A peptide-based interaction screen on disease-related mutations

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

im Fach Biophysik

eingereicht an der

Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von

M.Sc. Katrina Meyer

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
Prof. Dr. Bernhard Grimm

Gutachter:
1. Prof. Dr. Andreas Herrmann
2. Prof. Dr. Matthias Selbach
3. Prof. Dr. Markus Landthaler

Tag der mündlichen Prüfung: 11.02.2019

Alles Interesse an Krankheit und Tod ist nur ein anderer Ausdruck für das Interesse am Leben.

Thomas Mann 1875 - 1955

Table of contents

Zusammenfassung

Zahlreiche pathogene *„missense*"-Mutation führen dazu, dass Proteine nicht korrekt gefaltet werden können und dadurch ihre Funktionalität verlieren. Diese Mutationen finden sich häufig im hydrophoben Kern oder in geordneten Regionen von Proteinen. Eine andere Klasse von krankheitsrelevanten Mutationen befindet sich jedoch in ungeordneten Proteinregionen und beeinflusst somit wahrscheinlich nur begrenzt die Funktionalität, zum Beispiel durch Veränderungen kurzer linearer Sequenzmotive, die Protein-Protein Interaktionen vermitteln. In dieser Arbeit wird ein peptidbasierter Interaktionsscreen präsentiert mit dem sich Veränderungen im Interaktom identifizieren lassen. Die vorliegende Arbeit konzentriert sich dabei auf neurologische Krankheiten. Synthetische Peptide von wild-typ und zugehörigen mutierten Proteinregionen, die auf Zellulosemembran gespottet wurden, ermöglichen die gleichzeitige Untersuchung von mehr als hundert Mutationen mittels Massenspektrometrie. SILAC-basierte Quantifizierung ermöglicht den Vergleich von Interaktionspartnern von wild-typ und mutierten Peptiden. Mehr als ein Drittel aller getesteten Mutationen hatten veränderte Interaktionen zur Folge. Darunter befanden sich auch drei Prolin zu Leucin Mutationen in zytosolischen Regionen von Transmembranproteinen, die zusammen mit dem benachbarten Leucin einem Dileucinmotiv ergeben und dadurch verstärkt mit Clathrin interagieren. Verschiedenste Proteine mit Dileucinemotiven wurden bereits mit Clathrin-vermittelter Endozytose in Verbindung gebracht. Diese hinzugewonnene Endozytose könnte Krankheitsmechanismen erklären, da die Mislokalisation der betroffenen Transmembranproteine zum effektiven Verlust derer Funktion führen würde. Diese Hypothese wurde hier von verschiedenen *in vitro* und *in vivo* Experimenten bezüglich der P485L Mutation im Glukose Transporter-1 (GLUT1), die das GLUT1-Defizit-Syndrom hervorruft, bestätigt. Weitere Evidenz wurde außerdem für die Funktionalität anderer mutationsbedingter Dileucinmotive gewonnen und die systematische Analyse von pathogenen Mutationen hat gezeigt, dass Dileucinmotive signifikant und spezifisch in ungeordneten zytosolischen Regionen von Transmembranproteinen überrepräsentiert sind. Dieser Peptidescreen macht das Potenzial unvoreingenommener Analysen zur Aufklärung von Krankheitsmechanismen deutlich, die von Veränderungen in Protein-Protein Interaktionen hervorgerufen werden.

Summary

Many disease-associated missense mutations prevent proteins from folding correctly and lead to a complete loss-of-function. These mutations are often found in the hydrophobic core or in ordered regions of proteins. Another class of disease-related missense mutations, however, can be found in disordered regions. They are thought to impair only specific parts of a protein's functions. Those mutations could, for example, modify short linear motifs that mediate protein-protein interactions. Here, we designed a peptide-based interaction screen to identify interactions that are affected by disease-associated mutations in disordered regions. We used synthetic peptides corresponding to the wild-type and mutated protein regions surrounding the disease mutation spotted on cellulose membrane to pull-down interaction partners. This setup allows for the screening of more than hundred disease-associated mutations at a time via mass spectrometry. In this thesis, we focused on neurological diseases. SILAC-based quantification allowed us to compare the interaction partners of wild-type peptides and their mutated variants. More than one-third of tested variant pairs show differential interactions. Interestingly, three disease-related proline to leucine mutations in cytosolic tails of transmembrane proteins lead to gain of a dileucine sequence. Several dileucine-containing peptide motifs are involved in clathrin-mediated endocytosis. In line with this, the newly created motifs seem to mediate interaction with the clathrin-mediated endocytosis machinery, also in the presented screen. The gain of endocytosis could explain the disease mechanisms since mislocalization of the affected transmembrane proteins would lead to their loss of function. This hypothesis has been corroborated by several *in vitro* and *in vivo* experiments for glucose transporter-1 (GLUT1) P485L, a mutation causing GLUT1 deficiency syndrome. We were able to provide functional evidence for a set of additional gained dileucine motifs and a systematic analysis of pathogenic mutations revealed dileucine motifs to be significantly and specifically overrepresented in structurally disordered cytosolic regions of transmembrane proteins.

The data gained with the peptide screen highlights the power of differential interactome mapping as a generic approach to unravel disease mechanisms caused by changes in protein-protein interactions.

Statement of contributions

This work would not have been possible without the help and input from many people. I will shortly state here who participated in which parts of the project and mention it again throughout the text for a better overview. Establishment of the screen and selection of the candidates was carried out together with Marieluise Kirchner. Jing-Yuan Cheng conducted her Master`s thesis under my supervision. During that time, she contributed to the GLUT1 part of the project. Markus Landthaler's lab (MDC, Berlin, Germany) generated stable cell lines. Henrik Zauber helped to design and he analyzed the targeted proteomics approach. Validation of peptide-based interaction screen (with far-western blotting approach) was carried out together with Teresa Melder. Bora Uyar (Altuna Akalin lab, MDC, Berlin, Germany) carried out bioinformatics motif based analyses. Giulia Russo (Michael Krauss lab and Volker Haucke lab, FMP Berlin, Germany) carried out GST-pulldown assays. Juan M. Pascual (UT Southwestern Medical Center, Texas, USA) obtained fibroblasts from a GLUT1 deficiency syndrome patient. Sebastian Diecke's team (MDC, Berlin, Germany) generated inducible pluripotent stem cells (iPSCs) from the patient fibroblasts. Experiments on iPSCs were carried out together with Ina-Maria Rudolph (Thomas Willnow lab, MDC, Berlin, Germany). Ralf Kühn (MDC, Berlin, Germany) generated the GLUT1 P485L mouse. Luis R. Hernandez-Miranda (Carmen Birchmeier lab, MDC, Berlin, Germany) was responsible for maintenance of the mouse strain, mouse sample preparation, and consequent confocal imaging. Anna Szymborska (Holger Gerhardt lab, MDC, Berlin, Germany) performed super-resolution imaging on mouse samples. It is important to mention that the project is based on Matthias Selbach's idea.

Introduction

Mankind has, since its existence, suffered through diseases caused by bacteria, virus, fungi, parasites, toxic substances, malnutrition, or biologically based dysfunctions. For the longest time of human existence, however, the reasons for disease have remained elusive. Evil spirits, bad charms, and angry gods have been thought to be the causes and were tried to be fought in the most creative ways. Only after antibiotics had been discovered and the possibility to treat infectious diseases became available in the late nineteenth century, people realized that there must be another cause of disease. Also because unexplainably some diseases used to "run in families". By the 20th century, mysterious family maladies were beginning to be understood as genetic disorders. The first idea of human genetic diseases, in 1908 by Archibald Garrod, was that they were caused by faulty "ferments" (enzymes). Actually, his idea that enzymes, which would soon be identified as proteins, could be the active cause of genetic diseases led to the idea that genetic information could be in some way responsible for the production of proteins (Pasternak 2005). By the time that James D. Watson (b. 1928) and Francis H. C. Crick (1916-2004) discovered the double helical structure of DNA in 1953, researchers established that the information units in the DNA somehow encode information for the synthesis of enzymes - the "one gene - one enzyme" hypothesis. How much these enzymes needed to be changed to cause human genetic disease, though, was still completely elusive. In 1957, the first example to show that the change of a single amino acid could cause disease was hemoglobin where mutation of glutamic acid in position six to a valine would cause sickle cell anemia, a red blood cell deforming disease (Ingram 1957). A quarter of a century passed before researchers became able to sequence DNA and determine the genetic changes underlying disease. The most significant recent advance in human molecular genetics was the publication of the draft version of the human genome sequence in 2001 (Venter 2001).

The human genome can be changed and scrambled in many different ways. Whole blocks of DNA can be inserted or deleted, which, for chunks of more than 1,000 nucleotides, is called a copy number variation. The more common form of variation, however, are differences in the nucleotide composition at a single

position. Small insertions or deletions (indels) can have a big effect on the following protein sequence even if only one or two base pairs are inserted because they lead to frameshifts. Understanding the genetic code also made clear that mutations of single nucleotides in the protein coding regions can have different outcomes. They can either leave the resulting protein sequence intact, i.e. synonymous or silent mutations, or change the codons in a way that different amino acids are integrated, i.e. nonsynonymous or missense mutations. This is because the genetic code is degenerate which means that many amino acids are encoded by multiple codons.

The first platform in which disease-causing variants of the genome have been cataloged is the now called Online Mendelian Inheritance in Man (OMIM). It has been first established in 1966 by Victor McKusick (1921-2008), by then containing a mere 1,500 entries compared to the information about all known Mendelian diseases, and over 15,000 genes nowadays. This information has been partially integrated into the Humsavar database storing disease-causing genetic missense mutations in human UniProtKB/Swiss-Prot Entries (Famiglietti et al. 2014).

In many cases, it is not known whether newly discovered variants cause disease, and even when associations have been established, untangling the molecular mechanisms is often not a straightforward task (Cooper and Shendure 2011). Hence, disease mechanisms are well characterized only for very few of these entries, which leaves over 100,000 disease-associated variants without known functional annotation (Sahni et al. 2015).

Disease mutations in order and disorder

On a protein-structure centric point of view, missense mutations can fall into two different protein regions: Either into ordered/globular structured regions of proteins or into natively unstructured segments, so-called intrinsically disordered regions (IDRs). In the first case, pathogenic mutations likely cause disease due to loss of function of the protein by disrupting three-dimensional folding. In fact, most disease-causing missense mutations affect evolutionarily conserved amino acids within structured regions of proteins and destabilize their structure (Subramanian and Kumar 2006; Yue, Li, and Moult 2005). However,

22% of human disease mutations occur in IDRs (Uyar et al. 2014; Vacic and Iakoucheva 2012). Since mutations in IDRs probably do not alter the protein structure, explaining how mutations in disordered regions can cause disease is less straightforward.

The role of intrinsically disordered protein regions

Traditionally, the function of a protein has been thought to depend on a well-defined and folded three-dimensional structure of the polypeptide chain. This so-called structure-function paradigm is still prevalent in basic biology and biochemistry textbooks, and the abundance and functional significance of protein disorder in eukaryotes were largely unrecognized before the mid-1990s (Wright and Dyson 2015). Intrinsically disordered portions of proteins are often just thought to be passive, connective strings that hold the functional domains together. On the contrary to this common belief, it is nowadays well established that IDRs actively participate in different functions of proteins (Tompa 2011). While some proteins are predicted to be entirely disordered others might contain only some IDRs. Taken together they are usually referred to as intrinsically disordered proteins (IDPs). Almost one-third of the human proteome contains an IDR of at least 30 amino acids (van der Lee et al. 2014). Many IDPs have been collected into the DisProt database which provides the largest collection of proteins with disordered regions (Sickmeier et al. 2007).

However, protein disorder is a property that is mainly predicted from the protein sequence. This possibility has been first discovered by the group of Keith Dunker. They noted from a handful of examples that disordered protein regions were generally enriched in polar and charged amino acids and depleted in hydrophobic amino acids (Xie et al. 1998). Since then, more than 50 different disorder prediction methods have been developed by various research groups (He et al. 2009; Dosztanyi, Meszaros, and Simon 2009). One popular method is IUPred, which provides a prediction of protein disorder based on an energy estimation approach (Dosztányi et al. 2005; Dosztányi 2018).

IDPs are particularly associated with hubs in protein-interaction networks and play important roles in transcription, translation, signaling, and the cell cycle, all cellular processes that need flexible and fast adaptation (Tompa 2002; Wright

and Dyson 2015). Although there are exceptions to the rule: In a recent study, two IDPs have been found to interact with surprisingly strong affinity and without the formation of any transient secondary structure. The linker histone H1 and its nuclear chaperone prothymosin-alpha acquire this strong interaction via long-range electrostatic interactions that lead to rapid interconversion between different combinations of oppositely charged residues (Borgia et al. 2018).

Disorder shows a sharp increase associated with the transition from prokaryotic to eukaryotic cells. This suggests that the increased disorder content in eukaryotic proteomes might be used by nature to deal with the increased cell complexity due to the appearance of the various cellular compartments (Xue, Dunker, and Uversky 2012).

More and more IDPs are being revealed to play important functions in unexpected areas of life: they save tardigrades from desiccation (Boothby et al. 2017) or they assemble into a 3D extracellular organic matrix to form tooth enamel from first tetrapods to man (Wald et al. 2017).

The importance of disordered proteins is also underlined by the tight control on their transcriptional and translational levels. Altered abundances of IDPs are in fact connected to several diseases. Disordered regions are prone to make promiscuous molecular interactions when their concentration is increased, and it has been shown that this is the likely cause of pathology when genes are overexpressed (Vavouri et al. 2009; Babu et al. 2011). Fidelity in signaling may require that most IDPs are available in appropriate amounts and not present longer than needed (Gsponer et al. 2008).

IDPs are characterized by their biased amino acid composition, favoring charged and hydrophilic amino acids over bulky hydrophobic ones, and a general low sequence complexity (Wright and Dyson 2015). This makes them unable to form a well-organized hydrophobic core which is needed to form a structured domain (Uversky, Gillespie, and Fink 2000; Romero et al. 2001). It is the physical characteristics of IDPs that lead to their important roles in cellular signaling processes: A degree of flexibility, which enables them to interact promiscuously with different targets on different occasions and accessible sites for post-

translational modification (Wright and Dyson 2015). In fact, they are frequently modified by post-translational modifications (PTMs) (Iakoucheva et al. 2004; Collins et al. 2008). However, the one most important feature that gives IDPs the possibility to interact with other proteins is the presence of small recognition elements.

Intrinsically disordered regions can harbor interaction motifs

Disordered regions of signaling and regulatory proteins frequently contain multiple conserved sequence motifs that interact with nucleic acids or globular domains in other proteins (Dyson and Wright 2005, 2002). These so-called short linear motifs (SLiMs) are usually of about 3-11 amino acids in length (Diella et al. 2008; Davey et al. 2012; Dinkel et al. 2014). The average binding motif is 6–7 amino acids in length, with only 3–4 core positions conferring the majority of the interaction specificity (Tompa et al. 2014) and there are more than 200 domains known to interact with SLiMs (Stein, Mosca, and Aloy 2011).

Some of the earliest SLiMs to be defined have their function in cell cycle regulation. The original definition of linear motifs was provided by Tim Hunt (Hunt 1990):

"These motifs are linear, in the sense that three-dimensional organization is not required to bring distant segments of the molecule together to make the recognizable unit. The conservation of these motifs varies: some are highly conserved while others allow substitutions that retain only a certain pattern of charge across the motif."

Eukaryotic versions of these SLiMs are stored, for example, in the eukaryotic linear motif database (ELM) (Dinkel et al. 2016). The ELM resource was established in 2003 with the mission to collect, annotate, classify and detect short linear motifs (Puntervoll 2003). SLiM databases still only scratch the surface of existing SLiMs. The majority still waits to be discovered. Even if constantly growing, taken together the three most important linear motif databases: ELM, Linear Motif Mediated Protein-protein Interaction (Sarkar, Jana, and Saha 2015) and Minimotif Miner (Lyon et al. 2018), they contain less than 4,000 validated instances (as of September 2017, (Seo and Kim 2018)). This is

only a fraction of probably more than 100,000 SLiMs located within IDRs of the human proteome (Tompa et al. 2014).

SLiMs are also functionally diverse. They can, for example, be involved in facilitating protein localization, e.g. nuclear localization signal, (Kelly and Owen 2011; Fabbro and Henderson 2003), regulate protein degradation, e.g. polyubiquitination degrons KEN box or phosphodependent non canonical DSG degron (Davey et al. 2012; Castro et al. 2005; Fuchs, Spiegelman, and Kumar 2004) or directly regulate enzymatic activity, e.g. phosphodependent 14-3-3 motifs (Tompa et al. 2014). The same motif can be used in different contexts and in response to different signals to turn different signaling pathways on or off and cause different cellular responses (Wright and Dyson 2015). This is usually controlled at different stages: Pre-translational addition or removal of SLiM-containing exons, post-translational modification SLiM-containing peptides, allosteric SLiM inhibition or activation, and SLiM binding site competition (Dinkel et al. 2014). Peptide motifs are enriched in non-constitutive, alternatively spliced exons, tuning the regulatory potential of a protein by adding or removing peptide motifs (Buljan et al. 2012; Romero et al. 2006; Weatheritt and Gibson 2012) and also facilitating the rewiring of the interactome in different tissues. On a higher regulatory level, SLiMs often exhibit complex switching behavior for example by overlapping interfaces. In this way, they can cooperate with each other. Post-translational modifications can facilitate switching between different functional states of a protein, and thus, SLiMs function as key regulatory modules that allow for robust signaling networks (Van Roey, Dinkel, et al. 2013; Van Roey, Orchard, et al. 2013; Van Roey, Gibson, and Davey 2012; Davey et al. 2012).

SLiMs form one of the three major classes of interaction interfaces that IDRs can be split into: large serpentine disordered domains, multi-partite disordered interfaces, and compact monopartite, short linear motifs (SLiMs) (Davey et al. 2012). The majority of unstructured interfaces adapt to a template and secondary structures form after binding (Boehr, Nussinov, and Wright 2009). SLiMs differ from the other two large induced fit classes in two important attributes (Davey et al. 2012): Firstly, SLiMs are bound with lower affinity, usually between 1 and 150 micromolar (Seet and Pawson 2004; Neduva et al.

2005), for comparison, domains interact with relatively strong affinities even in the nanomolar range (Fuxreiter, Tompa, and Simon 2007). This allows SLiMs to engage in reversible and transient interactions (Wright and Jane Dyson 2009). It makes them crucial for many dynamic networks where large multi-protein complexes rapidly assemble and disassemble, but it also makes them inherently difficult to study.

Secondly, because of their short length, it takes only one or a few mutations to generate a new motif. SLiMs are evolutionary dynamic and their short and easy pattern lets them appear and disappear independently in different proteins, as evidenced by their ubiquitous presence in higher eukaryotes (Gould et al. 2010). The plasticity and adaptability this provides to the interactome, however, also has its downside. Existing and necessary SLiMs can be disrupted or novel, but malfunctioning SLiMs can be created by a single point mutation (Figure 1). A proteome-wide comparison of the distribution of missense mutations from disease and non-disease mutation datasets revealed that, in IDRs, disease mutations are more likely to occur within SLiMs than neutral missense mutations (Uyar et al. 2014). If mutations disrupt binding peptides and/or PTM sites, they may result in disease outcomes, such as Noonan syndrome, Usher syndrome, cherubism, and aberrant signaling in cancer (Guettler et al. 2011; Pajkos et al. 2012; Reimand and Bader 2013; Vacic et al. 2012).

Even viruses have discovered how important and powerful SLiMs are. They commonly take advantage of the intrinsic evolutionary plasticity of SLiMs by mimicking host motifs to hijack host pathways (Kadaveru, Vyas, and Schiller 2008; Davey, Travé, and Gibson 2011; Becerra, Bucheli, and Moreno 2017).

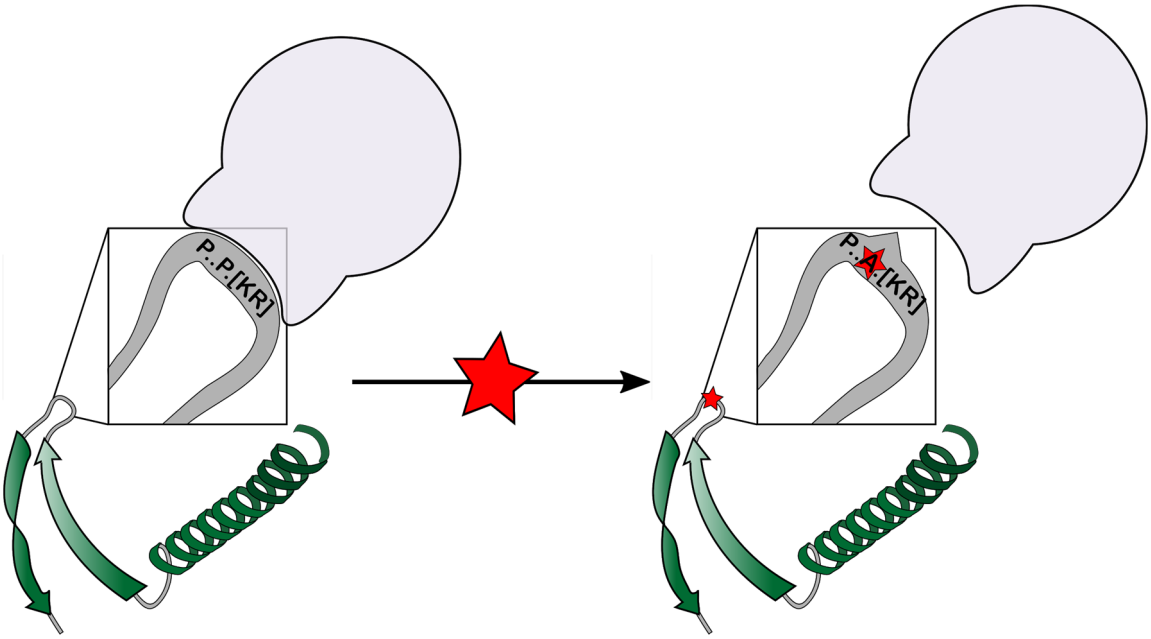


Figure 1 Mutations in disordered regions can impact protein-protein interactions.

Changes in protein-protein interactions in disease

Since the advent of molecular biology, we have learned that proteins do rarely act alone but interact with other proteins to fulfill their function. In the wealth of protein-protein interactions (PPIs), maps of proteins and their interactions can help to keep an overview. In PPI network graphs, nodes represent proteins while the lines connecting them represent their interactions. Zhong and colleagues have found that a set of disease-causing missense mutations can perturb a network by either complete loss of a gene product (node removal) or change in interaction (edgetic alterations). Node removal was likely to be caused by mutations affecting buried residues of the protein (comparable to ordered regions) and edgetic alterations were more often caused by mutations on the protein surface (comparable to disordered regions) (Zhong et al. 2009). In line with this finding, missense mutations were found to be enriched on the interaction interfaces of proteins associated with corresponding disorders (Wang et al. 2012). This loss of a distinct function compared to the loss of all functionality can also explain why different mutations in the same gene can cause different diseases. On the other hand, the analysis of protein interaction networks can also illustrate why mutations in different genes can lead to the same phenotype (Gandhi et al. 2006). The important role of PPIs mediated by IDRs leads to the conclusion that disruption of signaling- and regulatory

networks via interaction-specific defects is the most plausible mechanism for diseases that involve mutations in IDRs. Several human diseases involve abnormal protein-protein interactions (Schuster-Böckler and Bateman 2008; Marc Vidal, Cusick, and Barabási 2011; Sahni et al. 2015; Hosp et al. 2015).

Diverse studies have shown that changes in PPIs play an especially important role in neurological disorders.

Intrinsic disorder and protein-protein interactions in neurological diseases

Hosp and colleagues have quantitatively shown that changes in PPIs play a role in several neurodegenerative diseases, such as Alzheimer's disease, Huntington's disease, Parkinson's disease, and spinocerebellar ataxia type 1 (Hosp et al. 2015). A few years later, Malty and colleagues have shown that disruption of mitochondrial PPIs can lead to neurodegenerative diseases. They were able to show that several mutations at the protein interaction interface of SOD1 prevent its interaction with PRDX5, hence impairing their role in redox homeostasis, which in turn leads to the generation of reactive oxygen species (Malty et al. 2017). Also, mutations in either PARKIN or PINK disrupt the NF-κB complex which leads to its ectopic activation.

Neurodegenerative diseases are often complex multifactorial disorders characterized by the interplay of many dysregulated physiological processes (Keane et al. 2015). However, *de novo* mutations, and hence a monogenic basis of neurological disease, might play a bigger role than previously anticipated (Veltman and Brunner 2012).

In many of these monogenic cases, IDPs seem to play an important role. When individual proteins are involved in the pathogenesis of human neurodegenerative diseases, it appears that they are often either completely disordered or contained long disordered regions. These neurodegeneration-related IDPs are also characterized by high binding promiscuity, as they are able to interact with a large number of unrelated partners (Uversky 2015).

Taken together, this makes studying changes in PPIs as a potential cause of neurological diseases an especially appealing task.

Analyzing protein-protein interactions

Understanding PPIs is of crucial importance not only to help to unravel disease mechanisms but also to understand the basic functions of proteins. Several methods exist to study PPIs. X-ray crystallography and NMR provide detailed spatial information on interaction interfaces. Surface plasmon resonance (SPR), isothermal titration calorimetry (ITC) (Pierce, Raman, and Nall 1999), and Förster resonance energy transfer (FRET) (Kenworthy 2001) are some examples for methods that provide binding affinities and kinetics. All of these methods have in common that *a priori* knowledge about interaction partners is needed and they hardly allow for *de novo* interaction discovery. Higher throughput techniques that are available come with their own drawbacks: microarrays, phage display and the yeast two-hybrid system (Fields and Song 1989) rely on *in vitro* assays or heterologous biological systems. The currently most unbiased approach to study PPIs in a high-throughput manner is affinity purification followed by mass spectrometry (AP-MS). This technique allows analyzing interactions in their physiological context and in relevant organisms (Gingras et al. 2007). One important step in defining PPIs is setting apart specific from unspecific binders (background). One early idea was to get rid of background binders by stringent washes. In tandem affinity purification (TAP; (Puig et al. 2001)) the protein of interest is expressed with two affinity tags which allow two consecutive purifications to get rid of non-specific contaminants. Especially in the early times of AP-MS, this proved very useful because almost everything was yet to discover. Due to its extensive and stringent workflow, however, this method only allows for the identification of very stable complexes. Also, the sensitivity of modern mass spectrometers has become so high that they still identify even low abundant background binders. Comparable to the pigeons in Cinderella, quantitative AP-MS (q-AP-MS) came to the rescue and puts “the good ones into the pot, the bad ones into the crop”.

Mass spectrometry identifies proteins by assigning mass over charge ratios to peptides and, since these are not necessarily connected to abundance, it is not inherently a quantitative method. Several approaches have been developed to make proteomics become quantitative (Gstaiger and Aebersold 2009; Cox and Mann 2011; Bantscheff et al. 2012). One important advancement was the

metabolic incorporation of stable heavy isotopes in proteins in cell culture (SILAC). It permits different cellular populations to be mixed and analyzed together since the incorporated mass shifts make it possible to distinguish their proteomes. This allows for quantitative measurements. In the case that SILAC approaches are not feasible, there are also label-free methods to compare protein levels in different samples. Label-free quantification (LFQ) is a computational method to compare protein abundances between samples. It has to be borne in mind, though, that stable isotope-based methods can detect even minor changes while label-free methods usually require at least a twofold change (Cox et al. 2014). Stable isotope-based methods are generally more precise than label-free approaches since samples can be combined in an earlier step during sample preparation and they can be analyzed together (Sury, Chen, and Selbach 2010; Lau et al. 2014). With these approaches at hand, proteins that co-purify with a protein of interest can be compared to a negative control (Vermeulen, Hubner, and Mann 2008; Paul, Hosp, and Selbach 2011). Conveniently, the possibility to quantitatively compare samples not only allows to distinguish specific binders from the background, but it provides the unique possibility to even study dynamic changes in interaction due to perturbations of the system. This can, for instance, be used to study modification-dependent interactions, for example by studying binding to post-translationally modified immobilized peptides and their unmodified counterparts (Selbach et al. 2009; Bartke et al. 2010; Francavilla et al. 2013). Very importantly it also allows studying differences between the interactomes of wild-type and mutant proteins (Hosp et al. 2015). As stated before, the most interesting changes in interaction could concern weak and transient interactions. In standard AP-MS, the dilution step of producing cell lysate followed by washes might lead to the loss of exactly these weak or transient interactions. One way to deal with this is *in vivo* cross-linking (Kaake et al. 2014). Since the cross-linking reagents add a mass to an unknown position on the proteins and interconnect peptides, the search space is vastly increased and it might be due to these possible problems in identification that cross-linking is generally not the method of choice. A few years ago, another type of method was introduced that is able to identify protein connections *in vivo*. In these approaches, the protein of interest is equipped with a biotin ligase which enables the addition of a chemical handle (biotin) to

near-neighbors, including proximal and interacting proteins in their native cellular environment. These methods are called engineered ascorbate peroxidase (APEX) or proximity-dependent biotin identification (BioID) and, respectively, allow for different spatial or temporal resolution (Roux et al. 2012; Rhee et al. 2013). In a recent publication, AP-MS and BioID have been combined in a single construct to benefit from the advantages of both strategies and to even obtain interaction distances within a protein complex (Liu et al. 2018)

One class of PPIs that request to dig very deep in the mass spectrometry bag of tricks are PPIs mediated by SLiMs since their interaction sites are often used transiently. Indeed, high-throughput approaches such as affinity purification coupled to mass spectrometry (AP-MS) and yeast two-hybrid (Y2H) assays fail to capture these weak interactions, with only ~1% of Y2H associations relying on SLiMs (Neduva and Russell 2006). PPIs mediated by SLiMs have some technical advantages, though:

Peptide motifs are often sufficient to mediate interaction. They provide a compact and functionally autonomous module that does not interfere much with the structural and functional rest of the protein and hence can be studied even without the context of the whole protein (Tompa et al. 2014). This allows for the study of PPIs involving SLiMs by testing peptide-protein interactions. Schulze and Mann (2004) have shown that this feasible even for low-affinity interactions.

Approach in this thesis: A relatively high-throughput mass spectrometry-based screen for disease-causing changes in protein-protein interactions

"Identifying peptide motifs through sequence similarity searches is generally subject to high levels of statistical uncertainty, and they are elusive to identify experimentally. However, to gain a better and more complete description of the complex physiological and pathological processes of the cell, much more focus should be placed not only on identifying them, but also establishing their functionality through a combination of high- and low-throughput studies."
Tompa et al. 2014

In this thesis, we have decided to concentrate on missense mutations in IDRs implicated in neurological disease. By focusing on pathogenic mutations, we know that the affected IDR plays an important role in the protein's function. The pathogenicity reveals that changing the peptide sequence in this region has a broad impact. Since mutations in IDRs probably do not alter the protein structure, it is unlikely that this is based on the protein losing all its functionality.

We make use of the autonomous properties of IDRs and use synthetic peptides of 15 amino acid length surrounding the position of a disease mutation to pull down interaction partners from cell lysate. For the sake of cost-effectiveness and higher throughput, we employ peptide synthesis on a cellular membrane (Hilpert, Winkler, and Hancock 2007; Frank 1992). This approach is often used for identification of antibody binding sites with a far western set-up, but here single spots are excised and analyzed by mass spectrometry (Okada et al. 2012; Dittmar et al. 2017) to detect specific and differential interactors of IDRs and their disease variants. Such peptide pull-downs can maintain specificity even in the setting of low-affinity interactions (Schulze and Mann 2004). An increasing number of diseases have been connected to genomic mutations, thanks to extensive sequencing efforts. The disease mechanisms, however, are in most cases not known. This thesis will try to narrow the gap between genotype and phenotype by establishing a novel peptide-based interaction screen for disease-related mutations and proving its validity by following up on chosen candidates.

Materials and Methods

With some additions and some parts explained in greater detail, this section is mainly taken from Meyer et al. 2018.

Peptide-protein interaction screen

Candidate selection

Disease mutations in humans were taken from UniProt annotations (UniProt Consortium 2012) of Online Mendelian Inheritance in Man, OMIM®, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), (https://omim.org/).

This dataset consists of experimentally validated missense mutations that contribute to inherited diseases. Inherited disease mutations were downloaded from UniProt (http://www.uniprot.org/docs/humsavar.txt, release: 2015_07 of 24-Jun-2015, (Famiglietti et al. 2014)). Only mutations that were associated to 'Disease' were kept. 'Unclassified' mutations or 'Polymorphisms' were excluded. The 26,649 disease mutations were further filtered by applying a disorder cut-off. Disorder tendencies of 15 amino acids (AAs) long peptides, with the AA mutated in disease, if possible, located at position eight, were predicted using IUPred (Dosztányi et al. 2005) using the 'SHORT' (Famiglietti et al. 2014) profile considering the sequential neighborhood of 25 residues. IUPred disorder scores above 0.5 denote regions of the proteins that have 95% likelihood to be disordered. For filtering, the mean disorder score for all 15 AAs as well as the mutation position were required to be >0.5. This resulted in 1,878 disease mutations in disordered regions. Next, we assigned disease classes to 3,119 different diseases included in the Humsavar database by combining a manual approach (together with Marieluise Kirchner) with automatic annotation with the Human Phenotype Ontology database, HPO (Köhler et al. 2017). We selected 305 mutations causing neurological diseases. After manual inspection (together with Marieluise Kirchner), we remained with 128 mutations causing 124 distinct neurological diseases that were used for the peptide-protein interaction screen (Table 4).

Experimental setup

Peptides of 15 AAs, in total 128 wild-type peptide and 128 related peptides containing the disease-causing mutation (256 peptides) plus one control peptide pair were synthesized in situ on cellulose membrane using PepTrack™ techniques (JPT Peptide Technologies, Berlin, Germany) (Frank 2002). Peptides were bound to cellulose membrane with β-alanines performing as a linker. Control peptide: VPPPVPPRRR amino acids 1150-1158 of SOS1 contains classical polyproline type II motif PXPhiXR for SH3 domain recognition and binds GRB2 with an affinity of about 4 μM (Schulze and Mann 2004). To obtain control baits prolines were replaced with alanines to prevent proline-directed folding.

Two peptide filters were moistened in cell lysis buffer [50 mM HEPES pH 7.6 at 4°C, 150 mM NaCl, 1 mM EGTA, 1 mM MgCl2, 10% Glycerol, 0.5% Nonidet P-40, 0.05% SDS and 0.25% sodium deoxycholate, supplemented with protease inhibitor (Roche) and benzonase (Merck)]. In order to reduce nonspecific binding, the membrane was incubated with 1 mg/ml yeast t-RNA (Invitrogen) for 10 min and then washed twice with wash buffer. The entire peptide libraries were incubated with 15 ml of light or heavy SILAC labeled cell lysate (5 mg/ml) from SH-SY5Y cells for 2 h (details about cell culture and production of cell lysate see further down). Membranes were washed three times with wash buffer and air dried.

Cell culture

SH-SY5Y (for peptide array) and HEK T-REx™-293 (for follow-up studies) cells were cultured under standard cell culture conditions. In brief, cells were cultured in DMEM (Life Technologies) complemented with 10% fetal calf serum (Pan-Biotech) at 37°C and 5%CO2.

Cells used for SILAC based experiments were cultured in SILAC DMEM (Life Technologies) complemented with glutamine (Glutamax, Life Technologies), Pyruvate (Life Technologies), non-essential amino acids (Life Technologies) and 10% dialyzed fetal calf serum (Pan-Biotech). The SILAC DMEM was supplemented with standard L-arginine (Arg0, Sigma-Aldrich) and L-lysine (Lys0, Sigma-Aldrich) (“light”) as in (Schwanhäusser et al. 2011). Alternatively,

Arg6 and Lys4 (“medium-heavy”) or Arg10 and Lys8 (“heavy”) were added in place of their light counterparts.

