

## RESEARCH ARTICLE

## Deciphering the evolution of composite-type GSKIP in mitochondria and Wnt signaling pathways

Cheng-Yu Tsai<sup>1,2</sup>, Shean-Jaw Chiou<sup>3</sup>, Huey-Jiun Ko<sup>3,4</sup>, Yu-Fan Cheng<sup>3,4</sup>, Sin-Yi Lin<sup>3,4</sup>, Yun-Ling Lai<sup>3,4</sup>, Chen-Yen Lin<sup>3,4</sup>, Chihuei Wang<sup>5</sup>, Jiin-Tsuey Cheng<sup>6</sup>, Hsin-Fu Liu<sup>7</sup>, Aij-Li Kwan<sup>1,2,4</sup>, Joon-Khim Loh<sup>2,4\*</sup>, Yi-Ren Hong<sup>1,3,4,6\*</sup>

**1** College of Medicine, Kaohsiung Medical University and National Health Research Institutes, Kaohsiung, Taiwan, **2** Department of Neurosurgery, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan, **3** Department of Biochemistry, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan, **4** Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan, **5** Department of Biotechnology, Kaohsiung Medical University, Kaohsiung, Taiwan, **6** Department of Biological Sciences, National Sun Yat-Sen University, Kaohsiung, Taiwan, **7** Department of Medical Research, Mackay Memorial Hospital, Taipei, Taiwan

☯ These authors contributed equally to this work.

\* [m835016@kmu.edu.tw](mailto:m835016@kmu.edu.tw) (YRH); [jokhlo@kmu.edu.tw](mailto:jokhlo@kmu.edu.tw) (JKL)



## OPEN ACCESS

**Citation:** Tsai C-Y, Chiou S-J, Ko H-J, Cheng Y-F, Lin S-Y, Lai Y-L, et al. (2022) Deciphering the evolution of composite-type GSKIP in mitochondria and Wnt signaling pathways. *PLoS ONE* 17(1): e0262138. <https://doi.org/10.1371/journal.pone.0262138>

**Editor:** Michael Schubert, Laboratoire de Biologie du Développement de Villefranche-sur-Mer, FRANCE

**Received:** June 20, 2021

**Accepted:** December 16, 2021

**Published:** January 20, 2022

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0262138>

**Copyright:** © 2022 Tsai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting Information](#) files.

## Abstract

We previously revealed the origin of mammalian simple-type glycogen synthase kinase interaction protein (GSKIP), which served as a scavenger and a competitor in the Wnt signaling pathway during evolution. In this study, we investigated the conserved and nonconserved regions of the composite-type GSKIP by utilizing bioinformatics tools, site-directed mutagenesis, and yeast two-hybrid methods. The regions were denoted as the pre-GSK3 $\beta$  binding site, which is located at the front of GSK3 $\beta$ -binding sites. Our data demonstrated that clustered mitochondria protein 1 (CLU1), a type of composite-type GSKIP that exists in the mitochondria of all eukaryotic organisms, possesses the protein known as domain of unknown function 727 (DUF727), with a pre-GSK3 $\beta$ -binding site and a mutant GSK3 $\beta$ -binding flanking region. Another type of composite-type GSKIP, armadillo repeat containing 4 (ARMC4), which is known for cilium movement in vertebrates, contains an unintegrated DUF727 flanking region with a pre-GSK3 $\beta$ -binding site (115SPxF118) only. In addition, the sequence of the GSK3 $\beta$ -binding site in CLU1 revealed that Q126L and V130L were not conserved, differing from the ideal GSK3 $\beta$ -binding sequence of simple-type GSKIP. We further illustrated two exceptions, namely 70 kilodalton heat shock proteins (Hsp70/DnaK) and Mitofilin in nematodes, that presented an unexpected ideal GSK3 $\beta$ -binding region with a pre-GSK3 $\beta$  sequence; this composite-type GSKIP could only occur in vertebrate species. Furthermore, we revealed the importance of the pre-GSK3 $\beta$ -binding site (118F or 118Y) and various mutant GSK3 $\beta$ -binding sites of composite-type GSKIP. Collectively, our data suggest that the new composite-type GSKIP starts with a DUF727 domain followed by a pre-GSK3 $\beta$ -binding site, with the subsequent addition of the GSK3 $\beta$ -binding site, which plays vital roles for CLU1, Mitofilin, and ARMC4 in mitochondria and Wnt signaling pathways during evolution.

**Funding:** This work was supported by Kaohsiung Medical University Hospital, Taiwan, KMH109-9R28 (awarded to JKL), KMH106-M18 (awarded to CYT); NSYSU-KMU Joint Research Project, Taiwan NSYSUKMU-109-P006 and NSYSUKMU-110-P009 (awarded to YRH and JTC); Ministry of Science and Technology, Taiwan MOST109-2320-B037-032; 109-2320-B037-012; 110-2320-B037-029 (awarded to YRH)."

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

## Introduction

Glycogen synthase kinase interaction protein (GSKIP) was originally cloned and identified as a glycogen synthase kinase-3 $\beta$  (GSK3 $\beta$ ) interacting protein through yeast two-hybrid screening using GSK3 $\beta$  [1–4]. Subsequently, GSKIP was also characterized as a small A-kinase anchoring protein (AKAP) [5, 6]. Studies have suggested that GSKIP is a cytosolic scaffolding protein that contains protein kinase A (PKA) RII- binding sites at the N-terminal residue V41/L45 and a GSK3 $\beta$ -binding domain at the C-terminal residue L130 [1–3, 5, 6].

In a previous study, transfection of the neuronal-like SH-SY5Y cell line with wild-type GSKIP inhibited neurite outgrowth, implying the possible role of GSKIP in the neuronal system [7]. GSKIP was demonstrated to modulate dynamin-related protein 1 (Drp1) phosphorylation and to provide neuroprotection against oxidative stress for SH-SY5Y cells [8]. These findings indicate that GSKIP functions as an anchor protein for GSK3 $\beta$  in the cAMP/PKA/Drp1 signaling axis. Moreover, it was reported that GSKIP could interact with PKA and GSK3 $\beta$ , which is required for Wnt signaling regulation, via a cytoplasmic destruction complex with  $\beta$ -catenin, leading to the negative regulation of Wnt signaling [6, 9]. Using knockout animal models, Deák et al. demonstrated that GSKIP regulates GSK3 $\beta$  activity and controls palatal shelf fusion during mice embryonic development [10]. GSKIP is associated with a predisposition to familial myeloid malignancies and poses a challenge for hematologists in germ-line gene duplication studies [11–13]. GSKIP is also involved in the PPAR $\gamma$ -related angiogenic potential of endothelial pulmonary microvascular endothelial cells [14]. A recent study showed that miR-150-5p significantly inhibited Wnt- $\beta$ -catenin signaling by simultaneously targeting GSKIP and  $\beta$ -catenin in non-small-cell lung cancer (NSCLC) cells [15]. GSKIP was downregulated in disease-related cellular protein networks under different epidermal growth factor receptor (EGFR) mutations in NSCLC [16]. Moreover, long intergenic nonprotein coding RNA 173 (Linc00173) upregulated Etk by functioning as a competitive endogenous RNA “sponging” miRNA-218, leading to GSKIP upregulation and resulting in the translocation of  $\beta$ -catenin in small-cell lung cancer cells [17]. Additionally, miR-181c-5p was reported to mitigate the cancer cell characteristics and invasive properties of cervical squamous cell carcinoma by targeting the GSKIP gene [18]. In a previous study, we demonstrated that this protein PKA/GSKIP/GSK3 $\beta$  axis plays a role in Tau phosphorylation in Alzheimer’s disease [19]. Collectively, these findings indicate that (1) simple-type GSKIP may play a multifaceted role in several diseases in a context-dependent manner, and (2) it is involved in the regulation of mitochondria proteins and the Wnt signaling pathway.

GSKIP has two subtypes, simple and composite, which coexist in all prokaryotic and eukaryotic cells. In our previous study, we revealed the origin of mammalian simple-type GSKIP and its evolution and discovered that it is involved in the Wnt signaling pathway not only as a scavenger but also as a competitor [20]. However, the origin and evolution of composite-type GSKIP remain unknown. In the present study, we generated evidence to demonstrate that along with simple-type GSKIP-linked proteins, four groups of composite-type GSKIP, namely clustered mitochondria protein 1 (CLU1) and Mitofilin in mitochondria, heat shock protein 70 (Hsp70) chaperone, and Armadillo repeat-containing 4 (ARMC4), are involved in Wnt signaling. This can be used to uncover the origin of the GSK3-binding site during the evolution of the protein known as domain of unknown function 727 (DUF727) and four groups of composite-type GSKIP.

## Materials and methods

### Gene ontology and UniProt analysis

Gene Ontology (<http://geneontology.org>) and InterPro (<http://www.ebi.ac.uk/interpro/>) were used for data mining [21–23]. UniProt (<https://doi.org/10.1093/nar/gkaa1100>) was combined

with InterPro [21–25] to compare the results available at the National Center for Biotechnology Information (NCBI) website for domain enrichment by using the E-Value for parameters. Gene-based tests were performed using the following keywords as indicators: GSKIP, DUF727, GSK3 $\beta$ , CLU1, and ARMC4 separately and in combination with GSKIP-related data (S1 Fig). We used FlyBase version FB2014\_04 to identify conservative portions in InterPro. Panther data mining is a protein classification system that is applied for analysis based on evolutionary relationships. It is a large biology database of gene/protein families and their functionally related subfamilies. It can be used to classify and identify the function of gene products [26, 27].

### Phylogenetic analysis

We used ClustalW [28] for multiple GSKIP sequence alignment for DUF727 (50–100 amino acids), pre-GSK3 $\beta$  (115–118 amino acids), and the GSK3 $\beta$ -binding domain (122–130 amino acids). The maximum likelihood method was used to select CLU1 to generate an unrooted phylogenetic tree of 20 species; a phylogenetic tree of 18 species was also found for one bacterium and ARMC4 orthologs [29–31]. ClustalW was used to ensure the tree structure [28]. The accession numbers of SSF103107, PF05303, IPR007967, and IPR023231 of all species in the UniProt search (S1 Table) and the sequence alignment and phylogenetic analysis of CLU1 and ARMC4 (S1 Fig) were used as parameters. The T-Coffee method was used to align full-length CLU1 and ARMC4 proteins and their fragmentary proteins spanning the SSF103107 unintegrated superfamily [32–34]. The MEGA-X program was used to produce the best-fitting amino acid substitution model [29–31]. We used the maximum likelihood method from MEGA-X to reconstruct phylogenetic trees with the LG substitution model [35, 36]. Bootstrapping was used to evaluate the robustness of the phylogenetic trees.

