

# Supplementary Materials

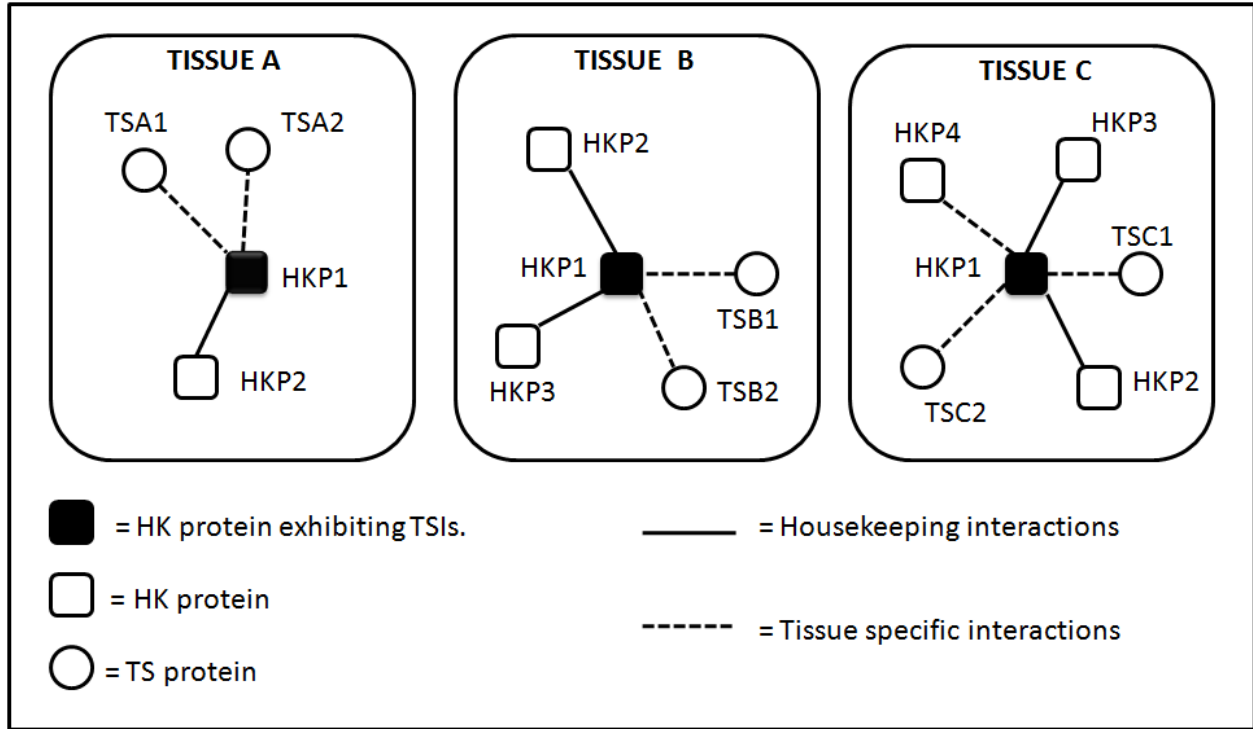
## **Structural Disorder: A tool for housekeeping proteins performing tissue-specific interactions**

*Sanghita Banerjee, Rajat K. De*

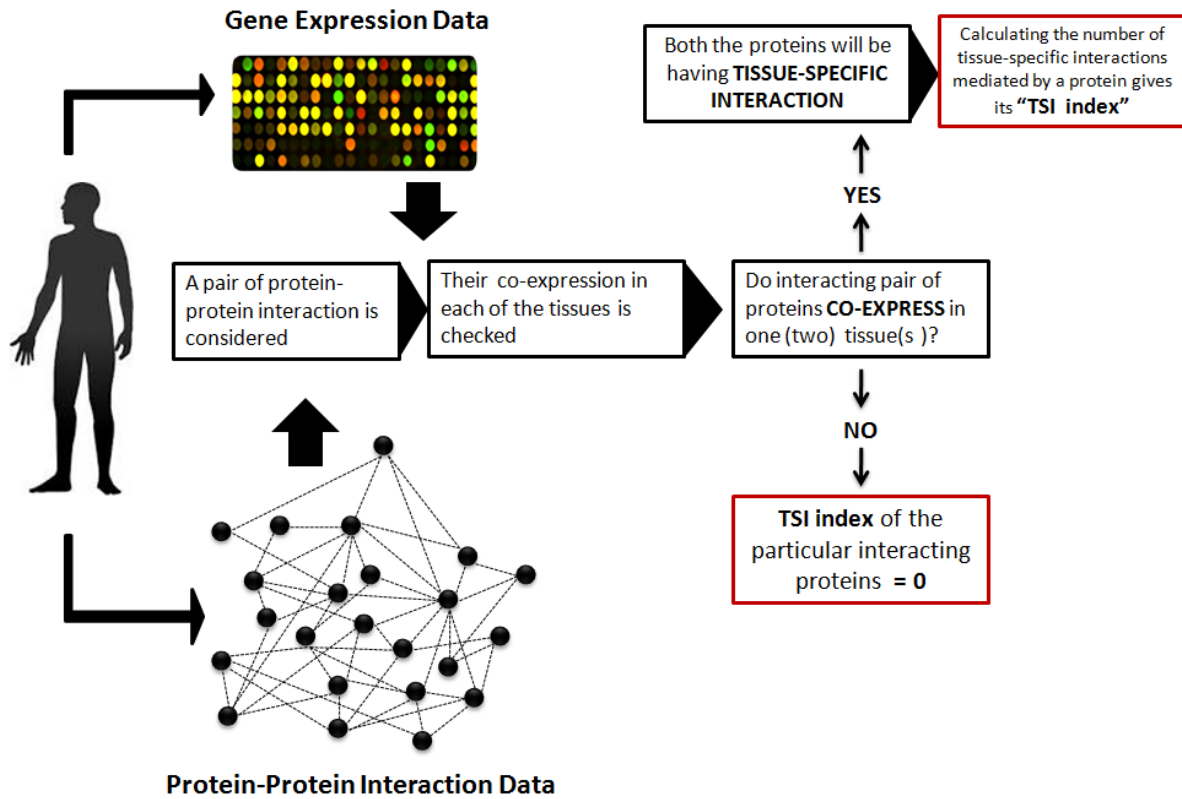
Machine Intelligence Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata  
700108, India

SB: (banerjee.sanghita@gmail.com)

RKD: (rajat@isical.ac.in)



**Figure S1: Tissue-specific interactions of housekeeping proteins.** Tissues A, B and C represent 3 unique tissues. The schematic diagram showing the interaction network of a particular housekeeping protein (HKP1) with both of its tissue-specific interacting proteins (TSA1 and TSA2 in Tissue A, TSB1 and TSB2 in Tissue B and TSC1 and TSC2 in Tissue C) and housekeeping proteins (HKP2 in tissue A, HKP2 and HKP3 in tissue B, and HKP2, HKP3 and HKP4 in tissue C). Thus the figure depicts that the interaction between HKP1-TSA1 and HKP1-TSA2 are tissue “A” specific interactions, interactions between HKP1-TSB2 and HKP1-TSB3 are tissue “B” specific interactions, while interactions between HKP1-TSC1 and HKP1-TSC2 are tissue “C” specific interactions. In tissue C, the interaction between HKP1-HKP4 apparently seems to be tissue “C” specific according to our diagram, but being housekeeping proteins, HKP4 and HKP1 will get expressed in other several tissues, and may interact with each other in other tissues also. So the probability of the interaction, i.e., HKP1-HKP4 to be a tissue-specific interaction is very low. On the other hand, the expression of TS proteins (TSA1, TSA2, TSB2, TSB3, TSC1 and TSC2) is selective to one(or two) tissue(s), hence tissue-specific interactions between a HK and TS proteins are more meaningful.



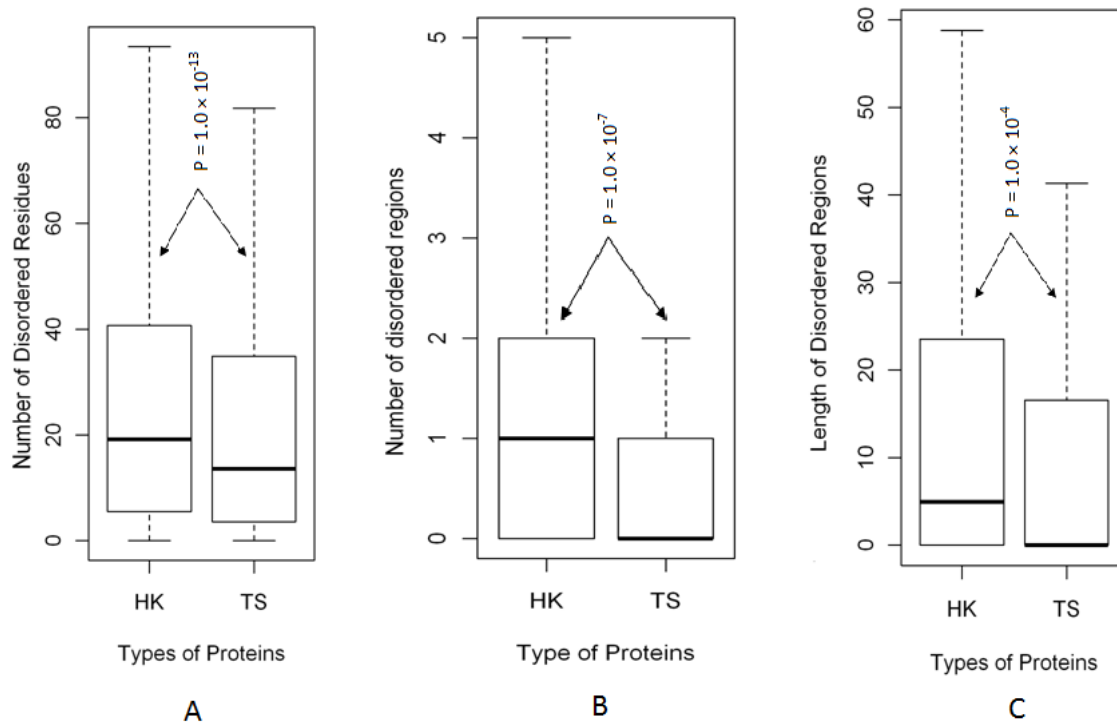
**Figure S2:** Steps showing the procedure of identification of tissue-specific interactions (TSIs) by integrating both gene expression and protein-protein interaction datasets of human.

