Determination and characterization of novel nuclear binding partners

of the neural cell adhesion molecule L1 and the HNK-1 glycan

DISSERTATION

Dissertation to achieve a doctoral degree

at the Faculty of

Mathematics, Informatics and Natural Sciences Department of Chemistry

of the Universität Hamburg

Gaston Pöllsner

Hamburg, 2021

Gutachter / Evaluators

Prof. Dr. Dr. h. c. Melitta Schachner

Prof. Dr. Andrew Torda

Day of oral defense

December 10th, 2021

I. Publications

Castillo G, Kleene R, Schachner M, Loers G, Torda AE. Proteins binding to the carbohydrate hnk-1: common origins? Int J Mol Sci. 2021;22(15):8116.

II.

Table of Contents

III. List of abbreviations

ab	antibody
Atf2	activating transcription factor 2
BCA	bicinchoninic acid
BSA	bovine serum albumin
CAF-1	chromatin assembly factor-1
CBX	chromobox homolog
CD	cluster of differentiation
cDNA	complementary deoxyribonucleic acid
CDS	coding sequence
CGC	cerebellar granule cell
Chek1	checkpoint kinase 1
CHL1	close homolog of L1 / cell adhesion molecule L1 like
CRASH	corpus callosum hypoplasia, retardation, adducted thumbs, spasticity, and hydrocephalus
CRISPR	clustered regularly interspaced short palindromic repeats
CSD	chromo shadow domain
DAPI	4′,6-diamidino-2-phenylindole
E. coli	Escherichia coli
ELISA	enzyme-linked immunosorbent assay
Fc	fragment crystallizable

FN3	fibronectin type-III
GEO	gene expression omnibus
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
HMGB	high mobility group box
HNK-1	human natural killer-1
HP1	heterochromatin protein 1
Ig	immunoglobulin
IPTG	isopropyl β-D-1-thiogalactopyranoside
kDa	kilodalton
KO	knockout
L1	cell adhesion molecule L1 (protein)
L1CAM	cell adhesion molecule L1 (gene)
L1- ICD	intracellular domain of L1
LIC	ligation independent cloning
MEME	multiple expectation maximizations for motif elicitation
MnM	Minimotif Miner
mRNA	messenger ribonucleic acid
nadl1	neural cell adhesion molecule L1
NFASC	neurofascin
Nipbl	nipped-B-like protein
NrCAM	neuronal cell adhesion molecule
PAGE	polyacrylamide gel electrophoresis
PCNA	proliferating-cell-nuclear-antigen

PCR	polymerase chain reaction
PLA	proximity ligation assay
pre-iPSCs	pre-induced pluripotent stem cells
PSI-BLAST	position-specific iterative basic local alignment search tool
qRT	Quantitative Reverse Transcription
Rad52	DNA repair protein Rad52 homolog
R-PTP	receptor-type tyrosine-protein phosphatase
SDS	sodium dodecyl sulfate
Sgo1	shugoshin 1
shRNA	short hairpin ribonucleic acid
siRNA	small interfering ribonucleic acid
SMART	simple modular architecture research tool
TIF1	transcriptional intermediary factor 1
Zmat3	zinc finger matrin-Type 3

1.   Zusammenfassung

Das neuronale Zelladhäsionsmolekül L1, auch bekannt als L1CAM, spielt eine wichtige Rolle bei der Entwicklung des Nervensystems und der ordnungsgemäßen Funktion des adulten Nervensystems. Das L1-Glykoprotein ist hauptsächlich in seiner Transmembranform bekannt, die eine wichtige Rolle bei Zell-Zell-Interaktionen spielt. Die Migration und Adhäsion von Neuronen sowie Synapsenbildung und Regulation der synaptischen Struktur hängt von deren Mikroumgebung ab. L1-Wechselwirkungen mit Glykanrezeptorproteinen wurden mit den zwei Kohlenhydratepitopen LewisX und HNK-1 in Verbindung gebracht. Darüberhinaus führen proteolytische Spaltungen des L1-Proteins zur Bildung von löslichen extrazellulären Fragmenten, Transmembranfragmenten sowie intrazellulären Fragmenten. Die löslichen intrazellulären Fragmente im Zytoplasma können in den Zellkern oder die Mitochondrien transportiert werden, wo noch nicht vollständig verstandene Wechselwirkungen und Signaltransduktionswege stattfinden. Meine Arbeit basierte darauf, mögliche Protein-Protein- und Glykan-Rezeptor-Wechselwirkungen innerhalb des Zellkerns zu finden. Ich habe biotechnische Techniken wie die Herstellung rekombinanter Proteine und Zellkulturen aus Kleinhirn-Körnerzellen der Maus sowie bioinformatische Techniken wie Sequenzvergleich und Sequenzkonservierung verwendet, um mögliche Wechselwirkungen innerhalb des Zellkerns zu finden. Ich fand heraus, dass die intrazelluläre Domäne von L1 mit Heterochromatin-Protein 1γ im Kern von Kleinhirn-Körnerzellen der Maus interagiert und dass die Interaktion eine Rolle bei der Regulierung von Genen spielen könnte, die an der Reaktion auf DNA-Schäden beteiligt sind. Außerdem konnte ich zeigen, dass die bekannten HNK-1-Rezeptoren kein gemeinsames Bindungsmotiv für das HNK-1 Kohlenhydrat aufweisen, sondern dass jede Rezeptorfamilie HNK-1 über einzigartige Sequenzen bindet.

2. Abstract

The neuronal cell adhesion molecule L1, also known as L1CAM, plays an important role in the development of the nervous system and the proper functioning of the adult nervous system. The L1 glycoprotein is mainly known and studied in its transmembrane form related to cell-cell interactions as neuron migration depends on the microenvironment of the neurons to guide their axons, form synapses, and regulate the synaptic structure. L1 interactions with glycan receptor proteins have been related to two main carbohydrate epitopes: LewisX and HNK-1. In the case of the mature transmembrane L1, proteolytic cleavages lead to the formation of soluble extracellular fragments, membrane-bound transmembrane fragments, and intracellular fragments. The soluble intracellular fragments in the cytoplasm can be transported to the nucleus or the mitochondria where interactions and pathways that are not yet fully understood take place. My work was based on finding possible protein-protein and glycan-receptor interactions inside of the nucleus. I used biotechnical techniques such as the production of recombinant proteins and cell culture from mouse cerebellar granule cells in addition to bioinformatics techniques such as sequence comparison and sequence conservation to try to find possible interactions happening inside of the nucleus. I found that the intracellular domain of L1 interacts with heterochromatin protein γ inside of the nucleus of cerebellar granule cells from mice and that the interaction might have a role in the regulation of genes involved in the DNA damage response. My work also revealed that receptors of the HNK-1 carbohydrate do not contain a common HNK-1 binding motif and that the HNK-1 receptors bind to the glycan via motifs unique for each receptor family.

3.   Introduction

3.1.   The cell adhesion molecule L1

The neural cell adhesion molecule L1 is a transmembrane protein that is mainly known for its multiple roles during neural development in vertebrates, such as axon development / growth, axon guidance / pathfinding, cell adhesion, myelination, cell-matrix adhesion, glial process formation, nervous system development, and neuronal migration (Hlavin and Lemmon, 1991; Jouet et al., 1994; Kudumala et al., 2013; Maness and Schachner, 2007; Schäfer et al., 2010; Tagliavacca et al., 2013). In the adult brain, L1's roles are not as diverse as during development but still comprise essential roles related to synaptic plasticity (Sytnyk et al., 2017).

In humans, the L1CAM gene is located in the long arm of the X chromosome at the Xq28 position and contains 29 exons. Two isoforms exist for the L1CAM gene, the neural variant, a product of the splicing of the exons 1 to 28, found in neurons and cells of neuronal origin, and the non-neuronal isoform that skips exons 2 and 27 (Reid and Hemperly, 1992; Takeda et al., 1996). The human neural variant mRNA codes for the 1 257 amino acid long (~200 kDa) neural cell adhesion molecule L1 transmembrane glycoprotein and the nonneural variant codes for a 1 253 amino acid long isoform that lacks residues 26 to 30, 1 177 to 1 180, and the residue 31 changes from valine to leucine when compared to the neural variant. L1CAM belongs to the L1 gene family (Maness and Schachner, 2007; Moos et al., 1988). The genes being part of the L1 gene family are L1CAM, CHL1 (located in 3p26.3), NrCAM (7q31.1), and NFASC (1q32.1) (Hortsch, 2000). Although they are all found in different chromosomes, they share a similar domain organization of six extracellular immunoglobulin domains (Ig), four to five extracellular fibronectin type III domains (FN3), a transmembrane region, and an intrinsically disordered intracellular domain at the C terminus (Figure 1 A). It has been hypothesized that this family is formed from 4 paralogs, resulting from two duplication events during the evolution of an ancestral L1 gene (Dehal and Boore, 2005; Mualla et al., 2013).

In vertebrates, the organization of the domains found in the L1 protein isoforms is also conserved among orthologs with the same distribution of six Ig domains and four to five FN3 (Figure 1 B). In zebrafish, a

duplication of the L1 gene nadl1 was shown on chromosome 23 dividing the nadl1 gene in nadl1.1 and nadl1.2 paralogs (Amores et al., 1998). Conservation in orthologs and a variety of paralogs of the gene coding for the neural cell adhesion molecule L1 in vertebrates and similar proteins among chordates (Mualla et al., 2013) may mirror the importance of L1 or L1-like proteins in neuronal development and activity.

A

| L1, *H.sapiens* | CHL1, *H.sapiens* | NRCAM, *H.sapiens* | Neurofascin, *H.sapiens* |
| 1257 aa | 1208 aa | 1180 aa | 1347 aa |

B

L1, *H.sapiens*
1257 aa

L1, *P.troglodytes*
1257 aa

L1, *M.musculus*
1259 aa

L1, *G.gallus*
1266 aa

L1, *X.tropicalis*
1250 aa

nadl1, D.rerio
1269 aa
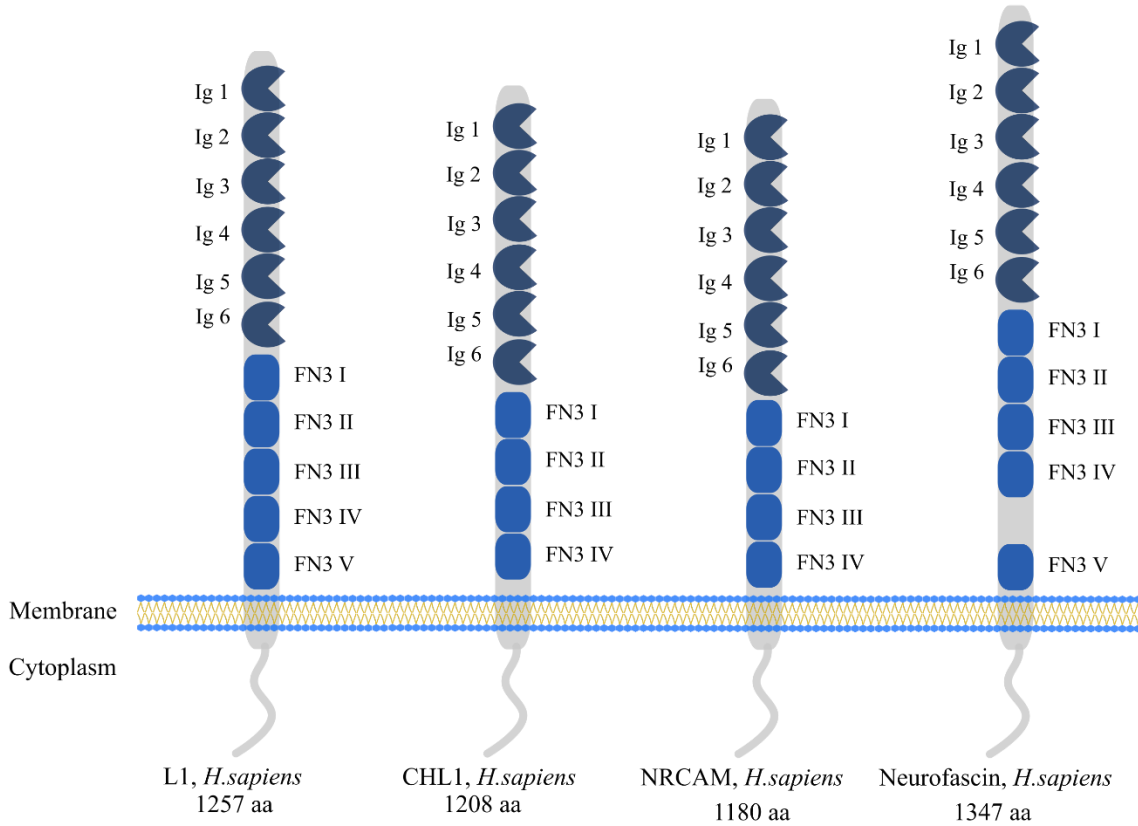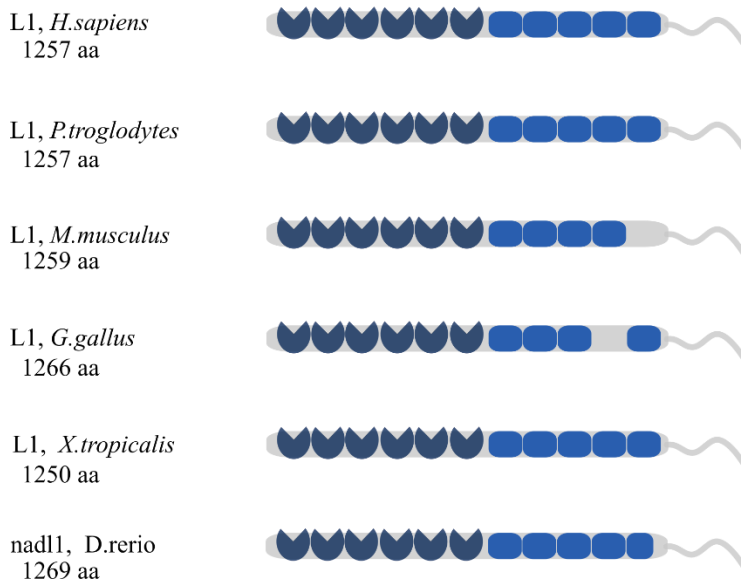
Figure 1. Schematic overview of L1 family members and conserved domains within the L1 protein. A. Structural comparison of the human L1 family of cell adhesion molecules. B. Structural comparison of the L1 molecule in different

organisms: L1CAM orthologs from vertebrate species (*H. sapiens* = human, *P. troglodytes* = chimpanzee, *M. musculus* = mouse, *G. gallus* = chicken, *X. tropicalis* = tropical clawed frog, and *D. rerio* = zebrafish). Shown are the immunoglobulin domains (Ig) as circular shapes, colored in deep blue and numbered 1-6, and the fibronectin type 3 domains (FN3) as rectangular shapes, colored in blue and numbered I-V.

3.2.   L1 and disease-causing mutations

Being such an important protein for neural development it is to expect that mutations affecting L1 might cause disease (Hortsch et al., 2014; Schäfer et al., 2010; Stumpel and Vos, 1993). X-linked hydrocephalus with stenosis of the aqueduct of Sylvius (HSAS), mental retardation, aphasia, spastic paraplegia, adducted thumbs (MASA), X-linked complicated hereditary spastic paraplegia type 1 (SPG1), and X-linked complicated corpus callosum agenesis are human neurodevelopmental disorders related to mutations in the L1CAM gene and collectively considered as L1 syndrome also known as CRASH syndrome. The L1 syndrome mostly affects homozygous males because of having just one copy of the X chromosome whereas heterozygous females are rarely affected with fewer than 5% showing symptoms (Vos et al., 2010). Hereditarily from a carrier female offspring, males have a 50% of chance of having one of the diseases from the L1 syndrome spectrum and females a 50% of chance of being a carrier of the disease. From the 212 disease-causing mutations in the dedicated database for L1CAM mutations, http://www.l1cammutationdatabase.info (Vos et al., 2010), 104 disease-causing mutations are either missense or nonsense point mutations (Figure 2), 51 are due to splicing defects, 46 are deletions or deletion-caused frameshifts, 7 are duplications or duplication-caused frameshifts, and the remaining 4 are due to other mutation types such as insertions and branch point mutations.
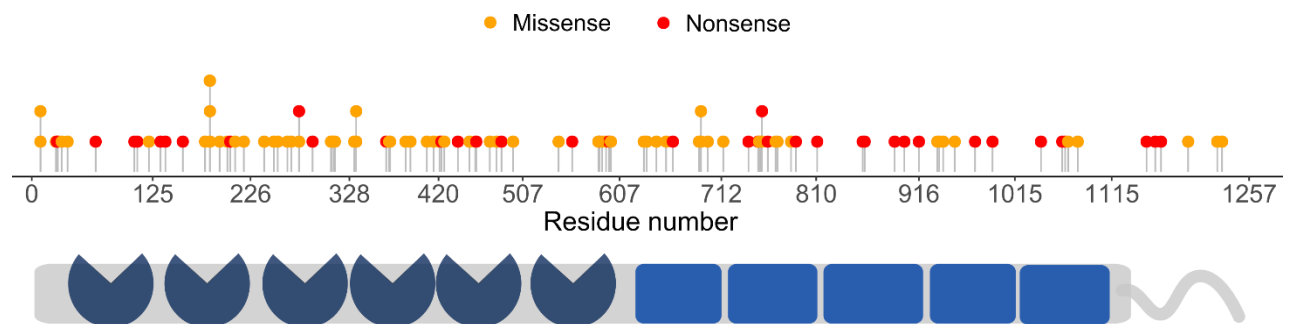
Figure 2. Independent disease-causing missense and nonsense mutations in human L1 protein. The lollipop plot on the top shows in the x-axis the residue numbers from the 1257 amino acids of the L1 protein. Shown are the locations of disease-causing missense (in orange) and nonsense (in red) mutations found in the database for L1 mutations (104 disease-causing missense and nonsense mutations) and mutations curated from the literature after 2012 (4 disease-causing missense and nonsense mutations). Staked are different mutations affecting the same amino acid. The diagram of the L1 protein shows the immunoglobulin domains (Ig) as circular shapes, colored in deep blue, and the fibronectin type 3 domains (FN3) as rectangular shapes, colored in blue.

From these 104 disease-causing point mutations in the database for L1 mutations, 6 are found in the N-terminus, 56 in the immunoglobulin domains, 38 in the fibronectin type 3 domains, and 4 in the intracellular domain. As most of the L1 interaction happens in the extracellular environment it is no surprise that mutations affecting the immunoglobulin domains or the fibronectin domains will affect the activity of the protein although it is noteworthy that a single point can have a strong phenotype such as hydrocephalus (Senat et al., 2001). Interestingly 4 disease-causing mutations are found in the intracellular domain of L1 (ICD, residues 1144 to 1257), one being a nonsense mutation (c.3496C->T; p.Arg1166X) "cleaving" most of the ICD and causing congenital hydrocephalus (Saugier-Veber et al., 1998). The other 3 disease-causing mutations are missense mutations of a serine residue leading to congenital hydrocephalus (p.Ser1194Leu, p.Ser1224Leu) (Fransen et al., 1994; Saugier-Veber et al., 1998) and a missense mutation of a tyrosine residue causing the MASA syndrome (p.Tyr1229His) (Fransen et al., 1997). As the last update of the database was in October 2012, a search in the PubMed database (NCBI, 2020) looking for disease-causing mutations in the L1 intracellular domain provided two additional publications dating later than 2012 introducing two novel nonsense mutations (p.Y1151X, p.Q1160X) (Adle-Biassette et al., 2013; Marín et al., 2015).

3.3. Post-translational modifications of L1

The L1 protein roles in neural development and functions are partially regulated by post-translational modifications such as glycosylations, the formation of disulfide bonds, phosphorylations, and proteolysis. The mature L1 human neural cell adhesion molecule has a mass of ~200 kDa although the theoretical mass

of the 1 257 of L1 is ~140 kDa. This means that 30% of the molecular weight of human L1 corresponds to post-translational modifications. Glycosylation represents the greater part of the extra mass. 20 glycosylation sites, 6 disulfide bonds, and 7 phosphorylation sites were curated from the human L1 protein using the UniProt knowledgebase (The UniProt Consortium, 2019) (Figure 3). The post-translational modifications are found in specific locations in L1 (Figure 3), disulfide bonds are specially located in conserved sites inside the immunoglobulin domains, N-glycosylation sites are found in the extracellular domain (Ig and FN3 domains) and phosphorylation sites are found in the intracellular domain.
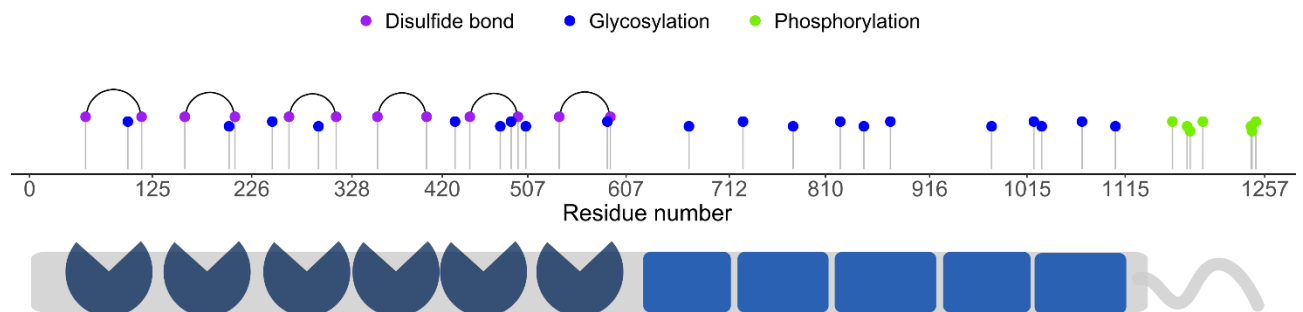
Figure 3. Post-translational modifications in human L1 protein. The lollipop plot on the top shows in the x-axis the residue numbers from the human 1 257 amino acids forming the L1 protein. Shown are the locations of post-translational modifications found in the UniProt knowledgebase. Disulfide bonds are shown in violet and linked by an arch, glycosylations are shown in blue, and phosphorylations in green. The diagram of the L1 protein shows the immunoglobulin domains (Ig) as circular shapes, colored in deep blue, and the fibronectin type 3 domains (FN3) as rectangular shapes, colored in blue.

Disease-causing missense point mutations affecting post-translational modified amino acids can be found independently in the human L1 protein. For example on the extracellular domain in one of the cysteines from the third Ig domain forming the disulfide bond where the cysteine is replaced by a tyrosine p.Cys264Tyr (Jouet et al., 1993) and p.Cys312Tyr (Kanemura et al., 2006) thus breaking the disulfide bond and causing the X-linked hydrocephalus phenotype. The fifth Ig domain disulfide bond is also affected by

a similar disease-causing point mutation p.Cys497Tyr (Finckh et al., 2000) leading to X-linked hydrocephalus and in the intracellular domain affecting the phosphosite p.Ser1194Leu (Fransen et al., 1994) leading to X-linked hydrocephalus. Although no direct mutation on a glycosylated residue has yet been found to lead to a known L1 disease, disease-causing defects in glycosylation of other proteins are known to cause disease (Freeze, 2006).

3.4.   Phosphorylation and L1

Protein phosphorylation is a reversible modification that is regulated via enzymes called kinases and phosphatases. Protein kinases add a phosphate group to serine or threonine residues mostly but can also phosphorylate tyrosines, histidines, aspartic acid, and glutamic acid. Phosphatases remove the phosphate group. Phosphorylation can allow temporary interactions and therefore regulate interactions leading to cell differentiation (Van Hoof et al., 2009), cell survival (Case et al., 2011), or apoptosis and nuclear transport (Jans and Hübner, 1996; Kuwahara et al., 2008) among others. In humans, L1 has been shown in-vitro that the L1 endocytosis is regulated by the dephosphorylation of the tyrosine residue situated at position 1 176 (Schaefer et al., 2002).

3.5.   Glycosylation and L1

Glycans are linear or branched chains of monosaccharides of varying lengths, linked by glycosyltransferases forming glycosidic bonds. Glycans can be bound to lipids and/or proteins (glycoproteins or proteoglycans) and serve as ligands for receptor proteins, called glycan-binding proteins or lectins (Kumar et al., 2012). Glycoproteins can be divided into two groups depending on the covalent bond between the glycan and the amino acid; N-glycans and O-glycans. N-glycans have a glycan chain usually connected via an N-acetylglucosamine (GlcNAc) terminus to the amide group of an asparagine residue. O-glycans have more variability concerning the covalent bond between a glycan and a serine or threonine. For example, glycoproteins commonly known as mucins have their O-glycans bound through an N-acetylgalactosamine link and the hydroxyl group of a serine or a threonine. Nevertheless, multiple other O-glycans exist with other linkages such as fucose, xylose, mannose to a serine or threonine (Varki et al., 2009). Glycans are involved in numerous biological processes, including pathogen-host recognition and interaction (Ielasi et al., 2016), tuning of inflammation and cancer (Dube and Bertozzi, 2005), cell-cell interactions, morphogenesis, cell adhesion, and cellular signaling. Proper glycosylation and glycan

recognition are very important for normal brain function during development and in the adult (Endo, 2005). It has been shown that protein-glycan interactions in the nervous system are involved in cell migration, neurite outgrowth and fasciculation, synapse formation and stabilization, and modulation of synaptic efficacy (Kleene and Schachner, 2004).

Chains of monosaccharides can be detected by antibodies as an antigen epitope. These epitopes have been found through the screening of cell-surface antigens on leukocytes via monoclonal antibodies called cluster of differentiation (Bernard et al., 1984). Each epitope was classified using a CD code. Later on, these epitopes received a common name depending on the type of cell or function they were linked to. Glycan epitopes can be found on both N- and O-linked glycans. The mature L1 protein carries two main glycan epitopes, HNK-1 / CD57 (Faissner et al., 1985) and LewisX / CD15 (Hennen et al., 2013; Streit et al., 1990). HNK-1 is highly expressed in the nervous system and plays a role in neural plasticity (Morise et al., 2017) and LewisX acts in neural development (Yagi et al., 2012) and cell-cell interaction (Sajdel-Sulkowska, 1998). Both epitopes are generated by two main enzymes, HNK-1 by the 3-beta-glucuronosyltransferase 1 (B3GAT1) and LewisX by the alpha (1,3) fucosyltransferase 4 (FUT4). Both epitopes are trisaccharides (Figure 4): in the case of HNK-1, N-acetylglucosamine is linked to galactose which is also linked to glucuronic acid, and in the case of LewisX, the N-acetylglucosamine is linked to galactose and fucose. HNK-1 and LewisX glycan epitopes can be modified to a sulfated form, especially when on N-Glycans. The biosynthesis of the sulfated moieties is performed by sulfotransferases in the third position of the glucuronic acid of HNK-1 and the third position of the galactose of LewisX. Glycan epitopes have been shown to play a role in the myelination of neurons (Díez-Revuelta et al., 2017) and to regulate myelination by N-glycan sulfation on glycoproteins (Yoshimura et al., 2017). The LewisX can also be modified into Sialyl-LewisX (CD15s) by linking sialic acid to the third position of the galactose of LewisX by enzymes of the sialyltransferase family.
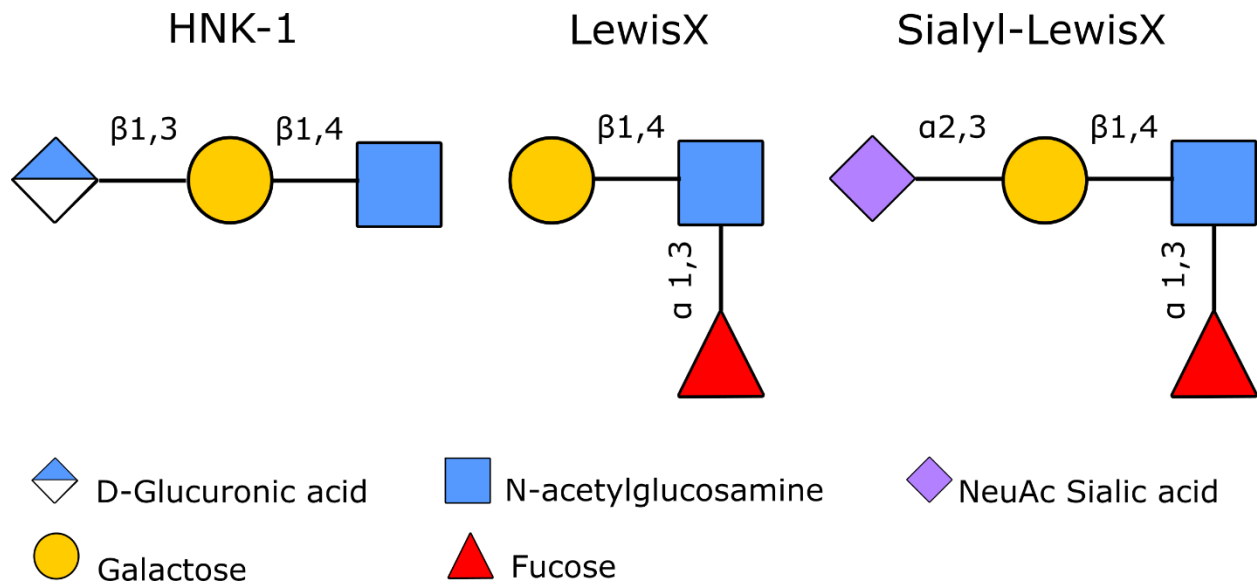
Figure 4. Diagram of the trisaccharide epitopes HNK-1 and LewisX. Glycan representation following the Symbol Nomenclature For Glycans (SNFG) (Neelamegham et al., 2019; Varki et al., 2015) of the HNK-1, LewisX, and Sialyl-LewisX epitopes. Circles represent hexoses, squares N-acetylhexosamines, the divided (part white-colored) diamond hexuronates, the diamond Deoxynonulosonates, and the tringle Deoxyhexoses. The color represents the monosaccharide type.

Receptors of glycan epitopes are mainly proteins called lectins (Sharon and Lis, 2004) and most notably C-type lectins. C-type lectins are known to bind a variety of glycans such as LewisX and sialyl-LewisX, and crystal structures from these lectins were obtained depicting the interaction (Guo et al., 2004; Pröpster et al., 2016; Somers et al., 2000). On the other hand, little is known about the recognition of the HNK-1, although efforts have been made to discover the binding domain of HNK-1 and it has been shown that a region from the G2 domain of α-laminin is responsible for binding HNK-1 (Hall et al., 1997, 1995). Known HNK-1 receptors are α-laminin (Hall et al., 1993), the high mobility group box (HMGB) proteins 1 and 2 (Chou et al., 2004; Fang et al., 2012; Rauvala and Rouhiainen, 2010) and cadherin-2 (also known as neural cadherin) (Morita et al., 2009). The HNK-1 epitope is not only found on glycoproteins but can be also expressed in glycolipids. Its functions in the nervous system are numerous and go from common neural roles such as cell adhesion, recognition, migration, synaptic plasticity (Hall et al., 1993; Kizuka and Oka,

2012; Morise et al., 2017; Schachner et al., 1995) to more unique and outstanding roles such as preferential motor re-innervation and regeneration in the peripheral nervous system and central nervous system after injury (Löw et al., 1994; Martini et al., 1994; Sahu et al., 2018; Simova et al., 2006).

3.6. L1 soluble fragments

Apart from the extracellular activity of the L1 molecule as a transmembrane protein, L1 can also be found in soluble forms as products from proteolytic cleavages. One of the fragments shown to be generated (Lutz et al., 2014, 2012) after sumoylation at the cytoplasmic domain of murine L1 and proteolyzed by the myelin basic protein at the arginine residue 687 is the L1 70 kDa fragment (L1-70). By this cleavage, two soluble fragments are produced: an ~130 kDa extracellular L1fragment containing the 6 immunoglobulin domains and an intracellular L1-70 fragment containing most of the 5 fibronectin type 3 domains, the transmembrane region, and the cytoplasmic domain. The L1-70 fragments were shown to be imported into the nucleus where they promote neuronal migration and neuritogenesis. Other fragments have been also found to be the product of the proteolysis of L1 such as the 28 kDa L1 fragment found in the nucleus of cancer cell lines, cleaved by the metalloprotease ADAM10 (Riedle et al., 2009), and the 140 kDa L1 fragment found in the hippocampus, generated by cleavage by the proprotein convertase PC5A (Kalus et al., 2003) and showing to have a role in neurite outgrowth. Soluble fragments of the L1 protein produced by proteases were found to be internalized into organelles such as the mitochondria regulating the mitochondrial complex I activity, and mitochondrial trafficking (Kraus et al., 2018a) or the nucleus where they regulate synaptic plasticity, and motor coordination (Kraus et al., 2018b).

3.7. The HP1 family of proteins

An interesting family of nuclear proteins is the Heterochromatin Protein 1 (HP1) family. In mammals, the highly conserved family consists of three paralogs HP1α encoded by the chromobox homolog 5 (CBX5) gene, HP1β by the CBX1 gene, and HP1γ by the CBX3 gene. The genes are all located on different chromosomes, for example in humans on chromosomes 12, 17, and 7 respectively. HP1 proteins are relatively small proteins (~20 kDa) and contain two related domains: the chromodomain and the chromo shadow domain, joined by a variable and unorganized hinge region. The chromodomain recognizes histone H3 methylated at lysine 9 (Bannister et al., 2001), the chromo shadow domain mediates dimerization of HP1 forming a binding surface for unorganized peptide ligands containing the peptide motif PXXVXL in

histone H3 or PXVXL in other proteins (P: proline, X: any amino acid, V: valine and L: leucine) (Cowieson et al., 2000; Smothers and Henikoff, 2000; Thiru et al., 2004) and the hinge region which contains a nuclear localization signal and interacts with DNA and RNA (Meehan et al., 2003; Muchardt et al., 2002).

Although being highly homogenous, each paralog has specific functions (Aucott et al., 2008; Brown et al., 2010) and has a specific location in the nucleus due to the nuclear localization signal located in the hinge region: HP1α is mostly located in centromeres, HP1β is generally located in the nucleus, and HP1γ is found in euchromatin (Minc et al., 2001, 2000, 1999). This specific location can be correlated to unique roles: HP1α for example is shown to promote nucleosome association driving chromatin condensation (Azzaz et al., 2014), HP1β is shown to regulate cell differentiation by deferentially allocating and associating with chromatin (Mattout et al., 2015) and HP1γ is shown to play a key role in transcription elongation (Zaidan and Sridharan, 2020). The binding region recognizing the PXVXL motif allows the HP1 proteins to interact with other proteins than the histone H3 (Liu et al., 2017) (Figure 5) such as the Chromatin assembly factor-1 (CAF-1) (Yan et al., 2018), the transcriptional intermediary factor 1 family (TIF1) (Nielsen et al., 1999) or the Shugoshin 1 (Sgo1).
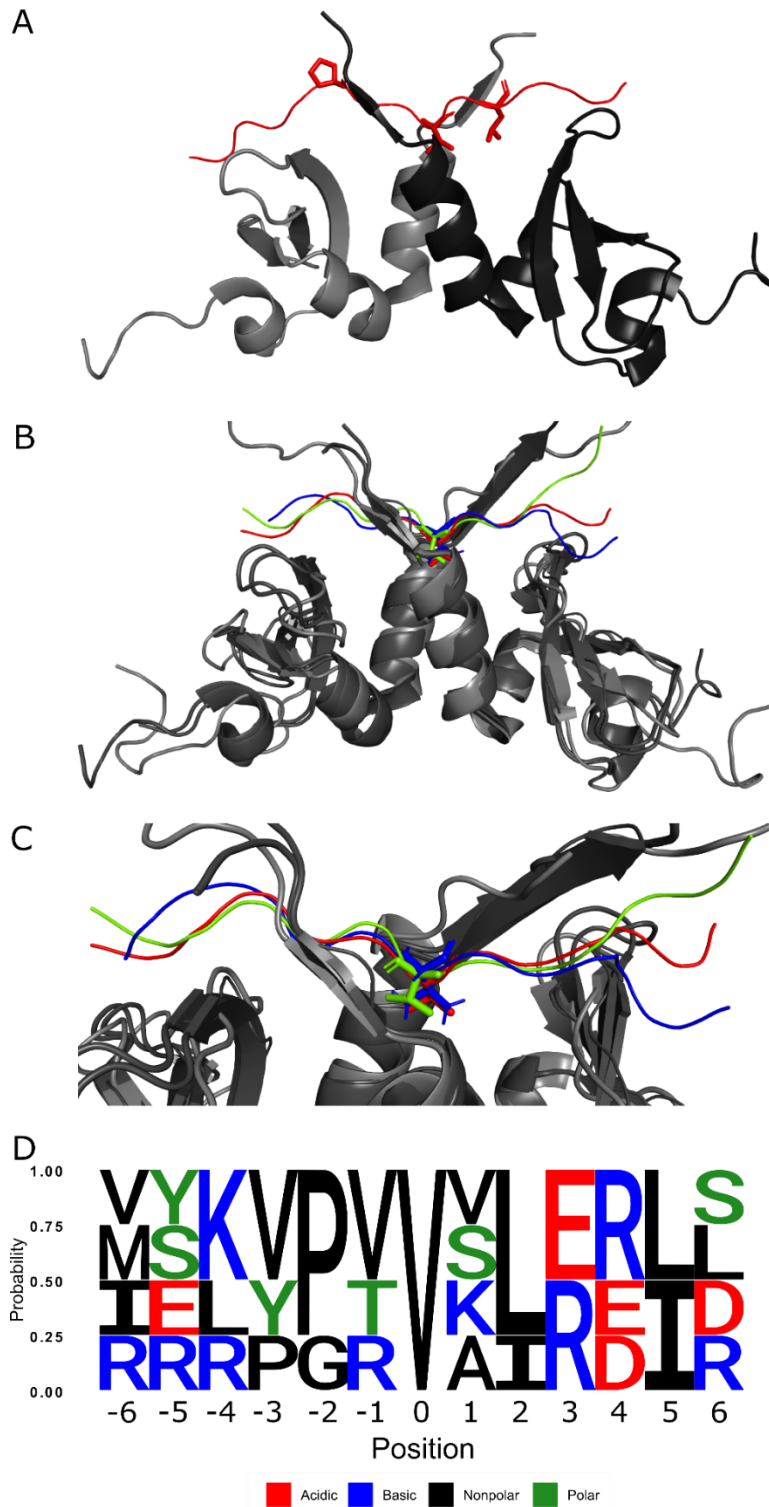
Figure 5. HP1 interaction with the PXXVXL and PXVXL/I motifs. A: Structure of the complex of the human chromo shadow domain (CSD) dimer from HP1γ (residues 110-176, in grey and black, PyMOL cartoon secondary structure

visualization) and the histone H3 peptide (residues 41-53, in red, PyMOL ribbon backbone visualization) containing the PXXVXL motif (showing the proline, valine, and leucine residues, PyMOL sticks side-chain visualization) from the PDB crystal structure 5T1I. B: Structural alignment (PyMOL, Align) of the complex of the CSD dimer from HP1 proteins (multiple shades of grey) and the PXVXL-like motifs. Shown are the structure of the human CSD dimer from HP1γ (residues 110-176) with the H3 ligand in red (residues 41-53) from the PDB 5T1I, the mouse CSD dimer from HP1β (residues 104-176) with the CAF-1 ligand in blue (residues 218-230) from the PDB 1S4Z, and the human CSD dimer from HP1β (residues 108-185) with the Sgo1 ligand in green (residues 447-459) from the PDB 3Q6S. The side chain from the valine residues from the PXVXL-like motifs is shown. C: Zoom inside the binding pocket of the CSD dimer D: Sequence logo from the alignment from histone H3 PXXVXL, CAF-1, TIF1β, and Sgo1 PXVXL/I motifs. Adapted from Liu et al., 2017.

