Chapter 4

Analysis of Tissue-Specific Protein- Protein Interaction Networks: Global versus Local Hubs

Publication arising out of the work presented in this chapter:

4.1. Introduction

Proteins with many interactions are known as hubs, which form the essential components of interaction network (Jeong et al. 2001; Chen & Xu 2005; Hahn & Kern 2005). Among hubs there are some which are found only in few tissues and can be referred to as tissue-specific hubs (TSHs) or local hubs. There are some other hubs which are universally expressed and therefore can be referred to as house-keeping hubs (HKHs) or global hubs. TSHs because of their local presence and high connectivity are expected to be destined for tissue-specific roles as compared with HKHs which are expected to play mostly core cellular functions. These two classes of hubs are, therefore, expected to have different features and different functions owing to differences in their expression breadths (i.e., the number of tissues they are expressed). This chapter gives a detailed comparison of various properties of local and global hubs.

4.2. Materials and Methods

4.2.1. Protein Interaction Data

The union of interactions stored in BioGrid (Stark et al. 2011), DIP (Salwinski et al. 2004), HPRD (Prasad et al. 2009), IntAct (Aranda et al. 2010) and MINT (Ceol et al. 2010) was compiled and the non-redundant human PPI interaction data thus obtained as discussed in Chapter 2 has been used in the present work. After several steps of curation which involved the removal of inferred and predicted information and retention of only those interactions that is validated by at least one experimental technique, 78356 unique undirected interactions could be retrieved for 12142 human proteins.

4.2.2. High confidence dataset

Only those interactions were retrieved which are supported by at least two different research publications. This subset was used as a high confidence dataset in order to reconfirm the results obtained from the other large interaction dataset. This dataset comprises of 14248 pair-wise interactions involving 5535 proteins.
4.2.3. Tissue-specific Networks (TSN)

The gene expression data retrieved from microarray was mapped on to curated Human PPI. Only those interactions were retained where the gene expression information of both the partners were available. TSNs used in this analysis are the same as those used in the analysis reported in the previous chapter. As mentioned in Chapter 2, 70 different networks corresponding to 70 normal tissues were constructed. All the analysis was done on the largest component of each tissue-specific network. Igraph, a publicly available R package was used (Dessau & Pipper 2008) for creating and analyzing networks/graphs (http://cneurocvs.rmki.kfki.hu/igraph/). The average value of the degrees corresponding to the top 20 percentile of highly connected proteins in every tissue was used as the minimum degree characterizing a node as hub. Thus, a protein with degree greater than 16 is characterized as hub in each tissue-specific network.

4.2.4. Gene Expression Abundance and Protein Abundance

Expression abundance (EA) is the mean expression value of a gene expressed in a number of tissue types. If a gene is not expressed in a particular tissue (with signal intensity < 9.85), the respective expression value is not considered for calculation of expression abundance. All the values derived for EA calculation were further rescaled to bring all variables to a common range. All the values derived were first log2 transformed and then normalized to get scores between 0 and 1 where 0 signifies least and 1 as the maximum expression abundance.

The protein abundance data were extracted from PaxDb (Wang et al. 2012), which provides integrated information on absolute protein abundance in ppm for 12 model organism including Homo sapiens. The organism level integrated data was used as abundance of the protein averaged for all tissues. The protein abundance values were first log2 transformed and were further rescaled to lie from 0.0 to 1.0.
4.2.5. Physico-chemical and structural properties of proteins

The number of charged residues (Asp, Glu, His, Lys and Arg) divided by the total length of the protein was calculated as the percentage of charged residues for every interacting protein. SSpro4 (Cheng et al. 2005) was used to predict exposed and buried residues and also the secondary structures (Helix, Strand and Coils) in proteins. Low-complexity regions in proteins were identified using Segmasker (Wootton & Federhen 1996) which is a part of blast+ package. Epestfind (Rogers et al. 1986), an algorithm in the EMBOSS (Rice et al. 2000) was used to identify PEST motifs. A standalone version of Disopred2 (Ward et al. 2004) was used to predict (at a 5% expected rate of false positives) the residues in the disordered regions of proteins. The % disorderness for each protein was calculated by counting the number of residues in the disordered regions and dividing that number with the total length of the protein. Locally installed IUPred (Dosztányi et al. 2005) was also used to make an independent disorder prediction. In this case % disorderness was calculated as the number of residues with IUPred score of greater than 0.5 divided by the number of residues in a protein.

4.2.6. Interacting domain and binding interfaces

HMMPfam (Eddy 1998), a built-in application of InterProScan (Zdobnov & Apweiler 2001) was used to scan Pfam (Finn et al. 2010) domains of all the interacting proteins. Out of all the Pfam domains predicted by HMMPfam for each protein, only the domains involved in interactions either in 3did (Stein et al. 2011), iPfam (Finn et al. 2005) or predicted as high confidence (HC) in DOMINE (Yellaboina et al. 2011) were considered for further analysis. Interacting domains for 7097 out of 9589 proteins were obtained. In addition to this, standalone version of ANCHOR (Mészáros et al. 2009) was used to predict binding interfaces in the disordered regions of the interacting proteins. Short linear motifs referred to as ELMs (Eukaryotic Linear Motifs) were also predicted using the standalone version of ANCHOR (Mészáros et al. 2009).
4.2.7. Protein post-translational modification (PTM)

The data for different types of PTMs were retrieved from dbPTM (Lee et al. 2006) which is an integrated repository from eleven public resources. dbPTM stores information with experimental evidence for 60 types of post translation modifications for 13073 proteins. Out of 9589 proteins considered in this study, PTMs are found to be associated with 8131 proteins. To estimate over-representation and under-representation of a given PTM \(i\), enrichment \((Z_i)\) in a group of proteins (TSH, HKH etc) \(j\) was calculated as the ratio of observed \((O_i)\) to the expected \((E_i)\) numbers of PTM \(i\) in that set of proteins \(j\).

\[
Z_i = \frac{O_i - E_i}{\sqrt{E_i}}
\]

where \(E_i = \frac{\text{Number of proteins with PTM} \ i}{\text{Total number of proteins}} \times \text{Total number of proteins in } j\)

When enrichment \(Z\) of a PTM \(i\) is >1.0 (when \(O > E\)) then that PTM is referred as over-represented and when enrichment is <1.0 (when \(O < E\)) then PTM type \(i\) is referred as under-represented for a set of proteins \(j\) (TSH, HKH etc.)

4.2.8. Protein isoforms due to alternative splicing

The number of protein products (evidence at the transcript level) possible due to alternative splicing of every gene was retrieved from Ensembl 69 (Flicek et al. 2012) (ftp://ftp.ensembl.org/pub/release69/fasta/homo_sapiens/pep/Homosapiens.GRCh37.69.pep.all.fa.gz). This information could be retrieved only for 9431 genes.

4.2.9. Evolutionary rate (dN)

The dN values for human genes which have been estimated with respect to mouse (\(Mus\) \(Musculus\)) orthologs were downloaded from BioMart (BioMart May 2012)
(www.ensembl.org/biomart/martview). Out of 9589 proteins considered in this study, dN values for only 8899 proteins could be obtained.

