)	1.93
WDFY1	WD repeat and FYVE domain containing 1	1.55
WDR81	WD repeat domain 81	2.51
WDSUB1	WD repeat, sterile alpha motif and U-box domain containing 1	1.91
WIPI1	WD repeat domain, phosphoinositide interacting 1	2.07
XYLB	xylulokinase homolog (H. influenzae)	1.57
YOD1	YOD1 OTU deubiquinating enzyme 1 homolog (S. cerevisiae)	1.50
ZBTB1	zinc finger and BTB domain containing 1	1.55
ZBTB25	zinc finger and BTB domain containing 25	1.90
ZC3H6	zinc finger CCCH-type containing 6	1.66
ZCCHC14	zinc finger, CCHC domain containing 14	2.42

ZFYVE16	zinc finger, FYVE domain containing 16	2.54
ZFYVE26	zinc finger, FYVE domain containing 26	1.53
ZNF330	zinc finger protein 330	1.90
ZNF438	zinc finger protein 438	1.70
ZNF57	zinc finger protein 57	1.90
ZNF610	zinc finger protein 610	1.91
ZNF616	zinc finger protein 616	1.93

Down-regulated genes (p-value < 0.01; Fold Change < -1.5)

Gene Name	Gene Description	Fold Change
ABCC3	ATP-binding cassette, sub-family C (CFTR/MRP), member 3	-1.53
ACPL2	acid phosphatase-like 2	-1.57
ADAMTS15	ADAM metalloproteinase with thrombospondin type 1 motif, 15	-1.54
ADAMTS9	ADAM metalloproteinase with thrombospondin type 1 motif, 9	-1.79
AMIGO2	adhesion molecule with Ig-like domain 2	-1.80
AMOTL2	angiominin like 2	-1.52
ARSJ	arylsulfatase family, member J	-1.85
ATF4	activating transcription factor 4 (tax-responsive enhancer element B67)	-1.65
B3GALT2	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 2	-1.56
BIRC3	baculoviral IAP repeat-containing 3	-1.82
BMP2	bone morphogenetic protein 2	-1.90
C10orf10	chromosome 10 open reading frame 10	-1.63
C10orf81	chromosome 10 open reading frame 81	-2.15
C1QTNF1	C1q and tumor necrosis factor related protein 1	-1.69
C1S	complement component 1, s subcomponent	-1.82
C21orf70	chromosome 21 open reading frame 70	-1.98
C3	complement component 3	-3.26
C5orf13	chromosome 5 open reading frame 13	-1.70
C6orf138	chromosome 6 open reading frame 138	-1.88
CCL5	chemokine (C-C motif) ligand 5	-1.76
CCNE2	cyclin E2	-1.65
CD96	CD96 molecule	-1.58
CDH11	cadherin 11, type 2, OB-cadherin (osteoblast)	-2.22
CDK1	cyclin-dependent kinase 1	-1.57
COL11A1	collagen, type XI, alpha 1	-1.56
COL11A2	collagen, type XI, alpha 2	-1.53
COL12A1	collagen, type XII, alpha 1	-1.58
CPA4	carboxypeptidase A4	-2.11
CTGF	connective tissue growth factor	-2.33
CXCL1	chemokine (C-X-C motif) lig 1 (melanoma growth stimulating activity, $\alpha$ )	-1.73
CYR61	cysteine-rich, angiogenic inducer, 61	-2.33
DEPDC7	DEP domain containing 7	-1.57
DGCR14	DiGeorge syndrome critical region gene 14	-1.51
DKK1	dickkopf homolog 1 (Xenopus laevis)	-1.86
DOCK4	dedicator of cytokinesis 4	-1.54
DSEL	dermatan sulfate epimerase-like	-1.90
EDIL3	EGF-like repeats and discoidin I-like domains 3	-1.59
EGR2	early growth response 2	-1.88
EHF	ets homologous factor	-1.67
ENC1	ectodermal-neural cortex 1 (with BTB-like domain)	-1.62
ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	-1.67
FAM111B	family with sequence similarity 111, member B	-1.63
FBN1	fibrillin 1	-1.68
FHDC1	FH2 domain containing 1	-1.74
FLT1	fms-related tyrosine kinase 1	-1.51
FREM2	FRAS1 related extracellular matrix protein 2	-1.56
FST	follistatin	-1.52
GBP1	guanylate binding protein 1, interferon-inducible, 67kDa	-1.87
GCM1	glial cells missing homolog 1 (Drosophila)	-1.61
GCNT1	glucosaminyl (N-acetyl) transferase 1, core 2	-1.60



