Protein Interactions Extrapolated from Feline Protein Complexes

Shawn Martin*[1], Zisu Mao[2], Linda S. Chan[3], and Suraiya Rasheed[2]

[1]Sandia National Laboratories, Department of Computational Biology, PO Box 5800, Albuquerque, NM 87185-0310; [2]Laboratory of Viral Oncology and Proteomics Research, Department of Pathology, University of Southern California, 1840 N. Soto St., Los Angeles, CA 90032-3626; [3]Department of Pediatrics, Keck School of Medicine, University of Southern California, 1840 N. Soto St., Los Angeles, CA 90032-3626

Shawn Martin: smartin@sandia.gov; Zisu Mao: zmao@usc.edu; Linda S. Chan: lschan@usc.edu; Suraiaya Rasheed: srasheed@usc.edu.

*Corresponding Author.

Abstract

The determination of protein-protein interaction networks is a difficult and important problem in biology. Present day approaches to this problem are usually based on two hybrid experimental measurements coupled with refinement and extrapolation using computational techniques. Here we consider a computational method for similar refinement and extrapolation using experimental data from which protein interactions can not be directly inferred. Our dataset was derived from an experiment designed to examine cell growth, tumorigenesis, and differentiation in a highly malignant cat melanoma cell line using 2-D gel electrophoresis and MALDI-TOF mass spectrometry. Instead of direct information about protein-protein interactions, this experiment yields protein complexes. We analyze and interpret these complexes to provide predictions of protein-protein interactions. We find that we are able to predict when a protein pair belongs to a complex with ~96% accuracy, suggesting that these protein pairs interact. We extrapolate the experimentally identified interaction pairs to the entire cat proteome in order to obtain a cat protein interaction network. This network has a scale free degree distribution, in agreement with previous observations about protein interaction networks.

Keywords: protein complexes, predicting protein-protein interactions, feline protein interaction network

1. Introduction

The determination of protein-protein networks is an important problem in biology which has been undertaken using both experimental and computational approaches. Common experimental approaches for identifying protein-protein interactions on a large scale include two hybrid systems (Fields and Song 1989), mass spectrometry (Ho, Gruhler et al. 2002), and protein chips (Zhu, Bilgin et al. 2001). Computational approaches for identifying protein-protein interactions include theoretical approaches based on genome sequences and phylogeny (Dandekar, Snel et al. 1998; Enright, Iliopoulos et al. 1999; Marcotte, Pellegrini et al. 1999; Goh, Bogan et al. 2000; Pazos and Valencia 2001), as well as approaches which refine and extrapolate directly from experimental data such as two hybrid (Bock and Gough 2001; Giot, Bader et al. 2003; Martin, Roe et al. 2005). In this paper we consider a computational method for inferring protein-protein networks using a different type of experimental data.

In previous work (Rasheed, Mao et al. 2005), we generated protein profiles for the study of the trans-differentiation of cat melanoma into neuronal cells using 2-D gel electrophoresis and Matrix Assisted Laser Desorption Ionization Time-of-flight (MALDI-TOF) mass spectrometry. The data we generated in this experiment does not directly provide information on protein-protein interactions and did not cover the entire cat proteome. It does, however, provide accurate information about certain protein groups in terms of spots in the 2-D gel. Each spot in the gel contains either a single protein per spot or exhibits multiple proteins in a complex.

In this paper, we consider a computational approach for inferring a proteome wide interaction network using these protein complexes. We start by investigating the hypothesis that these complexes are composed of functionally related and/or interacting proteins. We exhibit an ability to predict (with ~96% 10-fold cross-validation accuracy) which protein pairs are likely to occur together in a complex.

If we then assume that the protein complexes are composed of interacting proteins, we relate the

interactions to the underlying protein sequences to extrapolate from the 46 proteins present in our experiment to the entire cat proteome (569 proteins). We validate this network by confirming that it has a scale-free degree distribution and by verifying the existence of biologically meaningful protein pathways. The end result is a computationally inferred protein-protein interaction network for cat. To our knowledge this is the only feline protein interaction network (inferred or otherwise) presently available.

