# Comparative Genomics and Functional Divergence of Ficolin-2 (FCN2) Across Various Types of Species: A Bioinformatics Approach

N. Wisittipanit

***Abstract*— Ficolin-2 (FCN2) is a critical pattern-recognition molecule essential for the innate immune system and the lectin complement pathway. This study used a comprehensive bioinformatics approach, including comparative genomics, phylogenetic analysis, and physicochemical property profiling, to investigate the FCN2 protein across 20 vertebrate and invertebrate species. Our analysis revealed a dual evolutionary pattern defined by strong conservation alongside significant structural and adaptive divergence. The result confirmed that the FIBRINOGEN_C_2 domain, responsible for pathogen recognition, is highly conserved across all 20 species, indicating a fundamental and ancient immune function. However, the Collagen domain, which provides structural integrity in vertebrates, is completely absent in all invertebrate species. Phylogenetic analysis reinforced this distinction, clearly separating the highly conserved vertebrate FCN2 sequences from the highly divergent invertebrate sequences. Furthermore, physicochemical analysis showed distinct adaptive trends linked to both phylogeny and ecology. FCN2 in invertebrates and aquatic organisms displayed significantly higher molecular weights, greater pI variability, higher instability indices, and increased hydrophilicity compared to their vertebrate and land-based counterparts. These differences suggest that while the core immune function of FCN2 is maintained, its molecular structure and properties have undergone species-specific adaptation to meet diverse physiological and ecological demands.**

***Keywords*—Ficolin-2, lectin complement pathway, comparative genomics, phylogenetic tree, functional divergence**

## I. Introduction

The innate immune system serves as the first line of defense against invading pathogens, relying on a diverse array of germline-encoded Pattern-Recognition Molecules (PRMs) to detect conserved microbial structures. Among these, ficolins represent a family of soluble PRMs characterized by a collagen-like region and a C-terminal fibrinogen-like domain (FBG). In humans, three ficolins have been identified: Ficolin-1 (M-ficolin), Ficolin-2 (L-ficolin), and Ficolin-3 (H-ficolin or Hakata antigen), encoded by the FCN1, FCN2, and FCN3 genes, respectively [1].

Ficolin-2, the focus of this study, is predominantly synthesized in the liver and secreted into the plasma [2]. It serves as an essential pattern-recognition molecule in the innate immune system. And it also plays a critical role in activating the lectin complement pathway by binding to acetylated carbohydrate moieties and other pathogen-associated molecular patterns (PAMPs) on microbial surfaces, leading to the recruitment of MBL-associated serine proteases (MASPs) and subsequent complement cascade activation [3]. This process ultimately facilitates opsonization and clearance of pathogens. Moreover, FCN2 has been implicated in the recognition of apoptotic cells and interaction with other immune components like pentraxin 3 [4]. Furthermore, FCN2 has been found to be the potential biomarker for hepatocellular carcinoma [5], the most common type of primary liver cancer.

Although FCN2's significance in human immunity is well established, the evolutionary conservation and functional adaptation of FCN2 among land-based and aquatic vertebrate and invertebrate species has not been thoroughly explored. This study addresses this gap by employing a multi-step bioinformatics approach: retrieval and categorization of FCN2 sequences from 20 species, protein domain analysis, multiple sequence alignment, motif detection, and comparative physicochemical property profiling. Results reveal a largely conserved multi-domain structure and function, especially within vertebrates, but also highlight significant divergence in domain presence, molecular weight, and biochemical indices between land, aquatic, and invertebrate groups. These evolutionary differences suggest species-specific adaptations of FCN2, driven by ecological pressures and immune demands, and advance our understanding of its molecular diversity and specialization across the animal kingdom.

## II. Materials and Methods

To achieve a deep analysis of FCN2 across species, a variety of bioinformatics methods were employed.

### *A. Sequence Dataset*

The protein sequences of FCN2 genes were retrieved from NCBI (National Center for Biotechnology Information) [6] across 20 organisms. Table 1 shows details of the retrieved FCN2 protein sequences.

