Bioinformatic Analysis of Structure and Function of Lim Domains of Zyxin Family Proteins

Mohammad Quadir Siddiqui[1,4], Maulik D. Badmalia[1,4] and Trushar R. Patel[1,2,3]*

Alberta RNA Research and Training Institute, Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada.

[2] Department *of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, 3330 hospital drive NW Calgary, AB, T2N 4N1, Canada.*

[3] *Li Ka Shing Institute of Virology, University of Alberta, Edmonton, AB, T6G 2E1, Canada.*

*Corresponding Author: Trushar R. Patel
trushar.patel@uleth.ca

4 – equal author contribution.

Running Title: Zyxin Family LIM Domains and their Functions

Keywords: Cancer, Zyxin, Lim Domains, Leucine Rich Motifs, Protein Bioinformatics.

ORCID - Quadir 0000-0003-0891-3122
ORCID - Maulik 0000-0002-3384-6550
ORCID - Trushar 0000-0003-0627-2923

Abstract

Lim domains are one of the most abundant types of zinc-finger domains and are linked with diverse cellular functions ranging from cytoskeleton maintenance to gene regulation. Zyxin family Lim domains perform the critical cellular functions and are indispensable for cellular integrity. Despite having these important functions the fundamental nature of the sequence, structure, functions, and interactions of the Zyxin family Lim domains are largely unknown. Therefore, we have used a set of *in-silico* tools and bioinformatics databases to distill the fundamental properties of the Zyxin family proteins/Lim domains from their amino acid sequence, phylogeny, biochemical analysis, post-translational modifications, structure-dynamics, molecular interactions, and functions. Consensus analysis of the nuclear export signal suggests a conserved Leucine-rich motif composed of LxxLxL/LxxxLxL. Molecular modeling and structural analysis demonstrate that Lim domains of the members of the Zyxin family proteins share similarities with transcriptional regulators, suggesting they could interact with nucleic acids as well. Normal mode, Covariance, and Elastic Network Model analysis of Zyxin family Lim domains suggest only the Lim1 region has similar internal dynamics properties, compared to Lim2/3. Protein expression/mutational frequency studies of the Zyxin family demonstrated higher expression and mutational frequency rates in various malignancies. Protein-protein interactions indicate that these proteins could facilitate metabolic rewiring and oncogenic addiction paradigm. Our comprehensive analysis of the Zyxin family proteins indicates that the Lim domains function in a variety of ways and could be implicated in rational protein engineering and might lead as a better therapeutic target for various diseases, including cancers.

Introduction

Lim (***L**in-11, **I**sl-1, and **M**ec-3*) domains are **zin**c **f**inger domains (ZnFs), first identified as a DNA-binding motif in the transcription factor TFIIIA found originally in African Clawed Frogs (*Xenopus laevis*) (1). Individual Lim domains have independent functions and their binding properties are modulated by the number of zinc fingers in the Lim region (2). ZnFs have binding competency to diverse biomolecular substrates such as DNA, RNA, proteins, and lipids (3-6). These domains display distinct metal binding characteristics and require zinc/iron or other metal ions to stabilize the finger-like folds (7). ZnFs classification is based on the zinc-finger domain architecture. Presently, 35 different ZnFs are available in the HUGO database (8), including the more common and abundant types of ZnFs such as - **C**ysteine**2**-**H**istidine**2** (C2H2), **R**eally **I**nteresting **N**ew **G**ene (RING), **P**lant **H**omeo**d**omain (PHD), and Lim domains (9).

The ZnFs play important roles in many critical biological processes including: gene transcription, gene silencing, translation, RNA packaging, DNA repair, ubiquitin-mediated protein degradation, cytoskeleton organisation, actin targeting, cell apoptosis, autophagy, epithelial development, cell adhesion, cell motility, cell migration, cell proliferation and differentiation, cell signaling and signal transduction, mechanotransduction, protein folding, lipid binding, chromatin remodeling and metal-ion sensing (1,6,10-12). Due to their integral cellular functions, the ZnFs are also implicated in many pathologies such as neurological & psychiatric disorders, congenital heart diseases, diabetes, skin disorders, cancer progression, and tumorigenesis (13-19). ZnFs are emerging as therapeutic targets and have been studied as candidates for metal-based therapy (20), and are also investigated for genome editing technologies (21).

Lim domains are one of the most common types of ZnFs found in eukaryotic organisms and are thought to be originated from common ancestors of plants, fungi, amoeba, and animals (9,22). Lim-type zinc fingers (Lim domains) are named after three transcription factor proteins (*LIN-11*, *Isl-1* and *MEC-3*) identified in Lin-11 from Roundworms (*Caenorhabditis elegans*), *Isl-1* from Rat (*Rattus norvegicus*), and *Mec-3* from Roundworms (23-25). These domains are stabilized by one or more zinc ions. Although the amino acids sequence which holds the zinc ions are variable and can vary significantly amongst various Lim domain proteins, they always include cysteine, histidine, aspartate, or

glutamate (26,27). The Lim consensus sequence has been identified as CX_2CX_{16-23} $HX_2CX_2CX_2CX_{16-21}$ CX_2(C/H/D) (X=any amino acid) (28). However, the analysis of all the identified human Lim sequences shows a slightly broader consensus sequence. The number and spacing of highly conserved amino acids (such as cysteine/histidine/aspartate) specify the metal-binding characteristics of a particular Lim domain (26,29). The highly conserved amino acids coordinate Zn ion to stabilize the Zn finger topology. The semi-conserved amino acids (bulky/aliphatic) are also found in the invariant position, but their role has not been elucidated yet. The number and properties of amino acids other than conserved or semi-conserved are also variable in different Lim domains (28).

Based on their amino acid composition, domain architecture, function, and cellular localization, Lim domains are classified into four different categories: **A)** Nuclear only, **B)** Lim-only, **C)** Lim actin-associated, and **D)** Lim-catalytic **(Figure 1)**. Sequence and phylogenetic analysis show that Lim proteins played a major role in the evolution of multicellular organisms (22). The Nuclear Only Lim domains family typically contains proteins with two Lim domains and a C-terminal homeodomain. The proteins of this family function as a transcriptional activator and potential transcription factor such as Isl-1 and Lhx1, respectively (28). As their name suggests, these proteins are located into the nucleus and regulate the gene expression, which in turn modulates the cell lineage determination, neurogenesis, and cell differentiation (24,28,30). The Lim Only protein family (LMO, CRP, FHL, PINCH) harbours 1 to 5 Lim domains, and these Lim domains (especially in case of LMO/CRP) have a high similarity to Isl-1 and Lhx family **(Figure 1)**. The Lim Only family proteins have an evolutionary relationship with both Nuclear Only and Lim-actin associated members suggesting that it could have originated from common ancestors by divergent evolution (22,28). Proteins from this family function as regulators of cell proliferation, differentiation, apoptosis, and autophagy (11,31-35). The largest Lim superfamily containing proteins is the Lim actin-associated group. It is composed of eight families (Paxillin, Zyxin, Testin, Enigma, ALP, ABLIM, EPLIN, and LASP), which possess a total of 26 proteins identified so far (28), and are present in the nucleus as well as in the cytoskeleton (26,28). The other proteins in this family are not specified yet because of Lim sequence promiscuity (22). The Lim-catalytic family includes Lim domains that are found in catalytic proteins that act as monooxygenases and kinases (22,28) **(Figure 1)**. These proteins play an important role in cytoskeleton remodeling, dynamics, and reorganization. The proteins of this family also function as signal transducers, regulates various actin-dependent biological processes including cell-cycle progression, cell motility, and differentiation (28,36,37). Additionally, they also promote microtubule disassembly, axonal outgrowth, and brain development (38,39). Structural studies report that all Lim domains bind to two zinc atoms, and display the topology of zinc finger motifs (29,37).

The comparison of biochemical, structural, and dynamics features of the Lim domains of Zyxin family provides us information on how they are involved with similar roles as well as dedicated functions. These characteristics might contribute to rescuing or complimenting the functions of other proteins in the same family as many of the Zyxin family proteins are an obligate requirement of tumorigenic properties in the cells (40-42). A recent study suggests that Zyxin acts as a biomarker for the aggressive phenotypes of human brain cancer (glioblastoma multiforme) and is associated with poor prognosis of glioma patients (43). Our analysis also discusses the role of the Zyxin family in oncogenic addiction (44) and metabolic rewiring (45). In this article, we have provided detailed insights into the nature and properties of Lim domains from Zyxin family proteins (Lim actin-associated group), which have unique zinc finger motifs and are involved in diverse cellular processes, especially in carcinogenesis. Zyxin family members are key regulators of many pathways including the cellular developmental pathways (28). All the members of the Zyxin family Lim domain contain three Lim domains at the C-terminus which is devoid of the homeodomain but retains a common presence of **P**roline-**r**ich **r**epeats (PRR) at the N-terminus **(Figure 2)**. In this study, we have harnessed a set of *in-silico* tools and databases to gain

insights into the amino acid compositions, phylogeny and their biochemical analysis, post-translational modifications, structure and dynamics, molecular interactions, and functions of the Lim domains of Zyxin family proteins which include Zyxin, LimD1, Ajuba, LPP, WTIP, FBLIM1, TRIP6 **(Figure2)**. Moreover, we have also performed Normal Mode and Covariance analysis to obtain information on biologically functional motions of Lim domains. This comprehensive comparative study provides insights into the molecular basis of the versatility of Lim domains and their functions.

Results

Zyxin family proteins exhibit variations in PRR/Lim edges.

Lim domain-containing proteins are present in different cellular locations such as cell membranes, cytoskeleton, cytoplasm, and in the nucleus. Some Lim proteins such as Lim homeodomain (28) (LHX family: constitute by 11 proteins), and nuclear LMO (28) (Lim Only proteins) (LMO1/2/3/4/X) are exclusively located into nuclear milieu (**Figure 1**). However, the Zyxin family Lim proteins are present in cell junctions (46-49), cell membranes (46,48), focal adhesions (36,37,41,50), cytoplasm (28,46,51), cytoskeleton (28,36,37,41,47,48), P-body (47,48), centrosome (47,48,52), and the nucleus (28,42,46,48) (see **Table 1** for more details). The amino acid sequence and composition, peptide length, net charge (isoelectric point), and biochemical features of Zyxin family Lim domains vary considerably. However, the presence of cysteine/histidine residues and their spacing are quite uniform in these domains. These Lim domains also require Zn^{2+} ion coordination for folding and stability (28). Each Lim domain contains a set of four cysteine residues that individually coordinate Zn^{2+} ions leading to the formation of Zinc finger topology arranged in tandem (26,28) **(Figure2B)**. These two clusters of cysteine residues are separated by two amino acid-long spacers, of which one is a semi-conserved aliphatic/bulky amino acid, and the other is a non-conserved (charged/hydrophilic/hydrophobic) amino acid that is indispensable for the Lim domains' function (28). The topology of Lim domain zinc fingers shows treble-clef like folds. However, nuclear Lim domains have different amino acids compositions compared to the cytoskeletal Lim domains. Although, some regions of Lim1/2/3 domains of the cytoskeletal Lim-actin associated superfamily do share similarities with Nuclear/Lim Only families. These variations of the amino acids of Lim1/2/3 in the non-conserved region imparts the functional diversity to the Lim domains in their respective families (6,26,28).

At the sequence level, we observed that the amino acid composition of the Lim family proteins, particularly Zyxin related family (Zyxin, LimD1, Ajuba, LPP, WTIP, FBLIM1, TRIP6) is quite intriguing. The sequence analysis revealed that the N-terminal is proline-rich, devoid of cysteine/histidine residues. Furthermore, these polyproline regions contain repeated unit clusters of 2 to 6 amino acids at the N-terminus but it is absent at the C-terminus of the Lim domain. Alternatively, the C-terminus contains three cysteine-rich regions (Lim1/2/3 Zn finger domains), which coordinate Zn^{2+} ions to adopt a stable conformation. The length of these Zyxin family-related polypeptides revolves around 373 to 676 amino acids and constitutes ~48-69% of PRR as well as ~30-51% of cysteine-rich domains (**Table 3**).

All Zyxin family proteins have three Lim domains namely Lim1, Lim2, and Lim3 (Lim 1/2/3), located at the C-terminus (**Figure 2).** Our amino acid sequence analysis indicated that the hydrophobic residues are non-conserved and highly variable in Lim1/2/3 domains **(Supplementary Figure 1)**. Each Lim domain is ~55-65 amino acids long with a consensus signature repeat of cysteine/histidine residues which form two zinc fingers (each finger consists of ~30 amino acids). Sequence alignment of the Zyxin family Lim domains depicts the presence of conserved cysteine/histidine/aspartate (C/H/D) residues (**Supplementary Figure 1).** Due to the abundance of the hydrophobic residues in Lim domains we decided to perform a hydrophobicity analysis to study the distribution of hydrophobic residues in all three Lim domains of Zyxin family proteins as well as in individual Lim1/2/3 domains. Additionally, the hydrophobicity of the peptide is a basic

determinant of a protein's topology (53), and different amino acids have variable alpha-helical or beta-sheet propensity (53,54). The hydrophobic/charged/neutral amino acid proportion analysis would also shed light on understanding the specific localization (55), and thereby the functionality of different members of the Zyxin family. The hydrophobic plots presented in **Figure 3** demonstrate that the Lim1 and Lim2 domains of Zyxin, TRIP6, LimD1, LPP, and Ajuba are more hydrophobic as compared to the Lim3 domain. In the case of FBLIM1 and WTIP, we observed that the hydrophobicity of the Lim2 domain was higher than Lim1 and Lim3 domains. Such differential hydrophobicity in Lim domains could be linked to their distinct properties, structures, and functions, besides having a conserved zinc finger motif. Further phylogenetic analysis of the Zyxin family proteins was carried out to identify the most primitive member of this group. The dendrogram presented in **Supplementary Figure 2** suggests that LimD1 is the primitive member and all other members have evolved out of it, and Zyxin is the most evolved member of the Zyxin family.

Next, we implemented NetNES (56) algorithm to identify the nuclear shuttling trait via the **n**uclear **e**xport **s**ignal (NES) sequence, which could have recently evolved or might exist ancestrally (**Supplementary Figure 2)**. Both phylogenetic, as well as NetNES (56) analyses, revealed that the Zyxin family proteins have a common nuclear shuttling trait, which is ancestral in nature (**Supplementary Figure 2**). These results are also supporting the hypothesis of Lim domain diversification and functions bloomed by the evolution of the nuclear roles in addition to cytoskeletal properties (22). Furthermore, we have identified the consensus sequence of the NES sequence in the Zyxin family proteins. Our analysis revealed that FBLIM1/WTIP/LimD1 and TRIP6/Ajuba display LxxLxL and LxxxLxL consensus sequence, respectively (L=L/V/I/F/M; x=any amino acid). However, Zyxin and LPP have some promiscuity and show the LxxxLxxL/LxxLxxL consensus sequence (L=L/V/I/F/M; x=any amino acid) **(Figure 4)**. Interestingly, all NES sequences are found in the PRR of each of the Zyxin family proteins. These findings are consistent with the available report which shows that 75% proteins have LxxLxL, 11% proteins have LxxxLxL and 15% proteins do not comply either of these consensus sequences (56).

The isoelectric point (pI), post-translational modifications (PTMs), cellular locations, the functions of Zyxin family proteins can be altered.

In order to understand the basic properties of the Zyxin family proteins and their Lim domains (Zyxin, LimD1, Ajuba, LPP, WTIP, FBLIM1, TRIP6), we have calculated the pI of the unmodified full length as well as the Lim domain (**Figure 2 & Table 1**). We observed that the peptide length for Zyxin family proteins ranges from 373 to 676 amino acids, and has pI values of 5.71 to 8.52. However, their Lim domains vary from 185 to 195 amino acids, with the pI values that range from 5.81 to 7.81 **(Table 1)**. Our analysis suggests that FBLIM1, which is found exclusively at cell adhesion sites, has a peptide length of 371 amino acids and a pI of 5.71. The acidic pI of FBLIM1 supports the fact that many acidic proteins present in cytosol, cytoskeleton, vacuoles, and lysosomes (57). However, WTIP, which is predominantly present in the nucleus, has a peptide length of 430 amino acids and a pI of 8.53. Other Zyxin family proteins such as Zyxin, LimD1, Ajuba, LPP, WTIP, TRIP6 has a length of 430-676 amino acids and a pI of 6.20-7.19. Zyxin, LimD1, Ajuba, LPP, WTIP, TRIP6 proteins are found in the cell cytoskeleton and nucleus. The pI values for the nuclear proteins are distributed in a wide range from 4.5 to 10 unlike cytosolic/membrane proteins (57,58). Although a comparable number of nuclear proteins have basic pIs (58). It clearly indicates that the peptide length and distribution of isoelectric point could play an important role in the localization of Zyxin family proteins and their niches.