### Molecular modeling comparison

The initial three-dimensional (3D) NMR structural model of GSKIP (PDB ID: 1SGO) was obtained as previously described [20]. The NorthEast Structural Genomics consortium was used to obtain the human NMR structure of GSKIP (C14orf129, HSPC210; PDB ID: 1SGO). Next, the 3D structure of the ACT domain, which folds with a ferredoxin-like  $\beta\alpha\beta\beta\alpha\beta$  topology [37], was determined. The structure was minimized for 100,000 conjugate gradient steps and then subjected to 100-s isothermal, constant-volume MD simulation. The final structure was used in domain comparisons (GSKIP, ACT domain, and CLUH-KIAA0664-SSF103107) [1–4, 6, 7, 37].

### Consensus sequences and sequence logos: Weblogo 3.6.0

To illustrate the consensus sequence logos of GSKIP, we presented typical simple-type logos and several composite types. The rules were adapted from those used by the sequence comparison website PROSITE ([http://prosite.expasy.org/sequence\\_logo.html](http://prosite.expasy.org/sequence_logo.html)) [38], and standard one-letter codes were used for amino acids.

### NetWheels: Helical wheel diagram of peptides and net projections maker

The helices found in peptides and proteins are commonly modeled in two dimensions [39]. They can offer a view of the central axis in a protein. Wheel and net projections have been used to represent the two dimensions of 3D helical structures, and they enable the observation of the helical structural properties, especially in terms of residue polarity and intramolecular bonding. We used the helical wheel diagram shown here to determine the distribution of

amino acid residues in a helical segment within the sequence of simple-type GSKIP and various composite-type variants to distinguish the differences between the two types.

### Cloning, site-directed mutagenesis, and DNA sequencing

The plasmids of pACT2-GSKIP and pAS2-1-GSK3 $\beta$  were constructed for the yeast two-hybrid assay as described previously [40–42]. Briefly, GSK3 $\beta$  was cloned in-frame with the Gal4 DNA-binding domain in the pAS2-1 vector (MATCHMAKER Two-HybridSystem 2, Clontech) to yield the pAS2-1-GSK3 $\beta$  bait plasmid. In addition, DNA fragments encoding GSKIP were amplified through PCR using Taq polymerase (TaKaRa). The PCR fragments were then inserted into the BamHI and XhoI sites of the pACT2 (Clontech) vector to construct the pACT2-GSKIP plasmid. GSKIP Y118P, Y118A, Y118F, F122P, F122A, F122Y, L126P, F126A, L126V, L126Q, L130P, L130A, L130I, and L130V mutants were created through a site-directed mutagenesis technique by using the QuikChange Lightning kit (GE Healthcare, Sunnyvale, CA, USA). Mutated nucleotides were verified using an ABI PRISM 3730 Genetic Analyzer (Perkin-Elmer) for DNA sequencing. All experimental procedures were performed in accordance with the manufacturer's protocol.

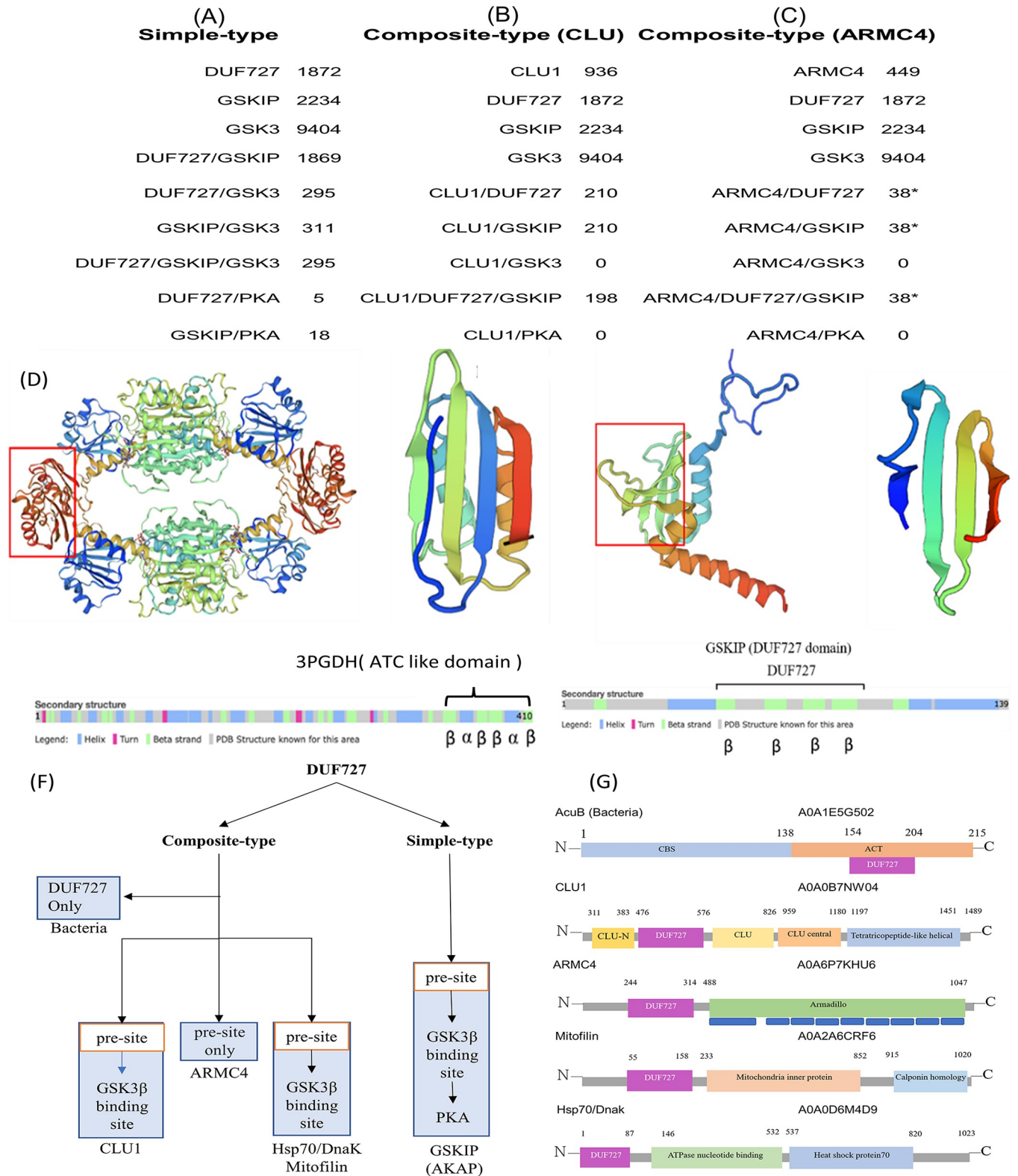
### Yeast two-hybrid system

Yeast two-hybrid screening was performed using the MATCHMAKER Two-Hybrid System 2 (Clontech) [40–42]. YRG-2 yeast host cells were purchased from Stratagene. pAS2-1 and pACT2 plasmids were cotransfected and selected on G2 plates deficient in tryptophan and leucine and on G3 plates deficient in histidine. The yeast host cells were MATa ura3–52 his3–200 ade2–101 lys2–801 trp1–901 leu2–3 112 gal4–542 gal80–538 LYS2::UASGAL1-TATA GAL1--HIS3 URA3::UASGAL4 17mers(x3)-TATACYC1-lacZ. A visible blue-color pattern in the colony filter lift assay on the G3 plates represented a positive interaction [41]. YRG-2 yeast cells were cotransfected with pACT2-GSKIP and an empty pAS2-1 vector and spread on G2 and G3 agar plates to determine the growth-inhibiting effect of GSKIP in yeast.

## Results

### Analysis of UniProt and GSKIP gene ontology: Identification of composite-type GSKIP

UniProt was used in combination with InterPro for domain enrichment analyses. The results were compared with the results in the NCBI database through stringent and extended gene-based tests for ranking the genes (Fig 1A, 1B and 1C). Our survey revealed that several domains were enriched for protein entries. The evolution of simple-type GSKIP and the GSK3 $\beta$ - and PKA RII-binding domains of 52 species has already been clearly demonstrated in previous reports [9, 20]. In this study, we extended the findings concerning four composite-type GSKIPs, namely CLU1 and ARMC4 (through the gene fusion mechanism presented in Fig 1A, 1B and 1C as well as the next subsection below) and another two proteins with sporadic occurrence (Mitofilin and Hsp70/Dnak), in nematodes (see the succeeding section on the hijack mechanism). CLU1 and ARMC4 contain DUF727; however, CLU1 contains other parts of CLU-N, CLU, CLU central, winged helix-like DNA-binding domains, and tetratricopeptide-like helical domains, but ARMC4 only contains 13X Armadillo repeats [43, 44]. DUF refers to a “domain of unknown function” [19–23, 45]. A short domain in the clustered mitochondria protein is involved in its mitochondrial cytoplasmic distribution [46, 47]. Moreover, ARMC4 in vertebrates has been identified as a multiprotein complex responsible for cilia movement; it is necessary for targeting and anchoring outer dynein arms [43, 44]. In the



**Fig 1. Summary of UniProt and InterPro data mining to reveal two types of GSKIP.** A, Evolved simple-type GSKIP from DUF727 to pre-GSK3 $\beta$  and then evolution to the GSK3 $\beta$ -binding site. The results of UniProtKB data retrieval using keywords are provided. When using GSKIP as a keyword, a total of 2234 protein entries were found, but the records were reduced when other terms such as DUF727, GSK3 $\beta$ , and PKA were added. B, When using CLU1 as a keyword, 936 protein entries were found, and 210 protein entries were convergent with the keywords “DUF727” or “GSKIP.” C, Using ARMC4 as a keyword resulted in