Manuscript received Nov. 9, 2025.

N. Wisittipanit is with School of science, Mae Fah Luang University, Chiang Rai, 57100 Thailand; and Center of Innovative Materials for Sustainability (iMatS), Mae Fah Luang University, Chiang Rai, 57100, Thailand

TABLE I: DETAILS OF FCN2 SEQUENCES (PROTEIN)

| No | Source Organism | Size (aa) | No | Source Organism | Size (aa) |
|---|---|---|---|---|---|
| **1** | Homo sapiens | 302 | **11** | Macaca mulatta | 313 |
| **2** | Mus musculus | 314 | **12** | Macrobrachium nipponense | 383 |
| **3** | Eschrichtius robustus | 323 | **13** | Pacifastacus leniusculus | 430 |
| **4** | Gorilla gorilla gorilla | 313 | **14** | Petaurus breviceps papuanus | 313 |
| **5** | Balaenoptera acutorostrata | 328 | **15** | Phacochoerus africanus | 323 |
| **6** | Canis lupus dingo | 322 | **16** | Pongo abelii | 313 |
| **7** | Halocynthia roretzi | 324 | **17** | Sus scrofa | 323 |
| **8** | Hippopotamus amphibius kiboko | 300 | **18** | Pan troglodytes | 313 |
| **9** | Magallana hongkongensis | 297 | **19** | Xenopus laevis | 314 |
| **10** | Lonchura striata | 319 | **20** | Zonotrichia leucophrys gambelii | 323 |

Those 20 organisms are further categorized using two methods: (1) ecology method (2) vertebrate/invertebrate method. The ecology method is based on where the organisms live (land or water). The vertebrate/invertebrate method is simply based on whether the organisms has backbone or spinal column or not. Table II shows the categories by the ecology method.

TABLE II: CATEGORIES BY THE ECOLOGY METHOD

| Category | Organism |
|---|---|
| **Land-based Organisms** | Homo sapiens, Gorilla gorilla gorilla, Pongo abelii, Pan troglodytes, Macaca mulatta, Mus musculus, Sus scrofa (Pig), Canis lupus dingo, Phacochoerus africanus (Warthog), Petaurus breviceps papuanus (Sugar Glider), Lonchura striata (Finch), Zonotrichia leucophrys gambelii (Sparrow) |
| **Aquatic-based Organisms** | Halocynthia roretzi (Sea Squirt), Magallana hongkongensis (Oyster), Macrobrachium nipponense (Prawn), Pacifastacus leniusculus (Crayfish), Eschrichtius robustus (Grey Whale), Balaenoptera acutorostrata (Minke Whale), Xenopus laevis (African Clawed Frog), Hippopotamus amphibius kiboko |

Table III shows the categories by the vertebrate/invertebrate method.

TABLE III: CATEGORIES BY THE VERTEBRATE/INVERTEBRATE METHOD

| Category | Organism |
|---|---|
| **Vertebrates** | Homo sapiens, Mus musculus, Eschrichtius robustus, Gorilla gorilla gorilla, Balaenoptera acutorostrata, Canis lupus dingo, Hippopotamus amphibius kiboko, Macaca mulatta, Petaurus breviceps papuanus, Phacochoerus africanus, Pongo abelii, Sus scrofa, Pan troglodytes, Lonchura striata, Zonotrichia leucophrys gambelii, Xenopus laevis |
| **Invertebrates** | Macrobrachium nipponense, Pacifastacus leniusculus, Magallana hongkongensis, Halocynthia roretzi |

### *B. Protein Classification Analysis*

The protein from "Homo sapiens" with the size of 302 amino acids is used to show the generic FCN2 protein architecture and classification. The protein was classified by InterPro [7], bioinformatics tool specialized in classification of protein families, that searches for related protein families and essential domains.

### *C. Phylogenetic Analysis*

Protein sequences of identified FCN2 genes were aligned using the Clustal Omega tool (Multiple Alignment using hidden markov model profiling technique and guided trees) [8] with default parameters, followed by manual inspection and refinement. Phylogenetic trees, shown in Fig. 1, were constructed using the neighbor-joining method; the number adjacent to an organism name is the bootstrap value. The higher the value, the higher confidence in the grouping (clade).