3.8.   The mouse as an animal model for L1

The mouse, *Mus musculus*, is the most used species for biochemical research (Hickman et al., 2017). Mice were selected as the animal model for this thesis for the advantage they represent when studying neurodegenerative disorders caused by gene mutations. The resemblance of phenotypes from disease-causing mutations of highly conserved genes in both humans and mice can elucidate biochemical pathways and cell signaling where the mutated protein is essential or partially essential for normal functioning. Human and mouse L1 proteins have an identity of 87.8% and a similarity of 93% (human L1 Swiss-Prot ID: P32004.2 and mouse L1 Swiss-Prot ID: P11627.1, alignment using the EMBOSS Needle program; (Sb and Cd, 1970)) and their intracellular domains (last 114 amino acids for both) are 100% identical. HP1γ human and mouse proteins are of the same size, 183 amino acids, and are almost identical with an identity of 99.5% (human HP1γ Swiss-Prot ID: Q13185.4 and mouse HP1γ Swiss-Prot ID: P23198.2, alignment using the EMBOSS Needle program), the only difference being the 75th residue belonging to the chromodomain, in humans asparagine and mouse aspartic acid. Therefore, both human and mouse chromo shadow domains are 100% identical.

L1 deficient mice have been engineered to successfully mimic the human L1syndrome (Dahme et al., 1997). These mice have helped to elucidate L1 functions. L1 knockout male mice (L1 -/y) are unable to reproduce, therefore it is not possible to obtain K L1 deficient females L1 (L1 -/-). The lineage is kept by back-crossing

heterozygous females L1-/+ with wild-type males L1 +/y from a genetically similar strain. The offspring mendelian ratios of crossing L1-/+ heterozygous females with L1 +/y males should be 25% for each of the 4 possible combinations: L1 -/+, L1 +/+, L1 +/y, and L1 -/y. The observed ratios in litters obtained during the years 2016 and 2017 were: for wild type genotypes L1 +/+ and L1 +/y approximatively 32%, heterozygous L1 -/+ approximatively 25%, and KO hemizygous L1 y/- approximatively 11%. It was often informed that the mothers tend to eat the hemizygous L1 y/- offspring and that the male -/y offspring are smaller and weaker which leads to the premature death of the animals, which among other factors could explain the sub-mendelian ratios.

Furthermore, drastically silencing or knocking out one of the CBX genes led to early lethality (Aucott et al., 2008; Brown et al., 2010; Thliveris et al., 2012) which makes it difficult to work with such mutants. Other alternatives such as silencing the CBX genes in cultured cells using a viral vector were selected as a study model.

4.   Objective

The focus of my research project was the identification and characterization of nuclear binding partners of soluble L1 fragments using bioinformatical, biochemical, immunohistochemical, and biotechnological approaches. The investigation took place in-vitro by experimenting with recombinant proteins and using primary cell cultures of mouse cerebellar granule cells as the neural model to study the protein interactions and the possible involvement of such interactions in gene regulation and neuroprotectant mechanisms. Using the mouse as the animal model could help to better understand homolog proteins interactions in humans.

This research was also focused on the finding of novel nuclear receptors of glycan epitopes possibly present in soluble L1 fragments such as the glycans HNK-1 and LewisX and the investigation of their function within the nucleus of brain cells. To achieve this goal, bioinformatical and molecular modeling approaches were used.

5. Results

5.1. L1 interaction with HP1γ

The cell adhesion molecule L1 plays important functions during the development of the nervous system and in nervous system functions in adulthood. In addition to homophilic L1-L1 interactions heterophilic cis- and trans-interactions of cell surface L1 with other proteins have been described. The recent discovery of the generation and transport of L1 fragments into the nucleus of neural cells of mice (Lutz et al., 2014, 2012) suggests that these nuclear L1 fragments can interact with nuclear proteins or the DNA and thereby modulate L1-mediated functions. To identify motifs within the L1 protein that might mediate interaction with nuclear proteins, the web tool database Minimotif Miner (MnM) was used. Diverse consensus motifs were found within the full-length mouse L1 protein (164 hits; data not shown). Since the L1-fragments present in the nucleus contain the intracellular domain and only small portions of the extracellular domain, for further analysis only protein binding motifs found in the intracellular domain of L1 (L1- ICD) comprising amino acids 1147 to 1260 were selected (Table 1). Four potential motifs were found in the L1-ICD which could be recognized by a binding partner.

Motif	Occurring Positions	Annotation	Gene Name	Score	Cellular Location
DxxDxxxD	1157-1164	This consensus motif binds the phosphatase domain of PP1	Ppp1cc	0.78	Nucleus, Mitochondrium
[FWY]G[DEP][FLMWY]	1175-1178	This consensus motif binds the ear domain of the AP1 gamma subunit	Ap1g1	0.86	Golgi apparatus
E[ST]D	1182-1184	This consensus motif in neurofascin binds Ankyrin G	Ank3	0.76	Lysosome, Cytoskeleton
PxVxL	1254-1258	This consensus motif binds the chromo shadow domain of HP-1	Cbx1,2,3	0.86	Nucleus

Table 1. Consensus sequences within the L1- ICD. Shown are 4 potential motifs in the L1- ICD that could be recognized by binding partners. The motif column describes the aminoacidic sequence, where each letter represents the one-letter amino acid code,

letters inside brackets mean "one of the letters inside of the brackets" and x meaning "any amino acid is accepted". The position column indicates the residue numbers in the protein sequence where the motif starts and ends. The annotation column describes the binding partner interaction details. The filter score is found between 0 and 1, where 1 is the score for a proven interaction. The closest the score is to 1 the more probable is that the interaction takes place. Cellular location was predicted using the Uniprot database for the gene name and species.

Binding motifs for two nuclear proteins were found in the L1- ICD: the serine / threonine-protein phosphatase PP1-gamma catalytic subunit and the chromobox protein homologs 1,2 and 3 family members. As the canonical binding motif for the HP1 family of proteins, PXVXL, has been widely studied and is well characterized in other proteins, I decided to focus on this motif and the potential interaction of L1 with the HP1 family proteins.

5.1.1. Interaction of HP1 family members with L1

To verify the *in-silico* results indicating a possible interaction between L1- ICD and the HP1 family of proteins via the PXVXL motif within L1, the C-terminus of the L1- ICD containing the canonical HP1-binding pentapeptide motif PXVXL at position 1254 to 1258 and a mutated sequence disrupting this motif was generated and purified. For this aim, the L1- ICD sequence comprising the 345 base pairs coding for the L1- ICD was cloned into the pET-28a(+) plasmid (Figure 6.). To disrupt the HP1-binding pentapeptide motif PXVXL, the sequence coding for this motif was changed from PAVALE to PAAAAE using the pET vector containing the L1- ICD cDNA as a template for site-directed mutagenesis.
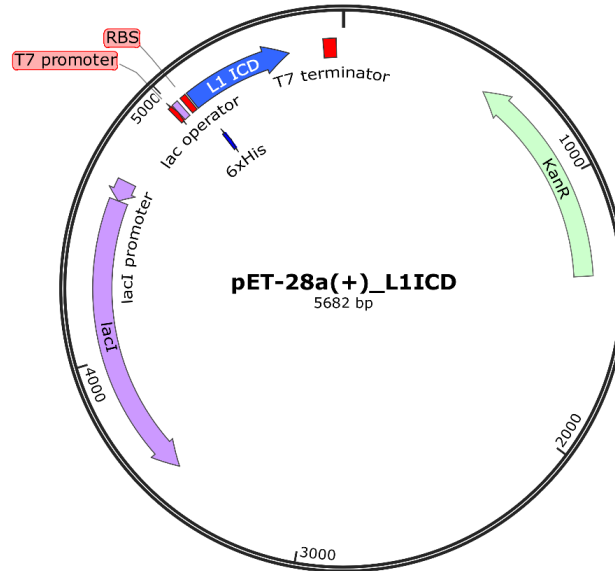
Figure 6. Map of the pET-2a(+) vector containing the cDNA coding for L1- ICD. In blue, the L1- ICD insert in the same reading frame with the 6X histidine tag is shown. In red, the T7 expression system with the T7 promoter for the T7 polymerase, the T7 terminator encoding an RNA stem-loop allowing efficient transcription termination, and ribosome binding site are shown. In violet, the lac operon is shown: lac operator sequence recruiting the lac repressor protein coded by the lacI promoter. The lac repressor binding to the DNA stops in presence of IPTG. In green, the Kanamycin resistance cassette is depicted.

Both constructs were amplified by transforming them into One Shot TOP10 chemically competent *E. coli*, sequenced to verify the site-directed mutagenesis in the mutant clone, and then transformed into the chemically competent *E. coli* One Shot BL21 Star (DE3) for protein production. Wild-type and mutated L1- ICD proteins were produced in bacteria, purified, and subjected to SDS-PAGE and subsequent western blot analysis using an L1 specific antibody for the detection of purified proteins (Figure 7.). Products were expected to have a molecular weight of approximately 15 kDa.

L1-ICD	+	-
Mutated L1-ICD	-	+
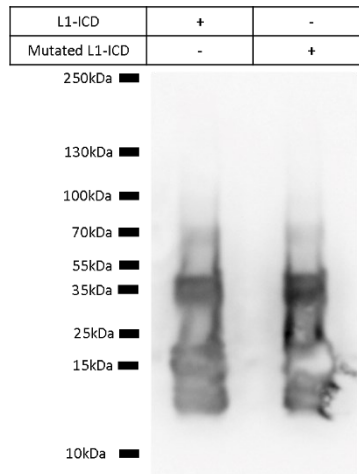
Figure 7. Western blot analysis of purified recombinant wild-type and mutated L1-ICD proteins. Purified proteins were separated on SDS-PAGE and subjected to western blot analysis with monoclonal L1 antibody C2 which recognizes an epitope in the L1- ICD.

For both recombinant proteins, two signals could be observed: one between 10 kDa and 25 kDa and the other one between 35kDa and 55kDa. The higher molecular weights could be explained to be dimers of L1-ICD.

To elucidate if L1 directly interacts with HP1 family members via its ICD an ELISA was performed. For this, commercially available recombinant full-length HP1α, β, and γ proteins were used as substrate coating, and different concentrations of wild-type and mutated L1- ICD were applied as a ligand (Figure 8.).
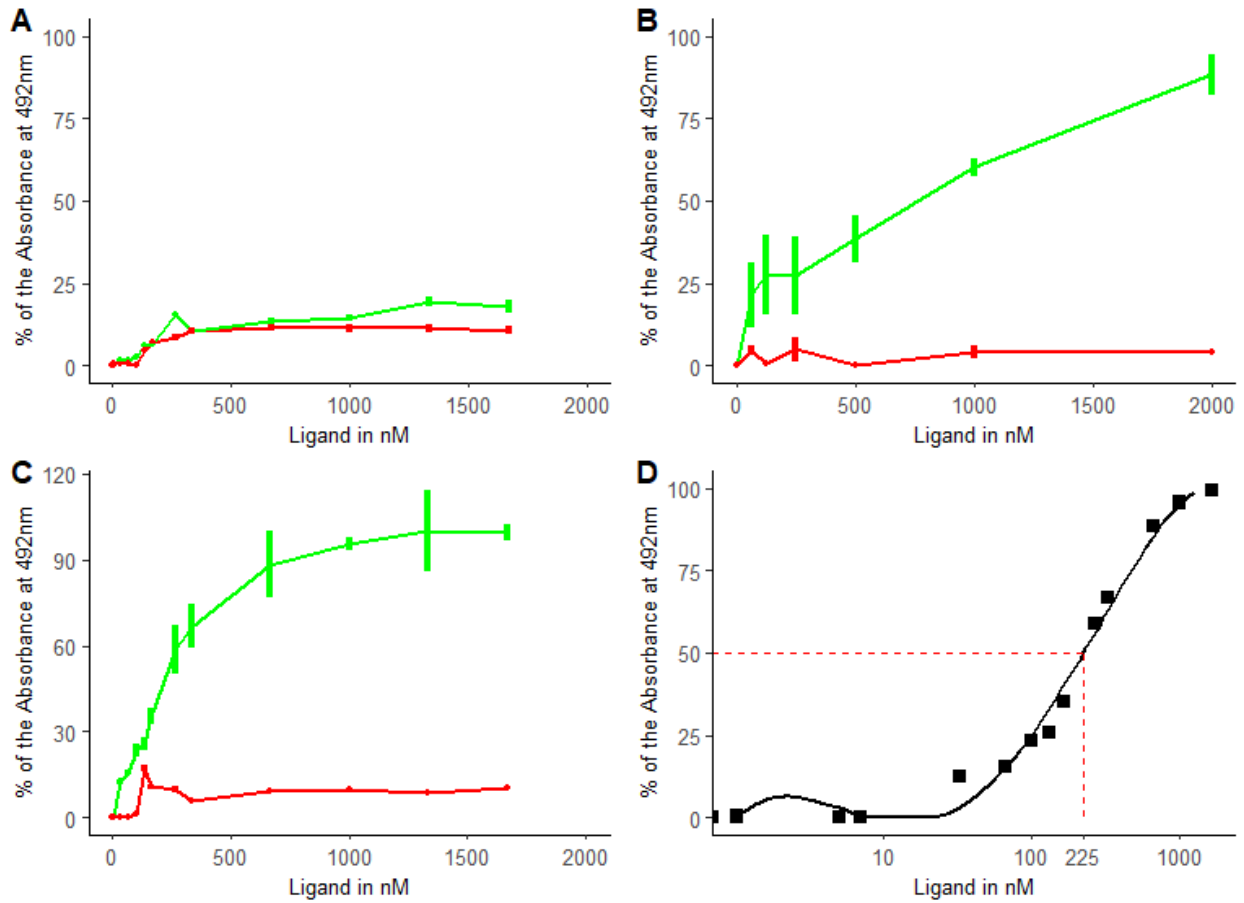
Figure 8. Binding of the HP1 family proteins to the intracellular domain of L1. HP1α (A), β (B), and γ (C)(1 μM) were immobilized and different concentrations of L1- ICD containing the PXVXL motif (in green) or mutated L1- ICD with the mutated motif PAAAA (in red) were applied as a binding partner. Bound L1- ICDs were detected using the L1 C2 monoclonal antibody. Error bars represent the standard error of the mean. The absorbance was normalized using a min-max scaling to obtain the absorbance. The experiment was performed once using triplicates for HP1α and γ and twice using triplicates for HP1β. D: Data from the % of the absorbance measurement of the binding of L1- ICD (C, the curve in green) replotted against a log10 scale (x-axis). The dotted red line shows the ligand concentration needed to obtain 50% of the absorbance (estimated Kd).

Results from the ELISA experiments show that only the L1- ICD binds to HP1γ in a concentration-dependent and saturable manner and confirm that this binding is mediated by the PXVXL motif present in the L1- ICD since the mutated L1- ICD does not bind to HP1γ (Figure 8.). The results also imply the need for the dimerization of the receptor to bind the ligand (50% of bound ligand ~225nM and the concentration

of the receiver as a dimer ~500nM). The non-mutated L1 -ICD containing the PXVXL motif also binds to HP1β however not in a concentration-dependent manner, and the binding is also mediated by the PXVXL motif as the mutated L1 -ICD does not bind to HP1β. In contrast, there was no interaction measurable for HP1α and the L1- ICD.

5.1.2. Binding of L1- ICD to the chromo shadow domain of HP1γ

Since the PXVXL motif was reported to interact with the chromo shadow domain (CSD) of HP1γ (Thiru et al., 2004) and the PXVXL motif within L1- ICD mediates binding to HP1γ, the next step was to determine if the L1- ICD also binds to the CSD of HP1γ. To analyze if this hypothesis is true, I produced recombinantly the CSD of HP1γ in *E. coli* using a commercial plasmid containing the human CBX3 mRNA. The reported sequence having the strongest affinity to the PXVXL motif (Liu et al., 2017) corresponding to the residues 110 to 176 of HP1γ CSD, was extracted and inserted in a pLATE52 vector. The recombinant protein was expressed in *E. coli*, purified, and used for further studies (Figure 9.).
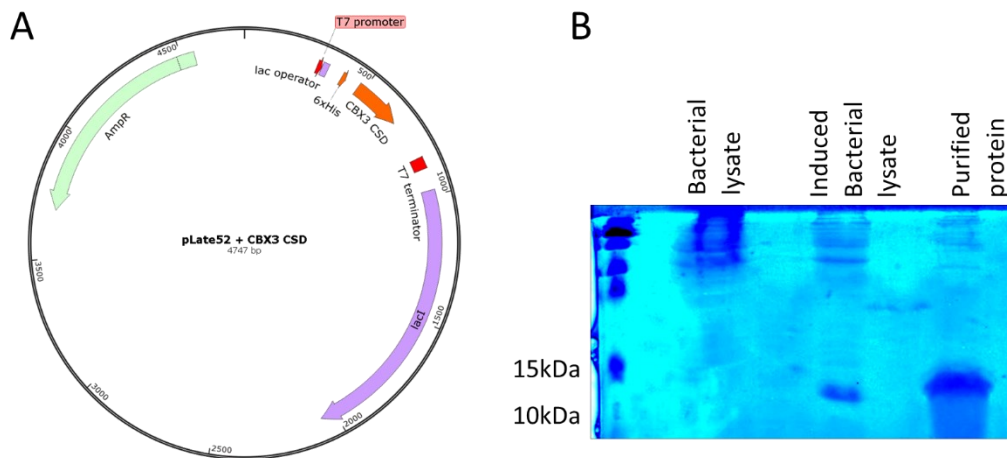
Figure 9. Production of the HP1γ chromo shadow domain. A: Map of the pLATE52 vector containing the cDNA coding for the CSD of HP1γ. In orange, the HP1γ CSD in-frame with the 6X histidine tag is shown. In red, the T7 promoter and terminator. In violet, the lac operon: lac operator and lacI. In green, the Ampicillin resistance cassette. B: Image of the Coomassie blue staining of an SDS gel with proteins from the bacterial lysate before IPTG induction (bacterial lysate), after 4 h of IPTG induction (induced bacteria lysate) and protein purification via nickel-beads and extracted by on-column proteolysis using the WELQut serine protease (purified protein).

A recombinant protein with a size between 10 and 15kDa was obtained (Figure 9.). Unfortunately, a polyclonal antibody against the C-terminus of HP1γ did not recognize the recombinant protein in western blot (data are not shown), and the histidine tag was removed after the proteolysis. The result of an overexpressed protein only visible after induction with IPTG and the interaction with the nickel matrix proves that a recombinant protein was produced.

To determine if HP1γ binds to the L1- ICD via its CSD an ELISA was performed by immobilizing the recombinant HP1γ CSD and applying different concentrations of L1- ICD and mutated L1- ICD ligands(Figure 10.).
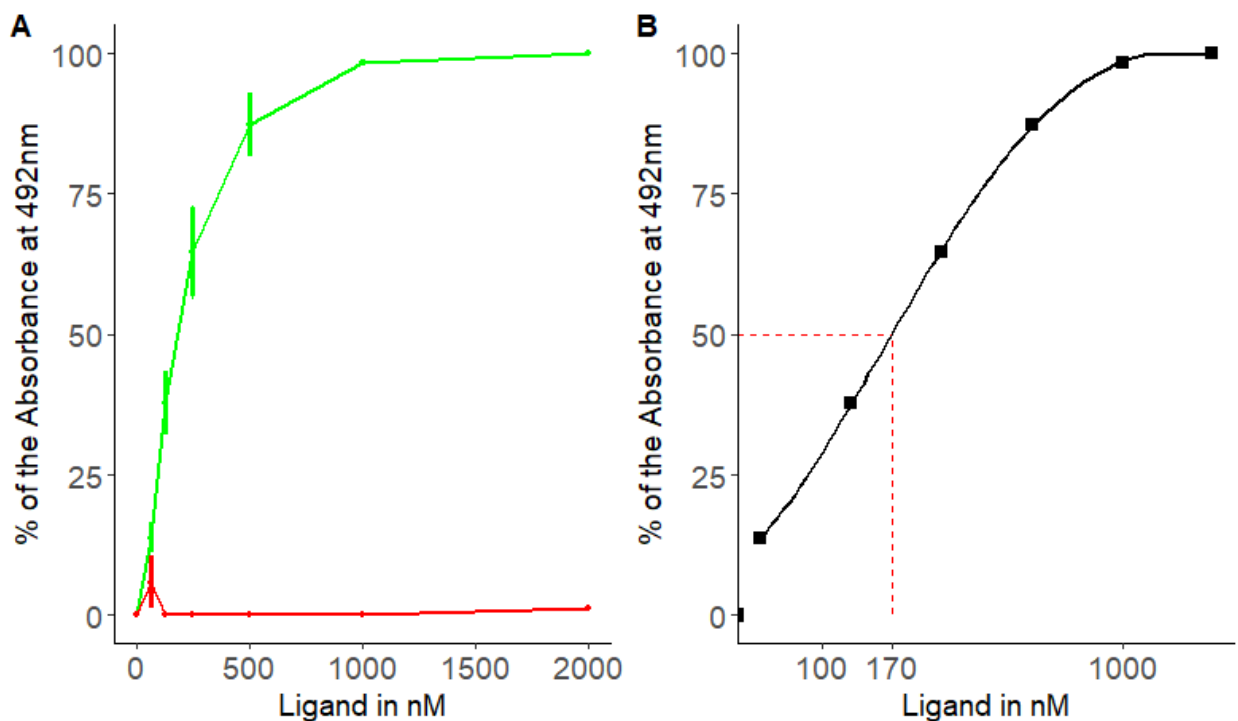
Figure 10. Interaction between the CSD of HP1γ and the intracellular domain of L1. The plate was coated with 1 µM of recombinant HP1γ CSD and different amounts of soluble L1- ICD and mutated PAAAA L1- ICD were applied as ligands. Monoclonal L1 antibody C2 was used to detect the binding between the L1- ICD and the recombinant HP1γ CSD. The experiment was repeated 9 times. A: percentage of absorbance with increasing ligand concentration (L1- ICD in green

and mutated L1- ICD in red). Error bars represent the standard error of the mean. B: L1- ICD data replotted against a log10 scale (x-axis). The dotted red line shows the ligand concentration needed to obtain 50% of the absorbance (estimated Kd).

L1- ICD, but not the mutated L1- ICD bound to the CSD of HP1γ in a concentration-dependent and saturable manner, verifying that L1- ICD binds to the CSD of HP1γ via its PXVXL motif (Figure 10.).

5.1.3. HP1γ and the L1- ICD are located in close proximity in the nucleus of neural cells

To determine if the interaction between L1- ICD and the CSD of HP1γ takes place in brain tissue or cultured neurons, fluorescence microscopy was used. Fragments of L1 are present in lower concentrations compared to full-length L1 under basal conditions (Lutz et al., 2014, 2012). Therefore, colocalization of both L1-ICD, and HP1γ might not be observed in tissue because L1 fragment levels might be below the detection limit. It is noteworthy to mention in this context that isolated cerebellar granule cells (CGCs) can be stimulated with the L1-specific function-triggering antibody 557 which leads to enhanced generation and thus the presence of L1 fragments inside of the cells. First, I stained for L1 and HP1γ in brain slices from early postnatal mouse brains (Figure 11. and Figure 12.). For this experiment, an antibody against L1 was chosen that reacts with the L1- ICD, and thus full-length L1, as well as L1 fragments L1-70 and L1-30, will be detected.
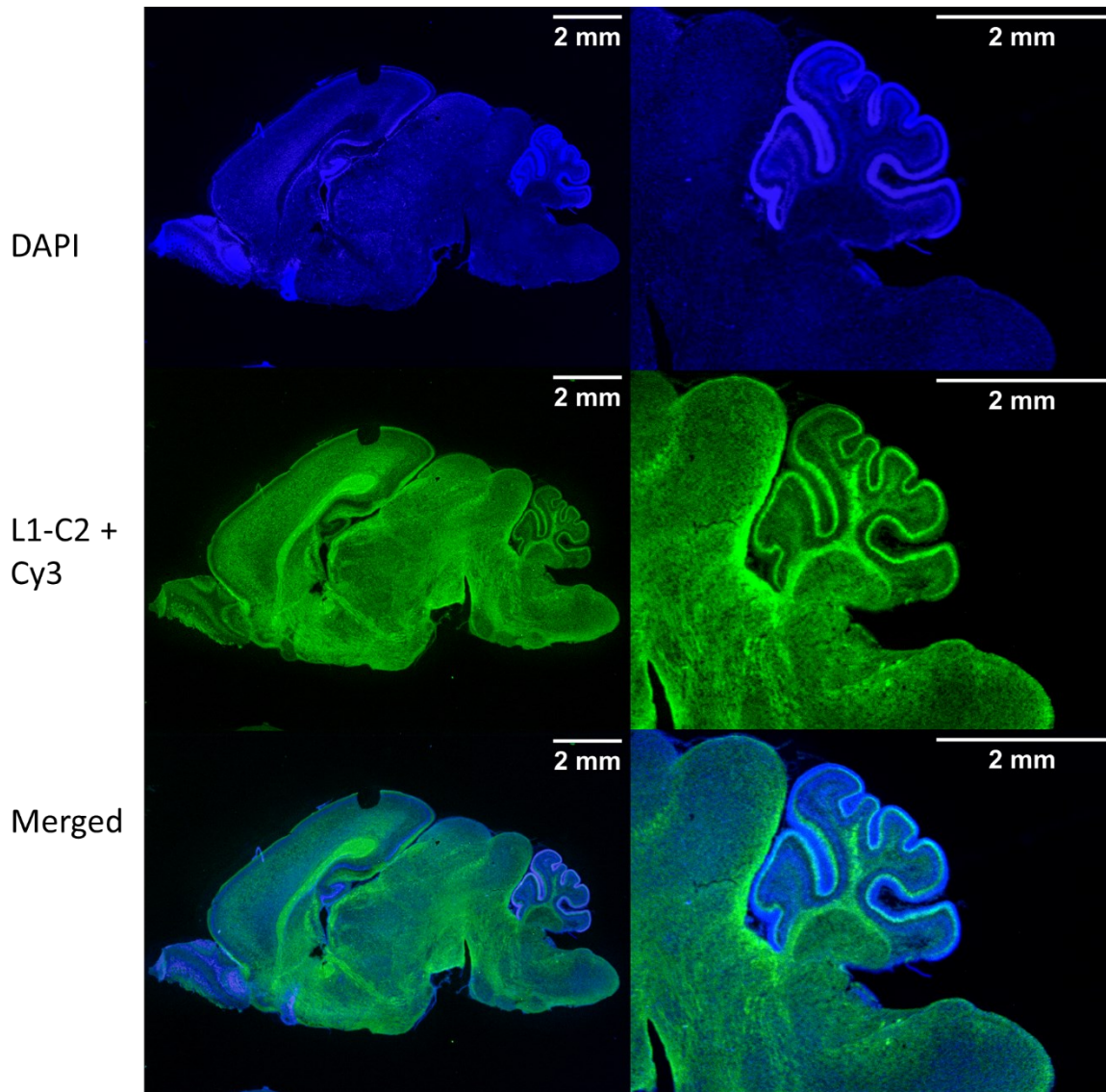
Figure 11. L1 expression in the brain at the postnatal day 7. Brains of 7-day-old wild-type mice were fixed in 4% formaldehyde, dehydrated in sucrose, and flash-frozen in liquid nitrogen. 15 µm brain slices were obtained by cutting the brain sagittally using a cryostat. The slices were then stained with DAPI (blue, nuclear staining) and against L1 using the monoclonal antibody L1 C2 (shown in green). Representative immunofluorescence images are shown.

L1 is mostly expressed in the corpus callosum and cerebellum, but it is also present in lower quantities in all other brain regions. In the cerebellum, a high number of cells and a substantial amount of L1 can be detected (Figure 11.).

After showing the presence of L1 in the cerebellum of 7-day-old mice, the localization of L1, and possible colocalization of L1 and HP1γ were determined (Figure 12.).
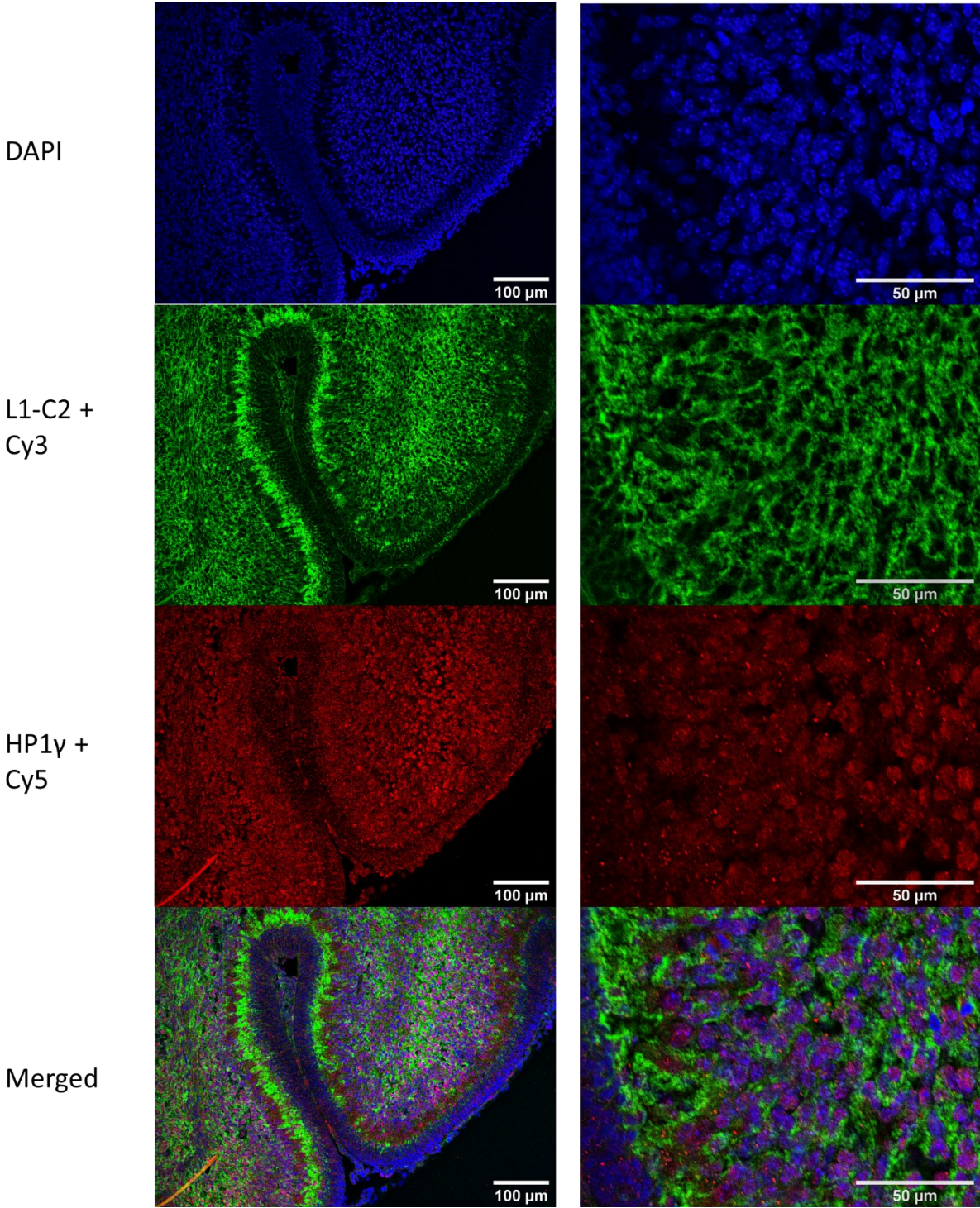
Figure 12. Expression of L1 and HP1γ in the cerebellum of 7-day old mice. Brains from 7-day-old wild-type mice were fixed in 4% formaldehyde, dehydrated in

sucrose, and flash-frozen in liquid nitrogen. 15 µm brain slices were obtained by cutting the brain on the sagittal plane using the cryostat. The slices were then stained using DAPI (blue; nuclear staining), the L1-C2 monoclonal antibody, and Hp1γ antibody as primary antibodies and Cy3 anti-mouse and Cy5 anti-rabbit as secondary antibodies. Immunofluorescence staining shows L1 in green and HP1γ in red. Representative immunofluorescence images are shown.

Neither L1 inside of the nucleus or the colocalization between L1 and HP1γ were detectable in cerebellar tissue under these conditions probably due to low amounts of nuclear L1 in comparison to the cell surface levels of L1 (Figure 12.).

To obtain a higher resolution of nuclei and L1 staining within nuclei of cerebellar neurons, neurons from 7-day-old L1 +/y and L1 -/y littermate mice were isolated and cultured overnight. Then, cells were fixed and stained for L1 and HP1γ.
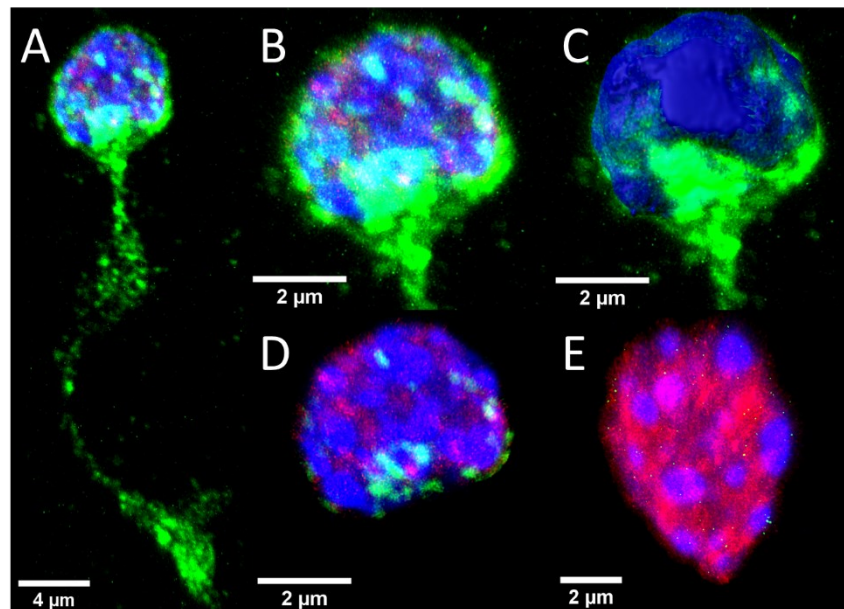
Figure 13. Localization of L1 and HP1γ in cerebellar granule cells of 7-day old mice. 3D image from stacks of immuno-fluorescence staining images for DAPI (blue), L1 (green), and HP1γ (red) using the Imaris software. A: L1 +/y neuron, soma, and single axon. B: Zoom of the soma. C: 3D mask of the nucleus (DAPI fluorescence in blue). Outside the mask the non-nuclear L1 signal (green). D: Masked fluorescence (signal inside of the nucleus was filtered from the rest of the fluorescence outside of the nucleus) of the nucleus volume showing L1 and HP1γ

signals inside of the nucleus. E: Same procedure applied for A-D but in neurons from L1 -/y mice. Representative immunofluorescence images are shown; the experiments were done in 18 CGCs (n=3).

Although only low levels of colocalization of L1 and HP1γ could be observed (in yellow, not easily observable at plain sight and only observable when calculating voxel overlapping using the Imaris software colocalization feature) in L1 +/y CGCs, L1 was clearly visible inside of the nucleus of L1 +/y CGCs. Higher levels of HP1γ in L1 -/y neurons could be observed in comparison with L1 +/y neurons (Figure 13.).