2. Materials and Methods

2.1 Proteomics analyses of Feline cells

The complete details of the experimental protocol can be found in (Rasheed, Mao et al. 2005). Here we present an overview for completeness. The biological and molecular characteristics of a highly malignant cat melanoma cell line (CT1413) have been described in (Rasheed 1983). For proteomics analyses, cells were grown in minimal media for 24 hours. After 24 hours, culture medium was removed and fresh growth medium containing RD114 virus was added. Cells were harvested after trans-differentiation (usually after 48 hours). RD114 is an endogenous retrovirus originally derived from the brain of a young cat and grown in human rhabdomyosarcoma cell line (McAllister, Nicolson et al. 1972). To validate protein profiles two independent cell culture experiments were conducted almost 12 months apart and proteomes of both cell types were analyzed separately.

Next, proteins were extracted from 2×10^7 cells from each set of experimental (trans-differentiated) and control melanoma cells. Two different reagents were used to solubilize proteins as rapidly as possible (approximately 10-15 seconds). The most soluble membrane proteins were removed in the first extraction and the less soluble proteins were separated in the second fraction. All cell lysates were sonicated for 2-seconds and clarified by centrifugation for 90 minutes. Proteins were separated by 2-D gel electrophoresis and stained with Coomassie blue and protein spots from each of the 15 gels were evaluated by the use of a CCD camera and an image-processing analytical program (PDQuest from BioRad). This program compared the quality and normalized quantity of each spot across 15 gels, then created a master gel-image of 3,129 well-calibrated, quantifiable spots using internal reference proteins. All differentially expressed (i.e. upregulated and down-regulated) proteins were identified in each gel and 467 spots (including several spots common to both cell types) were excised from multiple gels. Proteins were digested using trypsin and peptide fingerprints of each in-gel digest and peptide fingerprints were analyzed by MALDI-TOF mass spectrometry.

Finally, spectra were submitted to the Swiss-Prot protein database (Boeckmann, Bairoch et al. 2003) for protein identification. We used manual acquisitions of spectra, which yielded more reliable and reproducible results compared to automated acquisitions. The confidence level in our protein identification was high because almost all proteins were confirmed in corresponding spots in multiple gels and by duplicating the entire experiment for validation. Furthermore, we included only those proteins that were most reproducible in the high stringency Feline Protein database and results were confirmed in other mammalian species (Boeckmann, Bairoch et al. 2003). Altogether, we identified 46 proteins among 302 spots from multiple gels tested. All spots in which proteins were not identified reproducibly by mass spectrometry from the same or different gels were not included in any analysis.

2.2 Inferring Interactions from Complexes

The data consisted of groups (complexes) of proteins found together in single spots separated by 2-D gel electrophoresis. We removed redundant complexes and did not consider complexes with only one protein. This left sets of unique complexes with 2 to 8 proteins per complex.

Our initial assumption was that any protein pair within a complex interacts. However, it seems likely that a protein pair that occurs in more than one complex is more likely to interact. We therefore assigned confidence measures to the interactions within the complexes. Our measure is based on the probability that a protein pair will occur at random in multiple complexes.

We observe that the probability that proteins *A* and *B* occur in a complex of size *n* at random is P_n = *P*(*A,B* occurring in complex of size *n*) =

$$\binom{a-2}{n-2} \Big/ \binom{a}{n} = \frac{n(n-1)}{a(a-1)},$$

where *a* is the number of proteins. Next, we calculate the probability that *A* and *B* occur in exactly *m* of *c* complexes. This probability is given by $P_{m,n}$ = *P*(*A,B* occurring in exactly *m* of *c* possible complexes) =

$$\binom{c}{m} P_n^{\,m} (1-P_n)^{c-m}.$$

However, we are more interested in the probability that A and B occur in m or more complexes. This is given by