### *D. Multiple Sequence Alignment*

The multiple sequence alignment among 20 sequences were performed using the MUSCLE tool [9]. It is a fast and highly accurate alignment tool, used to highlight regions of similarity, essential for identifying conserved regions or domains (which often have important structural or functional roles) and inferring evolutionary relationships between the sequences.

### *E. Physicochemical Properties of FCN2 Protein*

The physicochemical properties of FCN2 proteins was analyzed by Expasy's ProtParam tool [10] which analyze several essential physicochemical parameters of a protein by only a sequence input. Those properties include molecular weight, theoretical isoelectric point (pl), amino acid composition, atomic composition, extinction coefficient, instability index, aliphatic index, and grand average of hydropathicity (GRAVY). In this study, all 20 protein sequences were analyzed by the ProtParam tool.

## III. RESULTS

This section shows the analysis results from protein classification, phylogenetic analysis, multiple sequence alignment and physicochemical properties of those 20 FCN2 proteins.

### *A. Protein Classification Analysis Result*

Our protein classification analysis revealed that the FCN2 protein of human (Homo sapiens), length = 302 aa, belongs to the family of "Fibrinogen C-terminal domain-containing protein" which has diverse functions in angiogenesis (formation of new blood vessels), metabolism, cell adhesion and innate immunity. The domains of FCN2 protein comprises "Collagen" domain at position 39 to 84 and "FIBRINOGEN_C_2" domain at position 85 – 302. The Collagen domain, a triple helix structure (repeats of G-X-Y), specifically provides structural integrity and high tensile strength to the protein. The other domain: "FIBRINOGEN_C_2", which is a specific part of the globular C-terminal domain found in Fibrinogen-related

proteins (FREPs) that contains a conserved tertiary structure, provides essential functions, depending on the specific protein it is part of, in both blood coagulation and pattern recognition for innate immunity.

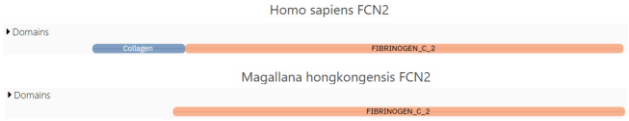


Fig. 1. Domains of FCN2 protein from Homo sapiens (vertebrates) and Magallana hongkongensis (invertebrates).

This two-domain architectural pattern of FCN2 proteins is largely preserved in most land-based and aquatic-based organisms, suggesting strong evolutionary pressure to maintain its fundamental modular structure. However, all four invertebrate organisms lack the collagen domain, suggesting that the FCN2 proteins of those organisms might not need the structural strength provided by the collagen domain, throughout their evolution. Fig. 1 shows domains of FCN2 protein from Homo sapiens (vertebrates) and Magallana hongkongensis (invertebrates). And Fig. 2 shows the 3D image of FCN2 protein from Homo sapiens.

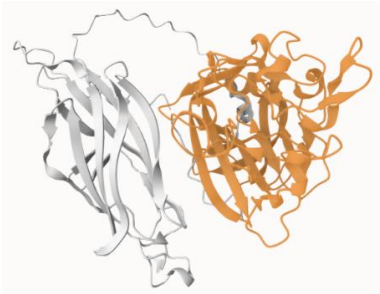

Fig. 2. 3D image of FCN2 protein from Homo sapiens; the grey part represents “Collagen” domain and the orange part represents “FIBRINOGEN_C_2” domain.

