

Research article

Open Access

Gene discovery in the hamster: a comparative genomics approach for gene annotation by sequencing of hamster testis cDNAs

Sreedhar Oduru¹, Janee L Campbell¹, SriTulasi Karri¹, William J Hendry³, Shafiq A Khan¹ and Simon C Williams*^{1,2}

Address: ¹Department of Cell Biology & Biochemistry, Texas Tech University Health Sciences Center, Lubbock, Texas, USA, ²Southwest Cancer Center at UMC, Lubbock, Texas, USA and ³Department of Biological Sciences, Wichita State University, Wichita, Kansas, USA

Email: Sreedhar Oduru - oduru.sreedhar@ttuhsc.edu; Janee L Campbell - janee.campbell@ttuhsc.edu; SriTulasi Karri - phrsk@ttuhsc.edu; William J Hendry - william.hendry@wichita.edu; Shafiq A Khan - shafiq.khan@ttuhsc.edu; Simon C Williams* - simon.williams@ttuhsc.edu

* Corresponding author

Published: 3 June 2003

Received: 31 January 2003

BMC Genomics 2003, 4:22

Accepted: 3 June 2003

This article is available from: <http://www.biomedcentral.com/1471-2164/4/22>

© 2003 Oduru et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Complete genome annotation will likely be achieved through a combination of computer-based analysis of available genome sequences combined with direct experimental characterization of expressed regions of individual genomes. We have utilized a comparative genomics approach involving the sequencing of randomly selected hamster testis cDNAs to begin to identify genes not previously annotated on the human, mouse, rat and Fugu (pufferfish) genomes.

Results: 735 distinct sequences were analyzed for their relatedness to known sequences in public databases. Eight of these sequences were derived from previously unidentified genes and expression of these genes in testis was confirmed by Northern blotting. The genomic locations of each sequence were mapped in human, mouse, rat and pufferfish, where applicable, and the structure of their cognate genes was derived using computer-based predictions, genomic comparisons and analysis of uncharacterized cDNA sequences from human and macaque.

Conclusion: The use of a comparative genomics approach resulted in the identification of eight cDNAs that correspond to previously uncharacterized genes in the human genome. The proteins encoded by these genes included a new member of the kinesin superfamily, a SET/MYND-domain protein, and six proteins for which no specific function could be predicted. Each gene was expressed primarily in testis, suggesting that they may play roles in the development and/or function of testicular cells.

Background

The initial publication of two draft versions of the human genome led to intense debate over the exact number of genes in the human genome [1,2]. Current estimates suggest that the human genome encodes approximately 35,000 to 38,000 although the final number must await the complete annotation of each genome sequence. The search for additional genes not discovered during early

annotation attempts has involved the use of several different approaches. These have included the sequencing of randomly selected cDNAs from various tissue sources, the development of computer-based prediction programs of ever-increasing accuracy, and the direct comparison between the human genome and the genome sequences of other vertebrates and invertebrates [3–9]. Using these

approaches, fully annotated genomes of numerous species will be available within a relatively short time.

We approached the problem of gene identification by using a combination of experimental and *in silico* techniques. Specifically, we initiated a project designed to sequence expressed sequence tags from the hamster testis and used these sequences to identify unannotated, or incompletely annotated, genes in the human and other vertebrate genomes. The hamster has not been used extensively in genomics research; however, it has been used extensively in various areas of investigation including circadian rhythm research [10] and also in investigations in a number of areas of research in reproductive biology. For example, the study of hamster gametes has revealed significant information concerning the mechanisms underlying species-specific sperm-egg interactions [11–13] and the deleterious effects of endocrine disruptors on male and female reproductive development [14–17]. The hamster, mouse and rat are all members of the family Muridae, however both mice and rats belong to the subfamily Murinae while hamsters belong to the subfamily Cricetinae. Three hamster species that are commonly used in research are *Mesocricetus auratus* (Syrian golden hamster), *Cricetus griseus* (Chinese hamster) and *Phodopus sungorus* (Siberian hamster). Therefore, sequence information from any hamster species should complement information gained from other closely related species.

The testis was chosen for these studies as it represents a viable source for the identification of novel genes. The adult testis is a complex organ consisting of numerous different somatic cell types as well as germ cells at all stages of spermatogenesis from the gonocyte stem cells to the mature sperm cells [18]. Consequently, several unique gene populations, including those involved in the regulation of meiosis, as well as those specific to the various testicular cell types, are expressed in the testis. A recent gene discovery study performed in the testis of *Drosophila melanogaster* found that 47% of greater than 1500 sequenced cDNAs did not match to ESTs previously identified in this organism [19]. Likewise the testis of the cynomolgus monkey has yielded several novel gene sequences [8,9]. Therefore, we reasoned that the sequencing of ESTs from hamster testis might reveal the existence of novel genes conserved in other species that may function in controlling testicular development and/or function. In this report, we describe our initial results from the sequencing of randomly-selected cDNAs from the testes of male Syrian golden hamsters. In particular we identified eight cDNAs that appear to be derived from genes that were not previously annotated in the human genome. We describe the detailed analysis of two of these genes, which encode a new member of the kinesin superfamily of microtubule-

based molecular motors and a protein likely to be involved in chromatin remodeling.

Results

Generation and sequence analysis of a Hamster testis cDNA library

Random clones from a hamster testis cDNA library were selected and sequenced as described in materials and methods. The sequences were screened to remove ribosomal RNA and vector sequences, which yielded 735 distinct sequences. The sequence of each clone was compared to sequences in public databases to identify its closest match. Each sequence was then assigned to a functional group based on this comparison (Table 1). The genes were distributed amongst all of the functional groups listed with the highest numbers in groups associated with protein synthesis and degradation (11%), metabolism (6%), gene regulation and RNA processing (8%) and intracellular signaling (5%). Overall, the data set contains examples of numerous testis-specific genes as well as genes that display less limited patterns of expression.

We next considered the overall complexity of the sequences in our database as our cloning strategy made it

Table 1: Classification of sequenced hamster cDNAs

| Classification | Number of clones |
|---------------------------|------------------|
| Transcription | 23 (21) |
| Chromatin associated | 20 (17) |
| RNA processing | 17 (15) |
| DNA synthesis | 2 |
| Cell cycle | 9 |
| Cell surface antigens | 1 |
| Cell adhesion | 4 |
| Receptors | 2 |
| Extracellular signaling | 2 |
| Intracellular signaling | 36 (33) |
| Stress proteins | 13 (6) |
| Trafficking | 11 (10) |
| Metabolism | 41 (37) |
| Translation | 31 (26) |
| Protein modification | 21 (18) |
| Protein turnover | 30 (29) |
| Apoptosis | 3 |
| Cytoskeleton | 30 (23) |
| Functionally unclassified | 239 (222) |
| Unknown | 200 |

Each of the hamster ESTs was compared against public sequence databases using BLAST. Each clone was then functionally classified based on the results of these searches. The number of unique fragments belonging to gene in each functional category is listed first, followed in parentheses by the total number of unique genes represented once the occurrence of multiple fragments from the same gene is accounted for.

possible to obtain multiple fragments from the same cDNA. We identified only 40 instances where more than one fragment was derived from the same gene (data not shown). In addition, only 17 fragments appeared more than two times in the database and only 6 fragments appeared more than three times. The most common sequences were derived from genes that are known to be highly expressed in the testis in other organisms and include heat shock protein 90A [20], chitinase [21], outer dense fiber of sperm tails [22] and kinectin [23]. Therefore, the majority of sequences were represented only once or at most two times in the data set. The fact that most sequences were not identified multiple times suggests that we have not exhaustively sequenced all of the *DpnII* fragments from cDNAs expressed in the hamster testis. This is not surprising as conservative estimates suggest that there are many thousands of unique transcripts expressed in the various somatic and germ cell lineages in the testis.

The largest group of sequences identified in this project falls into the functionally unclassified group (Table 1). This group includes a small number of named genes for which no function has yet been assigned but primarily contains the results of full-length cDNA sequencing projects or genes predicted by gene finding software. Approximately one third of these clones (78 out of 239) were originally isolated from testis libraries, suggesting that a large number of genes remain to be functionally characterized in the testis.

