Evolution and comparative genomics of subcellular specializations: EST sequencing of *Torpedo* electric organ

Javad Nazarian [a], Deborah L. Berry [b], Salar Sanjari [a], Mohammed Razvi [a], Kristy Brown [a], Yetrib Hathout [a], Akos Vertes [c], Sherry Dadgar [a], Eric P. Hoffman [a,*]

[a] *Research Center for Genetic Medicine, Children's National Medical Center, Department of Integrative Systems Biology, The George Washington University School of Medicine & Health Sciences, Washington, DC, United States*
[b] *Histopathology Tissue Shared Resource, Lombardi Cancer Center, Georgetown University, Washington, DC, United States*
[c] *Department of Chemistry, W.M. Keck Institute for Proteomics Technology and Applications, George Washington University, Washington, DC, United States*

ARTICLE INFO

ABSTRACT

Uncharacterized open reading frames (ORFs) in human genomic sequence often show a high degree of evolutionary conservation, yet have little or no tissue EST or protein data suggestive of protein product function. The encoded proteins may have highly restricted expression in specialized cells, subcellular specializations, and/or narrow windows during development. One such highly specialized and minute subcellular compartment is the neuromuscular junction (NMJ), where motorneurons contact muscle fibers. The electric *Torpedo* ray has evolved to expand the NMJ structure to the size of a large organ (electroplax organ), and we hypothesized that *Torpedo* electroplax proteins would be candidates for human ESTs expressed at the human NMJ. A total of 9719 primary electroplax cDNA clones were sequenced. We identified 44 human ORFs showing high (>63%) amino acid identity to *Torpedo* electroplax transcripts with enrichment for mRNA splicing motifs (SH2 and pre-mRNA splicing domains), an observation potentially important for the strict nuclear domains maintained by myonuclei underlying the NMJ. We generated antibodies against two uncharacterized human genes (*C19orf29* [*Drosophila* cactin] and *C15orf24*) and showed that these were indeed expressed at the murine NMJ. Cactin, a member of the Rel transcription factor family in *Drosophila*, localized to the postsynaptic cytosol of the NMJ and nuclear membrane. *C15orf24* protein localized to the murine postsynaptic sarcolemma. We show a novel approach towards identifying proteins expressed at a subcellular specialization using evolutionary diversity of organ function and cross-species mapping.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The availability of complete genomic sequences for a wide array of species has facilitated cross-species comparisons of sequence data, and identification of highly conserved transcript units. For many of the conserved transcript units, EST data supports the expression of the transcript, and a protein has been identified with some function assigned to that protein. However, there remain many highly conserved transcript units where EST support is limited, and there has been no characterization of the predicted protein product. Genomics methods of identifying transcripts have typically used cDNA libraries constructed from whole tissues, whole embryos (during development), or immortalized cell lines. With the emergence of the next-generation sequencing technologies, the possibility of identifying rare sequences is becoming a possibility (Metzker, 2010; Wang et al., 2009). There are many highly specialized subcellular elements of cells, and both the cells harboring the specializations and the specializations may show very low abundance in any tissue or organ, or be expressed at a narrow time window during development. Such cells and subcellular specializations may not be expressed in immortalized cell culture models and/or difficult to detect in whole tissue extracts unless subcellular profiling is employed such as laser microdissection coupled to molecular profiling (Nazarian et al., 2005).

We hypothesized that it would be possible to take advantage of evolutionary diversity of cell and organ function, where a small specialization in humans has been grown to very large size in specific species. Examples of such species-specific augmentation of cells and subcellular specializations include the giant axon of squid (Atchison et al., 1986; Clay, 2005), and the electric organ of certain rays and eels (Miller, 1982; Nazarian et al., 2007; Nitkin et al., 1987; Sealock, 1982). A cross-species genomics approach might enable the identification of components of the species-specific organ, and then the human orthologues identified via cross-species homology comparison.

Myofibers are large syncytial cells. Typically, thousands of myonuclei, each derived from distinct myogenic precursor cells (myoblasts) fuse into

* Corresponding author. Research Center for Genetic Medicine, Children's National Medical Center, 111 Michigan Av. NW, Washington, DC, 20010, United States. Tel.: +1 202 476-6011; fax: +1 202 476 6014.
E-mail address: EHoffman@cnmcresearch.org (E.P. Hoffman).

the large myofiber that is visible to the naked eye. A very small membrane specialization of the myofiber is the NMJ, the point at which the motor neuron innervates the myofiber. The myofiber nuclei immediately underlying the NMJ express a distinct transcriptional program (nuclear domain), including localized expression of the acetylcholine receptor subunit genes and other NMJ-specific proteins. The NMJ is a 30–50 μm localized region of the myofiber plasma membrane, and comprises only a very small fraction (0.1%) of the myofiber membrane surface area (Tsim and Barnard, 2002). Due to the limited size of the NMJ, biochemists turned to the *Torpedo* electroplax, where this species has hypertrophied the NMJ to the size of a kilogram organ. The electroplax organ consists of stacks of hexagonally shaped modified and specialized muscle fibers (each called an electroplaque) which have lost the ability to contract, composed largely of NMJ-like structures (Sealock, 1982; Sheridan, 1965). Instead of functioning as contractile muscle, the electroplax of electric fish has the primary function of generating electrical shock, up to 600 V through water, in response to outside stimuli such as a predator or prey (Kent, 1992). During the 1970s and 1980s, initial studies using the Pacific electric ray (*Torpedo californica)* electric organ led to the first biochemical identification, purification and visualization of the key transmembrane ion channel, the acetylcholine receptor (AChR) (Kistler and Stroud, 1981; Kistler et al., 1982; Mitra et al., 1989). Further molecular components of nerve synapses and neuromuscular junctions such as agrin and acetyl cholinesterase were consequently discovered using *Torpedo* electric organ (Miller, 1982; Nitkin et al., 1987; Sealock, 1982; Carr et al., 1989).

Despite many years of research, only about a dozen NMJ-associated proteins have been identified. The relative small size of the NMJ coupled with the lack of genome/proteome databases for *T. californica* have hindered attempts to identify a more complete picture of the NMJ. For example, based on observed physiological or biochemical cascades, the existence of two proteins/complexes (MASC, and RATL) have been hypothesized, yet the identity of these proteins remains unknown (Apel et al., 1997; Glass et al., 1996; Zhou et al., 1999). Also, the NMJ is a model for nuclear domains and plasma membrane specializations, with distinct mRNA and protein regulation required to maintain a single NMJ per cell, yet the molecular mechanics of establishing and maintaining the nuclear domains are not well understood. Identification of novel molecular constituents of the NMJ should provide insights into synapses, nuclear domains, regulation of membrane specializations, and disorders of these processes in poorly understood neuromuscular conditions such as the spinal muscular atrophy (SMA) and the amyotrophic lateral sclerosis (ALS).