### B. *Phylogenetic Analysis Result*

The FCN2 phylogenetic analysis is primarily driven by the distinction between Vertebrates and Invertebrates. The 16 vertebrate sequences (mammals, birds, amphibian) are tightly clustered with exceptionally short branch lengths (minimum ~ 0.003), indicating a highly conserved FCN2 gene likely maintained by strong purifying selection due to its critical immune function. Conversely, the four invertebrate sequences (Macrobrachium, Pacifastacus, Magallana, Halocynthia) are positioned on the deepest, most divergent branches (lengths exceeding 0.28), signifying a much greater accumulation of substitutions and evolutionary distance. The influence of habitat is secondary, with the extremely high conservation confined to land-based placental mammals, although the clustering of aquatic whales with the land-based Hippopotamus confirms that the FCN2 sequence evolution follows recent taxonomic ancestry rather than current environment. Fig. 3 shows the phylogenetic tree of all 20 FCN2 proteins built from the Clustal Omega tool.

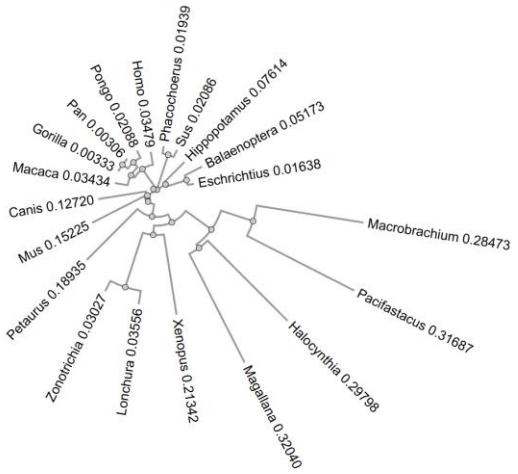


Fig. 3. Phylogenetic tree of all 20 FCN2 proteins.

### C. *Multiple Sequence Alignment Result*

The alignment result clearly shows that the front section of protein sequences (position 1 to 95 of the alignment) of all four invertebrate organisms (Table 3) align poorly to those of vertebrate organisms, which is precisely the collagen domain. However, all 20 organisms share the same FIBRINOGEN_C_2 domain (position 313 – 477).

### D. *Physicochemical Properties of FCN2 Protein*

The physicochemical analysis of all 20 FCN2 proteins shows distinct trends when comparing land-based organisms to aquatic-based organisms. In term of molecular weight, aquatic organisms like Eschrichtius robustus, Balaenoptera acutorostrata, Magallana hongkongensis, and Macrobrachium nipponense tend to have overall higher molecular weights for FCN2 compared to most land-based mammals such as Homo sapiens or Mus musculus. Moreover, several aquatic species display higher or more variable pI values (e.g., Magallana hongkongensis with a pI of 9.03), while land-based organisms generally cluster around a pI of 6-7, suggesting less variation in the charge properties of FCN2 in land-adapted proteins. Furthermore, aquatic organisms often present higher aliphatic indices (representing potentially increased protein thermostability) and instability indices, seen in species such as Magallana hongkongensis and Macrobrachium nipponense. On the other hands, land species tend to have intermediate aliphatic values with somewhat lower instability indices, indicating possible adaptation in different environments. Finally, across both groups, GRAVY values (hydrophobicity) tend to be negative but aquatic organisms like oysters and prawns display less negative values, suggesting slightly increased hydrophilicity for aquatic FCN2 proteins, possibly as an adaptation to water-rich habitats.

For the comparison of physicochemical properties between vertebrates and invertebrates organisms, invertebrates, such as Macrobrachium nipponense and Pacifastacus leniusculus often show higher FCN2 molecular weights (up to 48.39 kDa for crayfish) than those of vertebrates. Moreover, invertebrates

