BIOINFORMATICS ORIGINAL PAPER

Vol. 27 no. 24 2011, pages 3407–3414
doi:10.1093/bioinformatics/btr580

Systems biology

Advance Access publication October 20, 2011

A system-level approach for deciphering the transcriptional

Mattia Zampieri[1,2], Giuseppe Legname[3], Daniel Segrè[4,5,6] and Claudio Altafini[2,*]

[1]Institute of Molecular Systems Biology, ETH, Wolfgang-Pauli-Strasse 16, 8093 Zurich, Switzerland, [2]Functional Analysis and [3]Neurobiology Sector SISSA, Scuola Internazionale Superiore di Studi Avanzati, via Bonomea 265, 34136 Trieste, Italy, [4]Bioinformatics Program, [5]Department of Biology and [6]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Deciphering the response of a complex biological system to an insulting event, at the gene expression level, requires adopting theoretical models that are more sophisticated than a one-to-one comparison (i.e. *t*-test). Here, we investigate the ability of a novel reverse engineering approach (System Response Inference) to unveil non-obvious transcriptional signatures of the system response induced by prion infection.

Results: To this end, we analyze previously published gene expression data, from which we extrapolate a putative full-scale model of transcriptional gene–gene dependencies in the mouse central nervous system. Then, we use this nominal model to interpret the gene expression changes caused by prion replication, aiming at selecting the genes primarily influenced by this perturbation. Our method sheds light on the mode of action of prions by identifying key transcripts that are the most likely to be responsible for the overall transcriptional rearrangement from a nominal regulatory network. As a first result of our inference, we have been able to predict known targets of prions (i.e. PrP^C) and to unveil the potential role of previously unsuspected genes.

Contact: altafini@sissa.it

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on May 30, 2011; revised on August 17, 2011; accepted on October 14, 2011

1 INTRODUCTION

A major challenge in Systems Biology is to provide a global and quantitative understanding of complex diseases by means of high-throughput technologies. In a system-level approach, the synergistic response dictated by the cooperation and interplay of all the components within an organism is considered as a fundamental concept in the interpretation of 'omics' data. Recent efforts employ model-driven analysis and transcriptomic data for studying neurodegenerative diseases (Benetti *et al.*, 2010). However, while DNA microarrays have been extensively used to comprehensively characterize the differentially expressed genes (DEGs), the application of more sophisticated models, working beyond simple differential expression detection, is still an open challenge (Fig. 1). To this end, the ability of reverse engineering algorithms to infer putative gene–gene interactions, functional relationships and genome-scale regulatory networks has been already proven to be successful (Faith *et al.*, 2007; Gardner *et al.*, 2003). Although the applications to multicellular organisms are still sporadic, this class of algorithms represents a promising approach toward genome-wide studies of complex diseases (Basso *et al.*, 2005; Wang *et al.*, 2009). Our aim in this article is to propose a novel reverse engineering strategy in order to investigate transcriptional changes induced by prion propagation in mouse models.

Among the various neurodegenerative disorders, prion diseases have captured the attention of the scientific community due to their infectious nature and transmissibility (Prusiner, 1982). The cellular form of the prion protein (PrP^C) is the substrate for an autocatalytic reaction that, through its recruitment and modification, acts in the propagation of a disease-causing form denominated PrP^{Sc} or prion. This autocatalytic process leads to the formation of fibrils and amyloid deposits, and is associated to a progressive fatal damage of the brain tissue. Importantly, experimental evidence has demonstrated that PrP knockout and PrP anchorless (Chesebro *et al.*, 2005) mice are resistant to infection. In spite of the many approaches aimed at unveiling its physiological functions, the degree of overlap between the pathogenic dysfunction of the protein in prion diseases and its normal role is still largely unknown [for a review see Aguzzi *et al.* (2008)]. In such a complicated 'puzzle', it is widely accepted that since PrP^C is a key mediator for the neurotoxicity of PrP^{Sc}, studying its biological functions through the identification of its related molecular partners may help unraveling the mechanisms underling prion-dependent neurodegeneration.

In light of this observation, we integrate here a recently published transcriptional dataset (Hwang *et al.*, 2009), monitoring prion propagation in mouse brain tissues, with independent gene expression experiments performed under many diverse conditions. The computational framework introduced in this article aims at deciphering transcriptional variations in the context of a system-wide regulatory network, moving from a single gene perspective to a system-level analysis (Fig.1). The procedure is composed of three steps. In the first step, using a compendium of gene expression measurements, we obtain a genome-scale interaction graph of the mouse brain. Then, in a second step, we refine the initial model via a multiple regression scheme. In the third step, this linear regressive model is used to investigate expression changes associated to Prion infection, where for each gene we estimate the influence of prion

*To whom correspondence should be addressed.

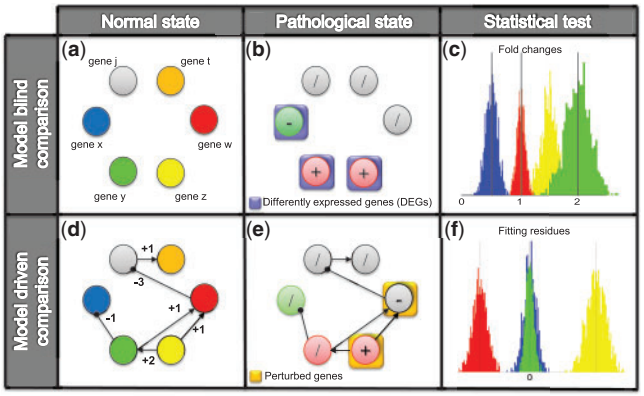

Fig. 1. System response inference: a toy genetic network consisting of six genes exemplifies the advantages of using a system-level data comparison (**a**). Standard statistical tests (i.e. *t*-test) unveil significant fold change in gene expression variations for each transcript individually (**b**), neglecting the underlying regulatory network. Such statistical test can identify whether the expression level of a transcript is significantly changed with respect to a reference. Putative gene expression changes are reported in panel (**c**). In this specific example, two genes are identified to be overexpressed [red/+ nodes] and one downregulated (green/- node), while the remaining three do not show any changes (grey nodes). By knowing the corresponding genetic regulatory network (**d**), we can discriminate the coherent variations from the unexpected ones. As shown in the example, two of the genes that showed a significant expression variations are consistent with model predictions i.e. the expression changes of genes *x* and *y* can be explained by the variation of gene *z*. This is reflected by a skew distribution of discrepancies (i.e. residues), between model predictions and observed data, centered around 0 (**f**). At the same time, one transcript, *w*, is not responding coherently to the initial model. The fact that its expression is unchanged, when it should have been increased, might relate to an anomalous direct effect of the pathology, preventing a synergistic response between all the genes in the system. Hence, the list of 'perturbed genes' can be sensibly different from the standard DEGs identified from individual fold change analysis (**b**/**e**).

