PROGRESS IN BIOLOGICAL SCIENCES

# Construction and Analysis of Tissue-Specific Protein-Protein Interaction Networks in Humans

Mitra Mirzarezaee[1, *], Mehdi Sadeghi[2, 3] and Babak N. Araabi[4, 5]

[1]*Department of Computer Engineering, Science and Research Branch , Islamic Azad University, Tehran, Iran.*
[2]*National Institute of Genetic Engineering and Biotechnology(NIGEB), Tehran, Iran.*
[3]*School of Computer Sciences , Institute for Research in Fundamental Sciences, IPM, Tehran, Iran.*
[4]*Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Iran.*
[5]*School of Cognitive Sciences, Institute for Research in Fundamental Sciences, IPM, Tehran, Iran.*

Received: July, 16, 2012; Accepted: October, 21, 2012

## Abstract

We have studied the changes in protein-protein interaction network of 38 different tissues of the human body. 123 gene expression samples from these tissues were used to construct human protein-protein interaction network. This network is then pruned using the gene expression samples of each tissue to construct different protein-protein interaction networks corresponding to different studied tissues of the body. This study is helpful for understanding how human protein interactions change in different tissues. In this way, similar tissues of the body and special functions of each tissue, corresponding to their individualized subnetworks, can be identified. We have calculated graph parameters for studying these protein-protein interaction networks and hubs and non-hubs of the studied protein-protein interaction networks are identified. We found a common subnetwork among protein-protein interaction networks of the studied tissues and a tree of tissue similarities has been constructed. We have also found that average correlation coefficient of hubs in human protein-protein interaction networks obeys a normal-like distribution though it is not possible to separate party and date hubs.

**Keywords:** Protein-Protein Interaction network; Normal human tissue; Human gene expressions; Hubs; Non-hubs.

## Introduction

Proteins play essential roles in performing various biological functions in a cell. The most important are the basic cellular processes that they perform within each cell. Studying Protein-Protein Interaction Networks (PINs) on a genome-scale has become possible through advances in high-throughput experimental research. These experiments have generated large amounts of interaction data for several species including S. cerevisiae (Uetz *et al*., 2000; Ito *et al*., 2001; Ho *et al*., 2002, Gavin *et al*., 2006;Krogan *et al*., 2006), Escherichia coli (Butland *et al*., 2005), Drosophila melanogaster (Giot *et al*., 2003), Caenorhabditid elegans (Li *et al*., 2004), and Homo sapiens(Rual *et al*., 2005; Stellzl *et al*., 2005). The corresponding PINs are accessible through databases such as IntAct (Hermjakob *et al*., 2004), DIP(Salwinski *et al*., 2004), and BioGrid (Bretkreutz *et al*., 2003).

In Humans, DNA microarrays (Schena *et al*., 1995; Lockhart *et al*., 1996) have been used to profile gene expression in cancer and other diseases. In cancer, for example, microarray profiling has been applied to classify tumors according to their sites of origin (Su *et al*., 2001; Ramaswamy *et al*., 2001; Bloom *et al*., 2004) to discover previously unrecognized subtypes of cancer (Alizadeh *et al*., 2000; Bittner *et al*., 2000; Perou *et al*., 2000; Bhattacharjee *et al*., 2001; Garber *et al*., 2001; Lapointe *et al*., 2004) to

* Corresponding author: mirzarezaee@srbiau.ac.ir
P.O. Box: 1477893855
Tel.: +98-21-44865100; Fax: +98-21-44817175

predict clinical outcome (van't Veer *et al*., 2002; Shipp *et al*., 2002; Leung *et al*., 2002), and to suggest targets for therapy (Armstrong *et al*., 2004; Stegmaier *et al*., 2004). However, the identification of improved markers for diagnosis and molecular targets for therapy will depend on knowledge not only of the expressed genes in the diseased tissues of interest, but also on detailed information about the expression of the corresponding genes across the gamut of normal human tissues. At present our knowledge on differences in gene expression in various human tissues and under different circumstances is incomplete (Su *et al*., 2002; Saito-Hisaminato *et al*., 2002; Hsiao *et al*., 2001; Warrington *et al*., 2000; Shyamsundar *et al*., 2005). In recent studies (Zoltan *et al*., 2008; Bossi and Lehner, 2009; Souiai *et al*., 2011; Lopes *et al*., 2011), the common tissue specific subnetworks of different parts of the human body and the housekeeping genes have been identified. These studies are either based on the use of gene expressions or data of PINs from other sources. The clear difference of our work with what they had done is in two parts: First, the method of PIN construction is different. We have proposed a new method; when available gene expression samples from each tissue is not enough to directly construct the PIN. Second, we have tried to find the similar PINs in studied tissues.

In the next sections the proposed methods of PIN construction based on gene expression profiles and methods for calculation of similarities between them are discussed and Analysis of graph related parameters, identification of common subnetworks and PINs similarities are explained.

## Materials And Methods

### Gene Expression Profiles

123 gene expression samples from 38 different tissues of normal human body were downloaded from Stanford Microarray Database[1]. In this database, 44064 human genes were studied (Shyamsundar *et al*., 2005). These gene expressions are available in the form of a matrix having 38 rows and 123 columns. The columns represent samples and the rows represent the gene profiles. The gene expression profiles are normalized in a z-score fashion such that the average expression ratio of one profile is 0 and the standard deviation is 1. In the case where gene expressions are not provided, the average of the gene expressions in the other samples are used instead.

It should be mentioned that they have named the Buffycoat as a tissue although it is not really a tissue of the body. We used those data in our experiments as gene expressions can be obtained from that blood layer.