4.2.10. Functional Enrichment

DAVID (Dennis et al. 2003) was used for GO functional enrichment and IPRSCAN (Zdobnov & Apweiler 2001) for domain enrichment studies of hubs and non-hubs.

4.3. Results and Discussion

4.3.1. Identification of global and local hubs in tissue-specific networks

Tissue-specific networks were constructed by mapping tissue-wise expression data pertaining to 70 tissues on to human global PPI. In total, there were 1979 hubs identified in tissue-specific networks. Proteins in tissue-specific networks can be classified into five distinct classes on the basis of their degree and expression breadth (EB; the number of tissues in which a protein is expressed) (Figure 4.1). They are: 1) House-keeping hubs (HKHs): Proteins expressed in $\geq 60$ tissues and also form hubs in $\geq 60$ tissues. Of 1979 hubs, 908 HKHs were found. 2) Tissue-specific hubs (TSHs): Proteins expressed in $\leq 10$ tissues and hubs in $\leq 10$ tissues. Of 1979 hubs, 220 TSHs were found. There is another class of hubs which although expressed in all tissues prefer only few tissues to become hub. These hubs are referred to as tissue-preferred hubs (TPHs). 3) TPHs are defined as proteins expressed in $\geq 60$ tissues but are hubs in $\leq 10$ tissues. Only 138 TPHs were found. 4) Housekeeping non-hubs (HKNHs): Proteins expressed in $\geq 60$ tissues and non-hubs in all of them. Of 7610 non-hubs, 3663 were HKNHs and 5) Tissue-specific non-hubs (TSNHs): Proteins expressed in $\leq 10$ tissues and are non-hubs in those tissues. Among 7610 non-hubs, 1903 were TSNHs. Comparative analysis of TSHs, TPHs and HKHs revealed that TSHs and HKHs exhibit distinct properties while TPHs exhibit properties similar to HKHs. Hence, in the following sections the results of comparative analysis of TSHs and HKHs are emphasized. Following is the list of number of hubs, HKHs, TSHs, TPHs and other hubs identified in 70 tissue-specific networks (Table 4.1).
Figure 4.1. Schematic representation of tissue-specific networks and various kinds of hubs and non-hubs. Hubs are shaded with red colour. Tissue-specific proteins are colored in violet and housekeeping proteins in yellow colour.
Table 4.1: Number of Hubs, HKHs, TSHs, TPHs and Other hubs in different tissues.

