



OPEN

Metazoans evolved by taking domains from soluble proteins to expand intercellular communication network

SUBJECT AREAS:
MOLECULAR EVOLUTION
GENOME INFORMATICSHyun-Jun Nam¹, Inhae Kim², James U. Bowie³ & Sanguk Kim^{1,2}Received
17 November 2014Accepted
9 March 2015Published
29 April 2015Correspondence and
requests for materials
should be addressed to
S.K. (sukim@postech.
ac.kr)

¹School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang, 790-784, Korea, ²Department of Life Sciences, Pohang University of Science and Technology, Pohang, 790-784, Korea, ³Department of Chemistry and Biochemistry, UCLA-DOE Institute of Genomics and Proteomics, Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90095-1570, United States.

A central question in animal evolution is how multicellular animals evolved from unicellular ancestors. We hypothesize that membrane proteins must be key players in the development of multicellularity because they are well positioned to form the cell-cell contacts and to provide the intercellular communication required for the creation of complex organisms. Here we find that a major mechanism for the necessary increase in membrane protein complexity in the transition from non-metazoan to metazoan life was the new incorporation of domains from soluble proteins. The membrane proteins that have incorporated soluble domains in metazoans are enriched in many of the functions unique to multicellular organisms such as cell-cell adhesion, signaling, immune defense and developmental processes. They also show enhanced protein-protein interaction (PPI) network complexity and centrality, suggesting an important role in the cellular diversification found in complex organisms. Our results expose an evolutionary mechanism that contributed to the development of higher life forms.

Transition from non-metazoa to multicellular animals is a pivotal event in the history of life. The evolution of multicellularity requires the development of stable cell adhesion and communication^{1,2} and the division of labor among different cell types^{3,4}. These developments enable enormous functional innovation, such as the immune system, the nerve system, and complex developmental processes^{5,6}. Indeed, comparative genomic analysis reveals dramatic increases in cell-adhesion receptors and extracellular matrix (ECM) associated proteins during metazoan genomic evolution⁷. Nevertheless, the evolutionary mechanisms that led to the rapid emergence of the genes required for the development of complex cellular interactions remain poorly understood.

The evolution of membrane proteins is an obvious place to look for mechanistic basis of the diversification that seen in the transition to multicellular life as they are directly positioned to interaction with other cells⁸. Yet membrane protein evolution is known to be constrained in several ways. In particular, the hydrophobic environment imposed by lipid bilayers restricts the amino acid composition and structural diversity of membrane proteins⁹ and the rate of divergence is constrained by the high level side chain burial in the transmembrane regions¹⁰. Furthermore, domain recombination, a major mechanism of soluble protein diversification^{11,12}, is not common for the transmembrane domains of membrane proteins¹³. How then did membrane proteins undergo the revolution in functional diversification required for the evolution of multicellular organisms?

Recently, we discovered membrane proteins do employ recombination as a major mechanism of diversification, but rather than exchanging parts between membrane proteins, they efficiently exchange domains with soluble proteins¹⁴. Thus, we reasoned that domain exchanges between membrane and soluble proteins at the extra-membrane region may have been a key factor in metazoan evolution.

Here, we examine the functional expansion of membrane proteins during the evolution of metazoan species. We found that membrane proteins frequently recruit domains from soluble proteins in metazoan species. Moreover, newly incorporated soluble domains became particularly important players in intercellular PPI network. Especially, they are enriched in functions critical for multicellularity, such as cell-adhesion, immune and developmental processes. Our results suggest that domain sharing between membrane and soluble proteins was a major mechanism for generating the panoply of proteins required for cellular cooperation in metazoans.



Results

Domain sharing between membrane and soluble proteins. To investigate functional expansion of membrane proteins during evolution, we identified (i) “*membrane protein domains*” that are found in membrane proteins, (ii) “*soluble protein domains*” that are found in soluble proteins (iii) “*shared domains*” that are found in both membrane and soluble proteins from each non-metazoan and metazoan genomes. We first classified membrane and soluble proteins from complete genomes of 5 non-metazoan and 5 metazoan species using the UniProt database¹⁴, and assigned domains into membrane and soluble proteins by using profile-HMMs (HMMERs) of Pfam database (see methods and materials). Table 1 shows the numbers of shared, membrane, and soluble protein domains of non-metazoan and metazoan species.

We found that membrane proteins share diverse functional domains with soluble proteins in metazoan species compared to non-metazoans. Among 4,715 human domains, 970 domains (20.5%) are shared by membrane and soluble proteins (Fig. 1a). In human genome, 1,276 membrane protein domains and 2,552 soluble protein domains were found. However, in yeast genome, among the total of 2,817 domains, only 137 domains (5.2%) are shared by membrane and soluble proteins. Thus, human membrane proteins have significantly more shared domains compared to yeast membrane proteins. We confirmed that this observation could not occur by random chance comparing them to datasets with randomly assigned membrane and soluble proteins (Kolmogorov-Smirnov test; p -value = 3.49×10^{-96} , Supplementary Fig. S1). We analyzed 59 non-metazoan and 43 metazoan genomes and found that metazoans have on average twice more diverse shared domains than non-metazoans (Supplementary Fig. S2). The number and fraction of shared domains in representative non-metazoan and metazoan species are shown in Fig. 1b.