### Protein-Protein Interaction Network Construction Method

Using the gene expression data, we constructed a sparse co-expression network using the k mutual nearest neighbor criterion (Agrawal, 2002). In this method, a list of k nearest neighbor profiles is produced for every gene expression profile. The nearest neighbor of one expression profile is defined as the most similar profile by the Euclidean similarity measure;

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1),$$

Where xi, and yi are the corresponding x and y values in an n dimensional space and n is the number of samples. In this way, a list of k nearest neighbors of each gene (protein) is constructed. Two nodes are connected if they are on each others' list. This way, a gene co-expression network is constructed. The optimal k is 15 (Agrawal and Domany, 2003).

### Pruning Human Protein-Protein Interaction Network In Different Tissues Using Different Cut-off Methods

Identifying active and inactive genes of each tissue needs a criterion. A gene is active if its gene expression value is above a specified threshold otherwise it is inactive. A good threshold for this purpose should be the one that is extracted from

1.http://smd.stanford.edu/cgi-bin/publication/viewPublication.pl?pub_no=426

the data itself. We used the statistics of the data for this purpose. We calculated average, average minus variance, average minus 0.5*variance, average plus variance, average plus 1.5*variance, and median thresholds. In this way, if the gene expression of a protein in a tissue of the body is above the calculated threshold, this protein is active, otherwise it is inactive and all of its interactions with the rest of the network are omitted.In this way, 38 PINs corresponding to 38 studied tissues are obtained by pruning the original constructed PIN.

### *Average Cut-off Method*

In finding the cut-off values using average method, we calculated average gene expressions of each gene in 123 available samples. We used these calculated averages as cut-off values. In each tissue of the body, we compared the expression profile of each gene with its average within available samples. If the average gene expression value in that tissue is above the average of its values in the whole samples, gene is active otherwise it is inactive.

### *Average Minus Variance Cut-off Method*

In finding the cut-off values using average minus variance method, we calculated average and variance of gene expression profiles for each gene in 123 available samples. For each tissue, we compared the expression profile of each gene with its average minus variance value of the samples. If the gene expression value in that tissue is above that threshold, gene is active otherwise it is inactive.
Cut-off values in Average plus 0.5*Variance, Average plus Variance, and Average plus 1.5* variance are calculated in a similar manner.

### *Median Cut-off Method*

In finding the cut-off values using median method, we calculated median of gene expression profiles for each gene using available samples. We used these calculated values as cut-off points. In each tissue of the body, we compared the expression profile of each gene with its median from samples. If the gene expression is above the median of its values in the whole samples, gene is active.

### Graph Related Parameters

The formal representation of PINs as undirected graphs makes it possible to utilize a variety of well-established graph measures. We have computed 12 individual graph measures that reflect the following graph properties: size, distribution, relevance, density, and modularity. The following parameters are calculated: closeness, graph diameter, index of aggregation, entropy of distribution of edges, connectivity, number of edges divided by the number of vertices, entropy, graph centrality, sum of the Wiener number, modified vertex distance number, and Eigen values. More information on these twelve parameters can be found in (Platzer *et al*., 2007).

### Dynamic Behavior of Interactions And Hubs In Different Human Tissues

We calculated number of hubs and interactions of the obtained PINs to see how they change among different tissues. Number of interactions is the summation of all interactions inside the whole PIN. If protein A has interaction with protein B, Protein B also interacts with protein A. Thus, numbers of actual interactions were calculated as summation of all vertex degrees.
Hub proteins are defined as the proteins with most interactions within the PINs. Usually proteins with more than eight interactions are called hubs (Komurov and White, 2007). We used this definition to separate hubs from non-hubs. Hubs are further divided into low connectivity and high connectivity groups. Hubs with up to ten interactions grouped in low connectivity hubs and the ones with more than ten interactions are called high connectivity hubs.

### Identifying Party Hubs And Date Hubs of Human Protein-Protein Interaction Networks

In comparison to date hubs, party hubs are defined as proteins that show high average Pearson Correlation Coefficients (PCC) with their interacting partners as defined by Han *et al*. (Hen *et al*., 2004). The point between two peaks in the bimodal distribution of the hubs and their

partners are selected as the threshold for separating party hubs from date hubs(Hen *et al*., 2004).
In another definition, if gene expression values of a protein vary significantly in different situations or conditions, this protein is assumed dynamic otherwise static(Komurov and White, 2007).
We calculated the average PCC of each hub and its interacting partners using available samples and then estimated the distribution of average PCCs for hubs. We have also calculated variance of gene expressions of each protein in the available samples of these 38 tissues. This way we have labeled proteins as static or dynamic.
We have used two cut-off methods for identifying static and dynamic proteins of humans. Here the methods are explained in more details.

### *Cut-off Method Based on Histogram of The Data*

Histogram of the gene-expression variances/variances divided by averages for all the genes in different samples is plotted. The threshold is placed where the plotted figure is almost smooth. Proteins are labeled dynamic if their gene expressions are above the obtained threshold otherwise they are called static.
In these methods, one cut-off value for all the genes is computed.

### Similarity of Pins of Human Tissues

To identify the most similar tissues of the human body based on their functions or PINs, we calculated pairwise similarity of PINs using common edges in the networks. This means that for each pair of PINs numbers of common edges/common functions are calculated. This similarity measure is then normalized and used for clustering the PINs and constructing a hierarchical tree of human similar tissues. We used Matlab hierarchical clustering tool[1] for this purpose.