We have recently reported in a pilot study, a partial proteome and transcriptome profile of the *Torpedo* electric organ (Nazarian et al., 2007). In this previous study, we constructed a cDNA library of the *Torpedo* electroplax, sequenced 607 cDNA clones, conducted cross-species homology searches, and generated a custom MS/MS spectral matching database for proteomics identification of proteins (cytoplasmic TOF/TOF MS/MS, and membrane LC-LTQ MS/MS). In this current report, we extend this analysis by sequencing additional 9719 primary cDNA clones. Our ability to identify novel components of the mammalian NMJ was validated by showing that the proteins encoded by two previously uncharacterized human open reading frames (*C19orf29* [*Drosophila* cactin], *C15orf24*) both localize to murine neuromuscular junctions.

2. Materials and methods

2.1. Mouse and Torpedo specimens

Two (4-mo-old C57Bl/10) mice were euthanized by cervical dislocation, and tibialis anterior (TA), intercostal (IC) muscles and spinal cords (SC) were collected. The tissue was immediately frozen in isopentane cooled in liquid nitrogen and stored at −80 °C. Specimens were used for, immunostaining, and western blott assays. All animal

procedures and experiments complied fully with the principles set forth in the "Guide for the Care and Use of Laboratory Animals" prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council, and were approved by the Children's National Medical Center's Institutional Animal Use and Care Committee. *T. californica* specimens were obtained from our collaborator Dr. Khurana (University of Pennsylvania).

2.2. cDNA library construction and sequencing, and custom proteome database

5′-biased cDNA library construction was done using polyA + mRNA from flash frozen *Torpedo* electroplax organ, as previously described (Nazarian et al., 2007). *Torpedo* clones were sequenced from the presumed 5′ end of the transcript/clone using the T3 primer (5′ ATTAACCCTCACTAAAGGGA 3′). Clones were sequenced by High-Throughput Sequencing Solutions (University of Washington High-Throughput Genomics Unit, Seattle, WA). Templates were generated by Rolling Circle Amplification (Templi-Phi—GE Healthcare) and sequenced using Big-Dye chemistry (ABI), with products resolved and read using an ABI 3730xl (50 cm Capillary). Sequencher software (Gene Codes, Ann Arbor, MI) was used to construct contigs of all sequences.

cDNA sequences (in FASTA format) were imported into GPS v3.5 Explorer software, translated into six reading frames, and *in silico* trypsin digested. This database was then searched using the same parameters as above [alkylation of Cys residues ($+57$ Da) and possible Met residues oxidation ($+16$ Da)]. The resulting file was indexed as a local library searchable by peptide mass spectra.

2.3. Antibody production

Affinity-purified polyclonal antibodies were produced against two ORFs, chromosome 15 ORF 24 [(*C15orf24*) human gene ID: 56851, epitope: LPKVVNTSDPDMRRE; *Torpedo* accession number: EH115594], and chromosome 19 ORF 29 (*C19orf29*, human gene ID: 58509, epitope: GTSVDPVEPVEPEE; Torpedo accession number EH115553) the putative human orthologue of *Drosophila* cactin. The peptides were conjugated to KLH and coimmunized into rabbits, with the resulting sera affinity purified using each immobilized peptide separately (Bethyl Laboratories).

2.4. Immunostaining

Mouse tibialis anterior muscles were dissected, snap frozen in isopentane cooled in liquid nitrogen, and sectioned at 8 μm. Sections were blocked with horse serum, and probed with *C19orf24* or cactin primary antibodies (1:1000) that were detected by anti-rabbit Alexa Fluor 594 secondary antibodies (Molecular Probes). Sections were counterstained with anti-bungarotoxin Alexa Fluor 488 (Molecular Probes) to label the neuromuscular junction and with DAPI to label nuclei and imaged on a Zeiss ApoTome (Thornwood, NY) microscope. Controls for specificity of the immunostaining included slides incubated with secondary antibodies alone, primary antibodies mixed with purified antigenic peptide (epitope blocking), and preimmune sera from the same animal (data not shown).

2.5. Immunoblotting

Frozen torpedo electroplax tissue was obtained from Aquatic Research Consultants (San Pedro, CA). Mouse tibialis anterior muscles, gastrocnemius muscles, and spinal cord tissues were dissected and snap frozen on dry ice. All tissues were pulverized in liquid nitrogen, homogenized in PBS with 1% Tween, sonicated, and spun at $800 \times g$ for 15 min to separate the cytosolic and membrane fractions. Protein concentration was determined by BioRad DC Protein Assay and 10 μg

of extracts were analyzed by SDS-PAGE. Proteins were transferred to hybond nitrocellulose (Amersham) and the blots were probed with primary antibodies to cactin and secondary anti-rabbit-HRP antibodies. Signal was detected using the ECL system (Amersham).

2.6. Phylogenic analyses

Alignments were done in Mesquite 2.74 using ClustalW (Huelsenbeck et al., 2001) with gap opening and extension costs set to 8/6. Bayesian analysis for each gene was performed using MrBayes V3.1.2 (Thompson et al., 1994). The Markov chain Monte Carlo was run with four chains for 1 million generations, sampling the Markov chain every 1000 generations, and the sample points of the first 250,000 generations were discarded as "burnin". Other settings in MrBayes were lsetnst = 6 rates = invgamma; unlink statefreq = (all) revmat = (all) shape = (all) pinvar = (all); prsetapplyto = (all) ratepr = variable. The taxon set, with the first one listed interpreted as primary outgroup, in analysis one was *Homo sapiens, T. californica, Mus musculus, Rattus norvegicus,* and *Danio rerio.* The taxon set in the second analysis was as above but with an addition of the primary outgroup *Drosophila melanogaster.*

Table 1
Torpedo electroplax cDNA sequences that show homology to human open reading frames (ORFs).

