

Tracking the Structure of Protein Interaction Network via Multiple Genetic Perturbations on Mouse Embryonic Stem Cells — Implementation of the Entropy Maximization Principle

Lei Mao¹, Rossella De Cegli², Mario Lauria², Grit Nebrich¹,
Jean Maurice Delabar³, Yann Herault⁴, Gilda Cobellis² and Joachim Klose¹

¹*Institute for Medical Genetics, Charité – Universitätsmedizin, Berlin,*

²*Telethon Institute of Genetics and Medicine, Napoli,*

³*Université Paris, Paris,*

⁴*Institut de Génétique Biologie Moléculaire et Cellulaire, IGBMC, Institut clinique de la Souris, ICS, CNRS, INSERM, Université de Strasbourg, UMR7104, UMR964, Illkirch*

¹*Germany*

²*Italy*

^{3,4}*France*

1. Introduction

The Human Genome Project (HUGO) opened a new era for biomedical research. However, a decade after the completion of HUGO, we are still standing at the very beginning regarding the understanding of cellular dynamics under development, disease and aging. To bridge between genotype and phenotype, it is imperative to track the underlying genetic interaction network, and perhaps more importantly, the protein-protein interaction (PPI) network, which forms the basal layer of the majority of fundamental cellular processes [Antal et al., 2009].

Current approaches for protein-protein interaction network structure detection include gene fusion experiments such as yeast two-hybrids, co-occurrence evidences via immunoprecipitation, as well as literature or orthology-based knowledge extrapolation [Stelzl and Wanker, 2006]. Albeit powerful and foreseeing incremental technical improvements, these methods either rely heavily on experimental detection thresholds, or could be biased as they mirror the focus of individual fields of research. Hence, it is vital to path-find the underlying cellular PPI network with hypothesis-free (*de novo*) methods [Janes and Yaffe, 2006].

A biological organism can be perceived as a complex system with a set of interacting elements. Basically, when a meta-stable system consisting of steady number of interacting elements is disturbed systematically, one is able to use the system outcome to puzzle back the underlying network structure. This is commonly known as “reverse engineering” in system sciences [Bansal et al., 2007].

Previously, the use of such reverse engineering approaches on the exploration of cellular PPI network has been hampered by limited amount of high confident expression profiling data.

Nowadays, however, techniques for the probing of gene expression profile have been witnessing swirling advances. mRNA microarray, mRNA-Seq, protein antibody chips, as well as high resolution protein electrophoresis combined with mass spectrometry have been catching-up rapidly to become high throughput and high confident methodologies. These solid technical platforms promote the use of experimental data in sophisticated *de novo* inferring of cellular interaction networks. Recently, such approach has been applied on diverse complex systems [Shmulevich et al., 2002]. In a pioneering study of di Bernardo et. al., the reverse engineering approach was successfully applied on yeast to prioritize the gene products and cellular pathways involved in drug responses [di Bernardo et al., 2005].

It should be noted that all current reverse engineering approaches have practical limitations. For example, Bayesian methods need a sufficiently large number of experiments comparable to the number of genes. Undersampled datasets will not be adequate to generate confident outcome. Multiple regression methods work well only when the identity of the perturbed genes is known, or can be reliably inferred from the data [Lauria et al., 2009]. Fortunately, this methodological constraint of gene-by-gene perturbation can be partially relaxed if entropy maximization principle is invited into play. The entropy maximization principle was based on the information theory of Shannon [Shannon, 1948]. By this mean, the most parsimonious PPI interaction network structure that is able to give rise to experimental expression profiles can be probed out despite non-complete system perturbation. This principle of entropy maximization has been proved useful in inferring the genetic and signaling networks of peripheral nerve development and yeast chemostat [Dhadialla et al., 2009; Lezon et al., 2006].

Nevertheless, in multicellular organism and especially in higher eukaryotes, each differentiated cellular system can exhibit alternative PPI networks. This has been prohibiting the probing of mammalian interaction network in a large extent. Embryonic stem (ES) cells can be considered as the starting point of multicellular life. Apart from the high relevance to biomedical research, the ES cells are homogeneous self-maintaining systems. Upon a genetic mutation, either an ES cell will balance the effect of genetic insult and remain in their pluripotent state, or else undergo cell death or differentiation should the impact of the genetic mutation surpasses a certain threshold [Amit et al., 2000; Mao et al., 2007]. Such "all or none" cell faith decision constitutes a solid genomic and proteomic stability, hence put forward the ES cells as a brilliant model system for *de novo* exploration of the PPI network structure.

In this study, we applied a hybrid approach of multiple genetic perturbations and entropy maximization principle on the mouse embryonic stem cells to probe the underlying cellular protein-protein interaction network structure. Our results suggest the essential role of antioxidative response, transcriptional regulation and protein degradation pathways in the rebalancing of the proteomic network. Moreover, our predicted protein interaction data indicates that overexpression of Ripk4, a protein kinase originally located on human chromosome 21, could severely perturb the cellular cholesterol syntheses pathways via an adenosyl-homocystein mediated mechanism. Our study demonstrates the value of mouse ES cells in protein network research in the context of system biology.

2. Methods

Mouse embryonic stem (ES) cells were genetically perturbed with a pallet of 14 single genes overexpressed individually in different subclones. We studied the impacts of these genetic

system stimuli on the PPI network at both transcriptomic and proteomic levels. The workflow of the current study is summarized in figure 1.