Membrane proteins often recruit domains from soluble proteins and expand functional diversity of metazoan genome. We examined shared domains in non-metazoan and metazoan genomes. We quantified how frequently membrane proteins in metazoan organisms acquired pre-existing domains in non-metazoans or how frequently shared domains were created with the transition to multicellular organisms. Shared domains were divided into two groups; (1) those are found from both non-metazoan and

metazoans, and (2) those are found only from metazoans. We found that shared domains exist in both non-metazoans and metazoans are more frequent than shared domains specific to metazoans. The number of pre-existing shared domains is larger than expected by random chance (Supplementary Fig. S3). In contrast, the number of shared domains specific to metazoan is smaller than expected. These results suggest that membrane proteins frequently reused pre-existing domains to gain new functions.

We found that membrane proteins frequently recruit domains from soluble proteins. In metazoan species, various domains are found from both membrane and soluble proteins, but in non-metazoan species, those domains are found only from soluble proteins (Supplementary Fig. S4 and Table S1). For example, in non-metazoan species, leucine-rich repeat (LRR) domains are found only from soluble proteins, but in metazoan species, they become a shared domain and are found in both membrane and soluble proteins (Fig. 2a, b). The LRR domain of membrane protein NGL-2 is a shared domain and functions to expand intercellular PPI network in metazoan nerve system. As a synaptic adhesion protein, NGL-2 regulates the formation of expiatory synapses through LRR domain which recruits pre-and-postsynaptic proteins, such as Lamint-2, DLG4 and NMDA receptors^{15,16}. Another example, von Willebrand factor type A (VWA) are also found from both membrane and soluble proteins of metazoan species, but they are found only from soluble proteins in non-metazoan species (Fig. 2c, d). A metazoan protein ITGAL recruits intercellular adhesion molecules, such as ICAM1, ICAM2, ICAM3 and ICAM4, and mediates adhesive interaction for immune response and surveillance^{17,18}. The VWA domain of membrane protein integrin alpha-L (ITGAL) become a shared domain and has an important role in metazoan immune system. These results suggest that recruiting domains from soluble proteins during the evolution may have contributed to the expansion of functional diversity of membrane proteins in metazoan species.

Shared domains expanded cell-cell communication networks of metazoan species. It has been suggested that increased intercellular network complexity is one of the major factors that contributes to the growth of organismal complexity^{19–22} so we asked whether shared domains contribute to this process. Connections of membrane proteins with or without shared domain were examined in the

Table 1 | Distribution of the shared, membrane and soluble protein domains in non-metazoan and metazoan genomes. The sum of the fractions of shared domains, membrane protein domains and soluble proteins domains exceed 100%, because shared domains are included in both membrane and soluble protein domains

Taxonomy	Species	The number of domain types				The fraction of domain types (%)		
		Shared domains	Membrane protein domains	Soluble protein domains	Total domains	Shared domains	Membrane protein domains	Soluble protein domains
Non-metazoan	<i>Neosartorya fumigata</i>	253	702	2920	3369	7.51	20.84	86.67
Non-metazoan	<i>Saccharomyces cerevisiae</i>	147	531	2433	2817	5.22	18.85	86.37
Non-metazoan	<i>Pichia angusta</i>	135	489	2430	2784	4.85	17.56	87.28
Non-metazoan	<i>Schizosaccharomyces pombe</i>	120	527	2470	2877	4.17	18.32	85.85
Non-metazoan	<i>Malassezia globosa</i>	123	413	2105	2395	5.14	17.24	87.89
Metazoan	<i>Caenorhabditis elegans</i>	426	940	2927	3441	12.38	27.32	85.06
Metazoan	<i>Anopheles darlingi</i>	426	977	2943	3494	12.19	27.96	84.23
Metazoan	<i>Drosophila melanogaster</i>	413	997	3148	3732	11.07	26.71	84.35
Metazoan	<i>Mus musculus</i>	753	1308	4097	4652	16.19	28.12	88.07
Metazoan	<i>Homo sapiens</i>	970	1399	4286	4715	20.57	29.67	90.90