propagation on its transcriptional regulation. In other words, we systematically identify the genes that result 'perturbed' with respect to the nominal model by a prion infection. A remarkable outcome of our network-based strategy is the ability to identify PrP^{C} among the predicted prion targets, in agreement with its role of primary substrate for prion replication. This result is even more important if we consider that to our knowledge prion infection has never been noted to interfere with the transcriptional regulation of the *Prnp* gene. Moreover, the comparison between the PrP^{C} partners inferred on the nominal model and the PrP^{Sc} targets predicted from the perturbed model allows to estimate the significance of the overlap between the two.

2 METHODS

2.1 Data collected

For *Mus Musculus*, we compiled a collection of microarrays (X_{Training}) containing experiments performed with Affymetrix chips (1366 experiments Affymetrix Mouse Genome 430 2.0), selecting Gene Expression Ominibus (GEO) experiments performed on brain tissue. Probe set intensities referring to a common gene were averaged, and only genes with a corresponding orthologous gene in Human were retained (for a total of 13 204 genes). Identical procedure was followed to collect a dataset of gene expression profiles for Human (Affymetrix Human Genome U133 Plus 2.0 Array) and Rat (Affymetrix Rat Genome 230 2.0 Array). See Supplementary Table S1 for the list of GEO codes. All datasets were preprocessed and normalized prior to network inference. Protein Complex network for human, rat and mouse were downloaded from the CORUM database (Ruepp *et al.*, 2010), whereas PPIs from the MppDB website (Li *et al.*, 2009).

2.2 SRI procedure

In this section, we describe the method used for reverse engineering large-scale gene–gene transcriptional dependencies and predicting systemic changes induced by a perturbing event (i.e. prions). This framework is inspired by the works of Bansal *et al.* (2006); di Bernardo *et al.* (2005); Gardner *et al.* (2003). In order to perform a truly systems-level analysis of gene expression profiles, we first seek for significant transcriptional inter-dependencies by means of a relevance network approach (Step I). We then use this relevance network to identify a nominal model based on a linear ODE description (Step II). Finally, the inferred model is used to detect gene expression changes caused by prion infection or *Prnp* deletion incompatible with the model (Step III).

Step I: Relevance network algorithms have the remarkable advantage of being computationally feasible for genome-wide applications (Basso *et al.*, 2005; Faith *et al.*, 2007). This class of methods assumes that gene–gene interactions can be identified by determining the gene pairs possessing the highest expression similarity over multiple conditions. A square matrix ($n \times n$, where n is the number of genes) of pairwise Pearson's correlations between gene expression patterns is computed. The (unsigned) absolute values of Pearson's correlation between gene i and j is:

$$C_{ij} = \left| \frac{E[(x_i - \bar{x}_i)(x_j - \bar{x}_j)]}{\sqrt{v_i v_j}} \right|, \tag{1}$$

where x_i is the expression vector for gene i (i.e. m gene expression measurements made under diverse conditions) and x_j is the expression vector for gene j. In Equation (1) $\bar{x}_i$, v_i and $\bar{x}_j$, v_j are means and variances of x_i and x_j over the m experiments and $E[\cdot]$ denotes expectation. Each row/column of C (i.e. C^i) can be considered as a distribution of pairwise similarity coefficients between a gene (i.e. i) and all the others. This distribution can be transformed in the corresponding vector of rankings, from 1 to n (W^i). As a consequence of this transformation, the correlation between i and j is associated to two 'ranking values'. Denote $W^i(j)$ and $W^j(i)$ the rank of C_{ij} with respect to C^i and C^j, respectively. Genes (i and j) are considered as putative interactors if and only if both their ranking indexes ($W^i(j)$ and $W^j(i)$) are above a certain common threshold θ [Equation (2)]. A graph of putative interactions, G_R, is obtained by the following relationship:

$$(i,j) \in G_R \Longleftrightarrow W^i(j) \geq \theta \;\&\; W^j(i) \geq \theta. \tag{2}$$

Step II: for each gene i, we assume that the change in concentration can be expressed as a linear combination of the concentrations of interacting genes [i.e. x_j such that $(i,j) \in G_R$], plus possibly an additive external stimulus/perturbation [Equation (3)]. This external input can be an environmental change (i.e. nutrient viability), a genetic perturbation (i.e. overexpression/deletion) or a toxic compound able to perturb the system (i.e. drive the system to a new steady state). Similar to Gardner *et al.* (2003), the responses of the system to the various perturbations can be formalized as a system of linear differential equations (ODEs) in which the input vector $u = [u_1, \ldots, u_p]$ is modeled as a linear combination of the external perturbations u_i

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{(i,j) \in G_R} a_{ij} x_j + \sum_{l=1}^{p} b_{il} u_l \tag{3}$$

or in matrix form (for $x = [x_1 \ldots x_n]^T$):

$$\dot{x} = Ax + Bu \tag{4}$$

where the coefficient a_{ij}, representing the influence of gene j on gene i, is non-zero if and only if $(i,j) \in G_R$. The variable u can be taken as a vector