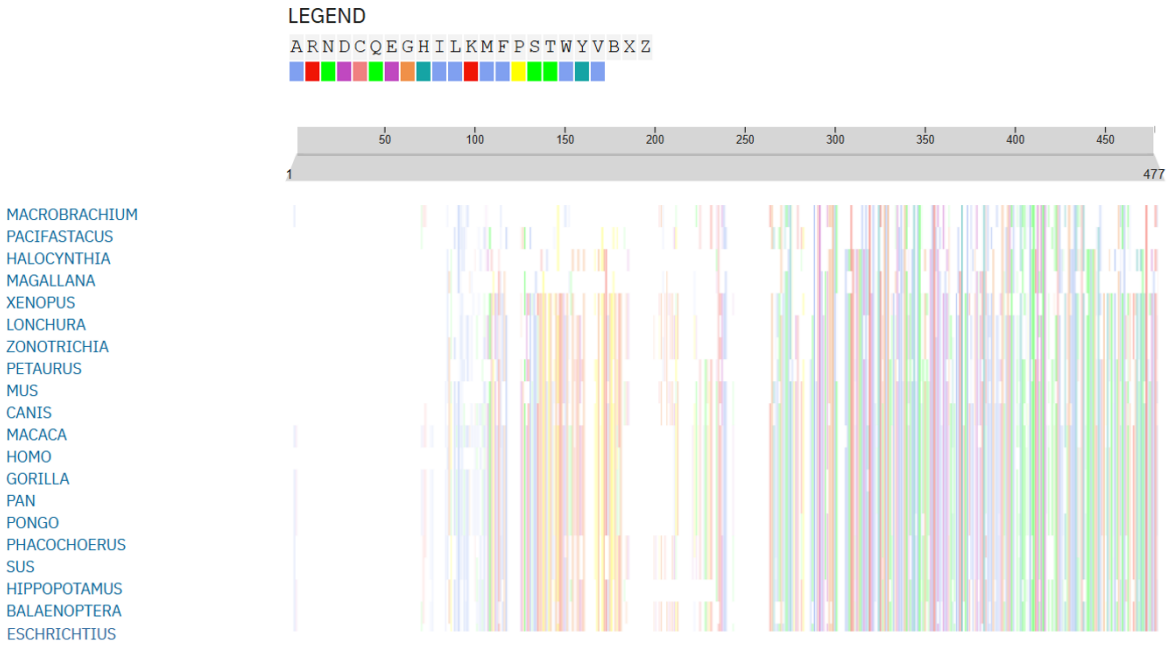


Fig. 3. Multiple alignment results of FCN2 protein sequences from 20 organisms

display greater variation in pI, e.g., oysters have a pI of 9.03, while crustaceans are lower (around 5.1–5.57); but vertebrates generally have a tighter pI range, 5 to 7.6. Furthermore, invertebrates often exhibit substantially higher instability indexes (up to 60.69 for crayfish) and aliphatic indices (up to 77.35 for crayfish), indicating potential adaptation for variable environments and structural differences in FCN2 proteins; while vertebrates remain between 20–38 for instability and around 60–66 for aliphatic index. Finally, both groups show negative GRAVY scores; but invertebrates, particularly aquatic species, the scores are less negative compared to those of vertebrates, showing slight increases in hydrophilicity.

## IV. Discussion

The bioinformatics methods employed in this study, including sequence retrieval from NCBI, protein classification with InterPro, phylogenetic analysis with Clustal Omega, alignment with MUSCLE, and physicochemical analysis with ProtParam, have provided a multifaceted view of FCN2 evolution across 20 distinct organisms.

The results strongly suggest a dual pattern of evolution. On one hand, the FIBRINOGEN_C_2 domain appears to be a core, ancestral component. This domain was identified in the Homo sapiens sequence as spanning positions 85-302 and is associated with the essential innate immunity function. The multiple sequence alignment confirmed that this domain is highly conserved, being shared across all 20 vertebrate and invertebrate organisms studied. This implies that the fundamental role of FCN2 in pattern recognition is deeply rooted and has been maintained throughout the evolution of these diverse species.

On the other hand, a major point of divergence is the collagen domain. In humans, this domain (positions 39-84) provides structural integrity. However, the protein classification analysis revealed that this domain is absent in all four invertebrate species analyzed. This finding was further corroborated by the multiple sequence alignment, which demonstrated that the invertebrate sequences aligned poorly in the corresponding N-terminal region (positions 1-95). This structural divergence suggests that the FCN2 protein in invertebrates may not require the specific tensile strength provided by the collagen domain, or that this feature is a specific adaptation of the vertebrate lineage. The phylogenetic analysis reinforces this clear split. The neighbor-joining tree not only grouped organisms largely according to established vertebrate phylogeny but also clearly separated the four invertebrate species into a distinct cluster.