Next, we analyzed the PTMs such as acetylation, mono-methylation, ubiquitylation, and phosphorylation of Zyxin family proteins which indicated that phosphorylation is the predominant PTM for Zyxin family proteins (**Table 2**). Since the pI has a critical role in protein distribution inside cells and phosphorylation impacts the pI of proteins (58,59), we also analyzed the pI of each Zyxin

family proteins considering their high phosphorylation levels (**Supplementary Figure 3 & Table 2**). Protein phosphorylation found in eukaryotes and prokaryotes drives the conformational switching of proteins/enzymes inside the cell and regulates the diverse cellular processes such as cell signaling, cell growth, metabolism, apoptosis, membrane transport, gene expression, and cell cycle activation (60,61). Various studies support the fact that altered phosphorylation is one of the hallmark features of carcinogenesis (62-64). Imatinib, gefitinib and trastuzumab are the novel anticancer drugs that prevent phosphorylation and mitigate the tumorigenic potential of the cancer cells (61,65,66). Phosphorylation modifications can also shift the pI of a target protein (59), and previous work has linked the phosphorylation levels that affect the net charge of the protein, with cancer predisposition (57-59). Phosphorylation also has the capability to mask and unmask the NES region which destines protein localization (67,68). The phosphorylation analysis of Zyxin family proteins (Zyxin, LimD1, Ajuba, LPP, WTIP, FBLIM1, TRIP6) by PhosphositePlus (69) indicated that the pI shift varied depending on the number of phosphorylation modifications and affects the pI considerably (**Table 2)**. It was observed that FBLIM1 which has a pI of 5.71 confined to only focal adhesion sites, whereas the WTIP with a pI of 8.53 is mainly confined to the nucleus. All other zyxin family proteins have a pI in the range of 6.20-7.19 which is found in focal adhesion sites, cell membranes, cytosol and the nucleus. Interestingly, LPP which is mainly involved in cell motility and gene transcription (46) undergoes the drastic range of a pI shift from 7.18 for unphosphorylated proteins to 2.84 with maximum of 83 phosphorylated sites as suggested by PhosphositePlus (69). Thus, our analysis established the link between the acidic, basic and near-neutral pI of Zyxin family proteins and their cellular localisation. In conclusion, it is plausible that fluctuation of pIs after phosphorylation/ dephosphorylation in the Zyxin family proteins ascertain their subcellular locations and niches (**Table 1 & 2**). These findings also support the fact that changes in pI is associated with the gain of additional activities, such as subcellular compartmentalization and changes in interacting partners that imparts functional differences of the proteins (70).

Structural features of Zyxin family Lim domains indicate their similarities with transcription factors.

Since it is established that the sequence of Zyxin family proteins has an important role in functioning and localization, we further sought to investigate their structural similarities and dissimilarities. Due to the absence of high-resolution structures of full-length proteins of this family, we decided to perform homology modeling for each sequence. The modeling performed using Robetta package as described in the methods section suggested that the majority of the Zyxin members irrespective of their length, except for Ajuba and TRIP6, adopt a hairpin shape such that the N- and C-terminus of the protein were on the same side and adjacent to each other **(Figure 5)**. In the case of TRIP6, both the terminals were positioned diametrically opposite to each other. However, for Ajuba, the C-terminus is located in the middle, whereas the N-terminus is present at the opposite end **(Figure 5).** Each model was then validated using Procheck analysis which employs the Ramachandran plot, which suggested that all the residues were in permitted regions of the Ramachandran plot implying that the models predicted for Lim domains are correctly folded (**Supplementary Figure 4**) (71). Absolute quality estimation is a parameter that allows a comparison of the modeled structure with reported PDB structures. The score of comparison is represented via Z score, where a Z-score in between 0.5 to 1 represents a model that closely resembles an experimentally determined native structure. **(Supplementary Figure 5**)(72). In **Q**uantitative **M**odel **E**nergy **AN**alysis (QMEAN) analysis of Lim domain models of Zyxin family proteins showed a Z-score ranging from 0.6 – 0.7 which is well in the acceptable range of QMEAN Z-score. The comprehensive model validation performed using PROCHECK and QMEAN analysis showed that all models have a good quality of stereochemistry as well as nativeness, respectively, making them reliable to be used for further studies. Investigation of structural similarities allows us to decipher the conserved domains which in turn helps to predict and

ascertain the functions of enigmatic proteins (73).

Lim domains of the Zyxin family are Zn fingers, which in turn are reported to be DNA binders. Therefore, we decided to investigate the presence of DNA binding stretches using DNAbinder webserver (74). DNAbinder trains a **S**upport **V**ector **M**achine (SVM) using a database to search for sequence similarity and motif finding approaches to discriminate DNA binding proteins from non-binders. For this study we implemented the analysis using amino acid composition mode, the webserver performed a PSI-BLAST analysis against a database containing DNA binding and non-binding proteins (1153 for each case). The SVM scoring was performed against a threshold of 0, with the resultant SVM indicating if the protein interacts with DNA or not. The negative SVM values suggest non-DNA binders whereas proteins with positive SVM scores are DNA binders. Our analysis indicated that all the members of the Zyxin family have positive SVM scores ranging from 0.11 (LPP) and 1.71 (WTIP), suggesting that all the proteins are DNA binding proteins **(Table 4)**. Using DNAbinder, we can identify if a protein interacts with DNA, however, we could not identify the regions or domains of a particular protein that specifically interact with the DNA. Therefore, we have used DALI server (73) to investigate the structural/domain similarities of zyxin family proteins with DNA binding proteins.

The DALI server analyzes protein structures in 3D by comparing the submitted structure queries against the deposited structures in the PDB server. This allows us to identify proteins with biological similarities that may not have been identified by mere sequence analysis (73,75). For the alignment of zyxin protein, the DALI server could successfully identify 88 structures from the PDB server. Interestingly, the majority of structures shared only 17 – 28 % sequence identity (% id), however, the Z score values showed high structural similarities with established transcription factors presented in **Supplementary Figure 6**. The Z-score depicts the similarity between the protein of interest with the proteins from the PDB database. It encompasses not only the structural similarities but the Structural Classification of Proteins (SCOP) similarities of the proteins as well. The SCOP database is a classification that organises proteins of the known three-dimensional structure according to common structural features and evolutionary relationships (76). The inclusion of SCOP helps us to remove biases introduced by size, substructures and divergence which may result in the server giving false-positive proteins showing similarity (73,77). Thus, we decided to further study structures with Z-score >9.4. This strategy provided us with a total of 6 structures for the subsequent analysis. Note that these structures showed only 23–28 % identity. The highest scoring structure in DALI analysis was a complex of Lmo4 protein and Lim domain (PDB ID: 1RUT) (78), followed by a complex of Lim Only protein 4 (LMO4) and Lim domain-binding protein (LDB1) (PDB ID: 2DFY) (79), the complex of Lim domain from homeobox protein LHX3 and islet-1 (ISL1) (PDB ID: 2RGT) (80), Lim/homeobox protein Lhx4, insulin gene enhancer (ISL2) (PDB ID: 6CME) (81) and an intramolecular as well as the intermolecular complex of Lim domains from homeobox protein LHX3 and islet-1 (ISL1) (PDB ID: 4JCJ) (82) **(**see **Supplementary Figure 6)**. The highest scorer 1RUT which included Lim domain transcription factor LMO4 (Lim Only) was first identified as a breast cancer autoantigen (83), which works as a transcriptional factor (84) and interacts with tumor suppressor Breast Cancer protein 1 (BRCA1) (85). The expression of LMO4 is developmentally regulated in the mammary gland and it is known to repress the transcriptional activity of BRCA1 (85). A common theme observed in all the highest scores protein complexes was that the participating Lim domains were part of transcription factors critical for developmental pathways. Furthermore, in case of candidates with Z-scores of <9.4, we observed the proteins such as LHX4 (Lim homeobox protein-4), ISL1 (Islet-1), CRP1 (**C**ysteine **R**ich **P**rotein 1), Rhombotin, TRIP6 (**T**hyroid **R**eceptor **I**nteracting **P**rotein 6), T-cell acute lymphocytic leukemia protein 1, LMO 2 (**L**im **O**nly 2), FHL2 (**F**our And A **H**alf **L**IM domains protein 2), Leupaxin, CRP2 (**C**ysteine **R**ich **P**rotein 2), PINCH2 (**P**articularly **I**nteresting **N**ew **C**ys-**H**is protein 2), chromatin structure remodeling complex subunit R5, RING finger protein (**Supplementary Figure 6**). The structural similarities of Zyxin Lim domains

with various transcriptional regulators suggest it may have a role in nucleic acid binding and its regulation. Interestingly, when we performed that DALI analysis for all the Zyxin family proteins (Zyxin, LimD1, Ajuba, LPP, WTIP, FBLIM1, and TRIP6), we observed a similar structural alignment pattern with common proteins such as Isl1, LHX3, LHX4, ISL1, CRP1, and Rhombotin. However, it is important to mention that TRIP6 showed a higher degree of similarity with PINCH, Paxillin, and Lim domain BI (**Table 5**). TRIP6 also displayed more unique interactions than other members; as described in protein-protein interactions section.

Zyxin family Lim domains show contrasting dynamics with large inter-domain motion of individual Lim1/2/3.

To understand the collective functional motions of Lim domains, NMA (**N**ormal **M**ode **A**nalysis) and Covariance analysis of Lim domains from Zyxin family proteins were performed using iMODS (86). This server calculates the normal modes from internal coordinates (torsion angles) which is more accurate than Cartesian approximations and implicitly maintains stereochemistry (86,87). The iMODs server performs the NMA of a biomolecule simulating a transition between two conformations, the output is generated using an affine model-based approach which helps us identify the rigid bodies and depicts their motion using simple curved arrows (86). The NMA analysis of Zyxin family proteins shows different Lim domain motions with respect to Lim1/2/3 (**Figure 6**). Covariance analysis indicates the variability of correlated, uncorrelated, and anti-correlated motions. The covariance matrix suggests the fluctuations for C-alpha atoms around their mean positions and coupling between pairs of the residues (88). It is depicted as correlated (red region), uncorrelated (white) and anti-correlated (blue) motions. However, it is important to point out that the difference in motion of Lim domains was observed in the ensemble motion of Lim1/2/3 domains of the protein, whereby the motion of individual domains within the protein strongly resembled each other (**Figure 7**). The Lim1 domain has the highest correlated motion as shown by red regions of the Covariance matrix, but in the case of Lim2, all proteins showed varied motion. On the other hand, the Lim3 domain of all proteins displayed similar motions (**Figure 6**).

Next, we employed iMODs for **E**lastic **N**etwork **M**odel (ENM) analysis of Zyxin family Lim domains to identify the pairs of atoms that are connected by springs. The results are depicted by a grayscale graph, where each dot represents one spring between two atom pairs. The degree of greyness for each dot represents the stiffness, dark grey represents stiffer springs and *vice versa*. Lim domains of all the Zyxin family proteins, specifically the Lim1 region show higher stiffness as compared to the Lim2 and Lim3 regions (**Figure 8)**. The areas which showed stiff grey regions were the same regions that showed correlated motions, suggesting the presence of stiff stretches of residues that gave correlated motion in Lim domains.

Mapping of protein-protein interactions of the Zyxin family proteins could provide insights into their crucial role in cancer progression.

The Zyxin family proteins are functionally diverse, playing an important role from cytoskeleton to transcriptional machinery, which is in turn possible only by protein-protein interactions. It is reported that the proline-rich region and Lim domain work as a docking site for the protein-protein interactions (28,36,89). Zyxin was first identified as a focal adhesion component in chicken embryo fibroblasts (90). It is present in a differential amount in cells like a smooth muscle cell, cardiac cells, brain & liver cells, and first interaction was identified with α-actinin (90,91). As mentioned above, zyxin contains N-terminal PRR, NES, and three Lim domains at the C-terminus (**Figures 1 & 2)**.

Zyxin is present in both the nucleus and the cytoplasm and is involved in influencing numerous signaling pathways. Furthermore, the cellular levels of Zyxin are increased in various malignant cancers (40,42,43,92). Therefore, we decided to explore the protein interactome for the Zyxin protein family using the available databases such as Bioplex (93), Genevestigator®, GeneMANIA (94), MINT (95), STITCH (96), Signor 2.0 (97) and STRING (98). First, we used the Genevestigator database that provides information on the

expression levels of proteins in various cancers. As presented in **Figure 9**, where the *x*-axis represents the protein expression level, the Zyxin family proteins are upregulated in almost all the different cancer types, including breast, colon, kidney, liver, and prostate cancers. Mutation frequency is another determinant of the protein function/nature exhibited by mutated protein causing subsequent consequences in different pathology and diseases (99). Therefore, we next performed a mutation frequency analysis using PhosphoSitePlus® databases. The resultant lollipop plots are shown for each Zyxin family proteins as a function of particular cancer types in **Figure 10**. These plots suggest that all members of the Zyxin family exhibit high levels of mutational frequency in different cancers. The mutation frequency is typically influenced by tissue types such as endometrial, stomach, lung, kidney, breast, and ovarian cancer (100). Our analysis suggests that all of the Zyxin family proteins have a higher mutational frequency in different cancer. However, their mutational frequency is varied in tissue types such as endometrial, stomach, lung, kidney, breast, and ovarian cancers **(Figure 10)**.

Using Bioplex analysis (93), we identified different cytoskeleton and gene regulatory proteins that interact with Zyxin family proteins. Bioplex depicts the protein-protein interactions and also the probabilities of association using the HEK293T cell line as a model. **Figure 11** presents the interacting proteins as prey (green circle) and bait (grey square) and the direction of interaction is represented by the arrow. The Bioplex analysis suggests that the Zyxin family proteins interact with each other. For example, Ajuba interacts with Zyxin, TRIP6, LIMD1, and LPP. LIMD1 was found to interact with TRP6 and Ajuba, whereas WTIP interacts with TRIP6.

Next, we performed physical association, co-expression, and pathway analysis for the Zyxin family proteins (Zyxin, LimD1, Ajuba, TRIP6, FBLIM, WTIP, and LPP) using Genemania (94), which manifested the involvement of each Zyxin family protein as a critical player that could alter the signaling pathways **(Figure 12)**. Using this pathway analysis, we found that Zyxin could be linked with NOLC1 (Nucleolar and coiled-body phosphoprotein 1), which facilitate ribosomal processing and modifications (101). Pathway analysis also suggests that Zyxin, FBLIM and LPP interact with VASP, indicating that their similar roles or compensatory nature where the effect of an absence of one of the proteins can be compensated by another protein of the same family. Pathway analysis by OmniPath (102) exhibits the role of Zyxin in the AKT and IL7R pathways (**Supplementary Figure 7**). Both of these pathways are linked with leukaemia (103) and T-cell development (104). This analysis also suggests that Zyxin regulates MAPK, GSK3B and CDK1 pathway indirectly. Using the Signor analysis, we found that Akt phosphorylates Zyxin on Ser142 and regulates apoptosis (92).

We have also used STITCH (**S**earch Tool for **I**nteractions of **C**hemicals), STRING (**S**earch **T**ool for **R**etrieval of **I**nteracting **G**enes/Proteins) and MINT (Molecular IN**T**eraction database) databases to mine the protein-protein interactions of the Zyxin family proteins (**Supplementary Figure 7)**. STITCH database provides information on interacting partners of a protein of interest, which could in turn provide a further understanding of their molecular/cellular functions (96). On the other hand, the STRING database provides enhanced coverage of protein associations with functional genome-wide discoveries (98). Upon performing the STITCH/STRING analysis, we observed that Zyxin interacts with BCAR1 (Breast cancer anti-estrogen resistance protein 1), which coordinates tyrosine kinase based signaling (105,106). Using the PHAROS server, we have analyzed the different transcription factors which harbour Lim domains. Lim domain occurrence in different transcription factors which have a clear and obligate role in gene expression of developmental pathways. These transcription factors are LHX and their isoforms, which we have also identified in our structural alignment section.

Discussion

Zyxin family of proteins are ubiquitously present in the cytoplasm and nucleoplasm (Figure 1). Despite belonging to the same family and having similar structures, they largely differ in terms of sequence, length, amino acid compositions, motif arrangement, biochemical, structural, and functional properties. Zyxin family members

could become an attractive target for cancer therapy, due to their multifaceted roles in cancer development and progression. Before exploring the avenues of drug discovery, it is important to understand the basic structure and function of a protein. A prior knowledge derived from computational work immensely aids in designing experiments which saves time and increases experimental efficiency. This work is an attempt towards the basic understanding of Zyxin family proteins especially Lim domains. Here, we have used various bioinformatics tools to explore the similarity and diversity in the protein as well as to understand what are the plausible roles that these proteins could play which can be further exploited for drug development

The Zyxin family proteins are functionally diverse, playing an important role from cytoskeleton to transcriptional machinery, which is in turn possible only by protein-protein interactions. It is reported that the proline-rich region and Lim domain work as a docking site for the protein-protein interactions (28,36,89). Zyxin was first identified as a focal adhesion component in chicken embryo fibroblasts (90). It is present in a differential amount in cells like smooth muscle cells, cardiac cells, brain & liver cells; their first interaction was identified with α-actinin (90,91). As mentioned above, zyxin contains N-terminal PRR, NES, and three Lim domains at the C-terminus (**Figures 1 & 2)**. Using the PhosphoSitePlus® webserver, we mapped the possible PTM sites on the proteins under study **(Supplementary Figure 3)**. Due to higher phosphorylation modification, Zyxin is also regarded as a phosphoprotein. Zyxin is required to undergo phosphorylation events to shuttle from the cytosol to the nucleus, and to modulate the apoptotic response (37,40,41,107). Moreover, a recent report demonstrated that zyxin is phosphorylated during mitosis and promotes colon cancer tumorigenesis in a mitotic-phosphorylation-dependent manner (42). Thus, we decided to investigate the phosphorylation sites of the Zyxin family proteins. Our analysis revealed that the majority of phosphorylation sites are located in the PRR of all the proteins observed (**Table 2**). As mentioned above, the PRR is a preferred docking site for multiple protein interactions (108), thus any changes in PTM might modulate the protein-protein interaction.

