

# Ancestral centriole and flagella proteins identified by analysis of *Naegleria* differentiation

Lillian K. Fritz-Laylin and W. Zacheus Cande\*

Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

\*Author for correspondence (zcande@berkeley.edu)

Accepted 25 August 2010

Journal of Cell Science 123, 4024-4031

© 2010. Published by The Company of Biologists Ltd

doi:10.1242/jcs.077453

## Summary

*Naegleria gruberi* is a single-celled eukaryote best known for its remarkable ability to form an entire microtubule cytoskeleton de novo during its metamorphosis from an amoeba into a flagellate, including basal bodies (equivalent to centrioles), flagella and a cytoplasmic microtubule array. Our publicly available full-genome transcriptional analysis, performed at 20-minute intervals throughout *Naegleria* differentiation, reveals vast transcriptional changes, including the differential expression of genes involved in metabolism, signaling and the stress response. Cluster analysis of the transcriptional profiles of predicted cytoskeletal genes reveals a set of 55 genes enriched in centriole components (induced early) and a set of 82 genes enriched in flagella proteins (induced late). The early set includes genes encoding nearly every known conserved centriole component, as well as eight previously uncharacterized, highly conserved genes. The human orthologs of at least five genes localize to the centrosomes of human cells, one of which (here named Friggin) localizes specifically to mother centrioles.

**Key words:** *Naegleria*, Assembly, Centriole, Evolution, Flagella

## Introduction

*Naegleria gruberi* grows as an amoeba without flagella, centrioles or even cytoplasmic microtubules; it relies on an actin-based cytoplasmic cytoskeleton for chemotaxis and motility, and its mitotic spindle is contained within an intact nuclear envelope (Fulton, 1970; Fulton, 1977; Fulton and Dingle, 1971). However, when exposed to stressors such as changes in temperature or nutrient availability, *Naegleria* rapidly differentiates into a flagellate, forming a complete cytoplasmic microtubule cytoskeleton from scratch (Fig. 1) (Fulton and Dingle, 1967). This differentiation occurs synchronously – approximately 90% of cells assemble basal bodies (structures equivalent to centrioles) within a 15-minute window, followed by flagella approximately 10 minutes later (Fig. 1) (Fulton and Dingle, 1971). Although *Naegleria* assembles basal bodies de novo, protein incorporation occurs in the same order as that occurring during assembly of human centrioles (Fritz-Laylin et al., 2010a). The evolutionary distance of *Naegleria* from animals means that genes shared between *Naegleria* and humans were probably present in the ancestor of all eukaryotes (Cavalier-Smith, 2002; Ciccarelli et al., 2006) (for a review, see Fritz-Laylin et al., 2010b). Thus, *Naegleria* differentiation affords a unique opportunity to study ancestral features of centriole and flagellum assembly.

Interphase animal cells contain numerous microtubules emanating from microtubule organizing centers (MTOCs) called centrosomes. Centrosomes contain centrioles that are primarily composed of nine microtubule triplets, and the surrounding amorphous pericentriolar material (PCM) that anchors cytoplasmic microtubules. Centrioles are called basal bodies when they are used to organize axonemes, the microtubule core of eukaryotic cilia and flagella. These whip-like structures propel single-celled organisms and move fluids within multicellular organisms. Metazoan cells also have nonmotile cilia that function as ‘cellular antennae’ by gathering information about the surrounding

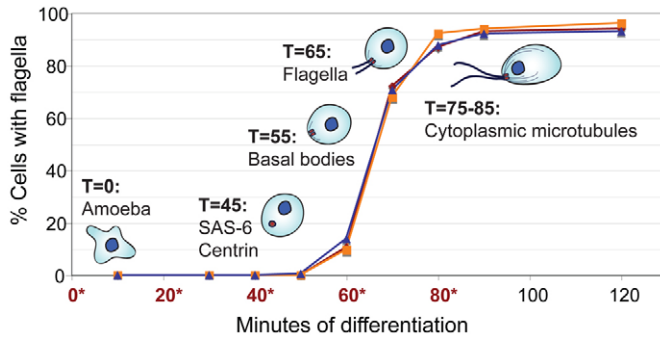
environment using their varied signaling receptors (Marshall and Nonaka, 2006).

Proteomic analyses indicate that centrosomes and basal bodies contain many of the same proteins, a large number of which are thought to be functional components of centrioles (Andersen et al., 2003; Keller et al., 2005; Kilburn et al., 2007). However, only a handful of these proteins have been characterized functionally (Strnad and Gönczy, 2008). This is in part due to the technical difficulties associated with studying centriole assembly in most organisms. First, new centrioles usually assemble in association with a mature centriole, hindering proteomic characterization of assembly intermediates. Second, centriole assembly is usually tied to the cell cycle, rendering it difficult to distinguish centriole-specific genes from other induced cell cycle genes. And finally, de novo assembly (where centrioles are built in the absence of preexisting ones) usually occurs in a single cell or embryo, making proteomic or microarray-based approaches unfeasible. Here, we used the synchronous de novo basal body assembly pathway of *Naegleria* to overcome these technical roadblocks and identify genes used specifically for building basal bodies and flagella.

## Results and Discussion

### Flagella and basal body gene transcripts are induced with different kinetics

We isolated total RNA at 20-minute intervals during *Naegleria* differentiation (at 0, 20, 40, 60 and 80 minutes; Fig. 1) from three biological replicates (supplementary material Fig. S1A). The relative abundance of transcripts from each time-point was quantified using custom full-genome *Naegleria* DNA microarrays. Approximately 24% of *Naegleria* genes are induced at least twofold, and an additional 39% are reduced by at least 50% during the amoeba-to-flagellate transition (4065 and 6484 genes, respectively;  $P < 0.01$ , after correction for multiple testing). Differentially regulated genes include those involved in stress



**Fig. 1. Selected events during differentiation of *Naegleria*.** Each of the three curves represents the percentage of cells with visible flagella during one differentiation replicate used for microarray analysis; the time-points collected are indicated with an asterisk (\*). Important events during assembly of the *Naegleria* basal body are indicated.

responses (including Hsp20 and Hsp 90; data not shown) and core metabolism (including glycolysis, Krebs cycle and pyruvate–acetate metabolism; data not shown), as well as cytoskeletal components.

Only a fraction of the thousands of induced genes are likely to be microtubule related. To aid in our search for uncharacterized and evolutionarily conserved centriole proteins, we focused on genes found in *Naegleria* and other flagellates but missing in non-flagellated organisms [the flagellar motility (FM) gene set (Fritz-Laylin et al., 2010b)]. FM members include genes specific to basal bodies and flagella but exclude genes such as that encoding  $\alpha$ -tubulin that are also used by organisms without flagella. To permit analysis of general microtubule proteins involved in basal body and flagella formation, we added *Naegleria* homologs of known microtubule cytoskeleton proteins (Fritz-Laylin et al., 2010b). Finally, we also added 63 genes conserved in organisms that undergo amoeboid movement and missing in organisms that do not undergo amoeboid locomotion [the amoeboid motility (AM) gene set] (Fritz-Laylin et al., 2010b), to serve as a specificity control.

Overall, 78% of the FMs and 60% of the AMs have at least twofold induction or repression, respectively ( $P < 0.01$ , after correction for multiple testing), providing large-scale confirmation of previous evidence that *Naegleria* differentiation is controlled at the transcriptional level (Lai et al., 1988; Levy et al., 1998). We next investigated whether the timing of gene expression was linked to function.

Cluster analysis of the expression data for these 310 genes (the AM and FM gene sets, and *Naegleria* homologs of known microtubule genes) resulted in five major gene clusters (A–E; Fig. 2). Clusters A and C consist primarily of genes found in the FM gene set and have increased expression during differentiation. However, the genes in cluster A reach peak expression levels by 20 minutes and begin decreasing in expression by 40 minutes, whereas the expression of genes in cluster C peaks by 40 minutes and remains high through to 80 minutes. Manual inspection revealed that the cluster with earlier expression contains many known centriole genes (Table 1), whereas the later expression cluster contains flagella genes (supplementary material Table S1). The general induction of basal body genes before flagella genes agrees with the fact that *Naegleria* assembles its basal bodies before it assembles its flagella ( $t = 55$  and  $t = 65$  minutes, respectively) (Fig. 1) (Fritz-Laylin et al., 2010a; Fulton and Dingle, 1971).

### Centriole-enriched gene cluster

The 55 genes found in the centriole gene cluster include *Naegleria* homologs of seven genes whose products are thought to be required for assembly of the centriole or basal body:  $\epsilon$ -,  $\delta$ - and  $\eta$ -tubulin, SAS-4 (CPAP), SAS-6, centrin (Cen2) and POC1 (for references, see Table 1). This set represents the majority of components shown to be required specifically for centriole assembly that are conserved outside animals (Carvalho-Santos et al., 2010; Hodges et al., 2010; Strnad and Gönczy, 2008). Other core centriole genes not found in the cluster either have not been identified in the *Naegleria* genome (*PLK4*) or were not included in the microarray (*BLD10*).

The centriole-enriched gene cluster also encodes homologs of microtubule nucleation factors [ $\gamma$ -tubulin, GCP3 and GCP6 (Raynaud-Messina and Merdes, 2007)], as well as proteins required for general microtubule functions, such as the microtubule-severing protein katanin p60, which is known to localize to centrosomes (Hartman et al., 1998). This gene set also includes several genes encoding centrosome-localized proteins of unknown function, and eight completely uncharacterized genes (Table 1).

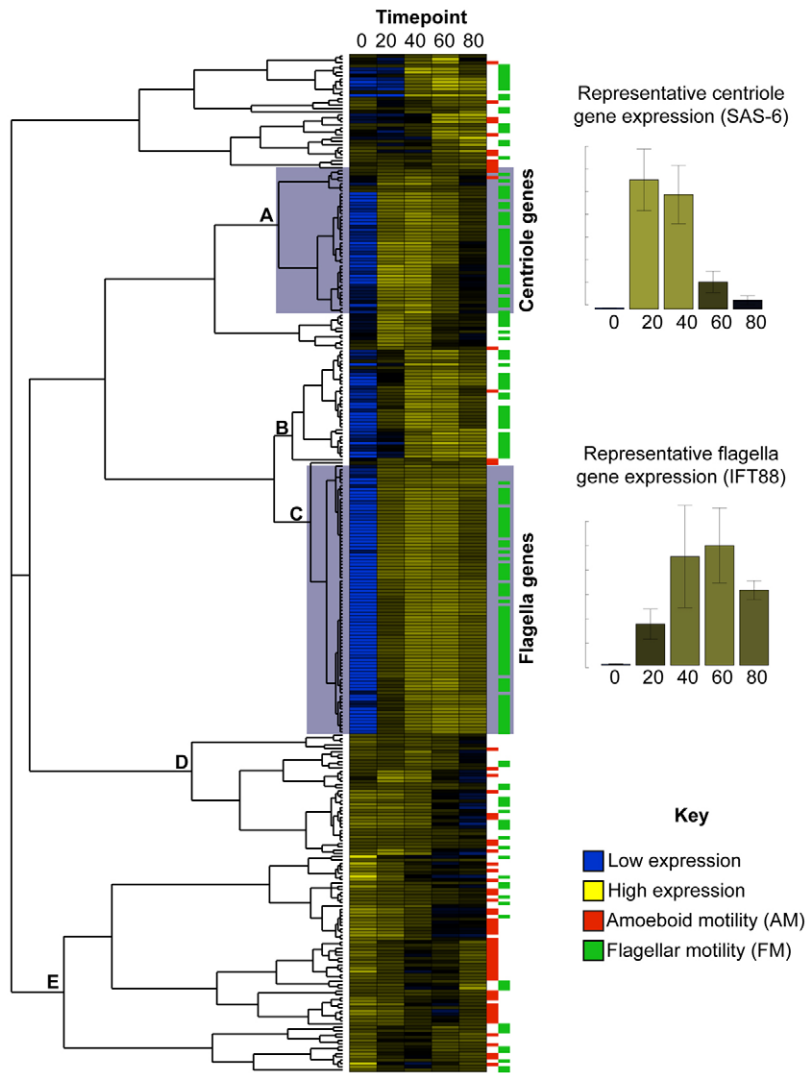
Axonemal dyneins are large protein complexes containing light, intermediate and heavy chains that slide microtubules past each other to produce flagellar movement. Surprisingly, the centriole-enriched cluster includes nine axonemal dynein light and intermediate chain homologs, as well as a homolog of kintoun (PF13) that is required for assembly of dynein arm complexes (Omran et al., 2008). By contrast, *Naegleria* dynein heavy chain genes are expressed later along with other flagella-specific genes. Assembly of dynein light and intermediate chain complexes can be genetically uncoupled from assembly of the dynein heavy chain in *Chlamydomonas* (Omran et al., 2008). Although there are many possible reasons for the early expression of dynein light and intermediate chains, but not heavy chains, it is possible that *Naegleria* pre-assembles flagellar intermediate and light chain dyneins before incorporating heavy chains, before flagellar outgrowth.

### Flagella-enriched gene cluster

The flagella-enriched gene cluster contains 82 genes (supplementary material Table S1), including genes encoding proteins used for transporting proteins to the base of the growing flagellum (BBS components BBS1–BBS5 and BBS7–BBS9) and within the flagellum to its growing tip (FLA3, kinesin 2, IFT20, IFT52, IFT57, IFT80, IFT88, IFT122 and IFT140), as well as structural components of the flagellum itself [including PF20 and PF16, RSP4 and Rib72 (Pazour et al., 2005)]. This gene set also includes 23 FM genes with homologs found in the *Chlamydomonas* flagella proteome (Pazour et al., 2005) but which are otherwise uncharacterized. Together, these data suggest that these proteins are probably core components of eukaryotic flagella and therefore prime candidates for future functional analyses.

To validate the putative flagellar components, we conducted a proteomic analysis of *Naegleria* flagella. We purified the flagella of  $\sim 4 \times 10^8$  flagellate cells using low-speed centrifugation followed by a sucrose step-gradient. The resulting sample contained flagella and no visible cell bodies and comprised largely two proteins of a size similar to that of  $\alpha$ - and  $\beta$ -tubulin (supplementary material Fig. S1, panel B), as is typical for clean flagellar preparations (e.g. Kowit and Fulton, 1974). MUDPIT mass spectrometry analysis of the sample identified 415 proteins (supplementary material Table S2).

Of the 82 genes in the flagellar-enriched gene cluster, 23 were also identified in our proteomics analysis (supplementary material



**Fig. 2. Centriole- and flagella-enriched gene clusters.** Each row represents one gene, with columns representing expression at the indicated time-point. Blue represents low expression, and yellow represents high expression. The cladogram groups genes based on similarity in expression across the five time-points; five major clusters are indicated (A–E). Red and green boxes on the right-hand side indicate membership of the AM (amoeboid motility) and FM (flagellar motility) gene sets, respectively. The graphs indicate relative gene expression – using as examples centriole-enriched cluster gene (*SAS-6*) and flagella-enriched cluster gene (*IFT88*). The data are plotted as the means  $\pm$  s.d. ( $n=3$ ).