SILAC amino acids were prepared in PBS to stock solutions at the concentrations listed in Table 1.

Table 1 SILAC amino acids

Amino acid	Mol weight (average)	Stock conc [g/L]
Arg0	210.6619	84
Arg6	216.6178	86
Arg10	220.5915	88
Lys0	182.6485	146
Lys4	223.1341	178
Lys8	190.5912	152

Amino acids were diluted 1:2000 in medium. All components were added to a vacuum filter system and sterile filtered.

Cell lysate for peptide array

Cell pellet from SH-SY5Y cells grown in 50x 15 cm dishes for each SILAC label were lysed in lysis buffer [50 mM HEPES pH 7.6 at 4°C, 150 mM NaCl, 1 mM EGTA, 1 mM $MgCl_2$, 20% Glycerol, 1% Nonidet P-40, 0.1% SDS and 0.5% sodium deoxycholate, supplemented with protease inhibitor (Roche) and benzonase (Merck)] after 15 total passages and 7 passages in SILAC medium. Volume of lysis buffer and cell pellet was almost 1:1. Cells were lysed 30 min on ice and then passed 5x through 20G needle and 1x through 26G needle. Lysate was centrifuged for 25 min at 4,600 x g at 4°C to get rid of cell debris. This resulted in ~7 ml lysate per SILAC label. Cell lysate was diluted 1:1 with wash buffer to yield a final concentration of 5mg/ml and to dilute detergent concentration that could inhibit peptide-protein interactions.

Sample preparation for mass spectrometric analysis

We got advice on the method from Daniel Perez Hernandez (MDC). Most importantly he introduced us to membrane blocking with yeast t-RNA.

Single spots were punched out from cellulose membrane with a 2 mm diameter ear punch (Carl Roth) and SILAC pairs were placed together in a 96-well plate (Thermo Scientific) prepared with 30 µl of denaturation buffer [6 M urea (Sigma-Aldrich), 2 M thiourea (Sigma-Aldrich), 10 mM HEPES, pH 8]. Samples were reduced by incubating with 10 µl of 3.3 mM DTT (Sigma-Aldrich) in 50 mM ammonium bicarbonate (ABC) buffer for 30 min at RT, followed by an alkylation step using 10 µl of 18.3 mM iodoacetamide (IAA) (Sigma-Aldrich) in 50 mM ABC for 60 min at RT. The samples were first digested using 1 µg endopeptidase LysC (Wako, Osaka, Japan) for 4 h. The samples were diluted by adding 100 µl of 50 mM ABC (pH = 8.5), and finally digested with 1 µg trypsin (Promega) for 16 h. The digestion was stopped by acidifying each sample to pH < 2.5 by adding 10% trifluoroacetic acid solution. The peptide extracts were purified and stored on stage tips according to (Rappsilber, Ishihama, and Mann 2003). Using a stage tip adaptor for high throughput stage tipping in a centrifuge (Kind gift from Nils Kulak, Max Planck Institute of Biochemistry).

LC-MS/MS analysis

Peptides were eluted using Buffer B (80% Acetonitrile and 0.1% formic acid) and organic solvent was evaporated using a speedvac (Eppendorf). Samples were diluted in Buffer A (5% acetonitrile and 0.1% formic acid). Peptides were separated on a reversed-phase column with 45 min gradient with a 250 nl/min flow rate of increasing Buffer B concentration on a High Performance Liquid Chromatography (HPLC) system (ThermoScientific). Peptides were ionized using an electrospray ionization (ESI) source (ThermoScientific) and analyzed on a Q-exactive plus Orbitrap instrument (ThermoScientific). Dynamic exclusion for selected precursor ions was 30 s. The mass spectrometer was run in data dependent mode selecting the top 10 most intense ions in the MS full scans, selecting ions from 300 to 1700 m/z (Orbitrap resolution: 70,000; target value: 1,000,000 ions; maximum injection time of 120 ms). The resulting MS/MS spectra from the Orbitrap had a

resolution of 17,500 after a maximum ion collection time of 60 ms with a target of reaching 100,000 ions.

Data analysis

The resulting raw files were analyzed using MaxQuant software version 1.5.2.8 (Cox and Mann 2008). Default settings were kept except that 'match between runs' and 're-quantify' was turned on. Lys8 and Arg10 were set as labels and oxidation of methionines and N-terminal acetylation were defined as variable modifications. Carbamidomethylation of cysteines was set as fixed modification. The *in silico* digests of the human Uniprot database (2015-12), a fasta file containing all peptides used for pull-down and a database containing common contaminants were done with Trypsin/P. The false discovery rate was set to 1% at both the peptide and protein level and was assessed by in parallel searching a database containing the reversed sequences from the Uniprot database. Following statistics and figures were done using R (R version 3.2.1, RStudio Version 1.0.143). The resulting text files from MaxQuant analysis were filtered to exclude reverse database hits, potential contaminants, and proteins only identified by site. We imputed missing LFQ-intensity values with random noise simulating the detection limit of the mass spectrometer (Keilhauer, Hein, and Mann 2015). To this end, imputed values were taken from a log normal distribution with 0.25× the standard deviation of the measured, logarithmized values, down-shifted by 1.8 standard deviations. In this way, we obtained a distribution of quantitative values for each protein across samples. For determination of specific interactions, two replicated pull-downs for the same peptide were tested against all other pull-downs, excluding the corresponding variant peptide, by the nonparametric Mann–Whitney U test. Resulting p-values (-log10) and fold-changes (log2 space) have been plotted as volcano plots to determine cut-offs. We used an approach that uses a graphical formula to combine fold-change and p-value cut-off (Keilhauer, Hein, and Mann 2015): $-log10(p) \geq \frac{c}{|x|-x_0}$ with x: enrichment factor of a protein, p: p-value of the Mann–Whitney U test calculated from replicates, x_0: fixed minimum enrichment, c: curvature parameter. The curvature parameter c determines the maximum acceptable p-value for a given enrichment x.

The parameters c and x_0 can be optimized based on prior knowledge of known true and false positives (Keilhauer, Hein, and Mann 2015). Here, cut-offs were chosen according to known interaction partners of the SOS1 control peptide (Schulze and Mann 2004). This resulted in x_0=0, c=8.

This cut-off was applied to all other pull-downs to separate specific binders from background. SILAC ratios were normalized by subtracting the median SILAC ratio of every experiment from all SILAC ratios in that experiment. To define interaction partners that bind differentially to wild-type and mutant peptide, a SILAC cut-off was defined. For wild-type specific interaction partners, the mean log2 SILAC ratio of the two replicates needed to be >1 and none of the two ratios <0 (mutant specific mean log2 SILAC ratio < -1 and none of the two ratios >0). Resulting figures were modified in Inkscape (0.91).

PRM on adaptor proteins that bind dileucine peptides

Experimental procedure was identical to general peptide-protein interaction screen. Only peptide variants from GLUT1_P485L, ITPR1_P1059L, CACNA1H_P648L and CACNA1H_A748V (control peptide) were used for experiment.

LC-MS/MS analysis

Peptides were separated by reverse phase chromatography on an effective 150 min gradient (0, 2, 100, 30, 15, 1 and 5 min with 2, 4, 20, 30, 60, 90 and 90% of buffer B with 90% acetonitrile) and analyzed on a Q-Exactive HFx (Thermo Fisher Scientific). The PRM settings were: 30,000 resolution; 5e5 AGC target; 1.6 m/z isolation window; 60 ms max ion injection time. The inclusion list for the PRM method was generated using Picky (Zauber, Kirchner, and Selbach 2018) with SILAC option enabled and a retention time window of 30 min. Predicted retention-times were calibrated in Picky with a complex sample of 100 ng Pierce HeLa Protein standard (Thermo Fisher Scientific) immediately before the PRM measurements.

Analysis of PRM data

By Henrik Zauber: Traces of all fragments from precursors in the spectral library (as exported from picky) were extracted from all rawfiles using the Thermo MSFileReader and the MSFileReader.py bindings written by François Allen. For each light or heavy scan the normalized spectral contrast angle (SCN) was calculated (Toprak et al. 2014). Peaks were manually selected and required a SCN > 0.4 and Fragment Matches > 4 in the light or heavy channel. Further peaks needed to be within a similar retention time range across all different measurements. Ratios for each fragment using the maximum intensity of each peak were calculated. The median log2 transformed ratio (log2FC) for each peptide in each raw-file was calculated from selected fragment ratios: The five highest abundant fragments were selected from the peak with the highest detected SCN. Peptide log2FC were plotted as boxplot distributions in a protein centric manner across the different experiments (Figure 9) (Analysis by Henrik Zauber).

Far western validation of peptide-protein interactions

Peptide Spot Array was stained with Ponceau solution (2% Ponceau in 30% TCA) for 2 min. The membrane was subsequently activated with MeOH for 5 min. The following steps were conducted at 4°C. Membrane was washed 3x 3 min with wash buffer [50 mM HEPES pH 7.6 at 4°C, 150 mM NaCl, 1 mM EGTA, 1 mM $MgCl_2$] (same as for peptide array experiment followed by mass spectrometry). Membrane was blocked with 1 mg/ml yeast t-RNA (Invitrogen) for 10 min and then washed twice with wash buffer. The entire peptide library was incubated with 14 ml of light SILAC labeled cell lysate (5 mg/ml) from SH-SY5Y cells for 2 h (recovered from peptide experiment followed by mass spectrometry). Membranes were washed three times shortly with wash buffer. Subsequent steps were again conducted at room temperature. The membrane was washed and blocked in 3% BSA in wash buffer for 10 min, followed by 1 h incubation in primary antibody (anti-CHC, 1:1500 (4 µl in 6 ml), rabbit, abcam, ab21679) in a sealed plastic bag and on a rotor. Following washing steps were kept short to minimize the time for peptide-protein interactions to resolve. Membrane was rinsed once and washed twice for 5 min with wash buffer. This was followed by incubation for 1 h with secondary antibody (anti-rabbit HRP, GE Healthcare) 1:10.000 (1 µl in 10 ml) again in a sealed plastic

bag on rotor. The membrane was washed like before. Proteins were detected with chemiluminescence substrate (Perkin Elmer) on a ChemiDoc MP Imaging System (Bio-Rad) and Image Lab 5.2.1 Software (Experiment with Teresa Melder).

Follow-up on GLUT1

Generation of stable cell line

We purchased SLC2A1 (GLUT1) from Harvard Plasmid repository (HsCD00378964). Q5® Site-Directed Mutagenesis Kit (NEB) was used for adding a stop codon and subsequent P485L mutation. The stop codon has been added to the gene with the following primers Fw:TCCCAAGTGTAATTGCCAACTTTCTTGTACAAAGTTG, Rev:ATCAGCCCCCAGGGGATG.

PCR CYCLING CONDITIONS:

STEP	TEMP	TIME
Initial Denaturation	98°C	30 s
25 Cycles	98°C	10 s
	67°C	20 s
	72°C	1 min 45 s
Final Extension	72°C	2 min
Hold	8°C	

The P485L Mutation was introduced by changing c.1454 C>T (Slaughter, Vartzelis, and Arthur 2009) with Fw:CTGTTCCATCtCCTGGGGGCT, Rev:CTCCTCGGGTGTCTTGTCAC.

PCR CYCLING CONDITIONS:

STEP	TEMP	TIME
Initial Denaturation	98°C	30 s
25 Cycles	98°C	10 s
	68°C	20 s
	72°C	1 min 45 s
Final Extension	72°C	2 min
Hold	8°C	

SLC2A1 and *SLC2A1* mutant have been further cloned into a destination vector with an N-terminal BirA-FLAG Tag (pDEST-pcDNA5-BirA-FLAG N-term (Couzens et al. 2013)) with Gateway cloning strategy (Thermo Fisher Scientific). HEK 293 Flp-In T-Rex cells (Invitrogen) that exhibit tetracycline-inducible expression of BirA-FLAG-GLUT1 or BirA-FLAG-GLUT1_P485L were generated using the Flp-In system developed by Life Technologies according to the manufacturer's protocol (Kindly generated for us by the Markus Landthaler lab, MDC Berlin). The Flp-In approach allows streamlined generation of stable mammalian cell lines by taking advantage of a Saccharomyces cerevisiae-derived DNA recombination system. This system uses a recombinase (Flp) and site-specific recombination (Craig 1988; Sauer 1994) to facilitate integration of the gene of interest into a specific site (FRT site) in the genome of mammalian cells.

BioID

BioID employs a promiscuous biotin ligase to biotinylate proteins based on proximity labeling (Roux et al. 2018). It is a powerful method to identify interacting and proximal proteins even for transient interactions.

Medium-heavy and heavy labelled HEK T-REx™-293 cells were induced for 24 h with 0.1 µg/ml doxycycline to induce expression of GLUT1 (wild-type, wt) or GLUT1_P485L (mutant, mut). Light labelled cell lines from both GLUT1 and GLUT1_P485L were left uninduced and served as a control for background binding. SILAC labeling allowed for quantitative comparison of proteins that were proximity labelled by the expressed GLUT1 constructs (Forward experiment: Light - Control, Medium-heavy - wt, Heavy - mut; Label swap experiment: Light - Control, Medium-heavy - mut, Heavy - wt). During the induction period all cell lines were incubated for 24 h in cell culture medium containing 50 µM biotin. BioID experiment was performed essentially as in (Couzens et al. 2013), with minor adaptations.

Mass spectrometry setup and analysis was done similarly as to samples from peptide pull-downs, but on bead digested peptides were separated on a 2,000 mm monolithic column with a 100 µm inner diameter filled with C18 material that was kindly provided by Yasushi Ishihama (Kyoto University) using a 4 h linear gradient with a 300 nl/min flow rate of increasing Buffer B

concentration on a High Performance Liquid Chromatography (HPLC) system (Thermo Scientific). The resulting raw files were analyzed using MaxQuant software version 1.5.2.8 (Cox and Mann 2008). Default settings were kept except that 'match between runs' and 're-quantify' was turned on. Lys4 and Arg6 or Lys8 and Arg10 were set as labels and oxidation of methionines and N-terminal acetylation were defined as variable modifications. Carbamidomethylation of cysteines was set as fixed modification. The *in silico* digests of the human Uniprot database (2015-12), a fasta file containing the sequence of BirA-FLAG-GLUT1 and a database containing common contaminants were done with Trypsin/P. The false discovery rate was set to 1% at both the peptide and protein level and was assessed by in parallel searching a database containing the reversed sequences from the Uniprot database. Biotinylated proteins with a wild-type to mutant enrichment ratio (log2FC) >1 or <-1 in both replicates were considered as significant. These proteins were analyzed for gene ontology enrichment of cellular components with http://metascape.org (Tripathi et al. 2015).

FLAG-GLUT1 localization

HEK 293 Flp-In T-Rex cells with BirA-FLAG-GLUT1 or BirA-FLAG-GLUT1_P485L were seeded on coverslips coated with poly-L-lysine (Sigma-Aldrich). After induction for 24 h in doxycycline (0.1 µg/ml) containing media, cells were fixed with 4% PFA (paraformaldehyde). Standard procedures were used for immunostaining. Cells were stained against FLAG 1:200 (F1804, Sigma). Nucleus was stained with DAPI (Sigma). Costainings were performed by Jing-Yuan Cheng: FLAG staining was accompanied by staining to one of the following endosomal markers and with the following dilutions: anti-EEA1 (Cell Signaling Technology, 1:100); anti-Rab4 (Cell Signaling Technology, 1:100); anti-Rab9 (Cell Signaling Technology, 1:100); anti-LAMP1 (Cell Signaling Technology, 1:100). Mouse anti-FLAG staining was substituted by rabbit anti-GLUT1 (Merck Millipore, 1:500) to costain mouse monoclonal antibodies: anti-VTI1A (BD Biosciences, 1:100); anti-VTI1B (BD Biosciences, 1:100). Secondary antibodies all come from Invitrogen. For colocalization analysis three z stacks of 5-10 cells each were quantified for each marker with Imaris v8.4.1. For details see "Fluorescence microscopy from cell culture", "Colocalization analysis".

FLAG-GLUT1 motif mutants

GLUT1_P485A mutant was generated with help of http://nebasechanger.neb.com/ from pENTRY_GLUT1 vector. Entry vectors containing GLUT1_wt, _P485L or _P485A were further cloned into a destination vector harboring an N-terminal FLAG-tag, via gateway cloning technology. HEK 293 were seeded on coverslips coated with poly-L-lysine (Sigma-Aldrich). After 12 h, they were transfected with either of the three plasmids with Jetprime (Polyplus-Transfection). After 24h, cells were fixed with 4% PFA. Standard procedures were used for immunostaining. Cells were stained against FLAG 1:200 (F1804, Sigma-Aldrich). Nucleus was stained with DAPI (Sigma-Aldrich). Samples were imaged on an upright epifluorescence microscope (Leica DM5000 B) and Leica DFC365 FX camera.

Transferrin uptake

Essentially as in "FLAG-GLUT1 localization". Additionally, after 24 h cells were serum-starved for 1 h and used for Transferrin (Tf) uptake. For Tf uptake, cells were incubated with 10 µg ml^{-1} Tf-Alexa568 (Life Technologies) for 10 min at 37°C. Three z stacks of more than 15 cells each have been quantified with Imaris v8.4.1. For details see "Fluorescence microscopy from cell culture", "Colocalization analysis".

FLAG-GLUT1 localization under AP-2 µ knock-down

To rescue the GLUT1_P485L phenotype, clathrin-mediated endocytosis (CME) was inhibited by knocking down AP-2 µ and hence the adaptor complex responsible for recognition of cargo for CME. AP-2 is a stable protein complex and hence two subsequent rounds of knock-down are needed (Motley et al. 2003).

On day 1 cells were seeded in 6-well plates. On day 2, cells were transfected with 25 nM final siRNA concentration (AP-2 µ: ON-TARGETplus Human AP2M1 (Dharmacon) and non-target: ON-TARGETplus Non-targeting Pool (Dharmacon)) according to DharmaFECT (Dharmacon) transfection protocol. 24 h after transfection, medium was replaced with complete medium to reduce cytotoxicity and incubated for another 24 h. On day 4, cells were transfected

again as on day 2. On day 5, cells were seeded in a 24-well plate onto coverslips coated with poly-L-lysine (Sigma-Aldrich) for microscopy and into a 6-well plate for western blot analysis. Doxycycline (0.1 µg/ml) was added to the medium to induce expression of the GLUT1 constructs. After induction for 48 h, cells in 24-well plates were fixed with 4% PFA. Standard procedures were used for immunostaining. Cells were stained with rabbit polyclonal GLUT-1 antibody 1:200 (Merck Millipore) and co-stained with mouse monoclonal anti-alpha adaptin antibody [AP6] 1:200 (Abcam). Secondary antibodies with Alexa fluorophores were all purchased from Invitrogen. Nucleus was stained with DAPI (Sigma).

Lysate from cells in 6-well plates was used for western blotting, α and µ2 subunits of AP-2 were detected using mouse monoclonal antibodies from Thermo Fisher Scientific and BD transduction, respectively. Profilin 1 was stained as a loading control with polyclonal rabbit antibody from CST. Horseradish peroxidase coupled secondary antibodies were purchased from GE Healthcare. Proteins were detected with chemiluminescence substrate (Perkin Elmer) on a ChemiDoc MP Imaging System (Bio-Rad) and quantified with Image Lab 5.2.1.

Fluorescence microscopy from cell culture

Images from FLAG-GLUT1 localization were acquired by Leica DMI6600 confocal laser scanning microscope with an HCX PL APO 63.0/1.40 oil objective. Transferrin uptake, GLUT1-localisation under AP-2 µ knock down, antibody feeding assay and GLUT1 in iPSCs were acquired by a Zeiss LSM 700 confocal laser scanning microscope with an EC Plan-Neofluar/NA1.3 40x oil objective or a EC Plan-Apochromat/NA1.4 63x oil objective. Images were further processed with Fiji (Schindelin et al. 2012).

Colocalization analysis

Imaris v8.4.1 was used for quantitative colocalization analysis. The original z-stack images were adjusted by adding an adequate mask on the respective red channel to subtract background noise (Costes et al. 2004). The threshold for the mask was uniformly adjusted in each staining experiment. Automatic thresholding was used to define the area where a colocalization would be determined and the statistics was calculated for each colocalization channel

(Costes et al. 2004). For the images whose observed correlation was not statistically significant in comparison to randomized images, the colocalization channel was built without additional thresholding on the masked dataset. The resulting thresholded Pearson's coefficients were exported. The number of images and cells in the analyses is stated in the respective Method sections.

Radioactive glucose uptake under AP-2 µ knock-down

AP-2 knock down was performed essentially as described before (FLAG-GLUT1 localization under AP-2 µ knock-down). In detail, here, initially, 500,000 cells were seeded in a 6-well plate and subsequently 200,000 cells were seeded in triplicates in a 24-well plate without coverslips. 1 µg/ml doxycycline was used to induce GLUT1 expression. Radioactive glucose uptake was performed mainly as in (Shi and Kandror 2008). Radioactive glucose cocktail was prepared by adding 10 µL of 3H-2-deoxy-D-glucose in ethanol:water solution (specific activity, 5–10 Ci (185–370 GBq)/mmol) (Perkin Elmer) to a 2.0-mL tube and left open for 5 min to evaporate ethanol. 1.6 mL of KRH(−) glucose buffer (121 mM NaCl, 4.9 mM KCl, 1.2 mM $MgSO_4$, 0.33 mM $CaCl_2$, 12 mM HEPES, pH 7.4) and 16 µL of cold 2-DOG (100X) stock solution (100 mM 2-deoxy-D-glucose in KRH (−) glucose) (Sigma-Aldrich) were added to the tube. Cells in each well were rinsed with DMEM (without serum, SFM) warmed to 37°C and SFM was added to cells slowly and carefully by the side of the well in order to avoid detachment of cells. Cells were incubated with 0.5 mL of SFM (in case of +dox containing 1µg/µl doxycycline) per well for 2 h at 37°C. Cells in each well were washed twice with 2 mL of KRH (−) glucose buffer at 37°C. 225µl of KRH (-) containing 25 µM final cytochalasin B (dissolved in DMSO) or 0.5% DMSO were added to each well. Immediately after, 25 µL of radioactive glucose cocktail was added to all wells. Samples were incubated at 37°C for 1 h and then transferred on ice. Radioactive glucose cocktail was aspirated, and ice-cold KRH (+) glucose (121 mM NaCl, 4.9 mM KCl, 1.2 mM $MgSO_4$ 0.33 mM $CaCl_2$: 12 mM HEPES, 25 mM D- (+)- Glucose, pH 7.4.) was added to terminate the reaction. Cells were washed once more with ice-cold KRH (+) glucose. Plate was transferred to room temperature, and 400 µL of 0.1% SDS in KRH (−) glucose were added to each well, incubated at room temperature for 10 min, and thoroughly resuspend to homogeneity. 100 µl of the lysate was kept to

measure protein concentration with DC protein assay kit I (BioRad). 300 µL of lysates were transferred in scintillation vials containing 4 mL of Rotiszint eco plus scintillation fluid (Carl Roth) and count in a Liquid Scintillation Analyzer (Tri-Carb 2800TR, PerkinElmer) for 1 min per vial. These numbers represent "Counts in the samples". In parallel, 10 µL of the radioactive glucose cocktail were mixed with 290 µL of 0.1% SDS in KRH (−) glucose and this mixture was measured under the same conditions. This number represents "Counts in the cocktail". The amount of intracellular 2-deoxyglucose was calculated using the following formula:

$$\left[\frac{[Counts\ in\ the\ sample]\times 1000}{[Counts\ in\ the\ cocktail]\times 0.03\times[C]\times t}\right]\ \text{pmol/mg}\times\text{min}$$

where [C] is protein concentration in mg/ml and t is the total time of incubation with radioactive glucose in min. All resulting values were divided by the overall mean value from all wild-type GLUT1 (+ doxycycline, - cytochalasin B) and multiplied by 100 to receive relative values for glucose uptake (%). For test of statistical significance, the mean values of three technical replicates were calculated from three biological replicates and determined by one-tailed t-test. Depicted values are mean values over all replicates and error bars show standard error of mean (SEM) over all replicates.

GST pulldown assay

GLUT1 cytoplasmic C-terminal tail was amplified from pDEST_pcDNA5_FLAG_BirA GLUT1 or GLUT1_P485L with Fw:tatatcGAATTCGTTCCTGAGACTAAAGGC, Rev:aacaatGCGGCCGCTTACACTTGGGAATCAGCC. This resulted in C-terminal tail amino acids 451-492 (UniProt P11166). Added EcoRI and NotI restriction sites were used to insert the PCR product into pGEX6P1 to generate GST-fusion chimera. Other cytoplasmic regions were ordered as gBlocks Gene Fragments (IDT) from the region ± 20AA of the mutation position, with an additional 5'-EcoRV restriction site and 3'-stop codon-NotI restriction site. After restriction, the gene fragments were inserted into pGEX6P2.

By Giulia Russo: Expression of GST-tagged proteins was induced for 5 h at 22°C by addition of isopropyl thio-β-d-galactoside (0.5 mM) to E.Coli BL21 in 2X YT medium (0.8 OD). To lyse the cells, bacterial pellets were resuspended in PBS

and left on ice for 15 min in presence of PMSF (1 mM), cyanase (4 U/µl) and lysozyme (1 mg/ml). Then, Triton X-100 was added to 0.5% and cells were sonicated for 2 min. Lysates were centrifuged for 15 min at 50,000 x g. 300 µl of glutathione-coupled beads were added to the supernatant and rotated end-over-end for 2 h at 4°C. Beads were washed three times with PBS / 0.1% Triton X-100 and once with PBS.

Pulldown experiments were performed using mouse brain extracts. Mouse brains were homogenized in 2.5 ml buffer (20 mM Hepes, 320 mM sucrose, pH 7.5) containing protease inhibitor cocktail (Sigma-Aldrich). The homogenate was centrifuged at 1000 x g for 10 min and the supernatant was supplemented with 1% Triton X-100, 50 mM KCl and 2 mM $MgCl_2$, and kept on ice for 10 min with occasional vortexing. The lysate was cleared by centrifugation at 17,000 x g for 15 min and at 178,000 x g for 15 min. The supernatant was recovered and used at a concentration of 7.5 mg protein/ml.

The pulldown experiment was performed using 85 µg of GST fusion proteins and 0.6 ml protein extract by end-over-end rotation for 3 h. The samples were washed four times with buffer containing 20 mM Hepes, 50 mM KCl, 2 mM $MgCl_2$, Triton X-100 (1%) and once in the same buffer without detergent. Proteins were eluted from the beads twice with Laemmli buffer and analysed by Western blotting. The following antibodies and dilutions were used: mouse anti-talin 1:1000 (Sigma-Aldrich), mouse anti-γ1 adaptin of AP-1 1:500 (BD Transduction), mouse anti α-adaptin of AP-2 1:200 (BD Transduction), horseradish peroxidase-conjugated goat anti-mouse 1:2000 (Jackson labs).

GLUT1 in patient-derived iPSCs

Generation of patient-derived iPSCs

By Juan M. Pascual: Fibroblasts were obtained from a GLUT1 deficient patient with the P485L mutation. The voluntary informed consent process was documented in writing as approved in advance by the University of Texas Southwestern Medical Center Institutional Review Board. This included information regarding the de-identification of the sample and the adherence to HIPAA regulations.

A 4 mm single-use, sterile skin punch was applied to the lateral surface of the left shoulder after the skin had been cleansed with iodine solution in aseptic fashion followed by injection of 0.5 ml of 1% unbuffered lidocaine with a vasoconstrictor. Prilocaine and lidocaine cream had been previously applied to the area. The punch was advanced by rotation under pressure and the explant was severed from its base and harvested in culture medium containing complete DMEM plus 20% FBS and placed on ice until the explant was divided for culture the same day. The explant was divided into 12-15 evenly sized pieces and each piece maintained in a 10 cm dish at 37°C until fibroblast confluence was reached. The cells were then treated with trypsin and passaged into a T-25 flask for further expansion. Fibroblasts were grown to approximately 50% confluence in the T-25 flask. They were then suspended with trypsin and frozen over dry ice in complete DMEM medium with 10% DMSO at a density of 10^6 cells/ml per vial prior to storage and shipment on dry ice.

By Sebastian Diecke: The patient fibroblast were reprogrammed using the mRNA reprogramming kit ReproRNA™-OKSGM from Stem Cell Technologies according to the instructions (by Sebastian Diecke lab, MDC/BIH). In brief, $1x10^5$ fibroblast cells were plated on Geltrex coated 6-well plate using regular DMEM media with 10% FBS. The day after the cells were transfected with the ReproRNA- OKSGM construct using the ReproRNA transfection reagents and growth Media with B18R. The next 5 days the growth media was changed every day and supplemented with B18R and 0.8 µg/ml puromycin. After 8 days the growth media was exchanged by ReproTeSR and first colonies appeared after 14 days. In total 5 clones were picked and established using mTESR-1 media.

As a control for the experiments the following fibroblasts (NHDF-Ad-Der Fibroblasts, C-2511, LONZA) were reprogrammed using the Epi5™ Episomal iPSC Reprogramming Kit from Thermo Fisher Scientific following the vendor's instructions. The established lines and clones were registered and named using the Human pluripotent stem cell registry (https://hpscreg.eu/): BIHi037-(A-E).

The iPSCs used for the experiments were characterized using the PSC 4-Marker immunocytochemistry kit from Thermo Fisher SCIENTIFIC following the instructions of the protocol. In addition to the 4 markers (OCT4, SOX2, TRA-1-60 and SSEA4) included in the kit, the expression of another pluripotency marker NANOG (Nanog PA1-097, Thermo Fisher Scientific) was analyzed.

iPSC culturing

By Ina-Maria Rudolph: Human iPSC cultures were maintained on plates coated with hESC-Qualified Matrigel™ (Corning) in mTESR-1 medium (Stem Cell Technologies) following the manufacturer's instruction. All cells were cultured at 37°C in humidified atmosphere containing 5% O2 and 5% CO2. Cells were passaged using StemPro Accutase (Thermo Fisher) and replated in mTESR1 medium with the addition of 10 µM ROCK inhibitor Y-27632 (LC Laboratories) (culturing and preparation of cells for experiments by Ina-Maria Rudolph).

GLUT1 localization in iPSCs

Human iPSCs were seeded on coverslips coated with hESC-Qualified Matrigel™ (Corning). Cells were fixed with 4% PFA, stained with rabbit polyclonal GLUT1 antibody 1:200 (Merck Millipore) and costained with mouse monoclonal VTI1A antibody 1:100 (BD Transduction). Secondary antibodies with Alexa fluorophores have all been purchased from Invitrogen.

GLUT1_P485L in a mouse model

By Ralf Kühn: GLUT1 P485L mice were produced by microinjection of C57BL/6N zygotes with Cas9 protein (IDT), synthetic guide RNA (IDT) (5'GAGGAGCTCTTCCACCCTCT3') and a mutagenic single stranded deoxyoligonucleotide (IDT) (5'TAGCTGCCTGTGCTCCAGAGAGATCCTTGGGCTGCAGGGAGCAGGCCGGGCTGG GTGTGGGGCTCCTCACACTTGGGAGTCCGCCCCCAacaaGTGGAAGAGCTCCTCGGG TGTCTTGTCACTTTGG3') as recombination template, as described (Wefers et al. 2017). Reagents were diluted in microinjection buffer (10 mM Tris, 0.1 mM EDTA, pH 7.2), filtered through a centrifugal filter (Millipore, UFC30LG25) and stored in single use aliquots at -80°C. For microinjections, zygotes were obtained by mating of C57BL/6N males with superovulated C57BL/6N females (Charles River, Sulzbach, Germany). Zygotes were injected into one pronucleus following standard procedures (Ittner and Götz 2007). Injected zygotes were transferred into pseudo-pregnant NMRI female mice to obtain live pups. All mice showed normal development and appeared healthy. Mice were handled according to institutional guidelines under experimentation license no. G0162/12 approved by the Landesamt für Gesundheit und Soziales (Berlin, Germany) and housed in standard cages in a specific pathogen-free facility on a 12 h light/dark cycle with *ad libitum* access to food and water.

Immunofluorescence in mouse tissue

Mouse keeping and experiments by Carmen Birchmeier lab (MDC) and especially Luis R. Hernandez-Miranda with help from Sven Buchert. E14-E15.5 embryos were obtained by Caesarian section from pregnant dam on day 14-15.5 post-coitus. Whole-mount embryos were dissected in ice-cold phosphate buffer and fixed for 2 h with a solution of 4% PFA in ice-cold phosphate buffer and cryoprotected overnight in 30% sucrose in phosphate buffer at 4°C. Whole embryo heads were sectioned in a horizontal plane using a cryostat to obtain 12-16 µm sections.

Sample preparation for confocal microscopy

Essentially as in Hernández-Miranda et al. 2011. Brain sections were incubated in blocking buffer 1 (5% horse serum and 0.1% Triton™-X 100 made in PBS) for 1 h at room temperature. Then, sections were incubated overnight in blocking buffer 1 containing the following antibodies: rabbit anti-Glut1 (1:200; Merck Millipore #07-1401), rat anti-ICAM2 (1:100; BD Biosciences #553326) and Isolectin GS-IB4 Alexa Fluor 488 conjugate (1:100, Thermo Fisher Scientific #I21411) at 4°C. Next, sections were washed three times in ice-cold PBS and incubated for 3 h in blocking buffer 1 containing Cy3 horse anti-rabbit (1:500; Jackson Lab), Cy5 horse anti-rat (1:500; Jackson Lab) and DAPI at room temperature. Fluorescence was imaged on a Zeiss LSM 700 (Jena, Germany) confocal microscope in a non-blind manner.

Sample preparation for STED microscopy

Sample staining and STED microscopy by Anna Szymborska (MDC). Sections were washed twice for 5 minutes with PBS to remove the embedding resin and incubated in 0.2% Triton X-100 in blocking buffer 2 (1% bovine serum albumin, 1% fetal calf serum in PBS) for 1 h at room temperature. Samples were incubated with primary antibodies anti-Glut1 (rabbit anti-human, Merck Millipore #07-1401) and anti-ICAM2 (rat anti-mouse CD102, BD Biosciences #553326) at 1:100 dilution in blocking buffer 2 overnight at 4°C. Next, samples were washed three times for 5 minutes in PBS and incubated with STAR Red goat anti-rabbit antibody (Abberior, #2-0012-011-9), Alexa Fluor 594 donkey anti-rat antibody (Thermo Fisher Scientific #A21209), and Isolectin GS-IB4 Alexa Fluor 488 conjugate (Thermo Fisher Scientific #I21411), all diluted at 1:500 in blocking buffer 2. Subsequently, samples were washed three times for 5 min in PBS and mounted in Abberior Mount Solid Antifade mounting reagent (Abberior, #4-0100-007-4) under #1.5 coverslips (22x50 mm, VWR #631-0138) and allowed to cure overnight at room temperature.

STED imaging and image analysis

IB4 and ICAM2 signals were used to assess the positions of luminal (IB4 and ICAM2 positive) and abluminal (IB4 positive, ICAM2 negative) vessel membranes. Cross-sections of vessels (10-20 per animal) were selected for imaging in areas where luminal and abluminal membranes were clearly distinguishable, typically in the vicinity of the endothelial cell nucleus.

STED images were acquired using Abberior STED microscope equipped with 640 nm, 561 nm and 485 nm pulsed excitation lasers, 775 nm and 595 nm pulsed depletion lasers, UPlanSApo 100x/1.40 Oil objective (Olympus), 509/22 (GFP), 605/50 (Cy3) and 685/70 (Cy5) bandpass emission filters and spatial light modulators for STED beam shaping and alignment. Emitted light was collected with avalanche photodiode detectors using 8 ns-wide detection time gates. 120 µs total pixel dwell time per channel was used. Cy3 and Cy5 channels were acquired by line switching followed by the acquisition of the GFP channel.