$$C_{m,n} = \sum_{i>=m} P_{i,n}.$$

Plots showing how $C_{m,n}$ varies with m for different values of n are shown in Figure 1. The curves shown in Figure 1 agree with intuition because larger complexes lead to greater probabilities that a given pair will occur in multiple complexes. Since our complexes vary from in size from 2 to 8, we use the curves $C_{m,2}$ to $C_{m,8}$ to estimate probability values for protein interactions. As an example, the curve $C_{m,8}$ tells us that a protein pair must occur in 11 complexes to be considered significant at the .05 level, while the curve $C_{m,4}$ tells us that a protein pair must occur in 5 complexes to be considered significant at .05.

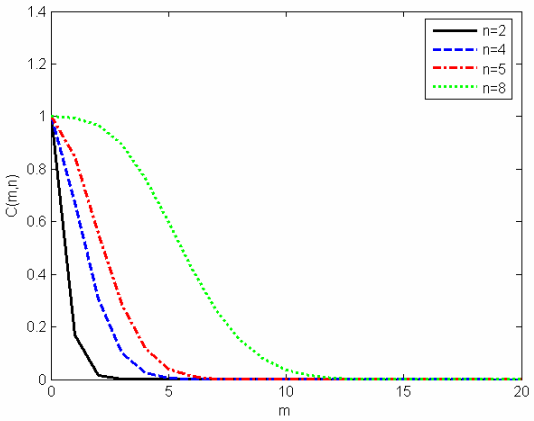

Figure 1: Probabilities of occurrence of a protein pair in m complexes, each of size n. The probabilities decrease as the number of complexes m increases and the complex size n decreases. These curves confirm intuition: larger complexes lead to greater probabilities that a given pair will occur in multiple complexes.

2.3 Relating Interactions to Sequence

After obtaining a set of protein-protein interactions, we seek to evaluate our ability to predict when two proteins will interact. There are a variety of methods available for accomplishing this goal, including the methods described in (Bock and Gough 2001; Sprinzak and Margalit 2001; Ben-Hur and Noble 2005; Martin, Roe et al. 2005). We use the most general of these methods (Ben-Hur and Noble 2005; Martin, Roe et al. 2005) so that we may extrapolate our dataset to the full proteome later in our analysis. The methods of (Ben-Hur and Noble 2005; Martin, Roe et al. 2005) relate protein interactions to sequence using pairwise kernels in Support Vector Machines (SVMs) (Burges 1998; Bennet and Campbell 2000; Cristianini and Shawe-Taylor 2000). We used protein sequences obtained from Swiss-Prot (Boeckmann, Bairoch et al. 2003) and the pairwise product kernel in (Martin, Roe et al. 2005). This method uses SVMs and tensor products (Faulon, Churchwell et al. 2003; Faulon, Visco et al. 2003), and has also been applied successfully to the prediction of β-strand ordering in protein structure prediction (Brown, Martin et al. 2005).

The method used here to relate interactions to sequnces is based on the use of a symmetric tensor product string kernel (Martin, Roe et al. 2005). We use this kernel in a Support Vector Machine to relate sequences to interactions. Support Vector Machines are binary classifiers which make decisions by locating a maximal margin hyperplane (Burges 1998; Bennet and Campbell 2000; Cristianini and Shawe-Taylor 2000). In this work, Support Vector classification was performed using SVM^{light} (Joachims 1999) with the kernel modifications in (Martin, Roe et al. 2005). A patch for SVM^{light}, along with the necessary software for calculating signature products for protein sequences can be obtained from http://www.cs.sandia.gov/~smartin/.