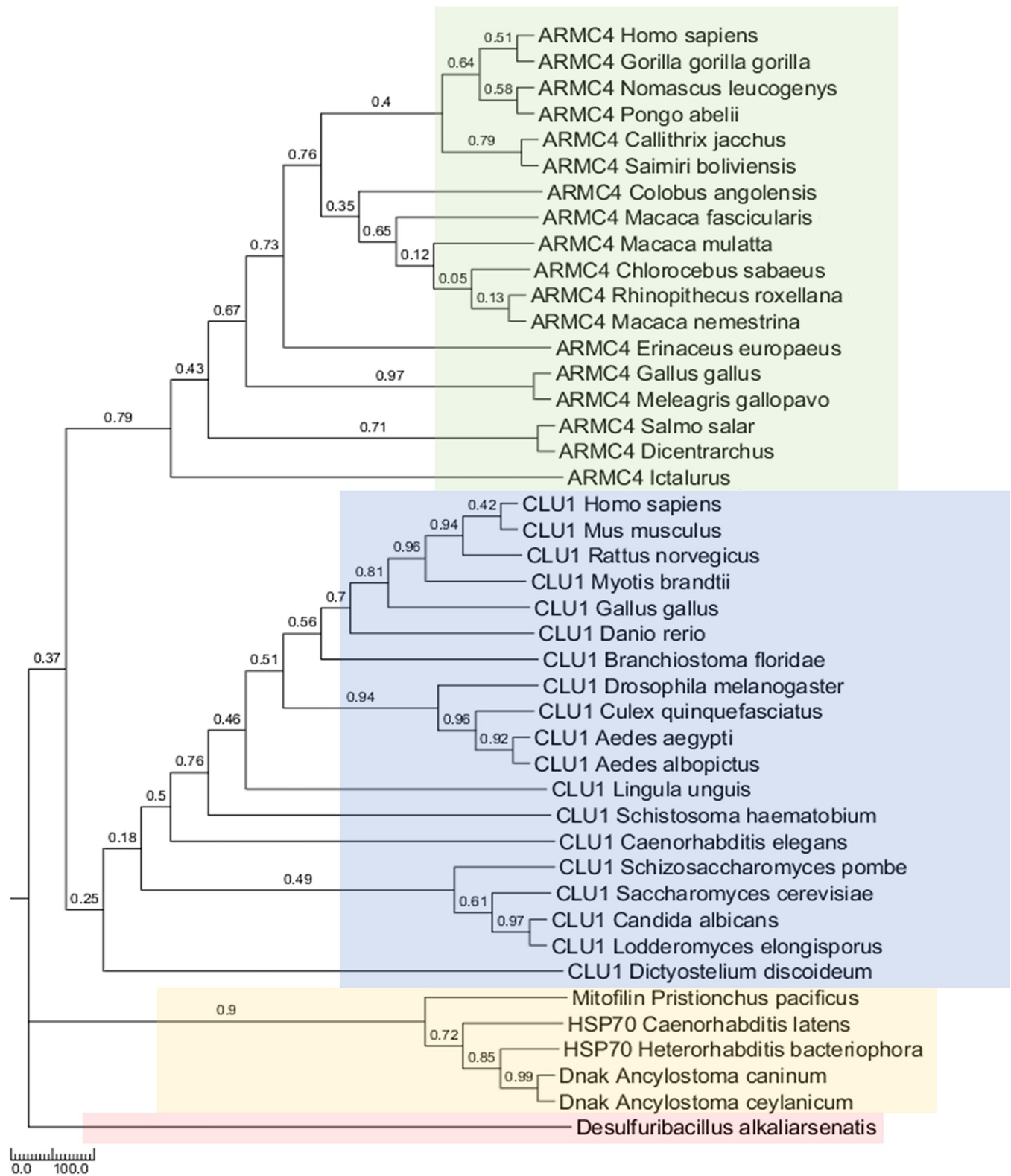
449 protein entries, but only 38 records remained as the SSF103107 hidden code (marked with “\*” as a cryptic code; see the text for details) when DUF727 or GSKIP was combined with ARMC4 for data mining. D, *E. coli* 3PGDH with a pair of ACT domains formed an eight-stranded antiparallel sheet whose 3D structure was determined (left) to fold into a ferredoxin-like  $\beta\alpha\beta\beta\alpha\beta$  topology (red box, right panel). E, NMR analysis of DUF727 (PDB ID: 1SGO) with the central part of antiparallel  $\beta\beta\beta\beta$  topology (red box, right panel) based on the Northeast Structural Genomics consortium (left). F, All queries of previous simple-type GSKIP were summarized using UniProt and Panther data mining (see the text for details). G, Four groups of composite-type GSKIP: AcuB, CLU1, ARMC4, and sporadic (Hsp70/DnaK and Mitofilin).

<https://doi.org/10.1371/journal.pone.0262138.g001>

present study, data mining was performed to search the available protein entries, with the results yielding 210 of 936 (22%) with respect to CLU1, 198 of 936 (21%) with respect to CLU1/DUF727, and 210 of 936 (22%) with respect to CLU1/GSKIP. Additionally, the hidden (cryptic) superfamily of DUF727 (SSF103107) was indicated to be an unintegrated family. The hypothetical protein c14orf129 hspc210 superfamily entry could enrich entries for the GSKIP (SSF103107) domain to 1869/1872 (99.8%; Fig 1A, 4 vs. 1). When ARMC4 was retrieved (all entries must be searched to find the hidden code SSF103107), only the hidden (cryptic) superfamily of DUF727 (identified as the unintegrated family SSF103107) with 38/449 entries (8%) was shown to contain a pre-GSK3-binding site (115SPxF118) containing F118 instead of Y118 as compared with the normal GSK3-binding site (more details on the comparisons are provided in subsequent discussion). We occasionally found composite-type GSKIP in two bacterial species, one of which was gram-positive *Desulfuribacillus alkaliarsenatis* (A0A1E5G502\_9BACL). The acetoin-utilization protein AcuB contained the CBS and ACT domains (CBS, 1–138 aa; ACT, 139–212 aa; GSKIP; DUF727, 154–204 aa), and DUF727 was located within the ACT domain as composite GSKIP. The ACT domain was identified in a PSI-BLAST search. *Escherichia coli* 3PGDH was discovered to be the first protein with an ACT domain that folds into a ferredoxin-like  $\beta\alpha\beta\beta\alpha\beta$  topology [37] (Fig 1D). The ACT domain is found in a variety of contexts and may be a conserved regulatory ligand-binding fold. However, DUF727 has antiparallel  $\beta\beta\beta\beta$  topology (PDB ID: 1SGO, the NorthEast Structural Genomics consortium was used for NMR analysis; Fig 1E), with a similar  $\beta$ -sheet homology (ferredoxin-like  $\beta\alpha\beta\beta\alpha\beta$  topology; Fig 1D right panel compared with Fig 1E) as the human CLUH-KIAA0664-SSF103107 superfamily domain (Fig 1E). We propose that during evolution, GSKIP may have originated from DUF727, as found in bacteria, and may have acquired the pre-GSK3 $\beta$ -binding motif and GSK3 $\beta$ -binding domain to become the ancestor of simple-type GSKIP. Through evolution, simple-type GSKIP later acquired the PKA RII-binding domain as GSKIP/AKAP in vertebrates, whereas simple-type GSKIP has been retained in invertebrates. For composite-type GSKIP, one bacterium was found to contain DUF727 with the ACT domain, whereas in all eukaryotes, DUF727 was incorrectly recognized as GSKIP. Although DUF727 lacks the ideal GSK3 $\beta$ -binding domain, it was still counted among the CLU/TIF31 and CLUH/KIAA0664 proteins. In some vertebrates, the GSKIP domain is also found in an ARMC4-containing protein with a pre-GSK3 $\beta$ -binding site only. Two sporadic Mitofilin and Hsp/DnaK proteins were found in invertebrates containing perfect pre-GSK3 $\beta$ -binding and GSK3 $\beta$ -binding sites (Fig 1F). We summarize four groups of composite-type GSKIP in Fig 1G.

### Two groups of composite-type GSKIP families may have evolved through a gene-fusion mechanism: CLU1 (mitochondria) and ARMC4 (Wnt pathway)

In different species, to determine the conserved sequences of GSK3 $\beta$ -binding regions and DUF727 in CLU1, we first selected one bacterial prototype and 20 GSKIP orthologs and used ClustalW and MEGA-X to construct a phylogenetic tree (Figs 2 and S1A). These 20 species



**Fig 2. Phylogenetic analysis of four composite-type GSKIP proteins, namely ARMC4, CLU1, HSP70/Dnak and Mitofilin, among various species.** The evolutionary history of GSKIP was inferred using the maximum likelihood method. Initial trees for the heuristic search were obtained automatically by applying the neighbor-joining and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model and by then selecting the topology with the superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 43

amino acid sequences. The final dataset included a total of 161 positions. Phylogenetic tree generated using MEGA-X [31]. The bootstrap values represented as the likelihood function for each species as indicated.

<https://doi.org/10.1371/journal.pone.0262138.g002>

comprised one bacteria [*D. alkaliarsenatis* (WP\_069641749.1)] (prototype), four yeasts [*Saccharomyces cerevisiae* (NP\_013725.1), *Schizosaccharomyces pombe* (NP\_595319.2), *Candida albicans* (XP\_710862.2), and *Lodderomyces elongisporus* (XP\_001526951.1)], one slime mold [*Dictyostelium discoideum* (XP\_629429.1)], one flatworm [*Schistosoma Haematobium* (XP\_012796625.1)], one Nematoda [*Caenorhabditis elegans* (NP\_499097.1)], four insects [*Drosophila melanogaster* (NP\_646986) and *Culex quinquefasciatus* (XP\_001843004.1)], including two species of mosquitoes [*Aedes aegypti* (XP\_001651007.1) and *Aedes albopictus* (A0A182G693)], one brachiopod [*Lingula anatina* (XP\_013383658.1)], and seven Chordata [*Branchiostoma lanceolatum* (XP\_002595966.1), *Myotis brandtii* (Brandt's bat, S7P578), *Danio rerio* (E7FE02), *Gallus gallus* (XP\_015151355.1), *Mus musculus* (NP\_001074627.1), *Rattus norvegicus* (NP\_001292142.1), and *Homo sapiens* (NP\_056044.3)]. Additionally, ClustalW was combined with the T-Coffee web-based program and MEGA-X to ensure a tree structure (Fig 3). Both a GSK3 $\beta$ -binding motif (122Fxxx126LxxR/K/QL130) and a pre-GSK3-binding motif (115SPxF118 rather than 115SPxY118 in simple-type GSKIP) were found in most of the GSKIP orthologs (Fig 3).