Furthermore, the analysis of physicochemical properties revealed clear adaptive trends linked to both phylogeny and ecology. Invertebrates, such as Pacifastacus leniusculus (crayfish), exhibited substantially higher molecular weights (up to 48.39 kDa), greater pI variability, and significantly higher instability and aliphatic indices compared to vertebrates. A similar, though less pronounced, trend was observed when comparing aquatic-based organisms to land-based organisms. For example, aquatic species like Magallana hongkongensis (oyster) showed high pI (9.03) and instability indices.

A particularly notable finding was the difference in hydrophobicity. While FCN2 proteins were generally hydrophilic (negative GRAVY scores), both invertebrate and aquatic organisms displayed less negative GRAVY values. This suggests a slight increase in hydrophilicity, which may be an

TABLE II: PHYSICOCHEMICAL PROPERTIES OF 20 FCN2 PROTEINS

| Source Organism | MW(kDa) | pI | Positive Charge | Negative Charge | Extinction coefficients | Instability Index | Aliphatic Index | GRAVY |
|---|---|---|---|---|---|---|---|---|
| Homo sapiens | 33.00 | 6.3 | 31 | 33 | 60390 | 24.14 | 60.73 | -0.518 |
| Mus musculus | 33.98 | 7.0 | 32 | 32 | 61880 | 23.97 | 64.33 | -0.503 |
| Eschrichtius robustus | 34.56 | 6.3 | 32 | 35 | 54890 | 27.91 | 60.46 | -0.515 |
| Gorilla gorilla gorilla | 34.05 | 6.3 | 32 | 34 | 58900 | 23.6 | 62.97 | -0.465 |
| Balaenoptera acutorostrata | 34.92 | 7.1 | 34 | 34 | 54890 | 23.14 | 61.07 | -0.495 |
| Canis lupus dingo | 34.53 | 6.6 | 33 | 34 | 54890 | 27.95 | 60.31 | -0.524 |
| Halocynthia roretzi | 36.83 | 6.6 | 36 | 37 | 83310 | 28.17 | 65.28 | -0.663 |
| Hippopotamus amphibius kiboko | 32.49 | 7.5 | 33 | 32 | 53400 | 22.57 | 62.13 | -0.518 |
| Magallana hongkongensis | 33.54 | 9.0 | 37 | 30 | 75860 | 20.19 | 75.15 | -0.362 |
| Lonchura striata | 35.33 | 4.9 | 30 | 41 | 77350 | 38.09 | 60.56 | -0.455 |
| Macaca mulatta | 34.06 | 7.5 | 35 | 34 | 60390 | 19.09 | 66.39 | -0.499 |
| Macrobrachium nipponense | 43.54 | 5.1 | 41 | 53 | 71850 | 46.83 | 75.82 | -0.553 |
| Pacifastacus leniusculus | 48.39 | 5.5 | 37 | 52 | 65890 | 60.69 | 77.35 | -0.584 |
| Petaurus breviceps papuanus | 34.59 | 7.5 | 37 | 36 | 64860 | 28.89 | 56.39 | -0.567 |
| Phacochoerus africanus | 34.89 | 5.7 | 31 | 37 | 63370 | 23.1 | 61.92 | -0.521 |
| Pongo abelii | 33.98 | 6.6 | 33 | 34 | 60390 | 21.85 | 62.97 | -0.521 |
| Sus scrofa | 34.68 | 5.6 | 29 | 36 | 61880 | 26.44 | 62.85 | -0.46 |
| Pan troglodytes | 34.1 | 6.3 | 32 | 34 | 60390 | 24.59 | 62.04 | -0.486 |
| Xenopus laevis | 35.08 | 6.5 | 35 | 37 | 92360 | 27.91 | 65.76 | -0.619 |
| Zonotrichia leucophrys gambelii | 35.91 | 5.1 | 31 | 41 | 78840 | 38.8 | 60.4 | -0.473 |

adaptation to their water-rich environments. The tighter clustering of pI values (6-7) in land-based organisms and the lower instability indices in vertebrates may indicate a protein structure that has stabilized in correspondence with a more consistent internal environment.

