Analyzing Intracellular Short Linear Motifs of AMIGO and NGL Orthologs

A Major Qualifying Project Report

Submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

In

Biology and Biotechnology

By

___________________________________
Doreet Nagatti

January 23, 2017

APPROVED:

________________________
Joseph B. Duffy, Ph. D.
Biology and Biotechnology
WPI Project Advisor
ABSTRACT

LIGs are a family of transmembrane proteins, containing a leucine-rich repeat (LRR) and an immunoglobulin-like (Ig) domain, important in cell interactions and signaling. There are 36 human LIG proteins, of which the AMIGO subfamily and NGL subfamily have sizeable intracellular domains for which minimal functional knowledge has been obtained. Within intracellular regions of transmembrane molecules short linear motifs (SLiMs) that function as targeting signals, modification sites, and protein binding sites often exist. Identification of motifs conserved across different species provides a phylogenetic approach to aid in the discovery of functional SLiMs. In this study, orthologs of the AMIGO and NGL human proteins were identified in Mus musculus (mouse), Gallus gallus (chicken), Callorhinchus milii (elephant shark) and used to identify putative SLiMs.
ACKNOWLEDGEMENTS

I would like to thank Professor Duffy for creating this amazing opportunity for me and contributing to the success and completion of my project. Thank you for clarifying my project goals, and giving me guidance, feedback, and support throughout the project. Additionally, I want to thank Duffy for his career and life advice as my academic advisor and friend for the past four years. Thank you for always believing in me and pushing me to challenge myself and reach my full potential.

Finally, I would like to thank the faculty and staff of the Biology and Biotechnology Department for all their support throughout my undergraduate career and Worcester Polytechnic Institute for giving me the opportunity to complete this project and receive my Bachelor of Science Degree.
# TABLE OF CONTENTS

ABSTRACT ................................................................. 2  
ACKNOWLEDGMENTS ..................................................... 3  
TABLE OF CONTENTS .................................................... 4  
INTRODUCTION ............................................................. 5  
MATERIALS AND METHODS ............................................. 10  
RESULTS ................................................................. 13  
DISCUSSION .............................................................. 26  
REFERENCES ............................................................... 28  
APPENDICES ............................................................... 30
INTRODUCTION

Transmembrane proteins are essential for cell-cell interactions and cell signaling. LIGs are a subset of transmembrane proteins that with an extracellular domain, containing a set of leucine-rich repeats (LRRs) followed by an immunoglobulin-like (Ig) domain (s) and an intracellular domain of varying length (MacLaren et al., 2004). Figure 1 below shows a graphic image of one LIG protein, Kek1, whose function in Epidermal Growth Factor signaling has been well documented (Ghiglione et al, 1999; Alvarado et al. 2004). The image shows the LRRs in red, as well as the single Ig domain in blue.

![Figure 1. Kek1 LIG protein structure](image)

There are 36 total human LIG proteins, including the LINGO, NGL, SALM, NLRR, Pal, ISLR, LRIG, GPR, Adlican, Peroxidasin-like proteins, Trk neurotrophin receptors, AAI11068, and AMIGO subfamilies (Homma et al. 2008). Figure 2 below shows the number proteins in the human proteome containing either LRRs only (350), Ig domains only (1100), or the combined presence of both (36).
Various studies have been completed on the functions of the proteins in cellular signaling and their extracellular domains, but not much is understood about the intracellular domains of these proteins. Of the human LIG proteins, the AMIGO subfamily and NGL subfamily have sizeable intracellular domains, and therefore were chosen for analysis in this study.

The AMIGO subfamily of LIGs consists of three proteins, AMIGO1, AMIGO2, and AMIGO3 (Kuja-Panula et al., 2003). Structurally, the AMIGO proteins contain seven LRRs and one Ig domain. Figure 3 below shows the structure of AMIGO1.

**Figure 2. Presence of LRRs and Ig domains in the human proteome.**

**Figure 3. AMIGO1 LIG protein**
AMIGO proteins appear to function as regulators for the phosphoinositide 3-kinase (PI3K) – 3-phosphoinositide-dependent kinase 1 (PDK1) – protein kinase B (Akt) signaling pathway (Park et al. 2015). This pathway is important for extracellular signaling that controls cell growth, survival, metabolism, angiogenesis, and protein translation. AMIGO2 specifically regulates localization of PDK1 at the plasma membrane, which in turn activates Akt. Improper regulation of this pathway is thought to be associated with various metabolic, cardiovascular, and neurological diseases. Additionally, other studies have found that improper activation of the pathway contributes to the likelihood of cancer, by tumor angiogenesis and metastasis (Park et al. 2015). Thereby, studying the cellular functions of the AMIGO proteins in the PI3K – PDK1 – Akt signaling pathway is valuable in research for cancer prevention and therapy.

The NGL subfamily of LIGs also consists of three proteins, NGL1, NGL2, and NGL3. In contrast to the AMIGO family structure, NGL proteins contain nine LRRs and one Ig domain (Woo et al., 2009). Figure 4 below shows the structure of NGL1.

![NGL1 LIG protein](image)

Figure 4. NGL1 LIG protein
NGL proteins function as trans-synaptic cell adhesion molecules (CAMs) that bind a family of Netrin-G ligands. NGL1 and NGL2 have been found to bind Netrin-G1 and Netrin-G2, respectively (Woo et al. 2009). Figure 5 below shows the proposed binding of NGL1 to Netrin-G1. Netrin-G1 and G2 are structurally related to Netrins, a family of molecules important in axon guidance, but are distinct in that they are linked to the membrane by a glycosyl phosphatidyl-inositol (GPI) lipid anchor, rather than secreted like Netrins (Woo et al. 2009). Moreover, they also do not bind to the classical Netrin receptors, Deleted in Colorectal Cancer (DCC) and Unc5 (Woo et al. 2009).

**Figure 5. Netrin-G1/NGL1 binding**

Cell adhesion molecules, or CAMs, have been found to be involved in many aspects of synapse development, including synapse formation, differentiation, trans-synaptic signaling, and structural and functional synaptic changes (Woo et al. 2009). As putative CAMs, due to their structure and membrane association, the Netrin-G/NGL complexes are thought to be associated with synaptic diversification and neural circuit functions in vertebrates, which is essential for
information processing in the brain (Matsukawa et al. 2014). Additionally, NGL1 and NGL2 have intracellular postsynaptic density-95/ disks large/ zona occludens-1 (PDZ) binding sites found in many synaptic proteins that interact with scaffolding proteins (Matsukawa et al. 2014). Understanding the cellular functions of the NGL proteins in the Netrin-G/NGL complexes may provide better insight to the functions of the vertebrate brain and open avenues for new therapies for diverse brain dysfunctions.

Because of the presence of LRRs and Ig domains, well-defined sequence elements, the extracellular regions of LIG proteins are generally well understood. In contrast, due to the lack of analyses on intracellular regions of LIG proteins the contribution of the intracellular regions to LIG function and cell signaling are less clear. In these intracellular regions, the absence of previously defined domains or repeats suggests that they may function in mechanisms distinct from many canonical signaling pathways. One hypothesis is that there are many short linear motifs (SLiMs), which are small regulatory interfaces of ~3-10 amino acid residues that function as targeting signals, modification sites, and protein binding sites (Edwards et al. 2007; Davey et al. 2012). Identifying such motifs, or SLiMs, within the intracellular region of LIGs is a key step in understanding the contribution of their intracellular regions to cell signaling.

In order to identify potential intracellular motifs conserved between members of the AMIGO and NGL families in different species, orthologs of the human proteins were identified and compared to several other jawed vertebrate species. Of these, the comparison included two bony vertebrates – a mammal, Mus musculus (mouse) and a reptile, Gallus gallus (chicken), and a cartilaginous fish, Callorhinchus milii (elephant shark). Evolutionary comparisons of the intracellular domains of these species were made for the both the AMIGO and NGL families in order to identify putative conserved motifs and potential cellular significance.
MATERIALS AND METHODS

Identification of members of the LIG family in jawed vertebrates

For approximately half of the human LIG family (appendices) accession numbers from Table 1 of *Gene Expression Patterns* (Homma et al. 2008) were entered into the NCBI Protein database ([https://www.ncbi.nlm.nih.gov/protein](https://www.ncbi.nlm.nih.gov/protein)) to obtain the respective LIG protein sequences. The FASTA sequence representing the complete sequence of each protein was downloaded and saved.

The human AMIGO and NGL FASTA protein sequences for each member of the subfamilies were then entered into the BLAST protein database to find putative orthologs in the mouse, chicken, and elephant shark genomes. The FASTA sequences were entered on [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and the following tax IDs representing sequence databases for the respective species were entered in the organism box for individual searches: *Mus musculus* (Tax ID10090), *Gallus gallus* (chicken) (Tax ID 9031), and *Callorhinchus milii* (elephant shark) (Tax ID7868). Putative orthologs for a particular LIG were identified as the cross species match with the highest identity to the human protein query and then further confirmed by a reciprocal BLAST of that highest scoring protein match back to the human genome. Matches that reciprocally identified the initial human LIG protein query as the highest match were defined as orthologs.

Identification of intracellular domains using CCTOP

The human, mouse, chicken, and shark AMIGO and NGL protein sequences were entered into the CCTOP prediction server ([http://cctop.enzim.ttk.mta.hu/?_=jobs/submit](http://cctop.enzim.ttk.mta.hu/?_=jobs/submit)) to identify the putative signal peptide, extracellular, transmembrane, and intracellular regions. Protein
sequences were color coded to distinguish the extracellular (green), transmembrane (blue), and intracellular (red) domains.

Alignment of intracellular domains using Clustal Omega and Boxshade

To create protein sequence alignments, the human, mouse, chicken and shark AMIGO and NGL intracellular domain sequences were entered into the Clustal Omega sequence alignment program (http://www.ebi.ac.uk/Tools/msa/clustalo/) with a >”LIG Name” before each intracellular domain. For example for NGL1 the following was entered for each species followed by their IC Domains: >HsNGL1, >MmNGL1, >GgNGL1, >CmNGL1

The Pearson/Fasta output format was selected and the alignments were run with standard parameters and saved.

Next, the AMIGO and NGL alignments were entered into the Boxshade software (http://embnet.vital-it.ch/software/BOX_form.html). Consensus line with letters was chosen and 1.0 was checked for the fraction of sequences option, representing 100% conservation among the input proteins. Boxshade was run and the output was saved for each LIG. Each boxshade alignment was subjected to visual analysis to identify motifs. Motifs were defined as being at least 4 consecutive conserved amino acids, motifs were considered distinct if there were several nonconserved amino acid amino acids between them. The motifs were highlighted in yellow in the consensus line and labeled for each LIG.