Identification of novel testis specific genes

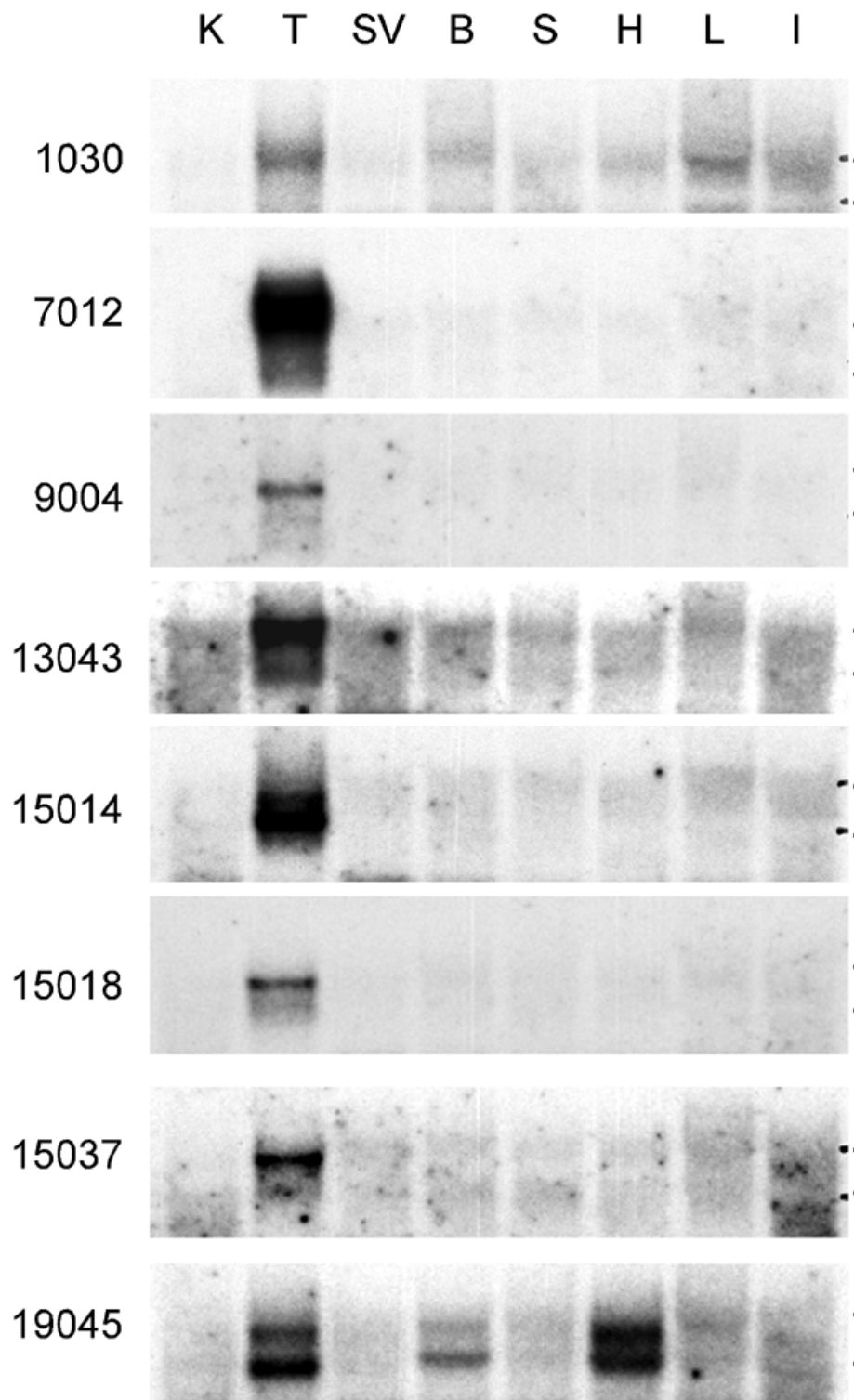
200 sequences could not be sorted into the various functional groups described above and represented a potential source of novel genes that had not previously been recognized by EST or genome sequencing. However, there were several other possible origins for such sequences. These included contamination of the original cDNA sample with genomic DNA (note that the testis total RNA was treated with DNase before mRNA purification) or artificial joining of unrelated sequences into chimeric fragments that resulted in matches that were not detected in our screen. As we synthesized cDNA from total RNA and not cytoplasmic RNA, it was possible that some fragments could be derived from RNAs that were incompletely processed [24]. Furthermore, these fragments could be derived from alternatively spliced exons of known genes or from genes that are only expressed in hamster. Therefore, several clones were selected for further examination based on their sequence similarity to unannotated regions of the human genome (see below) and for Northern analysis using hamster RNAs prepared from various tissues. Eight clones yielded specific signals in RNA from hamster testis while two clones also detected signals in other tissues (figure 1). Subsequent analysis has revealed that each clone is

derived from a bona fide gene and the evidence for this conclusion is provided below for each gene (see Table 2). The derived protein structures of the polypeptides encoded by each gene are shown in figure 2.

1030: A new member of the kinesin superfamily

The 1030 clone contained an ORF of 80 amino acids that displayed similarity to members of the kinesin superfamily of molecular motors [25,26]. Northern analysis using the hamster 1030 probe detected a single band in testis RNA and weak signals in several other tissues including brain, heart and lung (figure 1). Comparison of the hamster nucleotide sequence with genomic sequences revealed highly significant matches within chromosomal positions 9q21.33 in human, 13B2 in mouse and 17 in rat. Partial cDNAs from human and macaque also mapped to the same genetic locus (Table 2). Using the procedures described in materials and methods we mapped an 18 exon gene that was followed by a consensus polyadenylation signal in the human, mouse and rat genome sequences (figure 3A). Exon 1 is located within a strong CpG island, indicating the likely presence of a promoter in this region. PCR cloning of the human cDNA was employed to confirm the exonic structure predicted by the genomic sequences. These studies revealed the existence of three alternatively spliced products, one containing all 17 coding exons (exon 1 is non-coding) and two variants lacking either exon 11 or exons 12 and 13 (figures 3B and 3C).

Conceptual translation of the predicted full-length human mRNA revealed a 1401 amino acid protein that is highly conserved in macaque, mouse and rat (figures 2 and 4). Database searches revealed that it encodes a previously unreported member of the kinesin superfamily. The kinesin superfamily in humans and mice is comprised of at least 45 members, designated with the prefix KIF [26]. Based on this naming system we have assigned the name KIF27 to this new family member with the suffixes A, B and C to designate the splice variants described above (figure 3B). Domain mapping within KIF27 revealed an amino terminal kinesin motor domain and a putative topoisomerase domain located in the center of the protein (figure 2 and 3B). Phylogenetic analysis revealed that KIF27 belongs to the N-5 phylogenetic group of KIFs [26], which includes mouse KIF21A and 21B, and human KIF4 (a group also known as the chromokinesins). Comparisons with the other members of the N-5 subgroup revealed greatest similarity within the motor domain (47–50% identity) (figure 5). In addition, the sequence of the neck domain of KIF27 located at the C-terminal end of the motor domain conformed to the consensus sequence for the neck domain of chromokinesins (figure 3B) [27].

**Figure 1**

Northern analysis of expression patterns of putative novel genes in hamster tissues. 20 µg of total RNA from the indicated hamster tissues were electrophoresed on formaldehyde-agarose gels, transferred to Nylon membranes and hybridized with ³²P-labeled DNA probes derived from putative novel genes. The clone number of each gene is listed to the left of each panel and the positions of the 18S and 28S ribosomal RNAs are indicated to the right. Key: K, kidney; T, testis; SV, seminal vesicle; B, brain; S, spleen; H, heart; L, liver; I, intestine.

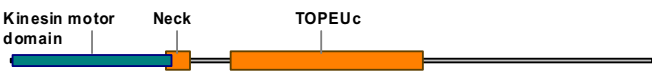
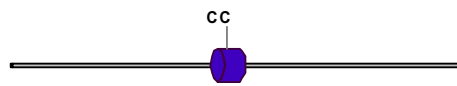
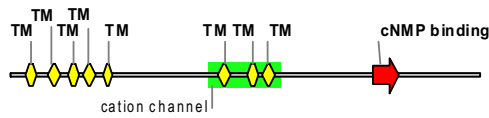
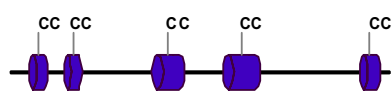
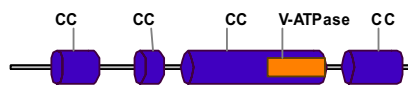

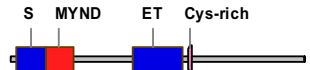
| Clone | Structural features | Size (aas) | Closest relative | Putative function | Approved Name |
|------------|-------------------------------------------------------------------------------------|------------|------------------|--------------------|---------------|
| 1030 (Hs) |  | 1401 | KIF4 | Chromokinesin | KIF27 |
| 7012 | No Motifs | 579 | B29 | Unknown | |
| 9004 (Hs) |  | 836 | None | Unknown | |
| 13043 (Hs) |  | 882 | None | Ion transport | |
| 15014 (Hs) |  | 604 | FLJ13386 | Unknown | |
| 15018 (Hs) |  | 688 | SCAP1B | Unknown | |
| 15037 (Mm) |  | 917 | Otg1 | Unknown | |
| 19045 (Hs) |  | 433 | HSKM-B | Methyl transferase | SMYD2 |

Figure 2

Structural features of proteins encoded by novel genes. The full length coding regions of eight new proteins were determined and the domain structure of the encoded proteins was determined using SMART and Profilescan. The structure of the predicted human protein is shown except for 7012 where no motifs were detected and 15037 where the mouse protein is shown. The size of each protein and its closest relative in protein databases is shown along with a putative function where one could be determined based on structural motifs. Names approved by the HUGO nomenclature committee are indicated for KIF27 and SMYD2, the other clones are referred to as hypothetical proteins in current database entries. The structural motifs detected are indicated on each linear protein. Key: kinesin motor: kinesin motor domain; neck: kinesin neck domain; TOPEUc; topoisomerase domain; CC: predicted coiled coil segment; TM transmembrane region; cNMP binding: cyclic nucleotide binding domain; V-ATPase: V-type ATPase catalytic domain; S and ET: interrupted Su(var)3-9, E(z), trithorax domain; MYND: myeloid transcription factor, nervy, DEAF-I domain.

7012

This clone contained an ORF of 73 amino acids that detected a strong signal by Northern blotting in testis RNA (figure 1). Genome analysis permitted the mapping of a common set of 10 exons on human, mouse and rat chromosomes 5, 15 and 2, respectively (Table 2), assembly of which resulted in translation products of 579, 555 and

563 amino acids, respectively (figure 6). These predictions were also corroborated by a macaque cDNA (AB070167). The encoded proteins did not contain any recognizable functional motifs (figure 3) and were most similar to an uncharacterized human protein named B29 (figure 2). B29 was originally identified as a gene located on human chromosome 18q21 in a search for candidate tumor sup-

Table 2: Putative novel genes identified by sequencing of random hamster testis cDNAs

| Clone number ¹ | Size (bp) ² | Accession Number ³ | Human ⁴ | Mouse ⁵ | Rat ⁶ | Human cDNA ⁷ | Macaque cDNA ⁷ | Mouse cDNA ⁷ | Rat cDNA ⁷ | Fugu cDNA ⁷ |
|---------------------------|------------------------|-------------------------------|--------------------|--------------------|------------------|--------------------------------------------------------|------------------------------------|-------------------------|-----------------------|------------------------|
| 1030 | 241 | BI431001 | 9q21.32 | 13B2 | 17 | AL133654 | AB072777 AB070138 | Cons | Cons | ND |
| 7012 | 221 | BI431002 | 5p12 | 15A1 | 2 | XM_059659 | AB070167 | Cons | Cons | ND |
| 9004 | 264 | BI431003 | 1p13.3 | 3F3 | 2 | AL832216 AK095517 XM_086235 | AB074456 | ND | ND | ND |
| 13043 | 169 | BI431004 | 3q13.2 | 16B5 | 11 | XM_120088 | AB070087 | XM_156183 | ND | ND |
| 15014 | 205 | BI431005 | 11q21 | 9A2 | 8 | AK058122 BC017929 | AB070180 | XM_134746 | ND | ND |
| 15018 | 142 | BI431006 | 4q28.1 | 3B | 2 | NM_144643 AK093361 | ND | AK011907 | NM_153740 | ND |
| 15037 | 178 | BI431007 | 10q26.11 | 19D2 | 1 | AL834368 | ND | XM_140750 | ND | ND |
| 19045 | 212 | BI431008 | 1q32.3 | 1H6 | 13 | NM_020197 | ND | XM_129660 | Cons | Cons |

The eight cDNAs that detected specific signals in Northern analysis were analyzed further. Key: 1: Clone number refers to the position of the particular clone in a 96 well plate, i.e. 1030 is clone 30 on plate 1. 2: Size of original hamster cDNA insert in base pairs (bp). 3: Each sequence has been submitted to dbEST at NCBI. 4, 5, 6: The location on the human, mouse or rat genome sequences that displayed the most significant match to the hamster sequence. 7: Accession numbers for orthologous cDNAs in public databases. ND: not detected. Cons: constructed in this study.

pressor proteins in lung [28]. However, the function of B29 is unknown and it does not contain any obvious functional domains. Interestingly, the apparent size of the detected mRNA is significantly larger than that of the assembled sequences, suggesting that additional exons are likely to remain to be discovered. In this regard, gene prediction software identified an additional 35 exons in the mouse genomic sequence that were partly conserved in the rat but not present in the human sequence. Final mapping of the genomic structure of 7012 will await further refinement of each genomic sequence.