<table>
<thead>
<tr>
<th>Tissues</th>
<th>Hubs</th>
<th>HKH</th>
<th>TSH</th>
<th>TPH</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adipocyte</td>
<td>1127</td>
<td>905</td>
<td>5</td>
<td>0</td>
<td>217</td>
</tr>
<tr>
<td>AdrenalCortex</td>
<td>1234</td>
<td>899</td>
<td>6</td>
<td>0</td>
<td>329</td>
</tr>
<tr>
<td>Adrenalgland</td>
<td>1005</td>
<td>887</td>
<td>1</td>
<td>0</td>
<td>117</td>
</tr>
<tr>
<td>Amygdala</td>
<td>1254</td>
<td>906</td>
<td>2</td>
<td>3</td>
<td>343</td>
</tr>
<tr>
<td>Appendix</td>
<td>1259</td>
<td>896</td>
<td>11</td>
<td>2</td>
<td>350</td>
</tr>
<tr>
<td>AtrioventricularNode</td>
<td>943</td>
<td>832</td>
<td>3</td>
<td>0</td>
<td>108</td>
</tr>
<tr>
<td>BDCA4+ DendriticCells</td>
<td>1422</td>
<td>893</td>
<td>53</td>
<td>30</td>
<td>446</td>
</tr>
<tr>
<td>Bonemarrow</td>
<td>1207</td>
<td>898</td>
<td>16</td>
<td>3</td>
<td>290</td>
</tr>
<tr>
<td>CardiacMyocytes</td>
<td>1483</td>
<td>901</td>
<td>33</td>
<td>31</td>
<td>518</td>
</tr>
<tr>
<td>Caudatenucleus</td>
<td>1092</td>
<td>898</td>
<td>2</td>
<td>0</td>
<td>192</td>
</tr>
<tr>
<td>CD105+ Endothelial</td>
<td>1345</td>
<td>899</td>
<td>44</td>
<td>20</td>
<td>382</td>
</tr>
<tr>
<td>CD14+ Monocytes</td>
<td>1352</td>
<td>895</td>
<td>38</td>
<td>19</td>
<td>400</td>
</tr>
<tr>
<td>CD19+ BCells (neg._sel.)</td>
<td>1340</td>
<td>892</td>
<td>34</td>
<td>18</td>
<td>396</td>
</tr>
<tr>
<td>CD34+</td>
<td>1537</td>
<td>903</td>
<td>72</td>
<td>44</td>
<td>518</td>
</tr>
<tr>
<td>CD4+ Tcells</td>
<td>1356</td>
<td>893</td>
<td>37</td>
<td>24</td>
<td>402</td>
</tr>
<tr>
<td>CD56+ NKCells</td>
<td>1487</td>
<td>895</td>
<td>53</td>
<td>50</td>
<td>489</td>
</tr>
<tr>
<td>CD8+ Tcells</td>
<td>1292</td>
<td>876</td>
<td>38</td>
<td>21</td>
<td>357</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>1010</td>
<td>874</td>
<td>2</td>
<td>1</td>
<td>133</td>
</tr>
<tr>
<td>CerebellumPeduncles</td>
<td>1295</td>
<td>905</td>
<td>5</td>
<td>6</td>
<td>379</td>
</tr>
<tr>
<td>CiliaryGanglion</td>
<td>907</td>
<td>811</td>
<td>2</td>
<td>0</td>
<td>94</td>
</tr>
<tr>
<td>CingulateCortex</td>
<td>1226</td>
<td>905</td>
<td>2</td>
<td>4</td>
<td>315</td>
</tr>
<tr>
<td>colon</td>
<td>1228</td>
<td>901</td>
<td>5</td>
<td>0</td>
<td>322</td>
</tr>
<tr>
<td>DRG</td>
<td>959</td>
<td>854</td>
<td>0</td>
<td>0</td>
<td>105</td>
</tr>
<tr>
<td>GlobusPallidus</td>
<td>937</td>
<td>838</td>
<td>2</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td>Heart</td>
<td>1449</td>
<td>892</td>
<td>42</td>
<td>32</td>
<td>483</td>
</tr>
<tr>
<td>Hypothalamus</td>
<td>1285</td>
<td>908</td>
<td>3</td>
<td>3</td>
<td>371</td>
</tr>
<tr>
<td>Kidney</td>
<td>983</td>
<td>873</td>
<td>3</td>
<td>0</td>
<td>107</td>
</tr>
<tr>
<td>Liver</td>
<td>1439</td>
<td>897</td>
<td>39</td>
<td>28</td>
<td>475</td>
</tr>
<tr>
<td>Lung</td>
<td>1321</td>
<td>899</td>
<td>18</td>
<td>11</td>
<td>393</td>
</tr>
<tr>
<td>Lymphnode</td>
<td>1136</td>
<td>890</td>
<td>11</td>
<td>2</td>
<td>233</td>
</tr>
<tr>
<td>MedullaOblongata</td>
<td>1100</td>
<td>897</td>
<td>3</td>
<td>3</td>
<td>197</td>
</tr>
<tr>
<td>OccipitalLobe</td>
<td>1113</td>
<td>900</td>
<td>5</td>
<td>1</td>
<td>207</td>
</tr>
<tr>
<td>OlfactoryBulb</td>
<td>1016</td>
<td>879</td>
<td>1</td>
<td>0</td>
<td>136</td>
</tr>
<tr>
<td>Ovary</td>
<td>861</td>
<td>800</td>
<td>1</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>Pancreas</td>
<td>1015</td>
<td>878</td>
<td>3</td>
<td>0</td>
<td>134</td>
</tr>
<tr>
<td>PancreaticIslet</td>
<td>1296</td>
<td>907</td>
<td>4</td>
<td>2</td>
<td>383</td>
</tr>
<tr>
<td>ParietalLobe</td>
<td>1235</td>
<td>902</td>
<td>1</td>
<td>1</td>
<td>331</td>
</tr>
<tr>
<td>Pineal day</td>
<td>1465</td>
<td>908</td>
<td>29</td>
<td>24</td>
<td>504</td>
</tr>
<tr>
<td>Pineal night</td>
<td>1424</td>
<td>905</td>
<td>26</td>
<td>17</td>
<td>476</td>
</tr>
<tr>
<td>Pituitary</td>
<td>1316</td>
<td>908</td>
<td>7</td>
<td>3</td>
<td>398</td>
</tr>
<tr>
<td>Placenta</td>
<td>1244</td>
<td>904</td>
<td>15</td>
<td>2</td>
<td>323</td>
</tr>
<tr>
<td>Pons</td>
<td>1151</td>
<td>903</td>
<td>4</td>
<td>3</td>
<td>241</td>
</tr>
<tr>
<td>PrefrontalCortex</td>
<td>1388</td>
<td>908</td>
<td>10</td>
<td>15</td>
<td>455</td>
</tr>
<tr>
<td>Prostate</td>
<td>1388</td>
<td>908</td>
<td>19</td>
<td>10</td>
<td>451</td>
</tr>
<tr>
<td>retina</td>
<td>1462</td>
<td>908</td>
<td>8</td>
<td>19</td>
<td>527</td>
</tr>
<tr>
<td>Salivarygland</td>
<td>1005</td>
<td>880</td>
<td>6</td>
<td>0</td>
<td>119</td>
</tr>
<tr>
<td>SkeletalMuscle</td>
<td>1343</td>
<td>897</td>
<td>32</td>
<td>10</td>
<td>404</td>
</tr>
<tr>
<td>Skin</td>
<td>967</td>
<td>855</td>
<td>5</td>
<td>0</td>
<td>107</td>
</tr>
<tr>
<td>Gene</td>
<td>Gene expression abundance</td>
<td>Protein abundance</td>
<td>transcript level</td>
<td>protein level</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>Small intestine</td>
<td>1192</td>
<td>899</td>
<td>6</td>
<td>4</td>
<td>283</td>
</tr>
<tr>
<td>Smooth Muscle</td>
<td>1278</td>
<td>900</td>
<td>11</td>
<td>5</td>
<td>362</td>
</tr>
<tr>
<td>Spinal cord</td>
<td>1266</td>
<td>907</td>
<td>3</td>
<td>3</td>
<td>353</td>
</tr>
<tr>
<td>Subthalamic nucleus</td>
<td>1092</td>
<td>895</td>
<td>3</td>
<td>1</td>
<td>193</td>
</tr>
<tr>
<td>Superior Cervical Ganglion</td>
<td>1362</td>
<td>891</td>
<td>29</td>
<td>14</td>
<td>428</td>
</tr>
<tr>
<td>Temporal Lobe</td>
<td>1080</td>
<td>895</td>
<td>2</td>
<td>1</td>
<td>182</td>
</tr>
<tr>
<td>Testis Germ Cell</td>
<td>1151</td>
<td>892</td>
<td>10</td>
<td>2</td>
<td>247</td>
</tr>
<tr>
<td>Testis Intersitial</td>
<td>1103</td>
<td>890</td>
<td>10</td>
<td>0</td>
<td>203</td>
</tr>
<tr>
<td>Testis Leydig Cell</td>
<td>1203</td>
<td>905</td>
<td>10</td>
<td>2</td>
<td>286</td>
</tr>
<tr>
<td>Testis Seminiferous Tubule</td>
<td>1105</td>
<td>891</td>
<td>11</td>
<td>0</td>
<td>203</td>
</tr>
<tr>
<td>Testis</td>
<td>1140</td>
<td>895</td>
<td>20</td>
<td>1</td>
<td>224</td>
</tr>
<tr>
<td>Thalamus</td>
<td>1204</td>
<td>907</td>
<td>2</td>
<td>1</td>
<td>294</td>
</tr>
<tr>
<td>Thymus</td>
<td>1153</td>
<td>881</td>
<td>25</td>
<td>2</td>
<td>245</td>
</tr>
<tr>
<td>Thyroid</td>
<td>1439</td>
<td>908</td>
<td>17</td>
<td>12</td>
<td>502</td>
</tr>
<tr>
<td>Tongue</td>
<td>1214</td>
<td>904</td>
<td>11</td>
<td>3</td>
<td>296</td>
</tr>
<tr>
<td>Tonsil</td>
<td>1260</td>
<td>906</td>
<td>14</td>
<td>3</td>
<td>337</td>
</tr>
<tr>
<td>Trachea</td>
<td>1041</td>
<td>888</td>
<td>6</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>Trigeminal Ganglion</td>
<td>1198</td>
<td>887</td>
<td>14</td>
<td>2</td>
<td>295</td>
</tr>
<tr>
<td>Uterus Corpus</td>
<td>1101</td>
<td>897</td>
<td>4</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Uterus</td>
<td>1124</td>
<td>889</td>
<td>4</td>
<td>3</td>
<td>228</td>
</tr>
<tr>
<td>Whole Blood</td>
<td>1356</td>
<td>902</td>
<td>42</td>
<td>14</td>
<td>398</td>
</tr>
<tr>
<td>Whole Brain</td>
<td>1099</td>
<td>870</td>
<td>6</td>
<td>2</td>
<td>221</td>
</tr>
</tbody>
</table>

Figure 4.2. Expression and Protein Abundance of TSHs and HKHs. A) Distribution plots illustrating Gene expression abundance, B) Scatter plot between protein abundance and expression abundance C) Distribution plots of abundance of TSH and HKH proteins. As shown, HKHs are produced in more number of copies at transcript level than TSHs (P value ~ 1.93e-34). There is a positive correlation between protein and transcript abundance (r =0.48, P value < 2.2e-16). HKH proteins are present in more number of copies than TSHs (P value ~ 1.1208e-15) (Kolmogorov-Smirnov test).
4.3.2. **Tissue-specific hubs are less abundantly expressed than housekeeping hubs**

Expression Abundance (EA) of a gene is calculated as the average of the expression values of that gene in the tissues it is found to be expressed (Park & Choi 2010; Tuller et al. 2008). In coherence with the earlier report it was found that hubs are more abundantly expressed than non-hubs (P value < 2.2 e -16) (Saeed & Deane 2006). Between the two classes of hubs, TSHs are less abundantly expressed as compared with HKHs (P value ~ 7.8e-35) (Figure 4.2 A).

