Genomics 110 (2018) 283–290

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Evolutionary rate heterogeneity between multi- and single-interface hubs across human housekeeping and tissue-specific protein interaction network: Insights from proteins' and its partners' properties

Kakali Biswas[a], Debarun Acharya[a], Soumita Podder[a,b], Tapash Chandra Ghosh[a,*]

[a] Bioinformatics Centre, Bose Institute, P-1/12, C.I.T. Scheme VII M, Kolkata 700 054, India
[b] Department of Microbiology, Raiganj University, Raiganj, Uttar Dinajpur 733134, India

ARTICLE INFO

Keywords:
Tissue-specific hubs
Housekeeping hubs
Multi-interface hubs
Single-interface hubs
dN/dS ratio
Functional divergence
Conformational diversity

ABSTRACT

Integrating gene expression into protein-protein interaction network (PPIN) leads to the construction of tissue-specific (TS) and housekeeping (HK) sub-networks, with distinctive TS- and HK-hubs. All such hub proteins are divided into multi-interface (MI) hubs and single-interface (SI) hubs, where MI hubs evolve slower than SI hubs. Here we explored the evolutionary rate difference between MI and SI proteins within TS- and HK-PPIN and observed that this difference is present only in TS, but not in HK-class. Next, we explored whether proteins' own properties or its partners' properties are more influential in such evolutionary discrepancy. Statistical analyses revealed that this evolutionary rate correlates negatively with protein's own properties like expression level, miRNA count, conformational diversity and functional properties and with its partners' properties like protein disorder and tissue expression similarity. Moreover, partial correlation and regression analysis revealed that both proteins' and its partners' properties have independent effects on protein evolutionary rate.

1. Introduction

Cells are the fundamental unit of life. Except for the unicellular ones, every living organism possesses diverse types of cells adapted to perform specialized functions. The functions of each cell are mediated by the molecular machinery, of which proteins play an essential part. Proteins interact with each other and perform almost all the fundamental life processes. Such interactions involve interfaces or domains, which execute the functions of the protein. Protein domains play a crucial part in molecular evolution since these are used as structural building blocks and may create proteins with discrete functions due to exon shuffling [1–3]. The advancement in high-throughput protein interaction data helps to analyze protein functions from the network perspective. Moreover, within the whole protein interaction network, there are some small, densely linked components formed by the interactions between proteins, nucleic acids, and other small molecules, and are weakly connected to the rest of the protein-interaction network. These components are termed as modules [4]. Recent advances in discovery and revision of the proteins in modules using computational biology have enabled us to model these protein-protein interactions as a network where proteins represent the nodes with interactions as links between the nodes.

Inside the protein-protein interactions network (PPIN), proteins with a high degree of connectivity are found to be essential and are likely to perturb the PPIN upon deletion, misfunction or misregulation [5]. These proteins, named as hub, are distinct from lesser connected proteins or non-hubs and are evolutionary more conserved. Although most of the earlier studies featuring hub proteins from evolutionary perspective compared hub and non-hub proteins in PPIN, more recent studies aim at detailed analysis of hub proteins. One such study by Han et al. classified hub proteins into two groups - multi-interface hub proteins (MI or party hubs) and single interface hub proteins (SI or date hubs) based on protein domain architecture and correlated expression of the interacting partners [6]. Comparing the evolutionary rate between these MI and SI hubs revealed discrete differences—MI proteins were found to be evolutionarily more conserved than SI proteins [7], which may be mainly due to selective constraint acting on a larger region in MI proteins, as it usually possesses more interacting surfaces. Additionally, the party hubs mediate within-module interactions (intra-module), whereas date hubs integrate between modules (inter-module) [7]. However, the SI proteins acting on various modules face stronger consequences when deleted than the less pervasive densely connected MI proteins, due to their association with diverse functions [8]. Besides, a few studies have been carried out to understand the structural

* Corresponding author.
E-mail address: tapash@jcbose.ac.in (T.C. Ghosh).

https://doi.org/10.1016/j.ygeno.2017.11.006
Received 23 June 2017; Received in revised form 10 November 2017; Accepted 29 November 2017
Available online 02 December 2017

(conformational) and functional role of these hub proteins [9,10]. The functions of a protein are mediated mainly by its structure. Although each protein is thought to possess definite three-dimensional conformation determined by its amino acid sequence, that may not be the only conformation adopted by the protein within a cellular system [11]. The magnitude of conformational diversity encompasses structural changes like fluctuation of protein's side chains and the movement of loops and secondary structures, even to the global rearrangement in protein tertiary structure [12].

Further insights into human PPIN classified topological variation based on gene expression data. Based on gene expression breadth, all genes are grouped as either tissue-specific (TS), or housekeeping (HK). Previous studies revealed many differences between the HK and TS genes in humans. Human HK genes are more compact in structure, containing shorter intron length, 5′ UTR length and coding sequence length [13]. Consistent with this, HK genes are enriched in shorter repetitive sequences such as Alu-elements, but depleted in longer repetitive sequences like Long Interspersed Nuclear Element 1 (LINE-1) elements [14]. Additionally, elucidation of evolutionary rate differences among these two groups resulted in similar findings across organisms as diverse as unicellular fungi to humans, the housekeeping genes (HK) evolve slower than tissue-specific genes (TS) [15]. Accordingly, the whole PPI network was also grouped into tissue-specific or local network and housekeeping or global network, where TS hubs (TSH) evolve faster than HK hubs (HKH). These TSH also feature longer genes, less protein expression abundance, tight regulation and greater protein intrinsic disorder content than HKH [54]. Additionally, within the PPI network, HK genes are more central and are associated with core cellular processes whereas TS genes are more peripheral with modified core cellular processes as well as regulatory and developmental functions [16–18]. However, these findings remain confounding as some TS genes are reported to evolve slower than even this HK class of genes [19–21]. To address this issue, Podder et al. classified human proteins into MI and SI counterparts and analyzed the evolutionary rate of TS and HK genes between these two groups. They found that within MI proteins, both TS- and HK-genes show similar evolutionary rates, whereas, within SI proteins, HK genes evolve slower than TS genes [10]. Furthermore, recent studies based on PPI-network properties highlights the impact of the partner proteins on proteins' evolutionary rate [16,22], as the interacting partners also contribute to the central node evolution via the domain-domain interaction [23]. Such analysis on HK- and TS-hubs revealed that interacting partners of the TSH are more conserved than HKH with diverse subcellular localization [22]. However, these studies lack detailed insights into the protein interaction network-based properties and the influence of interacting partners on the evolutionary rate. Therefore, a detailed spatially resolved analysis is required to explain the evolutionary rate variation between these two hub classes.