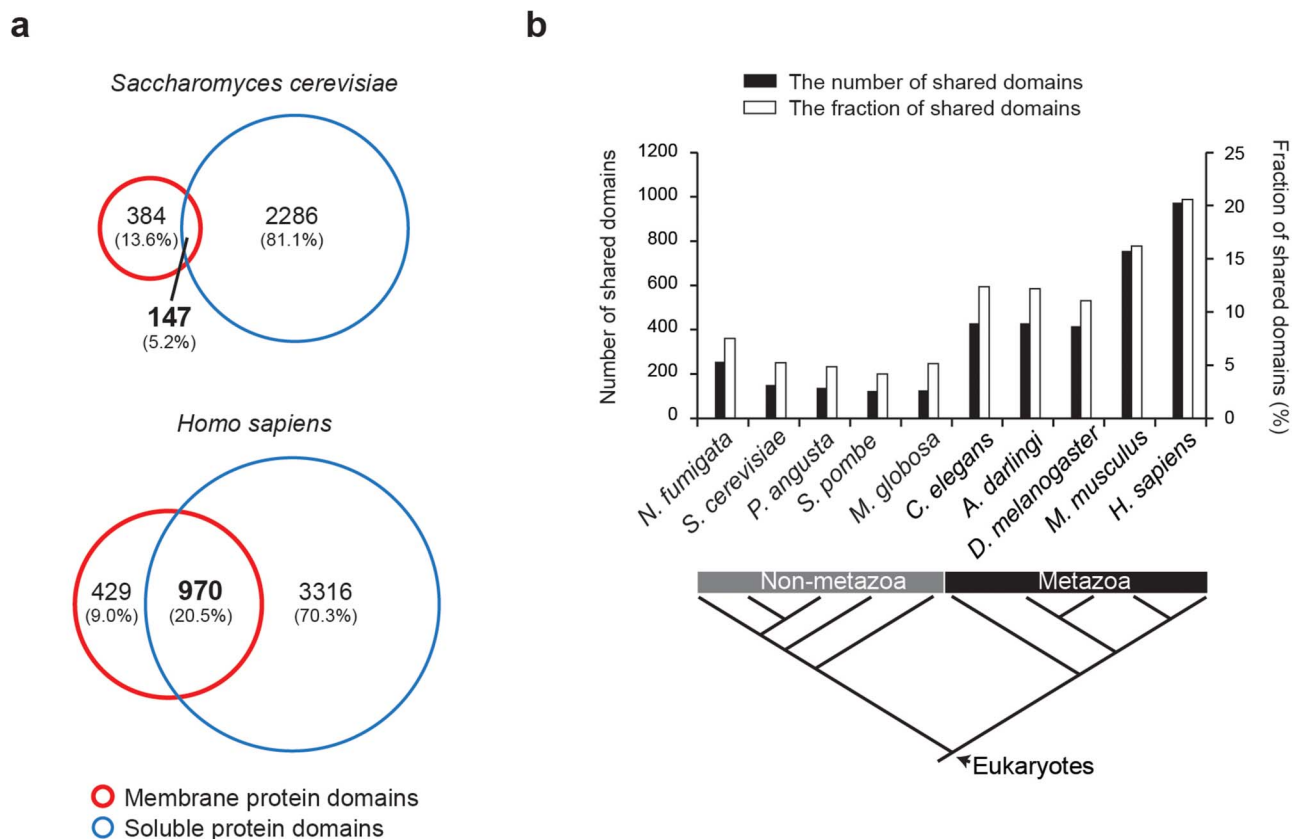


Figure 1 | Shared domains of membrane and soluble proteins in non-metazoan and metazoan genomes. (a) Overlap analysis of membrane and soluble protein domains in yeast and human genomes. (b) The number (black bars) and fraction (white bars) of shared domains of membrane and soluble proteins in 5 non-metazoan and 5 metazoan genomes.

human PPI network. We found that membrane proteins with shared domains have a significantly greater number of network connections than those without shared domains (p -value = 8.83×10^{-6} ; Fig. 3a). Moreover, membrane proteins with shared domains have a higher betweenness centrality than those without shared domains (p -value = 2.94×10^{-7} ; Fig. 3b). Betweenness centrality counts the number of shortest paths between all pairs of nodes passing through it²³. Thus, proteins with high betweenness centrality are important for information flow in the network^{24,25} and tend to interact with many different functional groups²⁶. In the topology analysis of fly PPI network, we also confirmed that membrane proteins with shared domains have a significantly greater number of network connection and a higher betweenness centrality than ones without shared domains (Supplementary Fig. S5). Our results suggest that membrane proteins with shared domains contributed to the expansion of diverse network connections and information flow in the cell-cell communication network.

Next, we examined the location of the interaction partners of membrane proteins with or without shared domains. Information of sub-cellular localization was taken from Gene Ontology²⁷ and UniProt database²⁸. Among the 1,751 membrane proteins with shared domains, about 80% (1,472) of them are localized at the plasma membrane (Supplementary Fig. S6). We found that membrane proteins with shared domains mostly interact with partners in the extracellular region rather than the cytoplasm (enrichment score = 8.27×10^{-8} and 9.08×10^{-2} for extracellular and cytosolic proteins, respectively; Fig. 3c), whereas the interaction partners of membrane proteins without shared domains are mainly located at cytoplasm (enrichment score = 6.49×10^{-1} and 6.66×10^{-6} for extracellular and cytosolic proteins, respectively; Fig. 3c). Moreover, we find that shared domains are generally presented at the extracellular side of membrane proteins.

As shown in Fig. 3d, 73.8% (2,124) of shared domains are found on the outside of membrane proteins (extracellular side) compared to only 26.1% (754) of the shared domains that are located at the inside of membrane proteins (cytosolic side). These results suggest that shared domains located at the extracellular side of membrane proteins play a particularly important role in generating the cell-cell communication network.

Shared domains participate in metazoan-specific functions. We next examined the function of shared domains that dramatically increased in metazoan species. We compared the functional difference of proteins with or without shared domains by functional enrichment analysis of Gene Ontology (GO). As shown in Fig. 4, membrane proteins that have shared domains are significantly enriched in five GO categories of biological functions that are important for multicellular life, such as “cell adhesion”, “regulation of signaling”, “defense response”, “immune system process” and “developmental process”. The biological process of “cell adhesion” is significantly enriched in membrane proteins with shared domains compared to the ones without shared domains (p -value = 9.46×10^{-22} ; hypergeometric test). Cell-adhesion gene families are critical for metazoan evolution because integrity of multicellular organisms is sustained by the stable adhesion of neighboring cells²⁹. For example, a shared domain of metazoan membrane protein, F5/8 type C domain of contactin-associated proteins-like 2, plays a critical role for the interactions between neurons and glia during nervous system development³⁰ and mediate the formation of an adhesion complex for salutatory conduction^{31,32}. Biological functions of “defense response” and “regulation of immune system process” are greatly overrepresented among membrane proteins with shared domains (p -value = 2.48×10^{-10} and p -value = 6.10×10^{-7} ,