Despite having varied amino acid composition and length, the zyxin family members show a common architecture with the PRR at the N-terminal followed by three Lim domains located at the C-terminus (**Figure 2A**). The Lim domains have a Zn-finger motif arrangement, with a Zn^{2+} ion held in place by four cysteines. Each Lim domain has two such cysteine clusters holding Zn^{2+} ions in place (**Figure 2B**). The multiple sequence alignment of Lim domains suggested that the hydrophobic residues present in zinc finger are non-conserved **(Supplementary Figure 1)**. The subsequent hydrophobic analysis showed that the Lim domains have different hydrophobicity levels. For example, the Lim1 and Lim2 domains of Zyxin, TRIP6, LimD1, and LPP have higher hydrophobicity properties than Lim3. However, in the case of FBLIM1 and WTIP, Lim2 had higher hydrophobicity than Lim1 and Lim3 (**Figure 3**). This correlation was also reflected in the pI values of the Lim domains, Zyxin, TRIP6, LimD1, and LPP have a pI value lower than 7.5 whereas both FBLIM1 and WTIP have a pI value higher than 7.5 (**Table 1**). These findings are consistent with the bimodal distribution of acidic and basic proteins which in turn is correlated with the subcellular localization and cellular niches (57). The pI of a protein is a key determinant of the distribution of protein in a cell. Phosphorylation of proteins influences the pI as well as the shuttling of proteins into the nucleus through masking and unmasking of NES (68,109). Additionally, phosphorylation is also involved in activating proteins and enzymes to enable them to perform various cellular functions. Furthermore, alterations in the phosphorylation of a protein is one of the hallmark features of cancer (110). The PRR is a preferred docking site for multiple protein interactions (108), thus any changes in PTM might modulate the protein-protein interaction. As PRR is a favoured region for post-translational modifications, thus it will be interesting to further explore the mechanisms underpinning higher post-translational modification (PTM) for PRR region compared to the C-terminal Lim domains. In the case of the Zyxin protein family, Zyxin exhibits higher phosphorylation than other members, thus it is regarded as a phosphoprotein. Zyxin is required to undergo phosphorylation in order to shuttle

from the cytosol to the nucleus which in turn modulates the apoptotic response (37,40,41,107). Moreover, a recent report demonstrated that Zyxin undergoes phosphorylation during mitosis and promotes colon cancer tumorigenesis in a mitotic-phosphorylation-dependent manner (42). Thus, we decided to investigate the phosphorylation sites of the Zyxin family proteins using the PhosphoSitePlus® webserver **(Supplementary Figure 3)**. Our analysis revealed that the majority of phosphorylation sites are located in the PRR of all the member proteins (**Table 2**). Our phosphorylation analysis showed that it affects phosphorylation and influences the pI, which might potentiate additional activities, such as subcellular compartmentalization, change in interacting partners, that provide functional plasticity to these proteins.

Zyxin family proteins are reported to be localized in both cytoplasm and nucleus. Furthermore, some zyxin family proteins are capable of shuttling between nucleus and cytoplasm, a feature governed by the presence of the NES sequence. The alignment of NES sequences for Zyxin family proteins revealed a consensus motif of LxxLxL/LxxxLxL with some promiscuity in zyxin/LPP as LxxxLxxL/LxxLxxL (**Figure 4**). This sequence is present in all of the Zyxin family proteins irrespective of their reported function and localization. This analysis supports the hypothesis that the Zyxin family proteins are evolved out of a single ancestor and have retained the NES sequence **(Supplementary Figure 2)**. The homology modeling of the members of Zyxin family showed a structure such that the protein folded onto itself forming a hairpin-like fold with N- and C-termini present on the same site adjacent to each other **(Figures 5 and 6)**. These modeled structures were rigorously validated using PROCHEK/QMEAN and subjected to structural alignment and dynamics studies in order to identify the biologically relevant functional motion (**Supplementary Figures 4 and 5)**. Our main objective was to generate 3D models of Zyxin family proteins that can be used for dynamics studies to understand the evolutionary relationships and their functions.

Subsequently, we performed the Lim domains that are known to interact with proteins as well as nucleic acids (18,26,28,36,49,51,111,112). Typically, Lim1 mainly interacts with proteins (36) whereas Lim2/3 primarily has nucleic acid binding properties (18,27,37,47,111,113). However, Lim domains can also function as a protein-binding interface (36,89,112-115). The high-resolution structures of some of the Lim domains have been determined using nuclear magnetic resonance and X-ray crystallography, which suggests that the Lim domain zinc finger consists of two antiparallel β-hairpins with a short α-helix (78-80,82,116-119). The conserved cysteine/histidine/aspartate residues coordinate with the zinc atom which helps to maintain the secondary and tertiary structure of the Lim domains (28) (as depicted in **Figure 2**).

NMA of Zyxin family proteins using 3D models, which suggested that all the proteins show a pinching motion. In order to make a finer observation of the movement of the protein, we performed a Covariance analysis which showed that Lim1 had the most correlated motion across all the members. However, Lim2 and Lim3 showed varied motion across all the members **(Figure 7)**. Through the Covariance matrix, we observed that TRIP6 shows the highest correlated motion whereas Ajuba and WTIP have relatively lower correlated motion. We also observed that Ajuba and WTIP show similar motion, whereas LPP, LIMD1, and FBLIM have similar motion but Zyxin and TRIP6 have a completely unique motion amongst the Zyxin family (**Figure 7)**. Correlated motions are important for allosteric regulation, catalysis, ligand binding, protein folding, and mechanics of motor proteins (120-123). The correlated motion of the Lim1 region indicates that it has a higher propensity for binding towards its binding partner. However, the Lim3 region of each protein shows more anti-correlated motion. These transformations in the covariance matrix of Lim domains reflect the differences in the internal structural dynamics of the Zyxin family proteins, and these dynamic behaviours could confer diverse functions to each Zyxin family member as well as each LIM regions. Similarly, stiffness of the protein was investigated using Elastic Network Modeling (ENM). The observations made for stiffness were in corroboration with the covariance analysis: the stretches that showed the highest correlated motion also showed a high degree of stiffness **(Figure 8)**.

Several studies have linked the roles of Zyxin and other Zyxin family member proteins in cancer development and metastasis. For example, since Zyxin is a shuttling protein, it is reported that Zyxin has differential expression levels in melanocytes/melanoma cells and modulates cell spreading, proliferation, and differentiation (41). Furthermore, Zyxin was found to be a target of β-amyloid peptide and was implicated in Alzheimer's disease (124). Zyxin regulates the cellular growth by Hippo-Yorkie/Yki signaling (42), a conserved mechanism in mammalian cancer cells. Studies have reported that in colon cancer altered phosphorylation pattern of Zyxin protein is essential (42,125). In human breast cancer cells, Zyxin promotes tumorigenesis by Hippo-yes-associated protein (YAP) signaling pathway which has been involved in cancer progression and plays a crucial role in different human malignancies as well (42). Although, Zyxin was initially regarded as a focal adhesion protein (28,90), subsequent studies established that it has multifunctional properties and acts as a signal transducer (28,36,37). As described earlier, each Lim domain of zyxin has variable non-conserved regions. The presence of Zyxin in the nucleus and in the cytosol indicates its intriguing role in transcription or gene regulation (28,126). WTIP has a crucial role in cell proliferation and the downregulation of WTIP is associated with poor prognosis and survival of non□small□cell lung cancer patients (127). Interestingly, LIMD1, Ajuba, and WTIP discovered as a novel mammalian processing body (P-body) components and implicated in novel mechanisms of miRNA-mediated gene silencing (47). FBLIM1 promotes cell migration and invasion in human glioma by altering the PLC-γ/STAT3 signaling pathway and could be used as a molecular marker for early diagnosis in glioma patients (128). LPP was required for TGF-β induced cell migration/adhesion dynamics and regulates the invadopodia formation with SHCA adapter protein cooperation (129), and loss of LPP/Etv5/MMP-15 may be implicated in the prognostic marker of lung adenocarcinoma (130).

Bearing this in mind, we performed extensive protein interactome analysis for Zyxin family proteins that provided insights into the relationships between the Zyxin family proteins and cancer, and it was observed that Zyxin family proteins are upregulated in a variety of cancers (**Figure 9**). Taking into account that Zyxin is involved in the upregulation of various oncogenic pathways, the increased levels of these proteins in cancer cells further highlights the importance of investigating Zyxin proteins in carcinogenesis, progression, and detection of cancer. Next, we investigated the role of the mutational frequency of Zyxin proteins in various cancers, which revealed that Zyxin, LPP, LIMD1, TRIP6, and FBLIM displays high levels of mutational frequency in stomach cancer (**Figure 10**). Similarly, Zyxin, LIMD1 and TRIP6 mutations are linked with endometrial cancer. In the case of lung cancer, LPP displayed the highest amount of mutation frequency. For bladder, kidney and head/neck cancers, proteins TRIP6, WTIP, and Ajuba, respectively, showed the highest mutational frequency. The increased mutation frequency again highlights the role of zyxin member proteins in various cancers. Both these experiments helped us understand that the Zyxin protein family is highly susceptible to mutation and are upregulated in many cancers, but neither of the studies presented the proteins that the Zyxin proteins interact in a cancerous cell.

We studied the Zyxin protein family interactions using Bioplex. We observed many interactions with different proteins. This relationship is represented as bait and prey, depending on the direction of interactions. Interestingly, we observed that the Zyxin family proteins interact amongst themselves as well, with Ajuba being the focal point showing interactions with TRIP6, LIMD1, Zyxin, and LPP. WTIP and FBLIM1 do not show such interactions (**Figure 11 & Figure 12**). Zyxin interacts with bait and prey mode with VASP, FHL3, and TANC2 (**Figure 11A**). LimD1 and TRIP6 act more closely to regulate the interaction network amongst themselves (**Figure 11).** TRIP6 has the most complex interactome compared to other zyxin family members (**Supplementary Figure 8)**. Alternatively, WTIP displayed the smallest interaction network consisting of two proteins only of which one is TRIP6, the other protein is PPP2R3A which is reported to negatively control cell growth and division. However, the presence of mutated forms of this protein is reported in multiple cancers (131,132). An

intriguing observation was made in the case of FBLIM1 which predominantly exists in focal cell adhesion sites, and was shown to interact with isoforms of HIST1H3. The predominant function of FBLIM is reported as the interaction and maintenance of the cytoskeleton, thus it would be worth investigating as to why the Bioplex server reports a high degree of interaction of the cytoplasmic protein with a nuclear protein. Furthermore, we also found that the known interacting partners of Zyxin, such as ENAH/VASP proteins and other Lim domain-containing proteins also interact with the other members of the Zyxin family proteins. Such analysis could provide insights into the various roles Zyxin family members are involved with. For example, Zyxin-Ajuba interaction studies could further explore how Ajuba negatively regulates the Hippo signaling pathway, positively regulates the mi-RNA mediated gene silencing, and is required for mitotic commitment (46,47,108).

Our protein-protein interactions (PPIs) analysis could provide new insights into the functional diversity of the Zyxin protein family. We have used a set of databases to extract the interaction network of each member of the zyxin family, which in turn helps in understanding the impact and insights of zyxin family proteins at the whole cellular proteome level. Interaction of Zyxin with NOLC1 (Nucleolar and coiled-body phosphoprotein 1) and BCAR1 (Breast cancer anti-estrogen resistance protein 1) suggest a novel role of zyxin in ribosomal processing and in tyrosine kinase-based signaling, respectively (**Supplementary Figures 7)**. GeneMANIA explores functionally similar genes from the genomics/proteomics dataset which provide its predictive value for the functions/interactions, co-localization, and co-expression of the query gene with the proteins in the database. Zyxin, LimD1, and Ajuba showed more physical interactions as compared to WTIP, LPP, and FBLIM1. But the same could not be said in case of co-localization/co-expression, where Zyxin, FBLIM1, Ajuba, and TRIP6 showed higher degrees of similarity. It is important to mention here that for all the Zyxin family members, a similar degree of co-expression and colocalization was observed except for LPP. LPP demonstrated to have higher degrees of co-expression where a direct and indirect co-expression was observed with many genes, particularly TAGLN, CNN1, and SMTN (**Figures 12)**. Out of this TAGLN and CNN1 are reported as a putative marker for bladder cancer, whereas SMTN is reported to interact with Breast Cancer Metastasis Suppressor 1 (BRMS1), a protein known to suppress metastasis in multiple carcinomas (133,134). Many zyxin family proteins such as Zyxin, WTIP, and LPP, are directly implicated in the cancer prognosis and acts as a biomarker for certain tumorigenic conditions (40,127,133).

The present *in-silico* studies and comprehensive analysis of the Zyxin family proteins clearly indicate that these proteins can function in a variety of ways from protein-protein to protein-nucleic acid interactions. This highlights the importance of targeting the Zyxin family proteins for the development of the better therapeutic intervention in different disease conditions, particularly for cancers.

Methods

Sequence, amino acid composition, phylogeny, and basic biochemical analysis.

All the Zyxin family (Zyxin, LimD1, Ajuba, LPP, WTIP, FBLIM1, TRIP6) sequences were retrieved from the UniProt database as outlined in **Table 1** (135). Amino acid sequence alignments were performed using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) which also offers identification of critical amino acid residues (136). Sequence PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool) was performed to find distant evolutionary related proteins. For the phylogenetic analysis, Zyxin family sequences were subjected to Clustal W (137), the output was then imported into Jalview (138) and a tree was generated through neighbour joining using a BLOSUM62 (139). **N**uclear **e**xport **s**ignal (NES) prediction analysis was performed using NetNES (56) which calculates the scores from the Hidden Markov Model (HMM) and Artificial Neural Network (ANN). To determine the consensus sequence, conserved Leucine-rich NES signal analyzed from NetNES (56) aligned by MUSCLE (140) and logo was generated using WEBLOGO (141). ProtParam was used to calculate amino acid composition and basic

biochemical characteristics (142). We used the program ProtScale to perform hydrophobicity analysis (142). Furthermore, PhosphoSitePlus v6.5.9.1 was used to calculate the number of phosphorylation modification sites as well as to determine the isoelectric point in the different phosphorylated states (69).

Structure modeling, validation and alignment of Zyxin family Lim domains

Zyxin family Lim domain amino acid sequence was retrieved from UniProtKB (135), as UniProt IDs detailed in **Table 1** were utilized to build homology models using the Rosetta server (143). The Robetta server performs the comparative modeling with the available structure using BLAST, PSI-BLAST, and 3D-Jury packages or employs *de novo* modeling approaches using the *de novo* Rosetta fragment insertion method (144). Next, all the Lim domain models of Zyxin family proteins were validated by SWISS-MODEL workspace (145,146) encompassing the Anolea, DFire, QMEAN, Gromos, DSSP, Promotif, and ProCheck packages. Structure alignment of Zyxin family Lim domains was performed by the Dali server (73,75).

Normal mode analysis (NMA), Elastic network modeling and Covariance analysis.

NMA and Covariance analyses were performed using the iMODS server using default settings (86). After structure submission in PDB format or PDB ID, iMODS calculates the lowest frequency modes (normal mode represents the biological motions) in internal coordinates of a single/trajectories structure. This server calculates the lowest frequency normal modes in internal coordinates and utilizes the improved faster NMA calculations by implementing the iterative Krylov-subspace method (147). iMODS also provides motion representations with a vector field, deformation analysis, eigenvalues, and covariance maps. It also has an affine model-based arrow representation of the dynamic regions (86). Furthermore, the iMODS server also helps us understand how correlated the motion of residues in the protein is as well as how flexible the movement of each residue is under the Normal Mode Analysis.

Protein-protein interactions of Zyxin family Lim domains.

To explore the interactome of the Zyxin family Lim domains, we have used Bioplex explorer (Bioplex 3.0) (93). Protein-protein interactions define the involvement of the desired protein in a particular pathway that would help to map the pathogenic network. We have utilized the Pharos database to extract the role of Lim domains in cancer and autoimmune diseases (148). Pharos interface provides the collective information of the query protein by extracting the information from other databases, particularly in a disease context. Pharos is helpful in browsing the relevant information of the desired target in a systematic manner (148). To get the information about the signaling pathways influenced by the Zyxin family Lim domains, we have used OmniPath (102) and Signor 2.0 (97) programs. We also used Genemania (94) that provides insights into protein-protein, protein-DNA and genetic interactions, pathways, gene and protein expression data, protein domains, and phenotypic screening profiles. We also explored the experimentally determined and predicted protein-protein interactions using the STRING database (98). Next, we performed an analysis of the chemical and protein interactions using the program STITCH (96). We also utilized the **M**olecular **INT**eraction database (MINT) database to identify the interactions mediated by the Zyxin family Lim domains (95). MINT focuses on experimentally determined protein-protein interactions mined from the available literature. Next, we used GENEVESTIGATOR to analyze transcriptomic expression data from repositories (149). The collection of gene expression data from different samples (tissue, disease, treatment, or genetic background) with graphics and visualization tools (149).

Acknowledgements
We thank Canada Research Chair program for their support with this work.

Funding and additional information
MDB is funded by Alberta Innovates Strategic Research Projects to TRP. MQS is funded by Canada Research Chairs program to TRP. TRP is a Canada Research Chair in RNA and Protein Biophysics.

Conflict of Interest: Authors declare no competitive interests in regard to this manuscript.

