BOP1 – a key player of ribosomal biogenesis

Nabajyoti Borah[1], Krishna Madhav Nukala[2], Varahalarao Vadlapudi[1], Satya Prakash Gubbala[1] and Ramars Amanchy[1,*]

[1]Applied Biology Division, CSIR – Indian Institute of Chemical Technology, Uppal Road, Hyderabad 500 007, India
[2]Department of Biology, University of Iowa, 129 E, Jefferson Street, Iowa City, 52242, USA

Ribosomal biogenesis involves coordination of protein complexes for translation, cell growth and differentiation. An analysis of ribosomal biogenesis factor ERB1/BOP1 in evolution and cancer was carried out using comparative bioinformatics approaches focusing on protein domain identification, phylogenetic analysis, homology modelling, analyses of gene expression and interaction networks. We have identified WD40 domain as an essential co-occurring domain in all the BOP1 proteins predominantly in eukaryotes and also identified some key structural motifs in BOP1. A strong correlation of BOP1 has been observed in multiple signalling pathways, dysregulation of which leads to cancer. Using interaction networks and data mining, which are literature-derived, we have identified important interaction partners of PeBoW complex that co-express with BOP1 in signalling pathways. Analysis of BOP1 differential gene expression and interaction pathways reveals that BOP1 plays an important role in regulating p53 signalling and cell-cycle networks, and provides a crosstalk with key cellular processes. Examination of cancer and tissue expression profiling points to the upregulation of BOP1 in a variety of cancers. Thus BOP1 can be considered as a potential cancer biomarker and therapeutic target.

Keywords: Cancer biomarker, molecular phylogeny, protein complexes, ribosomal biogenesis, therapeutic target.

ERB1/BOP1 is a protein involved in ribosomal biogenesis, an energetically challenging event and a hallmark of evolution[1]. Eukaryotic 80S ribosomes are composed of 60S and 40S subunits. Proper assembly of these subunits requires multiple proteins[2]. Eukaryotic ribosomal biogenesis factor (ERB1), commonly referred to as block of proliferation protein (BOP1), has a special function in the processing of rRNA precursors that give rise to the maturation of 5.8S, 25/28S to the 60S ribosomal subunit. The *BOP1* gene is found in the 8q24.3 region of chromosome 8 in *Homo sapiens*[3]. Increased copy number of *BOP1* gene alters the chromosomal segregation and is observed in 8q24 amplification in colorectal cancers[4].

*For correspondence. (e-mail: ramars@iict.res.in)

ERB1/BOP1 protein consists of primarily two domains, namely BOP1 and WD40 (ref. 5). The evolutionary loss or gain of function of these domains along with other associated functional domains makes BOP1 an interesting candidate to study evolution of the ribosomal assembly and rRNA processing. We performed domain analysis of BOP1 protein from over 1000 organisms, to see how the *BOP1* gene has evolved to gain biological functions and how gain of function works in favour of these organisms. We have examined adaptive evolutionary patterns and found that BOP1 domains occur only once per organism in ERB1 protein, and *BOP1* pseudogenes or paralogs were not observed. BOP1, was first isolated from the cDNA library screening for growth-related sequences in mouse embryonic fibroblasts[6] and was shown to interact with Pes1 (ref. 7) and p53 complex[8]. BOP1 forms PeBoW (PES1/BOP1/WDR12) protein complex in the maturation of rRNAs (28S, 5.8S) and biogenesis of the 60S ribosomal subunit. Phylogenetic analysis confirms this aspect to be consistent in multiple eukaryotes, thus signifying evolutionarily conserved features of ribosomal biogenesis. BOP1 domain, also referred to as BOP1NT (domain in the N-terminal region), is relatively similar across different organisms. However there is variation in the number of domain combinations and its association with WD40. The number of different domain architectures is variable and is related to the complexity and environmental adaptation of the organism. We applied molecular phylogenetic approach to gain insight into the evolutionary relationship of the BOP1 protein among different phyla. A 'supertree'[9] was constructed using both the BOP1 domain and the complete BOP1 sequence to address molecular diversity, and build evolutionary relationships and patterns[10] using hierarchical clustering[11]. The three-dimensional structure of BOP1 has not yet been identified and deposited in any structural genomics databases. Hence, using homology modelling of BOP1 followed by quality evaluation, we observed conservation of predicted structures in multiple species. Gene expression profiling of BOP1 overexpression data has uncovered many signalling pathways that are directly or indirectly regulated by BOP1. Functional module connectivity maps made using existing gene–gene interaction networks from various databases have allowed the generation of novel hypotheses regarding potential functional roles of BOP1 and its involvement in cancers.

Methods

Data mining and domain analysis

Non-redundant protein sequences from 1061 organisms across all life forms containing BOP1 domain were retrieved from the NCBI database[12], facilitated by Conserved Domain Architecture Retrieval Tool (CDART). BOP1 domain positions were then identified in the protein sequences and extracted using a Python script. The domain analysis was largely based on the resources available in CDD (accession: cl19763)[13], SMART (accession: SM01035)[14] and PFAM (accession: PF08145)[15].

Sequence alignment and phylogeny

Based on taxonomy, domain sequences were divided and subjected to multiple sequence alignment to visualize the conserved regions in the domain. A phylogenetic tree was constructed with the help of MEGA7 (ref. 16). Multiple sequence alignments were carried out using MAFFT (ref. 17) and the tree was constructed based on maximum likelihood method[18].

Co-expression profiling and heat map

Network analysis for protein–protein interaction and co-expression profiling was done using STRING. A coherent set-up in STRING iteratively merges the orthologous groups at the various clades and levels, until a fully consistent state is achieved. In co-expression analysis, STRING generates scores based on the interaction between different proteins with the input protein and predicts the functional partner. All the data are collected through text-mining, experiments, databases and co-expression results. We constructed a heat map using predicted gene expression values obtained from different model organisms. After giving BOP1 as input, *Saccharomyces cerevisiae* was chosen as the reference organism. In the data settings we selected active interaction sources and maintained high association score (confidence >0.700) to limit the false positive rate. In the interaction network we have divided this main network into sub groups based on their molecular function.

Differential gene expression and interaction network analysis

A list of differentially expressed (DE) genes was obtained using GEO2R analysis tool of NCBI, followed by meta analysis using on-line tool iPathwayGuide (AdvaitaBio) for normalized microarray data analysis[19]. The method begins with the searching of GEO datasets related to BOP1. The expression profiling array datasets having accession (GSE50841)[20], selected for the present analysis[20] and two groups of colon cancer cell lines were compared, of which one group had cells constitutively overexpressing BOP1. In order to identify the DE genes, the data generated by GE02R were retrieved and uploaded in iPathwayGuide for meta analysis, and 340 genes that were identified to be differentially expressed were selected for network analysis after stringent normalization. The DE genes based on their statistical significance were identified by analysing GEO datasets with the help of Advaita tool in iPathwayGuide[21,22]. The DE genes were searched for their involvement in different biological pathways relying on standalone repositories such as Reactome[23–25] and WikiPathways[26]. A volcano plot was generated in R studio (version 0.99.903) using a graphing package 'ggplot2' (cran.r-project.org/package=ggplot2). In the plot, fold change between the two groups (on a log scale) was plotted on the X-axis, while *t*-test significance *P*-values (on negative log scale) were plotted on the *Y*-axis. Colour threshold for –log10 *P*-value was set at 4.221. A heat map was generated in R using gplots package representing the degree of up- and down-regulation of genes in association with their different biological pathways. We divided interactions into functional sets and defined interactions as conserved sequences between gene sets. We also studied the degree of distribution of all genes in GGIN, to see whether or not they are scale-free. We have performed manual functional annotation of every gene (using http://www.genecards.org/) for their involvement in the biological pathways.

BOP1 along with the identified up- and down-regulated genes were searched in ConsensusPathDB[27] to find their interaction networks. In order to conjure up the interaction network, all the possible interactions were selected irrespective of their confidence level and mapping criteria. There were five interaction sources from which the network was generated: Biogrid[28], IntAct[29], Spike[30], Mint[31] and HPRD[32]. For proper visualization, the whole network was exported as a computer-readable file (BioPAX level 3), and it was used as an input to Cytoscape (version 3.4.0)[33].