There are many controversies regarding the correlation between mRNA and protein abundance (Greenbaum et al. 2003) but in the present study, a positive correlation between protein abundance and gene expression abundance was found (r = 0.48, P value < 2.2e-16) (Figure 4.2 B). The values of protein abundance were retrieved from PaxDb. Of the 220 TSHs and 908 HKHs the abundance values for 159 and 804 were obtained, respectively. Nonetheless, it was found that TSH proteins are less abundant than HKH proteins (Figure 4.2 C).

4.3.3. **Tissue-specific hubs are more tightly regulated class of disordered proteins than house-keeping hubs**

In eukaryotes, the hubs are commonly known to be disordered (Haynes et al. 2006). This study also found that hubs are significantly more disordered than non-hubs (P value ~ 1.1e-4) (Figure 4.3 A). Disordered proteins exhibit conformational variability and are capable of interacting with many partners simultaneously or at different time intervals. Between TSHs and HKHs, TSHs are significantly more disordered than HKHs (P value ~ 0.005) (Figure 4.3 B). It has been shown that the disordered proteins are tightly regulated set of proteins with less expression, less mRNA half-life together with frequent degradation (Gsponer & Babu 2009). PEST motifs are the sites for the proteolytic degradation and are known to reduce the half-lives of proteins. It was found that TSH proteins contain significantly higher number of PEST motifs than HKHs (P value ~ 0.003) (Figure 4.3 C) indicating shorter half-lives of the TSHs as compared to the HKHs. Since, TSH proteins are limited to tissue-specific roles so,
they are less abundantly expressed and are quickly degraded, whereas HKH proteins are required continuously for the cell, they are, therefore, abundantly expressed and are also endowed with longer half-lives. Non-hubs are similar to HKHs with least number of PEST motifs and thus seem to have longer half-lives.

4.3.4. Tissue-specific hubs are evolving at faster rates and are lengthier than housekeeping hubs

The major determinant of evolutionary rates of the genes in mammals are their size, essentiality, expression breadth and also their expression levels (Liao et al. 2006). It is very well known that hubs are under evolutionary pressure as compared with non-hubs (Fraser et al. 2002). Therefore, the evolutionary rates of TSHs and HKHs were compared which were estimated as the rate of non-synonymous mutations in human proteins as compared to mouse orthologs. It was found that hubs, in general, are evolving at slower rates than non-hubs and among hubs, HKHs are evolving at slower rates than TSHs (P value ~ 2.8e-21) (Figure 4.4 A). Hubs are known to show an over-representation of multi-domain proteins and therefore are known to be lengthier than non-hubs (Ekman et al. 2006). When the lengths of TSH and HKH proteins were compared vis-a-vis the non-hub proteins, it was found that HKH proteins are significantly shorter than TSH proteins (P value ~ 0.001) but longer than non-hub proteins (P value ~ 0.02) (Figure 4.4 B). During evolution there can be selection for gene/protein compactness due to the deletion of functionally redundant domains/segments hence, housekeeping genes and their translated protein products are shorter than the other genes and their protein products (Eisenberg & Levanon 2003).

4.3.5. Tissue-specific hubs and housekeeping hubs exhibit distinct structural properties

Small hubs are usually characterized by large charged surfaces, which facilitate them to make hydrophobic interactions with their partners whereas, large hubs are characterized by disordered regions, which facilitate them to adopt versatile conformations (Patil & Nakamura...
Figure 4.3. Distribution plots illustrating A) Percentage Disorderliness in proteins B) Percentage Disorderliness in TSHs and HKHs and C) Number of PEST motifs in TSHs and HKHs. Hubs are significantly more disordered than non-hubs (P value ~ 1.1e-4). Among TSHs and HKHs, TSHs are significantly more disordered (P value ~ 0.005) and highly enriched with PEST motifs (P value ~ 0.003) (Kolmogorov-Smirnov test). Median values (for TSHs 36.64 and for HKHs 31.38) are indicated by broken lines.

Figure 4.4. Boxplots illustrating distribution of A) Evolutionary Rates and B) Protein Lengths for hubs, non-hubs, TSHs, TSNHs, HKHs and HKNHs (outliers have been masked for clarity). It can be seen that hubs are evolutionarily more conserved than non-hubs (P value ~ 2.3e-62) and enriched by longer proteins than non-hubs (P value ~ 1.15e-06). TSHs are longer (P value ~ 0.0015) and fast evolving proteins (P value ~ 8.75e-22) as compared to HKHs. HKHs are the most evolutionarily conserved proteins and comprise of compact proteins. TSHs and HKHs are evolutionarily more conserved (P value < 0.004) and lengthier (P value < 0.0006) than their non-hubs counter parts i.e. TSNH and HKNH. (Kolmogorov-Smirnov test).
In general, hubs are enriched with charged interfaces and exposed surfaces as compared with non-hubs. The number of exposed residues in protein was predicted by SSpro4 (Cheng et al. 2005). When these structural features were compared between TSHs and HKHs, it was found that HKHs contain significantly higher number of charged residues but a lesser number of exposed residues than TSHs (Figure 4.5 A and B). SSpro4 (Cheng et al. 2005) was used to predict the secondary structures of hubs and non-hubs. When the % of residues in the loop regions was calculated, it was found that hubs have higher fraction of residues in loop regions as compared to non-hubs. Moreover, TSHs also show higher % of residues in loops regions than HKHs (Figure 4.5 C).

**4.3.6. Tissue-specific hubs are comprised of fewer low complexity regions as compared with housekeeping hubs**

Low Complexity Regions (LCRs) are known for increasing plasticity of the protein both by acquiring static (Tatham & Shewry 2000) and dynamic conformations (Tompa 2002; Dunker et al. 2000). The flexibility acquired by LCRs helps proteins to bind versatile partners. Low-complexity regions in hubs, TSHs and HKHs were identified using Segmasker (Wootton & Federhen 1996) which is a part of blast+ package. It has been also reported that the number of interacting partners of a protein is positively related to the LCR length (Coletta et al. 2010). Similarly, it was found that hubs contain more number of LCRs than non-hubs (P value \( \sim 1.12e-08 \)). Moreover, HKHs have higher fraction of residues in LCRs as compared to TSHs (P value 0.021) (Figure 4.5 D). Based on the position of LCRs, Coletta et al (2010) have defined LCRs as terminal-LCR and centre-LCR, and have also reported their position dependent roles. Taking the terminology as defined by Coletta et al, it was found that HKHs are enriched with terminal-LCRs than TSHs as they are associated with functions such as protein transport, protein complexes etc.
Figure 4.5. Comparison of distribution of A) Percentage of charged residues, B) Percentage of exposed residues, C) Fraction of residues in coils/loop regions and D) Percentage of residues in Low Complexity Regions in TSH, TSNH, HKH and HKNH. TSHs consist of less number of LCRs (P value ~ 0.021) and less number of charged residues (P value ~ 8.9e-06) than HKHs. Furthermore, TSHs comprise of more numbers of residues in coils (P value ~ 0.001) and more number of solvent exposed residues (P value ~ 0.01) as compared to HKHs. (Kolmogorov-Smirnov test)
4.3.7. **Tissue-specific and housekeeping hubs have similar propensities for interactions but distinctly differ in their interaction degrees**