9004

This clone encoded an ORF of 87 amino acids that did not display significant similarity to any known proteins in public databases. Northern blotting revealed a specific signal in hamster testis RNA (figure 1). The clone sequence mapped to human chromosome 1p13.3, mouse chromosome 3F3 and rat chromosome 2 (Table 2). Human and macaque cDNAs have recently appeared in the database that encompass this clone and the sequences of their encoded proteins are compared in figure 7. These polypeptides contain a predicted coiled coil region (figure 2) but do not contain other functional domains that might indicate their possible function(s).

13043

This clone encodes an ORF of 56 amino acids that did not display significant similarity to known proteins. Northern blotting revealed a strong signal in testis (figure 1). Genome comparisons revealed matches on human chromosome 3, mouse chromosome 16 and rat chromosome 11 (Table 2). Human, mouse and rat coding sequences were assembled using a macaque cDNA (AB070087) as

template (figure 8). Domain mapping identified several conserved transmembrane domains in each protein as well as a putative cyclic nucleotide binding site close to the C-terminus (figure 2). In addition a putative cation channel was identified in the center of the protein. Further analysis will be necessary to determine whether this protein functions as a regulatable cation channel.

15014

This clone encodes an ORF of 67 amino acids whose mRNA was specifically detected in hamster testis RNA (figure 1). Genome comparisons revealed strong similarities with regions of human chromosome 11, mouse chromosome 9 and rat chromosome 8 (Table 2). This clone matches to hypothetical proteins recently added to the annotation of the human and mouse genomes (FLJ13386 and XP_134746) as well as a protein encoded by a macaque cDNA (BAB63125). The alignment in figure 9 accounts for each of these clones as well as additional exons predicted in the mouse from inter-genome comparisons. The encoded protein contains several predicted coiled-coil regions but no other identifiable functional domains (figure 2).

15018

This clone encodes an ORF of 47 amino acids that detected a specific signal in hamster testis RNA (figure 1). It mapped to a predicted 20 exon gene on human chromosome 4, with orthologous (but incomplete) sequences on mouse chromosome 3 and rat chromosome 2 (Table 2). An orthologous protein has recently been reported from rat and named sodium channel associated protein 1A (SCAP1A). Alignment of the human and rat proteins is shown in figure 10. Although the specific function of this

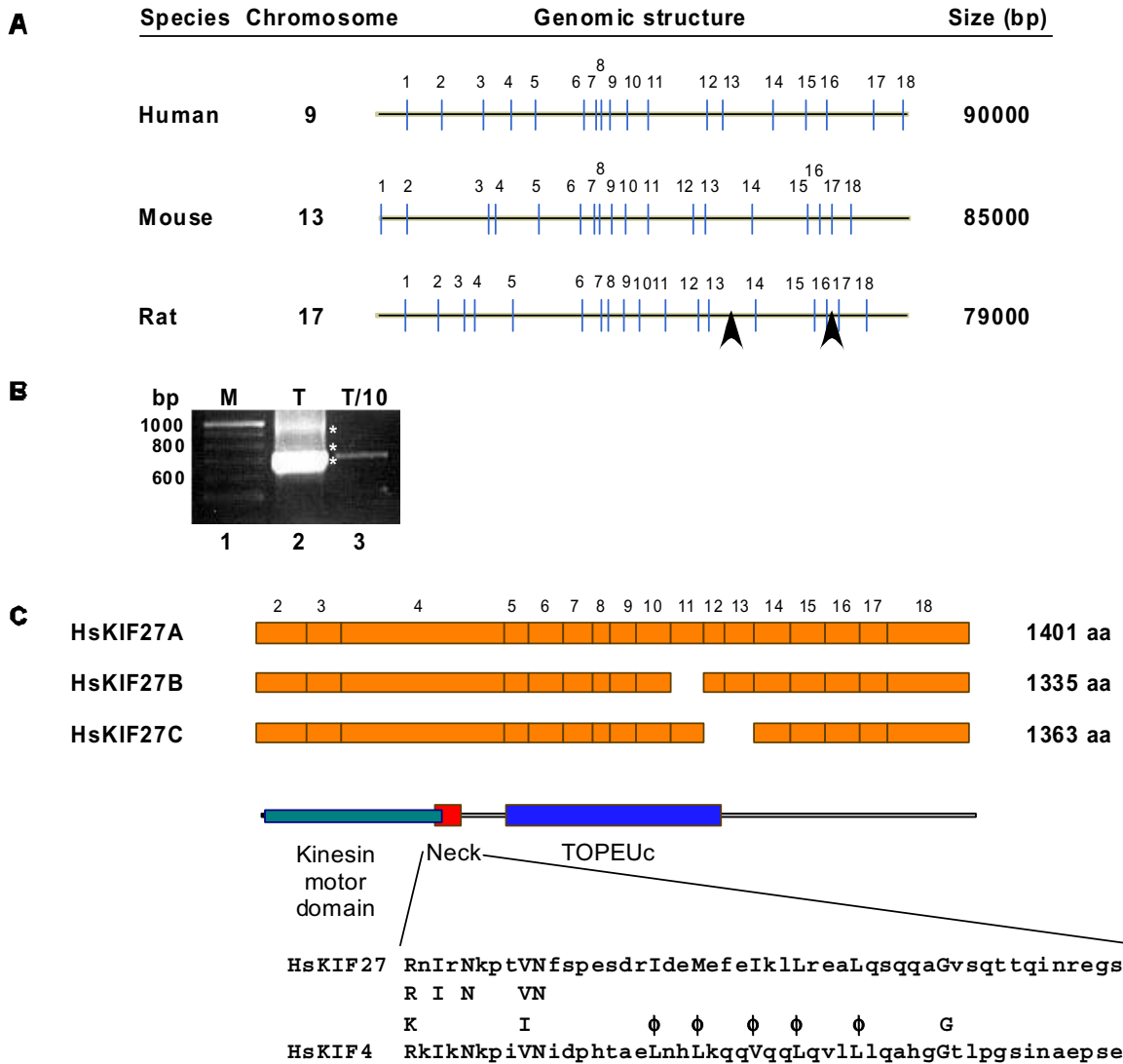


Figure 3

Clone 1030 is derived from a new member of the kinesin superfamily. A. The genomic structure of 1030-derived genes on chromosomes 9, 13 and 17 of human, mouse and rat, respectively, are depicted with numbered vertical bars representing exons. The rat sequence was constructed from four separate contigs (RNOR01052409, RNOR01052410, RNOR01052411 and RNOR01052412, Accession numbers: AABR02079933, AABR02039869, AABR02135902 and AABR02012418) and the arrowheads indicate the positions of short gaps in the genomic sequence. B. PCR analysis of human testis cDNA reveals the presence of three alternatively spliced isoforms. Human testis cDNA was amplified with primers designed to amplify bases 2118–3129 of the predicted full length cDNA based on the genomic structure described above. Three discrete products were observed and are indicated with asterisks. These products were isolated, subcloned and sequenced and corresponded to cDNAs resulting from alternatively spliced mRNAs. The relative proportions of the three products varied in different cDNA preparations although the shortest product was generally the predominant product (data not shown). Key: M: 100bp marker ladder; T: human testis cDNA; T/10: one-tenth of human testis cDNA loaded in previous lane. C. The structure of transcripts derived from the human gene was analyzed by RT-PCR yielding three mRNAs that result from alternative splicing. The exon distribution of each mRNA is indicated (exon 1 is non-coding and was omitted) along with the size of the predicted protein. This protein has now been named KIF27 with a prefix to indicate the species and a letter to indicate the splice variant. Kinesins are characterized by a ~300 amino acid motor domain that is located at the N-terminus of KIF27. The N-5 subgroup of kinesins also possesses a characteristic neck domain at the C-terminal end of the motor domain. A comparison of the KIF27 neck domain with the corresponding domain of KIF4 is shown, with conserved amino acids indicated between the sequences and in upper case letters within each sequence. Key: ϕ: hydrophobic residue.