Homology modelling, structure prediction and evaluation

Homology models of BOP1 proteins were generated using SWISS-MODEL[34]. The quality of the obtained model has been evaluated for internal consistency and reliability through *Z*-score using ERRAT[35] and RAMPAGE[36] Ramachandran plot analysis. Sequence pattern recognition was performed using WebLogo[37].

Normal and cancer tissue profiling

Absolute expression values, generated in Affymetrix Microarray U133 Plus2 platform, for BOP1, ACTB and

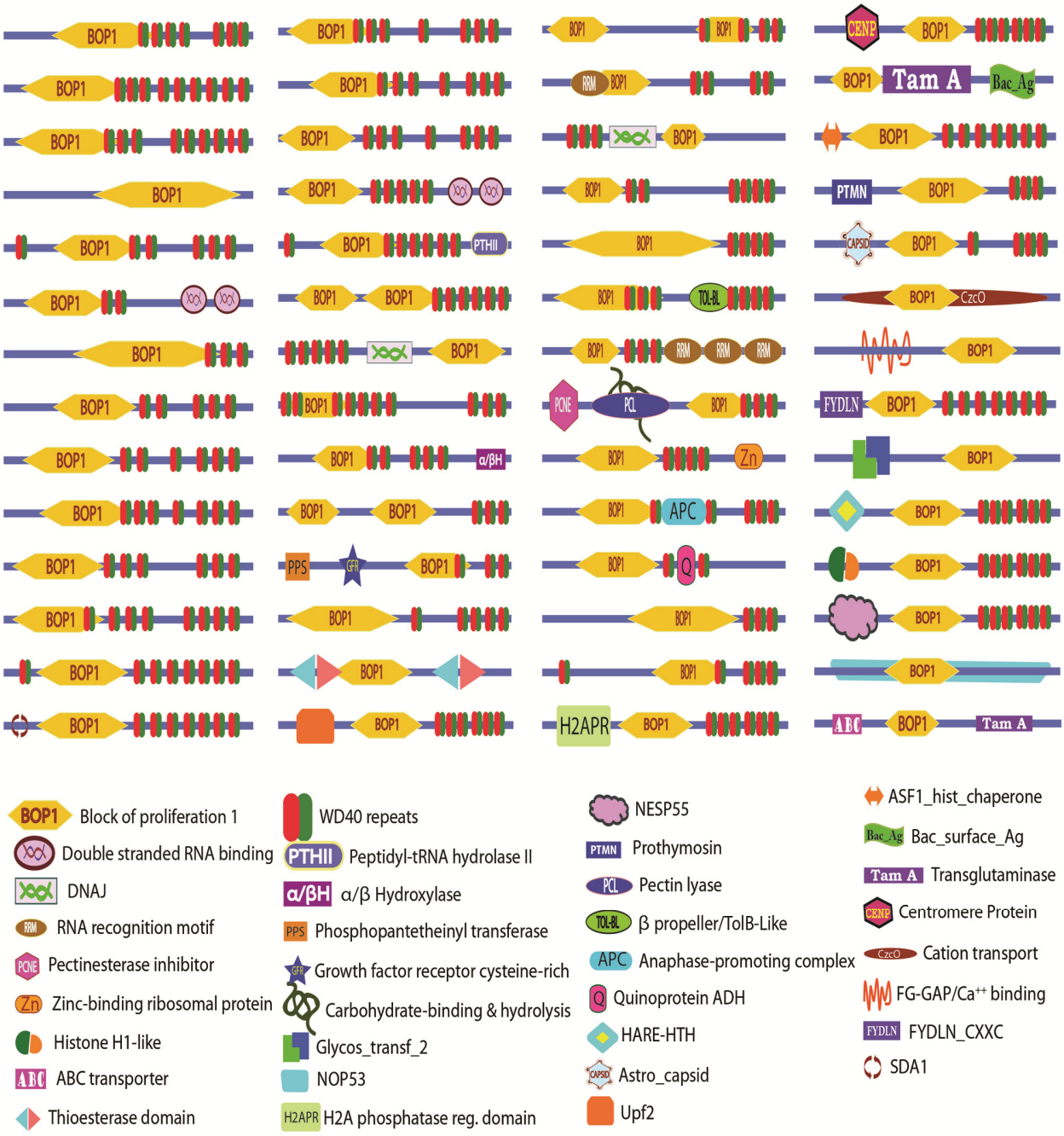

Figure 1. BOP1 domain architecture and association with WD40 and other functional protein domains as analysed by CDART, SMART and Pfam.

KRAS genes of normal and cancerous tissues were obtained from GENT (Gene Expression database of Normal and Tumour tissues) database[38]. For swift visualization and analysis of cancer outliers, log2 transformation was carried out and COPA package was used in R environment.

Post-translational modifications

Search in HPRD[32] and PhophoSitePlus[39] have allowed us to map post-translational modifications in BOP1.

Results and discussion

Association of BOP1 domain with other functional domains

BOP1 domain is found to be associated with WD40 repeats in a majority of species. WD40 domains are 40 amino acid peptide repeats terminating in tryptophan (W) and aspartate (D) residues, due to the presence of which the protein takes beta propeller fold-like conformation[40].

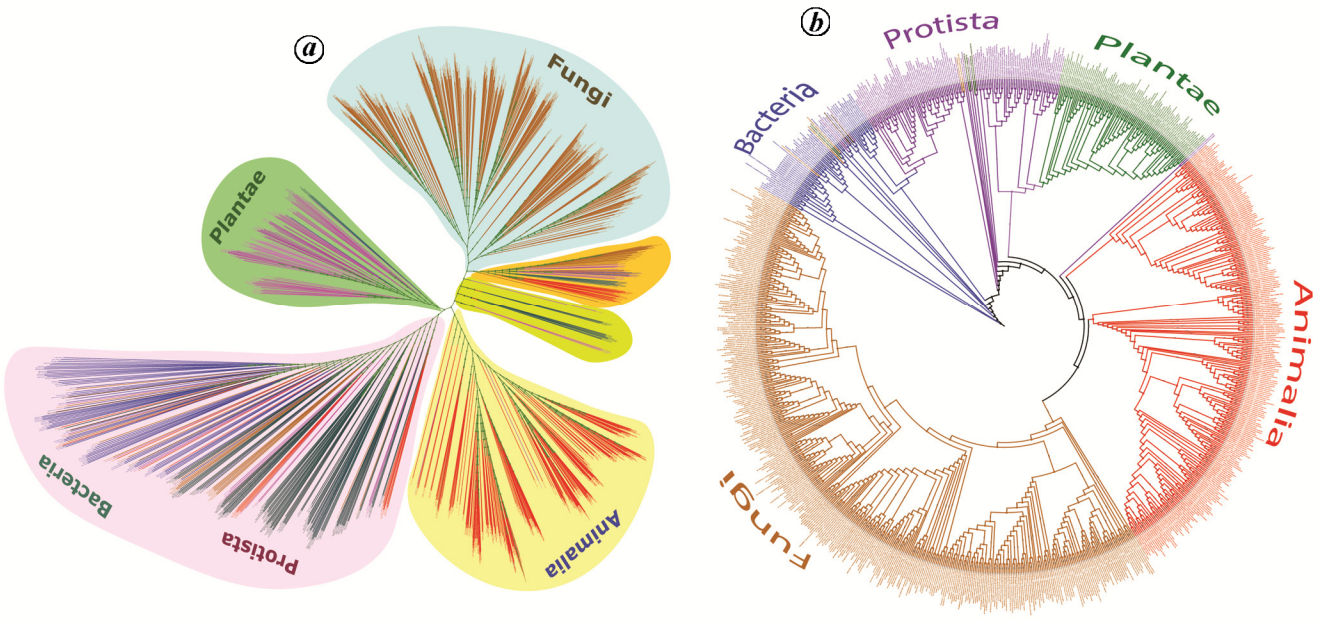

Figure 2. Phylogenetic tree of (***a***) BOP1 Protein and (***b***) BOP1 domain obtained by ML method with 1000 bootstrap replicates. The tree was displayed and annotated using web-based phylogenetic tool, iTOL. Each kingdom in the tree is depicted using discrete colours for proper visualization.