## EXTENDED RESULTS

### Extended Results using Human Protein Atlas (HPA) dataset

#### 1. Structural disorder in housekeeping proteins (Corresponding to Section 3.1)

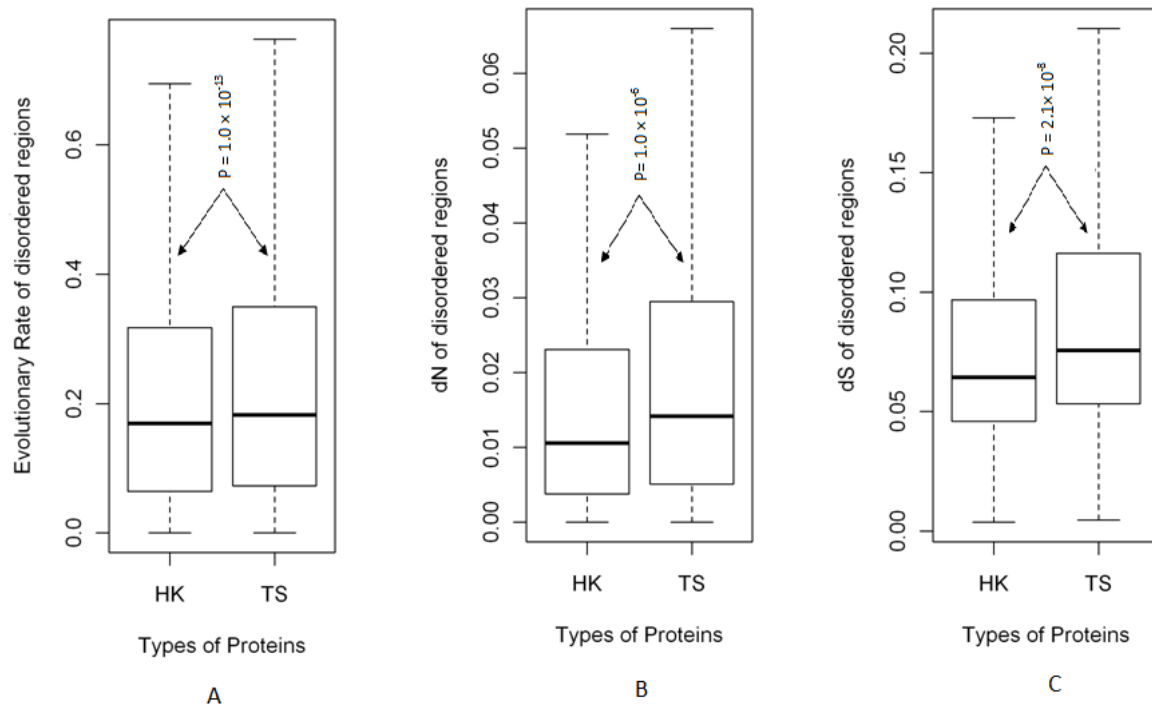
Housekeeping proteins exhibit a higher enrichment in structural disorder (measured using the parameters like i. number of disordered residues, ii. number of disordered regions, and iii. length of disordered regions) compared to those of tissue-specific ones (Figure S3). It indicates that the trend of HK proteins being structurally more disordered than TS proteins is independent of the expression datasets used in the study.



**Figure S3:** Difference in the A. number of disordered residues, B. number of disordered regions, and C. length of disordered regions between HK and TS proteins.

## 2. Evolutionary conservation of disordered regions within housekeeping and tissue specific proteins (Corresponding to Section 3.2)

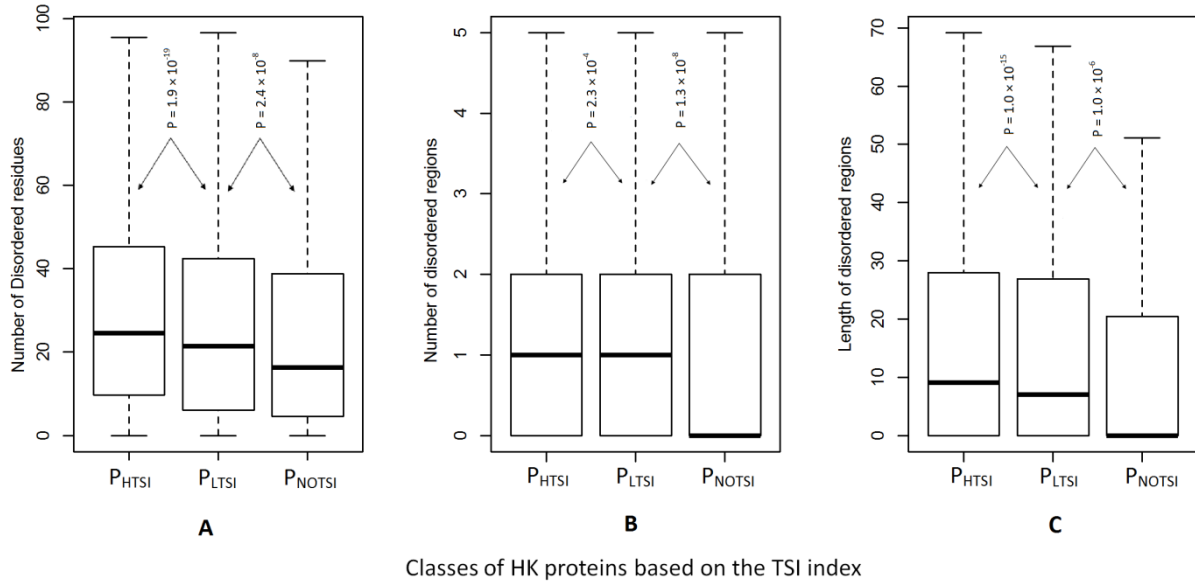
The stretches of disordered regions in HK proteins evolve slowly compared to those disordered regions within TS proteins. The rate of non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions within disordered regions is relatively less within the stretches of disordered regions residing within HK proteins, in contrast to TS proteins (Figure S4).



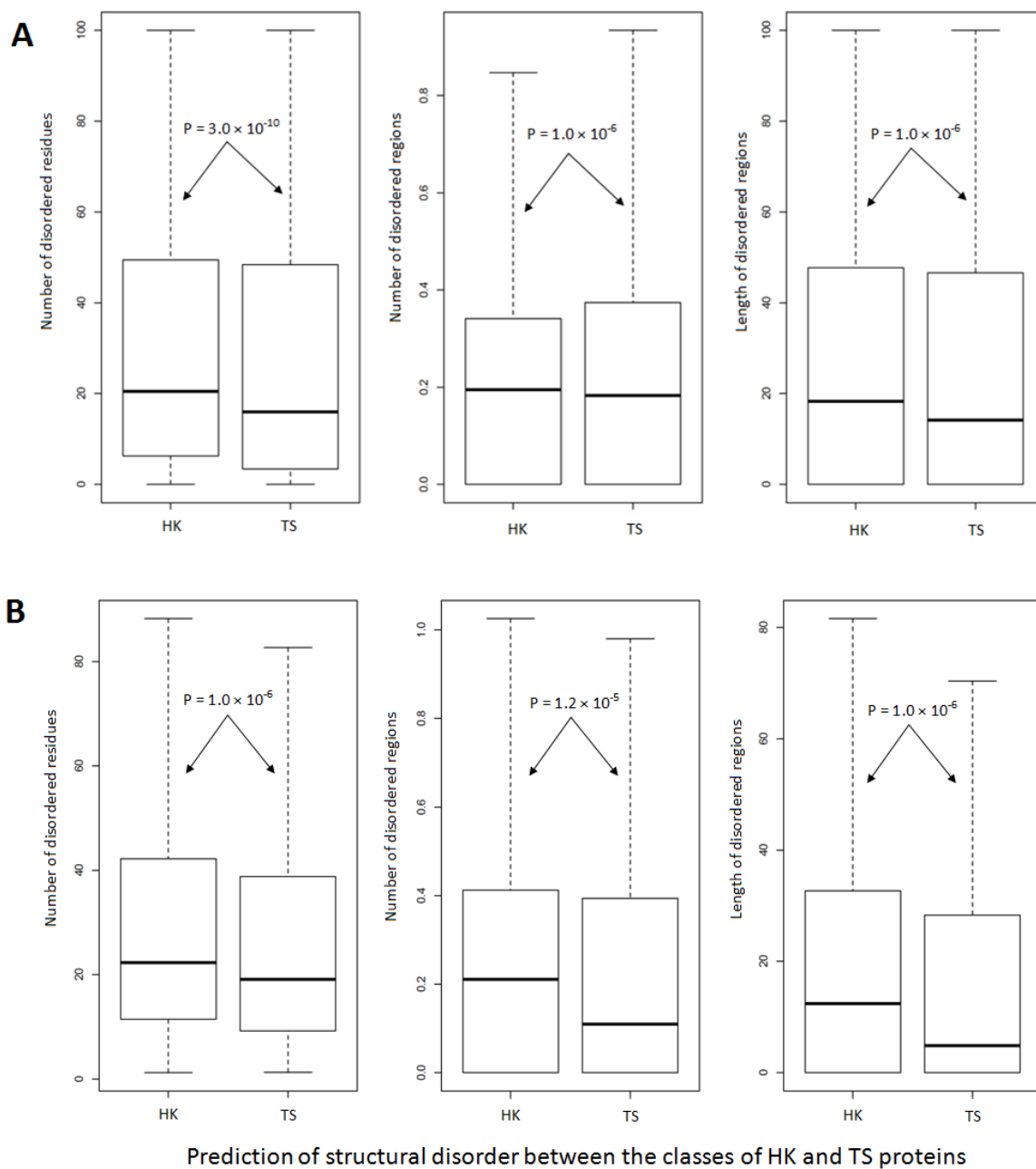
**Figure S4:** Differences in the distribution of A. the evolutionary rates ( $d_N/d_S$  ratio), B. the rate of non-synonymous ( $d_N$ ), and C. synonymous ( $d_S$ ) substitutions between housekeeping (HK) and tissue-specific (TS) proteins.