In this study, we delved deep into the understanding of protein evolutionary rate based on their expression breadth (whether housekeeping or tissue-specific) and the contribution of domain number (whether single or multiple) to it. We tried to identify at which level the evolutionary conservation endures. Furthermore, we sought to explore which among the two: protein's own property or partner properties influence the evolutionary rate of proteins the most.

2. Materials and methods

2.1. Retrieval of dataset

We obtained tissue specific gene expression data from EMBL-EBI expression atlas (https://www.ebi.ac.uk) for “baseline” expression where the expression level of each gene in normal and untreated conditions. Then we calculated tissue specificity index τ [24] of each gene for tissue specificity using the following formula [10]−

$$\tau = \frac{\sum_{j=1}^{\eta H}\left(1 - \left[\frac{\log_2 S_H(i,j)}{\log_2 S_H(i,max)}\right]\right)}{\eta_H - 1}$$

(where, ηH = number of human tissues examined and S_H (i, max) = highest expression signal of gene i across the ηH tissues). The value ranges from 0 to 1, where genes with τ-values close to ‘0’ are considered to be more towards housekeeping and those with τ-values close to ‘1’ are considered as tissue-specific (TS). τ = 0 represents equal expression of the gene across all tissue, i.e. housekeeping (HK) genes. We sorted our dataset according to an increasing τ values and obtained genes from extreme 20% of the population from both ends. Thereby, we obtained 1198 HK and 7767 TS genes.

2.2. Protein connectivity data retrieval and interacting domain identification

Protein-protein interaction data was obtained from BioGRID (release 3.4.130) (https://thebiogrid.org/) [25]. Genes with at least five interacting partners were considered to be highly connected or hub proteins. We obtained human protein sequences from the UCSC genome browser (http://genome.ucsc.edu). Interacting domains were retrieved from Pfam repository (http://pfam.sanger.ac.uk/) [55]. The hypothesis behind the Pfam data retrieving was that the interacting domains confer binding capability to protein regions. The cut-off values used for domain assignment are (1) e-value of alignment $e^{< 1.0 \times 10^{-4}}$; (2) domain length > 12; (3) matched sequence length > 80% of domain length [26]. In particular, single interface proteins were designated as having few interaction interfaces (two at most) and multi-interface proteins having more than two interacting interfaces [27]. The numbers of HKH_MI and HKH_SI proteins are respectively 303 and 895. The numbers of MI and SI proteins belonging to TSH PPIN are 1705 and 6062, respectively.

2.3. Estimation of evolutionary rate

The evolutionary rates of human genes were calculated by dividing non-synonymous substitution rate (dN) with synonymous substitution rate (dS). The dN and dS values were retrieved from BioMart interface of Ensembl Version 87 (http://www.ensembl.org/biomart/martview) [28] for *Homo sapiens* (GRCh37) using one to one Human-Mouse as well as Human-Chimpanzee orthologous pairs.

2.4. Prediction of miRNA targets sites and gene expression level assessment

The number of miRNA targets per gene were obtained from TargetScan (release 6.2) (http://www.targetscan.org) [29] for its more reliable data over other databases. Tissue-wise RNA-seq gene expression data was obtained from the human protein atlas [30]. Average gene expression level of HK genes was calculated by considering only those tissues where it shows higher than mean expression level calculated for all tissues. Expression level for TS genes represents only the tissue where the desired gene is expressed at its highest level.

2.5. Collection of conformational and functional annotation

Protein conformational diversity data was acquired from CoDNaS database [31]. The database utilizes a total of 70,467 PDB structures (Protein Data Bank, a repository of biological macromolecular structure) [32], representing a set of 9398 monomeric proteins of the protein data bank. Conformational diversity was measured as the maximum RMSD (root-mean-square deviation measuring the average distance between the superimposed atoms) between available conformers of a protein. RMSD values were normalized to RMSD100 for all proteins with > 40 residues [33]. This provided us with 1094 human proteins with corresponding conformational diversity values.

Next, we acquired the protein-coding human genes with functional annotation from Ensembl Genome Browser (http://www.ensembl.org/) [28] for "biological process" GO classification for individual gene and its paralog. Functional divergence was determined using Czekanowski-Dice distance formula [34].

$$\text{Functional distance (i, j)} = \frac{\text{Number of(Terms (i)}\Delta\text{Terms(j))}}{[\text{Number of(Terms(i)} \cup \text{Terms(j))} + \text{Number of (Terms(i)} \cap \text{Terms(j))}]}$$

Here, i represents GO terms of individual human genes, j represents GO terms of the paralogous genes, Δ corresponds to the symmetrical difference between the GO term sets of two genes, ∪ and ∩ represents the non-redundant and common GO terms, respectively.

2.6. Protein disorder content estimation

Protein disordered residues were predicted from one of the top disorder predictors: IUPred algorithm [35,36]. It provides a fair estimation of disorder residue by assigning disorder tendency score for each residue by their ability to form favorable pair-wise contacts with neighboring amino acids [37]. Protein disorder content was defined as the fraction of the total number of such disordered residues within a protein. Moreover, flexible loops were trimmed down from the calculation by taking only 30 or more consecutive predicted disordered residues at a stretch. Other stretches were denominated as ordered regions [38].

2.7. Protein tissue expression similarity calculation

As described earlier, proteins were designated as tissue-specific or housekeeping depending on their τ (tau) value. Furthermore, the name of each tissue where the protein is expressed with the highest level of expression along with the higher bin of tau value was denoted for that tissue-specific (Top 20%). Now, these tissue names for each gene data was integrated with protein-protein interaction data among protein and its partner. Tissue expression similarity between a protein (y) and its interacting partner (z) was calculated as

$$\text{Tissue expression similarity } (y, z) = 1 - \frac{\text{Number of } (Tissues(y)\,\Delta\, Tissues(z))}{[\text{Number of } (Tissues(y) \cup (Tissues(z)) + \text{Number of } (Tissues(y) \cap (Tissues(z)]}$$

Here, '*Tissues*(y)' and '*Tissues*(z)' represent the name of tissues where the protein 'y' and its interacting partner 'z' is expressed, respectively., Δ corresponds to the symmetrical difference between the tissues where the two proteins are expressed, ∪ and ∩ represents the nonredundant and common tissues, where proteins 'y' and 'z' are expressed.