BOP1 domain at the N-terminus followed by WD40 repeats forming seven-bladed β-propeller architecture, could stabilize BOP1 in its interaction with biogenesis assembly proteins. WD40 domain-containing proteins, although abundant in eukaryotes not restricted to BOP1 protein, are rarely present in prokaryotes[41]. In few organisms ERB1 also has other domains which are gained in different timeline of evolution. These domains have added new functions to ERB1. We observed WD40 repeats varying in number and/or position (Figure 1). This may be important in offering varying conformations in differential interactions of BOP1. Some of the additional functional domains include dsRNA binding domain, RNA recognition motif, zinc-binding ribosomal protein, peptidyl tRNA hydrolase II, phosphopantetheinyl transferase, pectin lyase, carbohydrate-binding and hydrolysis domains that are usually absent in other ribosomal assembly proteins. Such domains are present in a few organisms, pointing out to their adaptation and survival in different environments.

BOP1 homolog of *Oxytricha trifallax* (a protozoan with 16,000 chromosomes) has an anaphase-promoting complex domain having WD40 repeats. In *Nannochloropsis gaditana*, there is a chaperone domain DNAj/Hsp40 domain at the C-terminal end. *Trichinella spiralis* contains double-stranded RNA-binding domain. BOP1 of *Poeciliopsis prolifica* shows a domain architecture consisting of Toll B-like receptor. *Brassica rapa* has a complex of pectinesterase inhibitor (PMEI) and sugar hydrolysis (CASH) domains. BOP1 homolog of *Pyronema omphalodes* has a peptidyl-tRNA hydrolase (Pth2) domain. The occurrence of these functional domains along with the BOP1 domain presents evidence of a single protein carrying out concerted actions through evolutionary gain of advantage in divergent species.

Molecular phylogeny

Protein domains are discrete evolutionary units and are considered to have emerged only once during evolution[42]. Molecular phylogeny studies for taxonomical purposes are based on ribosomal RNA genes[43], but a few focusing on ribosomal proteins have revealed functional evolutionary aspects[44]. There are around 32 ribosomal proteins that are well conserved across eukaryotes, bacteria and archaea, representing a universal pool of ribosomal architecture and a common ancestor[45]. Evolutionary theory envisages a simple machinery for primitive protein synthesis[45]. The phylogenetic analysis for protein ERB1 in whole and BOP1 domains yielded similar trees of life for classifying most species. It is evident from the tree that ERB1 has emerged during the course of evolution and can be observed in 41 unique phyla (Figure 2 *a* and *b*). Interestingly, in the phylogenetic analysis, we observed BOP1-like domains in viruses, bacteria and protista. Tree analysis was conducted to study the significance of BOP1 domain as a phylogenetic marker to gain better understanding of its origin and evolution. This analysis draws evolutionary and taxonomical inference through the occurrence of BOP1 domain as well as other functional domains. The results generated by this BOP1-centric tree showcase the conservation of this gene across a majority

of species, while a few display divergence from their clade in the tree of life. Considering the conserved nature of the super tree, we consider BOP1 as a potential marker for constructing phylogenetic trees, especially of eukaryotes. This tree reveals the diversified evolutionary patterns and interestingly both BOP1 domain and ERB1 protein sequences give predictable, yet comparative and similar trees.

Co-expression profile of BOP1 in eukaryotic model organisms

Knowledge on biological systems can be gained by understanding their biological networks, mainly protein–protein interactions (PPIs)[46] and gene co-expression networks[47]. To elucidate ERB1 gene function on a global scale, we used a co-expression matrix consisting of a set of 51 genes in *Saccharomyces cerevisiae* and compared it with co-expression in other model organisms (*Arabiodopsis thaliana*, *Bos taurus*, *Caenor habditis elegans*, *Danio rerio*, *Drosophila melanogaster*, *Mus musculus*, *Oryza sativa japonica*, *Plasmodium falciparum* and *Schizosaccharomyces pombe*). One important aspect of the analysis was that co-expression of genes was first analysed in *S. cerevisiae*, and the promising genes were compared for their co-expression in nine other model organisms. We observed that *S. cerevisiae* expression of *ERB1* gene has positive correlations with 48 out of 50 genes and negative correlation with *TIF6* and *SPB4* (Figure 3 *a*). Whereas in other model organism, *BOP1* gene expression is negatively correlated with RIX7, RLP7, NSA2, and RIX1 along with TIF6 (Figure 3 *b*). TIF6 shows similar negative correlation of co-expression with ERB1 in both *S. cerevisiae* and model organisms. Mammalian homolog TIF6 in yeast has lost transcription initiation function, but shows involvement in ribosomal biogenesis. Similarly, SPB4 with a helicase function has also been reported to be involved in ribosomal biogenesis[48]. RRP12 shows good co-expression correlation in five different organisms (*A. thaliana*, *C. elegans*, *D. melanogaster*, *O. sativa japonica* and *S. pombe*), NOC2, NOC3, PUF6, YTM1 in three organisms (*A. thaliana*, *D. melanogaster* and *O. sativa japonica*), RPA135 in three organisms (*A. thaliana*, *D. melanogaster* and *S. pombe*), NSA2 in two organisms (*A. thaliana* and *O. sativa japonica*) and NOG1 in two organisms (*A. thaliana* and *P. falciparum*; Figure 3 *c*). In *S. cerevisiae*, CBF5P is the essential nucleolar protein in box H/ACA small nucleolar RNPs (snoRNPs) and codes for pseudouridine (Psi) synthase enzyme[49]. *RPA135* is a gene with good coorrelation and similar co-expression values as that of BOP1. RPA135 has a C-terminal Zn-binding domain and is the second-largest subunit of yeast RNA polymerase I (ref. 50). PUF6 has important functional roles in asymmetric mRNA distribution and ribosome biogenesis localized to the nucleolus[51]. RRP12 is an essential protein for the export of pre-40S and pre-60S particles out of the nucleus[52]. NOC2, NOC3 genes are involved in intranuclear transport of the pre-60S subunit[53]. NOG1 is a nucleolar GTP-binding protein functionally linked to ribosome biogenesis[54]. *Nsa2* gene is required for the maturation of 27 SB pre-rRNAs[55], while RLP7 is required for large ribosomal subunit biogenesis[56]. We found co-expression relationships, each of which has been conserved across certain species, which might confer a selective functional advantage in evolution. From this we conclude that these genes could be integral partners of ERB1 functional complex in ribosomal biogenesis. We confer that the co-expressed genes show positive correlation with BOP1 and have proven involvement in ribosomal biogenesis and show nucleolar localization.

BOP1 interactome

After identification of nucleolar proteins with good co-expression correlation to BOP1, we analysed BOP1 interaction networks in protein interaction databases. Four major classes of ERB1-interacting proteins have been identified based on biological functions (ribosomal biogenesis, regulation of gene expression, DNA synthesis and metabolic regulation; Figure 4). The interactome displays 40 closely knit proteins showing involvement in ribosomal biogenesis, 24 in gene regulation, 17 in metabolic regulation and 19 in DNA synthesis; all of them are connected to each other and to BOP1. Most of the interacting proteins belong to the ribosome biogenesis pathway as expected. From this observation it is evident that BOP1 is a critical protein in the synthesis and maturation of ribosomal subunits. The complex of PES1, WDR12 with BOP1 forms the important PeBOW complex whose significance lies in the maturation of the 60S subunit, and an aberration in any one of the interacting partners could lead to serious health implications[57]. Analysis revealed more number of interactomes of BOP1 than known before, participating in diverse cellular functions. We infer that there are several pathways in which BOP1 is involved, probably acting as a scaffold or as a downstream regulator of ribosomal biogenesis and gene regulation.