Larger binding interface and several short linear motifs facilitates hubs to interact with different partners. Hubs harbor significantly more number of residues at binding interfaces than non-hubs (P value ~ 2.15e-27). As mentioned in material and methods section, standalone version of ANCHOR (Mészáros et al. 2009) was used to predict binding interfaces in the disordered regions of the interacting proteins. For interacting domains, domains involved in interactions either in 3did (Stein et al. 2011), iPfam (Finn et al. 2005) or predicted as high confidence (HC) in DOMINE (Yellaboina et al. 2011) were considered. Short linear motifs were also predicted using the standalone version of ANCHOR (Mészáros et al. 2009). Short linear motifs are also over-represented in hubs as compared with non-hubs (P value ~ 5.11-09). It was found that both TSHs and HKHs harbour similar number of interacting domains (**Figure 4.6 A**). Their disordered regions also harbour similar number of binding interfaces (P value ~ 0.25) (**Figure 4.6 A**) suggesting similar propensities of the both hubs for making interactions with other proteins. However, it is interesting to note that the number of interactions made by TSHs is significantly less as compared to HKHs (P value ~ 3.274e-12) (**Figure 4.6 B**) which suggests that TSHs are probably the unsaturated hubs still evolving for interactions and (or) the binding sites in HKHs are probably promiscuous. It was further found that TSHs are extensively enriched with short linear motifs as compared to HKHs (P value < 2e-04) (**Figure 4.6 C**) indicating TSHs to be involved in transient and reversible interactions and therefore ideal for signalling pathways or dynamic interactions. Similar results were obtained when TSHs and HKHs were identified in TSNs constructed with high confidence PPI dataset.
Figure 4.6. Interaction domains and degree of TSHs, TPHs and HKHs A) Number of interacting domains and predicted binding sites found in TSHs, TPHs and HKHs: All the three categories i.e. TSHs, TPHs and HKHs possess similar number of interacting domains (P value ~ 0.46). TSHs and HKHs possess similar number of binding interfaces (P value ~ 0.25) in the disordered regions whereas TPHs possess least number binding interfaces compared to TSHs and HKHs (P value 0.0028). (Outliers have been masked for clarity). B) Degree (= max (degree, i = 1 to 70 tissues) ) distributions of TSHs, TPHs and HKHs. The number of interactions made by TSHs is significantly less as compared to HKHs (P value ~ 3.274e-12). The number of interactions made by TPH is significantly less as compared to HKHs and TSHs (P value < 10^-30). Median values are indicated by broken lines. C) Number of Short linear motifs found in TSHs, TPHs and HKHs: Between TSHs and HKHs, TSHs are more enriched with short linear motifs than HKHs (P value < 2e-04). TPHs is similar to HKHs with respect to number of short linear motifs (P value 0.09) but lesser than TSHs (P value ~ 1.837e-05). (Kolmogorov-Smirnov test) (Outliers have been masked for clarity).
4.3.8. Tissue-specific hubs are signalling proteins whereas housekeeping hubs are involved in core cellular machinery

Gene ontology functional enrichment analysis of HKHs and TSHs was carried out using DAVID (Dennis et al. 2003). Gene ontology terms related to transcription regulation (GO:0016563), signalling by protein kinase activity (GO:0004672), receptors like steroid hormone receptors (GO:0003707) or nuclear receptors (GO:0004879) are enriched in TSH proteins. TSHs are mostly kinases, receptors or secreted proteins. Some of the examples are prothrombin (UniProtKB: P00734) which is a hub specific to liver tissue and is secreted in plasma. Another example is the toll-like receptor TLR4 (UniProtKB: O00206), a hub specific to blood tissue and monocyte which gets activated during immune response. Another example is Retinoic acid receptor RXR-gamma (UniProtKB: P48443), a hub specific to pineal day and night tissues which binds to their target response elements in response to retinoic acid. Moreover, TSHs are enriched with domains and motifs associated with protein kinase and SH2 motif which clearly indicate the involvement of these hubs in intracellular signal transduction (Table 4.2). IPRSCAN (Zdobnov & Apweiler 2001) was used to scan domains in TSH and HKH proteins.

On the other hand, gene ontology terms like macromolecule metabolic process (GO:0032268), RNA splicing (GO:0008380), mRNA processing (GO:0006397) and complex assembly (GO:0065003) are enriched in HKHs. Thus, HKHs are the components of core cellular machinery such as transcription, enzymes, spliceosome complexes involved in regulation of metabolic processes. Since these molecular functions are always required in the cell, HKH proteins are abundant and also escape the degradation mechanisms required for the proper maintenance of the cell. Even, the domains associated with HKHs are components of various core ubiquitous functions of all types of cells (Table 4.3).
Table 4.2: Domain enrichment for tissue-specific hubs as obtained by DAVID.

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Fold Enrichment</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPR000719:Protein kinase, core</td>
<td>4.76</td>
<td>7.8e-9</td>
</tr>
<tr>
<td>IPR017441:Protein kinase, ATP binding site</td>
<td>4.81</td>
<td>1.5e-8</td>
</tr>
<tr>
<td>IPR008271:Serine/threonine protein kinase, active site</td>
<td>4.63</td>
<td>1.5e-5</td>
</tr>
<tr>
<td>IPR017442:Serine/threonine protein kinase-related</td>
<td>4.57</td>
<td>1.9e-5</td>
</tr>
<tr>
<td>IPR000980:SH2 motif</td>
<td>8.45</td>
<td>9.2e-5</td>
</tr>
<tr>
<td>IPR008946:Nuclear hormone receptor, ligand-binding</td>
<td>13.03</td>
<td>1.25e-3</td>
</tr>
<tr>
<td>IPR000536:Nuclear hormone receptor, ligand-binding, core</td>
<td>13.04</td>
<td>1.25e-3</td>
</tr>
<tr>
<td>IPR001245:Tyrosine protein kinase</td>
<td>7.29</td>
<td>1.43e-3</td>
</tr>
<tr>
<td>IPR001628:Zinc finger, nuclear hormone receptor-type</td>
<td>11.90</td>
<td>1.30e-2</td>
</tr>
<tr>
<td>IPR001723:Steroid hormone receptor</td>
<td>11.64</td>
<td>1.48e-2</td>
</tr>
<tr>
<td>IPR013088:Zinc finger, NHR/GATA-type</td>
<td>10.73</td>
<td>2.36e-2</td>
</tr>
</tbody>
</table>

Table 4.3: Domain enrichment for housekeeping hubs as obtained by DAVID.