We found that two distinct bacteria contained composite-type GSKIP based on the observation of the aligned sequences without the GSK3 $\beta$ -binding site, which revealed that they are primitive [5, 6, 20]. Why and how composite-type GSKIP evolved a pre-GSK3 $\beta$ -binding site prior to the GSK3 $\beta$  functional domain should be investigated, and such an investigation could provide some indication of the role of the GSK3 $\beta$ -binding site in composite-type GSKIP evolution.

When DUF727 is regarded as an initial domain, the CLU1 mitochondria family [48, 49] and the ARMC4 family of proteins that emerge [43, 44] are composite-type GSKIPs. By contrast, in simple-type GSKIP, the GSK3 $\beta$ -binding region is conserved with some residues, indicating that the 122Fxxx126L/QxxR/KV/G/A130 region is essential. The Leu130 residue has been characterized as being essential for GSK3 $\beta$  binding in humans [1, 6, 20], indicating its inactivation when this site is modified with GSK3 $\beta$ . In the present study, this always occurred in CLU1 (Table 1), indicating that CLU1 is still in the process of evolution. Of note, the modification of the ARMC4-pre-GSK3 site 118F to 118Y abolished its interaction with GSK3 (Table 1), indicating the same evolutionary process of the ARMC4-pre-GSK3 $\beta$  site as that of the CLU1 family. When using the helical wheel diagram, the comparisons revealed various V126L residues, but no 130L residue, in CLU1 (Fig 4, GSKIP: 122–139 aa with its various mutants) and Logo 3.6.0 (Fig 5A–5D). These data enabled the evaluation of whether the consensus sequence is conserved in pre-GSK3- (115SPxF/Y118; Fig 5B and 5C compared with Fig 5A) and GSK3 $\beta$ -binding sites (122Lxxx126LxxK/RL130, Fig 5D compared with Fig 5A). Crucially, three composite GSKIP variants (126Q, 126V, and 130V) from different species were found to bind GSK3 by Y2H (Table 1). In addition, the DUF727 domain seemed to gradually become a pre-GSK3 $\beta$ -binding site (115SPxF118) as a flanking region with the consensus sequence site 115SPxF118xxx122Fxxx126LxxR/KL130, whereas in vertebrates, ARMC4 contained a hidden code SSF103107 (unintegrated DUF727 superfamily, Fig 5C) flanking region with pre-GSK3 $\beta$ -binding sites at 115SPxF118 only (see the next subsection below).

### A hidden code SSF103107 domain as a metaphor of DUF727 in the ARMC4 family

In our previous study, we used the keywords DUF727 (PF05303) and GSKIP (IPR007967) to search for ARMC4. Subsequently, DUF727 and GSKIP were eliminated as keywords from the





**Fig 3. Sequence alignment of 20 composite-type GSKIP/CLU1 proteins among various species.** Multiple sequence alignment with respect to the DUF727 region of GSKIP orthologs using T-Coffee. The conserved residues are indicated with asterisks, and residues with high similarity among the orthologs are marked with dots at the bottom. The 115SPxF118xxx122FxxxQxxRV130 motif in the pre-GSK3 $\beta$ - and GSK3 $\beta$ -binding regions is also indicated by blue and green boxes, respectively. T-Coffee also estimates alignment accuracy and improves phylogenetic tree reconstruction.

<https://doi.org/10.1371/journal.pone.0262138.g003>

**Table 1. Mutagenesis assay of the pre-GSK3 $\beta$ -binding site (118F or 118Y) and various mutant GSK3 $\beta$ -binding sites of full-length GSKIP and GSKIptide (115–139).**

	GSKIP		GSKIptide		
	G2	G3		G2	G3
<u>WT</u>	+	+	WT	+++	+++
Y118P	+++	-	Y118P	+++	-
Y118A	+++	+++	Y118A	+++	-
Y118F	+++	+++	Y118F	+++	-*
F122P	+++	-	F122P	+++	-
F122A	+++	+++	F122A	+++	-
F122Y	+++	+++	F122Y	+++	-
L126P	+++	-	L126P	+++	-
L126A	+++	+++	L126A	+++	-
L126V	+++	+++	L126V	+++	-
<u>L126Q</u>	+++	+++	L126Q	+++	-
L130P	+++	-	L130P	+++	-
L130A	+++	+++	L130A	+++	-
L130I	+++	+++	L130I	+++	-
<u>L130V</u>	+++	+++	L130V	+++	-

The  $\beta$ -galactosidase filter assay was also used for semiquantitative analysis. (+) indicates positive interaction, (-) indicates the interaction is abolished, and (+++) indicates a strong interaction.

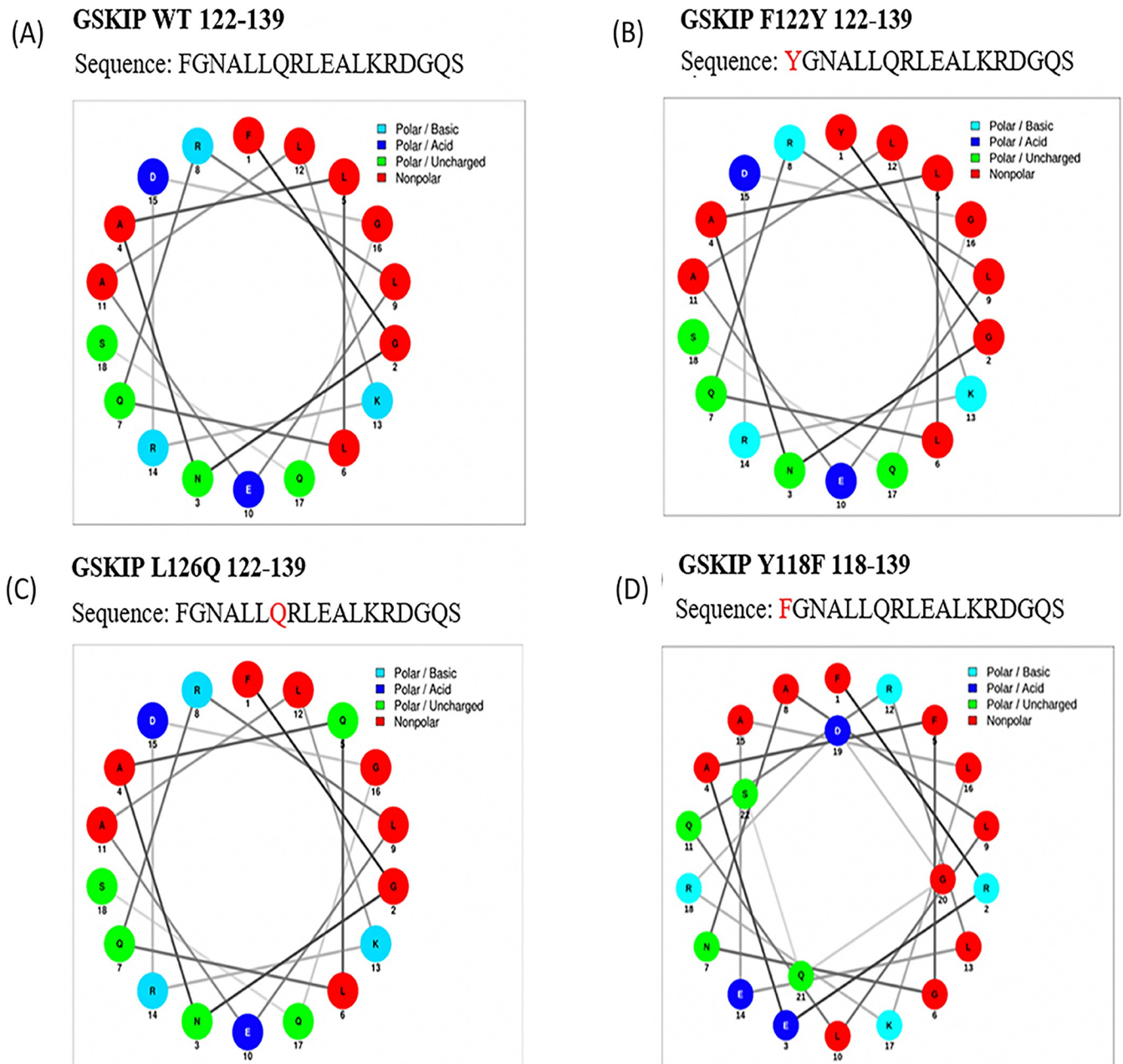
The red lines indicate the pre-GSK3 $\beta$ -binding site and various GSK3 $\beta$ -binding sites.

\* Data are comparable to those of Hundsrucker et al., 2010 [6].

<https://doi.org/10.1371/journal.pone.0262138.t001>

website. Only the hidden code SSF103107 was used as a metaphor for DUF727 and GSKIP for searching for ARMC4. Although the hidden code has been classified into the DUF727 superfamily, the hidden code SSF103107 domain shares homology with DUF727 (PF05303) and GSKIP (IPR007967). Therefore, we checked all ARMC4 family queries, and 38 of 449 species (8%) were found to contain the hidden code SSF103107, with protein evolution through gene (domain) fusion. Because the pre-GSK3 $\beta$ -binding site (115SPxF118) was retained only in the ARMC4 family in vertebrates, the GSK3 $\beta$ -binding site is not yet present at this stage. The pre-GSK3 $\beta$ -binding site (115SPxF118) may exist as a primitive type in vertebrates. We selected 18 SSF103107 orthologs from vertebrates to determine their alignment and performed T-Coffee phylogenetic tree construction of ARMC4 proteins evolved through gene fusion in 8% of vertebrate species (S1B Fig). We selected SSF103107 orthologs, and ClustalW and MEGA-X were used for phylogenetic tree construction (S1C Fig). Proteins encoded by ARMC4, 10 armadillo repeat motifs (ARMs), and one HEAT repeat were found in ARMC4, which is thought to be involved in ciliary and flagellar movement [43, 44]. The term armadillo is derived from the historical name of the  $\beta$ -catenin gene in the fruit fly *Drosophila*, where the armadillo repeat was first discovered. Although  $\beta$ -catenin was previously thought to be a protein involved in linking cadherin cell adhesion proteins to the cytoskeleton, a recent study indicated that  $\beta$ -catenin regulates the homodimerization of alpha-catenin, which in turn controls act branching and bundling [50]. However, the armadillo repeat has been found in a wide range of proteins with other functions. This protein domain plays a vital role in transducing Wnt signals during embryonic development [51].