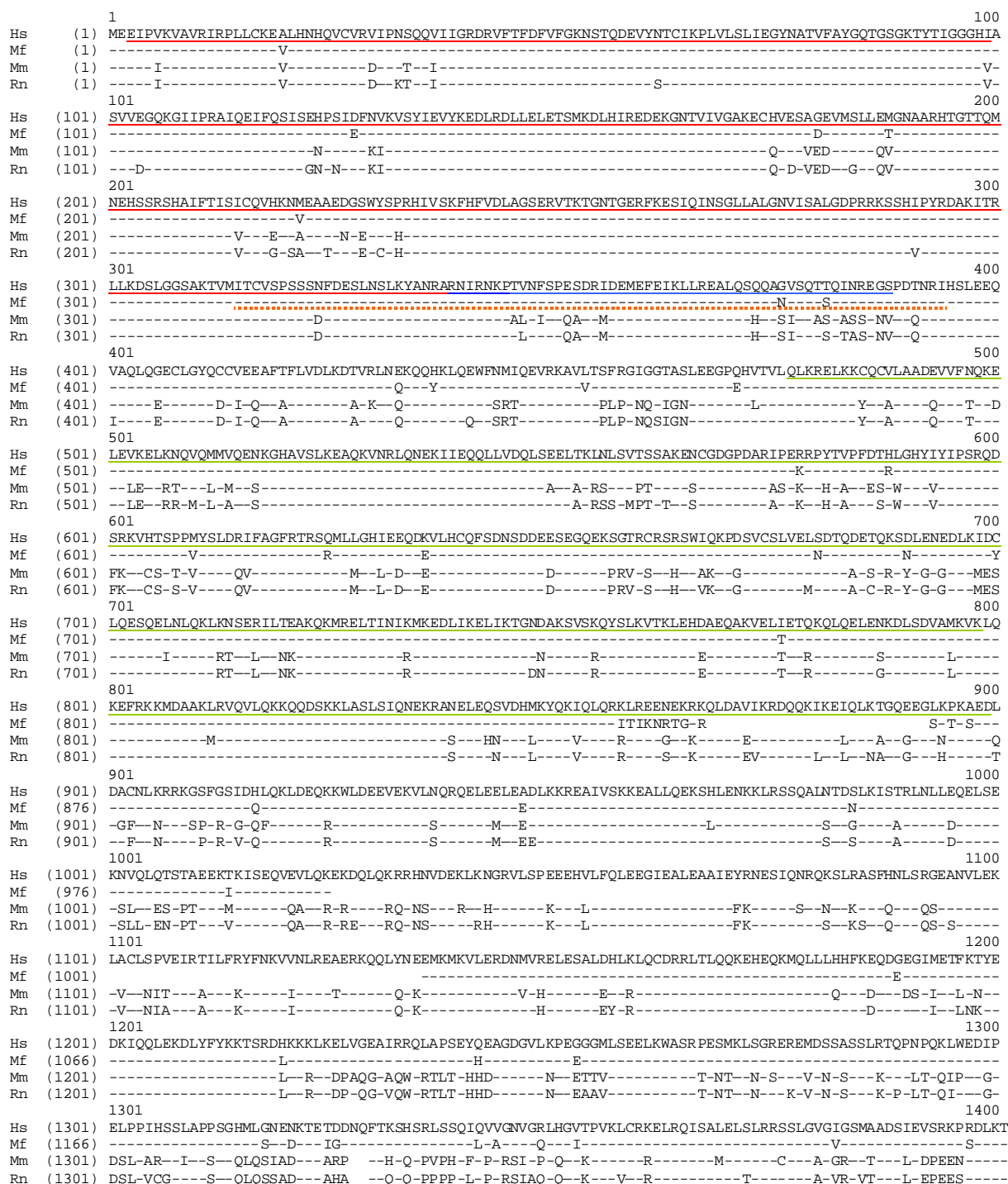


Figure 4

Alignment of human, macaque, mouse and rat KIF27 sequences. The derived amino acid sequences of human, macaque, mouse and rat KIF27 were aligned using the AlignX program from the Vector NTI suite of sequence analysis programs. The KIF27A sequence from human is shown and identical amino acids within the orthologous proteins are indicated with a dash. Gaps are indicated as spaces within the sequence. The motor domain is indicated with a red underline, the neck domain with a blue underline and the topoisomerase domain with a green underline. The segment corresponding to the original hamster clone is indicated with a red dotted underline. Key: Hs: *homo sapiens*; Mf: *macaca fascicularis*; Mm: *mus musculus*; Rn: *rattus norvegicus*.



Figure 5

Alignment of the motor domain of KIF27 and other members of kinesin N-5 subgroup. The motor domain of human KIF27 was aligned with the corresponding domains of human KIF4A (NP_046332) and mouse KIF21A (NP_057914) and KIF21B (NP_064346). The motor and neck domains are underlined in red and blue, respectively. The sequence of each protein is shown and blocks of similarity are shaded.

protein is currently unknown, it contains a potential V-type ATPase domain as well as several predicted coiled-coil regions (figure 2).

15037

This clone encodes an ORF of 59 amino acids that was detected specifically in hamster testis (figure 1). It mapped to specific regions on chromosomes 10, 19 and 1 in human, mouse and rat, respectively. A comparison of the predicted human protein based on several partial human cDNAs and a mouse protein named oocyte-testis gene 1 (Otg1) is shown in figure 11. The protein sequence is predicted to contain several coiled coil regions but no other potential functional domains (figure 2).

19045

This clone detected two distinct transcripts in testis, brain and heart by Northern blotting (figure 1). The sequence mapped to regions on chromosomes 1, 1 and 13 in human, mouse and rat, respectively and we have assembled the orthologous sequences from human, mouse, rat and Fugu (figure 12). The protein is 433 amino acids in length in human, mouse and rat (434 in Fugu) and contains two recognizable functional domains characteristic of proteins with chromatin remodeling activity (figure 2). The first is an interrupted SET (Su(var)3-9, E(z), trithorax) domain [29] and the second is a MYND (mye-

loid transcription factor, nervy, DEAF-1) domain [30]. A similar domain organization is also found in the BOP (CD8bopposite, recently renamed the SET and MYND domain protein (SMYD)) family of proteins and alignment of the human protein identified here with SMYD1 (BOP1) from mouse and chicken is shown in figure 13. The SMYD1 proteins are transcriptional regulatory proteins with chromatin-remodeling activities important in cardiac tissue, muscle and T lymphocytes [31-33]. Therefore, the new protein identified in this report is likely to play as yet undefined roles in transcription and has been assigned the official name SMYD2 to designate its relatedness to SMYD1.

Discussion

In this report, we describe the sequencing and initial characterization of greater than 700 randomly selected ESTs from the hamster testis. This represents the first such study carried out in hamster, as dbEST listed just twenty-seven entries from hamster in January 2003 (release 012403). It has been widely speculated that the sequencing of additional mammalian genomes will aid in the annotation of the human genome, particularly in the identification of previously unidentified coding regions through the mapping of conserved regions in different genomes [34]. We describe here our initial characterization of eight genes that were not annotated on human genome sequences at



Figure 6

Alignment of human, macaque, mouse and rat 7012 sequences. The sequences of predicted peptides for human (Hs), macaque (Mf), mouse (Mm) and rat (Rn) orthologues of 7012 were compared using AlignX. Portions of the human sequence were obtained from XP_059659, a peptide predicted from annotation of the human genome at NCBI. The macaque sequence was derived from Genbank entry BAB63112, the predicted translation product of a cDNA isolated from macaque testis. The mouse sequence was obtained by mapping the exon structure of the orthologous gene on the region of mouse chromosome 15. The rat coding sequence was constructed by assembling conserved exons on contigs RNOR01068092, RNOR01068093 and RNOR01068094 (Accession numbers AABR02034780 and AABR02122391). The sequence corresponding to the hamster clone obtained in this study is indicated with a dotted red underline in the human sequence.

the beginning of our study. Although predicted structures for some of these genes have appeared recently, our data represents the first experimental verification of their existence in several cases, particularly in the testis. We were unable to predict specific functions for the proteins encoded by six of the genes, however, the other two genes encoded a new member of the kinesin superfamily (KIF27) and a protein predicted to play a role in chromatin remodeling.

KIF27: a new kinesin family member

Our studies revealed the existence of a new member of the kinesin superfamily of microtubule-based molecular motors, which we have named KIF27. KIF27 RNA was detected primarily in hamster testis but weaker signals were also present in several other tissues, suggesting that this protein may function in numerous cell types. Significant characteristics of the KIF27 genes mapped in human, mouse and rat included a conserved 18 exon arrangement

and the existence of at least three mRNAs that resulted from alternative splicing, at least in human. Although partial cDNAs existed for the human and macaque genes in public databases, we were able to construct full length cDNA sequences for human, mouse and rat from genomic comparisons, and to assemble a corresponding sequence from macaque by joining two previously reported cDNAs. Kinesins are characterized by a conserved motor domain of approximately 350 amino acids that may be located at the amino terminus (KIN-N), carboxy-terminus (KIN-C) or within the polypeptide (KIN-I) [27] and the motor domain of KIF27 is located in the N-terminus of the protein. Based on this arrangement and the sequence of the adjacent neck domain, we assigned KIF27 to the N-5 phylogenetic group of kinesins [27]. The N-5 subgroup is defined by the human KIF4 protein, which is primarily localized to the nuclear matrix and associates with chromosomes during mitosis [35]. In cell division, nuclear kinesins of this chromokinesin class appear to be

```

1
Hs (1) MDEADFSHTTYKQEDLPYDGLDLSQIKIGNDYSFTSKKDGLEVLNQIIFIADDPQEKAMHNETCGNTAVTIPLGKITENAANKKDEKEKQCTAALHIPAN 100
Mf (1) -----D-----P--C-----N-AF--SS-----M-M---T-----ET
101
Hs (101) EGDASKSSISDILLHLSKEPFLRGGIDCETLPEISNADSFEEEAIIKSIISCYNKNSWPKEQTPPELTDQLNPKRDGENSNKPGSATTTEENTSDLEGP 200
Mf (101) -----C-----L-----K-----P-M-----ET
201
Hs (201) VAAGDSSHQENVNVLTKTKGPGDKQKSYQGQSPQKQTEKANSNGNTFKYGGQVHYQLPDFSKIAPKVKIPKNKIINKPLAIAKQASFSKSRDKPTLVQ 300
Mf (201) -----I-----N---A-P--H---S---R-----E-N---N-----V---L-E--AV--
301
Hs (301) DSLETMPESNCVEKQHQEQKGIKITEPSQIQMEPIVHIHQELLTGIESEASLSKLSPTSQKGTSSSSSYIFQKISQKQKMCQKLEQTDQLKTKVQEF 400
Mf (300) -----I-----E-----T-----I-----R-----
401
Hs (401) RIKQDSPYHLQDKKLVLEKLGHLLEQNFLATKDKHLTLQQQVHKHESTIVGDFDPERKVEGEIFKLEMLLEDVKEKMDSEKYTSAPSLPVSSPVTL 500
Mf (400) -----E-----S-----D-V-----
501
Hs (501) DLASAFSSLSNEIPKEHPGHPGPRGSGGSEVTGTPQGGPQEAPEELCELAPQTYLNGHYGDAAAQNKPDQVAMRLSSNSGEDPNGTPRRQDCAEMTAP 600
Mf (500) ---TS-----V-E-----P-----A-----K-----T-I-----TV--Q-----RC--G-----T---
601
Hs (601) SPSCAFRRLLEWKQKVEKKHGRINCGRFSIVLHEKAPHSDSTPNSDTGHSFCSDSGTEMQSNKCQDCGKTIPTSRACRKEPTKEFHYRYNTPGQNY 700
Mf (600) -----P-----R-H-----TV-----G-----P-----H---E---
701
Hs (701) NHSKRGAFFVPHSLDEKNSPSPFLKPKRICKSORVNSKSFKGEHEPTPGKKKLOAFMTYSSDPATPSPHFYSCRISGSKSLCDFDSTEBIKSEILNSAL 800
Mf (699) -----R---H-----LAS- F-CPGLMYSPTSKSS---WQEAELGLENMKSQ
801 836
Hs (801) HALRTATILKETTDQMIKTIAEDLAKAQRWRNRLKY

