Human Subtelomeric WASH Genes Encode a New Subclass of the WASP Family

Elena V. Linardopoulou[1], Sean S. Parghi[1], Cynthia Friedman[1], Gregory E. Osborn[2], Susan M. Parkhurst[2], Barbara J. Trask[1*]

1 Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, 2 Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America

Subtelomeres are duplication-rich, structurally variable regions of the human genome situated just proximal of telomeres. We report here that the most terminally located human subtelomeric genes encode a previously unrecognized third subclass of the Wiskott-Aldrich Syndrome Protein family, whose known members reorganize the actin cytoskeleton in response to extracellular stimuli. This new subclass, which we call WASH, is evolutionarily conserved in species as diverged as _Entamoeba_. We demonstrate that WASH is essential in _Drosophila_. WASH is widely expressed in human tissues, and human WASH protein colocalizes with actin in filopodia and lamellipodia. The VCA domain of human WASH promotes actin polymerization by the Arp2/3 complex in vitro. WASH duplicated to multiple chromosomal ends during primate evolution, with highest copy number reached in humans, whose WASH repertoires vary. Thus, human subtelomeres are not genetic junkyards, and WASH's location in these dynamic regions could have advantageous as well as pathologic consequences.

Introduction

Human chromosome termini are unusual in their sequence content and frequency of structural rearrangement. Their ends are capped with telomeres that protect chromosomes against degradation and fusion [1] and that can exert a silencing effect on nearby genes [2,3]. Patchworks of large DNA segments duplicated on various subsets of chromosomes lie just proximal of telomeres [4]. These subtelomeric regions comprise less than 0.1% of the human genome, but account for over 40% of interchromosomal duplicates in the genome assembly that formed since human and chimpanzee diverged. Other lines of evidence further support the notion that subtelomeres are hotspots of DNA breaks and repair [5–7].

While subtelomeric dynamics might disrupt gene function, they can also fuel rapid changes in subtelomeric gene repertoires. Subtelomeres of yeast and the malaria parasite, _Plasmodium_, harbor genes with important roles in adaptive processes [8,9]. Human subtelomeric genes vary in copy number and chromosomal distribution and include odorant and cytokine receptors, homeodomain proteins, secretoglobins, and other genes of unknown function [4]. Demonstrations of expression from more than one subtelomeric location for several human genes [4] suggest that interindividual variation in subtelomeric gene repertoires might underlie some phenotypic differences among humans. Indeed, subtle chromosomal rearrangements involving subtelomeres and neighboring chromosome-specific regions are detected in 3%–5% of individuals with unexplained mental retardation or developmental disorders [10]. However, no essential gene has yet been identified within human subtelomeres.

Here, we focus on the most telomerically duplicated human genes (currently arbitrarily named MGC52000), which heretofore had unknown function. A truncated copy of MGC52000 was first annotated in the pseudoautosomal region of Xqter/Yqter and called CXYorf1 [11]. A murine ortholog, orf19, encoding a 475-amino acid protein with similarity to proteins of unknown function from _Drosophila melanogaster_ and _Caenorhabditis elegans_, was subsequently identified, but no motifs that would disclose its biological function were reported [12]. The _C. elegans_ ortholog (Y48E1B.1) was recently named _ddl-2_ for its involvement in life-span regulation and reported to have several proline-rich domains [13].

We now show that MGC52000 encodes a new member of the Wiskott-Aldrich Syndrome Protein (WASP) family and is conserved from _Entamoeba_ to human. Known WASP family members participate in cytoskeleton reorganization and signal transduction by acting as effectors of Rho-GTPases and polymerizing actin via the Arp2/3 complex [14,15]. They are involved in cell motility, phagocytosis, and cytokinesis in diverse processes such as embryogenesis, angiogenesis, inflammatory immune response, microbial infection, and cancer metastasis [16–18]. Five WASP family members are currently known in mammals, and they fall into two subclasses, WASP/N-WASP and SCAR/WAVE [16,18]. We have renamed the human subtelomeric MGC52000 genes WASH, for <u>W</u>iskott-<u>A</u>ldrich <u>S</u>yndrome Protein and <u>S</u>CAR <u>H</u>omolog.

Our comparative and phylogenetic analyses of the WASP

* To whom correspondence should be addressed. E-mail: btrask@fhcrc.org

Author Summary

Human subtelomeres are rearrangement-prone regions near chromosome ends. They are concentrations of large, recent interchromosomal duplications. Over half of subtelomeric sequences changed copy number or location since humans and chimpanzee diverged, and subtelomeric content varies greatly among humans. Despite this dynamic activity, subtelomeres contain genes. We report the discovery of genes defining a previously unrecognized third subclass of the Wiskott-Aldrich Syndrome protein (WASP) family within human subtelomeres. The known WASP family members reorganize actin structures in cells in response to various signals, thereby causing cells to change shape and/or move. Representatives of this newly identified subclass, called WASH, exist in many other species, even in *Entamoeba* and slime mold. Like other WASP family members, WASH colocalizes with actin at the cell periphery and promotes actin polymerization in vitro. Flies lacking WASH die before becoming adults, demonstrating that WASH is critical for survival, and its function is distinct from that of the two other WASP subclasses, Wasp and Scar. Identification of the WASH subclass opens the way for future elucidation of WASH's role in the life cycles of diverse organisms, the implications of human variation in WASH copy number, and the consequences of WASH's location in dynamic telomere-adjacent regions.

family reveal that WASH proteins define a new subclass in the WASP family. We experimentally confirm that human WASH protein colocalizes with actin in cells and promotes Arp2/3-dependent actin polymerization in vitro and that WASH is transcribed in most tissues. We show that the single *Drosophila* WASH ortholog is essential for viability. It is therefore remarkable that human genomes harbor multiple functional WASH paralogs in highly dynamic subtelomeric regions.

Materials and Methods

Details of materials and methods are provided in Text S1.

Results

Widespread Chromosomal Distribution and Expression of Human WASH Genes

Seven copies of WASH are evident in the latest human genome assembly (March 2006). One is in 2q13–14 at the ancestral telomere–telomere fusion site (2qFS), and the others lie at different chromosomal ends (Figure 1). In contrast, the mouse genome contains only one ortholog

(AJ304796) at an internal location on chromosome 17 (which we verified by FISH, not shown). One human cDNA in GenBank, BC048328, contains a full-length 468-aa open reading frame (ORF) with 86.3% amino-acid identity to the mouse ortholog. Human WASH genes have 11 exons and span 15 kbp, commencing in a 1-kb CpG island and ending within 5 kbp of the telomere array at four locations in the assembly (Figure 1). Only the 9p copy in the assembly encodes a full-length intact protein, while other copies are prematurely truncated by frame-shifts or in-frame stop codons. Three chromosome termini carry truncated copies lacking the first two exons, owing to past translocation events that transferred only the distal portion of the gene [4]. Spliced transcripts also lacking the proximal portion and containing a shorter ORF (264 aa) have been reported to GenBank (e.g., AY217347, Figure 1).