### 3. Influence of structural disorder in housekeeping proteins mediating tissue specific interactions (Corresponding to Section 3.3)

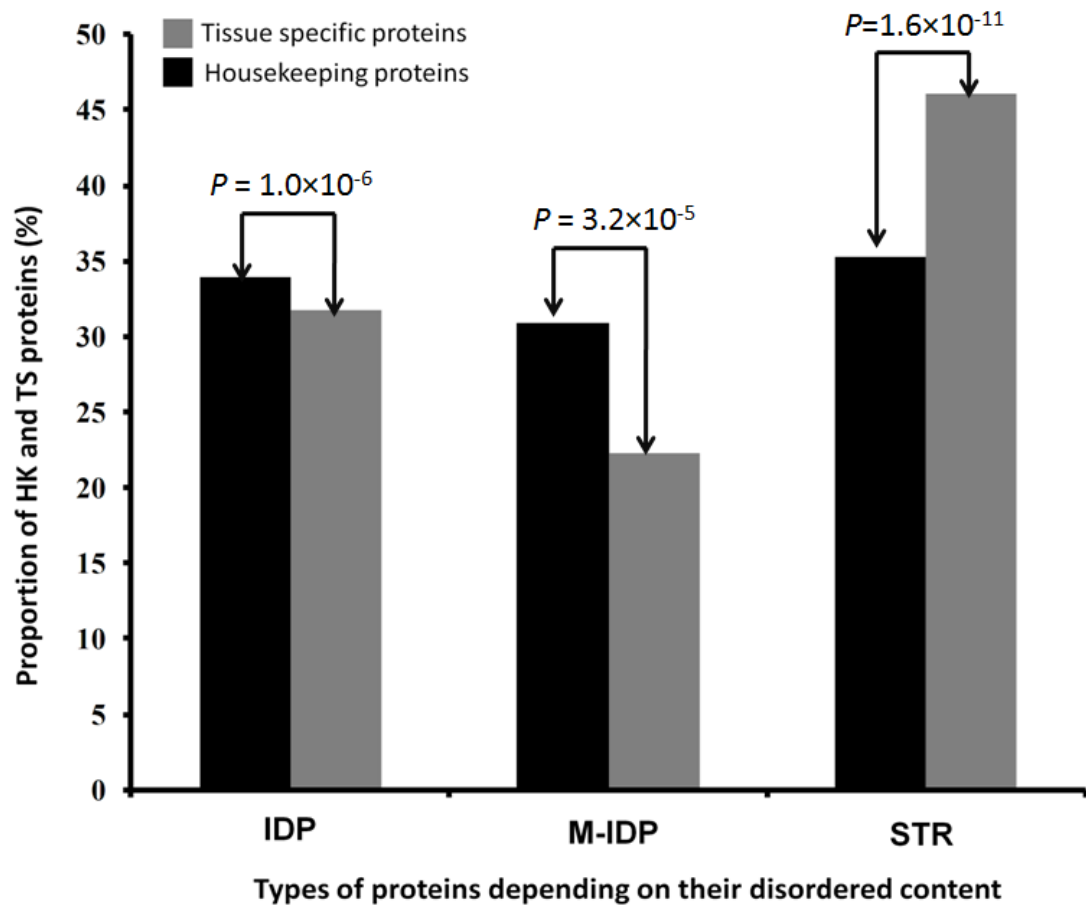
In the case of HPA dataset, we have categorized the HK proteins based on the threshold of average TSI index ( $\approx 5$ ) as: i)  $P_{HTSI}$  (TSI index  $\leq 5$ ), ii)  $P_{LTSI}$  (TSI index  $> 5$ ), and iii)  $P_{NOTSI}$  (TSI index = 0). The set of  $P_{HTSI}$  exhibits a higher enrichment in structural disorder in comparison to the other sets of HK proteins (i.e.,  $P_{LTSI}$  and  $P_{NOTSI}$ ).



**Figure S5:** Differences in the distributions of the parameters – A. number of disordered residues, B. number of disordered regions, and C. length of disordered regions measured within the groups of HK proteins having varying degrees of TSI index using Human Protein Atlas (HPA) dataset.

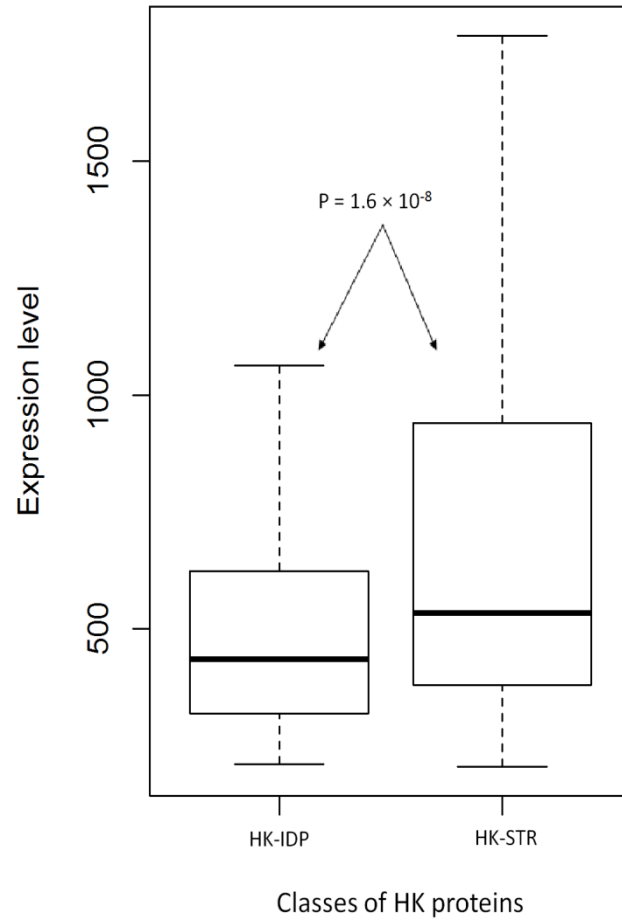


**Figure S6:** Prediction of structural disorder (i.e., number of disordered residues, number of disordered regions and length of the disordered regions) between the sets of HK and TS proteins using different disorder prediction tools like A) ESpritz and B) PONDR-FIT.

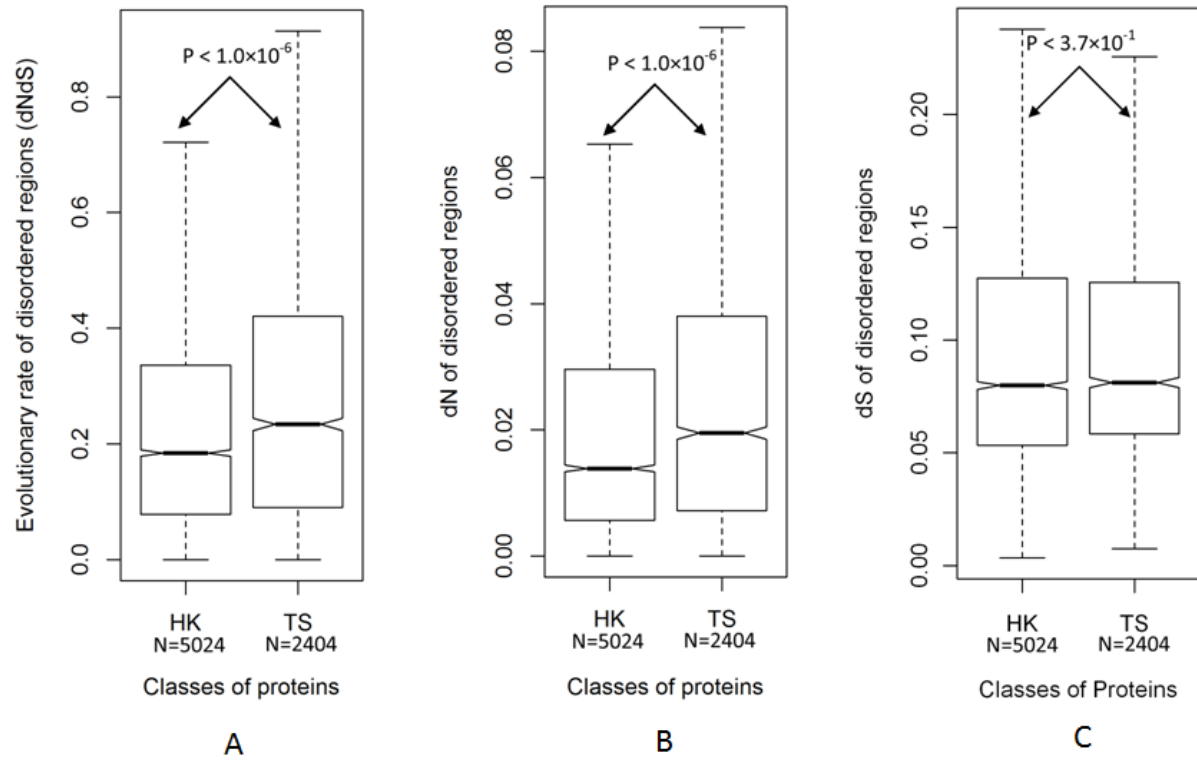


**Figure S7:** Bar plot showing the difference in the proportion (in percentage) of the highly disordered (IDPs), moderately disordered (M-IDPs) and well-structured proteins (STRs) within the groups of housekeeping (in black) and tissue-specific (in grey) proteins.