STED images in 488 nm and 640 nm channels were aligned using reference images of fluorescent beads (Tetraspeck, 100 nm, ThermoFisher Scientific #T7280). Centers of beads were determined by centroid fit and resulting positions were used as control points to calculate an affine transformation between the 488 nm and 640 nm channels.

To quantify the average amount of membrane-localized Glut1 per vessel, a measurement area containing all pixels within 300 nm of manually segmented abluminal membrane was created. Luminal membranes were not included in the analysis due to frequent collapse of vessels during sample preparation. The ratio between mean Glut1 and mean IB4 signal was used as a measure of Glut1 to account for the amount of membrane in the measurement area, imaging depth and antibody penetration differences between samples. The ratio of Glut1 to IB4 was further corrected using images of Tetraspeck beads for relative intensity fluctuations between Cy5 and GFP detection channels between imaging sessions. Statistical significance was assessed using unpaired Student’s t-tests of log2 transformed data. The analysis was performed using ImageJ and Matlab 2015 (Mathworks, Inc).

Analysis of human missense variants and short linear motifs (SLiMs)

By Bora Uyar:

SLiM regular expression patterns

262 annotated SLiM class definitions (regular expression patterns) were downloaded from the Eukaryotic Linear Motif (ELM) database (Dinkel et al. 2016). In order to analyze dileucine motifs, an additional motif '.LL.' was added to this compilation and named 'LIG_diLeu_1' in order to conserve the naming convention followed by the ELM database. The general formula for the dileucine motif is [DERQ]…L[LVI] as in http://elm.eu.org/elms/TRG_LysEnd_APsAcLL_1. Importantly, however, Kozik et al. 2010 have shown that even a "dileucine on its own can act as a weak internalization signal". In fact, the CACNA1H peptide does not carry a strict dileucine motif, but it still recruits clathrin in our hands. Literature and our finding convinced us to continue our follow-ups with LL as the only requirement for the motif.

Pathogenic and non-pathogenic missense variants

Humsavar dataset: For the analysis of the missense variants that lead to *de novo* SLiM instances in protein sequences Uniprot Humsavar dataset (version 12-Apr-2017) (Famiglietti et al. 2014) was downloaded and filtered for missense variants. Variants that are classified as 'Disease' or 'Polymorphism' in this dataset were selected.

ClinVar dataset: Clinically relevant genomic variation data annotated in the ClinVar database (Landrum et al. 2016) was downloaded from the ftp server (ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/variant_summary.txt.gz) in tab-delimited format (latest update on 25th of March, 2017). The downloaded table was filtered for assembly version GRCh38, and variants of type 'single nucleotide variant' were kept. In order to integrate the ClinVar annotations with other kinds of annotations available from the Uniprot database, these nucleotide variants were translated to the Uniprot protein sequences to obtain single amino-acid substitutions using the Ensembl Variant Effect Predictor (version 82) (VEP) (McLaren et al. 2016). The output of VEP tool was filtered to only keep missense variants such that the translated amino-acid substitution occurs at

exactly the same amino acid at the same position of the Uniprot sequence with the same gene name as those of the annotation in the ClinVar dataset (‘Name’ field). Thus, 98,219 unique single amino-acid substitutions (missense variants) from 4,298 Uniprot sequences were obtained. Variants primarily annotated with clinical significance levels ‘Pathogenic’, ‘Pathogenic/Likely pathogenic’, or ‘Likely pathogenic’ were grouped as ‘Disease’ variants, while variants annotated with ‘Benign’, ‘Benign/Likely benign’, or ‘Likely benign’ were grouped into the ‘Polymorphism’ variants.

Analysis of gain of SLiMs via missense variants in disordered regions

For each reviewed human protein from Uniprot (20,191 proteins), the disorder scores of each residue were calculated using IUPred (using the ‘short’ setting). Using an IUPred disorder score cut-off of 0.4, the missense variants in disordered regions were selected. The missense variants that overlap PFAM domains were further filtered out based on the PFAM domain annotations found in the protein feature files downloaded from Uniprot in GFF format (e.g. the link to the GFF file for GLUT1 is http://www.uniprot.org/uniprot/P11166.gff). These protein feature files were also used to detect the transmembrane proteins and their cytoplasmic/extracellular regions. The missense variants in disordered regions and not overlapping any PFAM domains were further classified as variants from 1) the whole proteome, 2) the transmembrane proteins (only those that have annotation of at least one cytoplasmic domain or an extracellular domain, in total 3,836 proteins), 3) the cytoplasmic domains of transmembrane proteins, and 4) extracellular domains of transmembrane proteins. For each of these classes, the number of disease-causing variants and the number of polymorphisms that lead to a gain of SLiMs was counted and a two-sided Fisher’s Exact Test was applied to see if there is a statistically significant difference for the likelihood of a given class of SLiMs to be gained via disease-causing variant compared to that of polymorphisms.

Peptide-protein interaction network analysis

180 peptide-protein interactions that passed the LFQ filter and showed significant differential SILAC ratios between wild-type and mutant forms of the

peptides were used to compose a peptide-protein interaction network. The network was visualized using Cytoscape 3.5.1 (Shannon et al. 2003). Enriched GO terms for each sub-graph were calculated using the topGO R package (Alexa and Rahnenfuhrer, 2016).

A functional gain of dileucine motifs

Antibody feeding assay

An antibody feeding assay was used to study the gain of endocytosis by gain of dileucine motifs. For antibody internalization assay, genes and cytoplasmic regions were chosen according to the following criteria: All 11 disease mutations from Humsavar and Clinvar ('Pathogenic' or 'Conflicting interpretations of pathogenicity', in case 'pathogenic' or likely pathogenic was included in the different interpretations) that lead to a gain of a dileucine motif were considered. All regions +/-7 AAs of the mutation were analyzed according to Eukaryotic Linear Motif (ELM) database (Dinkel et al. 2016). CACNA1H_P618L and RET_P1039L were not considered for the assay since wild-type variants of the peptides already harbor trafficking motifs. For GLUT1_P485L the whole cytoplasmic C-terminus was amplified via PCR adding EcoRV 5' and NotI 3'. All other seven constructs were generated by inserting the region surrounding the mutation position with Q5® Site-Directed Mutagenesis Kit (NEB) resulting in a 15 AA insert (we were not able to generate a construct for CACNA1H_P648L) (practical help with cloning by Martha Hergeselle). For primers see Table 2. Chimeras consisting of the respective cytoplasmic region and the human TAC antigen (interleukin-2 receptor α chain, CD25) were constructed based on a TAC construct (Diril et al. 2009). HeLa cells were transiently transfected with the TAC chimera constructs using jetPRIME (Polyplus-transfection). Two days after transfection, cells were labeled with anti-TAC IgG (Santa Cruz Biotechnology) (1:1000 diluted in Opti-MEM; Invitrogen) for 30 min at 4°C. After one change of medium (to Opti-MEM at 37 °C), plasma membrane antigens were allowed to internalize for 30 min at 37°C. The cells were then fixed with 4% PFA (Sigma-Aldrich) for 10 min on ice, and surface-bound TAC antibody was blocked using goat anti-mouse serum [goat anti-mouse IgG (Thermo Fisher Scientific at a 1:5 dilution in goat serum dilution buffer, consisting of 30% normal goat serum

(Sigma-Aldrich), 450 mm NaCl in 20 mM sodium phosphate buffer pH 7.4] for 2 h at room temperature. Cells were permeabilized and blocked with goat serum dilution buffer containing 0.2% saponin for 10 min. For detection of internalized TAC antibody, a goat anti-mouse Alexa Fluor 488-conjugated IgG (Invitrogen) was added for 1 h. Cells were then washed three times for 10 min each with sodium phosphate buffer containing 0.02% saponin. For total TAC staining, the specimens were incubated for 1 h with TAC antibody diluted 1:1000. As secondary antibody, an Alexa Fluor 594-conjugated goat anti-mouse IgG (Invitrogen) was added for 30 min, and nuclei were stained using DAPI (Sigma-Aldrich). Cells were washed, and coverslips were mounted in ProLong® Gold antifade reagent (Invitrogen). For imaging, cells with positive signal in the 594 channel were chosen. All cells shown are positive for total TAC staining (594). We have seen that the level of internalized TAC chimera (488) does not correlate with the amount of total TAC staining (594), between samples and in the same sample, and hence we have decided to exclude this channel from visualization.

Table 2 Primers used for generation of TAC-chimera proteins.

Primer Name	Sequence
CACNA1H_P648_fw	TTGAGCTTGAACAGCCCTGATTAGGATATCCAGCACAGTGGC
CACNA1H_P648wt_rev	CGGGCCGTGCCCCCCGGTGCCTGGGATTGTTCTTCTACTCTTCCTC
CACN1H_P648L_rev	CAGGCCGTGCCCCCCGGTGCCTGGGATTGTTCTTCTACTCTTCCTC
ITPR1_P1059_fw	CTGGACTTGGATGACCACGGCTAGGATATCCAGCACAGTGGC
ITPR1_P1059wt_rev	TGGGGTGTTCTCCTCACTTCCTCCGATTGTTCTTCTACTCTTCCTC
ITPR1_P1059L_rev	TAGGGTGTTCTCCTCACTTCCTCCGATTGTTCTTCTACTCTTCCTC
L1CAM_S1194_fw	CTCAACGGGGACATCAAGCCCTAGGATATCCAGCACAGTGGC
L1CAM_S1194wt_rev	CGATGGCTGGCTGCTGCCAAAGGCGATTGTTCTTCTACTCTTCCTC
L1CAM_S1194L_rev	CAATGGCTGGCTGCTGCCAAAGGCGATTGTTCTTCTACTCTTCCTC
RHBDF2_P189_fw	CTGGCCCGGGGCCGGGCCTTCTAGGATATCCAGCACAGTGGC
RHBDF2_P189wt_rev	CGGATCCACAATCTTGGGCATCTTGATTGTTCTTCTACTCTTCCTC
RHBDF2_P189L_rev	CAGATCCACAATCTTGGGCATCTTGATTGTTCTTCTACTCTTCCTC
CFTR_P5_fw	GCCAGCGTTGTCTCCAAACTTTAGGATATCCAGCACAGTGGC
CFTR_P5wt_rev	CTTTTCCAGAGGCGACCTCTGCATGATTGTTCTTCTACTCTTCCTC
CFTR_P5L_rev	CTTTTCCAGAAGCGACCTCTGCATGATTGTTCTTCTACTCTTCCTC
CFTR_P750_fw	CGCATCAGCGTGATCAGCACTTAGGATATCCAGCACAGTGGC
CFTR_P750wt_rev	AGGCAGTATCGCCTCTCCCTGCTCGATTGTTCTTCTACTCTTCCTC
CFTR_P750L_rev	AAGCAGTATCGCCTCTCCCTGCTCGATTGTTCTTCTACTCTTCCTC
KCNQ1_R452_fw	CTGGACCACTTCTCTGTCGACTAGGATATCCAGCACAGTGGC
KCNQ1_R452wt_rev	CCGCCGCTCTTCTGGGGGGTCGCAGATTGTTCTTCTACTCTTCCTC
KCNQ1_R452L_rev	CAGCCGCTCTTCTGGGGGGTCGCAGATTGTTCTTCTACTCTTCCTC
KCNQ1_R591_fw	CTGAACCGAGTAGAAGACAAGTAGGATATCCAGCACAGTGGC
KCNQ1_R591wt_rev	GCGGGCGCCGATCGTGTTGCTGCCGATTGTTCTTCTACTCTTCCTC
KCNQ1_R591L_rev	GAGGGCGCCGATCGTGTTGCTGCCGATTGTTCTTCTACTCTTCCTC

Results

A peptide-based interaction screen on disease-related mutations

We set out to develop a high-throughput screen to study the changes in protein-protein interactions (PPIs) provoked by disease-causing missense mutations. We were especially interested in mutations located in intrinsically disordered regions (IDRs), since their likelihood to impact isolated protein functionalities, like PPIs, is increased. Short peptide sequences, SLiMs, in IDRs are often sufficient to mediate PPIs. This means that they can maintain functionality even when isolated from the context of the whole protein. This made us reason that quantitative interaction proteomics with immobilized synthesized peptides should enable us to study the impact of disease-causing missense mutations in IDRs.

Detecting specific and differential peptide-protein interactions

As the first step, we assembled a list of promising candidates to include in our peptide-based interaction screen (together with Marieluise Kirchner) (Figure 2).

We turned to Uniprot's Humsavar disease database as a reliable source of disease-causing mutations (Uyar et al. 2014). We predicted disordered regions with the IUPred web tool (Dosztányi et al. 2005). This allowed us to filter for missense mutations that lye in IDRs. Additionally, we decided to concentrate on mutations that cause neurological diseases (for details see Methods). We ended up with more than 100 peptide pairs of 15 AAs in length that correspond to the IDRs in both wild-type and mutant variant. All final candidates were synthesized on cellulose membrane (by JPT peptide technologies, Berlin, Germany) and used to pull down interaction partners from cell lysate (Figure 3). To identify interacting proteins, each peptide-spot then had to be excised and the associated proteins were identified and quantified with shotgun mass spectrometry.

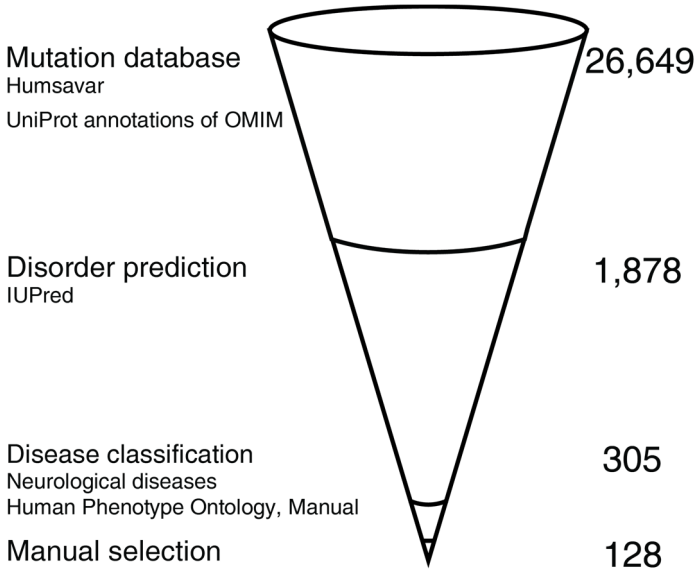


Figure 2 Candidate selection for peptide-protein interaction screen on neurological disease-causing mutations. Candidates were selected from missense mutations in the Humsavar database (Uniprot) by selecting mutations in disordered regions that cause neurological diseases (Disease classification and manual selection together with Marieluise Kirchner).

One important question that follows such a measurement is: which proteins truly interact with the peptide and which are promiscuous binders that interact non-specifically (i.e. background binders).

To address this point we applied two levels of filtering. First, we compared the normalized intensities (label-free quantification - LFQ) (Cox et al. 2014) coming from two replicate peptide spots (light and heavy SILAC replicate) with the associated proteins that had been identified for all other peptide spots. Specific interaction partners are thought to be significantly enriched for a limited number of peptides.

Second, we applied SILAC-based quantification (Mann 2006) to detect proteins binding differentially to wild-type and mutant peptide variants. To this end, two identical membranes were incubated with cell lysate from either light- or heavy-metabolically labeled cells. After excision, wild-type and mutant variant from forward (wt-light and mut-heavy) or reverse (mut-light and wt-heavy) experiment were combined in the same tube. To validate the general functionality of the screen and to provide significance thresholds for all other candidates, we included a control peptide from the literature.

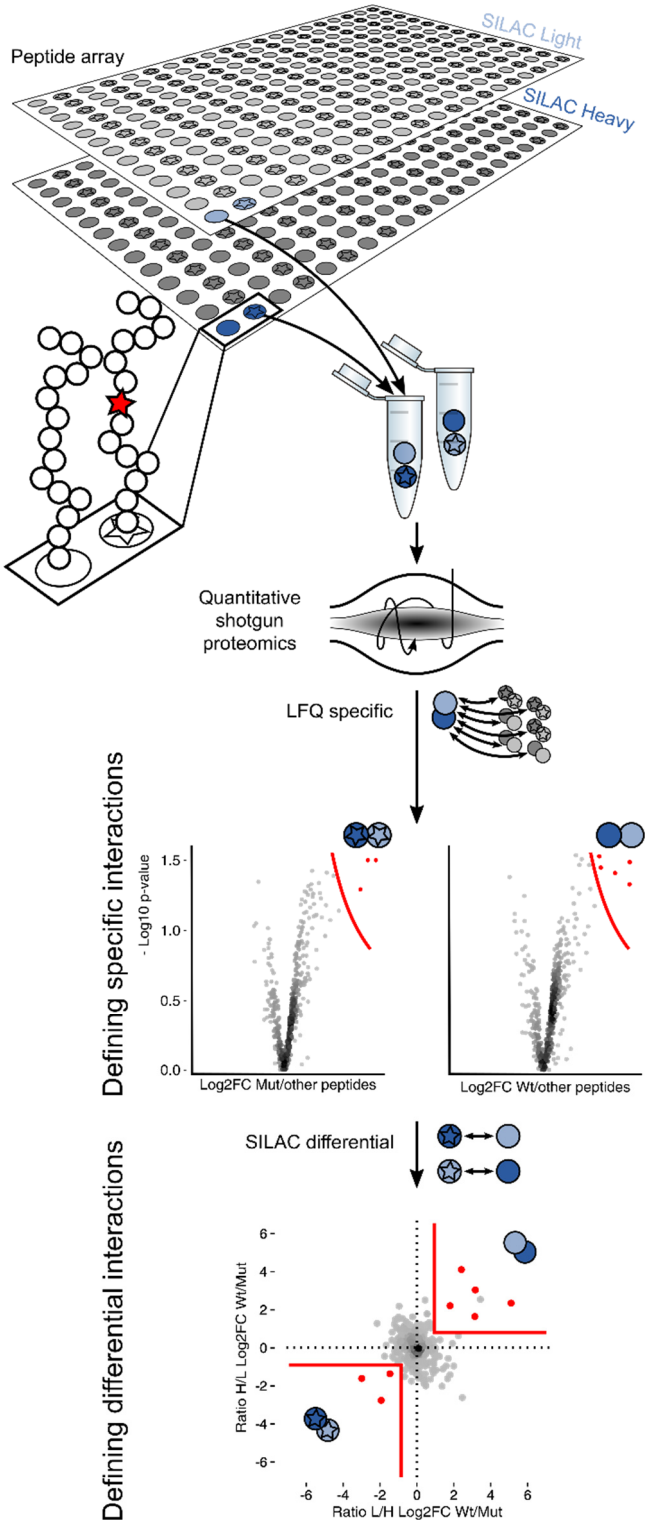


Figure 3 Quantitative interaction screen with disease-associated disordered regions. Cellulose membranes with synthetic wild-type (circles) and mutated (stars) peptides are incubated with lysate from light (light blue) or heavy (dark blue) SILAC-labeled cells to pull-down interacting proteins. Spots are excised, corresponding wild-type/mutant pairs are combined and analyzed by quantitative shotgun proteomics (represented by an Orbitrap). Middle: label-free quantification (LFQ) identifies specific interactors by comparing both replicates to all other pull-downs. Volcano plots depict protein enrichment in the two replicate pull-downs of a given peptide over all other peptide pull-downs, separately for the wild-type (left) and mutant peptide (right). The threshold (red lines) was derived from the benchmark experiment with the SOS1 peptide. LFQ-specific interactors are depicted in red. SILAC-based quantification identifies differential binders by directly comparing corresponding wild-type and mutant pairs. Differential binders of the wild-type and mutant peptide appear in the upper-right and lower-left quadrants, respectively.

PxxP motif-containing peptide recruits proteins with SH3-domains

To test the applicability of our screen we had to choose a benchmark to answer the following questions: (i) Can the peptide-based interaction screen capture known protein-protein or even known peptide-protein interactions and (ii) can it detect differential binding to wild-type and mutated peptides?

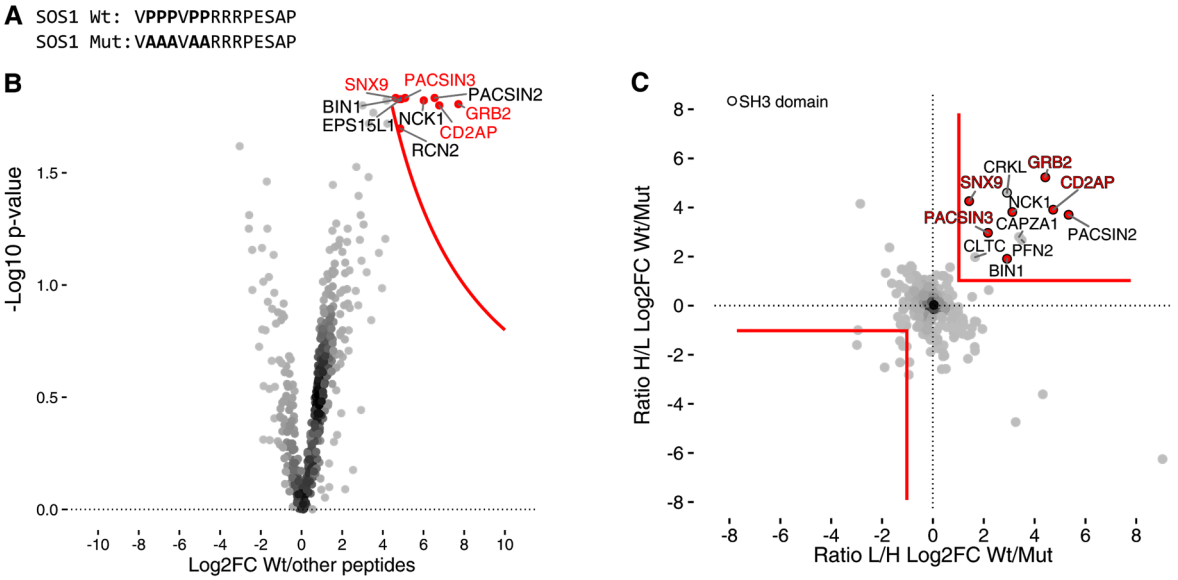


Figure 4 Proof of concept, on SH3-domain binding peptide, results in known specific and differential interaction partners. A, SOS1 derived peptide with an SH3 domain-binding PxxP motif. Prolines are substituted by alanines to disrupt binding motif. B, Volcano plot from LFQ data for wild-type SOS1 peptide. Specific binders are shown as red dots. Four out of five known binders (red gene names) are detected. C, Differential binders of the wild-type and mutant SOS1 peptide. Proteins with SH3 domains are shown with black outlines.

To this end, we included a peptide stemming from SOS1 (son of sevenless homolog 1) (Schulze and Mann 2004). This peptide represents amino acids 1150–1158 of SOS1 (SGSGVPPPVPPRRR) (Figure 4 A). It contains a proline-rich motif that is known to interact with SH3 (Src Homology 3) domain-containing proteins. Of its known interaction partners: GRB2, CD2AP, PACSIN3, SNX9 and SNX18 (Schulze and Mann 2004), only the latter could not be identified in our set-up and we identified five additional interactors. Strikingly, when comparing the wild-type to the mutated peptide in which all prolines were exchanged with alanines, all remaining significant and differential interaction partners in our setup contain SH3 domains. The SOS1 peptide is known to bind GRB2 with an

affinity of about 4 μM (Vidal et al. 1998). This made us confident that we would be able to identify other relevant low-affinity SLiM-domain interactions in our screen.

Peptide-protein interaction screen reveals possible disease mechanisms

Approximately, 400 interacting proteins were identified on average to be associated with every single peptide spot. If all of those would be considered specific interaction partners this would result in ~100,000 interactions for the whole array (Figure 5 A). One strength of shotgun proteomics, however, is that promiscuous background binders are not a problem, but can be used to normalize data and, when treated in the right way, can be distinguished from specific interactions (Keilhauer, Hein, and Mann 2015). Applying our filter criteria (LFQ filter) reduced the interactions to 618 that are considered to be specific (Figure 5 A). These were distributed over about half of the peptide candidates (120 out of 256) (Figure 5 B). The others did not show specific interactions when compared to the background. However, not all of these specific interactions were differential when comparing wild-type and mutant peptide (Figure 5 A). In 180 cases the interaction was affected by the mutation. 111 of these interactions are lost through mutations in the peptide, while 69 are gained. Of note, since pull-downs can also capture indirect binders, not all of these interactions are necessarily direct. This means that more than half of peptide-variant pairs that have significant interactors also show differential interaction of at least one of the specific interaction partners (44 out of 76) (Figure 5 B).

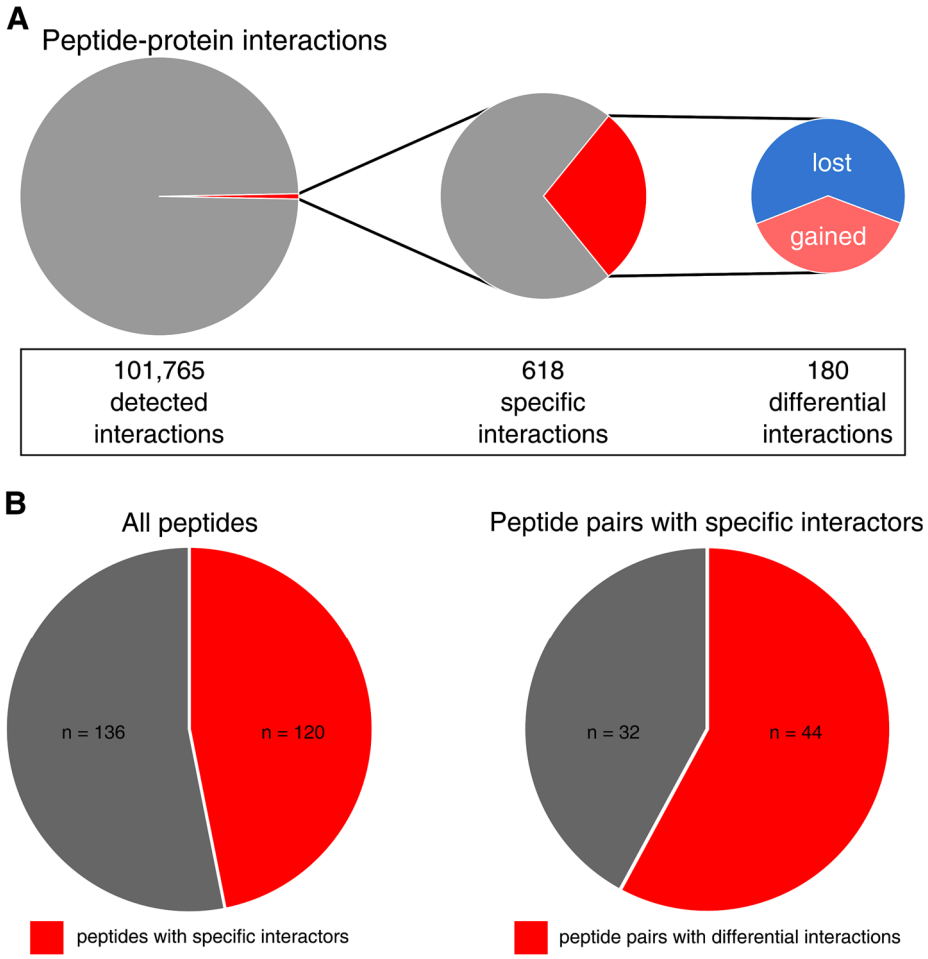


Figure 5 Specific and differential interactors of wild-type and mutant IDRs. A, Quantitative filter to select specific and differential interactors. Only a minor fraction of all detected interactions are specific (LFQ filter). Moreover, only a fraction of specific interactors is differential (SILAC filter), i.e. shows preferential binding to the wild-type or mutant form of a peptide. Mutation-induced interaction losses are more frequent than mutation-induced gains. B, Impact of the two filters on peptide candidates. After applying the LFQ filter on all interactions, about half of the 256 peptides (128 variant pairs) show at least one specific binder (left pie chart). These 120 peptides relate to 76 peptide pairs with specific interactions of wild-type and/or mutant peptide. More than half of these 76 peptide pairs show differential interactions (SILAC filter).

To illustrate all differential interactions in a compact manner, they are here displayed in a network (Figure 6). Some peptides share differential interactors, which might suggest similar disease mechanisms. It is, for example, interesting to note that in three cases the conversion of arginine to tryptophan leads to the loss of HSP90b interactions. None of these proteins, however, was known to interact with HSP90b before and so the significance of this finding is unclear.

It is also interesting to see that interactions of FLOT1 and FLOT2 are lost when arginine is converted to cysteine (GJB1_R264C), but gained when arginine is added through mutation (DKC1_G402R).

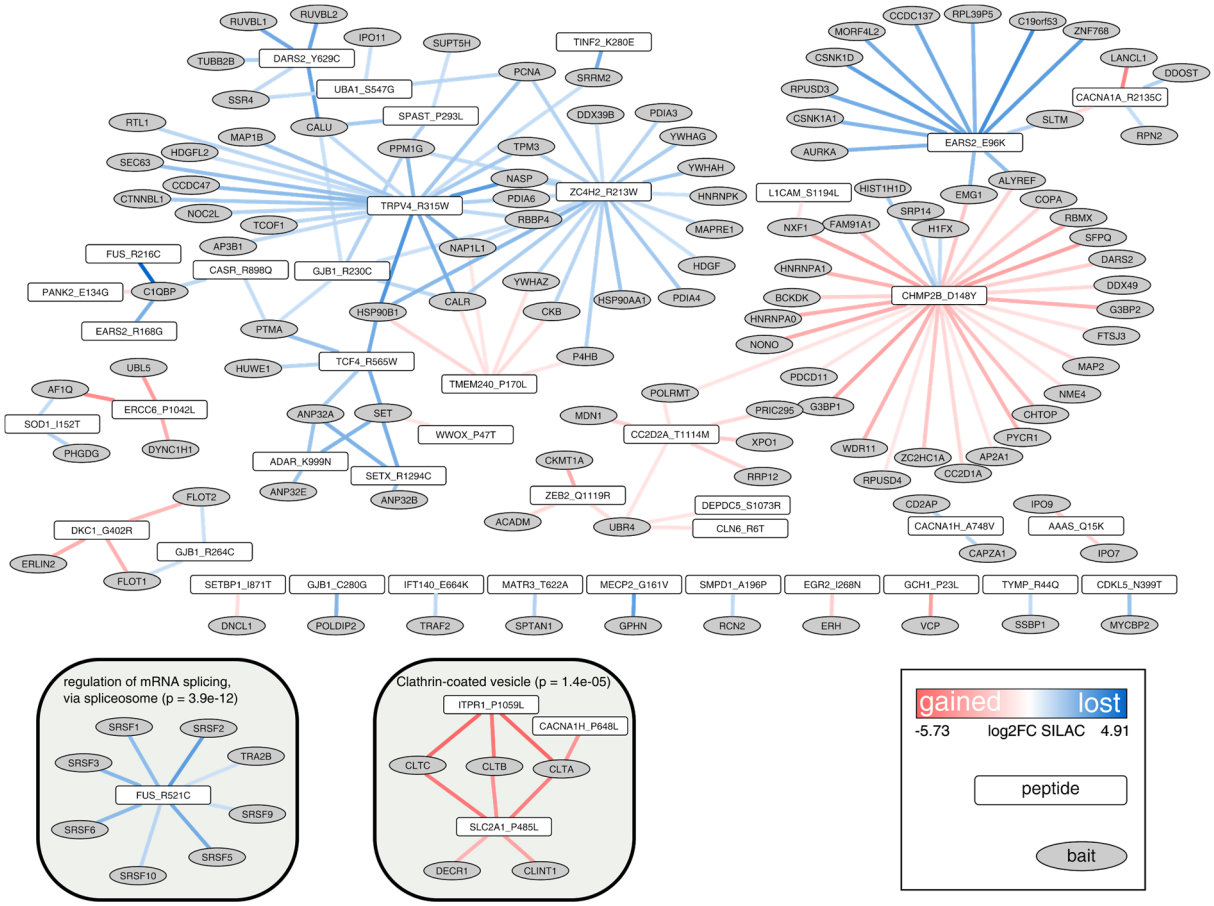


Figure 6 Differential interactions of wild-type and mutant IDRs. A network of all differential interactions. Peptides (rectangles) and interacting proteins (ovals) are represented as nodes. The edges indicate preferential binding to the wild-type (blue) or mutant (red) form of a peptide (edge-darkness indicates SILAC ratios). Highlighted subnetworks are enriched in splicing regulators and clathrin-coated vesicle proteins, respectively.

New insights into the role of FUS_R521C in amyotrophic lateral sclerosis (ALS)

One particularly interesting finding is the loss of interactions of the fused in sarcoma (FUS) R521C mutant (insert in Figure 6). FUS is an RNA-binding protein implicated in amyotrophic lateral sclerosis (ALS) (Deng, Gao, and Jankovic 2014). ALS is a fatal neurological disorder characterized primarily by the rapid degeneration of motor neurons (Sharma et al. 2016). There are over 50 mutations in FUS in ALS families (Deng, Gao, and Jankovic 2014). Most of these

mutations are dominant missense mutations that cluster in and around the C-terminal nuclear localization signal (Zakaryan and Gehring 2006). One of these mutations is R521C. In our screen, however, we found significant interaction of several splicing factors with the wild-type peptide stemming from the amino acids 512-526 of FUS. These interactions were decreased due to the R521C mutation. R521C has been shown to lead only to a mild cytosolic mislocalization (Dormann et al. 2010) and hence it is tempting to believe that there might be two disease mechanisms that add up to each other. This assumption is also corroborated by the fact that FUS is known to be involved in RNA splicing and to interact with SRSF10. Even more strikingly, this interaction has been mapped to the region of concern, since it is known that it is lost in truncated forms of FUS that lack the C-terminus (Yang et al. 1998).

Gain of dileucine motifs as a recurrent cause of disease

Another cluster stands out even more: Three peptides seem to gain interactions with clathrin when they carry a specific disease mutation (Figure 6 and Figure 7 A, B). Clathrin is a major player in intracellular trafficking, which leads to the assumption that erroneous protein trafficking might be the underlying disease mechanism. This would be particularly reasonable in the case of transmembrane proteins, and in fact, all three peptides stem from membrane spanning transporters: a calcium channel CACNA1H (calcium channel, voltage-dependent, T type, alpha 1H subunit) or $Ca_v3.2$, ITPR1 (inositol 1,4,5-trisphosphate receptor type 1) and SLC2A1 (solute carrier family 2, facilitated glucose transporter member 1) also known as GLUT1 (glucose transporter 1) (Figure 7 C). The disease mutations have something else in common: In all three cases, a proline had been mutated to a leucine, changing the sequence to two consecutive leucines (a simplified version of the dileucine motif, (D/E)…LL) (Figure 7 D). Dileucine motifs are known to be recognized by proteins involved in intracellular clathrin-mediated trafficking (Pandey 2009). To make the motif accessible to the trafficking machinery it has to face the cytosol. All three dileucine motifs are harbored in a cytosolic domain of the respective transporter. The idea suggests itself that, in all cases presented, a mislocalization of the transmembrane proteins mimics the loss of one copy of the gene, responsible for a haploinsufficient disease.