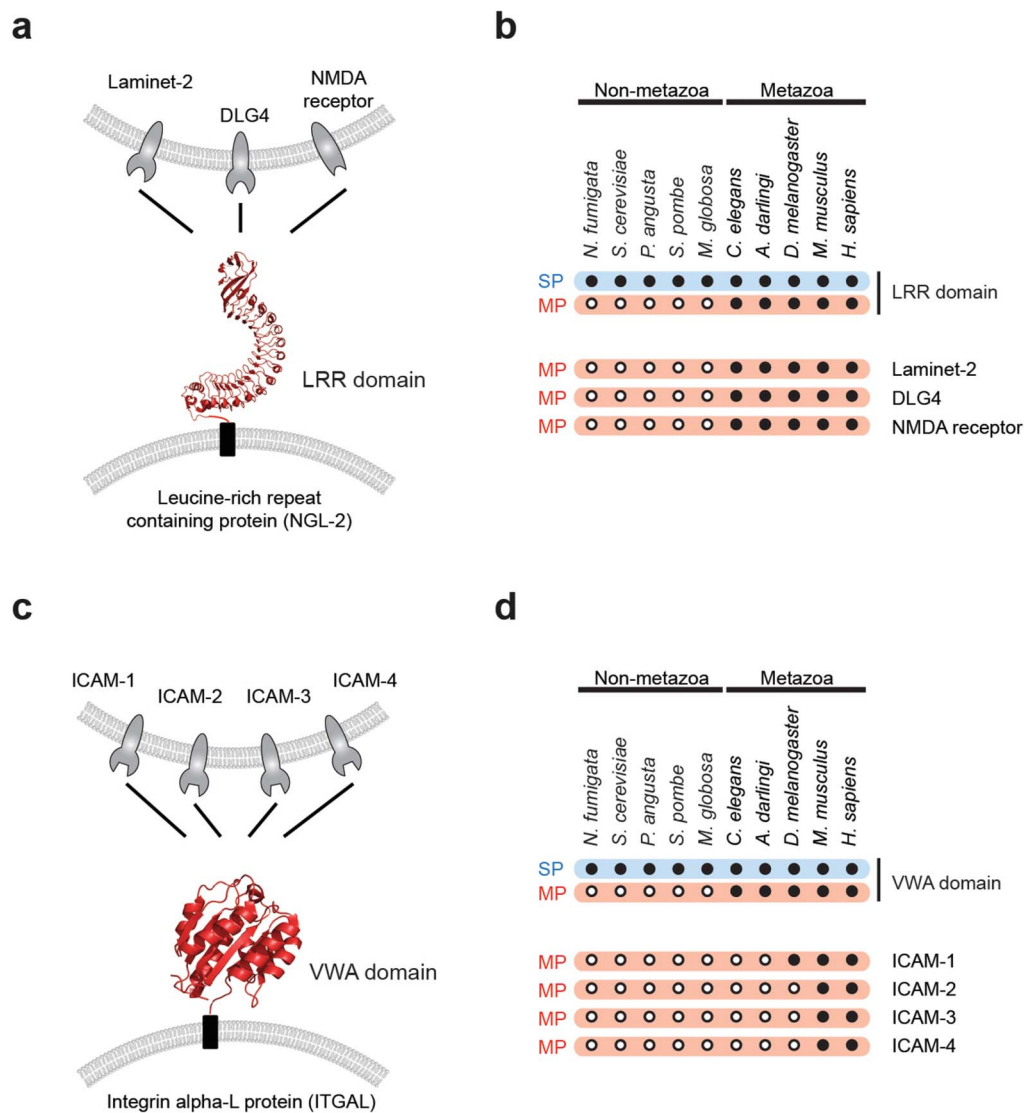


Figure 2 | Examples of membrane proteins with shared domains that connect intercellular networks. (a) LRR domains of leucine-rich repeat-containing proteins (NGL-2) interact with Laminet-2, DLG4 and NMDA receptors. (b) Phylogenetic profiles of LRR domains and their interaction partners in non-metazoan and metazoan species. A black dot indicates the presence of shared domains. A hollow dot indicates the absence of shared domains. (c) VWA domains of integrin alpha-L proteins (ITGAL) interact with ICAM1, ICAM2, ICAM3 and ICAM4. (d) Phylogenetic profiles of VWA domains and their interaction partners.

respectively; hypergeometric test). Multicellular organisms have developed effective defense and immune system to combat microbes and parasites³³. For example, Necrosis Factor domains and C-type lectin domains are found to be shared domains of metazoan membrane proteins and have functional roles in immune cells to contact to their cognate receptors in other cells^{34,35}. Moreover, the biological function of “developmental process” is significantly enriched in membrane proteins with shared domains (p -value = 5.79×10^{-7} ; hypergeometric test). Developmental process is a crucial progression for metazoan species to generate and organize specialized cell types and organs. Other GO categories significantly different between membrane proteins with or without shared domains are listed in Supplementary Table S2. These results suggest that shared domains are particularly important in metazoan specific functions and contribute to the transition to multicellularity.

Discussion

Our results show that the incorporation of soluble domains into membrane proteins is a major contributor to the development of cellular diversification and cooperation enabled metazoan evolution.

Previously, it has been suggested that membrane proteins expand their structural diversity by domain duplication^{36–38} or hetero-oligomerization¹³. However, domain recombination within membrane proteins is rare since the fold diversity of membrane proteins is limited¹⁰. Current approaches for understanding the functional diversity of membrane proteins have focused on transmembrane domains, rather than their extra-membrane domains⁹, but we recently showed that the incorporation of soluble protein domains is a major mechanism for the diversification of membrane proteins during evolution¹⁴. Our results suggest that domain sharing of membrane and soluble proteins is an important evolutionary pathway to obtain the functional diversity of membrane proteins.

