

# Rewiring of PDZ Domain-Ligand Interaction Network Contributed to Eukaryotic Evolution

Jinho Kim<sup>1</sup>, Inhae Kim<sup>1</sup>, Jae-Seong Yang<sup>2</sup>, Young-Eun Shin<sup>1</sup>, Jihye Hwang<sup>3</sup>, Solip Park<sup>2</sup>, Yoon Sup Choi<sup>4</sup>, Sanguk Kim<sup>1,2,3\*</sup>

**1** Division of Molecular and Life Science, Pohang University of Science and Technology, Pohang, Korea, **2** School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang, Korea, **3** Division of ITCE, Pohang University of Science and Technology, Pohang, Korea, **4** Cancer Research Institute, Seoul National University, Seoul, Korea

## Abstract

PDZ domain-mediated interactions have greatly expanded during metazoan evolution, becoming important for controlling signal flow via the assembly of multiple signaling components. The evolutionary history of PDZ domain-mediated interactions has never been explored at the molecular level. It is of great interest to understand how PDZ domain-ligand interactions emerged and how they become rewired during evolution. Here, we constructed the first human PDZ domain-ligand interaction network (PDZNet) together with binding motif sequences and interaction strengths of ligands. PDZNet includes 1,213 interactions between 97 human PDZ proteins and 591 ligands that connect most PDZ protein-mediated interactions (98%) in a large single network via shared ligands. We examined the rewiring of PDZ domain-ligand interactions throughout eukaryotic evolution by tracing changes in the C-terminal binding motif sequences of the PDZ ligands. We found that interaction rewiring by sequence mutation frequently occurred throughout evolution, largely contributing to the growth of PDZNet. The rewiring of PDZ domain-ligand interactions provided an effective means of functional innovations in nervous system development. Our findings provide empirical evidence for a network evolution model that highlights the rewiring of interactions as a mechanism for the development of new protein functions. PDZNet will be a valuable resource to further characterize the organization of the PDZ domain-mediated signaling proteome.

**Citation:** Kim J, Kim I, Yang J-S, Shin Y-E, Hwang J, et al. (2012) Rewiring of PDZ Domain-Ligand Interaction Network Contributed to Eukaryotic Evolution. *PLoS Genet* 8(2): e1002510. doi:10.1371/journal.pgen.1002510

**Editor:** Michael Snyder, Stanford University School of Medicine, United States of America

**Received:** September 7, 2011; **Accepted:** December 12, 2011; **Published:** February 9, 2012

**Copyright:** © 2012 Kim et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported in part by National Research Foundation grants (R31-10100 of the World Class University program, 20100028453 of the NCRC program, and 20100020528) funded by the Korean Ministry of Education, Science, and Technology, Seoul National University Cancer Research Institute (cri-10-2), and a POSTECH BSRI grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: sukim@postech.ac.kr

## Introduction

PDZ domains are linear motif-mediated protein-protein interaction modules. PDZ domain-ligand interactions have been greatly expanded in metazoans and are widely used to assemble signaling complexes, including those found in neuronal synapses [1]. Thus, an understanding of how PDZ domain-ligand interactions have evolved would help elucidate the design principle of the eukaryotic signaling network. Many studies have revealed the evolutionary history of PDZ domain families and their functional roles [2,3]. However, it remains poorly understood how PDZ domain-mediated interactions are rewired during the evolution of the protein interaction network.

Systematic analysis of interaction rewiring will provide new insights into eukaryotic evolution, which is not fully explained via only the expansion of protein families. Recently, it was suggested that rewiring of interactions is an important mechanism for the evolution of biological systems. Network comparison studies showed that protein interactions frequently change after gene duplication [4,5]. In particular, linear motifs were suggested to have great potential to rewire interactions because of their high rate of change [6,7]. Indeed, phosphorylation sites in one species are often lost in other species [8,9]. Moreover, human-specific

phosphorylation sites are recently examined to identify novel phenotypes in humans because the interaction rewiring of kinase interactions may contribute to the emergence of novel biological functions [10].

Structural information of interacting cellular components (i.e., structural interactome) would provide a more complete picture of a cell and help elucidate the evolutionary principle of the protein interaction network [11]. Recently, structural information of protein complexes were mapped onto protein interaction networks [12,13]. Indeed, such interface information of protein interactions would more clearly explain evolutionary principles, such as the network evolution model by gene duplication [12] and the role of residues surrounding linear motifs in terms of binding specificity [14]. Therefore, to understand the underlying design principle of the PDZ domain-ligand interaction network, detailed interface information at the amino acid level is needed.

In this work, we attempted the first systematic investigation of interaction rewiring in the PDZ domain-ligand interaction network and its role in eukaryotic evolution. We constructed a comprehensive human PDZ domain-ligand interaction network and traced the changes in interaction rewiring during evolution. We developed position weight matrices (PWMs) of human PDZ domains from the experimental data of PDZ domain-ligand

## Author Summary

Rewiring of interactions is a powerful tool for the evolution of organism complexity. Rewiring among preexisting proteins provides a simple mechanism for the development of new signaling circuits by redirecting information flows without a gain or loss of genes. Particularly, interactions mediated by short linear motifs can be easily changed by mutations during evolution, resulting in a rewiring of interactions. However, how interaction rewiring of linear motif interactions facilitates the emergence of new protein function during evolution is poorly understood. Here, we systematically investigated the rewiring of interactions mediated by PDZ domains, which are one of the most commonly found peptide recognition modules. We found that PDZ domain-ligand interactions are frequently rewired by C-terminal sequence mutations in PDZ ligands during evolution. Especially, rewiring of PDZ domain-ligand interactions was involved in neuronal function development, occurring concurrently with the emergence of vertebrates and suggesting that reorganization of signaling pathways by rewiring PDZ domain-ligand interactions significantly contributed to the evolution of nervous systems in vertebrates. Our findings highlight the rewiring of interactions as an effective means for functional innovation, providing new insight into eukaryotic evolution, which has not been fully explained by only the expansion of protein families.

interactions. The binding motif information of PDZNet helped to elucidate the changes in PDZ domain-ligand interactions. We found that PDZ domain-ligand interactions are frequently rewired throughout evolution via mutations of C-terminal PDZ ligand sequences. Particularly, interaction rewiring occurred concurrently with emergence of vertebrates whose rewired interactions were largely involved in neuronal signaling, suggesting that nervous system evolution might be achieved by the interaction rewiring of signaling components, such as PDZ protein-ligand interactions. Furthermore, the broad specificity of PDZ domains contributes to interaction rewiring by increasing the chance of acquiring PDZ binding motifs by sequence mutations. Our findings will prompt a new approach for the study of eukaryotic evolution by considering the rewiring of interactions as a major evolutionary process of domain-ligand interactions.

## Results

### Discovery of PDZ domain-ligand interactions and binding strengths

To elucidate how PDZ domain-ligand interactions have evolved, an accurate and detailed understanding of their interactions is essential. Furthermore, a network approach is useful to understand how evolution of PDZ domain-ligand interactions contributed to eukaryotic evolution, because protein functions may not be encoded in an individual protein but rather be encoded in the relationships between proteins in a protein-protein interaction network [15–17]. Therefore, we constructed a comprehensive network of PDZ protein-ligand interactions by integrating the experimental data of PDZ domain-ligand interactions and protein-protein interaction databases (Figure 1).

We developed a quantitative model of PDZ domain binding strengths from the experimental data of PDZ domain-ligand interactions, including interactions between 81 PDZ domains and 217 peptides from a protein array [18], the phage display of 86 PDZ domains [19,20], interactions between 147 PDZ domains

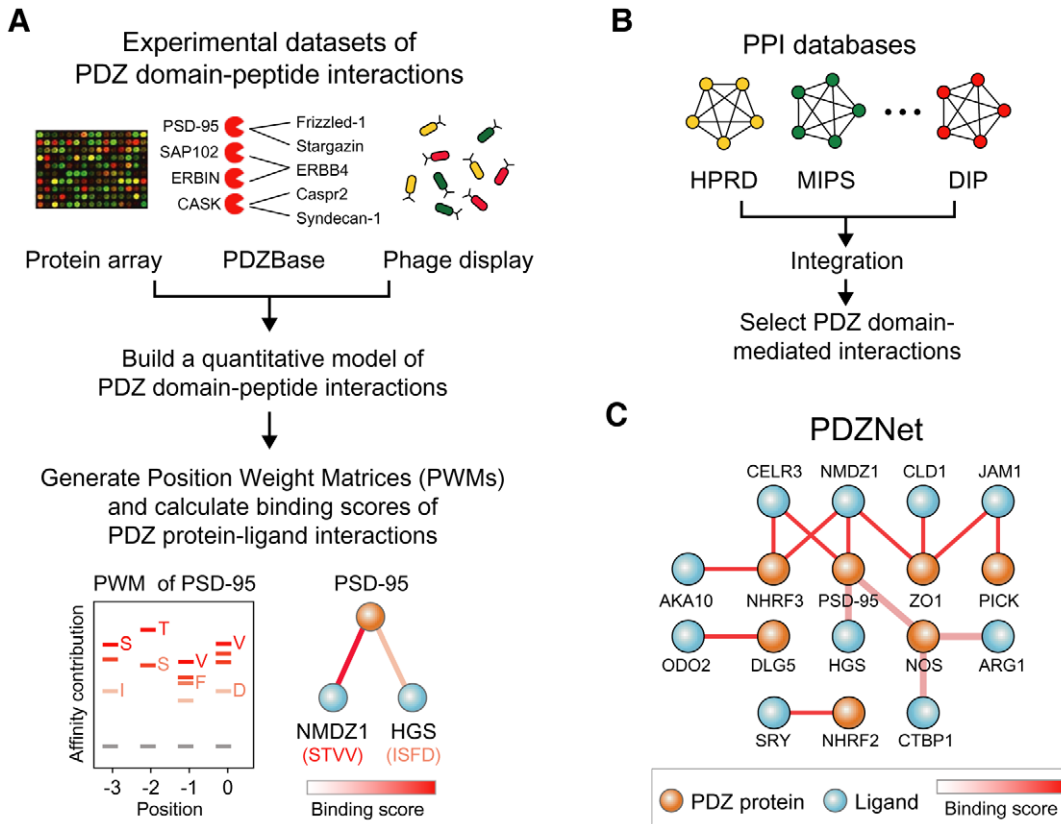
and 219 ligands from a database of *in vivo* PDZ domain-ligand interactions (PDZBase) [21], and literature mining [22,23] (Figure 1A). This model converts binary interactions between PDZ domains and ligands into PWMs, which can quantify the binding strengths of a given PDZ domain and peptide sequence based on the physical and chemical properties of binding pocket residues as well as the frequencies of amino acids found in the bound peptides. To capture the binding strengths of the PDZ domain-peptide interactions, we combined a machine-learning algorithm and an information theory-based PWM method. We provide the PWMs of human PDZ domains as a resource (Table S1). In this study, we focused on the C-terminal motifs of ligands for the analysis of PDZNet. Although several internal PDZ binding motifs have been reported, most PDZ domain-ligand interactions are mediated by C-terminal residues, owing to the structural constraint on the internal motifs that require the  $\beta$ -hairpin fold [24,25].