($p \times 1$) identical to 1. The $n \times p$ input matrix B collects the influences of the external perturbations u on each single gene. The k-th column of B indicates how effective/intense the k-th perturbation u_k is on the state vector x (di Bernardo *et al.*, 2005). Under steady state conditions, the variation in time of the mRNA concentration is by definition equal to zero. In this case, Equation (4) becomes a simpler system of algebraic equations:

$$Ax + Bu = 0. \tag{5}$$

In order to estimate the parameters of the model, we employ a large compendium of gene expression profiles (X_{Training}), containing many different external perturbations, none of them is related to prion infection. In Equation (5), we assume that the matrix A stays constant over all the different experimental conditions, and that the columns of B are sparse (i.e. few non-zero elements). This last assumption reflects that the different perturbations collected in the dataset are likely to affect just a small portion of genes, if compared with the entire genome. This is justified for example in case of mutants, where a specific gene has been knocked out. By collecting a vast selection of different experimental conditions (i.e. >1300 chips), we aim at avoiding that particular perturbations are overrepresented in the dataset, hence that the assumption of sparsity does not hold for some of the rows of B. Therefore, for each transcript, expression changes caused by the direct effect of an external stimulus are likely to occur just in few of the experimental conditions of the training dataset, and to be 'uniformly' distributed over the rows of B (i.e. different perturbations do not act always on the same transcripts). Under these assumptions, the contribution of B is marginal, and sufficiently accurate estimation of the non-zero terms of the A matrix (a_{ij}) can still be drawn from the following approximation:

$$Ax \approx 0. \tag{6}$$

Similar arguments are used in most reverse engineering algorithms, see e.g. in (Cosgrove *et al*., 2008). A multiple regression framework is here used to learn the network coefficients a_{ij} from the training dataset X_{Training}. This framework seeks for the solutions which minimize the L^2 norm between the predictions of the model and the experimental values of X_{Training}. For each gene, the result is the following vector of coefficients:

$$\hat{a}_{c_i} = \underset{a}{\operatorname{argmin}} \left(\| x_i - \sum_{(i,j) \in A_I} \frac{a_{ij}}{a_{ii}} x_j \|^2 \right) \tag{7}$$

where x_i is a vector of gene expression measurements of gene i (i.e. row of X_{Training}). The solution of Equation 7 is correct up to an undetermined scaling factor (the diagonal term a_{ii}) by which we rescale each row of A. The outcome of the n regression problems is a weighted asymmetric matrix A_c and, for each gene i, a vector of residues (r_i), resulting from the difference between observed and predicted values.

Selection of θ: In order to avoid an arbitrary selection of the cut-off parameter θ in Equation (2), we adopt a Bayesian Information Criterion (BIC). Since an increase in the number of free parameters in Equation (7) always reduces the absolute sum of residuals (r), we search for the best compromise between the model complexity, in terms of the number of parameters (i.e. non zero entries in the connectivity map G_R) and fitness quality (i.e. discrepancies between predicted and observed values). Therefore, we fitted multiple models corresponding to different choices of θ in Equation (2) and estimated the corresponding fitness qualities via the BIC. The BIC, Equation (8), is the most common criterion for the fitness of a mathematical model to observed data, leaning more toward lower dimensional models (i.e. sparse matrices)

$$\text{BIC}_i = m \left[ln \left(\frac{\text{RSS}_i}{m} \right) \right] + ln(m) K_i \tag{8}$$

where K_i is the number of parameters in the model for gene i (i.e. number of edges of i) and RSS_i the residual sum of squares ($\sum_{z=1}^{m} r_{i_z}^2$). The value of θ corresponding to the minimum of the mean of BIC_i indexes ($\hat{\theta}$) is selected for the final step of the analysis [Equation (2)]:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left(\sum_{i=1}^{n} BIC_i \right) \tag{9}$$

The identification of the graph G_R is the most underdetermined step of the reverse engineering approach. However, coherently with Bansal *et al.* (2006), we found that small changes of θ do not drastically change the structure of A_c and its numerical entities.

Step III: the resulting A_c model is used to predict the expression profiles associated to a new condition, here denoted as X_{Prion} [in our case, the dataset published in Hwang *et al.*, (2009)]. The procedure is to search for the genes for which the nominal model A_c is incorrectly describing the expression changes in the new experimental condition. The dataset X_{Prion} contains perturbative effects induced exclusively by prion propagation or *Prnp* absence: prion infection (143 experiments/columns in X_{Prion}), *Prnp* deletion (22 experiments) and both perturbations at the same time (27 experiments). According to Equation (6), we expect that if gene i is not directly influenced by these external agents, then its expression changes should be correctly predicted by the nominal model (i.e. $A_c X_{\text{Prion}} \approx 0$). This should be reflected in a skewed distribution of residues around 0 (named r_{Prion}). On the contrary, whenever gene i is directly influenced by a prion-related external stimulus we expect the initial model to be ineffective in predicting gene expression variations. This indicates that the equation should be corrected as $A_c x_{\text{Prion}} + B_{\text{Prion}} u_{\text{Prion}} = 0$. Non-zero entries in the i-th row of B_{Prion} unveil the input 'pathway' of the prion-related perturbation. For each gene i, we test this hypothesis by comparing the discrepancies that follow if we impose $A_c x_{\text{Prion}_i} \approx 0$ (i.e. r_{Prion_i}) with those obtained during the fitting procedure of Step II (r_i), through a two-sample t-test [Equation (10)] (Fig. 2):

$$t_i = \frac{\bar{r}_i - \bar{r}_{\text{Prion}_i}}{\sqrt{\frac{s_{r_i}^2}{m_t} + \frac{s_{r_{\text{Prion}_i}}^2}{m_{\text{Prion}_i}}}} \tag{10}$$

where $\bar{r}_i$ and $\bar{r}_{\text{Prion}_i}$ are the mean of the residues in the training and the 'prion-specific' dataset for gene i and s_{r_i}, $s_{r_{\text{Prion}_i}}$ are the corresponding SDs. Using the Student's t-probability distribution, a statistical significance (P-value) is assigned to each gene. Once the distribution of analytical P-values has been generated, the q-value is used to correct the measure of significance for multiple testings (Storey, 2002). When the two distributions of residues are significantly different, the gene i is considered to be a target of the applied perturbation.