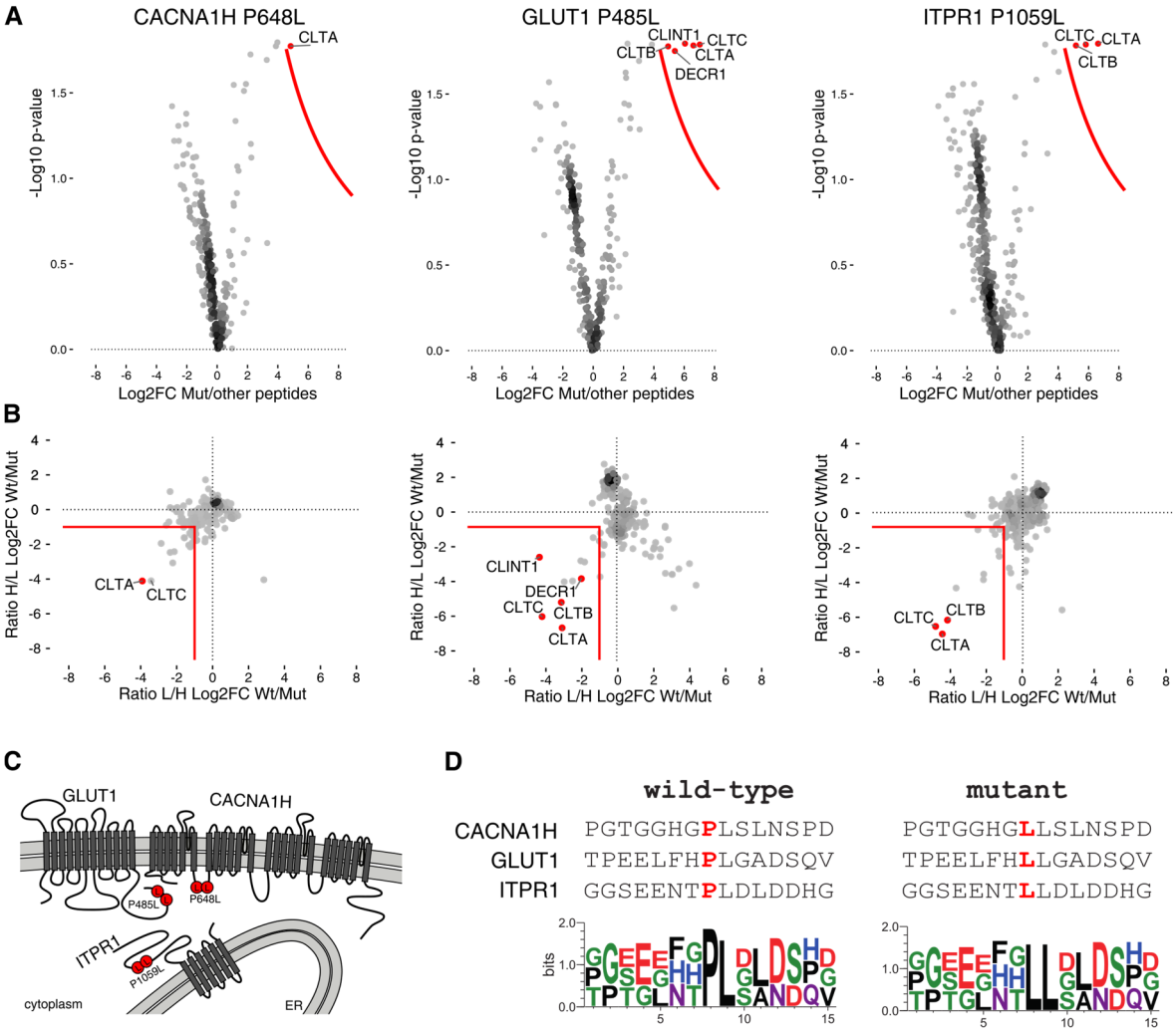


Figure 7 Recruitment of clathrin by recurrent gains of dileucine motifs. A, Volcano plots for pull-downs with mutated peptides derived from CACNA1H, GLUT1, and ITPR1. Specific binders (relative to all other pull-downs) are highlighted in red. All three peptides specifically interact with clathrin. B, Corresponding SILAC plots show that clathrin and related proteins preferentially bind to the mutated form of the peptides (relative to the wild-type). C, Graphical representation of the mutation sites. All three mutations affect cytosolic regions of transmembrane proteins. CACNA1H and GLUT1 are located mainly in the plasma membrane and ITPR1 mainly in the ER. D, Alignment of the three peptide sequences reveals a common gain of a dileucine motif.

Only the mutations in GLUT1 and ITPR1 actually lead to the creation of a textbook dileucine motif ((D/E)…LL) (ELM entry). CACNA1H_P648L does not create a classical trafficking motif. The P1059L mutation in ITPR1, however, also creates a so-called clathrin box (LLDLD). Even if the created motifs are not perfect, it has been shown repeatedly that variations are common and that dileucine can even act on its own (Staudt, Puissant, and Boonen 2016; Traub 2009; Kozik et al. 2010; Pandey 2009).

Mutations in the three transporters lead to different neurological diseases. CACNA1H_P648L is involved in childhood absence epilepsy (Chen et al. 2003). This disease manifests as loss of consciousness with sudden on- and offset. The causality of the mutation is under debate, but Chen and colleagues report that the P648L mutation in one copy of the gene leads to this specific form of epilepsy.

ITPR1_P1059L leads to spinocerebellar ataxia type 15/16 (SCA15/16) (Hara et al. 2008). SCA15/16 is characterized by pure cerebellar ataxia, very slow progression, and marked cerebellar atrophy (Gardner et al. 2005; Storey et al. 2001). Most SCA15/16 patients have heterozygous deletions at the 5’ end of the ITPR1 gene which lead to a dramatic decrease in ITPR1 expression levels (Yamazaki et al. 2011). ITPR1_P1059L was found to have indistinguishable channel characteristics from that of the wild-type protein. Anyways, heterozygous deletions and the P1059L mutation of ITPR1 exhibit similar clinical features in affected patients (Hara et al. 2008). The domain in which P1059 is localized is known to be highly regulated through splicing, phosphorylation, ATP binding but also interaction with regulatory proteins. Hence, it is accessible to protein-protein interaction.

The GLUT1_P485L mutation causes GLUT1 deficiency syndrome (G1DS), a disorder characterized by seizures and intellectual disability with onset in early infancy (Leen et al. 2010; Pascual et al. 2008; De Vivo et al. 1991). GLUT1 is a glucose transporter localized in the endothelial cells of the blood-brain barrier and in astrocytes and is mainly responsible for glucose transport into the brain and hence its energy supply. Also in the case of G1DS, loss of a single copy of GLUT1 is known to cause the disease (Seidner et al. 1998).

All three mutations lead to diseases that are known to be caused by haploinsufficiency of one gene. This is in good agreement with our theory that mislocalization leads to functional knock-out of one copy.

Peptide-protein interactions can be confirmed by far-western blotting approach

As an orthogonal method, a far-western blotting approach can be used to validate peptide-protein interactions (Figure 8). To verify interaction of clathrin

to dileucine containing peptides, we employed, in principle, the same experimental setup as before, but instead of cutting out single peptide spots to prepare them for mass spectrometry, we labeled the membrane with an antibody against clathrin, thereby making it possible to visualize clathrin interacting peptides. In agreement with our high-throughput data, we detect stronger signal at dileucine containing peptides when compared to most other peptides (specific) and also when comparing the mutant peptides to their wild-type counterparts (differential) (with Teresa Melder).

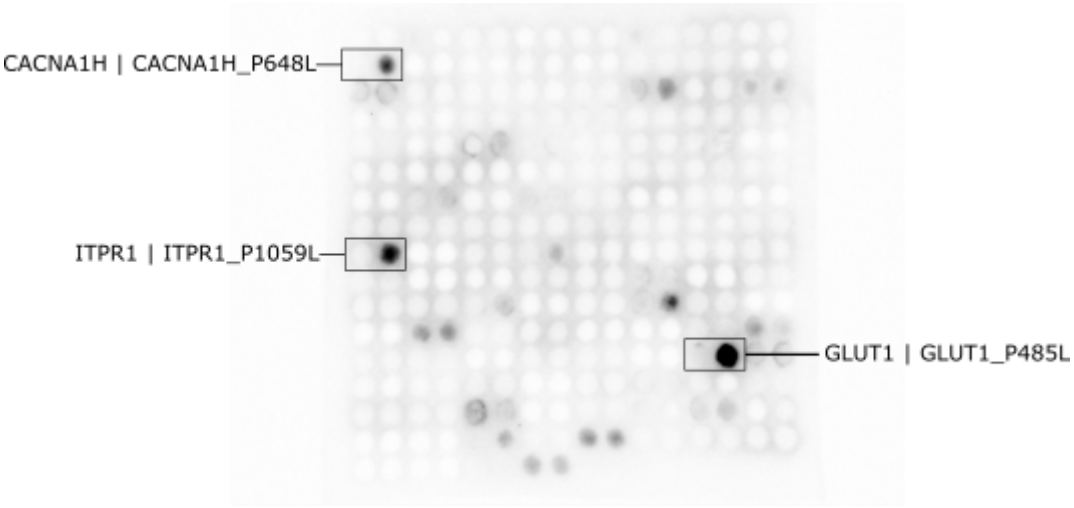


Figure 8 Recruitment of clathrin by dileucine motif-containing peptides can be confirmed by the far-western blotting approach. Whole peptide array was incubated with cell lysate and consequently probed with clathrin antibody. Dileucine motif-containing peptides show much stronger signal than all other peptides and mutant peptides are enriched compared to their wild-type counterparts (Experiment with Teresa Melder).

Adaptor proteins bind preferentially to dileucine containing peptides

I tiny flaw in the logic, however, is that dileucine motifs are generally not recognized by clathrin directly, but bind adaptor proteins that then recruit clathrin and lead to clathrin-mediated trafficking.

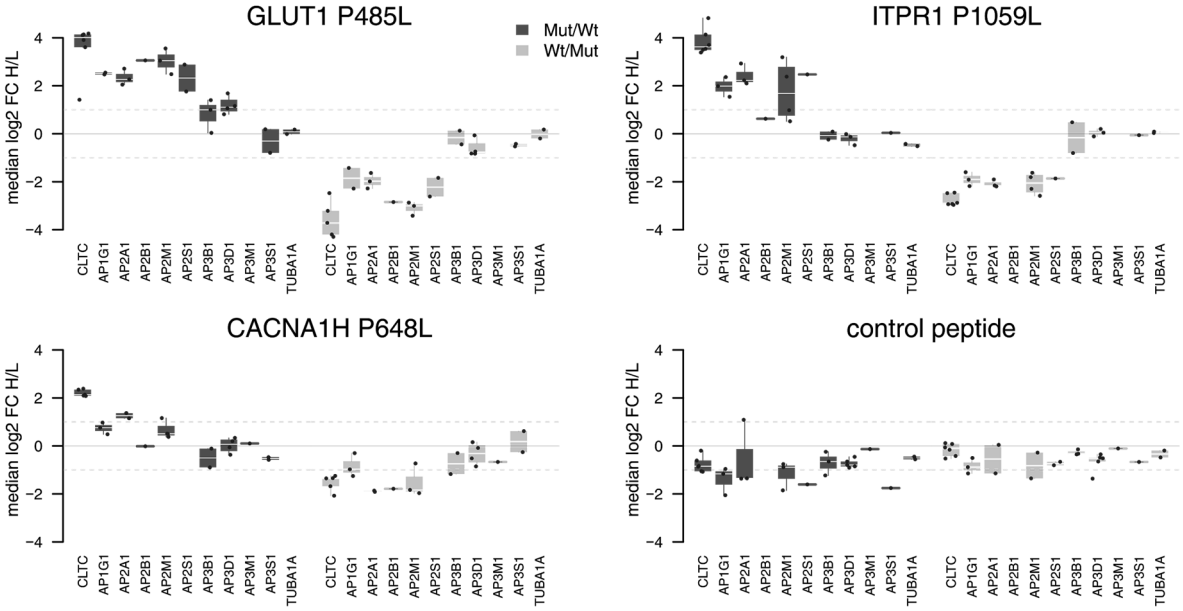


Figure 9 Adaptor proteins bind preferentially to mutant variant peptides carrying a dileucine. A highly sensitive, targeted mass spectrometry technique (parallel reaction monitoring, PRM) reveals that adaptor proteins (AP-1, AP-2, AP-3) bind preferentially to peptides carrying a dileucine when compared to their wild-type counterparts (Set-up with Henrik Zauber, analysis and figure by Henrik Zauber).

In mass spectrometry absence of evidence is not necessarily evidence of absence. Some peptides, and hence proteins, might escape identification because of their low abundance or the peptide`s characteristics. To increase the likelihood to detect adaptor proteins, in case they actually bind to the dileucine peptides, we made use of a more sensitive targeted mass spectrometry assay against peptides from several APs based on their known fragmentation spectra (Zauber, Kirchner, and Selbach 2018). Repeating the peptide pull-downs for GLUT1, CACNA1H and ITPR1 with targeted proteomics as read-out, confirmed that several APs preferentially interact with the mutated peptides (Figure 9). (Design with Henrik Zauber, Analysis by Henrik Zauber)

A dileucine-motif gain causes mislocalization of glucose transporter GLUT1

To confirm the expected mislocalization of GLUT1_P485L when compared to the wild-type, we generated a stable cell line (Kindly generated by Markus Landthaler's lab, MDC Berlin). Here, HEK cells equipped with an Flp-In site carry one copy of BirA-FLAG-tagged GLUT1 in the genome. Immunofluorescence

staining followed by confocal microscopy revealed that wild-type GLUT1 localizes mainly to the plasma membrane while GLUT1_P485L accumulates in intracellular vesicles (Figure 10).

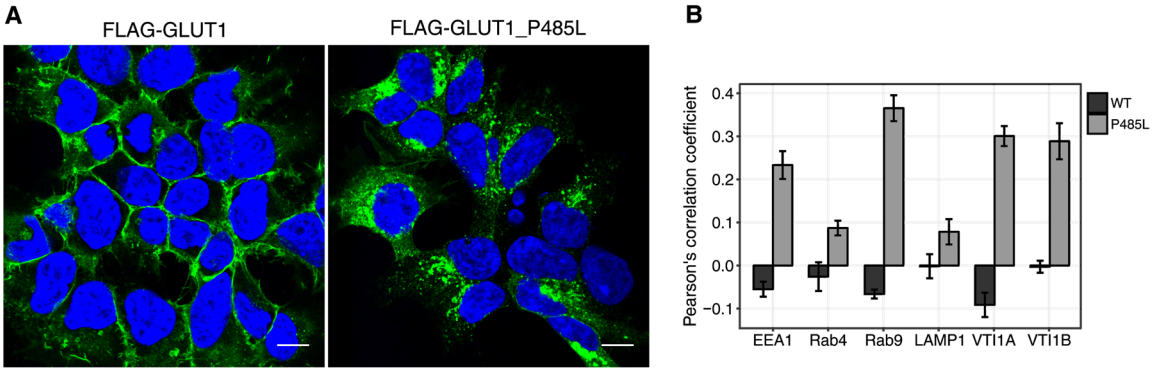


Figure 10 Mutation-induced dileucine motif gain causes mislocalization of the glucose transporter GLUT1 in stable HEK cells. A, Confocal images of GLUT1 localization in HEK cells stably expressing FLAG-GLUT1, reveal that the wild-type is localized mainly at the cell membrane while the P485L mutant is mislocalized to intracellular compartments (green, FLAG-GLUT1; blue, DAPI). Scale bars: 10 µm B, Colocalization analysis shows extensive colocalization of mutant, but not wild-type GLUT1 with markers of several endocytic compartments. Pearson's correlation coefficients (as implemented in the Imaris software) were determined for GLUT1 variants with the indicated marker proteins. Data represented as mean ± SD (Data obtained by Jing-Yuan Cheng).

To figure out more about the intracellular localization of the GLUT1 mutant and, potentially unravel trafficking proteins involved in the mislocalization, we applied a proximity labeling technique, BioID (Roux et al. 2012). When comparing GLUT1 wild-type and mutant, the biotin ligase (BirA) coupled to GLUT1 wild-type, biotinylated mainly proteins known to localize at the plasma membrane. The mutant-specific proteins, on the other hand, were significantly enriched in terms related to intracellular trafficking (Figure 11). Interestingly, subunits of three different adaptor proteins (APs) were also enriched by GLUT1_P485L (Figure 11 D).

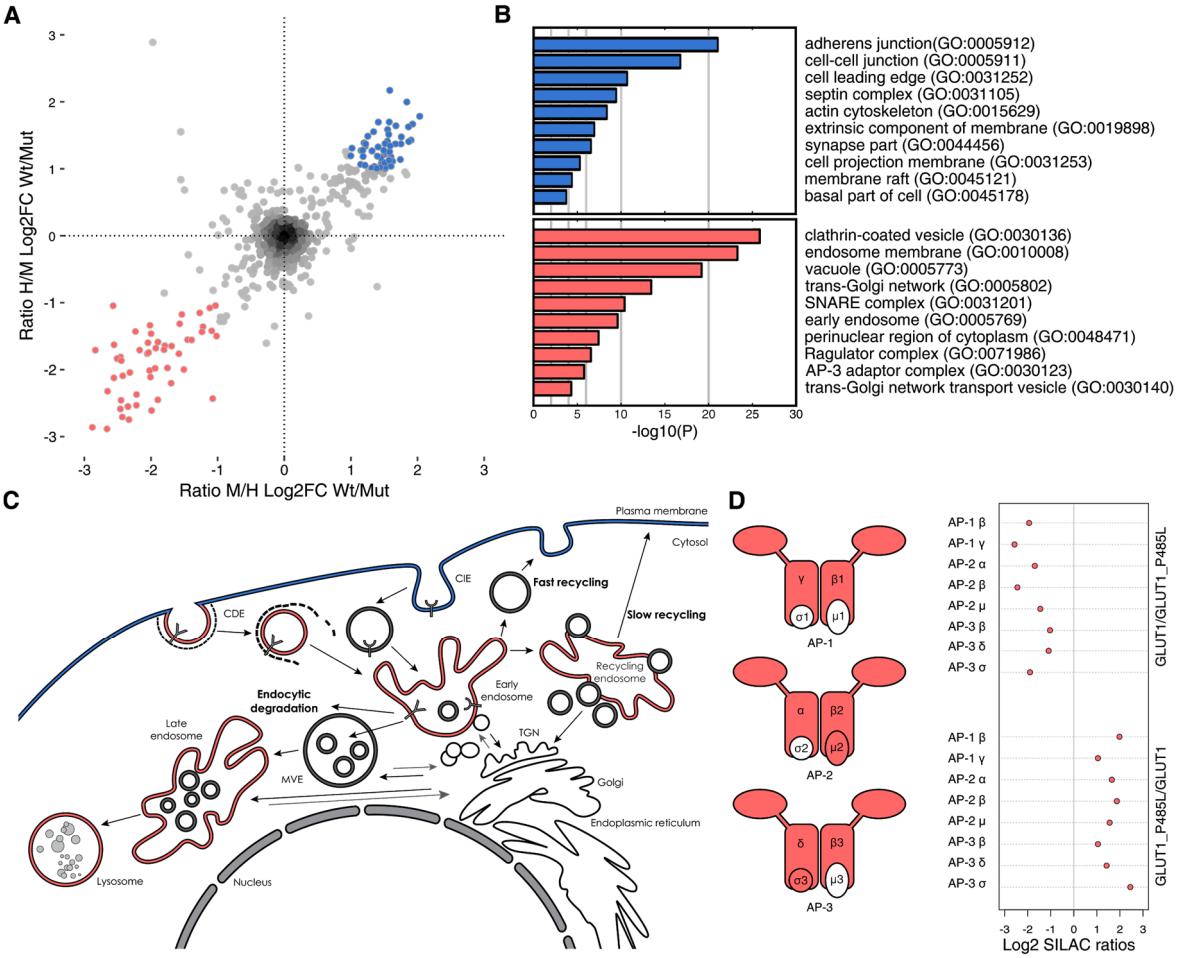


Figure 11 GLUT1_P485L mislocalizes to endocytic compartments. A, Comparison of proteins colocalizing with wild-type and mutant GLUT1 by proximity labelling (BioID). SILAC log2 fold changes (Log2FC) from two replicate experiments with swapped isotope labels. Blue and red labeled proteins are enriched by wild-type GLUT1 or mutant GLUT1, respectively. B, The ten most significant cellular component GO-terms reveal that mutated GLUT1 is involved in clathrin-dependent processes and endosomal trafficking. In contrast, wild-type GLUT1 colocalizes with plasma membrane-associated proteins. C, Coloured according to the top three enriched GO-terms and shows the variants' typical subcellular compartments. Figure adapted from (Raiborg and Stenmark 2009). D, Adaptor protein complexes AP-1, AP-2, and AP-3 show increased colocalization with GLUT1 due to P485L mutation in replicates of BioID experiment from (A). Identified subunits of APs are shown in red.

Binding of adaptor proteins causes mislocalization of GLUT1_P485L

Adaptor proteins are the connecting elements that recognize trafficking motifs on cargo proteins and recruit clathrin, so vesicles containing the cargo can be released from the membrane of origin (Traub and Bonifacino 2013). AP-1, AP-2, and AP-3 have been detected, each responsible for a different step in intracellular trafficking. AP-2 localizes to the plasma membrane and is responsible for the uptake of a wide range of cargo proteins via clathrin-mediated endocytosis. AP-1 plays a role in protein sorting in the trans-Golgi network (TGN) and endosomes (Bonifacino and Rojas 2006; Hirst et al. 2012). Also, AP-3 localizes to the TGN and endosomes, but to different endosomal buds (Peden et al. 2004; Theos et al. 2005). Here, it mediates cargo transport from tubular endosomes to late endosomes and is involved in the biogenesis of lysosome-related organelles (Park and Guo 2014).

Following up on this finding, we wanted to prove that APs can interact with the C-terminal cytoplasmic tail of GLUT1. To this end, we performed pull-downs of a GST fusion protein with the wild-type or P485L mutant cytoplasmic tail. In this *in vitro* set-up, the mutant but not the wild-type tail showed interaction with AP-1 and AP-2, further strengthening our disease hypothesis (Figure 12; Pull-down by Giulia Russo).

To test if GLUT1_P485L, in fact, reaches the plasma membrane and is then taken up again via endocytosis, we incubated cells expressing one of the GLUT1 variants with fluorescently labeled transferrin. Transferrin binds the transferrin receptor which is a well-known cargo for clathrin-dependent endocytosis (CDE). GLUT1_P485L but not GLUT1 wild-type showed extensive colocalization with endocytosed transferrin (Figure 13). This indicates that only GLUT1_P485L is involved in major trafficking by CDE from the plasma membrane.

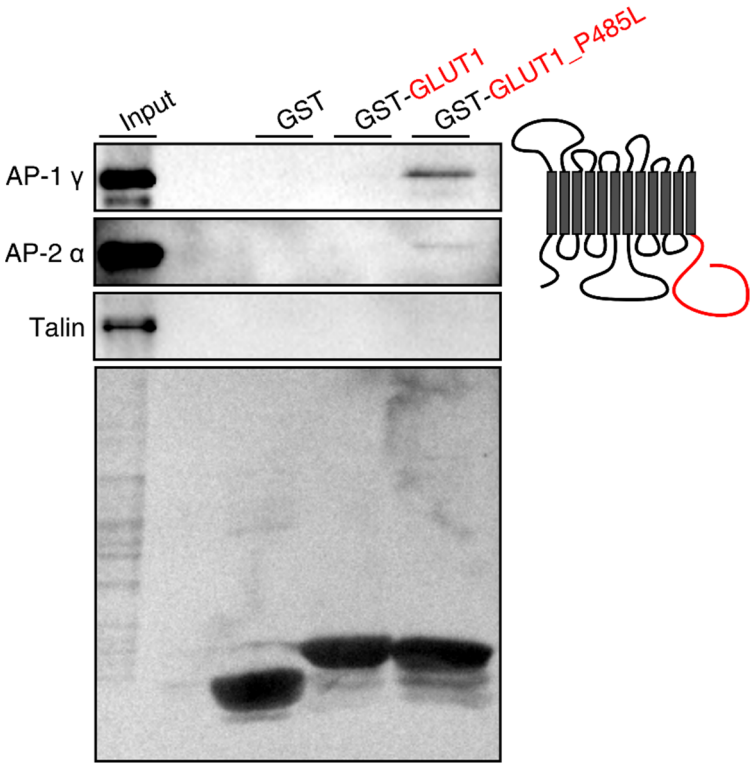


Figure 12 GST-pull down corroborates that GLUT1_P485L interacts with AP-1 and AP-2. P485L mutant but not wild-type GLUT1 C-terminal tail interacts with AP-1 and AP-2. Tails were tagged with GST to pull-down interaction partners from mouse brain lysate. Talin is shown as a negative control and is not pulled down from either of the two variants (Experiment by Giulia Russo).

Since AP-2 is responsible for endocytosis from the plasma membrane, we hypothesized that an AP-2 knock-down should result in a reconstitution of GLUT1_P485L levels at the plasma membrane and hence a rescue of the disease phenotype. To test this idea, we knocked down AP-2 expression levels via siRNA targeting the μ-subunit (see Methods). Consistently with our prediction, AP-2 knock-down restored GLUT1_P485L localization at the plasma membrane (Figure 14).

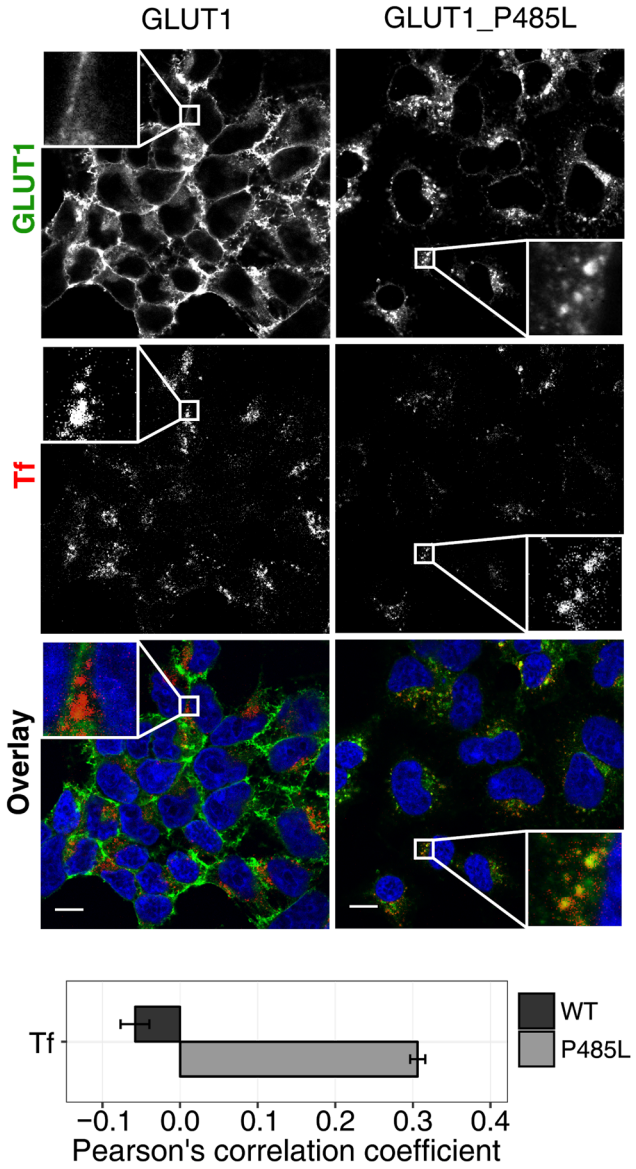


Figure 13 GLUT1_P485L colocalizes with endocytosed transferrin. Mutant but not wild-type GLUT1 extensively colocalizes with endocytosed transferrin. HEK cells stably expressing FLAG-GLUT1 were incubated with fluorescently labeled transferrin for 10 min before fixation. Scale bar: 10 µm. Quantification is represented as mean ± SD.

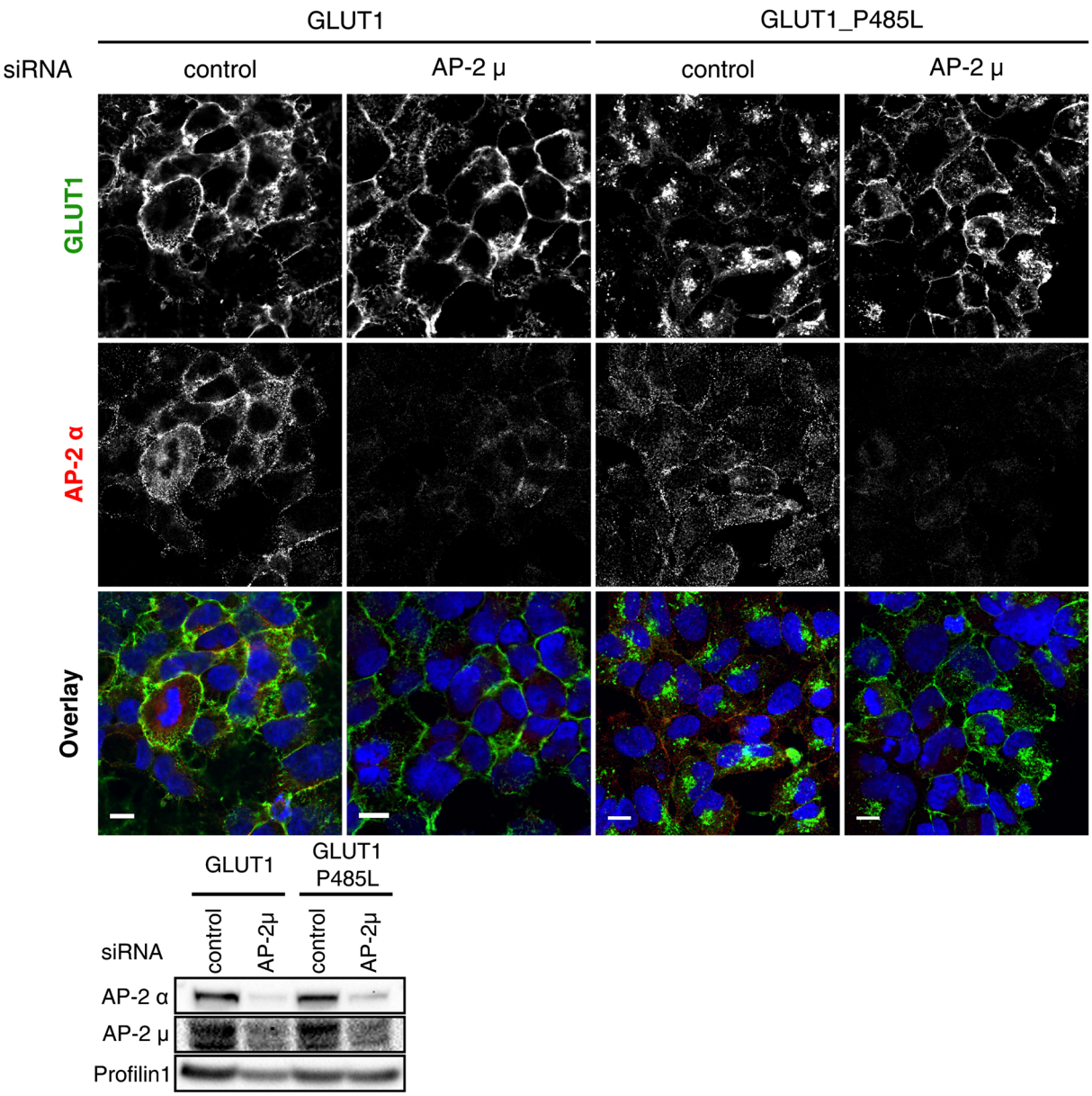


Figure 14 GLUT1_P485L mislocalization can be rescued by knockdown of AP-2. Western blot against AP-2 α- and μ-subunits shows downregulation after two rounds of siRNA transfection against AP-2 μ. AP-2 knockdown leads to relocalization of GLUT1_P485L to the plasma membrane and hence rescue of the mutation phenotype. Scale bar: 10 μm.

GLUT1_P485L is functional for glucose uptake

The next question posing itself is if GLUT1_P485L is indeed functional after its localization is restored to the plasma membrane. We decided to study this via a radioactive glucose uptake assay after knock-down of AP-2 (Figure 15). It is indeed possible to distinguish the additional amount of glucose that is taken up by the stable HEK cells via the additional copy of GLUT1 from their basal glucose

uptake (comparison of uptake -doxycycline with +doxycycline). In both cases, glucose uptake also ranges significantly over background measurements (comparison to values obtained after use of glucose uptake inhibitor, cytochalasin B). While glucose uptake in wild-type GLUT1 remained constant after knock-down of AP-2, it leads to a significant increase of glucose uptake in the P485L mutant. Since expression levels of GLUT1 variants in the two stable cell lines are not necessarily the same, this setup does not allow to draw conclusions about the comparison of wild-type and mutant levels. However, the findings make us confident, that mislocalization of an, at least partially, functional GLUT1 is the cause for the disease.

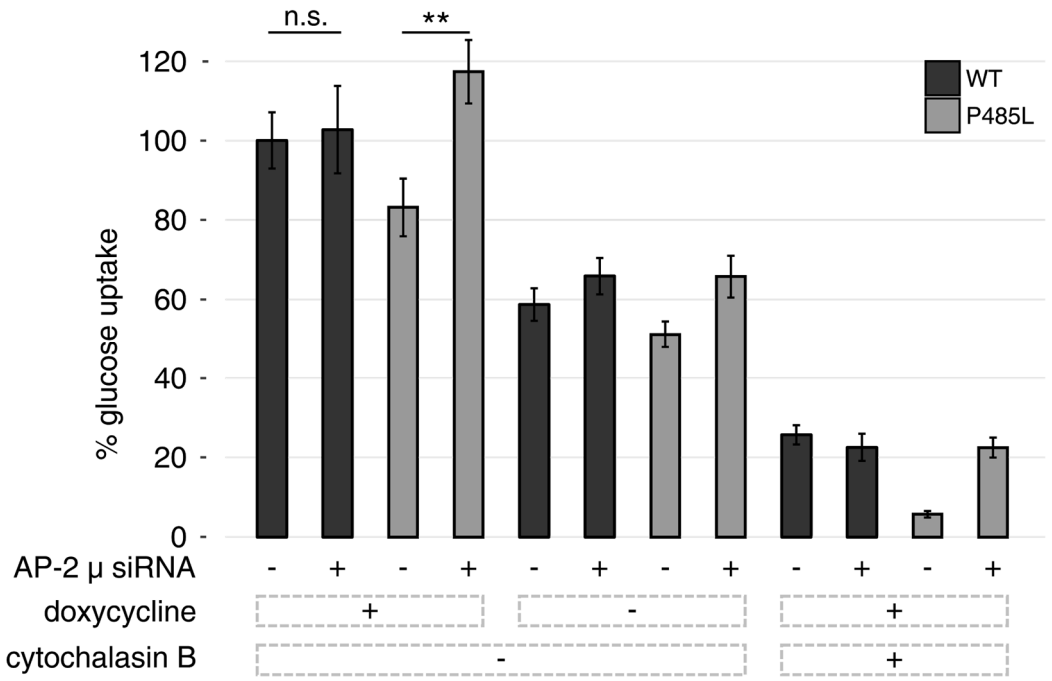


Figure 15 Glucose uptake of cells expressing GLUT1_P458L increases after AP-2 knockdown. GLUT1 expression of stable HEK cells was induced by doxycycline, cytochalasin B inhibition was used as a control. % glucose uptake is relative to GLUT1 wild-type, +doxycycline, -cytochalasin B. Mean values of technical triplicates from three independent experiments are shown. We only compare glucose uptake within and not between cell lines to account for possible differences between clones. Error bars: SEM. ** p-value < 0.01 from a paired, one-sided t-test.

GLUT1_P485L mutation leads to gain of function

In theory, it is still possible that mutation of proline 485 indirectly leads to mistrafficking by disrupting some existing SLiM. To test this, we exchanged the proline not by leucine but by alanine. In this way, we did not create a dileucine motif but made sure that a proline dependent motif would be disrupted. As expected, the alanine mutant mostly resembled the wild-type localization, providing additional evidence that the gain of dileucine motif is indeed causative for the mislocalization of GLUT1_P485L (Figure 16; Experiment together with Jing-Yuan Cheng).