Table S1), indicating that they are likely to be structural components of the flagellum itself (in contrast to proteins that might be required for flagellar function but are located within the cell body). Included in this overlap are seven flagellar-associated proteins (FAPs), which were identified in the proteomic analysis of *Chlamydomonas* flagella (Pazour et al., 2005) but remain otherwise uncharacterized. These proteins are therefore likely to be ancestral structural flagella components. The *Naegleria* flagellar proteome also includes a number of previously undescribed proteins (supplementary material Table S2), some of which might represent uncharacterized flagellar proteins.

#### Verification of putative centriole genes

The centriole-enriched gene cluster includes eight genes that have not previously been localized or otherwise characterized, which we refer to as ‘putative conserved centriole components’ (pCCCs; Table 1). Because orthologs of all centrosome-localized pCCCs can be found in a wide diversity of eukaryotes (supplementary material Table S3), they were probably present in the eukaryotic ancestor. To determine whether the pCCCs are likely to be centriole components, we transiently expressed N- and C-terminally GFP-tagged human orthologs of each pCCC in U2OS and HeLa human

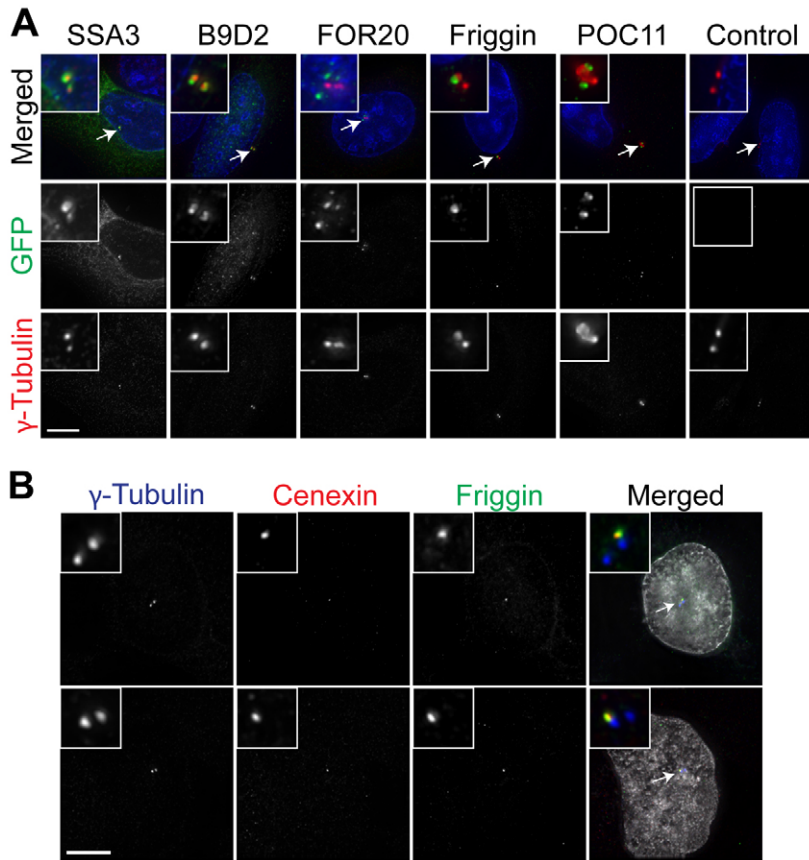
cell lines and used antibodies recognizing  $\gamma$ -tubulin to highlight centrosomes. To the eight unknown gene products, we added one whose homolog localizes to the base of the cilia of *Caenorhabditis elegans* (B9D2) and one that has only very recently been characterized in human cells (MOT52, also known as FOR20 or BBC20) (Sedjai et al., 2010). Five of the ten tagged proteins showed either diffuse cytoplasmic GFP or bright foci likely to be inclusion bodies (data not shown). This nonspecific localization neither confirms nor denies a possible centriole function. However, the remaining five localized within or near centrosomes using both N- and C-terminal GFP tags (Fig. 3A) and are described below.

First, MOT52 is found only in organisms with motile flagella (Merchant et al., 2007), and its homolog (BBC20) was found in the *Tetrahymena* basal body proteome (Kilburn et al., 2007). Recently, the human ortholog (FOR20) was reported to localize to pericentriolar satellites and to be involved in ciliary assembly (Sedjai et al., 2010). FOR20 is predicted to have a FOP dimerization domain (Pfam domain PF09398), required for the centrosomal localization of the FOP protein (Mikolajka et al., 2006). In transient transfections of both U2OS (Fig. 3A) and HeLa cells (data not shown), GFP-tagged FOR20 localized to multiple foci near centrosomes. A similar localization was reported using an antibody

Table 1. Centriole-enriched gene cluster

Class	Ortholog	Protein ID	Function	References	
BB, centriole assembly	$\epsilon$ -Tubulin	44774	BB, centriole assembly	(Dutcher et al., 2002)	
	$\delta$ -Tubulin	69007	BB, centriole assembly	(Dutcher and Trabuco, 1998; Garreau de Loubresse et al., 2001; O'Toole et al., 2003)	
	$\eta$ -Tubulin	65724	BB, centriole assembly	(Basto et al., 2006; Kirkham et al., 2003; Kleylein-Sohn et al., 2007; Leidel and Gonczy, 2003; Pelletier et al., 2006)	
	SAS-6	68996	BB, centriole assembly	(Culver et al., 2009; Dammermann et al., 2004; Leidel et al., 2005; Nakazawa et al., 2007; Peel et al., 2007; Pelletier et al., 2006; Rodrigues-Martins et al., 2007; Strnad et al., 2007)	
	SAS-4 (CPAP)	61107	BB, centriole assembly	(Basto et al., 2006; Kirkham et al., 2003; Kleylein-Sohn et al., 2007; Leidel and Gonczy, 2003; Pelletier et al., 2006)	
	Centrin	44488 56351	BB, centriole assembly	(Baum et al., 1986; Koblenz et al., 2003; Kuchka and Jarvik, 1982; Salisbury et al., 2002; Taillon et al., 1992; Winey et al., 1991)	
MT-specific centrosome function	POC1	33676	BB, centriole assembly	(Keller et al., 2009; Pearson et al., 2009)	
	$\gamma$ -Tubulin	56069	MT nucleation	(Raynaud-Messina and Merdes, 2007)	
	GCP6	61337	MT nucleation	(Raynaud-Messina and Merdes, 2007)	
	GCP3	434	MT nucleation	(Raynaud-Messina and Merdes, 2007)	
	SSA11 (TTL13)	80835	Tubulin tyrosine ligase	(Hammond et al., 2008)	
	Katanin P60	63871	MT severing	(Hartman et al., 1998)	
	$\alpha$ -Tubulin	71268	MTs	(Dutcher, 2001)	
		51830	MTs	(Dutcher, 2001)	
BB, centriole, centrosome localization	TBCE	81169	Putative tubulin chaperone		
	FOR20 (MOT52)	52938	Pericentriolar satellites, centrosome	(Sedjai et al., 2010)	
	POC12	29577	Unknown	(Keller et al., 2005)	
	PIBF	73664	Unknown	(Lachmann et al., 2004)	
BB-specific function	TTL1	33283	Tubulin tyrosine ligase	(Wloga et al., 2008)	
	MKS6 (CC2D2A)	77673	Cilium, BB function	(Gorden et al., 2008)	
	MKS3 (meckelin)	62841	BB migration, membrane docking	(Dawe et al., 2007)	
	B9D1	52666	Transition zone	(Williams et al., 2008)	
Flagellar dynein complex	B9D2	30379	Transition zone	(Williams et al., 2008)	
	MBO2	62959	Maintain direction of motility	(Tam and Lefebvre, 2002)	
	ODA6	60431	Dynein intermediate chain 2, axonemal	(Pazour et al., 2005)	
	IDA7	57343	Dynein intermediate chain, axonemal	(Pazour et al., 2005)	
	ODA6	79232	Dynein intermediate chain, axonemal	(Pazour et al., 2005)	
	DLC1	74922	Dynein light chain, axonemal	(Pazour et al., 2005)	
	IDA4 (P28)	82719	Dynein light intermediate chain, axonemal	(Pazour et al., 2005)	
	BOP5	63304	Dynein light intermediate chain, axonemal	(Pazour et al., 2005)	
	TCTEXT1	29177	Dynein light chain, axonemal	(Harrison et al., 1998; Kagami et al., 1998)	
	DLC1	54720	Dynein light chain, axonemal	(Pazour et al., 2005)	
Putative flagellar function	MOT24	32555	Dynein light chain, axonemal	(Merchant et al., 2007)	
	MOT45 (kintoun, PF13)	80717	Axonemal dynein complex assembly	(Omran et al., 2008)	
	FAP184	2066	Unknown	(Pazour et al., 2005)	
	FAP215	66643	Nucleotidase	(Pazour et al., 2005)	
	FAP127	44967	Unknown	(Pazour et al., 2005)	
	FAP57	61313	Unknown	(Pazour et al., 2005)	
	Previously uncharacterized (pCCCs)	Friggin (MOT37)	61232	Mother centriole localization	This study
		POC11	70454	Centrosome localization	This study
		SSA3	68814	Centrosome localization	This study
		Q6PH85	49668	Unknown	
POC16		62107	Unknown		
LRR6		31069	Unknown		
TECT3		65759	Unknown		
MOT39		72718	Unknown		
Other	?	73058	Dual-specificity phosphatase		
	PSK1	78247	Nucleotide kinase		
	?	82958	High-mobility-group protein		
	?	48624	Lipid synthesis		
	?	29264	HUS-1-like protein		
	?	81040	BTG-domain protein		
	MOT50	54684 71996	Chymotrypsin-like protein Chymotrypsin-like protein		

The 55 genes in the centrosome-enriched cluster are organized by function (or predicted function), followed by ortholog gene and/or protein names, with appropriate references in the last column. ?, no identifiable ortholog; BB, basal body; MT, microtubule. Sequence information for *Naegleria* protein IDs is available at [www.jgi.doe.gov/naegleria](http://www.jgi.doe.gov/naegleria).



**Fig. 3. Human orthologs of pCCCs localize to centrosomes, including the mother centriole protein Friggin.** GFP-tagged human orthologs of the indicated pCCC are shown in green after transient transfection of U2OS cells. **(A)** Centrosomes (arrows) are stained using an antibody against  $\gamma$ -tubulin (red), and DNA is shown in blue. **(B)** Centrosomes (arrows) are stained with an antibody against  $\gamma$ -tubulin (blue), mother centrioles with an antibody against cenexin (red), and DNA is shown in white. Scale bars: 5  $\mu$ m.

against FOR20 (Sedjai et al., 2010), thus validating our GFP tagging approach.

Second, B9D2 contains a B9 domain. The *C. elegans* homolog, TZA-1, localizes to the transition zone at the base of the cilium (Williams et al., 2008). Although the B9 domain has no known function, it is found in several proteins known to be localized to the centriole and/or basal body, including MKS1 (Dawe et al., 2007). Human B9D2 localized to centrosomes of U2OS cells, along with scattered foci throughout the cytoplasm (Fig. 3A).

Third, POC11 shows good conservation in many eukaryotes but has no identifiable domains other than a coiled-coil region. Localization of the human homolog (CCDC77) resulted in bright punctate spots within centrosomes of both U2OS and HeLa cells (Fig. 3A and data not shown, respectively), suggesting that POC11 represents a new family of centrosome proteins.

Fourth, SSA3 was so named for its predicted function in both motile and nonmotile flagella [SSA stands for 'sensory, structural and assembly' (Merchant et al., 2007)]. SSA3 has a conserved central region containing an 'ELMO/CED12' domain (Pfam domain PF04727), found in proteins that facilitate cytoskeletal rearrangements (Gumienny et al., 2001). SSA3-GFP-expressing cells contained diffuse cytoplasmic GFP, as well as centrosomal GFP in a small percentage (4%) of transfected cells that also displayed relatively small  $\gamma$ -tubulin foci (Fig. 3A). As  $\gamma$ -tubulin foci vary in size during the mammalian cell cycle (with small foci at G1- and early S-phases), SSA3 might localize to centrosomes in a cell-cycle-dependent manner.

Fifth, Friggin was originally named MOT37 for its predicted function in motile flagella (Merchant et al., 2007) and contains

leucine-rich repeats (LRRs), which typically mediate protein-protein interactions [Pfam clan CL0022 (Kobe and Deisenhofer, 1994)]. Unexpectedly, the 542-residue human ortholog of MOT37 localized to only one of two  $\gamma$ -tubulin foci in both U2OS (Fig. 3A) and HeLa cells (data not shown), suggesting that it might be a component specific to either mature or immature centrioles.

Centrioles develop over two cell cycles, acquiring the basic nine-triplet pinwheel structure during the first cell cycle and various appendages that allow it to function as a basal body for axonemal assembly during the second cycle. Several gene products have been shown to be involved in the assembly of appendages (Chang et al., 2003; Gromley et al., 2003; Lange and Gull, 1995; Mogensen et al., 2000; Ou et al., 2002), only one of which,  $\epsilon$ -tubulin (Chang et al., 2003), is conserved outside animals and likely to be ancestral to all extant eukaryotes.

To investigate whether MOT37 is a component of either mother or daughter centrioles, we expressed GFP-MOT37 and stained cells with an antibody recognizing the mother centriole component cenexin (Lange and Gull, 1995). GFP-MOT37 consistently colocalized with cenexin (Fig. 3B), indicating that MOT37 is a mother-centriole-specific protein that we predict is involved in the developmental transition from immature to mature centrioles. Because MOT37 probably represents a second ancestral mother centriole protein, we have named this eukaryotic protein family 'Friggin' after Frigg, the Norse goddess of motherhood.

### Concluding remarks

Understanding how centrioles and flagella assemble and function requires a full inventory of components. Previous studies have

used proteomic approaches to attempt to identify a complete parts list for centrioles (Andersen et al., 2003; Keller et al., 2005; Kilburn et al., 2007) or flagella (for a review, see Inglis et al., 2006). Theoretically, proteomic analyses can identify all stably localized proteins, including those that are species specific or required for unrelated biological functions. By contrast, our analysis has identified comprehensive sets of genes required specifically for centriole and flagellar function, independent of their localization. Of eight previously uncharacterized genes that we predict are involved in centriole assembly, at least three have human orthologs with a centrosome-related localization. The conservation of these proteins in both *Naegleria* and human, probably spanning over a billion years of eukaryotic evolution (Brinkmann and Philippe, 2007), indicates that these proteins are important for centriole function. Our study adds significantly to the list of conserved centriole components, including an additional protein (Friggin) that is apparently specific to mature centrioles.

Our analyses extend previous observations (Fulton et al., 1995; Levy et al., 1998) of two major programs of transcription during *Naegleria* differentiation: an early round of transcription of basal body genes and a later round of flagellar genes (Fig. 2), a timing that mirrors the assembly of basal bodies before flagella (Fig. 1). Although we have limited our analysis to centriole and flagella genes, this represents only one aspect of the *Naegleria* amoeba-to-flagellate transition. A cursory analysis indicates that genes from other core pathways, including basic metabolism and the stress response, are also regulated differentially. We have deposited our microarray data in the NCBI Gene Expression Omnibus (Edgar et al., 2002) under GEO Series accession number GSE21527 and encourage other scientists to take advantage of this rich data set.