GFPT2	glutamine-fructose-6-phosphate transaminase 2	-1.91
GJA5	gap junction protein, alpha 5, 40kDa	-1.77
GOLGA8E	golgin A8 family, member E	-1.53
GPR84	G protein-coupled receptor 84	-1.91
HAS2	hyaluronan synthase 2	-1.62
HOXA2	homeobox A2	-2.93
HSPG2	heparan sulfate proteoglycan 2	-1.68
ICAM4	intercellular adhesion molecule 4 (Landsteiner-Wiener blood group)	-3.15
IFIT2	interferon-induced protein with tetratricopeptide repeats 2	-1.57
IGFBP5	insulin-like growth factor binding protein 5	-1.94
IL1A	interleukin 1, alpha	-2.46
IL1B	interleukin 1, beta	-3.99
IL32	interleukin 32	-1.57
IL6	interleukin 6 (interferon, beta 2)	-1.92
IL7R	interleukin 7 receptor	-2.90
IL8	interleukin 8	-1.91
IRAK2	interleukin-1 receptor-associated kinase 2	-1.52
ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	-1.53
KCNN3	potassium intermediate/small conductance calcium-activated channel, N3	-1.82
KCNS3	potassium voltage-gated channel, delayed-rectifier, subfamily S, member 3	-1.52
KLF4	Kruppel-like factor 4 (gut)	-1.58
KYNU	kynureninase (L-kynurenine hydrolase)	-1.63
LOC100289612	arsenic transactivated protein 1	-1.64
MAMLD1	mastermind-like domain containing 1	-1.53
MAP3K5	mitogen-activated protein kinase kinase kinase 5	-1.51
MIR21	microRNA 21	-1.82
MIRLET7A2	microRNA let-7a-2	-1.80
MIRLET7I	microRNA let-7i	-1.61
MYADM	myeloid-associated differentiation marker	-1.56
NAV3	neuron navigator 3	-1.95
NEDD9	neural precursor cell expressed, developmentally down-regulated 9	-1.54
NIPAL1	NIPA-like domain containing 1	-1.50
NR2F1	nuclear receptor subfamily 2, group F, member 1	-1.52
NR4A1	nuclear receptor subfamily 4, group A, member 1	-1.66
NR4A2	nuclear receptor subfamily 4, group A, member 2	-1.56
NRG1	neuregulin 1	-1.53
NRXN3	neurexin 3	-1.80
NUAK2	NUAK family, SNF1-like kinase, 2	-2.49
OASL	2'-5'-oligoadenylate synthetase-like	-1.52
PAPPA	pregnancy-associated plasma protein A, pappalysin 1	-2.30
PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	-1.67
PCDH18	protocadherin 18	-1.84
PCDHB13	protocadherin beta 13	-1.67
PCDHB14	protocadherin beta 14	-1.75
PCDHB16	protocadherin beta 16	-1.64
PCDHB2	protocadherin beta 2	-1.56
PCDHB3	protocadherin beta 3	-1.51
PCDHB5	protocadherin beta 5	-1.59
PCDHB8	protocadherin beta 8	-1.67
PDE10A	phosphodiesterase 10A	-1.53
PDE1C	phosphodiesterase 1C, calmodulin-dependent 70kDa	-1.52
PDXDC2	pyridoxal-dependent decarboxylase domain containing 2	-1.75
PHLDB2	pleckstrin homology-like domain, family B, member 2	-1.56