```

Figure 7

Alignment of human and macaque 9004 sequences. Clone 9004 was found to display significant similarity to cDNAs derived from human (Hs) and macaque (Mf). The human protein is the predicted translation product of AL832216 and the macaque protein is the predicted product of AB074456. The complete human cDNA mapped to a 15 exon gene on chromosome 1. Only eleven of these fifteen exons could be detected in the mouse and rat genome sequences (not shown). The sequence corresponding to the hamster clone is underlined.

important for the maintenance of sister chromatids on the metaphase plate [36]. However, additional functions for nuclear kinesins have recently been uncovered. For example, KIF17a, a member of the N-4 subgroup that is predominantly expressed in germ cells, was recently shown to possess transcriptional regulatory properties by controlling access of the transcriptional activator protein CREM to a coactivator, ACT [37]. Clearly, functional analysis of the KIF27 polypeptide will be needed to determine the subcellular location of this protein to determine whether it may add to the growing number of kinesins that function in the nucleus. In this regard, although the carboxy terminal regions of KIF27 and KIF4 display little significant sequence similarity, both contain a putative topoisomerase domain that may be important for nuclear functions (figure 2). In addition, several clusters of basic amino acids that may function as nuclear localization signals are located in the KIF27 polypeptide.

SMYD2: a putative chromatin remodeling protein

The final clone isolated in our search encodes a 433 amino acid protein whose sequence was highly conserved from human to pufferfish. This protein, now named SMYD2, contains SET and MYND domains that are characteristic of proteins with chromatin remodeling capabil-

ities. The SET domain is a common feature of proteins with histone lysine methyl transferase (HKMT) activity and has been identified in hundreds of proteins in organisms ranging from bacteria and viruses to humans [38]. The SMYD2 SET domain is separated into two parts (i.e. a S-ET domain) and is followed by a short cyteine-rich region that is common in many SET domain proteins (figures 2 and 12). The MYND domain is located between the two halves of the SET domain and similar domains have been identified in a number of proteins that function as transcriptional repressors, including the ETO protein that is fused to the AML-1 transcription factor in the t(8;21) translocation in acute myeloid leukemias [39]. The MYND contains two zinc finger motifs and is a protein-protein interaction interface responsible for the recruitment of corepressors [39-41].

This domain organization is conserved in several proteins in public databases, including the recently described SMYD1 (aka BOP) family of proteins [31]. Three isoforms of SMYD1 have been reported thus far (referred to by their original names, m-BOP1, m-BOP2 and t-BOP) that are products of a single gene that result from either alternative splicing or promoter usage [31-33]. m-BOP is essential for cardiac differentiation and morphogenesis while t-

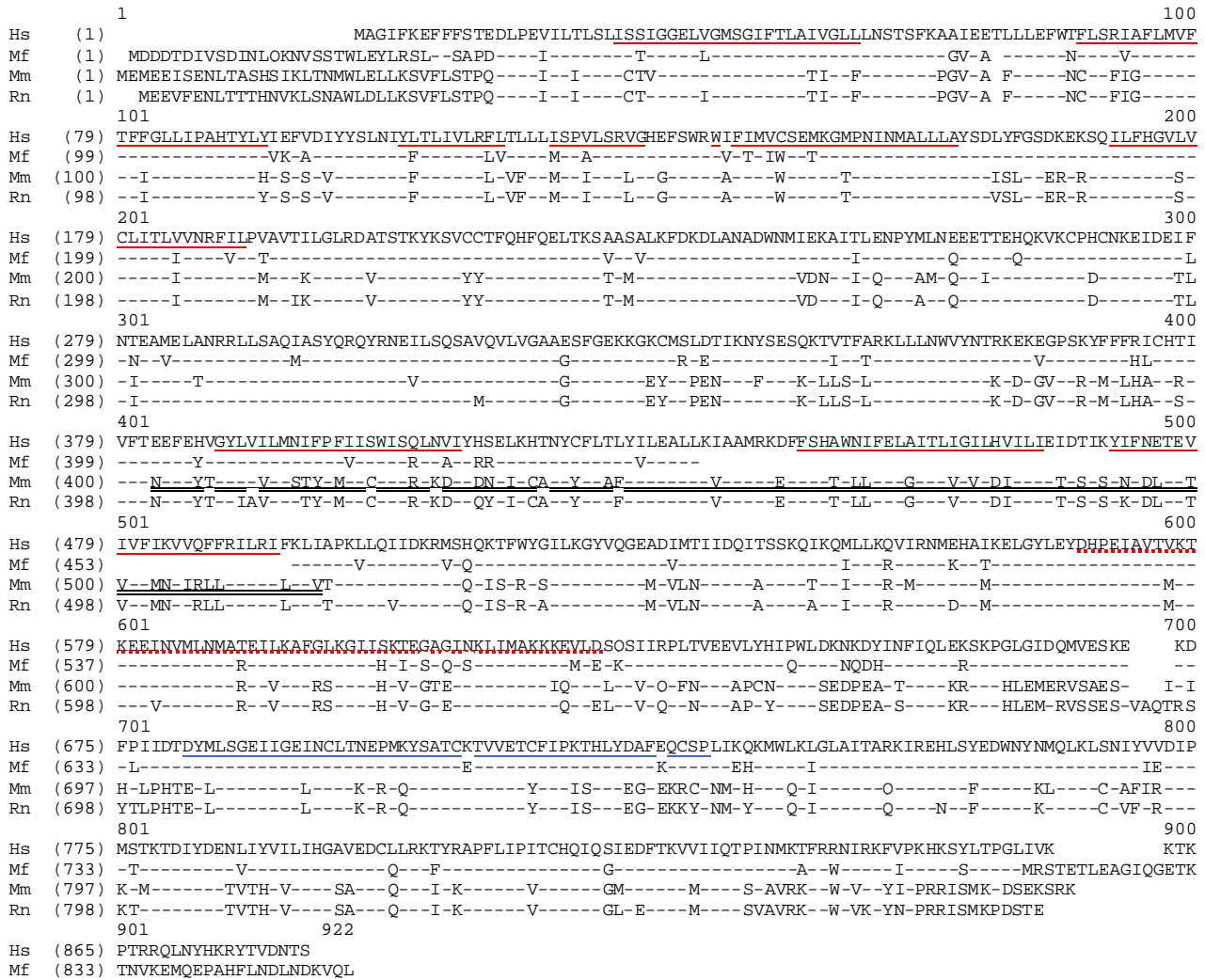


Figure 8
 Alignment of human, macaque, mouse and rat 13043 sequences. This alignment was anchored on a single cDNA isolated from macaque (AB0770087). This sequence was compared to the human (Hs), mouse (Mm) and rat (Rn) genomic sequences at Ensembl and individual exons were assigned. One additional potential exon was assigned based on similarity between genomic sequences based on PipMaker analysis. The human and mouse sequences listed here match partially with predicted proteins in both human and mouse (ENSP0000306627 and XP_156183, respectively) but conform more to the template predicted by the macaque sequence. The rat sequence was constructed from genomic sequences AABR02142366, AABR02119574, AABR02127843, AABR02117677 and AABR02097473. Predicted transmembrane domains are underlined in red in the human sequence and a potential nucleotide binding domain is underlined in blue. A putative cation channel in the human, mouse and rat proteins is double underlined in the mouse sequence. The region of similarity to the original hamster clone is indicated with a red dotted underline.

BOP is expressed in cytotoxic T lymphocytes. Studies are currently underway to examine the function(s) of SMYD2 in testis and heart.

Genomic studies in the hamster
 The impetus for this study arose from the need to perform microarray experiments to investigate the molecular changes elicited by environmental toxicants on male reproduction function in hamster. Comparisons performed between the limited numbers of hamster

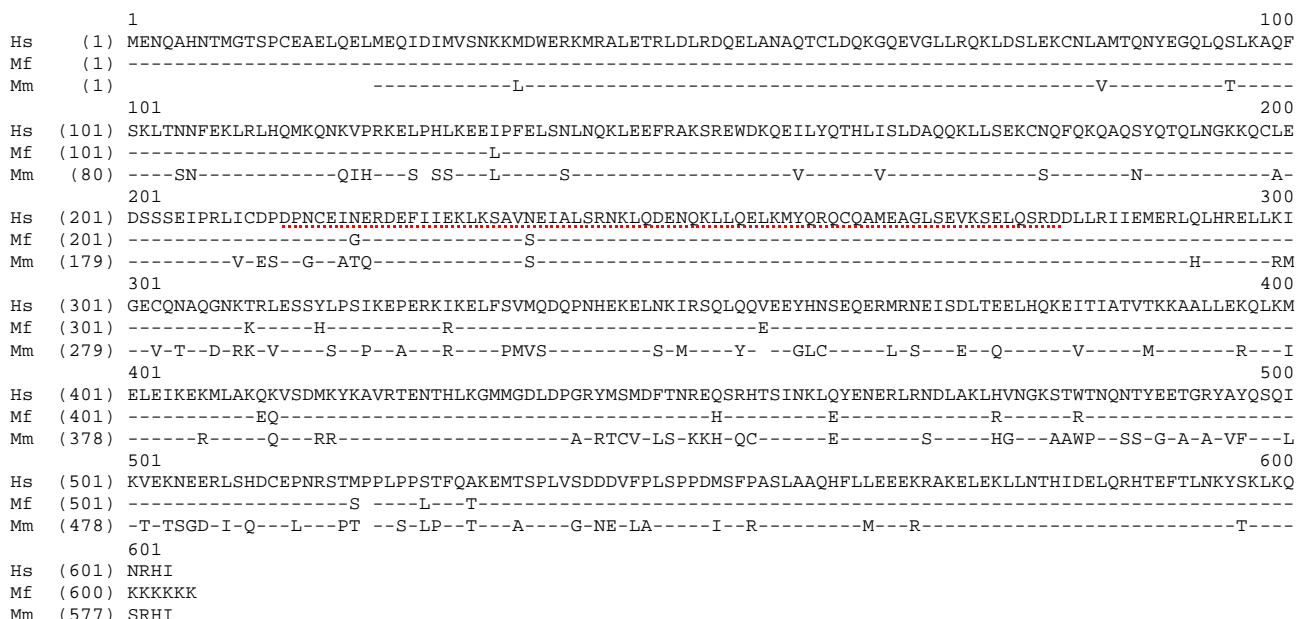


Figure 9
 Alignment of human, macaque and mouse 15014 sequences. The sequences of human (Hs), mouse (Mm) and macaque (Mf) 15014 orthologues were compared in Align X. The human sequence was derived from XP_089976 (LOC159989). The mouse sequence was obtained from a combination of XP_134746 (LOC234964), a hypothetical protein predicted from the NCBI annotation process along with additional exons predicted from comparisons between the human cDNA and the mouse genomic sequence. The macaque sequence was obtained from Genbank entry BAB63125. Gaps in the human sequence indicate the likely presence of alternative exons that may undergo alternative splicing. This hypothesis was supported by exon mapping in both the mouse and rat genome sequences (data not shown). We were unable to detect additional exons in the mouse genomic sequence that might encode the N-terminal sequences in the human protein sequence. The similarity to the hamster clone is underlined.