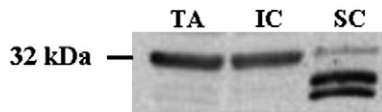| *Torpedo*		Human		
Torpedo cDNA GenBank accession number	No. of cDNA clones	Gene/protein ID official symbol	Amino acid % identity	Function/similar to
RNA processing				
EH115702	1	EAW75176.1; *C20orf14*	98%	PRP6 pre-mRNA processing factor
Hypothetical function				
EW689494	1	CAI14998.1	77%	Novel SH2 domain-containing protein
Contig of (EW692562; EW692829)	2	XP_935194	95%	Endosomal targeting
EW691177	1	EAW60225.1; *C22orf5*	90%	Transmembrane protein 184B
EW690431	1	NP_057039.1; *C2orf4*	94%	Mediator of ErbB2-driven cell motility
EW689749	1	NP_060387.2; *C14orf10*	92%	Serine/threonine-PP2A
Contig of (EW691292; EW692097)	2	EAW61630.1	75%	Isoform CRA_b
Unknown function				Unknown
Contig of (EW697079; EW695860; EW697008; EW696100; EW696827; EW696779; EW695774; EW695604; EW695604; EW696506; EW695652; EW695415; EW696081; EW695704; EW695756; EW695676; EW695943; EW695655; EW697364; EW696424; EH115873; EH115813)	22	AAH02531.1, *C20orf149*	63%	Unknown
Contig of (EW694144; EW694078; EW694237; EW694146;EW694206; EW694024)	6	EAW63210.1; *C8orf40*	82%	Unknown
Contig of (EW693112; EW693254; EW693537; EW693542)	4	EAW77312.1; *C15orf48*	72%	Unknown
Contig of (EW693253; EW693111; EW693495; EW693493)	4	AAI09304.1; *C10orf38*	76%	Unknown
Contig of (EW692974; EW692531; EW692798)	3	NP_077291.1; *C7orf23*	88%	Unknown
Contig of (EW692796; EW692530;EW693070)	3	NP_060932.2; *C3orf10*	100%	Unknown
Contig of (EW691293; EW692020)	2	NP_077270.1; *C6orf106*	92%	Unknown
Contig of (EW691291; EW691704)	2	AAH19351.2; *C4orf31*	89%	Unknown
Contig of (EW692064; EW691290)	2	NP_060366.1; *C20orf11*	97%	Unknown
Contig of (EW691287; EW692192)	2	EAW69978.1; *C1orf57*	71%	Unknown
Contig of (EW691286; EW691724)	2	AAH10908.1; *C1orf123*	90%	Unknown
Contig of (EW691289; EW691888)	2	AAH04818.1; *C14orf129*	81%	Unknown
Contig of (EW691288; EW691858)	2	AAH02750.1; *C11orf10*	100%	Unknown
EW689747	1	AAH02863.1; *C9orf156*	86%	Unknown
EW688678	1	EAW88338.1; *C9orf140*	64%	Unknown
EH115833	1	EAW94148.1; *C7orf25*	91%	Unknown
EH115528	1	AAH11709.1; *C6orf130*	80%	Unknown
EW688651	1	EAW62287.1; *C5orf15*	85%	Unknown
EW690380	1	EAW70869.1; *C2orf33*	73%	Unknown
EW689205	1	EAW59669.1; *C22orf13*	90%	Unknown
EW689145	1	CAM14447.1; *C20orf4*	78%	Unknown
EW689145	1	CAM14447.1; *C20orf4*	78%	Unknown
EW691211	1	CAH73931.1; *C1orf58*	89%	Unknown
EW690281	1	NP_110433.1; *C1orf21*	80%	Unknown
EW689261	1	EAW94893.1; *C1orf151*	90%	Unknown
EH115553	1	NP_001074012.1; *C19orf29*	96%	Unknown
EW689526	1	EAW94413.1; *C17orf71*	73%	Unknown
EH115594	1	NP_064539.1; *C15orf24*	91%	Unknown
EW688992	1	AAH30119.2; *C15orf109*	88%	Unknown
EW690655	1	AAH21701.1; *C14orf147*	95%	Unknown
EW690632	1	EAW80691.1; *C14orf101*	77%	Unknown
EW689791	1	EAX08708.1; *C13orf21*	89%	Unknown
EW691179	1	EAW58230.1; *C12orf44*	94%	Unknown
EW690914	1	AAH02750.1; *C11orf10*	100%	Unknown
EW689274	1	XP_946758.2	76%	Unknown
EW689251	1	XP_001129158.1	100%	Unknown
EW689442	1	NP_067037	94%	Unknown

Fig. 1. Immunoblot of the protein product of *C15orf24*. Shown is the immunoblot analysis of *C15orf24*, with detection of a band of approximately 32 kDa in mouse tibialis anterior (TA), intercostal muscles (IC) and spinal cord (SC).

3. Results

3.1. Sequencing of a cDNA library from T. californica electroplax

We previously described the construction of a *Torpedo* electroplax cDNA library containing 5.16×10^6 primary colony forming units (cfu/ml) in pBluescript II KS+, enriched for 5′ ends of transcripts. In order to expand our transcriptome and proteome database of *Torpedo* electroplax from the 607 clones previously reported (Nazarian et al., 2007), we sequenced an additional 9719 primary non-amplified colonies from this cDNA library. The average sequence length was 710 bp (uni-directional sequencing, average insert size 1.3 ± 0.86 kbp). All sequences were deposited in NCBI GenBank (GenBank sequence IDs EW688567-EW698285). We compiled the new cDNA sequences with the 607 clones sequenced previously (Nazarian et al., 2007), for a total cDNA clone number of 10,326. 28% of clone sequences showed no overlap with other sequences (singletons), although non-overlapping clones of the same subsequence could not be ruled out.

The 10,326 clones assembled into 4243 non-overlapping contigs. However, among singletons we detected non-overlapping clones (e.g. *AChE*) that represented the same gene. These 4243 sequences were searched against non-redundant (NCBInr) database, using first human sequences to define the highest homologies (>75% amino acid identity), then zebrafish, mice and rat for those not showing strong alignment to human. Of these, 1245 sequences showed homology to human.

3.2. Characterization of novel NMJ proteins

Within the 1245 *Torpedo* cDNAs that we found strong support for mapping of a human orthologue, 44 were uncharacterized human ORFs (Table 1). These were candidates for novel protein components of the mammalian NMJ. To test this, we produced affinity-purified polyclonal antisera against two of the ORFs (*C19orf29*; *C15orf24*) that we had previously identified (Nazarian et al., 2007). These two ORFs were selected based upon their particularly high conservation (*C19orf29* 96% identity between human and *Torpedo* over 55 amino acids; *C15orf24* 91% identity over 152 amino acids), and the suitability for antibody production (hydrophilic peptides). These two ORFs were detected by singleton ESTs from *Torpedo*.