PII5	peptidase inhibitor 15	-1.53
PLCB1	phospholipase C, beta 1 (phosphoinositide-specific)	-1.54
PMEPA1	prostate transmembrane protein, androgen induced 1	-1.60
PRDM1	PR domain containing 1, with ZNF domain	-1.90
PTPN22	protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	-1.75
PTPRZ1	protein tyrosine phosphatase, receptor-type, Z polypeptide 1	-1.71
PTX3	pentraxin 3, long	-2.47
PXDN	peroxidasin homolog (Drosophila)	-1.67
RALGPS2	Ral GEF with PH domain and SH3 binding motif 2	-1.65
RAPH1	Ras association (RalGDS/AF-6) and pleckstrin homology domains 1	-1.61
RCVRN	recoverin	-3.34
RNF144B	ring finger protein 144B	-1.53
RNU11	RNA, U11 small nuclear	-1.52
RNU1-1	RNA, U1 small nuclear 1	-1.70
RNU5D	RNA, U5D small nuclear	-2.11
S100A7	S100 calcium binding protein A7	-1.55
SAMD9	sterile alpha motif domain containing 9	-1.53
SDK1	sidekick homolog 1, cell adhesion molecule (chicken)	-1.71
SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	-1.51
SH3PXD2A	SH3 and PX domains 2A	-1.70
SHC4	SHC (Src homology 2 domain containing) family, member 4	-1.52
SLC14A1	solute carrier family 14 (urea transporter), member 1 (Kidd blood group)	-1.53
SLC5A3	solute carrier family 5 (sodium/myo-inositol cotransporter), member 3	-1.76
SLFN11	schlafen family member 11	-1.58
SLIT2	slit homolog 2 (Drosophila)	-1.96
SLITRK6	SLIT and NTRK-like family, member 6	-2.11
SNORA68	small nucleolar RNA, H/ACA box 68	-1.70
SNORD12C	small nucleolar RNA, C/D box 12C	-1.83
SNORD14E	small nucleolar RNA, C/D box 14E	-1.52
SNORD15B	small nucleolar RNA, C/D box 15B	-1.59
SNORD22	small nucleolar RNA, C/D box 22	-1.56
SNORD26	small nucleolar RNA, C/D box 26	-1.62
SNORD28	small nucleolar RNA, C/D box 28	-1.53
SNORD36B	small nucleolar RNA, C/D box 36B	-1.77
SNORD36C	small nucleolar RNA, C/D box 36C	-1.71
SNORD3A	small nucleolar RNA, C/D box 3A	-1.52
SNORD5	small nucleolar RNA, C/D box 5	-1.56
SNORD76	small nucleolar RNA, C/D box 76	-1.50
SNORD78	small nucleolar RNA, C/D box 78	-1.52
SOCS3	suppressor of cytokine signaling 3	-1.58
SOX2	SRY (sex determining region Y)-box 2	-1.54
SPATA5L1	spermatogenesis associated 5-like 1	-1.52
ST3GAL1	ST3 beta-galactoside alpha-2,3-sialyltransferase 1	-1.55
ST8SIA5	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 5	-1.54
STARD13	StAR-related lipid transfer (START) domain containing 13	-1.75
TFPI2	tissue factor pathway inhibitor 2	-1.51
TGFA	transforming growth factor, alpha	-1.59
TLR4	toll-like receptor 4	-1.57
TMEM171	transmembrane protein 171	-1.63
TMEM2	transmembrane protein 2	-1.60
TMOD1	tropomodulin 1	-1.52
TNF	tumor necrosis factor	-2.20
TNFAIP2	tumor necrosis factor, alpha-induced protein 2	-2.21

TNFAIP6	tumor necrosis factor, alpha-induced protein 6	-3.80
TNFRSF9	tumor necrosis factor receptor superfamily, member 9	-1.71
TNIK	TRAF2 and NCK interacting kinase	-1.57
TRIB1	tribbles homolog 1 (Drosophila)	-1.55
TUBB2B	tubulin, beta 2B	-1.52
UBE2C	ubiquitin-conjugating enzyme E2C	-1.59
UNC5C	unc-5 homolog C (C. elegans)	-1.74
VEGFC	vascular endothelial growth factor C	-1.79
WEE1	WEE1 homolog (S. pombe)	-1.72
WNT5A	wingless-type MMTV integration site family, member 5A	-2.03
ZFP36	zinc finger protein 36, C3H type, homolog (mouse)	-1.58

## Appendix C – General bacterial expression and purification protocols by PSF

Jeremy Keusch P.46 PSF

10/22/2014

Page 1 of 3

### Small-Scale Expression Screen in *E. coli* in DWB

#### Transformation of Expression Competent *E. coli*

A 20  $\mu$ l aliquot of expression competent *E. coli* eg Rosetta™2 (DE3) cells (Novagen, 71397-4) is removed from the -80°C and put on ice for 5 min.