2.8. Statistical analyses

All the statistical tests were performed using the SPSS (20.0) package [39]. Non-parametric Spearman's correlation test was used to evaluate the correlation coefficient between two parameters. Difference between parameters was calculated with Mann–Whitney U test. Linear and categorical regression analysis was performed using ANOVA model for understanding the relationship of the parameters with dN/dS ratio.

3. Results

3.1. Analysis on evolutionary rate difference among different hub protein classes

In this study, we explored the effect of tissue-specificity in modulating the evolutionary rate differences of human multi- and single-interface hubs. Previous studies suggest that highly connected or hub-proteins evolve slower than lowly connected or non-hub proteins [40]. Additionally, housekeeping genes are well-known for their slower evolutionary rate (than the tissue-specific genes) and so are the multi-interface hubs (than the single-interface hubs) [15,27]. In our analysis, we used the high-throughput RNA-seq data from the Human Protein Atlas [30] to obtain housekeeping and tissue-specific genes and physical protein interaction data from Biogrid [25] and co-expression data from [30] to obtain multi-and single-interface proteins and achieved similar trends for both the cases, that is, multi-interface (MI or party-hubs) evolve slower than single-interface (SI or date hub) proteins (Table 1). However, the evolutionary rate differences between MI and SI proteins within the housekeeping and tissue-specific groups are not yet clear. Therefore, we subdivided human housekeeping-(HKH) and tissue-specific hub (TSH) genes into MI and SI proteins and obtained four classes: HKH_MI, HKH_SI, TSH_MI and TSH_SI (Supplementary Tables S1A and S1B in Supplementary File 1). Comparing the evolutionary rate differences between MI and SI proteins in HKH and TSH groups using dN/dS ratio revealed that significant difference exists in the case of TSH_MI and TSH_SI (TSH_MI $<$ TSH_SI) but not in the case of HKH_MI and/SI (Table 1). However, as the MI proteins contain larger regions under selective constraint, we investigated the influence of protein size on our findings. We classified the proteins in our dataset into 'Small' and 'Large' classes depending on the median protein length. We found that the protein length has no influence in our dataset as the trend remains the same in both the length bins (Fig. 1). To explain this further, we studied the most probable parameters leading to such

Table 1
Average dN/dS ratio of different hub-proteins calculated using Human-Mouse and Human-Chimpanzee orthologs. *P*-value indicates significance level derived from Mann-Whitney *U* test ['*' denotes significant differences].

Orthologous gene pair	Category	Average dN/dS	Significance level
A. Difference between the evolutionary rate (dN/dS ratio) of tissue-specific hubs (TSH) and housekeeping hubs (HKH) using mouse and chimpanzee as outgroups			
Human-Mouse	TSH (n = 3691)	0.158	$P = 1.00 \times 10^{-6}$*, $\alpha < 0.001$
	HKH (n = 457)	0.094	
Human-Chimpanzee	TSH (n = 5248)	0.332	$P = 9.50 \times 10^{-4}$*, $\alpha < 0.001$
	HKH (n = 449)	0.287	
B. Difference between the evolutionary rate (dN/dS ratio) of multi-interface (MI) and single-interface (SI) hubs using mouse and chimpanzee as outgroups			
Human-Mouse	MI (n = 890)	0.133	$P = 1.00 \times 10^{-6}$*, $\alpha < 0.001$
	SI (n = 3258)	0.156	
Human-Chimpanzee	MI (n = 1292)	0.293	$P = 1.00 \times 10^{-6}$*, $\alpha < 0.001$
	SI (n = 4405)	0.339	
C. Difference between the evolutionary rate (dN/dS ratio) of MI- and SI-hubs within TSH and HKH genes using mouse and chimpanzee as outgroups			
Human-Mouse	TSH_MI (n = 770)	0.141	$P = 1.00 \times 10^{-6}$*, $\alpha < 0.001$
	TSH_SI (n = 2921)	0.163	
	HKH_MI (n = 120)	0.832	$P = 6.33 \times 10^{-2}$, $\alpha > 0.05$
	HKH_SI (n = 337)	0.991	
Human-Chimpanzee	TSH_MI (n = 1167)	0.214	$P = 1.00 \times 10^{-6}$*, $\alpha < 0.001$
	TSH_SI (n = 4081)	0.284	
	HKH_MI (n = 125)	0.132	$P = 4.38 \times 10^{-1}$, $\alpha > 0.05$
	HKH_SI (n = 329)	0.285	