To determine if L1 fragments are in close proximity to HP1γ inside of the nucleus allowing direct interaction, a proximity ligation assay was performed using L1 +/y CGCs, antibodies against the C-terminus of L1, antibodies directed against the chromodomain of HP1γ (which is not involved in the interaction with L1) and the respective proximity ligation probes (Figure 14.).
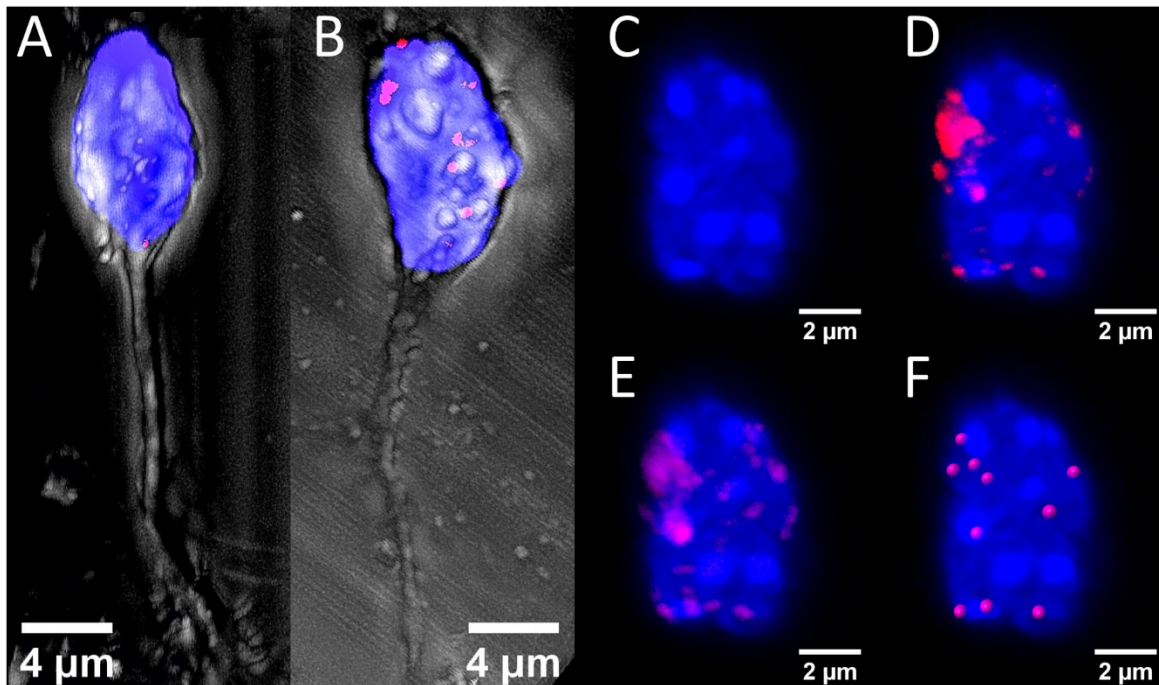
Figure 14. HP1γ and L1 fragments are present in cerebellar granule cells in close proximity. 3D image from a stack of immuno-fluorescence staining images for DAPI (blue) and PLA signal (red). A: Phase contrast and fluorescence information of a L1 -/y CGC. B: Phase contrast and fluorescence information of a L1 +/y CGC.

C: Close-up of the masked nucleus of a L1 +/y CGC (signal inside of the nucleus was filtered from the rest of the fluorescence outside of the nucleus) D: PLA signal found inside of the nucleus. E: Colocalization between the PLA signal (red) and the DAPI signal (blue). F: Spheres of 0.4μm of diameter representing the areas of PLA signal. Representative images were taken, the experiments were done in 12 CGCs (n=2).

Results showed that HP1γ and L1 fragments were close enough for the close proximity reaction to take place (Figure 14.). I could observe that the interaction of these two molecules is found in different places in the nucleus and that the number of PLA dots varies. I also observed singular occurrences of false-positive signals in L1 -/y CGCs, mostly near the axon but not inside the nucleus. On average, L1 +/y CGCs had 12 dots per nuclei (results not shown; 12 images from 2 experiments).

It would have taken an enormous amount of time to obtain single-cell 3D images to quantify the PLA signal in hundreds of cells. Therefore, I chose to use 2D imaging and automatized spot counting using ImageJ (Figure 15.). A transgenic mouse line was generated by the CRISPR gene-editing technique, carrying the L1 R687A point mutation (L1/R687A). This mouse line expresses L1 with a mutation in the MBP cleavage site responsible for the generation and nuclear import of the sumoylated 70 kDa transmembrane L1 fragment. This mutation will give a better insight into the effect of the mutation on the transport of L1 fragments inside of the nucleus.
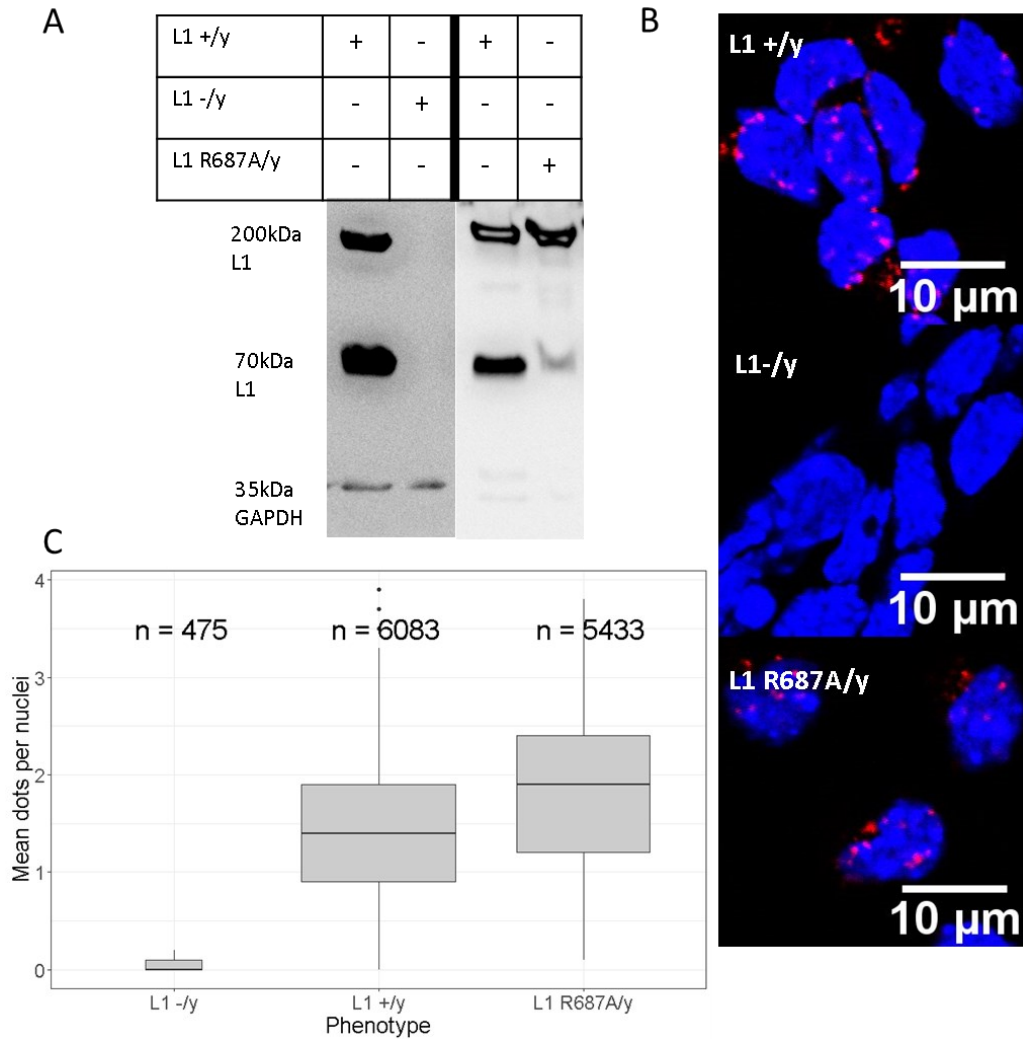
Figure 15. Quantification of the close proximity of HP1γ and L1 fragments in the nucleus of cerebellar granule cells. A: Western blot analysis of whole-brain soluble protein extract from 7-day-old littermates L1 +/y, L1 -/y and L1 +/y, L1 R687A/y mouse brains. The whole-brain extract was separated on SDS-PAGE, blotted and western blots were probed with L1 and GAPDH antibodies (L1 -/y does not give any signal for the L1 antibody). B: Immunofluorescence staining images for DAPI (blue) and PLA signal (red) of CGCs from L1 +/y, L1 -/y, and L1 R687A/y mice. C: Boxplots depicting the mean number of dots per nuclei per image for CGCs from L1 +/y, L1 -/y, and L1 +/y, L1 R687A/y. The regions in the image where DAPI was present were analyzed for circular red dots using the "Analyze Particles" feature of ImageJ. Only dots inside the area of the nucleus with an intensity superior to the cutoff were counted. The total DAPI area was divided by the area of a nucleus

to get the number of nuclei per image (237 images of L1 +/y, 32 images of L1-/y, and 176 images of L1 R687A/y CGCs from 3 independent experiments).

CGCs from L1 +/y mice have at least 1 dot in the nucleus representing close proximity between the L1 fragment and HP1γ (Figure 15.). This number varies from the 3D results (12 dots per nucleus) as the criteria to count dots inside of the nucleus was set not to count dots on the periphery of the nuclei and also not those with low intensity. As seen in Figure 14. the dots were found in different regions and depths of the nucleus. Therefore, it is less probable to find all of them in just one plane.

CGCs from L1 -/y mice do show background signal but the difference between L1 -/y and L1 +/y is clear. L1 R687A/y mice showed as much interaction signal as the L1 +/y what suggests that the fragments present in the nucleus interacting with HP1γ do not depend on the nuclear import of the sumoylated 70 kDa L1 fragment but maybe in other fragments such as the 28 kDa L1 fragment (Riedle et al., 2009). The 28kDa fragment was observed after stimulating CGCs from L1 +/y mice with Fc-fusion protein (Fc: fragment crystallizable region or tail region of an antibody) containing the extracellular domain of the mouse L1 protein for 1h (Figure 16.).

L1 +/y	+	+	+	+	+	+
L1 557 ab	-	-	+	+	-	-
L1-Fc	-	-	-	-	+	+

200 kDa
L1

70 kDa
L1

28 kDa
L1
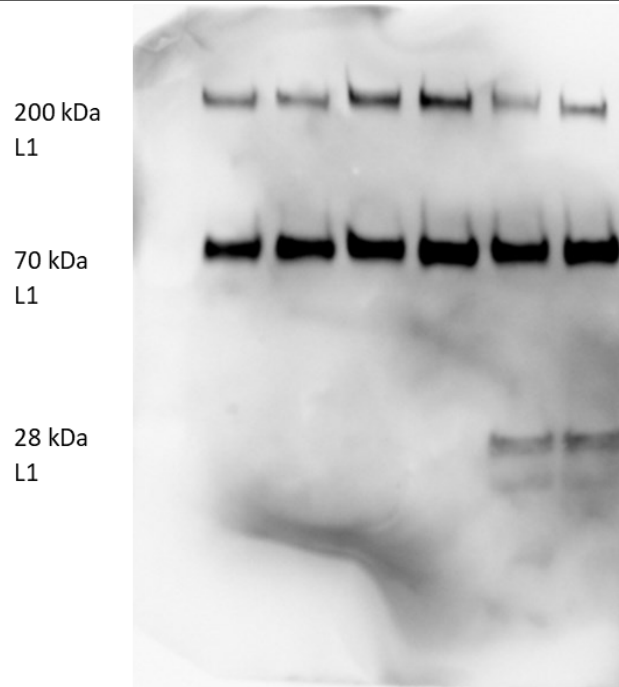
Figure 16. L1 in cerebellar granule cells from 7-day-old wild-type mice. Cerebellar granule cells (CGCs) from 6- to 8-day-old mice at a concentration of 1.10^6cells/mL were cultured in serum-free X1-medium for 24h. Cells were then stimulated with 50μg/mL of the L1 557 antibody or with 25μg/ml of the mouse L1 extracellular domain-Fc for 1h. CGCs were then subjected to subcellular protein fractionation using a commercial kit. Membrane, cytoplasm, chromatin-bound, and soluble nuclear fractions were obtained. The different fractions were separated on SDS-PAGE, blotted, and western blots were probed with the monoclonal L1-C2 antibody.

5.1.4. Possible effects of the interaction between the HP1γ CSD and the L1- ICD in gene regulation

One of the observations previously noted (5.1.3.Figure 13.) in immunofluorescence stainings of HP1γ in CGCs was that HP1γ appeared to be more abundant in the nucleus of CGCs of L1- /y mice in comparison

to L1 +/y mice. Therefore, I decided to quantify the volume of HP1γ inside of the nucleus of CGCs from L1 -/y and +/y mice and to also quantify the protein levels of HP1γ by doing a western blot of whole-brain lysate from L1 -/y and +/y mice (Figure 17.).
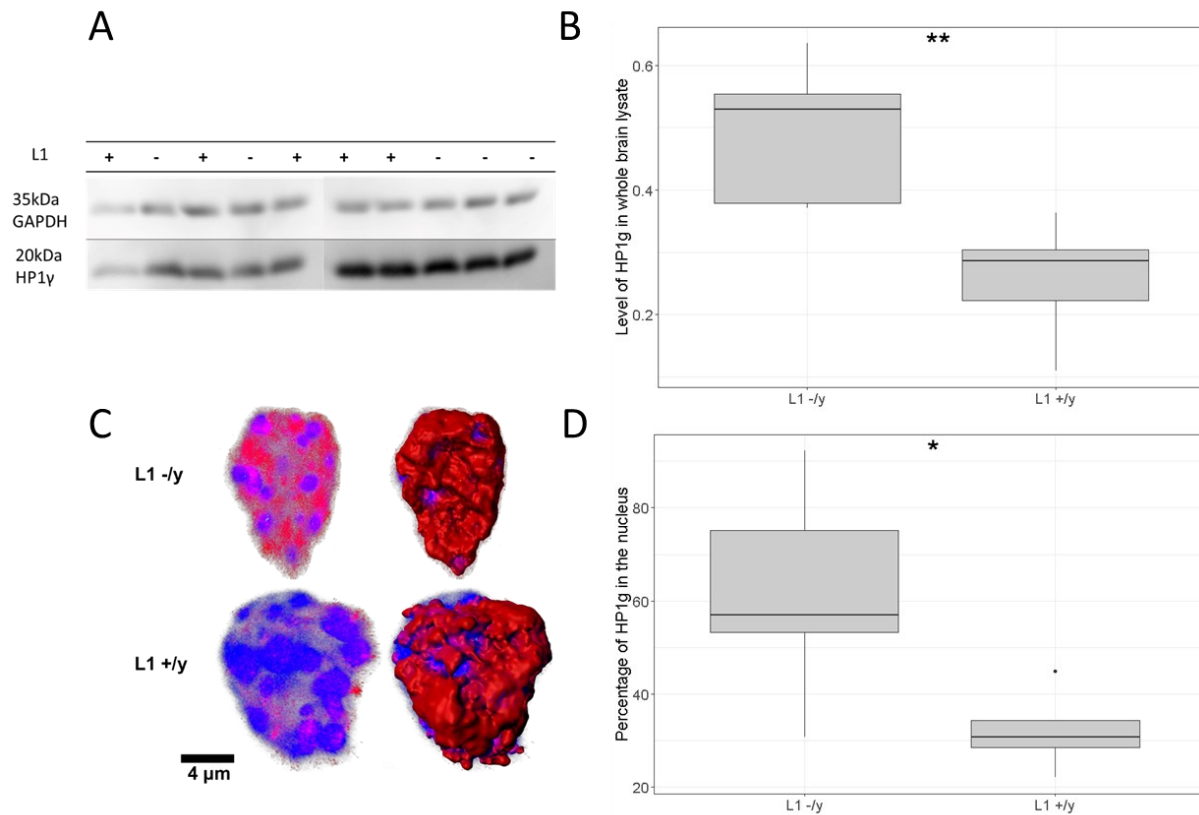
Figure 17. Higher levels of HP1γ are found in the nucleus of cerebellar granule cells of L1 -/y mice in comparison to cells from L1+/y mice. A: Western blot analysis of whole-brain soluble protein extract from brains of 7-day-old L1 +/y and L1 -/y mice. The whole-brain extract was separated on SDS-PAGE, blotted, and western blots were probed for HP1γ and GAPDH. B: Level of HP1γ in L1 +/y and L1 -/y whole-brain extracts (n=5 and T-test, ** p-value = 4e^{-05}). C: Representative images of 3D Immunofluorescence staining images of nuclei from CGCs of L1 -/y and L1 +/y mice. The volumes were built from image stacks using the Imaris software. DAPI staining is shown in blue and HP1γ in red; the HP1γ signal is represented as a solid volume. D: Quantification of HP1γ signal (red) inside of the nucleus (n=5 and T-test, * p-value = 2.5e^{-04}).

The levels of HP1γ were lower in whole-brain extracts and nuclei of CGCs of L1 +/y mice compared to levels in whole-brain extracts and nuclei of CGCs of L1 -/y mice (Figure 17. A & B). HP1γ seems to be upregulated or to be more stable and thus more abundant in brains of mice lacking L1. Furthermore, in CGCs the intensity of the HP1γ signal is higher in L1 -/y cells than in L1+/y cells (Figure 17. C & D). The distribution of HP1γ in the nucleus is the same in both L1 -/y and L1 +/y cells (Figure 17. C). L1 could directly or indirectly influence the levels of HP1γ in the whole brain and more precisely in CGCs of 7-day-old mice. One way in which L1 could influence HP1γ protein levels would be if L1 affects the transcription of the CBX3 mRNA (gene coding for HP1γ). To show how this interaction could affect gene transcription I designed an experiment using transcriptome information contained in public databases repositories such as Gene Expression Omnibus (GEO) (Figure 18.).

A

B
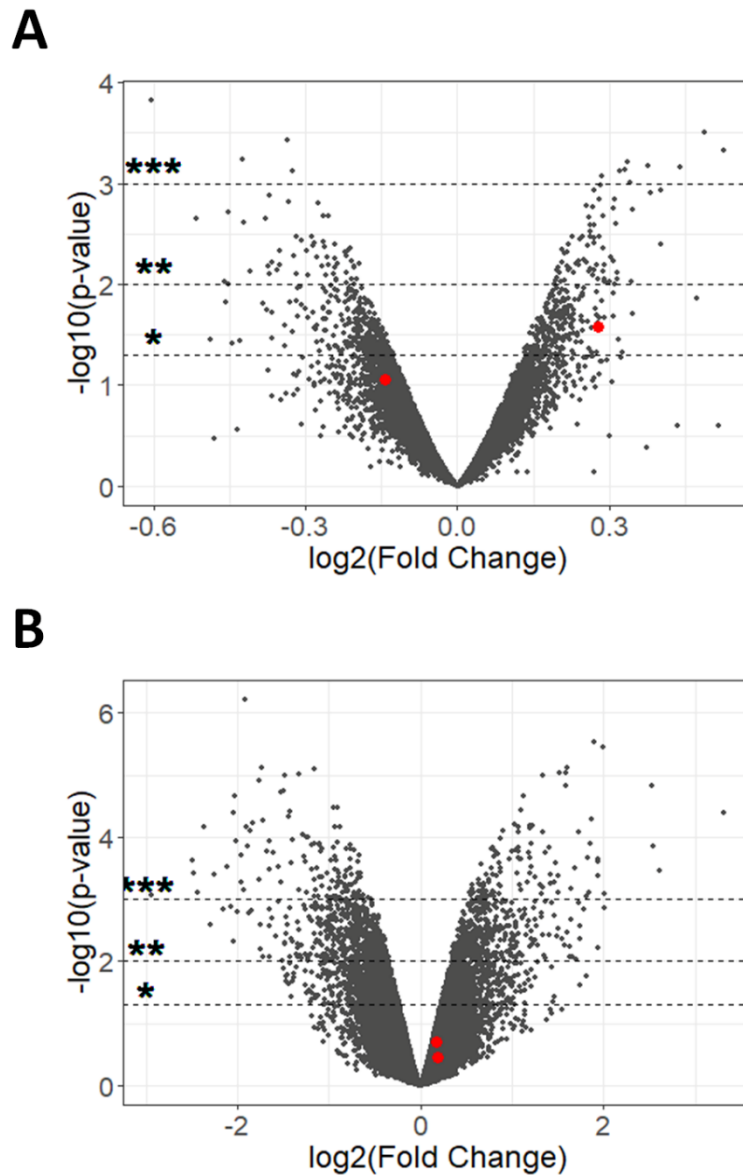
Figure 18. Gene expression in two Gene Expression Omnibus (GEO) datasets. A: Gene expression dysregulation after knocking out the L1 gene. Volcano plot of gene expression, showing statistical significance (-log10(p-value)) versus magnitude of fold change (log2(fold change)). In red, reporters corresponding to the CBX3 gene are depicted. Dotted lines show statistical significance. The microarray experiment was performed extracting RNA from the cerebellum from 11-day-old L1 -/y and L1+/y littermate mice (n=3) (Tapanes-Castillo et al., 2010). B: Gene expression dysregulation after knocking down the CBX3 gene. Volcano plot of gene expression, showing statistical significance (-log10(p-value)) versus magnitude of fold change (log2(fold change)). In red, reporters corresponding to

the L1 gene are shown. Dotted lines indicate statistical significance (* p-value = 0.05, ** p-value = 0.01, *** p-value = 0.001). The microarray experiment was performed in pre-induced pluripotent stem cells (iPSCs) where the CBX3 gene was knocked down using a siRNA transfection (n=3) (Sridharan et al., 2013).

Two gene reporters 160272_at and 100405_at (Figure 18. A), corresponding to the CBX3 gene can be found in the microarray experiment of extracted RNA from the cerebellum of 11-day-old mice L1 -/y in comparison to their L1 +/y littermates (Tapanes-Castillo et al., 2010). The number of target sequences matching 160272_at is statistically increased in L1 -/y cerebellum in comparison with L1 +/y cerebellum, while the number of targets matching 100405_at is not of statistical significance. The fold change of 160272_at is about 1.21 (2^0.2776053) which does not directly correlate with the two-fold protein levels of HP1γ found in L1 -/y mice compared with L1 +/y mice. Other factors might explain the higher levels of HP1γ in L1 -/y mice but the increased expression of CBX3 mRNA in L1-deficient mice could contribute to the higher levels of HP1γ found in L1 -/y mice.

It would be interesting to see if the lack of CBX3 could affect the expression of L1 in mice but as we do not have CBX3 deficient mice at our disposition I decided to study literature data on the transcriptome of cells knocked down (KD) for CBX3 (Sridharan et al., 2013). Knock-out of CBX3 in mice has been shown to lead to premature death.

Two reporters 1421958_at and 1450435_at, corresponding to the L1 mRNA can be found in the expression profiling by array experiment (Figure 18. B). Both of them are not statistically significantly dysregulated. In the case where CBX3 is knocked down, there seems to be no effect on the L1 expression in pre-induced pluripotent stem cells (pre-iPSCs), but the interaction between L1 and HP1γ could affect the regulation of other genes.

Dysregulated genes can be observed in the gene expression comparison between L1 -/y and L1 +/y and CBX3 KD pre-iPSCs and non-transfected pre-iPSCs. The interaction between L1 and HP1γ could be involved in the regulation of gene expression. If the same gene is dysregulated in the same direction when one of the partners is missing, it could be a candidate for gene regulation via the interaction of L1 with HP1γ.
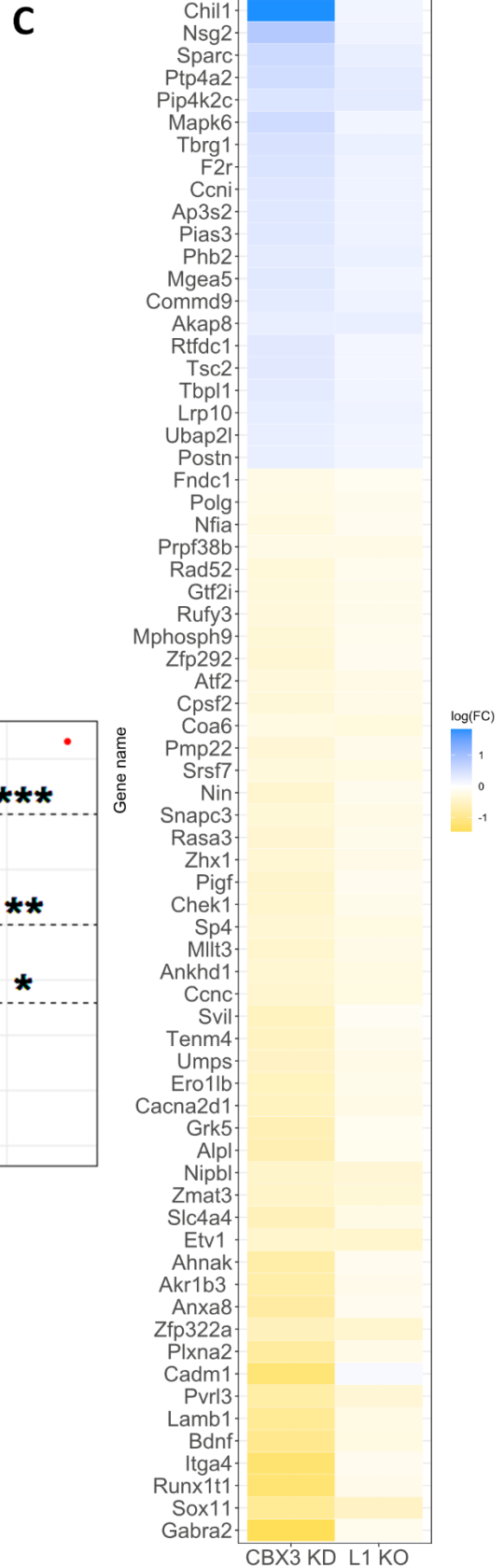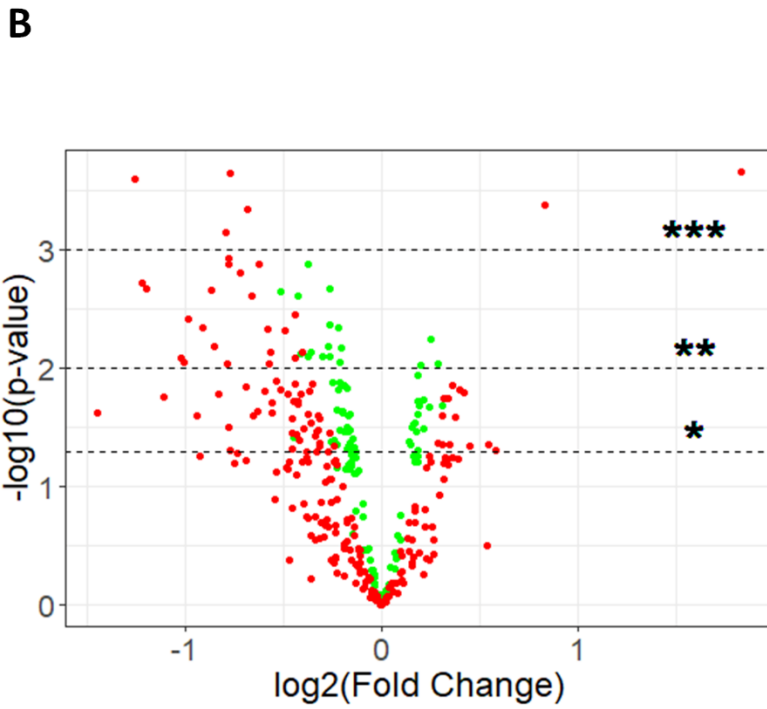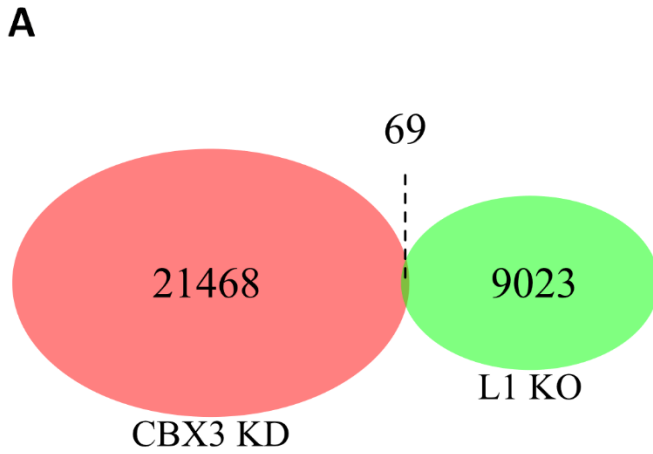
Figure 19. Commonly dysregulated genes found in both datasets, L1 -/y compared to L1 +/y mouse cerebella and CBX3 knockdown pre-iPSCs compared to non-transduced pre-iPSCs. A: Venn diagram representing each dataset in a circular form. The intersection of both circles represents common genes regulated in the same direction. B: Volcano plot of gene expression, showing statistical significance (-log10(p-value)) versus magnitude of fold change (log2(fold change)). Depicted in green are the reporters from the L1 dataset, and in red, the reporters from the CBX3 dataset. Dotted lines indicate statistical significance (* p-value = 0.05, ** p-value = 0.01, *** p-value = 0.001). C: Heatmap of the 69 up- (blue) and down-regulated (yellow) genes in both datasets.

69 genes (Figure 19.) were found to be dysregulated in the same direction for both datasets, L1 KO (L1 -/y) and CBX3 KD (CBX3 siRNA). 21 genes are found to be upregulated and 48 to be dysregulated. The down-regulated gene reporters were found to be more statistically significant and with a higher fold change than the upregulated genes. Although the fold change in the L1 dataset is not as strong as in the CBX3 dataset we can observe a similar pattern of dysregulation. To find similar biological functions that are affected by the dysregulation of genes when either L1 is missing or HP1γ is present in less quantity. I used the Database from Gene Ontology (GO) to do the correlation between the dysregulated genes and their biological function (Figure 20.).
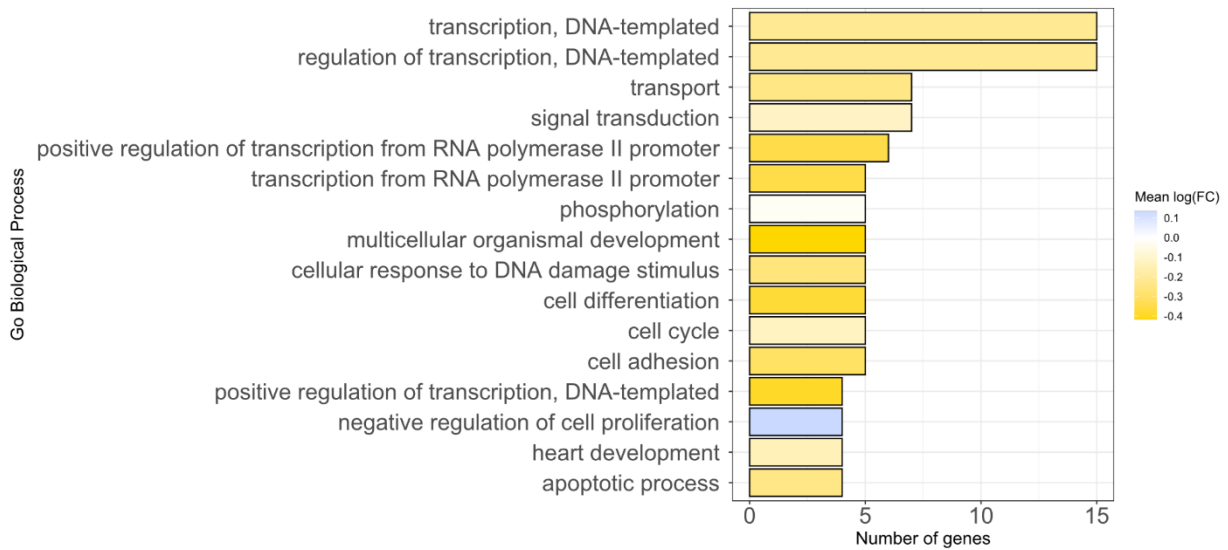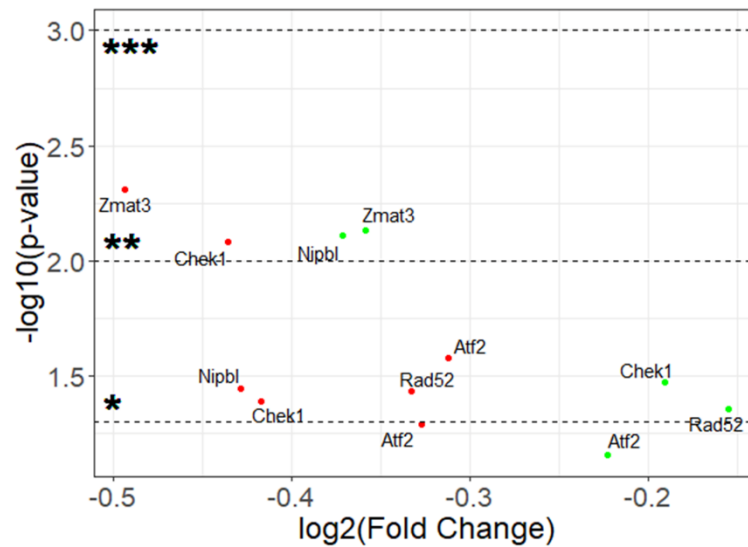
Figure 20. Biological processes related to dysregulated genes found in the cerebellum of L1 -/y mice compared to their L1 +/y littermates and CBX3 KD pre-iPSCs compared to non-transduced pre-iPSCs. A: Bar plot depicting the number of different dysregulated genes having a biological process in common (4 different genes within a biological function group as a cut-off). Shown in yellow are down-regulated genes and in blue up-regulated genes. B: Dot plot of gene expression from genes involved in the cellular response to DNA damage, showing statistical significance (-log10(p-value)) versus magnitude of fold change (log2(fold change)). Shown in green are the reporters from the L1 dataset, and in red are the

reporters from the CBX3 dataset. Dotted lines show statistical significance (* p-
value = 0.05, ** p-value = 0.01, *** p-value = 0.001).

After grouping the dysregulated genes by their gene ontology biological process, a list was made with 14 biological processes related to down-regulated genes (Figure 20. A). For up-regulated genes, one biological process related to the negative regulation of cell proliferation was found and the process of phosphorylation was formed by a mixture from up-and down-regulated genes.

From the 14 biological processes related to down-regulated genes, I chose to study the cellular response to DNA damage genes as it has been previously found that the L1- ICD was involved in the regulation of the DNA damage checkpoint response in glioblastoma stem cells (Cheng et al., 2011). The five down-regulated genes related to the cellular response to DNA damage found to be statistically significant were activating transcription factor 2 (Atf2), checkpoint kinase 1 (Chek1), nipped-B-like protein (Nipbl), DNA repair protein Rad52 homolog (Rad52), and zinc finger matrin-Type 3 (Zmat3) (Figure 20. B). All genes have at least one statistically significant reporter in one dataset. With the list of dysregulated genes, I decided to conduct a quantitative PCR to follow gene expression of these genes in CGCs from 7-day-old L1 +/y and L1 -/y mice.
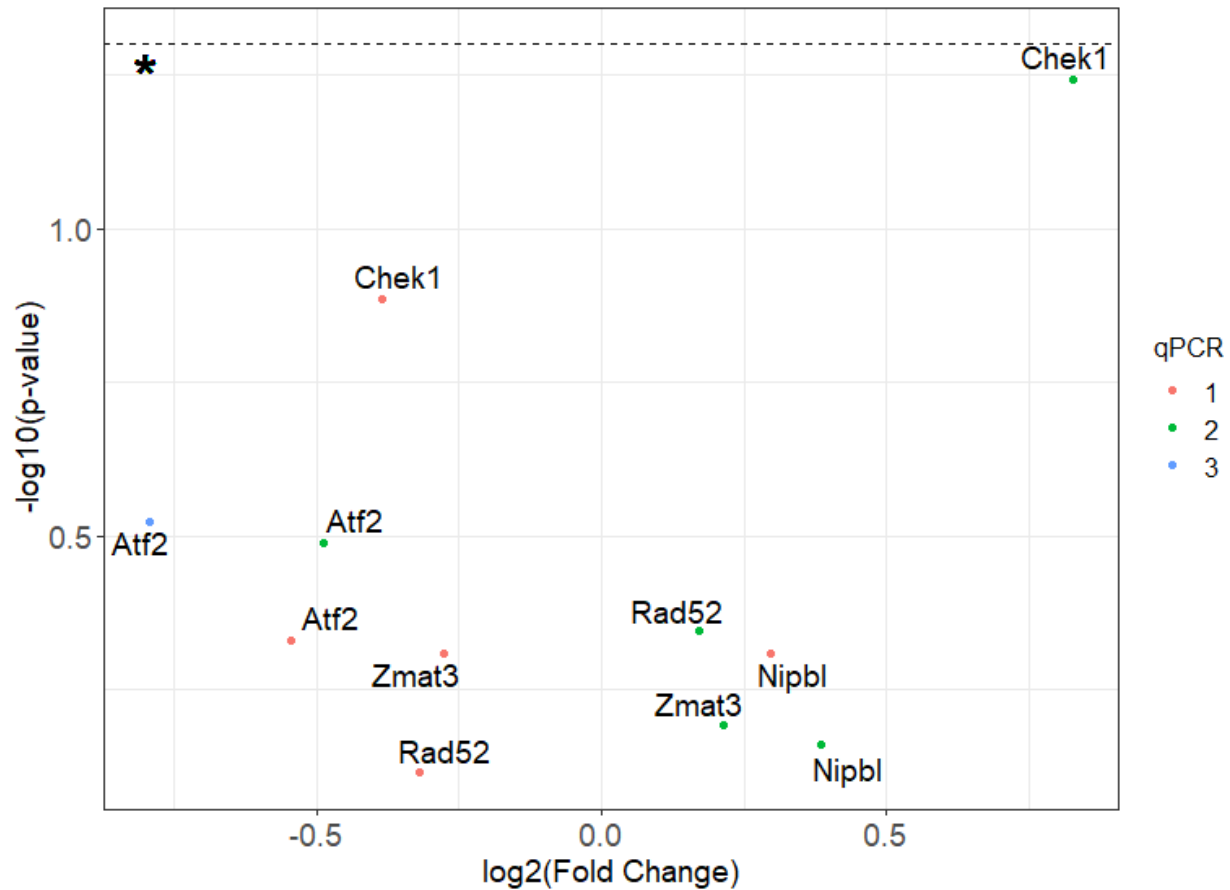
Figure 21. Quantitative PCR results for the DNA damage stimulus process-related genes from 3 experiments. Dot plot of gene expression from genes involved in the cellular response to DNA damage stimulus, showing statistical significance (-log10(p-value)) versus magnitude of fold change(log2(fold change)). In red (qPCR 1): results from mRNA extracted and retrotranscribed in cDNA from CGCs of four 7-day-old littermate mice (2 L1 -/y and 2 L1 +/y). In green (qPCR 2): results from mRNA extracted and retrotranscribed in cDNA from CGCs of two 7-day-old littermate mice (1 L1 -/y and 1 L1 +/y). In blue (qPCR 3): results from mRNA extracted and retrotranscribed in cDNA from CGCs of two 7-day-old littermate mice (1 L1 -/y and 1 L1 +/y). The Dotted line shows statistical significance (* p-value = 0.05).