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Fold Enrichment</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPR012677:Nucleotide-binding, alpha-beta plait</td>
<td>3.66</td>
<td>1.03e-9</td>
</tr>
<tr>
<td>IPR017441:Protein kinase, ATP binding site</td>
<td>2.65</td>
<td>1.88e-9</td>
</tr>
<tr>
<td>IPR000504:RNA recognition motif, RNP-1</td>
<td>3.61</td>
<td>3.29e-9</td>
</tr>
<tr>
<td>IPR000719:Protein kinase, core</td>
<td>2.57</td>
<td>4.84e-9</td>
</tr>
<tr>
<td>IPR001452:Src homology-3 domain</td>
<td>3.40</td>
<td>7.77e-8</td>
</tr>
<tr>
<td>IPR000717:Proteasome component region PCI</td>
<td>13.10</td>
<td>1.32e-7</td>
</tr>
<tr>
<td>IPR008271:Serine/threonine protein kinase, active site</td>
<td>2.73</td>
<td>1.61e-7</td>
</tr>
<tr>
<td>IPR017442:Serine/threonine protein kinase-related</td>
<td>2.64</td>
<td>8.18e-7</td>
</tr>
<tr>
<td>IPR000626:Ubiquitin</td>
<td>6.46</td>
<td>1.21e-5</td>
</tr>
<tr>
<td>IPR017998:Chaperone, tailless complex polypeptide 1</td>
<td>15.19</td>
<td>1.34e-5</td>
</tr>
<tr>
<td>IPR000608:Ubiquitin-conjugating enzyme, E2</td>
<td>6.79</td>
<td>1.94e-5</td>
</tr>
<tr>
<td>IPR002290:Serine/threonine protein kinase</td>
<td>2.79</td>
<td>2.34e-5</td>
</tr>
<tr>
<td>IPR000980:SH2 motif</td>
<td>4.01</td>
<td>2.70e-5</td>
</tr>
<tr>
<td>IPR019955:Ubiquitin supergroup</td>
<td>5.71</td>
<td>7.84e-5</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>E-value</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>IPR006649</td>
<td>Like-Sm ribonucleoprotein, eukaryotic and archaea-type</td>
<td>10.92</td>
</tr>
<tr>
<td>IPR002194</td>
<td>Chaperonin TCP-1, conserved site</td>
<td>14.86</td>
</tr>
<tr>
<td>IPR016135</td>
<td>Ubiquitin-conjugating enzyme/RWD-like</td>
<td>5.80</td>
</tr>
<tr>
<td>IPR001163</td>
<td>Like-Sm ribonucleoprotein, core</td>
<td>9.77</td>
</tr>
<tr>
<td>IPR002423</td>
<td>Chaperonin Cpn60/TCP-1</td>
<td>11.14</td>
</tr>
<tr>
<td>IPR014021</td>
<td>Helicase, superfamily 1 and 2, ATP-binding</td>
<td>3.64</td>
</tr>
<tr>
<td>IPR001650</td>
<td>DNA/RNA helicase, C-terminal</td>
<td>3.64</td>
</tr>
<tr>
<td>IPR014001</td>
<td>DEAD-like helicase, N-terminal</td>
<td>3.58</td>
</tr>
<tr>
<td>IPR000555</td>
<td>Mov34/MPN/PAD-1</td>
<td>11.43</td>
</tr>
<tr>
<td>IPR001353</td>
<td>Proteasome, subunit alpha/beta</td>
<td>8.80</td>
</tr>
<tr>
<td>IPR019954</td>
<td>Ubiquitin conserved site</td>
<td>6.01</td>
</tr>
<tr>
<td>IPR000308</td>
<td>14-3-3 protein</td>
<td>15.92</td>
</tr>
<tr>
<td>IPR005937</td>
<td>26S proteasome subunit P45</td>
<td>15.99</td>
</tr>
<tr>
<td>IPR011989</td>
<td>Armadillo-like helical</td>
<td>3.04</td>
</tr>
<tr>
<td>IPR011545</td>
<td>DNA/RNA helicase, DEAD/DEAH box type, N-terminal</td>
<td>4.26</td>
</tr>
</tbody>
</table>
4.3.9. Housekeeping hubs are subjected to more number of post-translational modifications than tissue-specific hubs

Specific protein-protein interactions are mediated by several post-translational modifications (PTMs). It is known that almost 60% of PTMs sites are located in the functional domains of proteins. It is also known that PTMs can affect protein conformational or functional specificities (Patil et al. 2010). Almost 60 different types of PTMs could be associated with proteins in the present study from dbPTM (Lee et al. 2006). PTMs such as phosphorylation, ubiquitylation, acetylation, N/O-linked glycosylation, and methylation are commonly found in interacting proteins. Hubs compared with non-hubs are found to be enriched with phosphorylation-dephosphorylation, S-Nitrosylation, sumolytion, acetylation, methylation, ubiquitylation and proteolytic cleavage PTMs. (Table 4.4). It has also been shown by earlier studies that the sites associated with post-translational modifications like phosphorylation (Iakoucheva et al. 2004) and ubiquitination (Radivojac et al. 2010) are known to have high levels of disorderness which itself is a common feature of hubs. In contrast, N-linked glycosylation and disulfide bonds are less prevalent in hubs.

When the number of PTM sites were compared between HKHs and TSHs, it was found that more number of PTM sites in HKHs than TSHs indicating HKHs as the major target for many of the PTMs (P value < 3.4e-09) (Figure 4.7). TSH proteins, on the other hand, are enriched with PTMs such as disulphide formation, N-linked glycosylation. Tissue-specific non-hubs and housekeeping non-hubs are less targeted by PTMs than TSHs and HKHs, respectively (P value <1.58e-20) (Figure 4.7).
Table 4.4: Types of post translational modifications (PTMs) in proteins.