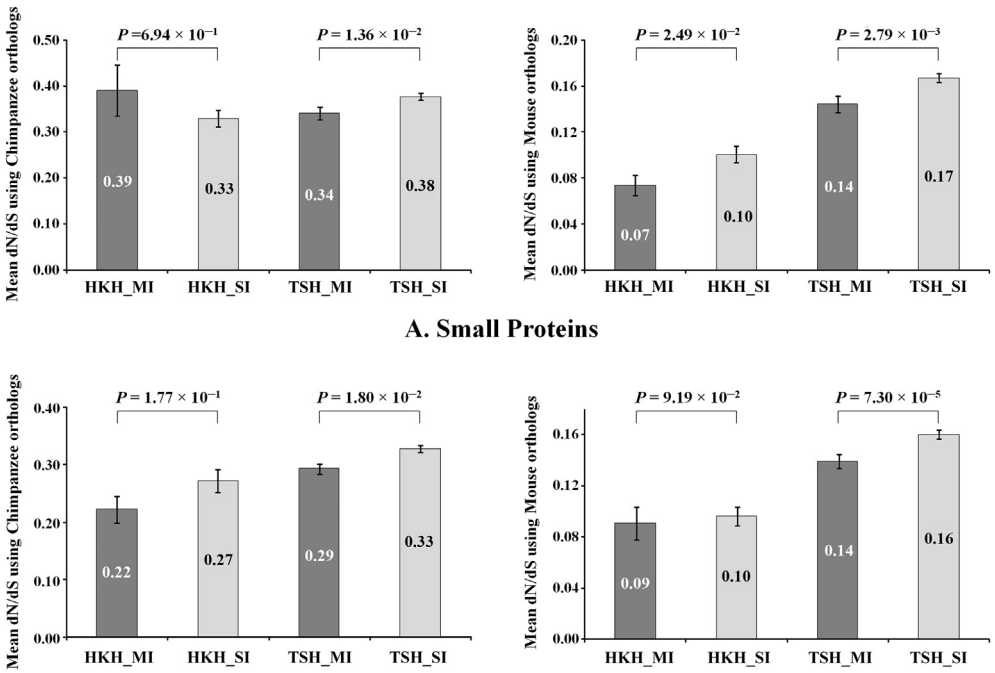

Fig. 1. Evolutionary rate (dN/dS ratio) differences between MI and SI proteins of HKH and TSH class of genes in 'small' (protein length = 24 amino acids - 473 amino acids; N = 3383) and 'large' (protein length = 474 amino acids - 8924 amino acids; N = 3391) proteins based on the median protein length (= 474 amino acids).

consequence. We focused on not only the protein but also their interacting partners' structural as well as functional properties guiding such evolutionary rate differences.

3.2. Role of gene expression and regulation on evolutionary rate of MI and SI hubs

One of the major determinants of protein evolution is expression level. Highly expressed genes evolve slowly, a phenomenon known as E-R anti-correlation [41]. We calculated the average expression level of MI and SI hubs across HKH and TSH classes and observed that MI proteins are significantly more highly expressed than SI proteins in TSH class whereas, no significant difference in average expression level was observed in these two groups of HKH class (Table 1, Fig. 2). Additionally, genes with a higher number of miRNA targets are also reported to be conserved [15]. Consistent with this, we found that

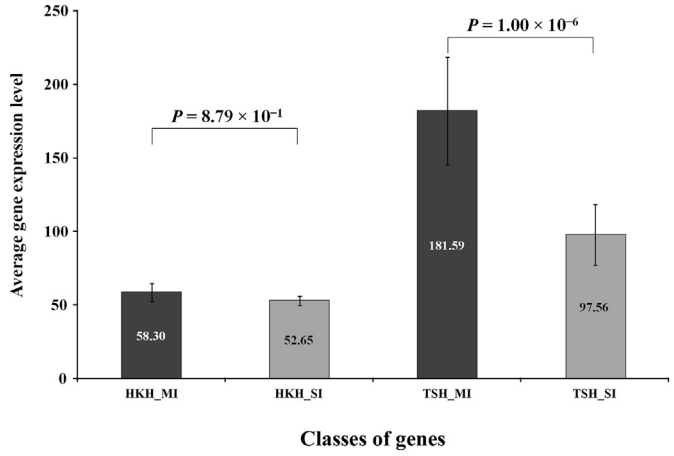

Fig. 2. Average gene expression level difference among MI and SI proteins of HKH and TSH class of genes. P value indicates significance level derived from Mann-Whitney U test.

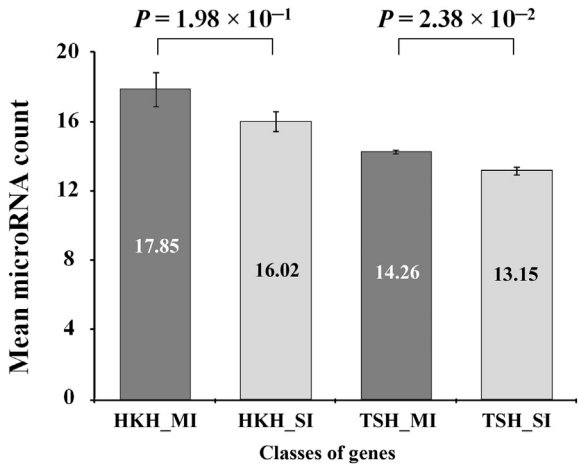

Fig. 3. Average number of miRNA target per gene difference among MI and SI proteins of HKH and TSH class of genes. P value indicates significance level derived from Mann-Whitney U test.

TSH_MI hubs are targeted by significantly more miRNAs than TSH_SI hubs (Fig. 3). However, the HKH_MI and HKH_SI did not show any difference in average miRNA number among them. Together, these results may serve as a probable reason for evolutionary discrepancy among TSH_MI/SI and HKH_MI/SI.

3.3. Role of structural and functional properties on the evolutionary rate of MI and SI hubs

Recent studies on protein functions primarily focused on this conformational diversity of proteins [11], which is found to be negatively correlated with evolutionary rate [42], mainly because it increases the functional diversity of proteins. Hence, we look for the conformational diversity of MI and SI hubs present in HK and TS PPIN. Accordingly, we found that MI proteins possess a significantly higher conformational diversity than SI proteins only for TSH class and not the HKH counterpart (Table 2).

Table 2
Average values for structural and functional properties of TSH and HKH. *P*-value indicates significance level derived from Mann-Whitney *U* test ['*' denotes significant differences].

Parameters	Classes of genes	Average value	Significance level
Average conformational diversity	TSH_MI (n = 198)	1.35	$P = 3.38 \times 10^{-2}$*, $\alpha < 0.05$
	TSH_SI (n = 229)	1.21	
	HKH_MI (n = 84)	1.33	$P = 4.54 \times 10^{-1}$, $\alpha > 0.05$
	HKH_SI (n = 137)	1.21	
Average functional diversity per gene	TSH_MI (n = 427)	0.65	$P = 8.89 \times 10^{-3}$*, $\alpha < 0.01$
	TSH_SI (n = 1267)	0.61	
	HKH_MI (n = 43)	0.65	$P = 3.72 \times 10^{-1}$, $\alpha > 0.05$
	HKH_SI (n = 82)	0.67	
Average core function per gene	TSH_MI (n = 820)	2.14	$P = 3.19 \times 10^{-2}$*, $\alpha < 0.05$
	TSH_SI (n = 2602)	2.05	
	HKH_MI (n = 209)	2.44	$P = 4.29 \times 10^{-1}$, $\alpha > 0.05$
	HKH_SI (n = 548)	2.32	