References

1. Klug, A. (1999) Zinc finger peptides for the regulation of gene expression. *J Mol Biol* **293**, 215-218
2. Hammarstrom, A., Berndt, K. D., Sillard, R., Adermann, K., and Otting, G. (1996) Solution structure of a naturally-occurring zinc-peptide complex demonstrates that the N-terminal zinc-binding module of the Lasp-1 LIM domain is an independent folding unit. *Biochemistry* **35**, 12723-12732
3. Hall, T. M. (2005) Multiple modes of RNA recognition by zinc finger proteins. *Curr Opin Struct Biol* **15**, 367-373
4. Brown, R. S. (2005) Zinc finger proteins: getting a grip on RNA. *Curr Opin Struct Biol* **15**, 94-98
5. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M., and Mackay, J. P. (2007) Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci* **32**, 63-70
6. Matthews, J. M., and Sunde, M. (2002) Zinc fingers--folds for many occasions. *IUBMB Life* **54**, 351-355
7. Krishna, S. S., Majumdar, I., and Grishin, N. V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res* **31**, 532-550
8. Eyre, T. A., Ducluzeau, F., Sneddon, T. P., Povey, S., Bruford, E. A., and Lush, M. J. (2006) The HUGO Gene Nomenclature Database, 2006 updates. *Nucleic Acids Res* **34**, D319-321
9. Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G., and Raschella, G. (2017) Zinc-finger proteins in health and disease. *Cell Death Discov* **3**, 17071
10. Baltz, R., Evrard, J. L., Domon, C., and Steinmetz, A. (1992) A LIM motif is present in a pollen-specific protein. *Plant Cell* **4**, 1465-1466
11. Han, S., Cui, C., He, H., Shen, X., Chen, Y., Wang, Y., Li, D., Zhu, Q., and Yin, H. (2020) FHL1 regulates myoblast differentiation and autophagy through its interaction with LC3. *J Cell Physiol* **235**, 4667-4678
12. Elbediwy, A., Vanyai, H., Diaz-de-la-Loza, M. D., Frith, D., Snijders, A. P., and Thompson, B. J. (2018) Enigma proteins regulate YAP mechanotransduction. *J Cell Sci* **131**
13. Ahmad, S., Wang, Y., Shaik, G. M., Burghes, A. H., and Gangwani, L. (2012) The zinc finger protein ZPR1 is a potential modifier of spinal muscular atrophy. *Hum Mol Genet* **21**, 2745-2758
14. Pao, P. C., Huang, N. K., Liu, Y. W., Yeh, S. H., Lin, S. T., Hsieh, C. P., Huang, A. M., Huang, H. S., Tseng, J. T., Chang, W. C., and Lee, Y. C. (2011) A novel RING finger protein, Znf179, modulates cell cycle exit and neuronal differentiation of P19 embryonal carcinoma cells. *Cell Death Differ* **18**, 1791-1804
15. Birnbaum, R. Y., Zvulunov, A., Hallel-Halevy, D., Cagnano, E., Finer, G., Ofir, R., Geiger, D., Silberstein, E., Feferman, Y., and Birk, O. S. (2006) Seborrhea-like dermatitis with psoriasiform elements caused by a mutation in ZNF750, encoding a putative C2H2 zinc finger protein. *Nat Genet* **38**, 749-751
16. Molkentin, J. D., Lin, Q., Duncan, S. A., and Olson, E. N. (1997) Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* **11**, 1061-1072

17. Senee, V., Chelala, C., Duchatelet, S., Feng, D., Blanc, H., Cossec, J. C., Charon, C., Nicolino, M., Boileau, P., Cavener, D. R., Bougneres, P., Taha, D., and Julier, C. (2006) Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet* **38**, 682-687
18. Kim, S. H., Kim, E. J., Hitomi, M., Oh, S. Y., Jin, X., Jeon, H. M., Beck, S., Jin, X., Kim, J. K., Park, C. G., Chang, S. Y., Yin, J., Kim, T., Jeon, Y. J., Song, J., Lim, Y. C., Lathia, J. D., Nakano, I., and Kim, H. (2015) The LIM-only transcription factor LMO2 determines tumorigenic and angiogenic traits in glioma stem cells. *Cell Death Differ* **22**, 1517-1525
19. Squassina, A., Meloni, A., Chillotti, C., and Pisanu, C. (2019) Zinc finger proteins in psychiatric disorders and response to psychotropic medications. *Psychiatr Genet* **29**, 132-141
20. Abbehausen, C. (2019) Zinc finger domains as therapeutic targets for metal-based compounds - an update. *Metallomics* **11**, 15-28
21. Li, H., Yang, Y., Hong, W., Huang, M., Wu, M., and Zhao, X. (2020) Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct Target Ther* **5**, 1
22. Koch, B. J., Ryan, J. F., and Baxevanis, A. D. (2012) The diversification of the LIM superclass at the base of the metazoa increased subcellular complexity and promoted multicellular specialization. *PLoS One* **7**, e33261
23. Freyd, G., Kim, S. K., and Horvitz, H. R. (1990) Novel cysteine-rich motif and homeodomain in the product of the Caenorhabditis elegans cell lineage gene lin-11. *Nature* **344**, 876-879
24. Karlsson, O., Thor, S., Norberg, T., Ohlsson, H., and Edlund, T. (1990) Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature* **344**, 879-882
25. Way, J. C., and Chalfie, M. (1988) mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in C. elegans. *Cell* **54**, 5-16
26. Michelsen, J. W., Schmeichel, K. L., Beckerle, M. C., and Winge, D. R. (1993) The LIM motif defines a specific zinc-binding protein domain. *Proc Natl Acad Sci U S A* **90**, 4404-4408
27. Li, P. M., Reichert, J., Freyd, G., Horvitz, H. R., and Walsh, C. T. (1991) The LIM region of a presumptive Caenorhabditis elegans transcription factor is an iron-sulfur- and zinc-containing metallodomain. *Proc Natl Acad Sci U S A* **88**, 9210-9213
28. Kadrmas, J. L., and Beckerle, M. C. (2004) The LIM domain: from the cytoskeleton to the nucleus. *Nat Rev Mol Cell Biol* **5**, 920-931
29. Kosa, J. L., Michelsen, J. W., Louis, H. A., Olsen, J. I., Davis, D. R., Beckerle, M. C., and Winge, D. R. (1994) Common metal ion coordination in LIM domain proteins. *Biochemistry* **33**, 468-477
30. Dong, W. F., Heng, H. H., Lowsky, R., Xu, Y., DeCoteau, J. F., Shi, X. M., Tsui, L. C., and Minden, M. D. (1997) Cloning, expression, and chromosomal localization to 11p12-13 of a human LIM/HOMEOBOX gene, hLim-1. *DNA Cell Biol* **16**, 671-678
31. McGuire, E. A., Davis, A. R., and Korsmeyer, S. J. (1991) T-cell translocation gene 1 (Ttg-1) encodes a nuclear protein normally expressed in neural lineage cells. *Blood* **77**, 599-606
32. Knierim, E., Hirata, H., Wolf, N. I., Morales-Gonzalez, S., Schottmann, G., Tanaka, Y., Rudnik-Schoneborn, S., Orgeur, M., Zerres, K., Vogt, S., van Riesen, A., Gill, E., Seifert, F., Zwirner, A., Kirschner, J., Goebel, H. H., Hubner, C., Stricker, S., Meierhofer, D., Stenzel, W., and Schuelke, M. (2016) Mutations in Subunits of the Activating Signal Cointegrator 1 Complex Are Associated with Prenatal Spinal Muscular Atrophy and Congenital Bone Fractures. *Am J Hum Genet* **98**, 473-489
33. Gueneau, L., Bertrand, A. T., Jais, J. P., Salih, M. A., Stojkovic, T., Wehnert, M., Hoeltzenbein, M., Spuler, S., Saitoh, S., Verschueren, A., Tranchant, C., Beuvin, M., Lacene, E., Romero, N. B., Heath, S., Zelenika, D., Voit, T., Eymard, B., Ben Yaou, R., and Bonne, G. (2009) Mutations of the FHL1 gene cause Emery-Dreifuss muscular dystrophy. *Am J Hum Genet* **85**, 338-353

34. Chiswell, B. P., Zhang, R., Murphy, J. W., Boggon, T. J., and Calderwood, D. A. (2008) The structural basis of integrin-linked kinase-PINCH interactions. *Proc Natl Acad Sci U S A* **105**, 20677-20682
35. Wang, N., Lin, K., Lu, Z., Lam, K., Newton, R., Xu, X., Yu, Z., Gill, G., and Andersen, B. (2007) The LIM-only factor LMO4 regulates expression of the BMP7 gene through an HDAC2-dependent mechanism, and controls cell proliferation and apoptosis of mammary epithelial cells. *Oncogene* **26**, 6431-6441
36. Schmeichel, K. L., and Beckerle, M. C. (1994) The LIM domain is a modular protein-binding interface. *Cell* **79**, 211-219
37. Schmeichel, K. L., and Beckerle, M. C. (1997) Molecular dissection of a LIM domain. *Mol Biol Cell* **8**, 219-230
38. Maekawa, M., Ishizaki, T., Boku, S., Watanabe, N., Fujita, A., Iwamatsu, A., Obinata, T., Ohashi, K., Mizuno, K., and Narumiya, S. (1999) Signaling from Rho to the actin cytoskeleton through protein kinases ROCK and LIM-kinase. *Science* **285**, 895-898
39. Schmidt, E. F., Shim, S. O., and Strittmatter, S. M. (2008) Release of MICAL autoinhibition by semaphorin-plexin signaling promotes interaction with collapsin response mediator protein. *J Neurosci* **28**, 2287-2297
40. Kotb, A., Hyndman, M. E., and Patel, T. R. (2018) The role of zyxin in regulation of malignancies. *Heliyon* **4**, e00695
41. van der Gaag, E. J., Leccia, M. T., Dekker, S. K., Jalbert, N. L., Amodeo, D. M., and Byers, H. R. (2002) Role of zyxin in differential cell spreading and proliferation of melanoma cells and melanocytes. *J Invest Dermatol* **118**, 246-254
42. Zhou, J., Zeng, Y., Cui, L., Chen, X., Stauffer, S., Wang, Z., Yu, F., Lele, S. M., Talmon, G. A., Black, A. R., Chen, Y., and Dong, J. (2018) Zyxin promotes colon cancer tumorigenesis in a mitotic phosphorylation-dependent manner and through CDK8-mediated YAP activation. *Proc Natl Acad Sci U S A* **115**, E6760-E6769
43. Wen, X. M., Luo, T., Jiang, Y., Wang, L. H., Luo, Y., Chen, Q., Yang, K., Yuan, Y., Luo, C., Zhang, X., Yan, Z. X., Fu, W. J., Tan, Y. H., Niu, Q., Xiao, J. F., Chen, L., Wang, J., Huang, J. F., Cui, Y. H., Zhang, X., Wang, Y., and Bian, X. W. (2020) Zyxin (ZYX) promotes invasion and acts as a biomarker for aggressive phenotypes of human glioblastoma multiforme. *Lab Invest*
44. Califano, A., and Alvarez, M. J. (2017) The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat Rev Cancer* **17**, 116-130
45. Halaoui, R., and McCaffrey, L. (2015) Rewiring cell polarity signaling in cancer. *Oncogene* **34**, 939-950
46. Petit, M. M., Fradelizi, J., Golsteyn, R. M., Ayoubi, T. A., Menichi, B., Louvard, D., Van de Ven, W. J., and Friederich, E. (2000) LPP, an actin cytoskeleton protein related to zyxin, harbors a nuclear export signal and transcriptional activation capacity. *Mol Biol Cell* **11**, 117-129
47. James, V., Zhang, Y., Foxler, D. E., de Moor, C. H., Kong, Y. W., Webb, T. M., Self, T. J., Feng, Y., Lagos, D., Chu, C. Y., Rana, T. M., Morley, S. J., Longmore, G. D., Bushell, M., and Sharp, T. V. (2010) LIM-domain proteins, LIMD1, Ajuba, and WTIP are required for microRNA-mediated gene silencing. *Proc Natl Acad Sci U S A* **107**, 12499-12504
48. Hirota, T., Kunitoku, N., Sasayama, T., Marumoto, T., Zhang, D., Nitta, M., Hatakeyama, K., and Saya, H. (2003) Aurora-A and an interacting activator, the LIM protein Ajuba, are required for mitotic commitment in human cells. *Cell* **114**, 585-598
49. Marie, H., Pratt, S. J., Betson, M., Epple, H., Kittler, J. T., Meek, L., Moss, S. J., Troyanovsky, S., Attwell, D., Longmore, G. D., and Braga, V. M. (2003) The LIM protein Ajuba is recruited to cadherin-dependent cell junctions through an association with alpha-catenin. *J Biol Chem* **278**, 1220-1228
50. Tu, Y., Wu, S., Shi, X., Chen, K., and Wu, C. (2003) Migfilin and Mig-2 link focal adhesions to filamin and the actin cytoskeleton and function in cell shape modulation. *Cell* **113**, 37-47

51. Ayyanathan, K., Peng, H., Hou, Z., Fredericks, W. J., Goyal, R. K., Langer, E. M., Longmore, G. D., and Rauscher, F. J., 3rd. (2007) The Ajuba LIM domain protein is a corepressor for SNAG domain mediated repression and participates in nucleocytoplasmic Shuttling. *Cancer Res* **67**, 9097-9106
52. Abe, Y., Ohsugi, M., Haraguchi, K., Fujimoto, J., and Yamamoto, T. (2006) LATS2-Ajuba complex regulates gamma-tubulin recruitment to centrosomes and spindle organization during mitosis. *FEBS Lett* **580**, 782-788
53. Minor, D. L., Jr., and Kim, P. S. (1994) Measurement of the beta-sheet-forming propensities of amino acids. *Nature* **367**, 660-663
54. Pace, C. N., and Scholtz, J. M. (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* **75**, 422-427
55. Acquaah-Mensah, G. K., Leach, S. M., and Guda, C. (2006) Predicting the subcellular localization of human proteins using machine learning and exploratory data analysis. *Genomics Proteomics Bioinformatics* **4**, 120-133
56. la Cour, T., Kiemer, L., Molgaard, A., Gupta, R., Skriver, K., and Brunak, S. (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng Des Sel* **17**, 527-536
57. Kiraga, J., Mackiewicz, P., Mackiewicz, D., Kowalczuk, M., Biecek, P., Polak, N., Smolarczyk, K., Dudek, M. R., and Cebrat, S. (2007) The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. *BMC Genomics* **8**, 163
58. Schwartz, R., Ting, C. S., and King, J. (2001) Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome Res* **11**, 703-709
59. Zhu, K., Zhao, J., Lubman, D. M., Miller, F. R., and Barder, T. J. (2005) Protein pI shifts due to posttranslational modifications in the separation and characterization of proteins. *Anal Chem* **77**, 2745-2755
60. Cohen, P. (2002) The origins of protein phosphorylation. *Nat Cell Biol* **4**, E127-130
61. Singh, V., Ram, M., Kumar, R., Prasad, R., Roy, B. K., and Singh, K. K. (2017) Phosphorylation: Implications in Cancer. *Protein J* **36**, 1-6
62. Hanahan, D., and Weinberg, R. A. (2000) The hallmarks of cancer. *Cell* **100**, 57-70
63. Appella, E., and Anderson, C. W. (2001) Post-translational modifications and activation of p53 by genotoxic stresses. *Eur J Biochem* **268**, 2764-2772
64. Radivojac, P., Baenziger, P. H., Kann, M. G., Mort, M. E., Hahn, M. W., and Mooney, S. D. (2008) Gain and loss of phosphorylation sites in human cancer. *Bioinformatics* **24**, i241-247
65. Slamon, D., and Pegram, M. (2001) Rationale for trastuzumab (Herceptin) in adjuvant breast cancer trials. *Semin Oncol* **28**, 13-19
66. Spector, N., Xia, W., El-Hariry, I., Yarden, Y., and Bacus, S. (2007) HER2 therapy. Small molecule HER-2 tyrosine kinase inhibitors. *Breast Cancer Research* **9**, 205
67. Engel, K., Kotlyarov, A., and Gaestel, M. (1998) Leptomycin B-sensitive nuclear export of MAPKAP kinase 2 is regulated by phosphorylation. *EMBO J* **17**, 3363-3371
68. Ohno, M., Segref, A., Bachi, A., Wilm, M., and Mattaj, I. W. (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**, 187-198
69. Hornbeck, P. V., Kornhauser, J. M., Latham, V., Murray, B., Nandhikonda, V., Nord, A., Skrzypek, E., Wheeler, T., Zhang, B., and Gnad, F. (2019) 15 years of PhosphoSitePlus(R): integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res* **47**, D433-D441
70. Alende, N., Nielsen, J. E., Shields, D. C., and Khaldi, N. (2011) Evolution of the isoelectric point of mammalian proteins as a consequence of indels and adaptive evolution. *Proteins* **79**, 1635-1648
71. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* **26**, 283-291

72. Benkert, P., Biasini, M., and Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343-350
73. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549
74. Kumar, M., Gromiha, M. M., and Raghava, G. P. (2007) Identification of DNA-binding proteins using support vector machines and evolutionary profiles. *BMC Bioinformatics* **8**, 463
75. Holm, L., and Laakso, L. M. (2016) Dali server update. *Nucleic Acids Res* **44**, W351-355
76. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**, 536-540
77. Sam, V., Tai, C. H., Garnier, J., Gibrat, J. F., Lee, B., and Munson, P. J. (2006) ROC and confusion analysis of structure comparison methods identify the main causes of divergence from manual protein classification. *BMC Bioinformatics* **7**, 206
78. Deane, J. E., Ryan, D. P., Sunde, M., Maher, M. J., Guss, J. M., Visvader, J. E., and Matthews, J. M. (2004) Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex. *EMBO J* **23**, 3589-3598
79. Jeffries, C. M., Graham, S. C., Stokes, P. H., Collyer, C. A., Guss, J. M., and Matthews, J. M. (2006) Stabilization of a binary protein complex by intein-mediated cyclization. *Protein Sci* **15**, 2612-2618
80. Bhati, M., Lee, C., Nancarrow, A. L., Lee, M., Craig, V. J., Bach, I., Guss, J. M., Mackay, J. P., and Matthews, J. M. (2008) Implementing the LIM code: the structural basis for cell type-specific assembly of LIM-homeodomain complexes. *EMBO J* **27**, 2018-2029
81. Stokes, P. H., Robertson, N. O., Silva, A. P. G., Estephan, T., Trewhella, J., Guss, J. M., and Matthews, J. M. (2019) Mutation in a flexible linker modulates binding affinity for modular complexes. *Proteins* **87**, 425-429
82. Gadd, M. S., Jacques, D. A., Nisevic, I., Craig, V. J., Kwan, A. H., Guss, J. M., and Matthews, J. M. (2013) A structural basis for the regulation of the LIM-homeodomain protein islet 1 (Isl1) by intra- and intermolecular interactions. *J Biol Chem* **288**, 21924-21935
83. Racevskis, J., Dill, A., Sparano, J. A., and Ruan, H. (1999) Molecular cloning of LMO41, a new human LIM domain gene. *Biochim Biophys Acta* **1445**, 148-153
84. Vu, D., Marin, P., Walzer, C., Cathieni, M. M., Bianchi, E. N., Saidji, F., Leuba, G., Bouras, C., and Savioz, A. (2003) Transcription regulator LMO4 interferes with neuritogenesis in human SH-SY5Y neuroblastoma cells. *Brain Res Mol Brain Res* **115**, 93-103
85. Sum, E. Y., Peng, B., Yu, X., Chen, J., Byrne, J., Lindeman, G. J., and Visvader, J. E. (2002) The LIM domain protein LMO4 interacts with the cofactor CtIP and the tumor suppressor BRCA1 and inhibits BRCA1 activity. *J Biol Chem* **277**, 7849-7856
86. Lopez-Blanco, J. R., Aliaga, J. I., Quintana-Orti, E. S., and Chacon, P. (2014) iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res* **42**, W271-276
87. Bray, J. K., Weiss, D. R., and Levitt, M. (2011) Optimized torsion-angle normal modes reproduce conformational changes more accurately than cartesian modes. *Biophysical journal* **101**, 2966-2969
88. Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) Essential dynamics of proteins. *Proteins* **17**, 412-425
89. Feuerstein, R., Wang, X., Song, D., Cooke, N. E., and Liebhaber, S. A. (1994) The LIM/double zinc-finger motif functions as a protein dimerization domain. *Proc Natl Acad Sci U S A* **91**, 10655-10659
90. Beckerle, M. C. (1986) Identification of a new protein localized at sites of cell-substrate adhesion. *J Cell Biol* **103**, 1679-1687
91. Miles, J. A., Frost, M. G., Carroll, E., Rowe, M. L., Howard, M. J., Sidhu, A., Chaugule, V. K., Alpi, A. F., and Walden, H. (2015) The Fanconi Anemia DNA Repair Pathway Is Regulated by an Interaction between Ubiquitin and the E2-like Fold Domain of FANCL. *J Biol Chem* **290**, 20995-21006