<table>
<thead>
<tr>
<th>Types of PTM</th>
<th>Hubs</th>
<th>NonHubs</th>
<th>HKH</th>
<th>HKNH</th>
<th>TSH</th>
<th>TSNH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphorylation</td>
<td>4.04</td>
<td>-2.26</td>
<td>3.33</td>
<td>-0.24</td>
<td>0.91</td>
<td>-3.00</td>
</tr>
<tr>
<td>Ubiquitylation</td>
<td>9.60</td>
<td>-5.54</td>
<td>9.76</td>
<td>4.77</td>
<td>-0.73</td>
<td>-11.55</td>
</tr>
<tr>
<td>Acetylation</td>
<td>14.18</td>
<td>-8.57</td>
<td>15.00</td>
<td>0.91</td>
<td>-0.67</td>
<td>-8.53</td>
</tr>
<tr>
<td>N-linked Glycosylation</td>
<td>-3.55</td>
<td>1.88</td>
<td>-6.45</td>
<td>-3.73</td>
<td>0.97</td>
<td>6.91</td>
</tr>
<tr>
<td>S-Nitrosylation</td>
<td>11.87</td>
<td>-8.19</td>
<td>13.70</td>
<td>-1.71</td>
<td>0.20</td>
<td>-8.23</td>
</tr>
<tr>
<td>Proteolytic Cleavage</td>
<td>7.00</td>
<td>-4.42</td>
<td>5.40</td>
<td>-5.95</td>
<td>0.94</td>
<td>1.93</td>
</tr>
<tr>
<td>Sumoylation</td>
<td>12.52</td>
<td>-9.27</td>
<td>11.32</td>
<td>-5.42</td>
<td>4.84</td>
<td>-3.55</td>
</tr>
<tr>
<td>Caspase</td>
<td>7.50</td>
<td>-4.86</td>
<td>8.41</td>
<td>-0.28</td>
<td>-1.53</td>
<td>5.00</td>
</tr>
<tr>
<td>O-linked Glycosylation</td>
<td>2.81</td>
<td>-1.64</td>
<td>0.17</td>
<td>-2.89</td>
<td>0.64</td>
<td>1.77</td>
</tr>
<tr>
<td>Methylation</td>
<td>12.11</td>
<td>-10.06</td>
<td>14.57</td>
<td>-4.22</td>
<td>2.09</td>
<td>-3.11</td>
</tr>
<tr>
<td>Disulfide bond</td>
<td>-0.31</td>
<td>0.17</td>
<td>-1.40</td>
<td>-2.58</td>
<td>2.07</td>
<td>5.15</td>
</tr>
<tr>
<td>Dephosphorylation</td>
<td>10.19</td>
<td>-9.67</td>
<td>8.16</td>
<td>-4.36</td>
<td>4.43</td>
<td>-2.23</td>
</tr>
<tr>
<td>Palmitoylation</td>
<td>1.78</td>
<td>-1.05</td>
<td>-1.56</td>
<td>-1.05</td>
<td>2.90</td>
<td>-0.14</td>
</tr>
<tr>
<td>Myristoylation</td>
<td>1.08</td>
<td>-0.62</td>
<td>0.98</td>
<td>-1.12</td>
<td>0.43</td>
<td>-0.08</td>
</tr>
<tr>
<td>Prenylation</td>
<td>4.29</td>
<td>-2.98</td>
<td>2.68</td>
<td>-1.17</td>
<td>1.52</td>
<td>-1.64</td>
</tr>
<tr>
<td>Pyrrolidone carboxylic acid</td>
<td>-1.63</td>
<td>0.83</td>
<td>-1.54</td>
<td>-0.21</td>
<td>-0.97</td>
<td>2.95</td>
</tr>
<tr>
<td>Hydroxylation</td>
<td>1.84</td>
<td>-1.13</td>
<td>1.99</td>
<td>-1.43</td>
<td>2.18</td>
<td>-0.53</td>
</tr>
<tr>
<td>Sulfation</td>
<td>-0.91</td>
<td>0.48</td>
<td>-1.71</td>
<td>-2.01</td>
<td>2.86</td>
<td>2.04</td>
</tr>
<tr>
<td>Amidation</td>
<td>-2.26</td>
<td>1.09</td>
<td>-1.54</td>
<td>-1.91</td>
<td>-0.74</td>
<td>4.25</td>
</tr>
<tr>
<td>Carbamidation</td>
<td>-0.58</td>
<td>0.31</td>
<td>-0.68</td>
<td>1.88</td>
<td>-0.67</td>
<td>-1.18</td>
</tr>
<tr>
<td>Isopeptide bond</td>
<td>4.05</td>
<td>-3.60</td>
<td>6.01</td>
<td>-2.58</td>
<td>-0.65</td>
<td>0.64</td>
</tr>
<tr>
<td>N6-malonyllysine</td>
<td>5.05</td>
<td>-5.80</td>
<td>6.01</td>
<td>-1.80</td>
<td>-0.65</td>
<td>-1.12</td>
</tr>
<tr>
<td>Nitration</td>
<td>0.53</td>
<td>-0.30</td>
<td>1.59</td>
<td>-1.80</td>
<td>-0.65</td>
<td>0.06</td>
</tr>
<tr>
<td>Oxidation</td>
<td>1.53</td>
<td>-0.96</td>
<td>0.85</td>
<td>-1.80</td>
<td>2.44</td>
<td>0.64</td>
</tr>
<tr>
<td>Citrullination</td>
<td>1.14</td>
<td>-0.70</td>
<td>1.34</td>
<td>-0.92</td>
<td>1.20</td>
<td>-0.15</td>
</tr>
<tr>
<td>GPI-anchor</td>
<td>-1.16</td>
<td>0.58</td>
<td>-0.34</td>
<td>0.86</td>
<td>-0.57</td>
<td>0.52</td>
</tr>
<tr>
<td>Pyridoxal phosphate</td>
<td>-1.67</td>
<td>0.81</td>
<td>-1.14</td>
<td>-1.24</td>
<td>-0.54</td>
<td>0.66</td>
</tr>
<tr>
<td>C-linked Glycosylation</td>
<td>-0.87</td>
<td>0.44</td>
<td>-0.08</td>
<td>-1.47</td>
<td>-0.50</td>
<td>2.52</td>
</tr>
<tr>
<td>ADP-ribosylation</td>
<td>1.32</td>
<td>-0.85</td>
<td>3.06</td>
<td>-0.27</td>
<td>-0.47</td>
<td>-0.43</td>
</tr>
<tr>
<td>Glutathionylation</td>
<td>2.01</td>
<td>-1.45</td>
<td>3.06</td>
<td>0.26</td>
<td>-0.47</td>
<td>-1.24</td>
</tr>
<tr>
<td>Neddylation</td>
<td>4.77</td>
<td>0.00</td>
<td>5.09</td>
<td>-1.87</td>
<td>-0.47</td>
<td>-1.24</td>
</tr>
<tr>
<td>Gamma-carboxyglutamic acid</td>
<td>0.84</td>
<td>-0.51</td>
<td>-0.93</td>
<td>-1.20</td>
<td>4.06</td>
<td>1.40</td>
</tr>
<tr>
<td>Deamidation</td>
<td>1.08</td>
<td>-0.69</td>
<td>0.28</td>
<td>-1.05</td>
<td>1.99</td>
<td>0.74</td>
</tr>
<tr>
<td>Nucleotide-binding</td>
<td>1.86</td>
<td>-1.37</td>
<td>2.57</td>
<td>0.16</td>
<td>-0.42</td>
<td>-1.09</td>
</tr>
<tr>
<td>Carboxylation</td>
<td>0.51</td>
<td>-0.30</td>
<td>-0.81</td>
<td>-1.53</td>
<td>4.81</td>
<td>1.95</td>
</tr>
<tr>
<td>Allylsine</td>
<td>2.63</td>
<td>-2.84</td>
<td>0.62</td>
<td>-1.40</td>
<td>2.50</td>
<td>0.16</td>
</tr>
<tr>
<td>Alkylation</td>
<td>0.07</td>
<td>-0.04</td>
<td>0.86</td>
<td>1.15</td>
<td>-0.31</td>
<td>-0.83</td>
</tr>
<tr>
<td>Deacetylation</td>
<td>3.18</td>
<td>0.00</td>
<td>3.90</td>
<td>-1.25</td>
<td>-0.31</td>
<td>-0.83</td>
</tr>
<tr>
<td>Glycation</td>
<td>-0.96</td>
<td>0.47</td>
<td>-0.66</td>
<td>1.15</td>
<td>-0.31</td>
<td>0.38</td>
</tr>
<tr>
<td>Transglutamination</td>
<td>0.07</td>
<td>-0.04</td>
<td>-0.66</td>
<td>-1.25</td>
<td>-0.31</td>
<td>2.80</td>
</tr>
<tr>
<td>Lipoprotein</td>
<td>1.56</td>
<td>-1.30</td>
<td>1.18</td>
<td>-1.08</td>
<td>3.40</td>
<td>0.68</td>
</tr>
<tr>
<td>Proteoglycan</td>
<td>-0.84</td>
<td>0.40</td>
<td>-0.57</td>
<td>-0.16</td>
<td>-0.27</td>
<td>2.08</td>
</tr>
<tr>
<td>Thioester bond</td>
<td>1.56</td>
<td>-1.30</td>
<td>1.18</td>
<td>-1.08</td>
<td>-0.27</td>
<td>0.72</td>
</tr>
<tr>
<td>FAD</td>
<td>-0.68</td>
<td>0.33</td>
<td>-0.47</td>
<td>0.25</td>
<td>-0.22</td>
<td>-0.58</td>
</tr>
<tr>
<td>Glycolys phoshatidyli inositol GPI</td>
<td>-0.68</td>
<td>0.33</td>
<td>-0.47</td>
<td>0.25</td>
<td>-0.22</td>
<td>-0.58</td>
</tr>
<tr>
<td>Pyruvate</td>
<td>-0.68</td>
<td>0.33</td>
<td>-0.47</td>
<td>0.88</td>
<td>-0.22</td>
<td>2.84</td>
</tr>
<tr>
<td>Retinal protein</td>
<td>-0.68</td>
<td>0.33</td>
<td>-0.47</td>
<td>0.62</td>
<td>-0.16</td>
<td>-0.41</td>
</tr>
<tr>
<td>Biotin</td>
<td>-0.48</td>
<td>0.23</td>
<td>-0.33</td>
<td>-0.62</td>
<td>0.16</td>
<td>-0.41</td>
</tr>
<tr>
<td>Deglycosylation</td>
<td>-0.48</td>
<td>0.23</td>
<td>-0.33</td>
<td>-0.62</td>
<td>0.16</td>
<td>2.01</td>
</tr>
<tr>
<td>Diphtalamide</td>
<td>1.59</td>
<td>0.00</td>
<td>2.71</td>
<td>-0.62</td>
<td>-0.16</td>
<td>-0.41</td>
</tr>
<tr>
<td>Glycoprotein</td>
<td>1.59</td>
<td>0.00</td>
<td>2.71</td>
<td>-0.62</td>
<td>-0.16</td>
<td>-0.41</td>
</tr>
</tbody>
</table>
Glycyl Serine isopeptide      -0.48  0.23 -0.33 -0.62 -0.16  2.01  
Hypusine                      -0.48  0.23 -0.33  0.98 -0.16 -0.41  
Iodination                     -0.48  0.23 -0.33 -0.62 -0.16  2.01  
Lipoil                        -0.48  0.23 -0.33  0.98 -0.16 -0.41  
N6-1-carboxyethyl lysine       1.59  0.00 -0.33 -0.62 -0.16 -0.41  
N6-carboxyllysine              -0.48  0.23 -0.33 -0.62 -0.16 -0.41  
S-linked Glycosylation         -0.48  0.23 -0.33 -0.62 -0.16  2.01  
Thioether bond                 -0.48  0.23 -0.33 -0.62 -0.16 -0.41  
TPQ                            -0.48  0.23 -0.33 -0.62 -0.16 -0.41  