Our study reveals that the incorporation of soluble extracellular domains was a particularly powerful mechanism for the evolution of multicellular life. Consistent with our finding, it has been shown that a large number of extracellular matrix proteins were exist in non-metazoans and expanded by domain shuffling during the transition to metazoans^{39,40}. Diverse domains of soluble proteins recombined with transmembrane domains and became a key player for cell-cell communications in metazoan species. Membrane-anchored

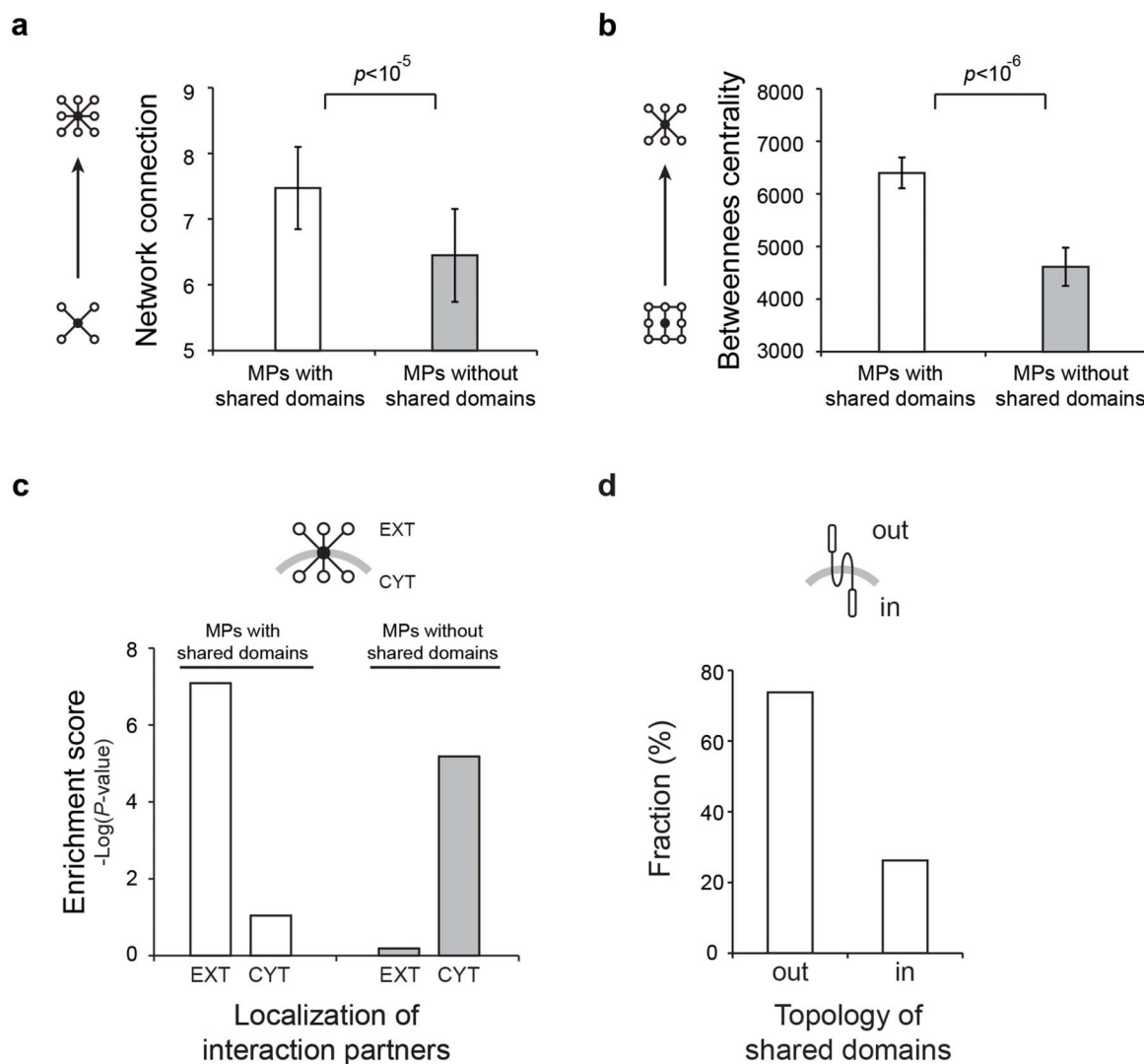


Figure 3 | Network properties of membrane proteins with shared domains. (a) The number of network connections (degree) of membrane proteins with and without shared domains were compared in the human PPI network. (b) Comparison of betweenness centrality of membrane proteins with and without shared domains. Error bars represent the standard error. (c) Comparison of the localization enrichment of the interaction partners of membrane proteins with and without shared domains. EXT indicates extracellular interaction partners and CYT indicates cytosolic interaction partners. (d) Topological orientation of shared domains in membrane proteins. ‘Out’ indicates the extracellular side and ‘in’ indicates the cytosolic side. The topology of shared domains was determined by combining the prediction results of TMHMM and location information of Pfam domains.

domains on cell surface are positioned to interact with proteins on neighboring cells and to detect various secreted molecules and to transfer the chemical information inside the cell. In unicellular organism, LRR domains usually function as an adapter for protein interactions of soluble proteins inside the cell^{41,42}. However, in multicellular organisms, LRR domains of membrane proteins anchored to cell surface, thereby creating a new function as cell-adhesion or cell-cell communication units in pre-and-post synaptic cells (Fig. 2a)^{15,16}. Moreover, membrane-anchored domains play an important role for self versus non-self recognition⁴³. For example, VWA domains of membrane proteins function on the surface of immune cells in metazoan species (Fig. 2c)^{17,18}. VWA domains of membrane-anchored integrin alpha-L proteins are known to enhance the binding affinity of immune complexes in leukocyte-endothelial cell-cell interaction and cytotoxic T-cell mediated immune responses^{44,45}. Our findings suggest that recombination of soluble and transmembrane domains confer new biological functions for metazoan multicellularity since the combination of two domains can make a new entity with both functions.