We found that the binding scores of PWMs well represent the experimental affinities of PDZ domain-ligand interactions (Figure 2). The large-scale binding affinities (Kd) of PDZ domain-peptide interactions are available for SNA1 and ERBIN PDZ domains [26]. The PWMs provided the binding scores of the interactions, which showed a strong positive correlation with the experimental affinities for both SNA1 ( $R^2=0.76$ ) and ERBIN ( $R^2=0.85$ ) PDZ domain-peptide interactions. Moreover, *in vivo* binding affinities of PSD-95\_1 (the first PDZ domain of PSD-95) with its ligands correlated well with its binding scores from the PWMs (Figure 3A and 3B). The Kds of Kv1.4 and GluR6 to PSD-95\_1 were measured experimentally [27]. Kv1.4 bound to PSD-95\_1 with high affinity (Kd = 1.5  $\mu$ M), whereas GluR6 bound to PSD-95\_1 with low affinity (Kd = 160  $\mu$ M). When we measured the binding scores of the PSD-95\_1 ligands based on PWMs, the binding score of Kv1.4 was found to be higher (binding score = 14.72) than that of GluR6 (binding score = 6.00; Figure 3B). Next, we measured how precisely the *in vivo* ligands of PDZ domains can be rediscovered by the binding scores obtained from PDZ domain-ligand interactions. We found that although the interaction data for the target PDZ domain were excluded from the training set, 290 of 320 (91%) of the known PDZ ligands were found in the top 10 percentile of binding scores (Figure 3C). We also found that our PWMs provided reliable predictions for PDZ domains derived from various species (Table S2). Furthermore, we found that our predicted PWMs agreed well with experimental data-based PWMs [20]. We compared the PWMs derived from phage display experiments with the predicted PWMs of the MAGI1\_2, DLG1\_2, and PTN13\_2 PDZ domains and confirmed that they were nearly identical (Figure S1).

To construct PDZNet with high-confidence interactions, we prioritized the experimentally validated PDZ protein-ligand interactions from the prediction results of the PWMs. It is a challenge to correlate the occurrence of amino acids in a linear motif to the binding specificity of peptide-binding domains [28]. The PWM method treats each amino acid position in a linear motif independently; thus, predicted interactions may include a fraction of false-positive results. Therefore, we only included interactions supported by experimental evidence. To assemble experimentally validated protein interactions, we integrated 22 different PPI databases containing 101,777 interactions among 11,043 proteins (Figure 1B and 1C).

### Properties of PDZNet

PDZNet is composed of 97 PDZ proteins and 596 partners with 1,212 interactions (Figure 4A), which can be accessed in Table S3. PDZ proteins interact with a various number of ligands (Figure S2)



**Figure 1. Construction of PDZNet.** (A) Building position weight matrices (PWMs) of human PDZ domains. Experimental data of the PDZ domain and peptide interactions were used to generate PWMs of PDZ domains. (B) Construction of the PDZ domain-ligand interaction network. Human protein interactions were collected by integrating existing PPI databases. (C) Integration of binding strengths into PDZNet. doi:10.1371/journal.pgen.1002510.g001

and most (98%) PDZ proteins are connected in a large single network via shared ligands. Beginning with PDZNet, we generated two network projections (Figure 4B), which displayed both PDZ protein-PDZ protein and ligand-ligand connections via common interacting partners. In the “PDZ protein network” (PPN; Figure 4B, left panel), nodes represent PDZ proteins; two PDZ proteins are connected if they share at least one ligand. Meanwhile, in the “PDZ ligand network” (PLN; Figure 4B, right panel), nodes are PDZ ligands; two PDZ ligands are connected if they share at least one PDZ protein. On average, a PDZ protein interacts with 17 partners, and a PDZ ligand interacts with three PDZ proteins. We further examined whether this multispecificity is also found at the domain level. For proteins with multiple PDZ domains, PDZNet specifies the interactions mediated by individual domains, yielding 2,643 PDZ domain-ligand interactions (Figure S3). On average, a PDZ domain interacts with 14 ligands, and a ligand interacts with four PDZ domains, suggesting that the complexity of PDZNet originated from the multispecificity of PDZ domains.

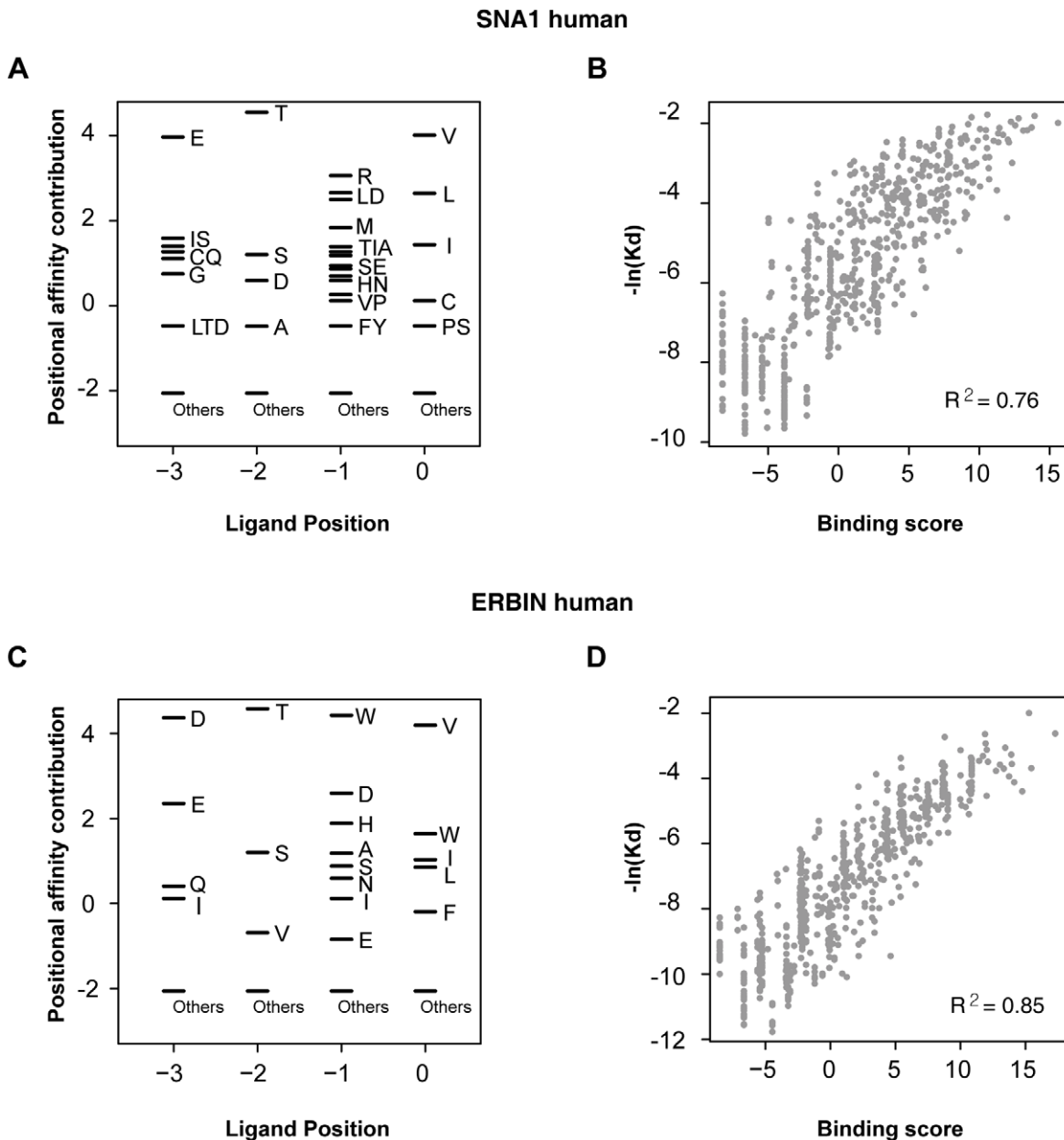
We discovered that an interface similarity exists among PDZ domains that share the same ligands. In the PPN, PDZ protein pairs connected by the same ligands tend to have similar pocket residues (Figure S4). For example, SAP97\_1, SAP97\_2, PSD-93\_1, PSD-93\_2, SAP102\_2, and PSD-95\_1 have similar binding pocket residues that bind the same ligand (AT2B4), suggesting that gene duplications contribute to the multispecific interactions in PDZNet and increase network complexity because interaction partners from gene duplication events tend to share the same interface [12]. Indeed, we found that SAP97, PSD-93, SAP102,

and PSD-95 PDZ proteins were paralogs, the products of gene duplication events. Interestingly, we also found cases of non-paralogous proteins that have similar binding specificities, which suggest that convergent evolution might also play a role in the development of network complexity. For example, two PDZ proteins, LAP2 and MAGI2, were found to interact with the same ligand, CTND2, although they are evolutionarily unrelated. Meanwhile, the PLN provides a complementary ligand-centered view of PDZNet. We found that the connected ligands in the PLN tend to have similar C-terminal sequences. As shown in Figure 4B, two PDZ ligands, ARVC and CTND2, interact with the LAP2 PDZ protein and have same binding motif (DSWV).

**Sequence mutations generate PDZ-binding motifs, driving the evolution of PDZNet**

We then asked how PDZ domains and ligands obtained multiple partners during evolution. Gene duplication and subsequent diversification events are considered major factors for network growth. Although gene duplication played a significant role in PDZ proteins and ligand evolution [29], it may not explain how a PDZ domain can interact with multiple, non-homologous ligands.

We found that sequence mutations played an important role for the attachment of non-homologous ligands to PDZ domains. On an evolutionary time scale, the compendium of PDZ ligands expands via two processes: (1) the introduction of new PDZ ligands by gene duplication of existing partners, or (2) the *de novo* evolution of new interactions via the acquisition of PDZ-binding motifs

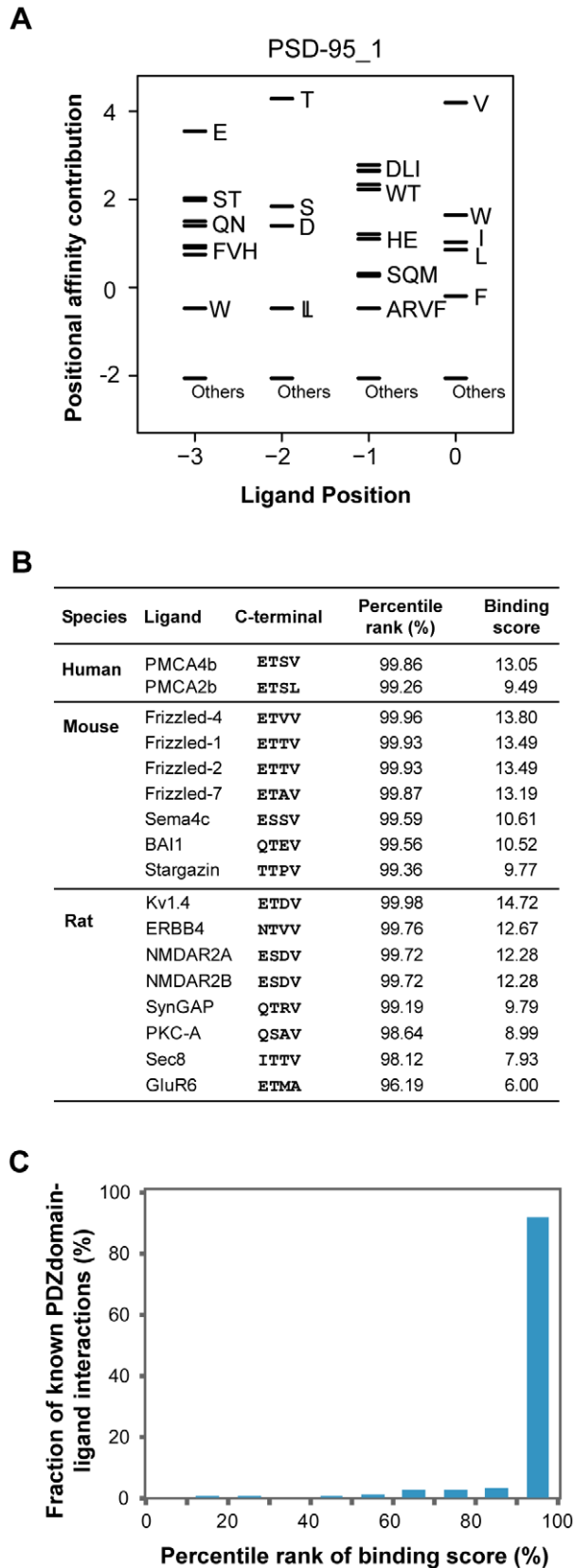


**Figure 2. Correlation between binding score and binding affinity.** (A and C) PWMs of the PDZ domains of SNA1 (human) and ERBIN (human). Black bars represent the affinity contribution of the binding scores to the corresponding amino acids. Clusters of amino acids with no preference are labeled “others.” (B and D) Scatter plots showing the correlation between binding score and binding affinity. doi:10.1371/journal.pgen.1002510.g002