SVMs require not only protein pairs that interact, but also pairs that do not interact. Non-interacting pairs were obtained by sampling at random from the complement of the set of protein pairs taken to interact (as calculated in Section 2.2). To test the ability of the SVMs to relate the protein interaction pairs to their underlying sequences, we used 10-fold cross-validation. We perform 10-fold cross-validation by dividing at random our dataset into ten equal size subsets. For each of the subsets, we obtain predictions from a SVM trained on the remaining nine subsets. We measure the accuracy, sensitivity, and specificity of our predictions. The accuracy is the ratio (TP+TN)/(TP+FP+TN+FN); the sensitivity is the ratio TP/(TP+FN); and the specificity is the ratio TN/(TN+FP). The accuracy measures the overall performance of a classifier; the sensitivity measures the performance of the classifier on the positive examples (interactions); and the specificity measures the performance of the classifier on the negative examples. Finally, to extrapolate from a subset of protein pairs to a full proteome, we simply train our SVM on the full set of known protein pairs (using negatives selected at random) and make predictions on the full proteome.

3. Results

We have compared 3129 protein spots in 15 gels derived from RD114-infected and uninfected melanoma cells. We used two independent experiments and analyzed peptide fingerprints of 467 differentially expressed (up-regulated and down-regulated) protein spots by MALDI-TOF mass

spectrometry. A total of 46 proteins were confirmed unambiguously from 302 spots excised from multiple gels in both experiments. The remaining 165 spots were discarded because they did not identify any protein from the Swiss-Prot database (Boeckmann, Bairoch et al. 2003), or proteins were not identified reproducibly from corresponding spots in different gels.

Each gel spot contained either a single protein per spot in a non-complexed form or a complex of 2–8 proteins in one spot. The frequency of distribution for the 46 proteins among the 302 spots included 82 single protein spots (32 in melanoma and 50 in neuronal cells) and 220 spots containing protein complexes (103 in melanoma and 117 in neuronal cells). From these complexes, we removed 36 redundant complexes to arrive at 184 complexes made up of 45 proteins.

We computed (as described in Materials and Methods, Section 2.2) the probability of protein pairs occurring by chance alone in multiple complexes. Using our most conservative estimate $C_{m,8}$ we identified 31 protein pairs at the .05 confidence level. These pairs are listed in the Appendix. Using our least conservative estimate $C_{m,2}$ we identified 142 protein pairs.

We next related the protein pairs to sequence for different probability estimates using our product Support Vector Machine (SVM) code (see Materials and Methods, Section 2.3). To test the ability of the SVMs to relate the protein pairs to their underlying sequences, we used 10-fold cross-validation to compute the accuracy, sensitivity, and specificity of the SVMs. The results of our calculation are shown in Table 1.

According to this analysis, our accuracies range from ~97% when assuming a complex size of 8 to 84% when assuming that any pair contained in a complex interacts. These accuracies show that the results improve when we use smaller but more reliable training sets. Since it is desirable to use larger training sets, we must choose between accuracy and training set size. A good balance is obtained using the top 69 protein pairs. These pairs occur in at least 6 complexes and have a p-value of < .05 when assuming that each complex has 5 proteins. This choice is also supported by the fact that the median complex size is 4 and the mean size is 4.5. The accuracy assuming each complex has 5 proteins is ~96%.

Num. Pairs	Num. Comps.	Comp. Size	Acc.	Spec.	Sens.
300	1		83.5	84.7	81.6
142	3	2	89.9	92.2	89.4
98	4	3	92.8	91.8	92.8
77	5	4	94.1	92.4	96.0
69	6	5	95.7	95.6	96.3
48	8	6	96.8	95.5	98.3
40	9	7	96.3	95.0	96.7
31	11	8	96.7	97.5	97.5

Table 1: Performance of the SVM model for the classification of protein pairs in complexes using 10-fold cross-validation. The first and second columns were computed using $C_{m,n}$ for the assumed complex size *n* in the third column. The first row uses all protein pairs, which occur in even one complex. Accuracy, specificity and sensitivity are all given as percentages.