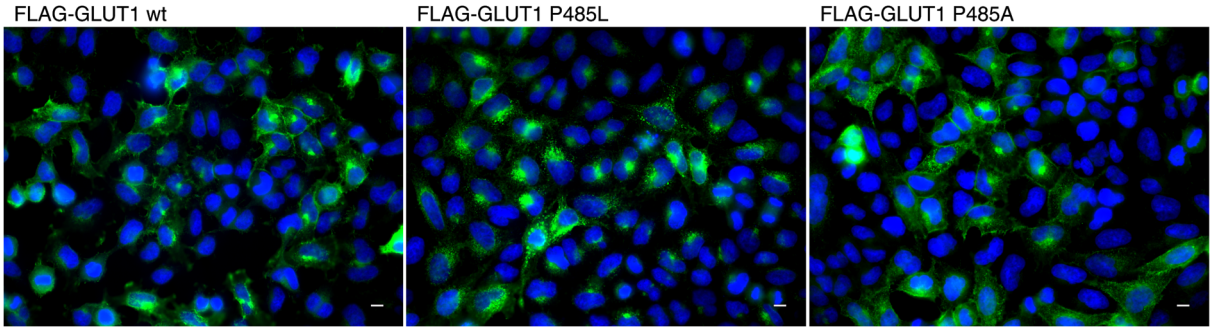


Figure 16 GLUT1_P485L is a gain of function mutation. Mutating proline 485 of GLUT1 to alanine instead of leucine maintains the wild-type phenotype in transiently transfected HEK 293 cells. Scale bar: 10 µm.

GLUT1 mislocalizes in patient-derived induced pluripotent stem cells (iPSCs)

All findings presented so far were either obtained *in vitro* or using stable cell lines that overexpress a tagged version of GLUT1. We were wondering if the same phenotype would be obtained also in a more physiological context. For this aim, we decided to turn to induced pluripotent stem cells from a patient carrying the GLUT1 P485L mutation. Patient fibroblasts were obtained via a skin punch from a GLUT1 deficient patient harboring the P485L mutation and were kindly obtained for us by her attending doctor (Juan M. Pascual, UT Southwestern Medical Center). Fibroblasts were then reprogrammed into iPSCs by RNA-based transfer of pluripotency markers (Sebastian Diecke, Core Facility Stem Cells, MDC) (Figure 17 A). Immunofluorescent staining of GLUT1 in control iPSCs showed predominant localization at the plasma membrane. Patient cells

additionally revealed characteristic intracellular GLUT1 accumulations (Figure 17 B). This shows that also untagged GLUT1_P485L at physiological levels is differently localized when compared to the wild-type variant. Patient cells still show partial staining of the plasma membrane, which is in agreement with the fact that the patient carries a heterozygous GLUT1 mutation (Leen et al. 2010; Pascual et al. 2008; Slaughter, Vartzelis, and Arthur 2009) and the cells are most likely populated with wild-type as well as GLUT1_P485L variants.

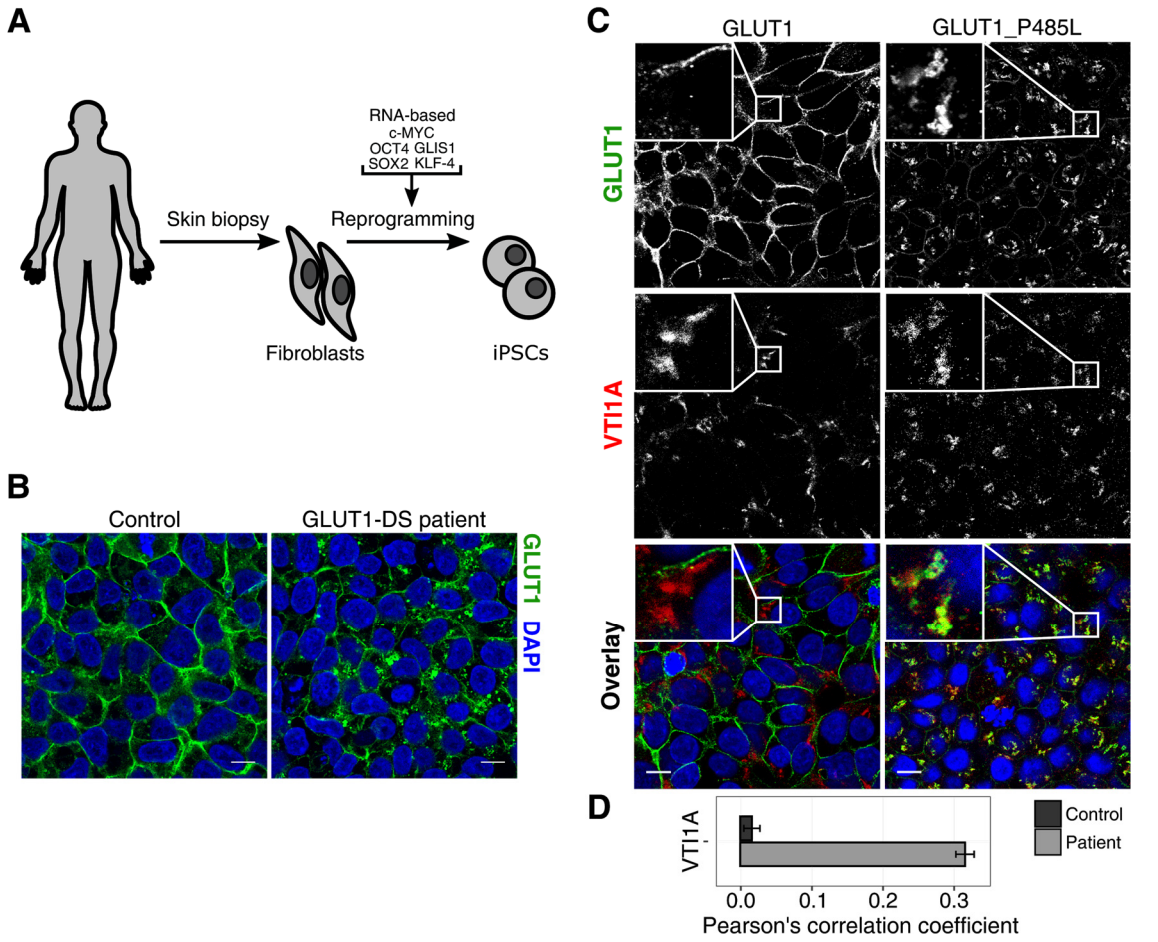


Figure 17 GLUT1_P485L shows intracellular mislocalization in iPSCs derived from a patient suffering from GLUT1 deficiency syndrome. A, A skin sample was taken from a GLUT1 deficiency patient with a heterozygous GLUT1_P485L mutation. Fibroblasts were grown and reprogrammed to iPSCs (Skin biopsy by Juan M. Pascual, generation of iPSCs by Sebastian Diecke). B, Heterozygous GLUT1_P485L mutation leads to partial mislocalization of GLUT1 in patient-derived iPSCs. Scale bars: 10µm. C and D, GLUT1_P485L colocalizes with the post-Golgi SNARE VTI1A. Scale bar: 10µm (C). Data represented as mean ± SD (D).

Intracellular GLUT1_P485L colocalizes with VTI1A, a post-Golgi-SNARE that functions throughout the endocytic pathway. This is consistent with data from

stable HEK cells and shows that GLUT1_P485L mislocalizes to endocytic compartments.

GLUT1 is not endocytosed from the plasma membrane in iPSCs

However, while in HEK cells colocalization with almost all endocytic trafficking compartments can be detected (data from Jing-Yuan Cheng), in patient iPSCs GLUT1 is concentrated in a pattern resembling the trans-Golgi network (TGN). There also is no colocalization of GLUT1_P485L with fluorescently labeled transferrin (data not shown). This indicates that GLUT1_P485L in steady state is not actively involved in endocytic trafficking from the plasma membrane. In agreement with these findings, AP-2 knockdown did not have the same effect on GLUT1_P485L localization in iPSCs as in HEK cells (data not shown). On the contrary, it did not seem to affect the mutant's localization at all. There could be different reasons for this (see Discussion).

GLUT1_P485L localization at the blood-brain barrier is perturbed *in vivo*

To show that GLUT1_P485L mislocalizes in its physiological environment and in a more physiological context of GLUT1 deficiency syndrome, we sought to study this mutant in endothelial cells of the blood-brain barrier in mice. To this end, we generated a mutant mouse harboring the GLUT1 P485L mutation via CRISPR/Cas9 technology (Figure 18 A; Mouse generated by Ralf Kuehn, Transgenic Core facility, MDC/BIH). Only heterozygous mice carrying the P485L mutation could be generated. These mice, however, appeared viable, fertile and did not display any obvious phenotype. We did not obtain any homozygous mice from six heterozygous crossings (genotyping with Martha Hergeselle). A more detailed inspection revealed that homozygous mutant pups come to term but are removed from the litter by the dams. To be able to compare the localization of GLUT1 in endothelial cells of the blood-brain barrier of wild-type, heterozygous and homozygous mutant mice, we carried out immunofluorescence stainings of the cerebral cortex in embryos of (E) 14.5-15.5 days (staining by Luis R. Hernandez-Miranda). These stainings reveal an

apparent decrease in GLUT1 levels from wt to homozygous mutants. Vessel morphology and length appears normal, as shown by anti-ICAM2 staining to label the luminal plasma membrane and isolectin B4 (IB4) staining to label the entire endothelial plasma membrane (Figure 18 B). Anyhow, endothelial cells only have very narrow space in between the luminal and abluminal membranes, which is in fact too narrow to be resolved with standard confocal microscopy (Wong et al. 2013; Cox and Sheppard 2004). We employed super-resolution, stimulated emission depletion (STED) microscopy, which is able to resolve the distance between the two membrane layers (Figure 18 C; STED imaging and analysis by Anna Szymborzka). Quantification revealed that GLUT1 levels seem to drop slightly in heterozygous mice and are significantly reduced in homozygous mutants (Figure 18 D).

This indicates that GLUT1 is removed from the plasma membrane *in vivo* and it disproves the possibility that GLUT1_P485L remains on the membrane causing disease because of its incapability to transport glucose.

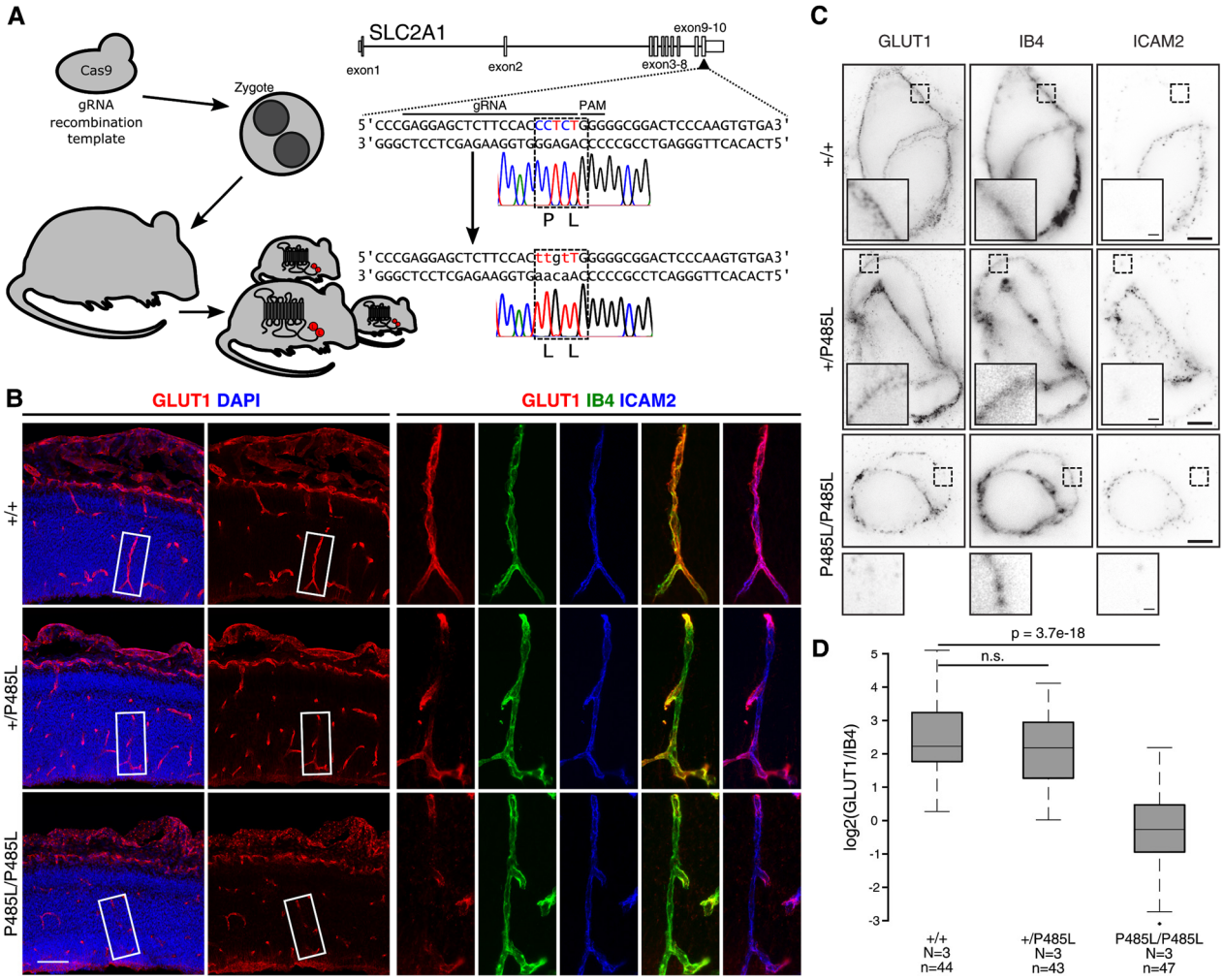


Figure 18 GLUT1_P485L in endothelial cells of the blood-brain barrier in mice. A, A mouse carrying the GLUT1 P485L mutation was created by CRISPR/Cas9-targeted method. PAM sequence and gRNA are marked in the targeted region of Slc2a1 (GLUT1 gene). Sanger sequencing confirmed insertion of mutation (chromatogram A = green, T = red, C = blue, G = black) (Mouse generated by Ralf Kühn). B, Immunohistological analyses of cortical slices of wild-type, heterozygous, and homozygous GLUT1 mutant mice using antibodies against GLUT1 (red) and DAPI (blue) as counterstain (left panels); a higher magnification of a vessel stained by antibodies against GLUT1 (red), IB4 (green), and ICAM2 (blue) is shown in the right panels. Scale bar: 100 µm (By Luis R. Hernandez-Miranda). C, Representative STED images of transverse cross-sections through brain vessels of wild-type (+/+), heterozygous (+/P485L) and homozygous (P485L/P485L) mutant mice stained with isolectin B4 and antibodies against GLUT1 and ICAM2. Insets show a fragment of the abluminal membrane (IB4 positive, ICAM2 negative) indicated with a black box. Scale bars: 2 µm (main panels); 0.25 µm (insets). D, Quantification of GLUT1 signal relative to IB4 signal in vessel membranes (n, number of vessels per genotype; N, number of animals per genotype). Boxplot central line indicates the median, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively (STED imaging and analysis by Anna Szymborzka).

How common are dileucine gains in disease?

Since we found three mutation-based dileucine gains in cytoplasmic tails of transmembrane proteins in our rather limited interaction screen, we were curious to see how many more potential "dileucineopathies" could be found in disease databases. We screened the Humsavar and Clinvar database for additional dileucine creating mutations in disordered cytosolic regions of transmembrane proteins. Taken together, we found eight additional mutations, four from Humsavar, and four from Clinvar (analysis by Bora Uyar).

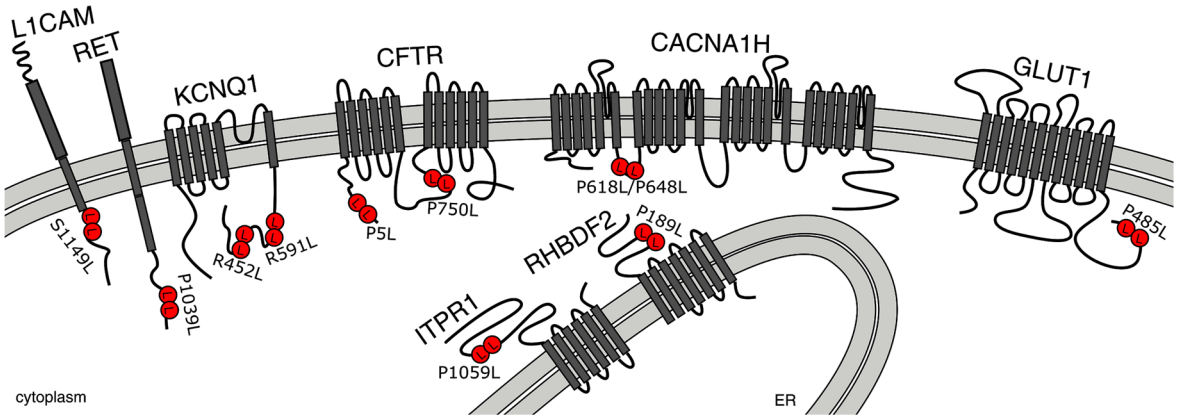


Figure 19 Dileucineopathies: dileucine motif gains in cytoplasmic regions of transmembrane proteins. A systematic bioinformatic search (in Humsvar and Clinvar databases) revealed a total of 11 pathogenic mutations in cytosolic segments of 8 different transmembrane proteins that create dileucine motifs (Analysis by Bora Uyar). See also: **Table 5**.

These mutations affect a total of five additional proteins to the ones found in our screen (Figure 19; Table 5). Among these mutations are also two dileucine gains in the cytoplasmic domain of the cystic fibrosis transmembrane conductance regulator (CFTR) that lead to cystic fibrosis. Interestingly, an increased endocytosis rate of CFTR has already been reported to lead to cystic fibrosis (Silvis et al. 2003). In this study, the disease is caused by increased removal from the cell surface due to a novel tyrosine-based endocytic sequence within an intracellular loop in CFTR. One important question remaining is if the creation of a dileucine is actually disease causative or a mere irrelevant byproduct of disease mutations that are actually based on a different mechanism. If the latter were the case, we would expect an equal distribution of novel dileucine motifs in disease and in non-pathogenic polymorphism databases. In the case of functional dileucine motifs, on the other hand, we would expect them to be especially

enriched in their region of action, which are disordered regions of transmembrane proteins facing the cytoplasm. A global survey of all disordered regions of the entire proteome revealed that dileucine motif gains occurred at about the same rate in disease and non-pathogenic variants (OR = 0.81, p-value = 0.319, two-sided Fisher's Exact Test) (Figure 20 B). In the cytosolic tails of transmembrane proteins, however, we observed a 3.7-fold enrichment of dileucine motifs implicated in disease (OR = 3.7, p-value = 0.017, two-sided Fisher's Exact Test, Figure 20 B). A comparison (disease-associated versus polymorphism) of all gained motifs in disordered regions of cytoplasmic tails of transmembrane proteins reveals the dileucine motif to have the most significant and specific enrichment (Figure 20 A). Interestingly, no double motif of any other amino acid is significantly enriched in the same kind of analysis (Table 3).

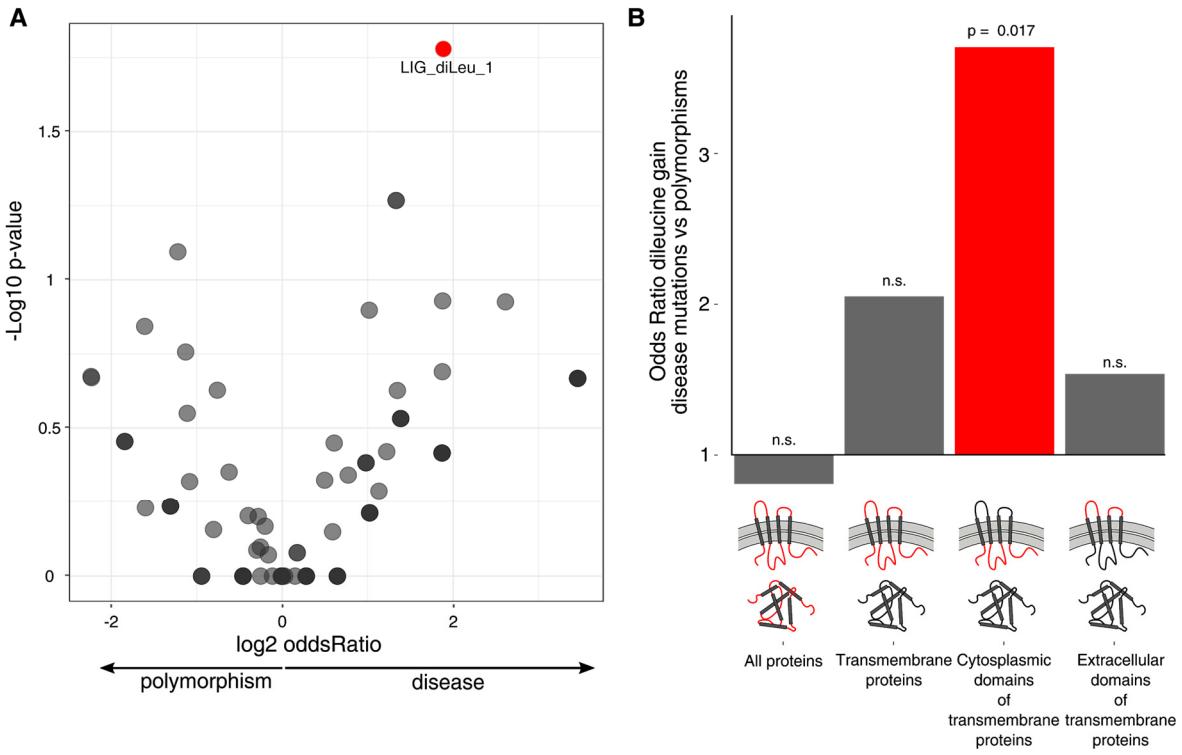


Figure 20 Mutation-induced dileucine motifs are a significant cause of disease. A, Comparison of all gained motifs (disease-associated versus polymorphism) in disordered regions of cytoplasmic tails of transmembrane proteins reveals the dileucine motif to have the most significant and specific enrichment (Analysis based on Humsavar, by Bora Uyar). B, Relative frequency of dileucine motif gains in disease mutations and polymorphisms in different disordered regions (IUPred score >= 0.4) of the proteome. Dileucine motif gain is significantly enriched only in disordered regions of the cytoplasmic domains of transmembrane proteins (two-sided Fisher's exact test; analysis by Bora Uyar).

Table 3 Enrichment of all amino acid duplicates in disease. Of all amino acids in cytoplasmic regions of transmembrane proteins, only double leucines are significantly enriched in disease (Analysis by Bora Uyar).

SLiM	Gain in disease	All other disease mutations	Gain in polymorphism	All other polymorphisms	p-value	oddsRatio	log2 oddsRatio
diLeu	7	302	7	1121	0.0168	3.707	1.8903
diArg	0	309	10	1118	0.1313	0.172	-2.5395
diMet	1	308	0	1128	0.2150	10.974	3.4560
diAsp	2	307	3	1125	0.2937	2.614	1.3863
diGly	1	308	12	1116	0.3207	0.434	-1.2042
diPro	1	308	12	1116	0.3207	0.434	-1.2042
diSer	9	300	24	1104	0.3956	1.425	0.5110
diLys	3	306	7	1121	0.4559	1.708	0.772
dille	0	309	5	1123	0.5911	0.33	-1.5995
diVal	2	307	4	1124	0.6149	2.032	1.0229
diAla	1	308	9	1119	0.6991	0.573	-0.803
diThr	3	306	8	1120	0.7113	1.505	0.5898
diGlu	4	305	12	1116	0.7597	1.316	0.3962
diAsn	0	309	2	1126	1	0.728	-0.4580
diCys	0	309	0	1128	1	1	0
diGln	1	308	7	1121	1	0.727	-0.4600
diHis	1	308	3	1125	1	1.564	0.6452
diPhe	0	309	1	1127	1	1.214	0.2798
diTrp	0	309	1	1127	1	1.214	0.2798
diTyr	0	309	1	1127	1	1.214	0.2798

Are dileucine mutations in other proteins functional?

The question remains if these additional dileucine motif gains can cause mistrafficking similar to GLUT1_P485L. To show if the newly formed motifs are active in endocytosis, we performed antibody feeding experiments. To this end, we created chimeric constructs consisting of the IL-2 receptor alpha chain (TAC) fused to mutated and wild-type cytosolic regions of the respective disease protein (practical assistance for cloning from Martha Hergeselle). TAC protein is brought to the plasma membrane and the additional tail decides if the protein is subsequently internalized or if it stays at the plasma membrane. Cells expressing these proteins were incubated with antibodies against the extracellular region of TAC and, in case of an active endocytosis motif, allowed to endocytose the chimeric proteins together with bound antibodies. A specific

staining protocol was then used to exclusively detect internalized antibodies (Diril et al. 2009). We expect no uptake of the wild-type fusion proteins, but a visible staining pattern for the mutant versions. Of the eleven dileucine candidates, two motif gains had been excluded from the screen because they already contain a predicted endocytic motif in the wild-type sequence (according to prediction with ELM). Of the remaining nine, for one we could not obtain the correct construct. Of the remaining candidates, one was the fusion proteins for GLUT1 that served as a positive control plus seven additional dileucine motif gains. Four of the seven tested mutations resulted in increased internalization relative to the corresponding wild-type sequences (Figure 21). Three mutations, however, did not behave as expected. KCNQ1_R591, does not acquire endocytosis. This might indicate that a dileucine cannot be surrounded simply by any other amino acid to lead to a functional motif. ITPR1_P1059 and CFTR_P750, both already show endocytosis of the wild-type construct. In the case of CFTR_P750 this could be explained by the fact that the wt peptide contains a degenerate endocytosis motif **E**QGEA**IL**PRISVIST. Kozik and colleagues have shown that in order to gain a functional motif the distance between acidic residue (E) and LL/IL can vary between two and four (Kozik et al. 2010).

ITPR1_P1059L is a peptide stemming from the original peptide screen. It is a special case as is does not only contain a dileucine motif (E…LL) but also a so-called clathrin box (L[IVLMF].[IVLMF][DE]). The wt peptide, however, does not even contain a degenerate version of the known motif. It is hence not clear why it shows endocytosis. However, both the original peptide screen and the PRM based analysis result in stronger recruitment of clathrin or APs by the mutant when compared to the wild-type. Even if the antibody uptake assay in its present form cannot be considered to be quantitative, it is intriguing to think that the mutant form seems to display a stronger staining pattern when compared to the wt.

Collectively, these results indicate that several additional pathogenic dileucine motif gains cause protein mistrafficking. This makes us think that there might be a general concept of “dileucineopathies”.

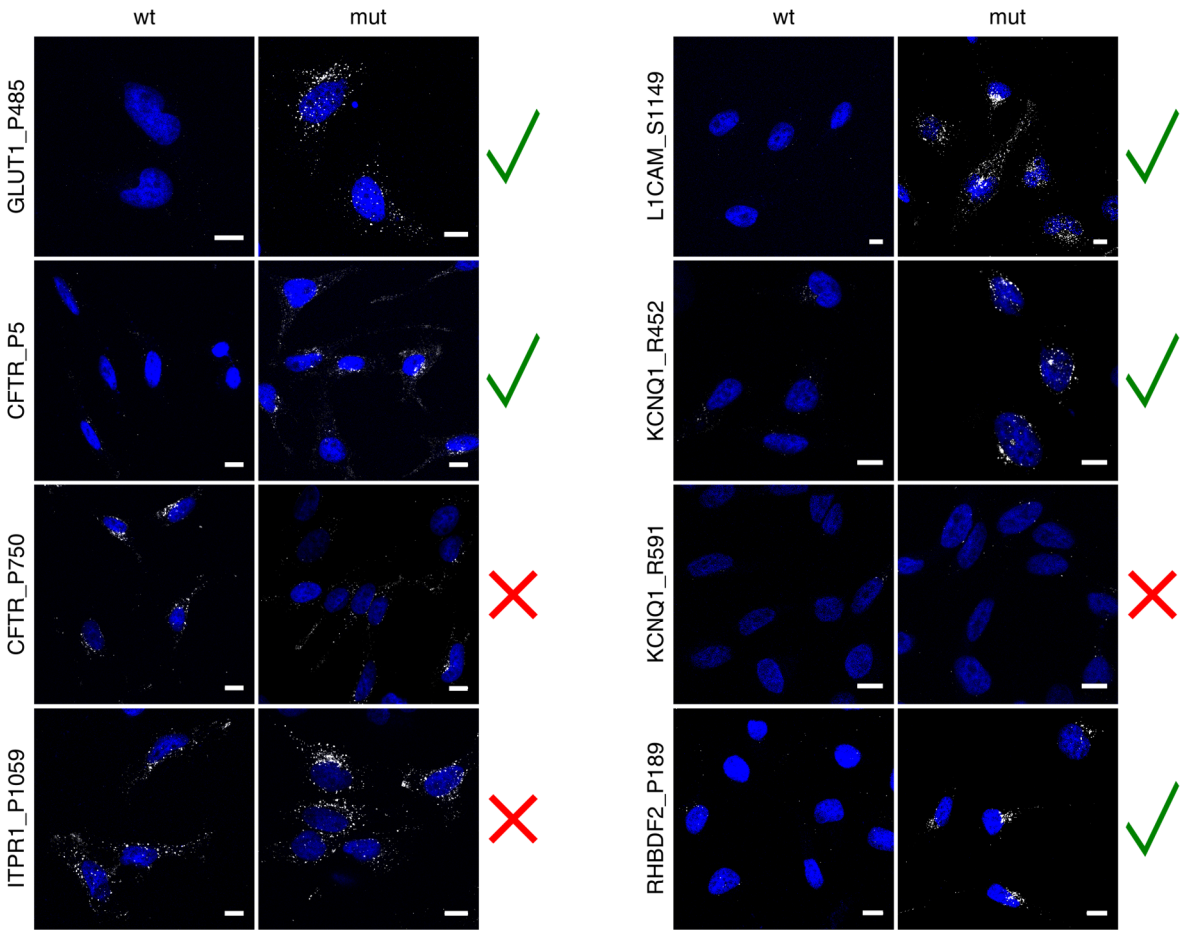


Figure 21 Mutation-induced gains of dileucine motifs are a recurrent cause of disease. Antibody feeding indicates that four out of seven tested mutations with a gain of dileucine motif lead to a gain in endocytosis. Fluorescence signal comes from the internalized antibody. Surface-exposed anti-TAC antigen was blocked prior to permeabilization. Scale bars: 10 µm.

Discussion

The work presented here is, to the best of my knowledge, the first systematic experimental analysis of how mutations in disordered regions affect protein-protein interactions. It includes a scalable proteomic screen that can (i) capture known interactions, (ii) detect how mutations in SLiMs affect binding of cognate domains and (iii) provide mechanistic insights into pathogenesis.

Potentials and limitations of the peptide screen

The two main advantages of the peptide-based interaction setup are its scalability, by using synthetic peptides, and specificity, by employing two quantitative filters. The importance of quantitative filters is illustrated by the control peptide. It is supposed to bind SH3 domain-containing proteins and application of the specificity filter got rid of non-SH3 containing proteins from otherwise SILAC differential interaction partners. The proteins that were discarded due to that threshold could be indirect interactors, at least they appear with low to medium confidence as interaction partners of remaining SH3 domain-containing proteins in the STRING database (Szklarczyk et al. 2015).

The approach presented here provides a powerful tool to study SLiM domain interactions that are typically weak or transient, for example, interactions between adaptors and sorting signals that have K_ds in the micromolar range (Owen, Collins, and Evans 2004). It is also a helpful approach for cytosolic domains of large transmembrane proteins that are otherwise intrinsically difficult to study. The screen is able to capture interactions that are often overlooked in classical proteomics, where protein-protein interaction discovery is often based on tandem affinity purification tag (TAP-tag) or yeast two-hybrid (Y2H) experiments. Such methods are typically biased toward stable interactions (Diella et al. 2008; Landry et al. 2013; B. A. Liu, Engelmann, and Nash 2012; Neduva and Russell 2006; Tompa et al. 2014). In fact, GLUT1_P485L has been part of a large scale Y2H study (Sahni et al. 2015) and has not been found to gain interaction partners. Even if we did detect clathrin-related proteins in our peptide-based interaction screen with shotgun analysis, we initially did not detect the APs responsible for the interaction. We were able

to detect them only with a more targeted approach. So the question remains how many other weak interactions we lost.

Between gain and loss of function, the more prevalent concept in PPI studies in disease is arguably loss-of-function and loss-of-interaction studies. However, additional to the discovery of the loss of physiological protein-protein interactions, we have found a significant fraction of gains of interaction. Our approach is open to considering both directions of changes in PPIs. Reports of gain in interaction are so far rare in the literature.

We made use of mutations stored in the OMIM database which contains many rare diseases. These are often understudied, but can in some way be considered "experiments of nature". By studying changes in protein behavior, and especially protein-protein interactions, this thesis can help understanding disease mechanisms but potentially also basic functions of proteins.

Even if the here employed peptide screen has many advantages, it is important to consider also the weaknesses of the approach.

Every screen is only as good as its input. In the case presented here, the screen depends on the disease databases used to extract disease-causing mutations. Even though Humsavar is thought to be a gold standard in disease databases and to contain only disease-causing mutations, the picture might become more dispersed when diving into the literature. One of the dileucineopathy candidates in the original screen, CACNA1H_P648L, seems to be disease-causing, but then the authors of the original publication state that the patient inherited the mutation from one parent, but the parents show no symptoms. The authors provide a valid explanation for this, which, however, remains to be proven (Chen et al. 2003). Heron and colleagues even challenge the findings of Chen et al. (Heron et al. 2004). On the other hand, in support of the pathogenicity, another study found several mutations in CACNA1H responsible for epileptogenesis (Khosravani et al. 2004).

In another incidence, L1CAM_S1194L has disappeared from the Humsavar variant database and is now listed as a natural variant in Uniprot, although still connected to disease and also linked to Clinvar, where it is categorized as pathogenic.

Inconsistencies like these highlight the difficulty of causal disease descriptions. However, the peptide array remains powerful in detecting changes in peptide-protein interactions. If these changes can be tolerated by the cell or if they lead to disease remains to be shown for every single case by more in-depth follow-up experiments.

Additionally, the approach presented here is clearly an *in vitro* technique. The usage of cell lysate disperses spatial resolution and the local concentration of peptide or ligand is not necessarily physiological. Additionally, the chosen cell line might not contain all involved interaction partners or might bring together proteins that actually never meet in the cell. Also taking IDRs out of their protein context can in some cases change their behavior.

Even if the screen harbors high potential for up-scalability a current bottleneck is the time needed for punching out single spots and the machine time needed to measure all samples in the mass spectrometer. A grid to punch out several spots at once could alleviate the first issue, while parallel injection could decrease run time up to 2-fold and TMT labeling would allow more samples to be analyzed in parallel, which could decrease measurement time up to 5-fold. Advances in machine performance will also decrease the runtime in the future. Very recently, it has been shown that 1000 proteins can be identified in a 5-minute gradient (Meier et al. 2018), which would be sufficient for a low complexity pull-down sample and relates to an almost 10-fold decrease in measurement time.

Potential applicability of the peptide screen

SLiMs are relatively simple sequence patterns and may arise convergently and be lost rapidly in a short timescale (Davey, Travé, and Gibson 2011; Davey et al. 2012; Holt et al. 2009). This easy emergence might lead to ‘‘noisy’’ interactions without functional consequences (Landry et al. 2013; Levy, Landry, and Michnick 2009; Tompa et al. 2014), thus raising an important question: how many of the binding motifs, PTM sites, and interactions mediated by these modules are functionally relevant? Since the peptide screen presented here employs disease mutations it allows studying which motifs are functionally relevant. One point that was not considered in the current analysis, is that IDRs

and SLiMs often involve PTMs and mutations often affect modification sites (Radivojac et al. 2008; Narayan, Bader, and Reimand 2016). With the current set-up, PTMs on peptides could be studied in a controlled manner, circumventing the issue of modifications by endogenous proteins, which might occur in the case of overexpression studies. A triple SILAC approach could be used to compare interaction partners of wt-unmodified, wt-modified and mutated peptides. In this way, it could be studied how PTMs regulate dynamic interactions of the interactome by creating a new binding interface, modulate the affinity of existing binding surfaces either positively or negatively, and how mutations change either of those functions. With probably more than a million PTM instances in the human proteome (Tompa et al. 2014), this leaves a lot of room for further studies.