(Figure 5A). To examine the extent of gene duplication in PDZNet growth, we calculated the paralog fractions of PDZNet because gene duplication products usually remain as homologous genes [30]. We discovered that the relatively small fraction of PDZ ligands that share a common partner were paralogs (33.6%), whereas a significantly larger portion of PDZ proteins that share a common partner were paralogs (54.5%; Wilcoxon’s rank-sum test;  $p = 1.24 \times 10^{-4}$ ; Figure 5B).

Next, we examined the sequence evolution of the binding motifs of human PDZ ligands and discovered that a large portion of PDZ ligands acquired their binding motifs via sequence mutations. We examined the C-terminal sequences of PDZ ligands in each PDZ domain-ligand interaction pair across 16 representative species.

We found that nearly one-third of human PDZ ligands gained their PDZ domain interactions by C-terminal mutations during evolution (Table S4; experimental evidence of human PDZ domain-ligand interactions are provided in Table S5). For example, NOS1AP obtained a PDZ-binding motif via sequence mutation and became an interaction partner with the NOS1 PDZ protein from vertebrates. We discovered that NOS1AP has orthologs in a wide range of species from yeast to human (Figure 5C). To examine whether the PDZ-binding motif of NOS1AP emerged from vertebrates, we compared the C-terminal sequences and binding scores of NOS1AP from invertebrate and vertebrate orthologs. The binding of mouse NOS1AP with NOS1 PDZ protein has been confirmed experimentally [31]. The



**Figure 3. Validation of PWMs on *in vivo* partners.** (A) The PWM of the PSD-95\_1 domain. (B) Known interacting partners of the PSD-95\_1 domain from three species are shown. (C) Fraction of known PDZ domain-ligand interactions are examined by percentile rank of binding scores.

doi:10.1371/journal.pgen.1002510.g003

C-terminal sequences of the vertebrate orthologs of NOS1AP are identical, whereas the C-terminal sequences of the invertebrate orthologs of NOS1AP vary across species and differ from those of the vertebrate NOS1AP orthologs. Moreover, we searched for evidence of a NOS1AP-NOS1 interaction in invertebrate PPI databases, including Databases of Interacting Proteins (DIP) [32], BioGrid [33], and Comprehensive Drosophila Interactions (Droidb) [34], but none was found. When we compared the binding scores of NOS1AP to NOS1, all invertebrates orthologs showed low binding scores (average binding score =  $-3.03$ ), whereas the binding scores of vertebrate orthologs were high (average binding score =  $5.27$ ; Figure 5D), indicating that NOS1AP was an invertebrate non-binder but gained the ability to bind the NOS1 PDZ protein in vertebrates.

### Rewiring of PDZ domain-ligand interactions plays an important role in the evolution of nervous systems in vertebrates

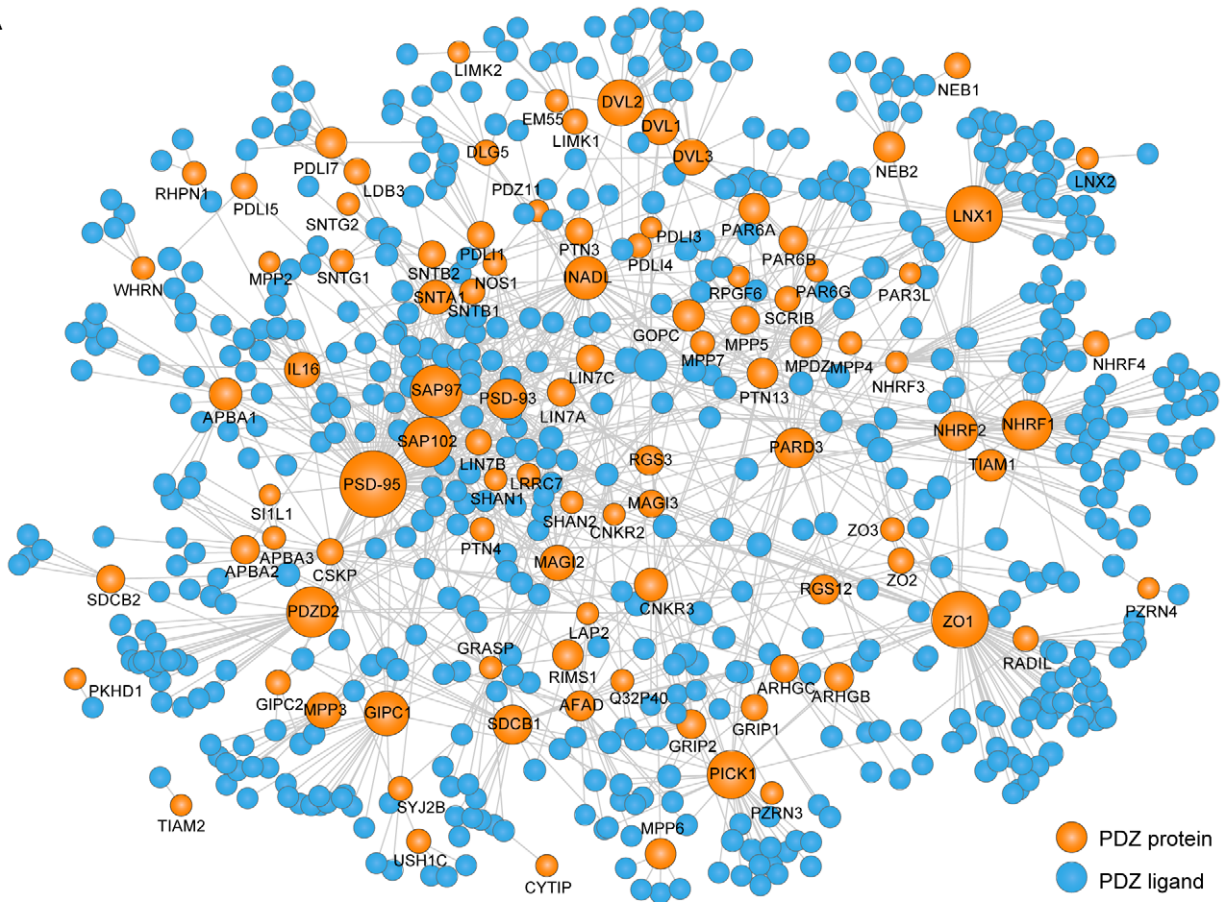
Interaction rewiring is an effective evolutionary mechanism given that it reconfigures molecular systems without a gain or loss of genes [35]. We hypothesized that the rewiring of PDZ domain-ligand interactions via sequence mutation contributed to the evolution of the vertebrate nervous system, in which PDZ proteins and ligands play an important role [1]. To determine whether the rewiring of PDZ domain-ligand interactions had a significant impact on vertebrate nervous systems, we calculated the rewiring rates between species from *Escherichia coli* to humans and examined the changes in the C-terminal sequences of PDZ ligands and the binding specificities of PDZ domains (Figure S5). We found that PDZ domain-ligand interactions were most frequently rewired between invertebrates and vertebrates (Figure 6).

We also examined the types of biological processes that are significantly affected by the rewiring events of PDZ domain-ligand interactions. We found that the PDZ ligands that arose in invertebrates and gained their PDZ-binding motifs in vertebrates participated significantly in the process of neurological system development (Table S6). For example, with the emergence of vertebrates, the breakpoint cluster region protein, BCR, acquired a PDZ-binding motif and began to interact with the AFADIN PDZ protein by changing its C-terminal sequence from ARLK (binding score =  $-2.03$ ) to STEV (binding score =  $5.87$ ). The binding of the PDZ domain interaction sites of BCR and the AFADIN PDZ domain was also confirmed by immunoprecipitation and NMR chemical shift perturbation experiments [36,37]. In vertebrates, BCR controls the interaction between AFADIN and RAS GTPase [36]. AFADIN also interacts with the vertebrate-specific receptors EPHA7 and EPHB3 of the Eph-receptor family, which regulate the morphology and motility of neuronal cells through the RAS GTPase [38,39]. Thus, the interaction between BCR and AFADIN may evolve to control EPH receptor signaling, which is greatly diversified in vertebrates. Meanwhile, we found that proteins that arose and acquired PDZ domain interaction sites in invertebrates tend to be involved in various cellular processes, such as vesicle-mediated transport, cell cycle, and RNA splicing (Table S6). These results suggest that the emergence of the vertebrate nervous system integrated preexisting functional units during the evolution of synapse complexity.