Creation of graphic for motifs with Weblogo

The AMIGO and NGL motifs were entered into the Weblogo3 software (http://weblogo.threeplusone.com/create.cgi). PDF output was selected, protein for sequence type, and error bars were unselected. Custom color scheme was selected and the symbols and
colors for the standard Chemistry amino acid classification scheme (Figure 6) was used for graphical representation were entered.

<table>
<thead>
<tr>
<th>Class</th>
<th>Amino Acids</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polar</td>
<td>G,S,T,Y,C</td>
<td>green</td>
</tr>
<tr>
<td>Neutral</td>
<td>Q,N</td>
<td>purple</td>
</tr>
<tr>
<td>Basic</td>
<td>K,R,H</td>
<td>blue</td>
</tr>
<tr>
<td>Acidic</td>
<td>D,E</td>
<td>red</td>
</tr>
<tr>
<td>Hydrophobic</td>
<td>A,V,L,I,P,W,F,M</td>
<td>black</td>
</tr>
</tbody>
</table>

**Figure 6. Chemistry Amino acid classification scheme**
RESULTS

To gain better insight to the mechanism by which LIG family members transduce extracellular cues into cellular responses, analyses of the intracellular domains of two LIG subfamilies, AMIGO and NGL, were performed to identify potential SLiMs. In order to identify any such motifs, a phylogenetic approach was undertaken and is outlined in a flow chart below (Figure 7). The overall approach relied on the notion that functionally important sequences are conserved, while sequences not under functional constraints diverge.

Figure 7. Steps in Identification of Motifs

Identification of members of LIG family in jawed vertebrates

In order to find orthologs for specific human LIGs, the full amino acid sequences of all LIGs was obtained (Materials and Methods and D. Anina, personal communication). The remainder of the analyses focused on the AMIGO and NGL subfamilies, both of which contain
three members in the human genome. The human AMIGO 1, 2, and 3 sequences can be seen in appendices A, B, and C, respectively. The human NGL 1, 2, and 3 sequences can be seen in appendices D, E, and F, respectively. The full amino acids sequences for the rest of the human LIGs identified can be found in appendix G.

In order to find orthologs of phylogenetic utility to the human LIG’s, species that were evolutionarily distinct enough to allow time for sequence selection and divergence needed to be selected. However, if the species that were too distant with respect to evolutionary time were chosen, this might prevent the accurate identification of a given set of orthologs for a specific LIG. To prevent this, a sequential approach was taken, initially orthologs for the human AMIGOs and NGLs were screened for in two bony vertebrates representing a close relative - the mouse genome (~75Myr), and a more distant relative - the chicken genome (~310Myr), followed by a significantly more distant vertebrate relative from the cartilaginous fishes – the elephant shark genome (~450Myr) (Waterston et al., 2002; ICGSC, 2004; Venkatesh et al., 2014). Figure 8 below shows a phylogeny of this vertebrate lineage, where sharks are represented by elephant shark in the figure, reptiles by chicken, and mammals by humans and mice.
Figure 8. Phylogeny of Vertebrate Lineage (Adapted from Smith et al. 2015)

In Figure 8, sharks, reptiles, and mammals are represented on different branches of the phylogenetic tree, denoting evolutionary relationships. Thereby, looking for orthologs of human LIG’s in these lineages provides time for protein divergence, but less comparatively to lampreys and lancelets. Since the aim was to identify orthologs that had significant divergence relative to human LIG’s, the mouse genome is a less useful comparison because they are still within the mammalian lineage and exhibit minimal sequence divergence. However, taking a stepwise approach by identifying a close ortholog in mouse followed by sequentially more distant orthologs allowed for more confidence in ortholog predictions in the chicken and elephant shark genomes. Looking at the chicken genome is a useful comparison within the reptile lineage as *Gallus gallus* was the first avian genome to be sequenced and has evolutionary distance to humans useful in determining functional elements (Schmutz et al. 2004). Finally, looking at the elephant shark genome is a useful comparison both due to its distance from humans and the fact
that *C.milli* has been found to have the slowest evolving genome of all vertebrates, making it a good model for understanding evolutionary change (Venkatesh et al. 2014).

Using the Blast protein database the human AMIGO and NGL protein sequences were used to find matches in the mouse, chicken, and elephant shark. These AMIGO 1, 2, and 3 orthologs can be seen in appendices A, B, and C, respectively. The NGL 1, 2, and 3 orthologs can be seen in appendices D, E, and F, respectively. Table 1 and 2 below summarize the AMIGO and NGL orthologs that were found across all four species. Interestingly, orthologs of Amigo 3, NGL 2, and NGL3 were not found in chicken, but were present in the more distantly related *C. milli*.

<table>
<thead>
<tr>
<th></th>
<th>Homo sapiens</th>
<th>Mus musculus</th>
<th>Gallus gallus</th>
<th>Callorhinchus milli</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AMIGO1</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>AMIGO2</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>AMIGO3</strong></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 1. AMIGO Orthologs (+/- is presence/absence of protein)

<table>
<thead>
<tr>
<th></th>
<th>Homo sapiens</th>
<th>Mus musculus</th>
<th>Gallus gallus</th>
<th>Callorhinchus milli</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NGL1</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>NGL2</strong></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>NGL3</strong></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 2. NGL Orthologs (+/- is presence/absence of protein)

**Identification of Intracellular Domains**

In order to identify conserved intracellular motifs within the orthologs of the AMIGOs and NGLs, the intracellular regions of the proteins had to be defined next. Conservation of
intracellular sequences among the different orthologs can help in discovery of functional SLiMs. Using the CCTOP protein software the intracellular domains of the orthologs were defined. CCTOP is a program used to predict the location of the transmembrane region of proteins, thereby predicting the location of the extracellular and intracellular domains. There are many programs that use different algorithms of hydrophobicity, structural information, and ranking residues to predict these domains. While the programs have different algorithms, a transmembrane domain is generally an alpha-helical stretch of about 18 hydrophobic residues; CCTOP finds a consensus among those algorithms.

The AMIGO and NGL protein sequences were color-coded using the CCTOP output to distinguish the extracellular (green), transmembrane (blue), and intracellular (red) domains. The color-coded sequences of the AMIGO and NGL orthologs can be seen in appendices A-F. The CCTOP was run for all the human LIGs and the corresponding color-coded sequences are in appendix G.

Looking at the length of the intracellular domains is useful because the longer the intracellular domain, the more likely SLiMs will be found in those intracellular domains. The lengths of the intracellular domains of the AMIGO and NGL orthologs were calculated and are summarized in Tables 3 and 4. Given a general length of ~100 residues for their intracellular domains and a typical length of ~3-10 residues for a SLiM, the AMIGOs and NGLs are likely to each contain a number of SLiMs.

<table>
<thead>
<tr>
<th>Homo sapiens</th>
<th>Mus musculus</th>
<th>Gallus gallus</th>
<th>Callorhinchus milii</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AMIGO1</strong></td>
<td>98</td>
<td>99</td>
<td>101</td>
</tr>
<tr>
<td><strong>AMIGO2</strong></td>
<td>102</td>
<td>99</td>
<td>101</td>
</tr>
<tr>
<td><strong>AMIGO3</strong></td>
<td>97</td>
<td>101</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 3. AMIGO IC Domain Lengths**
Identification of conserved Intracellular Sequences

With the intracellular regions defined, orthologs for a given set AMIGO or NGL proteins were aligned using the Clustal Omega alignment software to reveal conserved sequences that represent putative SLiMs. The Clustal alignments for the AMIGO and NGL intracellular domains can be seen in appendices H and I, respectively. After performing Clustal alignments, Boxshade was used to shade identical amino acids found in all four species black and similar amino acids grey. From the Boxshade outputs, putative motifs in each set of AMIGO and NGL orthologs were determined based on stretches of conserved amino acids between the four species. These were then compared between subfamily members to identify motifs conserved within the subfamily as well. The boxshade outputs and identified motifs for each set of AMIGO and NGL orthologs were highlighted and labeled (Figures 9-14).

<table>
<thead>
<tr>
<th></th>
<th>Homo sapiens</th>
<th>Mus musculus</th>
<th>Gallus gallus</th>
<th>Callorhinchus milii</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGL1</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>91</td>
</tr>
<tr>
<td>NGL2</td>
<td>106</td>
<td>106</td>
<td>0</td>
<td>93</td>
</tr>
<tr>
<td>NGL3</td>
<td>117</td>
<td>114</td>
<td>0</td>
<td>94</td>
</tr>
</tbody>
</table>

Table 4. NGL IC Domain Lengths
Figure 9. AMIGO1 Conservation and Motifs

Figure 10. AMIGO2 Conservation and Motifs

Figure 11. AMIGO3 Conservation and Motifs
This resulted in the identification of six distinct motifs in the AMIGO subfamily and six distinct motifs within the NGL subfamily as well. Tables 5 and 6 below show which of the six motifs can be found in each of the AMIGO and NGL proteins, respectively.

<table>
<thead>
<tr>
<th>Motif</th>
<th>Motif 1</th>
<th>Motif 2</th>
<th>Motif 3</th>
<th>Motif 4</th>
<th>Motif 5</th>
<th>Motif 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMIGO1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AMIGO2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AMIGO3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 5. AMIGO Motifs (+/- is presence/absence of motif)
Table 6. NGL Motifs (+/- is presence/absence of motif)

<table>
<thead>
<tr>
<th></th>
<th>Motif 1</th>
<th>Motif 2</th>
<th>Motif 3</th>
<th>Motif 4</th>
<th>Motif 5</th>
<th>Motif 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGL1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NGL2</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NGL3</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Graphical Representation of Motif Conservation

To better characterize conservation within the motifs, WebLogo was used to represent the frequency and biochemical properties of each amino acid in the motifs. The software generates a graphical representation that shows the amino acids with various sizes depending on how many of the four species they are found in. The amino acids are also color coded based on functional properties (corresponding colors can be seen in Figure 6 from the Methods). The resulting outputs are shown for each motif of the AMIGO’s and NGL’s in Figures 15-26.
Figure 18. AMIGO Motif 4 Weblogo

Figure 19. AMIGO Motif 5 Weblogo

Figure 20. AMIGO Motif 6 Weblogo
Figure 21. NGL Motif 1 Weblogo

Figure 22. NGL Motif 2 Weblogo

Figure 23. NGL Motif 3 Weblogo
Figure 24. NGL Motif 4 Weblogo

Figure 25. NGL Motif 5 Weblogo

Figure 26. NGL Motif 6 Weblogo
DISCUSSION

To gain a better understanding of the role of LIGs in cellular signaling, identification of SLiMs in two LIG subfamilies, AMIGO and NGL, was undertaken. Using a phylogenetic approach the intracellular domains of the AMIGO and NGL protein orthologs for four evolutionary distinct species, Homo sapiens (human), Mus musculus (mouse), Gallus gallus (chicken), and Callorhinchus milii (elephant shark), were compared. Based on conservation of amino acids within the intracellular domains between these species, motifs were determined for each LIG. A total of six different conserved motif sequences each were identified for the AMIGO and NGL proteins. All of the six motifs for each LIG display significant conservation across the four selected species. The identification of SLiMs and their degree of conservation strongly suggests that there is biological significance for these motifs.