Given that no interaction partner has been identified for ~75% of structural domain families there is a huge knowledge gap to be filled (Stein, Mosca, and Aloy 2011; Tompa et al. 2014). SLiMs are defined by such small information content that predicting them bioinformatically results in many false positives. Consequently, there are still rather few examples of bioinformatics discovery and subsequent experimental validation (Gibson et al. 2015). Large scale screens provide the opportunity to compare peptide sequences to protein domains in interacting proteins and hence the approach presented here could be used for motif discovery. A screen could be designed especially for that purpose and harbor many slight variations of the same sequence that could be tested for binding preferences. Here, TMT labeling could be of use again and provide the possibility to directly compare up to ten variations of similar sequences.

In this project, the screen has been applied to neurological diseases but it is not restricted to that and can in the future be applied to other diseases or even non-pathogenic polymorphisms.

Is the choice of control peptides representative for the screen?

Gibson and colleagues note that "protein-protein interactions they [SLiMs] mediate often depend on other factors, such as the PTM status of the motif and

the abundance of the interacting protein, both of which are not generally accounted for in experiments identifying interactions (Landry et al. 2013)" and that "this may not be an issue for certain protein domains, such as SH3 (Tonikian et al. 2009) and PDZ (Belotti et al. 2013), as the interactions they mediate are rather stable and are much better represented in HTSs [high-throughput studies]" (Tompa et al. 2014). This leads to the question if the choice of a poly-proline containing peptide really reflects the whole spectrum of interaction. It undoubtedly shows that the peptide array can, in general, detect low-affinity interactions but more controls might be needed to study the applicability under different settings and under the range of different affinities.

Open question on GLUT1_P485L

Even if we were able to show that GLUT1_P485L mislocalized when compared to the wild-type form and the phenotype is partially rescued by AP-2 knockdown, the deeper we dig the more open questions come up. For example, GLUT1_P485L mislocalizes in HEK 293 cells and also in patient-derived iPSCs. The apparent localization, though, is slightly different and colocalization analysis shows that the mutant resides in different compartments. GLUT1_P485L also colocalizes with endocytosed transferrin in HEK 293 but not in iPSCs. Consistently, we were able to rescue GLUT1_P485L localization via AP-2 knockdown in HEK 293 cells but not in iPSCs. It might be that due to different cellular conditions HEK and iPSC traffic wild-type GLUT1 differently. Since GLUT1_P485L carries only a single point mutation, it is likely that the characteristics of wt GLUT1 trafficking are maintained, while the newly gained dileucine motif adds clathrin-mediated trafficking. This would provide different access points for different APs at different locations in the cell. Wild-type GLUT1 is trafficked in a clathrin-independent manner (Eyster et al. 2009; Maldonado-Báez, Williamson, and Donaldson 2013). Although there is an exception when GLUT1 localization is controlled by thioredoxin-interacting protein (TXNIP). TXNIP itself contains a dileucine motif and binds to GLUT1 at the plasma membrane, where is recruits AP2 and clathrin to lead to endocytosis of GLUT1 (Wu et al. 2013).

At this point, it is difficult to say which cellular system provides more valid information about the actual disease mechanism, since neither of the cell lines accurately reflects endothelial cells of the blood-brain barrier. To continue the study in a more physiological context, endothelial cells of the blood-brain barrier could be differentiated from iPSCs (Lippmann et al. 2012) and used to study GLUT1_P485L trafficking. The closest approach, however, to studies in patients are animal disease models.

GLUT1_P485L *in vivo*

Studies of GLUT1 at the blood-brain barrier (BBB) *in vivo* present their very own challenges. The endothelial cell thickness ranges from about 0.2 µm away from the nucleus to about 0.9 µm in the vicinity of the nucleus (in rat brain capillaries) (Wong et al. 2013), revealing that due to its resolution limit confocal microscopy is infeasible to study GLUT1 localization in this system (Cox and Sheppard 2004). It is hence impossible to detect accumulations of GLUT1 in the cytosol or generally to distinguish the two different layers of the plasma membrane by confocal microscopy. We were, however, able to detect differences in GLUT1 levels in brain sections of mouse embryos with a visible decrease from wild-type over heterozygous to homozygous mutants. We employed super-resolution STED microscopy to be able to differentiate between the membrane layers of endothelial cells of the BBB. STED analysis confirmed downregulation of GLUT1 mutant levels, although it appeared to be significant only for the homozygous mutant. Anyways, we were not able to detect intracellular GLUT1. It might be that accumulations escaped our detection because of at least two different issues: (i) in comparison to cells in culture that grow flat on a surface and can easily be scanned entirely via z-stacks, endothelial cells in the brain span in different planes and it is much more challenging to image a cell in its entirety. Hence we might have been looking in the wrong planes. (ii) To detect a cell cut in the plane were accumulations might have formed, a microscopy technique that allows for scanning of the sample by eye is needed. This would be possible with confocal microscopy, which lacks sufficient resolution, and proved difficult with STED, where long wavelength emitting dyes are preferred that are not visible for the human eye. It is hence still possible that intracellular accumulations simply escaped our detection.

Interestingly, the apparent decrease in GLUT1 levels was not so apparent in adult mice, where there was no visible difference between wt and +/P485L mice with confocal and STED microscopy. P485L/P485L mice were not viable.

In the future, it will be interesting to see if GLUT1 levels are decreased in the whole brain, for example by Western blot or mass spectrometry. If they are not decreased that could mean that accumulations are present in the cytoplasm, whereas a general decrease could mean that the P485L mutation ultimately leads to degradation of GLUT1.

To validate the possibility to rescue the GLUT1_P485L phenotype. Mutant mice could be crossed with AP-2 deficient mice. The AP-2 knockout in mice is embryonically lethal (Mitsunari et al. 2005). However, a conditional AP-2 knockout (Kononenko et al. 2014) could be used to knock out AP-2 selectively in endothelial cells of the BBB and potentially rescue the phenotype. This model could also be used to study potential differences in GLUT1 trafficking at different stages in development.

Functionality of GLUT1 mutant

One important factor for the potential treatment of GLUT1 deficiency syndrome by relocalization of GLUT1_P485L to the plasma membrane is the actual functionality of the mutant. In a different study, glucose uptake had been tested in erythrocytes of a GLUT1_P485L patient. The K_M was found to be similar to the wild-type value but V_{max} was at roughly half of the wild-type capacity (Pascual et al. 2008). Estimating changes in K_M is raised difficult by factors such as subunit cooperativity arising from transporter polymerization (Coderre et al. 1995). However, at this point, it is unclear if endocytosis might play a role in that system, too. To rule out any influence the mutant should be tested, for example, in *Xenopus* oocytes as for other mutants in (Pascual et al. 2008) or as in (E. E. Lee et al. 2015).

In the study presented here, the rescue of GLUT1_P485L localization significantly increased the level of glucose uptake in a stable HEK 293 cell line. With the current set-up, however, no conclusions about the comparison of wild-type and mutant GLUT1 can be made. The stable cell line expressed mutant

GLUT1 generally in lower amounts than the wild-type, which makes a comparison inadequate.

Gain of dileucine motifs as a general disease mechanism

The so-called dileucine motif [D/E]…L[L/I] in cytosolic tails of transmembrane proteins can recruit different adaptor proteins, depending on its previous cellular location. The motif is not very complex and can easily emerge by chance. Additionally, even a single LL can be functional (Kozik et al. 2010).

In our peptide screen as well as in the bioinformatics analysis, dileucine motifs mostly emerged by mutation of proline to leucine. In our opinion this is due to two reasons. We were focusing on disordered regions and proline is the most disorder-promoting amino acid (Theillet et al. 2013). Also, proline codons can mutate into leucine codons by changes of a single nucleotide. Mutation of cytosine to uracil changes a proline C**C**[AUCG] to a leucine C**U**[AUGC]. Furthermore, the spontaneous deamination that converts cytosine to uracil or even worse 5-methylcytosine to thymine (Duncan and Miller 1980), is one of the most common DNA mutations.

Gain of functional dileucine motifs can have different effects on proteins. Either, as in GLUT1_P485L, which is not normally involved in clathrin-mediated trafficking, it adds a completely new aspect to protein transport. In other cases, a trafficking motif can also emerge where other such motifs are already present and functional, and only disturb the finely balanced trafficking. As for example in CFTR, where a gain of a tyrosine-based endocytosis motif enhances trafficking kinetics and leads to cystic fibrosis (Silvis et al. 2003).

Two of the proteins that we identified as candidates for dileucineopathies are mainly located at the ER. This raises the question about the functionality of the motif since the ER is not a target of clathrin-dependent trafficking. In general, trafficking of proteins is definitely not trivial and different variants (isoforms, modifications etc.) might be handled differently. The two proteins that mainly localize to the ER are RHBDF2 and ITPR1. Both of them, however, have been found also at other sites in the cell. RHBDF2 for example has also been shown to localize at the plasma membrane (Blaydon et al. 2012). ITPR1 was shown to

localize not only to the ER but also to the nuclear envelope, portions of the Golgi (Ross et al. 1989; Foskett, White, et al. 2007) and also to the plasma membrane (Tanimura, Tojyo, and Turner 2000; Dellis et al. 2006). The subcellular localization seems to be at least partially cell type dependent. The mutations might, therefore, affect the population of proteins that is not residing at the ER and thus cause disease. Of course, more detailed follow-up experiments are required to assess the impact of these mutations. Interestingly, a number of human neurodegenerative disorders such as Huntington's disease, SCA2, and SCA3 have recently been related to dysregulated ITPR1 function due to abnormal protein interactions (Bezprozvanny 2011; Foskett, Kevin Foskett, et al. 2007; Higo et al. 2010; Schorge et al. 2010). Of note, even if a dileucine is gained in a cytosolic tail of a transmembrane protein, this does not automatically mean it will be a functional trafficking motif.

There are several possibilities that might explain why we did not observe increased uptake for all dileucine candidates in the antibody feeding experiment. First, given the odds ratio of ~4 (Figure 20 B), it is not expected that all pathogenic dileucine motif gains cause disease by altering clathrin-dependent trafficking. Second, two of the mutation sites we tested (CFTR_P750, ITPR1_P1059) already showed uptake in the wild-type, which makes it difficult to assess the impact of the additionally gained dileucine motif. The CFTR_P750 peptide contains the sequence: E..E.IL, which closely resembles the dileucine motif and might explain endocytosis of the wild-type variant. In the disease context, these two mutations might still affect the kinetics of clathrin-dependent trafficking, as seen for a tyrosine-based internalization motif in CFTR (Silvis et al. 2003). A more quantitative method would be needed to assess differences in endocytosis rates.

Understanding disease mechanisms makes diseases druggable. Especially in this case, where relocalization of potentially functional proteins might lead to full recovery of protein function.

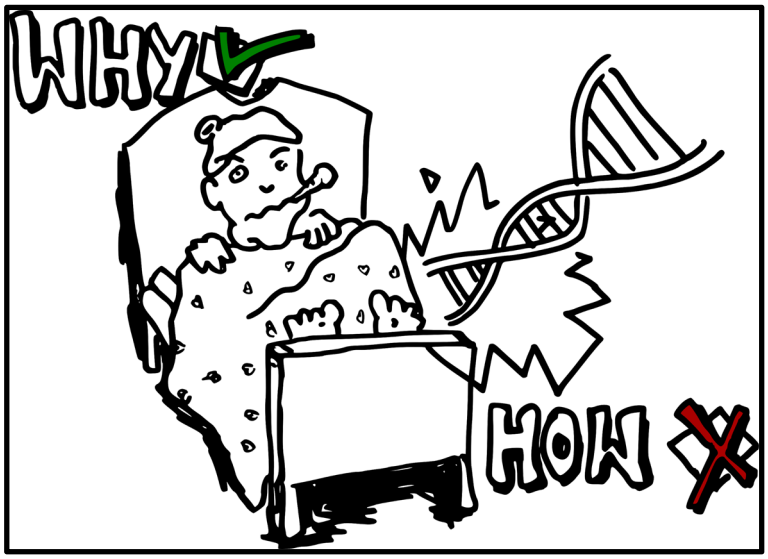
Conclusion

"It is becoming increasingly clear that we can often better understand the mechanistic impact of genomic variation on the functional proteome through using molecular interactions as a basic cellular functional unit, rather than functions of proteins as a whole." Woodsmith and Stelzl 2017

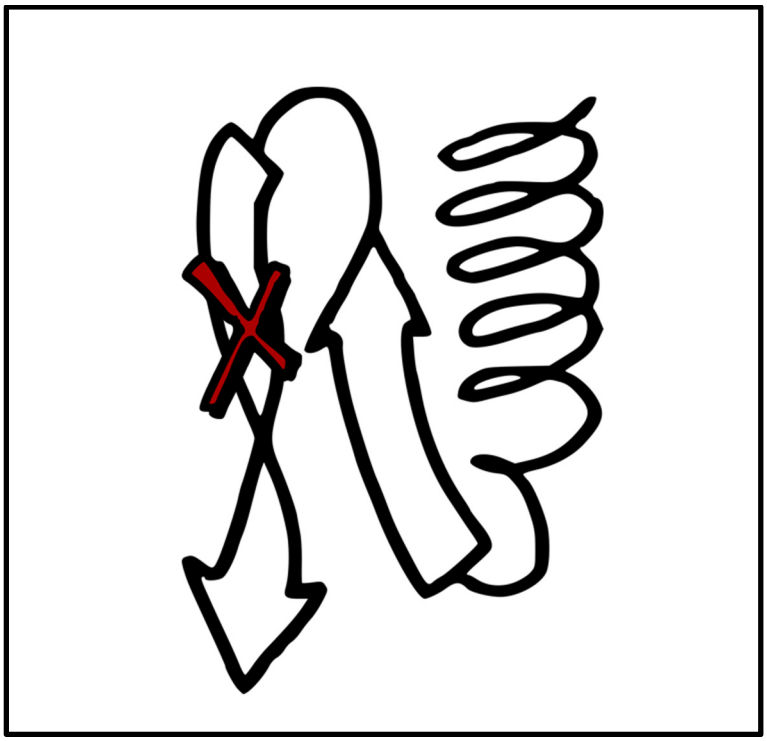
In this thesis, a screen on disease-causing mutations has revealed that the emergence of a leucine in cytoplasmic tails of transmembrane proteins can lead to the gain of a dileucine motif which can be recognized by the clathrin-mediated trafficking machinery and can lead to mislocalization of the protein.

The broad applicability and potential of the peptide-based interaction screen will hopefully continue to help to unravel disease mechanisms also in the future.

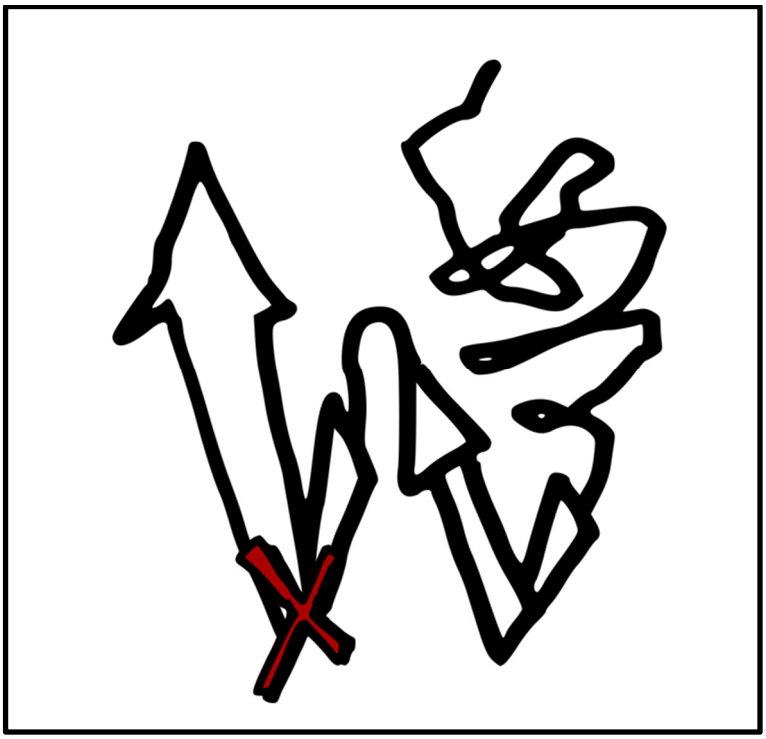
Graphic novel



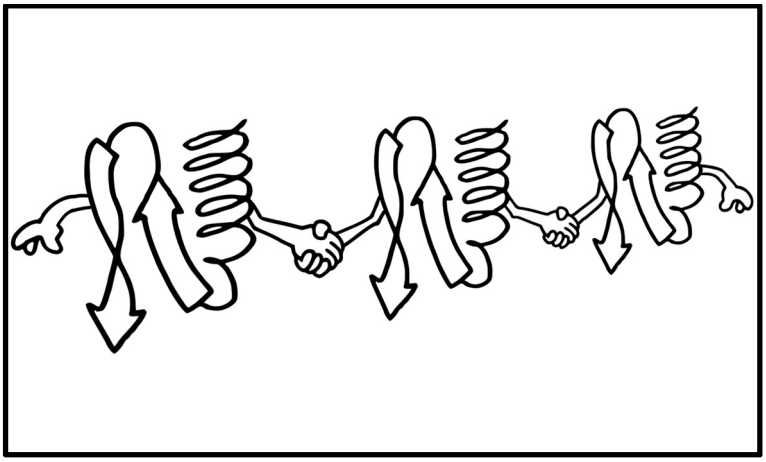
For more than hundred thousand known, pathogenic mutations the actual disease mechanism is not known.



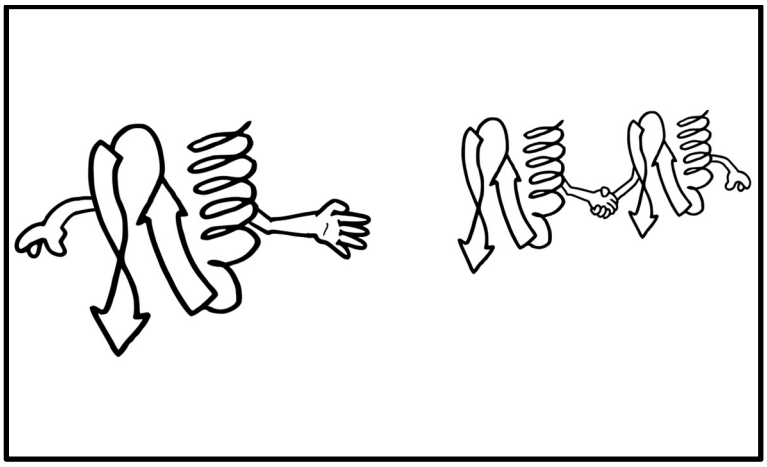
Mutations in structured regions of proteins most likely lead to misfolding and loss of function...



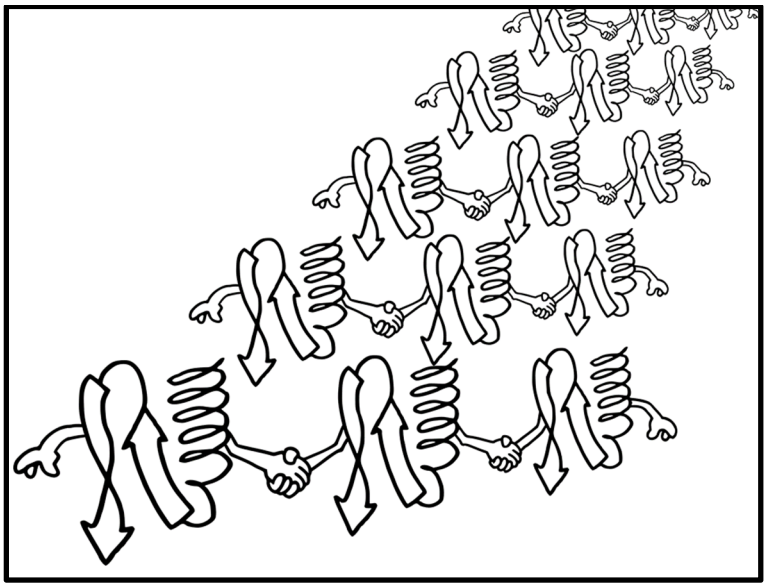
...but what about mutations that fall into the disordered segments of the protein?



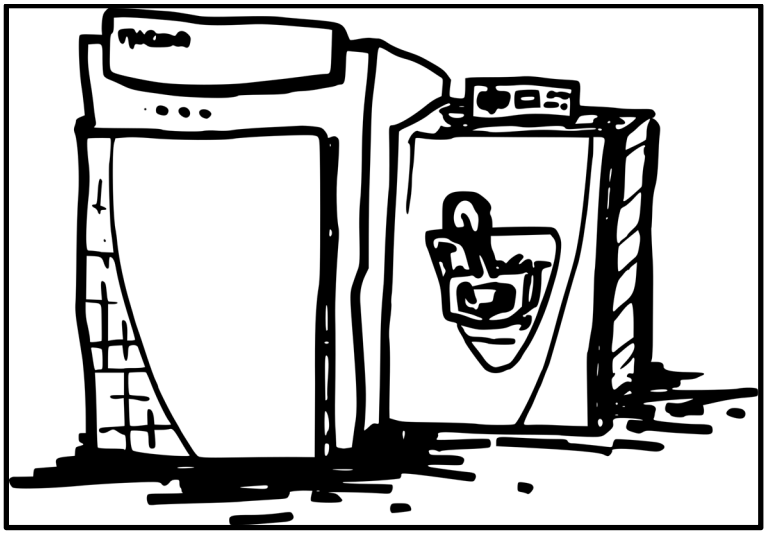
Proteins interact to fulfill their tasks in the cells. Interactions can be mediated by disordered regions.



Mutations can lead to disruption of these interactions and might be an explanation for disease.

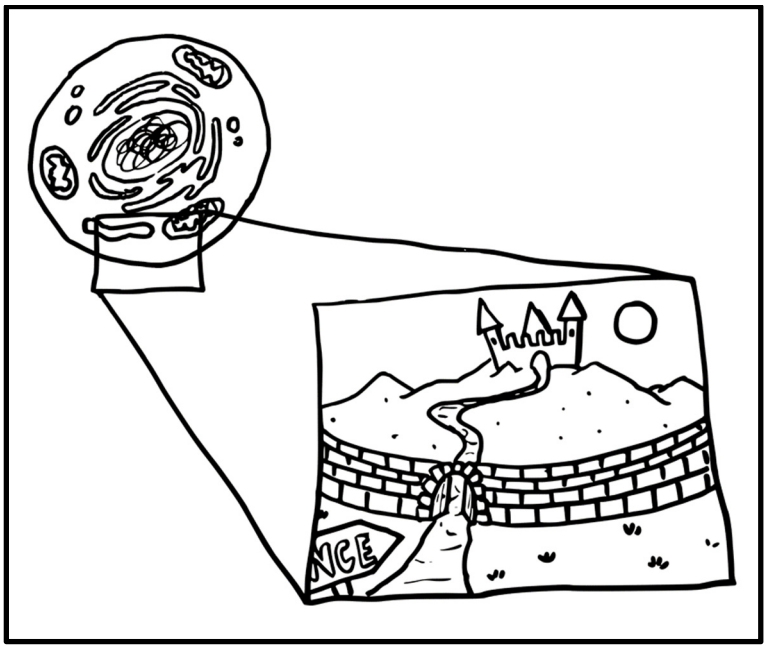


There are so many proteins and interactions in the cell that we need a powerful method to study them.

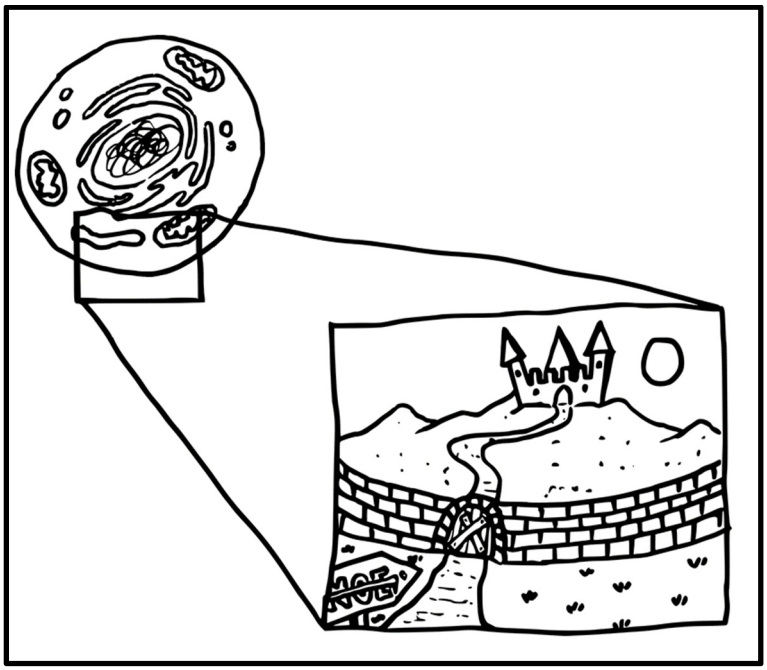


Mass spectrometry based proteomics is the perfect tool!

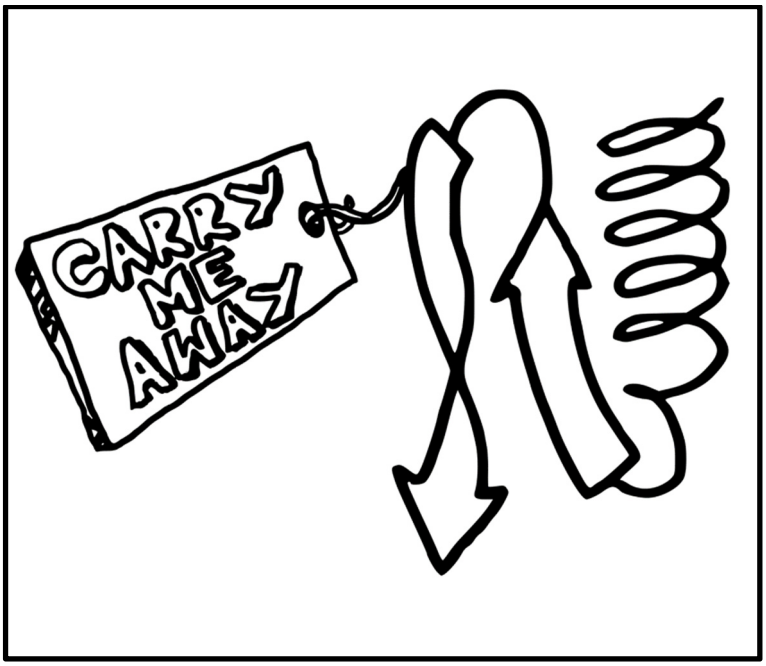
It helped us to find the needle in the haystack!



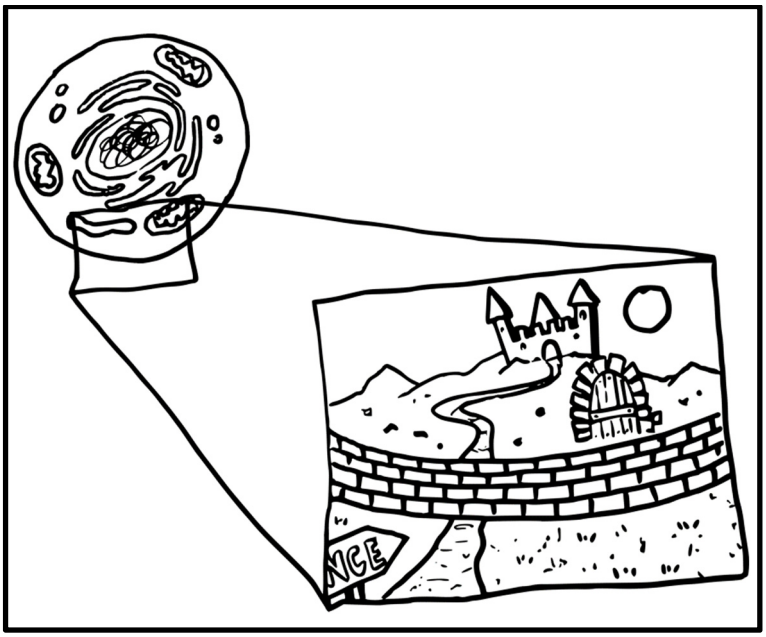
Some proteins function like gates to permit adequate supply of the cell with nutrients.



In many diseases, gates are blocked, which leads to disease.



Mutations don't always destroy; they can also create new things -- like tags that tell other proteins to transport the mutated protein to different places.



If this happens to a gate, it might still be fully functional but sitting in the wrong place. This does not permit nutrient transport anymore!

©ANZO MEYER

Supplementary information

Abbreviations

AA	Amino acid
ABC	Ammonium bicarbonate
BBB	Blood-brain barrier
BirA	Biotin-protein ligase
CACNA1H	Calcium Voltage-Gated Channel Subunit Alpha1 H
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/CRISPR associated protein 9
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle Medium
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N’,N’-tetraacetic acid)
ER	Endoplasmic reticulum
FBS	Fetal bovine serum
FMP	Leibniz-Forschungsinstitut für Molekulare Pharmakologie
GLUT1	Glucose transporter 1
HEK (293)	Human embryonic kidney cells 293
HRP	Horseradish peroxidase
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
iPSC	Inducible pluripotent stem cell
ITPR1	Inositol 1,4,5-trisphosphate receptor type 1
MDC	Max Delbrück Center for Molecular Medicine in the Helmholtz Association
mRNA	Messenger RNA
mut	Mutant
PCR	Polymerase chain reaction
PTM	Post-translational modification
SDS	Sodium dodecyl sulfate
siRNA	Small (or short) interfering RNA
SLC2A1	Solute carrier 2A1
wt	Wild-type

Supplementary tables

Table 4 All peptide candidates included in the peptide-based interaction screen

Disease name	Gene name	Peptide ID	Peptidesequence (wt)
Achalasia-addisonianism-alacrima syndrome (AAAS) [MIM:231550]	AAAS	Q9NRG9_Q15K	PPPPPRGQVTLYEHN
Aicardi-Goutieres syndrome 6 (AGS6) [MIM:615010]	ADAR	P55265_G1007R	LRTKVENGEGTIPVE
Aicardi-Goutieres syndrome 6 (AGS6) [MIM:615010]	ADAR	P55265_K999N	FENPKQGKLRTKVEN
Alternating hemiplegia of childhood 2 (AHC2) [MIM:614820]	ATP1A3	P13637_D220N	EPQTRSPDCTHDNPL
Episodic ataxia 2 (EA2) [MIM:108500]	CACNA1A	O00555_R2135C	VLGPKARRLDDYSLE
Episodic ataxia 2 (EA2) [MIM:108500]	CACNA1A	O00555_P897R	AELSREGPYGRESDH
Epilepsy, childhood absence 6 (ECA6) [MIM:611942]	CACNA1H	O95180_A748V	DPTRPPRATDTPGPG
Epilepsy, childhood absence 6 (ECA6) [MIM:611942]	CACNA1H	O95180_P648L	PGTGGHGPLSLNSPD
Epilepsy, childhood absence 6 (ECA6) [MIM:611942]	CACNA1H	O95180_G499S	PSAVQGQGPGHRQRR
Mental retardation and microcephaly with pontine and cerebellar hypoplasia (MICPCH) [MIM:300749]	CASK	O14936_P396S	KINTKSSPQIRNPPS
Epilepsy, idiopathic generalized 8 (EIG8) [MIM:612899]	CASR	P41180_A988V	EPQKNAMAHRNSTHQ
Epilepsy, idiopathic generalized 8 (EIG8) [MIM:612899]	CASR	P41180_R898Q	RSNVSRKRSSSLGGS
Joubert syndrome 9 (JBTS9) [MIM:612285]	CC2D2A	Q9P2K1_P1122S	TTAEGPNPSWNEELE
Joubert syndrome 9 (JBTS9) [MIM:612285]	CC2D2A	Q9P2K1_T1114M	FQRTVCHTTTAEGPN
Epileptic encephalopathy, early infantile, 2 (EIEE2) [MIM:300672]	CDKL5	O76039_V718M	RVGSFYRVPSPRPDN
Epileptic encephalopathy, early infantile, 2 (EIEE2) [MIM:300672]	CDKL5	O76039_N399T	STSKDLTNNNIPHLL
Joubert syndrome 5 (JBTS5) [MIM:610188]	CEP290	O15078_W7C	MPPNINWKEIMKVDP
Frontotemporal dementia and/or amyotrophic lateral sclerosis 2 (FTDALS2) [MIM:615911]	CHCHD10	Q8WYQ3_P34S	PPPSAAAPAPAPSGQ
Amyotrophic lateral sclerosis 17 (ALS17) [MIM:614696]	CHMP2B	Q9UQN3_T104N	KMAGAMSTTAKTMQA
Amyotrophic lateral sclerosis 17 (ALS17) [MIM:614696]	CHMP2B	Q9UQN3_Q206H	SDEEIERQLKALGVD
Frontotemporal dementia, chromosome 3-linked (FTD3) [MIM:600795]	CHMP2B	Q9UQN3_D148Y	MINDTLDDIFDGSDD
Ceroid lipofuscinosis, neuronal, 4A (CLN4A) [MIM:204300]	CLN6	Q9NWW5_R6T	MEATRRRQHLGATGG

Ceroid lipofuscinosis, neuronal, 8 (CLN8) [MIM:600143]	CLN8	Q9UBY8_E269V	DWNFAQPEAKSRPEG
Hypomyelination with brainstem and spinal cord involvement and leg spasticity (HBSL) [MIM:615281]	DARS	P14868_R494G	RQTSMFPRDPKRLTP
Leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation (LBSL) [MIM:611105]	DARS2	Q6PI48_Y629C	PPEELKPYHIRVSKP
Epilepsy, familial focal, with variable foci (FFEVF) [MIM:604364]	DEPDC5	O75140_S1073R	SSAQSAESSSVAMTP
Epilepsy, familial focal, with variable foci (FFEVF) [MIM:604364]	DEPDC5	O75140_S1162G	STNSSDSSSQQLVAS
Epilepsy, familial focal, with variable foci (FFEVF) [MIM:604364]	DEPDC5	O75140_A452V	SPKESENALPIQVDY
Dyskeratosis congenita, X-linked (DKCX) [MIM:305000]	DKC1	O60832_G402R	QGLLDKHGKPTDSTP
Dyskeratosis congenita, X-linked (DKCX) [MIM:305000]	DKC1	O60832_P409L	GKPTDSTPATWKQEY
Combined oxidative phosphorylation deficiency 12 (COXPD12) [MIM:614924]	EARS2	Q5JPH6_E96K	ENIEDMLEWAGIPPD
Combined oxidative phosphorylation deficiency 12 (COXPD12) [MIM:614924]	EARS2	Q5JPH6_R108W	PPDESPRRGGPAGPY
Combined oxidative phosphorylation deficiency 12 (COXPD12) [MIM:614924]	EARS2	Q5JPH6_R168G	PRYDNRCRNMSQEQV
Neuropathy, congenital hypomyelinating or amyelinating (CHN) [MIM:605253]	EGR2	P11161_I268N	PLTPLSTIRNFTLGG
Leukodystrophy with vanishing white matter (VWM) [MIM:603896]	EIF2B5	Q13144_S447L	NITLPEGSVISLHPP
Parkinson disease 18 (PARK18) [MIM:614251]	EIF4G1	Q04637_R1205H	PSQPEGLRKAASLTE
Xeroderma pigmentosum complementation group F (XP-F) [MIM:278760]	ERCC4	Q92889_G513R	EEGDVEEGYRREISS
Cockayne syndrome B (CSB) [MIM:133540]	ERCC6	Q03468_P1042L	HLKRRIQPAFGADHD
Pontocerebellar hypoplasia 1C (PCH1C) [MIM:616081]	EXOSC8	Q96B26_S272T	VKKLMDEVIKSMKPK
Amyotrophic lateral sclerosis 6 (ALS6) [MIM:608030]	FUS	P35637_R521C	DSRGEHRQDRRERPY
Amyotrophic lateral sclerosis 6 (ALS6) [MIM:608030]	FUS	P35637_G507D	GGDRGGFGPGKMDSR
Tremor, hereditary essential 4 (ETM4) [MIM:614782]	FUS	P35637_R216C	GQQDRGGRGRGGSGG
Cerebral palsy, spastic quadriplegic 1 (CPSQ1) [MIM:603513]	GAD1	Q99259_S12C	TPSSSATSSNAGADP
Leukodystrophy, globoid cell (GLD) [MIM:245200]	GALC	P54803_Y490N	SQPFPSTYKDDFNVD
Gaucher disease (GD) [MIM:230800]	GBA	P04062_R87W	ESTRSGRRMELSMGP
Glutaric aciduria 1 (GA1) [MIM:231670]	GCDH	Q92947_M191T	SGSDPSSMETRAHYN