We detect spliced transcripts of human WASH by RT-PCR in all tissues tested using RT-PCR, Northern blot analysis, and a 72-tissue dot-blot (Figures S1–S3). According to publicly available array data [19], human WASH is expressed in a variety of tissues, with somewhat higher levels in blood cells and some areas of the brain.

WASH Is an Evolutionarily Conserved New WASP Family Member

In order to deduce the function of the WASH protein, we compared orthologous sequences that we identified from 21 diverse organisms (Figures 2, 3, and S4). The resulting multiple sequence alignments reveal high conservation of this protein among vertebrates. Regions of local high similarity throughout the protein, but especially in the C-terminal portion of the alignment, are apparent between predicted vertebrate WASH proteins and sequences of fly, worm, and even evolutionarily distant *Dictyostelium* and *Entamoeba*. Notably, one of the two *Entamoeba* WASH genes encodes a short protein, containing only the C-terminal portion, and is like the human short form, indicating its potential functional significance. Vertebrate WASH proteins have conserved predicted nuclear localization and export signals and sumoylation sites (Figure S5) suggesting a possible role in the nucleus.

The most highly conserved C-terminal regions of predicted WASH proteins show similarity to the actin-binding domain (WH2) of WASP family members (Figures 2, 3, and S4–S9). WASH proteins also have a proline-rich stretch (P), followed by the WH2 domain (V), a central region (C), and an acidic

Figure 1. Organization of WASH Loci in the Current Genome Assembly and Representative Transcripts

Only the 9p copy (bold) has a full-length intact ORF; others are partial or disrupted by frameshifts or in-frame stop codons. One copy assigned to 1p is more likely to be a variant allele of 19p [4]. Coding exons are numbered 1 through 10; thin bars represent non-coding exons. Sequence indicated in gray is shared only by Xq/Yq and 16p and lacks the N-terminal portion of WASH. Black and gray arrowheads indicate terminal and degenerate internal telomere-repeat arrays, respectively.

doi:10.1371/journal.pgen.0030237.g001
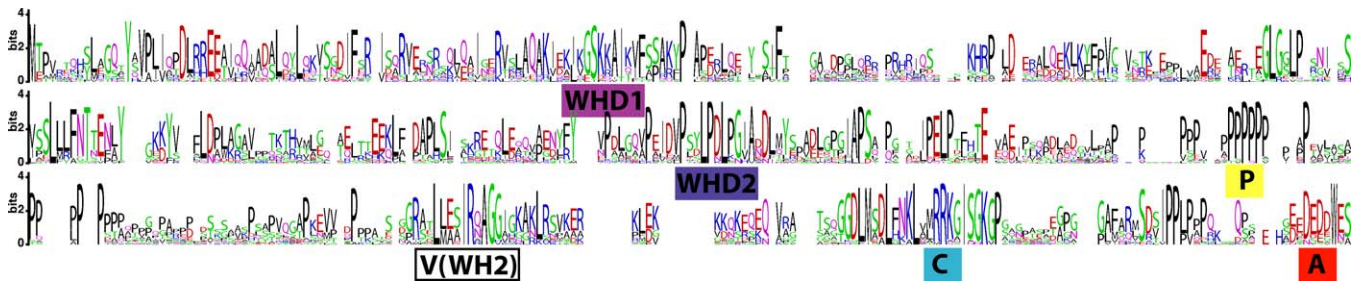
Figure 2. LogoPlot of Protein Alignment of WASH Orthologs from 21 Diverse Species from Mammals to *Entamoeba*

Heights of letters indicate degree of conservation in units of maximum entropy expressed in bits. Colors indicate amino-acid similarity: KRH, green; DE, blue; and AVLIPWFM, red. Conserved domains discussed in the text are labeled at their approximate midpoints. Gaps indicate regions in the alignment where only a minority of the sequences has residues present. The actual alignment is provided in Figure S4.

doi:10.1371/journal.pgen.0030237.g002

stretch (A) at the very C terminus. Together, these domains form the so-called VCA module found in all WASP family members [16] (Figure 3A). The VCA module is the minimal region required for binding and activation of actin polymerization via the Arp2/3 complex [14]. In known WASPs, the WH2 domain contains four conserved residues essential for actin binding [20], the A region has a conserved tryptophan residue and mediates binding to the Arp2/3 complex [14], and the amphipathic helical structure and conserved arginine in the C region induce conformational changes in Arp2/3 necessary to stimulate actin nucleation [21]. All these features are well conserved in WASH orthologs (Figures 2 and S4).

### WASH Proteins Define a New Subclass in the WASP Family

Mammalian WASH, WASP, and SCAR proteins separate into distinct clades supported by high bootstrap values in a phylogenetic tree (Figure 3B) based on the alignment of the VCA region of all known WASP family members (Figure S6). Based on their VCA sequences, WASP family orthologs from nonvertebrates do not cluster with statistical support into any one of these three clades. However, WASP family members from these organisms can be grouped together with the corresponding mammalian orthologs based on sequence homology within the N-terminal portion of the proteins (Figures 3A and S7).

Structural divergence of the known WASP family members outside the VCA region leads to variation in their activity and regulation and is used to subdivide them into WASPs and SCARs [15] (Figure 3A). WASP and N-WASP act as specific effectors of the small GTPase Cdc42 and have a WASP homology domain 1 (WH1) and a GTPase-binding domain (CRIB) at their N termini [22]. SCAR proteins lack these domains, but possess a specific SCAR-homology domain (SHD) and play a major role in Rac-induced actin dynamics [23].