Additionally, protein functional diversity between the paralogous pairs has long been treated as one of the key guiding factors of protein evolution [43–47]. Although gene duplication initially leads to the relaxation of purifying selection, the subsequent functional divergence between paralogs imposes selective constraints and slows down the evolutionary rate [48,49]. In this study, we noticed that MI proteins have a significantly higher functional divergence than SI proteins within the TSH class but not in HKH class (Table 2), indicating the selective constraints are higher for MI-TSH groups, which may be the underlying cause of their slower evolutionary rates. Furthermore, it was also reported that proteins performing core biological processes like metabolism, protein synthesis and its transport are largely conserved across species compared to the proteins involved in more regulatory processes like transcription factor binding or signal transduction [47]. Using gene ontology (GO) terms for the GO domain 'biological process' (GO-BP) [47] we noticed that number of core functions differ in MI and SI only within TSH but not in HKH (Table 2). However, the number of regulatory functions does not differ between the MI and SI proteins within both TSH and HKH classes. Thus, differences in conformational diversity along with functional diversity and association with core functions may impose higher selection pressure on TSH_MI compared to TSH_SI, whereas such differences are not attributable to MI and SI classes of HKH proteins.

3.4. Role of tissue-specificity similarity and protein intrinsic disorder content of protein partners on its evolutionary rate

Tissue-specific proteins making fewer interactions evolve faster than highly interacting housekeeping proteins [16]. An earlier study also deciphered the influence of interacting partners' properties on a protein's evolutionary rate [50]. Additionally, analysis of TSH and HKH genes' partners revealed that partners of TSH genes evolve slower than partners of HKH genes [22]. Thereby, we sought to investigate whether the tissue distribution of TSH genes and their interacting partners has any role in evolutionary rate. To do this, we constructed a tissue-specific similarity index according to the protein and its partner's tissue expression profile (explained in the Materials and methods section). Interestingly, we obtained a negative correlation ($\rho = -0.189$, $n = 4128$, $P = 1 \times 10^{-6}$) between tissue expression similarity with evolutionary rate, which also demonstrated that when a gene and its' interacting partner have a higher tissue-expression similarity, they are evolutionary more conserved than gene having interacting partners with lower tissue expression similarity (Fig. 4). Almost all of the housekeeping genes share similar tissue similarity with their partners as they are ubiquitously expressed in all tissue types.

Moreover, in an interaction network, SI proteins are more disordered than MI proteins [26,51] and perform transient interactions with their partners. However, when the interacting partners' intrinsic disorder content was analyzed in both TSH and HKH, we found significantly higher protein disorder content in interacting partners of TSH_SI than that of TSH_MI. Such a significant difference was not observed between the two HKH groups (Fig. 5). Thus, both partner proteins' tissue expression similarity, as well as intrinsic disorder content, may impact on a dissimilar evolutionary rate between TSH_MI and TSH_SI.

3.5. Influence of the studied factors on evolutionary rate

To examine whether each of the above mentioned parameters has a significant influence on evolutionary rate, we performed Spearman's rank correlation analysis by considering dN/dS as scalar dependent variable and all other parameters as explanatory variables. We found that dN/dS ratio upholds significant negative correlations with mean miRNA count, expression level, conformational diversity, functional diversity, core functional processes, domain similarity, partners' average disorder content and tissue similarity with partners (Table 3). Next, we intend to find out whether protein's own properties or its partners' properties are more influential in guiding protein evolutionary rate or if they act in a mutually exclusive way. For this we have performed partial correlation analysis in two ways— we have controlled all

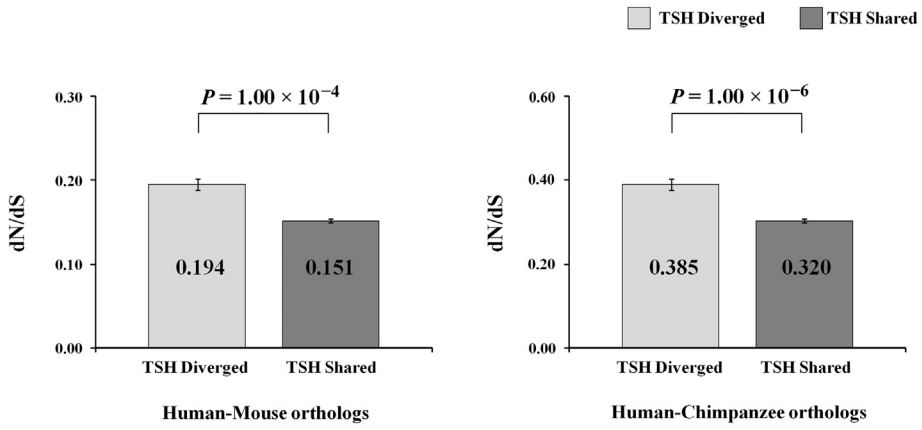

Fig. 4. Evolutionary rate (dN/dS ratio) differences between tissue-specific genes with similar (TSH_{shared}) and different ($TSH_{diverged}$) tissue-specificity similarity with their interacting partners. Human-Mouse and Human-Chimpanzee 1:1 orthologs were used to calculate the dN/dS ratio. *P*-values are provided in the figure.

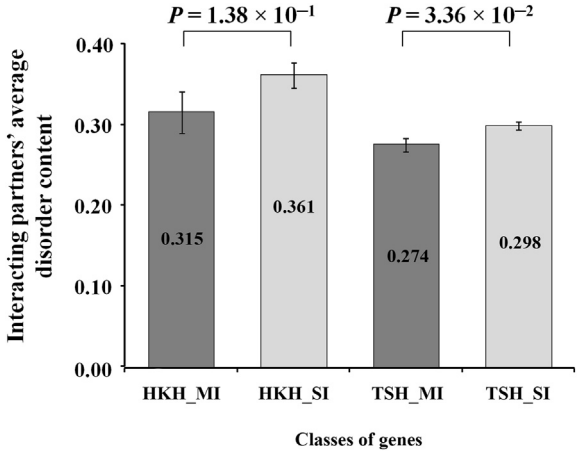

Fig. 5. Interacting partners' average disorder content for a given hub protein with difference among MI and SI proteins of HKH and TSH class of genes. *P* value indicates significance level derived from Mann-Whitney *U* test.