As described in the earlier text, we found that a substantial number of DUF727 domain insertions into CLU1 (in mitochondria) together with ARMC4 (in the Wnt pathway) may constitute the gene fusion recombination mechanism. We also detected ARMC4 in the pre-

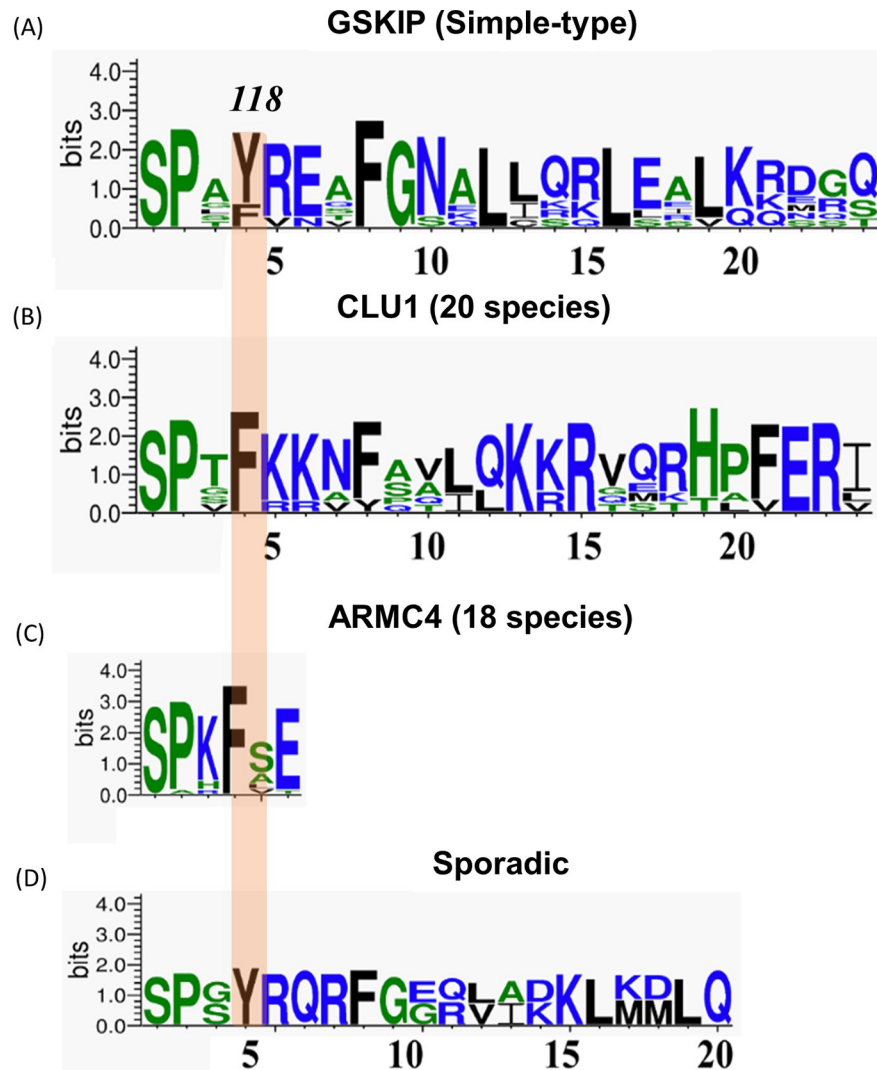


**Fig 4. Helical wheel analysis of GSKIP wt and various GSKIP mutants at amino acid sequence 122–139, which harbors the GSK3 $\beta$ -binding site.** A, GSKIP wt. B, GSKIP L126Q mutant. C, GSKIP F122Y mutant. D, GSKIP Y118F mutant.

<https://doi.org/10.1371/journal.pone.0262138.g004>

GSK3 $\beta$ -binding site (115SPxF118, Fig 5C), whereas the CLU1 family gradually formed from the pre-GSK3 $\beta$ -binding site to extend GSK3 $\beta$ -binding sites (115SPxF118xxxF122xxxQ/V126xxRV130, Fig 5B). However, no PKA-binding sites were found in composite-type GSKIP compared with simple-type GSKIP. We discovered evidence that the contribution of domain fusion to the evolution of multidomain proteins is bounded by the lower boundary of 63% in invertebrates and the upper boundary of 94% in vertebrates in the CLU1 family (Fig 3).

By contrast, in the vertebrate ARMC4 family, a cryptic (unintegrated) superfamily DUF727 (hidden code SSF103107) was also found to bind to the DUF727 domain in 8% of species with



Representation of simple-type (A) and Compositied-type Consensus sequences of GSKIP (B,C,D)

(A): S<sup>115</sup>Px[YF]<sup>118</sup>xxxL<sup>126</sup>xx[RKQ]<sup>129</sup>L<sup>130</sup>xx[L<sup>133</sup>

(B): S<sup>115</sup>PxF<sup>118</sup>xxx[FY]<sup>122</sup>xxx[QL]<sup>126</sup>xxR<sup>129</sup>[VGQT]<sup>130</sup>xx[HT]<sup>133</sup>

(C): S<sup>115</sup>PxF<sup>118</sup>xE<sup>120</sup>xM<sup>122</sup>

(D): S<sup>115</sup>PxY<sup>118</sup>xQxF<sup>122</sup>xxx[L<sup>126</sup>xxK<sup>129</sup>L<sup>130</sup>xxL<sup>133</sup>

[YF] means Tyr or phe; [RKQ] means Arg or Lys or Gln; x indicates any amino acid; ARMC4 contains pre-GSK3 $\beta$  binding sites in 20 species are also shown

**Fig 5. Differences in pre-GSK-binding site sequences.** Four consensus GSKIP sequences were represented using Weblogo 3.6.0. A, GSKIP (simple-type). B, CLU1 (CLUH). C, ARMC4. D, Sporadic (Mitofilin and Hsp70/Dnak).

<https://doi.org/10.1371/journal.pone.0262138.g005>

a pre-GSK3 $\beta$ -binding site (115SPxF118) (20 species compared with one bacterium). The association of DUF727 with the pre-GSK3 $\beta$ -binding site in vertebrate ARMC4 suggests that DUF727 (SSF103107) originates in a primitive stage of evolution. We suggest that the gene

fusion mechanism is a major contributor to the evolution of the CLU1 (in mitochondria) and ARMC4 (in the Wnt pathway) families in composite-type GSKIP.

### **Hijack (recombination) mechanism: A noncanonical order (strange order) of Hsp70/DnaK and Mitofilin in evolution**

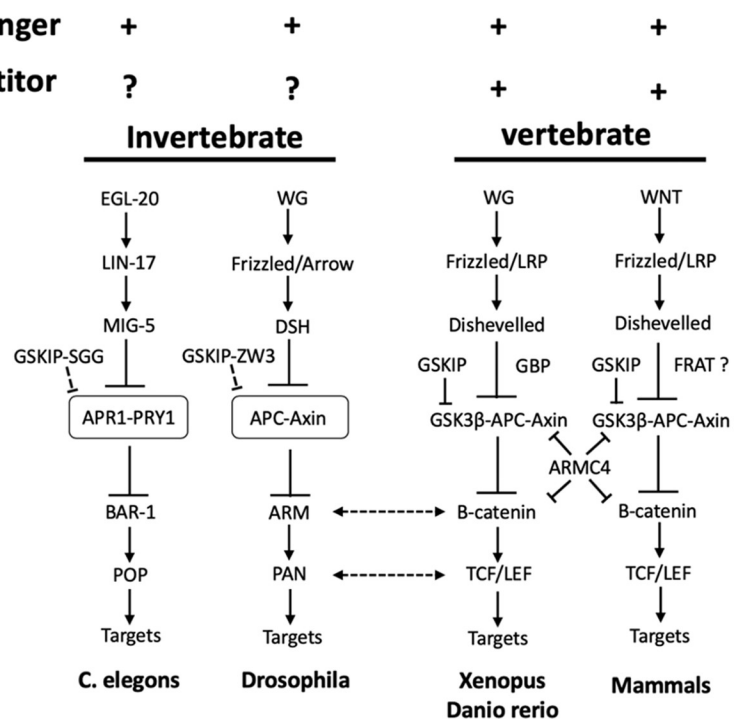
Members of the Hsp70 family are ubiquitously expressed and highly conserved; for example, the major form of Hsp70 from *E. coli*, termed DnaK, is approximately 50% identical to human Hsp70s. Hsp70 chaperone-assisted folding involves repeated cycles of substrate binding and release. Hsp70 activity is ATP-dependent. Hsp70 proteins comprise two regions: the amino terminus, which is the ATPase domain, and the carboxyl terminus, which is the substrate-binding region [52, 53].

Unexpectedly, in this study, we demonstrated the noncanonical order of the sporadic Hsp70/DnaK (A0A261CFR1\_9PELO, *C. latens*; A0A117XTM9\_HETBA, *Heterorhabditis bacteriophora*; A0A368GFB0\_ANCCA, *Ancylostoma caninum*; and A0A0D6M4D9\_9BILA, *A. ceylanicum*) and Mitofilin (H3EKR7\_PRIPA, *Pristionchus pacificus*) proteins found in the composite-type GSKIP of invertebrate nematodes; they exist in the mature forms 115SPxY118xxxFxxxLxxRL130 and 115SPxY118xxxFxxxV116xxKL130, respectively (Fig 5D). The pre-GSK3-binding site prefers 118Y instead of 118F in invertebrate nematodes compared with the CLU1 (all eukaryotes) and ARMC4 (vertebrate) families. These pre-GSK3 $\beta$ -binding sites utilizing 118Y instead of 118F to exhibit their GSK3 $\beta$ -binding activity were completely conserved in Mitofilin and Hsp70/DnaK in invertebrate nematodes but not vertebrates. The noncanonical order of these two composite-type GSKIPs in the evolutionary tree could be explained by a hijack (recombination) mechanism, as further evidenced by omega and T-Coffee comparisons. Apparently, the DUF727 domain gradually evolved as a pre-GSK3 $\beta$ -binding site (115SPxF118) flanking region with the consensus sequence site 115SPxF118xxx122FxxxLxxR/KL130 (Fig 5A), whereas in vertebrates, ARMC4 only contained the cryptic SSF103107 (unintegrated DUF727 superfamily, Fig 5C) flanking region with the pre-GSK3 $\beta$ -binding sites 115SPxF118 lacking the GSK3 $\beta$ -binding region (122FxxxLxxR/KL130). Altogether, these findings suggest that (1) at the pre-GSK3 $\beta$ -binding site, 118F is present prior to 118Y during evolution; (2) ARMC4 in higher organisms evolved through the GSK3-binding site of Hsp70/DnaK and Mitofilin; and (3) composite-type GSKIP evolved slower than simple-type GSKIP. We further performed site-directed mutagenesis and yeast two-hybrid assays to compare and reveal the importance of this pre-GSK3 $\beta$ -binding site with its flanking GSK3 $\beta$ -binding conserved sites in invertebrates and vertebrates (Table 1).