WASH proteins lack N-terminal domains characteristic of either WASP or SCAR proteins, but they possess two evolutionarily conserved regions that appear to be specific to WASH orthologs (named WASH homology domains 1 and 2 (WHD1 and −2)) (Figures 2 and 3A). Based on the distinguishing features of the N-terminal portions and our phylogenetic analyses of VCA regions, we conclude that WASH proteins define a new WASP family subclass that is conserved from *Entamoeba* to human.
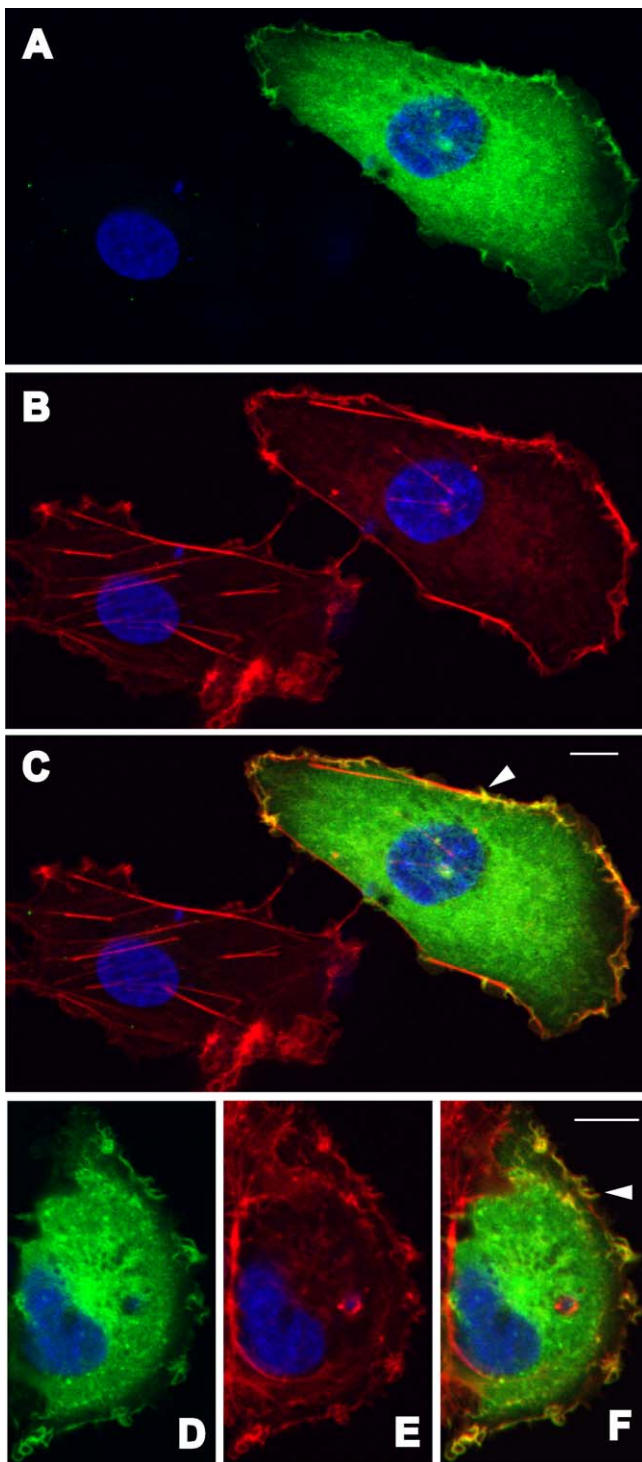
Human WASH Colocalizes with Actin

Human WASH protein colocalizes with actin in vivo, consistent with our prediction that WASH has a role in actin polymerization and cytoskeleton reorganization (Figure 4). We transiently expressed full-length WASH corresponding to
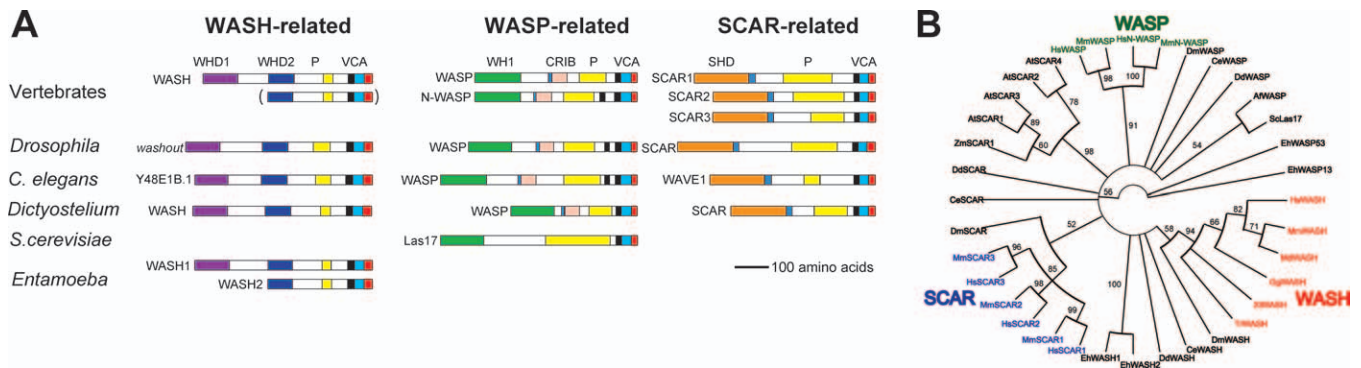
Figure 3. WASH Proteins Are Phylogenetically and Structurally Distinct from Known WASP Family Members

(A) WASP family members and characteristic domains, adapted from reference [16] to show the new subfamily of WASH orthologs. Humans possess multiple WASH genes distributed in their subtelomeres, including short forms like one of two *Entamoeba* orthologs. Excluding *Entamoeba*, all other non-primate species examined have one WASH ortholog, except *S. cerevisiae*, in which there is only a WASP-related homolog. P, proline-rich domain; VCA, actin-binding and polymerizing domains; WHD1 and WHD2, subfamily-specific N-terminal domains of WASH (Figure S7).

(B) Neighbor joining tree of WASP family members based on C-terminal (VCA region) alignment (Figure S6). Bootstrap values were calculated over 1,000 iterations. Vertebrate clades with high bootstrap support are in color. Species are indicated using two-letter abbreviations in front of each protein name (see Text S1).

doi:10.1371/journal.pgen.0030237.g003

Figure 4. Colocalization of Transiently Expressed, GFP-Tagged Full-Length Human WASH with Actin in Filopodia (Two Indicated by Arrowheads) and Lamellipodia.

(A) and (D), GFP-WASH (green) and DAPI to stain nuclei (blue); (B) and (E), Texas-red phalloidin to stain actin (red) and DAPI; (C) and (F), three-color overlay. Size bar, 10 μm.

doi:10.1371/journal.pgen.0030237.g004

the BC048328 cDNA sequence and carrying GFP at its C terminus in COS-7 cells. This fusion protein colocalizes with actin in filopodia and lamellipodia, which are actin-rich cell surface extensions, but not in actin-rich stress fibers.

## WASH Functions As Actin Nucleation Promoting Factor

WASP family proteins are known to stimulate actin polymerization by the Arp2/3 complex. The VCA module of previously characterized WASP family members is necessary and sufficient for this activity [14]. In order to test the effect of WASH on actin polymerization by the Arp2/3 complex, we performed pyrene actin-polymerization assays using the VCA region of the human WASH and N-WASP proteins. Without the Arp2/3 complex, neither the WASH VCA nor the N-WASP VCA has an effect on actin polymerization (Figure 5B and 5C). The combination of WASH VCA and Arp2/3 complex strongly stimulates spontaneous assembly of monomeric actin (Figure 5A and 5C), leading to reduction of the initial lag associated with nuclei formation, and increasing the maximum rate of filament elongation, which is dependent on the concentration of growing ends. WASH's effect on actin polymerization in the presence of Arp2/3 is less robust than that of the N-WASP VCA (Figure 5A and 5B), but is similar to that of other family members like WASP and SCAR, which each induce unique kinetics of actin assembly [24]. We conclude that WASH, like the known WASP family members, is an endogenous activator of de novo actin filament assembly.