3 RESULTS

In a recent work (Hwang *et al.*, 2009), more than 400 DNA chips were collected, monitoring the transcriptional changes associated to various disease stages, prion strains and mice backgrounds. The authors interpreted differentially regulated genes in light of the known network of protein–protein interactions and hypothesized a specific temporal order of the biological processes involved. In our study, the same dataset of Hwang *et al.* (2009) is integrated with a large compendium (i.e. 1366 Affymetrix experiments) of gene expression measurements associated to various conditions. We first trained a gene network model (A_c, see Section 2) on this large compendium. Next, on the background of this initial network, we evaluated the impact of an independent dataset (Hwang *et al.*, 2009) containing gene expression variations caused by a perturbing event, that can be either an external agent (i.e. prion inoculation) or an endogenous perturbation (i.e. *Prnp* knockout). In our interaction network, inoculated prions were modeled as hidden variables/nodes, while a gene knockout was interpreted as a node deletion (Fig. 3). We then investigated the behavior of the system upon Prion infection or *Prnp* deletion with special attention

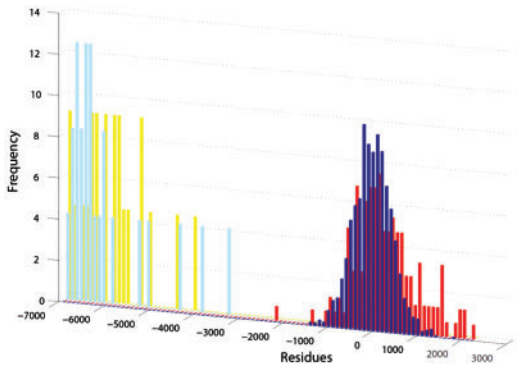

Fig. 2. Residue distributions: the results obtained for the *Prnp* gene are shown as an example. While for the training set (blue) a Gaussian distribution centered around zero was recovered (in agreement with the expectation of white noise residues), a heavier tail was emerging when predicting prion-infected mice measurements (red). The drastic opposite effect is expected when knocking out the *Prnp* gene (yellow) or when both perturbations are applied (light blue). It is worth to stress that the authors of Hwang *et al.* (2009) did not find any significant expression changes in the *Prnp* gene. This result was common to all the studies performing gene expression analysis in prion-infected systems [for a review see Benetti *et al.* (2010)]. Our outcome was different from previous studies [such as Hwang *et al.* (2009)] because we allowed a gene to be 'significant' regardless of its individual difference in expression between infected and normal samples. Rather, systemic gene expression alterations are here considered and used in order to estimate the expected expression changes of the singular transcript. In a system-level response, a global transcriptional rearrangement is expected to counter balance the new 'operating point/equilibrium'. If this regulation is not taking place, or in other words if the gene is not able to properly respond, enhancing or repressing its transcription, its contribution in the disease progression should be taken into account as well (Fig. 1). Therefore, our approach can be considered to be complementary to the analysis performed in Hwang *et al.* (2009).

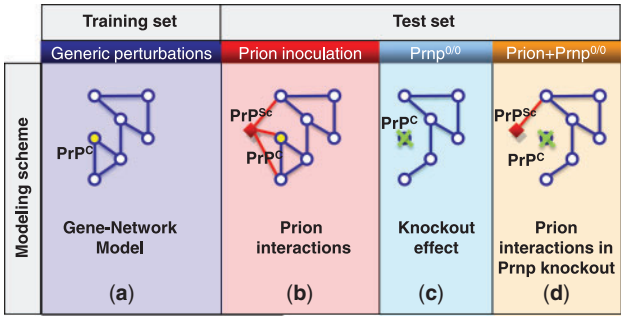

Fig. 3. Gene network approach: scheme of the analytical framework applied to investigate the PrP^{C} functions and PrP^{Sc} targets. A large compendium of gene expression measurements, comprehensive of a variety of different conditions, is first used to extrapolate a weighted network of 'gene-gene' interactions (**a**, PrP^{C} is highlighted in yellow). Nodes represent genes, whereas edges link 'interacting' genes. The rewiring of the network caused by self-replicating prions (red node, **b**) or by a deletion of the gene (**c**) will render our initial nominal model ineffective for the genes directly linked to the perturbation. We address this issue comparing the inconsistencies (i.e. residues) of the model predictions if applied to perturbation-specific (**b**–**d**) experimental values.

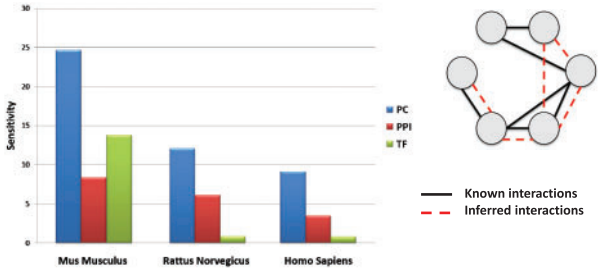

Fig. 4. Network *in−silico* validation: we assessed the ability to recover the network of protein–protein interactions (PPI), protein complexes (PCs) and transcriptional regulation (TF) in three different mammals (mouse, rat and human, see Section 2). We evaluated the overlap between the inferred networks (dashed red lines in the toy network) with the known networks (solid black lines) by computing the relative sensitivity index ($\frac{\text{true positive}}{\text{true positive} + \text{false negative}}$). The histogram shows that even in such underdetermined conditions (ratio experiments/genes ∼0.1), a reverse engineering approach on a multicellular system may extrapolate meaningful information.

to alterations in the gene expression that could not be explained by the initial nominal model. Such discrepancies are assumed to be the consequence of the investigated perturbing events (see Section 2). This method, called 'System Response Inference' (SRI), is inspired by computational frameworks published in Bansal *et al.* (2006); di Bernardo *et al.* (2005); Gardner *et al.* (2003). Results are organized in four subsections. In the first, we investigate the inferred co-expressed partners of *PrnP*. In the next two subsections, we analyze and interpret the data of Hwang *et al.* (2009). SRI predictions for putative PrP^{C} buffering genes and PrP^{Sc} targets are analyzed and the differences emerging from our system-level approach with respect to the results in Hwang *et al.* (2009) are discussed. Subsequently, we consulted a drug database in search for druggable targets among the first neighbors identified in the extrapolated network for PrPs. Finally, an analysis performed across human, mouse and rat samples revealed co-expression patterns that were highly conserved across these species.