### **Human pre-GSK3 $\beta$ - and GSK3 $\beta$ -binding sites of the GSKIP domain in a budding yeast (*S. cerevisiae*) model**

We used budding yeast (*S. cerevisiae*) as an ideal model organism for studying domains with pre-GSK3 $\beta$ - and GSK3 $\beta$ -binding regions to determine how a new composite-type GSKIP was conserved in multicellular organisms through natural selection and it did not interfere with endogenous simple-type GSKIPs. As described in the earlier text, the pre-GSK3 $\beta$ -binding site (118F or 118Y) and various mutant GSK3 $\beta$ -binding sites (from CLU1) were compared through site-directed mutagenesis, as in a previous study, using yeast two-hybrid methods. Our data revealed 118Y, 122F, 126F, and 130L to be essential for binding during evolution, which is consistent with our previous study results [20]. Taken together, these data imply that the pre-GSK3 $\beta$ -binding site with 118Y plays a crucial role for GSK3 $\beta$ -binding sites as well as 122F, 126L, and 130L (Table 1) [1, 3, 20].

## A competitor or scavenger: is GSKIP surely involved in the Wnt signaling pathway?



**Fig 6. Physiological implications and importance of both simple-type and composite-type GSKIP in Wnt signaling.** GSKIP may play different roles in the Wnt pathway in different species. For *C. elegans* and *Drosophila*, due to a lack of the Axin GSK3β-binding motif on either the APR1-PRY1 or APC-Axin complex, GSKIP orthologs (GSKIP-SGG and GSKIP-ZW3) may serve only as scavengers that prevent complex formation. For vertebrates such as *Xenopus*, *D. rerio*, and mammals, GSK3β is one of the proteins involved in the formation of the destruction complex with β-catenin. Thus, GSKIP as a negative regulator may play a dual role as a scavenger to prevent GSK3β binding to the destruction complex or as a competitor for the Axin GSK3β-binding site with GBP or FRAT involved in the canonical Wnt pathway. ARMC4 may function as a competitor to APC and β-catenin. In addition, the pangolin protein (TCF/LEF) is involved in the Wnt signaling pathway in pangolins, an intermediate host species for SARS (modified from [19, 51]).

<https://doi.org/10.1371/journal.pone.0262138.g006>

### Physiological implications and importance of GSKIP in Wnt signaling

Our study revealed that the bat CLUH protein of composite-type GSKIP possesses a dominant consensus sequence (115SPxF118xxxF122xxxQ126xxRV130) quite similar to all the CLUH species described in the earlier text (S2 Table).

Overexpression of ATG2B and GSKIP increased progenitor sensitivity to thrombopoietin, enhancing megakaryocyte progenitor differentiation [11, 13]. The presence of the AK7 gene results in a predisposition to ciliary dyskinesia [43, 44, 49]. The protein encoded by ARMC4 containing 10 ARMs has also been reported to be involved in ciliary movement [43, 44].

The molecular mechanism for the coevolution of these neighboring genes with both simple-type (GSKIP-AK7) and composite-type GSKIP (ARMC4) in vertebrates remains unclear. A study suggested that pangolins are natural hosts of beta-coronaviruses, and comprehensive surveillance of coronaviruses in pangolins could improve our understanding of the spectrum of COVID-19 (the pandemic arising from a type of coronavirus called Sars-CoV-2; see [54] and S2 Table). Urgent investigation of the bat genome is required and may reveal the orchestration of specific elements in the COVID-19 crisis. ARMC4 may function as a competitor to

APC and  $\beta$ -catenin. In addition, the protein TCF4 in pangolins, as an intermediate host species of severe acute respiratory syndrome, indicates the involvement of the Wnt signaling pathway. This may be a coincidence but also raises the interesting question of what is the root cause of COVID-19 infection. Previously, we proposed that the ARMC4 family (as a composite-type GSKIP with a physiological role) together with simple-type GSKIP could function as competitors and is involved in the Wnt signaling pathway [20]. Both simple-type GSKIP and composite-type GSKIP (ARMC4) have been implicated in the Wnt signaling pathway (Fig 6) [50, 51, 55, 56].

## Discussion

According to UniProt data, we found the domain/fusion mechanism of the CLU and ARMC4 families and the hijack (recombination) mechanism for two sporadic genes of Hsp70/DnaK [52, 53] and Mitofilin [46, 47] that were hijacked by the DUF727 domain. The evolution of these composite-type GSKIPs was also revealed by the gene location in the corresponding organelles.

In this study, we detected gene fusion events by using UniProt, T-Coffee, and ClustalW. A total of 1748 simple-type GSKIP proteins as components of gene fusion in composite-type GSKIP were also detected, many of which were from the CLUH and ARMC4 families. The predicted functional associations were with mitochondria and the Wnt signaling pathway. We demonstrated for the first time that gene fusion is a complex evolutionary process for CLUH and ARMC4, including phylogenetic distance. Gene fusion plays a key role in the evolution of gene architecture. We can observe its effect if gene fusion occurs in coding sequences [57]. When gene fusion occurs in the assembly of a new gene, new functions emerge through the addition of peptide modules to the multidomain protein.

The detection methods for gene fusion events on a large biological scale can provide insights into the multimodular architecture of proteins [26, 58, 59].

In bacteria, a novel ligand-binding domain (ATP-binding motif), named the ACT domain, was identified through a PSI-BLAST search. *E. coli* 3PGDH was the first protein with an ACT domain that was found to fold into a ferredoxin-like (potential ATP in photosynthetic bacteria)  $\beta\alpha\beta\beta\alpha\beta$  topology. The ACT domain is found in a variety of contexts and is proposed to be a conserved regulatory ligand-binding fold. However, DUF727 has an antiparallel  $\beta\beta\beta\beta$  topology (NMR PDF file), sharing a similar sheet homology (Fig 1D and [2]) as the human CLUH--KIAA0664-SSF103107 superfamily domain (Fig 1E). Thus, Hsp70/Dnak, which requires ATP, chaperones, mitochondria, and Mitofilin, requires an ATP source for binding to CLU1. In particular, this standalone ACT domain protein might form complexes upon binding to other proteins, such as kinases, which interact with and regulate ARMC4,  $\beta$ -catenin, and APC in the Wnt pathway. We also observed that DUF727 fused to CLU1 and ARMC4. Moreover, domain accretion through a gene fusion mechanism may be a major contributor to gene evolution [60].

Three GSKIP-containing isoforms in humans are located in different chromosomes: GSKIP in chromosome 14, CLUH/KIAA0664 in chromosome 17, and ARMC4 in chromosome 10, and GSKIPs, CLUH, and ARMC4 still retain a DUF727. This may elucidate certain events in the evolutionary process. In particular, the unique order of Hsp70/Dnak and Mitofilin evolution may be evidence of another mechanism (hijacking). Our queries resulted in several new findings. First, we found a highly primitive form of GSKIP in lower organisms, such as gram-positive bacteria, and neither a PKA-binding domain nor a GSK3 $\beta$ -binding site was found when tracing GSKIP homologs in vertebrates and invertebrates [27, 37]. Second, the DUF727 domain in CLU1 and ARMC4, regarded as the central fragment of GSKIP, can exist

alone or with other domains as part of multidomain structures. Third, DUF727 first extended to a pre-GSK3 $\beta$  site at 115SPXF118 in vertebrate ARMC4 during evolution. Fourth, two sporadic composite-type GSKIPs (Mitofilin and Hsp70/Dnak) resulting from a hijacking mechanism were found in invertebrate nematodes. As shown in Fig 1F, at an evolutionarily later time, simple-type GSKIP acquired the PKA RII-binding domain as GSKIP/AKAP in vertebrates. Simple-type GSKIP is still retained in invertebrates. For composite-type GSKIP, one bacterium containing DUF727 with the ACT domain was found, whereas in all eukaryotes, DUF727 was improperly recognized as GSKIP. Although DUF727 lacks the ideal GSK3 $\beta$ -binding domain, it is still counted in CLU/TIF31 and CLUH/KIAA0664 sequences. In some vertebrates, the GSKIP domain was found in an ARMC4-containing protein with a pre-GSK3 $\beta$ -binding site only. The two sporadic proteins of Mitofilin and Hsp/DnaK were found in invertebrates containing perfect pre-GSK3 $\beta$ -binding and GSK3 $\beta$ -binding sites.

In sum, this study generated evidence demonstrating that the composite-type GSKIP of CLU1 and Mitofilin in mitochondria and the Hsp70 chaperone and ARMC4 have functions in Wnt signaling and simple-type GSKIP-linked proteins have roles in the origin of GSK3 $\beta$ -binding sites during DUF727/GSKIP evolution. Both simple-type and composite-type GSKIP (ARMC4) attract the most attention due to their degree of involvement in the Wnt signaling pathway (Fig 1F and Fig 6). Moreover, the entire genome of *Rhinolophus ferrumequinum* (the greater horseshoe bat) should be sequenced because doing so can provide insights into the importance of the bat genome. The likely insights into the bat genome might contribute to the efforts to end the current COVID-19 outbreak in so far as the insights may elucidate how GSKIPs function in bat mitochondria and Wnt signaling pathways; particularly in immunology studies may also help to understand how mother nature of bat genome harboring coronavirus could be used to combat emerging variants of this pandemic virus.

Two questions remain unanswered: first, under what conditions did this recombination occur to form composite-type GSKIP (to mitochondria, chaperone proteins, or armadillo repeats)? Second, in which organism did this recombination occur? For evolutionary biologists, these composite-GSKIP proteins can reveal the key steps in the evolution of GSKIP, particularly for composite-type GSKIP.

## Conclusions

Composite-type GSKIPs demonstrated the coevolution of pre-GSK3 $\beta$ - and GSK3 $\beta$ -binding sites that extended to DUF727 in the CLU1 and ARMC4 families, and the study findings may provide insights into the importance of both simple-type and composite-type GSKIP for the mitochondrial and Wnt signaling pathways.