## Materials and Methods

### *Naegleria* differentiation and RNA isolation

*N. gruberi* strain NEG grown on *Klebsiella* was differentiated three separate times using standard protocols (Fulton, 1970). Synchrony was estimated by counting the percentage of flagellates after fixing in Lugol's iodine (Fulton and Dingle, 1967), using a phase-contrast microscope with a  $\times 40$  objective ( $n > 100$  for each time-point).  $10^7$  cells were harvested at each time-point, and RNA extracted using Trizol reagent (Invitrogen), purified using RNeasy (Qiagen), treated with Turbo DNase (Ambion), and repurified with RNeasy (Qiagen), according to the manufacturers' instructions. RNA purity was verified by means of gel electrophoresis (supplementary material Fig. S1) and spectrophotometry.

### NimbleGen expression oligoarrays

The *N. gruberi* whole-genome expression oligoarray version 1.0 (NimbleGen Systems) comprises 182,813 probe sets corresponding to 15,777 gene models predicted on the *N. gruberi* genome sequence version 1.0 (Fritz-Laylin et al., 2010a), and an additional 963 open reading frames (ORFs) identified in intergenic regions. For each gene, 11 unique 60-mer oligonucleotide probes were designed by NimbleGen Systems. The *Naegleria* V1.0 oligoarray is fully described at the Gene Expression Omnibus (GEO) (Edgar et al., 2002) under accession number GSE21527.

Preparation of samples, hybridization and scanning were performed by NimbleGen Systems (Madison, WI), following their standard operating protocol. The raw data were subjected to robust multi-array analysis (RMA) (Irizarry et al., 2003), quantile normalization (Bolstad et al., 2003) and background correction, as implemented in the NimbleScan software package, version 2.4.27 (Roche NimbleGen). Reproducibility between biological replicates was inspected using MA [log-intensity ratios (M) versus log-intensity averages (A)] and scatter plots of log intensities (supplementary material Fig. S1), and *P*-values were calculated in a simple paired-data comparison model and were corrected for multiple testing using the BH (false discovery rate controlled) procedure, all within the R statistical package (<http://www.r-project.org/>).

### Expression clustering

The log-transformed expression data for the 310 cytoskeleton-related genes were subjected to gene normalization followed by hierarchical clustering, with centered correlation and complete linkage in the Cluster program (Eisen et al., 1998).

### Proteomics of *Naegleria* flagella

Flagella were isolated using published methods (Kowitz and Fulton, 1974) and mass spectrometry performed by the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory at UC Berkeley, CA. A nano LC column was packed in a glass capillary, of 100  $\mu\text{m}$  inner diameter, with an emitter tip. The column comprised 10 cm of Polaris c18 5- $\mu\text{m}$  packing material (Varian), followed by 4 cm of Partisphere 5 SCX (Whatman). The column was loaded using a pressure bomb and washed extensively with buffer A [5% acetonitrile, 0.02% heptafluorobutyric acid (HBFA)], then directly coupled to an electrospray ionization source mounted on a Thermo-Fisher LTQ XL linear ion-trap mass spectrometer. An Agilent 1200 HPLC delivering a flow rate of 30 nl/min was used for chromatography. Peptides were eluted using a 14-step MudPIT procedure (Washburn et al., 2001), using buffer A, buffer B (80% acetonitrile, 0.02% HBFA), buffer C (250 mM ammonium acetate, 5% acetonitrile, 0.02% HBFA) and buffer D (250 mM ammonium acetate, 5% acetonitrile, 0.02% HBFA, 500 mM ammonium acetate). The programs SEQUEST and DTASELECT were used to identify peptides and proteins from the *Naegleria* genome (Eng et al., 1994; Tabb et al., 2002).

### Localization of pCCCs

The following mammalian cDNAs from the human ORF collection in the form of Gateway entry vectors were purchased from Open Biosystems (Rual et al., 2004): B9D2 (CV025994), MOT52 (FOR20; EL735575), POC16 (EL737049), FM14 (EL735863), SSA3 (EL735155), MOT39 (CV023936) and TECT3 (EL736819), and these were verified by sequencing. The following cDNAs (and corresponding accession numbers) were obtained from Open Biosystems: POC11 (BC006444), MOT37 (BC016439) and LRRC6 (BC047286). Each ORF was amplified (primer sequences available upon request), transferred into a Gateway donor vector (pDONR221) and verified by sequencing. All ORFs were then transferred into the C-terminal (pCDNA-DEST47) and N-terminal (pCDNA-DEST53) GFP-tagged Gateway Vectors according to the manufacturers' protocols.

A total of 20,000 U2OS cells (ATCC catalog number HTB-96) was inoculated in 0.5 ml medium [DMEM (GIBCO catalog number 10569) supplemented with 10% FBS, 1% nonessential amino acids, and 1% sodium pyruvate] in 24-well plates containing coverslips. Cells were transfected the next day with lipofectamine 2000, and 14 hours later fixed for three minutes with methanol and prepared for immunofluorescence using standard methods (<http://mitchison.med.harvard.edu/protocols/gen1.html>), and the following antibodies: monoclonal antibody 20H5 against centrin (antibody 20H5) (Sanders and Salisbury, 1994), used at 1:400, antibody CD1B4 against cenexin (antibody CD1B4) (Lange and Gull, 1995) used at 1:3, and mouse monoclonal antibody against GFP (catalog number 11814460001, Roche) at a 1:500 dilution. Alexa-Fluor-conjugated secondary antibodies were sourced from Invitrogen (Carlsbad, CA) and used at a 1:500 dilution.

### Fluorescence deconvolution microscopy

Images were collected with SoftWorX image acquisition software (Applied Precision, Issaquah, WA) on an Olympus IX70 wide-field inverted fluorescence microscope with an Olympus PlanApo  $\times 100$ , (NA 1.35), oil-immersion objective and Photometrics CCD CH350 camera (Roper Scientific, Tucson, AZ). Image stacks were deconvolved with the SoftWorX deconvolution software and flattened as maximum projections (Applied Precision, Issaquah, WA).

### Sequence analysis

Domain predictions were performed using default parameters and version 24.0 of the Pfam database (Finn et al., 2008). pCCC orthologs were collected from the nonredundant ('nr') database at NCBI (Benson et al., 2009) with BLAST (Altschul et al., 1990).

We thank Lori Kohlstaedt and the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory at UC Berkeley, CA, for help with mass spectrometry experiments, Lani Keller, Juliette Azimzadeh, Hellen Dawe and Tim Stearns for helpful advice, Pierre Gónczy, Rebecca Heald, Scott Dawson, and Chandler Fulton for valuable comments on the manuscript and Keith Gull for the generous gift of the antibody against cenexin. This study was supported by a grant from the University of California Cancer Research Coordinating Committee (CRCC) to W.Z.C. and NIH grant A1054693 to W.Z.C. Article deposited in PMC for release after 12 months.

Supplementary material available online at

<http://jcs.biologists.org/cgi/content/full/123/23/4024/DC1>

## References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
- Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A. and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570-574.

- Basto, R., Lau, J., Vinogradova, T., Gardiol, A., Woods, C. G., Khodjakov, A. and Raff, J. W. (2006). Flies without centrioles. *Cell* **125**, 1375-1386.
- Baum, P., Furlong, C. and Byers, B. (1986). Yeast gene required for spindle pole body duplication: homology of its product with Ca<sup>2+</sup>-binding proteins. *Proc. Natl. Acad. Sci. USA* **83**, 5512-5516.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W. (2009). GenBank. *Nucleic Acids Res.* **37**, D26-D31.
- Bolstad, B. M., Irizarry, R. A., Astrand, M. and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193.
- Brinkmann, H. and Philippe, H. (2007). The diversity of eukaryotes and the root of the eukaryotic tree. *Adv. Exp. Med. Biol.* **607**, 20-37.
- Carvalho-Santos, Z., Machado, P., Branco, P., Tavares-Cadete, F., Rodrigues-Martins, A., Pereira-Leal, J. B. and Bettencourt-Dias, M. (2010). Stepwise evolution of the centriole-assembly pathway. *J. Cell Sci.* **123**, 1414-1426.
- Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297-354.
- Chang, P., Giddings, T. H., Jr, Winey, M. and Stearns, T. (2003). Epsilon-tubulin is required for centriole duplication and microtubule organization. *Nat. Cell Biol.* **5**, 71-76.
- Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283-1287.
- Culver, B. P., Meehl, J. B., Giddings, T. H., Jr and Winey, M. (2009). The two SAS-6 homologs in Tetrahymena thermophila have distinct functions in basal body assembly. *Mol. Biol. Cell* **20**, 1865-1877.
- Dammermann, A., Muller-Reichert, T., Pelletier, L., Habermann, B., Desai, A. and Oegema, K. (2004). Centriole assembly requires both centriolar and pericentriolar material proteins. *Dev. Cell* **7**, 815-829.
- Dawe, H. R., Smith, U. M., Cullinane, A. R., Gerrelli, D., Cox, P., Badano, J. L., Blair-Reid, S., Sriram, N., Katsanis, N., Attie-Bitach, T. et al. (2007). The Meckel-Gruber Syndrome proteins MKS1 and meckelin interact and are required for primary cilium formation. *Hum. Mol. Genet.* **16**, 173-186.
- Dutcher, S. K. (2001). The tubulin fraternity: alpha to eta. *Curr. Opin. Cell Biol.* **13**, 49-54.
- Dutcher, S. K. and Trabuco, E. C. (1998). The UN13 gene is required for assembly of basal bodies of Chlamydomonas and encodes delta-tubulin, a new member of the tubulin superfamily. *Mol. Biol. Cell* **9**, 1293-1308.
- Dutcher, S. K., Morrissette, N. S., Preble, A. M., Rackley, C. and Stanga, J. (2002). Epsilon-tubulin is an essential component of the centriole. *Mol. Biol. Cell* **13**, 3859-3869.
- Edgar, R., Domrachev, M. and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207-210.
- Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863-14868.
- Eng, J. K., McCormack, A. L. and Yates, J. R., III (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976-989.
- Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L. et al. (2008). The Pfam protein families database. *Nucleic Acids Res.* **36**, D281-D288.
- Fritz-Laylin, L. K., Assaf, Z. J., Chen, S. and Cande, W. Z. (2010a). Naegleria de novo basal body assembly occurs via stepwise incorporation of conserved proteins. *Eukaryotic Cell* **9**, 860-865.
- Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J., Carpenter, M. L., Field, M. C., Kuo, A., Paredes, A., Chapman, J., Pham, J. et al. (2010b). The genome of Naegleria gruberi illuminates early eukaryotic versatility. *Cell* **140**, 631-642.
- Fulton, C. (1970). Amebo-flagellates as research partners: the laboratory biology of Naegleria and Tetramitus. *Methods Cell Physiol.* **4**, 341-476.
- Fulton, C. (1977). Cell differentiation in Naegleria gruberi. *Annu. Rev. Microbiol.* **31**, 597-629.
- Fulton, C. and Dingle, A. D. (1967). Appearance of the flagellate phenotype in populations of Naegleria amoebae. *Dev. Biol.* **15**, 165-191.
- Fulton, C. and Dingle, A. D. (1971). Basal bodies, but not centrioles, in Naegleria. *J. Cell Biol.* **51**, 826-836.
- Fulton, C., Lai, E. Y. and Remillard, S. P. (1995). A flagellar calmodulin gene of Naegleria, coexpressed during differentiation with flagellar tubulin genes, shares DNA, RNA, and encoded protein sequence elements. *J. Biol. Chem.* **270**, 5839-5848.
- Garreau de Loubresse, N., Ruiz, F., Beisson, J. and Klotz, C. (2001). Role of delta-tubulin and the C-tubule in assembly of Paramecium basal bodies. *BMC Cell Biol.* **2**, 4.
- Gorden, N. T., Arts, H. H., Parisi, M. A., Coene, K. L., Letteboer, S. J., van Beersum, S. E., Mans, D. A., Hikida, A., Eckert, M., Knutzen, D. et al. (2008). CC2D2A is mutated in Joubert syndrome and interacts with the ciliopathy-associated basal body protein CEP290. *Am. J. Hum. Genet.* **83**, 559-571.
- Gromley, A., Jurczyk, A., Sillibourne, J., Halilovic, E., Mogensen, M., Groisman, L., Blomberg, M. and Doxsey, S. (2003). A novel human protein of the maternal centriole is required for the final stages of cytokinesis and entry into S phase. *J. Cell Biol.* **161**, 535-545.
- Gumienny, T. L., Brugnera, E., Tosello-Trampont, A. C., Kinchen, J. M., Haney, L. B., Nishiwaki, K., Walk, S. F., Nemergut, M. E., Macara, I. G., Francis, R. et al. (2001). CED-12/ELMO, a novel member of the CrkII/Dock180/Rac pathway, is required for phagocytosis and cell migration. *Cell* **107**, 27-41.
- Hammond, J. W., Cai, D. and Verhey, K. J. (2008). Tubulin modifications and their cellular functions. *Curr. Opin. Cell Biol.* **20**, 71-76.
- Harrison, A., Olds-Clarke, P. and King, S. M. (1998). Identification of the t complex-encoded cytoplasmic dynein light chain tctex1 in inner arm II supports the involvement of flagellar dyneins in meiotic drive. *J. Cell Biol.* **140**, 1137-1147.
- Hartman, J. J., Mahr, J., McNally, K., Okawa, K., Iwamatsu, A., Thomas, S., Cheesman, S., Heuser, J., Vale, R. D. and McNally, F. J. (1998). Katanin, a microtubule-severing protein, is a novel AAA ATPase that targets to the centrosome using a WD40-containing subunit. *Cell* **93**, 277-287.
- Hodges, M. E., Scheumann, N., Wickstead, B., Langdale, J. A. and Gull, K. (2010). Reconstructing the evolutionary history of the centriole from protein components. *J. Cell Sci.* **123**, 1407-1413.
- Inglis, P. N., Borovchik, K. A. and Leroux, M. R. (2006). Piecing together a cilium. *Trends Genet.* **22**, 491-500.
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264.
- Kagami, O., Gotoh, M., Makino, Y., Mohri, H., Kamiya, R. and Ogawa, K. (1998). A dynein light chain of sea urchin sperm flagella is a homolog of mouse Tctex 1, which is encoded by a gene of the t complex sterility locus. *Gene* **211**, 383-386.
- Keller, L. C., Romijn, E. P., Zamora, I., Yates, J. R., 3rd and Marshall, W. F. (2005). Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. *Curr. Biol.* **15**, 1090-1098.
- Keller, L. C., Geimer, S., Romijn, E., Yates, J., 3rd, Zamora, I. and Marshall, W. F. (2009). Molecular architecture of the centriole proteome: the conserved WD40 domain protein POC1 is required for centriole duplication and length control. *Mol. Biol. Cell* **20**, 1150-1166.
- Kilburn, C. L., Pearson, C. G., Romijn, E. P., Meehl, J. B., Giddings, T. H., Jr, Culver, B. P., Yates, J. R., 3rd and Winey, M. (2007). New Tetrahymena basal body protein components identify basal body domain structure. *J. Cell Biol.* **178**, 905-912.
- Kirkham, M., Muller-Reichert, T., Oegema, K., Grill, S. and Hyman, A. A. (2003). SAS-4 is a C. elegans centriolar protein that controls centrosome size. *Cell* **112**, 575-587.
- Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habadanck, R., Stierhof, Y. D. and Nigg, E. A. (2007). Plk4-induced centriole biogenesis in human cells. *Dev. Cell* **13**, 190-202.
- Kobe, B. and Deisenhofer, J. (1994). The leucine-rich repeat: a versatile binding motif. *Trends Biochem. Sci.* **19**, 415-421.
- Koblenz, B., Schoppmeier, J., Grunow, A. and Lechtreck, K. F. (2003). Centrin deficiency in Chlamydomonas causes defects in basal body replication, segregation and maturation. *J. Cell Sci.* **116**, 2635-2646.
- Kowitz, J. D. and Fulton, C. (1974). Purification and properties of flagellar outer doublet tubulin from Naegleria gruberi and a radioimmuno assay for tubulin. *J. Biol. Chem.* **249**, 3638-3646.
- Kuchka, M. R. and Jarvik, J. W. (1982). Analysis of flagellar size control using a mutant of Chlamydomonas reinhardtii with a variable number of flagella. *J. Cell Biol.* **92**, 170-175.
- Lachmann, M., Gelbmann, D., Kalman, E., Polgar, B., Buschle, M., Von Gabain, A., Szekeres-Bartho, J. and Nagy, E. (2004). PIBF (progesterone induced blocking factor) is overexpressed in highly proliferating cells and associated with the centrosome. *Int. J. Cancer* **112**, 51-60.
- Lai, E. Y., Remillard, S. P. and Fulton, C. (1988). The alpha-tubulin gene family expressed during cell differentiation in Naegleria gruberi. *J. Cell Biol.* **106**, 2035-2046.
- Lange, B. M. and Gull, K. (1995). A molecular marker for centriole maturation in the mammalian cell cycle. *J. Cell Biol.* **130**, 919-927.
- Leidel, S. and Gönczy, P. (2003). SAS-4 is essential for centrosome duplication in C. elegans and is recruited to daughter centrioles once per cell cycle. *Dev. Cell* **4**, 431-439.
- Leidel, S., Delattre, M., Cerutti, L., Baumer, K. and Gönczy, P. (2005). SAS-6 defines a protein family required for centrosome duplication in C. elegans and in human cells. *Nat. Cell Biol.* **7**, 115-125.
- Levy, Y. Y., Lai, E. Y., Remillard, S. P. and Fulton, C. (1998). Centrin is synthesized and assembled into basal bodies during Naegleria differentiation. *Cell Motil. Cytoskeleton* **40**, 249-260.
- Marshall, W. F. and Nonaka, S. (2006). Cilia: tuning in to the cell's antenna. *Curr. Biol.* **16**, R604-R614.
- Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Marechal-Drouard, L. et al. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* **318**, 245-250.
- Mikolajka, A., Yan, X., Popowicz, G. M., Smialowski, P., Nigg, E. A. and Holak, T. A. (2006). Structure of the N-terminal domain of the FOP (FGFR1OP) protein and implications for its dimerization and centrosomal localization. *J. Mol. Biol.* **359**, 863-875.
- Mogensen, M. M., Malik, A., Piel, M., Bouckon-Castaing, V. and Bornens, M. (2000). Microtubule minus-end anchorage at centrosomal and non-centrosomal sites: the role of ninein. *J. Cell Sci.* **113**, 3013-3023.
- Nakazawa, Y., Hiraki, M., Kamiya, R. and Hirono, M. (2007). SAS-6 is a cartwheel protein that establishes the 9-fold symmetry of the centriole. *Curr. Biol.* **17**, 2169-2174.
- Omran, H., Kobayashi, D., Olbrich, H., Tsukahara, T., Loges, N. T., Hagiwara, H., Zhang, Q., Leblond, G., O'Toole, E., Hara, C. et al. (2008). Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. *Nature* **456**, 611-616.