Most of the interactomes of BOP1 (58%), irrespective of their metabolic functions, are located in the nucleus, nucleoplasm and nucleolus. The role of BOP1 in ubiquitination of proteins could be justified by the presence of a major part of its interactomes in protein ubiquitination (10%), proteasome degradation (10%), protein transport (7%) and polyubiquitination (5%; Figure 4). This information also shows how BOP1 could regulate the protein transport and ubiquitination-dependent degradation. The WDR family of proteins was considered as the possible target for anti-fungal drugs[57–60]. Since we show that BOP1 has interactions with proteins of the WDR family

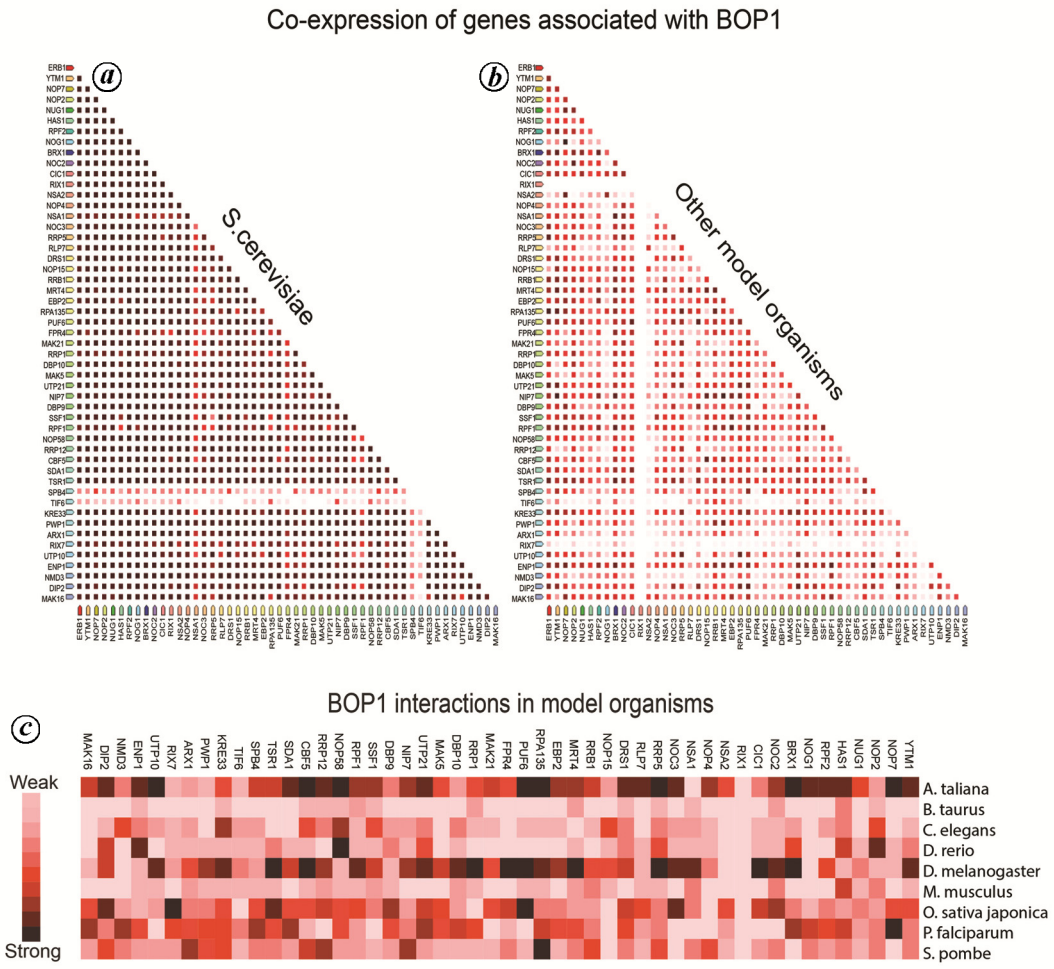

Figure 3. Co-expression matrix of a set of 51 genes in (***a***) *Saccharomyces cerevisiae* and (***b***) combined single-vector expression scores of other model organisms delineated as a heat map. (***c***) Expression analysis showing nine model organisms to be associated with most of the co-expressed genes. ERB1 centric in-depth analysis was further performed with these model organisms.

and also co-expressed with some WDR genes, BOP1 itself could be a novel anti-fungal therapeutic target. BOP1 also interacts with NOP family of proteins (NOP2, NOP34, NOP58 and NOP56) that are considered as a target for analgesic drugs[61], and studies focusing on NOP and BOP1 might lead to better understanding of infection and inflammation.

BOP1 gene expression in normal and cancer tissues

We carried out comparative gene expression analysis of BOP1 in normal and cancer tissues along with *K-Ras* oncogene and ACTB (actin-B)[62] as housekeeping gene. *KRAS* (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog) is the most common mutated gene of the RAS family in cancer[63]. The results obtained by comparative gene expression analyses show that BOP1 expression levels are upregulated in numerous cancer tissues (Figure 5 *a*). The plot shows clear comparison of BOP1 expression levels in both cancer and normal tissues, thus aiding in making a comprehensive conclusion about the role of BOP1 in cancer pathways. When comparing normal and cancer tissues, there is a pronounced increase of BOP1 expression in the latter. This remarkable observation was also more pronounced when BOP1 expression was compared to that of ACTB and KRAS. Block of proliferation protein (BOP1) was originally identified to be upregulated in hepatocellular cancinoma as part of 8q24 amplification, and the upregulation was observed to be much more than the candidate marker C-MYC[63–65]. This means that BOP1 might have the potential to act as a marker gene for most cancers and also a potential pharmacological target.

Especially higher expression levels of BOP1 were observed in blood, brain, colon, lung and skin cancer tissues when compared to other cancers. Evidences of BOP1 overexpression in cancer tissues were prominent and conclusive, although the mechanism of carcinogenesis is not understood. Overexpression of *BOP1* gene