We found that domain sharing between membrane and soluble proteins is coupled with the appearance of multicellular organisms to contribute their intercellular networks with higher complexity. Especially, metazoans turn out to have more diverse shared domains than non-metazoans (Fig. 1). One might ask that shared domains likely increase in other evolutionary transitions such as the appearance of vertebrates or mammals. However, such increase was not observed in vertebrates or mammals (Supplementary Fig. S2a). In the analysis of shared domain within metazoan species including 5 vertebrates and 12 mammals, significant increases of shared domains were not found.

We found that membrane proteins with shared extra-membrane domains have increased network connections and information flow for the extracellular PPI network (Fig. 3). Intercellular communication is the key to maintain homeostasis, developmental, defense and immune process of multicellular organism⁴⁶. Therefore, many efforts have been invested to identify novel extracellular interactions and in constructing cell-cell communication networks^{20–22}. Recently, comparative genomic studies on the nervous system discovered a few

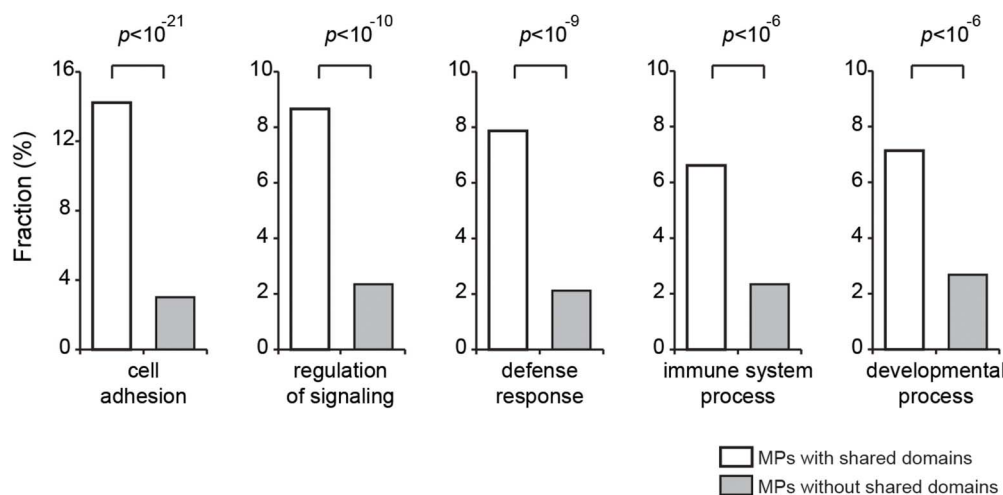


Figure 4 | Functional enrichment of membrane proteins with and without shared domains. The fraction of membrane proteins with and without shared domains was compared in functional categories using Gene Ontology (GO). Significantly enriched functional categories at GO level 2 and level 3 are shown (p -value $< 1.0 \times 10^{-4}$, hypergeometric test).

components of metazoan synaptic transmission and plasticity^{47–49}. However, it has been generally difficult to characterize molecular components that form intercellular communication system important for the evolution of multicellular organisms⁵⁰. Our findings suggest that the characterization of the domain diversity of membrane proteins may improve the understanding of the evolution of cell-cell communication network.

Methods

Classification of membrane and soluble proteins. We classified proteins of non-metazoan and metazoan genomes into membrane and soluble proteins by using experimental annotations and the transmembrane domains (TMDs) prediction program. The overall workflow is outlined in Supplementary Fig. S7. We first downloaded complete or reference genomes of 5 non-metazoan and 5 metazoan species from UniProt database²⁸. Membrane proteins were identified by having TMDs as indicated by the UniProt and Pfam databases⁵¹. Proteins annotated as ‘Single-/Multi-pass membrane proteins’ and ‘Transmembrane proteins’ were included, but ‘Peripheral membrane proteins’ were excluded. For the proteins without experimental annotation, we predicted TMDs by using TMHMM⁵², which is one of the most reliable TMD prediction programs for a large number of sequences^{53,54}. Proteins with predicted TMDs were considered to be membrane proteins¹⁴. We collected soluble proteins by excluding membrane proteins and putative membrane proteins. The total number of membrane and soluble proteins of each non-metazoan and metazoan genome are listed in Supplementary Table S3.

Identification of shared domains, membrane protein and soluble protein domains. We assigned 5,952 Pfam domains (Release 26)⁵¹ into membrane and soluble proteins of non-metazoan and metazoan genomes. All the sequences of membrane and soluble proteins were searched against the profile hidden Markov models of Pfam-A domains using pfam_scan.pl script and HMMER3⁵⁵. From the derived domains, we identified “shared domains”, “membrane protein domains”, and “soluble protein domains”. Shared domains are the domains that are found in both membrane and soluble proteins, whereas, membrane protein domains represent the domains which exist in membrane proteins. Soluble protein domains represent the domains which exist in soluble proteins. Then, we calculated the fraction of three domain classes by dividing the total number of Pfam domain in individual genome. Table 1 lists the numbers and fractions of shared domains, membrane protein and soluble protein domains of all 5 non-metazoan and 5 metazoan genomes.

Functional enrichment analysis. To investigate biological functions of proteins with shared domains, we utilized the Gene Ontology (GO) Biological Process (BP) annotations²⁷ that were derived from the entire GO (version 1.2). We examined which GO BP terms were enriched in the proteins with shared domains relative to the proteins without shared domains. First, we extracted 2,011 BP terms at GO level 2 and level 3, each of which annotates at most 20,000 human proteins. Next, we selected statistically overrepresented BP terms for the proteins with shared domains and compared with the proteins without shared domains using a hypergeometric test. Only GO BP terms that were overrepresented with p -value lower than 1.0×10^{-4} were employed. The same procedure was applied for membrane and soluble protein domains. Supplementary Table S2 lists all GO BP terms which are significantly enriched in the proteins with shared domains.