From the five down-regulated genes related to the cellular response to the DNA damage stimulus process, only Atf2 showed a constant down-regulation through all the experiments (Figure 21.). Nipbl showed upregulation, and Chek1, Rad52, and Zmat3 showed both up and down-regulation. In the third experiment, only the Atf2 primer was used from the 5 genes related to the DNA damage stimulus process.

Efforts were made to generate transduced mutant CGCs by either silencing the CBX3 gene in L1 +/y cells, by expressing L1, or the PXVXL deficient version of L1 in L1 -/y cells. Silencing 50% of the total HP1γ in the CGCs took more than 3 days to achieve (Figure 22.) which did not allow to do experiments such as migration and neurite outgrowth. The addition of full-L1 or the PXVXL deficient version of L1 in L1 -/y cells by transduction was not successful in CGCs as no fragments of L1 were not observed in these cells.
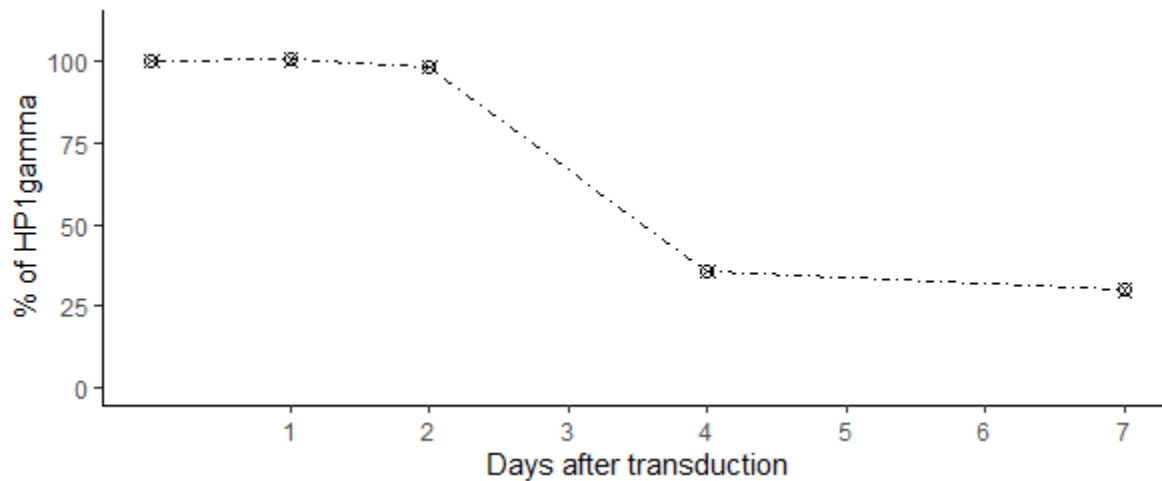
Figure 22. AAV-RNAi-mediated silencing of HP1γ in cerebellar granule cells. CGCs from 6- to 8-day-old mice at a density of 1.10^6 cells/ml were cultured in serum-free X1-medium for 24 h. Cells were then transduced using adeno-associated viruses containing a plasmid coding for the HP1γ shRNA with a multiplicity of infection (MOI) of 3000. The shRNA sequence used was the same as in Oshiro et al., 2015.

In our group we created a genetically modified mouse, with the help of the CRISPR gene-editing technique, having a mutation in L1 inside the motif binding to importin alpha KKXK. In these animals, the transport of L1 to the nucleus should be abolished. The mutant mouse line L1 K1147R was successfully obtained but the mutant mice did not show any difference to the L1 +/y littermates while comparing the interaction of L1 with HP1γ inside of the nucleus.

5.2.   Nuclear proteins carrying the glycan epitopes HNK-1 and/or Lewis^X and receptors

5.2.1.   Detection of nuclear proteins carrying the glycan epitopes HNK-1 and/or Lewis^X

Glycan recognition is very important for normal brain function during development and in the adult. It has been shown that protein-glycan interactions in the nervous system are involved in cell migration, neurite outgrowth and fasciculation, synapse formation and stabilization, and modulation of synaptic efficacy. Although many protein receptors for glycans on the cell membrane are known, only recently glycans have been detected in the nucleus of neurons and their receptors are mainly unknown as well as their roles in nuclear functions. To identify and characterize carriers of the glycans HNK-1 (also known as CD57 or LEU7) and Lewis^X (also known as CD15 or SSEA-1) in the nucleus of neural cells, brains of adult mice were homogenized and a soluble nuclear protein fraction was obtained by nuclear isolation using sucrose gradient centrifugation. Nuclear proteins were then subjected to SDS-PAGE and subsequent western blot analysis using HNK-1 and Lewis^X specific antibodies. Protein bands of approximately 200-350 kDa were detected with HNK-1 and LewisX antibodies (Figure 23.).
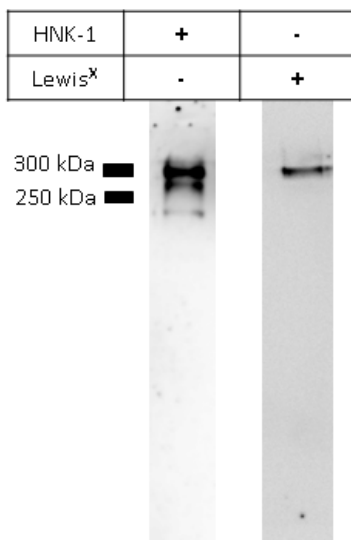
| HNK-1 | + | - |
| Lewis^X | - | + |

Figure 23. HNK-1 and Lewis^X carriers in the nucleus of adult mice. HNK-1 and Lewis^X glycan carrier proteins are high molecular weight proteins. Homogenized adult mouse brains were ultra-centrifuged in a sucrose gradient. The nuclear phase was extracted and centrifuged to separate the soluble protein fraction. The soluble

nuclear protein extract was separated on SDS-PAGE, blotted, and western blots were probed with HNK-1 and Lewis^X antibodies.

Previous studies have shown that a 70 kDa fragment of L1 (L1-70), which includes part of the extracellular domain with several glycosylation sides, can be imported into nuclei. Furthermore, levels of full-length L1 and the L1-70 fragment are altered when cells expressing L1 are stimulated with the L1 557 antibody (Lutz et al., 2014, 2012). In cultured cerebellar neurons of wild-type mice, L1-stimulation increases the amount of L1-70 in the nucleus allowing its detection in the nucleus. Thus, it is deemed interesting to stimulate neurons with 557 L1-antibody and to investigate if this L1-70 fragment carries HNK-1 and/or LewisX epitopes.
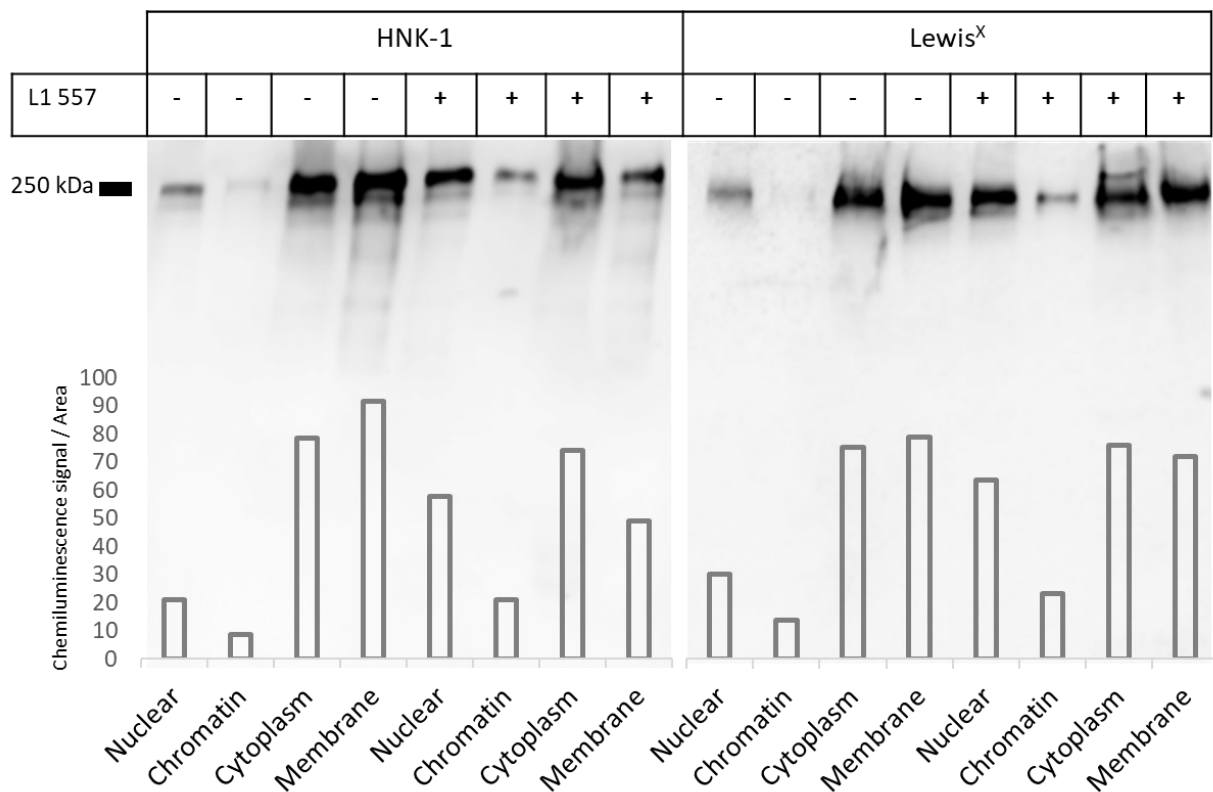
Figure 24. HNK-1 and Lewis^X carriers in cerebellar granule cells from 7-day-old wild-type mice. CGCs from 6- to 8-day-old mice at a density of 1.10^6cells/ml were cultured in serum-free X1-medium for 24h. Cells were then stimulated with 50μg/mL of the L1 557 antibody for 1h. CGCs were then subjected to subcellular protein fractionation using a commercial kit. Membrane, cytoplasm, chromatin-

bound, and soluble nuclear fractions were obtained. The different fractions were separated on SDS-PAGE, blotted, and western blots were probed with HNK-1 and Lewis^X antibodies. A histogram of the chemiluminescence signal per area measured in function of the cell fractions was overlayed to the western blot image (no loading control was used).

At equal cell density, CGCs stimulated with the L1 557 antibody show higher levels of HNK-1 and/or Lewis^X carrier protein(s). A signal at approximately 250-300 kDa was observed for both HNK-1 and Lewis^X (Figure 24.). A possible protein candidate with this molecular weight was R-PTP-zeta, also known as phosphacan (250 kDa).
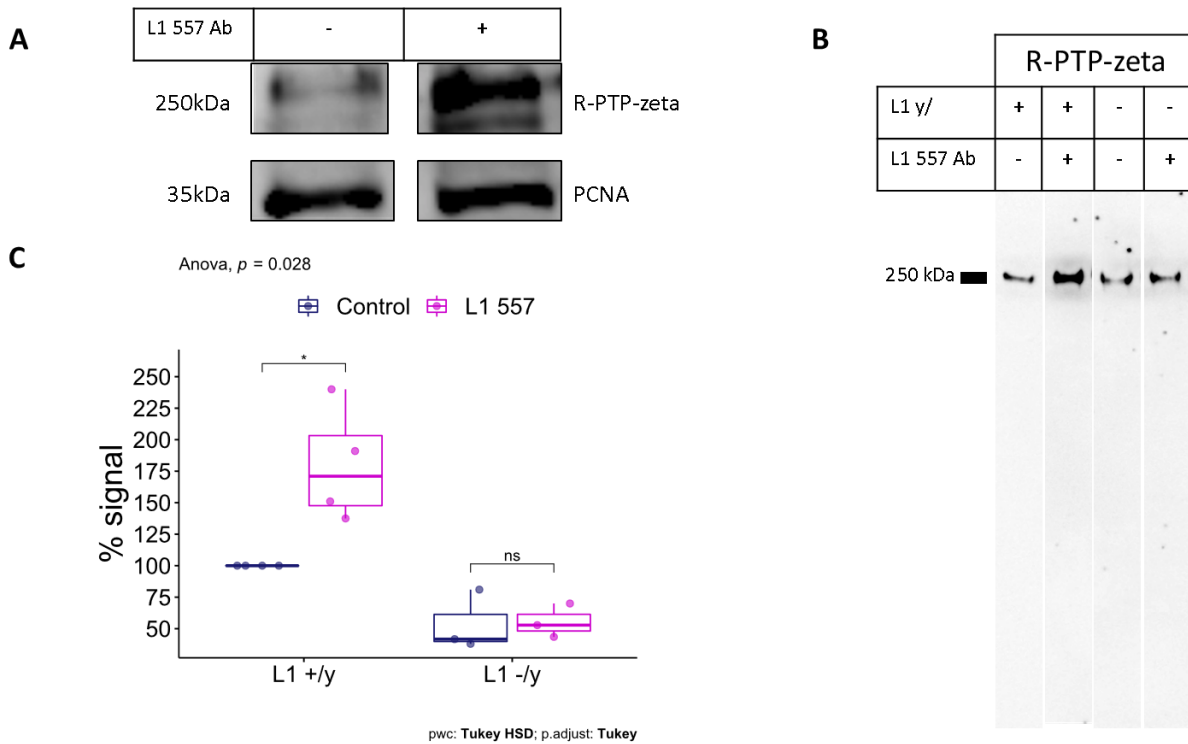
Figure 25. R-PTP-zeta in cerebellar granule cells from 7-day-old mice and L1 557 antibody stimulation. CGCs from 6- to 8-day-old wild-type (L1 +/y) and L1-deficient (L1 -/y) mice at a density of 1.10^6 cells/ml were cultured in a serum-free medium for 24h. Cells were then stimulated with 50 µg/mL of the L1 557 antibody for 1h (L1 557). CGCs were then subjected to subcellular protein fractionation. Membrane, cytoplasm, chromatin-bound, and soluble nuclear fractions were obtained. The different fractions of soluble protein extract were separated on SDS-PAGE, blotted, and western blots were probed with R-PTP-zeta specific antibody. A: Western blot of L1 +/y CGCs nuclear fraction non stimulated (-) and stimulated

with L1 557 antibody (+). As loading control proliferating-cell-nuclear-antigen (PCNA) was used in the nuclear fraction. B: Western blot of L1 +/y and L1 -/y CGCs nuclear fraction non stimulated (L1 557 Ab -) and stimulated with L1 557 antibody (L1 557 Ab +). C: Boxplot depicting the percentage of signal (chemiluminescence) from the nuclear fraction of L1 +/y and L1 -/y CGCs nuclear fraction non stimulated (Control) and stimulated with L1 557 antibody (L1 557). Total signal with R-PTP-zeta antibody was normalized using PCNA as the loading control. For each of the 4 independent experiments non-stimulated L1 +/y signal was set to represent 100 % of the signal (Tukey's honest significance test, * p-value = 0.028)

R-PTP-zeta protein levels in the nucleus of CGCs are ~ 50 % lower in L1 -/y than in L1 +/y CGC nuclei. The protein levels in L1 -/y CGCs stimulated with the L1 557 antibody do not change in contrast to L1 557 stimulated L1 +/y CGCs in which R-PTP-zeta protein levels in the nucleus of CGCs increase ~ 175%. Therefore the quantity of R-PTP-zeta in the nucleus of CGCs depends partially on L1.

5.2.2. Bioinformatic approaches to find and elucidate HNK-1 binding regions

R-PTP-zeta being the main carrier of both glycan epitopes HNK-1 and Lewis^X (5.2.1.Figure 25.) and being partially regulated by L1 in CGCs, it is interesting to know which nuclear receptors are known to interact with these epitopes and if unknown nuclear receptors can be found (Table 2).

Table 2. Receptors for HNK-1 and Lewis^X. List of receptors found in the literature for the glycan epitopes HNK-1 and Lewis^X. Evidence of interaction is based upon crystal structures from the protein data bank (PDB).

Protein name	Gene name	Receptor for	Evidence of interaction
E-selectin	SELE	Lewis^X	PDB: 4CSY
P-selectin	Selplg	Lewis^X	PDB: 1G1R
Mannose-binding protein A	Mbl1	Lewis^X	PDB: 2KMB
Collectin-12	Colec12	Lewis^X	PDB: 2OX9
Cadherin-2	CDH2	HNK-1	Not Available
Laminin subunit alpha-1	Lama1	HNK-1	Not Available

| High mobility group protein B1 | Hmgb1 | HNK-1 | Not Available |
| High mobility group protein B2 | Hmgb2 | HNK-1 | Not Available |

Lewis^X receptors bind the glycan epitope via their C-type lectin domain. It is shown in the literature and in crystallography data that Lewis^X binds to the receptors in a similar way. There is a great amount of literature describing the importance of the HNK-1 epitope in the development of the nervous system but in contrast, almost no information is available on how non-lectin HNK-1 receptors bind the glycan epitope. Therefore, I tried to identify possible regions of similarity between the non-lectin HNK-1 receptors, cadherin-2, laminin subunit alpha-1, and high mobility group proteins B1 and B2, that may be involved with the binding of the HNK-1 epitope. Similar proteins to the 4 HNK-1 receptors were found via a BLAST search setting. These datasets will be families of each of the HNK-1 receptors (Figure 26.).
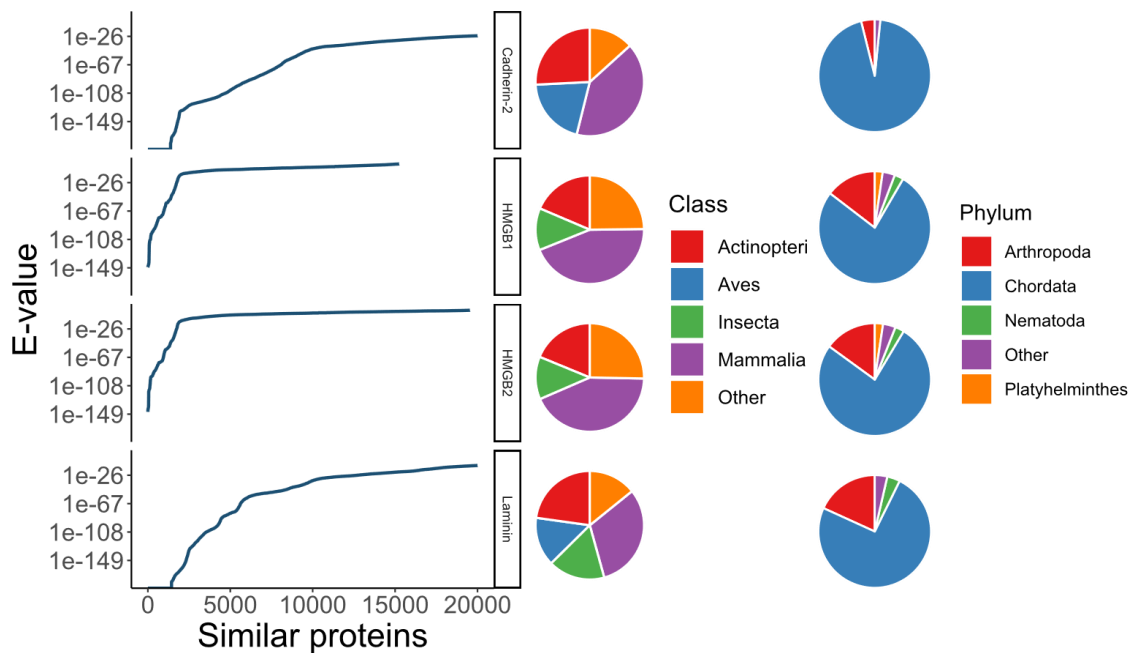
Figure 26. HNK-1 receptors proteins family's similarity and their taxonomy (class and phylum). The line plot depicts on the x-axis the ~20 000 similar proteins to each of the HNK-1 receptors in function to their expected value probability (the lower the better). The probability represented by the E-value is the given number of expected hits of similar quality. The circle charts show the percentages of each class and phylum in the different families.

Results from HMGB-1 and 2 were almost identical, as expected for two paralogs. HMGB-2 similar proteins having a larger number of similar proteins (HMGB1 family, 15 000 similar proteins and HMGB2 family, 20 000, see Figure 26.) was a better dataset to represent the HMGB family. Most of the similar proteins across the HNK-1 receptor families belong to the animal kingdom. In the three families, the mammals class forms the majority of similar proteins alongside other chordate classes such as bony fish and birds. The other phylum's represented in minority are tetrapods, nematodes, and Platyhelminthes. These different organisms share the characteristic of having a nervous system. This was reassuring for the studies to follow as I tried to find evolutionarily conserved regions among the 3 families. The ~20 000 similar proteins were aligned and reduced to 5 000, resulting in three datasets of 5 000 similar proteins. The three families were merged into a dataset of 15 000 proteins and were aligned (Figure 27.).
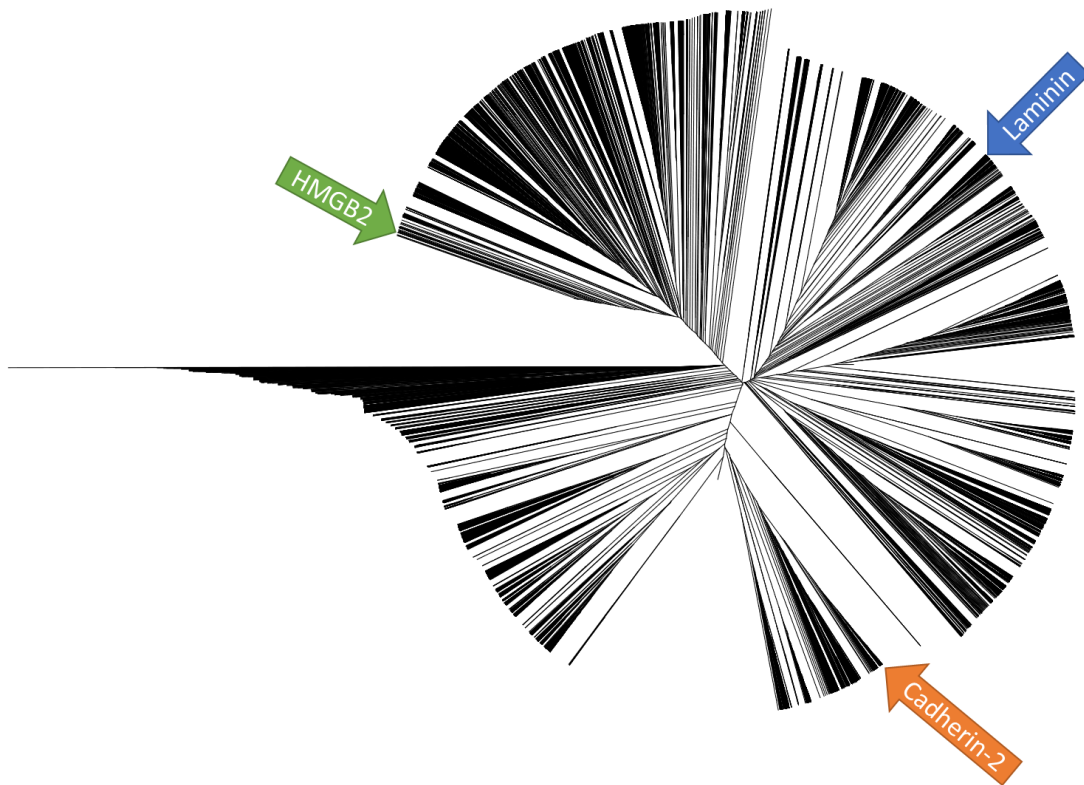
Figure 27. Radial phylogram of the aligned three HNK-1 receptor families of similar proteins. The guided tree was based on the distance matrix generated from the alignment of the 15 000 proteins (5 000 per receptor family) similar to the HNK-1 receptors cadherin-2, laminin, and HMGB2.

The guide tree showed that the three families of the HNK-1 receptors do not share any evolutionary similarity (Figure 27.). They belong to three different branches that diverge from the starting node. As no evolutionary relationship between the receptors was found the search for a common amino acid secondary structure or an amino acid motif was the next step. The three families had different maximum values of the number of expected hit scores (E-values, Figure 28.). In order to set a common second cut-off, I decided to remove the similar proteins where the E-value was greater than the lowest maximum E-value from the three families.

Protein name	Sequence length	*n homologues*
Cadherin-2	906	14 128
HMGB-2	199	1573
Laminin (subunit α-1)	3083	11 999
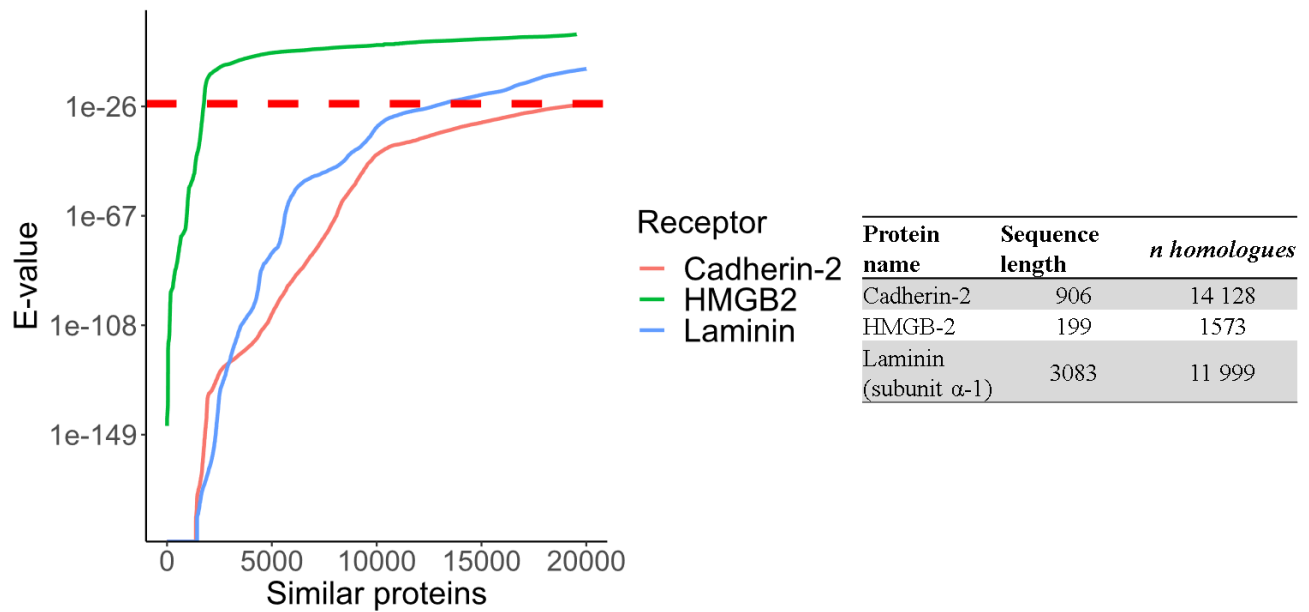
Figure 28. Cut-off applied to the datasets of the three HNK-1 receptor families of similar proteins. The line graph on the left depicts the datasets containing ~20 000 similar proteins to each of the HNK-1 receptors in function to their given number of expected hits of similar quality (E-value), which is the probability of the similar protein being found just by chance (the lower the better). The red dotted line marks the 1e-25 cut-off applied. The table on the right shows the remaining similar proteins (homologs) for each HNK-1 receptor family.

I decided to use all similar proteins with an E-value equal to or less than $1e^{-25}$ after noticing that the cadherin-2 family had its maximum E-value around this value whereas the other two families had higher E-values. This cutoff was selected for all 3 families independently on how it will affect the final number of similar proteins (Figure 28.) to have a general maximum E-value across families. Also, synthetic constructs and

protein tags were removed from the datasets to avoid false-positive similar domains. After applying the cut-off, the new datasets were aligned (Figure 29.).

Cadherin-2

HMGB2

Laminin

A

Domain
- LamNT
- EGF_Lam
- LamB
- Laminin_I
- LamG

B

- Nonpolar
- Acidic
- Basic
- Polar
- Other

C

Cluster
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
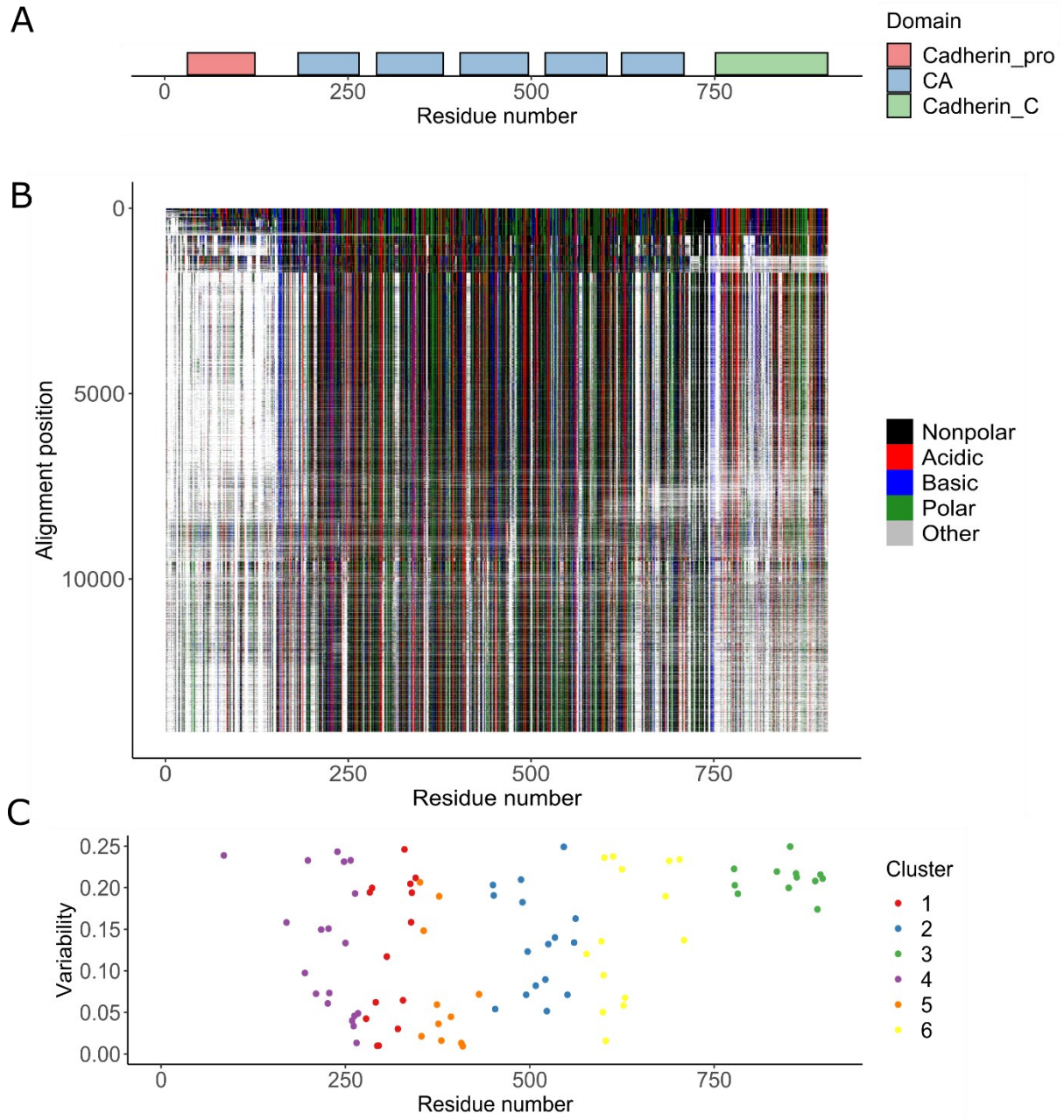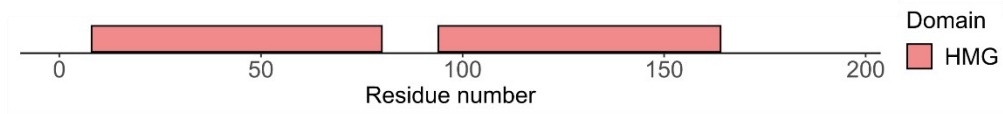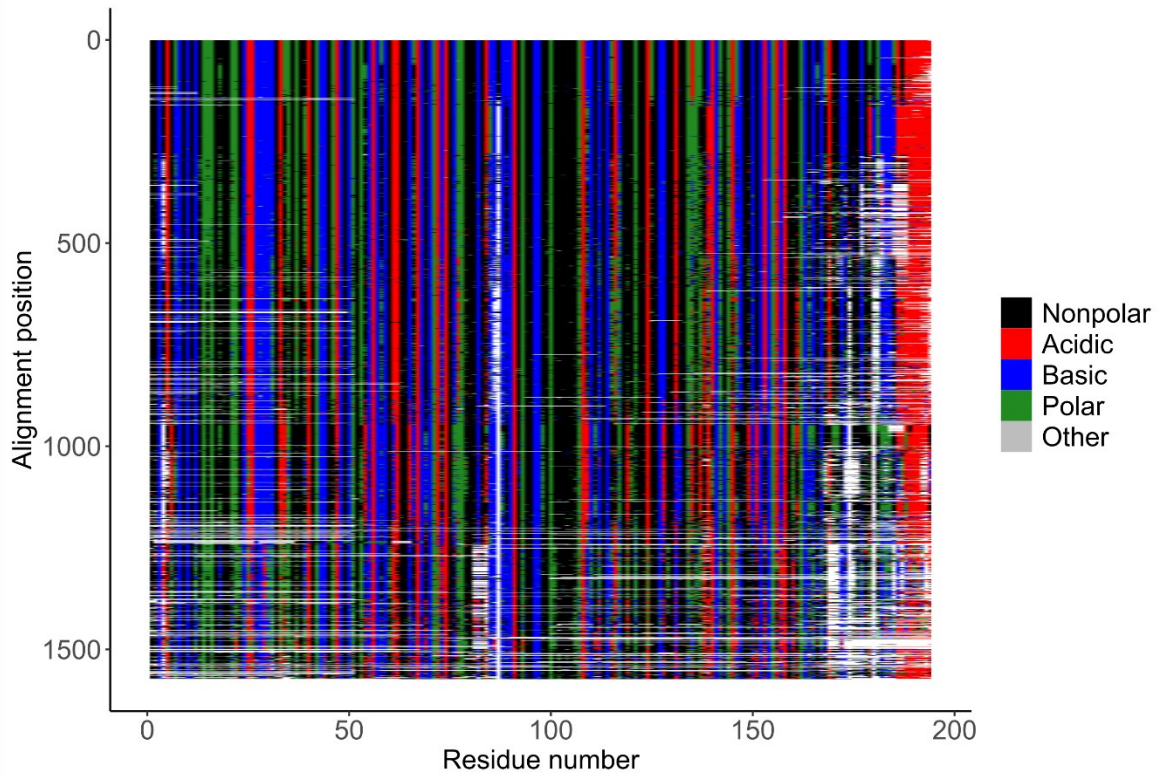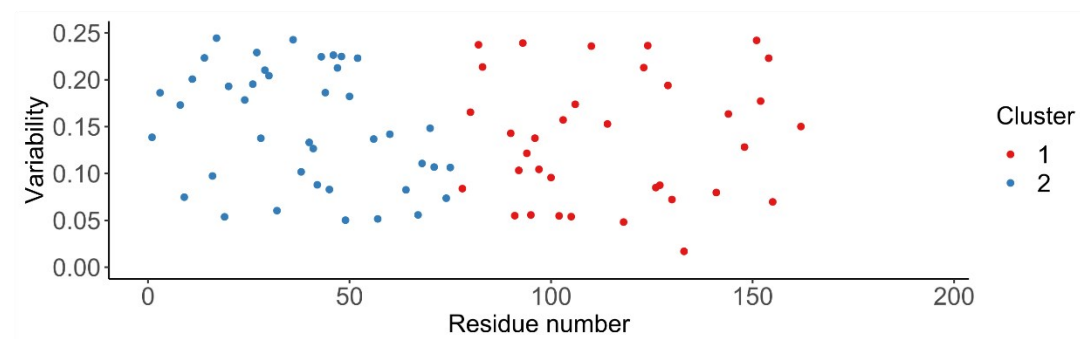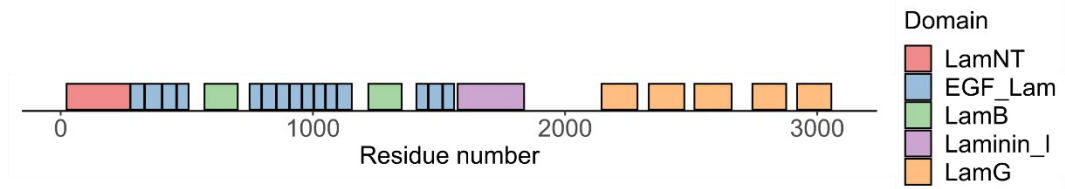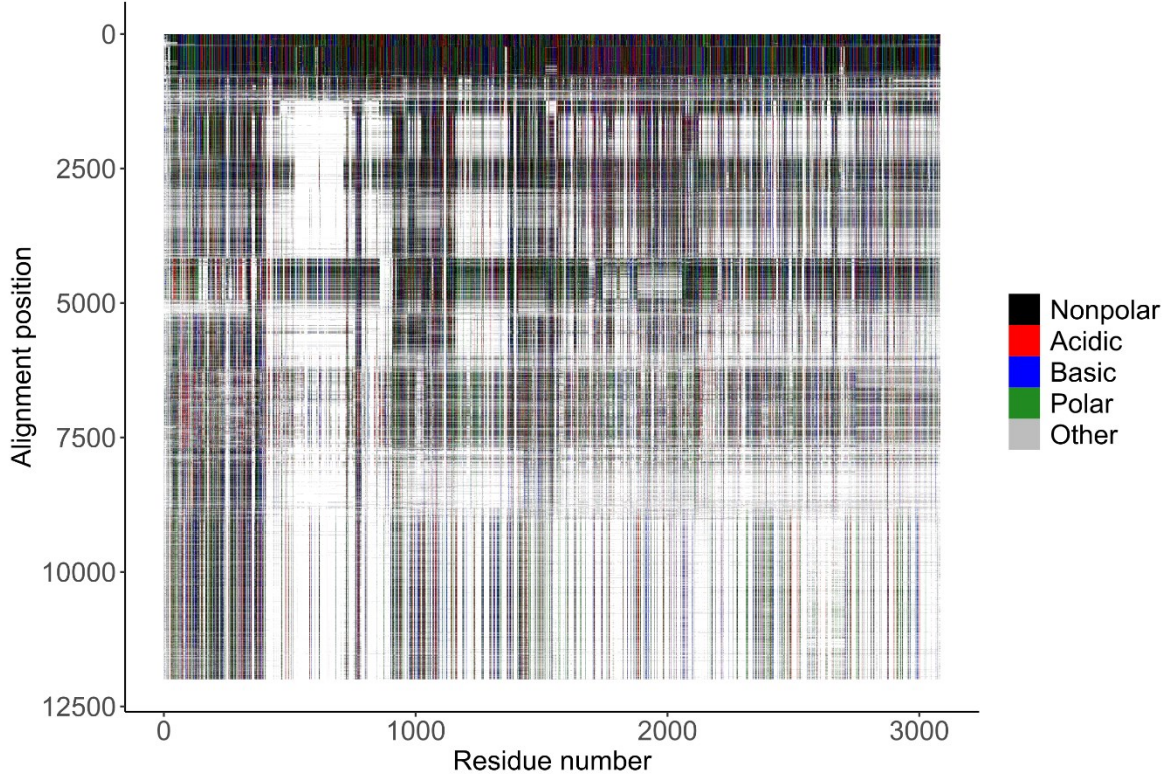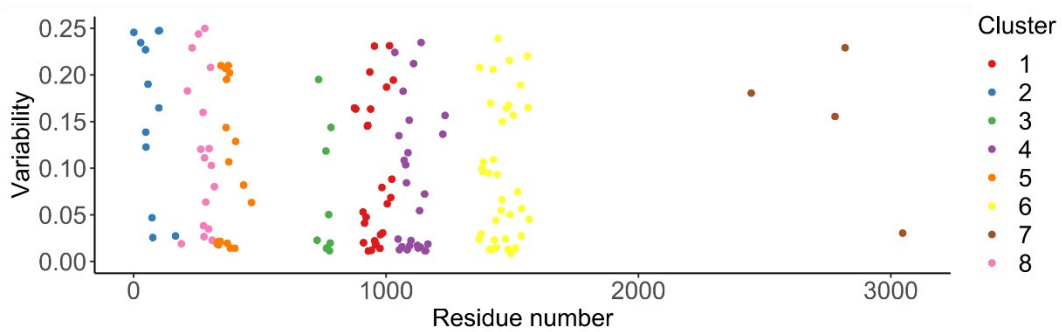
Figure 29. Conservation of the HNK-1 receptors residues inside of the alignment of their similar proteins. For each of the three families of HNK-1 receptors: A. Diagram showing known domains identified using SMART. The length of the line

is proportional to the protein number of residues, colored blocks represent the domains, the length of the block is representative of the motif's position in the protein sequence. B. Tile diagram of aligned sequences corresponding to the length and position of the representative HNK-1 receptor sequence residues without gaps. On the y-axis shown is the position of sequences inside of the alignment, the first position being the representative HNK-1 receptor followed by similar proteins in decreasing order from more similar to less similar proteins. On the x-axis shown are residue numbers of the representative HNK-1 receptor protein sequence; the first residue corresponds to the N-terminus and the last residue corresponds to the C-terminus. The color of the tiles corresponds to the chemical property of the amino acid. Amino acids were grouped by 4 chemical properties, nonpolar, acidic, basic, and polar (other represents amino acids outside of the canonical 20 amino acids, gaps are shown in grey). C. Clusters of less variable (more conserved) residues inside the alignment. Variability in order of the residue position of the representative HNK-1 receptor protein sequence, the first residue corresponding to the N-terminus and the last residue corresponding to the C-terminus. Variability was calculated using the Shannon entropy for the 20 amino acids, giving to each position of the residues inside the protein sequence a variability value (0 not variable to 1 very variable). A cut-off was made for positions with a variability less than 0.25 and with a presence inside of the alignment of more than 60% (maximum 40% of gaps for that residue position inside of the alignment). The remaining conserved positions were then clustered using the k-means approach based on their position on the protein sequence.