1  $\mu$ l of cDNA (1-10 ng) is gently added to the cells without pipetting up and down.

Gently flick the tube and incubate on ice for 30 min.

Heat-shock the cells for exactly 30 seconds in a 42°C water bath, without shaking.

Put the tube back on ice for 2 min.

Add 80  $\mu$ l of RT SOC medium.

Incubate the tube at 37°C shaking at 225 rpm for 1 hour

Plate the cells onto warm and dry LB Agar plates supplemented with appropriate antibiotics eg. 50  $\mu$ g/ml carbenicillin, and 34  $\mu$ g/ml chloramphenicol. Better if also supplemented with 1% glucose (special request to media kitchen).

Incubate o/night at 37°C

For the expression screen you need three consecutive working days.

#### Day 1

Prepare 96-well DWB (AbGene AB-0932) by the addition of 1.5 ml GS96(Q-Biogene/ MP Biomedicals Cat# 3101-132)/ 0.05% glycerol/ 1% glucose supplemented with appropriate antibiotics (eg 50  $\mu$ g/ml carbenicillin and 34  $\mu$ g/ml chloramphenicol). Using 200  $\mu$ l pipette tips pick individual colonies (numbered on plate) into each well. When plate is complete remove tips and seal the DWB with gas-permeable adhesive seal (AbGene AB-0718). Incubate these starter cultures o/night at 37°C with shaking at 225 rpm in the Kühner DWB holders.

#### Day 2

Prepare 24-well DWB by the addition of 2.5 ml appropriate medium\* (IPTG or AI\*\*)

IPTG cultures:

- A. GS96 / 0.05% glycerol/ 1% glucose supplemented with appropriate antibiotics (eg 50  $\mu$ g/ml carbenicillin and 34  $\mu$ g/ml chloramphenicol)

AI cultures:

- B. \*\*Autoinduction medium (see **Note** on page 2) supplemented with appropriate antibiotics (eg 50  $\mu$ g/ml carbenicillin and 34  $\mu$ g/ml chloramphenicol)

Dilute the overnight starter cultures (from the 96 DWB) 1:50 by transferring 50  $\mu$ l of overnight culture into the prepared 24-well DWB.

For "A" IPTG

Shake at 225 rpm, 37°C until "A" has reached an OD<sub>600</sub>~0.6.

Cool the culture by shaking at 20°C for 30 minutes.

Induce the culture by the addition of IPTG to a final conc. of 0.5 mM (12.5  $\mu$ l of 100mM IPTG)

Grow the cultures o/night (~20 hours) by shaking at 225 rpm at 20°C

For "B" Autoinduction

Shake at 225 rpm, 37°C for ~4-5 hours (cloudy) before reducing the temperature to 20°C.

Grow the cultures for a further 20-24 h by shaking at 225 rpm at 20°C.

#### Day 3

Re-array the cultures from the 24-well DWB into a new 96-well DWB maintaining the plate matrix/layout. Transfer 1.2 ml of each culture, using one channel of the 1.2ml LTS pipette, into the 96-well DWB and pellet the cells by centrifugation at 6000 x g at 4°C for 30 minutes using the Beckman JS5.3 rotor on the 2<sup>nd</sup> floor.

Carefully decant the media from the cell pellets and allow the plates to drain for a minute on some paper towels.

Freeze cell pellets on dry ice and leave for at least 30 minutes or store at -80°C.

**Note**

\*The recommended medium for the mini and large-scale cultures is GS96 (+glycerol/1%glucose and appropriate antibiotics). It also works with LB but the growth of the cultures is really variable, especially following the 1:50 dilution from the o/night starter cultures.

\*\*Autoinduction medium

See attached protocol (ZYP-5052 rich medium for auto-induction)

### Magnetic bead Ni<sup>2+</sup>-NTA screen protocol

Based on Qiagen's "Ni-NTA Magnetic Agarose Beads Handbook"

#### Day 3 Contd.

1. Re-suspend the cells in 280 ul of buffer NPI-10-Tween supplemented with fresh 1mg/ml lysozyme and 3 units/ml Benzonase (Sigma E-1014). (Note, the lysis volume used to be 230ul but this is insufficient if cell pellet is big).