Drosophila WASH Is Essential

The single WASH ortholog in *Drosophila*, which we have named *washout* (*wash*; CG13176), is structurally and phylogenetically distinct from the two WASP family members already characterized in this species (Figure 3A). Homozygous mutations in either Wasp or Scar result in zygotic lethality, although some of the Wasp mutants ("escapers") survive until adulthood and appear morphologically normal, but are lethargic and passive in their behavior [25,26]. *wash* gene products, like those of Wasp and Scar, are provided maternally, and *wash* transcripts appear to be distributed uniformly throughout the early embryo, based on RNA in situ hybridizations performed by the Berkeley Drosophila Genome Project (BDGP) [27] (http://www.fruitfly.org/cgi-bin/ex/insitu.pl).

To investigate *washout* function, we obtained a P-element insertion line, P{EPgy2}CG13176[EY15549] (ref [28]; BDGP, unpublished data), with a P-element insertion at the beginning of the *washout* coding region. These flies are homozygous viable with no apparent phenotype. We generated numerous imprecise excision alleles of this insertion line that are lethal when homozygous. In one of the alleles, *wash*[Δ185], more than half of the coding region (up to the VCA module) is deleted and two stop codons at positions 11 and 12 are introduced (Figure 6A); no flies homozygous for this allele survive to adulthood (Figure 6B). Flies bearing precise excisions of the P-element insertion in the homozygous state were viable (Figure 6B), indicating that the recessive lethality of *wash*[Δ185] is due to disruption of *washout*.

To determine the stage at which lethality occurs in *wash*[Δ185] homozygous flies, we crossed flies carrying the *wash*[Δ185] allele to a balancer chromosome (a multiply inverted chromosome that suppresses recombination) carrying GFP, which allows
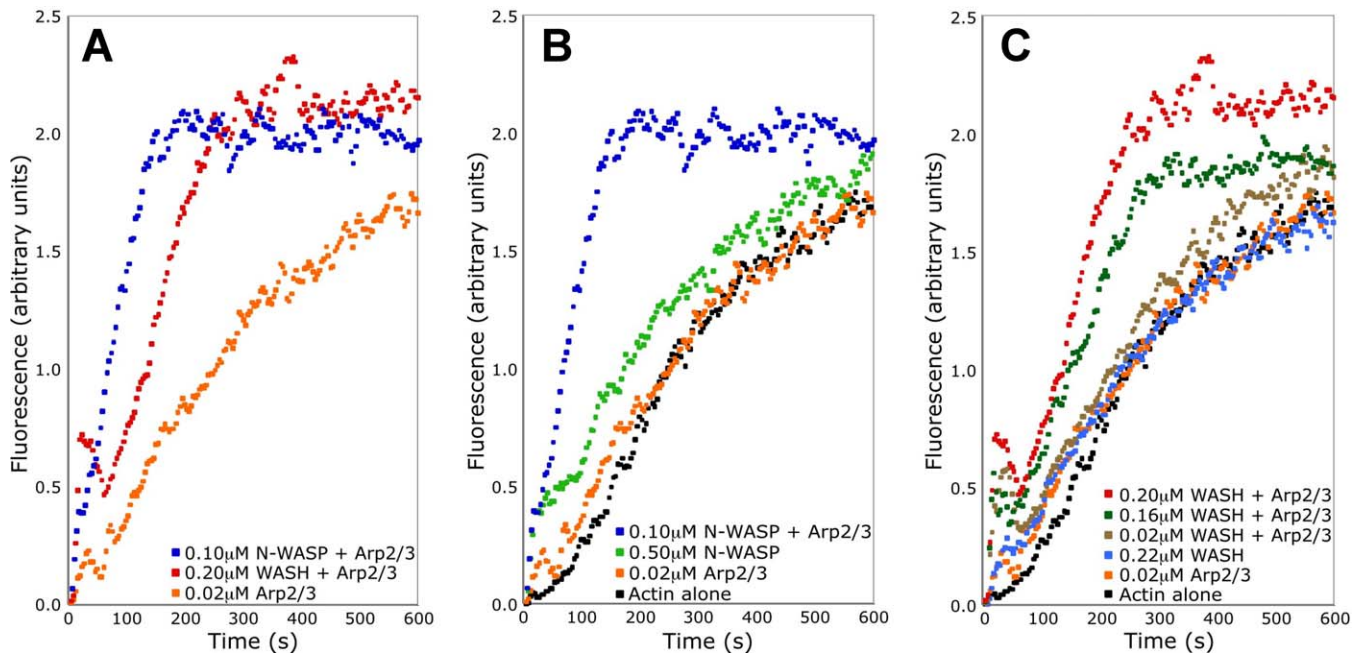
Figure 5. WASH Promotes Actin Nucleation by the Arp2/3 Complex

Pyrene-actin polymerization assays were conducted with the concentrations indicated for WASH VCA and N-WASP VCA (A). Similar to N-WASP VCA (B), WASH VCA (C) does not promote actin nucleation in the absence of Arp2/3. In the presence of Arp2/3, WASH promotes actin nucleation in a concentration-dependent manner (C).
doi:10.1371/journal.pgen.0030237.g005

for easy detection and separation of $wash^{\Delta185}$ homozygous embryos. Analyses of 200 $wash^{\Delta185}$ homozygous embryos revealed that the majority of them (98%) hatched, although none of them produced an adult fly. Further analysis showed that these $wash^{\Delta185}$ animals die at the transition from 3rd larval instar to prepupal stage. While flies heterozygous for the $wash^{\Delta185}$ allele have wild-type pupal morphology, $wash^{\Delta185}$ homozygotes fail to contract their body and evert their spiracles prior to secreting their pupal cuticle, resulting in an elongated appearance (Figure 6C).
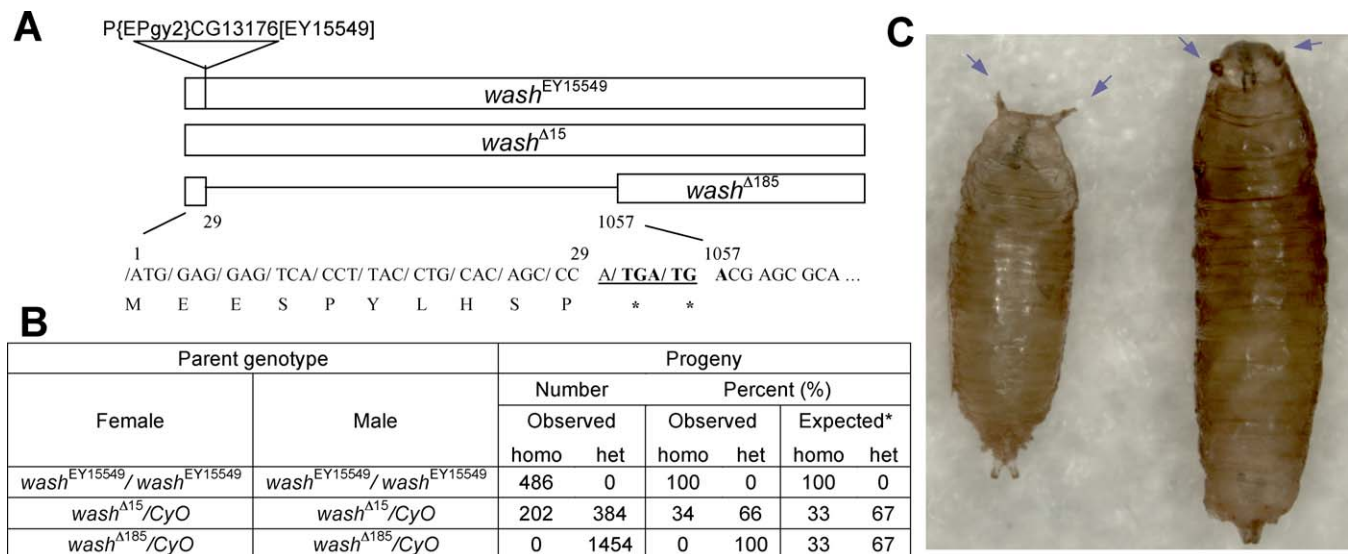
Figure 6. Analysis of *washout* Mutant Flies