Transmembrane topology of shared domains. To investigate whether shared domains are located on the inside or outside of cells, we first predicted the transmembrane (TM) topology of the membrane proteins by using TMHMM⁵⁴, and then assigned the sequence-position information of shared domains into the TM topology. We evaluated the reliability of localization prediction of shared domains. The results of TMHMM prediction and the experimentally confirmed annotations of UniProt database were compared. We examined the agreement of location (extracellular and cytoplasmic sides) of shared domains. The overall accuracy and precision of localization prediction of shared domains reached 92% and 94%, respectively (Supplementary Fig. S8).

Subcellular localization analysis. Subcellular localizations of membrane proteins with shared domains were derived from UniProt database²⁸. Five subcellular localizations (plasma membrane, ER membrane, Golgi membrane, mitochondrial membrane and nuclear membrane) and their transmembrane topology were listed in Supplementary Table S4.

The subcellular localization of membrane protein interaction partners was derived from Gene Ontology (GO) Cellular Component (CC) annotation (version 1.2)²⁷. Extracellular and cytosolic subcellular localization were assigned. Proteins with CC terms of “extracellular region” and “extracellular space” were assigned as extracellular proteins (EXT) and those with CC terms of “cytosol” and “cytoplasmic part” were assigned as cytosol proteins (CYT). Subcellular localization information was available for 1,018 proteins of 2,362 interaction partners of membrane proteins. We used the hypergeometric distribution to calculate the enrichment score for the observed fraction of extracellular and cytosol partners of membrane proteins with or without domains.

Construction of protein interaction network. We compiled human protein interactions from a total of 22 existing protein interaction databases⁵⁶: 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 literature⁵⁷. We removed low-confidence interactions that were not supported by direct experimental evidence, such as PPIs inferred from orthologous interactions or obtained by co-expression signals from microarray experiments. The final network comprised 101,777 interactions between 11,043 human proteins.

1. Eichinger, L. *et al.* The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**, 43–57 (2005).
2. King, N., Hittinger, C. T. & Carroll, S. B. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* **301**, 361–3 (2003).
3. Ruiz-Trillo, I. *et al.* The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet* **23**, 113–8 (2007).
4. Escudero, L. M. *et al.* Epithelial organisation revealed by a network of cellular contacts. *Nat Commun* **2**, 526 (2011).
5. Ekman, D., Bjorklund, A. K. & Elofsson, A. Quantification of the elevated rate of domain rearrangements in metazoa. *J Mol Biol* **372**, 1337–48 (2007).