92. Chan, C. B., Liu, X., Tang, X., Fu, H., and Ye, K. (2007) Akt phosphorylation of zyxin mediates its interaction with acinus-S and prevents acinus-triggered chromatin condensation. *Cell Death Differ* **14**, 1688-1699
93. Schweppe, D. K., Huttlin, E. L., Harper, J. W., and Gygi, S. P. (2018) BioPlex Display: An Interactive Suite for Large-Scale AP-MS Protein-Protein Interaction Data. *J Proteome Res* **17**, 722-726
94. Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., and Morris, Q. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214-220
95. Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A. P., Santonico, E., Castagnoli, L., and Cesareni, G. (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* **40**, D857-861
96. Szklarczyk, D., Santos, A., von Mering, C., Jensen, L. J., Bork, P., and Kuhn, M. (2016) STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res* **44**, D380-384
97. Licata, L., Lo Surdo, P., Iannuccelli, M., Palma, A., Micarelli, E., Perfetto, L., Peluso, D., Calderone, A., Castagnoli, L., and Cesareni, G. (2020) SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. *Nucleic Acids Res* **48**, D504-D510
98. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Mering, C. V. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613
99. Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B., Ng, P. K., Jeong, K. J., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E. M., Mularoni, L., Rubio-Perez, C., Nagarajan, N., Cortes-Ciriano, I., Zhou, D. C., Liang, W. W., Hess, J. M., Yellapantula, V. D., Tamborero, D., Gonzalez-Perez, A., Suphavilai, C., Ko, J. Y., Khurana, E., Park, P. J., Van Allen, E. M., Liang, H., Group, M. C. W., Cancer Genome Atlas Research, N., Lawrence, M. S., Godzik, A., Lopez-Bigas, N., Stuart, J., Wheeler, D., Getz, G., Chen, K., Lazar, A. J., Mills, G. B., Karchin, R., and Ding, L. (2018) Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385 e318
100. Iranzo, J., Martincorena, I., and Koonin, E. V. (2018) Cancer-mutation network and the number and specificity of driver mutations. *Proc Natl Acad Sci U S A* **115**, E6010-E6019
101. Chen, H. K., Pai, C. Y., Huang, J. Y., and Yeh, N. H. (1999) Human Nopp140, which interacts with RNA polymerase I: implications for rRNA gene transcription and nucleolar structural organization. *Mol Cell Biol* **19**, 8536-8546
102. Turei, D., Korcsmaros, T., and Saez-Rodriguez, J. (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods* **13**, 966-967
103. Cante-Barrett, K., Spijkers-Hagelstein, J. A., Buijs-Gladdines, J. G., Uitdehaag, J. C., Smits, W. K., van der Zwet, J., Buijsman, R. C., Zaman, G. J., Pieters, R., and Meijerink, J. P. (2016) MEK and PI3K-AKT inhibitors synergistically block activated IL7 receptor signaling in T-cell acute lymphoblastic leukemia. *Leukemia* **30**, 1832-1843
104. Johnson, S. E., Shah, N., Bajer, A. A., and LeBien, T. W. (2008) IL-7 activates the phosphatidylinositol 3-kinase/AKT pathway in normal human thymocytes but not normal human B cell precursors. *J Immunol* **180**, 8109-8117
105. Abassi, Y. A., Rehn, M., Ekman, N., Alitalo, K., and Vuori, K. (2003) p130Cas Couples the tyrosine kinase Bmx/Etk with regulation of the actin cytoskeleton and cell migration. *J Biol Chem* **278**, 35636-35643
106. Sakakibara, A., Ohba, Y., Kurokawa, K., Matsuda, M., and Hattori, S. (2002) Novel function of Chat in controlling cell adhesion via Cas-Crk-C3G-pathway-mediated Rap1 activation. *J Cell Sci* **115**, 4915-4924

107. Hervy, M., Hoffman, L. M., Jensen, C. C., Smith, M., and Beckerle, M. C. (2010) The LIM Protein Zyxin Binds CARP-1 and Promotes Apoptosis. *Genes Cancer* **1**, 506-515
108. Nix, D. A., Fradelizi, J., Bockholt, S., Menichi, B., Louvard, D., Friederich, E., and Beckerle, M. C. (2001) Targeting of zyxin to sites of actin membrane interaction and to the nucleus. *J Biol Chem* **276**, 34759-34767
109. Brunet, A., Kanai, F., Stehn, J., Xu, J., Sarbassova, D., Frangioni, J. V., Dalal, S. N., DeCaprio, J. A., Greenberg, M. E., and Yaffe, M. B. (2002) 14-3-3 transits to the nucleus and participates in dynamic nucleocytoplasmic transport. *J Cell Biol* **156**, 817-828
110. Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of cancer: the next generation. *Cell* **144**, 646-674
111. Sanchez-Garcia, I., and Rabbitts, T. H. (1994) The LIM domain: a new structural motif found in zinc-finger-like proteins. *Trends Genet* **10**, 315-320
112. Wu, R., Durick, K., Songyang, Z., Cantley, L. C., Taylor, S. S., and Gill, G. N. (1996) Specificity of LIM domain interactions with receptor tyrosine kinases. *J Biol Chem* **271**, 15934-15941
113. Osada, H., Grutz, G., Axelson, H., Forster, A., and Rabbitts, T. H. (1995) Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc Natl Acad Sci U S A* **92**, 9585-9589
114. Valge-Archer, V. E., Osada, H., Warren, A. J., Forster, A., Li, J., Baer, R., and Rabbitts, T. H. (1994) The LIM protein RBTN2 and the basic helix-loop-helix protein TAL1 are present in a complex in erythroid cells. *Proc Natl Acad Sci U S A* **91**, 8617-8621
115. Wadman, I., Li, J., Bash, R. O., Forster, A., Osada, H., Rabbitts, T. H., and Baer, R. (1994) Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia. *EMBO J* **13**, 4831-4839
116. Hamill, S., Lou, H. J., Turk, B. E., and Boggon, T. J. (2016) Structural Basis for Noncanonical Substrate Recognition of Cofilin/ADF Proteins by LIM Kinases. *Mol Cell* **62**, 397-408
117. Wilkinson-White, L., Gamsjaeger, R., Dastmalchi, S., Wienert, B., Stokes, P. H., Crossley, M., Mackay, J. P., and Matthews, J. M. (2011) Structural basis of simultaneous recruitment of the transcriptional regulators LMO2 and FOG1/ZFPM1 by the transcription factor GATA1. *Proc Natl Acad Sci U S A* **108**, 14443-14448
118. Boeda, B., Briggs, D. C., Higgins, T., Garvalov, B. K., Fadden, A. J., McDonald, N. Q., and Way, M. (2007) Tes, a specific Mena interacting partner, breaks the rules for EVH1 binding. *Mol Cell* **28**, 1071-1082
119. Gadd, M. S., Bhati, M., Jeffries, C. M., Langley, D. B., Trewhella, J., Guss, J. M., and Matthews, J. M. (2011) Structural basis for partial redundancy in a class of transcription factors, the LIM homeodomain proteins, in neural cell type specification. *J Biol Chem* **286**, 42971-42980
120. Kamberaj, H., and van der Vaart, A. (2009) Extracting the causality of correlated motions from molecular dynamics simulations. *Biophys J* **97**, 1747-1755
121. Goodey, N. M., and Benkovic, S. J. (2008) Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* **4**, 474-482
122. Hammes, G. G. (2002) Multiple conformational changes in enzyme catalysis. *Biochemistry* **41**, 8221-8228
123. Jarymowycz, V. A., and Stone, M. J. (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev* **106**, 1624-1671
124. Lanni, C., Necchi, D., Pinto, A., Buoso, E., Buizza, L., Memo, M., Uberti, D., Govoni, S., and Racchi, M. (2013) Zyxin is a novel target for beta-amyloid peptide: characterization of its role in Alzheimer's pathogenesis. *J Neurochem* **125**, 790-799
125. Oh, H., and Irvine, K. D. (2010) Yorkie: the final destination of Hippo signaling. *Trends Cell Biol* **20**, 410-417

126. Wang, Y., and Gilmore, T. D. (2003) Zyxin and paxillin proteins: focal adhesion plaque LIM domain proteins go nuclear. *Biochim Biophys Acta* **1593**, 115-120
127. Wu, Z., Qiu, M., Mi, Z., Meng, M., Guo, Y., Jiang, X., Fang, J., Wang, H., Zhao, J., Liu, Z., Qian, D., and Yuan, Z. (2019) WT1-interacting protein inhibits cell proliferation and tumorigenicity in non-small-cell lung cancer via the AKT/FOXO1 axis. *Mol Oncol* **13**, 1059-1074
128. Ou, Y., Ma, L., Dong, L., Ma, L., Zhao, Z., Ma, L., Zhou, W., Fan, J., Wu, C., Yu, C., Zhan, Q., and Song, Y. (2012) Migfilin protein promotes migration and invasion in human glioma through epidermal growth factor receptor-mediated phospholipase C-gamma and STAT3 protein signaling pathways. *J Biol Chem* **287**, 32394-32405
129. Kiepas, A., Voorand, E., Senecal, J., Ahn, R., Annis, M. G., Jacquet, K., Tali, G., Bisson, N., Ursini-Siegel, J., Siegel, P. M., and Brown, C. M. (2020) The SHCA adapter protein cooperates with lipoma-preferred partner in the regulation of adhesion dynamics and invadopodia formation. *J Biol Chem*
130. Kuriyama, S., Yoshida, M., Yano, S., Aiba, N., Kohno, T., Minamiya, Y., Goto, A., and Tanaka, M. (2016) LPP inhibits collective cell migration during lung cancer dissemination. *Oncogene* **35**, 952-964
131. Dunwell, T. L., Hesson, L. B., Pavlova, T., Zabarovska, V., Kashuba, V., Catchpoole, D., Chiaramonte, R., Brini, A. T., Griffiths, M., Maher, E. R., Zabarovsky, E., and Latif, F. (2009) Epigenetic analysis of childhood acute lymphoblastic leukemia. *Epigenetics* **4**, 185-193
132. Nam, R. K., Benatar, T., Amemiya, Y., Wallis, C. J. D., Romero, J. M., Tsagaris, M., Sherman, C., Sugar, L., and Seth, A. (2018) MicroRNA-652 induces NED in LNCaP and EMT in PC3 prostate cancer cells. *Oncotarget* **9**, 19159-19176
133. Hurst, D. R., Mehta, A., Moore, B. P., Phadke, P. A., Meehan, W. J., Accavitti, M. A., Shevde, L. A., Hopper, J. E., Xie, Y., Welch, D. R., and Samant, R. S. (2006) Breast cancer metastasis suppressor 1 (BRMS1) is stabilized by the Hsp90 chaperone. *Biochem Biophys Res Commun* **348**, 1429-1435
134. Liu, Y., Wu, X., Wang, G., Hu, S., Zhang, Y., and Zhao, S. (2019) CALD1, CNN1, and TAGLN identified as potential prognostic molecular markers of bladder cancer by bioinformatics analysis. *Medicine (Baltimore)* **98**, e13847
135. Magrane, M., and UniProt, C. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* **2011**, bar009
136. Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113
137. Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* **Chapter 2**, Unit 2 3
138. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191
139. Eddy, S. R. (2004) Where did the BLOSUM62 alignment score matrix come from? *Nat Biotechnol* **22**, 1035-1036
140. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797
141. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190
142. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* **40**, W597-603
143. Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic acids research* **32**, W526-W531

144. Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**, W526-531
145. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics* **22**, 195-201
146. Bordoli, L., and Schwede, T. (2012) Automated protein structure modeling with SWISS-MODEL Workspace and the Protein Model Portal. *Methods Mol Biol* **857**, 107-136
147. Lopez-Blanco, J. R., Reyes, R., Aliaga, J. I., Badia, R. M., Chacon, P., and Quintana-Orti, E. S. (2013) Exploring large macromolecular functional motions on clusters of multicore processors. *Journal of Computational Physics* **246**, 275-288
148. Nguyen, D. T., Mathias, S., Bologa, C., Brunak, S., Fernandez, N., Gaulton, A., Hersey, A., Holmes, J., Jensen, L. J., Karlsson, A., Liu, G., Ma'ayan, A., Mandava, G., Mani, S., Mehta, S., Overington, J., Patel, J., Rouillard, A. D., Schurer, S., Sheils, T., Simeonov, A., Sklar, L. A., Southall, N., Ursu, O., Vidovic, D., Waller, A., Yang, J., Jadhav, A., Oprea, T. I., and Guha, R. (2017) Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Res* **45**, D995-D1002
149. Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* **2008**, 420747

List of Tables

Table 1: List of Zyxin family proteins containing Lim domain, and functions and biochemical properties.

Protein (UNIPROT ID)	Cellular role *	Cellular localization	Isoelectric point[1]		Amino acids / Molecular weight (kDa)[1]	
			Full protein	LIM domain	Full protein	LIM domain
Zyxin (Q15942)	Adhesion plague protein, signal transducer	Cytoskeleton, cytosol and nucleus	6.22	7.16	572/ 61.27	189/21.10
LimD1 (Q9UGP4)	Cell fate determination and a tumour suppressor	Predominantly in nucleus, P- body	6.20	5.90	676/ 72.19	195/21.87
Ajuba (Q96IF1)	Cell fate determination and repression of gene transcription	Cell membrane, cytoskeleton, centrosome, nucleus, P- body	6.86	5.81	538/ 56.93	195/21.87
LPP (Q93052)	Maintains cell shape, motility, and activation of gene transcription	Cell membrane and nucleus	7.18	6.78	612/ 65.74	190/21.15
WTIP (A6NIX2)	Cell fate determination, repression of gene transcription, and miRNA-mediated gene silencing	Nucleus, P- body	8.53	7.50	430/ 45.12	195/21.80
FBLIM1 (Q8WUP2)	Cell-ECM adhesion proteins and filamin-containing actin filaments	Focal adhesion sites	5.71	7.79	373/ 40.66	190/21.70
TRIP6 (Q15654)	Relays signals from the cell surface to the nucleus	Focal adhesion sites, cytoplasm, nucleus	7.19	6.06	476/ 50.28	188/20.49

*= *Information retrieved from UniprotKB*
[1]= *Calculated using ProtParam*

Table 2: Effect of post-translational modifications on isoelectric point of Zyxin family proteins.