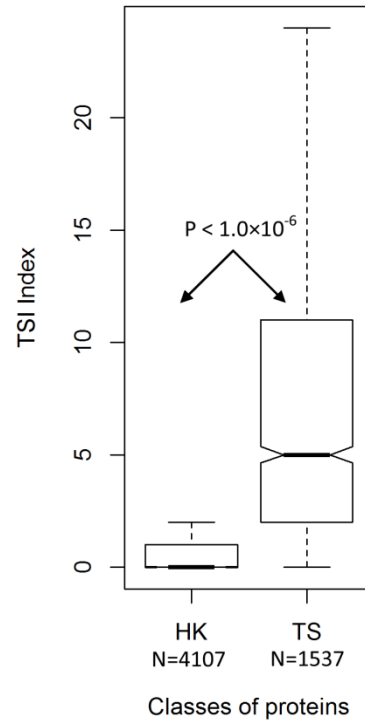




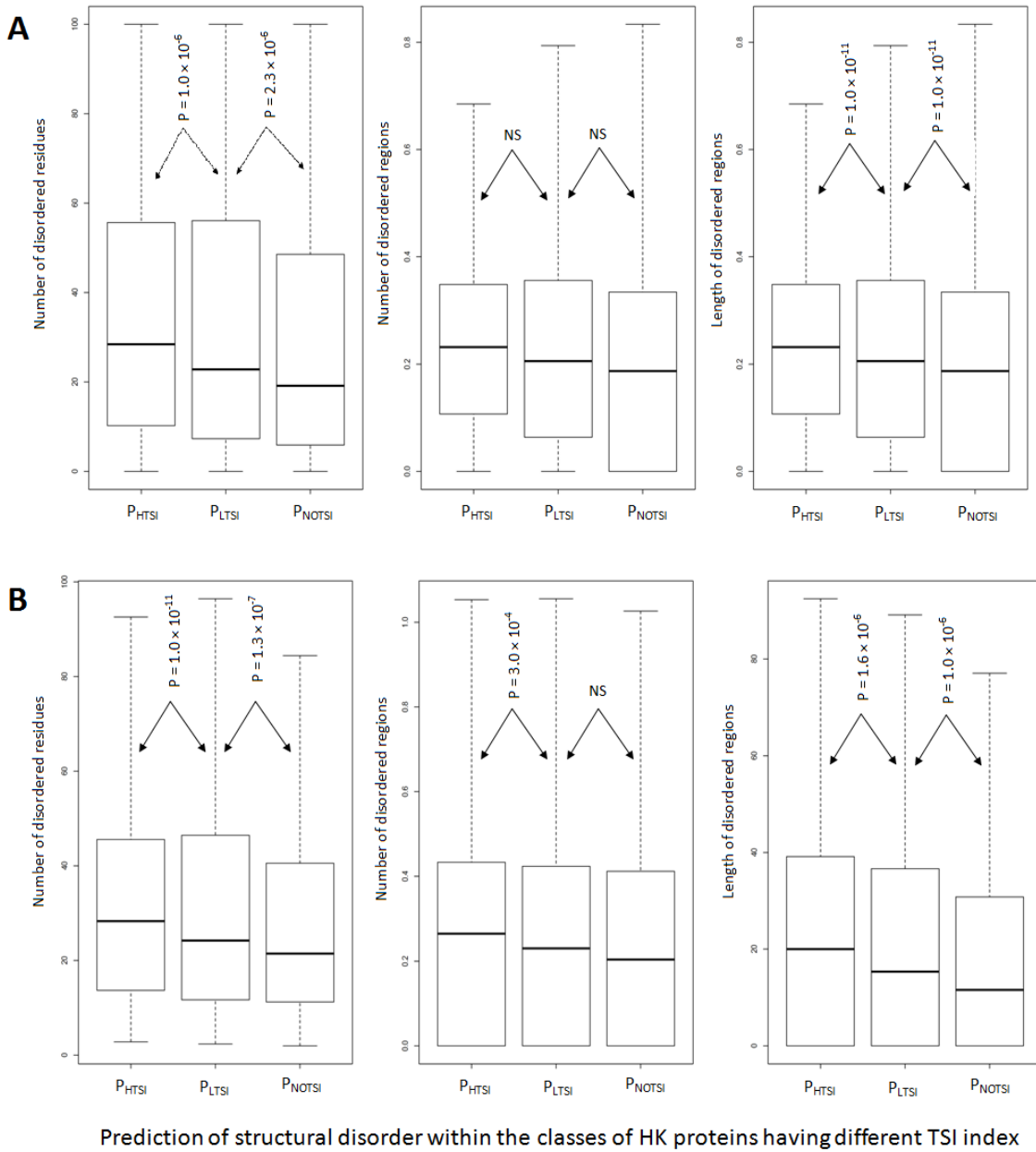
**Figure S8:** Boxplots showing the difference in the distribution of expression level between disordered (IDPs) and well-structured (STRs) HK proteins.



**Figure S9:** Boxplots showing the differences in the distributions of (A) evolutionary rate ( $d_N/d_S$ ), (B) rate of non-synonymous substitution ( $d_N$ ), and (C) rate of synonymous substitution ( $d_S$ ) of intrinsically disordered regions between the classes of housekeeping (HK) and tissue-specific (TS) proteins.



**Figure S10:** Boxplots showing the difference in the distribution of tissue-specific interaction index (TSI index) within the classes of housekeeping (HK) and tissue specific (TS) proteins.



**Figure S11:** Prediction of structural disorder (i.e., number of disordered residues, number of disordered regions and length of the disordered regions) between the groups of HK proteins having varying degrees of TSI index ( $P_{HTSI}$ ,  $P_{LTSI}$  and  $P_{NOTSI}$ ) using different disorder prediction tools like A) ESpritz and B) PONDR-FIT.

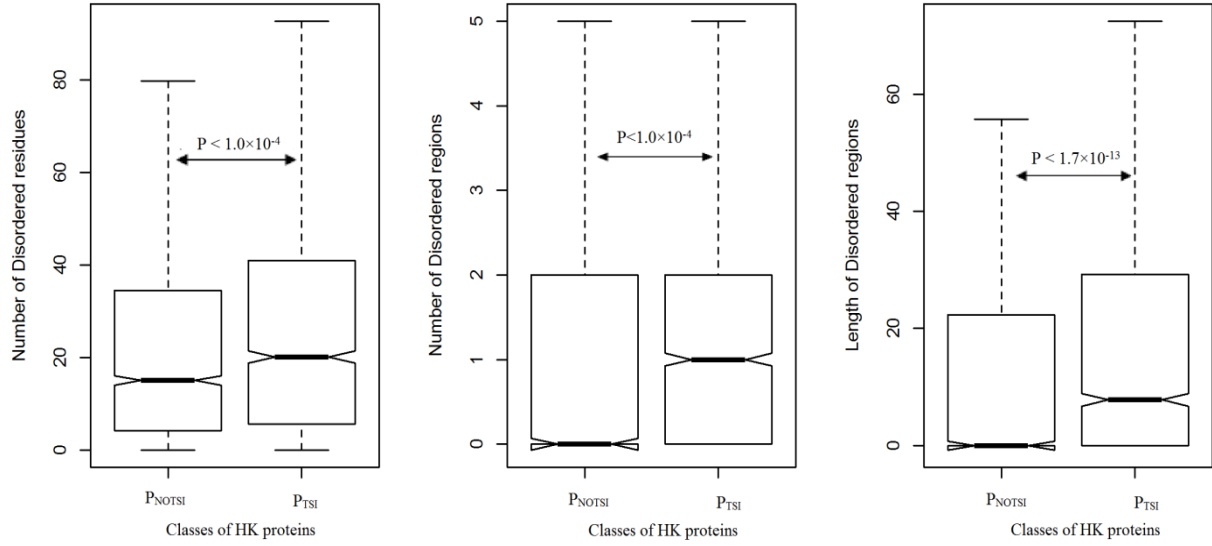
**Table S1: Categorization of housekeeping proteins based on flexible threshold values of TSI index.**

Case No	Group	Range of TSI index	Sample Size (N)	Figure (Showing distribution of structural disorder)
I	P <sub>TSI</sub>	81 to 1	1888	Figure S12
	P <sub>NOTSI</sub>	0	2220	
II	P <sub>HTSI (High)</sub>	81 to 11	90	Figure S13
	P <sub>LTSI (Low)</sub>	10 to 1	1798	
	P <sub>NOTSI</sub>	0	2220	
III	P <sub>HTSI (High)</sub>	10 to 5	133	Figure S14
	P <sub>LTSI (Low)</sub>	4 to 1	778	
	P <sub>NOTSI</sub>	0	2220	
IV	P <sub>HTSI (High)</sub>	21 to 11	47	Figure S15
	P <sub>LTSI (Low)</sub>	10 to 1	911	
	P <sub>NOTSI</sub>	0	2220	
V	P <sub>TSI-RANDOM1</sub>	Grouped Randomly, Not depending on any threshold values of TSI index.	1402	Figure S16
	P <sub>TSI-RANDOM2</sub>		1536	
	P <sub>TSI-RANDOM3</sub>		1170	