The fact that we are able to accurately predict when two proteins will occur in the same complex implies that there is some biological relevance to both the complexes and the protein pairs. This idea is further supported by the fact that we used rapid lyses (see Materials and Methods, Section 2.1) and that our previous work (Rasheed, Mao et al. 2005) suggested that most of the complexes that we isolated represented functionally active rather than randomly aggregated proteins. Our classification accuracies (Table 1) also confirm this notion. For these reasons we hypothesize that the complexes obtained in our experiment are in fact functionally related and that the protein pairs we have identified interact. Given this hypothesis, we can now extrapolate from the pairs identified in the experiment to a proteome wide interaction network for cat.

To obtain our inferred feline protein interaction network, we first downloaded sequence information for all available (569) feline proteins from Swiss-Prot/TrEMBL (Boeckmann, Bairoch et al. 2003). Using our model for relating sequences to interactions (trained on all 69 pairs occurring in 5 complexes as discussed above), we extrapolated to all possible pairs of feline proteins. We used a network drawing and visualization software (Cytoscape) to display and browse our network (Shannon, Markiel et al. 2003). The visualization is shown in Figure 2.

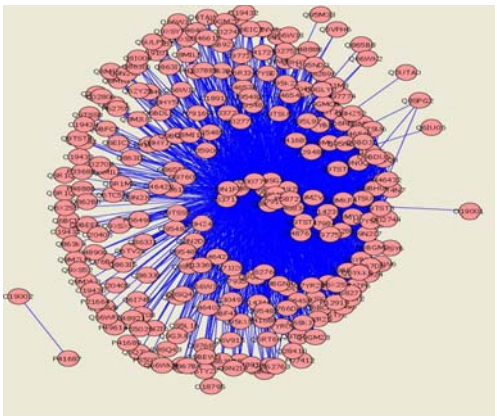

Figure 2: Visualization of the inferred protein-protein interaction network using Cytoscape.

We next examined the degree distribution of our inferred network. This distribution is shown in Figure 3, where we plot log (node degree) vs. log (number of nodes with that degree). The fact that these two quantities appear linearly related in our plot shows that our network has a scale-free degree distribution. This finding supports the validity of our network because it has been observed that biological networks, including protein-protein networks, are scale-free (Barabasi and Albert 1999; Jeong, Tombor et al. 2000; Jeong, Mason et al. 2001).

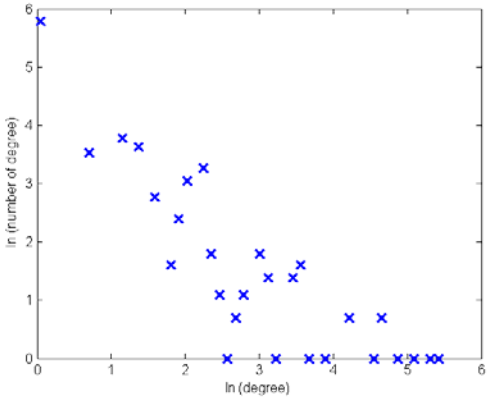

Figure 3: Degree distribution of the inferred interaction network for cat. The linear downward trend in this plot shows the scale-free nature of the degree distribution. This type of distribution indicates that a few proteins (hubs) interact with many other proteins, while most proteins (spokes) interact with only a few other proteins.

We also examined our inferred protein interaction network for biological relevance. Examination of the network has revealed that the most significant interactions in the cat melanoma cells involve signal-transducing proteins. The highest frequencies of interactions in these cells were detected with integrin beta-1 (ITGB1), mass/stem cell growth factor receptor (KIT), protooncogene tyrosine-protein kinase (FES) and zona pellucida sperm-binding protein (ZPB). Integrins are transmenbrane proteins present on the cell surface which interact with the extracellular matrix proteins, playing a critical role in cell proliferation, migration and differentiation. These proteins are also essential for focal adhesion, vesicle transport, regulation of actin cytoskeleton and targeting to other membrane proteins. Our results indicate that integrin beta-1 interacted most significantly with 15 signal transducing proteins including membrane bound kinases, enzymes, phosphatases, growth factors and receptors, which facilitate assembly of distinct complexes that regulate numerous cellular functions.