Table 3
Values of nonparametric correlation analysis using dN/dS ratio as a scalar dependent variable ['*' denotes significant differences].

Explanatory variables	Spearman's rho (ρ) correlation coefficient	Significance level (two-tailed)
Number of miRNA per gene (n = 3519)	− 0.271	$P = 1 \times 10^{-6}$*, $\alpha < 0.001$
Gene expression level (n = 4148)	− 0.073	$P = 2 \times 10^{-6}$*, $\alpha < 0.001$
Conformational diversity (n = 4148)	− 0.105	$P = 1 \times 10^{-6}$*, $\alpha < 0.001$
Average of disorder residue in interacting partner (n = 1710)	0.171	$P = 1 \times 10^{-6}$*, $\alpha < 0.001$
Average functional divergence (n = 742)	− 0.075	$P = 4.19 \times 10^{-2}$*, $\alpha < 0.05$
Tissue expression similarity of protein and its partner (4128)	− 0.113	$P = 1 \times 10^{-6}$*, $\alpha < 0.001$

the partners' properties (such as partners' average disorder content and tissue similarity with partners) and noticed the correlation between protein's own properties and evolutionary rate and also vice versa. The result is delineated in (Table 4) which indicates that both the protein's and its interacting partners' properties guide the evolutionary rate in a mutually exclusive way. By using linear regression analysis we have confirmed that evolutionary rate (dN/dS ratio) of proteins are independently influenced by its own properties like number of miRNAs per gene ($\beta = -0.142$, $P = 1.45 \times 10^{-2}$) as well as protein's interacting partner properties such as tissue expression similarity ($\beta = 0.080$, $P = 4.01 \times 10^{-2}$) are also important for determining the evolutionary rate of hub proteins across housekeeping and tissue specific genes.

4. Discussion

Integrating the protein-protein interaction (PPI) network with high-throughput gene expression data, researchers divided all human PPI network into sub-network of housekeeping or global and tissue-specific or local interacting parts. Highly connected (hub) proteins within PPI network are further divided into multi-interface (MI or party-hub) and single-interface (SI or date-hubs) hubs, based on the number of their interacting interface. MI-hubs, in general, evolve slower than SI hubs

Table 4
Values of partial correlation analysis using dN/dS ratio as a scalar dependent variable ['*' denotes significant differences].

Explanatory variables	Correlation value (r)	Significance level
Control for partners' properties		
miRNA (n = 307)	− 0.152	$P = 7.37 \times 10^{-3}$*, $\alpha < 0.01$
Core functions (n = 307)	− 0.116	$P = 4.17 \times 10^{-2}$*, $\alpha < 0.05$
Control for proteins' intrinsic properties		
Average disorder content in partners (n = 304)	0.115	$P = 4.38 \times 10^{-2}$*, $\alpha < 0.05$

[7] due to evolutionary constraints acting on larger surfaces. When the hub proteins from both the housekeeping (HK) and tissue-specific network (TS) were classified into MI and SI hubs, we found that MI proteins evolve slower than SI proteins in the TS PPIN, but not in the HK PPIN (Table 1), a trend slightly different from previous study [10]. Similar results were obtained after splitting all proteins in 'Small' (below-median) and 'Large' (above-median) bins, depending on their length, indicating the protein size has no significant impact on the observations (Fig. 1). As evolutionary rate exhibits a strong negative correlation with gene expression level ($\rho = -0.168$, $P = 9.75 \times 10^{-4}$), we presumed that comparison of gene expression level between MI and SI genes within HK- and TS-hubs might provide insight into their evolutionary rate difference. We found a significantly higher gene expression level in MI proteins in TSH class, whereas in HKH class, both MI and SI express at a similar level (Fig. 2). Thus, gene expression level seems to be a major determinant influencing the evolutionary rate of MI and SI proteins within HK and TS network. However, gene expression is regulated by numerous factors, of which miRNAs are a predominant regulator. Accordingly, the hub proteins are likely to have a high level of miRNA regulation with diverse local and global coordinated regulation [52]. Since regulatory stringency is supposed to be similar in all housekeeping genes, we did not get any significant difference in number of miRNA targets between MI and SI hubs. Whereas, tissue-restricted genes with diverse local sub-networks hold different regulatory constraint between MI and SI hubs, reflected by a greater number of miRNA per gene in MI_TSH proteins (Fig. 3), despite their higher gene expression, which is quite contradictory. However, our result is in agreement with the fact that genes with more miRNA target sites evolve slowly [53]. Now, the interaction between proteins in PPIN may be aided by multiple conformations of the same protein. This diverse conformation of a protein facilitates greater selection pressure on the protein-coding gene to maintain the structural domain/s via which the proteins interact. A strong negative correlation ($\rho = -0.186$, $P = 1.75 \times 10^{-4}$) between dN/dS and protein conformational diversity, as observed in our study also strengthen this hypothesis. Additionally, for duplicated genes, functional diversity between paralogs is a significant contributor to protein evolutionary rate, as it builds up selective constraints that were reduced immediately after gene duplication. It is fascinating to note that the global interacting proteins (HKH-MI and HKH-SI) with cellular maintenance purposes do not show a significant difference in conformational diversity or functional diversity. Conversely, local network of TSH proteins significantly differs in both conformational and functional diversity. This may be due to the fact that sub-networks within TS PPIN might encounter diverse selective pressure for maintaining these various expressional, conformational and functional similarities with their interacting partners.

Next, we intended to identify the contribution of interacting partners on proteins' evolutionary rate. As proteins collaborate to function as a unit, the impact of its partner on its evolutionary rate must be sought out. A significant negative correlation between tissue expression

similarities with evolutionary rate suggests that when a protein and its interacting partners possess a higher tissue expression similarity, it exhibit more evolutionary conservation, which is also supported by the differences within TSH proteins when they show similar (TSH_{shared}) and different ($TSH_{diverged}$) tissue-specificity similarity with their interacting partners (Fig. 4). The interacting partners of TSH_SI proteins was found to content higher protein intrinsic disorder content than TSH_MI class (Fig. 5), indicating their higher propensity to form transient tissue-specific interactions that are signatures of this group of proteins. Moreover, interactions involving proteins with lower tissue-expression similarities are also essential to maintain the connections required for the combined performance of the proteins in PPI network. Therefore, linking housekeeping and tissue-specific genes are much vital for maintaining the overall performance of a human body.