Dystonia, dopa-responsive (DRD) [MIM:128230]	GCH1	P30793_P23L	ARCSNGFPERDPPRP
Dystonia, dopa-responsive (DRD) [MIM:128230]	GCH1	P30793_R249S	EDPKTREEFLTLIRS
Parkinson disease 11 (PARK11) [MIM:607688]	GIGYF2	Q6Y7W6_T112A	TGRGGGGTVVGAPRG
Parkinson disease 11 (PARK11) [MIM:607688]	GIGYF2	Q6Y7W6_V1242I	NHSTLHSVFQTNQSN
Parkinson disease 11 (PARK11) [MIM:607688]	GIGYF2	Q6Y7W6_D606E	PPHMGELDQERLTRQ
Charcot-Marie-Tooth disease, X-linked dominant, 1 (CMTX1) [MIM:302800]	GJB1	P08034_R264C	GSLKDILRRSPGTGA
Charcot-Marie-Tooth disease, X-linked dominant, 1 (CMTX1) [MIM:302800]	GJB1	P08034_C280G	TGAGLAEKSDRCSAC
Charcot-Marie-Tooth disease, X-linked dominant, 1 (CMTX1) [MIM:302800]	GJB1	P08034_R230C	RRSNPPSRKGSGFGH
Epilepsy, focal, with speech disorder and with or without mental retardation (FESD) [MIM:245570]	GRIN2A	Q12879_D1251N	NLYDIDEDQMLQETG
Leukoencephalopathy, megalencephalic, with subcortical cysts, 2B (MLC2B) [MIM:613926]	HEPACAM	Q14CZ8_R288C	YMDQNDDRLKPEADT
Lesch-Nyhan syndrome (LNS) [MIM:300322]	HPRT1	P00492_V8G	MATRSPGVVISDDEP
Charcot-Marie-Tooth disease 2F (CMT2F) [MIM:606595]	HSPB1	P04792_T164A	LSPEGTLTVEAPMPK
Neuronopathy, distal hereditary motor, 2B (HMN2B) [MIM:608634]	HSPB1	P04792_T151I	LPPGVDPTQVSSSLS
Short-rib thoracic dysplasia 9 with or without polydactyly (SRTD9) [MIM:266920]	IFT140	Q96RY7_E664K	EPRLFVCEAVQETPR
Neuronopathy, distal hereditary motor, 6 (HMN6) [MIM:604320]	IGHMBP2	P38935_D974E	AQLQRRLDKKLSELS
Incontinentia pigmenti (IP) [MIM:308300]	IKBKG	Q9Y6K9_Q183H	AASEQARQLESEREA
Spinocerebellar ataxia 15 (SCA15) [MIM:606658]	ITPR1	Q14643_P1059L	GGSEENTPLDLDDHG
Hypogonadotropic hypogonadism 1 with or without anosmia (HH1) [MIM:308700]	KAL1	P23352_H672R	LKHRHPHHYKPSPER
Pallister-Hall syndrome (PHS) [MIM:146510]	KIF7	Q2M1P5_P632L	EEEEEEEPPRRTLHL
Hydrocephalus due to stenosis of the aqueduct of Sylvius (HSAS) [MIM:307000]	L1CAM	P32004_S1194L	AFGSSQPSLNGDIKP
Charcot-Marie-Tooth disease 1C (CMT1C) [MIM:601098]	LITAF	Q99732_T49M	GPTTGLVTGPDGKGM
Frontotemporal dementia (FTD) [MIM:600274]	MAPT	P10636_L583V	KIGSTENLKHQPGGG
Frontotemporal dementia (FTD) [MIM:600274]	MAPT	P10636_V654M	HKPGGGQVEVKSEKL
Progressive supranuclear palsy 1 (PSNP1) [MIM:601104]	MAPT	P10636_R5L	MAEPRQEFEVMEDHA
Amyotrophic lateral sclerosis 21 (ALS21) [MIM:606070]	MATR3	P43243_T622A	SQKTESSTEGKEQEE
Rett syndrome (RTT) [MIM:312750]	MECP2	P51608_P225R	GKLLVKMPFQTSPGG

Rett syndrome (RTT) [MIM:312750]	MECP2	P51608_G161V	DFDFTVTGRGSPSRR
Rett syndrome (RTT) [MIM:312750]	MECP2	P51608_E10Q	AGMLGLREEKSEDQD
Charcot-Marie-Tooth disease 2B2 (CMT2B2) [MIM:605589]	MED25	Q71SY5_A335V	LPPGPPGAPKPPPAS
Dejerine-Sottas syndrome (DSS) [MIM:145900]	MPZ	P25189_A221T	RQTPVLYAMLDHSRS
Neurodegeneration with brain iron accumulation 1 (NBIA1) [MIM:234200]	PANK2	Q9BZ23_E134G	GRLGAPMERHGRASA
Parkinson disease 2 (PARK2) [MIM:600116]	PARK2	O60260_A82E	RKGQEMNATGGDDPR
Parkinson disease 2 (PARK2) [MIM:600116]	PARK2	O60260_A92V	GDDPRNAAGGCEREP
Parkinson disease 6 (PARK6) [MIM:605909]	PINK1	Q9BXM7_P196L	GLLPGRGPGTSAPGE
Neurodegeneration with brain iron accumulation 2B (NBIA2B) [MIM:610217]	PLA2G6	O60733_R632W	PSDQLVWRAARSSGA
Boucher-Neuhauser syndrome (BNHS) [MIM:215470]	PNPLA6	Q8IY17_R1350W	EEEKSILRQRRCLPQ
Progressive external ophthalmoplegia with mitochondrial DNA deletions, autosomal dominant, 1 (PEOA1) [MIM:157640]	POLG	P54098_S511N	KKEPATASKLPIEGA
Mitochondrial DNA depletion syndrome 4B (MTDPS4B) [MIM:613662]	POLG	P54098_T251I	LIPLEVPTGASSPTQ
Sensory ataxic neuropathy dysarthria and ophthalmoparesis (SANDO) [MIM:607459]	POLG	P54098_G517V	ASKLPIEGAGAPGDP
Gerstmann-Straussler disease (GSD) [MIM:137440]	PRNP	P04156_P102L	THSQWNKPSKPKTNM
Noonan syndrome 1 (NS1) [MIM:163950]	PTPN11	Q06124_Y279C	ENKNKNRYKNILPFD
Noonan syndrome 1 (NS1) [MIM:163950]	PTPN11	Q06124_E139D	HGSFLVRESQSHPGD
Noonan syndrome 1 (NS1) [MIM:163950]	PTPN11	Q06124_L564F	TSGDQSPLPPCTPTP
Noonan syndrome 5 (NS5) [MIM:611553]	RAF1	P04049_R256S	GSLSQRQRSTSTPNV
Coffin-Lowry syndrome (CLS) [MIM:303600]	RPS6KA3	P51812_R729Q	GRSTLAQRRGIKKIT
Epileptic encephalopathy, early infantile, 6 (EIEE6) [MIM:607208]	SCN1A	P35498_L1514S	YYNAMKKLGSKKPQK
Generalized epilepsy with febrile seizures plus 2 (GEFS+2) [MIM:604403]	SCN1A	P35498_R27T	ESLAAIERRIAEEKA
Epileptic encephalopathy, early infantile, 6 (EIEE6) [MIM:607208]	SCN1A	P35498_F63L	EAGKNLPFIYGDIPP
Schinzel-Giedion midface retraction syndrome (SGMFS) [MIM:269150]	SETBP1	Q9Y6X0_G870S	ETIPSDSGIGTDNNS
Schinzel-Giedion midface retraction syndrome (SGMFS) [MIM:269150]	SETBP1	Q9Y6X0_I871T	TIPSDSGIGTDNNST
Schinzel-Giedion midface retraction syndrome (SGMFS) [MIM:269150]	SETBP1	Q9Y6X0_D868N	SEETIPSDSGIGTDN
Spinocerebellar ataxia, autosomal recessive, 1 (SCAR1) [MIM:606002]	SETX	Q7Z333_Q653K	SKEPMKVQDSVLIKA
Spinocerebellar ataxia, autosomal recessive, 1 (SCAR1) [MIM:606002]	SETX	Q7Z333_R1294C	LGLKKGPRKAYELSQ

GLUT1 deficiency syndrome 1 (GLUT1DS1) [MIM:606777]	SLC2A1	P11166_P485L	TPEELFHPLGADSQV
Niemann-Pick disease B (NPDB) [MIM:607616]	SMPD1	P17405_A196P	PSPPAPGAPVSRILF
Amyotrophic lateral sclerosis 1 (ALS1) [MIM:105400]	SOD1	P00441_D77Y	RKHGGPKDEERHVGD
Amyotrophic lateral sclerosis 1 (ALS1) [MIM:105400]	SOD1	P00441_I152T	NAGSRLACGVIGIAQ
Amyotrophic lateral sclerosis 1 (ALS1) [MIM:105400]	SOD1	P00441_G86R	ERHVGDLGNVTADKD
Spastic paraplegia 4, autosomal dominant (SPG4) [MIM:182601]	SPAST	Q9UBP0_P293L	PTTHKGTPKTNRTNK
Opitz GBBB syndrome 2 (GBBB2) [MIM:145410]	SPECC1L	Q69YQ0_T397P	EVSVACLTERIHQME
Microcephaly-capillary malformation syndrome (MICCAP) [MIM:614261]	STAMBP	O95630_R14P	VSLPPEDRVRALSQL
Spinocerebellar ataxia, autosomal recessive, 16 (SCAR16) [MIM:615768]	STUB1	Q9UNE7_E28K	EKSPSAQELKEQGNR
Isolated sulfite oxidase deficiency (ISOD) [MIM:272300]	SUOX	P51687_S427Y	IQELPVQSAITEPRD
Amyotrophic lateral sclerosis 10 (ALS10) [MIM:612069]	TARDBP	Q13148_G295S	GFGNSRGGGAGLGNN
Amyotrophic lateral sclerosis 10 (ALS10) [MIM:612069]	TARDBP	Q13148_G348C	ASQQNQSGPSGNNQN
Amyotrophic lateral sclerosis 10 (ALS10) [MIM:612069]	TARDBP	Q13148_G290A	FGNQGGFGNSRGGGA
Pitt-Hopkins syndrome (PTHS) [MIM:610954]	TCF4	P15884_R565W	KAEREKERRMANNAR
Pitt-Hopkins syndrome (PTHS) [MIM:610954]	TCF4	P15884_D535G	SSEDKKLDDDKKDIK
Pitt-Hopkins syndrome (PTHS) [MIM:610954]	TCF4	P15884_G358V	SPPSLSAGTAVWSRN
Hereditary motor and sensory neuropathy, proximal type (HMSNP) [MIM:604484]	TFG	Q92734_P285L	QQTGPQQPQQFQGYG
Dyskeratosis congenita, autosomal dominant, 3 (DKCA3) [MIM:613990]	TINF2	Q9BSI4_K280E	ASTRGGHKERPTVML
Spinocerebellar ataxia 21 (SCA21) [MIM:607454]	TMEM240	Q5SV17_P170L	KQKLYHNGHPSPRHL
Charcot-Marie-Tooth disease 2C (CMT2C) [MIM:606071]	TRPV4	Q9HBA0_R315W	PHKKADMRRQDSRGN
Mitochondrial DNA depletion syndrome 1, MNGIE type (MTDPS1) [MIM:603041]	TYMP	P19971_R44Q	PELIRMKRDGGRLSE
Spinal muscular atrophy X-linked 2 (SMAX2) [MIM:301830]	UBA1	P22314_S547G	NPHIRVTSHQNRVGP
Amyotrophic lateral sclerosis 15, with or without frontotemporal dementia (ALS15) [MIM:300857]	UBQLN2	Q9UHD9_A283T	QEPMLNAAQEQFGGN
Amyotrophic lateral sclerosis 15, with or without frontotemporal dementia (ALS15) [MIM:300857]	UBQLN2	Q9UHD9_P525S	PTGPAAPPGSTGSGG

Combined oxidative phosphorylation deficiency 20 (COXPD20) [MIM:615917]	VARS2	Q5ST30_A349T	LPGDVAVAVHPDDSR
Wolfram syndrome 1 (WFS1) [MIM:222300]	WFS1	O76024_A58V	GPGVRDAAAPAEPQA
Spinocerebellar ataxia, autosomal recessive, 12 (SCAR12) [MIM:614322]	WWOX	Q9NZC7_P47T	EKTQWEHPKTGKRKR
Wieacker-Wolf syndrome (WRWF) [MIM:314580]	ZC4H2	Q9NQZ6_R213W	CKAKSRSRNPKKPKR
Mowat-Wilson syndrome (MOWS) [MIM:235730]	ZEB2	O60315_Q1119R	YLQSITPQGYSDSEE

Table 5 Mutations in cytoplasmic regions of transmembrane proteins that lead to gain of dileucine motifs

dbSNP	Uniprot Accession	Gene name	Mutation	Flanking sequence	Disease
-	O95180	CACNA1H	P648L	PGTGGHGPLSLNSPD	Epilepsy, childhood absence 6 (ECA6) [MIM:611942]
rs60734921	O95180	CACNA1H	P618L	MNYPTILPSGVGSGK	Epilepsy, idiopathic generalized 6 (EIG6) [MIM:611942]
rs193922501	P13569	CFTR	P5L	MQRSPLEKASVV	Cystic fibrosis
rs140455771	P13569	CFTR	P750L	EQGEAILPRISVIST	Cystic fibrosis
-	Q14643	ITPR1	P1059L	GGSEENTPLDLDDH G	Spinocerebellar ataxia 15 (SCA15) [MIM:606658]
rs145229963	P51787	KCNQ1	R452L	CDPPEERRLDHFSVD	not provided
rs199472814	P51787	KCNQ1	R591L	GSNTIGARLNRVEDK	not provided
rs137852522	P32004	L1CAM	S1194L	AFGSSQPSLNGDIKP	Hydrocephalus due to stenosis of the aqueduct of Sylvius (HSAS) [MIM:307000]
rs79853121	P07949	RET	P1039L	GLSEEETPLVDCNNA	Congenital central hypoventilation syndrome (CCHS) [MIM:209880]
rs387907130	Q6PJF5	RHBDF2	P189L	KMPKIVDPLARGRAF	Tylosis with esophageal cancer (TOC) [MIM:148500]
-	P11166	SLC2A1	P485L	TPEELFHPLGADSQV	GLUT1 deficiency syndrome 1 (GLUT1DS1) [MIM:606777]

See Meyer et al. 2018 for additional tables.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-FLAG	Sigma-Aldrich	Cat# F3165; RRID:AB_259529
Rabbit polyclonal anti-GLUT1	Merck Millipore	Cat# 07-1401; RRID:AB_1587074
Rat monoclonal anti-ICAM2	BD Biosciences	Cat# 553326; RRID:AB_394784
Isolectin GS-IB4 Alexa Fluor 488 conjugate (IB4)	Thermo Fisher Scientific	Cat# I21411; RRID:AB_2314662
Mouse monoclonal anti-Vti1a	BD Biosciences	Cat# 611220; RRID:AB_398752
Mouse monoclonal anti-Vti1b	BD Biosciences	Cat# 611405; RRID:AB_398927
Rabbit polyclonal anti-EEA1	Cell Signaling Technology	Cat# 2411S; RRID:AB_2096814
Rabbit monoclonal anti-Rab4	Abcam	Cat# ab13252; RRID:AB_2269374
Rabbit monoclonal anti-Rab9	Cell Signaling Technology	Cat# 5118S; RRID:AB_10621426
Rabbit monoclonal anti-LAMP1	Cell Signaling Technology	Cat# 9091; RRID:AB_2687579
Mouse monoclonal anti-gamma Adaptin (AP-1 γ)	BD Biosciences	Cat# 610385; RRID:AB_397768
Mouse monoclonal anti-AP50 (AP-2 μ)	BD Biosciences	Cat# 611351; RRID:AB_398873
Mouse monoclonal anti- alpha Adaptin (AP-2 α)	Thermo Fisher Scientific	Cat# MA3-061; RRID:AB_2056321
Mouse monoclonal anti-alpha Adaptin (AP-2 α)	Abcam	Cat# ab2730; RRID:AB_303255
Mouse monoclonal anti - IL-2 R alpha (TAC)	Santa Cruz Biotechnology	Cat# sc-65258, RRID:AB_631112
Chemicals, Peptides, and Recombinant Proteins		
Deoxy-D-glucose, 2-[1,2-3H (N)]-, Specific Activity: 5-10 Ci (185-370 GBq)/mmol, 250 μCi (9.25 MBq)	Perkin Elmer	NET328250UC
L-arginine-HCl (Arg0)	Sigma-Aldrich	A6969; CAS: 1119-34-2
L-arginine-HCl($^{13}C_6$) (Arg6)	Sigma-Aldrich	643440; CAS: 201740-91-2
L-arginine-HCl($^{13}C_6$,$^{15}N_4$) (Arg10)	Sigma-Aldrich	608033; CAS: 202468-25-5
L-lysine-HCl (Lys0)	Sigma-Aldrich	L8662; CAS: 657-27-2
L-lysine-2HCl(4,4,5,5-D4) (Lys4)	Cambridge Isotope Laboratories	DLM-2640-PK; CAS: 657-26-1
L-lysine-HCl($^{13}C_6$,$^{15}N_2$) (Lys8)	Silantes	211604102

Cytochalasin B	Sigma-Aldrich	C2743; CAS: 14930-96-2
Deposited Data		
Peptide-protein interaction screen dataset	Meyer et al. 2018	PXD010027
PRM dataset	Meyer et al. 2018	PXD010005
BioID	Meyer et al. 2018	PXD010061
Experimental Models: Cell Lines		
Flp-In T-Rex GLUT1	Meyer et al. 2018	N/A
Flp-In T-Rex GLUT1_P485L	Meyer et al. 2018	N/A
Human: Patient-derived iPSCs	Meyer et al. 2018	https://hpscreg.eu/: BIHi037-(A-E)
Experimental Models: Organisms		
Mouse: C57BL/6N: GLUT1_P485L		N/A
Oligonucleotides		
ON-TARGETplus Human AP2M1 (1173) siRNA - SMARTpool	Dharmacon	Cat# L-008170-00-0005
ON-TARGETplus Non-targeting Pool	Dharmacon	Cat# D-001810-10-05
Recombinant DNA		
SLC2A1 (GLUT1)	Harvard Plasmid Repository	HsCD00378964
Software and Algorithms		
Imaris v8.4.1	Bitplane	N/A
MaxQuant v1.5.2.8	http://www.biochem.mpg.de/5111795/maxquant	Cox and Mann 2008

Acknowledgments

I want to start with thanking Matthias Selbach for having been such an inspiring leader, scientist, and person, throughout the entire time of my Ph.D. The positive and supportive work environment he has created is unique and I am eternally grateful to have had the opportunity to carry out my Ph.D studies in his lab. This positive vibe has been wonderfully filled with life by all current and past members of the lab: Amir, Boris, Carlos, Christian, Christopher, Damir, Djordje, Erik, Florian, Fritz, Henrik, Jimmy, Joao, Kamila, Katrin, Koshi, Marie, Martha, Michal, Mirjam, Teresa and Tommaso.

I would also like to thank all members of my Ph.D committee: Prof. Andreas Herrmann, Prof. Markus Landthaler, Prof. Matthias Selbach, Prof. Thomas Sommer and Dr. Bora Uyar for taking the time to read my dissertation and for participating in my defense.

I was blessed with a very diligent and scientifically curious master student, Jing-Yuan, who help me pushing the GLUT1 part of the project further. I want to especially thank Marie for starting off the project with me and being a great example on how to get things done; Martha, who always has always lent a helping hand and was the keeper of the cool in the most stressful situations; Christian, without whom the lab falls into pieces and who keeps everything (and everyone as Flying Ion Circus) running; Katrin, who took all the time in the world to introduce me to the lab and made my start in the lab really pleasant. She had to leave us way too early; Boris, who always had an open ear and the willingness to dive deep into projects to help out with new ideas.

It was a pleasure to work together with all collaboration partners on this project and so overwhelming to see with how much energy and enthusiasm they contributed. I am also grateful to the patient and her family. Seeing that our work reaches real people is priceless. So many different people from other labs at the MDC offered their help simply out of kindness: Ouidad and Nouhad from Markus Landthaler’s lab generated stable cell lines. Jane Reznick spent a whole afternoon with me in the isotope lab to introduce me to radioactive glucose uptake. Many others helped out with reagents or borrowed their machines.

Koshi was so kind to host me in Japan for 2 months and provided me with an unforgettable experience at the end of my Ph.D.

In general, this work would not have been possible without MDC's great scientific infrastructure and experts for everything just next door.

It's not only help with the daily scientific tasks, but also mental support during a Ph.D's ups and downs that make a real difference: Kamila, I would have never been able to do this without your hugs. Kamila and Michal, you are both my favorite conference/retreat roommates. Sharing experiences and talking with you and Erik about everything and anything made some big problems all of a sudden become small.

The "Wednesday's lunch" was a steady institution during the whole time of my Ph.D. Thank you, Anna, Giulia, Ina, and Julia. Sharing the dessert and the latest news with you was always a joy.

Ale, you are my life- and work-illustrator of choice. Thank you for always being there for me. The biggest thanks go to my family, a constant source of support, even if they sometimes wonder what in the world happened so I would end up in a white coat.

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt habe.

References

Alexa, A., and Rahnenfuhrer, J. 2016. "topGO: Enrichment Analysis for Gene Ontology. R package version 2.24.0. CRAN.

Babu, M. Madan, Robin van der Lee, Natalia Sanchez de Groot, and Jörg Gsponer. 2011. "Intrinsically Disordered Proteins: Regulation and Disease." Current Opinion in Structural Biology 21 (3): 432–40.

Bantscheff, Marcus, Simone Lemeer, Mikhail M. Savitski, and Bernhard Kuster. 2012. "Quantitative Mass Spectrometry in Proteomics: Critical Review Update from 2007 to the Present." Analytical and Bioanalytical Chemistry 404 (4): 939–65.

Bartke, Till, Michiel Vermeulen, Blerta Xhemalce, Samuel C. Robson, Matthias Mann, and Tony Kouzarides. 2010. "Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation." Cell 143 (3): 470–84.

Becerra, Andrés, Victor A. Bucheli, and Pedro A. Moreno. 2017. "Prediction of Virus-Host Protein-Protein Interactions Mediated by Short Linear Motifs." BMC Bioinformatics 18 (1): 163.

Belotti, Edwige, Jolanta Polanowska, Avais M. Daulat, Stéphane Audebert, Virginie Thomé, Jean-Claude Lissitzky, Frédérique Lembo, et al. 2013. "The Human PDZome: A Gateway to PSD95-Disc Large-Zonula Occludens (PDZ)-Mediated Functions." Molecular & Cellular Proteomics: MCP 12 (9): 2587–2603.

Bezprozvanny, Ilya. 2011. "Role of Inositol 1,4,5-Trishosphate Receptors in Pathogenesis of Huntington's Disease and Spinocerebellar Ataxias." Neurochemical Research 36 (7): 1186–97.

Blaydon, Diana C., Sarah L. Etheridge, Janet M. Risk, Hans-Christian Hennies, Laura J. Gay, Rebecca Carroll, Vincent Plagnol, et al. 2012. "RHBDF2 Mutations Are Associated with Tylosis, a Familial Esophageal Cancer Syndrome." American Journal of Human Genetics 90 (2): 340–46.

Boehr, David D., Ruth Nussinov, and Peter E. Wright. 2009. "The Role of Dynamic Conformational Ensembles in Biomolecular Recognition." Nature Chemical Biology 5 (11): 789–96.

Bonifacino, Juan S., and Raul Rojas. 2006. "Retrograde Transport from Endosomes to the Trans-Golgi Network." Nature Reviews. Molecular Cell Biology 7 (8): 568–79.

Boothby, Thomas C., Hugo Tapia, Alexandra H. Brozena, Samantha Piszkiewicz, Austin E. Smith, Ilaria Giovannini, Lorena Rebecchi, Gary J. Pielak, Doug Koshland, and Bob Goldstein. 2017. "Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation." Molecular Cell 65 (6): 975–84.e5.

Borgia, Alessandro, Madeleine B. Borgia, Katrine Bugge, Vera M. Kissling, Pétur O. Heidarsson, Catarina B. Fernandes, Andrea Sottini, et al. 2018. "Extreme Disorder in an Ultrahigh-Affinity Protein Complex." Nature 555 (7694): 61–66.

Buljan, Marija, Guilhem Chalancon, Sebastian Eustermann, Gunter P. Wagner, Monika Fuxreiter, Alex Bateman, and M. Madan Babu. 2012. "Tissue-Specific Splicing of Disordered Segments That Embed Binding Motifs Rewires Protein Interaction Networks." Molecular Cell 46 (6): 871–83.

Castro, Anna, Cyril Bernis, Suzanne Vigneron, Jean-Claude Labbé, and Thierry Lorca. 2005. "The Anaphase-Promoting Complex: A Key Factor in the Regulation of Cell Cycle." Oncogene 24 (3): 314–25.

Chen, Yucai, Jianjun Lu, Hong Pan, Yuehua Zhang, Husheng Wu, Keming Xu, Xiaoyan Liu, et al. 2003. "Association between Genetic Variation ofCACNA1H and Childhood Absence Epilepsy." Annals of Neurology 54 (2): 239–43.

Coderre, P. E., E. K. Cloherty, R. J. Zottola, and A. Carruthers. 1995. "Rapid Substrate Translocation by the Multisubunit, Erythroid Glucose Transporter Requires Subunit Associations but Not Cooperative Ligand Binding." Biochemistry 34 (30): 9762–73.

Collins, Mark O., Lu Yu, Iain Campuzano, Seth G. N. Grant, and Jyoti S. Choudhary. 2008. "Phosphoproteomic Analysis of the Mouse Brain Cytosol Reveals a Predominance of Protein Phosphorylation in Regions of Intrinsic Sequence Disorder." Molecular & Cellular Proteomics: MCP 7 (7): 1331–48.

Cooper, Gregory M., and Jay Shendure. 2011. "Needles in Stacks of Needles: Finding Disease-Causal Variants in a Wealth of Genomic Data." Nature Reviews. Genetics 12 (9): 628–40.

Costes, Sylvain V., Dirk Daelemans, Edward H. Cho, Zachary Dobbin, George Pavlakis, and Stephen Lockett. 2004. "Automatic and Quantitative Measurement of Protein-Protein Colocalization in Live Cells." Biophysical Journal 86 (6): 3993–4003.

Couzens, Amber L., James D. R. Knight, Michelle J. Kean, Guoci Teo, Alexander Weiss, Wade H. Dunham, Zhen-Yuan Lin, et al. 2013. "Protein Interaction Network of the Mammalian Hippo Pathway Reveals Mechanisms of Kinase-Phosphatase Interactions." Science Signaling 6 (302): rs15.

Cox, Guy, and Colin J. R. Sheppard. 2004. "Practical Limits of Resolution in Confocal and Non-Linear Microscopy." Microscopy Research and Technique 63 (1): 18–22.

Cox, Jürgen, Marco Y. Hein, Christian A. Luber, Igor Paron, Nagarjuna Nagaraj, and Matthias Mann. 2014. "Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ." Molecular & Cellular Proteomics: MCP 13 (9): 2513–26.

Cox, Jürgen, and Matthias Mann. 2008. "MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification." Nature Biotechnology 26 (12): 1367–72.

Cox, Jürgen, and Matthias Mann. 2011. "Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology." Annual Review of Biochemistry 80: 273–99.

Craig, N. L. 1988. "The Mechanism of Conservative Site-Specific Recombination." Annual Review of Genetics 22: 77–105.

Davey, Norman E., Gilles Travé, and Toby J. Gibson. 2011. "How Viruses Hijack Cell Regulation." Trends in Biochemical Sciences 36 (3): 159–69.

Davey, Norman E., Kim Van Roey, Robert J. Weatheritt, Grischa Toedt, Bora Uyar, Brigitte Altenberg, Aidan Budd, Francesca Diella, Holger Dinkel, and Toby J. Gibson. 2012. "Attributes of Short Linear Motifs." Molecular bioSystems 8 (1): 268–81.

Dellis, Olivier, Skarlatos G. Dedos, Stephen C. Tovey, Taufiq-Ur-Rahman, Stefan J. Dubel, and Colin W. Taylor. 2006. "Ca2+ Entry through Plasma Membrane IP3 Receptors." Science 313 (5784): 229–33.

Deng, Hao, Kai Gao, and Joseph Jankovic. 2014. "The Role of FUS Gene Variants in Neurodegenerative Diseases." Nature Reviews. Neurology 10 (6): 337–48.

De Vivo, Darryl C., Rosario R. Trifiletti, Ronald I. Jacobson, Gabriel M. Ronen, Ramin A. Behmand, and Sami I. Harik. 1991. "Defective Glucose Transport across the Blood-Brain Barrier as a Cause of Persistent Hypoglycorrhachia, Seizures, and Developmental Delay." The New England Journal of Medicine 325 (10): 703–9.

Diella, Francesca, Niall Haslam, Claudia Chica, Aidan Budd, Sushama Michael, Nigel P. Brown, Gilles Trave, and Toby J. Gibson. 2008. "Understanding Eukaryotic Linear Motifs and Their Role in Cell Signaling and Regulation." Frontiers in Bioscience: A Journal and Virtual Library 13 (May): 6580–6603.

Dinkel, Holger, Kim Van Roey, Sushama Michael, Norman E. Davey, Robert J. Weatheritt, Diana Born, Tobias Speck, et al. 2014. "The Eukaryotic Linear Motif Resource ELM: 10 Years and Counting." Nucleic Acids Research 42 (Database issue): D259–66.

Dinkel, Holger, Kim Van Roey, Sushama Michael, Manjeet Kumar, Bora Uyar, Brigitte Altenberg, Vladislava Milchevskaya, et al. 2016. "ELM 2016--Data Update and New Functionality of the Eukaryotic Linear Motif Resource." Nucleic Acids Research 44 (D1): D294–300.

Diril, Muhammed Kasim, Stefan Schmidt, Michael Krauss, Verena Gawlik, Hans-Georg Joost, Annette Schürmann, Volker Haucke, and Robert Augustin. 2009. "Lysosomal Localization of GLUT8 in the Testis--the EXXXLL Motif of GLUT8 Is Sufficient for Its Intracellular Sorting via AP1- and AP2-Mediated Interaction." The FEBS Journal 276 (14): 3729–43.

Dittmar, Gunnar, Daniel Perez-Hernandez, Elisabeth Kowenz-Leutz, Marieluise Kirchner, Guenther Kahlert, Radoslaw Wesolowski, Katharina Baum, et al. 2017. "Protein Interaction Screen on Peptide Matrix (PRISMA) Reveals Interaction Footprints and the PTM-Dependent Interactome of Intrinsically Disordered C/EBPβ." https://doi.org/10.1101/238709.

Dormann, Dorothee, Ramona Rodde, Dieter Edbauer, Eva Bentmann, Ingeborg Fischer, Alexander Hruscha, Manuel E. Than, et al. 2010. "ALS-Associated Fused in Sarcoma (FUS) Mutations Disrupt Transportin-Mediated Nuclear Import." The EMBO Journal 29 (16): 2841–57.

Dosztanyi, Z., B. Meszaros, and I. Simon. 2009. "Bioinformatical Approaches to Characterize Intrinsically Disordered/unstructured Proteins." Briefings in Bioinformatics 11 (2): 225–43.

Dosztányi, Zsuzsanna. 2018. "Prediction of Protein Disorder Based on IUPred." Protein Science: A Publication of the Protein Society 27 (1): 331–40.

Dosztányi, Zsuzsanna, Veronika Csizmok, Peter Tompa, and István Simon. 2005. "IUPred: Web Server for the Prediction of Intrinsically Unstructured Regions of Proteins Based on Estimated Energy Content." Bioinformatics 21 (16): 3433–34.

Duncan, Bruce K., and Jeffrey H. Miller. 1980. "Mutagenic Deamination of Cytosine Residues in DNA." Nature 287 (5782): 560–61.

Dyson, H. Jane, and Peter E. Wright. 2002. "Coupling of Folding and Binding for Unstructured Proteins." Current Opinion in Structural Biology 12 (1): 54–60.

Dyson, H. Jane, and Peter E. Wright. 2005. "Intrinsically Unstructured Proteins and Their Functions." Nature Reviews. Molecular Cell Biology 6 (3): 197–208.

Eyster, Craig A., Jason D. Higginson, Robert Huebner, Natalie Porat-Shliom, Roberto Weigert, Wells W. Wu, Rong-Fong Shen, and Julie G. Donaldson. 2009. "Discovery of New Cargo Proteins That Enter Cells through Clathrin-Independent Endocytosis." Traffic 10 (5): 590–99.

Fabbro, Megan, and Beric R. Henderson. 2003. "Regulation of Tumor Suppressors by Nuclear-Cytoplasmic Shuttling." Experimental Cell Research 282 (2): 59–69.