The KIT and FES protein-tyrosine kinases exhibit significant interactions with 10 and 3 distinct proteins respectively. These proteins are essential for phosphorylation, dephosphorylation, activation and deactivation of a wide range of molecular processes that are critical for cell differentiation and neurogenesis. The highly significant computational interactions of the sperm-binding protein ZPB in experimentally isolated complexes are particularly noteworthy as this domain of interaction has been shown to be required for connecting mechanosensory dendrites to sensory structures (Chung, Zhu et al. 2001).

4. Conclusion

We have performed a computational analysis of the results of an experimental technique that can be used to obtain protein profiles and protein complexes from 2-D gel electrophoresis and MALDI-TOF mass spectrometry. In previous work (Rasheed, Mao et al. 2005) this experiment was used to identify protein profiles showing that the trans-differentiation of melanoma into neuronal cells is directly associated with *de novo* expression of pro-inflammatory cytokines, neuro-regulatory enzymes/kinases, neuro-trophic factors, and concomitant suppression of growth-promoting proteins. In this work, we have provided computational evidence that the protein complexes obtained in the experiment also have biological relevance. We have extrapolated the inferred protein-protein interactions to the entire feline proteome.

The first evidence of this relevance consisted of the results of the derivation of a measure of the probability that a protein pair would occur at random in different groups of proteins (the complexes). Based on our measure, we found 31 protein pairs that occurred in multiple complexes at the .05 confidence level. We next used a previously established method (Martin, Roe et al. 2005) to obtain a relationship between the sequences of the proteins in the protein

pairs and the occurrence of the pairs in the complexes. We showed that we could predict with ~96% accuracy when a protein pair would occur in a complex.

Finally, we extrapolated the protein interactions inferred from the complexes to obtain a protein interaction network for cat. To our knowledge this is the only protein interaction network (inferred or otherwise) available for cat. Another observation in support of the biological relevance of the complexes and of our proposed interactions was the observation of a scale-free degree distribution in our inferred network. Further biological analysis was observed in the inferred network which emphasized the importance of several protein-protein interactions in the context of the original experiment.

5. Acknowledgements

We thank summer students of 2003-2004 (Steve Doo, Chrissy Hsiao, and Alex Lau) for their technical assistance and enthusiasm in working with the cat melanoma cell line. We thank Arpan Patel for analyses of protein complexes and David Wei for literature searches for protein functions. We thank W. Michael Brown for editing this manuscript. The Rasheed Research and Endowment Funds at USC supported this research. Shawn Martin was funded by the US Department of Energy's Genomics: GTL program (www.doegenomestolife.org) under project, 'Carbon Sequestration in *Synechococcus sp.*: From Molecular Machines to Hierarchical Modeling' (www.genomes-to-life.org). Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

6. References

Barabasi, A.-L. and R. Albert (1999). "Emergence of scaling in random networks." Science **286**(5439): 509-512.

Ben-Hur, A. and W. S. Noble (2005). "Kernel methods for predicting protein-protein interactions." Bioinformatics **21 Suppl 1**: i38-i46.

Bennet, K. P. and C. Campbell (2000). "Support vector machines: hype or hallelujah?" ACM SIGKDD Explorations **2**: 1-13.

Bock, J. R. and D. A. Gough (2001). "Predicting protein-protein interactions from primary structure." Bioinformatics **17**(5): 455-60.

Boeckmann, B., A. Bairoch, et al. (2003). "The Swiss-Prot protein knowledgebase and its supplement TrEMBL in 2003." Nucleic Acids Res. **31**: 365-370.

Brown, W. M., S. Martin, et al. (2005). "Prediction of B-strand Packing Interaction using the Signature Product." Journal of Molecular Modeling **12**(3): 355-361.