Place the DWB in an orbital shaker. Secure the plate with padding and lid and shake at 1400 rpm at 4°C for 30 min.

Clear the lysates by centrifugation at 6000 x g at 4°C for 30 min.

2. A few minutes before the end of the centrifugation, dispense 20 ul of the Ni-NTA bead suspension (Qiagen 36111) per well of a flat-bottomed 96-well plate (MTP, Qiagen 36985). Note: re-suspend the Ni-NTA magnetic beads by briefly vortexing the beads and immediately adding to the wells. Use pipette tips with wide openings.

3. Transfer the supernatant from step 2, taking care not to disturb the 'insoluble' pellet, to each well of the MTP containing the magnetic Ni-NTA beads. Transfer can be done using the Liquidator set at 180 ul and appropriate height adjustors.

Mix for 30 min at RT at 800 rpm using a vortex adapted for MTP.

\*\*\*Pellets may be re-suspended in 8 M urea buffered with 100 mM NaH<sub>2</sub>PO<sub>4</sub> and 10mM Tris, pH 8.0 for analysis of the 'insoluble' fraction on SDS-PAGE. Shake at 1400 rpm at 4°C for 30 min.

4. Place the MTP on the 96-well magnet (Qiagen Type-A) for 2 min and carefully remove the supernatant from the beads using the Liquidator and appropriate height adjustors.

5. Add 200 ul wash buffer to each well of MTP, remove from the magnet and mix on the vortex shaker for 5 min.

6. Place MTP on the 96-well magnet for 2 min and remove the buffer.

7. Repeat steps 5 and 6.

8. Add 25 ul of elution buffer to each well of MTP and mix on the vortex shaker for 2 min. Place on the magnet for 2 min and transfer the eluate to clean tubes for SDS-PAGE analysis.

15 ul soluble eluate mixed with 5 ul 4x sample buffer with DTT, load 15 ul per lane (29% of total)

\*\*\*Cell pellets re-suspended in 200 ul 8 M urea buffer. 18 ul mixed with 6 ul 4x sample buffer with DTT, load 5 ul per lane (1.9% of total)

17 lane 4-12% Bis-Tris SDS-PAGE. Run gel at 200V in MES for 35 min. Stain with InstantBlue (Expedeon via Lucerna).

#### Buffers used:

##### Buffer NPI-10-Tween:

50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, 1% v/v Tween 20, adjust pH to 8.0 using NaOH and filter before use. Store at 4°C.

##### Buffer NPI-20-Tween (Wash buffer):

50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, 0.05% v/v Tween 20, adjust pH to 8.0 using NaOH and filter before use. Store at 4°C.

##### Buffer NPI-250-Tween, 50 ml: (elution buffer):

50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM Imidazole, 0.05% Tween, adjust pH to 8.0 using NaOH and filter before use. Store at 4°C.

## **1. Large scale expression of soluble targets**

### **a. Native Expression with IPTG induction**

1. A single colony is inoculated into 20 ml of GS96 with 0.05% glycerol and 1% glucose and the appropriate antibiotics in a 250 ml flask. The culture is incubated at 37°C 225 rpm overnight
2. The overnight culture is diluted 1 in 100 into 1L of GS96 with 0.05% glycerol and 1% glucose plus the appropriate antibiotics. The 500 ml cultures in 2L shake-flasks are incubated at 37°C 225 rpm until an OD of ~0.6 is reached.
3. The temperature is reduced to 20°C. After 30min at 20°C, the cultures are induced with 500 µl of 0.5M IPTG and are incubated for a further 20h at 20°C, 225 rpm.
4. Harvest cells 6500 x g at 4°C for 30 min. Weigh pellets. Re-suspend pellet in 30 ml lysis buffer, freeze on dry-ice and store at -80° C. (see below)

### **b. Native Expression with Auto-Induction**

1. A single colony is inoculated into 20 ml of GS96 with glucose to 1% (w/v) and the appropriate antibiotics in a 50 ml falcon. The culture is incubated at 37°C 225 rpm overnight
2. The overnight culture is diluted 1 in 100 into 1L of auto-induction medium (see below). This is a home made version for expression in *E. coli*. It is based on the Overnight Express™ Auto-induction Technology from Novagen <http://www.emdbiosciences.com/html/NVG/onex.html>  
The development of the technique is described in FW. Studier, Protein Expression and Purification, 41, 2005, 207

#### **Stock solutions for Auto-induction medium**

Prepare stock solutions, filter for sterilization, store at 4°C.