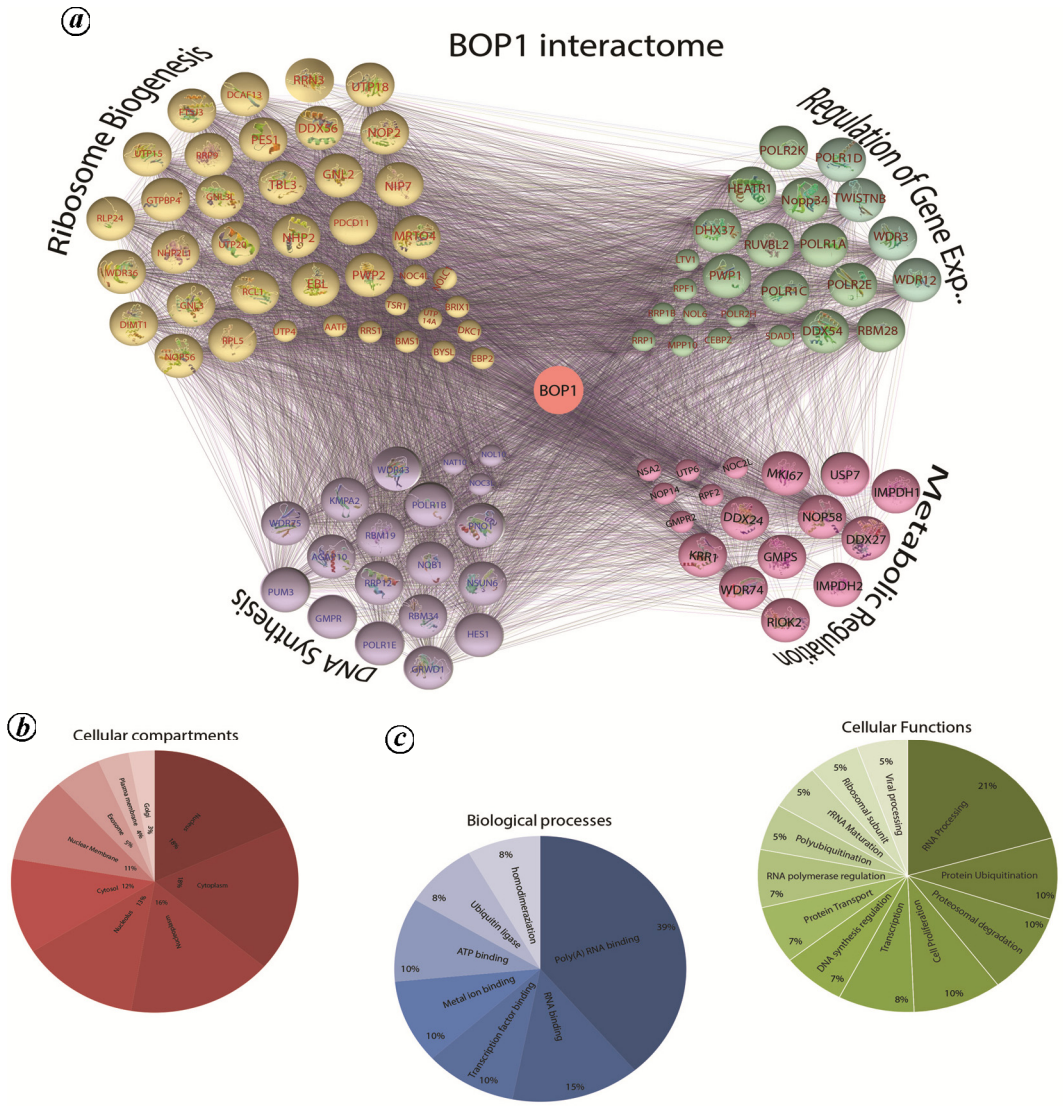

Figure 4. Interaction of BOP1 with different genes obtained after interactome analysis. Four clusters of genes associated with ribosome biogenesis, gene expression regulation, and metabolic regulation and DNA synthesis are found to interact with BOP1. Quantitative analysis of these genes based on their (***a***) cellular compartment of expression; (***b***) biological processes they are involved in and (***c***) cellular functions carried out by them, affirms that most of the genes interacting with BOP1 are involved in RNA processing, and they are highly expressed in the nuclear region of the cell.

accompanied by chromosomal aberrations like 8q24 amplification could be a lethal causal factor for colorectal cancer. Another significant aspect of BOP1 is that it is one of the candidates for methylation-dependent expression variations. It was shown to overexpress and had lower methylation rates when compared to other genes in ovarian cancer[66]. The present study also provides the same result, with BOP1 expression levels in Ovary-C way higher than Ovary-N. We have also examined if any SNPs and mutations in the coding regions are available in dbSNP[67]. We found a total of 1282 SNPs associated with the *BOP1* gene, of which 1158 were from the intronic region and 124 were from the exonic region. SNPs found in the exon regions were further analysed, and it was found that missense and synonymous type of SNPs were more prevalent in the coding regions. In order to find any interspecific conserved polymorphism in these regions, SNPs of *H. sapiens* were compared with those of *Rattus norvegicus*, *Mus musculus*, *Macaca fascicularis* and *Danio rerio*. However, no significant conservation match was found.

We have also searched for post-translational modifications of BOP1 available from the literature and proteomic databases. There was strong evidence for ubiquitination, again positively correlating with the interactome analysis. We also found that BOP1 protein is ubiquitinated on seven residues, thus adding to its already speculated role in the ubiquitination pathways. We also found BOP1 to be hyperphosphorylated on both serine/threonine as well as tyrosine. We found experimental evidence showing 18

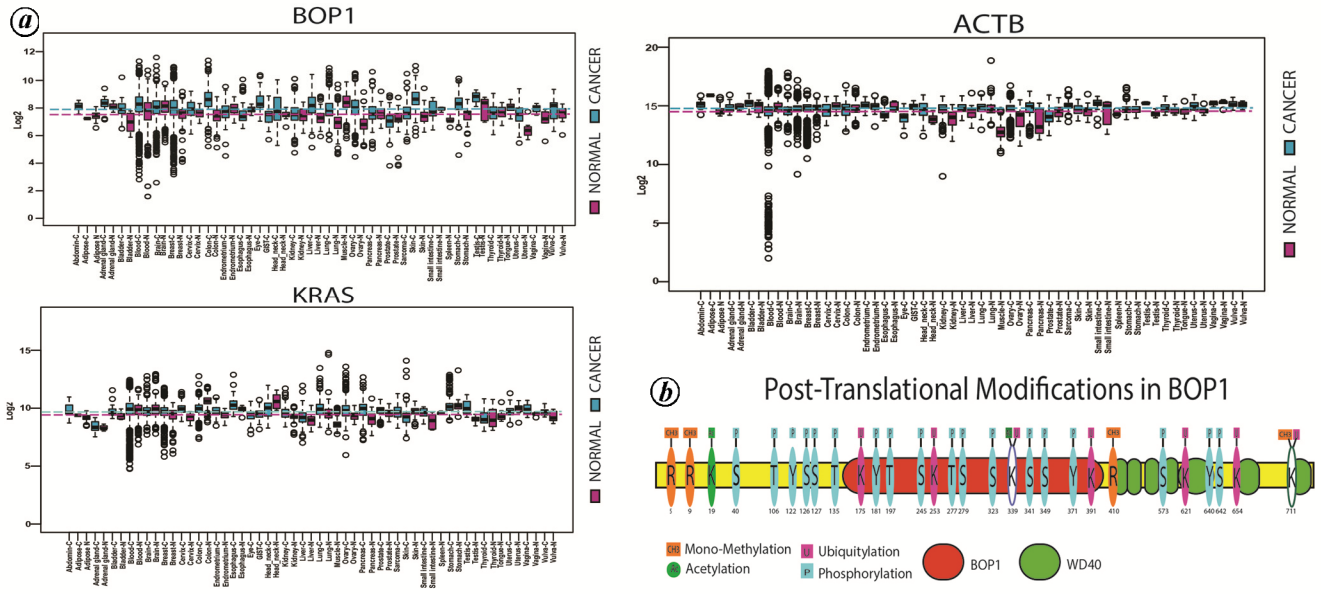

Figure 5. ***a***, A comparative gene expression analysis of BOP1 gene involved in normal and cancerous tissue. BOP1 is found to be overexpressed in most of the cancers when compared to KRAS oncogene: ACTB was taken as the housekeeping gene in this analysis. ***b***, Schematic representation of translational modifications of BOP1.

residues (10 serine, 4 threonine and 4 tyrosine) to be phosphorylated (Figure 5 *b*). Nine of the phosphorylated residues were observed in the BOP1 domain, indicating that there is a change in conformation and possibility of novel interactions. We analysed possible upstream kinases based on the phosphomotifs[68] and found that some of the phosphorylation sites had motifs specific to CDKs and casein kinase.

Homology modelling and structure prediction

Since ERB1 protein and BOP1 domain structures have not been solved, we chose to predict the structures using homology modeling and validate these predictions using computational approaches. Analysis of structural similarity of ERB1 protein in different organisms is an index of evolutionarily shared functional activity. The ERB1 protein, with its unique globular architecture consisting of seven-bladed β-propellers formed by seven WD40 domains[69], is well conserved across model organisms tested and across evolution, barring a few exceptions. This loop has no definitive secondary and tertiary structural features in the 60 amino acid stretch (Ser 554 to Ala 614; Figure 6 *a*). PEST motifs that are composed primarily of proline (P), glutamic acid (E), serine (S) and threonine (T) are clusters of charged amino acids that form flexible loops, and are linked to protein destabilization and degradation in eukaryotic systems[40]. BOP1 proteins are more accessible for proteases due to the presence of PEST region[70] for rapid cleavage, and this could be a mark of the many checkpoints in the control of ribosome biogenesis. We found that in the fungal BOP1 domains, two α-helices were formed by LDQLLDS and EELELIR sequence motifs that are highly conserved. We observed a small motif GTSLK (genus *Candida*), GQDVR and GRDLR motifs (Kingdom: animalia), KPP residues (class insecta) RKIY and DVEL (Kingdom: plantae) were involved in stabilizing two β-sheet formations (Figure 6 *b*). We found α-helix favouring sequences as LDSFLAS, KNWRKIY (plantae), and DELDQFL (animalia). We also observed proline to be well conserved as a residue at position 70 in the BOP1 domain. There is a unique C-terminal free hanging loop which is prominent in all the organisms and even remarkable in *C. albicans*. A motif representing H(I/X)GY(D/X)(X/I)(X/N)G was prominently conserved throughout the evolution, represented in the tree of life (Figure 6 *c*).