## Results

We studied the changes of human Protein Interaction Networks (PINs) within normal tissues. We constructed a general human PIN using expression profiles of human genes in 123 samples from 38 different tissues of the body. The constructed human PIN contains 44064 genes and 25245 edges among these nodes.
We have studied the differences of PINs in different tissues of the body. For this purpose, the constructed human PIN was pruned using the gene expression samples from each tissue to obtain a tissue specific PIN. To prune the original PIN for each tissue where more than one sample of that tissue gene expressions are available, the average of the gene expression from that tissue was used and the network is pruned according to the procedure explained in materials and methods. This way 38 different PINs corresponding to 38 different tissues of the body are obtained. Name of these tissues and available samples from each tissue to prune the original network is specified in table 1. The number of active genes in each tissue based on three applied cut-off methods is shown in table 2. As the obtained results show median, and average methods omit a lot of genes from each tissue. We wanted to study the dynamics where most of the genes are active therefore method of average minus variance was selected as the cut-off method to prune the original PIN.
For the first step in analyzing the differences of tissues PINs, number of edges and hubs of both the original network and the 38 tissue-specific networks were calculated. These results are shown in table 3 and 4. As it can be seen from the results, lymph node has the maximum number of edge (23752) and seminal vesicle has the maximum number of hubs (1041) among the studied tissues. The minimum number of edges (17368) belongs to occipital cortex of brain and the minimum number of hubs (485) is in colon tissue.
Graph related parameters including closeness, graph diameter, index of aggregation, entropy of edge distribution, connectivity, number of edges divided by the number of vertices, entropy, graph centrality, sum of the Wiener number, modified vertex distance number, and Eigen values were calculated as introduced by Plazer (Platzer *et al*., 2007). The results are shown in Table 5. It can be seen that most of the measured graph related

1. www.mathworks.com/help/toolbox/stats/linkage.html

**Table 1.** Gene Expression Samples from each tissue of the Normal Human body.

| Number | Tissue Name | Number of Samples | Sample Names |
|---|---|---|---|
| 1 | Brain, frontal cortex | 3 | GSM39887,GSM39907,GSM39946 |
| 2 | Brain, occipital cortex | 3 | GSM39888,GSM39908,GSM39937 |
| 3 | Uterus | 4 | GSM39889,GSM39902,GSM39952,GSM39988 |
| 4 | Colon | 3 | GSM39890,GSM39943,GSM39945 |
| 5 | Salivary gland | 4 | GSM39891,GSM39900,GSM39901,GSM39920 |
| 6 | Ovary | 5 | GSM39892,GSM39897,GSM39917,GSM39977, GSM39981 |
| 7 | Uterine corpus | 1 | GSM39893 |
| 8 | Vagina | 1 | GSM39894 |
| 9 | Bladder | 2 | GSM39895,GSM39918 |
| 10 | Adrenal | 4 | GSM39896,GSM39906, GSM39961,GSM39986 |
| 11 | Thymus | 2 | GSM39898,GSM39942 |
| 12 | Stomach | 4 | GSM39899,GSM39922,GSM39970,GSM39974 |
| 13 | Testes | 3 | GSM39903,GSM39958,GSM39973 |
| 14 | Diaphragm | 1 | GSM39904 |
| 15 | Brain, temporal cortex | 2 | GSM39905,GSM39935 |
| 16 | Lung | 4 | GSM39909,GSM39912,GSM39921,GSM39969 |
| 17 | Prostate | 6 | GSM39910,GSM39911,GSM39913,GSM39916, GSM39982,GSM40004 |
| 18 | Small bowel | 3 | GSM39914,GSM39936,GSM39959 |
| 19 | Kidney | 5 | GSM39915,GSM39924,GSM39926,GSM39956, GSM39983 |
| 20 | Cervix | 3 | GSM39919,GSM39923,GSM39949 |
| 21 | Parathyroid | 3 | GSM39925,GSM39940,GSM39953 |
| 22 | Seminal vesicle | 7 | GSM39927,GSM39928,GSM39932 |
| 23 | Tonsil | | GSM39929,GSM39931,GSM39944,GSM39951 |
| 24 | Lymph node | 5 | GSM39930,GSM39933,GSM39941,GSM39950, GSM39954 |
| 25 | Spleen | 3 | GSM39934,GSM39963,GSM39971 |
| 26 | Liver | 5 | GSM39947,GSM39948,GSM39968,GSM39976, GSM39984 |
| 27 | Placenta | 1 | GSM39955 |
| 28 | Pancreas | 2 | GSM39957,GSM39990 |
| 29 | Epididymus | 1 | GSM39960 |
| 30 | Fallopian tube | 4 | GSM39962,GSM39964,GSM39991,GSM40000 |
| 31 | Breast | 1 | GSM39965 |
| 32 | Esophagus | 3 | GSM39966,GSM39978, GSM39979 |
| 33 | Muscle | 2 | GSM39967,GSM39989 |
| 34 | Thyroid | 6 | GSM39972,GSM39985,GSM39995,GSM39996, GSM39997,GSM39998 |
| 35 | Heart | 6 | GSM39975,GSM39980,GSM39992,GSM39994, GSM39999,GSM40001 |
| 36 | Pericardium | 1 | GSM39987 |
| 37 | Gallbladder | 1 | GSM39993 |
| 38 | Buffycoat | 2 | GSM39938,GSM39939 |