#### 20 x NPS (500ml)

450 ml milli-q H<sub>2</sub>O  
33 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>  
68 g KH<sub>2</sub>PO<sub>4</sub>  
71 g Na<sub>2</sub>HPO<sub>4</sub>  
qsp to 500 ml with milli-q H<sub>2</sub>O

#### 50 x 5052 (500 ml) (5052 = 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose)

125 g glycerol  
365 ml milli-q H<sub>2</sub>O  
12.5 g glucose  
50 g α-lactose  
qsp to 500 ml with milli-q H<sub>2</sub>O

#### 1 M MgSO<sub>4</sub>

24.65 g MgSO<sub>4</sub>·7H<sub>2</sub>O  
Water to make 100 ml

#### **Medium for cell culture**

ZY (Sufficient for 500ml expression culture see ZYP5052 below)

5 g tryptone  
2.5g yeast extract  
463 ml water. (Do NOT adjust volume to 500ml).  
Autoclave

#### **ZYP-5052 rich medium for auto-induction** (Prepare just before use; 500 ml)

Component	500 ml
ZY	463 ml
1 M MgSO <sub>4</sub>	1ml
50 x 5052	10 ml
20 x NPS	25 ml

Protocols are derived from the Oxford Protein Production Facility (OPPF), Oxford, UK and BM14 (EMBL/ESRF), Grenoble, France.

- Add from stocks to autoclaved medium (500 ml per 2.5L flask).
- Add antibiotics, as needed (Kanamycin 50 µg/ml, chloramphenicol 34 µg/ml, carbenicillin 50 µg/ml)
- Inoculate from preculture
- Grow cells 4h, 37°C, 225 rpm
- After 4h reduce temperature to 20°C and incubate for another 20h
- Remove 1ml for small-scale expression test
- Harvest cells 30 min, 6500g, 4°C
- Weigh the pellets
- Resuspend cell pellet in 30 ml Lysis buffer (see below), freeze on dry-ice and store at -80° C.

## **2. Large scale purification of soluble targets**

### **a. Lysis**

1. The cell pellets are thawed for ~15 min at room temperature, this is supplemented with protease inhibitor tablets (e.g. Complete EDTA free from Roche Diagnostics) and 3 units/ml Benzonase (Sigma). A minimum volume of 100ml is needed for the cell disruptor. Work at 4°C from now on.
2. The sample is passed through the cell disruptor in the Thomae lab (P.52). A 20 µl sample may be taken for SDS-PAGE analysis.
3. The lysate is centrifuged in the JA-17 rotor at 30,000g for 30min, 4°C.
4. The cleared lysate is decanted away from the insoluble fraction and placed on ice.
5. The cleared lysate is filtered through a 0.45 µm minisart plus filter (Sartorius Stedim #17829) and a 20 µl sample may be taken for SDS-PAGE analysis.
6. The filtered, cleared lysate is loaded onto the Äkta Purifier or purified in batch mode with Ni-NTA agarose (Qiagen, Cat#30210; #1018240).

### **b. Nickel affinity purification (IMAC) and Gel filtration on the Äkta Purifier**

1. The 1 ml HisTrap columns are pre-charged with nickel (re-charging is required if the column has been re-generated) and equilibrated with 5 ml of Wash buffer (50 mM Tris pH 7.5, 500 mM NaCl, 20 mM Imidazole).
2. The Gel filtration columns are equilibrated prior to the Äkta run with ~140 ml of Gel filtration buffer.
3. For targets below ~15KDa a 16/60 HiLoad Superdex 75 is recommended, for larger targets a 16/60 HiLoad Superdex 200 may be used.

Buffers for IMAC and Gel Filtration made up with milli-Q H<sub>2</sub>O, 0.2 µM sterile-filtered and degassed (just before use):

Lysis buffer, 50 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole 0.2% Tween (Add 2 ml 10% Tween 20 to 100 ml of Wash buffer).