- O'Toole, E. T., Giddings, T. H., McIntosh, J. R. and Dutcher, S. K. (2003). Three-dimensional organization of basal bodies from wild-type and delta-tubulin deletion strains of *Chlamydomonas reinhardtii*. *Mol. Biol. Cell* **14**, 2999-3012.
- Ou, Y. Y., Mack, G. J., Zhang, M. and Rattner, J. B. (2002). CEP110 and ninein are located in a specific domain of the centrosome associated with centrosome maturation. *J. Cell Sci.* **115**, 1825-1835.
- Pazour, G. J., Agrin, N., Leszyk, J. and Witman, G. B. (2005). Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* **170**, 103-113.
- Pearson, C. G., Osborn, D. P., Giddings, T. H., Jr, Beales, P. L. and Winey, M. (2009). Basal body stability and ciliogenesis requires the conserved component Poc1. *J. Cell Biol.* **187**, 905-920.
- Peel, N., Stevens, N. R., Basto, R. and Raff, J. W. (2007). Overexpressing centriole-replication proteins in vivo induces centriole overduplication and de novo formation. *Curr. Biol.* **17**, 834-843.
- Pelletier, L., O'Toole, E., Schwager, A., Hyman, A. A. and Muller-Reichert, T. (2006). Centriole assembly in *Caenorhabditis elegans*. *Nature* **444**, 619-623.
- Raynaud-Messina, B. and Merdes, A. (2007). Gamma-tubulin complexes and microtubule organization. *Curr. Opin. Cell Biol.* **19**, 24-30.
- Rodrigues-Martins, A., Bettencourt-Dias, M., Riparbelli, M., Ferreira, C., Ferreira, I., Callaini, G. and Glover, D. M. (2007). DSAS-6 organizes a tube-like centriole precursor, and its absence suggests modularity in centriole assembly. *Curr. Biol.* **17**, 1465-1472.
- Rual, J. F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P. O. et al. (2004). Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* **14**, 2128-2135.
- Salisbury, J. L., Suino, K. M., Busby, R. and Springett, M. (2002). Centrin-2 is required for centriole duplication in mammalian cells. *Curr. Biol.* **12**, 1287-1292.
- Sanders, M. A. and Salisbury, J. L. (1994). Centrin plays an essential role in microtubule severing during flagellar excision in *Chlamydomonas reinhardtii*. *J. Cell Biol.* **124**, 795-805.
- Sedjai, F., Acquaviva, C., Chevrier, V., Chauvin, J. P., Coppin, E., Aouane, A., Coulter, F., Tolun, A., Pierres, M., Birnbaum, D. et al. (2010). Control of ciliogenesis by FOR20, a novel centrosome and pericentriolar satellite protein. *J. Cell Sci.* **123**, 2391-2401.
- Strnad, P. and Gönczy, P. (2008). Mechanisms of procentriole formation. *Trends Cell Biol.* **18**, 389-396.
- Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A. and Gönczy, P. (2007). Regulated HsSAS-6 levels ensure formation of a single procentriole per centriole during the centrosome duplication cycle. *Dev. Cell* **13**, 203-213.
- Tabb, D. L., McDonald, W. H. and Yates, J. R., 3rd (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**, 21-26.
- Taillon, B. E., Adler, S. A., Suhan, J. P. and Jarvik, J. W. (1992). Mutational analysis of centrin: an EF-hand protein associated with three distinct contractile fibers in the basal body apparatus of *Chlamydomonas*. *J. Cell Biol.* **119**, 1613-1624.
- Tam, L. W. and Lefebvre, P. A. (2002). The *Chlamydomonas* MBO2 locus encodes a conserved coiled-coil protein important for flagellar waveform conversion. *Cell Motil. Cytoskeleton* **51**, 197-212.
- Washburn, M. P., Wolters, D. and Yates, J. R., 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242-247.
- Williams, C. L., Winkelbauer, M. E., Schafer, J. C., Michaud, E. J. and Yoder, B. K. (2008). Functional redundancy of the B9 proteins and nephrocystins in *Caenorhabditis elegans* ciliogenesis. *Mol. Biol. Cell* **19**, 2154-2168.
- Winey, M., Goetsch, L., Baum, P. and Byers, B. (1991). MPS1 and MPS2: novel yeast genes defining distinct steps of spindle pole body duplication. *J. Cell Biol.* **114**, 745-754.
- Wloga, D., Rogowski, K., Sharma, N., Van Dijk, J., Janke, C., Edde, B., Bre, M. H., Levilliers, N., Redeker, V., Duan, J. et al. (2008). Glutamylation on alpha-tubulin is not essential but affects the assembly and functions of a subset of microtubules in *Tetrahymena thermophila*. *Eukaryotic Cell* **7**, 1362-1372.



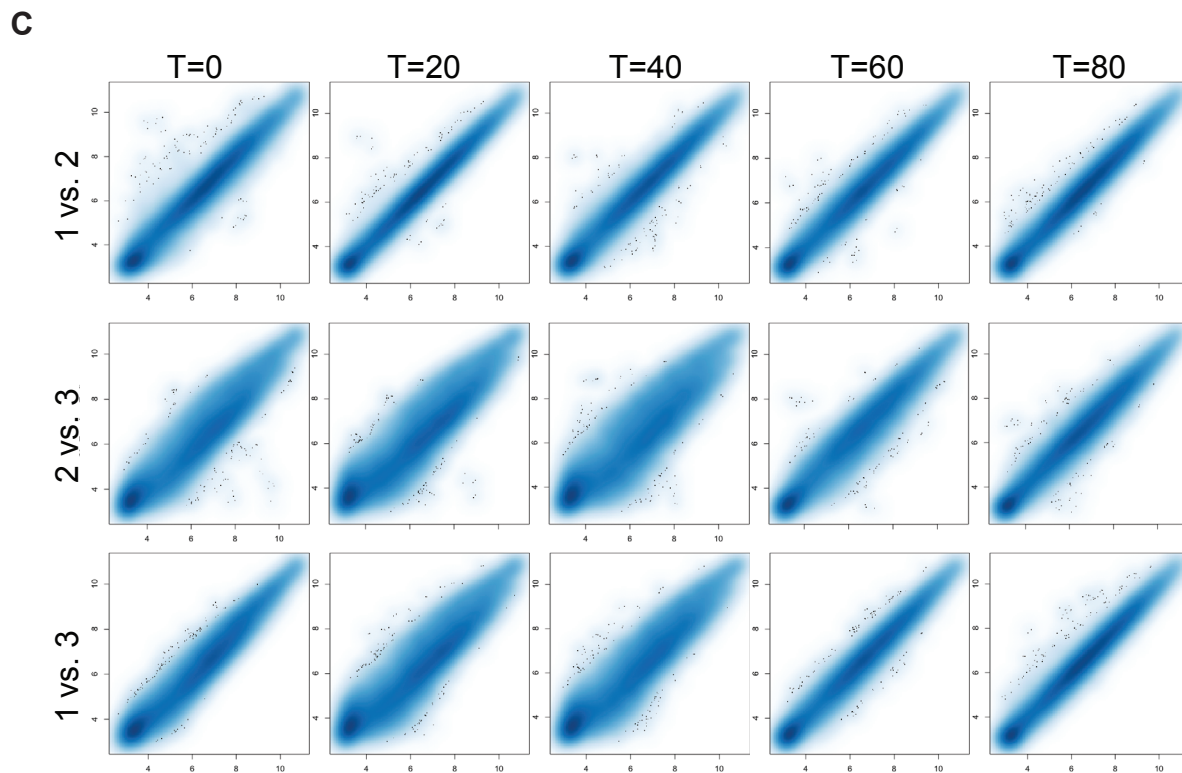
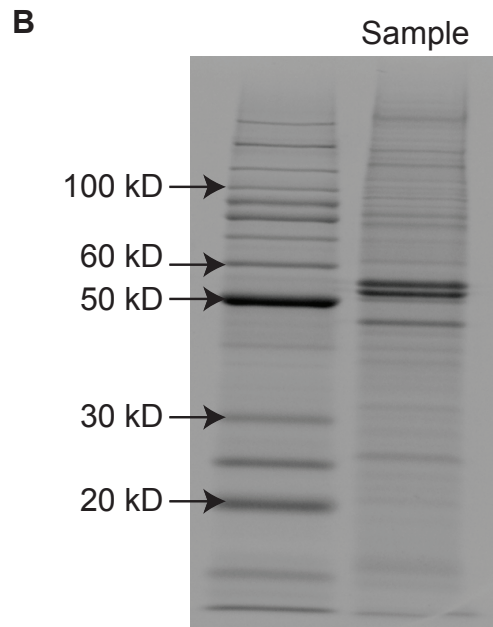
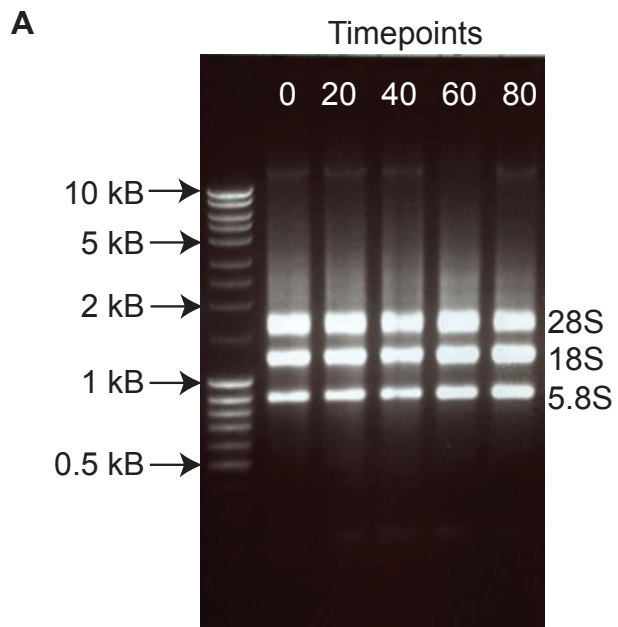


Table S1. Flagellar enriched gene cluster

Predicted function	JGI Peptide ID	Proteomics Hit	Predicted homolog
<b>IFT (Intraflagellar transport)</b>	79669		FLA3/Kinesin-associated protein 3
	63939		FLA2/FLA8 (kinesin 2 homolog)
	62977		IFT20
	77715		IFT52
	45002	Yes	IFT57
	77945		IFT80
	63280	Yes	IFT88
	71180		IFT122
	48798		IFT140
<b>BBS (Transport to flagellar base)</b>	65179		BBS1
	71257		BBS2
	44202		BBS3
	28891		BBS4
	34252		BBS5
	68114		BBS7
	80979		BBS8
	80972		BBS9
<b>Structural components</b>	39221	Yes	Alpha-tubulin
	56065		Alpha-tubulin
	56236		Alpha-tubulin
	83350		Beta-tubulin
	55423	Yes	Beta-tubulin
	56391		Beta-tubulin
	78637		ODA9 (outer arm dynein intermediate chain)
	80259	Yes	D1bLIC (dynein light intermediate chain)
	81548	Yes	ODA1 (p66 outer dynein arm docking complex protein)
	30532		ODA12
	81047	Yes	RIB72
	83064	Yes	RIB43A-domain containing protein
	49798	Yes	Radial spoke protein 4
	44954	Yes	Radial spoke-head-like
	30562	Yes	PF16
	952		PF20/SPG16
	82851	Yes	PACRG
	55564		Flagellar calmodulin (CAM1)
64648		KLP1 (kinesin-9)	
<b>General MT associated</b>	72175	Yes	Katanan P80
	83220		katanin p60
	62120		CLASP-like
	71898		DIP13
	64818	Yes	XRP2
<b>FAPP</b>	4601	Yes	FAP9
	50227		FAP14
	62998		FAP22

	81229		FAP32
	67854	Yes	FAP45
	68117		FAP50
	59637	Yes	FAP52
	380	Yes	FAP59
	50399		FAP60
	73596		FAP61
	29690		FAP66
	33146		FAP67
	62358		FAP69
	70274		FAP70
	79290		FAP82
	46605	Yes	FAP91
	79626	Yes	FAP116
	32701		FAP118.
	74561		FAP134
	65873		FAP161
	61993		FAP259
	54982		FAP253
	29188	Yes	FAP251
<b>Unknown</b>	32341		Kinase
	66608		Guanylate cyclase
	63921		MOT17
	56610		MOT25
	73917		MOT51
	63091	Yes	SSA4
	67046		SSA20 (chymotrypsin-like domain)
	74042		Leucine-rich repeats
	56805		Unknown
	60926		RABL2A
	79456	Yes	ARL3
	68064		ARL13
<b>Other</b>	80690		DYF13
	56340		POC7/UNC119/HRG4.
	29950		Nucleoside diphosphate kinase (flagellate-specific domain structure)
	69688	Yes	DPY30
	78184		FKB12
	78645	Yes	PDE14

Genes contained in the flagella-enriched gene cluster are classified by predicted function (left).