Burges, C. J. C. (1998). "A tutorial on support vector machines for pattern recognition." Knowl. Discov. Data Mining **2**: 121-167.

Chung, Y. D., J. Zhu, et al. (2001). "nompA encodes a PNS-specific, ZP domain protein required to connect mechanosensory dendrites to sensory structures." Neuron **29**(2): 415-28.

Cristianini, N. and J. Shawe-Taylor (2000). An Introduction to Support Vector Machines. Cambridge, UK, Cambridge University Press.

Dandekar, T., B. Snel, et al. (1998). "Conservation of gene order: a fingerprint of proteins that physically interact." Trends Biochem Sci **23**(9): 324-8.

Enright, A. J., I. Iliopoulos, et al. (1999). "Protein interaction maps for complete genomes based on gene fusion events." Nature **402**(6757): 86-90.

Faulon, J. L., C. J. Churchwell, et al. (2003). "The Signature molecular descriptor. 2. Enumerating molecules from their extended valence sequences." J. Chem. Inf. Comput. Sci. **43**(3): 721-734.

Faulon, J. L., D. P. Visco, Jr., et al. (2003). "The signature molecular descriptor. 1. Using extended valence sequences in QSAR and QSPR studies." J Chem Inf Comput Sci **43**(3): 707-20.

Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." Nature **340**(6230): 245-6.

Giot, L., J. S. Bader, et al. (2003). "A protein interaction map of Drosophila melanogaster." Science **302**(5651): 1727-36.

Goh, C. S., A. A. Bogan, et al. (2000). "Co-evolution of proteins with their interaction partners." J Mol Biol **299**(2): 283-93.

Ho, Y., A. Gruhler, et al. (2002). "Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry." Nature **415**(6868): 180-3.

Jeong, H., S. P. Mason, et al. (2001). "Lethality and centrality in protein networks." Nature **411**(6833): 41-2.

Jeong, H., B. Tombor, et al. (2000). "The large-scale organization of metabolic networks." Nature **407**(6804): 651-4.

Joachims, T. (1999). Making large-scale SVM learning pratical. Advances in Kernel Methods-Support Vector Learning. B. Scholkopf, C. J. C. Burges and A. J. Smola. Cambridge, MA, MIT Press**:** 169-184.

Marcotte, E. M., M. Pellegrini, et al. (1999). "Detecting protein function and protein-protein interactions from genome sequences." Science **285**(5428): 751-3.

Martin, S., D. C. Roe, et al. (2005). "Predicting protein-protein interactions using signature products." Bioinformatics **21**(2): 218-226.

McAllister, R. M., M. Nicolson, et al. (1972). "C-type virus released from cultured human rhabdomyosarcoma cells." Nat New Biol **235**: 3-6.

Pazos, F. and A. Valencia (2001). "Similarity of phylogenetic trees as indicator of protein-protein interaction." Protein Eng **14**(9): 609-14.

Rasheed, S. (1983). "Characterization of a differentiated cat melanoma cell line." Cancer Res **43**: 3379-3384.

Rasheed, S., Z. Mao, et al. (2005). "Is Melanoma a stem cell tumor? Identification of neurogenic proteins in trans-differentiated cells." J Transl Med **3**(1): 14.

Shannon, P., A. Markiel, et al. (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Res **13**(11): 2498-504.

Sprinzak, E. and H. Margalit (2001). "Correlated sequence-signatures as markers of protein-protein interaction." J Mol Biol **311**(4): 681-92.

Zhu, H., M. Bilgin, et al. (2001). "Global analysis of protein activities using proteome chips." Science **293**(5537): 2101-5.