Note: Hubs and non-hubs targeted by different PTMs. Table shows enrichment score (Z) representing over- and under representation of different PTMs sites as available from dbPTM in Hubs, Non-hubs, HKHs, HKNHs, TSHs and TSNHs. The over-represented enrichment score are shaded in red colour and the under-represented enrichment score are shaded in green colour.

Figure 4.7. Boxplot depicting distribution of number of post-translational modifications (PTMs) sites. Hubs, in general, have more number of PTM sites (P value < 2.2e-16) (outliers have been masked for clarity) compared to non-hubs. Among TSHs and HKHs, HKHs have more number of PTM sites (P value ~ 3.4e-09) than TSHs. (Kolmogorov-Smirnov test).
4.3.10. Genes encoding tissue-specific hubs produce less number of alternative spliced products than genes encoding housekeeping hubs

In general, a node in a PPI network represents the longest or the major isoform (splice variant) of the protein. However, a node is an ensemble of isoforms available for that protein due to the process of alternative splicing and therefore, degree of a node represents the union of the degrees of all the splice variants (Sinha & Nagarajaram 2013). Hubs are associated with more number of splice variants than non-hubs (Sinha & Nagarajaram 2013). The number of splice variants retrieved from Ensembl 69 (Flicek et al. 2012) associated with TSHs and HKHs showed that HKHs have more number of splice variants than TSHs (P value \( \approx 3.4e^{-09} \)) (Figure 4.8) which also explains the higher degrees associated with HKHs than TSHs.

4.3.11. Tissue-preferred hubs behave like housekeeping hubs but possess lesser number of partners

Comparative analyses of different types of hubs identified in tissue-specific networks showed distinct features of TSHs and HKHs. As mentioned earlier there is yet another group of hubs known as Tissue-preferred hubs (TPHs) that are expressed in more than 60 tissues but are hubs in not more than 10 tissues. It was found that TPHs mostly behave like HKHs but harbour less number of binding interfaces in the disordered region (P value 0.0034). Although TPHs are characterized as hubs by virtue of their degrees (>16) but they interact with significantly lesser number of partners as compared to TSHs and HKHs (P value \(<10^{-30}\)).
Figure 4.8. Boxplot depicting distribution of number of spliced variants. Hubs, in general are associated with more number of spliced variants (P value ~ 4.10e-09) compared to non-hubs. Among TSHs and HKHs, HKHs encode more number of spliced forms (P value ~ 0.02) than TSHs. (Kolmogorov-Smirnov test).
4.4. Summary

Highly interacting proteins can be classified based on their expression breadths as (a) ubiquitously present housekeeping hubs (HKHs) and (b) locally present tissue-specific hubs (TSHs). TSHs and HKHs are very distinct at the level of expression, sequence and structure (Figure 4.9). TSHs are lengthier, more disordered and quickly evolving proteins compared with HKHs. TSHs are also less abundantly expressed as compared with HKHs and enriched with PEST motifs indicating their tight regulation. Despite possessing similar number of binding interfaces, TSHs are associated with less degree centrality than HKHs suggesting that TSHs are ‘unsaturated hubs’ with regard to their binding capabilities and HKHs are highly ‘promiscuous hubs’. They play vital role in cell signalling, transcription regulation and act as transporters. These distinct characteristic features of hubs in tissue-specific network were otherwise ignored when global network instead of stratified network was used. With the advent of various kinds of gene expression information many context-specific network can be constructed which could lead to the identification of context-specific regulators and in future may help in drug discovery or targeting specific gene to a particular context. In the light of the hub dichotomy as date and party hubs which is highly argued and debated as mentioned in earlier studies (Batada et al. 2006; Agarwal et al. 2010; Ekman et al. 2006) I would like to propose a new dichotomy of hubs as “local (TSHs) and global hubs (HKHs)”’. As demonstrated, TSHs and HKHs as a consequence of their different expression breadths form distinct groups and each group is distinctly different from the other in terms of protein abundance, evolution, regulation, structural properties and functions. The proposed hub classification thus seems better than the vigorously disputed dichotomy of date and party hubs. In chapter 5, an attempt has been made to examine TSHs and HKHs based on co-expression pattern of their partners.
Figure 4.9. A graphical abstract showing a novel dichotomy of hubs: Tissue-specific hubs (TSH/local) and Housekeeping hubs (HKH/global).