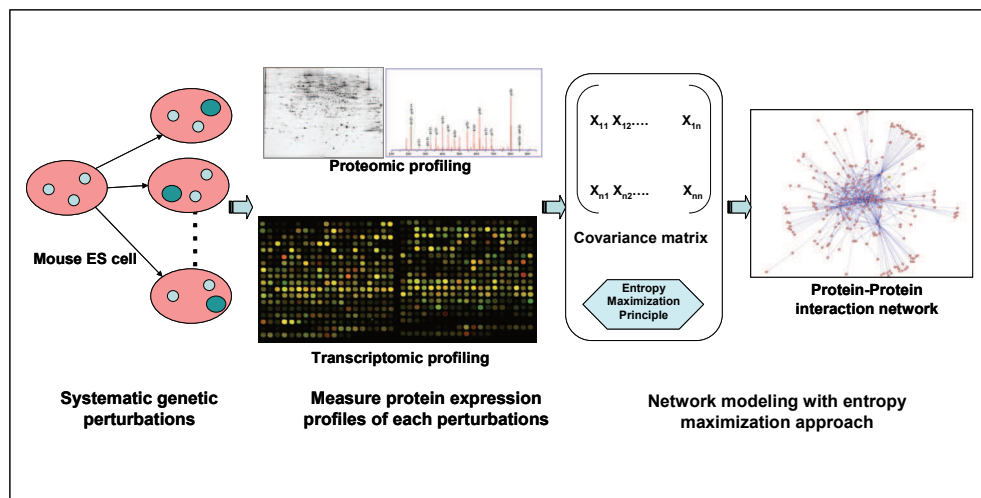


Fig. 1. Workflow of the current study. Mouse embryonic stem cells were genetically perturbed with a pallet of 14 single gene overexpressions. The impact of the system stimuli on the ES cell PPI network were readout with proteomic and mRNA microarray analyses. Gene expression profiling data undergo entropy maximization approach to reverse engineer the underlying PPI network structure (picture of microarray from Wikipedia)

2.1 Systematic genetic perturbations on mouse ES cells

This is a meta-study on a pallet of 14 transgenic mouse ES cell lines constructed in the frame of our European Research Grant “AnEUploidy”. Please refer to our previous publications for detailed methodology for the generation of transgenic cell lines and expression profiling [De Cegli et al., 2010; Mao et al., 2007]. Except for Snca-overexpression cell line, each transgenic ES cell line overexpresses one of the human chromosome 21 genes (Table 1). The impacts of these genetic system perturbations on the ES cells were examined at the proteomic and transcriptomic levels: All cell lines were measured by large-gel two-dimensional protein electrophoresis [Klose, 1999], whereas 8 of the cell lines were measured by mRNA microarray (Affymetrix GeneChip Mouse Genome 430_2 array) [De Cegli et al., 2010].

2.2 Theoretical framework of entropy maximum principle

We applied the principle of entropy maximization to identify the protein interaction network structure with the highest probability of giving rise to our experimentally observed proteomic and transcriptomic data [Lezon et al., 2006]. This relies on the Boltzmann’s concept of entropy maximization. In effect, as the number of ways of realizing a given macroscopic state can vary widely, the most likely state of the system is the one that corresponds to the largest number of microscopic states, or biggest entropy. This idea of entropy maximization provides a natural mechanism for revealing dominant protein-protein interaction structure inside of a quasi-stable system.

Gene Symbol	Protein name	Category	Original gene location
Aire	Autoimmune regulator	Transcription factor	21q22.3
Erg	Avian erythroblastosis virus E-26 (v-ets) oncogene related		21q22.3
Nrip1	Nuclear receptor interacting protein 1		21q11.2
Olig2	Oligodendrocyte transcription factor 2		21q22.11
Runx1	Runt related transcription factor 1		21q22.3
Sim2	Single-minded homolog 2		21q22.13
Pdxk	Pyridoxal (pyridoxine, vitamin B6) kinase	Protein kinase	21q22.3
Ripk4	Receptor-interacting serine-threonine kinase 4		21q22.3
Dyrk1A	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A		21q22.13
Hunk	Hormonally up-regulated Neu-associated kinase		21q22.1
Dopey2	Dopey family member 2		21q22.2
App	Amyloid beta (A4) precursor protein		21q21.3
Snca	Synuclein, alpha		4q21
MirLet7Cdel 99a	---	Micro RNA	21q21.1
Mir99adelLet 7C	---		21q21.1
Mir802	---		21q22.12

Table 1. Characteristics of the transgenes used to genetically perturb the mouse ES cell system

2.3 Data-driven *de novo* PPI network exploration

We consider each genetic modification as a distinct stimulus on the ES cell protein interaction network. In this sense, the protein expression profile of each transgenic cell line is a sampling on the perturbed system. Given a collection of \mathbf{n} genes, their expression profiles in a pellet of \mathbf{k} perturbation experiments can be represented by a $\mathbf{k} \times \mathbf{n}$ expression matrix. Here, each row vector of this expression matrix is the expression alteration (ratio of transgenic to control) of a give gene under different perturbations. We generated the expression matrixes from the proteomic and microarray data, respectively.

In the second step, we determined the $\mathbf{n} \times \mathbf{n}$ expression covariance matrix \mathbf{M} , whose entry M_{ij} is the expression covariance of gene X_i and X_j according to the following equation:

$$M_{ij} = Cov(X_i, X_j) = \sum_{m=1}^k \frac{(X_{i,m} - \overline{X_i}) \cdot (X_{j,m} - \overline{X_j})}{k} \quad (1)$$

where \mathbf{m} stands for index of genetic perturbation experiments (in our context: measurement of each transgenic cell line, and \mathbf{k} denotes the total number of perturbations ($\mathbf{k}=14$ for proteomic data, $\mathbf{k}=8$ for transcriptomic data)).