### Rewiring of PDZ domain-ligand interactions causes the proteins of premetazoan origin to engage in metazoan functions

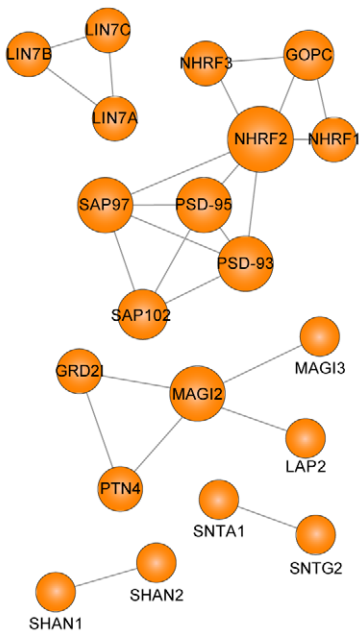
We found that metazoan-specific PDZ proteins adopted their ligands from proteins of premetazoan origin. The phylogenetic

**A**



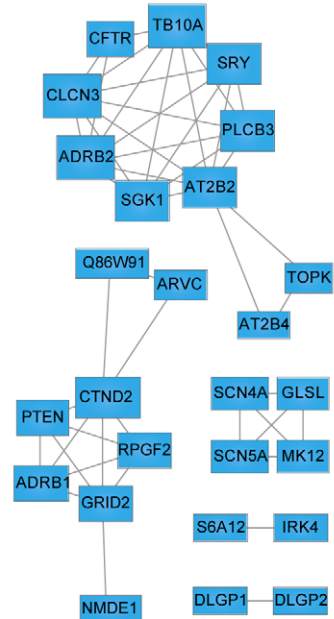
**B**

**PDZ Protein Network (PPN)**



PDZ proteins	Ligands	Motif
PTN4	NMDE1	ESDV
GRD21	GRID2	GTSI
	Q86W91	DSWV
LAP2	ARVC	DSWV
	CTND2	DSWV
MAGI2	RPGF2	VSAV
MAGI3	PTEN	ITKV
	ADRB1	ESKV
	CFTR	DTRL
NHRF1	ADRB2	DSLL
GOPC	TB10A	DTYL
NHRF3	CLCN3	IMFN
NHRF2	SGK1	DSFL
	PLCB3	NTQL
	SRY	WTKL
PSD-93	AT2B2	ETSL
PSD-95	AT2B4	ETSV
SAP97	TOPK	ETDV
SAP102	S6A12	ETHL
LIN7C	IRK4	ESAI
LIN7A	DLGP1	QTRL
LIN7B	DLGP2	QTRL
SHAN1	MK12	ETPL
SHAN2	SCN4A	ESLV
SNTA1	GLSL	ESMV
SNTG2	SCN5A	ESIV

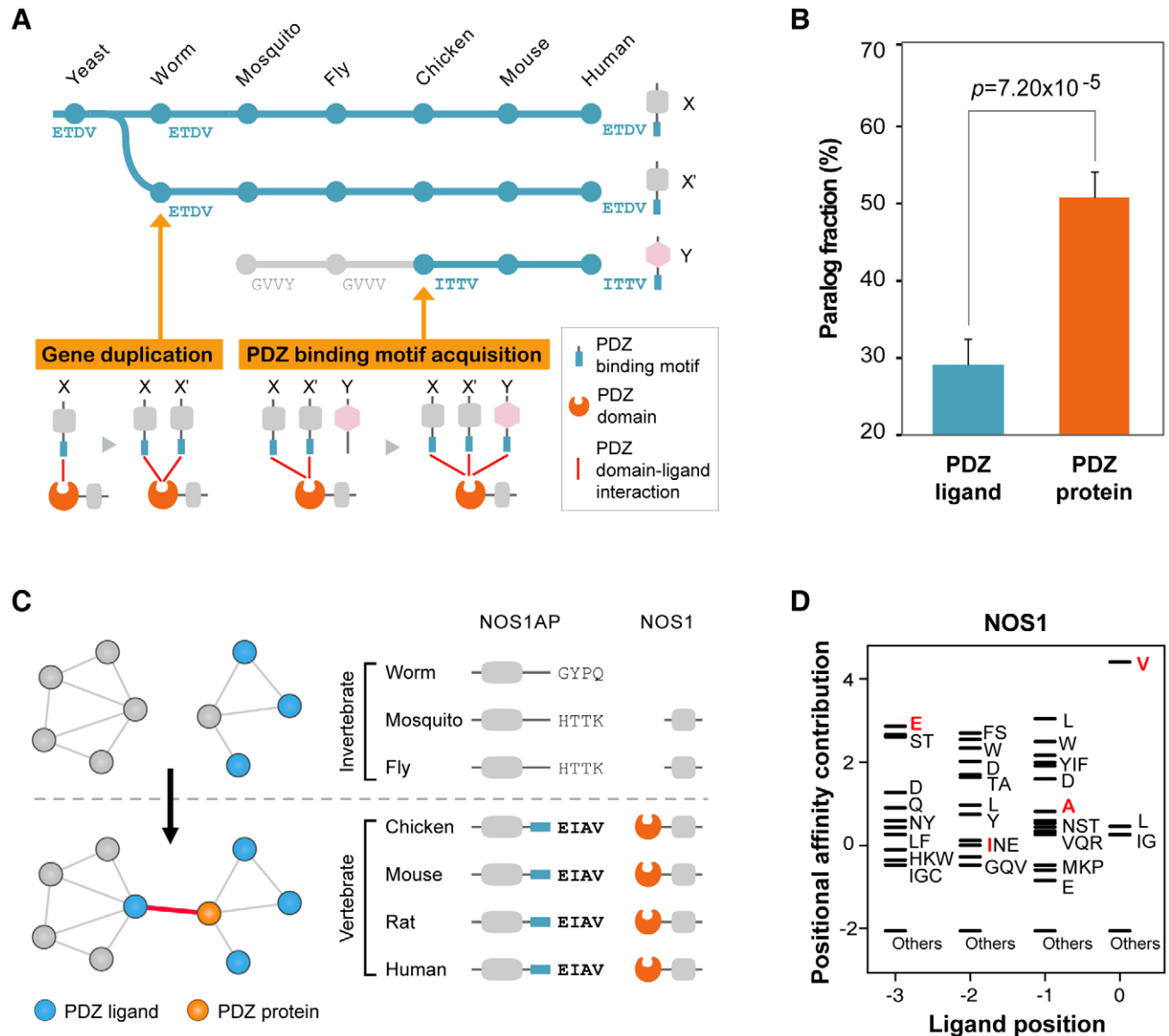
**PDZ Ligand Network (PLN)**



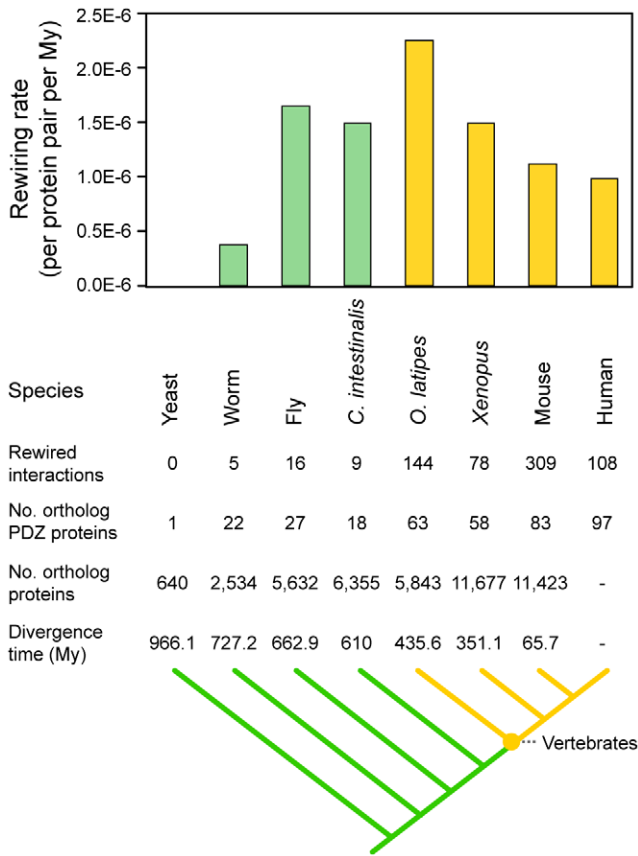
**Figure 4. Network analyses of human PDZ protein-ligand interactions.** (A) Network representation of PDZNet. Orange and blue circles correspond to PDZ proteins and ligands, respectively. The size of the node is proportional to the number of interacting partners. (B) Construction of the PDZ Protein Network (PPN) and the PDZ Ligand Network (PLN). (Center) A subset of PDZ protein-ligand interactions with experimental evidence of physical association. PDZ-binding motifs are presented on the right side of the ligands. (Left) The PPN projection of PDZNet in which two PDZ proteins are connected if they interact with the same ligand. (Right) The PLN projection in which two ligands are connected if they interact with the same PDZ protein. doi:10.1371/journal.pgen.1002510.g004

profile shows the origin of the PDZNet proteins (Figure S6). Many of the human PDZ ligands were not detected in premetazoans, whereas most human PDZ proteins were only found in metazoan species. The binding scores of nearly one-half of the premetazoan orthologs of the human PDZ ligands were less than zero,

indicating that these proteins were not the interaction partners of PDZ proteins in premetazoan species. However, these proteins acquired PDZ-binding motifs in metazoans and began to interact with metazoan PDZ protein partners. For example, EXOC4 is found in yeast and gained its PDZ-binding motif in vertebrates



**Figure 5. Growth of PDZNet by the acquisition of binding motifs.** (A) Two evolutionary models describe the expansion of PDZ domain-ligand interactions. In the gene duplication model, a new PDZ domain-ligand interaction is added by duplication of an existing PDZ ligand. In the sequence mutation model, a new interaction is added by mutations of the C-terminal sequence of the non-PDZ ligand. (B) Paralog fractions of PDZ ligands that share the same PDZ proteins (left) and PDZ proteins that share the same PDZ ligands (right). (C) An example of a PDZ domain-ligand interaction created by sequence mutations. Phylogenetic profiles of NOS1AP and NOS1 are presented. ‘-’ indicates that no ortholog was found in the corresponding species. Four C-terminal residues of NOS1AP orthologs are placed on the right side of the protein. (D) The PWM of NOS1. Four C-terminal residues of vertebrate NOS1AP orthologs (EIAV) are presented in red. ‘Others’ indicates amino acids that were not preferred in the binding pockets. doi:10.1371/journal.pgen.1002510.g005



**Figure 6. Rewiring rates of PDZ domain-ligand interactions throughout evolution.** Invertebrates and vertebrates are colored green and yellow, respectively.  
doi:10.1371/journal.pgen.1002510.g006

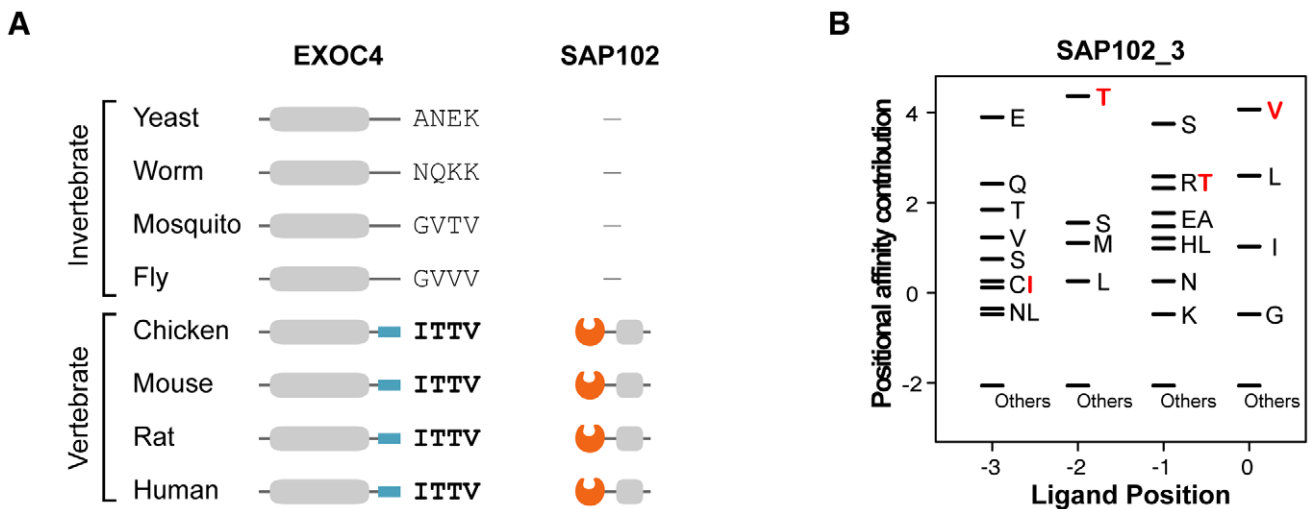
(Figure 7; the multiple sequence alignment of EXOC4 orthologs is shown in Figure S7). The binding of mouse EXOC4 with SAP102 via its PDZ domain was validated in a yeast two-hybrid system and

by pull-down assays [40]. The yeast ortholog of EXOC4 is a component of the exocyst complex, which transports vesicles to the plasma membrane. After gaining a PDZ-binding motif recognized by the SAP102 PDZ domain in vertebrates, it participates in NMDA receptor trafficking [40]. This finding suggests that the evolution of metazoan functions required the rewiring of functional modules that existed in premetazoans and contributed to network growth. Indeed, previous studies have noted that proteins of premetazoan origin played important roles in metazoan-specific functions, such as synaptic signaling [29]. Together, the premetazoan ancestry of PDZ ligands highlights the *de novo* occurrence of PDZ domain-ligand interactions in the rewiring of metazoan evolution.