For the three families, due to the large gap regions inside the alignments, the conservation of the amino acids of interest corresponding to the HNK-1 receptors was not possible to visualize. Therefore a compact form of the alignment only showing the conservation of the residues corresponding to the HNK-1 receptors without gaps was calculated (andrew-torda, 2020a)(Figure 29.). Also, Shannon's entropy of the 20 standard amino acids inside of the alignments without taking gaps into account and then filtering alignment positions with more than 40% of gaps was calculated and is shown as the variability (Figure 29.). The fact that the chemical properties can be conserved along the alignment does not guarantee that the position is conserved according to Shannon's entropy. The cadherin family clusters of conserved positions (Figure 29.) were mainly formed inside and around the cadherin (CA) domains and the cadherin C domain. I also observed in the alignment diagram that the residues found in the N-terminus were not present inside of the alignment. The HMBG2 family being the smallest showed fewer gaps in the alignment (Figure 29.). Two clusters of

conserved positions can be seen inside and around the high mobility group (HMG) domains. The laminin family was the one containing the most gaps in the alignment (Figure 29.). 7 from 8 clusters of conserved positions are found in the first half of the residue position (N-terminus side). Clusters of conserved positions can be mainly seen inside and around the laminin epidermal growth factor (EGF) and EGF-like domains and the N-terminal laminin domain (LamNT). A small cluster of 4 conserved positions was formed around the C-terminus.

After obtaining these results I was confident that the alignments contained sensible information about amino acid conservation that could help to find a common glycan-receptor motif. A domain recognizing the HNK-1 epitope inside the 2nd G-domain of laminin was identified (residues 2 454 - 2 474, KGVSSRSYVGCIKNLEISRST, NP_032506.2 laminin subunit alpha-1 precursor, *Mus musculus*) using biochemical and bioinformatic approaches (Hall et al., 1997, 1995).
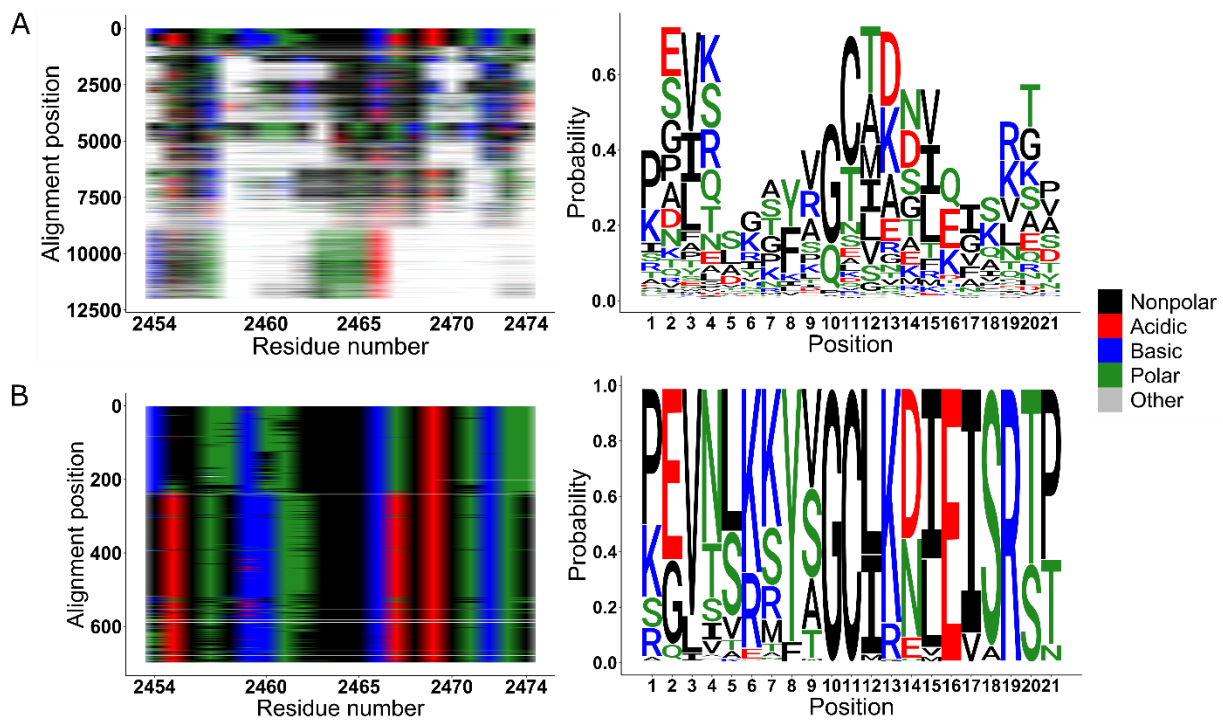
Figure 30. Conservation of the residues belonging to the laminin HNK-1 binding domain in the laminin family alignment. A. Tile diagram (left) of 12 000 aligned sequences corresponding to the length and position of the laminin HNK-1 binding domain (residues 2 454-2 474, KGVSSRSYVGCIKNLEISRST, NP_032506.2 laminin subunit alpha-1 precursor, *Mus musculus*). On the y-axis is shown the position of sequences conforming to the alignment, the first position being the

representative HNK-1 receptor followed by similar proteins in decreasing order from more similar to less similar. On the x-axis shown are residue numbers of the laminin HNK-1 binding domain corresponding to the representative HNK-1 receptor protein sequence. Sequence logo (right) of the 12 000 aligned sequences. The y-axis depicts the probability of finding a certain amino acid inside of the alignment (conservation). The x-axis depicts the position of the amino acid on the alignment. Stacked amino acid one-letter codes represent the presence of such an amino acid in the shown position inside of the alignment. B. Tile diagram (left) of the 700 most similar aligned sequences corresponding to the length and position of the laminin HNK-1 binding domain. Sequence logo (right) of the 700 most similar aligned sequences corresponding to the length and position of the laminin HNK-1 binding domain.

The 21 amino acids belonging to the region of the 2nd laminin G domain (residues 2 454-2 474) binding the HNK-1 epitope were mostly not conserved inside of the 12 000 sequences alignment (Figure 30. A). Although, when observed outside the 12 000-sequence alignments, the first 700 sequences had better residue conservation (Figure 30. A) with a consensus sequence:

RXX[VL]XX[KR][KR]YXGC[LI][KR]X[LI]EISR[TS]

To find other conserved regions within the HNK-1 receptor families, 30 motifs were calculated for each of the similar protein alignments of the three HNK-1 receptors using the Multiple Expression motifs for the Motif Elicitation program, MEME (Figure 31.).

To verify the validity of the found motifs, I decided to observe the motifs inside known domains (Figure 7A) specific to each family. Known domains had the advantage of having experimental, evolutionary, and structural data.
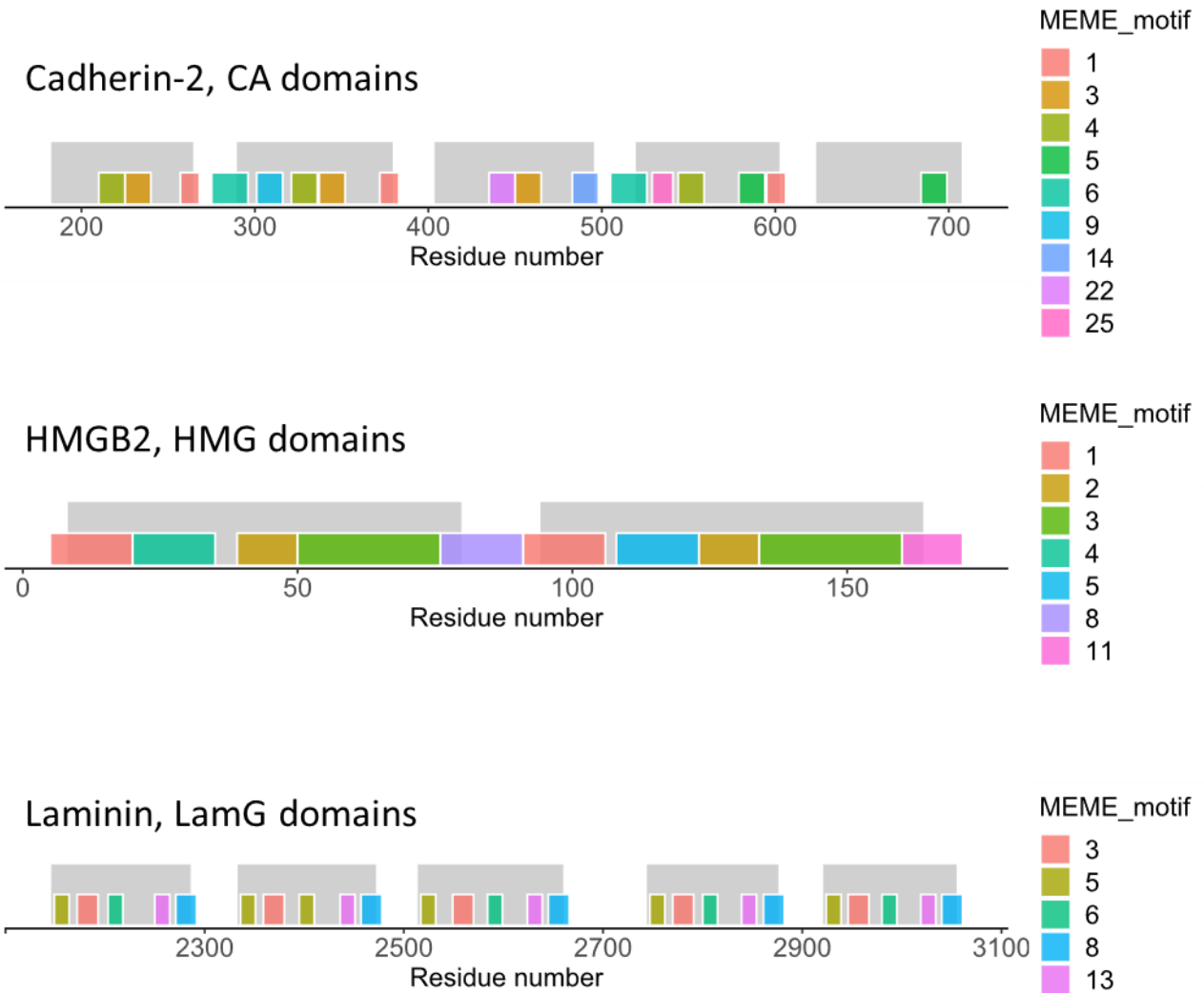
Figure 31. Sequence motifs (MEME motifs) found in the repeated domains of the three HNK-1 receptor families. In grey: cadherin-2 CA domains, HMGB2 HMG domains, and laminin G (LamG) domains. Shown in different colors are the MEME sequence motifs. MEME motifs found for each family are enumerated 1 to 30, MEME motif 1 being the most likely one to be a motif, and MEME 30 being the least likely one.

HNK-1 family alignments were reduced to 400 representative sequences. The MEME program was run to search for a maximum of 30 motifs, 6 to 30 residues long and with any number of repetitions.

The results from the MEME motifs were very promising as repeated motifs were observable surrounding and inside of protein domains defined in the literature. This is especially true for the laminin G domain (Figure 31.), where 4 of the 5 LamG domains contained MEME motifs 5, 3, 6, 13, and 8 in the same order

and location. Only the 2nd LamG domain did not follow this order of motifs although the only difference is that motif 5 is duplicated and replaces motif 6. The 2nd repetition of the 8th MEME motif for the laminin family (2457-2478) overlapped with the HNK-1 binding region inside the 2nd laminin G domain (residues 2454-2474) with a shift of 3 residues towards the C terminus (Figure 32.).
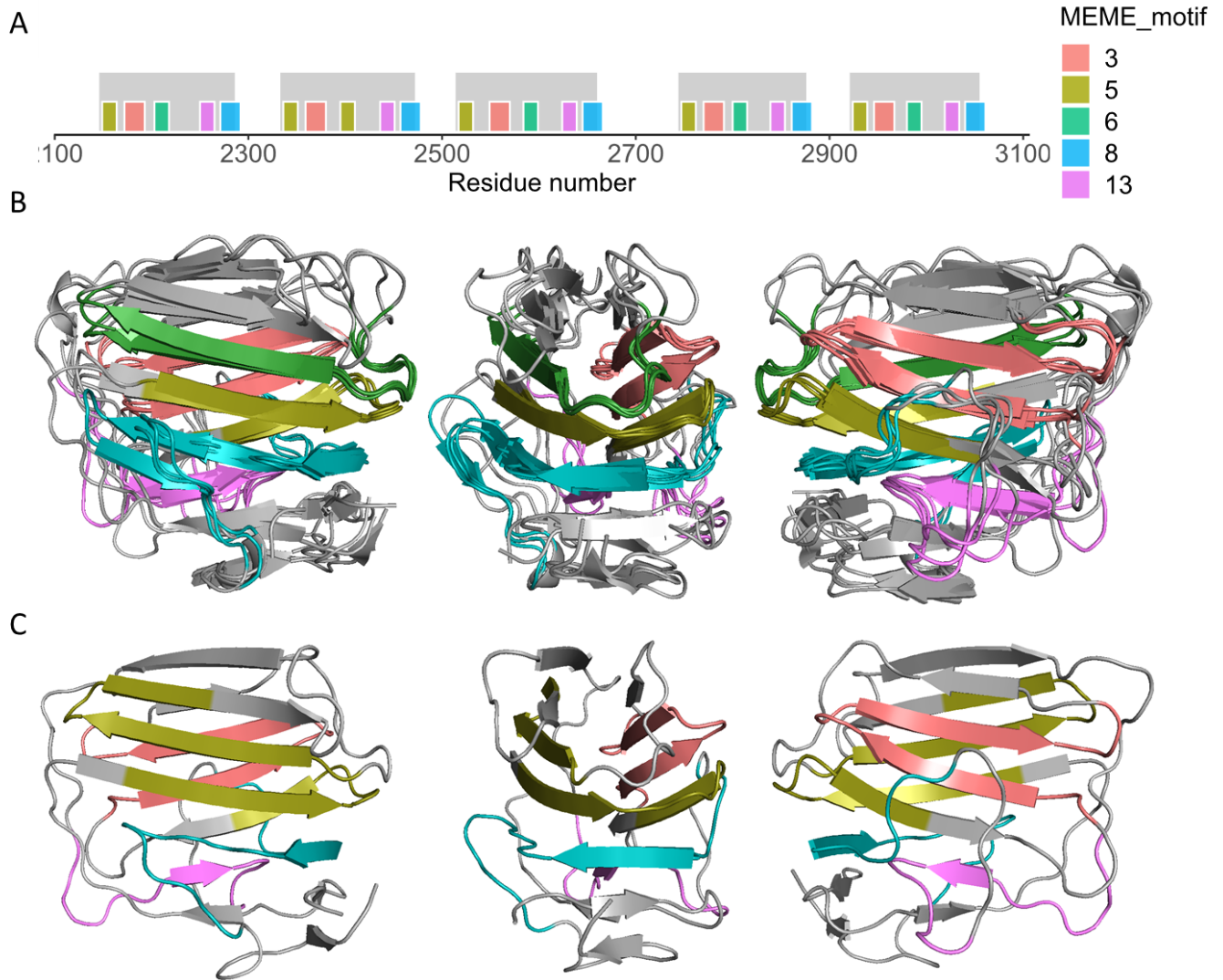
Figure 32. Sequence motifs (MEME motifs) were found in the 5 laminin G domains. A: Diagram depicting laminin domains and motifs found between residues 2 100 and 3 100. In grey, laminin G domains and in different colors: MEME sequence motifs. MEME motifs found for each family are enumerated 1 to 30, MEME motif 1 being the one to be most likely a motif, and MEME 30 being the least likely. B: 3D structure (cartoon view) of aligned laminin G domains 1 (residues 2 145 to 2 287, PDB 2JD4), 3 (2 513 to 2 661, PDB 5MC9), 4 (2743 to

2877, PDB 5MC9) and 5 (2 920 to 3 056, PDB 5MC9). Colored residues correspond to the MEME motifs colors. C: 3D structure (cartoon view) of the laminin G domain 2 (2 332 to 2 473, PDB 2JD4). This motif was found to bind HNK-1. Colored residues correspond to the MEME motifs colors.

Looking in more detail at the laminin G domains, MEME motifs found inside of this β-sandwich structures (Figure 32.) were found for motifs 5, 3, 6, 13, and 8 in the same positions when aligning the laminin G domains. The duplication of motif 5 and deletion of motif 6 represents the uniqueness of the laminin G domain 2 compared to the other 4 laminin G domains regarding the motif found by MEME.

As the 2nd laminin G domain is the only LamG domain found to be interacting with HNK-1, it was interesting to observe the conservation of the third repetition of the 5th MEME motif (residues 2 395 to 2 409) (Figure 33.).
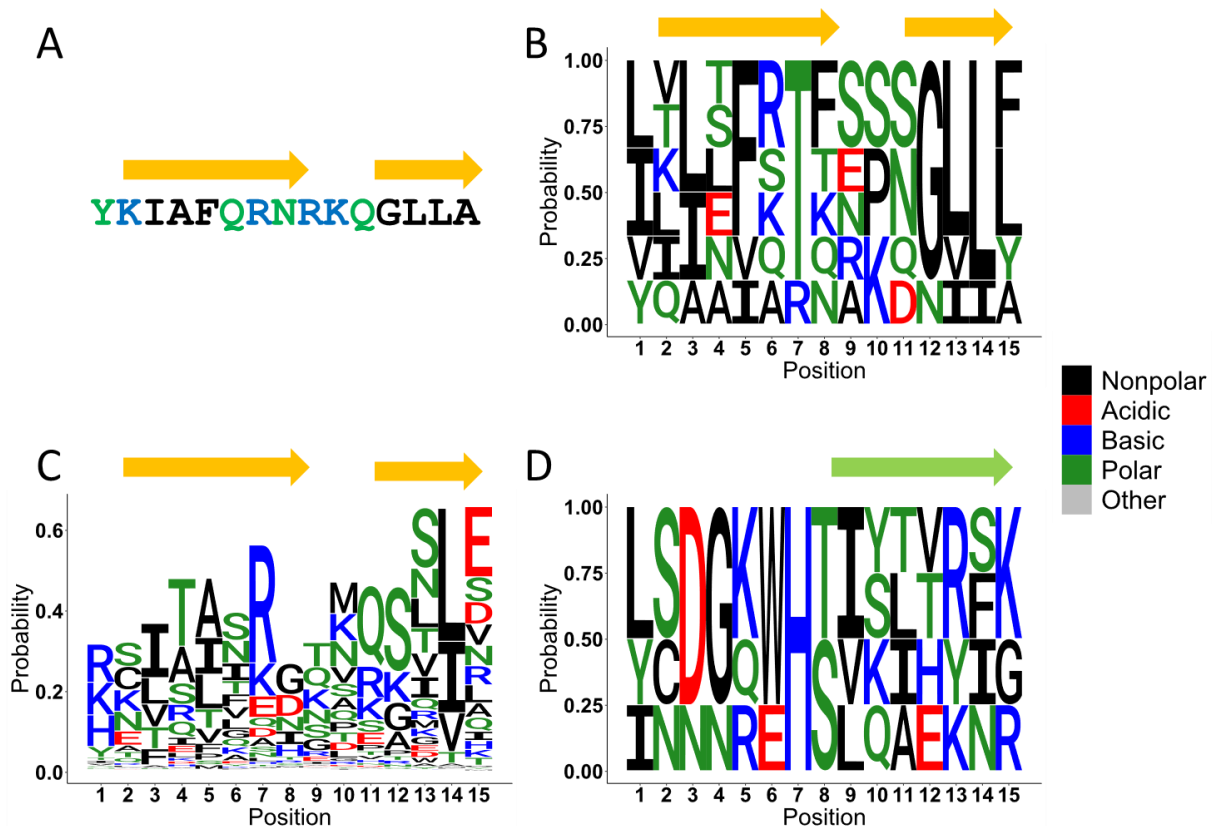
Figure 33. Conservation of the residues belonging to the laminin MEME motif number 5. A. Protein sequence from residues corresponding to the 3rd repeat of the MEME motif 5 in laminin (residues 2 395 to 2 409). Residues are colored according

to their chemical properties. The yellow arrows represent the beta strands found in the secondary structure. B. Sequence logo of the 6 repeats of the MEME motif 5 (residues 2 149 to 2 163, 2 336 to 2 350, 2 395 to 2 409, 2 517 to 2 531, 2 747 to 2 761, and 2 924 to 2 938). C. Sequence logo of 12 000 aligned sequences corresponding to the length and position of the 3rd repeat of the MEME motif 5. D. Sequence logo of the 4 repeats of the laminin MEME motif 6 (residues 2 203 to 2 163, 2 336 to 2 350, 2 395 to 2 409, 2 517 to 2 531, 2 747 to 2 761 and 2 924 to 2 938).

When comparing the 3rd repeat of the MEME motif 5 to the other 5 repeats, the main difference found was that the first residue of the motif was not neutral (Figure 33. B). The residue on the 7th position (residue arginine 2401) was not a polar threonine, which was conserved in the other 5 repeats. This arginine was conserved in the family alignment (Figure 33. C) when focusing on aligned residues to laminin residues 2395 to 2409. While comparing the "long" beta strand of the 3rd repeat of the MEME motif 5 (KIAFQRN; Figure 33. A positions 2 to 8) to the sequence logo of the "long" beta strand of the MEME motif 6 (Figure 33. D positions 8 to 14) we notice a difference in the chemistry of both "long" beta strands. Most notably at the beginning of the strands one basic residue followed by three non-polar residues for the 3rd repeat of the MEME motif 5 and two pairs of polar non-polar residues for the sequence logo of the MEME motif 6. The 3rd repeat of the MEME motif 5 differs from its motif group MEME 5 and also from the motif group MEME 6. Although these differences could be minor I decided to keep studying this special sequence.

To search for similar motifs now between HNK-1 receptor families, I decided to compare motifs from one family against the motifs from the other two families looking for identity with no gaps and ignoring any fragments length smaller than 6 (Figure 34.) in a "sliding window" manner. Motif sequences are consensus sequences from the alignments.
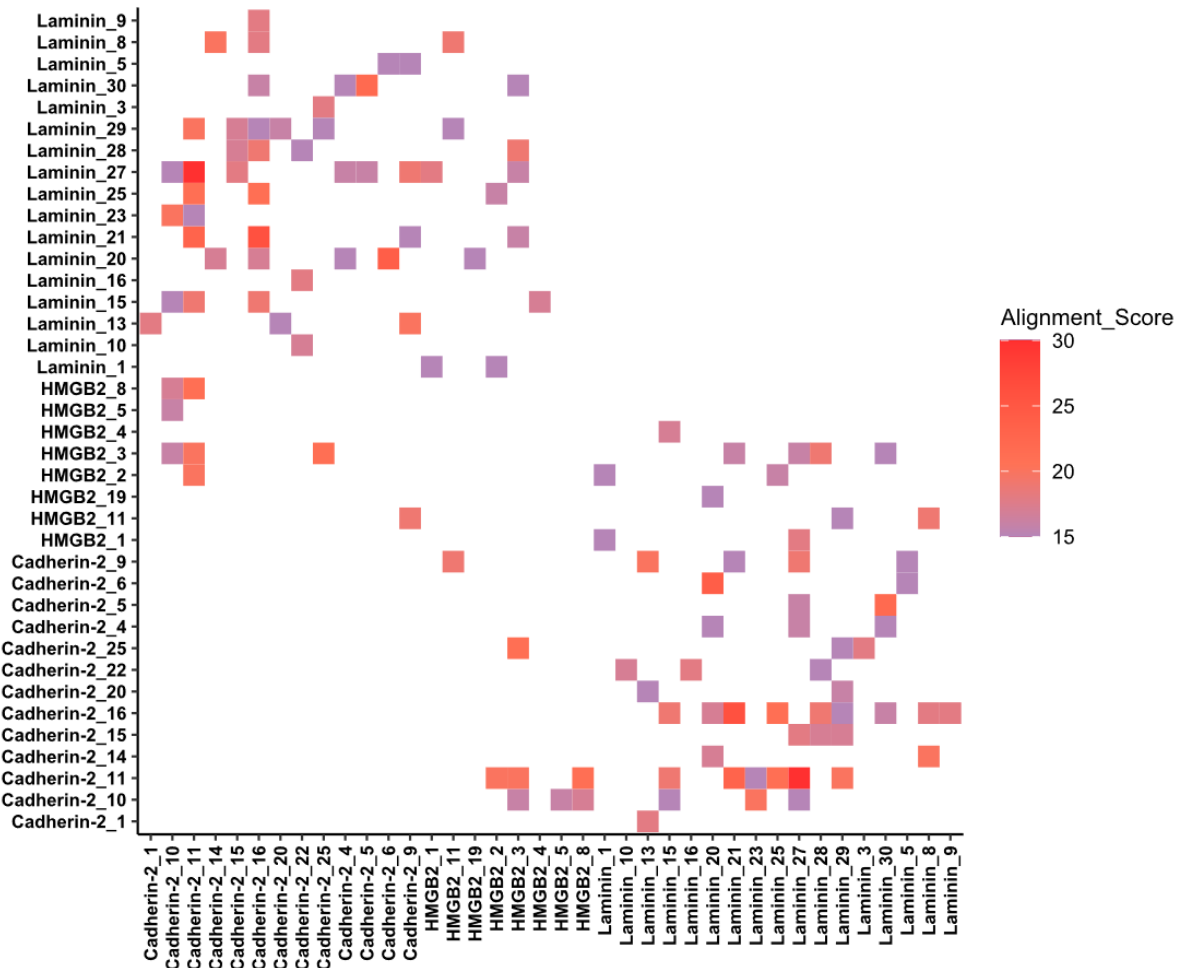
Figure 34. Sliding window comparison between motifs found in the 3 families of HNK-1 receptors. Heat map depicting the 38 most similar motifs (after an alignment score cutoff of greater or equal to 15 was set). The higher the alignment score the more similar are the sequences.

After comparing all the MEME motifs from one HNK-1 receptor family 1 versus 1 against the MEME motifs from the other families, multiple MEME motifs that are similar to MEME motifs from the other 2 families were found (MEME motifs from the same family were not compared).

From all the motifs across the three families, the MEME motif 8 stood out as it overlapped with the HNK-1 binding domain found in the literature (Hall et al., 1997, 1995, 1993) almost in its totality (3 amino acid shift).

Hall et al. HNK-1 binding domain	`KGVSSRSYVGCIKNLEISRST---`
MEME motif 8, consensus	`---STSGFVGCIRBLKINGKPVDL`
MEME motif 8, NP_032506.2	`---SSRSYVGCIKNLEISRSTFDL`

Similar motifs to the laminin MEME motif 8 with a high alignment score were found in cadherin-2 MEME motif 14 and HMGB2 MEME motif 11 (Figure 34.). The laminin MEME motif 8 was found inside of each of the 5 repeated laminin G domains, the cadherin-2 MEME motif 14 was found once inside the third cadherin-2 CA domain, and HMGB2 MEME motif 11 was found once inside the second HMGB2 HMG domain (Figure 31.). I decided to compare the sequence alignment similarities with the known structural data of these domains (Figure 35.). Cadherin-2 MEME motif 16 was also similar to the laminin MEME motif 8 but as it was found in the C terminus and there was no structural data available about this region of the protein and as the alignment score was lower to the cadherin motif 14 the cadherin motif 16 was not taken into account.

A

```
Cadherin-2_14    -------------TVSVTVIDVNENPYF
Laminin_8        SSRSYVGCIKNLEISRSTFDL------
HMGB2_11         --AAYRAKGKS-EA-------------
```

B

C

D

E
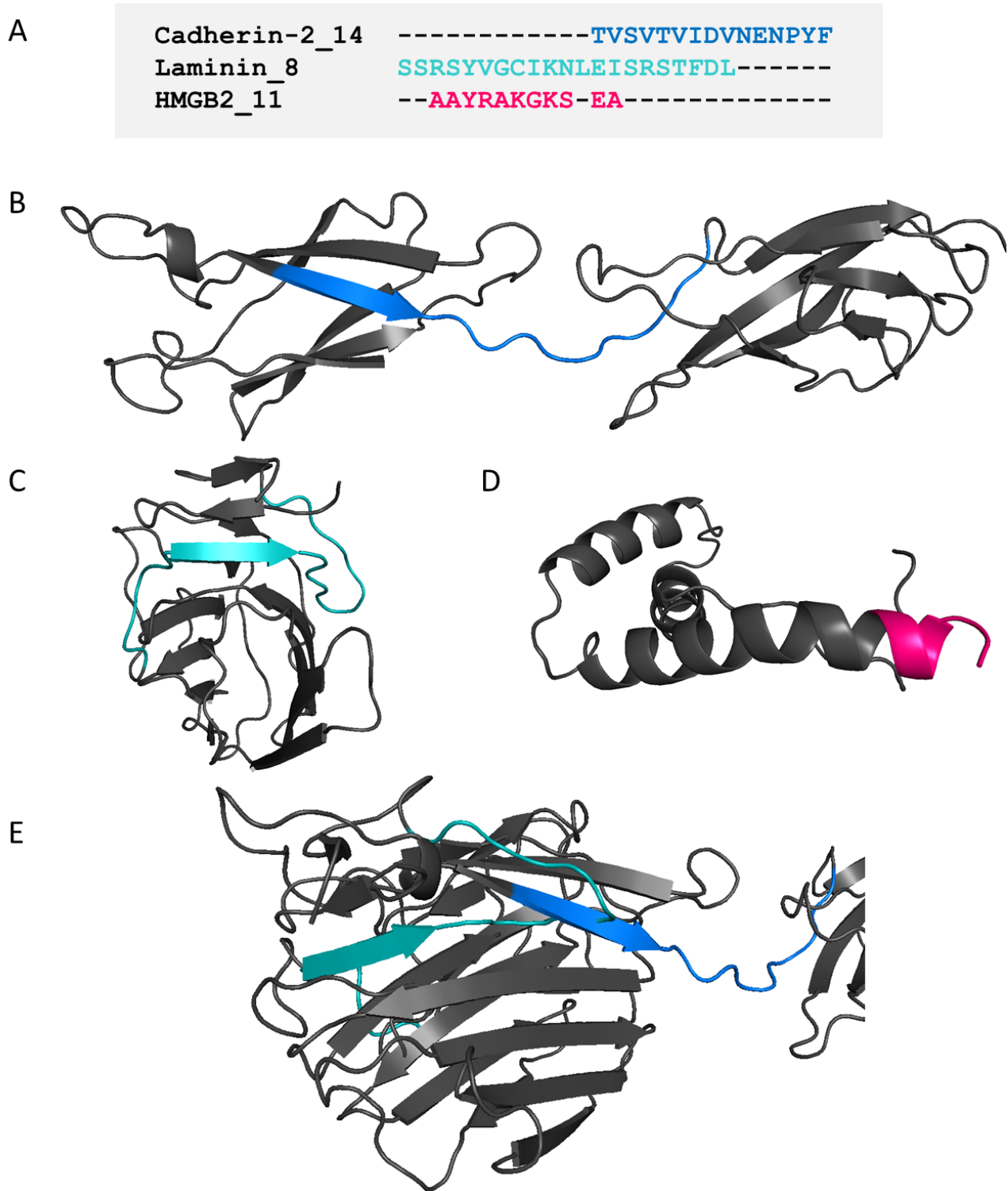
Figure 35. Similarities between the laminin MEME motif overlapping the predicted HNK-1 binding domain, cadherin-2, and HMGB2. A. Alignment of mouse cadherin-2, laminin, and HMGB2 sequences relative to the position of cadherin-2 MEME motif 14 (483 to 497) in blue, laminin MEME motif 8 (residues

2457 to 2477) in cyan and HMGB2 MEME motif 11 (160 to 170) in pink. B. 3D structure (cartoon view) of cadherin-2 residues 400 to 600 (PDB 3Q2W), in blue the cadherin-2 MEME motif 14. C. 3D structure (cartoon view) of laminin residues 2313 to 2489 (PDB 2JD4), in cyan the laminin MEME motif 8. D. 3D structure (cartoon view) of HMGB2 residues 88 to 165 (PDB 1J3D), in pink the HMGB2 MEME motif 11. E. 3D structure (cartoon view) of the alignment of the laminin residues 2472 to 2477 (RSTFDL) with the cadherin-2 residues 483 to 488.

Cadherin-2 and HMGB2 MEME motifs 14 and 11 aligned to different residues of the laminin MEME motif 8 (Figure 35. A). Cadherin-2 MEME motif 14 aligned to the C-terminus of the laminin MEME motif 8 and the HMGB2 MEME motif 11 aligned to the N terminus of the laminin MEME motif 8. The N terminus of the cadherin motif 14 was located in a beta-sheet (Figure 35. B) as well as MEME motif 8 from laminin (Figure 35. B and C). The HMGB2 MEME motif 11 was located at the end of an alpha helix, only the first 6 residues (160 to 165, AAYRAK) were visible in the 3D structure (Figure 35. D). HMGB2 3D structure is the only structure not belonging to a mouse (*Mus musculus*) protein sequence but to a wild boar (*Sus scrofa*) protein sequence. These two sequences were 96.2% identical (residues 89 to 165), the asparagine 109 was changed to histidine and the isoleucine 115 was changed to serine. When the cadherin-2 and laminin structures were aligned according to the previously aligned sequences (Figure 35. A), it is noticeable that in the alignment, the beta strands were not overlapping (Figure 35. E).