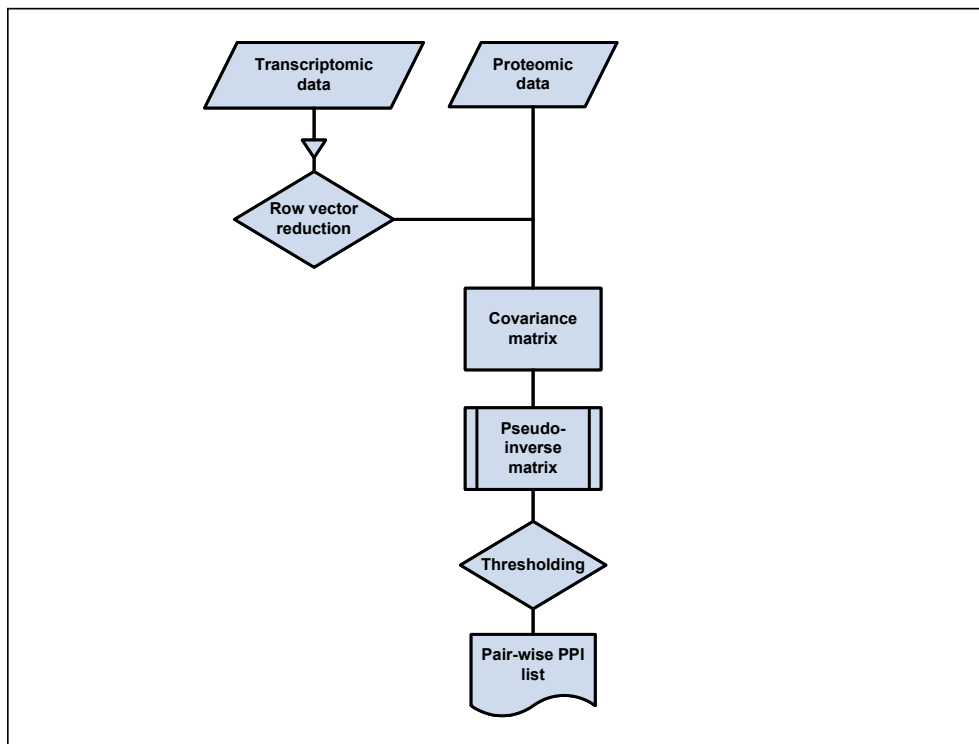


Fig. 2. Programming algorithms for the de novo inferring of protein-protein interaction information from protein expression data of multiple genetic perturbation experiments on mouse ES cells

According to the maximum entropy hypothesis, the corresponding pseudo-inverse matrix of the expression covariance matrix represents the best fit of the underlying protein-protein interaction network, while maximizing the entropy of the system. The strength of interactions is expressed by the value of M_{ij}^{-1} . Positive interaction values indicate the cis-action of the two genes. For instance, an up-regulation of gene A will lead to up-regulation of gene B, and *vice versa*. In contrast, negative values indicate the gene-gene trans-action. Subsequently, the pair-wise protein interaction information was extracted from the pseudo-inverse matrix of the covariance matrix. By thresholding this PPI information, we obtained the dominant PPI network information from the experimental data solely, without any prior information of the genes.

The algorithm described above was implemented in a Java application. This Java program takes the input data of the expression matrix, and process it with three consecutive modules: In module one, the expression matrix was reduced in terms of redundant gene symbols. Gene expression profiles bearing the same gene symbol, but of different probeset identity,

were averaged. In order to increase fidelity, only genes with multiple probesets were retained for further data evaluation. In module two, the covariance matrix was calculated according to equation 1. Subsequently, the pseudoinverse matrix of the covariance matrix was calculated using the BlueBit (<http://www.bluebit.gr/matrix-calculator/>) or ARARCNE software (for microarray data) [Margolin et al., 2006]. Finally, the last module extracts the PPI information in form of a list of pair-wise interaction: Gene A - Gene B, together with their corresponding edge weight obtained from the covariance values. We set the threshold for interaction to consider only those protein interactions with weight over 0.5 as significant. The overview of the programming structure is given in figure 2.

2.4 Investigation on the obtained PPI network

The protein interaction outcome from our entropy maximization approach was compared to the publicly available PPI databases via three online meta-databases:

- String (<http://string-db.org>),
- UniHI (<http://theoderich.fb3.mdc-berlin.de:8080/unihi>) and
- ConsensusPathDB (<http://cpdb.molgen.mpg.de>).

Together, they encompass a total of 29 individual PPI interaction databases. Network visualization was realized using the Cytoscape software (www.cytoscape.org). The online graph analysis tool CFinder (cfinder.org) was used to grab the community structures (closely interlinked sub-graphs) in our predicted PPI network. This method first locates all cliques of the network and then identifies the communities by carrying out component analysis of the clique-clique overlap matrix [Adamcsek et al., 2006]. Gene Ontology functional enrichment analysis was performed using Webgestalt (bioinfo.vanderbilt.edu/webgestalt).