**Mutations of PDZNet proteins are highly associated with neurological diseases**

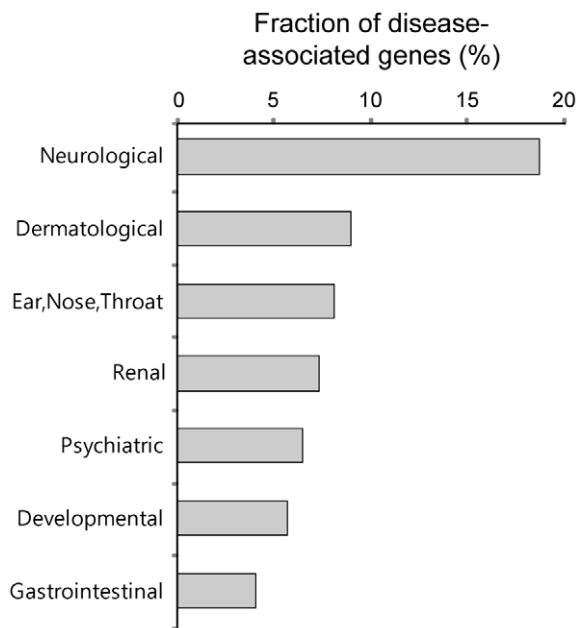
Next, we asked which physiological system was most affected by the mutations of PDZNet proteins. Mutations could affect the binding specificity of PDZ-ligand interactions via the replacement of interfacial residues or the destabilization of PDZ domain and ligand structure. If an interaction gained from the evolution of PDZNet had contributed to the development of a certain physiological system, an alteration of the interaction could be associated with genetic diseases caused by a malfunction of the system.

We investigated the disease associations of the PDZNet components and found that many PDZNet proteins are significantly associated with neurological diseases (Figure 8). Human genetic diseases were mapped to the components of PDZNet using disease-gene association data from the Online Mendelian Inheritance in Man (OMIM) [41]. Genetic diseases were classified into 20 disease classes based on the physiological system affected [42]. We examined whether a certain disease class was more enriched in the PDZNet components than the other proteins in the human interactome. Of the 20 disease classes examined, the neurological disease class was the most highly associated with mutations of the PDZNet components (Table S7). For example, a mutation in the PDZ protein, NLGNX, perturbed its PDZ domain interaction with the ligand protein, SNTG2,



**Figure 7. An example of a PDZ domain-ligand interaction created by sequence mutations.** (A) Phylogenetic profiles of EXOC4 and SAP102 are presented. ‘-’ indicates that no ortholog was found in the corresponding species. Four C-terminal residues of EXOC4 orthologs are placed on the right side of the protein. (B) The PWM of SAP102\_3. Four C-terminal residues of vertebrate EXOC4 orthologs (ITTV) are presented in red. ‘Others’ indicates amino acids that were not preferred in the binding pockets.  
doi:10.1371/journal.pgen.1002510.g007





**Figure 8. Fractions of the proteins in PDZNet whose mutations are associated with specific disease classes.** Disease classes with a  $p$  value less than 0.05 are shown.  
doi:10.1371/journal.pgen.1002510.g008

which is suggested to be a cause of mental retardation and autism [43–45]. This finding reconfirms the importance of PDZ domain-ligand interactions in the evolution of the nervous system. A morbid map of PDZNet components with the classification of genetic diseases is provided in Table S8 as a resource.

## Discussion

In this study, we describe the first PDZ protein-ligand interaction network coupled with quantitative binding strength. Our network approaches elucidated how PDZ domains have diversified their binding partners in the organization of various signaling complexes from receptors to downstream signaling relays. Moreover, we showed that *de novo* evolution of PDZ domain-ligand interactions played an important role in the growth of PDZNet. These findings provide empirical evidence for a network evolution model that highlights the rewiring of interactions as a mechanism of functional innovation.

PDZNet provides information beyond just the state of interaction binding. First, PDZNet provides information regarding the binding interface. High-throughput experiments provided large-scale PPI information; however, the identification of which amino acids were used in the interactions has been difficult. The quantitative model of PDZ domain-ligand interactions provides sequence information on domains and linear motifs, enabling a deeper understanding of the mechanisms involved in their interactions. Second, PDZNet provides the binding strengths of the interactions. The quantitative binding strengths of PDZ domain-ligand interactions enable us to understand the competition among interaction partners for switching between signaling flows.

The multispecificity of PDZ domain-ligand interactions has unique advantages in the evolution of PDZ domain function in the cell signaling network. First, the multispecificity of PDZ domains contributes to the frequent rewiring of PDZ domain-ligand interactions and broadens the extent of recognizable sequences,

thus increasing the chance that a protein gains a suitable sequence to interact with its partners. Indeed, we found that PDZ domain pockets prefer multiple amino acids for interactions. We analyzed amino acid preference patterns from the PWMs of human PDZ domains (Figure S8) and found that the degeneracy of binding motifs facilitate the binding of different PDZ ligands to the same PDZ domain. This finding is consistent with those of a recent study that revealed the specificities of PDZ domains lie on a continuum [18]. Second, the multispecificity of PDZ domains enables the combinatorial assembly of signaling complexes that control signaling processes. PDZ proteins interact with many signaling proteins and form preassembled complexes, which are important for the precision of information flow and the fidelity of cell signaling events [46]. An interesting observation from our network approach is that a PDZ protein is connected to many ligands. These ligands may interact with a PDZ protein in a tissue-specific manner; the subsequent cell type-dependent expression of the PDZ ligands may lead to an alternative assembly of signaling complexes, thus enabling cell type-specific responses for extracellular signals. Indeed, we observed that the ligands of the SAP97 PDZ protein showed tissue-specific expression patterns, allowing the formation of tissue-specific cell signaling complexes (Figure S9). Third, the multispecific interactions of PDZ domains may enhance the robustness of the signaling processes mediated by PDZ domains. The robustness of the cell signaling network is known to increase because several means often exist to achieve one function as the failure of one can be compensated by others [47]. In PDZNet, PDZ domains tend to interact with a series of homologous proteins, particularly cell surface receptors. This interaction may ensure reliable transmission of signals mediated by PDZ proteins to the plasma membrane.

We found that almost one-third of human PDZ ligands obtained their PDZ-binding motifs via C-terminal sequence mutations, providing evolutionary advantages to the PDZ domain-mediated interactions. First, the formation of linear motifs is an efficient mechanism to increase the number of interactions. Emergence of short linear motifs rarely disrupts the protein structure and can be accompanied by few amino acid changes [6]. Second, the *de novo* evolution of interactions via sequence mutation provides an effective means for functional innovation. Gene duplication is known to have a limited role in the molecular innovation of biochemical function but facilitates the modularization of functional networks by specialization [30]. In contrast, the *de novo* evolution of interactions allows connections between evolutionarily unrelated functional modules, thus enabling the reconfiguration of the molecular system. For instance, gain of the PDZ domain-ligand interaction between the EXOC4 PDZ ligand and the SAP102 PDZ protein demonstrated an innovation by bridging two different functional modules. We examined species-specific functional annotations of PDZ ligands and found that yeast EXOC4 participates in vesicle transport with other exocyst complex members, but vertebrate EXOC4 regulates NMDAR transport to the postsynaptic membrane by interacting with the SAP102 PDZ domain [40]. Third, when a PDZ protein gains ligands by sequence mutation, it avoids a loss of fitness caused by an increase in dosage. The *de novo* evolution of PDZ domain-ligand interactions does not increase the copy number of the PDZ ligand genes, avoiding an unfavorable increase in protein concentration. In contrast, gain of interactions by duplication may cause a loss of fitness because proteins that contain linear motifs tend to be intrinsically disordered and dosage sensitive [48].

We were also interested in whether new PDZ domain interaction sites were acquired via C-terminal point mutations or DNA insertions. After careful observation of DNA modifications

in newly acquired PDZ ligands, we found instances of both. For example, protein PBK of *Macaca mulatta* acquired PDZ domain interaction motif “ETDV” via C-terminal point mutations in which a single nucleotide substitution (T→C) changed Ile to Thr and another mutation (C→T) changed the codon for Gln to a stop codon (Figure S10A). On the other hand, EXOC4 acquired new PDZ domain interaction sites via DNA insertion in *Oryzias latipes* (Figure S10B). A large section of DNA inserted near the C-terminus of EXOC4 caused a frame shift mutation, which in turn became the PDZ domain-binding motif “ITTV.”

We found that the rewiring of PDZ domain-ligand interactions most frequently occurred between invertebrates and vertebrates. This massive rewiring may be connected to repeated rounds of whole-genome evolution in ancestral vertebrates. According to Ohno’s model [49], when a gene is duplicated, mutations freely accumulate in the redundant duplicate copy under no selection. Therefore, the duplicate copy has a greater chance of developing new functions without altering existing functions. This evolutionary mechanism may facilitate network rewiring in early vertebrates.

We found that the components of PDZNet are largely associated with neurological diseases. We then asked whether we could identify mutations affecting PDZ-ligand binding, which causes genetic diseases. The disruption of the PDZ domain interaction between PICK1 and GluR7 is known to cause seizures, a chronic neurological disease [50]. Mutations in the C-terminal sequence of GluR7 disrupted its PDZ domain interaction with PICK1. To examine whether our quantitative model can predict the effects of mutations in GluR7, we generated the PWM of the PICK PDZ domain and calculated the binding scores for both the wild-type and mutant forms of GluR7 (Figure S11). We found that the wild type had a high binding score (5.98), and the mutant had a much lower binding score (−0.02). This example illustrates how our method can be applied to characterize genetic diseases that are caused by mutations affecting PDZ domain-ligand interactions.

An important issue of the present biological network study is its incompleteness [51]. PDZNet has room for improvement regarding network coverage in two respects: shortage of nodes and links in the current network. To test whether the conclusions obtained in this work are sufficiently robust with regard to both, we constructed smaller random networks from PDZNet and repeated the analyses. In each trial, 20% of the proteins or interactions were randomly removed from PDZNet. We found that in all tests, the overall organization of the rescaled PDZNet remained largely unaltered, and the conclusions and the differences between the paralog fractions of the PDZ proteins and ligands were retained (Figures S12, S13, S14, S15, S16, S17), supporting the robustness of our findings to the future expansion of PDZ domain-ligand interactions.