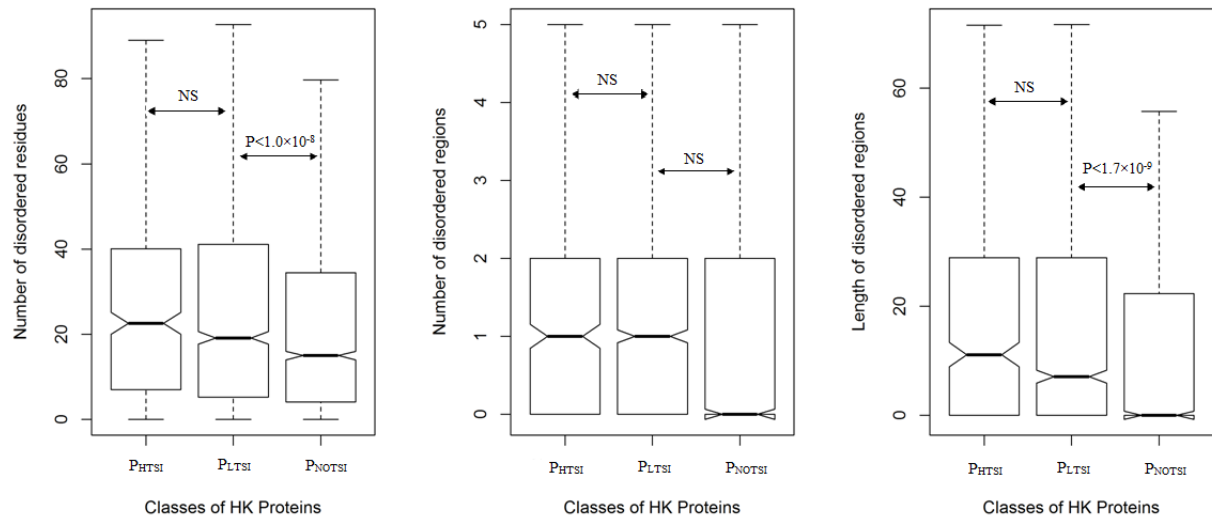
**Explanation:** We have reported five cases (I, II, III, IV and V) that compare the distribution of the measures of structural disorder (i.e., number of disordered residues, number of disordered regions and length of disordered regions) among different sets of HK proteins categorized based on their TSI index values. Each of these sets is classified based on the flexible range of TSI index values mentioned in Table S1. Sample size (N), mentioned in Table S1, indicates the number of HK proteins in each of the categories.

In Cases I (Figure S12) and II (Figure S13), we have classified the entire set of HK proteins based on different TSI threshold values, and measured the differences in the distributions of various features estimating structural disorder among the groups (Figure S12). The results have shown a similar trend with the corresponding results obtained in Section 3.3. However, the

difference in the distribution of structural disorder between  $P_{HTSI}$  and  $P_{LTSI}$  in Case II is not significant due to incomparable sample size (N).

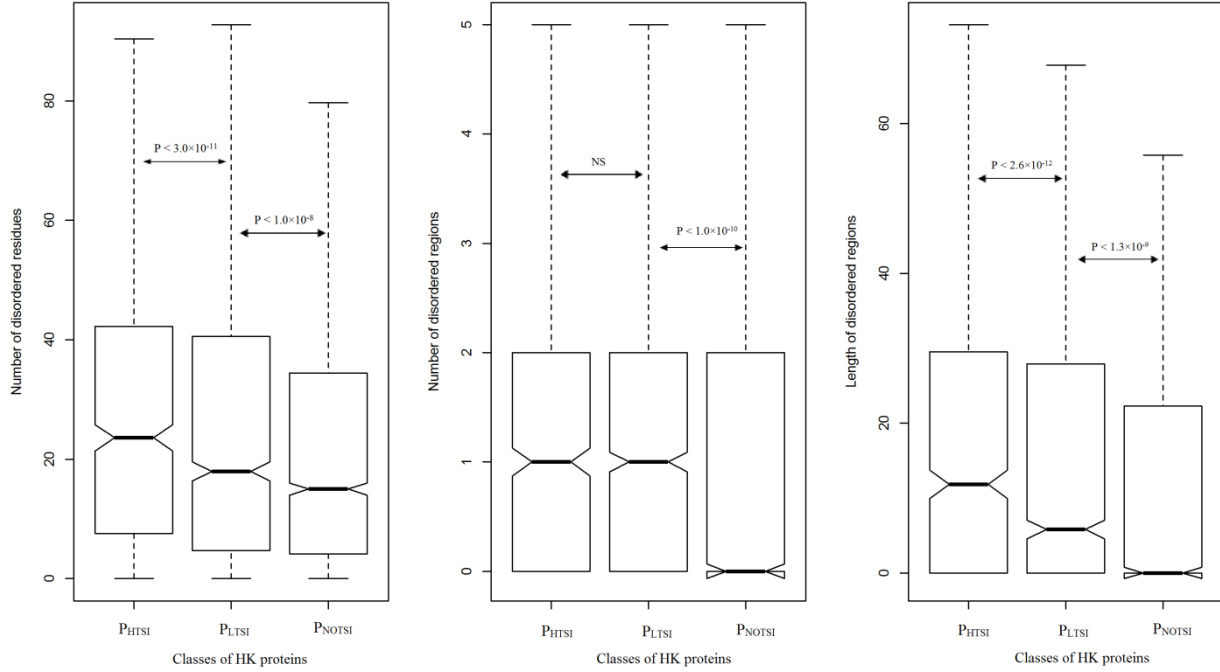


**Figure S12:** Boxplots showing the distribution of structural disorder between HK proteins undergoing TSIs and those not undergoing any TSIs.



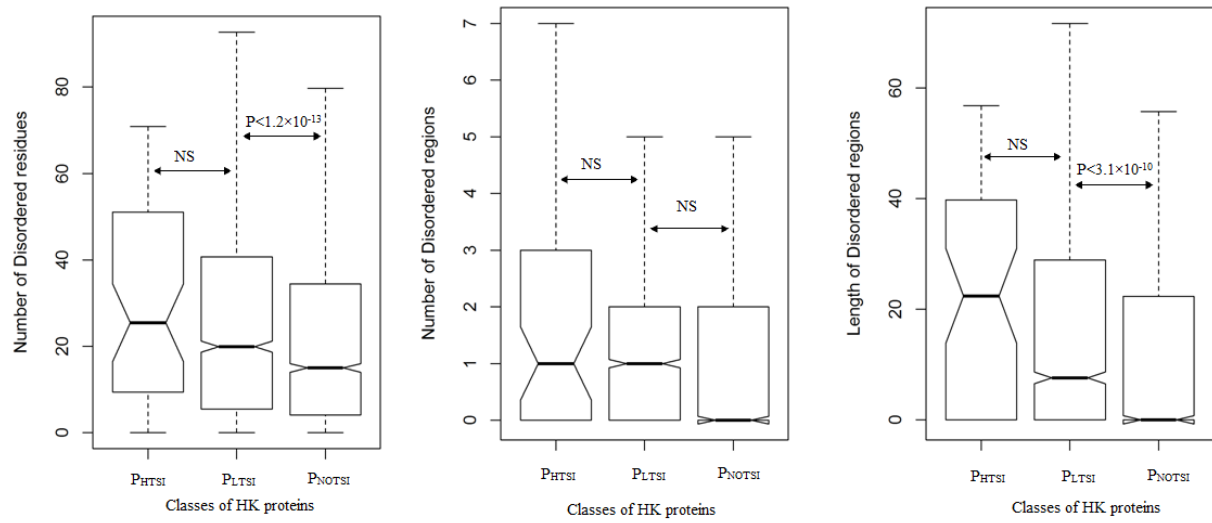
**Figure S13:** Boxplots showing the distribution of structural disorder between HK proteins undergoing high TSIs ( $P_{HTSI}$ ), low TSIs ( $P_{LTSI}$ ), and those not undergoing any TSIs ( $P_{NOTSI}$ ).

In Case III (Figure S14), we have ignored the set of HK proteins having TSI index  $>10$ , as the sample size of HK proteins having TSI index  $>10$  is too small in comparison with the number of proteins having TSI index = 10 to 1. We have further categorized the class of HK proteins into different sets depending on the different range of TSI index and have analyzed the distribution of structural disorder among them.



**Figure S14:** Boxplots showing the distribution of structural disorder between HK proteins undergoing high TSIs (P<sub>HTSI</sub>), low TSIs (P<sub>LTSI</sub>), and those not undergoing any TSIs (P<sub>NOTSI</sub>).