3.1 Prion protein co-expressed partners

By applying SRI to the collected gene expression database on mouse brain tissue samples (see Supplementary Table S1 for GEO codes), we first identified 450 significantly *Prnp*-related genes, of which 37 were anticorrelated (see Supplementary Table S2-partners and Fig. 3a). As we have already observed in a previous study (Zampieri *et al.*, 2008) for unicellular prokaryotes and eukaryotes (i.e. *Escherichia coli* and *Saccharomyces cerevisiae*), the strongest source of co-expression was the co-participation of gene products in the same protein complex (Fig. 4). This observation reflects the tendency to establish 'stable-bindings' between gene products of the inferred interacting gene pairs. Indeed, many predictions were in agreement with previous experimental studies seeking for proteins capable of binding PrP^{C} such as *Cltc*, *Gpm6a*, *Gpm6b* (Rutishauser *et al.*, 2009), *Lsamp*, *SparcL*1 (Schmitt-Ulms *et al.*, 2004), *Ywhag*, *Ywhah* and *Gnao*1 (Watts *et al.*, 2009) or to co-localize with prion (i.e. *ApoE*) (Nakamura *et al.*, 2000) (see Supplementary Material for further discussion). In addition, *Aplp*1, the precursor-like amyloid protein, has been identified, in association with *App*, as a probable ligand of PrP (Yehiely *et al.*, 1997). Such prediction was supported by their similar proteolytic processing and ability to bind copper

and, in case of *Aplp*1, to inhibit the cleavage of *App* by β-secretase (*Bace*1). It is worth noting the connection with *Apbb*1, thought to modulate the internalization of β-amyloid precursor protein and *Rtn*3, also inhibiting *Bace*1 activity. Furthermore, synaptojanin1 (*Synj*1), the major inositol 5-phosphatase (jointly with *Pitpna* and *Cadps*) is known to reduce the levels of cellular PIP2 and it has been shown to have an important impact in determining the level of Aβ42 (Berman *et al.*, 2008). The co-regulation with *Rtn*4, which reduces the anti-apoptotic activity of Bcl-xl and Bcl-2, might relate to the observed inverse transcriptional proportionality between *Prnp* and *Bcl2l2* (e.g. *Casp*6, *Prdx*4), pointing to a role of PrP in apoptosis (Kim *et al.*, 2004). *Btf*3, *Cdk*2 and *Baz*1*a*, all involved in transcriptional regulation, are candidates for inhibiting the *Prnp* expression.

By performing an ontology enrichment (Supplementary Table S3), we identified some of the biological functions already associated to PrP^C, suggesting the potential molecular partners responsible for these roles. For instance, the presence of small GTPase molecules such as Rab proteins might emerge because of clathrin-mediated endocytosis, reflecting their control role in PrP^C trafficking. Furthermore, the enrichment of categories like Golgi vesicle transport and endocytosis may be related to the transit of PrP^C from the endoplasmic reticulum to the cell surface through the Golgi apparatus. Several other genes were involved in nerve impulse transmission, such as again Synaptojanin-1, which in association with *Lrp*11, reinforced the evidences that PrP^C is internalized via clathrin-coated vesicles (Taylor and Hooper, 2007). In parallel, caveolin-mediated signaling (Pantera *et al.*, 2009) (here represented by *Cav*2) and the localization of PrP^C on membrane lipid rafts might explain the occurrence of genes related to lipid transport. In light of the many studies observing behavioral changes in *Prnp* knockout mice [i.e. reduced level of anxiety (Nico *et al.*, 2005)], we highlighted the presence of genes related to anxiety, such as *Gnao*1. Impaired hippocampal-dependent spatial learning was observed in *Prnp* null-mice as well, supporting the inferred relationship with *Ncdn*, required for spatial learning processes and Cpeb1 involved in long-term memory (Si *et al.*, 2003a) (see next section for further discussion). Furthermore, several sources of experimental evidences have shown the effect of PrP on cell proliferation and differentiation (Salès *et al.*, 2002) and were here confirmed by the enriched categories such as the nervous system development and differentiation (i.e. *Ids*, responsible for mental retardation, *Trim*2 contributing to the alteration of neural cellular mechanisms, and *Nedd*1). These properties were related to the activation of PrP^C by cAMP and MAP proteins kinases (which are also identified as significant ontological categories) through the interaction with *Ncam* (Santuccione *et al.*, 2005). Our analysis did not retrieve the *Ncam* interaction, but identified a closely related gene *Nrcam*, alongside many other cell adhesion molecules. Furthermore, the role of PrP in synaptic activity, neurotransmission and neuronal excitability was highlighted by the identification of *Syn*1. The previously observed relationship between PrP and Erk activity might be just a secondary effect of the relationship here identified with Ras signal transduction, MapK, calcium signaling pathways and PKC signaling (i.e. *Gnb2l*1, *Prkd*3). Furthermore, the high percentage of genes related to homeostatic processes and ion transport (i.e. *Adam*11 a metalloprotease, probably reflecting the known interaction with *Adma*23) indicated a possible involvement in protection against oxidative stress (i.e. *Oxr*1) and metal ion dysmetabolism. Finally, it is interesting to notice that melanogenesis emerged as a highly significant biological pathway (Supplementary Table S3), hence indicating the possibility for a relationship between non-pathogenic presence of proteinsase-resistant PrP and the mechanism of production and formation of melanin (Fowler *et al.*, 2006).