**Table 2.** Number of proteins in human tissues based on different cut-off methods.

| Cut-off Method/ Human Tissue | Avg.-Var. | Median | Avg. | Correlation Based |
|---|---|---|---|---|
| Brain, frontal cortex | 38979 | 17898 | 17913 | 18111 |
| Brain, occipital cortex | 37436 | 19849 | 19741 | 19976 |
| Uterus | 41608 | 21710 | 21531 | 22174 |
| Colon | 39853 | 17680 | 18122 | 17583 |
| Salivary gland | 39581 | 24066 | 23548 | 23329 |
| Ovary | 41453 | 24755 | 24236 | 24033 |
| Uterine corpus | 40190 | 34507 | 29644 | 29337 |
| Vagina | 39067 | 33842 | 31688 | 29806 |
| Bladder | 39264 | 26707 | 25772 | 25934 |
| Adrenal | 40580 | 24100 | 24059 | 23681 |
| Thymus | 39607 | 18036 | 17953 | 18877 |
| Stomach | 40992 | 20295 | 20244 | 20452 |
| Testes | 39023 | 20689 | 20482 | 20871 |
| Diaphragm | 37757 | 22984 | 22732 | 22756 |
| Brain, temporal cortex | 39628 | 17340 | 16475 | 17556 |
| Lung | 40002 | 20479 | 20281 | 20799 |
| Prostate | 41723 | 22654 | 22429 | 22347 |
| Small bowel | 40461 | 16634 | 17342 | 18162 |
| Kidney | 41531 | 20080 | 19828 | 20282 |
| Cervix | 40839 | 29822 | 30104 | 29635 |
| Parathyroid | 40638 | 20724 | 20473 | 21801 |
| Seminal vesicle | 40337 | 30058 | 26940 | 27264 |
| Tonsil | 39418 | 22861 | 22032 | 22312 |
| Lymph node | 41354 | 19033 | 19375 | 20248 |
| Spleen | 37658 | 21641 | 21640 | 21634 |
| Liver | 38871 | 21526 | 21367 | 21571 |
| Placenta | 40708 | 15231 | 15864 | 18513 |
| Pancreas | 39572 | 20227 | 20604 | 19998 |
| Epididymus | 39894 | 14657 | 15972 | 17844 |
| Fallopian tube | 37490 | 22003 | 21867 | 22032 |
| Breast | 36563 | 20010 | 19728 | 20701 |
| Esophagus | 37181 | 22216 | 21960 | 21902 |
| Muscle | 36257 | 23110 | 23128 | 22651 |
| Thyroid | 37760 | 19878 | 19539 | 20165 |
| Heart | 39598 | 19549 | 19647 | 19667 |
| Pericardium | 38951 | 31793 | 28930 | 27473 |
| Gallbladder | 37697 | 13598 | 13813 | 14861 |
| Buffycoat | 40895 | 21317 | 20182 | 21539 |

**Table 3.** Number of Edges and Hubs in the Original Human PPI.

| Cut-off Method/ Tissue | Edges Number | Hubs Number | 8<=Hubs<=10 | Hubs<=10 |
|---|---|---|---|---|
| Human | 25245 | 1106 | 699 | 407 |

**Table 4.** Number of Edges and Hubs in Tissue Networks with Average – Variance cut-off method.

| Cut-off Method/ Human Tissue | Edges Number | Hubs Number | 8<=Hubs<=10 | Hubs<=10 |
|---|---|---|---|---|
| Brain, frontal cortex | 19330 | 607 | 436 | 171 |
| Brain, occipital cortex | 17368 | 499 | 377 | 122 |
| Uterus | 22191 | 817 | 598 | 219 |
| Colon | 18407 | 485 | 383 | 102 |
| Salivary gland | 22255 | 1015 | 639 | 376 |
| Ovary | 23210 | 980 | 655 | 325 |
| Uterine corpus | 22962 | 1020 | 636 | 384 |
| Vagina | 20962 | 896 | 549 | 347 |
| Bladder | 21819 | 953 | 624 | 329 |
| Adrenal | 22709 | 958 | 643 | 315 |
| Thymus | 21228 | 810 | 555 | 255 |
| Stomach | 20837 | 724 | 500 | 224 |
| Testes | 19709 | 684 | 512 | 172 |
| Diaphragm | 20511 | 872 | 569 | 303 |
| Brain, temporal cortex | 20961 | 798 | 551 | 247 |
| Lung | 19920 | 659 | 488 | 171 |
| Prostate | 22946 | 955 | 631 | 324 |
| Small bowel | 22540 | 971 | 610 | 361 |
| Kidney | 21197 | 684 | 497 | 187 |
| Cervix | 21627 | 870 | 580 | 290 |
| Parathyroid | 22781 | 978 | 628 | 350 |
| Seminal vesicle | 23293 | 1041 | 655 | 386 |
| Tonsil | 21755 | 906 | 594 | 312 |
| Lymph node | 23752 | 1034 | 662 | 372 |
| Spleen | 18342 | 606 | 464 | 142 |
| Liver | 19756 | 756 | 539 | 217 |
| Placenta | 22591 | 971 | 611 | 360 |
| Pancreas | 18492 | 532 | 370 | 162 |
| Epididymus | 22124 | 951 | 604 | 347 |
| Fallopian tube | 17697 | 581 | 428 | 153 |
| Breast | 17999 | 663 | 501 | 162 |
| Esophagus | 18743 | 725 | 525 | 200 |
| Muscle | 17977 | 626 | 461 | 165 |
| Thyroid | 17945 | 567 | 428 | 139 |
| Heart | 20605 | 781 | 541 | 240 |
| Pericardium | 20298 | 798 | 542 | 256 |
| Gallbladder | 18530 | 599 | 440 | 159 |
| Buffycoat | 22099 | 934 | 574 | 360 |