Proteins of the β-propeller fold are ubiquitous in nature and widely used as structural scaffolds for ligand-binding and enzymatic activity. All β-propeller proteins with different blade numbers adopt disc-like shape; they are involved in a diverse set of functions, and defects in this family of proteins have been associated with human diseases[71]. We observed several PEST motifs in BOP1 proteins both in the BOP1 domain and N-terminal loop just prior to the BOP1 domain, in the central region that overlaps with the predicted NOP7 binding sequence[72].

Differential gene expression and networks analysis in BOP1 overexpression

We analysed how BOP1 overexpression affects cellular signalling pathways. Gene–gene integrative network (GGIN) analysis has been used to build different

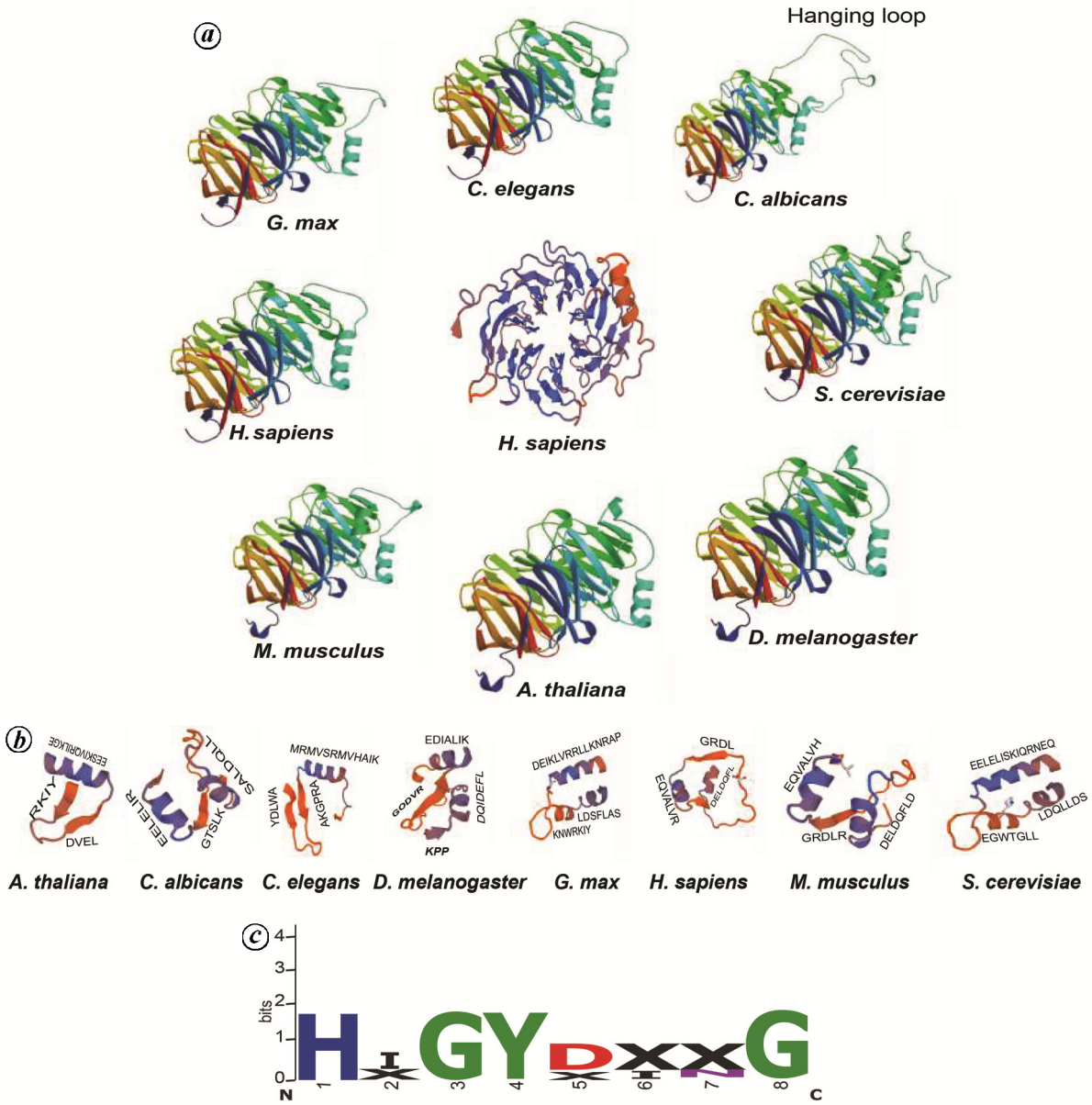

Figure 6. ***a***, Structure of WD40 domains in BOP1 protein often folds into seven-bladed β-propellers with a funnel-like shape in different model organisms including *S. cerevisiae*, *D. melanogaster*, *A. thaliana*, *M. musculus*, *H. sapiens*, *G. max* and *C. elegans*. In *C. albicans* WD40 domain takes the hanging-loop structure, while top view of seven-bladed β-propeller in *H. sapiens* can be visualized in the centre. ***b***, BOP1 domain structure in *A. thaliana*, *C. albicans*, *C. elegans*, *D. melanogaster*, *G. max*, *H sapiens*, *M. musculus* and *S. cerevisiae*. α-Helix is shown in purple colour and β-sheet in orange. Identified domain motifs are shown in single-letter code. ***c***, Sequence logo for BOP1 domain repeats. The letter plots represent amino acid conservation at each position. The logo was designed based on structural alignment of conserved motif in the BOP1 domain.

networks of the upregulated and down-regulated genes. Also, 340 DE genes were selected from the NCBI gene expression dataset (GSE50841) for network analysis[70]. GGIN analysis revealed molecular networks of various functional roles for the *BOP1* gene.

GGIN analysis also shows that when the *BOP1* gene is upregulated, its counterparts *WDR12* and *PES1* also get upregulated. Other upregulated genes that are part of this network are *KLF10*, *ARRB1*, *ARRB2*, *LTE1* and *NUDC*, which are part of cell division. *BOP1* also upregulates several genes involved in cellular growth and proliferation genes like *CHEK2*, *RAF1*, *POLR1D*, *MAPKAPK2*, *SOX5* and *CUL3*. Ribosomal genes like *NME1*, *RPL36A* and *RPS26* are found to be upregulated along with genes with significant roles in rRNA maturation, like *RpS7*, *RpS24*, *RPL29*, *RPS15* and *RPL36A*. Other significant networks formed through *BOP1* connect to networks with involvement in DNA replication (*MCM10*), transcription regulation (*TAF1A*), chromatin remodelling and gene expression (*JMJD6*). A strong link to ubiquitination through

BOP1 interactors *ITCH* and *UBQLN1*, *TANK*, *RELA*, *SMAD2* was observed in the genes.

The present study also shows that *BOP1* overexpression causes downregulation of the expression of cyclin-dependent kinases (CDKs), thereby impairing the cell cycle process. It has been already demonstrated that inactivation of BOP1 causes pre-RNA processing and thus causes multiple blocks in cell cycle progression[21]. Although it is clear from *BOP1* interactome and co-expression data that *BOP1* has a strong role to play in the cell cycle, the *BOP1* overexpression data represented by networks reveal several key players of the cell cycle that include *CDC5L*, *CDC7*, *CDC14A*, *CDC23*, *CDC25*, *CDC25B*, *CDC34*, *CDC37*, *CDC42*, *CDC73*, *CDKN1A*, *CDKN2A*, *CDKN2D*, *CDK2*, *CDK4*, *CDK4*, *CDK6* and *CDK9*. We observed the interaction of BOP1 with protein TP53 and ITCH, an ubiquitnation regulating protein ligase. A p53-dependent signalling axis was earlier established that connects ribosomal biogenesis and cell-cycle progression[72], in which BOP1 might play a regulatory role. Critical to all cancer cells is the regulation of p53. It is known that p53 is involved in the induction of apoptosis and cell-cycle arrest through an intricate pathway of signalling molecules[8]. In this pathway, BOP1 was involved in G1/S transition of the cell cycle[73]. It was interesting to note that the genes involved in regulation of TP53 activity (*L3MBTL1*, *DUSP4*, *PERP*, *BCL6*, *RRAGB*, *TP53I3* and *TP53INP1*) were down-regulated, while *TP53BP2* was the only gene that was up-regulated. Using a predictive approach to identify the pathways in which BOP1 may play an influential role from the networks built, a heat map containing fold change values of all the up- and down-regulated genes and their respective pathways was critically analysed. Most of the genes are involved in protein transport, trafficking and regulation of genetic processes, which is obvious from the fact that BOP1 has a role predominately in ribosomal biogenesis.