Gene information for each protein ID in the second column can be found at: <http://genome.jgi-psf.org/NaeGr1/NaeGr1.home.html>. Genes with hits in the proteomic analysis of *Naegleria* flagella are indicated in the second column with a “Yes”. Predicted orthologs of each gene are listed on the far right.

## Table S2. Proteomics analysis of *Naegleria flagella*

The *Naegleria* JGI protein ID and the number of peptides with strong confidence. *Naegleria* proteins that also belong to the FM gene set are indicated by a "Yes". The *Chlamydomonas* JGI protein ID for the top blast hit to version 4 of the *Chlamydomonas* genome and the NCBI accession for the top blast hit to the human genome, and corresponding e-values are given. The designation "nf" is used to indicate that no homolog with a e-value less than 1e-5 was identified. Pfam protein domain predictions for the *Naegleria* protein are based on version 24 of the Pfam database using gathering thresholds. For proteins with multiple Pfam domains, domains are listed sequentially, separated by a vertical line.

Naegleria protein ID	No. peptides	FM gene family member	Chlamydomonas top Blast hit	Chlamydomonas e-value	Human top blast hit	Human e-value	Human blast hit description	Naegleria protein Pfam domains
55423	71		129876	3.80E-217	NP_006079.1	1.70E-211	tubulin, beta, 2	PF00091: Tubulin/FtsZ, GTPase   PF03953: Tubulin/FtsZ, C-terminal
39221	38		128523	7.90E-217	NP_006000.2	3.90E-205	tubulin, alpha 1a	PF00091: Tubulin/FtsZ, GTPase   PF03953: Tubulin/FtsZ, C-terminal
80962	22		150766	1.90E-76	NP_904325.2	1.70E-117	kinesin family member 1B isoform alpha	PF00225: Kinesin, motor region   PF00498: Forkhead-associated
56359	22		nf	n/a	NP_003079.1	1.40E-10	fascin 1	PF06268: Fascin
61055	13		24114	1.80E-61	NP_001091.1	3.70E-120	actin, alpha 1, skeletal muscle	PF00022: Actin/actin-like
55913	13		132905	6.30E-89	NP_001393.1	1.10E-170	eukaryotic translation elongation factor 1 alpha 1	PF00009: Protein synthesis factor, GTP-binding   PF03144: Translation elongation factor EFTu/EF1A, domain 2   PF03143: Translation elongation factor EFTu/EF1A, C-terminal
88138	12		24114	1.20E-85	NP_001605.1	4.20E-176	actin, gamma 1 propeptide	PF00022: Actin/actin-like
56300	11		134171	2.20E-299	NP_009057.1	5.40E-302	valosin-containing protein	PF02359: AAA ATPase VAT, N-terminal   PF02933: Cell division protein 48, CDC48, domain 2   PF00004: AAA ATPase, core   PF00004: AAA ATPase, core
50528	11		145401	5.20E-19	NP_997200.1	7.70E-26	zinc finger protein 474	nf
55628	11	Yes	130849	0	XP_002345123.1	0	PREDICTED: dynein, axonemal, heavy chain 17	PF08385: Dynein heavy chain, N-terminal region 1   PF08393: Dynein heavy chain, N-terminal region 2   PF07728: ATPase associated with various cellular activities, AAA-5   PF07728: ATPase associated with various cellular activities, AAA-5   PF03028: Dynein heavy chain
58174	10		196241	7.00E-10	NP_001107829.1	1.30E-10	hypothetical protein LOC441054	nf
56244	10		16794	5.20E-11	NP_116243.2	6.00E-97	coronin 6	PF08953: Domain of unknown function DUF1899   PF00400: WD40 repeat   PF00400: WD40 repeat   PF08954: Domain of unknown function DUF1900
56254	10		24524	8.20E-270	NP_001952.1	3.40E-277	eukaryotic translation elongation factor 2	PF00009: Protein synthesis factor, GTP-binding   PF03144: Translation elongation factor EFTu/EF1A, domain 2   PF03764: Translation elongation factor EFG/EF2, domain IV   PF00679: Translation elongation factor EFG/EF2, C terminal
35692	9		129809	3.40E-35	NP_150254.1	9.60E-35	ribosomal protein L13	PF01294: Ribosomal protein L13e
78965	9		132719	1.20E-16	NP_659482.3	1.10E-25	coiled-coil domain containing 151	nf
81126	9		nf	n/a	nf	n/a	nf	nf
59363	9		206694	2.30E-201	NP_057392.1	3.00E-42	UMP-CMP kinase 1 isoform a	PF01326: Pyruvate phosphate dikinase, PEP/pyruvate-binding   PF00391: PEP-utilising enzyme, mobile region   PF02896: PEP-utilizing enzyme   PF00406: Adenylate kinase
44741	8		185673	1.60E-211	NP_006588.1	2.40E-221	heat shock 70kDa protein 8 isoform 1	PF00012: Heat shock protein 70
69418	8		342942	1.90E-39	NP_004046.2	9.00E-55	calpain 5	PF00648: Peptidase C2, calpain   PF01067: Peptidase C2, calpain
70087	8		137559	2.00E-23	NP_570967.1	7.40E-28	dynein, light chain, roadblock-type 2	PF03259: Roadblock/LC7
56405	8		185673	2.30E-217	NP_006588.1	1.90E-228	heat shock 70kDa protein 8 isoform 1	PF00012: Heat shock protein 70
82851	7	Yes	97201	3.50E-65	NP_001073847.1	1.30E-60	parkin co-regulated gene protein isoform 2	PF10274: ParcG
35785	7		172406	8.80E-76	NP_150644.1	2.20E-74	ribosomal protein L8	PF00181: Ribosomal protein L2   PF03947: Ribosomal protein L2
30562	7	Yes	103782	3.60E-118	NP_036575.1	1.20E-132	sperm associated antigen 6 isoform 1	PF00514: Armadillo   PF00514: Armadillo   PF02985: HEAT   PF00514: Armadillo
82397	6		96566	4.90E-66	NP_008965.2	1.30E-69	DnaJ (Hsp40) homolog, subfamily B, member 4	PF00226: Heat shock protein DnaJ, N-terminal   PF01556: Chaperone DnaJ, C-terminal
59029	6		nf	n/a	nf	n/a	nf	PF00736: Translation elongation factor EF1B, beta and delta chains, guanine nucleotide exchange
81457	6		408539	4.40E-15	NP_001400.3	3.60E-36	EGF-like-domain, multiple 3	PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF02010: PKD/REJ-like protein
81553	6		nf	n/a	nf	n/a	nf	nf

39465	5		138697	3.20E-46	NP_000967.1	2.20E-53	ribosomal protein L12	PF03946: Ribosomal protein L11   PF00298: Ribosomal protein L11
78704	5	Yes	116240	0	NP_116210.2	0	hydrocephalus inducing isoform a	nf
81047	5	Yes	126286	1.30E-99	NP_060570.2	1.10E-102	EF-hand domain (C-terminal) containing 1	PF06565: Protein of unknown function DUF1126   PF06565: Protein of unknown function DUF1126   PF06565: Protein of unknown function DUF1126
83254	5		181887	8.50E-07	nf	n/a	nf	PF00235: Profilin/allergen
56516	5		150605	4.10E-14	NP_055111.1	3.70E-16	calpain 7	PF00648: Peptidase C2, calpain
77087	5		182408	3.70E-52	NP_000420.1	1.40E-56	methionine adenosyltransferase I, alpha	PF02773: S-adenosylmethionine synthetase
58590	5		9715	4.00E-08	NP_002464.1	7.20E-11	myosin, heavy polypeptide 9, non-muscle	nf
80070	5		57302	6.90E-12	NP_001019.1	1.40E-12	ribosomal protein S25	PF03297: S25 ribosomal protein   PF00376: Bacterial regulatory protein, MerR
59812	4		398336	5.20E-16	nf	n/a	nf	nf
60431	4	Yes	188612	1.20E-85	NP_075462.3	4.80E-88	dynein, axonemal, intermediate polypeptide 2	nf
56558	4		106571	5.40E-15	NP_543136.1	5.70E-18	testis-specific gene A2	PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif
82967	4		159282	1.20E-53	NP_000969.1	3.40E-55	ribosomal protein L23	PF00238: Ribosomal protein L14b/L23e
80333	4		195598	9.60E-86	NP_000959.2	2.10E-85	ribosomal protein L4	PF00573: Ribosomal protein L4/L1e
83065	4		158129	1.50E-87	NP_005908.1	3.10E-100	cytosolic malate dehydrogenase	PF00056: Lactate/malate dehydrogenase   PF02866: Lactate/malate dehydrogenase
44954	4		54360	1.10E-26	NP_001010892.1	1.50E-35	radial spoke head 4 homolog A isoform 1	PF04712: Radial spokehead-like protein
57391	4		118745	1.20E-92	NP_005712.1	1.50E-118	ARP3 actin-related protein 3 homolog	PF00022: Actin/actin-like
37059	4		60432	0	NP_997320.2	0	dynein, axonemal, heavy chain 10	PF08385: Dynein heavy chain, N-terminal region 1   PF08393: Dynein heavy chain, N-terminal region 2   PF07728: ATPase associated with various cellular activities, AAA-5   PF03028: Dynein heavy chain
49798	4	Yes	24016	1.80E-27	NP_001010892.1	6.20E-23	radial spoke head 4 homolog A isoform 1	PF04712: Radial spokehead-like protein
78488	4		107530	8.30E-95	NP_872270.1	4.70E-90	pyruvate kinase, muscle isoform M1	PF00224: Pyruvate kinase, barrel   PF02887: Pyruvate kinase, alpha/beta
78559	4	Yes	130324	0	NP_056327.4	0	dynein, axonemal, heavy chain 1	PF08393: Dynein heavy chain, N-terminal region 2   PF07728: ATPase associated with various cellular activities, AAA-5   PF07728: ATPase associated with various cellular activities, AAA-5   PF07728: ATPase associated with various cellular activities, AAA-5   PF03028: Dynein heavy chain
44910	4		185571	2.80E-54	NP_000933.1	5.10E-52	peptidylprolyl isomerase B precursor	PF00160: Peptidyl-prolyl cis-trans isomerase, cyclophilin-type
31874	4		129362	9.20E-42	NP_057392.1	7.20E-46	UMP-CMP kinase 1 isoform a	PF00406: Adenylate kinase
82392	4		24114	9.40E-79	NP_001605.1	2.00E-162	actin, gamma 1 propeptide	PF00022: Actin/actin-like
63673	4		nf	n/a	nf	n/a	nf	nf
81845	4	Yes	134599	0	NP_061720.2	0	dynein, axonemal, heavy chain 7	PF08393: Dynein heavy chain, N-terminal region 2   PF07728: ATPase associated with various cellular activities, AAA-5   PF07728: ATPase associated with various cellular activities, AAA-5   PF03028: Dynein heavy chain
78928	4		nf	n/a	nf	n/a	nf	nf
75449	4		128099	1.20E-190	NP_068768.2	1.70E-170	plasma membrane calcium ATPase 3 isoform 3a	PF00122: E1-E2 ATPase-associated region   PF00702: Haloacid dehalogenase-like hydrolase   PF00689: ATPase, P-type cation-transporter, C-terminal
83145	4		179706	3.00E-43	NP_001002.1	9.50E-44	ribosomal protein S7	PF01251: Ribosomal protein S7E
36356	4		195590	2.50E-46	NP_000972.1	2.10E-46	ribosomal protein L19	PF01280: Ribosomal protein L19e
76335	3		nf	n/a	NP_003079.1	2.50E-09	fascin 1	PF06268: Fascin
80069	3		nf	n/a	nf	n/a	nf	nf
80996	3		nf	n/a	nf	n/a	nf	PF00036: Calcium-binding EF-hand
55960	3		195602	2.70E-74	NP_002943.2	1.20E-73	ribosomal protein S2	PF00333: Ribosomal protein S5, N-terminal   PF03719: Ribosomal protein S5, C-terminal
35035	3		174711	5.90E-22	NP_000975.2	3.20E-27	ribosomal protein L23a	PF03939: Ribosomal protein L23, N-terminal   PF00276: Ribosomal protein L25/L23
78595	3		284442	3.40E-08	nf	n/a	nf	nf
36068	3		182408	4.90E-130	NP_005902.1	2.60E-135	methionine adenosyltransferase II, alpha	PF00438: S-adenosylmethionine synthetase   PF02772: S-adenosylmethionine synthetase   PF02773: S-adenosylmethionine synthetase
79456	3	Yes	128761	1.00E-49	NP_004302.1	2.70E-53	ADP-ribosylation factor-like 3	PF00025: ARF/SAR superfamily
80939	3		297279	1.00E-35	NP_003960.1	2.10E-32	ubiquitin-conjugating enzyme E2M	PF00179: Ubiquitin-conjugating enzyme, E2
62019	3		nf	n/a	nf	n/a	nf	nf
56829	3		402056	9.10E-42	NP_006357.1	1.40E-58	adenylyl cyclase-associated protein 2	PF01213: Adenylate cyclase-associated CAP, N-terminal   PF08603: Adenylate cyclase-associated CAP, C-terminal