2.5 Label-free mass spectrometric relative protein quantification

The global protein expression profile of the Ripk4-overexpressing ES cells was additionally monitored by label-free mass spectrometry quantification [Ishihama et al., 2005]. For this purpose, 200 µg of total protein extract from Ripk4-overexpressing and control ES cells (as control we consider the same mouse ES clone in which the expression of the protein Ripk4 is as basal expression level) were first separated by SDS-PAGE in a gel format of 13 x 25 cm. Subsequently, the gel strip was cut into 16 homogeneous gel slides. These gel slides were subjected to in-gel trypsin digestion as previously described [Mao et al., 2010]. LC/ESI-MS/MS was performed on a LCQ Deca XP ion trap instrument (Thermo Finnigan, Waltham, MA, USA). The eluting gradient is formed by 0.1 % (v/v) formic acid (FA) in water as solvent A and 0.1 % (v/v) FA in acetonitrile (ACN) as solvent B and run at a flow rate of 200 nL per minute. The gradient is linear starting with 5 % B increasing to 70 % B in 180 minutes and additional 10 min to 95 % B. ESI-MS data acquisition is performed throughout the LC run. The raw data were extracted by TurboSEQUENT algorithm, trypsin autolytic fragments and known keratin peptides were filtered out. These files were searched using our in-house licensed Mascot Version 2.1 (Matrix Sciences, London, UK). The MS/MS ion searches are performed with the following set of parameters: database = Swiss-Prot, taxonomy = *Mus musculus*, Proteolytic enzyme = trypsin, the maximum of accepted missed cleavages = 1, mass value = monoisotopic, peptide mass tolerance = ±0.8 Da, fragment mass tolerance = ±0.8 Da and as variable modifications oxidation of methionine and acrylamide adducts (propionamide) on cysteine are expected. Only protein identifications with over 4 spectral counts were retained for further analysis.

The “Experimentally Modified Protein Abundance Index”, or “emPAI”, is a indexing value on the relative quantification of the proteins in a protein mixture [Ishihama et al., 2005]. The of emPAI ratios of the identified proteins from Ripk4-overexpression to that of control were used as a relative indication of protein expression alteration in ES cells bearing Ripk4 overexpression. We used the threshold of emPAI ratio >2 or emPAI ratio < 0.5 as an arbitrary cut-off for up-regulation or down-regulation of proteins, respectively.

3. Results

3.1 De novo protein interaction network attained from experimental data

We performed the entropy maximization based PPI exploration approach on our experimental data. For this purpose, we combined the proteomic and transcriptomic expression data of the transgenic mouse ES cell lines, whose protein expression profile have been investigated previously in our laboratories. This built up a 14×690 expression matrix at the proteomic data level (8×8383 for transcriptomic data), where each row vector of the

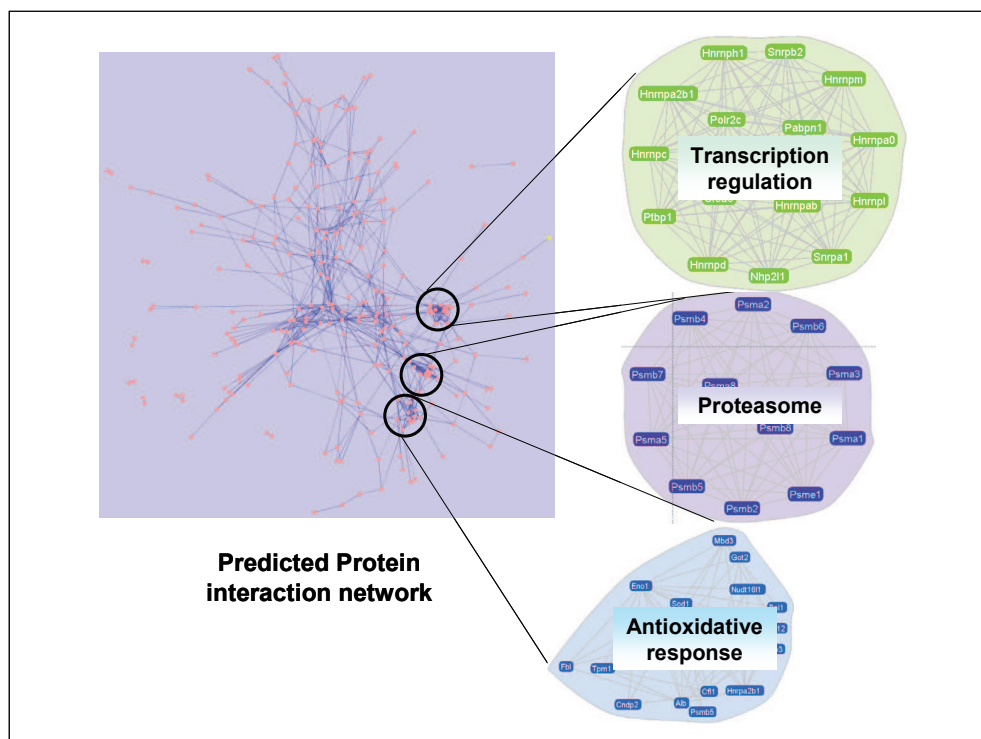


Fig. 3. Global network of protein interactions predicted from our experimental data using the entropy maximization approach. Three major community structures could be detected, which correspond to transcriptional regulation, proteasome, and antioxidative response processes. Nodes represent proteins, edges represent predicted functional interaction. Illustrated using Cytospace

expression matrix represents the expression ratio (transgene vs. parental control lines) of a distinct protein under divergent single gene overexpressions.

In order to obtain the protein co-regulation data, we calculated the covariance matrix of these genes using the self-implemented Java program. Following equation 1, the program returns a 690x690 (8383 x 8383 for transcriptomic data) symmetrical matrix, with each element representing the covariance of gene X_i and X_j . This matrix contains information of the co-action of genes under genetic system perturbation. Here, the covariance values were normalized to dimensionless values.