The next interesting motif to be compared was the laminin MEME motif number 5. The second laminin G domain differed from the other laminin G domains in that in the conserved position of the MEME motif 6 there was a repetition of MEME motif 5 (Figure 31.). This difference and the fact that it was slightly different from the other iterations of the motif was enough reason to pursue a more detailed analysis as the second laminin G domain is the one interacting with the HNK-1 epitope experimentally. For this motif, two cadherin-2 motifs had an alignment score higher or equal to 15 which indicates a possible similarity, cadherin MEME motifs 6 and 9 (Figure 34.); no HMGB2 MEME motifs were found after the cutoff of number of aligned residues higher or equal to 6.

Cadherin-2 MEME motif 6 was found repeated in CA domains 2 and 4 (Figure 31.) at the N-terminus. This motif was located mostly outside of the CA domains. In addition, cadherin-2 MEME motif 9 is unique (no repeats in the whole protein) and it is found in a beta strand of an anti-parallel beta-sheet just as the laminin

MEME motif 5 (Figure 36.). Although having an alignment score of 15, one of the lowest scores shown on the MEME motifs comparison heatmap (Figure 12), this sequence similarity showed the most interesting structural similarity.

A

```
Cadherin-2_9    NGMLRYRILSQAP-----ST
Laminin_5       -----YKIAFQRNRKQGLLA
```
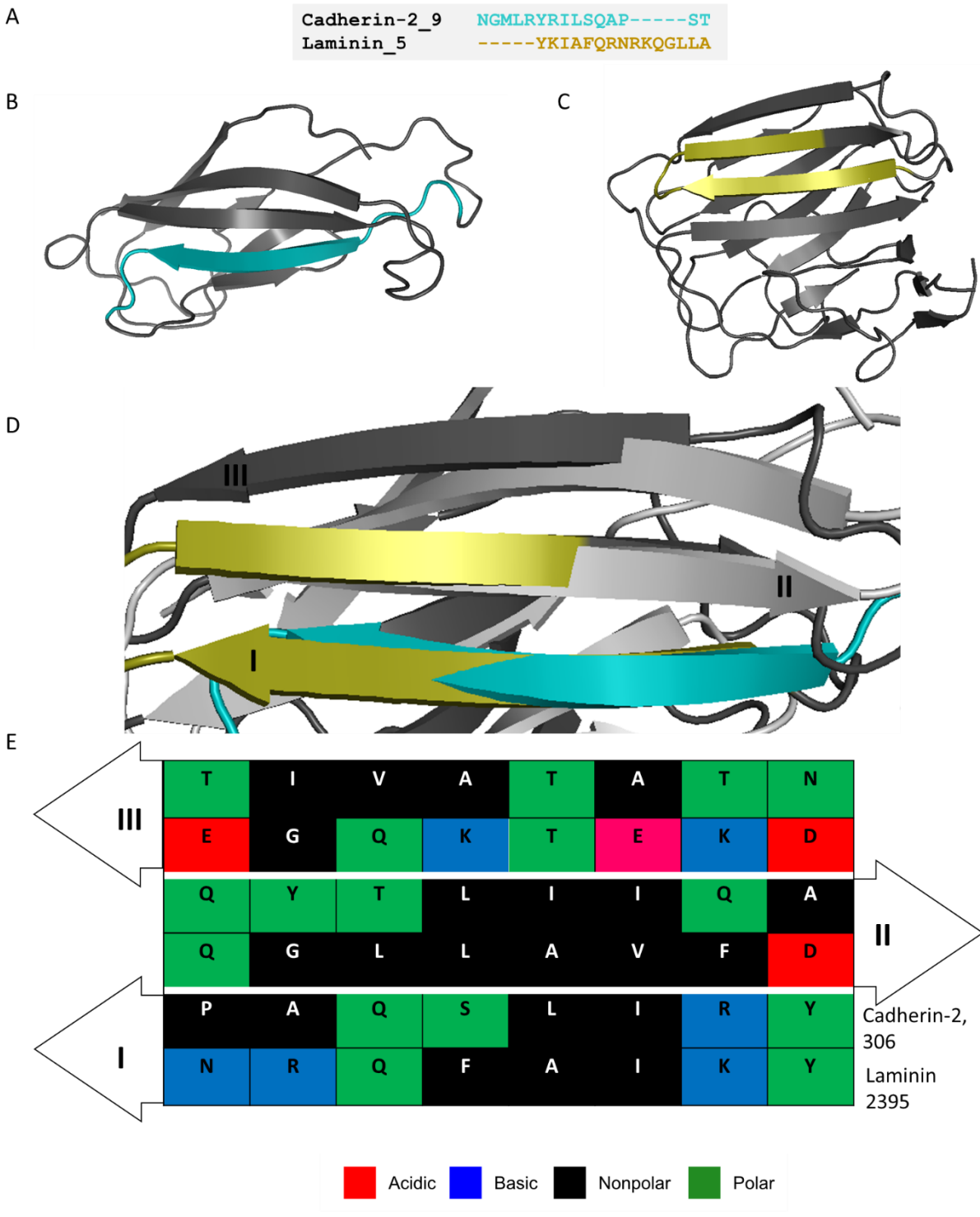
Figure 36. Similarities between the laminin MEME motif 5 and cadherin-2 MEME motif 9. A. Alignment of mouse cadherin-2 and laminin sequences relative to the position of cadherin-2 MEME motif 9 (residues 301 to 315) in cyan and

laminin MEME motif 5 (residues 2 395 to 2 409) in yellow. B. 3D structure (cartoon view) of cadherin-2 residues 265 to 374 (PDB 3Q2W), in cyan the cadherin-2 MEME motif 5. C. 3D structure (cartoon view) of laminin residues 2 313 to 2 489 (PDB 2JD4), in yellow the laminin MEME motif 5. D. 3D structure (cartoon view) of the alignment of laminin in grey and cadherin-2 in light grey, residues 2 395 to 2 402 (YKIAFQRN) with the cadherin-2 residues 305 to 313 (YRILSQAP). Beta strands in the upper foreground were named I, II, and III. E. Diagram depicting amino acids forming an anti-parallel beta-sheet structure. Beta strands I, II, and III were calculated using PyMOL cartoon representation and the positions of amino acids correspond to the polar contacts between residues (two residues from different beta strands that have polar contacts will be facing each other). Arrows show the direction of the beta strand and colors show the chemical properties of the side chains.

Aligned amino acids from cadherin-2 motif 9 and laminin motif 5 represent both two beta strands (beta strand I, Figure 36. D) in their corresponding 3D structures. Although the 3D structures of the anti-parallel beta-sheets from beta strands I, II, and III do not align optimally (RMSD = 2.165), I noticed that not only the first beta strand (Figure 36. E, beta strand I) is conserved but also the second one (II) retains some chemical resemblance. The third beta strand (III) was not conserved.

HMGB2 motifs aligned to the laminin motif 5 were either too short, 2 to 3 aligned amino acids, or the alignment score was too low. The best candidate was HMGB2 MEME motif 4; from these 15 amino acids long motif, 8 residues (residues 26 to 33, EHKKKHPD) aligned to the laminin MEME motif 5 (residues 2398 to 2405, AFQRNRKQ). These two sequences are nevertheless too different to be considered for further analysis. The 3D structure corresponding to the HMG domains consists of a bundle of three helices which challenges the alignment with beta strands.

6. Discussion

6.1. Concentration-dependent binding of the L1 intracellular domain and the chromo shadow domain of HP1γ

To be able to understand the functions of L1 in the nervous system it is important to know its signaling pathways and interaction partners. Especially for the L1-fragments present inside of neurons very few interaction partners are known. Therefore, I tried to identify novel intracellular interaction partners that shed light on the functions of these L1-fragments. In my research, I identified a PXVXL motif in the intracellular domain of L1 that could mediate the interaction of L1 with the chromoshadow domain of HP1 family proteins. To evaluate if L1 can indeed interact with HP1 family proteins via its intracellular domain ELISA was performed. Results showed that the L1- ICD was binding to HP1γ in a concentration-dependent and saturable manner. In contrast, the binding of L1- ICD to HP1β was not saturable and a binding to HP1α was not observed. The PXVXL motif binding has been depicted to be needing the dimerization of the CSD domain of HP1 proteins to interact with the ligand containing the motif (Cowieson et al., 2000; Thiru et al., 2004). In other words, two moles of HP1 proteins are needed to bind 1 mole of the ligand. This is also observable while looking at the binding of L1- ICD to HP1γ. The signal increases as the ligand increases and saturates once no more ligand can be bound by the receptor. The measured dissociation constant of 225 nM fits the expected theoretical value of 500 nM of ligand needed to saturate 1 µM of the receptor. On the contrary, while observing the HP1β curve no saturation is observable even after adding the ligand two times more concentrated than the receptor. This depicts a non-specific binding between the receptor and the ligand. This experiment proved that the full HP1γ protein binds in a concentration dependant manner and seems to need two molecules to bind one molecule of L1- ICD ligand. In addition, results showed that mutant L1- ICD lacking the PXVXL motif did not bind to HP1γ or HP1β. To verify that binding is mediated by the CSD of HP1, the ELISA experiment was repeated using the recombinant HP1γ CSD domain (figure 5 from chapter 5.1.2) while following results from a coincident paper (Liu et al., 2017). Liu et al. depicted and measured the binding affinity of the HP1γ CSD domain using isothermal titration calorimetry. The recombinant HP1γ CSD showing the highest binding affinity consisted of the residues 110 to 176 of HP1γ. Using the most optimal recombinant HP1γ CSD as a receptor demonstrated that the domain from HP1γ

needed to bind the ICD of L1 was indeed the CSD domain. The measured dissociation constant of 170 nM was lower as expected and could be due to improved accessibility by the ligand to the binding site of HP1γ CSD. Whether or not the recombinant protein was correctly produced stays an unanswered question as in opposition to the recombinant L1- ICD protein which was detectable by a monoclonal antibody, the recombinant HP1γ CSD was not detected by our available antibodies and other means to prove the fidelity of the produced recombinant protein, other than the size and activity were not available. The relative low standard error of the mean after repeating the experiment nine independent times, the similarity to the results obtained while using the validated commertial full recombinant HP1γ protein and the usage of a protocol used in a recent publication led me to accept that the recombinant HP1γ CSD was correctly produced.

6.2.   Binding of the L1 intracellular domain and the chromo shadow domain of HP1γ in tissue and cells

After showing that purified L1- ICD and purified HP1γ interact in a biochemical assay it is important to show that this interaction can take place in a cellular context. For this aim, co-localization studies and proximity ligation assays were performed. Co-localization obtained by overlaying two spectrally separated images whose spatial coordinates are the same but where different fluorophores were exited independently, is a common and practical technique used in fluorescence microscopy. The fluorophores are usually coupled to an antibody that directly (primary antibody) or indirectly (secondary antibody) targets a specific molecule. How specific the antibodies are and how little cross-reaction they have are factors that lower false-positive results. This technique helps to elucidate if two fluorescent signals overlap with each other and therefore that both molecular targets co-distribute in the same manner (Dunn et al., 2011). Through figures 6 to 8 from chapter 5.1.3, I looked for co-distribution of L1 inside of the nucleus in different levels: in the nuclei of cells in brain slices, the nuclei of the granule cells in the cerebellum, and the nucleus of cultured cerebellar granule cells. Only at the single-cell level (figure 8 D from chapter 5.1.3), I could observe the codistribution of the L1 and the nucleus and L1 and HP1γ. While implementing algorithms for colocalization analysis such as the Pearson correlation coefficient, the results could not confirm that the colocalization result was non-random. Also, one could argue that the distance separating both molecules could be greater as the distance needed for the two proteins to be considered as being able to interact and

the interaction can not be quantified. To verify that L1 and HP1γ are in close contact I performed PLA. PLA allows detecting proteins that are less than 40 nm away from each other by amplifying the signal created when two antibody-coupled complementary probes (plus and minus) are close enough to each other. In other words, each PLA signal means that targets from the probes at a distance less than 40nm. Since PLA signals could be observed in the nuclei of cultured wild-type but not L1-deficient cerebellar granule cells, results are specific and show that L1 and HP1γ are present in a close contact allowing them to interact in a cellular context.

The observable signal of L1- ICD being close to HP1γ inside the nucleus of CGCs of L1 R687A/y mutant mice showed unaltered interaction of L1 and HP1γ inside of the nucleus compared to the interaction in wild-type neurons. This indicated that the L1-70 fragment, which is not present in the neurons from L1 R687A/y mice is not the only L1-fragment within the nucleus and suggests that other L1 fragments are interacting with HP1γ in the nucleus. Western blot analysis showed only full-length L1 and the L1-70 fragment in subcellular fractions of cerebellar neurons from L1 wild-type or L1 R687A/y mice, therefore the L1 fragments found in the nucleus of neurons from L1 R687A/y mice are probably not visible with this technique. The most probable hypothesis is that the method of nuclear isolation used (Subcellular Protein Fractionation Kit for Cultured Cells; Thermo Fischer Scientific) is not appropriate to observe/isolate these fragments. Another hypothesis can be that these fragments are found in such a low concentration in comparison to full-length L1 and L1-70 so that signals that they are not observable in the western blot. Smaller L1 fragments have been observed in the nucleus of cancer cell lines (Riedle et al., 2009), which I also observed in one experiment using the nuclear fraction of CGCs stimulated with an Fc-fusion protein containing the extracellular domain of the mouse L1 protein (5.1.3.Figure 16.). Stimulation with the mouse L1 extracellular domain fused to Fc from human IgG (L1-Fc) induces L1 homophilic interactions and thus enhances L1-dependent cellular responses like axonal recovery after injury (Roonprapunt et al., 2003). Needless to say, the extracellular L1- Fc-fusion protein cannot be detected by the monoclonal antibody L1-C2 as it is specific to the intracellular domain of L1. Following these hypotheses, it is highly probable that the L1 28 kDa is the fragment interacting with the HP1γ chromo shadow domain. It is also arguable that the fusion protein could have been digested and the secondary antibody could have had cross-reacted. Nevertheless the molecular of the human IgG fragment weights ~50 kDa.

6.3. Interaction of HP1α and HP1β with the L1- ICD

HP1 protein family chromo shadow domains are identical in humans and mice (Figure 37.) and even between each other, there is high conservation. Although the first results depicted the binding of the L1 intracellular domain (figure 3 from chapter 5.1.1) with HP1α as not observable and with HP1β as not concentration-dependent or unspecific, the question about whether or not the small sequence variations in the chromo shadow domains of the three HP1 proteins determine if HP1γ is the only one to bind the intracellular domain of L1 was not answered. Similar recombinant chromo shadow domains from HP1α and HP1β could have been produced and their binding with the intracellular domain of L1 could have been compared. As well proximity ligation assays could have also been established to depict in vivo the distance between the intracellular domain of L1 and HP1α or HP1β.
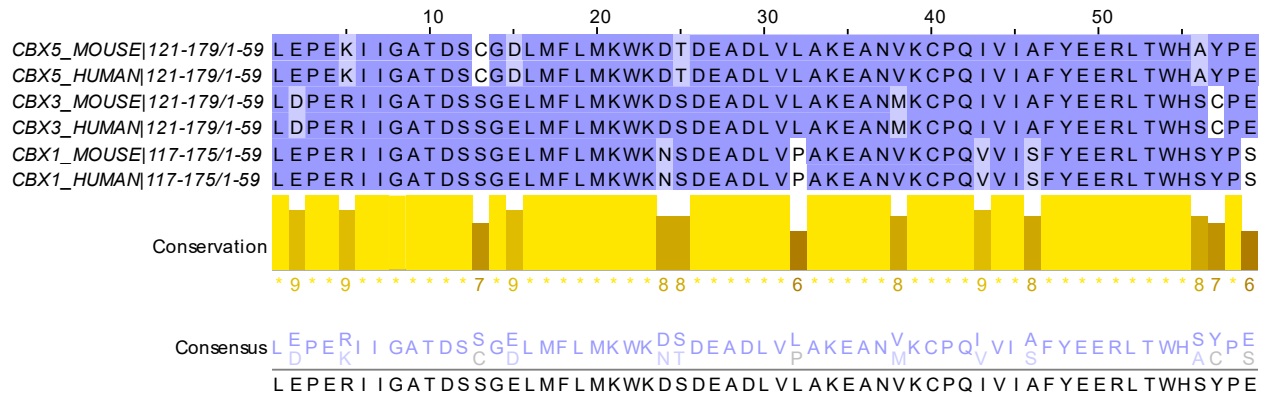
Figure 37. Multiple alignments of the chromo shadow domains of HP1 protein families of humans and mice. Jalview (Waterhouse et al., 2009) visualization of the alignment of chromo shadow domains from mouse and human of HP1α (CBX5; UniProtKB: Q61686 and P45973), HP1γ (CBX3; UniProtKB: P23198 and Q13185), and HP1β (CBX1; UniProtKB: P83917and P83916) obtained from the Uniprot knowledgebase (The UniProt Consortium, 2019). In shades of violet is shown the BLOSUM 62 conservation. A histogram with a conservation score based on a quantitative numerical index reflecting the conservation of the physicochemical properties for each column of the alignment. The best score being an 11 is represented by a "*". The lowest the index the less conserved. A sequence logo depicting the consensus sequence. In shades of violet is shown the BLOSUM 62 conservation.

6.4. Commonly dysregulated genes in absence of L1 or presence of low levels of HP1γ

Based on the gene regulation activity of HP1γ the question arose which genes could be regulated by HP1γ and if the correlation of these genes has been also dysregulated in a model lacking the L1 gene. If a gene that is dysregulated when HP1γ is found in low levels and is dysregulated in the same direction when L1 is not present it could hint at a functional connection between the two proteins (figures 12 to 14 from chapter 5.1.4). It is clear that the result does not show direct causality but it is likely to be a good start point to filter genes. The relationship of the dysregulation of genes related to the cellular response to DNA damage is a really interesting result although the main gene is related to the publication Cheng et al., 2011, nibrin (EntrezID: 27354), was not shown as dysregulated in both datasets. Despite the fact of only being able to reproduce the results of the ATF2 down-regulation in L1 -/y CGCs (figure 15 from chapter 5.1.4), the true experimental model would be to create a mutant mouse line having a non-functional L1 ICD PXVXL motif.

The transcription factor ATF2 has been described as being an essential protein during development and knocking down the ATF2 gene leads to postnatal death due to incorrect development of lungs and brain (Ackermann et al., 2011; Maekawa et al., 1999). Disease-causing mutations in the ATF2 gene have been related to skeletal disorders such as the Kashin-Beck disease (Han et al., 2013) and maybe more interestingly the dysregulation of ATF2 has been related to cancer. In hepatocellular carcinoma cases where the hepatitis C virus is present (Behnke et al., 2012) ATF2 was found to be upregulated (possible as a secondary effect of the virus activity in the liver cell) and thus to be possibly part of the tumor genesis. The DNA repair-related role of ATF2 can also be related to promoting resistance to therapeutic agents in melanomas (Lau et al., 2015), and the liver (Rudalska et al., 2014).

6.5. Importance of L1 fragments inside of the nucleus of cells

In this thesis, I have shown that L1 fragments containing the L1 intracellular domain interact with the chromo shadow domain of HP1γ via the PXVXL domain present at the C-terminus of the L1 intracellular domain in the nucleus from cultured cerebellar granule cells from mice. I also showed that L1 fragments other than or in addition to the L1 70 kDa fragment produced by MBP cleavage (Lutz et al., 2014, 2012)

are present inside of the nucleus of mouse cerebellar granule cells and are responsible for the interaction between the L1 intracellular domain and HP1. Furthermore, in cerebellar granule cells of mice lacking L1, HP1γ is present in higher quantities around the nucleus of these cells. I also tried to show the possible implication of this interaction for gene regulation and I found a lead pointing to a possible down-regulation of genes related to the cellular response to DNA damage while one of the interaction partners is knocked out or knocked down.

The main question that remained unanswered was: Is the inverse case true? Does overexpression of L1 and higher levels of L1 fragments in the nucleus upregulate the genes related to the cellular response to DNA damage?

L1 has been described as a biomarker for poor prognosis in cancer as it promotes tumor migration and tumor metastasis after epithelial-mesenchymal transition and for granting chemoresistance (Giordano and Cavallaro, 2020). For distant metastatic colonization, multiple factors need to be facilitated such as the detachment of primary tumor cells, the survival of single tumor cells or clusters of tumor cells during migration, and the spreading and growth of the secondary tumor. L1 expression modulates the capacity of intestinal cells to regenerate after epithelial-mesenchymal transition (Ganesh et al., 2020) imitating the healing processes after injury most notably by stopping the E-cadherin–REST transcriptional repression of L1, thus upregulating L1 expression. L1 is also known to increase the DNA damage checkpoint activation through L1- ICD-mediated nibrin upregulation thus increasing radioresistance in glioblastoma (Cheng et al., 2011). Two important factors of poor prognosis in cancer, invasiveness (stage III and IV) and therapeutic resistance, are directly related to the L1 activity. A better understanding of L1 interactions and binding partners is key to the research against cancer.

6.6. Glycosylated proteins inside of the nucleus of neural cells

After multiple experiments (5.2.1.Figure 23. and 5.2.1.Figure 24.) the presence of an HNK-1 or LewisX glycosylated L1 fragments inside of the nucleus of neural cells could not be shown with the techniques used in this thesis. In addition, R-PTP-zeta was shown as the main HNK-1 and LewisX carrier in the nucleus of neural cells. While trying to increase the level of the L1 70 kDa fragment inside of the nucleus after

stimulating CGCs with the L1 557 antibody, I could show that R-PTP-zeta levels increased. Whole-cell levels of R-PTP-zeta could have been measured to determine if L1 557 stimulation upregulated the production of R-PTP-zeta or if it regulated the transport of R-PTP-zeta to the nucleus but measured were only nuclear levels and compared between before and after stimulation (5.2.1.Figure 25.). Efforts were made to observe R-PTP-zeta with fluorescence microscopy in the nucleus of neural cells as this protein has been mainly depicted as being mostly located in the cytoplasm or being secreted but no antibody that was tested was found being able to allow proper imaging of this protein. The objective to find HNK-1 and LewisX receptors that could be interacting with glycosylated L1 fragments was thus not accomplished. Therefore, it would be interesting to follow the possible stakes of increasing or decreasing HNK-1 or LewisX epitopes levels presented to the HNK-1 or LewisX receptors by carriers.

6.7.   The HNK-1 binding region

After literature and database analysis (Table 2 from chapter 5.2.2) it became clear that the LewisX receptors are mostly linked to C-type lectins. The structure and conservation of proteins containing the C-type lectin domain have been almost fully described in the literature (Drickamer and Taylor, 2015). In contrast, HNK-1 receptors have not been described as having a common binding domain. No crystal structures of the receptors bound to the HNK-1 glycan are available although some regions of the HNK-1 receptors have been crystallized. Also, it is not even known whether the receptors could be specific to the sulfated or non-sulfated form of HNK-1. Therefore I decided to look into the conservation of each receptor between different species to be able to find common amino acids conserved among multiple species and hopefully between the HNK-1 receptors. As observed in 5.2.2.Figure 26. around 20 000 similar proteins from different species were found for each of the HNK-1 receptors. The species found to have similar proteins were mostly from the chordates phylum and in a lower percentage from the arthropods phylum. Both phyla are known to develop neuronal cells with the difference of chordates developing a centralized nervous system and arthropods a segmented one. Interesting phylums hidden behind the "other" connotation are multiple phyla also known for developing a nervous system such as mollusks and annelids. This is no surprise as laminin, cadherin-2, and the HMGB1 and 2 are highly involved in cell adhesion and neural plasticity (Chou et al., 2003; Morita et al., 2009; Schmitz et al., 1994).

The neighbor-joining tree obtained by aligning a representative number of similar proteins to the 3 HNK-1 receptors, does not show any common branching among the 3 HNK-1 receptors as observed in chapter 5.2.2.Figure 27. No evidence could be shown for a common ancestor protein or domain among the HNK-1 receptors. This result was not surprising as sequence alignment is the most direct approach to compare protein sequences and a visible conserved region would have long been found. I tried to look into more discrete conservation such as the conservation of specific amino acids through evolution using an expectation-maximization search within each family and then between each family of candidate receptors (Bailey et al., 2009). This search led to the identification of multiple sequence motifs that are commonly found in known domains available in proteomics databases as shown in 5.2.2.Figure 31. While comparing these conserved and probably function-related conserved motifs 1 versus 1 across HNK-1 receptor families, I could find related motifs between the HNK-1 receptor families as seen in 5.2.2.Figure 34.

The HNK-1 binding site of one of the three receptors has been described in the literature through multiple publications. The structure of the laminin HNK-1 binding site was depicted using recombinant proteins and glycan mimetics (Hall et al., 1997, 1995, 1993) and was found to be located in the second lamin G domain.

A repeated motif found in the five G laminin domains using expectation-maximization search (laminin MEME motif 8) overlapped the HNK-1 binding site described by Hall et al. (5.2.2.Figure 35.). Therefore it could be compared to the motifs from other families. No similarities were found between the amino acid sequence and structure corresponding to the laminin HNK-1 binding site and other similar conserved amino acid sequences from the other two HNK-1 receptors. No other similar motifs led to a possible common motif and at this point, it is clear that the three HNK-1 receptors are most likely binding the HNK-1 epitope either with different domains binding to the same epitope or with different domains binding the modified sulfated or non-sulfated HNK-1 epitope. The laminin G domains could be specific to the binding of different glycan epitopes as it has been shown that the second LG domain could be specific to bind HNK-1 (Hall et al., 1997, 1995, 1993) and the fourth LG domain to dystroglycan (Hohenester, 2019).

I found a personal interest in the only MEME motif that was different between the 5 laminin G domains, the 3rd repeat of the laminin MEME motif 5 found in the second laminin G domain (5.2.2.Figure 32. ; laminin residues 2 395 to 2 409). It was also shown to have differences to the other sequences belonging to the MEME 5 group and also had differences to the repeat MEME 6 (5.2.2.Figure 33.). The motif found in

laminin residues 2 395 to 2 409 was similar to the cadherin-2 motif MEME 9 found only in the second CA domain (5.2.2.Figure 36.). It could be a good point to test the binding of the second cadherin-2 CA domain to the HNK-1 epitope but this would have taken much more time than planned for this study at the moment of its finding.

The lack of information about the binding of the HNK-1 glycan and the possibility of glycosylated L1 fragments being transported into the nucleus of neural cells is reason enough to try to find nuclear binding partners of HNK-1. As an "unexpensive" in-silico study to find a possible shared binding domain between HNK-1 receptors, two extracellular proteins (laminin and cadherin-2) and one nuclear (HMGB1 and -2) could not produce an answer or give strong indications for a common motif. A biochemical screening approach using a glycan microarray, for example, might be the best approach to find novel receptors which could ultimately help to elucidate the binding of the HNK-1 glycan epitope.

7. Outlook

This research gives first ideas about the possible roles that the fragments of L1 could have inside the nucleus of neural cells. The possible role of gene regulation of genes related to the response of DNA damage by interaction with a member of the HP1 family is not only limited to the neuroscience field of study but also the oncology field can gain much-needed insight on L1 fragments inside of the nucleus. The possible upregulation of genes related to the DNA damage response regulated by increased amounts of soluble L1 fragments could be of major importance while choosing the treatment of a patient. Therapies that are highly dependent on damaging the DNA to stop the division of cancer cells might have less effect on tumors with high amounts of soluble L1 fragments transported inside of the nucleus of their cells.

Not only the protein-protein interactions of the L1 fragments happening in the nucleus are of interest but also the recognition of glycans possibly carried by L1 fragments. One of them being the glycan epitope HNK-1, where little is known about its binding domain and whether or not the domain recognizing this epitope has being conserved among different HNK-1 receptors.

The understanding of the proteolysis of L1, nuclear transport of L1 fragments, and nuclear activity of the L1 fragments have been shown in the last few years to have a considerable impact on medicine. Although being present in neurons in small concentrations the understanding of the regulation, transport, and binding partners of the L1 fragments might be key to find a therapeutic approach for certain cancer types.

8. Materials and methods

8.1. Materials

8.1.1. Laboratory animals

L1-deficient mice were generated by interruption of the mouse L1 gene by inserting in the exon 8 the thymidine kinase and neomycin-resistance genes and introducing the resulting DNA constructs into embryonic stem (ES) cells as described by Dahme et al., 1997. L1-deficient male mice and their age-matched wild-type L1 +/y male littermates were obtained from heterozygous breeding pairs on a 129SVJ genetic background.

L1 R867A mutant mice were generated using the CRISPR/Cas9 genome editing technology (Kleene et al., 2020). L1 R867A/y mice and their age-matched wild-type L1 +/y male littermates were obtained from heterozygous breeding pairs.

Mice were maintained under standard laboratory conditions with food and water supply ad libitum and with a 12-h light/dark cycle. All experiments were conducted in accordance with the German and European Community laws on the protection of experimental animals, and all procedures were approved by the local authorities of the State of Hamburg (Org 679) and conform to the guidelines set by the European Union. Experiments were carried out and documented following the ARRIVE guidelines for animal research (Kilkenny et al., 2010).

8.1.2. Reagents, chemicals, and equipment

All chemicals, reagents, kits and eqipment were purchased from the following companies: Abnova (Taipei, Taiwan), Addgene Inc. (Teddington, UK), Applied Biosystems (Foster City, CA, USA), Biolegend (Fell, Germany), BIOMOL (Hamburg, Germany), Bio-Rad Laboratories (Munich, Germany), Biozol (Eching, Germany), Carl Roth (Karlsruhe, Germany), Corning (Wiesbaden, Germany), Creative BioMart (Shirely, NY, USA), Dako/ Agilent Technologies (Santa Clara, CA, USA), Dianova (Hamburg, Germany), Enzo Life Sciences (Farmingdale, NY, USA), Eppendorf AG (Hamburg, Germany), LifeTechnologies

(Darmstadt, Germany), INVIVO BioTech Service (Hennigsdorf, Germany), Jackson ImmunoResearch (West Grove, PA, UK), Macherey-Nagel (Düren, Germany), Merck Chemicals (Darmstadt, Germany), New England Biolabs (NEB, Massachusetts, USA), Nunc (Roskilde, Denmark), PAA Laboratories (Cölbe, Germany), PAN Biotech (Aidenbach, Germany), Santa Cruz Biotechnologies (Dallas, TX, USA), Sigma-Aldrich (Taufkirchen, Germany), T. H. Geyer (Hamburg, Germany), ThermoFisher Scientific (Waltham, MA, USA), Roche Diagnostics (Mannheim, Germany), VWR International GmbH (Darmstadt, Germany). If not indicated otherwise, all chemicals were used from Sigma-Aldrich.

8.1.3. Solutions and buffers

Agarose gel 2.5% agarose gel for gel electrophoresis	2.5 g agarose Fill until 100ml with 1x TAE Buffer 5 µl Roti Safe (after agarose is dissolved)
Ampicillin stock solution for drug-resistant cell growth	10 g Ampicillin sodium salt Fill until 100ml with dH$_2$O
Antigen retrieval solution for histology	1% SDS PBS pH 7.3
Blocking buffer & Antibody dilution buffer for western blot	TBS-T 4% skim milk
Blocking solution for ELISA	DPBS 2% fatty acid-free bovine serum albumin
Blotting buffer for western blot	25 mM Tris 192 mM glycine

	0.01% SDS
	20% methanol
Carrageenan solution 0.5% for histology	0.5% carrageenan
	0.05% sodium azide
	PBS pH 7.3
Cryoprotectant solution for histology	30% sucrose
	PBS pH 7.3
DNase solution for cell culture	0.5% DNase I
	2.5% glucose
	Neurobasal A
Elution buffer pH 8.0 for native purification	50mM Tris
	250mM Imidazole
	cOmplete EDTA-free Protease Inhibitor Cocktail Tablets (according to final volume)
	NaOH until pH 8.0
Fixing solution 4% for histology	4% formaldehyde
	PBS pH 7.3
Fixing solution 8% for histology	8% formaldehyde
	PBS pH 7.3
Homogenization buffer	Isotonic buffer

for nuclear protein isolation	320 mM sucrose
	1 mM PMSF
Isotonic buffer	20 mM Tris
for nuclear protein isolation	10 mM KCl
	2 mM MgCl$_2$
	2 mM CaCl$_2$
	HCl until pH 7.5
Kanamycin stock solution	5 g Kanamycin sulfate
for drug-resistant cell growth	Fill until 100ml with dH$_2$O
Lysis buffer pH 8.0	50 mM NaH$_2$PO$_4$
for native purification	300 mM NaCl
	10 mM imidazole
	cOmplete EDTA-free protease inhibitor cocktail tablets (according to final volume)
	NaOH until pH 8.0
Lysis buffer	Homogenization buffer
for nuclear protein isolation	0.1% NP40
OPD substrate	0.5mg/ml ortho-phenylene diamine
for ELISA	1:10 stable peroxidase buffer
	dH20
PBS pH 7.3	13.7 mM NaCl
for histology	0.27 mM KCl
	0.8 mM Na$_2$HPO$_4$

	0.15 mM KH$_2$PO$_4$
	HCl until pH 7.3
Permeabilization solution	1% Triton X-100
for histology	0.1 M HCL
	PBS pH 7.3
Poly-L-lysine PLL Coating solution	0.01% PLL
for histology	dH20
RIPA buffer	20 mM Tris
for whole-brain lysate	150 mM NaCl
	1 mM EGTA
	1% NP-40
	1% sodium deoxycholate
	2.5 mM sodium pyrophosphate
	1 mM β-glycerophosphate
	25 Units/ml benzonase (Sigma-Aldrich)
	cOmplete EDTA-free Protease Inhibitor Cocktail Tablets (according to final volume)
	HCl until pH 7.5
Roeder C	20 mM Tris
for nuclear protein isolation	420 mM KCl
	2 mM MgCl$_2$
	2 mM CaCl$_2$
	0.2 mM EDTA
	0.5 mM DTE
	5-10% glycerol

	HCl until pH 7.5
SDS 10 % for SDS-PAGE	10 g SDS Fill until 100ml with dH_2O
SDS Sample buffer (4x) for SDS-PAGE	0.24 M Tris 8% sodium dodecyl sulfate (SDS) 40% glycerol 0.04% bromophenol blue 5% beta-Mercaptoethanol
SDS-PAGE 10% Resolving gel	1.7 ml dH_2O 2.0 ml 30% acrylamide-bisacrylamide (29:1) 2.3 ml 1 M Tris-HCl, pH 8.8 60 µl 10% SDS 15 µl 10% ammonium persulfate (APS) 6 µl tetramethylethylenediamine (TEMED)
SDS-PAGE Stacking gel	1.6 ml dH_2O 0.4 ml 30% acrylamide- bisacrylamide (29:1) 0.3 ml 1 M Tris-HCl pH 6.8 30 µl 10% SDS 15 µl 10% APS 6 µl TEMED
SDS-PAGE Running buffer	250 mM Tris 192 mM glycine 1% SDS

Stripping buffer for western blot	200 mM glycine 0.1% SDS 1% Tween-20 HCl until pH 2.2
Sucrose solution 2.0 M for nuclear protein isolation	Isotonic buffer 2.0 M sucrose 1 mM PMSF
Sucrose solution 1.6 M for nuclear protein isolation	2.0 M sucrose solution Isotonic buffer 1 mM PMSF
Sucrose solution 1.2 M for nuclear protein isolation	2.0 M sucrose solution Isotonic buffer 1 mM PMSF
TAE Buffer (50x) for gel electrophoresis	2 M Tris 1 M acetic acid 50 mM EDTA HCl until pH 8.0
TBS-T for western blot	TBS 0.01% Tween-20
Tris-buffered saline (TBS; 10x) for western blot	10 mM Tris 150 mM NaCl

	HCl until pH 7.5
Trypsin/DNase solution for cell culture	1% trypsin 0.1% DNase I 0.8 mM $MgCl_2$ HBSS
Washing buffer (DPBS-T) for ELISA	DPBS 0.01% Tween-20
Washing buffer pH 8.0 for native purification	50 mM NaH_2PO_4 300 mM NaCl 30 mM imidazole cOmplete EDTA-free Protease Inhibitor Cocktail Tablets (according to final volume) NaOH until pH 8.0
X-1 medium for cell culture	0.1% BSA 10 μg/ml insulin 100 μg/ml transferrin holo 30 nM sodium-selenite 4 nM L-thyroxine 1x penicillin/streptomycin 1x sodium-pyruvate 1x L-glutamine 1x B27 supplement Neurobasal A

8.2. Methods

8.2.1. Protein binding motif search

The protein sequence NP_032504.3, corresponding to the mouse (*Mus musculus*) neural cell adhesion molecule L1 isoform 1 precursor, was used as query protein sequence on the Minimotif Miner 3.0 and later 4.0 (http://cse-mnm.engr.uconn.edu/). The Multi-Score Ranking option was selected.

8.2.2. L1- ICD and mutated L1- ICD plasmids production

The sequence from the reference mRNA sequence NM_008478.3 (Mus musculus L1 cell adhesion molecule (L1cam), mRNA) was retro transcribed in cDNA (8.2.29). The region corresponding to the last 345 base pairs of the C terminus of L1 was amplified by PCR with primers containing the restriction sequence overhangs corresponding to NdeI and BamHI restriction sites (8.2.12). NdeI and BamHI are two restriction sites contained in the pET-28a(+) vector. The PCR product and the vector were then digested with NdeI and BamHI restriction enzymes (New England Biolabs) following the manufacturer's instructions, and the product was run at 150V in a 2.5% agarose gel containing the Roti-GelStain (Carl Roth) nucleic acid staining solution. The stain solution, while bound to nucleic acids in agarose gels and exited with ultraviolet light (290-320 nm), emits green, fluorescent light (515 nm). The gel was cut surrounding the fluorescent light emission and DNA was extracted using the NucleoSpin Gel and PCR Clean-up (Macherey-Nagel). The vector and the insert were ligated using the T4 DNA Ligase (New England Biolabs) following the manufacturer's instructions.