However, contrary to the assumption that the species more closely related would likely have more conserved regions with each other, chickens, which are more closely related to humans than elephant shark, did not have orthologs to the human AMIGO 2, NGL 2, and NGL 3 proteins, while the elephant shark did. This suggests the possible loss of functionality/necessity of these proteins for the chickens, while the proteins were still biologically relevant for the elephant shark.

Additionally the six motifs for each LIG were found and compared amongst the three proteins of each LIG subfamily (AMIGO 1, 2, and 3; NGL 1, 2, and 3). It was found that only AMIGOs 1 and 2 contained motifs 1, 2, and 3, suggesting that AMIGO 3 did not need these motifs for functionality. On another note, only AMIGO 2 contained motif 4 and only AMIGO 3 contained motifs 5 and 6, suggesting that these motifs are functionally significant, specifically for those AMIGO proteins.
Similarly for the NGL subfamily, only NGL 1 contained motif 2, only NGL 2 contained motif 4, and only NGL 3 contained motifs 5 and 6, suggesting their biological relevance to those proteins. The NGL motifs 1 and 3 were found in all three NGL proteins so they are conserved among NGLs and are not likely to serve a specific significance to just one NGL, but may represent a level of functional redundancy among the proteins.

The identification and characterization of a set of SLiMs in this study confirmed that there are significant areas of evolutionary conservation within the AMIGO and NGL LIG proteins’ intracellular domains. This conservation strongly supports a functional role for these motifs and provides insight into possible functional specificity and redundancy across family members. Guided by their identification, further analysis on each of the motifs will help in determining the essential role of these motifs in the functionality of each protein and its relevance to cellular communication.
REFERENCES


APPENDICES

Appendix A AMIGO 1 Orthologs

Adhesion molecule with Ig-like domain 1 [Homo sapiens]

>gi|26454838|gb|AAH40879.1| Adhesion molecule with Ig-like domain 1 [Homo sapiens]

MHPHRDRGPLWLFLSLFLSLLLFEVARAGAVSCPAAACLASNLSCKQQLPNVPHLSLPSYTALLDLSHLNSLSLRL
AEWTPTTRLQHLLSLHLHFISSEAEFPVPNLRLYDLSNQLRTLDEFLFSDLVQLHEVLVLLYNHYAMADEVCAF
DMDAQLKLYLSQONISRFPELELVKAEGLPMLTLALLDSLSSNKLKLPLPDQLKLPAL1KNGLYHNNPLNCDELYQ
LFSHQYQLSSMVDFQEDVCMNNSKLHNVFLSNLCNGYERKAWAEALGLDTLIIKCDTQEQGMTKVWTFSNERS
VLDEVTNQTVSVDGSLLFQQVQVEDGGVYCTAMGETFNETLSVELKVNFTLHGHHTLNTAYTTLVG
CILSVVLVLILYLT

amphoterin-induced protein 1 isoform a precursor [Mus musculus]

>gi|51988879|ref|NP_001004293.1| amphoterin-induced protein 1 isoform a precursor [Mus musculus]

MOPQRLRGLWLSSLFLVFLFLFEVARAGSVSCPANCLASNLCSKQQLPNVPSLPSYTALLDLSHLNSLSLRL
AEWTPTTRLNHLHSLSHLNFISSEAFVPVNLRLYDLSNHLHTDEFLFSDLVQALEVLLLYNYNHIVVDRNAF
EDMAQLKLYLSQONISRFPEVCLNKLKGPLDDLSSNKLKLPLTLQKLPAPWKNGLYLHNPPLECSPLYQ
LFSHQYQRCLSSMVDFQEDVCMHSHALHLNFSCDFQNEISKESNWAEALGLDTLIIKCDTQEQGMTVWPSNSEQ
VLSQGNSGSSVRNGDLFFKQVQVEDGGVYCTAMGETFNETLSVELKVYNFTLHGHHTLNTAYTTLV
GGCILSVVLVLILYLT

Ampoterin-induced protein 1-like [Gallus gallus]

>gi|971451924|ref|XP_015130504.1| PREDICTED: LOW QUALITY PROTEIN: amphoterin-induced protein 1-like [Gallus gallus]

MAVPGAVLAVLAVPAPVAGSCCPFRVCASNLSCSRAALSSVPAFLRFSVLDLSSHNNISLRADWAAGRLA
HLHALLLHANGLAVFSTEVAGPHVPHRLDHLLSNRRLALEENFLSDPELVEVLFLYNEISAIDRASADFNLSSLRKL
YLGRNHIARPFLLELRDGRSPQQLLDDLLSNRLLSPLAAPELQALPAFLRLDYVHGNPLGDCDPLYLVARGRHR
LWAVLDFQEEFLQLPAPAAGRPAIVLELSPELDCACAAREAVLAEAYLDGSVTLCDSRLRAAHGRHWFFGGRDV
PEEGNGSSAVALANGSLOLRALRPEDGGTGYACVRSGPAFNETLYVELVHNFTLHGPDGLNTAYTTLV
CILSVVLVLILYLT

Ampoterin-induced protein 1 [Callorhinchus milii]

>gi|632962942|ref|XP_007897606.1| PREDICTED: amphoterin-induced protein 1 [Callorhinchus milii]

MAVPGAVLAVLAVPAPVAGSCCPFRVCASNLSCSRAALSSVPAFLRFSVLDLSSHNNISLRADWAAGRLA
HLHALLLHANGLAVFSTEVAGPHVPHRLDHLLSNRRLALEENFLSDPELVEVLFLYNEISAIDRASADFNLSSLRKL
YLGRNHIARPFLLELRDGRSPQQLLDDLLSNRLLSPLAAPELQALPAFLRLDYVHGNPLGDCDPLYLVARGRHR
LWAVLDFQEEFLQLPAPAAGRPAIVLELSPELDCACAAREAVLAEAYLDGSVTLCDSRLRAAHGRHWFFGGRDV
PEEGNGSSAVALANGSLOLRALRPEDGGTGYACVRSGPAFNETLYVELVHNFTLHGPDGLNTAYTTLV
CILSVVLVLILYLT
Appendix B AMIGO 2 Orthologs

Adhesion molecule with Ig-like domain 2 [Homo sapiens]

>gi|28839672|gb|AAH47595.1| Adhesion molecule with Ig-like domain 2 [Homo sapiens]
MSLRVHTLPTLGLAVRPGCRLELLCLLMITTVGGPASGVCPTACICATDVCSTKNKLKVPGNLFRLLKRLDLYN
RRIGLLDSEWIPVSFAKLNTILLRNHNTISSTGFSSTPPNLKCLDLSLSNLKTLTQVKAQVFELLVVYLLHNHHIS
YLDPSAFGGLSQKLYLJGNTFQFMDLVGRFLAEMLMFLDVSYRNPWPMHMINLVPQKHRLGYILHGNPRVC
DCSLSYSLVFYRHRFSVWMSDFKNDYTCTRLWSRHSQVLLQDSFMNCDSNISNGFRAFGFHEAQVEERLMV
HCDSTKGNANDTDFIWGPNNLRLEPDGAGMENFVYHVGLNSVIESPFEDAVGVSICAMNKQRLLNETVDVTINVNSF
TVSRSHAHEAFNATTTA

Amphoterin-induced protein 2 precursor [Mus musculus]

>gi|30017449|ref|NP_835215.1| amphoterin-induced protein 2 precursor [Mus musculus]
MSLRFHTLPTLPRAKPGCRELLCLLVIAVMSPSASGSCTACICATDVCSTKNNLSVPGNLERLLKRLDYSN
RIGLLDADWIPSVFKLSTILLRNHNITSSTGFSSTPPNLKCLDLSNLKSVKSATFQELKALVEVLLHNHYIS
LDPAAFGQLSHQKLYLSNGFLTQFPMDLTFGRKLDLTFLDVSYRNPSPIMHNLLNPQOQLGFIHLHGPNFVC
DCSLLSILIFVRHRFFSVMSDFKNDYTCTRLWSRHSQHQLQLESFLCNCSYVGHSVHALGFIHEAQVEARAH
CDSTGNINTDFIWGPNNLRLEPDKGMNFRVFVYHLNSVIEPNGFEDAVGVSICAMNQRLLNETVDIMVSNF
TNRSHAHEAFNATTTTA

Amphoterin-induced protein 2 precursor [Gallus gallus]

>gi|313760565|ref|NP_001186479.1| amphoterin-induced protein 2 precursor [Gallus gallus]
MSLCRTTLPTLQIGAKVMCR NelCVLFFAVSVGSAPGMCPTTICICADICSTKNNLSVPGNLYRMKRLDYSN
RIGFLPEWIPVLFKNTILSITGFSSTTPNLKYLLDSNLKTLTGFVPQELGTLTEVLLKNQITH
IESSAFGGLYKQKFLYSLNFLLHFPDDLDFVGHKLTELILLDIFHQNISMPIQRLSVPAKLHSGLYVLHNPY
DCTLSMLIFWQHRFSVDFQTKYCTRSLPFRGYNQKLLHDNFLNCSESTIINSSQAFGFIHDAQVGRDLH
VDSRISDAGTHFVWVSPENKLELPDMETDKRFRFVHNSLESITDAOLEDGLYSAITAIKRLLNETIVRINSNF
VNRSHAHEAFNATTTTA

Notes:
- The sequences provided are amino acid sequences.
- These sequences are orthologs to AMIGO 2, which is associated with adhesion molecules and immune system functions.\n- The sequences are aligned to show similarity and differences in amino acid composition and structure.\n- The alignment includes species-specific sequences to highlight evolutionary relationships.