sequences in public databases with orthologous sequences from the mouse suggested that, despite the close taxonomic relationship between hamster and mouse, evolutionary divergence in coding sequences was sufficiently great in certain cases that reagents developed for the mouse would be of limited use for genomic studies in the hamster. This conclusion was supported by experimental observations in our laboratories indicating that probes derived from rat and mouse cDNAs yield inconsistent results in Northern blotting under stringent hybridization conditions (data not shown). The reagents described here will now permit the initiation of genomic studies in the hamster.

Methods

RNA, cDNA and plasmid preparation

Total RNA was prepared from testes of adult Syrian golden hamsters (*Mesocricetus auratus*). Poly A⁺ RNA was isolated using the poly A Spin[™] mRNA isolation kit (New England Biolabs, Beverly, MA). 5 µg of polyA⁺ RNA was converted

to double stranded cDNA using the cDNA Synthesis Kit (Life Technologies, Gaithersburg, MD). The cDNA was then digested to completion with *DpnII* and electrophoresed on a 1.5% agarose gel. Five populations of digested cDNAs in size ranges between 100 and 800 bp were excised from the gel and purified using Qiaex II resin (Qiagen, Valencia, CA). Each population of cDNA was ligated into a pBluescript vector that had been digested with *BamHI* and alkaline phosphatased to decrease the rate of self-ligation. The ligations were transformed into DH5α supercompetent cells (Life Technologies) and positive clones were identified by blue-white selection. The success of the cloning procedure was initially monitored by picking 5 clones from each group, preparing plasmid DNA and sequencing. These preliminary studies indicated that the most useful clone sets were those derived from cDNAs in the 100–300 bp range. White colonies from these two sets were carefully picked and used to inoculate single cells of a 96 well culture block with each well containing 1.2 ml of TB (1.2% Tryptone, 2.4% yeast

```

1
Hs (1) MAAEIDFLREQNRRLNEDFRRYQMESFSKYSSVQKAVCQEGEDDTFENLVFDQSFAPLVTEYDKHLGELNGQLKYYQKQVGMKQLQLENVIKENERLHS 100
Rn (1) --T--L--D--VK--DIL-QH-I-HIFRDPAM-NSMSK-GRG--LT-S-N--A-P--IA--E--E--R--T--HM-----T--T-----
101
Hs (101) ELKDAVEKKLEAFPLGTEVGTDIYADDETVRNLQEQLQLANQEKTQAVELWQTVSQELEDRLHKLKLYQEHMTEAQIHFVESQKQKQDLDFDQQLTKQLHVTN 200
Rn (101) K-----Q---L-F--GI-N--C-----I-----NW--Q---A---ESVQ-----NR-----NN-----K--A-
201
Hs (201) ENMEVTNQQLKTVTEQSVIEQLRKKLRQAKLELRVAVAKVEELTNVTEDLQGMKKKKEKDVVSAHGREASDRRLQQLQSSIKQLEIRLCVTIQEANQ 300
Rn (201) --I-M--HH-----NME--K---Q---D---T---K---G--E--L--E-IM--Q-K-----V-----S---IA---V
301
Hs (301) LRTENTHLEKQTRRELQAKCNELENEREYEAIVRARNMQLLEEANLQKSQALLEEKQKEEDIEKMKETVSRFVQDATIRTKKEVANTKKQCNIQISRLTEE 400
Rn (301) -K-GK-Q---IK-----S-S--K---S--D-----IKQN-I-----VER---K-I-HLI--A-KAR--ES---YEV-L-Q-K--
401
Hs (401) LSAQMECAEKQGGIERVIKKKAVVEEELEKLYREGFCNECDYRKLEMHOFFLVSRSKDDLQRLRLTRAENRIKQLETDSSEEISRYQEMIQKLQNVLE 500
Rn (401) -----D-D-----D-A-RG-R-----KQD-G-----C-AA-----KT-----IN-----SH-----T---
501
Hs (501) SERENCGLVSEQRLLKQENKQLRKETESLRKIALEAQQKAKVKISTMEHEFSIKERGFQVQREMEDSNRNSIVELRHLLATQQKAANRWKEETKKLTF 600
Rn (501) -----F-----E-Q---D---V-----L-V---Q---H-----A---T-----
601
Hs (601) SAEIRINNLSKSELSRQLHTQELLSQLEMANEKVAENEKLILEHQEKANRLQRRLSQAEEAASASQQLSVITVQRKAASLMNLENI 688
Rn (601) ---M--SS-----M-----

```

Figure 10

Alignment of human and rat 15018 sequences. The human (Hs) sequence is derived from the predicted protein FLJ30655 and the rat (Rn) sequence from SCAP1A (NP_714962). The similarity to the hamster clone is underlined.

```

1
Hs (1) MSETDHIASTSSDKNVGKTPELKEDSCNLFSGNESSKLENEKLLSLNTDKTLCQPNEHNRIEAQENYIPDHGGGEDSCAK 100
Mm (1) MKIRSRFEEMQSELVPS---E---I---ATTE--S--RD---ISV--D---R--TGAE---DS-RI---T---CSQ--V--SH--C-S--N---N
101
Hs (83) TDTGSENSEQIANFPSPGNFAKHISKTNETEQKVTQLIVLELRSSTFPPEANEKTYSESPYDTCTCKFISKIKSVASEDLLEEIESELSTEFAE HRVP 200
Mm (101) ---CP-D-G--DD--G-D-TEQV---K-P--T---A--K--APA-A--P--A-A-L-----R-L--EM-T---D---G-----A---G-Q--
201
Hs (182) NGMNKGEHALVLFEKCVQDKYLQQEHIIKLKIKENKKHQELFVDICSEKNDLREELKKRTETEKQHMNTIKQLESRIEELNKEVKASRDQLIAQVDTAKN 300
Mm (201) --L---Q--A-----HSR-----LTV-Q-----N---ILN-----S---R-----L-----I---K---V-----
301
Hs (282) AVQQLHKEMAQRMEQANKKCEEARQEKEAMVMKYVRGEKESLDLRKEKETLEKKLRDANKELEKNTNKIKQLSQEKGRLHQLYETKEGETTRLIREIDKL 400
Mm (301) -I--I-----D-----A-----R---S-----Q---S-----E--
401
Hs (382) KEDINSHVIKVKWAQNKLKAEMDSHKETKDKLKETTTKLQAKEEADQIRKNCQDMIKTYQESEEIKSNELDAKLRVTKGELEKQMQEKSDQLEMHHAKI 500
Mm (401) --EM--Q-----E---Q-----
501
Hs (482) KELEDLKRTFKEGMDELRTLRTKVCLEDERLRTEDELSKYKEIINRQKAEIQNLLDKVKTADQLQEQLQRGKQEIENLKEEVESLNSLINDLQKDIEGS 600
Mm (501) -----A-----R-----S-----IT--H---S-----H---M-----
601
Hs (582) RKRESELLFTERLTSKNAQLQSESNLSQSPDKVSCSESQLQSQCEQMKQTNINLESRLLKEEELRKEEVQTQAELACRQTEVKALSTQVEELKDELV 700
Mm (601) -----K-----SA---V-NL-----Q--G-A-R---K-----SAA-----
701
Hs (682) TQRRKHASSIKDLTKLQLQARRKLDQVESGSYDKEVSSMGSRSSSSGSLNARSSAEDRSPENTGSSVAVDNFPQVDKAMLIERIVRLQKAHARKNEKIEF 800
Mm (701) -----NV---S-----E-T-N-NH--DI-----I-----S-----E-----
801
Hs (782) MEDHIKQLVEEIRKTKIIQSYILREESGTLSSEASDFNKVHLSRRGGIMASLYTSHPADNGLTLELSLEINRKLQAVLEDTLLKNITLKENLQTLGTETI 900
Mm (801) -----V-----S-----
901
Hs (882) ERLIKHQHELEQRTKKT 917
Mm (901) -----A-----

```

Figure 11

Alignment of human and mouse 15037 sequences. The human (Hs) sequence is based partly on cDNA sequences AL834368, AF273054 and AK057508 with 5' coding sequences based on conservation to mouse cDNA and genomic sequences. The mouse (Mm) sequence is derived from oocyte-testis gene I (Otg1, NM_170757). The region of similarity to the hamster clone is underlined.