Famiglietti, Maria Livia, Anne Estreicher, Arnaud Gos, Jerven Bolleman, Sébastien Géhant, Lionel Breuza, Alan Bridge, et al. 2014. "Genetic Variations and Diseases

in UniProtKB/Swiss-Prot: The Ins and Outs of Expert Manual Curation." Human Mutation 35 (8): 927–35.
Fields, S., and O. Song. 1989. "A Novel Genetic System to Detect Protein-Protein Interactions." Nature 340 (6230): 245–46.
Foskett, J. Kevin, J. Kevin Foskett, Carl White, King-Ho Cheung, and Don-On Daniel Mak. 2007. "Inositol Trisphosphate Receptor Ca2 Release Channels." Physiological Reviews 87 (2): 593–658.
Foskett, J. Kevin, Carl White, King-Ho Cheung, and Don-On Daniel Mak. 2007. "Inositol Trisphosphate Receptor Ca2+ Release Channels." Physiological Reviews 87 (2): 593–658.
Francavilla, Chiara, Kristoffer T. G. Rigbolt, Kristina B. Emdal, Gianni Carraro, Erik Vernet, Dorte B. Bekker-Jensen, Werner Streicher, et al. 2013. "Functional Proteomics Defines the Molecular Switch Underlying FGF Receptor Trafficking and Cellular Outputs." Molecular Cell 51 (6): 707–22.
Frank, Ronald. 1992. "Spot-Synthesis: An Easy Technique for the Positionally Addressable, Parallel Chemical Synthesis on a Membrane Support." Tetrahedron 48 (42): 9217–32.
Frank, Ronald. 2002. "The SPOT-Synthesis Technique. Synthetic Peptide Arrays on Membrane Supports--Principles and Applications." Journal of Immunological Methods 267 (1): 13–26.
Fuchs, Serge Y., Vladimir S. Spiegelman, and K. G. Suresh Kumar. 2004. "The Many Faces of Beta-TrCP E3 Ubiquitin Ligases: Reflections in the Magic Mirror of Cancer." Oncogene 23 (11): 2028–36.
Fuxreiter, Monika, Peter Tompa, and István Simon. 2007. "Local Structural Disorder Imparts Plasticity on Linear Motifs." Bioinformatics 23 (8): 950–56.
Gandhi, T. K. B., Jun Zhong, Suresh Mathivanan, L. Karthick, K. N. Chandrika, S. Sujatha Mohan, Salil Sharma, et al. 2006. "Analysis of the Human Protein Interactome and Comparison with Yeast, Worm and Fly Interaction Datasets." Nature Genetics 38 (3): 285–93.
Gardner, Mckinlay, Melanie Knight, Kenju Hara, Shoji Tsuji, Susan Forrest, and Elsdon Storey. 2005. "Spinocerebellar Ataxia Type 15." Cerebellum 4 (1): 47–50.
Gibson, Toby J., Holger Dinkel, Kim Van Roey, and Francesca Diella. 2015. "Experimental Detection of Short Regulatory Motifs in Eukaryotic Proteins: Tips for Good Practice as Well as for Bad." Cell Communication and Signaling: CCS 13 (November): 42.
Gingras, Anne-Claude, Matthias Gstaiger, Brian Raught, and Ruedi Aebersold. 2007. "Analysis of Protein Complexes Using Mass Spectrometry." Nature Reviews. Molecular Cell Biology 8 (8): 645–54.
Gould, Cathryn M., Francesca Diella, Allegra Via, Pål Puntervoll, Christine Gemünd, Sophie Chabanis-Davidson, Sushama Michael, et al. 2010. "ELM: The Status of the 2010 Eukaryotic Linear Motif Resource." Nucleic Acids Research 38 (Database issue): D167–80.
Gsponer, Jörg, Matthias E. Futschik, Sarah A. Teichmann, and M. Madan Babu. 2008. "Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation." Science 322 (5906): 1365–68.
Gstaiger, Matthias, and Ruedi Aebersold. 2009. "Applying Mass Spectrometry-Based Proteomics to Genetics, Genomics and Network Biology." Nature Reviews. Genetics 10 (9): 617–27.

Guettler, Sebastian, Jose LaRose, Evangelia Petsalaki, Gerald Gish, Andy Scotter, Tony Pawson, Robert Rottapel, and Frank Sicheri. 2011. "Structural Basis and Sequence Rules for Substrate Recognition by Tankyrase Explain the Basis for Cherubism Disease." Cell 147 (6): 1340–54.

Hara, K., A. Shiga, H. Nozaki, J. Mitsui, Y. Takahashi, H. Ishiguro, H. Yomono, et al. 2008. "Total Deletion and a Missense Mutation of ITPR1 in Japanese SCA15 Families." Neurology 71 (8): 547–51.

He, Bo, Kejun Wang, Yunlong Liu, Bin Xue, Vladimir N. Uversky, and A. Keith Dunker. 2009. "Predicting Intrinsic Disorder in Proteins: An Overview." Cell Research 19 (8): 929–49.

Hernández-Miranda, Luis R., Anna Cariboni, Clare Faux, Christiana Ruhrberg, Jin Hyung Cho, Jean-François Cloutier, Britta J. Eickholt, John G. Parnavelas, and William D. Andrews. 2011. "Robo1 Regulates Semaphorin Signaling to Guide the Migration of Cortical Interneurons through the Ventral Forebrain." The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 31 (16): 6174–87.

Heron, Sarah E., Hilary A. Phillips, John C. Mulley, Aziz Mazarib, Miriam Y. Neufeld, Samuel F. Berkovic, and Ingrid E. Scheffer. 2004. "Genetic Variation of CACNA1H in Idiopathic Generalized Epilepsy." Annals of Neurology 55 (4): 595–96.

Higo, Takayasu, Kozo Hamada, Chihiro Hisatsune, Nobuyuki Nukina, Tsutomu Hashikawa, Mitsuharu Hattori, Takeshi Nakamura, and Katsuhiko Mikoshiba. 2010. "Mechanism of ER Stress-Induced Brain Damage by IP(3) Receptor." Neuron 68 (5): 865–78.

Hilpert, Kai, Dirk F. H. Winkler, and Robert E. W. Hancock. 2007. "Peptide Arrays on Cellulose Support: SPOT Synthesis, a Time and Cost Efficient Method for Synthesis of Large Numbers of Peptides in a Parallel and Addressable Fashion." Nature Protocols 2 (6): 1333–49.

Hirst, Jennifer, Georg H. H. Borner, Robin Antrobus, Andrew A. Peden, Nicola A. Hodson, Daniela A. Sahlender, and Margaret S. Robinson. 2012. "Distinct and Overlapping Roles for AP-1 and GGAs Revealed by the 'Knocksideways' System." Current Biology: CB 22 (18): 1711–16.

Holt, Liam J., Brian B. Tuch, Judit Villén, Alexander D. Johnson, Steven P. Gygi, and David O. Morgan. 2009. "Global Analysis of Cdk1 Substrate Phosphorylation Sites Provides Insights into Evolution." Science 325 (5948): 1682–86.

Hosp, Fabian, Hannes Vossfeldt, Matthias Heinig, Djordje Vasiljevic, Anup Arumughan, Emanuel Wyler, Genetic and Environmental Risk for Alzheimer's Disease GERAD1 Consortium, et al. 2015. "Quantitative Interaction Proteomics of Neurodegenerative Disease Proteins." Cell Reports 11 (7): 1134–46.

Hunt T. 1990 Protein sequence motifs involved in recognition and targeting: a new series. Trends Biochem. Sci.15:305.

Iakoucheva, L. M., P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, and A. K. Dunker. 2004. "The Importance of Intrinsic Disorder for Protein Phosphorylation." Nucleic Acids Research 32 (3): 1037–49.

Ingram, V. M. 1957. "Gene Mutations in Human Hæmoglobin: The Chemical Difference Between Normal and Sickle Cell Hæmoglobin." Nature 180 (4581): 326–28.

Ittner, Lars M., and Jürgen Götz. 2007. "Pronuclear Injection for the Production of Transgenic Mice." Nature Protocols 2 (5): 1206–15.

Kaake, Robyn M., Xiaorong Wang, Anthony Burke, Clinton Yu, Wynne Kandur, Yingying Yang, Eric J. Novtisky, et al. 2014. "A New in Vivo Cross-Linking Mass

Spectrometry Platform to Define Protein-Protein Interactions in Living Cells." Molecular & Cellular Proteomics: MCP 13 (12): 3533–43.
Kadaveru, Krishna, Jay Vyas, and Martin R. Schiller. 2008. "Viral Infection and Human Disease--Insights from Minimotifs." Frontiers in Bioscience: A Journal and Virtual Library 13 (May): 6455–71.
Keane, Harriet, Brent J. Ryan, Brendan Jackson, Alan Whitmore, and Richard Wade-Martins. 2015. "Protein-Protein Interaction Networks Identify Targets Which Rescue the MPP+ Cellular Model of Parkinson's Disease." Scientific Reports 5 (November): 17004.
Keilhauer, Eva C., Marco Y. Hein, and Matthias Mann. 2015. "Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) rather than Affinity Purification Mass Spectrometry (AP-MS)." Molecular & Cellular Proteomics: MCP 14 (1): 120–35.
Kelly, Bernard T., and David J. Owen. 2011. "Endocytic Sorting of Transmembrane Protein Cargo." Current Opinion in Cell Biology 23 (4): 404–12.
Kenworthy, A. K. 2001. "Imaging Protein-Protein Interactions Using Fluorescence Resonance Energy Transfer Microscopy." Methods 24 (3): 289–96.
Khosravani, Houman, Christophe Altier, Brett Simms, Kevin S. Hamming, Terrance P. Snutch, Janette Mezeyova, John E. McRory, and Gerald W. Zamponi. 2004. "Gating Effects of Mutations in the Cav3.2 T-Type Calcium Channel Associated with Childhood Absence Epilepsy." The Journal of Biological Chemistry 279 (11): 9681–84.
Köhler, Sebastian, Nicole A. Vasilevsky, Mark Engelstad, Erin Foster, Julie McMurry, Ségolène Aymé, Gareth Baynam, et al. 2017. "The Human Phenotype Ontology in 2017." Nucleic Acids Research 45 (D1): D865–76.
Kononenko, Natalia L., Dmytro Puchkov, Gala A. Classen, Alexander M. Walter, Arndt Pechstein, Linda Sawade, Natalie Kaempf, et al. 2014. "Clathrin/AP-2 Mediate Synaptic Vesicle Reformation from Endosome-like Vacuoles but Are Not Essential for Membrane Retrieval at Central Synapses." Neuron 82 (5): 981–88.
Kozik, Patrycja, Richard W. Francis, Matthew N. J. Seaman, and Margaret S. Robinson. 2010. "A Screen for Endocytic Motifs." Traffic 11 (6): 843–55.
Landrum, Melissa J., Jennifer M. Lee, Mark Benson, Garth Brown, Chen Chao, Shanmuga Chitipiralla, Baoshan Gu, et al. 2016. "ClinVar: Public Archive of Interpretations of Clinically Relevant Variants." Nucleic Acids Research 44 (D1): D862–68.
Landry, Christian R., Emmanuel D. Levy, Diala Abd Rabbo, Kirill Tarassov, and Stephen W. Michnick. 2013. "Extracting Insight from Noisy Cellular Networks." Cell 155 (5): 983–89.
Lau, Ho-Tak, Hyong Won Suh, Martin Golkowski, and Shao-En Ong. 2014. "Comparing SILAC- and Stable Isotope Dimethyl-Labeling Approaches for Quantitative Proteomics." Journal of Proteome Research 13 (9): 4164–74.
Lee, Eunice E., Jing Ma, Anastasia Sacharidou, Wentao Mi, Valerie K. Salato, Nam Nguyen, Youxing Jiang, et al. 2015. "A Protein Kinase C Phosphorylation Motif in GLUT1 Affects Glucose Transport and Is Mutated in GLUT1 Deficiency Syndrome." Molecular Cell 58 (5): 845–53.
Leen, Wilhelmina G., Joerg Klepper, Marcel M. Verbeek, Maike Leferink, Tom Hofste, Baziel G. van Engelen, Ron A. Wevers, et al. 2010. "Glucose Transporter-1 Deficiency Syndrome: The Expanding Clinical and Genetic Spectrum of a Treatable Disorder." Brain: A Journal of Neurology 133 (Pt 3): 655–70.

Lee, Robin van der, Marija Buljan, Benjamin Lang, Robert J. Weatheritt, Gary W. Daughdrill, A. Keith Dunker, Monika Fuxreiter, et al. 2014. "Classification of Intrinsically Disordered Regions and Proteins." Chemical Reviews 114 (13): 6589–6631.

Levy, Emmanuel D., Christian R. Landry, and Stephen W. Michnick. 2009. "How Perfect Can Protein Interactomes Be?" Science Signaling 2 (60): e11.

Lippmann, Ethan S., Samira M. Azarin, Jennifer E. Kay, Randy A. Nessler, Hannah K. Wilson, Abraham Al-Ahmad, Sean P. Palecek, and Eric V. Shusta. 2012. "Derivation of Blood-Brain Barrier Endothelial Cells from Human Pluripotent Stem Cells." Nature Biotechnology 30 (8): 783–91.

Liu, Bernard A., Brett W. Engelmann, and Piers D. Nash. 2012. "High-Throughput Analysis of Peptide-Binding Modules." Proteomics 12 (10): 1527–46.

Liu, Xiaonan, Kari Salokas, Fitsum Tamene, Yaming Jiu, Rigbe G. Weldatsadik, Tiina Öhman, and Markku Varjosalo. 2018. "An AP-MS- and BioID-Compatible MAC-Tag Enables Comprehensive Mapping of Protein Interactions and Subcellular Localizations." Nature Communications 9 (1): 1188.

Lyon, Kenneth F., Xingyu Cai, Richard J. Young, Abdullah-Al Mamun, Sanguthevar Rajasekaran, and Martin R. Schiller. 2018. "Minimotif Miner 4: A Million Peptide Minimotifs and Counting." Nucleic Acids Research 46 (D1): D465–70.

Maldonado-Báez, Lymarie, Chad Williamson, and Julie G. Donaldson. 2013. "Clathrin-Independent Endocytosis: A Cargo-Centric View." Experimental Cell Research 319 (18): 2759–69.

Malty, Ramy H., Hiroyuki Aoki, Ashwani Kumar, Sadhna Phanse, Shahreen Amin, Qingzhou Zhang, Zoran Minic, et al. 2017. "A Map of Human Mitochondrial Protein Interactions Linked to Neurodegeneration Reveals New Mechanisms of Redox Homeostasis and NF-κB Signaling." Cell Systems 5 (6): 564–77.e12.

Mann, Matthias. 2006. "Functional and Quantitative Proteomics Using SILAC." Nature Reviews. Molecular Cell Biology 7 (12): 952–58.

McLaren, William, Laurent Gil, Sarah E. Hunt, Harpreet Singh Riat, Graham R. S. Ritchie, Anja Thormann, Paul Flicek, and Fiona Cunningham. 2016. "The Ensembl Variant Effect Predictor." Genome Biology 17 (1): 122.

Meier, Florian, Andreas-David Brunner, Scarlet Koch, Heiner Koch, Markus Lubeck, Michael Krause, Niels Goedecke, et al. 2018. "Online Parallel Accumulation – Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer." https://doi.org/10.1101/336743.

Meyer, Katrina, Marieluise Kirchner, Bora Uyar, Jing-Yuan Cheng, Giulia Russo, Luis R. Hernandez-Miranda, Anna Szymborska, et al. 2018. "Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs." Cell, August. https://doi.org/10.1016/j.cell.2018.08.019.

Mitsunari, Takashi, Fubito Nakatsu, Noriko Shioda, Paul E. Love, Alexander Grinberg, Juan S. Bonifacino, and Hiroshi Ohno. 2005. "Clathrin Adaptor AP-2 Is Essential for Early Embryonal Development." Molecular and Cellular Biology 25 (21): 9318–23.

Motley, Alison, Nicholas A. Bright, Matthew N. J. Seaman, and Margaret S. Robinson. 2003. "Clathrin-Mediated Endocytosis in AP-2–depleted Cells." The Journal of Cell Biology 162 (5): 909–18.

Narayan, Soumil, Gary D. Bader, and Jüri Reimand. 2016. "Frequent Mutations in Acetylation and Ubiquitination Sites Suggest Novel Driver Mechanisms of Cancer." Genome Medicine 8 (1): 55.

Neduva, Victor, Rune Linding, Isabelle Su-Angrand, Alexander Stark, Federico de Masi, Toby J. Gibson, Joe Lewis, Luis Serrano, and Robert B. Russell. 2005. "Systematic Discovery of New Recognition Peptides Mediating Protein Interaction Networks." PLoS Biology 3 (12): e405.

Neduva, Victor, and Robert B. Russell. 2006. "Peptides Mediating Interaction Networks: New Leads at Last." Current Opinion in Biotechnology 17 (5): 465–71.

Okada, Hirokazu, Akiyoshi Uezu, Erik J. Soderblom, M. Arthur Moseley 3rd, Frank B. Gertler, and Scott H. Soderling. 2012. "Peptide Array X-Linking (PAX): A New Peptide-Protein Identification Approach." PloS One 7 (5): e37035.

Owen, David J., Brett M. Collins, and Philip R. Evans. 2004. "ADAPTORS FOR CLATHRIN COATS: Structure and Function." Annual Review of Cell and Developmental Biology 20 (1): 153–91.

Pajkos, Mátyás, Bálint Mészáros, István Simon, and Zsuzsanna Dosztányi. 2012. "Is There a Biological Cost of Protein Disorder? Analysis of Cancer-Associated Mutations." Molecular bioSystems 8 (1): 296–307.

Pandey, Kailash N. 2009. "Functional Roles of Short Sequence Motifs in the Endocytosis of Membrane Receptors." Frontiers in Bioscience 14 (June): 5339–60.

Park, Sang Yoon, and Xiaoli Guo. 2014. "Adaptor Protein Complexes and Intracellular Transport." Bioscience Reports 34 (4): 381–90.

Pascual, Juan M., Dong Wang, Ru Yang, Lei Shi, Hong Yang, and Darryl C. De Vivo. 2008. "Structural Signatures and Membrane Helix 4 in GLUT1." The Journal of Biological Chemistry 283 (24): 16732–42.

Pasternak, Jack J. 2005. An Introduction to Human Molecular Genetics: Mechanisms of Inherited Diseases. John Wiley & Sons.

Paul, Florian E., Fabian Hosp, and Matthias Selbach. 2011. "Analyzing Protein-Protein Interactions by Quantitative Mass Spectrometry." Methods 54 (4): 387–95.

Peden, Andrew A., Viola Oorschot, Boris A. Hesser, Cary D. Austin, Richard H. Scheller, and Judith Klumperman. 2004. "Localization of the AP-3 Adaptor Complex Defines a Novel Endosomal Exit Site for Lysosomal Membrane Proteins." The Journal of Cell Biology 164 (7): 1065–76.

Pierce, M. M., C. S. Raman, and B. T. Nall. 1999. "Isothermal Titration Calorimetry of Protein-Protein Interactions." Methods 19 (2): 213–21.

Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Séraphin. 2001. "The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification." Methods 24 (3): 218–29.

Puntervoll, P. 2003. "ELM Server: A New Resource for Investigating Short Functional Sites in Modular Eukaryotic Proteins." Nucleic Acids Research 31 (13): 3625–30.

Radivojac, P., P. H. Baenziger, M. G. Kann, M. E. Mort, M. W. Hahn, and S. D. Mooney. 2008. "Gain and Loss of Phosphorylation Sites in Human Cancer." Bioinformatics 24 (16): i241–47.

Raiborg, Camilla, and Harald Stenmark. 2009. "The ESCRT Machinery in Endosomal Sorting of Ubiquitylated Membrane Proteins." Nature 458 (7237): 445–52.

Rappsilber, Juri, Yasushi Ishihama, and Matthias Mann. 2003. "Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics." Analytical Chemistry 75 (3): 663–70.

Reimand, Jüri, and Gary D. Bader. 2013. "Systematic Analysis of Somatic Mutations in Phosphorylation Signaling Predicts Novel Cancer Drivers." Molecular Systems Biology 9: 637.

Rhee, Hyun-Woo, Peng Zou, Namrata D. Udeshi, Jeffrey D. Martell, Vamsi K. Mootha, Steven A. Carr, and Alice Y. Ting. 2013. "Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging." Science 339 (6125): 1328–31.

Romero, Pedro R., Saima Zaidi, Ya Yin Fang, Vladimir N. Uversky, Predrag Radivojac, Christopher J. Oldfield, Marc S. Cortese, et al. 2006. "Alternative Splicing in Concert with Protein Intrinsic Disorder Enables Increased Functional Diversity in Multicellular Organisms." Proceedings of the National Academy of Sciences of the United States of America 103 (22): 8390–95.

Romero, P., Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker. 2001. "Sequence Complexity of Disordered Protein." Proteins 42 (1): 38–48.

Ross, Christopher A., Jacopo Meldolesi, Teresa A. Milner, Tomohide Satoh, Surachai Supattapone, and Solomon H. Snyder. 1989. "Inositol 1,4,5-Trisphosphate Receptor Localized to Endoplasmic Reticulum in Cerebellar Purkinje Neurons." Nature 339 (6224): 468–70.

Roux, Kyle J., Dae In Kim, Brian Burke, and Danielle G. May. 2018. "BioID: A Screen for Protein-Protein Interactions." Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [et Al.] 91 (February): 19.23.1–19.23.15.

Roux, Kyle J., Dae In Kim, Manfred Raida, and Brian Burke. 2012. "A Promiscuous Biotin Ligase Fusion Protein Identifies Proximal and Interacting Proteins in Mammalian Cells." The Journal of Cell Biology 196 (6): 801–10.

Sahni, Nidhi, Song Yi, Mikko Taipale, Juan I. Fuxman Bass, Jasmin Coulombe-Huntington, Fan Yang, Jian Peng, et al. 2015. "Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders." Cell 161 (3): 647 60.

Sarkar, Debasree, Tanmoy Jana, and Sudipto Saha. 2015. "LMPID: A Manually Curated Database of Linear Motifs Mediating Protein–protein Interactions." Database 2015. https://doi.org/10.1093/database/bav014.

Sauer, B. 1994. "Site-Specific Recombination: Developments and Applications." Current Opinion in Biotechnology 5 (5): 521–27.

Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, et al. 2012. "Fiji: An Open-Source Platform for Biological-Image Analysis." Nature Methods 9 (7): 676–82.

Schorge, Stephanie, Joyce van de Leemput, Andrew Singleton, Henry Houlden, and John Hardy. 2010. "Human Ataxias: A Genetic Dissection of Inositol Triphosphate Receptor (ITPR1)-Dependent Signaling." Trends in Neurosciences 33 (5): 211–19.

Schulze, Waltraud X., and Matthias Mann. 2004. "A Novel Proteomic Screen for Peptide-Protein Interactions." The Journal of Biological Chemistry 279 (11): 10756–64.

Schuster-Böckler, Benjamin, and Alex Bateman. 2008. "Protein Interactions in Human Genetic Diseases." Genome Biology 9 (1): R9.

Schwanhäusser, Björn, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen, and Matthias Selbach. 2011. "Global Quantification of Mammalian Gene Expression Control." Nature 473 (7347): 337–42.

Seet, Bruce T., and Tony Pawson. 2004. "MAPK Signaling: Sho Business." Current Biology: CB 14 (17): R708–10.

Seidner, G., M. G. Alvarez, J. I. Yeh, K. R. O'Driscoll, J. Klepper, T. S. Stump, D. Wang, N. B. Spinner, M. J. Birnbaum, and D. C. De Vivo. 1998. "GLUT-1 Deficiency Syndrome Caused by Haploinsufficiency of the Blood-Brain Barrier Hexose Carrier." Nature Genetics 18 (2): 188–91.

Selbach, Matthias, Florian Ernst Paul, Sabine Brandt, Patrick Guye, Oliver Daumke, Steffen Backert, Christoph Dehio, and Matthias Mann. 2009. "Host Cell Interactome of Tyrosine-Phosphorylated Bacterial Proteins." Cell Host & Microbe 5 (4): 397–403.

Seo, Moon-Hyeong, and Philip M. Kim. 2018. "The Present and the Future of Motif-Mediated Protein–protein Interactions." Current Opinion in Structural Biology 50: 162–70.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." Genome Research 13 (11): 2498–2504.

Sharma, Aarti, Alexander K. Lyashchenko, Lei Lu, Sara Ebrahimi Nasrabady, Margot Elmaleh, Monica Mendelsohn, Adriana Nemes, Juan Carlos Tapia, George Z. Mentis, and Neil A. Shneider. 2016. "ALS-Associated Mutant FUS Induces Selective Motor Neuron Degeneration through Toxic Gain of Function." Nature Communications 7 (February): 10465.

Shi, Jun, and Konstantin V. Kandror. 2008. "Study of Glucose Uptake in Adipose Cells." Methods in Molecular Biology 456: 307–15.

Sickmeier, M., J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, et al. 2007. "DisProt: The Database of Disordered Proteins." Nucleic Acids Research 35 (Database): D786–93.

Silvis, Mark R., John A. Picciano, Carol Bertrand, Kelly Weixel, Robert J. Bridges, and Neil A. Bradbury. 2003. "A Mutation in the Cystic Fibrosis Transmembrane Conductance Regulator Generates a Novel Internalization Sequence and Enhances Endocytic Rates." The Journal of Biological Chemistry 278 (13): 11554–60.

Slaughter, Laurel, George Vartzelis, and Todd Arthur. 2009. "New GLUT-1 Mutation in a Child with Treatment-Resistant Epilepsy." Epilepsy Research 84 (2-3): 254–56.

Staudt, Catherine, Emeline Puissant, and Marielle Boonen. 2016. "Subcellular Trafficking of Mammalian Lysosomal Proteins: An Extended View." International Journal of Molecular Sciences 18 (1). https://doi.org/10.3390/ijms18010047.

Stein, Amelie, Roberto Mosca, and Patrick Aloy. 2011. "Three-Dimensional Modeling of Protein Interactions and Complexes Is Going 'Omics." Current Opinion in Structural Biology 21 (2): 200–208.

Storey, E., R. J.M. Gardner, M. A. Knight, M. L. Kennerson, R. R. Tuck, S. M. Forrest, and G. A. Nicholson. 2001. "A New Autosomal Dominant Pure Cerebellar Ataxia." Neurology 57 (10): 1913–15.

Subramanian, Sankar, and Sudhir Kumar. 2006. "Evolutionary Anatomies of Positions and Types of Disease-Associated and Neutral Amino Acid Mutations in the Human Genome." BMC Genomics 7 (December): 306.

Sury, Matthias D., Jia-Xuan Chen, and Matthias Selbach. 2010. "The SILAC Fly Allows for Accurate Protein Quantification in Vivo." Molecular & Cellular Proteomics: MCP 9 (10): 2173–83.

Szklarczyk, Damian, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, et al. 2015. "STRING v10: Protein-Protein Interaction Networks, Integrated over the Tree of Life." Nucleic Acids Research 43 (Database issue): D447–52.

Tanimura, A., Y. Tojyo, and R. J. Turner. 2000. "Evidence That Type I, II, and III Inositol 1,4,5-Trisphosphate Receptors Can Occur as Integral Plasma Membrane Proteins." The Journal of Biological Chemistry 275 (35): 27488–93.

Theillet, Francois-Xavier, Lajos Kalmar, Peter Tompa, Kyou-Hoon Han, Philipp Selenko, A. Keith Dunker, Gary W. Daughdrill, and Vladimir N. Uversky. 2013. "The Alphabet of Intrinsic Disorder: I. Act like a Pro: On the Abundance and Roles of Proline Residues in Intrinsically Disordered Proteins." Intrinsically Disordered Proteins 1 (1): e24360.

Theos, Alexander C., Danièle Tenza, José A. Martina, Ilse Hurbain, Andrew A. Peden, Elena V. Sviderskaya, Abigail Stewart, et al. 2005. "Functions of Adaptor Protein (AP)-3 and AP-1 in Tyrosinase Sorting from Endosomes to Melanosomes." Molecular Biology of the Cell 16 (11): 5356–72.

Tompa, Peter. 2002. "Intrinsically Unstructured Proteins." Trends in Biochemical Sciences 27 (10): 527–33.

Tompa, Peter. 2011. "Unstructural Biology Coming of Age." Current Opinion in Structural Biology 21 (3): 419–25.

Tompa, Peter, Norman E. Davey, Toby J. Gibson, and M. Madan Babu. 2014. "A Million Peptide Motifs for the Molecular Biologist." Molecular Cell 55 (2): 161–69.

Tonikian, Raffi, Xiaofeng Xin, Christopher P. Toret, David Gfeller, Christiane Landgraf, Simona Panni, Serena Paoluzi, et al. 2009. "Bayesian Modeling of the Yeast SH3 Domain Interactome Predicts Spatiotemporal Dynamics of Endocytosis Proteins." PLoS Biology 7 (10): e1000218.

Toprak, Umut H., Ludovic C. Gillet, Alessio Maiolica, Pedro Navarro, Alexander Leitner, and Ruedi Aebersold. 2014. "Conserved Peptide Fragmentation as a Benchmarking Tool for Mass Spectrometers and a Discriminating Feature for Targeted Proteomics." Molecular & Cellular Proteomics: MCP 13 (8): 2056–71.

Traub, Linton M. 2009. "Tickets to Ride: Selecting Cargo for Clathrin-Regulated Internalization." Nature Reviews. Molecular Cell Biology 10 (9): 583–96.

Traub, Linton M., and Juan S. Bonifacino. 2013. "Cargo Recognition in Clathrin-Mediated Endocytosis." Cold Spring Harbor Perspectives in Biology 5 (11): a016790.

Tripathi, Shashank, Marie O. Pohl, Yingyao Zhou, Ariel Rodriguez-Frandsen, Guojun Wang, David A. Stein, Hong M. Moulton, et al. 2015. "Meta- and Orthogonal Integration of Influenza 'OMICs' Data Defines a Role for UBR4 in Virus Budding." Cell Host & Microbe 18 (6): 723–35.

UniProt Consortium. 2012. "Reorganizing the Protein Space at the Universal Protein Resource (UniProt)." Nucleic Acids Research 40 (Database issue): D71–75.

Uversky, Vladimir N. 2015. "Intrinsically Disordered Proteins and Their (disordered) Proteomes in Neurodegenerative Disorders." Frontiers in Aging Neuroscience 7 (March): 18.

Uversky, V. N., J. R. Gillespie, and A. L. Fink. 2000. "Why Are 'Natively Unfolded' Proteins Unstructured under Physiologic Conditions?" Proteins 41 (3): 415–27.

Uyar, Bora, Robert J. Weatheritt, Holger Dinkel, Norman E. Davey, and Toby J. Gibson. 2014. "Proteome-Wide Analysis of Human Disease Mutations in Short Linear Motifs: Neglected Players in Cancer?" Molecular bioSystems 10 (10): 2626–42.

Vacic, Vladimir, and Lilia M. Iakoucheva. 2012. "Disease Mutations in Disordered Regions--Exception to the Rule?" Molecular bioSystems 8 (1): 27–32.

Vacic, Vladimir, Phineus R. L. Markwick, Christopher J. Oldfield, Xiaoyue Zhao, Chad Haynes, Vladimir N. Uversky, and Lilia M. Iakoucheva. 2012. "Disease-Associated Mutations

Disrupt Functionally Important Regions of Intrinsic Protein Disorder." PLoS Computational Biology 8 (10): e1002709.
Van Roey, Kim, Holger Dinkel, Robert J. Weatheritt, Toby J. Gibson, and Norman E. Davey. 2013. "The switches.ELM Resource: A Compendium of Conditional Regulatory Interaction Interfaces." Science Signaling 6 (269): rs7.
Van Roey, Kim, Toby J. Gibson, and Norman E. Davey. 2012. "Motif Switches: Decision-Making in Cell Regulation." Current Opinion in Structural Biology 22 (3): 378–85.
Van Roey, Kim, Sandra Orchard, Samuel Kerrien, Marine Dumousseau, Sylvie Ricard-Blum, Henning Hermjakob, and Toby J. Gibson. 2013. "Capturing Cooperative Interactions with the PSI-MI Format." Database: The Journal of Biological Databases and Curation 2013 (September): bat066.
Vavouri, Tanya, Jennifer I. Semple, Rosa Garcia-Verdugo, and Ben Lehner. 2009. "Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity." Cell 138 (1): 198–208.
Veltman, Joris A., and Han G. Brunner. 2012. "De Novo Mutations in Human Genetic Disease." Nature Reviews. Genetics 13 (8): 565–75.
Venter, J. C. 2001. "The Sequence of the Human Genome." Science 291 (5507): 1304–51.
Vermeulen, Michiel, Nina C. Hubner, and Matthias Mann. 2008. "High Confidence Determination of Specific Protein-Protein Interactions Using Quantitative Mass Spectrometry." Current Opinion in Biotechnology 19 (4): 331–37.
Vidal, Marc, Michael E. Cusick, and Albert-László Barabási. 2011. "Interactome Networks and Human Disease." Cell 144 (6): 986–98.
Vidal, M., J. L. Montiel, D. Cussac, F. Cornille, M. Duchesne, F. Parker, B. Tocqué, B. P. Roques, and C. Garbay. 1998. "Differential Interactions of the Growth Factor Receptor-Bound Protein 2 N-SH3 Domain with Son of Sevenless and Dynamin. Potential Role in the Ras-Dependent Signaling Pathway." The Journal of Biological Chemistry 273 (9): 5343–48.
Wald, Tomas, Frantisek Spoutil, Adriana Osickova, Michaela Prochazkova, Oldrich Benada, Petr Kasparek, Ladislav Bumba, et al. 2017. "Intrinsically Disordered Proteins Drive Enamel Formation via an Evolutionarily Conserved Self-Assembly Motif." Proceedings of the National Academy of Sciences of the United States of America 114 (9): E1641–50.
Wang, Xiujuan, Xiaomu Wei, Bram Thijssen, Jishnu Das, Steven M. Lipkin, and Haiyuan Yu. 2012. "Three-Dimensional Reconstruction of Protein Networks Provides Insight into Human Genetic Disease." Nature Biotechnology 30 (2): 159–64.
Weatheritt, Robert J., and Toby J. Gibson. 2012. "Linear Motifs: Lost in (pre)translation." Trends in Biochemical Sciences 37 (8): 333–41.
Wefers, Benedikt, Sanum Bashir, Jana Rossius, Wolfgang Wurst, and Ralf Kühn. 2017. "Gene Editing in Mouse Zygotes Using the CRISPR/Cas9 System." Methods 121-122 (May): 55–67.
Wong, Andrew D., Mao Ye, Amanda F. Levy, Jeffrey D. Rothstein, Dwight E. Bergles, and Peter C. Searson. 2013. "The Blood-Brain Barrier: An Engineering Perspective." Frontiers in Neuroengineering 6 (August): 7.
Woodsmith, Jonathan, and Ulrich Stelzl. 2017. "Understanding Disease Variants through the Lens of Protein Interactions." Cell Systems 5 (6): 544–46.

Wright, Peter E., and H. Jane Dyson. 2015. "Intrinsically Disordered Proteins in Cellular Signalling and Regulation." Nature Reviews. Molecular Cell Biology 16 (1): 18–29.

Wright, Peter E., and H. Jane Dyson. 2009. "Linking Folding and Binding." Current Opinion in Structural Biology 19 (1): 31–38.

Wu, Ning, Bin Zheng, Adam Shaywitz, Yossi Dagon, Christine Tower, Gary Bellinger, Che-Hung Shen, et al. 2013. "AMPK-Dependent Degradation of TXNIP upon Energy Stress Leads to Enhanced Glucose Uptake via GLUT1." Molecular Cell 49 (6): 1167–75.

Xie, Q., G. E. Arnold, P. Romero, Z. Obradovic, E. Garner, and A. K. Dunker. 1998. "The Sequence Attribute Method for Determining Relationships Between Sequence and Protein Disorder." Genome Informatics. Workshop on Genome Informatics 9: 193–200.

Xue, Bin, A. Keith Dunker, and Vladimir N. Uversky. 2012. "Orderly Order in Protein Intrinsic Disorder Distribution: Disorder in 3500 Proteomes from Viruses and the Three Domains of Life." Journal of Biomolecular Structure & Dynamics 30 (2): 137–49.

Yamazaki, Haruka, Hiroaki Nozaki, Osamu Onodera, Takayuki Michikawa, Masatoyo Nishizawa, and Katsuhiko Mikoshiba. 2011. "Functional Characterization of the P1059L Mutation in the Inositol 1,4,5-Trisphosphate Receptor Type 1 Identified in a Japanese SCA15 Family." Biochemical and Biophysical Research Communications 410 (4): 754–58.

Yang, L., L. J. Embree, S. Tsai, and D. D. Hickstein. 1998. "Oncoprotein TLS Interacts with Serine-Arginine Proteins Involved in RNA Splicing." The Journal of Biological Chemistry 273 (43): 27761–64.

Yue, Peng, Zhaolong Li, and John Moult. 2005. "Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease." Journal of Molecular Biology 353 (2): 459–73.

Zakaryan, Rouzanna P., and Heinz Gehring. 2006. "Identification and Characterization of the Nuclear Localization/Retention Signal in the EWS Proto-Oncoprotein." Journal of Molecular Biology 363 (1): 27–38.

Zauber, Henrik, Marieluise Kirchner, and Matthias Selbach. 2018. "Picky: A Simple Online PRM and SRM Method Designer for Targeted Proteomics." Nature Methods 15 (3): 156–57.

Zhong, Quan, Nicolas Simonis, Qian-Ru Li, Benoit Charloteaux, Fabien Heuze, Niels Klitgord, Stanley Tam, et al. 2009. "Edgetic Perturbation Models of Human Inherited Disorders." Molecular Systems Biology 5 (November): 321.