(A) A schematic of the *washout* coding region showing the original P-element insertion allele ($wash^{EY15549}$) and a precise ($wash^{exc15}$) and imprecise ($wash^{\Delta185}$) excision allele. Sequence analysis of $wash^{\Delta185}$ reveals two stop codons at positions 11 and 12 due to the internal deletion of 1,029 nucleotides and insertion of six nucleotides (underlined) at the junction.
(B) Genetic analysis of *washout* alleles. No flies homozygous for the $wash^{\Delta185}$ excision allele were recovered. *CyO* is a balancer chromosome that is homozygous lethal.
(C) The $wash^{\Delta185}$ mutant phenotype. Bright-field micrograph of pupae heterozygous (left) or homozygous (right) for the $wash^{\Delta185}$ allele. The homozygous mutant displays an elongated phenotype and spiracle-eversion (arrows) defects, while the heterozygote is phenotypically normal.
doi:10.1371/journal.pgen.0030237.g006

chr. site	human						chimpanzee		gorilla	oran-gutan	macaca
	H1	H2	H3	H4	H5	H6	C1	C2	G1	O1	M1
1pter			■			■					
2qFS	■■	■■	■■	■■	■■	■					
3pter									■■		
3qter	■■	■■	■■	■■	■■	■■					
4qter							■■	■■			
5pter							■■	■■			
5qter					■						
6pter	■	■■	■	■■	■■	■■			■■		
6qter							■■	■■			
7pter	■										
7qter							■	■			
8qter							■				
9pter	■■	■■	■■	■■	■■	■■					
9qter		■■			■■						
11pter			■■								
12pter	■■	■■	■■	■■	■■	■■	■■	■■	■■	■■	■■
15qter	■■	■■	■■	■							
16pter	■		■■								
16qter	■										
19pter	■■	■	■	■■	■■	■■					
20pter	■■	■■	■■	■■	■■	■■	■■	■■	■■		
20qter	■■	■■	■	■■	■■	■■					
Xqter								■			
Yqint						■					

Figure 7. Summary of Locations of WASH Detected by FISH in Unrelated Individuals

The signal-scoring criteria used to deem a location as homozygous (two bars) or heterozygous (one bar) for the presence of WASH sequence are supplied in Text S1. Here, we indicate human WASH locations excluding partial duplicates on Xqter, Yqter, and some 16pter alleles, which were detected only with FISH probes including the distal portion of WASH. For the nonhuman primates, we used a pool of probes completely spanning WASH in order to report all WASH locations in these species. Nonhuman primate chromosomes are numbered according to the corresponding chromosome in the human karyotype.
doi:10.1371/journal.pgen.0030237.g007

WASH Genes Duplicated in Primate Genomes

While a single-copy gene encodes WASH in most species analyzed, WASH genes experienced extensive duplication and dispersal to multiple chromosome ends during primate evolution. Our FISH assays detect many chromosomal sites, as well as extensive variation in the copy number and location of WASH genes between primate species and among human individuals (Figure 7). The total number of WASH copies, excluding partial duplications, ranges from >15 to >20 in the six analyzed human genomes. Collectively, these copies are found in 16 different sites and represent both WASH pseudogenes and intact WASH genes (see below). Our FISH analyses also support the conclusion that termini of 16p, Xq, and Yq typically carry a truncated WASH.

Nonhuman primates have far fewer copies of WASH than humans. WASH was detected at only eight or nine sites in the two chimpanzees analyzed by FISH and four sites in gorilla (Figure 7). WASH appears to be single-copy in orangutan and rhesus macaque, where it resides on 12p. Since the 12p location is shared among primates, it is likely to be the ancestral location of WASH before it duplicated during hominid diversification.

Human Variation in WASH Coding Potential

Given the large number of human WASH loci detected by FISH and the relative paucity of sequenced WASH paralogs in public databases, we characterized the protein-coding potential of additional copies. We sequenced long-range PCR products encompassing coding exons 2–10 from three unrelated individuals. We have so far identified up to five potentially functional WASH variants and multiple pseudo-genes per genome, although deeper sampling would be required to account for all copies detected by FISH in these individuals. Our sequence analyses also reveal that two of the eight human WASH copies captured in a panel of mono-chromosomal hybrid cell lines are full-length, intact ORFs.

Intact and null WASH alleles are segregating in the human population at some chromosomal loci. For example, the hybrid panel's chromosome 20 carries the two full-length ORFs, one mapping to each end by FISH, but the presence of WASH at 20qter is polymorphic (Figure 7). Furthermore, the chromosome-9 allele captured in a panel of monochromoso-mal hybrid cell lines contains a stop codon in coding exon 5, but is 99.9% identical to the sequenced 9pter allele in the genome assembly, which appears intact.

Our survey found a total of 12 different intact WASH ORFs, in addition to BC048328, assigned by our analyses to chromosome 20, and the 9p copy in the assembly. These 14 ORFs differ from each other across exons 2–10 by up to 18 amino acids (≥95.8% identity) and one indel of three amino acids in the P region (Table S1). It is possible that some of these variants have slightly different functions, as they exhibit up to 11 nonconservative amino acid differences, although we detect no evidence for positive selection in mammalian WASH genes by PAML or K-estimator analyses (Text S1).