Protein (Uniprot ID)	pI		Number of sites for PTM					
	Native[1]	Post PTM*[2]	Phosphorylation	Acetylations*	Ubiquitylation*	Monomethylation*	Dimethylation*	Caspase cleavage sites*
Zyxin (Q15942)	6.22	6.12-3.17	66	5	4	3	2	1
LimD1 (Q9UGP4)	6.20	6.12-3.85	65	-	8	4	-	-
Ajuba (Q96IF1)	6.86	6.63-4.63	30	-	2	1	-	-
LPP (Q93052)	7.18	6.89-2.84	83	3	5	1	1	-
WTIP (A6NIX2)	8.53	8.36-6.22	11	-	-	-	-	-
FBLIM1 (Q8WUP2)	5.71	5.60-4.96	9	1	-	-	-	-
TRIP6 (Q15654)	7.19	6.89-3.61	56	-	-	22	4	-

Data obtained from Phosphosite plus
[1]= Calculated from ProtParam
[2]= Range is from single phosphorylation to maximum number of phosphorylation

Table 3:
Percentage composition of Proline Rich Region (PRR) and Lim domain in the full-length proteins

Protein (Uniprot ID)	Percentage of PRR vs Full length*	Percentage of Lim domain vs. Full length*
Zyxin (Q15942)	66.9%	33.1%
LimD1 (Q9UGP4)	69.3%	30.7%
Ajuba (Q96IF1)	62.8%	37.2%
LPP (Q93052)	67.4%	32.6%
WTIP (A6NIX2)	51.6%	48.4%
FBLIM1 (Q8WUP2)	48.2%	51.8%
TRIP6 (Q15654)	58.4%	41.6%

**= Calculated by considering full length (all amino acids) as a 100% of the respective protein*
PRR= Proline Rich Repeat

Table 4:

SVM based prediction of DNA binding of Zyxin family proteins as calculated by DNAbinder webserver

Protein (Uniprot ID)	SVM score*	Prediction	Job Number on webserver
Zyxin (Q15942)	1.2919473	DNA binding protein	1710
LimD1 (Q9UGP4)	0.90588128	DNA binding protein	4069
Ajuba (Q96IF1)	1.4081202	DNA binding protein	7780
LPP (Q93052)	0.11482763	DNA binding protein	1668
WTIP (A6NIX2)	1.7520286	DNA binding protein	1032
FBLIM1 (Q8WUP2)	0.39872591	DNA binding protein	1215
TRIP6 (Q15654)	0.83398047	DNA binding protein	3994

**= Calculated for full length (all amino acids) under amino acid composition mode*

Table 5: DALI structural alignment of the Lim domain of Zyxin family proteins

Lim Domains	Description of structural similarities identified by DALI server	
	Unique	**Common**
Zyxin	Lim domain-bin, Lim domain-BI, CRP1, Four and Half Lim domain protein 2, Rhombotin-2	Fusion protein of LMO4 protein, Lim/Homeobox protein LHX4, Fusion protein of Lim domain transcription factor Fusion of Lim/Homeobox protein LHX3 Insulin gene enhancer protein ISL-1 Insulin gene enhancer P Fusion protein of Lim domain transcription factor Actin Like protein 7A
LimD1	Lim domain-BI, Lim/Homeobox protein LHX4, Rhombotin-2, CRP1	
Ajuba	Thyroid receptor interacting protein 6, T-cell acute lymphocytic leukemia protein 1	
LPP	Lim domain-BI, Rhombotin-2, Thyroid receptor interacting protein 6	
WTIP	Lim domain-bin, Lim domain-BI, Four and Half Lim domain protein 2, Rhombotin-2	
FBLIM1	Lim domain-BI, Four and Half Lim domain protein 2, Rhombotin-2, Thyroid receptor interacting protein 6	
TRIP6	Rhombotin-2, T-cell acute lymphocytic leukemia protein 1, Four and Half Lim domains 2, Thyroid receptor interacting protein 6, Rhombotin-2, Paxillin, PINCH, Lim domain-BI	

Figure Legends

Figure 1: The Lim family proteins are depicted with the Lim domains; X presents the number of proteins each family has. Abbreviations are Leucine-aspartate repeat (LD), PINCH motif (PM), TES motif A1 (TMA1), TES motif A2 (TMA2), ZYX motif (ZyM), EPLIN motif (EM), calponin homology domain (CH), glycine-rich region (Gly), SRC homology 3 domain (SH3), Zasp motif (ZM), Alp motif (AM), Villin headpiece domain (VHP).

Figure 2: Organization of domains and motifs of Zyxin family proteins. (A) The organization of the proline-rich region (PRR) and Lim1/2/3 domains. (B) A schematic representation of zinc-finger formed in Lim domains is shown, where a Zn^{2+} ion is chelated by four Cysteines (orange hexagon) the spacer amino acids between Cysteines are shown as yellow circles.

Figure 3: Hydrophobicity plots for Lim domains of Zyxin family proteins. The hydrophobic analyses were performed using the ProtScale server and plotted against the amino acid position. The scoring is varied from -2.0 to 2.5 on the y-axis reflect hydrophobicity (higher positive (+ve) score depicts more hydrophobicity). This analysis indicates that each Lim domain and Lim1/Lim2/Lim3 region has a distinct hydrophobicity score.

Figure 4: Sequence alignment of the Leucine-rich nuclear export signal (NES) sequence of Zyxin family proteins. (A) the NES sequence for each protein was identified through NetNES webserver, and the sequences were aligned using MUSCLE, (B) WEBLOGO was used to investigate the conservation of residues in the sequences, which suggested the presence of signature LxxLxL/LxxxLxL, (L=L/V/I/F/M; x=any amino acid) repeat, in NES sequence across all the members of Zyxin family except zyxin and LPP which shows LxxxLxxL/LxxLxxL consensus sequence.

Figure 5: Structures of all the Lim domains of the Zyxin family proteins were generated using comparative/homology modeling in the Robetta server. Robetta server generates structure using structures identified by BLAST and PSI-BLAST, the regions of the protein that could not be modeled using BLAST are modeled *de-novo* using Rosetta fragment insertion method. For each model, N-terminal is depicted in blue and the C-terminal is depicted by red. All models were validated using PROCHECK and SWISS-MODEL workspace; the structures with the highest Z-score are depicted here.

Figure 6: Normal mode analysis of Zyxin family Lim domain proteins. The NMA for each protein model was performed using the iMODS server under default settings. The motion of each domain is depicted by the arrows arising in the direction of motion. Additionally, the models are coloured on the basis of NMA mobility of each structure, where the mobility is represented using a colour spectrum (blue represents the lowest mobility whereas red represents the highest mobility, and intermediate mobility is depicted by interim colours, blue>green>yellow>red). Interestingly, the direction of motion of all proteins is such that the distal ends of the protein move in the direction towards each other producing a motion that resembles clamping.

Figure 7: Covariance analysis of the Zyxin family Lim domain. Covariance analysis was performed to understand the correlated movement of protein residues in the secondary structure which can shed light on the function of the protein. Each panel represents the covariance matrix for each member of the Zyxin family, in the matrix correlated motion is depicted as red, anti-correlated as blue, and uncorrelated as a white dot. High levels of correlated motions were observed in case of Lim1 domains of all protein proteins, varied amount of motion was observed in case of Lim2 with highest in TRIP6, again Lim3 showed similar kind of motion in all proteins. Proteins with similar localization and function showed similarity in motion. For example, Ajuba, LIMD1, and WTIP showed relatively similar covariance matrix of Lim3 domain these proteins are localized at the same site and constitute

P-Body components, primarily perform gene regulation and mediate miRNA silencing. LPP and FBLIM1 showed similar correlated, uncorrelated and anticorrelated motions in Lim1/2/3, and both proteins involved mainly in cell shape maintenance; supporting the fact that similar dynamics may have similar localization and functions, each Lim region has its own characteristic features and believed to perform functions independently.

Figure 8: The linking matrix of the Elastic network model shows the stiffness of the springs connecting the corresponding pair of atoms. The dots are coloured as a function of their stiffness, darker colours reflect stiffer springs**.** Interestingly all proteins showed similar kind of stiffness except TRIP6.

Figure 9: Expression analysis of the Zyxin family proteins in cancer. Protein expression profiles of Zyxin family proteins in different cancers (retrieved from Genevestigator) are presented. The bar graphs describe the levels of expression in cancer types. Evidently, the expression of Zyxin family proteins was high in all types of cancer.

Figure 10: Mutational frequency profiles of the Zyxin family proteins with different cancer malignancies. All proteins exhibit a considerable amount of mutational frequency which varied for each member. The highest mutation frequency was observed for Ajuba and the lowest was observed for LIMD1.

Figure 11: Protein-protein interaction analysis by Bioplex 3.0. The arrow indicates the bait and prey relationship of each protein. The bait protein is shown as a green circle whereas the prey protein is shown as grey squares. Clearly, TRIP6 has the highest number of interacting partners and WTIP has the least interacting partners. It is important to point out that certain zyxin family members were observed to interact with each other.

Figure 12: Pathway analysis by GeneMANIA. GeneMANIA highlighted the pathways that each member protein interacts within a cell. This helped us identify various proteins that interact with the query protein. The connections are shown by blue lines, the width of each line represents the degree of interaction through weighted representation. The thicker line represents more interaction and vice versa.

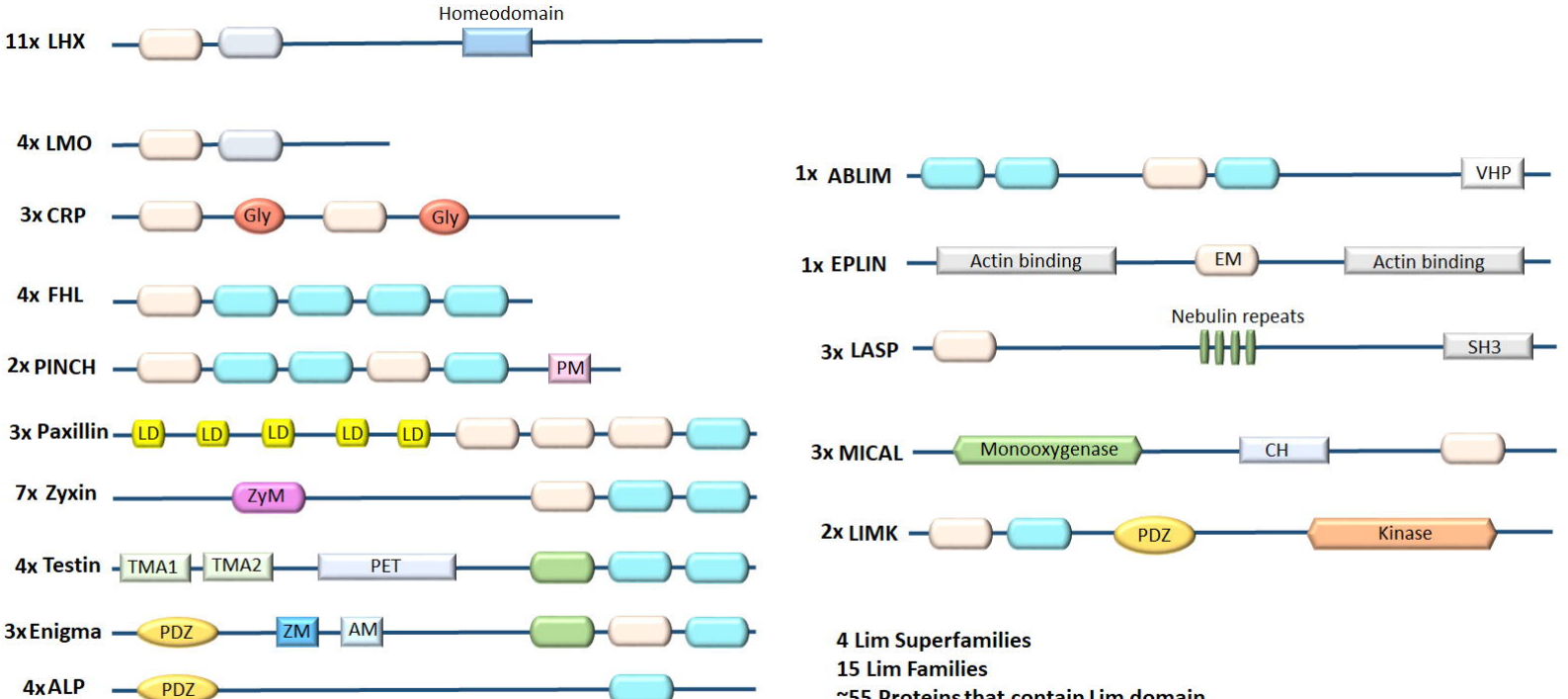
Homeodomain
11x LHX
4x LMO
3x CRP
Gly
Gly
4x FHL
2x PINCH
PM
3x Paxillin
LD
LD
LD
LD
LD
7x Zyxin
ZyM
4x Testin
TMA1
TMA2
PET
3x Enigma
PDZ
ZM
AM
4x ALP
PDZ
1x ABLIM
VHP
1x EPLIN
Actin binding
EM
Actin binding
Nebulin repeats
3x LASP
SH3
3x MICAL
Monooxygenase
CH
2x LIMK
PDZ
Kinase
4 Lim Superfamilies
15 Lim Families
~55 Proteins that contain Lim domain

Figure 1

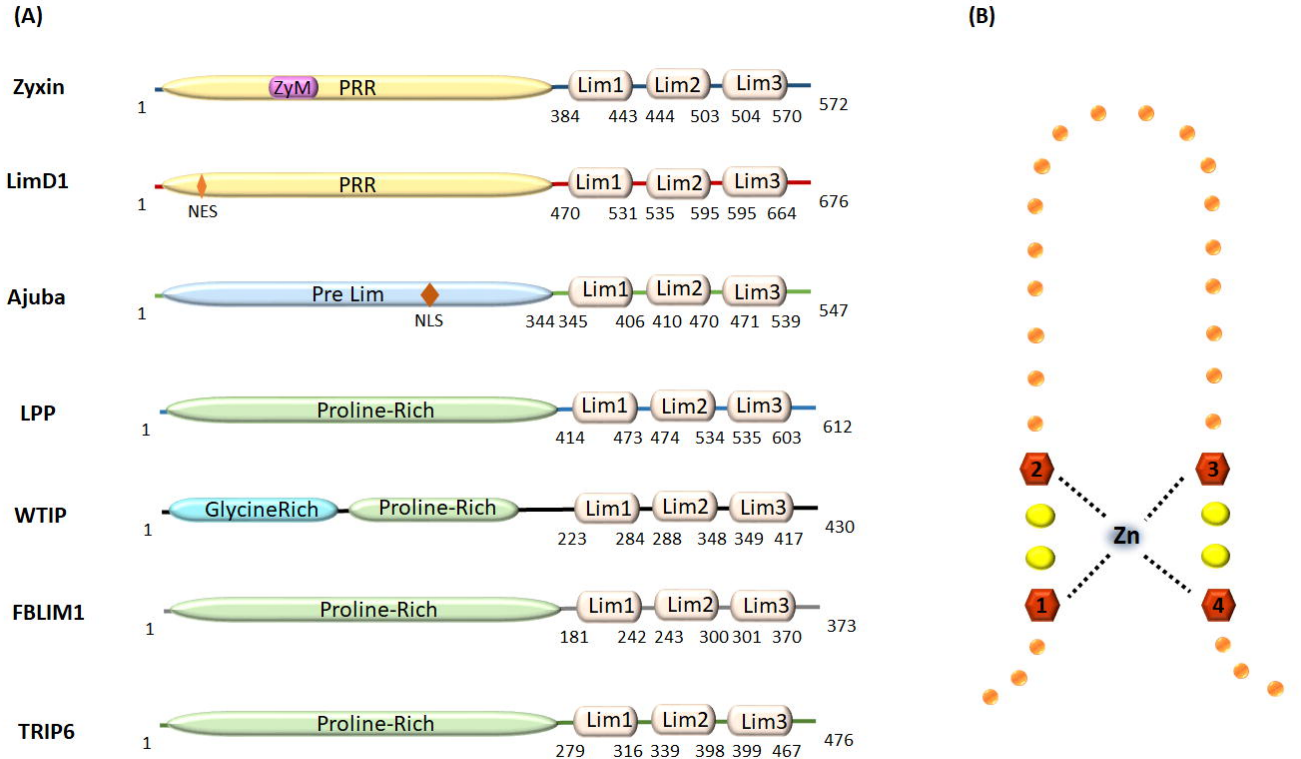
(A)
Zyxin
ZyM
PRR
Lim1
Lim2
Lim3
1
384
443
444
503
504
570
572
LimD1
PRR
NES
Lim1
Lim2
Lim3
1
470
531
535
595
595
664
676
Ajuba
Pre Lim
NLS
Lim1
Lim2
Lim3
1
344
345
406
410
470
471
539
547
LPP
Proline-Rich
Lim1
Lim2
Lim3
1
414
473
474
534
535
603
612
WTIP
GlycineRich
Proline-Rich
Lim1
Lim2
Lim3
1
223
284
288
348
349
417
430
FBLIM1
Proline-Rich
Lim1
Lim2
Lim3
1
181
242
243
300
301
370
373
TRIP6
Proline-Rich
Lim1
Lim2
Lim3
1
279
316
339
398
399
467
476
(B)
2
3
Zn
1
4

Figure 2

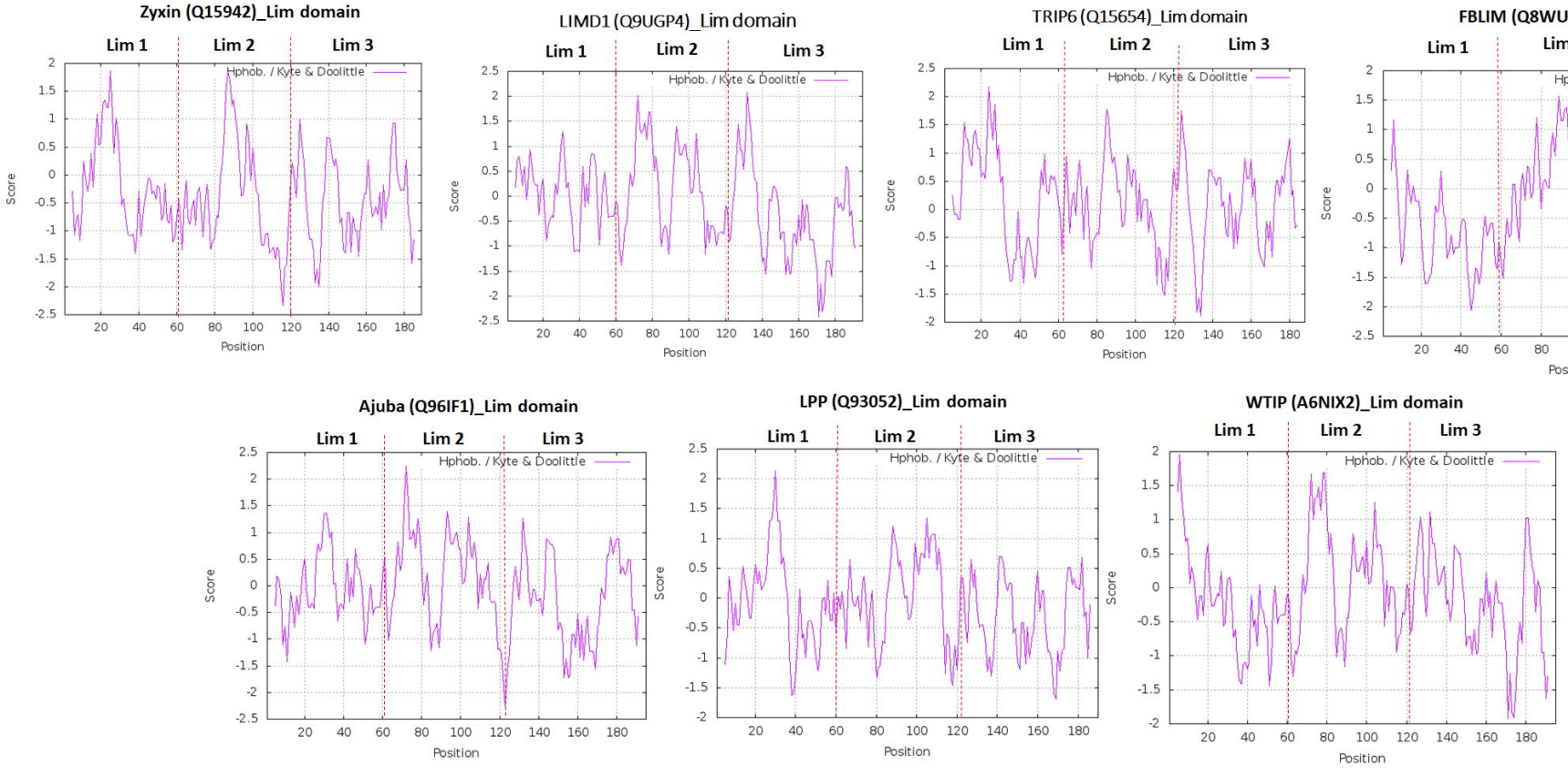
Zyxin (Q15942)_Lim domain
LIMD1 (Q9UGP4)_Lim domain
TRIP6 (Q15654)_Lim domain
FBLIM (Q8WUP2)_Lim domain
Ajuba (Q96IF1)_Lim domain
LPP (Q93052)_Lim domain
WTIP (A6NIX2)_Lim domain
Lim 1
Lim 2
Lim 3
Hphob. / Kyte & Doolittle
Score
Position