78686	3		107474	8.60E-156	NP_006421.2	4.10E-153	chaperonin containing TCP1, subunit 4 (delta)	PF00118: Chaperonin Cpn60/TCP-1
2066	3	Yes	192295	5.90E-38	NP_699207.1	2.50E-31	coiled-coil domain containing 96	nf
66931	3		nf	n/a	nf	n/a	nf	nf
78645	3	Yes	101278	7.10E-58	NP_006194.2	4.90E-63	phosphodiesterase 4D isoform 2	PF00233: 3'5'-cyclic nucleotide phosphodiesterase
81548	3	Yes	132719	2.40E-57	NP_689804.1	8.50E-35	coiled-coil domain containing 63	nf
56285	3		188144	5.30E-15	NP_004915.2	2.10E-110	actinin, alpha 4	PF00307: Calponin-like actin-binding   PF00307: Calponin-like actin-binding   PF00435: Spectrin repeat   PF00036: Calcium-binding EF-hand   PF00036: Calcium-binding EF-hand   PF08726: EF-hand, Ca insensitive
34423	3		nf	n/a	NP_006127.1	1.70E-44	capping protein (actin filament) muscle Z-line, alpha 2	PF01267: F-actin capping protein, alpha subunit
64186	3		182635	3.70E-08	NP_114122.2	9.30E-08	ropporin 1-like	nf
81906	3		195587	3.20E-55	NP_000966.2	4.70E-58	ribosomal protein L11	PF00281: Ribosomal protein L5   PF00673: Ribosomal protein L5
82992	3		83064	2.30E-91	NP_001966.1	6.50E-84	enolase 2	PF03952: Enolase   PF00113: Enolase
67854	3	Yes	206096	6.90E-44	NP_036469.2	4.20E-29	nasopharyngeal epithelium specific protein 1	nf
58587	3		183782	3.10E-73	NP_000960.2	2.50E-84	ribosomal protein L5	PF00861: Ribosomal protein L18P/L5E
81952	3		192162	2.10E-33	NP_005960.1	4.00E-29	nucleosome assembly protein 1-like 4	PF00956: Nucleosome assembly protein (NAP)
60866	3		136865	3.60E-70	NP_002852.1	1.70E-76	phosphate cytidyltransferase 2, ethanolamine	PF01467: Cytidyltransferase   PF01467: Cytidyltransferase
56310	3		180130	5.50E-35	NP_001006.1	6.80E-34	ribosomal protein S11	PF00366: Ribosomal protein S17
81218	3		132210	1.00E-95	NP_000282.1	4.40E-117	phosphoglycerate kinase 1	PF00162: Phosphoglycerate kinase
35538	3		195613	1.40E-15	NP_003320.2	1.90E-20	thioredoxin	PF00085: Thioredoxin domain
77219	3		138493	5.40E-163	NP_005539.1	1.70E-163	lysyl-tRNA synthetase isoform 2	PF01336: Nucleic acid binding, OB-fold, tRNA/helicase-type   PF00152: Aminoacyl-tRNA synthetase, class II (D, K and N)
82731	3		196567	7.70E-132	NP_002159.2	6.70E-130	isocitrate dehydrogenase 2 (NADP+), mitochondrial precursor	PF00180: Isocitrate/isopropylmalate dehydrogenase
57012	3		168484	9.50E-47	NP_000997.1	6.70E-43	ribosomal protein S3a	PF01015: Ribosomal protein S3Ae
61583	3		nf	n/a	nf	n/a	nf	nf
68583	3		nf	n/a	nf	n/a	nf	nf
59268	3		152943	8.20E-06	nf	n/a	nf	nf
67893	2		nf	n/a	NP_001138890.1	6.60E-10	tetratricopeptide repeat domain 28	nf
77772	2		188942	7.40E-63	NP_037396.3	5.20E-75	DEAD (Asp-Glu-Ala-Asp) box polypeptide 25	PF00270: DNA/RNA helicase, DEAD/DEAH box type, N-terminal   PF00271: DNA/RNA helicase, C-terminal
59931	2		166203	1.90E-13	NP_689993.1	1.10E-06	MAP7 domain containing 2	nf
77735	2		295585	2.30E-66	NP_064502.9	1.40E-264	leucyl-tRNA synthetase	PF09334: Aminoacyl-tRNA synthetase, class I (M)   PF08264: Valyl/Leucyl/Isoleucyl-tRNA synthetase, class I, anticodon-binding
67441	2		nf	n/a	nf	n/a	nf	nf
82648	2		195597	5.40E-12	NP_378669.1	9.80E-10	ribosomal protein L36	PF01158: Ribosomal protein L36E
69163	2		58843	2.00E-110	NP_005897.1	4.50E-108	male germ cell-associated kinase	PF00069: Protein kinase, core
80491	2		nf	n/a	nf	n/a	nf	nf
55929	2		127904	1.40E-178	NP_005989.3	7.00E-174	chaperonin containing TCP1, subunit 3 isoform a	PF00118: Chaperonin Cpn60/TCP-1
70276	2		nf	n/a	NP_115509.4	8.30E-06	Myc-binding protein-associated protein	nf
58709	2		195595	3.80E-11	NP_000986.2	4.90E-08	ribosomal protein L34	PF01199: Ribosomal protein L34e
37563	2		128522	3.00E-82	NP_006316.1	1.80E-90	ras-related nuclear protein	PF00071: Ras
75565	2		187517	6.60E-24	NP_003986.2	7.50E-24	natriuretic peptide receptor B precursor	PF00989: PAS fold   PF00211: Adenylyl cyclase class-3/4/guanylyl cyclase
81386	2		nf	n/a	NP_116164.2	4.30E-08	kelch-like	PF00651: BTB/POZ
82381	2		402056	1.40E-17	NP_006357.1	1.70E-25	adenylyl cyclase-associated protein 2	PF08603: Adenylate cyclase-associated CAP, C-terminal
53883	2		140618	8.70E-108	NP_002037.2	1.30E-110	glyceraldehyde-3-phosphate dehydrogenase	PF00044: Glyceraldehyde 3-phosphate dehydrogenase   PF02800: Glyceraldehyde 3-phosphate dehydrogenase
64917	2		nf	n/a	nf	n/a	nf	nf
74744	2		nf	n/a	nf	n/a	nf	nf
56509	2		77886	2.20E-35	NP_776254.2	9.10E-37	MORN repeat containing 3	PF00225: Kinesin, motor region   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif
72250	2		58944	5.60E-49	NP_001018146.1	1.10E-47	NME1-NME2 protein	PF00334: Nucleoside diphosphate kinase, core
79690	2		395908	2.00E-07	NP_006624.2	3.20E-88	IQ motif containing GTPase activating protein 2	PF00307: Calponin-like actin-binding   PF00630: Filamin/ABP280 repeat   PF00630: Filamin/ABP280 repeat   PF00630: Filamin/ABP280 repeat   PF00630: Filamin/ABP280 repeat   PF00616: Ras GTPase-activating protein   PF03836: RasGAP protein, C-terminal

55664	2		126382	8.30E-190	NP_001684.2	1.40E-198	vacuolar H <sup>+</sup> ATPase B2	PF02874: ATPase, F1/V1/A1 complex, alpha/beta subunit, N-terminal   PF00006: ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding   PF00306: ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal
77798	2		376980	4.20E-122	NP_004981.2	7.80E-98	methionyl-tRNA synthetase	PF09334: Aminoacyl-tRNA synthetase, class I (M)   PF01588: t-RNA-binding region
64930	2	Yes	138046	1.10E-34	NP_114130.3	6.20E-35	radial spoke 3 homolog	PF06098: Radial spoke 3
46570	2		nf	n/a	nf	n/a	nf	nf
82706	2		188837	1.10E-79	NP_000998.1	1.10E-83	ribosomal protein S4, X-linked X isoform	PF08071: Ribosomal protein S4E, N-terminal   PF01479: RNA-binding S4   PF00900: Ribosomal protein S4E, central   PF00467: KOW
63500	2		342942	2.10E-19	NP_004046.2	1.70E-33	calpain 5	PF00515: Tetratricopeptide TPR_1   PF00648: Peptidase C2, calpain
82819	2		131113	6.20E-185	NP_036205.1	7.10E-181	chaperonin containing TCP1, subunit 5 (epsilon)	PF00118: Chaperonin Cpn60/TCP-1
82964	2		137286	3.00E-130	NP_001885.1	2.20E-122	casein kinase 1 epsilon	PF00069: Protein kinase, core
80991	2		184001	3.00E-178	NP_002794.1	3.40E-174	proteasome 26S ATPase subunit 2	PF00004: AAA ATPase, core
64979	2		143072	1.70E-17	NP_148982.1	1.80E-17	ribosomal protein S24 isoform a	PF01282: Ribosomal protein S24e
60612	2		24114	1.60E-83	NP_001092.1	2.50E-171	beta actin	PF00022: Actin/actin-like
59586	2		129025	6.20E-30	NP_002188.1	2.20E-266	aconitase 1	PF00330: Aconitate hydratase, N-terminal   PF00694: Aconitate hydratase, C-terminal
69699	2		380616	1.00E-120	NP_005947.3	1.30E-117	methylenetetrahydrofolate dehydrogenase 1	PF01268: Formate-tetrahydrofolate ligase, FTHFS
82759	2		195592	3.60E-38	NP_001014.1	3.90E-38	ribosomal protein S20 isoform 2	PF00338: Ribosomal protein S10
36339	2		188144	4.30E-35	NP_008819.1	4.60E-35	calmodulin 1	PF00036: Calcium-binding EF-hand   PF00036: Calcium-binding EF-hand
66723	2		nf	n/a	nf	n/a	nf	PF00612: IQ calmodulin-binding region   PF00612: IQ calmodulin-binding region   PF00612: IQ calmodulin-binding region
57212	2		nf	n/a	nf	n/a	nf	nf
29544	2		138117	1.70E-239	NP_031381.2	5.90E-225	heat shock 90kDa protein 1, beta	PF02518: ATP-binding region, ATPase-like   PF00183: Heat shock protein Hsp90
82184	2		177203	7.80E-36	NP_006400.2	3.00E-72	actin related protein 2/3 complex subunit 1A	PF00400: WD40 repeat   PF00400: WD40 repeat
59485	2		nf	n/a	NP_003881.2	1.70E-07	Fc fragment of IgG binding protein	nf
74041	2		nf	n/a	NP_004230.2	3.30E-11	thyroid hormone receptor interactor 11	nf
70498	2		nf	n/a	nf	n/a	nf	nf
80201	2		nf	n/a	NP_004921.1	1.70E-67	F-actin capping protein beta subunit	PF01115: F-actin capping protein, beta subunit
80999	2		127247	3.10E-25	NP_000987.2	2.60E-25	ribosomal protein L35a	PF01247: Ribosomal protein L35Ae
63284	2		nf	n/a	NP_001034669.1	6.90E-06	aster-associated protein	nf
36352	2		196616	8.10E-192	nf	n/a	nf	PF01326: Pyruvate phosphate dikinase, PEP/pyruvate-binding   PF00391: PEP-utilising enzyme, mobile region   PF02896: PEP-utilizing enzyme
66239	2		188056	4.10E-09	XP_001722448.1	1.80E-32	PREDICTED: hypothetical protein LOC285754	nf
59818	2		185104	5.60E-134	NP_004989.2	1.40E-239	myosin IE	PF00063: Myosin head, motor region   PF06017: Myosin tail 2   PF00018: Src homology-3
69688	2	Yes	295278	1.00E-14	NP_115963.1	4.70E-13	dpy-30-like protein	PF05186: Dpy-30
59838	2		129593	1.00E-152	NP_000678.1	7.70E-129	adenosylhomocysteinase isoform 1	PF05221: S-adenosyl-L-homocysteine hydrolase   PF00670: S-adenosyl-L-homocysteine hydrolase, NAD binding
38656	2		155464	3.10E-64	NP_005800.3	1.00E-71	peroxiredoxin 2 isoform a	PF00578: Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen   PF10417: C-terminal domain of 1-Cys peroxiredoxin
70153	2		24114	1.10E-27	NP_005713.1	1.90E-31	actin-related protein 2 isoform b	PF00022: Actin/actin-like
82072	1		149311	1.70E-22	NP_004805.2	7.70E-10	WD repeat domain 57 (U5 snRNP specific)	PF00400: WD40 repeat   PF00400: WD40 repeat   PF00400: WD40 repeat
49131	1		nf	n/a	nf	n/a	nf	nf
66413	1		nf	n/a	nf	n/a	nf	nf
79795	1		nf	n/a	nf	n/a	nf	PF02845: Ubiquitin system component Cue
63304	1	Yes	158948	3.60E-78	NP_079039.3	2.10E-78	WD repeat domain 78 isoform 1	PF00400: WD40 repeat
78218	1		nf	n/a	NP_620073.2	2.10E-32	guanine nucleotide binding protein, alpha activating polypeptide O isoform b	PF00503: Guanine nucleotide binding protein (G-protein), alpha subunit
80282	1		nf	n/a	NP_001019837.1	1.20E-12	SH3-domain kinase binding protein 1 isoform b	PF00018: Src homology-3   PF00018: Src homology-3
70994	1		97466	2.40E-54	NP_004818.2	1.50E-40	ATP-binding cassette, subfamily G, member 2	PF00005: ABC transporter related   PF01061: ABC-2 type transporter
56722	1		nf	n/a	nf	n/a	nf	PF03114: BAR
59646	1		154307	1.80E-132	nf	n/a	nf	PF08267: Cobalamin (vitamin B12)-independent methionine synthase MetE, N-terminal   PF01717: Methionine synthase, vitamin-B12 independent