**Table 5.** Graph Related Parameters in Different Tissues of the Normal Human Body.

| Module Name | TotalCC | VarCC | GD | IOA | Hdist | Conn.* | NeNv | Hent | meanGC | varGC | meanW | varW | mVD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Human PIN | 286.8164 | 0.1762 | 0.0156 | 1 | NaN | 0.0027 | 1.0299 | 1.2393e+3 | 0.5335 | 0.1302 | 27.2731 | 3.2856e+3 | 2.3789e+3 |
| Brain, frontal | 3.8685e+3 | 0.1564 | 0.0034 | 1 | 1.5833 | 1.4832e-4 | 1.1725 | 1.3542e+5 | 0.3091 | 0.1620 | 4.9440e+4 | 1.7430e+9 | 5.8259e+5 |
| Brain, occipital | 3.8540e+3 | 0.1616 | 0.0043 | 1 | 1.5242 | 1.5134e-4 | 1.1199 | 1.0973e+5 | 0.3273 | 0.1675 | 4.2958e+4 | 1.5011e+9 | 4.5566e+5 |
| Uterus | 4.0127e+3 | 0.1506 | 0.0029 | 1 | 1.6450 | 1.4171e-4 | 1.2313 | 1.7069e+5 | 0.2931 | 0.1576 | 5.8034e+4 | 2.1699e+9 | 7.5982e+5 |
| Colon | 4.0536e+3 | 0.1603 | 0.0037 | 1 | 1.5157 | 1.3989e-4 | 1.1100 | 1.1594e+5 | 0.3216 | 0.1662 | 4.9096e+4 | 1.8175e+9 | 5.0640e+5 |
| Salivary gland | 3.7898e+3 | 0.1510 | 0.0030 | 1 | 1.7101 | 1.5971e-4 | 1.3091 | 1.8180e+5 | 0.2942 | 0.1577 | 5.3947e+4 | 1.8626e+9 | 7.3209e+5 |
| Ovary | 3.9403e+3 | 0.1491 | 0.0031 | 1 | 1.6891 | 1.4641e-4 | 1.2810 | 1.8755e+5 | 0.2856 | 0.1560 | 5.9679e+4 | 2.1662e+9 | 8.4466e+5 |
| Uterine corpus | 3.7604e+3 | 0.1468 | 0.0030 | 1 | 1.7102 | 1.5396e-4 | 1.3063 | 1.8851e+5 | 0.2832 | 0.1537 | 6.0190e+4 | 2.1489e+9 | 7.8929e+5 |
| Vagina | 3.8487e+3 | 0.1549 | 0.0036 | 1 | 1.6622 | 1.5818e-4 | 1.2630 | 1.6285e+5 | 0.3034 | 0.1619 | 5.4395e+4 | 2.0430e+9 | 6.0441e+5 |
| Bladder | 3.8247e+3 | 0.1524 | 0.0034 | 1 | 1.6961 | 1.5856e-4 | 1.2911 | 1.7540e+5 | 0.2962 | 0.1600 | 5.4145e+4 | 1.9332e+9 | 7.1519e+5 |
| Adrenal | 3.8453e+3 | 0.1483 | 0.0044 | 1 | 1.6851 | 1.4866e-4 | 1.2763 | 1.8206e+5 | 0.2799 | 0.1579 | 6.1214e+4 | 2.4539e+9 | 8.0592e+5 |
| Thymus | 3.8625e+3 | 0.1532 | 0.0034 | 1 | 1.6514 | 1.5085e-4 | 1.2415 | 1.6305e+5 | 0.2969 | 0.1592 | 5.2729e+4 | 1.8524e+9 | 7.0608e+5 |
| Stomach | 4.0141e+3 | 0.1548 | 0.0046 | 1 | 1.6068 | 1.4285e-4 | 1.1965 | 1.5231e+5 | 0.2981 | 0.1638 | 5.8086e+4 | 2.5094e+9 | 6.4933e+5 |
| Testes | 3.8188e+3 | 0.1551 | 0.0042 | 1 | 1.6042 | 1.5031e-4 | 1.1923 | 1.4208e+5 | 0.3029 | 0.1623 | 4.9833e+4 | 1.8523e+9 | 5.8375e+5 |
| Diaphragm | 3.6622e+3 | 0.1536 | 0.0032 | 1 | 1.6780 | 1.6431e-4 | 1.2728 | 1.6053e+5 | 0.2997 | 0.1597 | 5.0616e+4 | 1.6715e+9 | 6.4933e+5 |
| Brain, temporal | 3.8822e+3 | 0.1535 | 0.0038 | 1 | 1.6411 | 1.5030e-4 | 1.2311 | 1.5900e+5 | 0.2988 | 0.1600 | 5.1768e+4 | 1.8775e+9 | 6.7556e+5 |
| Lung | 3.9698e+3 | 0.1561 | 0.0032 | 1 | 1.5903 | 1.4546e-4 | 1.1794 | 1.4169e+5 | 0.3117 | 0.1616 | 4.8307e+4 | 1.7279e+9 | 5.9418e+5 |
| Prostate | 3.9554e+3 | 0.1493 | 0.0031 | 1 | 1.6814 | 1.4659e-4 | 1.2742 | 1.8385e+5 | 0.2907 | 0.1562 | 5.6464e+4 | 2.0842e+9 | 7.9575e+5 |
| Small bowel | 3.8902e+3 | 0.1513 | 0.0031 | 1 | 1.6888 | 1.5174e-4 | 1.2845 | 1.8126e+5 | 0.2901 | 0.1583 | 5.9482e+4 | 2.1667e+9 | 7.4512e+5 |