8.2.3. Chemical transformation of competent cells

100ng of the plasmid vector was pipetted into a thawed vial of *E. coli* One Shot™ (ThermoFisher) chemically competent cells. The vial was mixed gently by tapping and was incubated for 30 minutes on ice. After the incubation, a heat shock was performed by placing the vial inside a 42°C water bath for 30 seconds. After the heat shock, the vial was placed back in ice for 5 minutes. 250µl of pre-warmed rich medium S.O.C was added to the vial while working in a sterile manner around a Bunsen burner. The vial was then incubated at 37°C for 1 hour in a shaking incubator at 225 RPM. 100 µL from the transformation

vial were pipetted into the center of the surface of a Luria broth (LB) agar Petri dish containing kanamycin and spread evenly over the surface. The plate was incubated at 37°C for 24 hours. After observing separated distinguishable single colonies, multiple colonies were selected to be inoculated each in 10 ml of room temperature LB liquid medium containing 50µg/ml kanamycin and incubated. The cultures were incubated at 37°C overnight in a shaking incubator at 225 RPM. 100 µl from the cultures were pipetted each into the center of the surface of a lysogeny broth (LB) agar Petri dish containing kanamycin and spread evenly over the surface. The isolated cell clone plates were incubated at 37°C for 24 hours. The rest of the isolated cell clone culture was used for plasmid isolation using the NucleoSpin Plasmid, Mini kit for plasmid DNA purification (Macherey-Nagel). Plasmid isolation was performed according to the manufacturer's instructions and the concentration was measured using the NanoDrop 1000 spectrophotometer (ThermoFisher). The isolated cell clone plasmids were sequenced, and a cell clone was selected for further experiments, and the corresponding plate was stored at 4°C after observing separated distinguishable single colonies.

8.2.4. The setting of a 300 ml *E. coli* culture

A single colony from an isolated cell clone plate was selected to be inoculated in 10 ml of room temperature LB liquid medium containing 50 µg/ml kanamycin and incubated as a starter culture. The starter culture was incubated overnight at 37°C (14-16 h) with continuous shaking at 225 RPM. 5 ml from the starter culture were inoculated to 300 ml of room temperature LB liquid medium containing 50µg/ml kanamycin.

8.2.5. Plasmid amplification

Plasmids pET-28a(+) containing the coding sequence for L1- ICD and mutated L1- ICD were transformed into *E. coli* One Shot™ TOP10 (ThermoFisher) and 300 ml cultures were set and incubated overnight at 37 °C with continuous shaking at 225 RPM. The cultures were used for plasmid isolation using the NucleoBond Xtra Maxi kit for transfection-grade plasmid DNA (Macherey-Nagel). Plasmid DNA isolations were followed according to the manufacturer's instructions, the concentration was measured using the NanoDrop 1000 spectrophotometer (ThermoFisher) and was stored at -20° C.

8.2.6. Recombinant L1- ICD and mutated L1- ICD protein production

Plasmids pET-28a(+) containing the coding sequence for L1- ICD and mutated L1- ICD were transformed into *E. coli* One Shot™ BL21 Star™ (DE3) (ThermoFisher, Massachusetts, USA) chemically competent cells. A 300 ml culture was set and incubated until the OD_{600} reached 0.5-0.6 (3-4 hours approximatively) with continuous shaking at 225 RPM. When the OD_{600} reached 0.5-0.6, protein expression was induced by adding IPTG to a final concentration of 1.0 mM. The culture was grown for additional 4 hours and cells were harvested by centrifugation at 4000 x g for 10 min. The cell pellet was then frozen at -20°C. The frozen cell pellet was resuspended in lysis buffer for native purification (5 volumes lysis buffer per 1 volume cell pellet) and kept on ice for 10-15 min until thawed. The lysate was then sonicated 6 times for 10 seconds avoiding overheating and keeping the lysate on ice as often as possible. The lysate was centrifuged at 10 000 x g at 4°C for 30 min, the supernatant, soluble protein solution, was saved on ice and the pellet, insoluble protein extract, was frozen.

8.2.7. Recombinant L1- ICD and mutated L1- ICD protein purification

The soluble protein solution was incubated with Ni-NTA agarose beads (Qiagen, Hilden, Germany; 1ml beads per 5ml soluble protein solution) overnight at 4ºC with an upside-down movement. The agarose beads were separated from the soluble protein solution inside a column by vacuum filtering. The beads were then washed 3 times using ice-cold washing buffer (30 mM imidazole) for 30 min at 4°C and vacuum filtrating the solution. After washing, the beads were incubated with elution buffer (250 mM imidazole) for 2 h at 4°C and the eluate was collected by vacuum filtration. The purified protein eluate was concentrated by centrifugation using the Vivaspin 20 columns (Sartorius, Göttingen, Germany) with a cutoff of 5kDa for 1-2h at 5 000 x g. DPBS with magnesium and calcium was added to the concentrated purified protein solution, to exchange the elution buffer for the PBS solution and was again concentrated by centrifugation using the same column with a cutoff of 5 kDa for 1-2 h at 5 000 x g. The concentrated purified protein solution was stored at -20°C.

8.2.8. Recombinant L1 and mutant L1 protein concentration determination

Protein concentration was determined using the Pierce™ BCA™ Protein-Assay (ThermoFisher). The chelation reaction was followed according to the producers' protocol. Multiple dilutions of the protein solution were prepared with a dilution factor of 10 as well as multiple dilutions of bovine serum albumin (BSA) solution ranging from 2 000 µg/ml to 100 µg/ml. The absorption from the multiple dilutions of the protein solution was measured at 562 nm using the µQuant microplate reader (Bio-Tek, Vermont, USA) and compared to a standard curve calculated from the absorption of the known BSA dilution to quantify approximately the protein concentration.

8.2.9. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of recombinant L1- ICD and mutated L1- ICD protein solutions

Purified L1- ICD and mutated L1- ICD were denatured by adding to 3 volumes of protein solution 1 volume of 4 times concentrated Laemmli buffer for a final concentration of 1mg/ml of purified protein. The Laemmli buffer protein solution was further denatured by heating the solution to 95 °C for 5 min. 10 µl of the solution were pipetted into a well of a Mini-PROTEAN® TGX™ pre-casted 4 - 15 % polyacrylamide gel (Bio-Rad, California, USA) mounted inside an electrophoresis chamber (Bio-Rad) and submerged in running buffer. The electrophoresis was performed at 120 V until the colored markings of the smaller sizes of the protein ladder were separated enough.

8.2.10. Western blot analysis

The SDS-PAGE gel (8.2.9) was overlaid over two layers of filter paper and a sponge in a Mini Trans-Blot® blotting system (Bio-Rad) cassette submerged in ice-cold blotting buffer. This was followed by a 0.2 µm nitrocellulose sheet Amersham Protran (GE Healthcare) and finishing with two layers of filter paper and a sponge. The cassette was inserted inside the ice-covered blotting chamber containing ice-cold blotting buffer. The blotting was performed at 90 V for 90 min.

The nitrocellulose membrane was blocked using a 4 % skim milk powder in TBS-T (TBS, 0.01% Tween 20) solution for 1 h at room temperature while shaking. The blocking solution was discarded, a primary antibody solution containing the monoclonal mouse L1-C2 antibody (Santa Cruz Biotechnology) diluted

1:1000 in fresh blocking solution was added and the membrane was incubated at 4 °C overnight. The membranes were washed three times with TBS-T solution for 5 min while shaking and were incubated with a primary antibody solution containing HRP-conjugated anti-mouse antibodies diluted 1:10 000 in fresh blocking solution for 1h at room temperature while shaking. The membranes were washed three times with TBS-T solution for 5 min while shaking and visualized using the chemiluminescence mode of the LAS4000 Mini f (GE Healthcare, Freiburg, Germany) camera and the chemiluminescence subtract ECL select (Amersham, GE Healthcare).

8.2.11. ELISA binding of HP1 family proteins and L1- ICD and mutant L1- ICD

A high binding surface 384-well clear microplate (Corning, Tewksbury, MA, USA) was coated with 1 µM of recombinant proteins corresponding to the full-length his-tagged HP1α, HP1β, and HP1γ (Creative BioMart, New York, USA) and incubated at 4°C overnight. The recombinant protein solution was extracted, the plate was blocked using blocking solution (2% fatty acid-free BSA in PBS with magnesium and calcium) and incubated for 1.5 h at room temperature. The plate was washed with PBS for 30 sec at room temperature, L1- ICD and mutant L1- ICD recombinant proteins, diluted in PBS with magnesium and calcium with an increasing range from 1 nM to 2 000 nM, were added each as a ligand to triplicate wells coated with the three different HP1 recombinant proteins and the plate was incubated for 2 h at room temperature with shaking. The plate was washed 3 times for 5min with PBS-T (PBS, 0.01% Tween 20), L1-C2 antibody (Santa Cruz Biotechnology) diluted 1:1 000 in blocking solution was added as primary antibody and the plate was incubated for 2 h with shaking. The plate was washed 3 times for 5min with PBS-T, anti-mouse secondary antibody (Dianova, Hamburg, Germany) diluted 1:1 000 in blocking solution was added as secondary antibody and the plate was incubated for 2 h with shaking. The plate was washed 3 times for 5min at room temperature with PBS-T, OPD substrate solution (0.5mg/ml ortho-phenylene diamine in stable peroxidase buffer) was added and incubated for 5min with shaking. A stopping solution (2.5 M sulfuric acid) was added to cease the oxidation reaction producing a yellow-brown coloring. Colorimetric measurement of the solution inside the wells was measured at 492 nm using the µQuant microplate reader (Bio-Tek). The positive control was made by coating with recombinant L1- ICD or mutated L1- ICD and adding primary and secondary antibodies. The negative control was made by coating with recombinant L1- ICD or mutated L1- ICD and adding only secondary antibodies.

8.2.12. Polymerase chain reaction (PCR) amplification of the DNA sequence coding for the HP1γ chromo shadow domain

The pGEM-T plasmid containing the CBX3 gene sequence according to the reference sequence NM_007276.4 (Sino Biological, Beijing, China) was used as a template for the amplification of the 204 nucleotide sequence corresponding to the 68 amino acids of the HP1γ chromo shadow domain (CSD; RefSeq NM_007276.4: nucleotides 717 to 921; RefSeq NP_009207.2: amino acid residues 106 to 178). 7.5 pmol of each of the forward and reverse primers with overhangs specific for the aLICator LIC Cloning, Expression Kit 4 (N-terminal His-tag/WQ) (ThermoFisher Scientific) designed following manufacturer instructions, in line with the reading frame, with a melting temperature lower than 70° C and had a GC content between 40 and 50%.

Forward   5'   GGTTGGGAATTGCAAGCTGCTGACAAACCAAGAGGAT   3'

Reverse   5'   GGAGATGGGAAGTCATTAAGAATGCCAAGTTAGTCTC   3'

Primers were mixed with 100 ng of the template, 12.5 µl of CloneAmp™ HiFi PCR Premix (Takara Bio, Shiga, Japan), and Nuclease-Free Water (Qiagen) for a final volume of 25 µl. The PCR program (ThermoFisher) followed the manufacturer recommendations:

98 °C	10 sec	
55 °C	5 or 15 sec	35 cycles
72 °C	5 sec/kb	

8.2.13. PCR amplification product clean up

PCR products were mixed 1:6 with DNA Gel Loading Dye (ThermoFisher) and were subjected to electrophoresis (Bio-Rad) in a 2.5% agarose gel with 0.005% Roti Safe Gelstain (Carl Roth) submerged in TAE buffer at 100 V until the ladder is resolved. Using the E.A.S.Y. UV-light documentation system (Herolab, Wiesloch, Germany) and a scalpel an agarose gel cube was cut surrounding the luminescent DNA band corresponding to the 200-base marker. The amplified DNA was extracted from the agarose gel cube with the NucleoSpin Gel and PCR Clean-up, Mini kit for gel extraction, and PCR clean-up (Macherey-Nagel) following the manufacturer's instructions.

8.2.14. HP1γ chromo shadow domain ligation independent cloning (LIC)

The amplified sequence corresponding to the HP1γ CSD containing the overhangs specific for the aLICator LIC Cloning, Expression Kit 4 (N-terminal His-tag/WQ) (ThermoFisher) were inserted inside the pLATE52 vector according to the aLICator LIC Cloning, Expression Kit 4 (N-terminal His-tag/WQ) manufacturer's instructions.

8.2.15. Recombinant HP1γ chromo shadow domain production

The pLATE52 vector containing the HP1γ CSD coding sequence was amplified (8.2.5) using ampicillin as a selection antibiotic at 100 μg/ml and 1 clone was selected for recombinant protein production (8.2.6) using also ampicillin as a selection antibiotic. The recombinant protein purification was done by incubating the soluble protein solution with Ni-NTA agarose beads (Qiagen; 1ml beads per 5ml soluble protein solution) overnight at 4ºC with an upside-down movement. The agarose beads were separated from the soluble protein solution inside a column by vacuum filtering. The beads were then washed 3 times using ice-cold washing buffer (10 mM imidazole) for 30 min at 4°C and vacuum filtrating the solution. After washing, the beads were incubated with a 50 mM Tris-HCl (pH 8.0) solution containing 1:10 units WELQut Protease (ThermoFisher Scientific) per µg protein and was incubated for 16 h at 4°C. The protein solution was vacuum filtrated, and the protein concentration was measured.

8.2.16. Recombinant HP1γ chromo shadow domain coomassie staining

SDS-PAGE electrophoresis (8.2.9) was performed on the digested recombinant HP1γ chromo shadow domain protein solution and the gel was stained using the Roti-Blue kit (Carl Roth) overnight at room temperature with shaking. The gel was washed with deionized water until the bands could be differentiated from the background. An image was taken while visualizing the gel using a light table.

8.2.17. ELISA binding of recombinant HP1γ chromo shadow domain and L1- ICD and mutant L1-ICD

A high binding surface 384-well clear microplate (Corning, Tewksbury, USA) was coated with 1 µM of recombinant HP1γ chromo shadow domain and ELISA was performed (8.2.11) using the recombinant L1-ICD and mutant L1- ICD as ligand.

8.2.18. Cryosectioning of 7-day-old mouse brains for immunofluorescence

7-day-old L1 +/y wildtype mice were euthanized by decapitation. The brains were dissected on a petri dish over a cool plate and fixed in a solution of 4% formaldehyde in 0.1M PBS pH7.3 for one week. The fixed brains were then cryoprotected by soaking them in a 30% sucrose solution in 0.1M PBS for 1-2 days until the brains sink to the bottom of the recipient. The brains were snapped frozen in -80°C 2-methyl-butane. Fixed frozen brains were mounted inside the Leica CM3050 S cryostat (Leica, Buffalo Grove, USA) at -30°C using the Tissue-Tek O.C.T. Compound (Sakura Finetek, Torrance, USA) embedding medium. Sagittal cryo-slices with a thickness of 20 µm were obtained and directly mounted in a Superfrost Plus (ThermoFisher Scientific) adhesion microscope slide.

8.2.19. Immunofluorescence staining of brain sections

Brain cryo-sections mounted on a microscope slide were permeabilized in a solution with 1% Triton X-100 and 0.1 M HCl in PBS for 30 min. The slide was washed with PBS and incubated with an antigen retrieval solution of 1% sodium dodecyl sulfate (SDS) in PBS for 10 min. The slide was washed 4 times with PBS and incubated in a 0.5% carrageenan solution containing primary antibodies diluted 1:100 at 4°C overnight with light movement. The slide was washed 4 times with PBS and incubated in a 0.5% Carrageenan solution containing fluorescent secondary antibodies (with non-overlapping emission wavelengths) diluted 1:200 at room temperature for 2 h with light movement while protecting from light. Images were obtained using a fluoresce microscope.

8.2.20. Isolation and culture of cerebellar granule cells

Seven-day-old L1 +/y wildtype and L1 -/y mutant littermate mice were euthanized by decapitation. The brains were dissected from the head on a petri dish over a cool plate and the cerebellar area was separated from the rest of the brain and submerged in ice-cold Hanks' Balanced Salt Solution (HBSS) for further microdissection of the cerebellum under a stereomicroscope (Carl Zeiss, Oberkochen, Germany). Clean cerebella were divided into three pieces and moved to a biological fume hood (ThermoFisher). The cerebella were incubated with Trypsin/DNase solution for 15 min at room temperature. The cerebella were washed 3 times with ice-cooled HBSS and DNase solution was added. The cerebella were homogenized using fire-polished Pasteur pipettes with decreasing tip diameters. Ice cold HBSS was added and the cell solution was centrifuged for 15 min at 100 g and 4 °C. The remaining liquid was extracted, and the cell pellet was resuspended in X-1 medium at 37 °C. Cell concentration was measured using a 0.4% Trypan blue solution mixed 1:1 with the sample solution and a hemocytometer (Neubauer chamber). Cells were seeded in 0.01% PLL-coated surfaces and cultured for 24 h in a cell culture incubator set at 37 °C, 5% CO_2, and 90% humidity.

8.2.21. Immunofluorescence staining of cerebellar granule cells

Cerebellar granule cells of 7-day-old L1 +/y wildtype and L1 -/y mutant littermate mice were isolated and approximately 250 000 cells were cultured in PLL-coated round coverslips. The cells that had adhered to the coverslips were then fixed adding 1:1 a solution of 8% formaldehyde for a final concentration of 4% for 30 min at room temperature. The coverslips were washed with PBS and cells adhered to the coverslips were permeabilized in a solution with 1% Triton X-100 and 0.1 M HCl in PBS for 30 min. The slide was washed with PBS and incubated with an antigen retrieval solution of 1% sodium dodecyl sulfate (SDS) in PBS for 10 min. The slide was washed 4 times with PBS and incubated in a 0.5% Carrageenan solution containing primary antibodies diluted 1:100 at 4°C overnight with light movement. The slide was washed 4 times with PBS and incubated in a 0.5% carrageenan solution containing fluorescent secondary antibodies diluted 1:200 at room temperature for 2 h with light movement while protecting from light. The secondary antibody solution was removed, and the coverslips were mounted over microscopy slides using the ROTI Mount FluorCare DAPI (Carl Roth) medium.

8.2.22. 3D imaging of fluorescence microscopy z-stacked images

The F1000 confocal microscope (Olympus, Tokyo, Japan) optical axes (width x, height y, and dept z) were calculated using the Microscopy Nyquist rate calculator web application (minimal sampling density calculation to avoid information loss) to get an ideal voxel (3D pixel) size for the image acquisition for 3D imaging according to the objective numerical aperture and fluorophores used. Laser intensity was regulated to avoid photobleaching due to the z-stacks excitation cycles. The sequential acquisition mode was used and the wavelength acquisition window from the channel filters was narrowed to avoid spectral bleed-through between fluorophores. Once the z-stacks were obtained the images were processed using the Imaris 9.3.0 (Oxford Instruments, Abingdon, United Kingdom) software.

8.2.23. Proximity ligation assay (PLA) of L1- ICD, and HP1γ inside of the nucleus of cerebellar granule cells from mice

Cerebellar granule cells of 7-day-old L1 +/y wildtype and L1 -/y littermate mice or 7-day-old L1 +/y wildtype and L1 R687A/y mutant littermate mice were isolated (8.2.20) and approximatively 250 000 cells were cultured in PLL-coated glass coverslips. The cells adherend to the coverslips were then fixed adding 1:1 a solution of 8% formaldehyde for a final concentration of 4% for 30 min at room temperature. The coverslips were washed with PBS and cells adhered to the coverslips were permeabilized in a solution with 1% Triton X-100 and 0.1 M HCL in PBS for 30 min. The slide was washed with PBS and incubated with an antigen retrieval solution of 1% sodium dodecyl sulfate (SDS) in PBS for 10 min. The slide was washed 4 times with PBS and the Duolink In Situ Red Starter Kit Mouse/Rabbit (Sigma-Aldrich) was used following the manufacturer instructions using the anti-L1 mouse monoclonal antibody C2 (Santa Cruz Biotechnology) and the HP1γ rabbit polyclonal antibody PA5-30954 (ThermoFisher) at a 1:100 concentration. The coverslips were mounted over microscopy slides using the ROTI Mount FluorCare DAPI (Carl Roth) medium instead of the manufacturer's mounting medium. The PLA reaction will amplify a signal recognizable by detection probes containing fluorophore excitable at 594nm and emitting at 624nm if the two protein-antibody complexes containing one of both + and - probes are close enough to allow the reaction to happen, 0-40 nm.

8.2.24. Quantification of PLA signal in 3D fluorescence microscopy images

Z-stacks from fluorescence microscopy images were acquired (8.2.22) and processed via the Imaris 9.3.0 (Oxford Instruments, Abingdon, United Kingdom) software using the colocalization tool and forming spheres 0.4µm of diameter relative to the PLA signal voxels. The spheres per nuclei were counted.

8.2.25. Quantification of PLA signal in 2D fluorescence microscopy images

The microscope z-axis was located on the middle point of the nuclei and images were collected in different locations of the microscopic slide containing cells while avoiding overlapping. The images were automatically processed using Fiji (Fiji Is Just ImageJ; Schindelin et al., 2012) software locating and measuring the total volume of the nuclei signal, removing the signal outside the nucleus, and counting circular spots of 10-80 pixels of diameter corresponding to the PLA signal. After the program has iterated through all the images the nucleus volume and number of PLA signal spots were analyzed with the R programming language to calculate the median PLA signal spots per nuclei per image.

8.2.26. Soluble protein extract from mice whole brain lysates

7-day-old L1 +/y wildtype and L1 -/y mutant littermate mice were euthanized by decapitation. The brains were dissected from the head on a petri dish over a cool plate. The whole brain was homogenized in 1 ml RIPA buffer using an ice-cold Potter-Elvehjem Tissue Grinder. The homogenate was centrifuged at 17 000 x g for 15 min at 4 °C dividing the soluble from the insoluble protein.

8.2.27. Western blot protein quantification

Western blot analysis was made (8.2.10) from whole-brain soluble protein extract of 7-day-old L1 +/y wildtype and L1 -/y mutant littermate mice using the HP1γ rabbit polyclonal antibody PA5-30954 (ThermoFisher Scientific) and the GAPDH mouse monoclonal antibody 6C5 (Santa Cruz Biotechnology) as a loading control. The images were processed using the Fiji software by selecting rectangular areas corresponding to the chemiluminescent signal and measuring the area of the signal intensity curve. The area of the protein of interest was divided by the area of the loading control to obtain the level of protein of interest.

8.2.28. Gene expression from expression profiling by array experiments from the GEO database

Expression profiling by array datasets GSE13984 (Tapanes-Castillo et al., 2010) for the microarray (Affymetrix Murine Genome U74AV2 gene chip) of 11-day-old L1 +/y and L1 -/y mice and GSE44084 (Sridharan et al., 2013) for the microarray (Affymetrix GeneChip Mouse Genome 430 2.0) of pre-induced pluripotent stem cells (pre-iPSCs) and pre-iPSCs subjected to siRNA mediated knockdown of the CBX3 gene from the Gene Expression Omnibus (GEO) database were analyzed via R. The R script used mainly the Linear Models for Microarray and RNA-Seq Data limma R package (Ritchie et al., 2015) from the Bioconductor project (Gentleman et al., 2004) to determinate the fold change, p-value and link to GO functions. Further analysis and visualizations were obtained using the R package toolbox Tidyverse (Wickham et al., 2019).

8.2.29. Reverse transcription and cDNA production

Cerebellar granule cells of 7-day-old L1 +/y wildtype and L1 -/y mutant littermate mice or 7-day-old L1 +/y wildtype and L1 R687A/y mutant littermate mice were isolated (8.2.20) and approximatively 1 000 000 cells were cultured in a PLL-coated 6 well plates. After 24h of culture, the plates were placed on 4 °C cold plates and the culture medium was removed. RLT Plus (Qiagen, Hilden, Germany) lysis buffer containing ß-mercaptoethanol (10 µl/ml) was added, and cells were detached from the bottom of the well using a cell scraper. The solution was homogenized by centrifugation through the QIAshredder (Qiagen, Hilden, Germany) spin-column according to the manufacturer's instructions. RNA was purified from the homogenate using the RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. RNA concentration was measured using the NanoDrop 1000 spectrophotometer (ThermoFisher Scientific). Complementary DNA (cDNA) was synthesized using the M-MLV Reverse Transcriptase (Merck, Darmstadt, Germany) kit with an oligo dT_{18} primer to hybridize with the 3' poly adenine region from mRNAs. The reaction was done following the manufacturer's instructions. cDNA concentration was measured using the NanoDrop 1000 spectrophotometer (ThermoFisher Scientific).

8.2.30. Two-step quantitative real-time PCR (qRT-PCR) analysis

The qRT-PCR was performed with the cDNA library obtained in 8.2.29 and multiple primers targeting genes of interest for expression quantification using the qPCR Core kit for SYBR Green I Only dUTP kit (Eurogentec, Liège, Belgium). The reaction was done following manufacturer instructions by using the same proportion of reagents for a final volume of 20 µl instead of the 50 µl indicated by the manufacturer. The qRT-PCR reaction was measured by an Applied Biosystems ABI 7900HT Real-Time Thermo Cycler 96 Well (Applied Biosystems, California, USA) following the following 3-step protocol:

$$
\begin{array}{lll}
95\,°C & 10\ \text{min} & \\
95\,°C & 15\ \text{sec} & \\
60\,°C & 20\ \text{sec} & \Big\}\ 35\ \text{cycles} \\
72\,°C & 40\ \text{sec} & \\
4\,°C & \text{forever} & \\
\end{array}
$$

Relative expression quantification was calculated using the 2(-Delta Delta C(T)) method (Livak and Schmittgen, 2001) using the PCR R package.

The primer sequences were obtained from the PCR primer database for quantitative gene expression analysis, PrimerBank (Wang et al., 2012):

Gene name	PrimerBank ID	Primer	Sequence 5'->3'	NCBI GeneID
Atf2	387487a1	Forward	AAGTCTGGCTATCATACTGCTGA	11909
Atf2	387487a1	Reverse	GCCATGACAATCTGTGAAAGTGC	11909
Chek1	31542385a1	Forward	GTTAAGCCACGAGAATGTAGTGA	12649
Chek1	31542385a1	Reverse	GATACTGGATATGGCCTTCCCT	12649
Nipbl	26381017a1	Forward	ACATGGCTAATTCCAAACTGACG	71175
Nipbl	26381017a1	Reverse	CGAGGTAATTGCTTTGTTCCGAG	71175
Rad52	6755280a1	Forward	CTTTGTTGGTGGGAAGTCTGT	19365
Rad52	6755280a1	Reverse	CGGCTGCTAATGTACTCTGGAC	19365
Zmat3	6678585a1	Forward	TTCCTTTACCTAATCGGCCTTCA	22401
Zmat3	6678585a1	Reverse	TTCCTGCCCAAAAGCCTTCTG	22401

8.2.31. Soluble nuclear protein fraction isolated from adult mice brains

Adult wild-type mice were euthanized by CO_2 inhalation. The brains were dissected from the head on a petri dish over a cool plate. The whole brain was homogenized in 2 ml homogenization buffer (320nM sucrose, 1mM PMSF in isotonic buffer) using an ice-cold Potter-Elvehjem Tissue Grinder. The homogenate was centrifuged at 200 x g for 15 min at 4°C. The supernatant was discarded, and the pellet was incubated with 5 times its volume of lysis buffer (0.1 % NP-40) for 15 min at 4°C. The lysate was centrifuged at 1 000 x g for 15 min at 4°C, the supernatant was discarded, and the pellet was resuspended gently in 5 ml of homogenization buffer. A sucrose step gradient was prepared with a 2.0 M sucrose solution in isotonic buffer containing 1mM PMSF on the bottom of a 35 ml open-top ultracentrifugation tube, a 1.6 M sucrose solution in the middle, and a 1.2 M sucrose solution on the top. The resuspended pellet was carefully placed on the top of the sucrose gradient avoiding disturbing or mixing the step solutions. The step gradient was centrifuged in an Optima XE ultracentrifuge (Beckman Coulter, California, USA) with a SW32Ti Swing-out rotor (Beckman Coulter, California, USA) at 82 705 x g for 2 h at 4°C. The gradients were carefully discarded, and the nuclear pellet was resuspended in homogenization buffer for washing. The nuclear solution was centrifuged at 1 000g for 20min at 4°C. The supernatant was discarded, and the nuclear pellet was resuspended in Roeder C buffer (25µl per brain sample) and homogenized using an ice-cold Potter-Elvehjem Tissue Grinder. The homogenate was ultra-centrifuged at 100 000 g for 30 min at 4°C separating the solution into the soluble nuclear protein extract and the insoluble nuclear protein fraction. The soluble nuclear protein extract was used for the experiments while the insoluble nuclear protein fraction was stored at -80°C.

8.2.32. Cerebellar granule cells stimulation with L1 557 antibody

Cerebellar granule cells of 6-to8-day-old L1 +/y wildtype and L1 -/y mutant littermate mice were isolated (8.2.20) and approximately 1 000 000 cells were cultured in PLL-coated 6 well plates. After 24h of culture, the cells were either stimulated with 50µg/ml of the L1 557 antibody or given the same volume of X-1 medium and were further incubated for 1 hour in a cell culture incubator set at 37 °C,5% CO_2, and 90% humidity. The plates were placed on 4 °C cold plates and the culture medium was removed. 100 µl CEB

buffer containing protease inhibitors (volume advised for 10 µl of packed cell volume or 1 000 000 HeLa cells) from the Subcellular Protein Fractionation Kit for Cultured Cells (ThermoFisher Scientific) was added to the cerebellar granule cells adhered at the bottom of the plate well. The cells were resuspended by physically detaching them using a cell scraper. The following steps were done according to the manufacturer's instructions. Four soluble protein subcellular fractions were obtained, the cytoplasmic extract, the membrane extract, the nuclear extract, and the chromatin-bound nuclear extract.

8.2.33. Similar protein sequences search

Similar protein sequences for each of the HNK-1 binding proteins: cadherin-2 (RefSeq: NP_031690.3, Mus musculus), HMGB-1 (RefSeq: NP_034569.1, Mus musculus), HMGB-2 (RefSeq: NP_032278.1, Mus musculus) and laminin subunit α-1 (RefSeq: NP_032506.2, Mus musculus) were collected using psi-blast (Altschul et al., 1997) with up to 30 iterations, a maximum number of results set to 20 000 and accepting homologs with an expectation value of less than 10^{-21} is the probability of having by chance an unrelated sequence in our data set.

8.2.34. Similar protein sequences alignment of all families and guided tree calculation

The HNK-1 receptor similar protein families were cleaned (Tags were removed) using the clean_seqs C++ program from the sequtils toolbox (andrew-torda, 2020b) and aligned using the MAFFT (Katoh and Standley, 2013) multiple alignment program for amino acid or nucleotide sequences. Due to the length of amino acids and the large number of sequences making the alignment highly demanding in time and resources, cadherin-2 and laminin families were aligned with the default algorithm (--6merpair) and a maximum of five iterations (mafft --6merpair --amino --quiet – maxiterate 5 --distout). The HMGB-2 family had short-length amino acid sequences therefore it allowed this family to be aligned using a more accurate algorithm with affine gap penalties (--genafpair) (Altschul, 1998). The aligned sequences were reduced to 5 000 representative sequences using the distance matrix generated by the alignment (--distout) by the reduce C++ program from the sequtils toolbox (andrew-torda, 2020b). The reduced sequences from cadherin-2, laminin, and HMGB-2 were joined to form a 15 000 sequences dataset and was aligned with

MAFFT using the 6mer default algorithm and with the option of having a guide tree (--treeout) which divides the aligned sequences into multiple related groups, based on the number of shared 6mers, as an output file in addition to the alignment file.

8.2.35. Literature domain Search with SMART

Known domains mined from the literature and databases were taken with the help of the web tool SMART (Simple Modular Architecture Research Tool; http://smart.embl-heidelberg.de/) (Letunic and Bork, 2018; Schultz et al., 1998). Using the default search on the sequences of cadherin-2 (RefSeq: NP_031690.3, Mus musculus), HMGB-2 (RefSeq: NP_032278.1, Mus musculus) and laminin (RefSeq: NP_032506.2, Mus musculus).

8.2.36. HNK-1 receptor families new cutoff and independent alignment

Each HNK-1 receptor protein family (~20 000 sequences) was submitted to a second cutoff with an expectation value of less than 10^{-25} to have a general maximum expectation value across families. Cadherin-2 and laminin families were aligned with the MAFFT default 6mer algorithm (--6merpair) and HMGB-2 with the affine gap penalties algorithm (--genafpair) (Altschul, 1998). Alignments were depicted by a tile diagram where each colored tile represents a residue and white tiles gaps. The figures were obtained using the ggplot2 package from the tidyverse collection (Wickham et al., 2019). Sequence logos diagrams were obtained using the ggseqlogo package (Wagih, 2017) and the residue frequencies inside of the alignment.

8.2.37. Alignments entropy calculations

Variability was calculated from the alignment files and the HNK-1 receptors mouse query sequences: cadherin-2 (RefSeq: NP_031690.3, Mus musculus), HMGB-2 (RefSeq: NP_032278.1, Mus musculus), and laminin (RefSeq: NP_032506.2, Mus musculus) using the entropy Perl program (andrew-torda, 2020a).

$$S_i = \sum_{a=1}^{a=20} p_a \log_{20} p_a$$

The entropy program is based on the Shannon entropy S_i at each site i where p_a is the probability of amino acid type a in a column of the alignment and the summation runs over the 20 amino acid types. The output is two files, one containing the calculated entropy for each amino acid from

the HNK-1 receptors mouse query sequences inside of the alignment and an alignment file containing a "compressed" version of the alignment where columns containing gaps in the HNK-1 receptors mouse query row were removed leaving only columns corresponding to aligned similar proteins residues to the HNK-1 receptors mouse query residues (the first column corresponds to the first amino acid of the sequence and the last column to the last residue without gaps in between).

8.2.38. Clustering of conserved residues of the HNK-1 receptor families

Residues with 25 % or less variability and a maximum of 50 % of gaps were roughly clustered using the k-means algorithm (MacQueen, 1967) using the core R function with the default Hartigan-Wong (Hartigan and Wong, 1979) algorithm on a list of residue numbers corresponding to conserved residues. The number of clusters was determined using the NbClust package (Charrad et al., 2014) using the default settings, a minimum of 2 clusters, a maximum of 20 clusters, and the "kmeans" cluster analysis method. The number of clusters was not always selected according to the majority rule of the number of indices proposing the best number of clusters.

8.2.39. Sequence motif search

Similar protein sequence alignments obtained in 8.2.36 were reduced to 400 representative sequences using the distance matrix generated by the alignment (--distout) by the reduce C++ program from the sequtils toolbox (andrew-torda, 2020a). The reduced sequences containing for each family the query sequences, cadherin-2 (RefSeq: NP_031690.3, Mus musculus), HMGB-2 (RefSeq: NP_032278.1, Mus musculus), and laminin (RefSeq: NP_032506.2, Mus musculus), were run using the command-line version of the MEME Multiple EM for Motif Elicitation program (Bailey et al., 2009) looking for maximum 30 motifs, 6 to 50 residues long and any number of motif repetitions (-nmotifs 30 -minw 6 -maxw 50 -mod anr).

8.2.40. Motif comparison between families

The calculations result from 8.2.39 was screened for motifs found in the query sequences for each family. Finding a motif inside of the alignment does not imply that the motif is found in the query sequence. Motifs consensus sequences were compared between families using the pairwise alignment function from the

Biostrings package (H. Pagès et al., 2020) part of the Bioconductor project (Gentleman et al., 2004) with a blosum62 substitution matrix in a local alignment mode. The alignment could be defined as a "sliding window" alignment as the gap opening and gap extension cost was set to be extremely high, therefore gaps inside of the motif alignments will result in a poor score. The aligned residues that were smaller than 6 residues long or had an alignment score less than 15 had their alignment scores removed. Also, scores from alignments inside of the same family were removed from the heatmap.

8.2.41. 3D visualization of crystal structures

Structural data (cadherin-2, PDB: 3Q2W; HMGB2, PDB: 1J3D; laminin, PDB: 2JD4) was obtained from the Worldwide Protein Data Bank (wwPDB) and was visualized using the PyMOL Molecular Graphics System (Schrödinger, New York, USA).