82749	1		185012	9.10E-10	NP_077747.1	9.70E-10	AMP-activated protein kinase gamma2 subunit isoform b	PF00571: Cystathionine beta-synthase, core
81242	1		nf	n/a	NP_001014436.1	2.20E-08	drebrin-like isoform b	PF00241: Actin-binding, cofilin/tropomyosin type
81102	1		nf	n/a	nf	n/a	nf	nf
57099	1		159792	7.10E-10	NP_001106673.1	7.50E-12	nascent polypeptide-associated complex alpha subunit isoform b	PF01849: Nascent polypeptide-associated complex NAC
52995	1		nf	n/a	nf	n/a	nf	nf
35679	1		196624	1.80E-17	NP_002618.1	1.00E-13	phosphofructokinase, platelet	PF00365: Phosphofructokinase
49466	1		nf	n/a	NP_001106678.1	4.40E-07	putative thiosulfate sulfurtransferase KAT isoform 1	PF00581: Rhodanese-like
59934	1		144027	2.10E-12	NP_001035982.1	6.10E-46	diaphanous homolog 3 isoform a	PF06367: Diaphanous FH3   PF02181: Actin-binding FH2
59701	1		81593	2.70E-49	NP_002801.1	2.00E-44	proteasome 26S non-ATPase subunit 4	PF02809: Ubiquitin interacting motif
79871	1		nf	n/a	nf	n/a	nf	nf
78090	1		106571	6.50E-26	NP_001093167.1	1.60E-41	radial spoke head 10 homolog B2	PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif
56657	1		206446	2.50E-09	NP_055970.1	1.20E-24	La ribonucleoprotein domain family, member 4B	PF05383: RNA-binding protein Lupus La
63046	1		nf	n/a	nf	n/a	nf	nf
80379	1		nf	n/a	nf	n/a	nf	nf
59416	1		54608	4.20E-227	NP_001681.2	1.80E-223	ATPase, H+ transporting, lysosomal V1 subunit A	PF02874: ATPase, F1/V1/A1 complex, alpha/beta subunit, N-terminal   PF00006: ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding   PF00306: ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal
79158	1		196289	6.70E-23	NP_006338.1	1.10E-18	peptidylprolyl isomerase H	PF00646: Cyclin-like F-box   PF00160: Peptidyl-prolyl cis-trans isomerase, cyclophilin-type
81508	1		nf	n/a	nf	n/a	nf	nf
78159	1		154370	2.30E-07	XP_001719981.1	7.80E-08	PREDICTED: hypothetical protein, partial	nf
66586	1		nf	n/a	nf	n/a	nf	PF02204: Vacuolar sorting protein 9
71918	1		nf	n/a	nf	n/a	nf	nf
66323	1		nf	n/a	nf	n/a	nf	nf
80683	1		195601	1.30E-40	NP_001012.1	4.50E-37	ribosomal protein S17	PF00833: Ribosomal protein S17e
59279	1		142494	1.90E-08	NP_689540.2	1.30E-64	adenylate kinase 7	PF05186: Dpy-30
63128	1		196311	3.40E-43	NP_005613.2	3.80E-56	SA hypertension-associated homolog isoform 1	PF00501: AMP-dependent synthetase and ligase
75988	1		147457	4.30E-18	nf	n/a	nf	PF00211: Adenylyl cyclase class-3/4/guanylyl cyclase
55119	1		302551	1.30E-47	NP_000849.1	2.50E-50	guanylate kinase 1 isoform b	PF00625: Guanylate kinase
79455	1		nf	n/a	nf	n/a	nf	nf
57833	1		117918	1.60E-14	NP_997002.1	5.00E-08	leucine rich repeat containing 20 isoform 1	PF00560: Leucine-rich repeat   PF00560: Leucine-rich repeat   PF00560: Leucine-rich repeat
55855	1		132486	1.10E-139	NP_002711.1	2.00E-146	protein phosphatase 4, catalytic subunit	PF00149: Metallophosphoesterase
46605	1	Yes	196748	9.30E-19	NP_203528.2	1.30E-34	AAT1-alpha	nf
80103	1		60490	2.10E-17	NP_036381.2	9.30E-37	ras-like protein TC10	PF00071: Ras
59104	1		126059	1.80E-59	NP_001012321.1	1.20E-61	ribosomal protein SA	PF00318: Ribosomal protein S2
67954	1		nf	n/a	NP_001395.1	6.30E-15	eukaryotic translation elongation factor 1 gamma	PF00647: Translation elongation factor EF1B, gamma chain, conserved
78187	1		nf	n/a	NP_079413.3	5.40E-34	spatacsin isoform 1	nf
52303	1		283187	9.40E-06	NP_002896.2	4.50E-14	retinol dehydrogenase 5 (11-cis and 9-cis)	nf
80881	1		nf	n/a	NP_859067.2	5.60E-23	proline rich 6	PF04828: Glutathione-dependent formaldehyde-activating, GFA
63912	1		400848	1.50E-34	NP_005709.1	1.70E-53	actin related protein 2/3 complex subunit 4 isoform a	PF05856: ARP23 complex 20 kDa subunit
49678	1		nf	n/a	nf	n/a	nf	PF00036: Calcium-binding EF-hand
81767	1		183518	1.70E-31	NP_000961.2	1.30E-23	ribosomal protein L6	PF01159: Ribosomal protein L6E
36459	1		129742	8.30E-16	NP_001015.1	1.30E-14	ribosomal protein S21	PF01249: Ribosomal protein S21e
71952	1		nf	n/a	nf	n/a	nf	nf
72942	1		342942	2.30E-12	NP_008989.2	2.00E-18	calpain 11	PF00648: Peptidase C2, calpain
78714	1		134846	5.00E-42	NP_000101.2	1.3E-314	dihydropyrimidine dehydrogenase isoform 1	PF07992: FAD-dependent pyridine nucleotide-disulphide oxidoreductase   PF01180: Dihydroorotate dehydrogenase, core   PF00037: 4Fe-4S ferredoxin, iron-sulphur binding   PF00037: 4Fe-4S ferredoxin, iron-sulphur binding
59622	1		115373	2.90E-13	NP_057315.3	2.60E-13	nemo like kinase	PF07714: Tyrosine protein kinase
50704	1		nf	n/a	nf	n/a	nf	nf
60003	1		284442	4.40E-15	NP_115821.2	2.60E-39	multiple EGF-like-domains 11	PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF00053: EGF-like, laminin   PF00053: EGF-like, laminin   PF02010: PKD/REJ-like protein
45594	1		194382	1.00E-159	NP_036558.3	2.90E-287	splicing factor 3b, subunit 3	PF03178: CPSF A subunit, C-terminal
63520	1		nf	n/a	nf	n/a	nf	nf
71766	1		nf	n/a	nf	n/a	nf	nf





4601	1	Yes	195877	5.70E-17	NP_073614.1	4.90E-24	RAB, member RAS oncogene family-like 5 isoform a	nf
83014	1		137945	9.00E-193	NP_002799.3	8.90E-206	proteasome 26S non-ATPase subunit 2	PF01851: Proteasome/cyclosome, regulatory subunit   PF01851: Proteasome/cyclosome, regulatory subunit   PF01851: Proteasome/cyclosome, regulatory subunit   PF01851: Proteasome/cyclosome, regulatory subunit   PF01851: Proteasome/cyclosome, regulatory subunit
49382	1		71905	8.80E-08	NP_115993.3	6.40E-18	abhydrolase domain-containing protein 1	nf
61990	1		147457	1.60E-20	NP_997555.1	7.40E-09	GRIP1 associated protein 1 isoform 2	PF00211: Adenylyl cyclase class-3/4/guanylyl cyclase
48710	1		393582	1.40E-146	NP_006286.1	1.30E-190	valyl-tRNA synthetase	PF00133: Aminoacyl-tRNA synthetase, class Ia   PF08264: Valyl/Leucyl/Isoleucyl-tRNA synthetase, class I, anticodon-binding
62706	1		nf	n/a	NP_003557.2	1.50E-07	early endosome antigen 1, 162kD	PF00169: Pleckstrin-like   PF02759: RUN
51352	1		nf	n/a	NP_981932.1	5.60E-07	iodotyrosine dehalogenase 1	PF00881: Nitroreductase
36164	1		141272	3.60E-158	NP_056988.3	5.70E-168	eukaryotic translation initiation factor 5B	PF00009: Protein synthesis factor, GTP-binding   PF03144: Translation elongation factor EFTu/EF1A, domain 2
45945	1		nf	n/a	nf	n/a	nf	nf
78582	1		nf	n/a	nf	n/a	nf	nf
79908	1		172711	2.30E-12	NP_116200.2	1.00E-17	transforming growth factor beta regulator 1	PF05964: FY-rich, N-terminal   PF05965: FY-rich, C-terminal
45355	1		nf	n/a	nf	n/a	nf	nf
52250	1		150075	2.80E-09	NP_777612.1	9.70E-15	hypothetical protein LOC285555	nf
380	1	Yes	189109	5.20E-94	NP_852091.1	4.40E-101	coiled-coil domain containing 39	nf
82341	1		155455	3.00E-34	NP_000981.1	1.10E-31	ribosomal protein L27a	PF00256: Ribosomal protein L15
63901	1		nf	n/a	nf	n/a	nf	PF01063: Aminotransferase, class IV
81857	1		nf	n/a	NP_060795.1	2.10E-15	hypothetical protein LOC203228 isoform a	nf
45002	1	Yes	98642	7.80E-36	NP_060480.1	1.30E-46	estrogen-related receptor beta like 1	PF10498: Intra-flagellar transport protein 57
32656	1		196947	8.00E-98	NP_004659.2	9.10E-126	maltase-glucoamylase	PF01055: Glycoside hydrolase, family 31
80100	1		nf	n/a	nf	n/a	nf	nf
32474	1		287617	1.80E-80	NP_036450.1	7.00E-64	leucine zipper-EF-hand containing transmembrane protein 1	PF07766: LETM1-like   PF00036: Calcium-binding EF-hand
66832	1		168908	1.70E-54	NP_001034842.2	2.50E-71	echinoderm microtubule associated protein like 6	PF00400: WD40 repeat   PF00400: WD40 repeat
39446	1		191758	2.80E-24	NP_000985.1	1.50E-22	ribosomal protein L32	PF01655: Ribosomal protein L32e
63764	1	Yes	183240	1.4e-310	NP_056477.1	1.30E-298	selective LIM binding factor homolog	nf
64367	1		nf	n/a	nf	n/a	nf	nf
70007	1		nf	n/a	NP_002060.4	2.80E-30	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1	PF00503: Guanine nucleotide binding protein (G-protein), alpha subunit
66643	1	Yes	114106	3.70E-45	NP_002517.1	7.20E-35	5' nucleotidase, ecto	PF00149: Metallophosphoesterase   PF02872: 5'-Nucleotidase, C-terminal
50077	1		nf	n/a	nf	n/a	nf	nf
83081	1		nf	n/a	nf	n/a	nf	nf
36700	1		131550	6.10E-18	NP_005125.1	2.70E-63	protein phosphatase 4, regulatory subunit 1 isoform b	PF02985: HEAT   PF02985: HEAT   PF02985: HEAT   PF02985: HEAT   PF02985: HEAT
60892	1		421037	8.40E-39	NP_742128.1	8.70E-34	potassium voltage-gated channel, shaker-related subfamily, beta member 2 isoform 2	PF00248: Aldo/keto reductase
79521	1		145485	1.30E-19	NP_056377.1	9.80E-29	UBX domain containing 7	PF02809: Ubiquitin interacting motif   PF00789: UBX
54755	1		140045	8.90E-67	NP_066289.2	1.60E-66	ubiquitin C	PF00240: Ubiquitin   PF00240: Ubiquitin
79700	1		148829	1.30E-96	NP_006411.2	1.30E-181	ADP-ribosylation factor guanine nucleotide-exchange factor 2	PF01369: SEC7-like   PF09324: Protein of unknown function DUF1981, SEC7 associated
71591	1		nf	n/a	NP_976328.2	8.10E-16	E2F transcription factor 7	PF02319: Transcription factor E2F/dimerisation partner (TDP)   PF02319: Transcription factor E2F/dimerisation partner (TDP)
70051	1	Yes	193577	2.50E-31	NP_872383.1	3.50E-11	outer dense fiber of sperm tails 3-like 2	nf
63598	1		nf	n/a	nf	n/a	nf	nf
78148	1		419259	2.30E-33	nf	n/a	nf	PF00082: Peptidase S8 and S53, subtilisin, kexin, sedolisin
80259	1	Yes	130394	1.20E-20	NP_057092.2	1.50E-29	dynein 2 light intermediate chain isoform 1	nf
73968	1		297357	8.60E-14	NP_114162.2	3.40E-18	tetratricopeptide repeat domain 29	nf
49224	1		nf	n/a	nf	n/a	nf	nf

63235	1		284442	6.40E-21	NP_002151.2	7.70E-49	tenascin C precursor	PF01436: NHL repeat   PF01436: NHL repeat   PF01436: NHL repeat   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular
59012	1		nf	n/a	NP_061897.1	9.70E-26	HEAT repeat containing 5B	nf
51044	1		nf	n/a	nf	n/a	nf	nf
78215	1		191319	2.10E-129	NP_001484.1	3.50E-133	GDP dissociation inhibitor 1	PF00996: Rab GTPase activator
79232	1	Yes	188612	1.70E-102	NP_075462.3	7.90E-111	dynein, axonemal, intermediate polypeptide 2	nf
29188	1	Yes	145396	2.40E-99	NP_653269.3	6.70E-98	WD repeat domain 66	nf
79626	1	Yes	185392	1.10E-23	NP_056465.2	9.00E-29	TNF receptor-associated factor 3 interacting protein 1 isoform 1	PF10243: Microtubule-binding protein MIP-T3
83175	1		162845	4.40E-42	NP_036555.1	5.80E-44	ribosomal protein L13a	PF00572: Ribosomal protein L13
80363	1		120423	1.50E-09	NP_004954.2	2.30E-06	guanylate cyclase 2C	PF08376: Nitrate and nitrite sensing   PF00211: Adenylyl cyclase class-3/4/guanylyl cyclase
83261	1		380549	1.80E-22	NP_055112.2	1.30E-53	ETHE1 protein	PF00753: Beta-lactamase-like
74980	1		nf	n/a	NP_006176.2	3.50E-09	nuclear mitotic apparatus protein 1	nf
34772	1		179025	7.10E-42	NP_258412.1	4.40E-39	inosine triphosphatase isoform a	PF01725: Ham1-like protein
78103	1		2332	4.80E-06	NP_620073.2	1.10E-63	guanine nucleotide binding protein, alpha activating polypeptide O isoform b	PF00503: Guanine nucleotide binding protein (G-protein), alpha subunit
52811	1		nf	n/a	nf	n/a	nf	nf
76504	1		120079	8.60E-08	NP_085096.1	2.20E-51	sirtuin 2 isoform 2	PF02146: Silent information regulator protein Sir2
47729	1		nf	n/a	nf	n/a	nf	nf
79803	1		nf	n/a	nf	n/a	nf	nf
82900	1		126528	3.20E-101	NP_037473.3	3.20E-107	Obg-like ATPase 1 isoform 1	PF01926: GTP-binding protein, HSR1-related   PF06071: Protein of unknown function DUF933
74757	1		184661	8.50E-08	NP_037534.3	5.80E-09	fatty acid desaturase 1	PF00173: Cytochrome b5
80180	1		186878	7.50E-41	NP_001129521.2	7.20E-21	leishmanolysin-like isoform 1	PF01457: Peptidase M8, leishmanolysin
45058	1		24512	1.40E-58	NP_079143.3	4.70E-37	KPL2 protein isoform 1	nf
59665	1		78109	2.70E-10	NP_000984.1	1.10E-12	ribosomal protein L31 isoform 1	PF01198: Ribosomal protein L31e
67195	1		nf	n/a	nf	n/a	nf	nf
77359	1		nf	n/a	nf	n/a	nf	nf
80987	1		64643	3.60E-14	NP_751943.1	8.80E-14	kelch domain containing 1	PF01344: Kelch repeat type 1   PF01344: Kelch repeat type 1   PF01344: Kelch repeat type 1   PF07646: Kelch repeat type 2   PF01344: Kelch repeat type 1
69338	1		nf	n/a	nf	n/a	nf	nf
73610	1		nf	n/a	nf	n/a	nf	nf
65812	1		nf	n/a	nf	n/a	nf	nf
56351	1		159554	4.00E-55	NP_004057.1	5.40E-57	centrin 1	PF00036: Calcium-binding EF-hand   PF00036: Calcium-binding EF-hand   PF00036: Calcium-binding EF-hand   PF00036: Calcium-binding EF-hand
81954	1		nf	n/a	nf	n/a	nf	nf
81194	1		nf	n/a	nf	n/a	nf	nf
78039	1		nf	n/a	NP_071345.2	2.20E-09	PBX/knotted 1 homeobox 2	nf
34640	1		174103	8.40E-55	NP_066952.1	3.00E-12	pyrophosphatase 1	PF00719: Inorganic pyrophosphatase
80494	1		194063	1.00E-97	NP_078836.1	1.80E-145	acyl-CoA synthetase short-chain family member 3	PF00501: AMP-dependent synthetase and ligase
81635	1		nf	n/a	nf	n/a	nf	nf
80249	1		190633	1.50E-13	NP_078934.3	3.40E-71	importin 4	PF03810: Importin-beta, N-terminal   PF02985: HEAT   PF02985: HEAT   PF02985: HEAT   PF02985: HEAT
78612	1		141319	4.20E-147	NP_001055.1	3.00E-30	transketolase isoform 1	PF00456: Transketolase, N-terminal   PF02779: Transketolase, central region   PF02780: Transketolase, C-terminal
78847	1		421068	1.80E-126	NP_004651.2	6.50E-124	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	PF00270: DNA/RNA helicase, DEAD/DEAH box type, N-terminal   PF00271: DNA/RNA helicase, C-terminal
79189	1		195584	1.70E-72	NP_006004.2	9.70E-74	ribosomal protein L10	PF00252: Ribosomal protein L16
81288	1		nf	n/a	NP_004300.1	1.30E-28	Rho GDP dissociation inhibitor (GDI) alpha	PF02115: RHO protein GDP dissociation inhibitor
50947	1		nf	n/a	NP_055111.1	1.20E-07	calpain 7	nf
34873	1		181887	5.20E-07	nf	n/a	nf	PF00235: Profilin/allergen
77826	1		nf	n/a	NP_056000.1	8.60E-14	Cdc42 guanine exchange factor 9	nf
71179	1		nf	n/a	nf	n/a	nf	nf
78273	1		82172	1.00E-17	NP_000994.1	1.30E-14	ribosomal protein P1 isoform 1	PF00428: Ribosomal protein 60S
55791	1		140045	2.10E-74	NP_066289.2	2.10E-151	ubiquitin C	PF00240: Ubiquitin   PF00240: Ubiquitin   PF00240: Ubiquitin   PF00240: Ubiquitin
75436	1		nf	n/a	nf	n/a	nf	nf