| Kidney | 4.0440e+3 | 0.1527 | 0.0033 | 1 | 1.6045 | 1.3913e-4 | 1.1914 | 1.5493e+5 | 0.2995 | 0.1598 | 5.7373e+4 | 2.2005e+9 | 6.7249e+5 |
| Cervix | 3.9885e+3 | 0.1533 | 0.0027 | 1 | 1.6417 | 1.4651e-4 | 1.2354 | 1.6539e+5 | 0.3028 | 0.1583 | 5.1045e+4 | 1.7892e+9 | 7.3944e+5 |
| Parathyroid | 3.8486e+3 | 0.1487 | 0.0030 | 1 | 1.6885 | 1.4965e-4 | 1.2827 | 1.8343e+5 | 0.2855 | 0.1550 | 5.9782e+4 | 2.1780e+9 | 7.8357e+5 |
| Seminal vesicle | 3.8170e+3 | 0.1494 | 0.0031 | 1 | 1.7170 | 1.5342e-4 | 1.3138 | 1.9279e+5 | 0.2805 | 0.1562 | 6.1506e+4 | 2.1442e+9 | 8.5533e+5 |
| Tonsil | 3.7481e+3 | 0.1504 | 0.0045 | 1 | 1.6802 | 1.5481e-4 | 1.2738 | 1.7240e+5 | 0.2836 | 0.1595 | 5.9921e+4 | 2.3588e+9 | 7.1146e+5 |
| Lymph node | 3.9103e+3 | 0.1485 | 0.0031 | 1 | 1.7073 | 1.4776e-4 | 1.3024 | 1.9573e+5 | 0.2798 | 0.1556 | 6.2454e+4 | 2.2477e+9 | 8.9473e+5 |
| Spleen | 3.8512e+3 | 0.1613 | 0.0044 | 1 | 1.5685 | 1.5363e-4 | 1.1610 | 1.2506e+5 | 0.3198 | 0.1662 | 4.4513e+4 | 1.6417e+9 | 4.7469e+5 |
| Liver | 3.8739e+3 | 0.1577 | 0.0038 | 1 | 1.6105 | 1.5293e-4 | 1.2041 | 1.4389e+5 | 0.3101 | 0.1633 | 4.8637e+4 | 1.7817e+9 | 5.6783e+5 |
| Placenta | 3.9080e+3 | 0.1510 | 0.0028 | 1 | 1.6854 | 1.5074e-4 | 1.2818 | 1.8128e+5 | 0.2918 | 0.1571 | 5.8001e+4 | 2.0682e+9 | 7.6277e+5 |
| Pancreas | 3.9871e+3 | 0.1604 | 0.0042 | 1 | 1.5379 | 1.4512e-4 | 1.1332 | 1.2104e+5 | 0.3195 | 0.1671 | 5.0935e+4 | 2.0439e+9 | 4.6567e+5 |
| Epididymus | 3.8430e+3 | 0.1520 | 0.0035 | 1 | 1.6867 | 1.5446e-4 | 1.2835 | 1.7709e+5 | 0.2915 | 0.1585 | 5.6846e+4 | 2.0121e+9 | 7.2657e+5 |
| Fallopian tube | 3.7914e+3 | 0.1615 | 0.0059 | 1 | 1.5596 | 1.5763e-4 | 1.1542 | 1.1845e+5 | 0.3219 | 0.1697 | 4.4895e+4 | 1.8169e+9 | 4.7934e+5 |
| Breast | 3.6460e+3 | 0.1605 | 0.0049 | 1 | 1.5988 | 1.6538e-4 | 1.1928 | 1.2703e+5 | 0.3144 | 0.1673 | 4.6333e+4 | 1.7335e+9 | 4.5691e+5 |
| Esophagus | 3.7528e+3 | 0.1598 | 0.0039 | 1 | 1.6126 | 1.6229e-4 | 1.2069 | 1.3542e+5 | 0.3179 | 0.1653 | 4.3875e+4 | 1.5495e+9 | 4.9810e+5 |
| Muscle | 3.6026e+3 | 0.1571 | 0.0053 | 1 | 1.5938 | 1.6371e-4 | 1.1860 | 1.2592e+5 | 0.3112 | 0.1663 | 4.7467e+4 | 1.7931e+9 | 4.9476e+5 |
| Thyroid | 3.7355e+3 | 0.1575 | 0.0043 | 1 | 1.5548 | 1.5346e-4 | 1.1472 | 1.1943e+5 | 0.3145 | 0.1652 | 4.6476e+4 | 1.6922e+9 | 5.1296e+5 |
| Heart | 3.8338e+3 | 0.1529 | 0.0033 | 1 | 1.6309 | 1.5041e-4 | 1.2206 | 1.5424e+5 | 0.2988 | 0.1602 | 5.3928e+4 | 1.9642e+9 | 6.5225e+5 |
| Pericardium | 3.7379e+3 | 0.1526 | 0.0049 | 1 | 1.6388 | 1.5516e-4 | 1.2301 | 1.5274e+5 | 0.2953 | 0.1616 | 5.4636e+4 | 2.1745e+9 | 5.9670e+5 |
| Gallbladder | 3.7408e+3 | 0.1573 | 0.0043 | 1 | 1.5802 | 1.5421e-4 | 1.1694 | 1.2823e+5 | 0.3118 | 0.1629 | 4.4047e+4 | 1.5710e+9 | 5.0181e+5 |
| Buffycoat | 3.9433e+3 | 0.1529 | 0.0033 | 1 | 1.6673 | 1.5017e-4 | 1.2648 | 1.7397e+5 | 0.2953 | 0.1594 | 5.8527e+4 | 2.2395e+9 | 6.9140e+5 |