Through the pseudo-inverse matrix calculation, we generated a PPI network encompassing distinct proteins according to gene symbol. By taking the threshold of ± 0.5 , we obtained a matrix where 22,206 elements were non-zero. This reduces the number of genes showing significant interaction to 490. Specifically, five genes (Cdv3, Fbl, Got2, Hspb1 and Set) showed self inhibitory effect. The obtained PPI network is illustrated in figure 3.

Figure 4 shows the sub-graph build by the 50 most significant pair-wise protein interactions. Among them, 21 gene-pairs showed strong positive interaction, whereas 29 gene-pairs showed mutual inhibitory effect. However, the exact nature of inhibition is not to be deduced from our data.

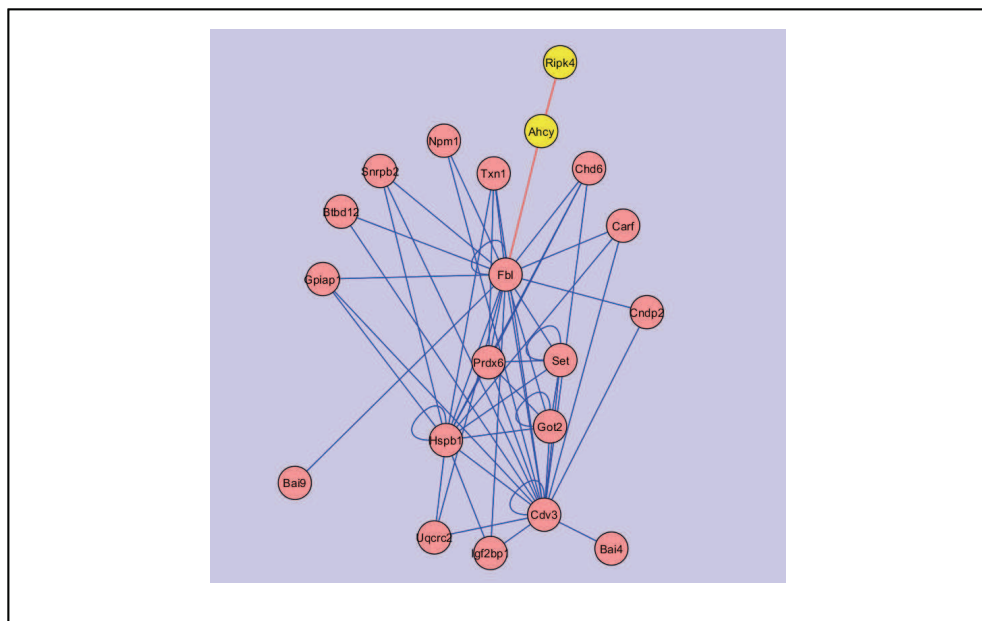


Fig. 4. Sub-graph showing the strongest pair-wise protein interactions in ES cells predicted by multiple genetic perturbations. Our result suggests that Ahcy represents one of the Ripk4 targets

3.2 Community structure analysis revealed key pathways involved in system rebalancing

In the next step, we investigated the community structures (sub-networks) inside our predict protein-protein interaction network. Community structures (CS) can be loosely defined as subsets of nodes that are more densely interconnected among each other than with the rest of

the network [Newman and Girvan, 2004]. As shown in figure 3, three dominant community structures could be discovered in our predicted PPI network of ES cell. Respecting their molecular function, protein nodes of these community structures are representative for transcriptomic regulation, proteasome, stress-response pathways, respectively.

3.3 Predicted High degree nodes are reminiscent to “balancer” proteins

As can be seen in figure 3, our predicted PPI network demonstrates the small-worldness property and scale-free degree distribution. Specifically, the network nodes of our predicted PPI network vary substantially in their connectivity, with a small number of proteins exhibiting strong pair-wise interactions with many other genes. Table 2 lists the 20 protein nodes with more than three direct interaction partners.

Node	Degree	Overlap to Balancer proteins
Taldo1	3	Yes
Nudt16l1	3	Yes
Cndp2	3	
RGD1308600	3	
NSFL1C	4	
Cdv3	4	
Prdx6	4	
Psmb5	4	
Fbl	5	
Bai1	5	
Got2	5	Yes
Tpm1	7	Yes
Cfl1	8	
Mbd3	9	
Npm1	10	Yes
Alb	13	Yes
Mis12	13	
Hnrpa2b1	13	Yes
Eno1	15	Yes
Sod1	19	

Table 2. High degree nodes in the predicted PPI network show significant overlap to our previously documented “balancer” proteins

In our previous communication, we reported our hypothesis that upon genetic perturbations on mouse ES cells, there are so-called “balancer” proteins, defined as proteins that buffer or cushion the system, that act against the system stimuli [Mao et al., 2007]. These central network elements could mediate network remodeling upon perturbation [Bode et al., 2007]. Consequently, we presume that part of these proteins with hub-like behavior could

have system functionalities to prevent from severe proteomic shift. Indeed, among the 20 high degree nodes in our predicted PPI network (degree ≥ 3), eight of the proteins belong to our previously detected “balancer” proteins. A Gene Ontology functional enrichment analysis revealed their close involvement in RNA-binding, oxidative response and cellular transport processes (Figure 5).