Due to the incompleteness of the interactome networks, expansion of network coverage is of significant value. PDZ domain-ligand interactions were relatively difficult to detect using current experimental techniques because transient interactions are often lost during experimental washing steps. Furthermore, a PDZ domain-ligand interaction often depends on phosphorylation [24], so it can be missed when screening for protein interactions performed in a single condition. Therefore, many PDZ domain-ligand interactions remain to be discovered. We anticipate that putative PDZ domain-ligand interactions with high-binding scores from PWMs, expression correlations, and similar phylogenetic profiles may be used to uncover novel interactions. Therefore, we provide a candidate list of PDZ domain-ligand interactions to assist in the discovery of novel PDZ domain-ligand interactions ([sbi.postech.ac.kr/pdz](http://sbi.postech.ac.kr/pdz)).

## Materials and Methods

### Experimentally confirmed PDZ domain-ligand interactions

We assembled experimentally confirmed PDZ domain-ligand interactions from various data sources. In detail, we obtained PDZ domain-peptide binding data from a high-throughput binding assay between 81 mouse PDZ domains and 217 peptides derived from genome-encoded receptors by protein array [18]. We collected *in vivo* PDZ domain-ligand interactions from the published literature, including peptide binding data of *Drosophila* INAD [22,23] and human AF-6 [52] PDZ domains. Additionally, a PDZ domain-ligand interaction database, PDZBase [21], which currently lists 339 *in vivo* interactions between 145 PDZ domains and 217 ligands, was used. Finally, we obtained 54 human and 28 worm PDZ domains in a high-throughput binding assay [20] and four N-terminal PDZ domains of human INADL using phage display [19]. The collection of these data resulted in 4,467 experimentally confirmed PDZ domain-ligand interactions.

### Human PDZ domains

We collected 563 human PDZ domain sequences from the Pfam repository [53]. After eliminating redundancy, we obtained 268 sequences. We then examined pocket residues of the PDZ domains using hidden Markov model (HMM) alignment, removed the sequences that did not align in the pocket region, and finally obtained 241 distinct human PDZ domains.

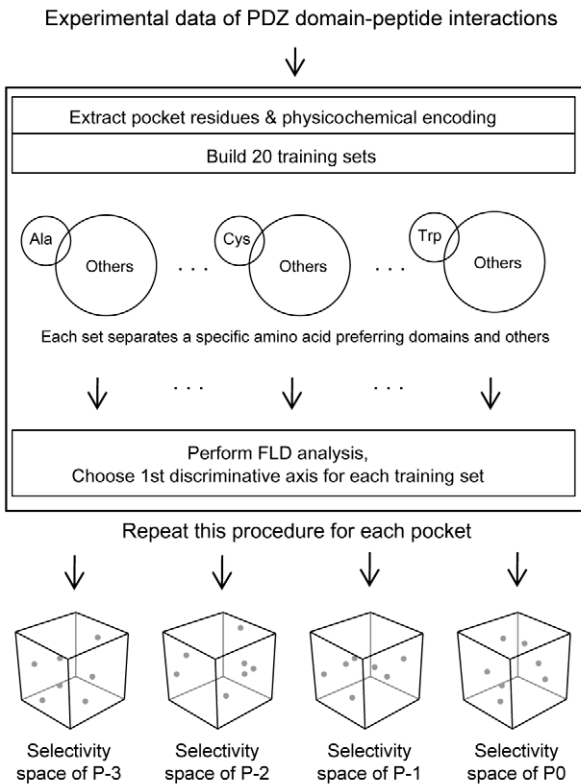
### Building a quantitative model of PDZ domain-ligand interactions

We developed a two-step approach to quantify the strength of binding between the PDZ domains and ligands. Using this approach, the binding affinity between each PDZ pocket and its corresponding ligand position was predicted individually based on the idea that the contribution of each ligand position to the binding affinity is additive [54], which is a widely accepted view in the modeling of linear motif interactions [55,56]. The workflow of our approach is summarized in Figure 9.

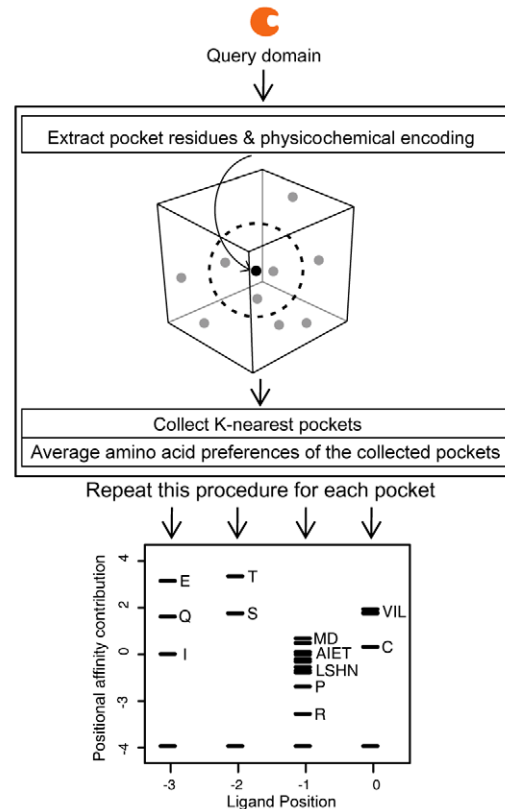
In the first step, we designed the selectivity space of each pocket (Figure 9, left panel) to contain 20 axes, representing preferences for the corresponding amino acids in the peptide ligand (Figure S18). To obtain the ligand selectivity of the PDZ domains, three types of interaction data were used, namely protein arrays of mouse PDZ domains against synthesized peptides [18], collections of individual studies of PDZ domain-ligand interactions [21–23], and high-throughput binding assays using phage display [19]. We made a multiple sequence alignment (MSA) of PDZ domains using a HMM for the PDZ domain. We then extracted pocket residues from the MSA and encoded them into feature vectors based on their physicochemical properties. With the feature vectors, we constructed 20 training sets. In each training set, the feature vectors from specific amino acid-preferring pockets were used as a positive set, and the remainder was used as a negative set. We then applied Fisher’s Linear Discriminant (FLD) analysis to these training sets such that discriminative axes were trained to distinguish specific amino acid-preferring pockets, resulting in a projection matrix composed of the axes’ direction vectors. By multiplying the feature vectors with the projection matrix, we located the pockets of PDZ domains in the selectivity space. Thus, the selectivity spaces for each pocket capture intrinsic amino acid preferences from binary interaction data.

In the second step, to build a PWM of a query PDZ domain, we generated an affinity profile that represents the relative affinity contributions of 20 amino acids to the PDZ domain pocket

## Training step



## Prediction step



**Figure 9. Construction of a quantitative model of PDZ domain-ligand interactions and generation of PWMs of human PDZ domains.** (Left) Training step. The procedure to build selectivity spaces for the four pockets of the PDZ domain. From multiple sequence alignments of PDZ domain sequences, pocket residues were extracted and converted into feature vectors. The feature vectors were alternatively assembled into 20 training sets, each comprising a specific amino acid-preferring group (positive set) and the remainder (negative set). Groups are represented by circles, and the preferred amino acids are shown within the circles. FLD analysis was performed on these training sets to generate 20 selectivity axes that assemble into a pocket selectivity space. Pockets are shown as spheres in the selectivity space. This procedure was repeated for each pocket, resulting in four selectivity spaces that correspond to each ligand position. We note that the selectivity “dots” have 20 dimensions but are represented by three-dimensional cubes for convenience. (Right) Prediction step. The procedure to build a PWM of the query domain. Pocket residues were extracted from the query domain and converted into a feature vector. This feature vector was projected on the selectivity space. The nearest 40 pockets from the query were collected, and their amino acid preferences were averaged. The averaged preferences were then converted into affinity contribution profiles. This procedure was repeated for each pocket, producing a PWM. doi:10.1371/journal.pgen.1002510.g009

(Figure 9, right panel). Based on the assumption that closely residing pockets in the selectivity space are similar in their amino acid preferences, we gathered the nearest neighbors of a query domain in a selectivity space to establish an affinity profile from their preferred amino acid sets. Pocket residues of a query PDZ domain were encoded into feature vectors using physicochemical properties and then located on a selectivity space using the projection matrix described above. We gathered ligand sets preferred by the nearest neighbors of the query pocket and estimated the binding affinity contributions of each position.

## Generating feature vectors

We converted the pocket residue sequences of a PDZ domain into vector representations by replacing all 20 amino acids with 10 physicochemical properties (amino acid indices) that describe the number of hydrogen bond donors [57], polarity [58], volume [58], bulkiness [59], hydrophobicity [59,60], isoelectric point [59], positive charge [57], negative charge [57], electron ion interaction potential [61], and free energy in water [62]. We normalized the values such that the standard deviation is 1 and the average is 0.

## Extraction of pocket residues

Our goal was to predict the specificities of a PDZ domain without knowledge of its structure. As such, a method to extract pocket residues from the sequences of PDZ domains was designed. To identify the positions of pocket residues within the PDZ domain sequence, an MSA was constructed, and the known structure of the PSD-95\_1 domain was referenced. We performed a multiple alignment of the PDZ domain sequences using a HMM [63] and an HMM that was optimized for PDZ domains from Pfam [53] and aligned the secondary structure profile of the PSD-95\_1 domain with the sequence alignment. Pocket residues were subsequently extracted according to the pocket definitions described in Wiedemann et al. [26] (Figure S19).

## Position weight matrix

To estimate the PDZ domain-ligand binding affinity, we adopted an information theory-based PWM method that is widely used to estimate protein-DNA binding affinities [64,65]. In each selectivity space, 40 preferred amino acids of the neighboring pockets of the query were gathered. A PWM was calculated using

the four sets of collected amino acids in which amino acid frequencies were calculated at each ligand position; these frequencies were compared to the background frequency that we expected to observe for the C-terminal sequences of the ligands.

$$PWM(a,i) = \log_2 \frac{k f_{a,i}}{p_{a,i}}$$

$PWM(a, i)$  is the affinity contribution of amino acid  $a$  at the  $i$ th position,  $f_{a,i}$  is the frequency of amino acid  $a$  at the  $i$ th position in the collected set, and  $p_{a,i}$  is the background frequency, defined as the probability of observing amino acid  $a$  at the  $i$ th position in any ligand protein. The constant  $k$  was empirically determined to be 1.921, so that PDZNet includes all experimentally confirmed PDZ domain-ligand interactions as positive binding scores.

A PWM was used to calculate the binding score of a potential interaction partner with a given sequence by summing the corresponding amino acids for the affinity contribution of each position. The binding score of each peptide was calculated according to the following formula:

$$binding\ score = \sum_{i=-3}^0 PWM(S_i, i)$$

where  $PWM(S_i, i)$  is the affinity contribution of the amino acid  $S_i$  at the  $i$ th position in the matrix and  $S_i$  is the amino acid at the  $i$ th position of the peptide.

### Affinity values of 5,257 peptides

Affinity values of the 5,257 peptides against both the SNA1 and ERBIN PDZ domains were obtained from Wiedemann et al. [26], who assessed the affinity values of the peptides with a combination of experiments (i.e., surface plasmon resonance and Boehringer light unit) and statistical analyses.

### Domain-wise cross-validation to identify known *in vivo* binding partners of PDZ domains

To evaluate the performance of our method, we measured the ability to identify the 217 known binding partners of 145 PDZ domains in the PDZBase [21]. Using a standard leave-one-out procedure, we generated PWMs and genome-wide rank lists of interaction candidates for each domain using their corresponding PWMs. Our method successfully predicts the binding partners of PDZ domains for which no interaction data are available. When we examined the percentile ranks of experimentally confirmed PDZ domain-ligand interactions, most were enriched at high positions in the rank lists (90<sup>th</sup>~100<sup>th</sup> percentile of the binding score; Figure 3C).