Figure 3

A

LIMD1	KFIEDLNMYEA
LPP	SLDAEIDSLTS
TRIP6	SLDAEIDLLSS
FBLIM1	DVLPDLDLLPP
Zyxin	LTLKEVEELEQ
WTIP	LLLAGLALREL
Ajuba	ELTALLRLTVG

B

Figure 4

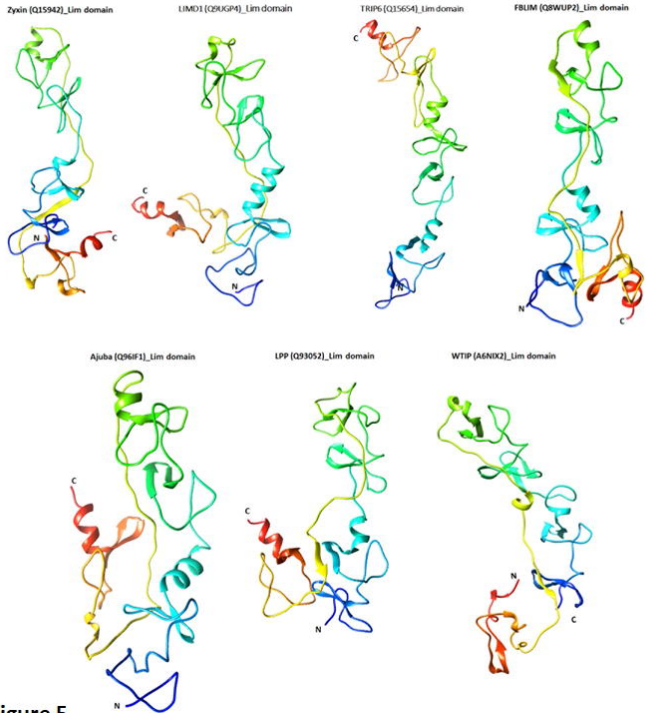
Zyxin (Q15942)_Lim domain
N
C
LIMD1 (Q9UGP4)_Lim domain
C
N
TRIP6 (Q15654)_Lim domain
C
N
FBLIM (Q8WUP2)_Lim domain
N
C
Ajuba (Q96IF1)_Lim domain
C
N
LPP (Q93052)_Lim domain
C
N
WTIP (A6NIX2)_Lim domain
N
C

Figure 5

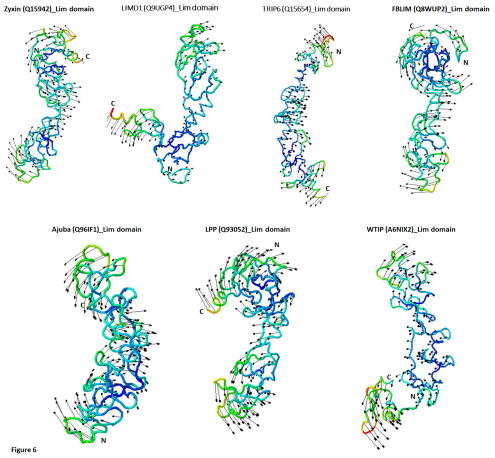
Zyxin (Q15942)_Lim domain
LIMD1 (Q9UGP4)_Lim domain
TRIP6 (Q15654)_Lim domain
FBLIM (Q8WUP2)_Lim domain
Ajuba (Q96IF1)_Lim domain
LPP (Q93052)_Lim domain
WTIP (A6NIX2)_Lim domain
C
N

Figure 6

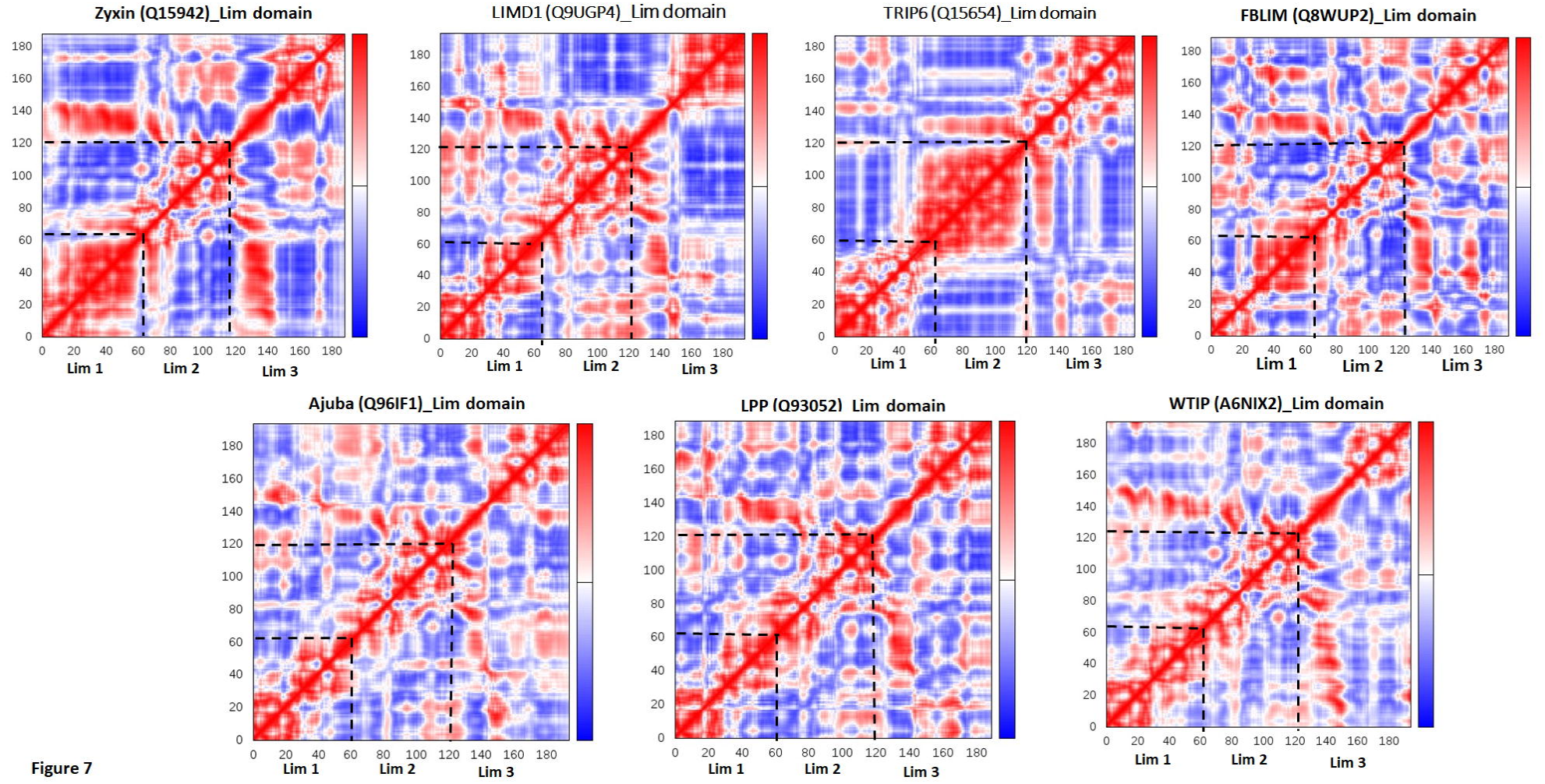
Zyxin (Q15942)_Lim domain
LIMD1 (Q9UGP4)_Lim domain
TRIP6 (Q15654)_Lim domain
FBLIM (Q8WUP2)_Lim domain
Ajuba (Q96IF1)_Lim domain
LPP (Q93052) Lim domain
WTIP (A6NIX2)_Lim domain
0 20 40 60 80 100 120 140 160 180
Lim 1
Lim 2
Lim 3

Figure 7

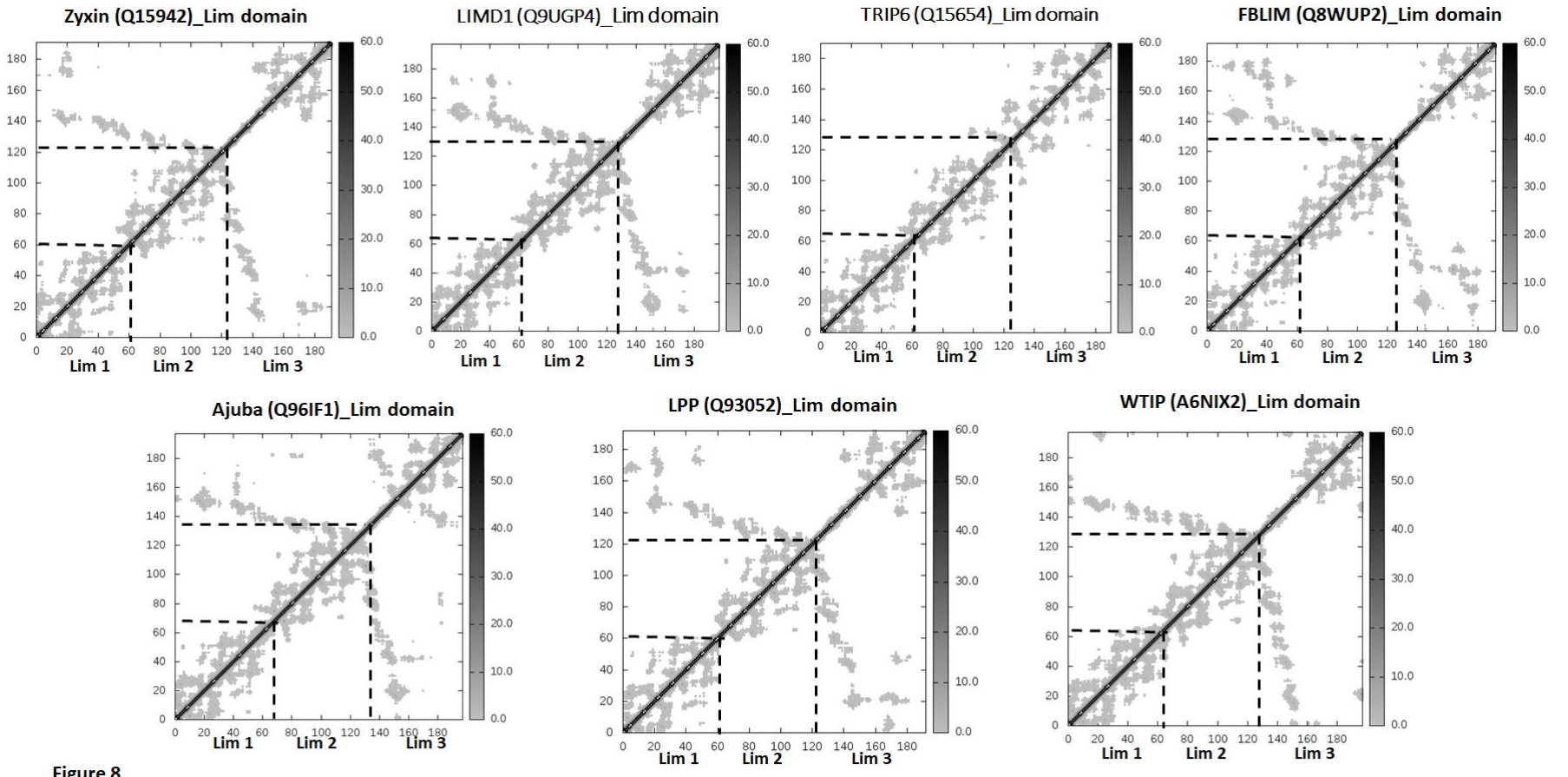
Zyxin (Q15942)_Lim domain
LIMD1 (Q9UGP4)_Lim domain
TRIP6 (Q15654)_Lim domain
FBLIM (Q8WUP2)_Lim domain
Ajuba (Q96IF1)_Lim domain
LPP (Q93052)_Lim domain
WTIP (A6NIX2)_Lim domain
Lim 1
Lim 2
Lim 3