3.2 PrP knockout compensation

Even if the PrP sequence is very well conserved among mammals, its ablation in mice results in no clear phenotype, nor is essential for their survival. This might be due to several reasons, such as the presence of buffering genes masking the PrP^C absence or to an adaptive process during central nervous system development. We used the inferred mouse transcriptional gene network model to interpret the response to a *Prnp* deletion from the same dataset of (Hwang *et al.*, 2009) (see Section 2 and Fig. 3c). We immediately noticed that, as expected, the perturbation intensity, as well as the *Prnp* associated q-value, was the lowest (see Supplementary Table S2, knockout). In addition, 27 genes show an equal incoherent variation (i.e. q-value = 0), which was not predicted by the initial gene network model. It is worth observing that a simple comparison of these transcripts with their reference counterparts (i.e. normal state) does not show any relevant expression changes. Indeed, in Hwang *et al.* (2009) no significant expression changes were found in PrP-null mice. This highlights the difference between our approach and a standard comparative study, emphasizing the ability of our method to extrapolate non-obvious transcriptional signatures.

Among these genes, the enriched biological processes were tRNA metabolism, intracellular transport and synaptic transmission (Supplementary Table S4). We could observe for example the presence of *Dstn*, encoding for an actin-binding protein belonging to the ADF family responsible for enhancing the turnover rate of actin. In addition, *Gls* catalyzes the conversion of glutamate from glutamine and its potential role for compensating PrP absence may derive from its implication in behavior disturbances in which glutamate acts as a neurotransmitter (Prusiner, 1981). *Gls* is also involved in Alzheimer, just like *Chgb*, a tyrosine-sulfated secretory protein. Given their co-expression patterns with PrP^C (see previous section) and their high significance as targets of PrP^{Sc} (next section), *Sirpa* and *Cpeb* are of particular interest (see also Supplementary Fig. S2). The former is a transmembrane glycoprotein receptor involved in the negative regulation of receptor tyrosine kinases, being a substrate for tyrosine phosphatases (PTP). This gene may participate to the signaling cascade involving *Fyn* (Pantera *et al.*, 2009) and their concomitant activation with *Prnp* was observed in bovine monocytes infected by Theileria (Jensen *et al.*, 2008). The latter has been recently discovered to be able to aggregate as a prion-like protein (Heinrich and Lindquist, 2011; Si *et al.*, 2003b). Additionally, in Si *et al.* (2003a) the authors showed its vital role in long-term memory, and it is tempting to speculate that a similar function may exist for PrP. It is worth noting that both transcripts are similarly reacting to a Prnp deletion and to prion infection (Supplementary Table S2), thereby suggesting that in case of prions replication the system is attempting to compensate for a lack of PrP^C. By comparing the residues in *Prnp*-null infected mice with mock-inoculated $Prnp^{0/0}$ (Fig. 3d), no significant genes were identified ($q < 0.05$). This result was consistent with the insensitivity

of prion knockout mice to prion infection and it was in agreement with the conclusions in Hwang *et al.* (2009).

3.3 Prion targets

The same procedure has been used to determine the portion of the inferred network which is most 'sensitive' to prion perturbation. Hence, in order to unveil the nodes (i.e. genes) which are likely affected by prions, the predictions of the initial model were tested on the experimental data collected in Hwang *et al.* (2009) and the inconsistencies were detected through a rigorous statistical test (see Section 2 and Fig. 3b). Model-based inference on prion-infected mice resulted in 3255 genes with an associated q-value lower than 0.01 (see Supplementary Table S2, targets). As already mentioned, previous results showed that prions bind with high specificity to PrP^C, and that further effects were mediated by its autocatalytic recruitment and conversion, although so far, no transcriptional variations have been extrapolated for the *Prnp* gene during infection (Benetti *et al.*, 2010). Notwithstanding these studies, the *Prnp* gene was here identified among the most significant prion 'targets'. Therefore, as a reference threshold for the q-value significance, we selected the one associated to *Prnp* ($7.12*10^{-9}$), as we did previously for the *Prnp* deletion. This led to the selection of 389 genes (out of the total 13 204 considered). In Hwang *et al.* (2009), the authors identified ∼7400 DEGs in at least one of the five combinations prion-strain/mouse-background, and among these genes, 333 were identified in all the combinations. Again, only a small fraction of genes (i.e. 58) are in common with our list, mainly those involved in immune system processes (Supplementary Table S8).

Among the detected prion targets, several were in common with the previously observed co-expressed partners (Section 2.1) of the cellular prion form ($P = 1.07 \cdot 10^{-6}$ hypergeometric test): for example, *Magi*2, encoding a protein interacting with Astrophin1 (modulating activin-mediated signaling), *Snap*25, already detected in scrapie-infected GT1 cells (Sandberg and Löw, 2005) or *Cntn*1, also involved in nervous system development. This in turn, reinforced the evidences of the central role of PrP^C as a mediator of prion neurotoxicity. Various genes in the list were involved in apoptosis, inflammatory response and leukocyte activation (Supplementary Table S5). It is not clear what are the primary damage causes, nor whether such damage is due to a gain or a loss of PrP functions, but our findings suggested apoptosis as a mechanism directly influenced by prion propagation. Genes involved in signaling, neurotransmitter transport and integrin-mediated signaling were perturbed during prion infection, in agreement with the reduced synaptic neurotransmission and dysfunction observed during TSE propagation (Ferrer *et al.*, 1999). In this respect, it has previously been reported that scrapie-infected mice showed higher *Bace*1 activity, suggesting an impairment of the regulatory role of PrP^C in its modulation. Our results indicated an unbalanced response of *Bace*2, which stimulates *App* processing in a non-amyloidogenic pathway suppressing the level of Aβ production, probably in an attempt to compensate for *Bace*1 dysfunction. It is also worth noting the Atf6 gene, one of the main stress sensor of the ER membrane and the related Eif4e, which has been recently found to potentially play a main role in prion, induced neurodegeneration (Roffé *et al.*, 2010). In addition, it has been observed that transgenic mice expressing PrP^C deletion mutants, or overexpressing Dpl are characterized by vacuolation of the myelinated fibers (Bremer *et al.*, 2010), where PrP^C presence was discovered (Radovanovic *et al.*, 2005), bringing back our analysis to genes involved in myelination (i.e. *Lgi*4, *Klk*6) and the several integrins among the most significant hits (*Itgax*, *Itgb*2, *Itgb*4). The enrichment of genes involved in the glycan degradation pathway and lipid transport and sequestering might reflect the response of the system in controlling the amount of PrP on the membrane surface, where its attachment is a necessary condition for prions to induce neurotoxic signals (Chesebro *et al.*, 2005). We also noticed the presence of several apolipoproteins, such as *Apoc*1, shown to play a role in Alzheimer, whereas *ApoD* and *Apold*1 may play a role in brain vasculature, affecting brain–blood permeability.