Intact, and possibly more divergent, copies of WASH might be found at other chromosomal ends in other individuals, since subtelomeres undergo interchromosomal sequence exchange [4]. Indeed, by evaluating the sequences of WASH copies in the current assembly, we detect two apparent exchange events that resulted in the transfer of different segments of several hundred base pairs from one chromo-

some to another (Figure S10). Thus, WASH genes evolve through a combination of acquired mutation, recombination between alleles and/or paralogs, and selection, and WASH repertoires can be expected to vary extensively among individuals.

Discussion

This study shows that the most telomerically located human gene family encodes evolutionarily conserved proteins with orthologs in vertebrates, flies, worms, slime mold, and entamoeba. Human WASH colocalizes with actin filaments in lamellipodia and filopodia and stimulates actin polymerization in vitro, corroborating our sequence-based predictions that WASH orthologs are new members of the WASP family. Our sequence comparisons show that these newly identified WASH proteins form a previously unrecognized subclass in the WASP family, distinct from the known N-WASP/WASP and SCAR/WAVE proteins.

Known WASP family members have been implicated in the formation of lamellipodia and filopodia and in membrane-trafficking processes such as endocytosis, intracellular pathogen motility, and vesicle motility [16,18,29]. Yeast has only one WASP-related protein; its deletion causes a severe growth phenotype [30]. *WASp* and *SCAR* are important for cell-fate decisions and cell morphology, respectively, during *Drosophila* embryonic development [25,26]. Drosophila *WASp* is also essential for myoblast fusion [31–33]. We show here that the newly identified third WASP family member in *Drosophila*, *washout*, is essential, indicating that this gene also has an important role in early development and is not redundant with other WASP family members.

The *Drosophila* WASH protein was recently identified as a component of a nuclear complex containing various transcriptional factors and chromatin modifiers [34]. Consistent with this finding, the vertebrate WASH consensus sequence possesses predicted nuclear localization, nuclear export, and sumoylation signals (Figure S5). Actin and actin-binding proteins are found in the nucleus, where they might be involved in chromatin remodeling, RNA processing, or gene expression [15,35,36]. N-WASP has been shown to participate directly in transcriptional regulation [36,37] in addition to its role in the cytoplasm. It is therefore likely that WASH also has both nuclear and cytoplasmic functions.

We anticipate that loss of WASH gene function in humans will have pathological consequences, particularly since human WASH is ubiquitously expressed, with highest levels in hematopoietic tissues and some brain areas. Knockout of either N-WASP or Scar2, both of which are also ubiquitously expressed, is embryonic lethal in mice [38,39]. Scar1 is most highly expressed in brain, and, accordingly, Scar1-null mice exhibit defects in brain function [40]. The WASP gene defective in human Wiskott-Aldrich syndrome is expressed in hematopoietic cells [41]; its loss causes eczema, thrombocytopenia, and immunodeficiency [42]. Human WASH genes are at heightened risk for deletion and rearrangement, since subtelomeres are hotspots of meiotic interchromosomal sequence transfers [4]. Furthermore, somatic variation in subtelomeric organization and/or telomere length could influence WASH gene expression. WASH was reported to be overexpressed in a breast cancer cell line [43] and might,

like overexpression of N-WASP and the SCARs, contribute to metastasis [17].

On the other hand, subtelomeric dynamics might contribute to normal human phenotypic variation and, more generally, to diversification of the WASP family. Subtle intra-species phenotypic variation might result from variation in WASH repertoires (gene number, location, and/or sequence). Intact and defective/missing WASH alleles segregate in the human population at some loci (e.g., 9pter, 20qter). We have shown that subtelomeric exchanges can create WASH loci that appear to combine sequence differences accrued by copies on different chromosomes. Finally, the identified human sequences with intact long ORFs differ from one another by as many as 11 nonconservative amino acid changes, raising the possibility that the encoded proteins might have slightly different functions.

Remarkably, the genes encoding this evolutionarily conserved protein multiplied within subtelomeric regions during primate evolution and thus might contribute to interspecies phenotypic differences. Maintenance of subtelomeric segmental duplications in yeast requires selective pressure [44]. Although genetic drift cannot be excluded as the explanation for the rapid recent expansion of WASH copies in the human lineage, WASH expansion might be important for fast evolving processes such as immune response and/or brain function.

The presence of WASH orthologs in *Entamoeba* also raises the possibility of WASH involvement in pathogenic infection. A number of unrelated pathogens hijack actin-polymerization pathways in host cells to facilitate their infection [45,46]. Rho GTPases and WASP family members are implicated in these pathogenic processes [45–47]. It is possible that the primate-specific subtelomeric expansion of WASH genes is associated with host response to pathogen infection.

Given the critical roles of the WASP family in diverse cellular processes, elucidation of the specific role(s) of this newly identified and evolutionarily conserved WASH subclass should provide important insights into actin dynamics in response to external signals. The location of WASH in highly dynamic human subtelomeric regions predisposes it to duplication, deletion, and rearrangement. Certain rearrangements could have pathological consequences if WASH plays as important a role in humans as it does in *Drosophila*, where it is essential. Further characterization of WASH protein function(s) should shed light on how WASH genes and their proximity to telomeres contribute to normal human variation as well as pathology.

Supporting Information

Dataset S1. Sequences of WASH Gene Variants and the Computationally Derived, Predicted WASH Protein Sequences of Various Species in FASTA File Format

Found at doi:10.1371/journal.pgen.0030237.sd001 (29 KB PDF).

Figure S1. RT-PCR Confirmation of Human WASH Expression Using exon 8 and exon 10 Primers

L, liver; LN, lymph node; H, heart; FB, fetal brain; g, genomic DNA. Reverse transcriptase (RT)-negative controls were also performed for some tissues. The 0.3-kb band is the expected size for fully spliced WASH transcript.

Found at doi:10.1371/journal.pgen.0030237.sg001 (226 KB PDF).

Figure S2. Multiple Tissue Expression Array Autoradiograph after Hybridization Using a WASH Probe (A), the Corresponding Key to

Tissues (B), and (C) Relative WASH Hybridization Signals Normalized to the Ubiquitin Signal, Organized by Tissue Type

Found at doi:10.1371/journal.pgen.0030237.sg002 (743 KB PDF).

Figure S3. Northern Blot Showing Expression of Spliced WASH Transcripts in Multiple Human Tissues

The ~1.8-kb band corresponds to the expected size of fully spliced, full-length WASH transcripts. The same blot was probed with β-actin as a loading control (bottom). The intensities of the ~1.8-kb WASH band in each tissue are expressed as a fraction of intensity of the β-actin band(s) in the numbers at the bottom of the gel. Where two numbers are given, the top one normalizes to the top actin band, and the bottom normalizes to the sum of the two actin bands.

Found at doi:10.1371/journal.pgen.0030237.sg003 (278 KB PDF).