### Construction of the protein interaction network

We compiled human protein interactions from a total of 22 existing protein interaction databases: the Bio-molecular Interaction Network Database (BIND), the Human Protein Reference Database (HPRD), the Molecular Interaction database (MINT), DIP, IntAct, BioGRID, Reactome, the Protein-Protein Interaction Database (PPID), BioVerse, CCS-HI1, the comprehensive resource of mammalian protein complexes (CORUM), IntNetDB, the Mammalian Protein-Protein Interaction Database (MIPS), the Online Predicted Human Interaction Database (OPHID), Ottawa, PC/Ataxia, Sager, Transcriptome, Complexex, Unilever, protein-protein interaction database for PDZ-domains (PDZBase), and a protein interaction dataset from the literature [66]. We

removed low-confidence interactions that were not supported by direct experimental evidence. The final network comprises 101,777 interactions between 11,043 human proteins.

### Construction of the PDZ domain-mediated interaction set

We collected all physical interactions mediated by the PDZ proteins from the integrated PPI network. This PDZ protein-mediated interaction set may have some interactions that are mediated by interaction domains other than PDZ domains, because many PDZ proteins have various domains other than PDZ domains. Therefore, we removed such interactions that were connected by domain-domain interactions rather than PDZ domain-ligand interactions. First, we confirmed that PDZ domain-mediated interactions are rarely augmented by other interaction domains. We found that domain-domain interactions are not present in the experimentally confirmed PDZ protein-ligand interactions from the PDZBase [21]. Furthermore, we found that domain-domain interactions are only enriched in low-scoring PDZ protein interactions (Figure S20). Based on these observations, we removed domain-domain interactions from PDZNet.

We also removed interactions that could be mediated by other peptide-binding domains, such as SH3 and WW domains, rather than PDZ domains. We searched the known peptide-binding motifs and removed interactions mediated by peptide-binding domains that had low binding scores. The cut-off binding score was set to the lowest binding score of the experimentally confirmed PDZ domain-peptide interactions from the PDZBase [21]. The binding score represents the predicted binding strength between a PDZ domain and the C-terminal sequence of its partner. Subcellular localization information was taken from Swiss Prot and consensus localization annotations [67].

### Measurement of the rewiring rates of PDZ domain-ligand interactions

Let two species,  $i$  and  $j$ , be in a common tree with humans, and species  $i$  is more distant from humans. If a human PDZ interaction is absent in species  $i$  and present in species  $j$ , we define the interaction as rewired. Thus, the rewiring occurred during the time interval between the emergence of species  $i$  and  $j$ . We also consider that all proteins in species  $i$  could be rewired to PDZ proteins in species  $j$ . Thus, we define the rewiring rate as the following:

$$r = \frac{n_j}{((t_j - t_i) \times p_{all,i} \times p_{pdz,j})}$$

where  $n_j$  is the number of rewired interaction found in species  $j$ ;  $t$  is the divergence time from human;  $p_{all,i}$  is the number of proteins orthologous to human protein in species  $i$ ; and  $p_{pdz,j}$  is the number of proteins orthologous to human PDZ proteins. Divergent time was obtained from the timetree (<http://www.timetree.net>).

To analyze the interactions between orthologous PDZ domains, we calculated the binding scores of the C-terminal sequences of orthologous PDZ ligands and the predicted PWMs with orthologous PDZ domain sequences.

### Gene ontology analysis of the rewiring events in PDZNet

We examined whether particular protein functions were enriched for protein categories that were defined based on the time of protein emergence and PDZ-binding motif acquisition. We systematically classified PDZ ligands into two categories: (1)

proteins arose in invertebrates and acquired PDZ domain interaction sites in vertebrates; (2) proteins arose and acquired PDZ domain interaction sites in invertebrates; we then analyzed the overrepresented functional terms of each group. We used DAVID [68] for gene set enrichment analysis. All ligand proteins were used as background.

### Analysis of disease associations of PDZNet proteins

Mutations of PDZNet proteins were mapped to genetic diseases using disease-gene association databases from OMIM. The OMIM database lists gene-disease associations between 2,929 disease types defined by Morbid Map (MM) and 1,777 genes associated with particular disease types. Disease types were further categorized into 1,340 distinct diseases by joining disease subtypes into a single disease if similar disease names were used. These disease types were further classified into 20 disease classes based on the physiological system affected [42]. The  $p$  values for over- or under-representation of the disease-associated genes in PDZNet were obtained using a hypergeometric distribution. We independently calculated the probability of the disease-associated genes in each class.

### Web server

We created a user-friendly web service that provides a PWM and rank list of interaction candidates of a given PDZ domain sequence (Figure S21). The automated pipeline of the web service extracts pocket residues from the query PDZ domain sequence, predicts binding specificity (represented as a PWM), and generates a genome-wide rank list of potential ligands. The web service can handle various exceptions. For example, if a query is an incorrect PDZ domain sequence or an incorrect alignment was made in the pocket residues, the web service provides messages with explanations.

### Supporting Information

**Figure S1** Comparisons of the quantitative model- and phage display data-derived PWMs of MAGI1\_2, DLG1\_2, and PTN13\_2. (PDF)

**Figure S2** Distribution of the interaction partners of 97 human PDZ proteins. The maximum number of ligands per PDZ protein is 102. The average interaction partner of the human PDZ protein is 12. (PDF)

**Figure S3** Network representation of domain-level interactions in PDZNet (best viewed by magnification in a PDF viewer). Domain numbers are presented on the right side of the PDZ protein names with a delimiter ('\_'). The network is composed of 2,643 interactions between 190 PDZ domains and 593 ligands. (PDF)

**Figure S4** Dendrogram of PDZ domains based on the identity of pocket residues. Domain numbers are presented on the right side of the PDZ protein names with a delimiter ('\_'). (PDF)

**Figure S5** Relationship between specificity determining residue (SDR) identity and PWM similarity. Each point represents an orthologous PDZ domain pair. (PDF)

**Figure S6** Phylogenetic profile of human PDZ proteins and ligands across 13 fully sequenced species. The presence (yellow)

and absence (black) of orthologs for the 104 PDZ proteins and 554 PDZ ligands are presented. (PDF)

**Figure S7** Multiple sequence alignment (MSA) of EXOC4 orthologs. The MSA was generated using Muscle with default options. C-terminal PDZ binding motifs are shown in bold. (PDF)

**Figure S8** Amino acid preference patterns of human PDZ domain pockets. (A–D) Clustering of amino acid preference profiles of 241 human PDZ domain pockets is shown. (PDF)

**Figure S9** Alternative expression of SAP97 ligands across three human tissues. The protein expression levels of SAP97 PDZ protein and its 13 ligands were compared across brain, bone, and epidermis. Protein expression was measured by quantitative mass spectrometry [69]. The protein abundance ratio was defined by the normalized mass spectrometry intensity value relative to the maximum intensity per protein. (PDF)

**Figure S10** Types of DNA modifications that gain PDZ-binding motifs. (A) A point mutation generated a PDZ-binding motif in the C-terminal amino acids of the *Macaca mulatta* PBK protein. The binding motif is highlighted in the PWM of SAP97\_1 (right). Mutations in the PDZ-binding motif are shown in the alignment of DNA sequences (bottom). (B) A DNA segment insertion generated a PDZ-binding motif in the C-terminal amino acids of the *Oryzias latipes* EXOC4 protein. The binding motif is highlighted in the PWM of SAP102\_1 (right). The inserted DNA segment is shown in the alignment of DNA sequences (bottom). (PDF)

**Figure S11** Mutation effects of the C-terminal GluR7 sequence. (A) C-terminal sequences and binding scores of wild-type and mutation forms of GluR7. (B) The PWM of the PICK1 PDZ domain. Four C-terminal residues of wild-type GluR7 are highlighted. (PDF)

**Figure S12** Repeated analysis of PDZNet by randomly removing 20% of proteins (trial 1). (A) Network representation of PDZNet. (B) Paralog fractions of PDZ ligands that share the same PDZ proteins (left) and PDZ proteins that share the same PDZ ligands (right). (PDF)

**Figure S13** Repeated analysis of PDZNet by randomly removing 20% of proteins (trial 2). (A) Network representation of PDZNet. (B) Paralog fractions of PDZ ligands that share the same PDZ proteins (left) and PDZ proteins that share the same PDZ ligands (right). (PDF)

**Figure S14** Repeated analysis of PDZNet by randomly removing 20% of proteins (trial 3). (A) Network representation of PDZNet. (B) Paralog fractions of PDZ ligands that share the same PDZ proteins (left) and PDZ proteins that share the same PDZ ligands (right). (PDF)

**Figure S15** Repeated analysis of PDZNet by randomly removing 20% of interactions (trial 4). (A) Network representation of PDZNet. (B) Paralog fractions of PDZ ligands that share the same PDZ proteins (left) and PDZ proteins that share the same PDZ ligands (right). (PDF)

**Figure S16** Repeated analysis of PDZNet by randomly removing 20% of interactions (trial 5). (A) Network representation of PDZNet. (B) Paralog fractions of PDZ ligands that share the same PDZ proteins (*left*) and PDZ proteins that share the same PDZ ligands (*right*). (PDF)

**Figure S17** Repeated analysis of PDZNet by randomly removing 20% of interactions (trial 6). (A) Network representation of PDZNet. (B) Paralog fractions of PDZ ligands that share the same PDZ proteins (*left*) and PDZ proteins that share the same PDZ ligands (*right*). (PDF)

**Figure S18** Discriminating power of selectivity axes. Each boxplot shows distributions of binders and non-binders of an amino acid, which are presented at the top of the plot. Binders are PDZ domain pockets that prefer the amino acid, and non-binders are those domain pockets that do not prefer the amino acid. The vertical axis corresponds to an axis of a selectivity space. Fisher's score (FS) is presented at the top of each plot, indicating the discriminating power of the selectivity axes. (PDF)

**Figure S19** Procedure for extracting pocket residues. (A) Schematic drawing of the PSD-95\_1 domain structure. (B) Position of each pocket residue on the structure. (C) The MSA of three representative PDZ domains was constructed using the hidden Markov model that was optimized for the PDZ domain. By adjusting the secondary structural profile on the MSA, the positions of pocket residues were identified. Gray boxes indicate the positions of pocket residues. (PDF)

**Figure S20** Fraction of domain-domain interactions according to the binding scores of all PDZ protein-mediated interactions. The PDZ protein-mediated interactions were binned based on binding score. The fraction of domain-domain interactions were measured for each bin. (PDF)

**Figure S21** Flow chart of web server and a sample output. The web server takes a query PDZ domain sequence and a species

name. The outputs are pocket residues, a PWM of the query PDZ domain, and a genome-wide rank list of proteins from the species chosen by the user. (PDF)

**Table S1** Position weight matrices (PWMs) for 515 human PDZ domains. For resource purposes, homologous PDZ domains are included in the list. (XLS)

**Table S2** Validation of PWMs on *in vivo* partners derived from various species. (XLS)

**Table S3** PDZ domain-ligand interactions in PDZNet. (XLS)

**Table S4** C-terminal sequences of human PDZ ligand orthologs. (XLS)

**Table S5** Experimental evidence of human PDZ domain-ligand interactions that emerged via sequence mutations. '—' indicates the absence of an ortholog (XLS)

**Table S6** Over-represented gene ontology (GO) terms of PDZNet proteins based on the time point of acquiring PDZ domain interaction sites. (XLS)

**Table S7** Disease classes associated with mutations of PDZNet components. (XLS)

**Table S8** A morbid map of PDZNet components with the classification of genetic diseases. (XLS)

## Author Contributions

Conceived and designed the experiments: JK SK. Performed the experiments: JK IK J-SY SP. Analyzed the data: JK IK J-SY Y-ES JH SP YSC SK. Wrote the paper: JK IK SK.