3.4 Druggable targets

Prion molecular targets may pave the way to the development and identification of compounds able to disrupt the mechanism of action of prion neurotoxicity. By targeting the inferred gene products, it might be possible to modulate disease progression, delaying it in time or to interrupt a fundamental neurotoxic chemical reaction. We consulted a repository database (*www.Drugbank.ca*) and linked approved drugs to the genes identified in our analysis (Supplementary Table S6). Drugs acting on cholesterol by inhibiting its biosynthesis, such as statins, have been previously used (Taraboulos *et al.*, 1995) for the treatment of Prion diseases. In particular, here we identified simvastatin (Kempster *et al.*, 2007) its ability to interact with *Itgb*2. We also noticed Spermine (polycataionic compound) and N-Acetyl-D-glucosamine, which have been shown to prolong the incubation time (Trevitt and Collinge, 2006). Compounds such as antihemophilic factors, Tenecteplase and Bevacizumab, are instead acting on the coagulation process, removal of blood clots and the inhibition of blood vessel growth, respectively. The second one in particular seems to be a promising treatment since it prevents the interactions of VEGF with its receptors. Abnormal regulation of VEGF expression has been implicated in several neurodegenerative disorders, including Prion disease (Koperek *et al.*, 2002), where neurodegeneration might be caused by impairing neural tissue perfusion. Another promising compound is choline, a major component of the polar head group of phosphatidylcholine, which plays a vital role in basic biological processes, including the maintenance of cell structure and function. The recent hypothesis that prions can form permeable pores and influence ion channels (Kourie *et al.*, 2003), points to the predicted drugs acting on sodium (like Lamotrigine, Amiloride, Quinidine) or potassium channels (Glibenclamide), while Iron Dextran could in principle be used to counteract metal ions dysmetabolism, a common effect of several neurodegenerative pathologies. We were also able to extrapolate three drugs acting on tyrosine kinase activity, Aldesleukin , Insuline and Methymazole, in conjunction with antioxidant compounds, such as L-cysteine, vitamin C, NADH, succininc, L-glutamic acids and others, shown to be able to reduce PrP^{Sc} propagation in infected cell (Kocisko *et al.*, 2003).

3.5 Co-expression patterns across multiple species

The SRI approach provides an analysis of transcription changes that is not restricted to individual isolated genes, but rather considers the system as a whole. This method is hence appropriate to characterize the prion protein interactome in its genetic network

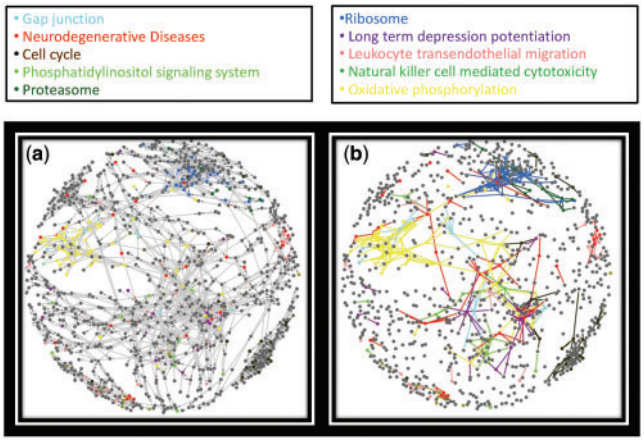

Fig. 5. Conserved co-expression patterns: the connected component emerging from the selection of the conserved graph of co-expression patterns across three different species is here represented. This subnetwork consists of 943 genes and 2409 interactions (see Fig. 5a and Supplementary Materials). Gene centric view (**a**): genes belonging to the identified significant enriched pathways are highlighted in different colors. Edge centric view (**b**): by selecting only the predicted edges of genes belonging to enriched categories and using the same color code, we highlighted the 'compactness' of relationships characterizing specific pathways, such as oxidative phosphorylation, ribosome, long-term depression and proteosome (see Supplementary Figs S3–S4 for further details). Genes related to neurodegenerative processes seemed to have many 'promiscuous' interactions linking to diverse biological processes (red edges).

context. Three nominal networks for mouse, rat and human (Fig. 4 and Section 2) were inferred from three large compendia of transcriptional data. By retaining only the parts of these networks with consistent signs, we were able to extrapolate the co-expression patterns which are conserved among the three species. An analysis of the connectivity of the resulting network clearly revealed the presence of a major connected component (Supplementary Fig. S3). Surprisingly, this conserved connected component included several genes related to various neurodegenerative diseases (i.e. *Prnp*/Prion diseases, *Snca*, *Uchl*1/Parkinson's disease, *Bace*1/Alzheimer's disease), making neurodegeneration one of the most significantly represented biological process (Fig. 5a and Supplementary Table S7). It is tempting to speculate on the fact that the conservation across species of these transcriptional interdependencies, might reflect the important role of coordination in their transcriptional regulation. The fact that a large fraction of genes is related to neurodegenerative diseases suggests the existence of a delicate balance at the transcriptional level, whose alteration may have severe consequences. Focusing only on the relationships identified for the genes belonging to the significantly enriched processes, we evinced an interesting property of this evolutionarily conserved graph (Fig. 5b and Supplementary Fig. S3). While genes related to cell cycle, ribosome and oxidative phosphorylation were clustered in compact modules, genes related to neurodegeneration (red dots) were sparser in the network, with many cross-talking edges. Our results are complementary to other studies performing global analysis of human protein interactions (Limviphuvadh *et al.*, 2007). In a recent work (Goñi *et al.*, 2008), investigating the known network of PPIs, the authors showed that genes characterized by an altered expression level in neurodegenerative diseases take part in many different pathways, reinforcing the concept of a multifactorial pathogenesis of neurodegenerative disorders. In our study, relying exclusively on transcriptional dependencies, similar conclusions are drawn. This in turn supports our first claim that key genes might be responsible for mediating and triggering the systematic response to a perturbing event (i.e. *Prnp*, see Supplementary Figs S3–S4) and that even in such complex systems, valuable information can be extrapolated by their transcriptional analysis.