This information is crucial for understanding the functional diversity and evolutionary history of adhesion molecules and related proteins across different species.
Amphoterin-induced protein 2 [Callorhinchus milii]
>gi|632937805|ref|XP_007901161.1| PREDICTED: amphoterin-induced protein 2 [Callorhinchus milii]
MTCSHKKAYSADVRAALTLCRQFVLLLLCVMAGNAALICFPVICASDIVCTCNRRNLNSVPRTLHKVATSOLDSY
SISLLTNSWAPVSLDRLRTLNLHNHNNIAISRGAFCAPQLKYLDDLSSNRLTALDDSLFEDLSLETLLLNYNQIAR
VSTGAFEGLKLQKLQLSQNLISHFPLQLYMGSRKLPLEELLDDLDFNLKLTSPVQLQASLAPRQLQSLYHANFPTC
DCSFYTMVTYWKRFVSTMDKDYSCNLQOLDKRTVSSLLMRDDDLNCSNSTFSLNALGMYEAHHIRGVYNN
CDSKILDNLNTVNLWVTPTENSLQSGIQYQGLQLVFLNGSLEIQVQPEDEGIYSCIAINSRRMLNETIVTLKHNFT
QERHRQTFNTAFTTLSAC
LASSIVLVIYLYLTPC
RCWCKSKQRRHRKPGNSARSSILSTTPHDVTNERKASTCKRVVFLEPVEKPLGQNGKIKFQPHPHIVTEKILRAK
RACKSDSSISSVFSDNLIVA

Appendix C AMIGO 3 Orthologs

Adhesion molecule with Ig-like domain 3 [Homo sapiens]
>gi|111493932|gb|AAI10419.1| Adhesion molecule with Ig-like domain 3 [Homo sapiens]
MTWLVLLLGLCMRLVRGTFDSEGGFPFPRLHNCPYKCIAADLSCCTGGLQDVPAILPAAPATAADLDLSHNALQTRL
PGWALPFLQRALHDLHNDALGRGVFVNASGLRLLDLSSNTLRALGRHDLDGLGAKEKLFLFNRLVHLEDEAHF
GLRLSSHLYGCNLASFSFDHLGLSATHLDLDDLSSNLGRHISVFELAALPAFLKNGLYHNNLPCDRCRTYHL
QRWHQRLGQLAVRDWAREYVLKAPKVASRVRFPQHSSVRVFENCSSAPALGERPEHLYALVGRSLRLCINTSVPAMR
IAWVSQPEQELLRAGPSRDGIAVLADGSLAIGNVQEIQHYAGLFVLCLATGPRLHNNQTREYNVSVHFPREPEAFNTG
TTLLGC
AVGLVLVLYLFPFPCC
RCCRRACRCRRWQPSTPSLQELSAQSSVLSTTPDAPSRKASVHKHVFWLEPGRGRGRLNGVRQVALAEFDLYNPGL
QLKAGSESASSGEGEPMITT

Amphoterin-induced protein 3 precursor [Mus musculus]
>gi|288933532|ref|NP_796249.1| amphoterin-induced protein 3 precursor [Mus musculus]
MAWLVLLGILLCMLAGGLTSLEDVVPPAPHNCPIICIAADVLSACGRGLQDLFVALPTTAELDLSSNHNLKRLH
PGWALPSLRLHRLHYNKEVNLHGHAPNASKLRLDLSSNLRLMHTHLDLGELEEKLFLFNNSLMHDLDFAFQ
GLRLSHEYLCNSELSSFSFNHLHGLGLTRLRTLDDLSSLKHIPELAALTPLYKNRLYHNNPLPCDSLHLYHL
RRWHQRLGQLALHFDREETVLKVFESRVSFREHRSRSVFKCNSVAAAPGELPEQLLAHQVGQSLRLFCNTSVPATR
VAWVSQKNLLVAPASQDSGIAVLADGSLAIGNVQEIQHYAGLVCLASGPRLHNNQTLEYNVVSQKARPETPFNTG
TTLLGC
IVGLVLVLLLYLFAPPCC
RGCCHCQRACRNRCPASSPLQELSAQSSMLSTTPPDAPSRKASVHKHVFWLEPGBKGLNGVRQVALAEFDLYNPGL
PMGLQKAGSESASSGEGEMLVSN

No gallus gallus Amigo 3
Amphoterin-induced protein 3 [Callorhinchus milii]

>gi|632946798|ref|XP_007888736.1| PREDICTED: amphoterin-induced protein 3 [Callorhinchus milii]
MRGPGSAGSVLWLSVGLWQFIGKSQASLHCPAVGCASDLSCSVQNSVPLPARLPEATSLLSHNLLLQLH
DNRLSHLPRTTLRANHNRHAIATKAAFSQGLTHLDSLQRTNLYSVEKHFRELTHLELLELYNNQIARVEDGAL
RLSSLQKVLNSQNLTHFFGSQLESTLRPLKVIDISSNWSSFPQVIALSNKGLYHLNNPLVDCVLYSML
LHHWKEQFSIYDFQEEHTCTAGQPRVSLRLHRKLFDNCTYASHGLGLVDNVATVGESLLCVNQLEHL
TTYWITENKELIGYFGSNFKFMYNGSLEIRRTQKDSGIGYICMATNQLMRNESQEVNVTYLRSKDEGFTNT
GLTILL
GCVVLSSLVLVMYLYLT
PCRCWCKTPHPHPTPNEACAQSSLISLATSPPCNEDANRTGGKKHVVFLEPVKDSQNKGKRLAVSDFPDVKNPKIL
QLKSDSESITSVFSDTPIMS

Appendix D NGL 1 Orthologs

NGL1, LRRC4C protein [Homo sapiens]

>gi|73909151|gb|AAH41374.3| LRRC4C protein [Homo sapiens]
MLNKMTLHPQQIMIGPRFNRALFDPLLVLALLQLLVLVAGLVRAQTCSVSCSNQFSKVICVRKLRVPGDSTN
TRLLNLHENQIQIITKNSFKHLRHELQLRSNHRTIEIGAFNLQLNLEFLDNRLLTIPNGAFVYLSKELW
LRNRPIESISYAFNRIPSRLDLGELKRLSYISEGAFEGSLRNYLMNLREIPNLKDELDLSGNHL
SARIPGSQGLMHLQWMIQSIQIIVERNAFNDLQSLVEINLHNNLTLPLLDFPLTLHLRRHLHHLHHPWNCND
ILWLSWIKDAMSNTACCARCNTPPNLKGRYGLEDQNYFTPCTYAPVIVEPPADLNVCMAELKRCRSTLSVS
WITPNTVMTGHAYKVRIAVLSGTLNFTNVTQDTGMYTCMVSNSVNTASATLNTAATTHTPFSYFSTVETSM
EPSQDEARTTDNNVGPVPVTDWETNVTTSLQPSTRKTEKTFIPVTDINSIGPIDEMKTTK
IIIGCFVAILMAAAMVLFIF
YKMRKQHHQHNNHPTRTVETEINVDDEITGTDPMESHPMAIEHEHLNHNSYSKPFFNHTTVNTINSIHSSVHEP
LLIRMNSKDNVQETQI

Leucine-rich repeat-containing protein 4C precursor [Mus musculus]

>gi|224944244|ref|NP_848840.3| leucine-rich repeat-containing protein 4C precursor [Mus musculus]
MLNKMTLHPQQIMIGPRFNRALFDPLLVLALLQLLVLVAGLVRAQTCSVSCSNQFSKVICVRKLRVPGDSTN
TRLLNLHENQIQIITKNSFKHLRHELQLRSNHRTIEIGAFNLQLNLEFLDNRLLTIPNGAFVYLSKELW
LRNRPIESISYAFNRIPSRLDLGELKRLSYISEGAFEGSLRNYLMNLREIPNLKDELDLSGNHL
SARIPGSQGLMHLQWMIQSIQIIVERNAFNDLQSLVEINLHNNLTLPLLDFPLTLHLRRHLHHLHHPWNCND
ILWLSWIRDMAPSNTACCARCNTPPNLKGRYGLEDQNYFTPCTYAPVIVEPPADLNVCMAELKRCRSTLSVS
WITPNTVMTGHAYKVRIAVLSGTLNFTNVTQDTGMYTCMVSNSVNTASATLNTAATTHTPFSYFSTVETSM
EPSQDEARTTDNNVGPVPVTDWETNVTTSLQPSTRKTEKTFIPVTDINSIGPIDEMKTTK
IIIGCFVAILMAAAMVLFIF
YKMRKQHHQHNNHPTRTVETEINVDDEITGTDPMESHPMAIEHEHLNHNSYSKPFFNHTTVNTINSIHSSVHEP
LLIRMNSKDNVQETQI
Leucine-rich repeat-containing protein 4C [Gallus gallus]

>XP_004941608.1 PREDICTED: leucine-rich repeat-containing protein 4C [Gallus gallus]
MLNKMTLHPQMIGPKFNRALDPDLFLVVLALQQLLVAVGLVRAQTCPSCSNQFSKVICTRMLREVPSGSLTVSCTRNLNLNQEMIQQIKVSFKHLRHELQLSLRNHIRTIEIGAFNGLANNNTLFNRLDITPNAGAFVLYLSKLKELW
LMNNPESIPSYAFRNPILRLDLGELKREIYSDAIFDSLSNLNLNLQMCNLQMRPSLMVLKLEELCGNLRL SQIRPGFQGQLNLQKLMMAIPIQVIERNADDDQLSLILELNNNTLLLPHDLTPPLFRHELHVLHNNPSCEC
ILWLSWIKDRAPSNCOMMCHTPPNLKGISICEDQKNCNPFLNCYPVIEAPTDLNLTHEAGMAELKCRASMSRVSW
ITPNSVMVTHGAYVRVIAVLDGTLLNKTVQDHTLGYTCMVSNSAGNTTASATLNVXTDNPGYTFSTVTETVE
PSQDAAQTTSEQGPVTSTWENNTSTLSFTPSTRSTETKFTITPVSTANIPGIDEMVKTTK
LIIICGFAITALMAAAMAVLMIF
YKMRKQQHRQRNHHAPRTVIEINVDDELTGDTPIESLHPMAIEHEHNLHNSYKSFNPHTTTVNTINSIHSSVHEP
LLIRMNSKDNVQETQI

Leucine-rich repeat-containing protein 4C [Callorhinchus milii]

>XP_007885838.1 PREDICTED: leucine-rich repeat-containing protein 4C [Callorhinchus milii]
MLNKMTLHPQMIGPKFNRALDPDLFLVVLALQQLLVAVGLVRAQTCPSCSNQFSKVICTRMLREVPSGSLTVSCTRNLNLNQEMIQQIKVSFKHLRHELQLSLRNHIRTIEIGAFNGLANNNTLFNRLDITPNAGAFVLYLSKLKELW
LMNNPESIPSYAFRNPILRLDLGELKREIYSDAIFDSLSNLNLNLQMCNLQMRPSLMVLKLEELCGNLRL SQIRPGFQGQLNLQKLMMAIPIQVIERNADDDQLSLILELNNNTLLLPHDLTPPLFRHELHVLHNNPSCEC
ILWLSWIKDRAPSNCOMMCHTPPNLKGISICEDQKNCNPFLNCYPVIEAPTDLNLTHEAGMAELKCRASMSRVSW
ITPNSVMVTHGAYVRVIAVLDGTLLNKTVQDHTLGYTCMVSNSAGNTTASATLNVXTDNPGYTFSTVTETVE
PSQDAAQTTSEQGPVTSTWENNTSTLSFTPSTRSTETKFTITPVSTANIPGIDEMVKTTK
LIIICGFAITALMAAAMAVLMIF
YKMRKQQHRQRNHHAPRTVIEINVDDELTGDTPIESLHPMAIEHEHNLHNSYKSFNPHTTTVNTINSIHSSVHEP
MRVNSKDNVQETQI