* Connectivity

parameters are in the same range for different tissues. That means the global properties of these networks do not change significantly from one tissue to the others. This result shows that topology of PINs in different tissues of the body are similar. Therefore, the differences among cells are because of the constructing elements of the PINs not their topology.

As the next step in analyzing the differences between PINs of tissues, intersection of all PINs were calculated. A network with 831 edges among 792 genes is the common network/modules among these 38 PINs. This network can be seen in Figure 1. It is the part that exists in all the tissues so it must show a general function that all the tissues handle. They are a small fraction of the whole interactions (3.1%). If one needs to study a tissue specifically, this general part can be omitted for simplicity and it can be assumed that the specific functions of tissues are not in the common part.

As the last step in identifying the similarities and differences among different tissues, a hierarchal tree based on these network similarities is constructed. It can be observed that some of the studied network tissues are more similar. That means these PINs have some similar functions. This finding can be used in the case where some simulations or artificial tissues are designed or in the study of similar and dissimilar tissues. This tree is shown in Figure 2. The numbers in the leaves of the tree are corresponding to different tissue names as labeled in table 1. We found ovary, tonsil and epididymus in the same cluster. Salivary gland and bladder are most similar in the interactions. Also it has been revealed that different parts of the brain including frontal, occipital, and temporal cortex are also similar in the interactions and functions.

For identifying party hubs and date hubs of the human PINs, we calculated the average PCC of each hub and its interacting partners using available samples and then we estimate the distribution of average PCCs for hubs. The estimated distribution is shown in Figure 3. This distribution is normal-like distribution though it is not possible to separate party hubs and date hubs using this data and its associated distribution. We have also calculated variance of gene expressions of each protein in the available samples of these 38 tissues. We have calculated the cut-off points for separating static and dynamic proteins using two different methods of histogram of variance and histogram of variance divided by average of the gene expression values. This way we have labeled proteins as static or dynamic according to their variances. This data is provided in supplementary file I. As the result shows, using the histogram of variance, 123 dynamic proteins are identified whereas histogram of variance divided by average separates 138 dynamic proteins. Choice of method will depend on the number of dynamic genes that is needed to study.

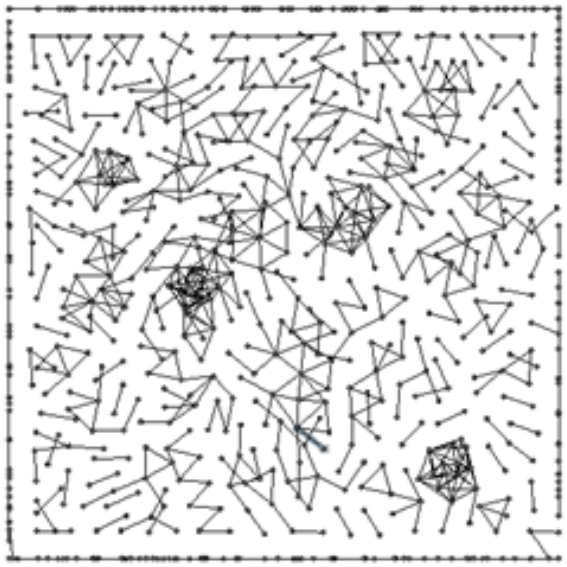

**Figure 1.** The common network among 38 body tissues.

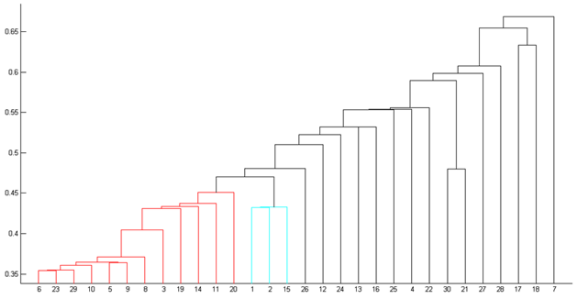

**Figure 2.** Hierarchical tree of tissue similarities.

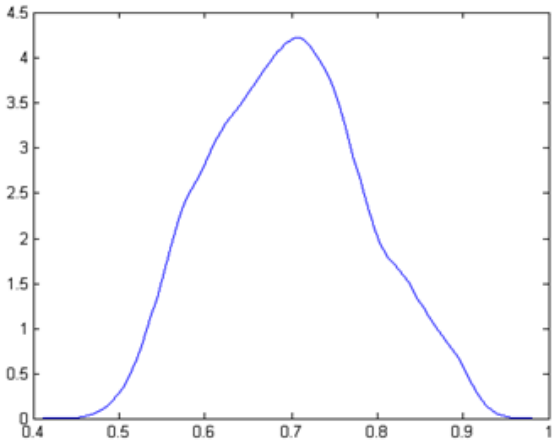


**Figure 3.** Distribution of average PCC of hubs in human proteins interaction network.

## Discussion And Conclusions

In this paper, we studied the human proteins in different tissues of the body. We were looking to see how functions may change from one tissue to the other. In this study, it has been shown that some proteins are not expressed in some of the tissues while they are active in the others. This observation shows that different tissues have different functionalities while they use the same set of genes. That is because different genes are activated in different tissues of the body and their interaction patterns will change from one tissue to the other. A common subnetwork is found in the studied tissues which is responsible for housekeeping and transferring. We have constructed a tree based on similarities of these tissue specific PINs. Some of the tissues are most similar according to this study that means they must have similar functionalities such as three different parts of brain.

We have also studied the changes in the number of hubs and edges of the PINs of the human tissues. We have calculated some of the graph related features of the PINs to see how they are changing from one tissue to the other. We have also estimated the distribution of average PCC among hubs. This distribution does not show a bimodal distribution to allow separating party hubs from date hubs. This result confirms the other studies in which bimodal distributions for separating party and date hubs have not been found (Agarwal *et al*., 2010).

Finally it should be mentioned that we have used the dataset provided by Shyamsunder *et al*. (Shyamsundar *et al*., 2005). It has been reported that the choice of dataset can influence the tissue-specific protein-protein interaction studies(Lopes *et al*., 2011), so we cannot generalize our results before testing them on other datasets as well.

## Acknowledgements

Funding: This work was in part supported by a grant from IPM (No.CS1389-0-01).