C15orf24 is predicted to transcribe a 726 bp mRNA corresponding to a 242 amino acid protein according to human genomic DNA data. Based on the predicted protein sequence, Gene Ontology (GO) annotation suggested that *C15orf24* codes for an extracellular or membrane protein. Antibodies to *C15orf24* were used for immunoblotting assays using mouse tibialis anterior (TA), intercostal (IC) muscle and spinal cord (SC) protein extracts (Fig. 1). A 32 kDa band was identified in TA and IC muscle extracts whereas a less prominent band was present in the spinal cord protein extract, with smaller molecular weight species that represented isoforms, degraded products, or cross-reactive species. The apparent discrepancy in the calculated (26 kDa) and detected (32 kDa) molecular weight could be attributed to post-translational modifications. *C15orf24* is predicted to code for a membrane or secreted protein. Such proteins are often highly post-translationally modified. For example, each glycosylated site could contribute up to 2 kDa of protein mass. As such more investigation is necessary to identify potential post-translational modification.

Alpha-bungarotoxin (α-BTX) binds with high affinity to α-acetylcholine receptor (α-AChR) in the myofiber membrane of the NMJ and was therefore used for co-staining of NMJs in mouse TA muscle. Co-localizations of α-BTX and the ORF proteins provide sub-structure information (plasma membrane vs. other subcellular localization; myofiber vs. motorneuron localization). Immunostaining assays using the *C15orf24* antibody showed specific staining of the murine NMJ (Fig. 2). Co-staining with α-bungarotoxin showed complete co-localization of the *C15orf24* protein product and bungarotoxin (Fig. 2), suggesting that the protein encoded by *C15orf24* localizes to the NMJ at the post-synaptic membrane, in the myofiber plasma membrane. Thus, a singleton electroplax EST with high sequence similarity between the *Torpedo* and human is shown to be a novel component of the myofiber membrane specialization of the NMJ.

For *C19orf29* (NCBI Accession number AY917150), sequence analysis predicts that this gene is the human orthologue of the *Drosophila* cactin
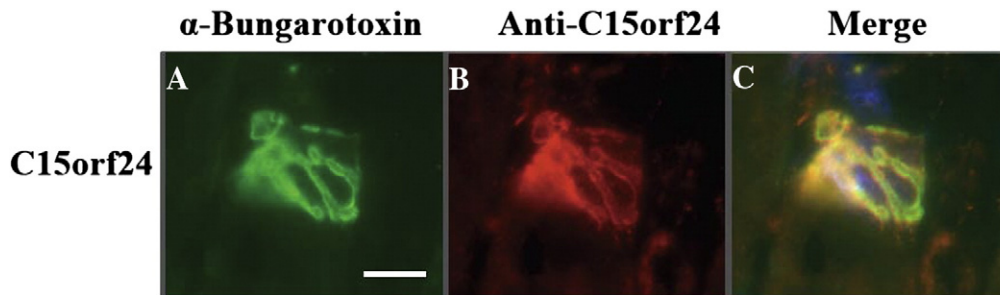

Fig. 2. *C15orf24* antibody detects a protein localized to the murine neuromuscular junction. Affinity purified antibodies against *C14orf24* peptide were tested on longitudinal cryosections from murine tibialis anterior muscle. Fluorescently labeled bungarotoxin labels the post-synaptic (myofiber) membrane (Panel A). Antibodies to *C15orf24* protein showed a pattern very similar to bungarotoxin (Panel B), and double labeling of the same section showed precise co-localization (Panel C). Sections were also stained with Hoechst (blue) to localize subsynaptic myonuclei (arrow in Panel C). Scale bar = 20 μm.
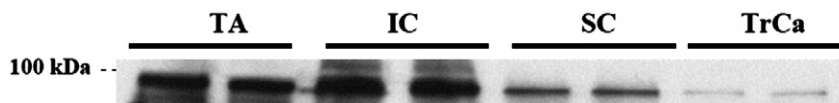

Fig. 3. Immunoblot of the *C19orf29* ORF (cactin). Shown is the immunoblot analysis of murine tibialis anterior (TA), intercostal (IC), spinal cord (SC) muscle extracts and *Torpedo* electroplax (TrCa) using polyclonal antibody against protein encoded by *C19orf19*. The protein product of *C19orf29* is detected at the predicted molecular weight (100 kDa), and is seen enriched in membrane fractions of both the *Torpedo* electroplax, and murine TA and gastrocnemius.

gene, an inhibitor of kappa B (IκB) interacting protein (Lin et al., 2000). The human cactin is predicted to code for a protein with 758 amino acids, and consistent with this, we detected a band of around 100 kDa on a Western blot of mouse muscle (Fig. 3). Immunoblot analysis of the *Torpedo* electroplax tissue, murine muscle (gastrocnemius, tibialis anterior), and spinal cord showed that cactin is detected most strongly in muscle tissues (Fig. 3).

Co-immunostaining of cactin and bungarotoxin in frozen sections of murine muscle showed NMJ-specific localization of cactin (Fig. 4). Control sections tested negative, including secondary antibodies alone, primary antibodies mixed with purified antigenic peptide (epitope blocking), and preimmune sera from the same animal (data not shown). Within the NMJ substructures, cactin was localized to the myofiber cytoplasm immediately underlying the NMJ myofiber membrane, and also to myonuclei underlying the NMJ (subsynaptic nuclei) (Fig. 4). The nuclear staining appeared to be at the periphery of the nuclei, likely at the myonuclear envelope. This data shows that the human *C19orf29* (presumed orthologue of *Drosophila* cactin) is expressed at the NMJ, and shows highly restricted expression to cytoplasm and nuclear envelop only within the NMJ nuclear domain (subcellular specialization).

4. Discussion

We present a cross-species genomics approach where a highly specialized organ, the electric organ of *T. californica,* was used as a means to identify novel components of the mammalian neuromuscular junction. We sequenced a total of 10,326 cDNA clones from the electric organ. We used Blast2Go suit software (Gotz et al., 2008) to analyze generated *T. californica* sequences. Blast2Go integrates functional annotation of DNA or protein sequences based on the Gene Ontology (GO). Our analyzing using Blast2Go showed that the majority of the sequences (~5000) match whether to UniProt/TrEMBL or UniProtKB/Swiss-Prot databases. Species distribution showed that identified sequences have the highest similarity to *H. sapiens, M. musculus,* and *Danio rerio* respectively (Fig. 5A). Among these sequences, top hits belonged to *T. californica, Gallus gallus, Xenopus* and *D. rerio* respectively (Fig. 5B). Since our sequences are already deposited in GenBank we expected to identify *T. californica* as the top hit. However, identifying *G. gallus* as the second best hit was not expected and may indicate a more comprehensive existing *G. gallus*