## References

- Good MC, Zalatan JG, Lim WA (2011) Scaffold proteins: hubs for controlling the flow of cellular information. *Science* 332: 680–686.
- Sakarya O, Conaco C, Eggecioglu O, Solla SA, Oakley TH, et al. (2010) Evolutionary expansion and specialization of the PDZ domains. *Mol Biol Evol* 27: 1058–1069.
- Jim J, Xie X, Chen C, Park JG, Stark C, et al. (2009) Eukaryotic protein domains as functional units of cellular evolution. *Sci Signal* 2: ra76.
- Beltrao P, Serrano L (2007) Specificity and evolvability in eukaryotic protein interaction networks. *PLoS Comput Biol* 3: e25. doi:10.1371/journal.pcbi.0030025.
- He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169: 1157–1164.
- Neduva V, Russell RB (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett* 579: 3342–3345.
- Wang X, Sommer RJ (2011) Antagonism of LIN-17/Frizzled and LIN-18/Ryk in Nematode Vulva Induction Reveals Evolutionary Alterations in Core Developmental Pathways. *PLoS Biol* 9: e1001110. doi: 10.1371/journal.pbio.1001110.
- Tan CS, Bodenmiller B, Pasulescu A, Jovanovic M, Hengartner MO, et al. (2009) Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci Signal* 2: ra39.
- Beltrao P, Trinidad JC, Fiedler D, Roguev A, Lim WA, et al. (2009) Evolution of phosphoregulation: comparison of phosphorylation patterns across yeast species. *PLoS Biol* 7: e1000134. doi: 10.1371/journal.pbio.1000134.
- Kim DS, Hahn Y (2011) Identification of novel phosphorylation modification sites in human proteins that originated after the human-chimpanzee divergence. *Bioinformatics*.
- Aloy P, Russell RB (2006) Structural systems biology: modelling protein interactions. *Nat Rev Mol Cell Biol* 7: 188–197.
- Kim PM, Lu LJ, Xia Y, Gerstein MB (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* 314: 1938–1941.
- Mosca R, Pons C, Fernandez-Recio J, Aloy P (2009) Pushing structural information into the yeast interactome by high-throughput protein docking experiments. *PLoS Comput Biol* 5: e1000490. doi:10.1371/journal.pcbi.1000490.
- Stein A, Aloy P (2008) Contextual specificity in peptide-mediated protein interactions. *PLoS ONE* 3: e2524. doi:10.1371/journal.pone.0002524.
- Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5: 101–113.
- Ramani AK, Li Z, Hart GT, Carlson MW, Boutz DR, et al. (2008) A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. *Mol Syst Biol* 4: 180.
- Lage K, Karlberg EO, Stirling ZM, Olason PI, Pedersen AG, et al. (2007) A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* 25: 309–316.
- Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, et al. (2007) PDZ domain binding selectivity is optimized across the mouse proteome. *Science* 317: 364–369.
- Vaccaro P, Brannetti B, Montecchi-Palazzi L, Philipp S, Helmer Citterich M, et al. (2001) Distinct binding specificity of the multiple PDZ domains of INADL, a human protein with homology to INAD from *Drosophila melanogaster*. *J Biol Chem* 276: 42122–42130.
- Tonikian R, Zhang Y, Szazinsky SL, Currell B, Yeh JH, et al. (2008) A specificity map for the PDZ domain family. *PLoS Biol* 6: e239. doi:10.1371/journal.pbio.0060239.

21. Beuming T, Skrabanek L, Niv MY, Mukherjee P, Weinstein H (2005) PDZBase: a protein-protein interaction database for PDZ-domains. *Bioinformatics* 21: 827–828.
22. Wes PD, Xu XZ, Li HS, Chien F, Doberstein SK, et al. (1999) Termination of phototransduction requires binding of the NINAC myosin III and the PDZ protein INAD. *Nat Neurosci* 2: 447–453.
23. van Huizen R, Miller K, Chen DM, Li Y, Lai ZC, et al. (1998) Two distantly positioned PDZ domains mediate multivalent INAD-phospholipase C interactions essential for G protein-coupled signaling. *EMBO J* 17: 2285–2297.
24. Harris BZ, Lim WA (2001) Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci* 114: 3219–3231.
25. Gisler SM, Kittanakom S, Fuster D, Wong V, Bertic M, et al. (2008) Monitoring protein-protein interactions between the mammalian integral membrane transporters and PDZ-interacting partners using a modified split-ubiquitin membrane yeast two-hybrid system. *Mol Cell Proteomics* 7: 1362–1377.
26. Wiedemann U, Boisguerin P, Leben R, Leitner D, Krause G, et al. (2004) Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J Mol Biol* 343: 703–718.
27. Piserchio A, Pellegrini M, Mehta S, Blackman SM, Garcia EP, et al. (2002) The PDZ1 domain of SAP90. Characterization of structure and binding. *J Biol Chem* 277: 6967–6973.
28. Gfeller D, Butty F, Wierzbicka M, Verschuere E, Vanhee P, et al. (2011) The multiple-specificity landscape of modular peptide recognition domains. *Mol Syst Biol* 7: 484.
29. Emes RD, Pocklington AJ, Anderson CN, Bayes A, Collins MO, et al. (2008) Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nat Neurosci* 11: 799–806.
30. Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449: 54–61.
31. Jaffrey SR, Snowman AM, Eliasson MJ, Cohen NA, Snyder SH (1998) CAPON: a protein associated with neuronal nitric oxide synthase that regulates its interactions with PSD95. *Neuron* 20: 115–124.
32. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, et al. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 32: D449–451.
33. Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, et al. (2011) The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res* 39: D698–704.
34. Yu J, Pacifico S, Liu G, Finley RL, Jr. (2008) DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC genomics* 9: 461.
35. Yamada T, Bork P (2009) Evolution of biomolecular networks: lessons from metabolic and protein interactions. *Nat Rev Mol Cell Biol* 10: 791–803.
36. Radziwill G, Erdmann RA, Margelisch U, Moelling K (2003) The *Bcr* kinase downregulates Ras signaling by phosphorylating AF-6 and binding to its PDZ domain. *Mol Cell Biol* 23: 4663–4672.
37. Zhou H, Xu Y, Yang Y, Huang A, Wu J, et al. (2005) Solution structure of AF-6 PDZ domain and its interaction with the C-terminal peptides from Neurexin and *Bcr*. *J Biol Chem* 280: 13841–13847.
38. Elowe S, Holland SJ, Kulkarni S, Pawson T (2001) Downregulation of the Ras-mitogen-activated protein kinase pathway by the EphB2 receptor tyrosine kinase is required for ephrin-induced neurite retraction. *Mol Cell Biol* 21: 7429–7441.
39. Noren NK, Pasquale EB (2004) Eph receptor-ephrin bidirectional signals that target Ras and Rho proteins. *Cell Signal* 16: 655–666.
40. Sans N, Prybylowski K, Petralia RS, Chang K, Wang YX, et al. (2003) NMDA receptor trafficking through an interaction between PDZ proteins and the exocyst complex. *Nat Cell Biol* 5: 520–530.
41. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 33: D514–517.
42. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, et al. (2007) The human disease network. *Proc Natl Acad Sci U S A* 104: 8685–8690.
43. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, et al. (2003) Mutations of the X-linked genes encoding neurologins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 34: 27–29.
44. Yamakawa H, Oyama S, Mitsuhashi H, Sasagawa N, Uchino S, et al. (2007) Neurologins 3 and 4X interact with syntrophin-gamma2, and the interactions are affected by autism-related mutations. *Biochem Biophys Res Commun* 355: 41–46.
45. Laumonnier F, Bonnet-Brihault F, Gomot M, Blanc R, David A, et al. (2004) X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet* 74: 552–557.
46. Scott JD, Pawson T (2009) Cell signaling in space and time: where proteins come together and when they're apart. *Science* 326: 1220–1224.
47. Kitano H (2004) Biological robustness. *Nat Rev Genet* 5: 826–837.
48. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* 138: 198–208.
49. Ohno S, Wolf U, Atkin NB (1968) Evolution from fish to mammals by gene duplication. *Hereditas* 59: 169–187.
50. Bertaso F, Zhang C, Scheschonka A, de Bock F, Fontanaud P, et al. (2008) PICK1 uncoupling from mGluR7a causes absence-like seizures. *Nat Neurosci* 11: 940–948.
51. Vidal M, Cusick ME, Barabasi AL (2011) Interactome networks and human disease. *Cell* 144: 986–998.
52. Hock B, Bohme B, Karn T, Yamamoto T, Kaibuchi K, et al. (1998) PDZ-domain-mediated interaction of the Eph-related receptor tyrosine kinase EphB3 and the ras-binding protein AF6 depends on the kinase activity of the receptor. *Proc Natl Acad Sci U S A* 95: 9779–9784.
53. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* 26: 320–322.
54. Ranganathan R, Ross EM (1997) PDZ domain proteins: scaffolds for signaling complexes. *Curr Biol* 7: R770–773.
55. Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G (2008) Predicting PDZ domain-peptide interactions from primary sequences. *Nat Biotechnol* 26: 1041–1045.
56. Yaffe MB, Leparc GG, Lai J, Obata T, Volinia S, et al. (2001) A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat Biotechnol* 19: 348–353.
57. Fauchere JL, Charton M, Kier LB, Verloop A, Pliska V (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int J Pept Protein Res* 32: 269–278.
58. Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185: 862–864.
59. Zimmerman JM, Eliezer N, Simha R (1968) The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol* 21: 170–201.
60. Argos P, Rao JK, Hargrave PA (1982) Structural prediction of membrane-bound proteins. *Eur J Biochem* 128: 565–575.
61. Cosic I (1994) Macromolecular bioactivity: is it resonant interaction between macromolecules?—Theory and applications. *IEEE Trans Biomed Eng* 41: 1101–1114.
62. Charton BI (1982) The structural dependence of amino acid hydrophobicity parameters. *J Theor Biol* 99: 629–644.
63. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14: 755–763.
64. D'Haeseleer P (2006) What are DNA sequence motifs? *Nat Biotechnol* 24: 423–425.
65. Stormo GD, Hartzell GW, 3rd (1989) Identifying protein-binding sites from unaligned DNA fragments. *Proc Natl Acad Sci U S A* 86: 1183–1187.
66. Bromberg KD, Ma'ayan A, Neves SR, Iyengar R (2008) Design logic of a cannabinoid receptor signaling network that triggers neurite outgrowth. *Science* 320: 903–909.
67. Park S, Yang J-s, Shin Y-e, Park J, Jang SK, et al. (2011) Protein localization as a principal feature of the etiology and comorbidity of genetic diseases. *Molecular Systems Biology* 7: 1–11.
68. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44–57.
69. Lundberg E, Fagerberg L, Klevebring D, Matic I, Geiger T, et al. (2010) Defining the transcriptome and proteome in three functionally different human cell lines. *Mol Syst Biol* 6: 450.