## References

1. Agarwal S, Deane CM, Porter MA, Jones NS. (2010) Revisiting Date and Party Hubs: Novel Approaches to Role Assignment in Protein Interaction Networks, PLOS Computational Biology, Vol. 6
2. Agrawal, H. and Domany, E. (2003) Potts ferromagnets on coexpressed gene networks: identifying maximally stable partitions. Phys. Rev. Lett., 90, 158, 102-106
3. Agrawal, H., (2002) Extreme self-organization in networks constructed from gene expression data. Phys. Rev. Lett., 89, 268, 702-706
4. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, *et al*. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, 403:503-511
5. Armstrong SA, Kung AL, Mabon ME, Silverman LB, Stam RW, Den Boer ML, Pieters R, Kersey JH, Sallan SE, Fletcher JA, *et al*..( 2003) Inhibition of FLT3 in MLL. Validation of a therapeutic target identified by gene expression based classification. Cancer Cell, 3:173-183
6. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, *et al*.(2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci USA, 98:13790-13795
7. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z,

Ben-Dor A, *et al*. (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature, 406:536-540

8. Bloom G, Yang IV, Boulware D, Kwong KY, Coppola D, Eschrich S, Quackenbush J, Yeatman TJ. (2004) Multi-platform, multi-site, microarray-based human tumor classification. Am J Pathol,164:9-16
9. Bossi A., Lehner B. (2009) Tissue specificity and the human protein interaction network, Molecular System biology, 5:260
10. Breitkreutz BJ, Stark C, Tyers M. (2003) The GRID: the General Repository for Interaction Datasets. Genome Biol., 4(3):R23
11. Butland G, Peregrín-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A. (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature, 433:531–537
12. Dezso Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, Dosymbekov D, Bugrim A, Rakhmatulin E, Brennan RJ, Guryanov A, Li K, Blake J, Samaha RR, Nikolskaya T. (2008) A comprehensive functional analysis of tissue specificity of human gene expression, BMC Bioinformatics, 6:49
13. Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van De Rijn M, Rosen GD, Perou CM, Whyte RI, *et al*. (2001) Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci USA, 98:13784-13789
14. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature, 440:631–636
15. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, *et al*. (2003) A protein interaction map of Drosophila melanogaster. Science, 302(5651):1727-1736
16. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature, 430:88-93
17. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, Margalit H, Armstrong J, Bairoch A, Cesareni G, Sherman D, Apweiler R. (2004) IntAct: an open source molecular interaction database. Nucleic Acids Res, D452-455
18. Ho Y, Gruhler, A, Heilbut A, Bader GD, Moore L, Adams S, Millar A, Taylor P, Bennett K., Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, *et al*. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature, 415:180–183
19. Hsiao LL, Dangond F, Yoshida T, Hong R, Jensen RV, Misra J, Dillon W, Lee KF, Clark KE, Haverty P, *et al*. (2001) A compendium of gene expression in normal human tissues. Physiol Genomics, 7:97-104
20. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA. , 98:4569–4574
21. Komurov K., White M. (2007) Revealing static and dynamic modular architecture of the eukaryotic protein interaction network. Molecular System Biology, 3:110-122
22. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrín-Alvarez JM, Shales M, Zhang X, *et al*. (2006), Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature, 440:637–643
23. Lapointe J, Li C, Higgins JP, Van De Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, *et al*. (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci

USA, 101:811-816

24. Leung SY, Chen X, Chu KM, Yuen ST, Mathy J, Ji J, Chan AS, Li R, Law S, Troyanskaya OG, *et al*. (2002) Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis. Proc Natl Acad Sci USA, 99:16203-16208

25. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, *et al*. (2004) A map of the interactome network of the metazoan C. elegans. Science, 303(5657):540-543

26. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol, 14:1675-1680

27. Lopes TJ, Schaefer M, Shoemaker J, Matsuoka Y, Fontaine JF, Neumann G, Andrade-Navarro MA, Kawaoka Y, Kitano H. (2011) Tissue-specific sub networks and characteristics of publicly available human protein interaction databases, Bioinformatics, 27:2414-21

28. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, *et al*.(2000) Molecular portraits of human breast tumours. Nature, 406:747-752

29. Platzer A, Perco P, Lukas A, Mayer B. (2007) Characterization of protein-interaction networks in tumors, BMC Bioinformatics, 8:224, doi

30. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, *et al*. (2001) Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci USA, 98:15149-15154

31. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, *et al*. (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature, 437(7062):1173-1178

32. Saito-Hisaminato A, Katagiri T, Kakiuchi S, Nakamura T, Tsunoda T, Nakamura Y (2002) Genome-wide profiling of gene expression in 29 normal human tissues with a cDNA microarray. DNA Res, 9:35-45

33. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. (2004) The database of interacting proteins, NucleicAcids Res, 1;32 D449-451

34. Schena M, Shalon D, Davis RW, Brown PO. (1995) Quantitative monitoringof gene expression patterns with a complementary DNAmicroarray. Science, 270:467-470

35. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, *et al*.(2002) Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med, 8:68-74

36. Shyamsundar R, *et al*. (2005) A DNA microarray survey of gene expression in normal human tissues, Genome Biol, 6(3):R22

37. Souiai O, Becker E, Prieto C, Benkahla A, De Las Rivas J, *et al*. (2011) Functional Integrative Levels in the Human Interactome Recapitulate Organ Organization. PLoS ONE, 6(7): e22051

38. Stegmaier K, Ross KN, Colavito SA, O'Malley S, Stockwell BR, Golub TR. (2004) Gene expression-based high-throughput screening(GEHTS) and application to leukemia differentiation. Nat Genet, 36:257-263

39. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, *et al*. (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell, 122(6):957-968

40. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A, *et al*. (2002) Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci USA, 99:4465-4470

41. Su AI, Welsh JB, Sapinoso LM, Kern SG,

Dimitrov P, Lapp H, Schultz PG, Powell SM, Moskaluk CA, Frierson HF Jr, Hampton GM. (2001) Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res, 61:7388-7393

42. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature, 403:623–627

43. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, *et al*. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature, 415:530-536

44. Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M. (2000) Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiol Genomics, 2:143-147.