## Supporting information

**S1 Table. Quote terms: SSF103107 (T3), PF05303 (T1), IPR007967 (G1), and IPR023231 (G3) of all species in the UniProt search.** T: type, G: group.  
(DOCX)

**S2 Table. Quote terms: “Clustered mitochondria protein” (CLUH) of Bat in UniProt search.**  
(DOCX)

**S1 Fig. Phylogenetic analysis of composite-type GSKIP, CLU1, and ARMC4.** Sequence alignment and phylogenetic analysis of ARMC4 from 18 vertebrate species along with a bacterium are shown. A and B, Phylogenetic tree of CLU1 and ARMC4 generated using MEGA-X, respectively. Boot strapping values represented the likelihood function for each other species



as indicated. The bootstrapping test measures the internal consistency of data produced above than 0.5 (50%) of the bootstrap replicates are consistent. C, Multiple sequence alignment with respect to the DUF727 region of GSKIP orthologs was conducted using ClustalW. D, T-Coffee estimates of alignment accuracy improved phylogenetic tree reconstruction. The conserved residues are indicated with asterisks, and residues with high similarity among the orthologs are marked with dots at the bottom. The label \* indicates the possible region of DUF727 in ARMC4.  
(TIF)

## Acknowledgments

We thank Gary Mawyer and Wallace Academic Editing for English edition on the manuscript.

## Author Contributions

**Conceptualization:** Cheng-Yu Tsai, Shean-Jaw Chiou, Huey-Jiun Ko, Chen-Yen Lin, Jiin-Tsuey Cheng, Joon-Khim Loh.

**Data curation:** Huey-Jiun Ko, Aij-Li Kwan.

**Formal analysis:** Chen-Yen Lin, Chihuei Wang.

**Project administration:** Yi-Ren Hong.

**Software:** Yu-Fan Cheng, Sin-Yi Lin, Yun-Ling Lai, Chen-Yen Lin, Chihuei Wang, Hsin-Fu Liu.

**Validation:** Hsin-Fu Liu.

**Visualization:** Sin-Yi Lin.

**Writing – original draft:** Cheng-Yu Tsai, Huey-Jiun Ko, Chen-Yen Lin.

**Writing – review & editing:** Cheng-Yu Tsai, Shean-Jaw Chiou, Huey-Jiun Ko, Yu-Fan Cheng, Yun-Ling Lai, Chen-Yen Lin, Jiin-Tsuey Cheng, Aij-Li Kwan, Joon-Khim Loh.

## References

1. Chou H-Y, Howng S-L, Cheng T-S, Hsiao Y-L, Lieu A-S, Loh J-K, et al. GSKIP Is Homologous to the Axin GSK3 $\beta$  Interaction Domain and Functions as a Negative Regulator of GSK3 $\beta$ . *Biochemistry*. 2006; 45(38):11379–89. <https://doi.org/10.1021/bi061147r> PMID: 16981698
2. Tang XN, Lo CW, Chuang YC, Chen CT, Sun YC, Hong YR, et al. Prediction of the binding mode between GSK3 $\beta$  and a peptide derived from GSKIP using molecular dynamics simulation. *Biopolymers*. 2011; 95(7):461–71. Epub 2011/02/18. <https://doi.org/10.1002/bip.21603> PMID: 21328310.
3. Howng S-L, Hwang C-C, Hsu C-Y, Hsu M-Y, Teng C-Y, Chou C-H, et al. Involvement of the residues of GSKIP, AxinGID, and FRATtide in their binding with GSK3 $\beta$  to unravel a novel C-terminal scaffold-binding region. *Molecular and Cellular Biochemistry*. 2010; 339(1):23–33. <https://doi.org/10.1007/s11010-009-0366-0> PMID: 20043192
4. Pilot-Storck F, Chopin E, Rual JF, Baudot A, Dobrokhotov P, Robinson-Rechavi M, et al. Interactome mapping of the phosphatidylinositol 3-kinase-mammalian target of rapamycin pathway identifies deformed epidermal autoregulatory factor-1 as a new glycogen synthase kinase-3 interactor. *Mol Cell Proteomics*. 2010; 9(7):1578–93. Epub 2010/04/07. <https://doi.org/10.1074/mcp.M900568-MCP200> PMID: 20368287; PubMed Central PMCID: PMC2938100.
5. Skroblin P, Grossmann S, Schäfer G, Rosenthal W, Klussmann E. Mechanisms of protein kinase A anchoring. *Int Rev Cell Mol Biol*. 2010; 283:235–330. Epub 2010/08/31. [https://doi.org/10.1016/S1937-6448\(10\)83005-9](https://doi.org/10.1016/S1937-6448(10)83005-9) PMID: 20801421.
6. Hundsrucker C, Skroblin P, Christian F, Zenn HM, Popara V, Joshi M, et al. Glycogen synthase kinase 3beta interaction protein functions as an A-kinase anchoring protein. *J Biol Chem*. 2010; 285(8):5507–

21. Epub 2009/12/17. <https://doi.org/10.1074/jbc.M109.047944> PMID: 20007971; PubMed Central PMCID: PMC2820779.
7. Lin CC, Chou CH, Howng SL, Hsu CY, Hwang CC, Wang C, et al. GSKIP, an inhibitor of GSK3beta, mediates the N-cadherin/beta-catenin pool in the differentiation of SH-SY5Y cells. *J Cell Biochem.* 2009; 108(6):1325–36. Epub 2009/10/16. <https://doi.org/10.1002/jcb.22362> PMID: 19830702.
  8. Loh JK, Lin CC, Yang MC, Chou CH, Chen WS, Hong MC, et al. GSKIP- and GSK3-mediated anchoring strengthens cAMP/PKA/Drp1 axis signaling in the regulation of mitochondrial elongation. *Biochim Biophys Acta.* 2015; 1853(8):1796–807. Epub 2015/04/30. <https://doi.org/10.1016/j.bbamcr.2015.04.013> PMID: 25920809.
  9. Dema A, Schröter MF, Perets E, Skroblin P, Moutty MC, Deák VA, et al. The A-Kinase Anchoring Protein (AKAP) Glycogen Synthase Kinase 3 $\beta$  Interaction Protein (GSKIP) Regulates  $\beta$ -Catenin through Its Interactions with Both Protein Kinase A (PKA) and GSK3 $\beta$ . *J Biol Chem.* 2016; 291(37):19618–30. Epub 2016/08/04. <https://doi.org/10.1074/jbc.M116.738047> PMID: 27484798; PubMed Central PMCID: PMC5016695.
  10. Deák VA, Skroblin P, Dittmayer C, Knobloch KP, Bachmann S, Klussmann E. The A-kinase Anchoring Protein GSKIP Regulates GSK3 $\beta$  Activity and Controls Palatal Shelf Fusion in Mice. *J Biol Chem.* 2016; 291(2):681–90. Epub 2015/11/20. <https://doi.org/10.1074/jbc.M115.701177> PMID: 26582204; PubMed Central PMCID: PMC4705389.
  11. Saliba J, Saint-Martin C, Di Stefano A, Lenglet G, Marty C, Keren B, et al. Germline duplication of ATG2B and GSKIP predisposes to familial myeloid malignancies. *Nat Genet.* 2015; 47(10):1131–40. Epub 2015/08/19. <https://doi.org/10.1038/ng.3380> PMID: 26280900.
  12. Duployez N, Lejeune S, Renneville A, Preudhomme C. Myelodysplastic syndromes and acute leukemia with genetic predispositions: a new challenge for hematologists. *Expert Rev Hematol.* 2016; 9(12):1189–202. Epub 2016/11/08. <https://doi.org/10.1080/17474086.2016.1257936> PMID: 27819178.
  13. Plo I, Bellanné-Chantelot C, Vainchenker W. ATG2B and GSKIP: 2 new genes predisposing to myeloid malignancies. *Mol Cell Oncol.* 2016; 3(2):e1094564. Epub 2016/06/17. <https://doi.org/10.1080/23723556.2015.1094564> PMID: 27308616; PubMed Central PMCID: PMC4905402.
  14. Vattulainen-Collanus S, Akinrinade O, Li M, Koskenvuo M, Li CG, Rao SP, et al. Loss of PPAR $\gamma$  in endothelial cells leads to impaired angiogenesis. *J Cell Sci.* 2016; 129(4):693–705. Epub 2016/01/09. <https://doi.org/10.1242/jcs.169011> PMID: 26743080; PubMed Central PMCID: PMC5108588.
  15. Dai FQ, Li CR, Fan XQ, Tan L, Wang RT, Jin H. miR-150-5p Inhibits Non-Small-Cell Lung Cancer Metastasis and Recurrence by Targeting HMG2 and  $\beta$ -Catenin Signaling. *Mol Ther Nucleic Acids.* 2019; 16:675–85. Epub 2019/05/24. <https://doi.org/10.1016/j.omtn.2019.04.017> PMID: 31121479; PubMed Central PMCID: PMC6529773.
  16. Nishimura T, Nakamura H, Yachie A, Hase T, Fujii K, Koizumi H, et al. Disease-related cellular protein networks differentially affected under different EGFR mutations in lung adenocarcinoma. *Scientific Reports.* 2020; 10(1):10881. <https://doi.org/10.1038/s41598-020-67894-7> PMID: 32616892
  17. Zeng F, Wang Q, Wang S, Liang S, Huang W, Guo Y, et al. Linc00173 promotes chemoresistance and progression of small cell lung cancer by sponging miR-218 to regulate Etk expression. *Oncogene.* 2020; 39(2):293–307. <https://doi.org/10.1038/s41388-019-0984-2> PMID: 31477834
  18. Li N, Cheng C, Wang T. MiR-181c-5p Mitigates Tumorigenesis in Cervical Squamous Cell Carcinoma via Targeting Glycogen Synthase Kinase 3 $\beta$  Interaction Protein (GSKIP). *Onco Targets Ther.* 2020; 13:4495–505. Epub 2020/06/18. <https://doi.org/10.2147/OTT.S245254> PMID: 32547080; PubMed Central PMCID: PMC7247609.
  19. Ko HJ, Chiou SJ, Wong YH, Wang YH, Lai Y, Chou CH, et al. GSKIP-Mediated Anchoring Increases Phosphorylation of Tau by PKA but Not by GSK3beta via cAMP/PKA/GSKIP/GSK3/Tau Axis Signaling in Cerebrospinal Fluid and iPS Cells in Alzheimer Disease. *J Clin Med.* 2019; 8(10). Epub 2019/10/24. <https://doi.org/10.3390/jcm8101751> PMID: 31640277; PubMed Central PMCID: PMC6832502.
  20. Chou C-H, Yang M-C, Hsiao B-X, Wang Y-H, Liu H-F, Chiou S-J, et al. The origin of GSKIP, a multifaceted regulatory factor in the mammalian Wnt pathway. *Biochimica et Biophysica Acta (BBA)—Molecular Cell Research.* 2018; 1865(8):1046–59. <https://doi.org/10.1016/j.bbamcr.2018.04.008> PMID: 29694914
  21. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* 2017; 45(D1):D190–d9. Epub 2016/12/03. <https://doi.org/10.1093/nar/gkw1107> PMID: 27899635; PubMed Central PMCID: PMC5210578.
  22. Universal Protein Resource (Uniprot). Available from: <https://www.uniprot.org/>.
  23. Interpro database website. Available from: <https://www.ebi.ac.uk/interpro>.
  24. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019; 47(D1):D506–d15. Epub 2018/11/06. <https://doi.org/10.1093/nar/gky1049> PMID: 30395287; PubMed Central PMCID: PMC6323992.