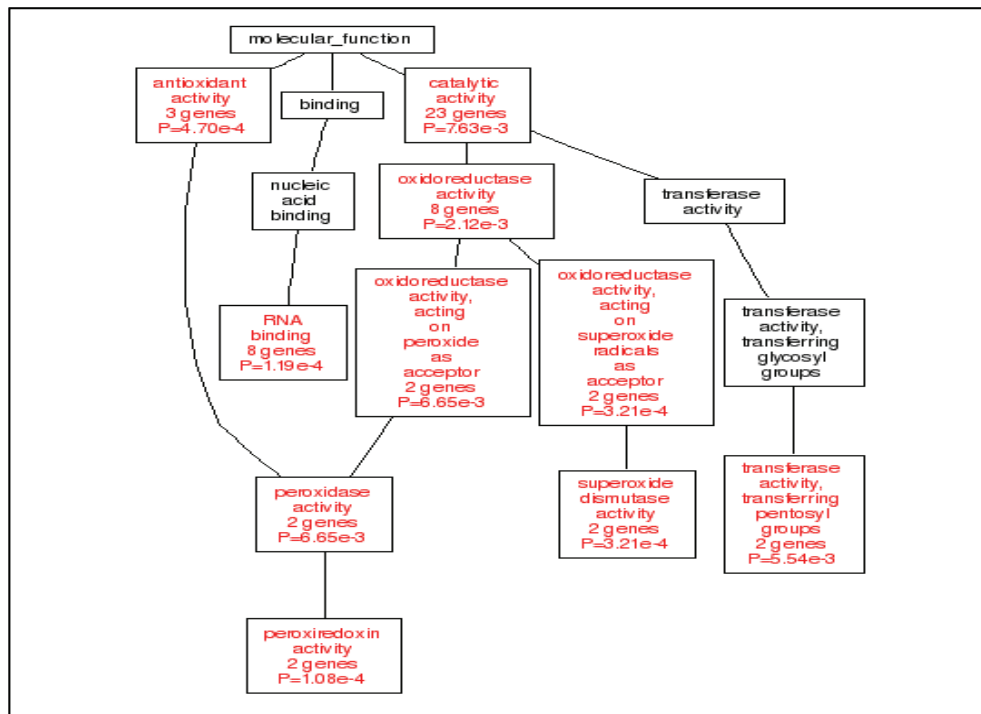


Fig. 5. Gene Ontology analysis revealed molecular function terms enriched in high degree protein nodes in our predicted PPI network

3.4 Comparison of our result to public PPI data reveals overlaps and novel predictions

In light of a comparison between our predicted PPI networks to publically available PPI databases, twelve direct pair-wise interactions predicted by our data-driven entropy maximization approach could be validated by public PPI databases (Table 3). For instance, one important protein interaction partner of the transcription factor single-minded homolog 2 (Sim2), the Ttc3 (tetratricopeptide repeat domain 3), was also predicted by our approach. However, another commonly known PPI partner of Sim2, aryl hydrocarbon receptor nuclear translocator (Arnt), did not appear in our list of predicted PPI partners of Sim2, although Arnt is present in the probset list of the microarray analysis.

In addition, 18 predicted links are similar to that documented in public databases (Table 4). For example, UniHI predicted the protein interaction of Fbl to many proteasome subunits including Psma4, Psma6, Psma7, Psma8, Psmb6 and Psmd8bp1. Here, we attempted to enrich this collection with an additional proteasome subunit, Psmb5, as a potential PPI

partner for Fbl. Moreover, the pair-wise interactions Ahcy–Npm1 was predicted by our data. This is coherent to the UniHI protein interaction database, which has documented an indirect PPI relation of “Ahcy–Fbl–Nmp1”.

Direct pair-wise protein interaction	Source database
Acad8 – Hadh	String
Cfl1 – Tmp1	String
Prdx1 – Sod2	String
Sod1 – Sod2	String
Ahcy – Sod2	ORTHO
Eno1 – Eno1	ORTHO
Hnra2b1 – Snrpa1	Reactome
Hnra2b1 – Snrpb	Reactome
Hspb1 – Hspb1	IntAct
Mis12 – Mis12	HPRD-Binary
Ahcy – Mtap	String
Sim2 – Ttc3	String

Table 3. Consensus direct pair-wise interactions between our predicted PPI and public available PPI databases

Our prediction	Previous documentation	Source Database
Alb – Slc8A3	Alb – Slc1A5, Slc25A13, Slc9A8	IntAct
Cfl1 – Got2	Cfl1 – Got1	OPHID, ORTHO
Cfl1 – Hspb1	Cfl1 – Hsph1	HPRD-Binary, CCSB-LIT
Cfl1 – Psmb5	Cfl1 – Psme4	ORTHO
Eno1 – Psmb5	Eno1 – Psm2	ORTHO
Fbl – Psmb5	Fbl – Psma4, Psma6, Psma7, Psma8, Psmb6 and Psm8bp1	ORTHO, OPHID, BioGrid, HPRD-Binary
Fbl – Rpl23A	Fbl – Rpl30, Rpl4, Rpl6, Rpl8, Rplp0, Rplp2	HPRD-Complex, OPHID, ORTHO
Fbl – Snrpb2	Fbl – Snrpn	BioGrid, HPRD-Binary
Got2 – Psmb5	Got2 – Psm10	ORTHO
Hnra2b1 – Rbm14	Hnra2b1 – Rbm5, Rbm8A	Reactome
Hnra2b1 – Txn1	Hnra2b1 – Txndc10, TxnL4A	ORTHO, Reactome
Ahcy – Got2	Ahcy – Got1	ORTHO, OPHID
Ahcy – Prdx6	Ahcy – Prdx1, Prdx2, Prdx4	ORTHO
Got2 – Prdx6	Got2 – Prdx5	ORTHO
Pdxk – Kcna7	Pdxk – Kcna1	String
Pdxk – Prkg2	Pdxk – Prkab	String
Pdxk – Zfp469	Pdxk – Zfp295	String
Ahcy – Npm1	Ahcy – Fbl – Npm1	OPHID, ORTHO

Table 4. Similar protein pair-wise interactions between our predicted PPI and public available databases

Similarly, our approach predicted three PPI partners for Pdxk: Kcna7 (potassium voltage-gated channel, member 7), Prkg2 (Protein kinase cGMP dependent Type II), and Zfp469 (Zinger finger protein 469), whereas analogous protein PPI partners have been documented in public PPI databases: Kcnma1 (potassium large conductance calcium-activated channel, subfamily M, alpha member 1), Prkab (protein kinase, AMP-activated) and Zfp295 (Zinger finger protein 295), respectively. Such overlaps increase the confidence of our predicted interaction list, and thus support the entropy-maximization approach as a useful method for the *de novo* PPI prediction.