Conclusion

We have performed a comprehensive systems analysis by identifying several orthologs of BOP1 across various genomes and establishing a strong evolutionary significance to the conservation of BOP1 domain and BOP1 protein. We have built a supertree based on both domain and complete BOP1 protein sequences to establish a phylogenetic evolutionary relationship among the organisms in relation to ribosomal function. In this study we have defined the presence of additional domains with varying functions along with the BOP1 domain, and have shown strong association of the BOP1 domain to WD40 domains. Homology modelling has provided insights into conservation of conformation of BOP1 in various organisms and occurrence of BOP1 as well as PEST motifs in BOP1 sequences. The analysis has revealed crucial players which co-express, interact and are upregulated by BOP1 that might be important in the role of BOP1 in ribosomal biogenesis and its other signalling roles. We have also shown BOP1 as a potential biomarker, highlighting its expression levels in cancer tissues. A critical correlation established between p53 and BOP1 might show that BOP1 is now one of the important molecules in regulating apoptosis and cell cycle. This study is unique as it involves application of several datasets in public repositories to implicate a specific gene hidden in databases as a potential evolutionary and cancer biomarker. The emerging role of ribosome biogenesis in cancer research and the dearth for potential marker genes pitches BOP1 as a primary candidate for further studies in all types of cancer signalling pathways.

Conflict of interest: The authors declare no conflict of interest and no competing financial interests.

1. Thomas, G., An encore for ribosome biogenesis in the control of cell proliferation. *Nature Cell Biol*., 2000, **2**, E71–E72.
2. Thomson, E., Ferreira-Cerca, S. and Hurt, E., Eukaryotic ribosome biogenesis at a glance. *J. Cell Sci*., 2013, **126**, 4815–4821.
3. Huret, J. L. *et al.*, Atlas of genetics and cytogenetics in oncology and haematology in 2013. *Nucleic Acids Res*., 2013, **41**, D920–D924.
4. Killian, A. *et al.*, Contribution of the *bop1* gene, located on 8q24, to colorectal tumorigenesis. *Genes, Chromosomes Cancer*, 2006, **45**, 874–881.
5. Strezoska, Z., Pestov, D. G. and Lau, L. F., BOP1 is a mouse WD40 repeat nucleolar protein involved in 28s and 5. 8s rRNA processing and 60S ribosome biogenesis. *Mol. Cell. Biol*., 2000, **20**, 5516–5528.
6. Pestov, D. G., Grzeszkiewicz, T. M. and Lau, L. F., Isolation of growth suppressors from a cDNA expression library. *Oncogene*, 1998, **17**, 3187–3197.
7. Lapik, Y. R., Fernandes, C. J., Lau, L. F. and Pestov, D. G., Physical and functional interaction between PES1 and BOP1 in mammalian ribosome biogenesis. *Mol. Cell*, 2004, **15**, 17–29.
8. Pestov, D. G., Strezoska, Z. and Lau, L. F., Evidence of p53-dependent cross-talk between ribosome biogenesis and the cell cycle: effects of nucleolar protein BOP1 on g(1)/s transition. *Mol. Cell. Biol*., 2001, **21**, 4246–4255.
9. Akanni, W. A., Wilkinson, M., Creevey, C. J., Foster, P. G. and Pisani, D., Implementing and testing Bayesian and maximum-likelihood supertree methods in phylogenetics. *R. Soc. Open Sci*., 2015, **2**, 140436.
10. Bininda-Emonds, O. R., Supertree construction in the genomic age. *Methods Enzymol*., 2005, **395**, 745–757.
11. Uchiyama, I., Hierarchical clustering algorithm for comprehensive orthologous-domain classification in multiple genomes. *Nucl. Acids Res.*, 2006, **34**, 647–658.
12. Geer, L. Y., Domrachev, M., Lipman, D. J. and Bryant, S. H., CDART: protein homology by domain architecture. *Genome Res*., 2002, **12**, 1619–1623.
13. Marchler-Bauer, A. *et al.*, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*., 2017, **45**, D200–D203.
14. Letunic, I., Doerks, T. and Bork, P., SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res*., 2012, **40**, D302–D305.