Appendix E NGL 2 Orthologs

NGL2 Leucine rich repeat containing 4 [Homo sapiens]

>gi|109730363|gb|AA11562.1| Leucine rich repeat containing 4 [Homo sapiens]
MKLLNQVTNHHWTNWAINLPPFVYLTACQPWILCAAAIAASAPQNPCSNSQFSKVCTRGLSEVPQPISNT
RYLNLMEUQIQRQIADFTFRMLHLHLEVLQGLRNSISRQIEVAGFNLASLNTLELFDNWTIVPSGAEYLSKHRELLW
RNPPESIPSYAFRNPILRLDLGELKREIYSEAFGELFLONLYNLGDNCLNMDPMLTPLVGEELEMSSNGHP
EIRPGFSFHLSSLKDNMNSQVSLIERNADDDQLVEIENLNNLSPLLHDLTPPLRTLYELVEHLHNNPSCEC
ILWLSWIKDRAPSNCOMMCHTPPNLKGISICEDQKNCNPFLNCYPVIEAPTDLNLTHEAGMAELKCRASMSRVSW
ITPNSVMVTHGAYVRVIAVLDGTLLNKTVQDHTLGYTCMVSNSAGNTTASATLNVXTDNPGYTFSTVTETVE
PSQDAAQTTSEQGPVTSTWENNTSTLSFTPSTRSTETKFTITPVSTANIPGIDEMVKTTK
LIIICGFAITALMAAAMAVLMIF
YKLRKRRQQRSTVTAARTEVEIIVQVDEPIATSAATAAAPSGVSSEGANVLPTIHDTNHYNTYPKPAHGAGHTENSGL
NSLHTVTITISEPYIIQTHTKDKVQETQI
Leucine-rich repeat-containing protein 4 precursor [Mus musculus]

>gi|124339785|ref|NP_619623.2| leucine-rich repeat-containing protein 4 precursor [Mus musculus]
MKLLNWQTVHHTWNAVLLPVPYTLATQWVILCAATAAASAGPGQNPSC5VSCSNQFSKSVCTRGLSEVPQGIPSNTRYNLMEMNIQMIQADTFRHRHLEQLQLQRGRNSRIRQIEVGAFLAGNLSLNTLELFDNLTVIPSAGFEYLSKRLRELLWRRNPISIPSYAFNPVRSLRDLGDGKYLLESEGAGFLQNLKLYNLGMNCTMPNLPFVGLVEELEMGSNHFPEIRPGSFHGLSSLKWLWMNSQVSLIEIRNAFDGLASLVELNLAHENLSSLPHDLFTPLRYVELHELHHLNPWNCDDIL

WLAWLREYIPNTSCCGRCPAMHRGRYLVDELVDQAFCSAIPMDAPRLNISREDAMAELCRTMPMSSVWKLLPNGTMLHASRHPISVLNDGTNLNFSRLVLIIDTVGYTCMVTNVAGNSASAYLNVSSAELENPFSFTTVTETEISFEDTRQKVPTTTSTQPAYTTSTVTTLIQQTTRVPKVQVPFSTDTDDMQTSLDEVMFTK
IIIGCFVAVTLLLAAALIVF

YKLRRKHHQQRSTVAAARTVEIIQVQDEIDPAAPAAATAAAPSGVSSEGAVVLPTIHDHINYTYKPAHGAHWENSLGNLHPTVTISSEPYYIQHTHDKVQETQI

No gallus gallus NGL 2

Leucine-rich repeat-containing protein 4 [Callorhinchus milii]

>NP_007907340.1 PREDICTED: leucine-rich repeat-containing protein 4 [Callorhinchus milii]
MCHTMNLWQTVHHTWNAALVLFYLSARMSVCAASGREQSCPTICSCSNQFSKVSCTRGLREVPQGIPSNTRYLNLMENIQLIQADTRHLVEQLQRGRNSRIRQIEVGAFLAGNLSLNTLELFDNLTVIPSAGFEYLSKRLRELLWRNNPISIPSYAFNPVRSLRDLGDGKYLLESEGAGFLQNLKLYNLGMNCTMPNLPFVGLVEELEMSGHFPIQGFSFLGKSLRKLWMNSQISVIERNFDDLTELDLVELNLASHNLSLPHDLFPRLRYLVQLHHLNPWNCDDIL

LAWLREYIPNNTFCGCRCHTPAHMRGYKVTEVDPGSFCSPGFVILEPPSNVINISERGTVKLRCTADMASVRNLPPNTFELHSGSAPRLSVFNGQTLHFLHLLTDGTYTCTVANMVAGASASALLHVTMAEINTANYTFSTTVTETTE

TVRTKVPFFLSTPTYKVFPSTPTVLUQSTRSRPRALPVPTDSDLIRASLDEVMFTK
IIIGCFVAVTLLLAAALIVF

YKLRRKHHQQRSTVAAARTIEIIINVEEMAGGGPGEGGGGSVPSVHDHMNYTYKPAHRAHWTDNSLGNLHHTTEPFPITQTHNKDVKQETQI

Appendix F NGL 3 Orthologs

NGL3 NP_001073926.1 [Homo sapiens]

>gi|122937309|ref|NP_001073926.1| leucine-rich repeat-containing protein 4B precursor [Homo sapiens]
MARARGSQCPLPPGRMSVPHGALLFLWLFSPPLAGGGAATVSAAGGGSPATSCPVACSCSNQASRVICTRDDLAEVPASIPVTRTNYLQENIQGVIRTDTFKHLRHRHEIQLSKNLVRKIEVAGFNLSLNTLELFDNLTVPTQAF

EYLSKRLRELLWNNPISIPSYAFNPVRSLRDLGDGKYLLESEGAGFLQNLKLYNLGMNCTMPNLPFVGLVEELEMGSNHFPIQGFSFLGKSLRKLWMNSQISVIERNFDDLTELDLVELNLASHNLSLPHDLFPRLRYLVQLHHLNPWNCDDIL

LAWLREYIPNNTFCGCRCHTPAHMRGYKVTEVDPGSFCSPGFVILEPPSNVINISERGTVKLRCTADMASVRNLPPNTFELHSGSAPRLSVFNGQTLHFLHLLTDGTYTCTVANMVAGASASALLHVTMAEINTANYTFSTTVTETTE

TVRTKVPFFLSTPTYKVFPSTPTVLUQSTRSRPRALPVPTDSDLIRASLDEVMFTK
IIIGCFVAVTLLLAAALIVF

YKLRRKHHQQRSTVAAARTIEIIINVEEMAGGGPGEGGGGSVPSVHDHMNYTYKPAHRAHWTDNSLGNLHHTTEPFPITQTHNKDVKQETQI
Leucine-rich repeat-containing protein 4B precursor [Mus musculus]

MAQAHIRGSPCPPLLPGRMWSPHAGALLLLWFLSFPLLAGGGVATSSAAGGSPPTATSCPAACSCSNQASRVICTRERLAEVPSAPIVNLRYNLQENISIQVIRTDTFKHLRLEILQLSKNLVRKIEVGAFNGLPSLNTLLEFNDERLTTVPTQAFEYLSKLRLELRLRNPSIESIPSAYAFNRVSPSLRLDDLGEKRLREYISEAAFEGLVNLRLYNLGMNLCIDPINTLALVRELEELSGNLDDLIRPGFSQGLTSLRLKLWMLHAQVATIERNAFDDDLSLEELNLSSHNLMSLPDLFTPLHLRLEVLNLHNPWHCNCDVLSLWSWLKETVPSNNTCCARCHAPAGLKRIGELDQHSTCFVAPVIVEPPDLNVTEGMAAELKCRTGTSMTSVNLTPNLMLTHSGYVRISVLHDFTLNFNTQDGTQYCTMNTSAGNTTASATLNVSAVDPAAGGPGGGGPGGGGGAGGAGAGGAGGAGGVETVETLETQPGEAAQPRGTIEKEPPPGPTTDGAWGGGRPDAAPAPASTTAPAPSSRPEASFPITPVTENALKDLDVMKTTKIIGCFVAITMMAAIAVMLVAFYKLKOHQLKHHGPTVTVEIINVEDELPAAASVSSVAAAAAVAGGAGVGDSHLALPALERDHLNHHHYVAAAFKAHYGGNPGGCGAKGPKLNSHEIPLFHKSGSKENVETQI

No gallus gallus NGL 3

Leucine-rich repeat-containing protein 4B-like [Callorhinchus milii]

>XP_007882747.1 PREDICTED: leucine-rich repeat-containing protein 4B-like [Callorhinchus milii]

MNULLMKVHSQRMASLGRTVSLVLLSWSALSGAGLAGVHCCEFCSWQFSKIVCRTEHRELREVPAISTNTNLQENVICVIKADTFQKRLHLIELQLSMMNLRHIEVGAFLGSLNNTLELFDDNRLTVPQGAEFLSSKRLEWLRRNPSIESIPSAYAFNRVQLRLDGLGELKLYEISDAAFEGLMNLCNLVIMPNLSTRELLESQNLREIIQPGFSQQLTSLRLKLWMLHAQVQLIERNAFFDLQSEELNLSNNTLSLHDLTFPLHRLRDLVHLNHPWHCNCDVLSLWSWLKETVPSNNTCCARCHAPAGLKRIGELDQHSTCFVAPVIVEPPDLNVTEGMAEALKCRAATSMTSVMWMTPNGTLMTHGYSRISVLHDFTLNFNTQDGTQYCTMNTSAGNTTASATLNVSAVDPTNSYFTTFTVETEVVDEPKAGFFEPGPTSSGDASYSTSTLSAPRSTRTTERVFVTIPITEVMNIMAGLDDVMKTTKIIGCFVAITMMAAIAVMLVAFYKLKOHQLKHHGQARTIIIEINVEDELPAAASVSSVAAAAAVAGGAGVGDSHLALPALERDHLNHHHYVAAAFKAHYGGNPGGCGAKGPKLNSHEIPLFHKSGSKENVETQI

APPENDIX G

GPR124 protein [Homo sapiens] also called ADGRA2

>gi|300934750|ref|NP_116166.9| adhesion G protein-coupled receptor A2 precursor [Homo sapiens]