genome database compared to *Xenopus* and *D. rerio.* We also used Blast2Go to determine the distribution of *T. californica* sequences based on their role in biological processes (Supplemental Table 1). These analyses showed a high association with neuromuscular process, neurological system process, protein binding, and membrane fractions (Supplemental Table). We inspected the 10,326 cDNA sequences by homology search for human ORFs of unknown function. We found that at least 44 transcripts showed a significant (>63%) degree of amino acid identity to uncharacterized human ORFs (Table 1). Although the main function of most of these proteins is unknown (i.e. contain conserved domains of unknown function), some of these transcripts contain domains involved in RNA processing, chromatin binding, or are homologous to Src homology 2 (SH2) domain (Table 1). Detection of these proteins emphasizes the importance of molecular machinery that is important for the establishment and maintenance of myonuclear domains. Furthermore bi-directional exchange of molecular information between muscle and nerve is important for stability of the NMJ (Rezaee et al., 1993). Muscle derived growth factors for example influence motorneuron survival where trophic factors such as agrin influence specific gene expression in subsynaptic myonuclei. Agrin, initially identified in *Torpedo* electric organ, is one of the main factors involved in initiating the clustering of postsynaptic scaffolding via MuSK signaling (Nitkin et al., 1987; Godfrey et al., 1984). Despite the fact that agrin is present in the adult synaptic regions in both nerve and muscle, it is expressed throughout the basal lamina during myofiber development (Sanes and Lichtman, 1999). However, it was later discovered that adult muscle fibers (and not motorneurons), express alternatively spliced agrin which lacks the C-terminal domain important for AChR clustering (Ferns and Hall, 1992; Ferns et al., 1992; Ruegg et al., 1992; Tsim et al., 1992). More importantly, tyrosin phosphatases such as SH2 are important for facilitating post-synaptic agrin/MuSK signaling and more importantly AChR clustering at the NMJ (Madhavan et al., 2005; Zhao et al., 2007; Zhao and Zhang, 2006).

Alternative splicing of α-dystrobrevin (DB) is a valid example of subsynaptic myonuclear domain specialization. DB exists in multiple isoforms including DB1 and DB2 (Balasubramanian et al., 1998; Wagner et al., 1993). DB1, strictly expressed at NMJs and myotendinous junctions (MTJs), has an extended C-terminal domain, which is tyrosine phosphorylated, where as DB2 lacks this domain and is
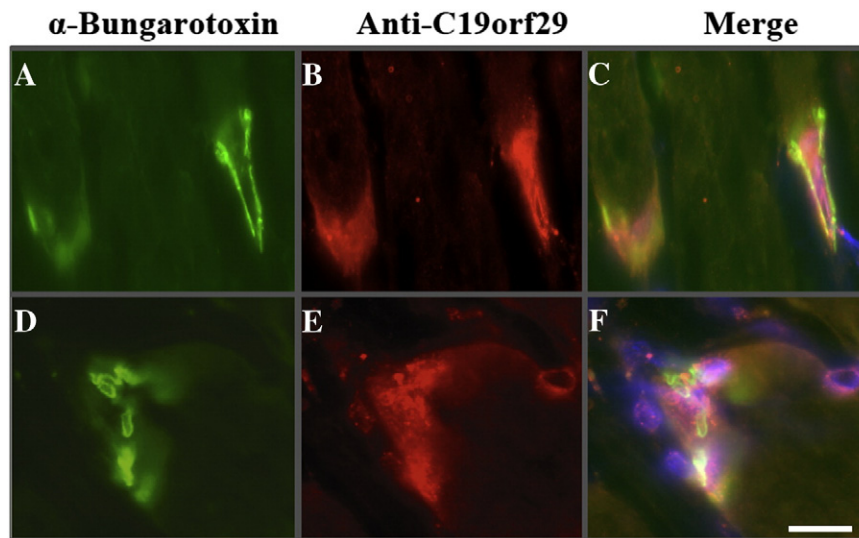

Fig. 4. Immunostaining of the frozen sections of murine tibialis anterior with antibodies to the protein encoded by *C19orf29* showed localization to the NMJ. Co-localization with bungarotoxin (Panels A, and C) showed different localization for *C19orf29* and bugarotoxin (Panel C). Hoechst staining (blue) was also done to localize subsynaptic myonuclei (Panel F, arrow). *C19orf29* was localized to the myofiber cytoplasm adjacent the NMJ (Panel B), and also the myonuclei immediately underlying the NMJ (Panel E, and F), with apparent nuclear envelope localization (arrowhead Panel E, and F). Scale bar = 20 μm.
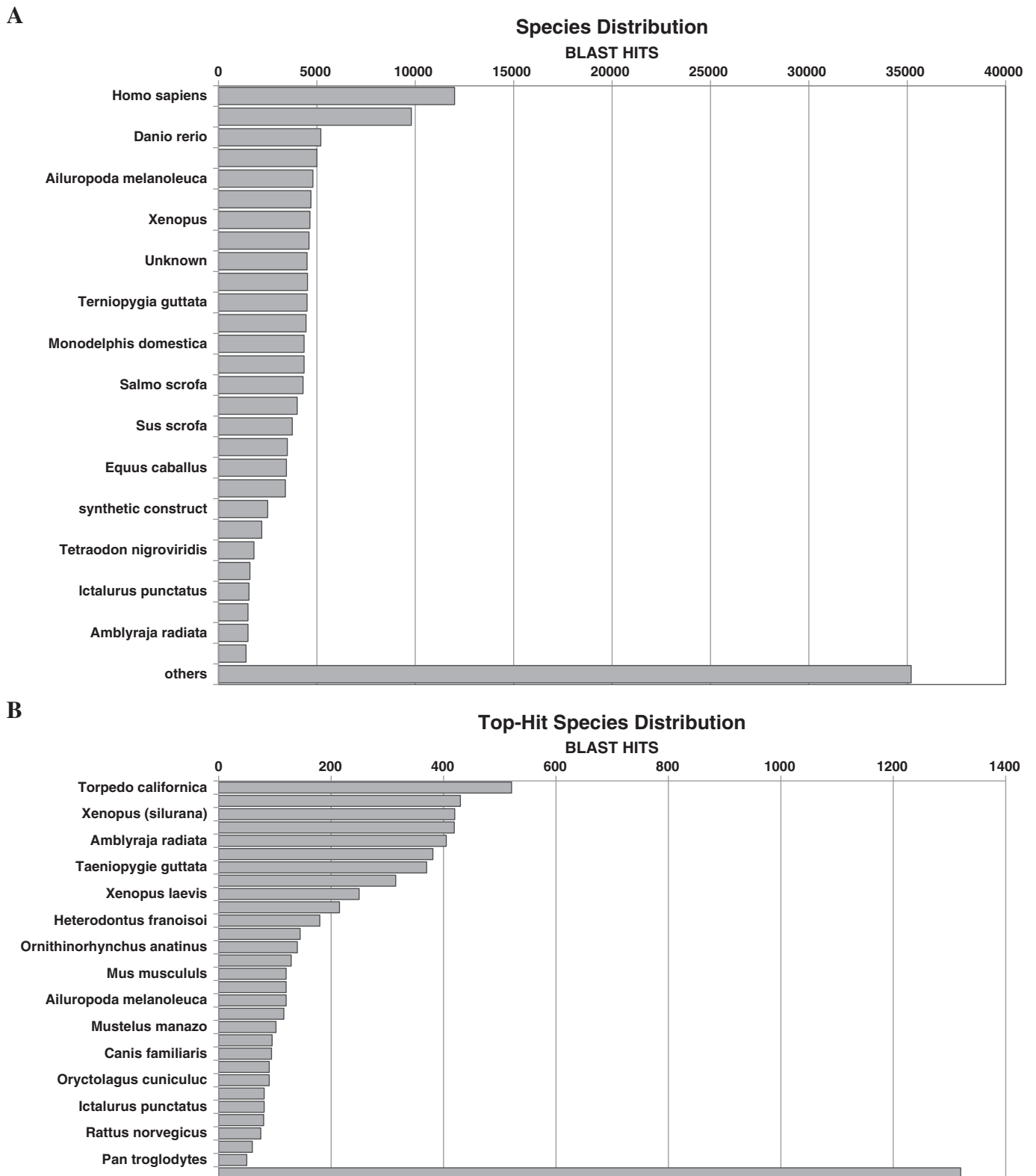
A