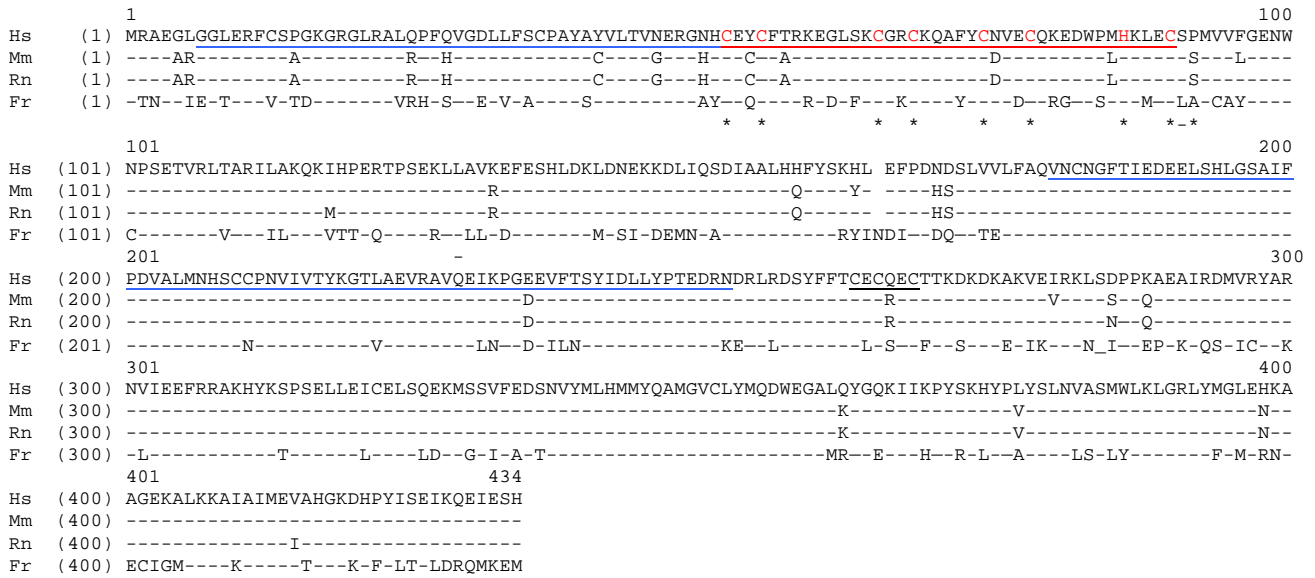


Figure 12

Alignment of human, mouse, rat and pufferfish SMYD2 sequences. The human (Hs) sequence was derived from HSKM-B (NM_020197) and the mouse (Mm) sequence from BC023119. The rat (Rn) sequence was established by mapping exons in rat genomic sequences (RNOR01035016 and RNOR01035017; Accession numbers: AABR02111753 and AABR02114803) and the Fugu (Fr, *fugu rubripes*) sequence using a similar strategy in Fugu genomic sequences (Scaffold_6078.1 and Scaffold 1370.1). The interrupted SET domain (S-ET) is indicated with a blue underline and the MYND domain with a red underline. The Cys and His residues in the MYND domain predicted to form two zinc fingers are indicated in red type and with an asterisk and the Cys-rich region following the SET domain is underlined in black.

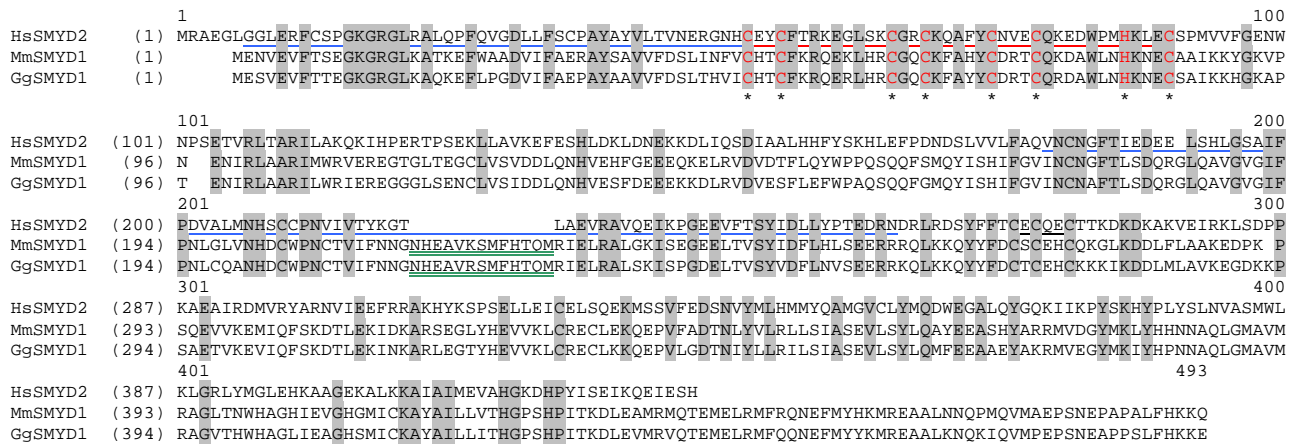


Figure 13

Comparison of SMYD2 to members of the SMYD1/BOP family of chromatin remodeling proteins. The human SMYD2 sequence was aligned with the sequence of mouse BOPI (AAC53021) and chicken BOPI (AAL31880). The S-ET and MYND domains and the Cys-rich region are labelled as in figure 12. Conserved residues are shaded and the sequences lacking from the BOP2 splice variants in mouse and chicken are indicated with a double green underline.

extract, 0.4% glycerol) supplemented with 50 µg/ml carbenicillin. The bacteria were grown for 20 hours at 37°C and glycerol stocks were subsequently prepared using a Biomek 3000 robot (Beckman-Coulter, Inc., Fullerton, CA). Fresh cultures (in 1 ml of 2X LB (2% Tryptone, 1% yeast extract, 1% NaCl) plus 50 µg/ml carbenicillin) were inoculated from the glycerol stocks using a 96-well needle transfer device and grown as before. Plasmid DNA was prepared using the full lysate protocol of the Montage™ Plasmid Miniprep96 Kit (Millipore, Bedford, MA). Typical yields of plasmid DNA ranged from 5–10 µg in 50 µl volume.

Sequencing

Plasmid DNA was diluted four-fold in water and 2 µl of the diluted sample was sequenced in 96-well format using the Dye Terminator Cycle Sequencing Kit (Beckman Coulter, Inc.). The reactions were cleaned up by an ethanol precipitation step in which 35 µl of 10 mg/ml glycogen, 70 µl of 0.5 M EDTA, 70 µl of 1.5 M Sodium Acetate (pH 4.8) was first mixed with the sequencing reaction (6.6 µl). 20 µl of 95% ethanol was added and the plate was centrifuged at 4,000 rpm for 45 minutes. The pellets were washed twice with 100 µl of 70% ethanol, air dried and dissolved in 25 µl of molecular biology grade formamide. A drop of mineral oil was overlaid on each reaction and these were then run on a CEQ2000 capillary array sequencer (Beckman Coulter, Inc.).

Sequence analysis

Raw sequence data was first imported into the Contig Express component of the Vector NTI suite of sequence analysis programs (InforMax, Inc. Bethesda, MD). Each clone was named according to its position in the original 96-well plate, for example, clone 1030 came from position 30 in plate 1. Vector sequences were first scanned in batch mode for the presence of vector-derived sequences using Contig Express and these sequences were trimmed before proceeding. In some cases, manual processing of the sequences was necessary using the VecScreen program at the National Center for Biotechnology Information (NCBI) <http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html>. The presence of *DpnII* sites at either end of the clone simplified the removal of vector sequences. The sequences were concatenated in batches of 96 into a text file using FASTA format and submitted for a batch BLASTN search using the interface program BLASTCL3. This search returned the three top matches into an output file. No further analysis was performed in those cases where a match with a known gene was clearly established, and the clone was annotated as the hamster orthologue of a known gene. A series of additional comparisons were performed if a clear relationship between a hamster clone and known sequence was not established by this initial search. The most useful additional comparison was a

BLASTX search, which compares the translated sequence of the input clone in all six frames to the peptide sequences in GenBank. This search clarified the identity of several additional hamster clones. 200 clones did not match to known sequences and were analyzed to determine whether they might represent potentially novel genes. First, the output sequences were translated in all frames to determine whether a complete open reading frame (ORF) was present. If so, the clones were resequenced on both strands to ensure that the sequence was correct. Promising sequences were compared to the public version of the annotated human sequence (v 9.30a.1) at Ensembl <http://www.ensembl.org> to find matches with human chromosomal sequences. Further comparisons were performed against the mouse (v 9.3a.1), rat (v 9.1.1), zebrafish (v 9.08.1), pufferfish (v 9.1.1) and mosquito (v 9.1a.1) genome sequence in the same database. When matching genomic loci were identified, approximately 100,000 bp of genome sequence surrounding the match was imported into Vector NTI and submitted to gene prediction software programs to determine if the hamster sequences were located within exons of predicted genes. Two gene prediction programs were used FGENESH <http://www.softberry.com/berry.phtml> and GENEMARK <http://opal.biology.gatech.edu/GeneMark/eukhmm.cgi?org=H.sapiens>. Predicted cDNA sequences were then assembled into contiguous files and subjected to further comparisons against cDNA and EST databases at NCBI. Genome structures were further examined using PipMaker, a program that supports comparisons of large DNA fragments and identifies short conserved regions of similarity, such as exons. The predicted protein sequences of the derived cDNAs were analyzed for the presence of functional domains using the conserved domain function of BLAST at NCBI as well as the Simple Modular Architecture Research Tool at <http://smart.embl-heidelberg.de>. Protein structures were annotated in Vector NTI and published in Canvas v7.0 (Deneba Systems, Inc., Miami, FL). Sequences were submitted to the appropriate databases at the National Center for Biotechnology Information (NCBI). Specifically, EST sequences were submitted to dbEST under accession numbers BI431001-BI431008 and CB884447-CB885166. Human KIF27 sequences were submitted to Genbank under accession numbers AY237536-AY237538. Sequences defined by annotation of previously available sequences were submitted to the Third Party Annotation database under accession numbers BK001053-BK001057 and BK001326-BK001332. The Human Genome Organization (HUGO) nomenclature committee has approved proposed gene names.

Northern analysis

Total RNA was purified from various hamster tissues using TriZol reagent (Life Technologies, Inc.). 20 µg of each RNA was electrophoresed through a 1% agarose-formal-

dehyde gel and transferred to a Nylon membrane (Micron Separations, Westborough, MA). The RNA was cross-linked to the membrane by exposure to UV light and hybridized with specific probes labeled with ^{32}P -dCTP (Perkin Elmer Life Sciences, Boston, MA) using the Prime-It random prime labeling kit (Stratagene, La Jolla, CA). The hybridization procedure was performed as described before [42]. Several duplicate membranes were prepared and the radioactive probe was stripped after each round of hybridization by boiling for 10 minutes in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1% SDS. The specific probes were prepared by digesting the appropriate pBluescript plasmid with *Xba*I and *Eco*RI, separating the fragment on a 1.5% agarose gel followed by extraction using the Qiaex II extraction kit (Qiagen).