Furthermore, we performed a statistical analysis combining the impact of both proteins' own property (such as expression level, number of miRNA count, conformational diversity and other functional properties) and its partners' properties (like intrinsic protein disorder and tissue expression similarity of the interacting protein partners) on proteins' evolutionary rate. Our findings suggest that genomic novelties are more introduced by intermodular hubs or SI-hubs in the tissue-specific network only. Whereas, MI proteins remain highly conserved within this network for performing core cellular processes and are under more stringent regulation. Conversely, the housekeeping genes with greater cellular maintenance functions might not permit the HKH_SI to undergo mutation, as it could be lethal to the system. Our findings illustrate that evolutionary rate of proteins is equally governed by both its partner properties along with protein's own properties.

5. Conclusion

Our study demonstrates that lower evolutionary rate of MI hubs than SI hubs is only present in the TSH but not in HKH of human PPIN, an analysis done for the first time. We here, provide statistical evidence to establish that both structural and molecular properties of protein as well as interacting partners implicated for determining protein evolutionary rate. Thus, our study makes new findings in exploring interacting partner's properties in the conservation of global and local protein interaction networks.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2017.11.006.

Abbreviations

TS	tissue-specific
HK	housekeeping
PPIN	protein-protein interaction network
TSH	tissue-specific hub
HKH	housekeeping hub
MI	multi-interface protein
SI	single interface protein
GO	gene ontology
BP	biological processes
dN	nonsynonymous nucleotide substitution per nonsynonymous site
dS	synonymous nucleotide substitution per synonymous site
miRNA	micro-RNA

Contributors

Conceived and designed the experiments: KB, SP, TCG. Performed the experiments: KB. Analyzed the data: KB, DA, SP, TCG. Wrote the paper: KB, DA. All authors read and approved the final article for submission.

Conflict of interests

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the UGC: Rajiv Gandhi National Fellowship (Sanction No. RGNF-2012-13-SC-WES-32829).

References

[1] P. Bork, Building protein-structure and function from modular units, FEBS Lett. 286 (1991) 47–54.
[2] H. Hegyi, M. Gerstein, The relationship between protein structure and function: a comprehensive survey with application to the yeast genome, J. Mol. Biol. 288 (1999) 147–164.
[3] I.D. Campbell, A.K. Downing, Building protein-structure and function from modular units, Trends Biotechnol. 12 (1994) 168–172.
[4] V. Spirin, L.A. Mirny, Protein complexes and functional modules in molecular networks, Proc. Natl. Acad. Sci. 100 (2003) 12123–12128.
[5] X. He, J. Zhang, Why do hubs tend to be essential in protein networks? PLoS Genet. 2 (2006) 826–834.
[6] J.D.J. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J.M. Walhout, M.E. Cusick, F.P. Roth, M. Vidal, Evidence for dynamically organized modularity in the yeast protein–protein interaction network, Nature 430 (2004) 88–93.
[7] H.B. Fraser, Modularity and evolutionary constraint on proteins, Nat. Genet. 37 (2005) 351–352.
[8] H.B. Fraser, Coevolution, modularity and human disease, Curr. Opin. Genet. Dev. 16 (2006) 637–644.
[9] B. Kahali, S. Ahmad, T.C. Ghosh, Exploring the evolutionary rate differences of party hub and date hub proteins in *Saccharomyces cerevisiae* protein–protein interaction network, Gene 429 (2009) 18–22.
[10] S. Podder, P. Mukhopadhyay, T.C. Ghosh, Multifunctionality dominantly determines the rate of human housekeeping and tissue specific interacting protein evolution, Gene 439 (2009) 11–16.
[11] A.M. Monzon, D.J. Zea, M.S. Fornasari, T.E. Saldano, S. Fernandez-Alberti, S.C.E. Tosatto, G. Parisi, Conformational diversity analysis reveals three functional mechanisms in proteins, PLoS Comput. Biol. 13 (2017).
[12] L.C. James, D.S. Tawfik, Conformational diversity and protein evolution–a 60-year-old hypothesis revisited, Trends Biochem. Sci. 28 (2003) 361–368.
[13] E. Eisenberg, E.Y. Levanon, Human housekeeping genes are compact, Trends Genet. 19 (2003) 362–365.
[14] C.D. Eller, M. Regelson, B. Merriman, S. Nelson, S. Horvath, Y. Marahrens, Repetitive sequence environment distinguishes housekeeping genes, Gene 390 (2007) 153–165.
[15] L.Q. Zhang, W.H. Li, Mammalian housekeeping genes evolve more slowly than tissue-specific genes, Mol. Biol. Evol. 21 (2004) 236–239.
[16] A. Bossi, B. Lehner, Tissue specificity and the human protein interaction network, Mol. Syst. Biol. 5 (2009).
[17] O. Souiai, E. Becker, C. Prieto, A. Benkahla, J. De Las Rivas, C. Brun, Functional integrative levels in the human interactome recapitulate organ organization, PLoS One 6 (2011).
[18] X.F. Zhang, L. Ou-Yang, D.Q. Dai, Wu MY, Y. Zhu, H. Yan, Comparative analysis of housekeeping and tissue-specific driver nodes in human protein interaction networks, BMC Bioinforma. 17 (2016).
[19] K. Biswas, S. Chakraborty, S. Podder, T.C. Ghosh, Insights into the dN/dS ratio heterogeneity between brain specific genes and widely expressed genes in species of different complexity, Genomics 108 (2016) 11–17.
[20] L.D. Hurst, N.G.C. Smith, Do essential genes evolve slowly? Curr. Biol. 9 (1999) 747–750.
[21] R. Nielsen, C. Bustamante, A.G. Clark, S. Glanowski, T.B. Sackton, M.J. Hubisz, A. Fledel-Alon, D.M. Tanenbaum, D. Civello, T.J. White, et al., A scan for positively selected genes in the genomes of humans and chimpanzees, PLoS Biol. 3 (2005) 976–985.
[22] M. Kiran, H.A. Nagarajaram, Interaction and localization diversities of global and local hubs in human protein-protein interaction networks, Mol. BioSyst. 12 (2016) 2875–2882.
[23] Z. Itzhaki, E. Akiva, Y. Altuvia, H. Margalit, Evolutionary conservation of domain–domain interactions, Genome Biol. 7 (2006).
[24] I. Yanai, H. Benjamin, M. Shmoish, V. Chalifa-Caspi, M. Shklar, R. Ophir, A. Bar-Even, S. Horn-Saban, M. Safran, E. Domany, Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification, Bioinformatics 21 (2005) 650–659.
[25] A. Chatr-aryamontri, B.J. Breitkreutz, R. Oughtred, L. Boucher, S. Heinicke, D.C. Chen, C. Stark, A. Breitkreutz, N. Kolas, L. O'Donnell, et al., The BioGRID interaction database: 2015 update, Nucleic Acids Res. 43 (2015) D470–D478.
[26] P.M. Kim, A. Sboner, Y. Xia, M. Gerstein, The role of disorder in interaction networks: a structural analysis, Mol. Syst. Biol. 4 (2008).
[27] P.M. Kim, L.J. Lu, Y. Xia, M.B. Gerstein, Relating three-dimensional structures to protein networks provides evolutionary insights, Science 314 (2006) 1938–1941.
[28] A. Yates, W. Akanni, M.R. Amode, D. Barrell, K. Billis, D. Carvalho-Silva,