Figure S4. Multiple Sequence Alignment of WASH Orthologs

Conserved domains are named above the alignment, using same colors as in Figures 2 and 3A. Boxed residues in red letters are conserved in >50% of sequences, and the >50% consensus is indicated at the bottom. Amino acids showing conservation across >80% of the sequences are also shown at the bottom. Amino acids showing 100% conservation are noted against red background. In the consensus sequences, ! is I or V; $ is L or M; % is F or Y; and # is any of NDQEBZ. Asterisks mark conserved residues important for function in known WASP family members. The line "Change" gives the residues changed by nonsynonymous SNPs detected in our survey of human WASH paralogs; underlined changes are nonconservative.

Found at doi:10.1371/journal.pgen.0030237.sg004 (594 KB PDF).

Figure S5. Logo Plot of Vertebrate WASH Based on Alignment Provided in Figure S4

Positions of predicted sumoylation sites and nuclear localization and export signals (SUMO, NLS, NES) are indicated. Conserved domains are named above the alignment, using the same colors as used in Figure 3A; the domains extend beyond the labels. Heights of letters indicate degree of conservation in maximum entropy expressed in bits. Similar amino acids are given the same color: KRH, green; DE, blue; and AVLIPWFM, red. Gaps in the logo plots are regions with residues in only the minority of sequences.

Found at doi:10.1371/journal.pgen.0030237.sg005 (992 KB PDF).

Figure S6. Multiple Sequence Alignment of the VCA Region in the C-Terminal Portion of WASP Family Members

See Figure S4 for details. Species are indicated using two-letter abbreviation before each protein name; the code is provided in Text S1 (note: Mm is *Mus musculus*, not *Macaca mulatta*).

Found at doi:10.1371/journal.pgen.0030237.sg006 (1.2 MB PDF).

Figure S7. Logo Plots of (A) WASH, (B) WASP/N-WASP, and (C) SCAR/WAVE Subclasses of the WASP Family to Illustrate Protein Conservation across Species and Differences among These Subclasses in Their N-Terminal Portions

The logo plots are based on alignments provided in Figures S4, S8, and S9, respectively. Conserved domains are named above the alignment, using the same colors as used in Figure 3A; the domains extend beyond the labels. Heights of letters indicate degree of conservation in maximum entropy expressed in bits. Similar amino acids are given the same color: KRH, green; DE, blue; and AVLIPWFM, red. Gaps in the logo plots are regions with residues in only the minority of sequences.

Found at doi:10.1371/journal.pgen.0030237.sg007 (2.5 MB PDF).

Figure S8. Multiple Alignment of WASP and N-WASP Orthologs from Diverse Species

See Figure S4 legend for details.

Found at doi:10.1371/journal.pgen.0030237.sg008 (402 KB PDF).

Figure S9. Multiple Alignment of SCAR (WAVE) Orthologs from Diverse Species

See Figure S4 legend for details.

Found at doi:10.1371/journal.pgen.0030237.sg009 (485 KB PDF).

Figure S10. Percent Identity Plots for Pairs of WASH Genomic Sequences That Appear to Have Undergone Recent Sequence Exchange (Chromosomes 1 versus 9 and 9 versus 16, Blue and Green Lines, Respectively)

The plot for chromosome 16 versus chromosome X (red lines) is shown for comparison. Percent identity is plotted in 500-bp windows moving in 10-bp increments. Exchange events appear as shifts to very high sequence similarity. These events are also detected by GeneConv (reference 7 in Text S1), with corresponding Bonferroni-corrected Karlin-Altschul *p*-values of 0.0016 (g1 setting) and 0.0065 (g0 setting). Position 0 is in intron 1 of WASH, 110 nt preceding the start of the 2nd coding exon, corresponding to the start of the partial duplication on 16 and X/Y. The end of the region shown is at nt position 4207 in intron 10.

Found at doi:10.1371/journal.pgen.0030237.sg010 (245 KB PDF).

Table S1. Locations of Sequence Variants Identified in 14 Different WASH ORFs That Appear to Be Intact, at Least in Coding Exons 2–10

We used the cDNA sequence BC048328 as reference; "-" indicates where a sequence is not different from BC048328. These variants were derived from chromosomes captured in a monochromosomal hybrid panel, certain chromosomes isolated by flow-sorting (20 or 9–12) from three individuals, and copies PCR amplified from genomic DNA (without chromosomal assignment) from the same individuals. Note: Two ORFs in one individual (with asterisks) were identical in coding sequence, but differed by 3% in intronic sequence implying that they derive from different chromosomal locations. n.a., not applicable; n.d., not determined. The SNP highlighted in blue is unconfirmed, as it was observed only once in this survey and is not a variant site in EST sequences in public databases. The variant sequences are provided in FASTA format in Dataset S1 and have been submitted to GenBank with accession numbers EU240546–EU240557.

Found at doi:10.1371/journal.pgen.0030237.st001 (348 KB PDF).

Text S1. Supplementary Information and Methods

Found at doi:10.1371/journal.pgen.0030237.sd002 (204 KB DOC).

References

1. Cenci G, Ciapponi L, Gatti M (2005) The mechanism of telomere protection: a comparison between *Drosophila* and humans. Chromosoma 114: 135–145.
2. Koering CE, Pollice A, Zibella MP, Bauwens S, Puisieux A, et al. (2002) Human telomeric position effect is determined by chromosomal context and telomeric chromatin integrity. EMBO Rep 3: 1055–1061.
3. Pedram M, Sprung CN, Gao Q, Lo AW, Reynolds GE, et al. (2006) Telomere position effect and silencing of transgenes near telomeres in the mouse. Mol Cell Biol 26: 1865–1878.
4. Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, et al. (2005) Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. Nature 437: 94–100.
5. Matise TC, Sachidanandam R, Clark AG, Kruglyak L, Wijsman E, et al. (2003) A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. Am J Hum Genet 73: 271–284.

6. d'Adda di Fagagna F, Reaper PM, Clay-Farrace L, Fiegler H, Carr P, et al. (2003) A DNA damage checkpoint response in telomere-initiated senescence. Nature 426: 194–198.

7. Rudd MK, Friedman C, Parghi SS, Linardopoulou EV, Hsu L, et al. (2007) Elevated rates of sister chromatid exchange at chromosome ends. PLoS Genet 3: e32. doi:10.1371/journal.pgen.0030032

8. Barry JD, Ginger ML, Burton P, McCulloch R (2003) Why are parasite contingency genes often associated with telomeres? Int J Parasitol 33: 29–45.

9. Fabre E, Muller H, Therizols P, Lafontaine I, Dujon B, et al. (2005) Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. Mol Biol Evol 22: 856–873.

10. Ravnan JB, Tepperberg JH, Papenhausen P, Lamb AN, Hedrick J, et al. (2006) Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. J Med Genet 43: 478–489.

11. Ciccodicola A, D'Esposito M, Esposito T, Gianfrancesco F, Migliaccio C, et al. (2000) Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. Hum Mol Genet 9: 395–401.