Nickel Wash buffer, 50 mM Tris pH7.5, 500 mM NaCl, 20 mM imidazole (For 1L, 50 ml 1M Tris pH 7.5; 100 ml 5M NaCl; 10 ml 2M Imidazole)

Nickel Elution buffer, 50 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole ( For 1L, 50 ml 1M Tris pH 7.5; 100 ml 5M NaCl; 250 ml 2M Imidazole)

Gel filtration buffer, 20 mM Tris pH 7.5, 200 mM NaCl (For 1L, 20 ml 1M Tris pH7.5; 40 ml 5M NaCl). 5 mM TCEP should be added when appropriate.

Protocols are derived from the Oxford Protein Production Facility (OPPF), Oxford, UK and BM14 (EMBL/ESRF), Grenoble, France.



## Appendix D – General limited proteolysis protocol by PAF

### Limited proteolysis for determination of stable domains

**Protease:** Weigh the subtilisin powder  
Dissolve it in 20mM Tris pH 8.0 + 0.2M NaCl  
Measure protein concentration by Bradford or NanoDrop  
Do dilutions...

Can also be done with trypsin, chymotrypsin in appropriate buffer

**Protocol:** Mix 10µg of your protein with ≠ [subtilysin]  
Incubate 5-10 minutes on ice  
Stop reaction by adding 1µl PMSF  
Add 5µl of 5x βMe-SDS-loading buffer and boil or freeze immediately  
If sending for mass spec, you can add GuHCl (add until no more dissolve)  
Load everything on a SDS-PAGE gel

10µg protein + ≠ % of subtilisin	in 20µl volume	
0.001%	0.1ng	10µl @ 0.01ng/µl (µg/ml)
0.003%	0.3ng	10µl @ 0.03ng/µl (µg/ml)
0.01%	1ng	10µl @ 0.1ng/µl (µg/ml)
0.03%	3ng	10µl @ 0.3ng/µl (µg/ml)
0.1%	10ng	10µl @ 1ng/µl (µg/ml)
0.3%	30ng	10µl @ 3ng/µl (µg/ml)
1%	100ng	10µl @ 10ng/µl (µg/ml)
3%	300ng	10µl @ 30ng/µl (µg/ml)

## Appendix E – Curriculum Vitae

**Name:** Jonas Schöler **Date of Birth:** July 6<sup>th</sup>, 1984  
**Civil Status:** Single, 1 child **Place of Birth:** Heidelberg, Germany  
**Nationality:** German

### Work Experience:

- 2009-current **Mechanisms of Cancer Group of Professor Ruth Chiquet-Ehrismann  
Friedrich Miescher Institute for Biomedical Research (FMI), affiliated with  
Novartis Institutes for BioMedical Research (NIBR), Basel, Switzerland**  
PhD, December 2009 – December 2014  
Topic: Characterizing the Teneurin 1 intracellular domain (human), by combining state-of-the-art screening methods with more in-depth biochemical approaches
- Cancer and Neurobiology research at a renowned institute
  - Wide variety of Molecular Biology techniques used, like Yeast-2 Hybrid screen, quantitative real-time PCR, Microarrays, and Protein Purification
  - Managing several projects and collaborations simultaneously
- 2009 **Cellular Machines Group of Professor Daniel Müller  
Biotechnology Center (BIOTEC), TU Dresden, Dresden, Germany**  
Master thesis, January 2009 – November 2009  
Topic: Using AFM Single Molecule Force Spectroscopy to study the protein Tau
- Neurobiology (e.g. Alzheimer disease) research
  - Using the cutting-edge technology AFM Single Molecule Force Spectroscopy
  - Successful collaboration with the Mandelkow laboratory of the Max-Planck Institute of Structural Biology in Hamburg

### Education:

- 2009-current **International PhD Program  
Friedrich Miescher Institute for Biomedical Research (FMI), affiliated with  
University of Basel, Basel, Switzerland**  
PhD in Cell Biology combines research and seminars at the FMI, with coursework at the University of Basel. **Expected Graduation: December 2014.**
- 2007-09 **Master of Science in Molecular Bioengineering  
Biotechnology Center (BIOTEC), TU Dresden, Dresden, Germany**  
The Molecular Bioengineering M.Sc. focuses on Biomedicine and Bio-nanotechnology.
- 2002-07 **Bachelor of Science in Biotechnology  
Minors in Economics and Microbiology  
The Pennsylvania State University, University Park, PA, USA**  
The Biotechnology B.Sc. teaches a strong foundation in the sciences and focuses on techniques used in Biotechnology and Pharmaceutical Industries.
- 1999-2002 **Unionville High School, Kennett Square, PA, USA**