4. Discussion

We predicted a PPI network that contains a list of possible pair-wise protein interactions in mouse ES cells. Notice that this PPI network resulted solely from the gene co-regulation experiments under multiple genetic perturbations. This demonstrates the usefulness of ES cells as a uniform, standardized cell system for system perturbation experiments. Intrinsically, our predicted PPI information is by default weighted, which infers the strength and nature of protein-protein interaction. This could be more superior to some other experimental approaches.

In contrast to random network, the presence of community structures in our predicted PPI network is a signature of the hierarchical nature of intrinsic cellular PPI network. Being able to identify such community structures could help us explore the interplay inside the networks upon genetic perturbation. It should be noted that such perturbation approach reveals predominantly those part of the network structure that is affected by the system stimuli. Thus, the detected community structures in our predicted PPI network reflect the most significantly involved cellular pathways under genetic mutations.

In addition, our data predicted significant protein-protein trans-interaction between Ripk4, a protein kinase originally located on human chromosome 21, and Ahcy (S-adenosylhomocystein hydrolase). This could represent novel knowledge. Unfortunately, Ahcy was not revealed by the proteomic analysis, whereas the expression profile of Ahcy in microarray analysis was heterogeneous among different probesets. In order to in-depth analyse this issue, we performed an additional proteomic analysis. Indeed, using the label-free mass spectrometry protein quantification, we observed a 53 % concentration decrease in the Ripk4-overexpression ES cells. This supports our predicted PPI relation between Ahcy and Ripk4.

It has been reported previously that in rats fed by the Ahcy enzyme inhibitor, the total plasma cholesterol level decreases significantly [Yamada et al., 2007]. Moreover, drug-induced Ahcy inhibition can also lead to anemia due to low erythrocyte membrane fluidity [Altintas and Sezgin, 2004]. However, the direct link between Ripk4 and Ahcy has not been documented explicitly so far. Deduced from our *de novo* inferred PPI network topology, the overexpressing of Ripk4 could inevitably lead to the inhibition of Ahcy enzyme activity, which in turn leads to steroid metabolism disturbance.

In line with this, it has been previously shown that drug induced blocking of sterol conversion to cholesterol in *C. elegans* causes serious defect in germ cell development and motor function [Choi et al., 2003]. This suggests the significance of cholesterol synthesis in neuronal function.

Indeed, an additional *post hoc* functional analysis on the expression profile of Ripk4 transgenic ES cells showed significant down-regulation in proteins involved in lipid metabolism. This includes several key nuclear receptors such as retinoic acid receptor, retinoid X receptor, peroxisome proliferators-activated receptor and steroid hormone receptor ERR2.

Taken together, we may hypothesize that the overexpression of Rlpk4 severely inhibits Ahcy's activity. Moreover, this interaction may be important for the cholesterol synthesis pathways. How the interaction of Rlpk4 with Ahcy could be correlated to Down syndrome pathology and neuronal dysfunction need to be further investigated.

5. Conclusion

In conclusion, this study demonstrated the feasibility of *de novo* tracking the structure of protein interaction network using a combined experimental and entropy-maximization approach using mouse ES cells as a model system. Albeit useful, this approach also has some limitations: Firstly, biological organisms are rather complex systems with vast number of system components. This engenders high experimental and calculation workloads. Secondly, the entropy maximization approach applied in this study considers only pair-wise protein-protein interaction. More compound interactions, such as triple node interactions or loop effects, which could also be relevant for the cellular PPI network, are not considered. Finally, it needs to be noted that like all other data-dependent modeling approaches, the performance of this method is highly dependent on data quality. In particular, too much of system reaction such as oscillation leads to high system noise, and can deleteriously influence the modeling outcome. In this sense, mouse ES cells could represent a warrant mammalian cell model of such system biological approaches due to their homogeneous and stable system behavior.

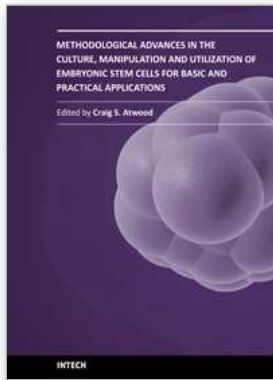
6. Acknowledgements

This Study was co-supported by the European Research Grant 37627 and the German Research Society (DFG) grant KL237/12-1.