25. The UniProt C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*. 2021; 49(D1):D480–D9. <https://doi.org/10.1093/nar/gkaa1100> PMID: 33237286
26. Enright AJ, Ouzounis CA. Functional associations of proteins in entire genomes by means of exhaustive detection of gene fusions. *Genome Biology*. 2001; 2(9):research0034.1. <https://doi.org/10.1186/gb-2001-2-9-research0034> PMID: 11820254
27. Bateman A. The structure of a domain common to archaeobacteria and the homocystinuria disease protein. *Trends Biochem Sci*. 1997; 22(1):12–3. Epub 1997/01/01. [https://doi.org/10.1016/s0968-0004\(96\)30046-7](https://doi.org/10.1016/s0968-0004(96)30046-7) PMID: 9020585.
28. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22(22):4673–80. Epub 1994/11/11. <https://doi.org/10.1093/nar/22.22.4673> PMID: 7984417; PubMed Central PMCID: PMC308517.
29. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Big-ger Datasets. *Mol Biol Evol*. 2016; 33(7):1870–4. Epub 2016/03/24. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904.
30. <https://www.megasoftware.net>. Available from: <https://www.megasoftware.net>
31. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018; 35(6):1547–9. Epub 2018/05/04. <https://doi.org/10.1093/molbev/msy096> PMID: 29722887; PubMed Central PMCID: PMC5967553.
32. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. 2000; 302(1):205–17. Epub 2000/08/31. <https://doi.org/10.1006/jmbi.2000.4042> PMID: 10964570.
33. Di Tommaso P, Moretti S, Xenarios I, Orbitig M, Montanyola A, Chang JM, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*. 2011; 39(Web Server issue):W13–7. Epub 2011/05/12. <https://doi.org/10.1093/nar/gkr245> PMID: 21558174; PubMed Central PMCID: PMC3125728.
34. <http://tcoffee.org.cat/apps/tcoffee/do:expresso>. Available from: <http://tcoffee.org.cat/apps/tcoffee/do:expresso>
35. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012; 29(8):1969–73. Epub 2012/03/01. <https://doi.org/10.1093/molbev/mss075> PMID: 22367748; PubMed Central PMCID: PMC3408070.
36. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Available from: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
37. Chipman DM, Shaanan B. The ACT domain family. *Curr Opin Struct Biol*. 2001; 11(6):694–700. Epub 2001/12/26. [https://doi.org/10.1016/s0959-440x\(01\)00272-x](https://doi.org/10.1016/s0959-440x(01)00272-x) PMID: 11751050.
38. <http://weblogo.threeplusone.com>. Available from: <http://weblogo.threeplusone.com>
39. NetWheels: Peptides Helical Wheel and Net projections maker. Available from: <http://lbqp.unb.br/NetWheels>
40. Fields S, Song O-k. A novel genetic system to detect protein–protein interactions. *Nature*. 1989; 340(6230):245–6. <https://doi.org/10.1038/340245a0> PMID: 2547163
41. Chien CT, Bartel PL, Sternglanz R, Fields S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A*. 1991; 88(21):9578–82. Epub 1991/11/01. <https://doi.org/10.1073/pnas.88.21.9578> PubMed Central PMCID: PMC52761. PMID: 1946372
42. Zhu L. Yeast GAL4 two-hybrid system. A genetic system to identify proteins that interact with a target protein. *Methods Mol Biol*. 1997; 63:173–96. Epub 1997/01/01. <https://doi.org/10.1385/0-89603-481-X:173> PMID: 9113650
43. Hjej R, Lindstrand A, Francis R, Zariwala MA, Liu X, Li Y, et al. ARMC4 mutations cause primary ciliary dyskinesia with randomization of left/right body asymmetry. *Am J Hum Genet*. 2013; 93(2):357–67. Epub 2013/07/16. <https://doi.org/10.1016/j.ajhg.2013.06.009> PMID: 23849778; PubMed Central PMCID: PMC3738828.
44. Milara J, Armengot M, Mata M, Morcillo EJ, Cortijo J. Role of adenylate kinase type 7 expression on cilia motility: possible link in primary ciliary dyskinesia. *Am J Rhinol Allergy*. 2010; 24(3):181–5. Epub 2010/06/12. <https://doi.org/10.2500/ajra.2010.24.3468> PMID: 20537283.
45. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015; 43(Database issue):D204–12. Epub 2014/10/29. <https://doi.org/10.1093/nar/gku989> PMID: 25348405; PubMed Central PMCID: PMC4384041.
46. John GB, Shang Y, Li L, Renken C, Mannella CA, Selker JM, et al. The mitochondrial inner membrane protein Mitofilin controls cristae morphology. *Mol Biol Cell*. 2005; 16(3):1543–54. Epub 2005/01/14. <https://doi.org/10.1091/mbc.e04-08-0697> PMID: 15647377; PubMed Central PMCID: PMC551514.

47. Mun JY, Lee TH, Kim JH, Yoo BH, Bahk YY, Koo HS, et al. *Caenorhabditis elegans* Mitofilin homologs control the morphology of mitochondrial cristae and influence reproduction and physiology. *J Cell Physiol.* 2010; 224(3):748–56. Epub 2010/06/26. <https://doi.org/10.1002/jcp.22177> PMID: 20578245.
48. Fields SD, Conrad MN, Clarke M. The *S. cerevisiae* CLU1 and *D. discoideum* cluA genes are functional homologues that influence mitochondrial morphology and distribution. *Journal of Cell Science.* 1998; 111(12):1717–27. <https://doi.org/10.1242/jcs.111.12.1717>
49. Cox RT, Spradling AC. Clueless, a conserved *Drosophila* gene required for mitochondrial subcellular localization, interacts genetically with parkin. *Dis Model Mech.* 2009; 2(9–10):490–9. Epub 2009/07/28. <https://doi.org/10.1242/dmm.002378> PMID: 19638420.
50. Nusse R, Clevers H. Wnt/ $\beta$ -Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell.* 2017; 169(6):985–99. Epub 2017/06/03. <https://doi.org/10.1016/j.cell.2017.05.016> PMID: 28575679.
51. van Amerongen R, Berns A. Knockout mouse models to study Wnt signal transduction. *Trends Genet.* 2006; 22(12):678–89. Epub 2006/10/19. <https://doi.org/10.1016/j.tig.2006.10.001> PMID: 17045694.
52. Susin MF, Baldini RL, Gueiros-Filho F, Gomes SL. GroES/GroEL and DnaK/DnaJ Have Distinct Roles in Stress Responses and during Cell Cycle Progression in *Caulobacter crescentus*. *Journal of Bacteriology.* 2006; 188(23):8044. <https://doi.org/10.1128/JB.00824-06> PMID: 16980445
53. Xavier IJ, Mercier PA, McLoughlin CM, Ali A, Woodgett JR, Ovsenek N. Glycogen synthase kinase 3 $\beta$  negatively regulates both DNA-binding and transcriptional activities of heat shock factor 1. *J Biol Chem.* 2000; 275(37):29147–52. Epub 2000/06/17. <https://doi.org/10.1074/jbc.M002169200> PMID: 10856293.
54. Liu P, Jiang JZ, Wan XF, Hua Y, Li L, Zhou J, et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog.* 2020; 16(5):e1008421. Epub 2020/05/15. <https://doi.org/10.1371/journal.ppat.1008421> PMID: 32407364; PubMed Central PMCID: PMC7224457.
55. Zhang C, Zheng W, Huang X, Bell EW, Zhou X, Zhang Y. Protein Structure and Sequence Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique Similarity between Its Spike Protein Insertions and HIV-1. *J Proteome Res.* 2020; 19(4):1351–60. Epub 2020/03/24. <https://doi.org/10.1021/acs.jproteome.0c00129> PMID: 32200634; PubMed Central PMCID: PMC7099673.
56. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new coronavirus epidemic: Evidence for virus evolution. *J Med Virol.* 2020; 92(4):455–9. Epub 2020/01/30. <https://doi.org/10.1002/jmv.25688> PMID: 31994738; PubMed Central PMCID: PMC7166400.
57. Durrens P, Nikolski M, Sherman D. Fusion and Fission of Genes Define a Metric between Fungal Genomes. *PLOS Computational Biology.* 2008; 4(10):e1000200. <https://doi.org/10.1371/journal.pcbi.1000200> PMID: 18949021
58. Yanai I, Derti A, DeLisi C. Genes linked by fusion events are generally of the same functional category: a systematic analysis of 30 microbial genomes. *Proc Natl Acad Sci U S A.* 2001; 98(14):7940–5. Epub 2001/07/05. <https://doi.org/10.1073/pnas.141236298> PMID: 11438739; PubMed Central PMCID: PMC35447.
59. Pasek S, Risler J-L, Brézellec P. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics.* 2006; 22(12):1418–23. <https://doi.org/10.1093/bioinformatics/btl135> PMID: 16601004
60. Eichler EE. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.* 2001; 17(11):661–9. Epub 2001/10/24. [https://doi.org/10.1016/s0168-9525\(01\)02492-1](https://doi.org/10.1016/s0168-9525(01)02492-1) PMID: 11672867.