## Courses:

### **Pharmaceuticals – from Research to Market**, Quality/GMP Course by **ETH Zürich**.

- Quality Assurance, Quality Management and GMP overview
- Discussion of rules and regulations in Switzerland, EU and USA

### **AIM4 Patients** course, Drug Discovery and Development Overview by **NIBR**.

- Overview of the Drug Discovery Process at NIBR
- Insight into the different phases of Clinical Trials

### **D3** course – Drug Discovery and Development Project Simulation by **NIBR**.

- Computer simulation from Drug Discovery until Market Release of a Drug
- Project team work with different expertises to reach a common goal

## Responsibility:

Student representative of the FMI PhD program, 2009-2010.

Student representative of the Master Course and on the Examination Committee & the Scientific Board, BIOTEC, 2007-2009.

## Activities:

Offensive Tackle in American Football, Dresden Monarchs (German Football League, 1<sup>st</sup> Bundesliga), 2007-08; Radebeul Suburban Foxes 2008-09 (Verbandsliga Ost); Basel Meanmachine 2010-2012 (Swiss National League B, then A).

Extracurricular activities are an important complement to the professional life. For me, this includes enjoying the outdoors, like Biking and Inline Skating, or relaxing with a good novel. Traveling and experiencing new cultures has also been a very important part of my life, though my priorities have shifted a bit since my daughter was born.

## Skills:

**Bilingual in German and English;** basic level of understanding in French and Spanish.

**Comprehensive Computer-Related Knowledge;** working with Microsoft, Macintosh, and Linux-based Operating Systems, Microsoft Office, Relevant Scientific Software like CLC Main Workbench, and handling Large Volumes of Data.

## References:

Prof. Dr. Ruth Chiquet-Ehrismann  
Senior Group Leader at the Friedrich-Miescher Institute for Biomedical Research  
+41 (0) 61 697 2494, [ruth.chiquet@fmi.ch](mailto:ruth.chiquet@fmi.ch)

Prof. Dr. Daniel Müller  
Head of Biophysics Group and Chair of Bionanotechnology at the Department of Biosystems Science and Engineering of the ETH Zürich  
+41 (0) 61 387 3307, [daniel.mueller@bsse.ethz.ch](mailto:daniel.mueller@bsse.ethz.ch)

Dr. Roberto Iacone  
Group Head at the NORD (Neurobiology, Ophthalmology, and Rare Diseases) department, Hoffmann-La Roche AG  
+41 (0) 61 687 1783, [roberto.iacone@roche.com](mailto:roberto.iacone@roche.com)

## Publications:

**Schöler, J.**, Ferralli, J., Thiry, S., Chiquet-Ehrismann, R. accepted. The intracellular domain of Teneurin-1 influences MITF-dependent transcription by HINT1 binding. *The Journal of Biological Chemistry*.

Tucker, R. P., Beckmann, J., Leachman, N. T., **Schöler, J.**, Chiquet-Ehrismann, R. (2012). Phylogenetic Analysis of the Teneurins: Conserved Features and Premetazoan Ancestry. *Molecular Biology and Evolution*, 29(3): 1019-1029.

Wegmann, S., **Schöler, J.**, Bippes, C. A., Mandelkow, E., Muller, D. J. (2011). Competing interactions stabilize pro- and anti-aggregant conformations of human tau. *The Journal of Biological Chemistry*, 10.1074.

## Poster presentations/ Conferences:

**Schöler, J.**, Beckmann, J., Chiquet, R. (Apr 2011). Genome-wide analysis of teneurin1 ICD. *PhD Meeting Barcelona*.

More posters presented at the FMI Annual Meetings 2010-2014, taking place at varying locations around Switzerland

Attending further seminars and conferences (e.g. 2<sup>nd</sup> International Congress on Stem Cells and Tissue Formation, July 2008; Federation of European Connective Tissue Societies (“FECTS”) meeting, July 2010; Federation of Biochemical Societies Congress (“FEBS”) “Mechanisms in Biology”, July 2013)