7. Appendix

Pair	Prot. 1 Name	Abbr.	Acc. #	Prot. 2 Name	Abbr.	Acc. #	p-value
1	Integrin beta-1 precursor	ITGB1	P53713	Mast/stem cell growth factor recept	KIT	Q28889	<.001
2	Integrin beta-1 precursor	ITGB1	P53713	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	<.001
3	Integrin beta-1 precursor	ITGB1	P53713	Transferrin receptor protein 1	TFRC	Q9MYZ3	<.001
4	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	Mast/stem cell growth factor recept	KIT	Q28889	<.001
5	Zona pellucida sperm-binding protei	ZPB	P48834	Integrin beta-1 precursor	ITGB1	P53713	<.001
6	Serum albumin precursor	ALB	P49064	Integrin beta-1 precursor	ITGB1	P53713	<.001
7	Proto-oncogene tyrosine-protein kin	FES	P14238	Integrin beta-1 precursor	ITGB1	P53713	<.001
8	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	Transferrin receptor protein 1	TFRC	Q9MYZ3	<.001
9	Serum albumin precursor	ALB	P49064	Mast/stem cell growth factor recept	KIT	Q28889	<.001
10	Mast/stem cell growth factor recept	KIT	Q28889	Transferrin receptor protein 1	TFRC	Q9MYZ3	<.001
11	Pyruvate kinase, M1 isozyme	PKM2	P11979	Integrin beta-1 precursor	ITGB1	P53713	<.001
12	Integrin beta-1 precursor	ITGB1	P53713	Aminopeptidase N	APN	P79171	<.001
13	Serum albumin precursor	ALB	P49064	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	<.001
14	Proto-oncogene tyrosine-protein kin	FES	P14238	Mast/stem cell growth factor recept	KIT	Q28889	<.001
15	Sodium/calcium exchanger 1 precurso	NCX1	P48767	Integrin beta-1 precursor	ITGB1	P53713	<.001
16	Pyruvate kinase, M1 isozyme	PKM2	P11979	Mast/stem cell growth factor recept	KIT	Q28889	0.002
17	Sodium/calcium exchanger 1 precurso	NCX1	P48767	Mast/stem cell growth factor recept	KIT	Q28889	0.002
18	Proto-oncogene tyrosine-protein kin	FES	P14238	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	0.007
19	Zona pellucida sperm-binding protei	ZPB	P48834	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	0.007
20	Zona pellucida sperm-binding protei	ZPB	P48834	Mast/stem cell growth factor recept	KIT	Q28889	0.007
21	Integrin beta-1 precursor	ITGB1	P53713	Alkaline phosphatase, tissue-nonspe	ALPL	Q29486	0.007
22	Zona pellucida sperm-binding protei	ZPB	P48834	Serum albumin precursor	ALB	P49064	0.016
23	Beta-glucuronidase precursor	GUSB	O97524	Integrin beta-1 precursor	ITGB1	P53713	0.016
24	Glutamate decarboxylase, 67 kDa iso	GAD67	P14748	Integrin beta-1 precursor	ITGB1	P53713	0.016
25	Aminopeptidase N	APN	P79171	Mast/stem cell growth factor recept	KIT	Q28889	0.016
26	Zona pellucida sperm-binding protei	ZPB	P48834	Transferrin receptor protein 1	TFRC	Q9MYZ3	0.016
27	Mast/stem cell growth factor recept	KIT	Q28889	Interleukin-1 beta convertase precu	CASP1	Q9MZV6	0.016
28	Integrin beta-1 precursor	ITGB1	P53713	Cathepsin W precursor	CTSW	Q9TST1	0.016
29	Lysosomal alpha-mannosidase precurs	MANB	O46432	Integrin beta-1 precursor	ITGB1	P53713	0.038
30	Integrin beta-1 precursor	ITGB1	P53713	Toll-like receptor 4 precursor	TLR4	P58727	0.038
31	Serum albumin precursor	ALB	P49064	Transferrin receptor protein 1	TFRC	Q9MYZ3	0.038

Protein pairs occurring in multiple complexes. Probabilities of occurrence (p-values) are computed using the assumption that a protein pair is significant if it occurs in 11 complexes (.05 level for $C_{m,8}$).