6. Copley, R. R., Schultz, J., Ponting, C. P. & Bork, P. Protein families in multicellular organisms. *Curr Opin Struct Biol* **9**, 408–15 (1999).
7. King, N. *et al.* The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–8 (2008).
8. Nam, H. J., Jeon, J. & Kim, S. Bioinformatic approaches for the structure and function of membrane proteins. *BMB Rep* **42**, 697–704 (2009).
9. Elofsson, A. & von Heijne, G. Membrane protein structure: prediction versus reality. *Annu Rev Biochem* **76**, 125–40 (2007).
10. Oberai, A., Ihm, Y., Kim, S. & Bowie, J. U. A limited universe of membrane protein families and folds. *Protein Sci* **15**, 1723–34 (2006).
11. Todd, A. E., Orengo, C. A. & Thornton, J. M. Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* **307**, 1113–43 (2001).
12. Vogel, C., Teichmann, S. A. & Pereira-Leal, J. The relationship between domain duplication and recombination. *J Mol Biol* **346**, 355–65 (2005).
13. Liu, Y., Gerstein, M. & Engelman, D. M. Transmembrane protein domains rarely use covalent domain recombination as an evolutionary mechanism. *Proc Natl Acad Sci U S A* **101**, 3495–7 (2004).
14. Nam, H. J., Han, S. K., Bowie, J. U. & Kim, S. Rampant exchange of the structure and function of extramembrane domains between membrane and water soluble proteins. *PLoS Comput Biol* **9**, e1002997 (2013).
15. Biederer, T. Hooking up new synapses. *Nat Neurosci* **9**, 1203–4 (2006).
16. Kim, S. *et al.* NGL family PSD-95-interacting adhesion molecules regulate excitatory synapse formation. *Nat Neurosci* **9**, 1294–301 (2006).
17. Shimaoka, M. *et al.* Structures of the alpha L I domain and its complex with ICAM-1 reveal a shape-shifting pathway for integrin regulation. *Cell* **112**, 99–111 (2003).
18. Zhang, H. *et al.* An unusual allosteric mobility of the C-terminal helix of a high-affinity alphaL integrin I domain variant bound to ICAM-5. *Mol Cell* **31**, 432–7 (2008).
19. Stumpf, M. P. *et al.* Estimating the size of the human interactome. *Proc Natl Acad Sci U S A* **105**, 6959–64 (2008).
20. Martin, S. *et al.* Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling. *Mol Cell Proteomics* **9**, 2654–65 (2010).
21. Kirouac, D. C. *et al.* Dynamic interaction networks in a hierarchically organized tissue. *Mol Syst Biol* **6**, 417 (2010).
22. Wu, Y., Garmire, L. X. & Fan, R. Inter-cellular signaling network reveals a mechanistic transition in tumor microenvironment. *Integr Biol (Camb)* **4**, 1478–86 (2012).
23. Goh, K. I., Kahng, B. & Kim, D. Universal behavior of load distribution in scale-free networks. *Phys Rev Lett* **87**, 278701 (2001).
24. Wagner, A. & Fell, D. A. The small world inside large metabolic networks. *Proc Biol Sci* **268**, 1803–10 (2001).
25. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–4 (2000).
26. Dunn, R., Dudbridge, F. & Sanderson, C. M. The use of edge-betweenness clustering to investigate biological function in protein interaction networks. *BMC Bioinformatics* **6**, 39 (2005).
27. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–9 (2000).
28. Magrane, M. & Consortium, U. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* **2011**, bar009 (2011).
29. Gumbiner, B. M. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell* **84**, 345–57 (1996).
30. Poliak, S. *et al.* Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K⁺ channels. *Neuron* **24**, 1037–47 (1999).
31. Li, X., Zhang, J., Cao, Z., Wu, J. & Shi, Y. Solution structure of GOPC PDZ domain and its interaction with the C-terminal motif of neuroligin. *Protein Sci* **15**, 2149–58 (2006).
32. Barrett, J. C. & Shelby, M. D. Mechanisms of human carcinogens. *Prog Clin Biol Res* **374**, 415–34 (1992).
33. Muller, W. E. & Muller, I. M. Origin of the metazoan immune system: identification of the molecules and their functions in sponges. *Integr Comp Biol* **43**, 281–92 (2003).
34. Sun, M. & Fink, P. J. A new class of reverse signaling costimulators belongs to the TNF family. *J Immunol* **179**, 4307–12 (2007).
35. Cambi, A. & Figdor, C. Necrosis: C-type lectins sense cell death. *Curr Biol* **19**, R375–8 (2009).
36. Shimizu, T., Mitsuke, H., Noto, K. & Arai, M. Internal gene duplication in the evolution of prokaryotic transmembrane proteins. *J Mol Biol* **339**, 1–15 (2004).
37. Rapp, M., Granseth, E., Seppala, S. & von Heijne, G. Identification and evolution of dual-topology membrane proteins. *Nat Struct Mol Biol* **13**, 112–6 (2006).
38. Bowie, J. U. Flip-flopping membrane proteins. *Nat Struct Mol Biol* **13**, 94–6 (2006).
39. Hynes, R. O. The evolution of metazoan extracellular matrix. *J Cell Biol* **196**, 671–9 (2012).
40. Murray, P. S. & Zaidel-Bar, R. Pre-metazoan origins and evolution of the cadherin adhesome. *Biol Open* **3**, 1183–95 (2014).
41. Kobe, B. & Kajava, A. V. The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* **11**, 725–32 (2001).
42. Gay, N. J., Packman, L. C., Weldon, M. A. & Barna, J. C. A leucine-rich repeat peptide derived from the *Drosophila* Toll receptor forms extended filaments with a beta-sheet structure. *FEBS Lett* **291**, 87–91 (1991).
43. Litman, G. W. & Cooper, M. D. Why study the evolution of immunity? *Nat Immunol* **8**, 547–8 (2007).
44. Hynes, R. O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673–87 (2002).
45. Springer, T. A. Complement and the multifaceted functions of VWA and integrin I domains. *Structure* **14**, 1611–6 (2006).
46. Gerhart, J. 1998 Warkany lecture: signaling pathways in development. *Teratology* **60**, 226–39 (1999).
47. Sakarya, O. *et al.* A post-synaptic scaffold at the origin of the animal kingdom. *PLoS One* **2**, e506 (2007).
48. Garcia, M. L. & Strehler, E. E. Plasma membrane calcium ATPases as critical regulators of calcium homeostasis during neuronal cell function. *Front Biosci* **4**, D869–82 (1999).
49. Muller, D., Buchs, P. A., Stoppini, L. & Boddeke, H. Long-term potentiation, protein kinase C, and glutamate receptors. *Mol Neurobiol* **5**, 277–88 (1991).
50. Ryan, T. J. & Grant, S. G. The origin and evolution of synapses. *Nat Rev Neurosci* **10**, 701–12 (2009).
51. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res* **42**, D222–30 (2014).
52. Moller, S., Croning, M. D. & Apweiler, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–53 (2001).
53. Chen, C. P., Kernysky, A. & Rost, B. Transmembrane helix predictions revisited. *Protein Sci* **11**, 2774–91 (2002).
54. Cuthbertson, J. M., Doyle, D. A. & Sansom, M. S. Transmembrane helix prediction: a comparative evaluation and analysis. *Protein Eng Des Sel* **18**, 295–308 (2005).
55. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195 (2011).
56. Kim, J. *et al.* Rewiring of PDZ domain-ligand interaction network contributed to eukaryotic evolution. *PLoS Genet* **8**, e1002510.
57. Bromberg, K. D., Ma'ayan, A., Neves, S. R. & Iyengar, R. Design logic of a cannabinoid receptor signaling network that triggers neurite outgrowth. *Science* **320**, 903–9 (2008).

Acknowledgments

We thank Prof. James U. Bowie and SBI members for helpful discussion throughout the entire project. This work was supported in part by Korean National Research Foundation grants (2013018606), a Pohang University of Science and Technology (POSTECH) Basic Science Research Institute (BSRI) grant and National Institutes of Health (NIH) Grant RO1 GM063919.

Author contributions

H.J.N. and S.K. designed the study. H.J.N., J.U.B. and S.K. wrote the paper. H.J.N., I.K. and S.K. analyzed data.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Nam, H.-J., Kim, I., Bowie, J.U. & Kim, S. Metazoans evolved by taking domains from soluble proteins to expand intercellular communication network. *Sci. Rep.* **5**, 9576; DOI:10.1038/srep09576 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>