## V. Conclusion

This study analyzed the FCN2 protein across 20 species by comparing protein architecture, phylogenetic relationships, sequence conservation, and physicochemical properties.

The results reveal two primary conclusions. First, the FCN2 protein possesses a highly conserved FIBRINOGEN_C_2 domain, which was present in all 20 species, including vertebrates and invertebrates. This strongly supports its fundamental and ancient role as a pattern-recognition molecule in the innate immune system.

Second, the protein exhibits significant functional divergence and adaptation. This is most evident in the complete absence of the structural collagen domain in the four invertebrate species studied. This structural divergence is accompanied by distinct physicochemical properties. Invertebrates and aquatic organisms, when compared to vertebrates and land-based organisms, respectively, tend to have higher molecular weights, more variable pI values, higher instability indices, and slightly increased hydrophilicity. These findings suggest that while FCN2's core immune function is preserved, its structure and biochemical properties have been adapted to meet the distinct physiological and ecological demands of different species.

## Acknowledgment

The author is grateful for Assistant Professor Dr. Ekachai Chukeatirote who help guide this bioinformatics study of FCN2 proteins.

## References

[1] M. Matsushita, and T. Fujita, "Ficolins and the lectin complement pathway," *Immunological Reviews*, vol. 180, pp. 78-85, 2001. https://doi.org/10.1034/j.1600-065X.2001.1800107.x

[2] U. Holmskov, S. Thiel, and JC. Jensenius, "Collectins and Ficolins: Humoral Lectins of the Innate Immune Defense," *Annual Review of Immunology*, vol. 21, pp. 547-578, 2003. https://doi.org/10.1146/annurev.immunol.21.120601.140954

[3] S. Bidula, H. Kenawy, Y.M. Ali, D. Sexton, W.J. Schwaeble, and S. Schelenz, "Role of Ficolin-A and Lectin Complement Pathway in the Innate Defense against Pathogenic Aspergillus Species," *Infection and Immunity*, vol. 81(5), pp. 1730-1740, 2013. https://doi.org/10.1128/IAI.00032-13

[4] Y.J. Ma, A. Doni, T. Hummelshoj, C. Honore, A. Bastone, A. Mantovani, N.M. Thielens, and P. Garred, "Synergy between Ficolin-2 and Pentraxin 3 Boosts Innate Immune Recognition and Complement Deposition," *J Biol chem*, vol. 284(41), pp. 28263-28275, 2009. https://doi.org/10.1074/jbc.M109.009225

[5] N. Wisittipanit, and E. Chukeatirote, "Identifying FCN2 as a Potential Biomarker for Hepatocellular Carcinoma Using a Novel Supervised Classification Method," *Genomics and Genetics*, vol. 12(1), pp. 19-27, 2019.

[6] Protein [Internet], Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2025 Oct 15]. Available from: https://www.ncbi.nlm.nih.gov/protein/

[7] M. Blum, A. Andreeva, L.C. Florentino, and S.R. Chuguransky, et.al, "InterPro: the protein sequence classification resource in 2025," *Nucleic Acids Research*, vol. 53(1), pp. 444-456, 2025. https://doi.org/10.1038/msb.2011.75

[8] F. Sievers, A. Wilm, D. Dineen, and T.J. Gibson, et.al, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," *Mol Syst Biol*, vol. 11(7), pp. 539, 2011. https://doi.org/10.1038/msb.2011.75

[9] R.C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Res*, vol. 32(5), pp. 1792-1797, 2004. https://doi.org/10.1093/nar/gkh340

[10] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, and A. Bairoch, "Protein Identification and Analysis Tools on the Expasy Server," in *The Proteomics Protocols Handbook*, J.M. Walker, Ed, Humana Press, 2005, pp. 571-607. https://doi.org/10.1385/1-59259-890-0:571