15. Finn, R. D. *et al.*, The PFAM protein families database: towards a more sustainable future. *Nucleic Acids Res*., 2016, **44**, D279–D285.
16. Kumar, S., Stecher, G. and Tamura, K., MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol*., 2016, **33**, 1870–1874.
17. Katoh, K. and Standley, D. M., MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol*., 2013, **30**, 772–780.
18. Felsenstein, J., Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol*., 1981, **17**, 368–376.
19. Szklarczyk, D. *et al.*, String v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*., 2015, **43**, D447–D452.
20. Williams, K. E., Lemieux, G. A., Hassis, M. E., Olshen, A. B., Fisher, S. J. and Werb, Z., Quantitative proteomic analyses of mammary organoids reveals distinct signatures after exposure to environmental chemicals. *Proc. Natl. Acad. Sci. USA*, 2016, **113**, E1343–E1351.
21. Qi, J. *et al.*, New wnt/beta-catenin target genes promote experimental metastasis and migration of colorectal cancer cells through different signals. *Gut*, 2016, **65**, 1690–1701.
22. Ahsan, S. and Drăghici, S., Identifying significantly impacted pathways and putative mechanisms with ipathwayguide. *Curr. Proto. Bioinform*., 2017, **57**, 7.15.11–17.15.30.
23. Draghici, S. *et al.*, A systems biology approach for pathway level analysis. *Genome Res*., 2007, **17**, 1537–1545.
24. Fabregat, A. *et al.*, The reactome pathway knowledgebase. *Nucleic Acids Res.*, 2016, **44**, D481–D487.
25. Bohler, A. *et al.*, Reactome from a wikipathways perspective. *PLOS Comput. Biol*., 2016, **12**, e1004941.
26. Slenter, D. N. *et al*., Wikipathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res*., 2017, **46**, D661–D667.
27. Herwig, R., Hardt, C., Lienhard, M. and Kamburov, A., Analyzing and interpreting genome data at the network level with consensuspathdb. *Nature Protoc.*, 2016, **11**, 1889–1907.
28. Chatr-Aryamontri, A. *et al.*, The bioGRID interaction database: 2015 update. *Nucleic Acids Res*., 2015, **43**, D470–D478.
29. Orchard, S. *et al.*, The mintact project – intact as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*., 2014, **42**, D358–D363.
30. Paz, A. *et al.*, SPIKE: a database of highly curated human signaling pathways. *Nucleic Acids Res*., 2011, **39**, D793–D799.
31. Licata, L. *et al.*, MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res*., 2012, **40**, D857–D861.
32. Peri, S. *et al.*, Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res*., 2003, **13**, 2363–2371.
33. Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O. and Bader, G. D., Cytoscape. js: a graph theory library for visualisation and analysis. *Bioinformatics*, 2016, **32**, 309–311.
34. Biasini, M. *et al.*, SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*., 2014, **42**, W252–W258.
35. Colovos, C. and Yeates, T. O., Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.*, 1993, **2**, 1511–1519.
36. Wang, W., Xia, M., Chen, J., Deng, F., Yuan, R., Zhang, X. and Shen, F., Data set for phylogenetic tree and RAMPAGE Ramachandran plot analysis of SODS in *Gossypium raimondii* and *G. arboreum*. *Data Brief*, 2016, **9**, 345–348.
37. Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res*., 2004, **14**, 1188–1190.
38. Shin, G., Kang, T. W., Yang, S., Baek, S. J., Jeong, Y. S. and Kim, S. Y., GENT: gene expression database of normal and tumor tissues. *Cancer Informat.*, 2011, **10**, 149–157.
39. Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V. and Skrzypek, E., PhosphoSitePlus, 2014: mutations, ptms and recalibrations. *Nucleic Acids Res*., 2015, **43**, D512–D520.
40. Li, D. and Roberts, R., WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cell. Mol. Life Sci*., 2001, **58**, 2085–2097.
41. Stirnimann, C. U., Petsalaki, E., Russell, R. B. and Muller, C. W., WD40 proteins propel cellular networks. *Trends Biochem. Sci*., 2010, **35**, 565–574.
42. Gough, J., Convergent evolution of domain architectures (is rare). *Bioinformatics*, 2005, **21**, 1464–1471.
43. Olsen, G. J. and Woese, C. R., Ribosomal RNA: a key to phylogeny. *FASEB J.*, 1993, **7**, 113–123.
44. Lecompte, O., Ripp, R., Thierry, J. C., Moras, D. and Poch, O., Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res*., 2002, **30**, 5382–5390.
45. Tamura, K. and Alexander, R. W., Peptide synthesis through evolution. *Cell. Mol. Life Sci.*, 2004, **61**, 1317–1330.
46. Schwikowski, B., Uetz, P. and Fields, S., A network of protein–protein interactions in yeast. *Nature Biotechnol*., 2000, **18**, 1257–1261.
47. Stuart, J. M., Segal, E., Koller, D. and Kim, S. K., A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 2003, **302**, 249–255.
48. Sachs, A. B. and Davis, R. W., Translation initiation and ribosomal biogenesis: involvement of a putative rRNA helicase and RPL46. *Science*, 1990, **247**, 1077–1079.
49. Zebarjadian, Y., King, T., Fournier, M. J., Clarke, L. and Carbon, J., Point mutations in yeast CBF5 can abolish *in vivo* pseudouridylation of rRNA. *Mol. Cell. Biol*., 1999, **19**, 7461–7472.
50. Naryshkina, T., Bruning, A., Gadal, O. and Severinov, K., Role of second-largest RNA polymerase i subunit Zn-binding domain in enzyme assembly. *Eukaryotic Cell*, 2003, **2**, 1046–1052.
51. Yang, Y. -T., Ting, Y. -H., Liang, K. -J. and Lo, K. -Y., The roles of puf6 and loc1 in 60S biogenesis are interdependent and both are required for efficient accommodation of rpl43. *J. Biol. Chem*., 2016, **291**, 19312–19323.
52. Oeffinger, M., Dlakic, M. and Tollervey, D., A pre-ribosome-associated heat-repeat protein is required for export of both ribosomal subunits. *Genes Dev*., 2004, **18**, 196–209.
53. Milkereit, P. *et al.*, Maturation and intranuclear transport of pre-ribosomes requires NOC proteins. *Cell*, 2001, **105**, 499–509.
54. Jensen, B. C., Wang, Q., Kifer, C. T. and Parsons, M., The NOG1 GTP-binding protein is required for biogenesis of the 60S ribosomal subunit. *J. Biol. Chem*., 2003, **278**, 32204–32211.
55. Lebreton, A., Saveanu, C., Decourty, L., Jacquier, A. and Fromont-Racine, M., Nsa2 is an unstable, conserved factor required for the maturation of 27 SB Pre-rRNAs. *J. Biol. Chem*., 2006, **281**, 27099–27108.
56. Dunbar, D. A., Dragon, F., Lee, S. J. and Baserga, S. J., A nucleolar protein related to ribosomal protein l7 is required for an early step in large ribosomal subunit biogenesis. *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 13027–13032.
57. Grimm, T. *et al.*, Dominant-negative PES1 mutants inhibit ribosomal RNA processing and cell proliferation via incorporation into the PeBow-complex. *Nucl. Acids Res*., 2006, **34**, 3030–3043.
58. Holzel, M. *et al.*, Mammalian wdr12 is a novel member of the PES1-BOP1 complex and is required for ribosome biogenesis and cell proliferation. *J. Cell Biol*., 2005, **170**, 367–378.
59. Rohrmoser, M. *et al.*, Interdependence of PES1, BOP1, and WDR12 controls nucleolar localization and assembly of the pebow complex required for maturation of the 60s ribosomal subunit. *Mol. Cell. Biol.*, 2007, **27**, 3682–3694.
60. Guerriero, G., Silvestrini, L., Obersriebnig, M., Hausman, J. F., Strauss, J. and Ezcurra, I., A *wdr* gene is a conserved member of a

chitin synthase gene cluster and influences the cell wall in *Aspergillus nidulans*. *Int. J. Mol. Sci.*, 2016, **17**, 1031.
61. Schroder, W., Lambert, D. G., Ko, M. C. and Koch, T., Functional plasticity of the n/ofq-nop receptor system determines analgesic properties of nop receptor agonists. *Br. J. Pharmacol*., 2014, **171**, 3777–3800.
62. Leal, M. F. *et al.*, Identification of suitable reference genes for investigating gene expression in anterior cruciate ligament injury by using reverse transcription-quantitative pcr. *PLOS ONE*, 2015, **10**, e0133323.
63. Mann, K. M., Ying, H., Juan, J., Jenkins, N. A. and Copeland, N. G., KRAS-related proteins in pancreatic cancer. *Pharmacol. Therapeut.*, 2016.
64. Suzuki, Y., Orita, M., Shiraishi, M., Hayashi, K. and Sekiya, T., Detection of *ras* gene mutations in human lung cancers by single-strand conformation polymorphism analysis of polymerase chain reaction products. *Oncogene*, 1990, **5**, 1037–1043.
65. Chung, K. Y., Cheng, I. K., Ching, A. K., Chu, J. H., Lai, P. B. and Wong, N., Block of proliferation 1 (BOP1) plays an oncogenic role in hepatocellular carcinoma by promoting epithelial-to-mesenchymal transition. *Hepatology*, 2011, **54**, 307–318.
66. Wrzeszczynski, K. O. *et al.*, Identification of tumor suppressors and oncogenes from genomic and epigenetic features in ovarian cancer. *PLOS ONE*, 2011, **6**, e28503.
67. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K., dbsNP: the NCBI database of genetic variation. *Nucleic Acids Res*., 2001, **29**, 308–311.
68. Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R., Tattikota, S. G. and Pandey, A., A curated compendium of phosphorylation motifs. *Nature Biotechnol*., 2007, **25**, 285–286.
69. Rechsteiner, M. and Rogers, S. W., Pest sequences and regulation by proteolysis. *Trends Biochem. Sci*., 1996, **21**, 267–271.
70. Rechsteiner, M., Pest sequences are signals for rapid intracellular proteolysis. *Sem. Cell Biol*., 1990, **1**, 433–440.
71. Chen, C. K., Chan, N. L. and Wang, A. H., The many blades of the beta-propeller proteins: conserved but versatile. *Trends Biochem. Sci*., 2011, **36**, 553–561.
72. Strezoska, Z., Pestov, D. G. and Lau, L. F., Functional inactivation of the mouse nucleolar protein BOP1 inhibits multiple steps in pre-rRNA processing and blocks cell cycle progression. *J. Biol. Chem*., 2002, **277**, 29617–29625.
73. Riley, T., Sontag, E., Chen, P. and Levine, A., Transcriptional control of human p53-regulated genes. *Nature Rev. Mol. Cell Biol*., 2008, **9**, 402–412.

ACKNOWLEDGEMENTS. R.A. thanks the Science and Engineering Research Board, Government of India for grant support (EMR/2016/005994).

Received 13 February 2019; revised accepted 22 May 2019

doi: 10.18520/cs/v117/i3/422-433