7. References

- Adamcsek, B., Palla, G., Farkas, I. J., Derenyi, I., and Vicsek, T. (2006). CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 22, 1021-1023.
- Altintas, E., and Sezgin, O. (2004). S-adenosylhomocysteine hydrolase, S-adenosylmethionine, S-adenosylhomocysteine: correlations with ribavirin induced anemia. *Med Hypotheses* 63, 834-837.
- Amit, M., Carpenter, M. K., Inokuma, M. S., Chiu, C. P., Harris, C. P., Waknitz, M. A., Itskovitz-Eldor, J., and Thomson, J. A. (2000). Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol* 227, 271-278.
- Antal, M. A., Bode, C., and Csermely, P. (2009). Perturbation waves in proteins and protein networks: applications of percolation and game theories in signaling and drug design. *Curr Protein Pept Sci* 10, 161-172.
- Bansal, M., Belcastro, V., Ambesi-Impiombato, A., and di Bernardo, D. (2007). How to infer gene networks from expression profiles. *Mol Syst Biol* 3, 78.
- Bode, C., Kovacs, I. A., Szalay, M. S., Palotai, R., Korcsmaros, T., Csermely, P., Szalay, M. S., Kovacs, I. A., Korcsmaros, T., Bode, C., and Csermely, P. (2007). Network analysis of protein dynamics. *FEBS Lett* 581, 2776-2782.
- Choi, B. K., Chitwood, D. J., and Paik, Y. K. (2003). Proteomic changes during disturbance of cholesterol metabolism by azacoprostane treatment in *Caenorhabditis elegans*. *Mol Cell Proteomics* 2, 1086-1095.

- De Cegli, R., Romito, A., Iacobacci, S., Mao, L., Lauria, M., Fedele, A. O., Klose, J., Borel, C., Descombes, P., Antonarakis, S. E., et al. (2010). A mouse embryonic stem cell bank for inducible overexpression of human chromosome 21 genes. *Genome Biol* 11, R64.
- Dhadialla, P. S., Ohiorhenuan, I. E., Cohen, A., and Strickland, S. (2009). Maximum-entropy network analysis reveals a role for tumor necrosis factor in peripheral nerve development and function. *Proc Natl Acad Sci U S A* 106, 12494-12499.
- di Bernardo, D., Thompson, M. J., Gardner, T. S., Chobot, S. E., Eastwood, E. L., Wojtovich, A. P., Elliott, S. J., Schaus, S. E., and Collins, J. J. (2005). Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat Biotechnol* 23, 377-383.
- Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 4, 1265-1272.
- Janes, K. A., and Yaffe, M. B. (2006). Data-driven modelling of signal-transduction networks. *Nat Rev Mol Cell Biol* 7, 820-828.
- Klose, J. (1999). Large-gel 2-D electrophoresis. *Methods Mol Biol* 112, 147-172.
- Lauria, M., Iorio, F., and di Bernardo, D. (2009). NIREst: a tool for gene network and mode of action inference. *Ann N Y Acad Sci* 1158, 257-264.
- Lezon, T. R., Banavar, J. R., Cieplak, M., Maritan, A., and Fedoroff, N. V. (2006). Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc Natl Acad Sci U S A* 103, 19033-19038.
- Mao, L., Romer, I., Nebrich, G., Klein, O., Koppelstatter, A., Hin, S. C., Hartl, D., and Zabel, C. (2010). Aging in mouse brain is a cell/tissue-level phenomenon exacerbated by proteasome loss. *J Proteome Res* 9, 3551-3560.
- Mao, L., Zabel, C., Herrmann, M., Nolden, T., Mertes, F., Magnol, L., Chabert, C., Hartl, D., Herault, Y., Delabar, J. M., et al. (2007). Proteomic shifts in embryonic stem cells with gene dose modifications suggest the presence of balancer proteins in protein regulatory networks. *PLoS One* 2, e1218.
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7 Suppl 1, S7.
- Newman, M. E., and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 69, 026113.
- Shannon, C. E. (1948). Prediction and entropy of printed English. *The Bell System Technical Journal* 30, 50-64.
- Shmulevich, I., Dougherty, E. R., Kim, S., and Zhang, W. (2002). Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18, 261-274.
- Stelzl, U., and Wanker, E. E. (2006). The value of high quality protein-protein interaction networks for systems biology. *Curr Opin Chem Biol* 10, 551-558.
- Yamada, T., Komoto, J., Lou, K., Ueki, A., Hua, D. H., Sugiyama, K., Takata, Y., Ogawa, H., and Takusagawa, F. (2007). Structure and function of eritadenine and its 3-deaza analogues: potent inhibitors of S-adenosylhomocysteine hydrolase and hypocholesterolemic agents. *Biochem Pharmacol* 73, 981-989.



Methodological Advances in the Culture, Manipulation and Utilization of Embryonic Stem Cells for Basic and Practical Applications

Edited by Prof. Craig Atwood

ISBN 978-953-307-197-8

Hard cover, 506 pages

Publisher InTech

Published online 26, April, 2011

Published in print edition April, 2011

Pluripotent stem cells have the potential to revolutionise medicine, providing treatment options for a wide range of diseases and conditions that currently lack therapies or cures. This book describes methodological advances in the culture and manipulation of embryonic stem cells that will serve to bring this promise to practice.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Lei Mao, Rossella De Cegli, Mario Lauria, Grit Nebrich, Jean Maurice Delabar, Yann Herault, Gilda Cobellis and Joachim Klose (2011). Tracking the Structure of Protein Interaction Network via Multiple Genetic Perturbations on Mouse Embryonic Stem Cells — Implementation of the Entropy Maximization Principle, Methodological Advances in the Culture, Manipulation and Utilization of Embryonic Stem Cells for Basic and Practical Applications, Prof. Craig Atwood (Ed.), ISBN: 978-953-307-197-8, InTech, Available from: <http://www.intechopen.com/books/methodological-advances-in-the-culture-manipulation-and-utilization-of-embryonic-stem-cells-for-basic-and-practical-applications/tracking-the-structure-of-protein-interaction-network-via-multiple-genetic-perturbations-on-mouse-em>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.