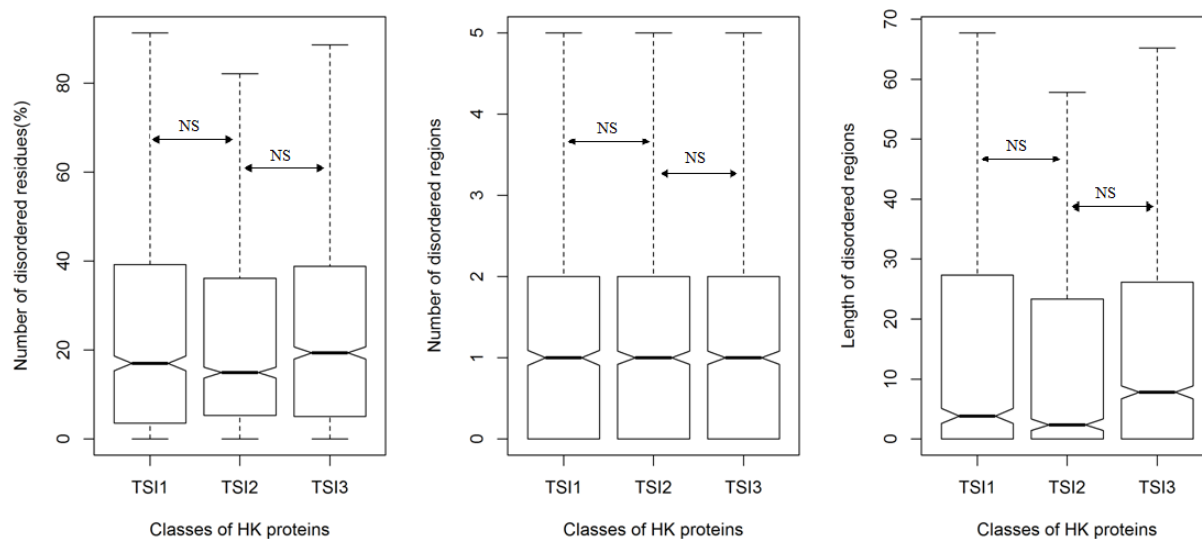
In Case IV (Figure S15), we have ignored the set of HK proteins having TSI index  $> 21$ . Then, we have categorized the remaining HK proteins based on their TSI index values and compared the distributions of three parameters measuring structural disorder. The boxplots exhibit the differences in the distributions of the parameters of structural disorder (except the difference in number of disordered regions between  $P_{HTSI}$  and  $P_{LTSI}$ ). However, some of the differences are not significant (NS), perhaps due to incomparable sample sizes.



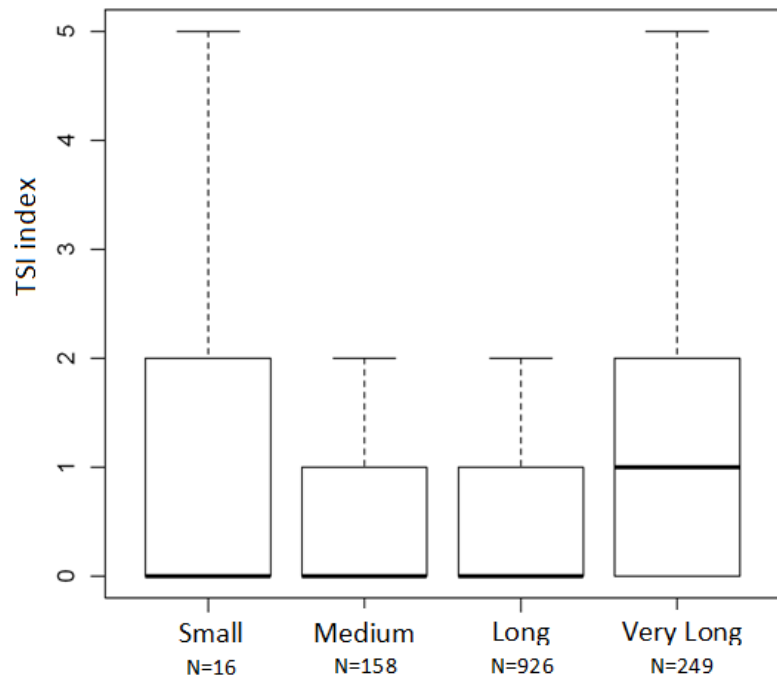
**Figure S15:** Boxplots showing the distribution of structural disorder between HK proteins undergoing high TSIs ( $P_{HTSI}$ ), low TSIs ( $P_{LTSI}$ ), and those not undergoing any TSIs ( $P_{NOTSI}$ ).



In Case V (Figure S16), we have randomly grouped the set of HK proteins into three sets (TSI1, TSI2, and TSI3) of almost equal sample size. We have done this sampling in order to test whether the former grouping is at all meaningful. As expected, the differences in the distributions are not significant among the sets of TSI1, TSI2 and TSI3 (Figure S16), in spite of having comparable sample sizes. Moreover, the distribution does not even reflect any relationship between the extent of structural disorder and the TSI index of HK proteins.

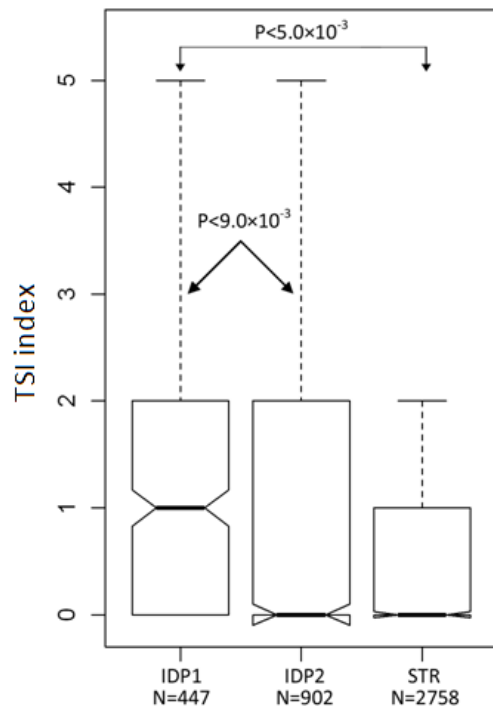


**Figure S16:** Boxplots showing the distribution of structural disorder between different groups of HK proteins (TSI1, TSI2 and TSI3) based on random TSI index values.



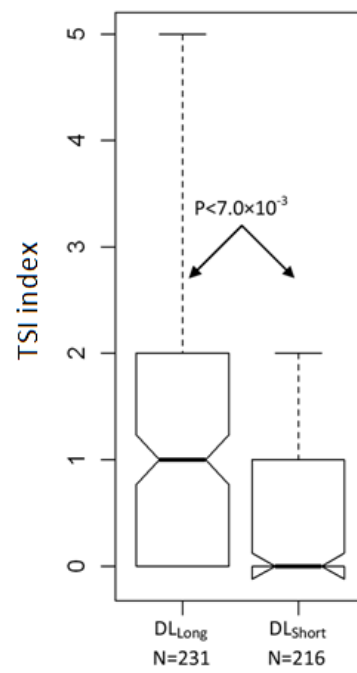
Disordered Proteins depending on the length of disordered regions