12. Gianfrancesco F, Falco G, Esposito T, Rocchi M, D'Urso M (2001) Characterization of the murine orthologue of a novel human subtelomeric multigene family. Cytogenet Cell Genet 94: 98–100.

13. Hansen M, Hsu AL, Dillin A, Kenyon C (2005) New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a Caenorhabditis elegans genomic RNAi screen. PLoS Genet 1: e17. doi:10.1371/journal.pgen.0010017

14. Marchand JB, Kaiser DA, Pollard TD, Higgs HN (2001) Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. Nat Cell Biol 3: 76–82.

15. Bompard G, Caron E (2004) Regulation of WASP/WAVE proteins: making a long story short. J Cell Biol 166: 957–962.

16. Millard TH, Sharp SJ, Machesky LM (2004) Signalling to actin assembly via the WASP (Wiskott-Aldrich syndrome protein)-family proteins and the Arp2/3 complex. Biochem J 380: 1–17.

17. Yamaguchi H, Condeelis J (2007) Regulation of the actin cytoskeleton in cancer cell migration and invasion. Biochim Biophys Acta 1773: 642–652..

18. Takenawa T, Suetsugu S (2007) The WASP-WAVE protein network: connecting the membrane to the cytoskeleton. Nat Rev Mol Cell Biol 8: 37–48.

19. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 101: 6062–6067.

20. Paunola E, Mattila PK, Lappalainen P (2002) WH2 domain: a small, versatile adapter for actin monomers. FEBS Lett 513: 92–97.

21. Panchal SC, Kaiser DA, Torres E, Pollard TD, Rosen MK (2003) A conserved amphipathic helix in WASP/Scar proteins is essential for activation of Arp2/3 complex. Nat Struct Biol 10: 591–598.

22. Caron E (2003) Regulation by phosphorylation. Yet another twist in the WASP story. Dev Cell 4: 772–773.

23. Pollard TD, Borisy GG (2003) Cellular motility driven by assembly and disassembly of actin filaments. Cell 112: 453–465.

24. Zalevsky J, Lempert L, Kranitz H, Mullins RD (2001) Different WASP family proteins stimulate different Arp2/3 complex-dependent actin-nucleating activities. Curr Biol 11: 1903–1913.

25. Ben-Yaacov S, Le Borgne R, Abramson I, Schweisguth F, Schejter ED (2001) Wasp, the Drosophila Wiskott-Aldrich syndrome gene homologue, is required for cell fate decisions mediated by Notch signaling. J Cell Biol 152: 1–13.

26. Zallen JA, Cohen Y, Hudson AM, Cooley L, Wieschaus E, et al. (2002) SCAR is a primary regulator of Arp2/3-dependent morphological events in Drosophila. J Cell Biol 156: 689–701.

27. Tomancak P, Beaton A, Weiszmann R, Kwan E, Shu S, et al. (2002) Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol 3: RESEARCH0088.

28. Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, et al. (1999) The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. Genetics 153: 135–177.

29. Vartiainen MK, Machesky LM (2004) The WASP-Arp2/3 pathway: genetic insights. Curr Opin Cell Biol 16: 174–181.

30. Li R (1997) Bee1, a yeast protein with homology to Wiscott-Aldrich syndrome protein, is critical for the assembly of cortical actin cytoskeleton. J Cell Biol 136: 649–658.

31. Kim S, Shilagardi K, Zhang S, Hong SN, Sens KL, et al. (2007) A critical function for the actin cytoskeleton in targeted exocytosis of prefusion vesicles during myoblast fusion. Dev Cell 12: 571–586.

32. Massarwa R, Carmon S, Shilo BZ, Schejter ED (2007) WIP/WASp-based actin-polymerization machinery is essential for myoblast fusion in Drosophila. Dev Cell 12: 557–569.

33. Schafer G, Weber S, Holz A, Bogdan S, Schumacher S, et al. (2007) The Wiskott-Aldrich syndrome protein (WASP) is essential for myoblast fusion in Drosophila. Dev Biol 304: 664–674.

34. Hochheimer A, Zhou S, Zheng S, Holmes MC, Tjian R (2002) TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. Nature 420: 439–445.

35. Shumaker DK, Kuczmarski ER, Goldman RD (2003) The nucleoskeleton: lamins and actin are major players in essential nuclear functions. Curr Opin Cell Biol 15: 358–366.

36. Suetsugu S, Takenawa T (2003) Translocation of N-WASP by nuclear localization and export signals into the nucleus modulates expression of HSP90. J Biol Chem 278: 42515–42523.

37. Wu X, Yoo Y, Okuhama NN, Tucker PW, Liu G, et al. (2006) Regulation of RNA-polymerase-II-dependent transcription by N-WASP and its nuclear-binding partners. Nat Cell Biol 8: 756–763.

38. Snapper SB, Takeshima F, Anton I, Liu CH, Thomas SM, et al. (2001) N-WASP deficiency reveals distinct pathways for cell surface projections and microbial actin-based motility. Nat Cell Biol 3: 897–904.

39. Yan C, Martinez-Quiles N, Eden S, Shibata T, Takeshima F, et al. (2003) WAVE2 deficiency reveals distinct roles in embryogenesis and Rac-mediated actin-based motility. EMBO J 22: 3602–3612.

40. Soderling SH, Langeberg LK, Soderling JA, Davee SM, Simerly R, et al. (2003) Loss of WAVE-1 causes sensorimotor retardation and reduced learning and memory in mice. Proc Natl Acad Sci U S A 100: 1723–1728.

41. Zhu Q, Watanabe C, Liu T, Hollenbaugh D, Blaese RM, et al. (1997) Wiskott-Aldrich syndrome/X-linked thrombocytopenia: WASP gene mutations, protein expression, and phenotype. Blood 90: 2680–2689.

42. Ochs HD, Thrasher AJ (2006) The Wiskott-Aldrich syndrome. J Allergy Clin Immunol 117: 725–738.

43. Leirdal M, Shadidy M, Rosok O, Sioud M (2004) Identification of genes differentially expressed in breast cancer cell line SKBR3: potential identification of new prognostic biomarkers. Int J Mol Med 14: 217–222.

44. Koszul R, Dujon B, Fischer G (2006) Stability of large segmental duplications in the yeast genome. Genetics 172: 2211–2224.

45. Voigt H, Guillen N (1999) New insights into the role of the cytoskeleton in phagocytosis of Entamoeba histolytica. Cell Microbiol 1: 195–203.

46. Munter S, Way M, Frischknecht F (2006) Signaling during pathogen infection. Sci STKE 2006: re5.

47. Franco-Barraza J, Zamudio-Meza H, Franco E, del Carmen Dominguez-Robles M, Villegas-Sepulveda N, et al. (2006) Rho signaling in Entamoeba histolytica modulates actomyosin-dependent activities stimulated during invasive behavior. Cell Motil Cytoskeleton 63: 117–131.