Species Distribution
BLAST HITS

B

Top-Hit Species Distribution
BLAST HITS

Fig. 5. Distribution of *Torpedo californica* sequences based on their percent identity when compared to existing databases. A) *Torpedo californica* sequences show a high similarity to human sequences, however, when accounted for percent identity B) *Torpedo* sequences are more similar to other species including the puffer fish *(Danio rerio)*. *Torpedo californica* appears in on the top-hit species because all of our sequences are already deposited in GenBank.

therefore expressed throughout the muscle fiber including NMJ and MTJ. Together, our findings provide suitable candidate human ORFs for future analysis towards identification of novel components of the NMJ.

One of the most interesting contigs of cDNA sequences is one comprised of 22 clones (Table 1). The translated contig codes for 111 amino acids which has a 63% sequence identity (total of 71 amino acids) to translated human chromosome 20 ORF 149 (*C20orf149*). The

high cDNA clone number may correlate with the degree of importance of this transcript in function and maintenance of the electric organ suggesting it is an ideal potential NMJ candidate.

Two ORFs were selected for antibody production based on their cDNA library source, computed subcellular localization and presence of functional domains. *C15orf24* contains a domain of unknown function (DUF 2012) which belongs to the eukaryotic family of uncharacterized proteins. Our immunohistochemistry experiments showed that the protein coded by *C15orf24* (*C15orf24_P*) localized to mouse NMJs (Fig. 2). *C15orf24_P* localized to the postsynaptic membrane following a close proximity to acetylcholine receptors (AChRs). Further investigations are important to define the molecular functions of this protein at the NMJ and other electrical synapses.

We have previously reported sequencing *T. californica* cDNA clone homologs to a human open reading frame on chromosome 19 (*C19orf29*) (Nazarian et al., 2007), which is homologous to the *Drosophila* cactin protein. Fly cactin was originally identified by yeast two hybrid system as a coil–coil protein with N-terminal Arg-Asp (RD)-like domain that directly interacts with the fly cactus protein (Lin et al., 2000). In flies, cactin functions with the Rel family of transcription factors which are involved in vertebrate and invertebrate embryonic development and dorsal–ventral embryonic polarity (Lin et al., 2000; Belvin and Anderson, 1996; Rusch and Levine, 1996). Nuclear factor kappaB (NF-κB) is a member of the Rel family which is important for neuronal synaptic plasticity (Meffert et al., 2003). In Drosophila, three genes, dorsal (dl), dorsal related immunity factor (Dif) and relish (Rel), encode NF-kB/Rel proteins (Ayyar et al., 2007; Moussian and Roth, 2005; Royet et al., 2005). Dif and dl are sequestered in the cytoplasm by dimerization with cactus. Upon phosphorylation and degradation of cactus, Dif and dl are released and translocated to the nucleus (Bate, 1990). To test the subcellular localization of *C19orf29* (human cactin), we developed polyclonal antibodies against this protein and detected its localization to postsynaptic scaffolding as well as nuclear membrane of both subsynaptic and non-subsynaptic myonuclei (Fig. 4). Our results, while adding cactin to the list, reinstate previous studies describing synaptic activities of Rel family proteins (Beramendi et al., 2005; Heckscher et al., 2007). Our observation that cactin is localized to the subsynaptic cytoplasm suggests that it is regulated by synaptic electrical activity. Indeed, the role of electrical impulses and their role in differential gene expression program of subsynaptic myonuclei is well studied (Sanes and Lichtman, 1999;

Fig. 6. The Bayesian tree for each gene with branch lengths proportional to inferred divergence times. A) *C15orf24* exhibits substantial divergence that is shared between *Torpedo californica* and *Danio rerio*, with an additional evolutionary burst in the latter. B) *C19orf29* on the other hand, shows extremely long and islolated divergence in *Torpedo californica*, which is not shared with any other taxon in this analysis.

Sanes and Lichtman, 2001). We conducted phylogenic Bayesian study of both *C15orf24* and *C19orf24* with branch lengths proportional to inferred divergence times (Fig. 6). *C15orf24* sequence was compared to *H. sapiens, D. rerio, M. musculus, Rattus norvegicus* and *D. melanogaster* orthologues. We found that *C15ORF24* showed a substantial divergence that is shared between *T. californica* and *D. rerio* (puffer fish), with additional evolutionary burst in the latter (Fig. 5A). We analyzed *C19orf29* with sequences available from *H. sapiens, D. rerio, M. musculus, R. norvegicus* and *D. melanogaster*. *C19orf29* showed extremely long and isolated divergence in *T. californica*, which is not shared with any other taxon in the analysis (Fig. 5B). More studies are required to reveal the specific role of these two ORFs at the NMJ and neuronal synapses.

This study, to our best knowledge, provides the largest number of molecules potential for synapse formation and stability. Furthermore, we show the NMJ association of two ORF-coded proteins to murine neuromuscular junctions. Future studies should focus first on localizing potential targets introduced in this study and further characterizing their specific function at the neuromuscular junction.

5. Accession numbers

The 9719 *Torpedo* electroplax cDNA sequences have been submitted to NCBI under accession numbers: *EW688567-EW698285*.