A



Classes of proteins

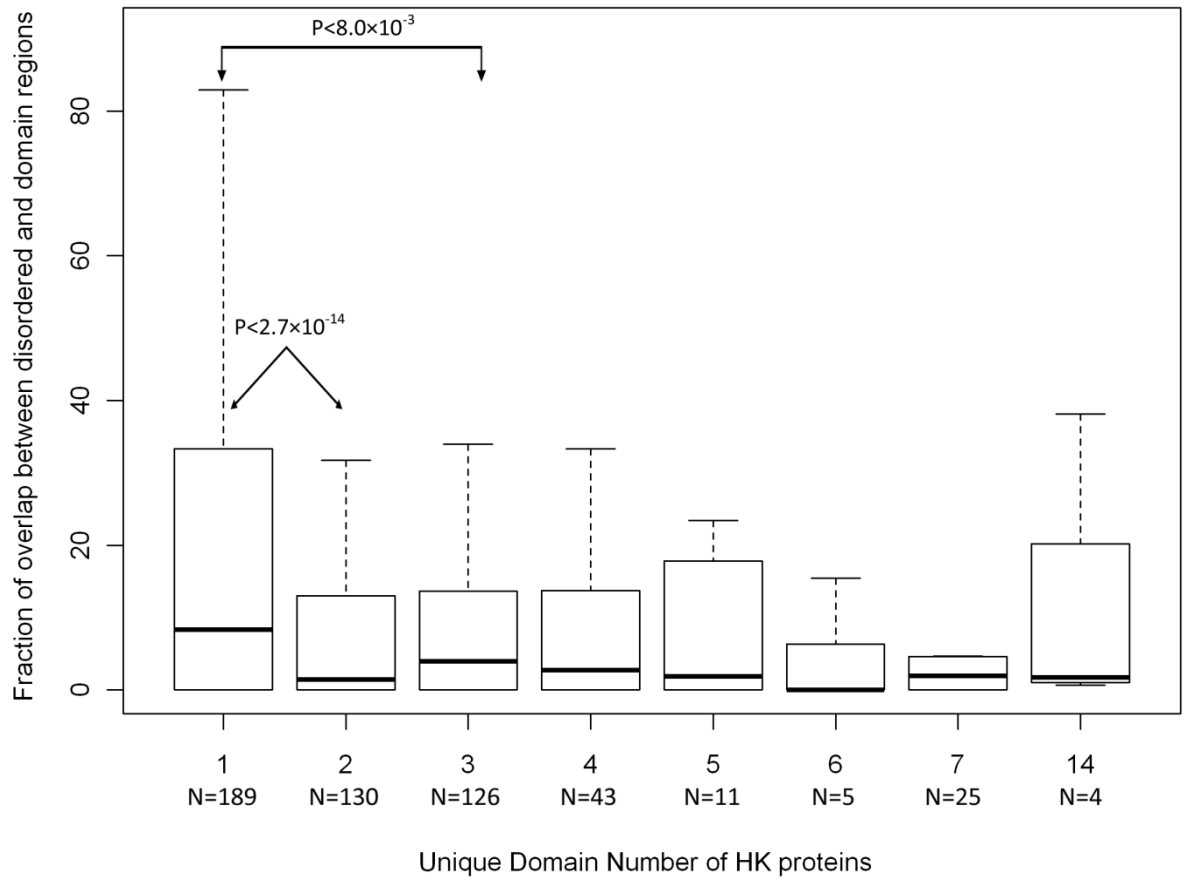
B



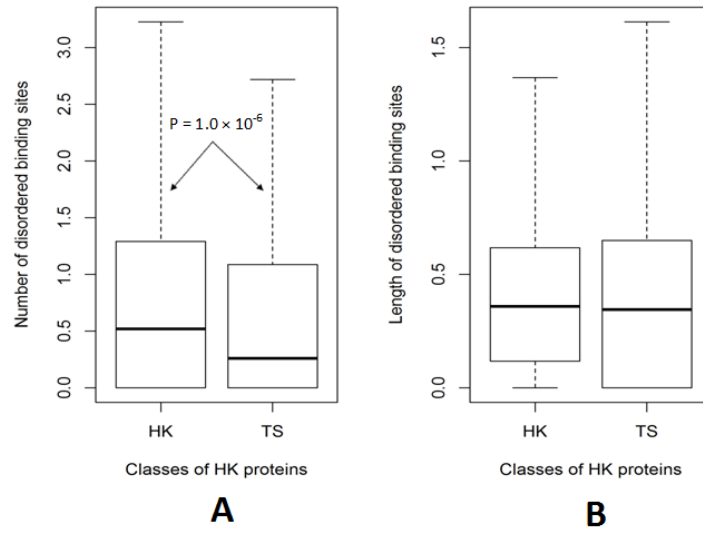
IDPs long and short

C

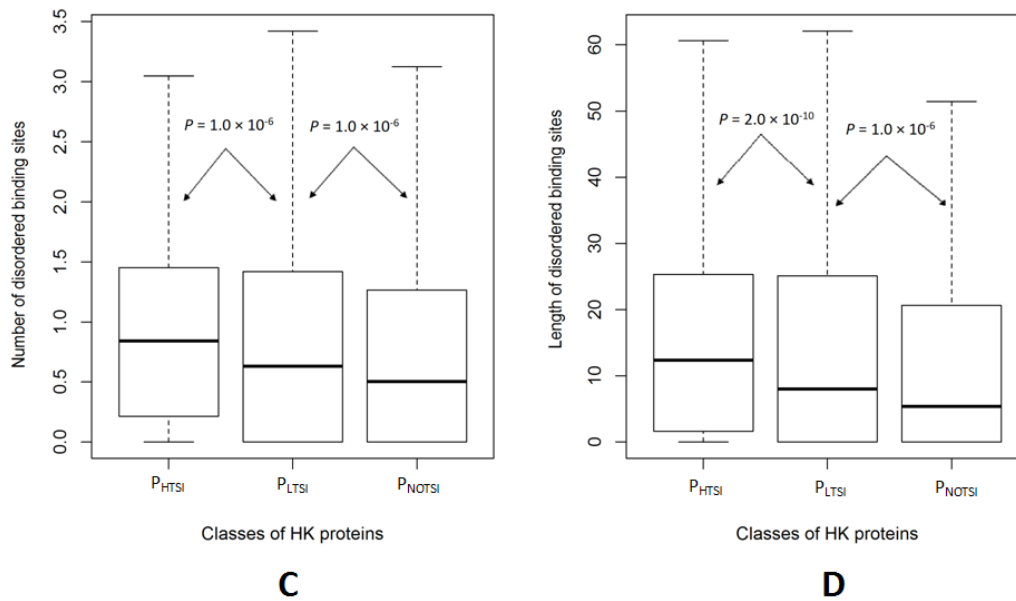
**Figure S17:** Boxplots showing A) the distribution of tissue-specific interaction index (TSI index) among the four groups (Small, Medium, Long, Very Long) of IDPs classified on the basis of the length of their disordered regions, B) Distribution of TSI index among three groups of intrinsically disordered proteins (IDPs) on the basis of the number of disordered regions (i.e., DR count) as IDP1 (DR count = 1), IDP2 (DR count > 1), and STR (DR count = 0), C) Distribution of TSI index within the two groups: DL<sub>Long</sub> (DL > 120) and DL<sub>Short</sub> (DL ≤ 120) categorized from the entire set of IDP1, based on the average length (≈120 residues) of the disordered region.



**Figure S18:** Boxplot showing the distribution of the fraction of disordered regions that overlaps with the adjacent protein domains among the groups of housekeeping (HK) proteins categorized on the basis of their unique domain number (ranging from 1 to 14). N denotes the number of proteins in each group.

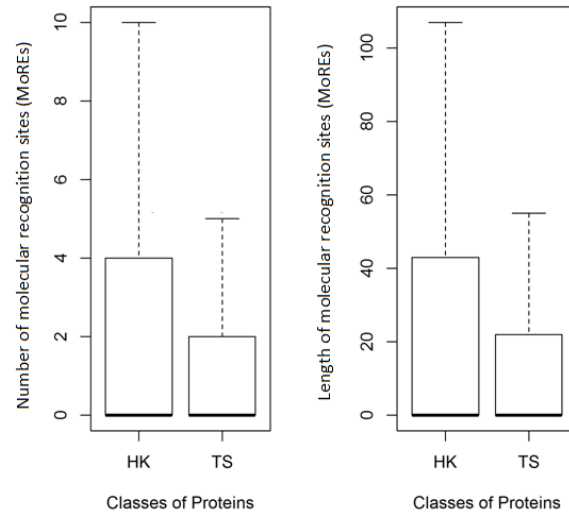


Prediction of disordered binding regions between the classes of HK and TS proteins



Prediction of disordered binding regions within the classes of HK proteins having high TSIs, low TSIs and no TSI.

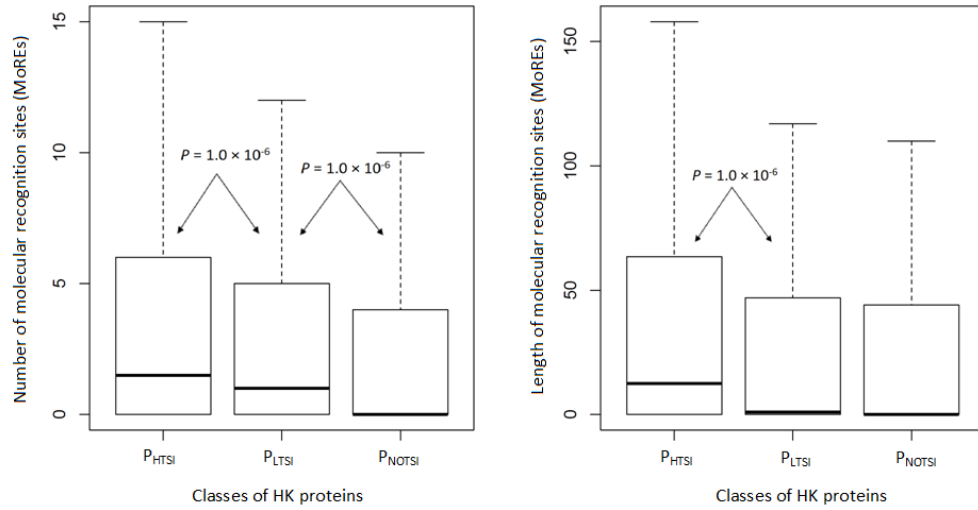
**Figure S19: Prediction of disordered binding regions using ANCHOR method.** Boxplots showing the differences in the distribution of the number and length of the disordered binding sites present between the classes of housekeeping (HK) and tissue specific (TS) proteins. NS stands for non-significant.



**A**

**B**

Prediction of molecular recognition elements between the classes of HK and TS proteins

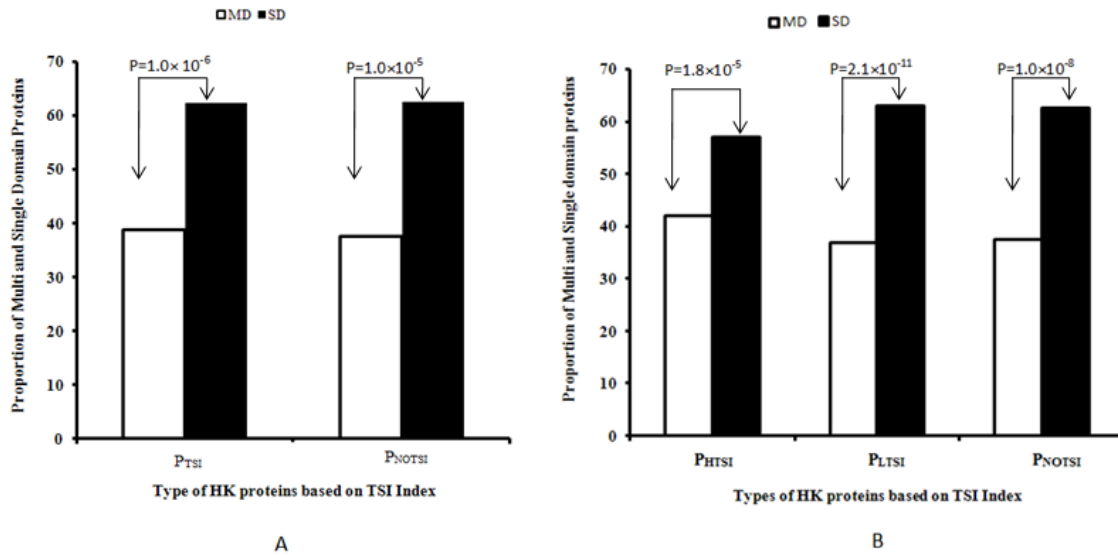


**C**

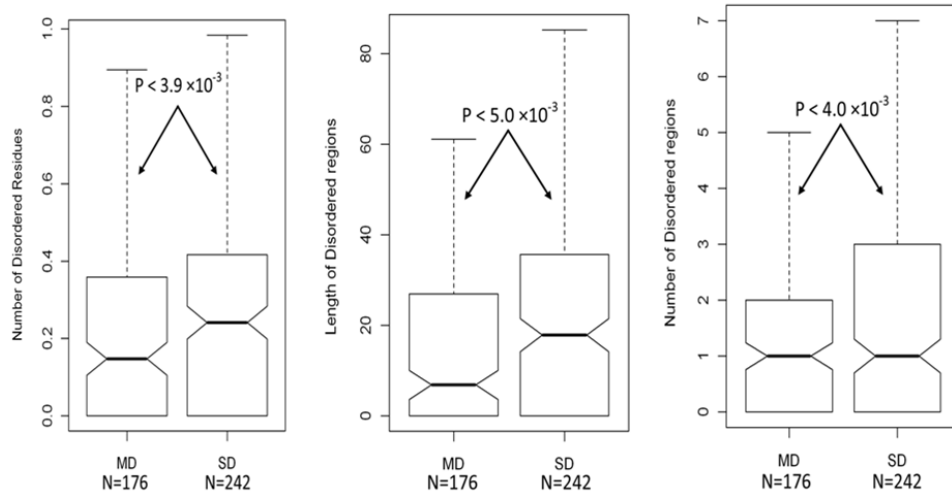
**D**

Prediction of disordered binding regions within the classes of HK proteins having high TSIs, low TSIs and no TSI.