MGACGRMRARGARPLPLPLLPLLLLAPEARAPCPCLLSIRSCCKCGERPKGLSGVPAPPARRVVCSDGDLPEPSEPFLGPNTFTLLSSNKTIGLRNGSFGLGLSLLEKDLRNNIISTVQPGAFLGGLGELKRLDSNNRIGCLSTEPSQGFLPRLRNLSNISGIFSSQGFVGPDELKTVDGEFLTCDCRLWLLFPAQNRSQLQSEHTLCAYPHQAQLGSLQEAQCLCEGAEALHHTHLISPRQVFQGDRPLFQCSASYLGNDRIRWYHNRFPFGEGDEEQALLLAESLHIDCFTTSBELSLSHIQVWAGSEWECTVSMAGNASKKVETVLETSASYCPAERVNNRDFWNPRTLGTIAYQQSLQFPPFSTVPLGGAPGTRASRKRDCRARGWEPGDYSHCLYTNDITRVLTVFLMPINASHALTLAHQLRFYTAxAASFSMDMVVYVAQMIQKFLGVYDDYQIKELVEVMVMDNASLMLDEHLLWLAQREDKACRIQLASCPMPHHGAPGVQHALSLPHAQHISVNNVALENAYLHPSYVLSCRTFQREGRVGPGGSSPGNPFEPEPEPDQQLRHFCTTGRFNPNSLSSPHKNSVALASIQLPLFLSSLPAALAPPPFPDCDQLQLVFRNRGLFHSHNTSRFGAAPGPKRGRVATPVIAGTSGCCVGNLTFEVPASLRRHWAEAGAEFPVAAWSQEGPGEAGGWTSEGQLRSSQNPVSLAHQHNLGVNVAMLSEAPRFEGGAGAGLHP
VVYPCTALLLCLFATIITYIL
NHSSIRSVRKGWH
MLLNLCFHIAMTSAVFAGGIT
LTNYMQV
CQAVGITLHYSSLSTLLWMGV
KARVHKELTRAPPPQEGDLAGTLPSPMLR
FYLIAGGIPIIICICTAANV
HNRYDHSVCWLVWRLPLG
AFYIPVALILLITWIYFLCAGL
RLRPQAPNQPKAGNSRASLEAGEELRGSTRLSGSGLPLLDSGSOLLATGSAVTPGPPEDGDSLYSPGQVLALVTT
HFLYLYMAWACGALAV
SQWLPR
VVCSCLYGVAASALGLFVFT
HCARRRDVRASWRACCPASPAHAPPAPRALPAALADDGESPVPGEPPSSPGLKPLGALGPGPKCLTLNLQAQSQ
VCEAGAAGGEGEGEPAPTRCRNLARHHPNVHHRARRAKSRAGHAGEAGCKNLKALRLGGAAGALELLSSESGSLL
HNSPTDYSLGSSRNSPGAGLQLEGEMLTPESGDSDTSAAPLSEARGARQRRASRDSLKGGALEKESHRRSYPLNA
ASLNGAPKGKGYDDVTLMGAEVASGGCMKTGWLKSETTV

GenBank: AAI46775.1 (splicing isoform?)

>gi|148922284|gb|AAI46775.1| GPR124 protein [Homo sapiens]
MRGAPARLLPLLPLLLPLLPLLPLLAPAGPCPLSIRCSCKCGERPGLGKLGSGVPGPARRRVCSSGDLPEPP
EGPLLPNSTVTLSSNKITGLRNPSGLGSLKLEKLDRNRLNNISTVQPGAFGLGLGEKLRLDLSNRIGCL
TSCETFQGLPRLRLLISQNEFSLQLQGFVDELPAKVVDDLETFDClCHRLWLLPMQAENQSLQSEHTL
CAYPSLAHAGLQSLQEAQLCECELLEHHHLNQVQVFQDRLPFQCSASYLGNTRHRNYHRAP
VEGDEQAGILLAEIHDCTFITSELTLSHIGWAVSEGCTVSMQAQGSKKEIVVPLTASAQCPAER
VANNRQDFRWRPTLAIGTAQYELCQLQYPTSVLPGGGAPGRASRRCRAWPEPDYGSHLYTDTRV
YTFLVMIPNASNLTLHAQLRVLRTAEEAASFDMDVYVQQMIKFLGYYDVEKELVEMVMDASNLMLV
DEHLWLAQREDKACIRVAGELDRGGAALSAPMAQHISVELSAAPFREVGGAGALHP
VVYPCTALLLCLFATIITYILNHSSIRSVRKGWHMLLNLCFHIAMTSAVFAGGITLTNYQMVCQAVGITLHYSSLS
TLLWMGVKARVHKELTRAPPPQEGDLAGTLPSPMLRFLYIAGGIPIIICICTAANVIIIHNYRDSVCWLVWRLPLG
GAFYIPVALILLITWIYFLCAGLRLRPQAPNQPKAGNSRASLEAGEELRGSTRLSGSGLPLLDSGSOLLATGSAV
PGPPEDGDSLYSPGQVLALVTHFYLAMWACGALAVSQRWRVCSSCLGVAASALGLFVFT
HCARRRDVRASWRACCPASPAHAPPAPRALPAALADDGESPVPGEPPSSPGLKPLGALGPGPKCLTLNLQAQSQ
VCEAGAAGGEGEGEPAPTRCRNLARHHPNVHHRARRAKSRAGHAGEAGCKNLKALRLGGAAGALELLSSESGSLL
HNSPTDYSLGSSRNSPGAGLQLEGEMLTPESGDSDTSAAPLSEARGARQRRASRDSLKGGALEKESHRRSYPLNA
ASLNGAPKGKGYDDVTLMGAEVASGGCMKTGWLKSETTV

GPR125 protein [Homo sapiens] also called ADGRA3
>gi|59823631|ref|NP_660333.2| adhesion G protein-coupled receptor A3 precursor [Homo sapiens]
MEPPGGRRQPPQLPLLPLLALLALLLGGGGGGGGAGCCKGDRPGGARAGAAEGFKVCSSLLEL
AQVLPFPDTLPNRVTTLISRLNKISEKLNGSFSLGLISLRLDLRLNLSLROSSDHGFWSLGLSSLKRLDLTNNR
ICGLNADIFSGQFTLNLRLNLNSGLNFLSSQFTFDYLASLRLEFQTYELLCDCNLMHRWVEKKNITV
DTCRVPYKSLAQQPVTFQGQLLTDCDPLLLEPSYMPSTHRQVPGQVESLPSLFCMAMYSDQMQLVWYQD
GRI vedETDSQGFVEKNNHCSLIALASLTISQAGQSTGNWGHVQKTRKNTRTVDINVLESSAQCP
PERVWWKSKFRWRPTLRAGITAYLQCTRNTNGHSGIYPNGQQDIERAKWARCRDGFGWADDYSRCQYANDV
TRVLVYFMQNPLNLTNAVATQRLALLAYTEVAANFSKDMDVEIFVAEMIEKFRFTEEKSKELGDVMVIA
SNLMADERVVLNPSRAEIKCIVQCSLQRCIRATYLRAGGHAHVVYSTNPIALEAYVIKTSGFTGMTCTVF
QKVAAARDDTGLDSYRGRDPEMLNKLQLSFCVNVTSFFTSLALKNTIVEAISQLPPFLSFPQKRELPTD
DSLYQLKAIFPHNCPLPTGNCWTDGDCLTVPVTILKDIGNVTDTHIPVNTLRLIAHADAV
AARWDFFDLNQGQGWKKDGCCHILYSENITEDQCYSLSNYAMLDTGEYLTQAASSLHP
VVYTTAILLLCLAVISYIY
HHSRISLKSWH
MLVNLCFHIPLTCVVFVGGIT
ISLR/Meflin Immunoglobulin superfamily containing leucine-rich repeat [Homo sapiens]
>gi|83405860|gb|AAI11014.1| Immunoglobulin superfamily containing leucine-rich repeat [Homo sapiens]
MQELHLLWWALLLGLAQACPEPCDCGEKYGFQIADCAYRDLESVPFPFANVTTLSSANRLPGLPEGAFAFPERVPLLQLWAAGALSSLHKLDDLSSLNLIDFSDWLHNLSLQKQLMDSTNFLIPRDAFRSLRALRSLQLNHRNLHTLAEGTFTPTLTALSHLQLINENPFDCTCGIVWLKWTALTTAVISPEQDIACATSPHLKGTFLSLRLLPLPCSAQVQDLYSVPDSQDAGELRGFLALACVHCDVDQPAQPLHWHIQIPSIGVEITSPNVTGDRLGPTVASSQQPRFQAPANGSSLIPDGKLEEGTITYSCLATNELGSAESSVDVALATFEGEGDTRLGRFFHKGAVEGKCGTYTVDSQVPVEDMVNVUITLRSANGPEAAAVEPGQ
LPPGLLLLQSLLLFFFL
TSF

Linx Immunoglobulin superfamily containing leucine-rich repeat 2 [Homo sapiens]
>gi|156230954|gb|AAI52430.1| Immunoglobulin superfamily containing leucine-rich repeat 2 [Homo sapiens]
MFPLRALWLVWALLGVAGSCPEPCACVDKYAHQFADCAKELREVPEGLPANVTTLSSANKITVLRGAGADVTVQTVSLWLAHANEVRTVEPGALAVLSQLKLNDLSKFLHNFISFPWSDLRNLSDLQKLHMKNNRNLGSLPRDALGALPDRLRQINNRRLRTALPQFTDALSALSHQLYHNPFCGCCGLWLQWAATRVSFLPESDICASPPALQGVPVPYRPALPACSAPVHSALAEPEELAPGPTLRAGLAFVHLCHADGHPFTRLQWQLIPFGTVVELPEVLSGLDEGEGGEDDQTAPAPTAPAPAPAPTARFLALANGSSLVLPFLSALAEKAGVYTCRAHNLGANSSTRVAVAAGGPKHAPAGGEPDQAPTSERKSTAKGRNSVLPSKPEGIKQGLAKVSIQGETETEPEETDSTEDEEAEAOQAEQCGNQDPSRYVSNAFEKPACKHPFVLEGVICAFLVAEAREAVQLFLANPQGPAGGAGPRPGRPRLLLYLCPAGGGAAVQWSRVEEGVNAYWFGRLRPGTNYSVCLALAGEACHVQVFSTKKEPSL
LVIVAVVFVFLVTATVPPLAAC
CHLLAHHPGKYPRLRPQAPDMPMPKREAADFDPRASYLESEKYPAGGEAGGEEPEVDVQGEGLEDDEAEQGDPSDLQREESSLACSLVESQSQKANEQEEFAQGEYSDRSLPLGAEAVNIAEINGNYRQTAG

LINGO1 protein [Homo sapiens]
>gi|46250264|gb|AAH68558.1| LINGO1 protein [Homo sapiens]
MLAGGVRSMPSPLACCWQIILLVLSGVSATCCPPYECASAOQDRAVLCLHRKRFVAVPEGIPETRLLDLGKRNKTLQNQDFEASFFEELEELNLNVSAYEPEAFGPFNFLTNRNLGRLNLGKLYNQGFI
KLDISENKLIVLDDYMFQDLYNRLSEVGDNLVISHRASFLNSLEQTLTKECNLSITPEALSMLHLGLVLRRLHLNINAIREDYSFRRKLYRKLVEISWHFLDMMTNFCLNYLNLTSITHCNLATAVFLRHLTYRLRFNLSSNYRIETIESGMLHELRLQIEQLVLQGGQLAVVEFYAFLPNLNLVRVNSGNQTLLEESVFHSGVNLLETILDSNPLACDRCRLLVWFFRRWRLNFNRQPTCATPEFVQKEFKDFPDVLPNYFTCRRARI
LINGO2 protein [Homo sapiens]