59637	1	Yes	128114	3.90E-144	NP_659491.4	1.30E-135	WD40-repeat protein upregulated in HCC isoform b	PF00400: WD40 repeat   PF00400: WD40 repeat   PF00400: WD40 repeat
61455	1		189584	8.10E-06	nf	n/a	nf	PF02197: cAMP-dependent protein kinase regulator, type II PKA R subunit
71175	1		400743	6.70E-19	NP_849172.2	4.70E-26	NOD3 protein	PF00560: Leucine-rich repeat
54734	1		188942	1.20E-46	NP_001958.2	1.30E-48	eukaryotic translation initiation factor 4A2	PF00270: DNA/RNA helicase, DEAD/DEAH box type, N-terminal
51975	1		nf	n/a	nf	n/a	nf	nf
38344	1		376728	2.50E-117	NP_003671.1	4.10E-43	tyrosyl-tRNA synthetase	PF00579: Aminoacyl-tRNA synthetase, class Ib
37722	1		191903	2.40E-224	NP_695012.1	3.30E-270	ubiquitin-activating enzyme E1	PF00899: UBA/THIF-type NAD/FAD binding fold   PF00899: UBA/THIF-type NAD/FAD binding fold   PF10585: Ubiquitin-activating enzyme active site   PF02134: Ubiquitin-activating enzyme repeat   PF02134: Ubiquitin-activating enzyme repeat   PF09358: Ubiquitin-activating enzyme e1 C-terminal domain
79747	1		nf	n/a	NP_758953.1	2.40E-12	hypothetical protein LOC145376	nf
51126	1		nf	n/a	nf	n/a	nf	nf
60048	1		195585	2.50E-55	NP_009035.3	2.50E-59	ribosomal protein L10a	PF00687: Ribosomal protein L1
68988	1		nf	n/a	nf	n/a	nf	nf
66818	1		nf	n/a	NP_149101.1	1.30E-16	AIG2-like domain 1	PF06094: AIG2-like
69743	1		nf	n/a	nf	n/a	nf	nf
53120	1		nf	n/a	NP_001129396.1	1.50E-12	von Willebrand factor D and EGF domains	PF00188: SCP-like extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular   PF07974: EGF, extracellular
81941	1		193661	7.80E-20	NP_055300.1	3.10E-13	prostaglandin-D synthase	PF02798: Glutathione S-transferase, N-terminal   PF00043: Glutathione S-transferase, C-terminal
36951	1		184170	1.20E-37	NP_002873.1	3.20E-36	RAN binding protein 1	PF00638: Ran Binding Protein 1
73592	1		nf	n/a	nf	n/a	nf	nf
73783	1		392133	2.90E-08	nf	n/a	nf	nf
39045	1		191776	1.00E-74	NP_001000.2	2.00E-73	ribosomal protein S5	PF00177: Ribosomal protein S7
78491	1		420861	6.70E-149	NP_055568.3	3.60E-151	BMS1-like, ribosome assembly protein	PF08142: AARP2CN   PF04950: Protein of unknown function DUF663
67649	1		121702	2.60E-11	NP_001128582.1	8.90E-10	DnaJ (Hsp40) homolog, subfamily A, member 3 isoform 2	PF00226: Heat shock protein DnaJ, N-terminal
71354	1		nf	n/a	nf	n/a	nf	nf
55551	1		184847	8.00E-82	NP_002781.2	6.20E-79	proteasome alpha 5 subunit	PF10584: Proteasome subunit A N-terminal signature   PF00227: 20S proteasome, A and B subunits
44967	1	Yes	152806	9.70E-38	NP_060835.1	1.20E-41	meiosis-specific nuclear structural 1	nf
64053	1	Yes	138649	1.90E-18	NP_113661.2	1.20E-29	intraflagellar transport 81-like isoform 2	nf
82821	1		114649	3.20E-85	NP_003959.3	1.40E-99	ubiquitin-activating enzyme 3 isoform 1	PF00899: UBA/THIF-type NAD/FAD binding fold   PF02134: Ubiquitin-activating enzyme repeat   PF08825: E2 binding
56360	1		152438	5.30E-14	NP_000837.3	4.60E-12	glutathione S-transferase alpha 2	PF02798: Glutathione S-transferase, N-terminal   PF00043: Glutathione S-transferase, C-terminal
70637	1		nf	n/a	NP_001035197.1	1.60E-11	mutL homolog 3 isoform 1	PF08676: MutL, C-terminal, dimerisation
61077	1		nf	n/a	nf	n/a	nf	PF00651: BTB/POZ
66714	1		nf	n/a	nf	n/a	nf	nf
67163	1		195613	1.30E-17	NP_003320.2	4.30E-23	thioredoxin	PF00085: Thioredoxin domain   PF00789: UBX
73943	1		347422	9.30E-13	NP_006401.3	1.60E-17	HIV-1 Tat interactive protein 2, 30kDa isoform b	nf
75841	1		nf	n/a	nf	n/a	nf	nf
59795	1		nf	n/a	NP_001034200.1	7.70E-14	tripartite motif-containing 71	PF00643: Zinc finger, B-box   PF01436: NHL repeat   PF01436: NHL repeat   PF01436: NHL repeat   PF01436: NHL repeat
83064	1	Yes	77703	4.90E-27	NP_056468.1	1.20E-29	RIB43A domain with coiled-coils 2	PF05914: RIB43A
62763	1		293546	3.40E-35	NP_742128.1	1.60E-32	potassium voltage-gated channel, shaker-related subfamily, beta member 2 isoform 2	PF00248: Aldo/keto reductase
63237	1		nf	n/a	nf	n/a	nf	nf
58925	1		nf	n/a	NP_055191.2	3.40E-165	cytoplasmic FMR1 interacting protein 2	PF05994: Cytoplasmic fragile X mental retardation protein interacting protein
54368	1		120324	5.20E-13	NP_057062.1	2.10E-13	TNNI3 interacting kinase isoform b	PF07714: Tyrosine protein kinase
79448	1		126189	3.80E-123	NP_000958.1	1.40E-127	ribosomal protein L3 isoform a	PF00297: Ribosomal protein L3
65473	1		nf	n/a	NP_001129955.1	1.60E-06	hypothetical protein LOC284067	nf
62234	1		nf	n/a	nf	n/a	nf	nf
76540	1		196612	6.70E-36	nf	n/a	nf	PF01293: Phosphoenolpyruvate carboxykinase, ATP-utilising
69737	1		nf	n/a	NP_653303.2	3.50E-08	hypothetical protein LOC149499	nf
79230	1		nf	n/a	nf	n/a	nf	nf
56305	1		127172	1.60E-06	nf	n/a	nf	PF01459: Porin, eukaryotic type
80937	1		nf	n/a	nf	n/a	nf	PF00646: Cyclin-like F-box
69783	1		nf	n/a	nf	n/a	nf	nf
72989	1		114898	1.60E-07	NP_115737.1	2.00E-07	THO complex 3	PF00400: WD40 repeat   PF00400: WD40 repeat
38811	1		320624	1.80E-13	NP_057216.2	2.60E-11	myelin expression factor 2	PF00076: RNA recognition motif, RNP-1
76059	1		nf	n/a	nf	n/a	nf	nf
65386	1		nf	n/a	nf	n/a	nf	nf

63280	1	Yes	24421	4.70E-132	NP_006522.2	3.00E-127	intraflagellar transport 88 homolog isoform 2	PF00515: Tetratricopeptide TPR_1   PF00515: Tetratricopeptide TPR_1   PF00515: Tetratricopeptide TPR_1   PF00515: Tetratricopeptide TPR_1   PF00515: Tetratricopeptide TPR_1
62959	1	Yes	192170	1.20E-87	NP_065930.2	1.40E-51	coiled-coil domain containing 146	nf
45350	1		413646	2.00E-18	NP_000050.2	5.50E-51	breast cancer 2, early onset	PF09169: BRCA2, helical   PF09103: BRCA2, oligonucleotide/oligosaccharide-binding 1
80902	1		nf	n/a	nf	n/a	nf	nf
79216	1		nf	n/a	nf	n/a	nf	nf
62660	1		60490	6.50E-16	NP_001782.1	2.50E-43	cell division cycle 42 isoform 1	PF00071: Ras
67392	1		nf	n/a	nf	n/a	nf	nf
78419	1		380622	4.90E-120	NP_001986.2	1.20E-114	acyl-CoA synthetase long-chain family member 1	PF00501: AMP-dependent synthetase and ligase
67691	1		nf	n/a	NP_079228.4	8.60E-25	hypothetical protein LOC80017 isoform a	nf
88119	1		283552	1.50E-40	NP_060887.2	2.40E-23	adenylate cyclase 10	PF00211: Adenylyl cyclase class-3/4/guanylyl cyclase
58163	1		nf	n/a	nf	n/a	nf	PF02205: Actin-binding WH2
61418	1		77886	1.40E-35	NP_776254.2	5.50E-37	MORN repeat containing 3	PF00225: Kinesin, motor region   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF03671: Ubiquitin-like, Ufm1
81993	1		185792	2.50E-19	NP_543136.1	3.10E-14	testis-specific gene A2	PF04969: CS   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif   PF02493: MORN motif
59962	1		378877	5.90E-22	NP_005498.1	1.20E-18	cofilin 1 (non-muscle)	PF00241: Actin-binding, cofilin/tropomyosin type
33131	1		195599	2.30E-43	NP_000963.1	1.10E-51	ribosomal protein L7a	PF01248: Ribosomal protein L7Ae/L30e/S12e/Gadd45
52838	1		188144	3.90E-14	NP_008819.1	2.10E-14	calmodulin 1	nf
80067	1		nf	n/a	nf	n/a	nf	nf
63091	1	Yes	107835	9.20E-33	NP_659484.3	5.60E-46	oxidored-nitro domain-containing protein isoform 1	PF10188: Organic solute transport protein 1
59552	1		nf	n/a	nf	n/a	nf	nf
35893	1		152525	1.70E-33	NP_003330.1	6.30E-31	ubiquitin-conjugating enzyme E2D 2 isoform 1	PF00179: Ubiquitin-conjugating enzyme, E2
82965	1		206048	4.80E-52	NP_003784.2	8.00E-63	cathepsin F	PF08246: Proteinase inhibitor I29, cathepsin propeptide   PF00112: Peptidase C1A, papain C-terminal
73911	1		106046	1.40E-31	NP_000897.3	9.20E-37	natriuretic peptide receptor 1	PF00989: PAS fold   PF00211: Adenylyl cyclase class-3/4/guanylyl cyclase
88122	1		81259	5.30E-08	NP_001782.1	6.50E-29	cell division cycle 42 isoform 1	PF00071: Ras
29287	1		26265	6.40E-57	NP_000356.1	6.20E-56	triosephosphate isomerase 1 isoform 1	PF00121: Triosephosphate isomerase
39155	1		377723	2.50E-85	NP_075266.1	7.00E-103	acyl-CoA synthetase long-chain family member 4 isoform 2	PF00501: AMP-dependent synthetase and ligase
44726	1		206042	7.70E-116	NP_006829.1	9.50E-115	methionyl aminopeptidase 2	PF00557: Peptidase M24, catalytic core
34303	1		182183	9.80E-13	NP_036465.2	2.60E-16	c-myc binding protein	nf
74109	1		nf	n/a	nf	n/a	nf	PF00043: Glutathione S-transferase, C-terminal
66740	1		nf	n/a	nf	n/a	nf	nf
53758	1		nf	n/a	nf	n/a	nf	nf
35026	1		184701	3.10E-167	NP_006422.1	2.70E-165	chaperonin containing TCP1, subunit 2	PF00118: Chaperonin Cpn60/TCP-1
33387	1		24114	1.50E-80	NP_001605.1	1.40E-154	actin, gamma 1 propeptide	PF00022: Actin/actin-like
81437	1		nf	n/a	nf	n/a	nf	nf
59723	1		nf	n/a	XP_002344344.1	1.50E-07	PREDICTED: hypothetical protein, partial	nf
82086	1		nf	n/a	nf	n/a	nf	nf
55800	1		205854	5.70E-56	NP_001016.1	9.40E-60	ribosomal protein S23	PF00164: Ribosomal protein S12/S23
73065	1		nf	n/a	nf	n/a	nf	nf

Table S3. Orthologs of pCCCs in different eukaryotes

Common name	Species	POC11		Friggin/MOT37		B9D2		SSA3		FOR20/MOT52	
		Accession	Length	Accession	Length	Accession	Length	Accession	Length	Accession	Length
Human	<i>Homo sapiens</i>	Q9BR77	488	Q8IYG6	542	NP_085055	175	NP_001128493	381	NP_653201	174
Mouse	<i>Mus musculus</i>	NP_080304	489	NP_722472	552	NP_742160	175	NP_659166	381	NP_079621	174
Frog	<i>Xenopus laevis</i>	NP_001086567	496	nf	n/a	NP_001085984	176	NP_001084770	376	NP_001090045	169
Zebra fish	<i>Danio rerio</i>	NP_001020649.1	427	XP_692700	569	NP_001002394	175	nf	n/a	XP_696350	146
Sea anemone	<i>Nematostella vectensis</i>	EDO34896	453	XP_001641243	446	XP_001638897	175	XP_001623228	310	XP_001633880	168
Trichoplax	<i>Trichoplax adherins</i>	XP_002108951	450	XP_002108579	557	XP_002109118	175	XP_002114557	199	XP_002108148	169
Choanoflagellate	<i>Monosiga brevicola</i>	XP_001743324	528	XP_001746946	717	XP_001742148	175	nf	n/a	nf	n/a
Naegleria	<i>Naegleria gruberi</i>	EFC41717	387	XP_002683426	613	XP_002681671	181	XP_002675911	290	XP_002671358	335
Green algae	<i>Chlamydomonas reinhardtii</i>	XP_001693122	451	XP_001690181	735	XP_001693128	176	XP_001697124	699	XP_001697797	133
Paramecium	<i>Paramecium tetraurelia</i>	CAK72449	502	XP_001442263	545	XP_001432714	183	XP_001427368	405	XP_001449123	165
Tetrahymena	<i>Tetrahymena thermophila</i>	XP_001023528	511	XP_001017079	865	XP_001010073	188	XP_001033010	720	XP_001032688	172

Species names, NCBI accessions, and protein lengths (number of amino acids) for each centrosome-localized pCCC are as indicated.

nf indicates no clear ortholog was identified in the given organism.