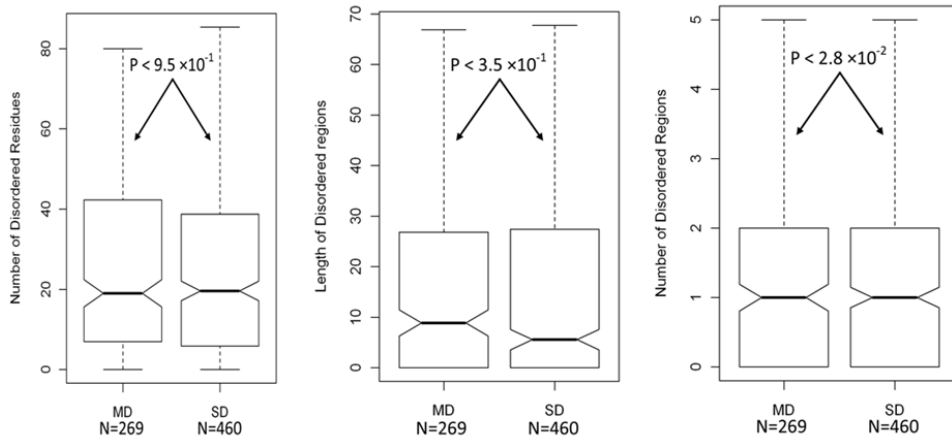
**Figure S20: Prediction of molecular recognition elements (MoREs) using MoRFPred method.** Boxplots showing the differences in the distribution of the number and length of the molecular recognition elements (MoREs) present between the classes of HK proteins undergoing a high number of TSIs (P<sub>HTSI</sub>), low number of TSIs (P<sub>LTSI</sub>) and no TSI (P<sub>NOTSI</sub>).



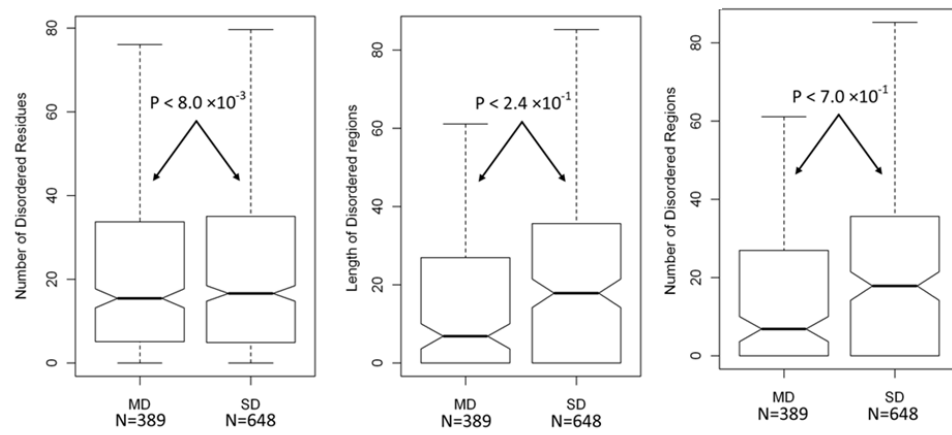
**Figure S21:** Histogram showing A) the proportion of multi-domain (MD) proteins and single domain (SD) proteins within the classes of HK proteins mediating TSIs and those that does not and B) the difference in the proportion of single domain and multi domain proteins within the groups of HK proteins mediating a high number of TSIs ( $P_{HTSI}$ ), low number of TSIs ( $P_{LTSI}$ ), and those not mediating any TSI ( $P_{NOTSI}$ ).



**A. Classes of proteins having High TSIs ( $P_{HTSI}$ )**



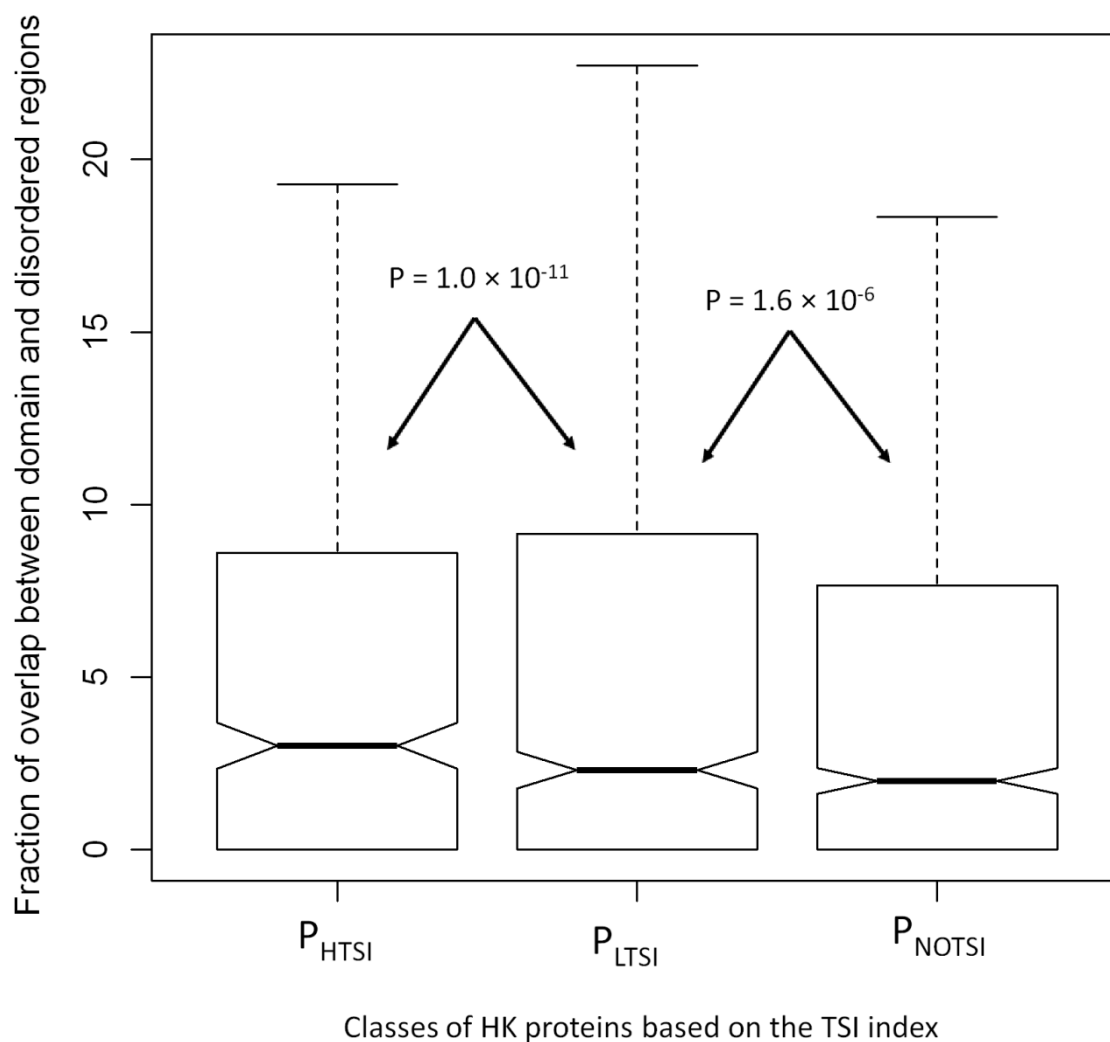
**B. Classes of proteins having Low TSIs ( $P_{LTSI}$ )**



**C. Classes of proteins having No TSIs ( $P_{NOTSI}$ )**



**Figure S22:** Boxplots showing the distribution of three parameters 1) number of disordered residues, 2) number of disordered regions, and 3) length of disordered regions between the groups of multi-domain (MD) proteins and single domain (SD) proteins within the classes of housekeeping (HK) proteins having A) high TSI index ( $P_{\text{HTSI}}$ ), B) Low TSI index ( $P_{\text{LTSI}}$ ) and C) No TSI ( $P_{\text{NOTSI}}$ ).



**Figure S23:** Boxplot showing the distribution of the fraction of disordered regions that overlap with the adjacent protein domains among the groups of housekeeping (HK) proteins having high TSI ( $P_{HTSI}$ ), low TSI index ( $P_{LTSI}$ ) and those not undergoing any TSI ( $P_{NOTSI}$ ).