>gi|187953591|gb|AAI37515.1| LINGO2 protein [Homo sapiens]
MLHTAISCWQPFLGLAVVLIFMGSTIGCPARCECSAQNKSVSCHHRLLIAIPEGIPIETKILDLSKNRLSKVNPVEFISYPLLEIDSDNIIANVEPGAFTNQLNLSRLRKLQNLKLVPLGVFTGLSNLTKLDISENKIVLIDDYMFDQLHNLKSLKLEVDNLVYISNSRAFSGLLSLEQLTKECNLTAVTEALSHLRISLSLHLKHLNINMPVYAFRFLHLKEIDYFPLLDMPANSLYGLNLTLSVTNTLSSTVPLAFKHVLYTTLNLNSYNPISTIEAGMFSDLIRLQELHIVGQAQLRTEIPEFSGQLRFLRFLNVSQNLLETLEENVFSSPAVELSLINSNPLACDRCWLLWIQRTQFLQFGQQFGGPMGAGPDTIERSKDFSTALSIFYTCCKPKIREKKLQHNLVDGQTQVQLECSADGDQPVISWTTPRRFITTNSKNGRATVLDGTELIRFAQDQDSGMVYVCIAASNAANGFSTSLTVKGFSADRFYАНТПМЯТДСНТСИСНГТНФТСЛДЛКТILVSTAMGCTFTPLGVLVLCFVLLFV
WSRGGKQHKNSIDLEYVPRKNNGAVVEGEVAGPFRNMKI

LINGO3 protein [Homo sapiens]

Genbank: NP_001094861.1

>gi|157426829|ref|NP_001094861.1| leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 3 precursor [Homo sapiens]
MTCWLCVLSLPLLPPPAPPPAPGCCPARCECTVQTRAVACTRRLTATVPGIPATAETRLLELSNRIRICLNPGLAIPAEEELLSNATAIHEPEGAFANLRLVLRLRGNQQLKIPPGVTRDLNLTLLDLSENKLVILLLDYTFQDLHLRRLRLEVDNLVFSRRAFAGLLALEELTLERCNLTSAGSLGHLRSLGALRLHHAIASLEDQNRHLPPGLLLHIDNWPLLEEVAAGSLRGRLNLTSLVTHTNTIVAPAARHRQAHLTCNLSSHNPSTIVPRSFGSRDLVRLRELHLAGALLAVAPEQAPFGLQRILRLNSLNNLSTLEESTHFSVNTLELTVRDGNPLACDCRLMWIVQRKTNDFGRLPACPATAEAVGDALRLNFDSVLFPEYVCRKPKIKERRLQRVTAVAGEVDVRFCLRADGEPAPTAVAVTQHRPYTATSAAGRIALIZVLIRLQOLEVRLRSLQGNRIRMPLACDCRLLWIVQRRKTLNFDGRLPCAPSVAEAVGDALRLNFDSDVLFEYFVCRKPKIKERRLQRVTATAGEVDVRFCLRADGEPAPTAVAVTQHRPYTATSAAGRIALIZVLIRLQ
WSRGGQTHKHKNSIDLEYVPRKNNGAVVEGEVAGPFRNMKI

LINGO4 Leucine rich repeat and Ig domain containing 4 [Homo sapiens]

GenBank: AA17221.1 (NP_001004432.1)

>gi|187953489|gb|AA17221.1| Leucine rich repeat and Ig domain containing 4 [Homo sapiens]
MDAATAPKQAPWPPPLLLPLLPGGSGCIAVCDCTSQFOAVLCHQRLEAVPGGLPDLTEDLDSGNRLWGLQCMSSLLQLSDNLLQNLSTLEPGAFHLQSLRTLRQNLRRIMPGCVFSGLSALTCLUDRNQIVLFDLGAFEGLGQLKEVGDNHLLVFVFAGFAGLAKLSTLTLCRNLSSTVPGLARLARLVALCRLALREDLIGRIRGGEQLEHRLWSEPAGDGPSGVLGLNLSLAITRCNLSVFPQFYLHSLFLVRDDLSNPISAPARRSLPILVRQELRSLGAQSTISAHAFGHLTAFHLLDVAADNLQTLLETAGFPSPDKLVTQRTSPLGCLLQLLRLRRHDFMSPACAPAGPHVOQLSKLEKFSDLPPGHFTCKPALIPSARGWVIAEDEGHAFTFSCCSDGDAPTVSWMRRHFQAWLGARAGRVRLEDGTLEISVQDLRDRGAYCVCVSNA
WSRGGQTHKHKNSIDLEYVPRKNNGAVVEGEVAGPFRNMKI

SALM1 Leucine rich repeat and fibronectin type III domain containing 2 [Homo sapiens]
SALM2 Leucine rich repeat and fibronectin domain containing 1 [Homo sapiens]
Genbank: Q9P244.2

SALM3 Leucine rich repeat and fibronectin domain containing 3 [Homo sapiens]
Genbank: AAH15581.2 (NP_076941)
SALM4 Leucine rich repeat and fibronectin type III domain containing 3 [Homo sapiens]  
GenBank: AAH03578.1 (NP_078785.1)

>gi|13097762|gb|AAH03578.1| Leucine rich repeat and fibronectin type III domain containing 3 [Homo sapiens]  
MAILPLLCLLPLAPASSPPQASATPSCPAPRRCRCQTQSLPSVLCPGAAILLVPPSDDRRAELRADDNF  
IASVRRRDLANMTGLLLSNRNTIRHVARAGFAIDLRLARHLSDGRTLSNLGECGLQLRVLNLRLSN  
NLQALALAAGALDCAETLEDLLSYNNLEQLPWEALRGLGNVLGNDNLLASVPAAGFSRLHLKRLD  
MTSNRRTIPPPDPLFSLPPLAPRPGSPASALVLAFGPNLPHCNCVLLWRLRAIREDDELEACASPPALGG  
RYFVANGEEEVCEEPAPVTHRSPLLAVPGRPAALRCAVEGDPERPVRWSFQGRLLGNNSSARAPFNT  
LELVTNTEPPGGITFCITAIANTEGRATAVELTVGGPQQPQLNSTSCDPSPRDGPDLATPFAASASAKV  
ADTPGPITDTRQVTQTGHATAALQVQPDRQIPGIRMQYIQYNSADDILVYRMIPAESRSSLDLASGR  
TYLDCVLAVYEDSATGTLATTPVCGARFSTPEALRPCGAPAHFP  
LGGTMIIALGGVASVLVFLFVLL  
MRYKVRHGGQPPGAKIPAPVSSVCQSTNGALGPTPTAPPAPAPEALRAHTVVLDCEPWGPHEVPGP

SALM5 Leucine rich repeat and fibronectin type III domain containing 5 [Homo sapiens]  
GenBank: AAH43165.1 (NP_689660.2)

>gi|28175743|gb|AAH43165.1| Leucine rich repeat and fibronectin type III domain containing 5 [Homo sapiens]  
MEKILFYLFLIGIAVKAQICPKRVCQILSPNLATLCAKKGLLFVPPNIDRRTVELRADNFVTNIKRKD  
FAMNTSILVTLFSIFTPHAFPLDNARNLRLNLHNRNRLKTIDMFSGLSNLHNLHNNQLTLLS  
TAPDDVFALEELDSYNNLETPIDWAVVEKMVSLHTLSLDHMDNIPFKTGFSHLHKMTRLDVTSMKQLK  
PPDLPFQAQLATSGISSPSTFAFGPGNPLHCNCWLLRRLSREDDELETASCPLLTRGLRYWSIPEE  
EFLCEPPLTTIRTHMVRCFQGPAPLICITCRANGRDEPAIIHWISPEGKLSNATRSLVYDNGTDILLITTVK  
DTGAFTCISANPGATQIVDLHIIKLPHELLMSTNIHEIPDPGSDISTSTKGSNTSSNGDKLQSDK  
IVVVAEASTTALISKFFNPQINTIPGIRMFQIQYNTYDDTTLVYRMIPPTSTKFLVNNLAAAMTYDLCVLAILY  
DDGGTSLLATRVTVCIGIQFTTQDYVRCHFMQSQFL  
GGTTMIIIGVIVASVLFVIIILMI  
RYKVCNNNQGHKVTKVSVNYSQNTQAIOQCGSVLTPQSVSKQAVGHEENASCCCATDSNVIQSSSETCSSQDSSTTS  
ALPPSWTSTSSTVSQKQRGTKTPSTEPQNEAVTNVESQNTNRNSTALQASPPDSVEGTPSKRAHIKPNALLT  
NVDQIVQETQRELEI