Supplementary materials related to this article can be found online at doi:10.1016/j.margen.2010.12.004.

References

Apel, E.D., Glass, D.J., Moscoso, L.M., Yancopoulos, G.D., Sanes, J.R., 1997. Rapsyn is required for MuSK signaling and recruits synaptic components to a MuSK-containing scaffold. Neuron 18 (4), 623–635.

Atchison, W.D., Luke, V.S., Narahashi, T., Vogel, S.M., 1986. Nerve membrane sodium channels as the target site of brevetoxins at neuromuscular junctions. Br. J. Pharmacol. 89 (4), 731–738.

Ayyar, S., Pistillo, D., Calleja, M., Brookfield, A., Gittins, K., Goldstone, C., Simpson, P., 2007. NF-kappaB/Rel-mediated regulation of the neural fate in *Drosophila*. PLoS ONE 2 (11), e1178.

Balasubramanian, S., Fung, E.T., Huganir, R.L., 1998. Characterization of the tyrosine phosphorylation and distribution of dystrobrevin isoforms. FEBS Lett. 432 (3), 133–140.

Bate, M., 1990. The embryonic development of larval muscles in *Drosophila*. Development 110 (3), 791–804.

Belvin, M.P., Anderson, K.V., 1996. A conserved signaling pathway: the Drosophila toll-dorsal pathway. Annu. Rev. Cell Dev. Biol. 12, 393–416.

Beramendi, A., Peron, S., Megighian, A., Reggiani, C., Cantera, R., 2005. The inhibitor kappaB-ortholog cactus is necessary for normal neuromuscular function in *Drosophila melanogaster*. Neuroscience 134 (2), 397–406.

Carr, C., Fischbach, G.D., Cohen, J.B., 1989. A novel 87, 000-Mr protein associated with acetylcholine receptors in *Torpedo* electric organ and vertebrate skeletal muscle. J. Cell Biol. 109 (4 Pt 1), 1753–1764.

Clay, J.R., 2005. Axonal excitability revisited. Prog. Biophys. Mol. Biol. 88 (1), 59–90.

Ferns, M.J., Hall, Z.W., 1992. How many agrins does it take to make a synapse? Cell 70 (1), 1–3.

Ferns, M., Hoch, W., Campanelli, J.T., Rupp, F., Hall, Z.W., Scheller, R.H., 1992. RNA splicing regulates agrin-mediated acetylcholine receptor clustering activity on cultured myotubes. Neuron 8 (6), 1079–1086.

Glass, D.J., Bowen, D.C., Stitt, T.N., Radziejewski, C., Bruno, J., Ryan, T.E., Gies, D.R., Shah, S., Mattsson, K., Burden, S.J., DiStefano, P.S., Valenzuela, D.M., DeChiara, T.M., Yancopoulos, G.D., 1996. Agrin acts via a MuSK receptor complex. Cell 85 (4), 513–523.

Godfrey, E.W., Nitkin, R.M., Wallace, B.G., Rubin, L.L., McMahan, U.J., 1984. Components of Torpedo electric organ and muscle that cause aggregation of acetylcholine receptors on cultured muscle cells. J. Cell Biol. 99 (2), 615–627.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36 (10), 3420–3435.

Heckscher, E.S., Fetter, R.D., Marek, K.W., Albin, S.D., Davis, G.W., 2007. NF-kappaB, IkappaB, and IRAK control glutamate receptor density at the *Drosophila* NMJ. Neuron 55 (6), 859–873.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294 (5550), 2310–2314.

Kent, G.C., 1992. Comparative Anatomy of the Vertebrates. Mosby-Year Book, St. Louis, MO.

Kistler, J., Stroud, R.M., 1981. Crystalline arrays of membrane-bound acetylcholine receptor. Proc. Natl Acad. Sci. USA 78 (6), 3678–3682.

Kistler, J., Stroud, R.M., Klymkowsky, M.W., Lalancette, R.A., Fairclough, R.H., 1982. Structure and function of an acetylcholine receptor. Biophys. J. 37 (1), 371–383.

Lin, P., Huang, L.H., Steward, R., 2000. Cactin, a conserved protein that interacts with the *Drosophila* IkappaB protein cactus and modulates its function. Mech. Dev. 94 (1–2), 57–65.

Madhavan, R., Zhao, X.T., Ruegg, M.A., Peng, H.B., 2005. Tyrosine phosphatase regulation of MuSK-dependent acetylcholine receptor clustering. Mol. Cell. Neurosci. 28 (3), 403–416.

Meffert, M.K., Chang, J.M., Wiltgen, B.J., Fanselow, M.S., Baltimore, D., 2003. NF-kappa B functions in synaptic signaling and behavior. Nat. Neurosci. 6 (10), 1072–1078.

Metzker, M.L., 2010. Sequencing technologies—the next generation. Nat Rev Genet. 11(1): p. 31–46.

Miller, C., 1982. Open-state substructure of single chloride channels from *Torpedo* electroplax. Philos. Trans. R. Soc. Lond. B Biol. Sci. 299 (1097), 401–411.

Mitra, A.K., McCarthy, M.P., Stroud, R.M., 1989. Three-dimensional structure of the nicotinic acetylcholine receptor and location of the major associated 43-kD cytoskeletal protein, determined at 22 A by low dose electron microscopy and x-ray diffraction to 12.5 A. J. Cell Biol. 109 (2), 755–774.

Moussian, B., Roth, S., 2005. Dorsoventral axis formation in the *Drosophila* embryo—shaping and transducing a morphogen gradient. Curr. Biol. 15 (21), R887–R899.

Nazarian, J., Bouri, K., Hoffman, E.P., 2005. Intracellular expression profiling by laser capture microdissection: three novel components of the neuromuscular junction. Physiol. Genomics 21 (1), 70–80.

Nazarian, J., Hathout, Y., Vertes, A., Hoffman, E.P., 2007. The proteome survey of an electricity-generating organ (*Torpedo californica* electric organ). Proteomics 7 (4), 617–627.

Nitkin, R.M., Smith, M.A., Magill, C., Fallon, J.R., Yao, Y.M., Wallace, B.G., McMahan, U.J., 1987. Identification of agrin, a synaptic organizing protein from *Torpedo* electric organ. J. Cell Biol. 105 (6 Pt 1), 2471–2478.

Rezaee, M., Isokawa, K., Halligan, N., Markwald, R.R., Krug, E.L., 1993. Identification of an extracellular 130-kDa protein involved in early cardiac morphogenesis. J. Biol. Chem. 268 (19), 14404–14411.