9. Bibliography

Ackermann, J., Ashton, G., Lyons, S., James, D., Hornung, J.-P., Jones, N., Breitwieser, W., 2011. Loss of ATF2 function leads to cranial motoneuron degeneration during embryonic mouse development. PloS One 6, e19090. https://doi.org/10.1371/journal.pone.0019090

Adle-Biassette, H., Saugier-Veber, P., Fallet-Bianco, C., Delezoide, A.-L., Razavi, F., Drouot, N., Bazin, A., Beaufrère, A.-M., Bessières, B., Blesson, S., Bucourt, M., Carles, D., Devisme, L., Dijoud, F., Fabre, B., Fernandez, C., Gaillard, D., Gonzales, M., Jossic, F., Joubert, M., Laurent, N., Leroy, B., Loeuillet, L., Loget, P., Marcorelles, P., Martinovic, J., Perez, M.-J., Satge, D., Sinico, M., Tosi, M., Benichou, J., Gressens, P., Frebourg, T., Laquerrière, A., 2013. Neuropathological review of 138 cases genetically tested for X-linked hydrocephalus: evidence for closely related clinical entities of unknown molecular bases. Acta Neuropathol. (Berl.) 126, 427–442. https://doi.org/10.1007/s00401-013-1146-1

Altschul, S.F., 1998. Generalized affine gap costs for protein sequence alignment. Proteins Struct. Funct. Bioinforma. 32, 88–96. https://doi.org/10.1002/(SICI)1097-0134(19980701)32:1<88::AID-PROT10>3.0.CO;2-J

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., Westerfield, M., Ekker, M., Postlethwait, J.H., 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282, 1711–1714. https://doi.org/10.1126/science.282.5394.1711

andrew-torda, 2020a. andrew-torda/sequence-entropy: basic code for sequence entropy. Zenodo. https://doi.org/10.5281/zenodo.4053030

andrew-torda, 2020b. andrew-torda/sequtils 1.0. Zenodo. https://doi.org/10.5281/zenodo.4066305

Aucott, R., Bullwinkel, J., Yu, Y., Shi, W., Billur, M., Brown, J.P., Menzel, U., Kioussis, D., Wang, G., Reisert, I., Weimer, J., Pandita, R.K., Sharma, G.G., Pandita, T.K., Fundele, R., Singh, P.B., 2008. HP1-beta is required for development of the cerebral neocortex and neuromuscular junctions. J. Cell Biol. 183, 597–606. https://doi.org/10.1083/jcb.200804041

Azzaz, A.M., Vitalini, M.W., Thomas, A.S., Price, J.P., Blacketer, M.J., Cryderman, D.E., Zirbel, L.N., Woodcock, C.L., Elcock, A.H., Wallrath, L.L., Shogren-Knaak, M.A., 2014. Human heterochromatin protein 1α promotes nucleosome associations that drive chromatin condensation. J. Biol. Chem. 289, 6850–6861. https://doi.org/10.1074/jbc.M113.512137

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202-208. https://doi.org/10.1093/nar/gkp335

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., Kouzarides, T., 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410, 120–124. https://doi.org/10.1038/35065138

Behnke, M., Reimers, M., Fisher, R., 2012. The Expression of Embryonic Liver Development Genes in Hepatitis C Induced Cirrhosis and Hepatocellular Carcinoma. Cancers 4, 945–968. https://doi.org/10.3390/cancers4030945

Bernard, A., Boumsell, L., Dausset, J., Milstein, C., Schlossman, S.F. (Eds.), 1984. Leucocyte Typing: Human Leucocyte Differentiation Antigens Detected by Monoclonal Antibodies. Specification — Classification — Nomenclature. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-68857-7

Brown, J.P., Bullwinkel, J., Baron-Lühr, B., Billur, M., Schneider, P., Winking, H., Singh, P.B., 2010. HP1γ function is required for male germ cell survival and spermatogenesis. Epigenetics Chromatin 3, 9. https://doi.org/10.1186/1756-8935-3-9

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. J. Stat. Softw. 61, 1–36. https://doi.org/10.18637/jss.v061.i06

Cheng, L., Wu, Q., Huang, Z., Guryanova, O.A., Huang, Q., Shou, W., Rich, J.N., Bao, S., 2011. L1CAM regulates DNA damage checkpoint response of glioblastoma stem cells through NBS1. EMBO J. 30, 800–813. https://doi.org/10.1038/emboj.2011.10

Chou, D.K.H., Henion, T.R., Jungalwala, F.B., 2003. Regulation of expression of sulfoglucuronyl carbohydrate (HNK-1), Amphoterin and RAGE in retinoic acid-differentiated P19 embryonal carcinoma cells. J. Neurochem. 86, 917–931. https://doi.org/10.1046/j.1471-4159.2003.01911.x

Chou, D.K.H., Zhang, J., Smith, F.I., McCaffery, P., Jungalwala, F.B., 2004. Developmental expression of receptor for advanced glycation end products (RAGE), amphoterin and sulfoglucuronyl (HNK-1) carbohydrate in mouse cerebellum and their role in neurite outgrowth and cell migration. J. Neurochem. 90, 1389–1401. https://doi.org/10.1111/j.1471-4159.2004.02609.x

Cowieson, N.P., Partridge, J.F., Allshire, R.C., McLaughlin, P.J., 2000. Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. Curr. Biol. CB 10, 517–525. https://doi.org/10.1016/s0960-9822(00)00467-x

Dahme, M., Bartsch, U., Martini, R., Anliker, B., Schachner, M., Mantei, N., 1997. Disruption of the mouse L1 gene leads to malformations of the nervous system. Nat. Genet. 17, 346–349. https://doi.org/10.1038/ng1197-346

Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3, e314. https://doi.org/10.1371/journal.pbio.0030314

Díez-Revuelta, N., Higuero, A.M., Velasco, S., Peñas-de-la-Iglesia, M., Gabius, H.-J., Abad-Rodríguez, J., 2017. Neurons define non-myelinated axon segments by the regulation of galectin-4-containing axon membrane domains. Sci. Rep. 7, 12246. https://doi.org/10.1038/s41598-017-12295-6

Drickamer, K., Taylor, M.E., 2015. Recent insights into structures and functions of C-type lectins in the immune system. Curr. Opin. Struct. Biol., Carbohydrate-protein interactions • Biophysical and molecular biological methods 34, 26–34. https://doi.org/10.1016/j.sbi.2015.06.003

Dube, D.H., Bertozzi, C.R., 2005. Glycans in cancer and inflammation — potential for therapeutics and diagnostics. Nat. Rev. Drug Discov. 4, 477–488. https://doi.org/10.1038/nrd1751

Dunn, K.W., Kamocka, M.M., McDonald, J.H., 2011. A practical guide to evaluating colocalization in biological microscopy. Am. J. Physiol. - Cell Physiol. 300, C723–C742. https://doi.org/10.1152/ajpcell.00462.2010

Endo, T., 2005. Glycans and glycan-binding proteins in brain: galectin-1-induced expression of neurotrophic factors in astrocytes. Curr. Drug Targets 6, 427–436. https://doi.org/10.2174/1389450054021909

Faissner, A., Teplow, D.B., Kübler, D., Keilhauer, G., Kinzel, V., Schachner, M., 1985. Biosynthesis and membrane topography of the neural cell adhesion molecule L1. EMBO J. 4, 3105–3113.

Fang, P., Schachner, M., Shen, Y.-Q., 2012. HMGB1 in development and diseases of the central nervous system. Mol. Neurobiol. 45, 499–506. https://doi.org/10.1007/s12035-012-8264-y

Fransen, E., Schrander-Stumpel, C., Vits, L., Coucke, P., Van Camp, G., Willems, P.J., 1994. X-linked hydrocephalus and MASA syndrome present in one family are due to a single missense mutation in exon 28 of the L1CAM gene. Hum. Mol. Genet. 3, 2255–2256. https://doi.org/10.1093/hmg/3.12.2255

Fransen, E., Van Camp, G., Vits, L., Willems, P.J., 1997. L1-associated diseases: clinical geneticists divide, molecular geneticists unite. Hum. Mol. Genet. 6, 1625–1632. https://doi.org/10.1093/hmg/6.10.1625

Ganesh, K., Basnet, H., Kaygusuz, Y., Laughney, A.M., He, L., Sharma, R., O'Rourke, K.P., Reuter, V.P., Huang, Y.-H., Turkekul, M., Er, E.E., Masilionis, I., Manova-Todorova, K., Weiser, M.R., Saltz, L.B., Garcia-Aguilar, J., Koche, R., Lowe, S.W., Pe'er, D., Shia, J., Massagué, J., 2020. L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer. Nat. Cancer 1, 28–45. https://doi.org/10.1038/s43018-019-0006-x

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y.H., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80. https://doi.org/10.1186/gb-2004-5-10-r80

Giordano, M., Cavallaro, U., 2020. Different Shades of L1CAM in the Pathophysiology of Cancer Stem Cells. J. Clin. Med. 9. https://doi.org/10.3390/jcm9051502

Guo, Y., Feinberg, H., Conroy, E., Mitchell, D.A., Alvarez, R., Blixt, O., Taylor, M.E., Weis, W.I., Drickamer, K., 2004. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. Nat. Struct. Mol. Biol. 11, 591–598. https://doi.org/10.1038/nsmb784

H. Pagès et al., 2020. Biostrings: Efficient manipulation of biological strings.

Hall, H., Deutzmann, R., Timpl, R., Vaughan, L., Schmitz, B., Schachner, M., 1997. HNK-1 carbohydrate-mediated cell adhesion to laminin-1 is different from heparin-mediated and sulfatide-mediated cell adhesion. Eur. J. Biochem. 246, 233–242. https://doi.org/10.1111/j.1432-1033.1997.t01-1-00233.x

Hall, H., Liu, L., Schachner, M., Schmitz, B., 1993. The L2/HNK-1 carbohydrate mediates adhesion of neural cells to laminin. Eur. J. Neurosci. 5, 34–42. https://doi.org/10.1111/j.1460-9568.1993.tb00202.x

Hall, H., Vorherr, T., Schachner, M., 1995. Characterization of a 21 amino acid peptide sequence of the laminin G2 domain that is involved in HNK-1 carbohydrate binding and cell adhesion. Glycobiology 5, 435–441. https://doi.org/10.1093/glycob/5.4.435

Han, J., Guo, X., Tan, W., Zhang, F., Liu, J., Wang, W., Xu, P., Lammi, M.J., 2013. The expression of p-ATF2 involved in the chondeocytes apoptosis of an endemic osteoarthritis, Kashin-Beck disease. BMC Musculoskelet. Disord. 14, 209. https://doi.org/10.1186/1471-2474-14-209

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28, 100–108. https://doi.org/10.2307/2346830

Hennen, E., Safina, D., Haussmann, U., Wörsdörfer, P., Edenhofer, F., Poetsch, A., Faissner, A., 2013. A LewisX Glycoprotein Screen Identifies the Low Density Lipoprotein Receptor-related Protein 1 (LRP1) as a Modulator of Oligodendrogenesis in Mice*,. J. Biol. Chem. 288, 16538–16545. https://doi.org/10.1074/jbc.M112.419812

Hickman, D.L., Johnson, J., Vemulapalli, T.H., Crisler, J.R., Shepherd, R., 2017. Commonly Used Animal Models. Princ. Anim. Res. Grad. Undergrad. Stud. 117–175. https://doi.org/10.1016/B978-0-12-802151-4.00007-4

Hlavin, M.L., Lemmon, V., 1991. Molecular structure and functional testing of human L1CAM: an interspecies comparison. Genomics 11, 416–423. https://doi.org/10.1016/0888-7543(91)90150-d

Hohenester, E., 2019. Laminin G-like domains: dystroglycan-specific lectins. Curr. Opin. Struct. Biol., Sequences and Topology ● Carbohydrates 56, 56–63. https://doi.org/10.1016/j.sbi.2018.11.007

Hortsch, M., 2000. Structural and functional evolution of the L1 family: are four adhesion molecules better than one? Mol. Cell. Neurosci. 15, 1–10. https://doi.org/10.1006/mcne.1999.0809

Hortsch, M., Nagaraj, K., Mualla, R., 2014. The L1 family of cell adhesion molecules: a sickening number of mutations and protein functions. Adv. Neurobiol. 8, 195–229. https://doi.org/10.1007/978-1-4614-8090-7_9

Ielasi, F.S., Alioscha-Perez, M., Donohue, D., Claes, S., Sahli, H., Schols, D., Willaert, R.G., 2016. Lectin-Glycan Interaction Network-Based Identification of Host Receptors of Microbial Pathogenic Adhesins. mBio 7. https://doi.org/10.1128/mBio.00584-16

Jouet, M., Rosenthal, A., Armstrong, G., MacFarlane, J., Stevenson, R., Paterson, J., Metzenberg, A., Ionasescu, V., Temple, K., Kenwrick, S., 1994. X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked hydrocephalus result from mutations in the L1 gene. Nat. Genet. 7, 402–407. https://doi.org/10.1038/ng0794-402

Kalus, I., Schnegelsberg, B., Seidah, N.G., Kleene, R., Schachner, M., 2003. The proprotein convertase PC5A and a metalloprotease are involved in the proteolytic processing of the neural adhesion molecule L1. J. Biol. Chem. 278, 10381–10388. https://doi.org/10.1074/jbc.M208351200

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M., Altman, D.G., 2010. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLOS Biol. 8, e1000412. https://doi.org/10.1371/journal.pbio.1000412

Kizuka, Y., Oka, S., 2012. Regulated expression and neural functions of human natural killer-1 (HNK-1) carbohydrate. Cell. Mol. Life Sci. CMLS 69, 4135–4147. https://doi.org/10.1007/s00018-012-1036-z

Kleene, R., Lutz, D., Loers, G., Bork, U., Borgmeyer, U., Hermans-Borgmeyer, I., Schachner, M., 2020. Revisiting the proteolytic processing of cell adhesion molecule L1. J. Neurochem. n/a. https://doi.org/10.1111/jnc.15201

Kleene, R., Schachner, M., 2004. Glycans and neural cell interactions. Nat. Rev. Neurosci. 5, 195–208. https://doi.org/10.1038/nrn1349

Kraus, K., Kleene, R., Braren, I., Loers, G., Lutz, D., Schachner, M., 2018a. A fragment of adhesion molecule L1 is imported into mitochondria, and regulates mitochondrial metabolism and trafficking. J. Cell Sci. 131. https://doi.org/10.1242/jcs.210500

Kraus, K., Kleene, R., Henis, M., Braren, I., Kataria, H., Sharaf, A., Loers, G., Schachner, M., Lutz, D., 2018b. A Fragment of Adhesion Molecule L1 Binds to Nuclear Receptors to Regulate Synaptic Plasticity and Motor Coordination. Mol. Neurobiol. 55, 7164–7178. https://doi.org/10.1007/s12035-018-0901-7

Kudumala, S., Freund, J., Hortsch, M., Godenschwege, T.A., 2013. Differential effects of human L1CAM mutations on complementing guidance and synaptic defects in Drosophila melanogaster. PloS One 8, e76974. https://doi.org/10.1371/journal.pone.0076974

Kumar, K.K., Chandra, K.L.P., Sumanthi, J., Reddy, G.S., Shekar, P.C., Reddy, B.V.R., 2012. Biological role of lectins: A review. J. Orofac. Sci. 4, 20. https://doi.org/10.4103/0975-8844.99883

Lau, E., Sedy, J., Sander, C., Shaw, M.A., Feng, Y., Scortegagna, M., Claps, G., Robinson, S., Cheng, P., Srivas, R., Soonthornvacharin, S., Ideker, T., Bosenberg, M., Gonzalez, R., Robinson, W., Chanda, S., Ware, C., Dummer, R., Hoon, D., Kirkwood, J.M., Ronai, Z.A., 2015. Transcriptional repression of IFNβ1 by ATF2 confers melanoma resistance to therapy. Oncogene 34, 5739–5748. https://doi.org/10.1038/onc.2015.22

Letunic, I., Bork, P., 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 46, D493–D496. https://doi.org/10.1093/nar/gkx922

Liu, Y., Qin, S., Lei, M., Tempel, W., Zhang, Y., Loppnau, P., Li, Y., Min, J., 2017. Peptide recognition by heterochromatin protein 1 (HP1) chromoshadow domains revisited: Plasticity in the pseudosymmetric histone binding site of human HP1. J. Biol. Chem. 292, 5655–5664. https://doi.org/10.1074/jbc.M116.768374

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif 25, 402–408. https://doi.org/10.1006/meth.2001.1262

Löw, K., Orberger, G., Schmitz, B., Martini, R., Schachner, M., 1994. The L2/HNK-1 carbohydrate is carried by the myelin associated glycoprotein and sulphated glucuronyl glycolipids in muscle but not cutaneous nerves of adult mice. Eur. J. Neurosci. 6, 1773–1781. https://doi.org/10.1111/j.1460-9568.1994.tb00570.x

Lutz, D., Loers, G., Kleene, R., Oezen, I., Kataria, H., Katagihallimath, N., Braren, I., Harauz, G., Schachner, M., 2014. Myelin Basic Protein Cleaves Cell Adhesion Molecule L1 and Promotes Neuritogenesis and Cell Survival. J. Biol. Chem. 289, 13503–13518. https://doi.org/10.1074/jbc.M113.530238

Lutz, D., Wolters-Eisfeld, G., Joshi, G., Djogo, N., Jakovcevski, I., Schachner, M., Kleene, R., 2012. Generation and Nuclear Translocation of Sumoylated Transmembrane Fragment of Cell Adhesion Molecule L1. J. Biol. Chem. 287, 17161–17175. https://doi.org/10.1074/jbc.M112.346759

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. Presented at the Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, The Regents of the University of California.

Maekawa, T., Bernier, F., Sato, M., Nomura, S., Singh, M., Inoue, Y., Tokunaga, T., Imai, H., Yokoyama, M., Reimold, A., Glimcher, L.H., Ishii, S., 1999. Mouse ATF-2 null mutants display features of a severe type of meconium aspiration syndrome. J. Biol. Chem. 274, 17813–17819. https://doi.org/10.1074/jbc.274.25.17813

Maness, P.F., Schachner, M., 2007. Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. Nat. Neurosci. 10, 19–26. https://doi.org/10.1038/nn1827

Marín, R., Ley-Martos, M., Gutiérrez, G., Rodríguez-Sánchez, F., Arroyo, D., Mora-López, F., 2015. Three cases with L1 syndrome and two novel mutations in the L1CAM gene. Eur. J. Pediatr. 174, 1541–1544. https://doi.org/10.1007/s00431-015-2560-2

Martini, R., Schachner, M., Brushart, T.M., 1994. The L2/HNK-1 carbohydrate is preferentially expressed by previously motor axon-associated Schwann cells in reinnervated peripheral nerves. J. Neurosci. Off. J. Soc. Neurosci. 14, 7180–7191.

Mattout, A., Aaronson, Y., Sailaja, B.S., Raghu Ram, E.V., Harikumar, A., Mallm, J.-P., Sim, K.H., Nissim-Rafinia, M., Supper, E., Singh, P.B., Sze, S.K., Gasser, S.M., Rippe, K., Meshorer, E., 2015. Heterochromatin Protein 1β (HP1β) has distinct functions and distinct nuclear distribution in pluripotent versus differentiated cells. Genome Biol. 16, 213. https://doi.org/10.1186/s13059-015-0760-8

Meehan, R.R., Kao, C.-F., Pennings, S., 2003. HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. EMBO J. 22, 3164–3174. https://doi.org/10.1093/emboj/cdg306

Minc, E., Allory, Y., Courvalin, J.C., Buendia, B., 2001. Immunolocalization of HP1 proteins in metaphasic mammalian chromosomes. Methods Cell Sci. Off. J. Soc. Vitro Biol. 23, 171–174. https://doi.org/10.1007/978-94-010-0330-8_18

Minc, E., Allory, Y., Worman, H.J., Courvalin, J.C., Buendia, B., 1999. Localization and phosphorylation of HP1 proteins during the cell cycle in mammalian cells. Chromosoma 108, 220–234. https://doi.org/10.1007/s004120050372

Minc, E., Courvalin, J.C., Buendia, B., 2000. HP1gamma associates with euchromatin and heterochromatin in mammalian nuclei and chromosomes. Cytogenet. Cell Genet. 90, 279–284. https://doi.org/10.1159/000056789

Moos, M., Tacke, R., Scherer, H., Teplow, D., Früh, K., Schachner, M., 1988. Neural adhesion molecule L1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin. Nature 334, 701–703. https://doi.org/10.1038/334701a0

Morise, J., Takematsu, H., Oka, S., 2017. The role of human natural killer-1 (HNK-1) carbohydrate in neuronal plasticity and disease. Biochim. Biophys. Acta Gen. Subj. 1861, 2455–2461. https://doi.org/10.1016/j.bbagen.2017.06.025

Morita, I., Kakuda, S., Takeuchi, Y., Itoh, S., Kawasaki, N., Kizuka, Y., Kawasaki, T., Oka, S., 2009. HNK-1 glyco-epitope regulates the stability of the glutamate receptor subunit GluR2 on the neuronal cell surface. J. Biol. Chem. 284, 30209–30217. https://doi.org/10.1074/jbc.M109.024208

Mualla, R., Nagaraj, K., Hortsch, M., 2013. A phylogenetic analysis of the L1 family of neural cell adhesion molecules. Neurochem. Res. 38, 1196–1207. https://doi.org/10.1007/s11064-012-0892-0

Muchardt, C., Guilleme, M., Seeler, J.-S., Trouche, D., Dejean, A., Yaniv, M., 2002. Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. EMBO Rep. 3, 975–981. https://doi.org/10.1093/embo-reports/kvf194

NCBI, N., 2020. National Center for Biotechnology Information (US). PubMed [WWW Document]. PubMed. URL https://pubmed.ncbi.nlm.nih.gov/

Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lütteke, T., O'Boyle, N., Packer, N.H., Stanley, P., Toukach, P., Varki, A., Woods, R.J., The SNFG Discussion Group, 2019. Updates to the Symbol Nomenclature for Glycans guidelines. Glycobiology 29, 620–624. https://doi.org/10.1093/glycob/cwz045

Nielsen, A.L., Ortiz, J.A., You, J., Oulad-Abdelghani, M., Khechumian, R., Gansmuller, A., Chambon, P., Losson, R., 1999. Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. EMBO J. 18, 6385–6395. https://doi.org/10.1093/emboj/18.22.6385

Oshiro, H., Hirabayashi, Y., Furuta, Y., Okabe, S., Gotoh, Y., 2015. Up-regulation of HP1γ expression during neuronal maturation promotes axonal and dendritic development in mouse embryonic neocortex. Genes Cells Devoted Mol. Cell. Mech. 20, 108–120. https://doi.org/10.1111/gtc.12205

Pröpster, J.M., Yang, F., Rabbani, S., Ernst, B., Allain, F.H.-T., Schubert, M., 2016. Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8. Proc. Natl. Acad. Sci. U. S. A. 113, E4170-4179. https://doi.org/10.1073/pnas.1602214113

Rauvala, H., Rouhiainen, A., 2010. Physiological and pathophysiological outcomes of the interactions of HMGB1 with cell surface receptors. Biochim. Biophys. Acta 1799, 164–170. https://doi.org/10.1016/j.bbagrm.2009.11.012

Reid, R.A., Hemperly, J.J., 1992. Variants of human L1 cell adhesion molecule arise through alternate splicing of RNA. J. Mol. Neurosci. MN 3, 127–135. https://doi.org/10.1007/BF02919404

Riedle, S., Kiefel, H., Gast, D., Bondong, S., Wolterink, S., Gutwein, P., Altevogt, P., 2009. Nuclear translocation and signalling of L1-CAM in human carcinoma cells requires ADAM10 and presenilin/γ-secretase activity. Biochem. J. 420, 391–402. https://doi.org/10.1042/BJ20081625

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47. https://doi.org/10.1093/nar/gkv007

Roonprapunt, C., Huang, W., Grill, R., Friedlander, D., Grumet, M., Chen, S., Schachner, M., Young, W., 2003. Soluble cell adhesion molecule L1-Fc promotes locomotor recovery in rats after spinal cord injury. J. Neurotrauma 20, 871–882. https://doi.org/10.1089/089771503322385809

Rudalska, R., Dauch, D., Longerich, T., McJunkin, K., Wuestefeld, T., Kang, T.-W., Hohmeyer, A., Pesic, M., Leibold, J., von Thun, A., Schirmacher, P., Zuber, J., Weiss, K.-H., Powers, S., Malek, N.P., Eilers, M., Sipos, B., Lowe, S.W., Geffers, R., Laufer, S., Zender, L., 2014. In vivo RNAi screening identifies a mechanism of sorafenib resistance in liver cancer. Nat. Med. 20, 1138–1146. https://doi.org/10.1038/nm.3679

Sahu, S., Li, R., Kadeyala, P.K., Liu, S., Schachner, M., 2018. The human natural killer-1 (HNK-1) glycan mimetic ursolic acid promotes functional recovery after spinal cord injury in mouse. J. Nutr. Biochem. 55, 219–228. https://doi.org/10.1016/j.jnutbio.2018.01.016

Sajdel-Sulkowska, E.M., 1998. Immunofluorescent detection of CD15-fucosylated glycoconjugates in primary cerebellar cultures and their function in glial-neuronal adhesion in the central nervous system. Acta Biochim. Pol. 45, 781–790.

Saugier-Veber, P., Martin, C., Le Meur, N., Lyonnet, S., Munnich, A., David, A., Hénocq, A., Héron, D., Jonveaux, P., Odent, S., Manouvrier, S., Moncla, A., Morichon, N., Philip, N., Satge, D., Tosi, M., Frébourg, T., 1998. Identification of novel L1CAM mutations using fluorescence-assisted mismatch analysis. Hum. Mutat. 12, 259–266. https://doi.org/10.1002/(SICI)1098-1004(1998)12:4<259::AID-HUMU7>3.0.CO;2-A

Sb, N., Cd, W., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453. https://doi.org/10.1016/0022-2836(70)90057-4

Schachner, M., Martini, R., Hall, H., Orberger, G., 1995. Functions of the L2/HNK-1 carbohydrate in the nervous system. Prog. Brain Res. 105, 183–188. https://doi.org/10.1016/s0079-6123(08)63294-x

Schäfer, M.K.E., Nam, Y.-C., Moumen, A., Keglowich, L., Bouché, E., Küffner, M., Bock, H.H., Rathjen, F.G., Raoul, C., Frotscher, M., 2010. L1 syndrome mutations impair neuronal L1 function at different levels by divergent mechanisms. Neurobiol. Dis. 40, 222–237. https://doi.org/10.1016/j.nbd.2010.05.029

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Schmitz, B., Schachner, M., Ito, Y., Nakano, T., Ogawa, T., 1994. Determination of structural elements of the L2/HNK-1 carbohydrate epitope required for its function. Glycoconj. J. 11, 345–352. https://doi.org/10.1007/BF00731208

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. Proc. Natl. Acad. Sci. 95, 5857–5864. https://doi.org/10.1073/pnas.95.11.5857

Senat, M.V., Bernard, J.P., Delezoide, A., Saugier-Veber, P., Hillion, Y., Roume, J., Ville, Y., 2001. Prenatal diagnosis of hydrocephalus-stenosis of the aqueduct of Sylvius by ultrasound in the first trimester of pregnancy. Report of two cases. Prenat. Diagn. 21, 1129–1132. https://doi.org/10.1002/pd.184

Sharon, N., Lis, H., 2004. History of lectins: from hemagglutinins to biological recognition molecules. Glycobiology 14, 53R-62R. https://doi.org/10.1093/glycob/cwh122

Simova, O., Irintchev, A., Mehanna, A., Liu, J., Dihné, M., Bächle, D., Sewald, N., Loers, G., Schachner, M., 2006. Carbohydrate mimics promote functional recovery after peripheral nerve repair. Ann. Neurol. 60, 430–437. https://doi.org/10.1002/ana.20948

Smothers, J.F., Henikoff, S., 2000. The HP1 chromo shadow domain binds a consensus peptide pentamer. Curr. Biol. CB 10, 27–30. https://doi.org/10.1016/s0960-9822(99)00260-2

Somers, W.S., Tang, J., Shaw, G.D., Camphausen, R.T., 2000. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. Cell 103, 467–479. https://doi.org/10.1016/s0092-8674(00)00138-0

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., Carey, M., Garcia, B.A., Plath, K., 2013. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat. Cell Biol. 15, 872–882. https://doi.org/10.1038/ncb2768

Streit, A., Faissner, A., Gehrig, B., Schachner, M., 1990. Isolation and biochemical characterization of a neural proteoglycan expressing the L5 carbohydrate epitope. J. Neurochem. 55, 1494–1506. https://doi.org/10.1111/j.1471-4159.1990.tb04931.x

Stumpel, C., Vos, Y.J., 1993. L1 Syndrome, in: Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J., Stephens, K., Amemiya, A. (Eds.), GeneReviews®. University of Washington, Seattle, Seattle (WA).

Sytnyk, V., Leshchyns'ka, I., Schachner, M., 2017. Neural Cell Adhesion Molecules of the Immunoglobulin Superfamily Regulate Synapse Formation, Maintenance, and Function. Trends Neurosci. 40, 295–308. https://doi.org/10.1016/j.tins.2017.03.003

Tagliavacca, L., Colombo, F., Racchetti, G., Meldolesi, J., 2013. L1CAM and its cell-surface mutants: new mechanisms and effects relevant to the physiology and pathology of neural cells. J. Neurochem. 124, 397–409. https://doi.org/10.1111/jnc.12015

Takeda, Y., Asou, H., Murakami, Y., Miura, M., Kobayashi, M., Uyemura, K., 1996. A nonneuronal isoform of cell adhesion molecule L1: tissue-specific expression and functional analysis. J. Neurochem. 66, 2338–2349. https://doi.org/10.1046/j.1471-4159.1996.66062338.x

Tapanes-Castillo, A., Weaver, E.J., Smith, R.P., Kamei, Y., Caspary, T., Hamilton-Nelson, K.L., Slifer, S.H., Martin, E.R., Bixby, J.L., Lemmon, V.P., 2010. A modifier locus on chromosome 5 contributes to L1 cell adhesion molecule X-linked hydrocephalus in mice. Neurogenetics 11, 53–71. https://doi.org/10.1007/s10048-009-0203-3

The UniProt Consortium, 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515. https://doi.org/10.1093/nar/gky1049

Thiru, A., Nietlispach, D., Mott, H.R., Okuwaki, M., Lyon, D., Nielsen, P.R., Hirshberg, M., Verreault, A., Murzina, N.V., Laue, E.D., 2004. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. EMBO J. 23, 489–499. https://doi.org/10.1038/sj.emboj.7600088

Thliveris, A.T., Clipson, L., Sommer, L.L., Schoenike, B.A., Hasenstein, J.R., Schlamp, C.L., Alexander, C.M., Newton, M.A., Dove, W.F., Amos-Landgraf, J.M., 2012. Regulated Expression of Chromobox Homolog 5 Revealed in Tumors of ApcMin/+ ROSA11 Gene Trap Mice. G3 GenesGenomesGenetics 2, 569–578. https://doi.org/10.1534/g3.112.002436

Varki, A., Cummings, R.D., Aebi, M., Packer, N.H., Seeberger, P.H., Esko, J.D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J.J., Schnaar, R.L., Freeze, H.H., Marth, J.D., Bertozzi, C.R., Etzler, M.E., Frank, M., Vliegenthart, J.F., Lütteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K.F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., Kornfeld, S., 2015. Symbol Nomenclature for Graphical Representations of Glycans. Glycobiology 25, 1323–1324. https://doi.org/10.1093/glycob/cwv091

Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (Eds.), 2009. Essentials of Glycobiology, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

Vos, Y.J., de Walle, H.E.K., Bos, K.K., Stegeman, J.A., Ten Berge, A.M., Bruining, M., van Maarle, M.C., Elting, M.W., den Hollander, N.S., Hamel, B., Fortuna, A.M., Sunde, L.E.M., Stolte-Dijkstra, I., Schrander-Stumpel, C.T.R.M., Hofstra, R.M.W., 2010. Genotype-phenotype correlations in L1 syndrome: a guide for genetic counselling and mutation analysis. J. Med. Genet. 47, 169–175. https://doi.org/10.1136/jmg.2009.071688

Wagih, O., 2017. ggseqlogo: a versatile R package for drawing sequence logos. Bioinforma. Oxf. Engl. 33, 3645–3647. https://doi.org/10.1093/bioinformatics/btx469

Wang, X., Spandidos, A., Wang, H., Seed, B., 2012. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. Nucleic Acids Res. 40, D1144-1149. https://doi.org/10.1093/nar/gkr1013

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686

Yagi, H., Saito, T., Yanagisawa, M., Yu, R.K., Kato, K., 2012. Lewis X-carrying N-glycans regulate the proliferation of mouse embryonic neural stem cells via the Notch signaling pathway. J. Biol. Chem. 287, 24356–24364. https://doi.org/10.1074/jbc.M112.365643

Yan, H., Xiang, X., Chen, Q., Pan, X., Cheng, H., Wang, F., 2018. HP1 cooperates with CAF-1 to compact heterochromatic transgene repeats in mammalian cells. Sci. Rep. 8, 14141. https://doi.org/10.1038/s41598-018-32381-7

Yoshimura, T., Hayashi, A., Handa-Narumi, M., Yagi, H., Ohno, N., Koike, T., Yamaguchi, Y., Uchimura, K., Kadomatsu, K., Sedzik, J., Kitamura, K., Kato, K., Trapp, B.D., Baba, H., Ikenaka, K., 2017. GlcNAc6ST-1 regulates sulfation of N-glycans and myelination in the peripheral nervous system. Sci. Rep. 7. https://doi.org/10.1038/srep42257

Zaidan, N.Z., Sridharan, R., 2020. HP1γ regulates H3K36 methylation and pluripotency in embryonic stem cells. Nucleic Acids Res. 48, 12660–12674. https://doi.org/10.1093/nar/gkaa1091

10. Appendix

List of hazardous substances according to GHS

Substance name	CAS number	GHS Pictogram	Hazard statements	Precautionary statements
acrylamide-bis acrylamide	110-26-9		H302 H315 H317 H319 H340 H350 H361f H372	P201 P280 P301+P312 P302+P352 P305+P351+P338 P308+P313
Ammonium persulfate	7727-54-0		H273 H302 H315 H319 H317 H334 H335	P210 P221 P285 P302+P351+P338 P405 P501
Ampicillin	69-52-3		H317 H334	P261 P280 P342+P311
Beta-Mercaptoethanol	60-24-2		H301+H331 H310 H315 H317 H318 H361d H373 H410	P201 P262 P280 P301+P310+P330 P302+P352+P310 P305+P351+P338+P310
Bromophenol blue	115-39-9		H332 H302 H319	P261 P264 P280 P304+340 P312 P301+312 P330 P305+351+338 P337+313
DAPI	28718-90-3		H315 H317 H335	P261 P264 P271 P272 P280 P302+P352 P304+P340 P312 P321 P332+P313 P333+P313 P362 P363 P403+P233 P405 P501
DTE	6892-68-8		H315 H319 H335	P261 P264 P271 P280 P302+P352 P304+P340

				P305+P351+P338 P312 P321 P332+P313 P337+P313 P362 P403+P233 P405 P501
EDTA	6381-92-6		H302 H312 H315 H319 H332 H335 H373 H412	P260 P261 P264 P270 P271 P273 P280 P301+P312 P302+P352 P304+P312 P304+P340 P305+P351+P338 P312 P314 P321 P322 P330 P332+P313 P337+P313 P362 P363 P403+P233 P405 P501
Formaldehyde	50-00-0		H301 H311 H314 H317 H331 H341 H350	P201 P202 P260 P261 P264 P270 P271 P272 P280 P281 P301+P310 P301+P330+P331 P302+P352 P303+P361+P353 P304+P340 P305+P351+P338 P308+P313 P310 P311 P312 P321 P322 P330 P333+P313 P361 P363 P403+P233 P405 P501
HCL	7647-01-0		H314 H331	P260 P261 P264 P271 P280 P301+P330+P331 P303+P361+P353 P304+P340 P305+P351+P338 P310 P311 P321 P363 P403+P233 P405 P501
Imidazole	288-32-4		H302 H314 H360D	P201 P202 P260 P264 P270 P280 P281 P301+P312 P301+P330+P331 P303+P361+P353 P304+P340

				P305+P351+P338 P308+P313 P310 P321 P330 P363 P405 P501
Methanol	67-56-1		H225 H301 H331 H311 H370	P210 P233 P240 P241 P242 P243 P260 P261 P264 P270 P271 P280 P301+P310 P302+P352 P303+P361+P353 P304+P340 P307+P311 P311 P312 P321 P322 P330 P361 P363 P370+P378 P403+P233 P403+P235 P405 P501
NaOH	1310-73-2		H314	P260 P264 P280 P301+P330+P331 P303+P361+P353 P304+P340 P305+P351+P338 P310 P321 P363 P405 P501
OPD	95-54-5		H301 H312 H317 H319 H332 H341 H351 H400 H410	P201 P202 P261 P264 P270 P271 P272 P273 P280 P281 P301+P310 P302+P352 P304+P312 P304+P340 P305+P351+P338 P308+P313 P312 P321 P322 P330 P333+P313 P337+P313 P363 P391 P405 P501
PMSF	329-98-6		H301 H314	P260 P264 P270 P280 P301+P310 P301+P330+P331 P303+P361+P353 P304+P340 P305+P351+P338 P310 P321 P330 P363 P405 P501

SDS	151-21-3		H228 H302 H318 H315 H319 H332 H335 H412	P210 P240 P241 P261 P264 P270 P271 P273 P280 P301+P312 P302+P352 P304+P312 P304+P340 P305+P351+P338 P310 P312 P321 P330 P332+P313 P337+P313 P362 P370+P378 P403+P233 P405 P501
Sodium azide	26628-22-8		H300 H400 H410	P264 P270 P273 P301+P310 P321 P330 P391 P405 P501
Tetramethylethylenediamine	110-18-9		H225 H302 H314 H332	P210 P233 P240 P241 P242 P243 P260 P261 P264 P270 P271 P280 P301+P312 P301+P330+P331 P303+P361+P353 P304+P312 P304+P340 P305+P351+P338 P310 P312 P321 P330 P363 P370+P378 P403+P235 P405 P501

11. Acknowledgments

I express my gratitude to Prof. Dr. Dr. h. c. Melitta Schachner for giving me the opportunity to write this thesis in her laboratory. I am most grateful for her role as a scientific mentor as well as a character mentor which has helped me in my professional life.

I would also like to express my gratitude to Prof. Dr. Andrew Torda for giving me the tools and opportunity to develop my skills in bioinformatics. Thank you for always being supportive and sharing your knowledge.

I would like to thank Dr. Gabriele Loers and Dr. Ralf Kleene for their guidance and good supervision. Thank you for all, from teaching lab techniques to advising the follow-up of the experiments. Your helpful commentaries during the presentation of results, written communications of results, and during the writing of our publication were key during my thesis. I am also very grateful to Ute Bork for the useful technical pieces of advice.

I am grateful to my previous colleagues, Agnieszka Kotarska, Jelena Katic, Kristina Kraus, Luciana Fernandes, Maria Girbes-Minguez, Nina Westphal, Patricia García-Jareño, Viviana Granato, and Volodimir Serdiuk for their teamwork and friendship.
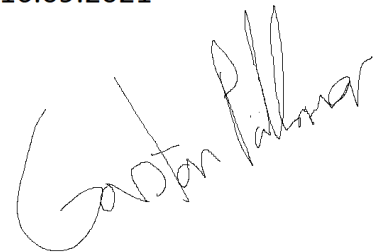
Finally, I would like to thank my wife for all her support.

12.   Declaration on Oath*:

„I hereby declare on oath that this doctoral dissertation is written independently and solely by my own based on the original work of my PhD and has not been used other than the acknowledged resources and aids. The submitted written version corresponds to the version on the electronic storage medium. I declare that the present dissertation was prepared maintaining the Rules of Good Scientific Practice of the German Research Foundation and it has never been submitted in the present form or similar to any other University or board of examiners..."

Date, Signature

16.09.2021