LRIG1 protein [Homo sapiens]  
GenBank: AAH71561.1

>gi|48734697|gb|AAH71561.1| LRIG1 protein [Homo sapiens]  
MARVPRGLGAPPSSCPCLCLLLWLVRLVPETFAAAAGPRACAACACTAGSLDCGRGLAAZIPDLPSWT  
RSLNLSYNKFSEIDPAFPAGFELPNLQEVYLNNELTAVPSLGAASSHHVSLFLQHNIKIRSVFGSQLKAYL  
LEVDLSLNNITEVRNCTUGHPFPIDKELNLRALGGPFDGLSRSSLTRLSKRNITIQLVPAVKFL  
PRLTQLDLNRRNLRLNRLEIGLTEFQGLNLSLEVLKLQRNNISKLTDGAFLWGLSMHVLHLEYNSLVEVNSGLY  
GLTALHQHLHSNSIARIRHKWSFCQKHLHLSFNNLRTDEEASLAELSSLVRLSRLSNHNDISIHEAGA  
PKGKRLSRLVLDHDHNEETTEDSAFGSDLSLKLLELLEPSQAGCSPSQPHSSAGGRTGFKNKSI  
VAKRAFFSLEGELHNLGNAIRSVQFDAFVKMNKLHELISDSDFSLCDQKWLWPILWRLMRQAFVTA  
TCAHPSLEQGQISFPFVCPOQPITQETPMTAMAVGKDIRTCSAA簋inputEmail.com  
LTNADMFHVHQAOGDVEHYTTTLHRVRQTFGEHGYRQCVNTHGFTSYHKARLTVLVPFSTKTPH  
DITiritTMMARLECAATGHPQIAWWKDGGTDFAAQTFTSLVPELDRVSVEGETVALQCKATGNPPFR  
ITWFXGDRPLSLTERHLIPDPQLLLVQQVQVNADEAGRYTCEMSNTLTERAHQVSLVPAACRGKDGT  
VQFTIIVSSVLTSWLWWCII  
YQTRKKSEYSESVTTNTDETVFDVPSLYSGLSQTLSRQDQETQTVRTEGPQPANGHESNVCGPRADSHFEPDTHSVC  
RQPCLKCAGSAYHEKPKAMEKAEGTPGPHMENHGRVRVCSDDNTEVDCYSGRGAFQPQPVMRSDAQPSANPGEPG
LRIG2 Leucine-rich repeats and immunoglobulin-like domains 2 [Homo sapiens]
GenBank: AAI17371.1 (NP_055628)
>gi|109658890|gb|AAI17371.1| Leucine-rich repeats and immunoglobulin-like domains 2 [Homo sapiens]
MAPAPLGVPEQLLCRSLVLTLQQATALLLLPAAAGLCPAPCSRLPLDLCRRKKLAPSRWALS
GLLPPTAILDSNHRSNWSINESLQTLQEVKMUNELTEIPYFEGPTSNITLLLSLVNHPEINAQAL
QFYPALDSLSNIISEIKTSFSFPRMQOLKYLNLSSNRRTIILEAGCFDNLSSSLVVKLNRNRMSP1PPK
IFKLPALQLFELRKRNIKIVEGLTFQGLDSLRSLKMQRNGISKLKDAGFGLNMELELEHNMLTRVNK
GWLGLRLQLQQLVSYOANAIERISPDAWEFCRQSERDLSONQLTLRDESAFVGSLSSLRELRLNGDNRTIVI
ADGVVFLSNQDRLRNLNISWIAEDESAFAGLTSLTQKLIIQGNQIKSTIKAFIGESLEHLDNNN
AIMSIQENAFSQTLHKLINTSLLCDCLKLQLEKLWVSNFQVHSNVSCPAWELAGQSLNVLDKDF
VCDDFQKIQFRTHEPI1ALRGMNVLTLTCTAVSSDPSMSTWKRDEIYVDTSENVYQWAAGEALE
YTSILHFLNVTDEKGQFCVTNHFSNSYQAKLTWNPESFTKMPDLTIRTGAMARLECAAEHPPA
PQISWQKGDTDFPAAERMMHPEVFFIANVIKEDIAGMMICMAQTAGGLASANLTVELEPSFR
PLEDKTVTRGETAVLQCIAGGSPSAPRLNWTKGDGPLVLTERHFFAANQLLIIVDAGLEDAGYCTIMSN
TLGTERHIYLNIPSSCDQSSIHGDDGWTTVG
IVIIVVVCVCGTSLIIWIVI
YHMRKKNEDYSTNTTEELNLPIPDSYLSQOTLSEPEQEGYNSAEGSHQQLMPANGYIHKGTDGGTGRVICS
YDNANISYRTRECYPITYIAEEDVLQDTLSLVSVMQKETYLVHPQDTTALSLELPANREPAFNTNHERI
LKPTQMSGETLQRFPVWNINRELGLPFFPSQPQVHPHQPLONEGRLAGREPDCCASSMSCHRLQDHA
FDRTNRIQDGSEG

LRIG3
Leucine-rich repeats and immunoglobulin-like domains 3 [Homo sapiens]
GenBank: AA216170.1 (NP_700356)
>gi|116496819|gb|AA216170.1| Leucine-rich repeats and immunoglobulin-like domains 3 [Homo sapiens]
MSAPSRAAGLCLLLCAVLGARGDSGGRGSQPSGTAERPCPTTCLGDLLDCSRKLARLPE
PLPSWVARLDSHNRSLSFKASSMSHLQLSREVKLNNNEITPINLGPVSANITLSSLAGNRTVEILPEH
LKEFQSLETDLDDSNQSFAPLQLYLNSVTMSPEGFDNLATTLVKLNRNRIAI
PKMKFLPQEHLEEKRNKNKIKVQDGLQGALKLKMQRNGTVKLMGDFGWSNLMEILQDHLNTEI
GWLGLRLQLHLSQANFIRSDAWFECQRLSEDLDTFNHLSSDLFSGLFLNLTHNHGNNRVS1
ADCAFPRGSLSLTDNLKNREISWTEIDMNGAFSGLKLRLLSQNRINSIITKAFGTDLAREHLDLS
AIMLSQNGAFSMMKQLQLHHLNTSLLCQCKWLPQWAENFQSFVNASCAHPQLKGRSIFAVSDPG
FVCDDFPKQITQVEQPETSQAIKGNSLFIASCAASSDSMTPFASKNDLHDAEMENYAHLRAQGE
YETTTLRLREVFASEGYQCVISNHFGSSYVKAKILTWNMLSPSTKPMDELIRAGMARLECAAVGH
APQIAWQKDDGTDFFAPRAERMVHEVMPEDDFVIVDKIEDIGYSCTAQNSAGISANATLVTLEPSFL
RPLLDRVTQGRTAVLQIAGSPPKLLNWKTDDSDPVVTRHHFAAGNQLLIIVDSDVSDAGYCTEMS
NTGTERGNVRLSVIPTTDSCMTAPSLDDG
ATVGVIIAVVCVCGTSLIIWIVI
YHRRRNEDCSSTNTDETLNLPAIPSYLSQOTLQADDQGYVSSESGSHHQFVTSAGGFFLPQHDSSGCHIDNSS
EADVAAATDLFCPLFGLSTGPMYLKGNVGYDPFETYHTGCSPPDRTVLMDHYEFPYIKKECYP
ECHPSEECERS
FSNISWPVHVRKLNNNTSYSHNEGPGMNKLCLMKSSLDFSANPEAPASVASMSTFMGFKALRPRLDAYSFQPSD
CQPRAFYKLAHHSPDLDDSSGEDKGERTDFOEENHICFTKQTPNHTFQSYDLD
AMIGO1

>GgAmigo1
-CRCCCRRAEKSAPPATTAASFVRAQRHPR--PRRRGAAAP--PRARGGAGHRGAGGQN
GRYKAGGSFPTAAVAVGAPREGPRAQRKVDSPDSVSSVFSDTPIVV
>CmAmigo1
PCHCWCRRKKA--TQQESIHSSSLSTTPTHQAE---AEKEALDRMVFIPARCLGQNGKVQPNFEPDRR----
LSATSRKKSDSESFSTVLLDSPVVV
>HaAmigo1
-CRCCCRGVEKPSSHQQDSLSSLSTTPNHDMPAGDKKDDGFDRVAFLEPGQGQSGKLKPNTLPVEA-------
TGGQQRMSVESVSVFSVDTPIVV
>MmAmigo1
PCCRCRVEKPSSHQQDSLSSLSTTPNHDMPAGDKKDDGFDRVAFLEPGQGQNGKLKPNTLPVEA-------
TGGQQRMSVESVSVFSVDTPIVV

AMIGO2

>CmAmigo2
--RCWCKSKQRHKKPPGNSARSSILSTTPSDV---TERKASTCKVRVFLEPVKEPLKQN
GK1KFQPNNVIAESILKRTTRKSDSNVSFSVIPMS
>HaAmigo2
-CPCCCKTKRKRLPQSSASILSTTPPPNECSAQPSSILSATPCNEDANKTGGKVHLV
GVKLFPSNVIAESILKRTTRKSDSNVSFSVIPMS
>MmAmigo2
-CPCCKCAKRKKNSQSSASILSTTPPPNECSAQPSSILSATP-----GKRVVFLEPLKDTAAGQN
GVKLFPSNVIAESILKRTTRKSDSNVSFSVIPMS

AMIGO3

--------PCRCWCRRTPPPHTPPPNECSAQPSSILSATPCNEDANKTGGKHLV
EPVKSQNGKRLAVSEEFPDVKPSPKQIKLKSITSVFSVDTPTMS
>HaAmigo3
----RCRRACRCRRWPO----TSSPLQELSAQSSVLSTTPDA--PSRKASVHKVVFL EPGRGRLNRQVQAVAAEDF--
LYNPGGLQKAGSEASASSIGSEGMPTT
>MmAmigo3
RGCCCHCQRCRRWPO--------ASSPLQELSAQSSVLSTTPDA--PSRKASVHKVVFL EPGRGRLNRQVQAVAAEDF--
LCNPMLQKAGSEASASSIGSEGLVMS

APPENDIX I NGL Clustal Alignments

NGL1

>CmNGL1
YKMRKQHHHQNHAPRTIEIINVDEEIAARTAGAVSHLTPMPAIEHEDMNHYNYSYKAPFNH
T--VNTINSIHSSVHEPLMRVMVNSKDNVQETQI
>HaNGL1
YKMRKQHHHQNHAPRTIEIINVDEEIAARTAGAVSHLTPMPAIEHELMHNYSYKSPFNH
TTTVTINSIHSSVHEPLMIRNSKDNVQETQI
>MmNGL1
YKMRKQHHHQNHAPRTIEIINVDEEIAARTAGAVSHLTPMPAIEHELMHNYSYKSPFNH
TTTVTINSIHSSVHEPLMIRNSKDNVQETQI
>GgNGL1
YKMRKQHHHQNHAPRTIEIINVDEEIAARTAGAVSHLTPMPAIEHELMHNYSYKSPFNH
TTTVTINSIHSSVHEPLMIRNSKDNVQETQI

NGL2

>HaNGL2
YKLRKQRSTVRSTAARTEIVQVEDIPAAATSAATAAAPTSGVSQEGAVVPLTPHIDHNY
NTYKFAQHAGWHISLSLPTTVTTISEPYIQTHTKVQETQI
>MmNGL2
YKLRKRHQQRSTVTAARTVEIIQVDEDIPAAAPAAATAAPSGVSGEGAVVLPTIHDDHNY
NTYKPAHGAHTENSLGNSLHPTVTTISEPYYIQTHKDKVQETQI
>CmNGL2
YKLRKRHQQRSTVAAARTIEIIINVEEMAGG-------------GPGEVGGGSVPSVHDHMNY NTYKPAHRAHWTDNSLGNLHT--
-TIPEPFIIQTHNKKDKVQETQI

**NGL3**

> CmNGL3
- KL RKQHQLKHGQARTIEIIINVEEDLG---------------------------PTTGDNCALPAVE
HGPNHY-----TAYKHYNNNTSNCTK--NPLHNSVHEPFLKSSKENVQETQI
> HaNGL3
YKLRKQHQLKHGPTRTVEIIINVEDELPAASAVSVAASSGSVGGGDSHLALPALE
RDHLNHVAAAFKHYSNSPNPSGGCGGKGPGLNSIHEPFLKSGEVENVQETQI
> MmNGL3
- KL RKQHQLKHGPTRTVEIIINVEDELPAASAVSVAASSGSVGGGDSHLALPALE RDHLHNYVAAAFKHYSNPSGCGGKGPGLNSIHEPFLKSGEVENVQETQI

GPGEGGGSVPSVHDHMNY NTYKPAHRAHWTDNSLGNLHT--TIPPEPFIQTHNKKDKVQETQI