Royet, J., Reichhart, J.M., Hoffmann, J.A., 2005. Sensing and signaling during infection in *Drosophila*. Curr. Opin. Immunol. 17 (1), 11–17.

Ruegg, M.A., Tsim, K.W., Horton, S.E., Kroger, S., Escher, G., Gensch, E.M., McMahan, U.J., 1992. The agrin gene codes for a family of basal lamina proteins that differ in function and distribution. Neuron 8 (4), 691–699.

Rusch, J., Levine, M., 1996. Threshold responses to the dorsal regulatory gradient and the subdivision of primary tissue territories in the *Drosophila* embryo. Curr. Opin. Genet. Dev. 6 (4), 416–423.

Sanes, J.R., Lichtman, J.W., 1999. Development of the vertebrate neuromuscular junction. Annu. Rev. Neurosci. 22, 389–442.

Sanes, J.R., Lichtman, J.W., 2001. Induction, assembly, maturation and maintenance of a postsynaptic apparatus. Nat. Rev. Neurosci. 2 (11), 791–805.

Sealock, R., 1982. Visualization at the mouse neuromuscular junction of a submembrane structure in common with *Torpedo* postsynaptic membranes. J. Neurosci. 2 (7), 918–923.

Sheridan, M.N., 1965. The fine structure of the electric organ of *Torpedo marmorata*. J. Cell Biol. 24, 129–141.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22 (22), 4673–4680.

Tsim, K.W., Barnard, E.A., 2002. The signaling pathways mediated by P2Y nucleotide receptors in the formation and maintenance of the skeletal neuromuscular junction. Neurosignals 11 (1), 58–64.

Tsim, K.W., Ruegg, M.A., Escher, G., Kroger, S., McMahan, U.J., 1992. cDNA that encodes active agrin. Neuron 8 (4), 677–689.

Wagner, K.R., Cohen, J.B., Huganir, R.L., 1993. The 87 K postsynaptic membrane protein from *Torpedo* is a protein-tyrosine kinase substrate homologous to dystrophin. Neuron 10 (3), 511–522.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10 (1), 57–63.

Zhao, X.T., Zhang, Z., 2006. The role of protein tyrosine phosphatases Shp-2 involved in the formation of the neuromuscular junction. Zhonghua Yi Xue Za Zhi 86 (15), 1052–1056.

Zhao, X.T., Qian, Y.K., Chan, A.W., Madhavan, R., Peng, H.B., 2007. Regulation of ACh receptor clustering by the tyrosine phosphatase Shp2. Dev. Neurobiol. 67 (13), 1789–1801.

Zhou, H., Glass, D.J., Yancopoulos, G.D., Sanes, J.R., 1999. Distinct domains of MuSK mediate its abilities to induce and to associate with postsynaptic specializations. J. Cell Biol. 146 (5), 1133–1146.

Supplemental Table 1. Distribution of *Torpedo californica* sequences based on their role in biological processes. Blast2Go was used generate relationships based on gene ontology (GO) information.

GO Term	Score	Parents
neuromuscular process	1.00	neurological system process
myofibril assembly	1.00	cellular component assembly involved in morphogenesis, actomyosin structure organization, striated muscle cell development
skeletal muscle tissue growth	1.00	skeletal muscle tissue development, developmental growth
muscle contraction	1.00	muscle system process
regulation of membrane potential	1.00	cellular ion homeostasis
signal transduction	1.00	regulation of cellular process, signal transmission
cation transport	1.00	ion transport
musculoskeletal movement	1.00	multicellular organismal movement
multicellular organismal process	0.82	biological_process
biological_process	0.83	
system process	0.72	multicellular organismal process
cellular process	0.62	biological_process
neurological system process	0.60	system process
skeletal muscle tissue development	0.60	skeletal muscle organ development, striated muscle tissue development
developmental growth	0.60	growth, developmental process
striated muscle cell development	0.60	striated muscle cell differentiation, muscle cell development
muscle system process	0.60	system process
regulation of cellular process	0.60	regulation of biological process, cellular process
ion transport	0.60	transport
cellular component assembly involved in morphogenesis	0.60	anatomical structure formation involved in morphogenesis, cellular component assembly, cellular component morphogenesis
signal transmission	0.60	signaling process
cellular ion homeostasis	0.60	ion homeostasis, cellular chemical homeostasis
multicellular organismal movement	0.60	multicellular organismal process
actomyosin structure organization	0.60	actin cytoskeleton organization
striated muscle tissue development	0.36	muscle tissue development
anatomical structure formation involved in morphogenesis	0.36	anatomical structure morphogenesis, developmental process
skeletal muscle organ development	0.36	muscle organ development
ion homeostasis	0.36	chemical homeostasis
transport	0.36	establishment of localization
cellular component assembly	0.36	cellular component biogenesis, cellular component organization
cellular component morphogenesis	0.36	cellular developmental process, anatomical structure morphogenesis, cellular component organization

striated muscle cell differentiation	0.36	muscle cell differentiation
muscle cell development	0.36	cell development, muscle cell differentiation
growth	0.36	biological_process
cellular chemical homeostasis	0.36	chemical homeostasis, cellular homeostasis
actin cytoskeleton organization	0.36	cytoskeleton organization, actin filament-based process
regulation of biological process	0.36	biological regulation, biological_process
signaling process	0.36	biological_process, signaling
biological regulation	0.26	biological_process
muscle structure development	0.26	anatomical structure development
chemical homeostasis	0.22	homeostatic process
establishment of localization	0.22	biological_process, localization
muscle organ development	0.22	organ development, muscle structure development
cell development	0.22	cellular developmental process, cell differentiation
muscle tissue development	0.22	tissue development, muscle organ development
anatomical structure morphogenesis	0.22	anatomical structure development, developmental process
actin filament-based process	0.22	cellular process
muscle cell differentiation	0.22	cell differentiation, muscle structure development
cytoskeleton organization	0.22	organelle organization
cellular homeostasis	0.22	homeostatic process, cellular process
cellular component biogenesis	0.22	biological_process
cellular component organization	0.22	biological_process
signaling	0.22	biological_process
cellular developmental process	0.22	developmental process, cellular process
anatomical structure development	0.18	developmental process
homeostatic process	0.13	regulation of biological quality
cell differentiation	0.13	cellular developmental process
tissue development	0.13	organ development, anatomical structure development
localization	0.13	biological_process
organ development	0.13	system development, anatomical structure development
organelle organization	0.13	cellular component organization, cellular process
developmental process	0.11	biological_process
regulation of biological quality	0.08	biological regulation
system development	0.08	multicellular organismal development, anatomical structure development
multicellular organismal development	0.05	multicellular organismal process, developmental process