C. Cummins, P. Clapham, S. Fitzgerald, L. Gil, et al., Ensembl 2016, Nucleic Acids Res. 44 (2016) D710–D716.
[29] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (2005) 15–20.
[30] M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, et al., Tissue-based map of the human proteome, Science 347 (2015).
[31] A.M. Monzon, C.O. Rohr, M.S. Fornasari, G. Parisi, CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state, Database-J. Biol. Databases Curation Jan. 1 2016 (2016), http://dx.doi.org/10.1093/database/baw038 (baw038).
[32] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[33] O. Carugo, S. Pongor, A normalized root-mean-spuare distance for comparing protein three-dimensional structures, Protein Sci. 10 (2001) 1470–1473.
[34] D. Acharya, D. Mukherjee, S. Podder, T.C. Ghosh, Investigating different duplication pattern of essential genes in mouse and human, PLoS One 10 (2015) e0120784.
[35] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (2005) 3433–3434.
[36] Z. Dosztanyi, M. Sandor, P. Tompa, I. Simon, Prediction of protein disorder at the domain level, Curr. Protein Pept. Sci. 8 (2007) 161–171.
[37] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins, J. Mol. Biol. 347 (2005) 827–839.
[38] S. Light, R. Sagit, D. Ekman, A. Elofsson, Long indels are disordered: a study of disorder and indels in homologous eukaryotic proteins, Biochim. Biophys. Acta Proteins Proteomics 2013 (1834) 890–897.
[39] N.H. Nie, D.H. Bent, C.H. Hull, SPSS: Statistical Package for the Social Sciences, McGraw-Hill, New York, 1970.
[40] I.K. Jordan, L. Mariño-Ramírez, Y.I. Wolf, E.V. Koonin, Conservation and coevolution in the scale-free human gene coexpression network, Mol. Biol. Evol. 21 (2004) 2058–2070.
[41] D.A. Drummond, J.D. Bloom, C. Adami, C.O. Wilke, F.H. Arnold, Why highly expressed proteins evolve slowly, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 14338–14343.
[42] D.J. Zea, A.M. Monzon, M.S. Fornasari, C. Marino-Buslje, G. Parisi, Protein conformational diversity correlates with evolutionary rate, Mol. Biol. Evol. 30 (2013) 1500–1503.
[43] Gu ZL, D. Nicolae, Lu HHS, W.H. Li, Rapid divergence in expression between duplicate genes inferred from microarray data, Trends Genet. 18 (2002) 609–613.
[44] X. Gu, Functional divergence in protein (family) sequence evolution, Genetica 118 (2003) 133–141.
[45] R.A. Studer, M. Robinson-Rechavi, Large-scale analysis of orthologs and paralogs under covarion-like and constant-but-different models of amino acid evolution, Mol. Biol. Evol. 27 (2010) 2618–2627.
[46] P. Zhang, Gu ZL, W.H. Li, Different evolutionary patterns between young duplicate genes in the human genome, Genome Biol. 4 (2003).
[47] N. Lopez-Bigas, S. De, S.A. Teichmann, Functional protein divergence in the evolution of *Homo sapiens*, Genome Biol. 9 (2008).
[48] I.K. Jordan, Y.I. Wolf, E.V. Koonin, Duplicated genes evolve slower than singletons despite the initial rate increase, BMC Evol. Biol. 4 (2004).
[49] D. Acharya, T.C. Ghosh, Global analysis of human duplicated genes reveals the relative importance of whole-genome duplicates originated in the early vertebrate evolution, BMC Genomics 17 (2016) 1–14.
[50] T. Makino, T. Gojobori, The evolutionary rate of a protein is influenced by features of the interacting partners, Mol. Biol. Evol. 23 (2006) 784–789.
[51] S. Podder, T.C. Ghosh, Exploring the differences in evolutionary rates between monogenic and polygenic disease genes in human, Mol. Biol. Evol. 27 (2010) 934–941.
[52] H. Liang, W.-H. Li, MicroRNA regulation of human protein–protein interaction network, RNA-Publ. RNA Soc. 13 (2007) 1402–1408.
[53] C. Cheng, N. Bhardwaj, M. Gerstein, The relationship between the evolution of microRNA targets and the length of their UTRs, BMC Genomics 10 (2009).
[54] M. Kiran, H.A. Nagarajaram, Global versus local hubs in human protein–protein interaction network, J. Proteome Res. 12 (12) (2013) 5436–5446.
[55] A. Bateman, E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S.R. Eddy, S. Griffiths-Jones, K.L. Howe, M. Marshall, E.L. Sonnhammer, The Pfam protein families database, Nucleic Acids Res. 30 (1) (2002) 276–280.