Figure 8

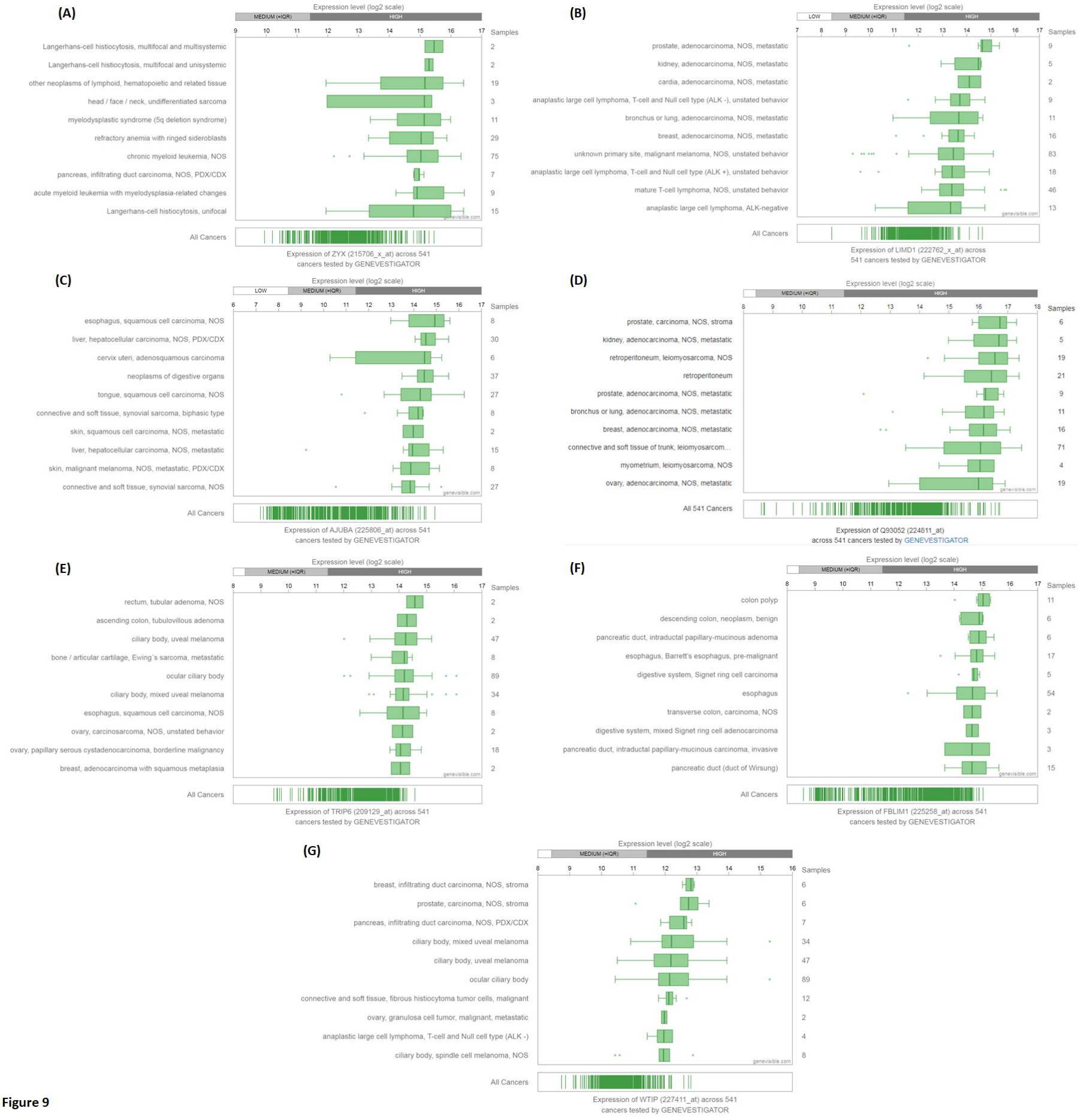
(A)
Expression level (log2 scale)
MEDIUM (±IQR)
HIGH
Samples
Langerhans-cell histiocytosis, multifocal and multisystemic 2
Langerhans-cell histiocytosis, multifocal and unisystemic 2
other neoplasms of lymphoid, hematopoietic and related tissue 19
head / face / neck, undifferentiated sarcoma 3
myelodysplastic syndrome (5q deletion syndrome) 11
refractory anemia with ringed sideroblasts 29
chronic myeloid leukemia, NOS 75
pancreas, infiltrating duct carcinoma, NOS, PDX/CDX 7
acute myeloid leukemia with myelodysplasia-related changes 9
Langerhans-cell histiocytosis, unifocal 15
All Cancers
Expression of ZYX (215706_x_at) across 541 cancers tested by GENEVESTIGATOR
(B)
Expression level (log2 scale)
LOW
MEDIUM (±IQR)
HIGH
Samples
prostate, adenocarcinoma, NOS, metastatic 9
kidney, adenocarcinoma, NOS, metastatic 5
cardia, adenocarcinoma, NOS, metastatic 2
anaplastic large cell lymphoma, T-cell and Null cell type (ALK -), unstated behavior 9
bronchus or lung, adenocarcinoma, NOS, metastatic 11
breast, adenocarcinoma, NOS, metastatic 16
unknown primary site, malignant melanoma, NOS, unstated behavior 83
anaplastic large cell lymphoma, T-cell and Null cell type (ALK +), unstated behavior 18
mature T-cell lymphoma, NOS, unstated behavior 46
anaplastic large cell lymphoma, ALK-negative 13
All Cancers
Expression of LIMD1 (222762_x_at) across 541 cancers tested by GENEVESTIGATOR
(C)
Expression level (log2 scale)
LOW
MEDIUM (±IQR)
HIGH
Samples
esophagus, squamous cell carcinoma, NOS 8
liver, hepatocellular carcinoma, NOS, PDX/CDX 30
cervix uteri, adenosquamous carcinoma 6
neoplasms of digestive organs 37
tongue, squamous cell carcinoma, NOS 27
connective and soft tissue, synovial sarcoma, biphasic type 8
skin, squamous cell carcinoma, NOS, metastatic 2
liver, hepatocellular carcinoma, NOS, metastatic 15
skin, malignant melanoma, NOS, metastatic, PDX/CDX 8
connective and soft tissue, synovial sarcoma, NOS 27
All Cancers
Expression of AJUBA (225806_at) across 541 cancers tested by GENEVESTIGATOR
(D)
Expression level (log2 scale)
MEDIUM (±IQR)
HIGH
Samples
prostate, carcinoma, NOS, stroma 6
kidney, adenocarcinoma, NOS, metastatic 5
retroperitoneum, leiomyosarcoma, NOS 19
retroperitoneum 21
prostate, adenocarcinoma, NOS, metastatic 9
bronchus or lung, adenocarcinoma, NOS, metastatic 11
breast, adenocarcinoma, NOS, metastatic 16
connective and soft tissue of trunk, leiomyosarcom... 71
myometrium, leiomyosarcoma, NOS 4
ovary, adenocarcinoma, NOS, metastatic 19
All 541 Cancers
Expression of Q93052 (224811_at) across 541 cancers tested by GENEVESTIGATOR
(E)
Expression level (log2 scale)
MEDIUM (±IQR)
HIGH
Samples
rectum, tubular adenoma, NOS 2
ascending colon, tubulovillous adenoma 2
ciliary body, uveal melanoma 47
bone / articular cartilage, Ewing's sarcoma, metastatic 8
ocular ciliary body 89
ciliary body, mixed uveal melanoma 34
esophagus, squamous cell carcinoma, NOS 8
ovary, carcinosarcoma, NOS, unstated behavior 2
ovary, papillary serous cystadenocarcinoma, borderline malignancy 18
breast, adenocarcinoma with squamous metaplasia 2
All Cancers
Expression of TRIP6 (209129_at) across 541 cancers tested by GENEVESTIGATOR
(F)
Expression level (log2 scale)
MEDIUM (±IQR)
HIGH
Samples
colon polyp 11
descending colon, neoplasm, benign 6
pancreatic duct, intraductal papillary-mucinous adenoma 6
esophagus, Barrett's esophagus, pre-malignant 17
digestive system, Signet ring cell carcinoma 5
esophagus 54
transverse colon, carcinoma, NOS 2
digestive system, mixed Signet ring cell adenocarcinoma 3
pancreatic duct, intraductal papillary-mucinous carcinoma, invasive 3
pancreatic duct (duct of Wirsung) 15
All Cancers
Expression of FBLIM1 (225258_at) across 541 cancers tested by GENEVESTIGATOR
(G)
Expression level (log2 scale)
MEDIUM (±IQR)
HIGH
Samples
breast, infiltrating duct carcinoma, NOS, stroma 6
prostate, carcinoma, NOS, stroma 6
pancreas, infiltrating duct carcinoma, NOS, PDX/CDX 7
ciliary body, mixed uveal melanoma 34
ciliary body, uveal melanoma 47
ocular ciliary body 89
connective and soft tissue, fibrous histiocytoma tumor cells, malignant 12
ovary, granulosa cell tumor, malignant, metastatic 2
anaplastic large cell lymphoma, T-cell and Null cell type (ALK -) 4
ciliary body, spindle cell melanoma, NOS 8
All Cancers
Expression of WTIP (227411_at) across 541 cancers tested by GENEVESTIGATOR

Figure 9

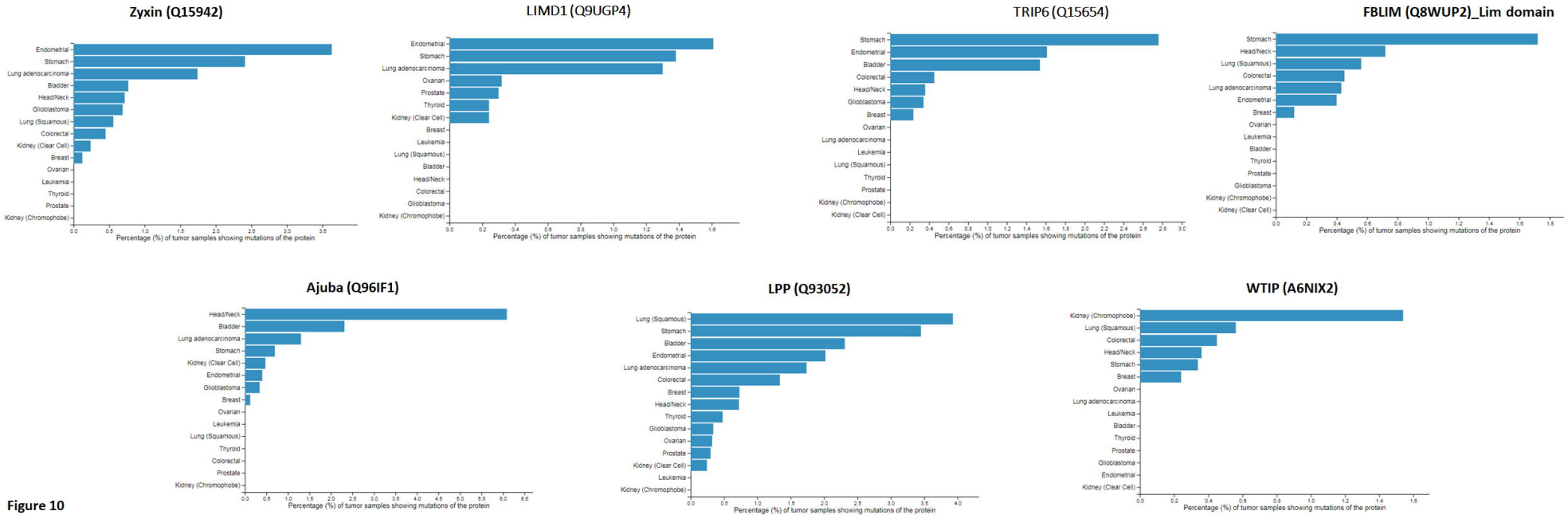
Zyxin (Q15942)
Endometrial
Stomach
Lung adenocarcinoma
Bladder
Head/Neck
Glioblastoma
Lung (Squamous)
Colorectal
Kidney (Clear Cell)
Breast
Ovarian
Leukemia
Thyroid
Prostate
Kidney (Chromophobe)
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Percentage (%) of tumor samples showing mutations of the protein
LIMD1 (Q9UGP4)
Endometrial
Stomach
Lung adenocarcinoma
Ovarian
Prostate
Thyroid
Kidney (Clear Cell)
Breast
Leukemia
Lung (Squamous)
Bladder
Head/Neck
Colorectal
Glioblastoma
Kidney (Chromophobe)
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6
Percentage (%) of tumor samples showing mutations of the protein
TRIP6 (Q15654)
Stomach
Endometrial
Bladder
Colorectal
Head/Neck
Glioblastoma
Breast
Ovarian
Lung adenocarcinoma
Leukemia
Lung (Squamous)
Thyroid
Prostate
Kidney (Chromophobe)
Kidney (Clear Cell)
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0
Percentage (%) of tumor samples showing mutations of the protein
FBLIM (Q8WUP2)_Lim domain
Stomach
Head/Neck
Lung (Squamous)
Colorectal
Lung adenocarcinoma
Endometrial
Breast
Ovarian
Leukemia
Bladder
Thyroid
Prostate
Glioblastoma
Kidney (Chromophobe)
Kidney (Clear Cell)
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8
Percentage (%) of tumor samples showing mutations of the protein
Ajuba (Q96IF1)
Head/Neck
Bladder
Lung adenocarcinoma
Stomach
Kidney (Clear Cell)
Endometrial
Glioblastoma
Breast
Ovarian
Leukemia
Lung (Squamous)
Thyroid
Colorectal
Prostate
Kidney (Chromophobe)
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5
Percentage (%) of tumor samples showing mutations of the protein
LPP (Q93052)
Lung (Squamous)
Stomach
Bladder
Endometrial
Lung adenocarcinoma
Colorectal
Breast
Head/Neck
Thyroid
Glioblastoma
Ovarian
Prostate
Kidney (Clear Cell)
Leukemia
Kidney (Chromophobe)
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
Percentage (%) of tumor samples showing mutations of the protein
WTIP (A6NIX2)
Kidney (Chromophobe)
Lung (Squamous)
Colorectal
Head/Neck
Stomach
Breast
Ovarian
Lung adenocarcinoma
Leukemia
Bladder
Thyroid
Prostate
Glioblastoma
Endometrial
Kidney (Clear Cell)
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6
Percentage (%) of tumor samples showing mutations of the protein

Figure 10

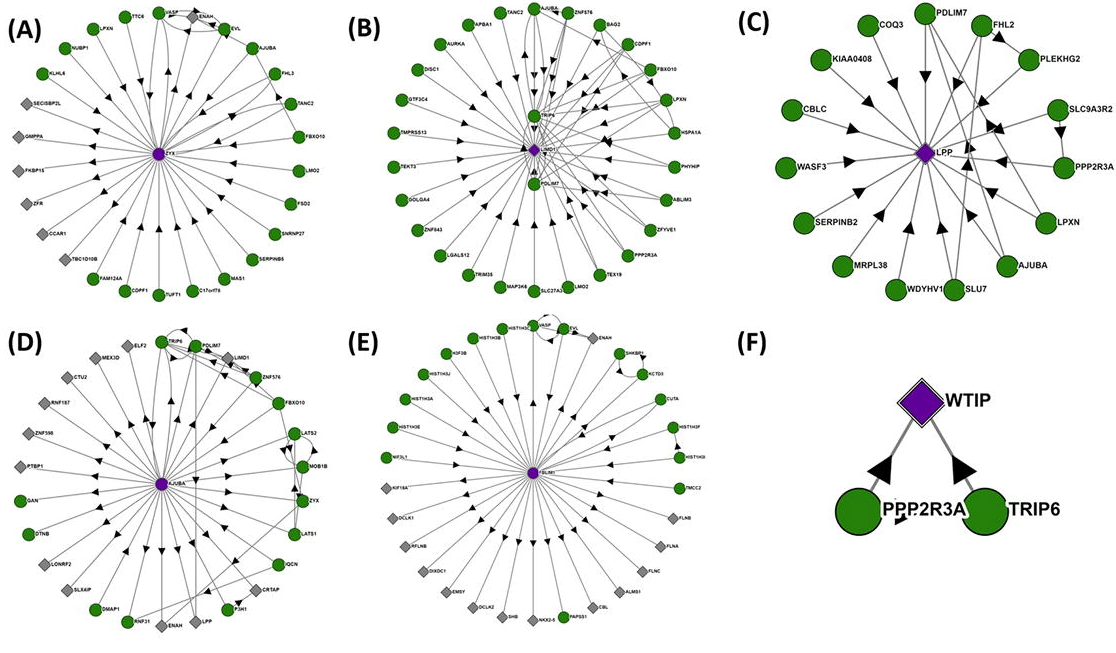
(A)
(B)
(C)
(D)
(E)
(F)
PDLIM7
COQ3
FHL2
KIAA0408
PLEKHG2
CBLC
SLC9A3R2
LPP
WASF3
PPP2R3A
SERPINB2
LPXN
MRPL38
AJUBA
WDYHV1
SLU7
WTIP
PPP2R3A
TRIP6

Figure 11

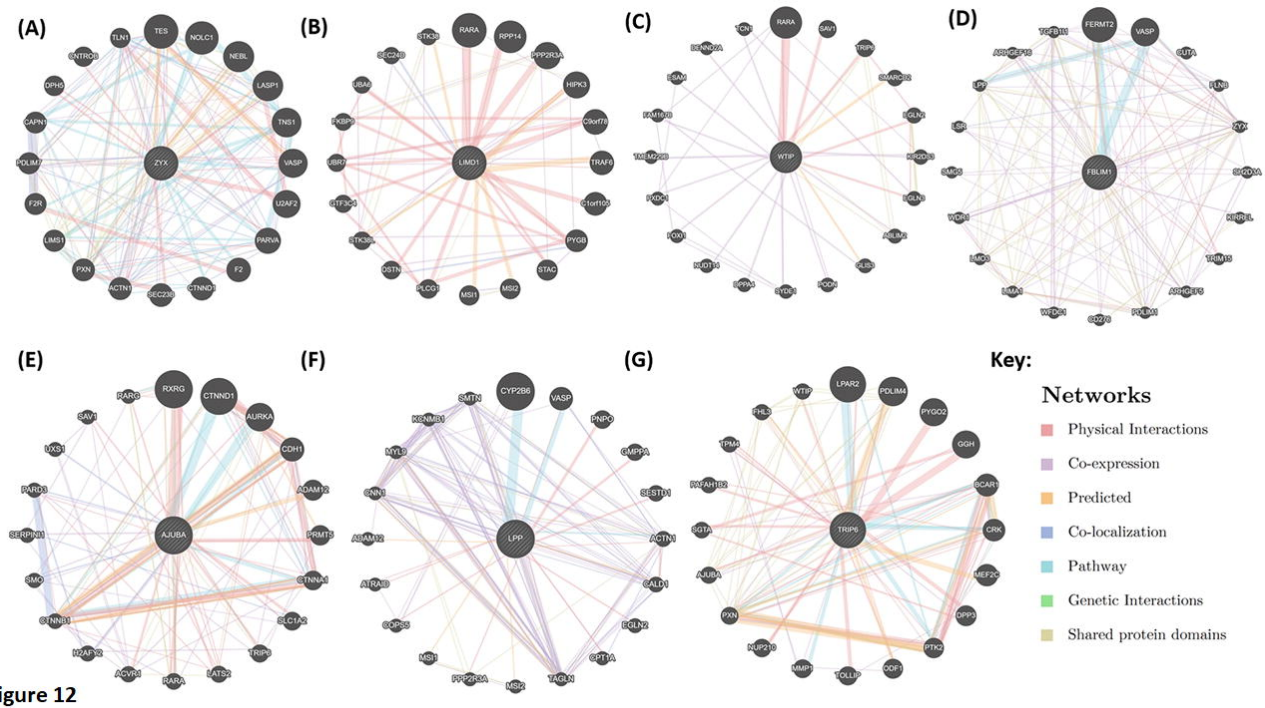
(A)
ZYX
TES
NOLC1
NEBL
LASP1
TNS1
VASP
U2AF2
PARVA
F2
CTNND1
SEC23B
ACTN1
PXN
LIMS1
F2R
PDLIM7
CAPN1
DPH5
SNTRCB
TLN1
(B)
LIMD1
RARA
RPP14
PPP2R3A
HIPK3
C9orf78
TRAF6
C1orf103
PYGB
STAC
MSI2
MSI1
PLCG1
OSTN
STK36
SEC24B
JBA6
FKBP9
UBR7
GTF3C4
STK36
(C)
WTIP
RARA
SAV1
TRIP6
SMARCB2
EGLN2
KRCDD3
EGLN3
ABLIM2
SLIS3
PCDH1
SYDE1
PPPA4
KLOT14
POXN1
PXDC1
TMEM22B
FAM16B
ISAM
BCNNDZA
TCN1
(D)
FBLIM1
FERMT2
VASP
CUTA
FLNB
KYX
SPCDA
KIRREL
TRIM15
ARHGEF5
PDLIM1
CDZ78
WFDC1
LIMA1
LIMD3
WDR1
SMC5
LSR
LPP
ARHGEF15
TGFB1I1
(E)
AJUBA
RXRG
CTNND1
AURKA
CDH1
ADAM12
PRMT3
CTNNA1
SLCTA2
TRIP6
LATS2
RARA
ACVR1
H2AFY2
CTNNB1
SMO
SERPINI1
PARD3
UXS1
SAV1
RARG
(F)
LPP
CYP2B6
VASP
PNPO
GMPPA
SESTD1
ACTN1
CALD1
EGLN2
CPT1A
TAGLN
MSI2
PPP2R3A
MSI1
COPS5
ATRAID
ADAM12
CNN1
MYL9
KCNMB1
SMTN
(G)
TRIP6
LPAR2
PDLIM4
PYGO2
GGH
SCAR1
CRK
MEF2C
DPP3
PTK2
CQF1
TOLLIP
MMP1
KUP210
PXN
AJUBA
SGTA
PAFAH1B2
TPM4
FHL3
WTIP
Key:
Networks
Physical Interactions
Co-expression
Predicted
Co-localization
Pathway
Genetic Interactions
Shared protein domains

Figure 12