PCR cloning of human KIF27 cDNAs

PCR primers based on the predicted human KIF27 genomic sequence were designed for PCR cloning of overlapping regions of human KIF27 cDNAs. PCR reactions were performed using human testis Marathon RACE-Ready cDNA (Clontech, Palo Alto, CA) using conditions described previously [43]. PCR products were subcloned into the pGEM-Teasy plasmid and sequenced. The sequences were then assembled into full length cDNAs using the Vector NTI sequence analysis suite of programs. The primers used for the reaction shown in figure 2 were: 5' AACTAGATGTAGAAGTCGTTTCATGGATTCC 3' and 5'TTCCAGTAAGTTCAGGCGAGTTG 3'.

Authors' contributions

S.O. performed cDNA cloning of the KIF27 isoforms and performed sequence analysis

J.L.C. characterized and analyzed the cDNA library

S.K. performed Northern analysis

W.J.H. provided hamster tissues and assisted in sequence analysis

S.A.K. assisted in sequence analysis

S.C.W. designed the project, performed cDNA synthesis and sequence analysis

All authors read and approved the manuscript

Acknowledgements

We appreciate the assistance of Natalya Klueva in the Texas Tech University Center for Biotechnology and Genomics for assistance with sequencing and sequence analysis. We thank Dan Hardy for providing human testicular cDNA for PCR cloning of human KIF27. We also thank Curt Pfarr and Demet Nalbant for critical reading of the manuscript. This project was supported by National Institutes of Health, NIEHS grant number ES 10232 to S.A.K.

References

1. IHGS Consortium: **Initial sequencing and analysis of the human genome** *Nature* 2001, **409**:860-920.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA and Holt RA: **The sequence of the human genome** *Science* 2001, **291**:1304-51.
3. Aparicio S, Chapman J, Stupka E, Putnam N, Chia J-m, Dehal P, Christoffels A, Rash S, Hoon S and Smit A: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*** *Science* 2002, **297**:1301-1310.
4. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M and An P: **Initial sequencing and comparative analysis of the mouse genome** *Nature* 2002, **420**:520-62.
5. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V and Down T: **The Ensembl genome database project** *Nucl Acids Res* 2002, **30**:38-41.
6. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R and Miller W: **PipMaker – a web server for aligning two genomic DNA sequences** *Genome Res* 2000, **10**:577-86.
7. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M and Goodstein DM: **The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins** *Science* 2002, **298**:2157-67.
8. Osada N, Hida M, Kusuda J, Tanuma R, Hirata M, Hirai M, Terao K, Suzukisur Y, Sugano S and Hashimoto K: **Prediction of unidentified human genes on the basis of sequence similarity to novel cDNAs from cynomolgus monkey brain** *Genome Biol* 2002, **3**:RESEARCH0006.
9. Osada N, Hida M, Kusuda J, Tanuma R, Hirata M, Suto Y, Hirai M, Terao K, Sugano S and Hashimoto K: **Cynomolgus monkey testicular cDNAs for discovery of novel human genes in the human genome sequence** *BMC Genomics* 2002, **3**:36.
10. Goldman BD: **The Siberian hamster as a model for study of the mammalian photoperiodic mechanism** *Adv Exp Med Biol* 1999, **460**:155-64.
11. Hartmann JF and Hutchison CF: **Nature of the pre-penetration contact interactions between hamster gametes in vitro** *J Reprod Fertil* 1974, **36**:49-57.
12. Yanagimachi R: **Mammalian Fertilization** *In: The Physiology of Reproduction* Edited by: Knobil E, Neill JD. New York: Raven Press; 1994:189-317.
13. Bi M, Wassler MJ and Hardy DM: **Sperm adhesion to the extracellular matrix of the egg** *In: Fertilization* Edited by: Hardy DM. San Diego: Academic Press; 2002:153-180.
14. Hendry WJ 3rd, Branham WS and Sheehan DM: **The hamster cheek pouch as a convenient ectopic site for studies of uterine morphogenesis and endocrine responsiveness** *Differentiation* 1992, **51**:49-54.
15. Hendry WJ 3rd, Zheng X, Leavitt WW, Branham WS and Sheehan DM: **Endometrial hyperplasia and apoptosis following neonatal diethylstilbestrol exposure and subsequent estrogen stimulation in both host and transplanted hamster uteri** *Cancer Res* 1997, **57**:1903-8.
16. Hendry WJ 3rd, DeBrot BL, Zheng X, Branham WS and Sheehan DM: **Differential activity of diethylstilbestrol versus estradiol as neonatal endocrine disruptors in the female hamster (*Mesocricetus auratus*) reproductive tract** *Biol Reprod* 1999, **61**:91-100.
17. Khan SA, Ball RB and Hendry WJ 3rd: **Effects of neonatal administration of diethylstilbestrol in male hamsters: disruption of reproductive function in adults after apparently normal pubertal development** *Biol Reprod* 1998, **58**:137-42.
18. Eddy EM: **Regulation of gene expression during spermatogenesis** *Semin Cell Dev Biol* 1998, **9**:451-7.
19. Andrews J, Bouffard GG, Cheadle C, Lu J, Becker KG and Oliver B: **Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis** *Genome Res* 2000, **10**:2030-43.
20. Yue L, Karr TL, Nathan DF, Swift H, Srinivasan S and Lindquist S: **Genetic analysis of viable Hsp90 alleles reveals a critical role in *Drosophila* spermatogenesis** *Genetics* 1999, **151**:1065-79.
21. Fisher KJ and Aronson NN Jr: **Cloning and expression of the cDNA sequence encoding the lysosomal glycosidase di-N-acetylchitobiase** *J Biol Chem* 1992, **267**:19607-16.

22. Brohmann H, Pinnecke S and Hoyer-Fender S: **Identification and Characterization of New cDNAs Encoding Outer Dense Fiber Proteins of Rat Sperm** *J Biol Chem* 1997, **272**:10327-10332.
23. Leung E, Print CG, Parry DA, Closey DN, Lockhart PJ, Skinner SJ, Batchelor DC and Krissansen GW: **Cloning of novel kinectin splice variants with alternative C-termini: structure, distribution and evolution of mouse kinectin** *Immunol Cell Biol* 1996, **74**:421-33.
24. Das M, Harvey I, Chu LL, Sinha M and Pelletier J: **Full-length cDNAs: more than just reaching the ends** *Physiol Genomics* 2001, **6**:57-80.
25. Goldstein LSB: **Kinesin molecular motors: Transport pathways, receptors, and human disease** *PNAS* 2001, **98**:6999-7003.
26. Miki H, Setou M, Kaneshiro K and Hirokawa N: **All kinesin superfamily protein, KIF, genes in mouse and human** *PNAS* 2001, **98**:7004-7011.
27. Vale RD and Fletterick RJ: **The design plan of kinesin motors** *Annu Rev Cell Dev Biol* 1997, **13**:745-777.
28. Yanaihara N, Kohno T, Takakura S, Takei K, Otsuka A, Sunaga N, Takahashi M, Yamazaki M, Tashiro H and Fukuzumi Y: **Physical and transcriptional map of a 311-kb segment of chromosome 18q21, a candidate lung tumor suppressor locus** *Genomics* 2001, **72**:169-79.
29. Jenuwein T and Allis CD: **Translating the histone code** *Science* 2001, **293**:1074-80.
30. Gross CT and McGinnis W: **DEAF-1, a novel protein that binds an essential region in a Deformed response element** *Embo J* 1996, **15**:1961-70.
31. Hwang I and Gottlieb PD: **The Bop gene adjacent to the mouse CD8b gene encodes distinct zinc-finger proteins expressed in CTLs and in muscle** *J Immunol* 1997, **158**:1165-74.
32. Sims RJ 3rd, Weihe EK, Zhu L, O'Malley S, Harriss JV and Gottlieb PD: **m-Bop, a repressor protein essential for cardiogenesis, interacts with skNAC, a heart- and muscle-specific transcription factor** *J Biol Chem* 2002, **277**:26524-9.
33. Gottlieb PD, Pierce SA, Sims RJ, Yamagishi H, Weihe EK, Harriss JV, Maika SD, Kuziel WA, King HL and Olson EN: **Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis** *Nat Genet* 2002, **31**:25-32.
34. O'Brien SJ, Menotti-Raymond M, Murphy WJ, Nash WG, Wienberg J, Stanyon R, Copeland NG, Jenkins NA, Womack JE and Marshall Graves JA: **The promise of comparative genomics in mammals** *Science* 1999, **286**:458-62.
35. Lee YM, Lee S, Lee E, Shin H, Hahn H, Choi W and Kim W: **Human kinesin superfamily member 4 is dominantly localized in the nuclear matrix and is associated with chromosomes during mitosis** *Biochem J* 2001, **360**:549-56.
36. Wittmann T, Hyman A and Desai A: **The spindle: a dynamic assembly of microtubules and motors** *Nat Cell Biol* 2001, **3**:E28-34.
37. Macho B, Brancorsini S, Fimia GM, Setou M, Hirokawa N and Sassone-Corsi P: **CREM-dependent transcription in male germ cells controlled by a kinesin** *Science* 2002, **298**:2388-90.
38. Jenuwein T: **Re-SET-ting heterochromatin by histone methyltransferases** *Trends Cell Biol* 2001, **11**:266-73.
39. Lutterbach B, Westendorf JJ, Linggi B, Patten A, Moniwa M, Davie JR, Huynh KD, Bardwell VJ, Lavinsky RM and Rosenfeld MG: **ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors** *Mol Cell Biol* 1998, **18**:7176-84.
40. Ansieau S and Leutz A: **The conserved Mynd domain of BS69 binds cellular and oncoviral proteins through a common PXLXP motif** *J Biol Chem* 2002, **277**:4906-10.
41. Lutterbach B, Sun D, Schuetz J and Hiebert SW: **The MYND motif is required for repression of basal transcription from the multidrug resistance 1 promoter by the t(8;21) fusion protein** *Mol Cell Biol* 1998, **18**:3604-11.
42. Nalbant D, Williams SC, Stocco DM and Khan SA: **Luteinizing hormone-dependent gene regulation in Leydig cells may be mediated by CCAAT/enhancer-binding protein-beta** *Endocrinology* 1998, **139**:272-9.
43. Du Y, Campbell JL, Nalbant D, Youn H, Bass AC, Cobos E, Tsai S, Keller JR and Williams SC: **Mapping gene expression patterns during myeloid differentiation using the EML hematopoietic progenitor cell line** *Exp Hematol* 2002, **30**:649-58.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