4 DISCUSSION

In this study, we employed previously published gene expression datasets to recognize the function of PrP^C and its role in responding to prion inoculation inducing incoherent gene expression variations in a gene network model. A reverse engineering approach was here adopted to analyze the data of Hwang *et al.* (2009) starting from the basic assumption that a global transcriptional rearrangement must take place in order to adapt to a new condition (i.e. prion infection). Specific gene expression changes were expected to trigger other expression variations to finally reach an internal state of balance possibly different from the original one (see Figs 1 and 3). Our model-based results restricted the spectrum of potential key interactions, performing an educated selection of transcripts that show an incoherent response with respect to the overall systemic reaction to prion infection. The proposed framework offers new insights into the physiological function of PrP^C and the molecular mechanisms underlying prion disease pathogenesis, unveiling a potential key role of previously unsuspected genes. Our results suggest that in order to fully explore the potential and advantages of microarray technologies, the size of the collected datasets must be large enough to capture the complete range of variations associated with the studied phenomenon. In this perspective, the data derived from a single experimental condition are complemented by the large independent dataset. Here, we show that SRI is not only a valuable tool to perform genome-wide studies in mammals, but is also capable of predicting the complex effects of endogenous and exogenous perturbations in a biological system, restricting the spectrum of plausible relationships that have to be experimentally investigated.

Funding: MIUR (PRIN project) (to C.A.) and PRIORITY FP7 project (contract no.: 222887 to G.L.).

Conflict of interest: none declared.

REFERENCES

Aguzzi,A. *et al.* (2008) The prion's elusive reason for being. *Annu. Rev. Neurosci.*, **31**, 439–477.

Bansal,M. *et al.* (2006) Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics*, **22**, 815–822.

Basso,K. *et al.* (2005) Reverse engineering of regulatory networks in human b cells. *Nat. Genet.*, **37**, 382–390.

Benetti,F. *et al.* (2010) Gene expression profiling to identify druggable targets in prion diseases. *Exp. Opin. Drug Discov.*, **5**, 177–202.

Berman,D.E. *et al.* (2008) Oligomeric amyloid-beta peptide disrupts phosphatidylinositol-4,5-bisphosphate metabolism. *Nat. Neurosci.*, **11**, 547–554.

Bremer,J. *et al.* (2010) Axonal prion protein is required for peripheral myelin maintenance. *Nat. Neurosci.*, **13**, 310–318.

Chesebro,B. *et al.* (2005) Anchorless prion protein results in infectious amyloid disease without clinical scrapie. *Science*, **308**, 1435–1439.

di Bernardo,D. *et al.* (2005) Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat. Biotechnol.*, **23**, 377–383.

Faith,J.J. *et al.* (2007) Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol.*, **5**.

Ferrer,I. *et al.* (1999) Expression of proteins linked to exocytosis and neurotransmission in patients with Creutzfeldt-Jakob disease. *Neurobiol. Dis.*, **6**, 92–100.

Fowler,D.M. *et al.* (2006) Functional amyloid formation within mammalian tissue. *PLoS Biol.*, **4**.

Gardner,T.S. *et al.* (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, **301**, 102–105.

Goñi,J. *et al.* (2008) A computational analysis of protein-protein interaction networks in neurodegenerative diseases. *BMC Syst. Biol.*, **2**, 52.

Heinrich,S.U. and Lindquist,S. (2011) Protein-only mechanism induces self-perpetuating changes in the activity of neuronal aplysia cytoplasmic polyadenylation element binding protein (cpeb). *Proc. Natl Acad. Sci. USA*, **108**, 2999–3004.

Hwang,D. *et al.* (2009) A systems approach to prion disease. *Mol. Syst. Biol.*, **5**, 252–252.

Jensen,K. *et al.* (2008) Differences in the transcriptional responses induced by theileria annulata infection in bovine monocytes derived from resistant and susceptible cattle breeds. *Int. J. Parasitol.*, **38**, 313–325.

Kempster,S. *et al.* (2007) Simvastatin treatment prolongs the survival of scrapie-infected mice. *Neuroreport*, **18**, 479–482.

Kim,B.H. *et al.* (2004) The cellular prion protein (prpc) prevents apoptotic neuronal cell death and mitochondrial dysfunction induced by serum deprivation. *Brain Res. Mol. Brain Res.*, **124**, 40–50.

Kocisko,D.A. *et al.* (2003) New inhibitors of scrapie-associated prion protein formation in a library of 2000 drugs and natural products. *J. Virol.*, **77**, 10288–10294.

Koperek,O. *et al.* (2002) Disease-associated prion protein in vessel walls. *Am. J. Pathol.*, **161**, 1979–1984.

Kourie,J.I. *et al.* (2003) Copper modulation of ion channels of prp[106-126] mutant prion peptide fragments. *J. Membr. Biol.*, **193**, 35–45.

Limviphuvadh,V. *et al.* (2007) The commonality of protein interaction networks determined in neurodegenerative disorders (NDDs). *Bioinformatics*, **23**, 2129–2138.

Li,X. *et al.* (2009) A mouse protein interactome through combined literature mining with multiple sources of interaction evidence. *Amino Acids*, **38**, 1237–1252.

Nakamura,S. *et al.* (2000) Immunohistochemical detection of apolipoprotein E within prion-associated lesions in squirrel monkey brains. *Acta Neuropathol.*, **100**, 365–370.

Nico,P.B. *et al.* (2005) Altered behavioural response to acute stress in mice lacking cellular prion protein. *Behav. Brain Res.*, **162**, 173–181.

Pantera,B. *et al.* (2009) Prp(c) activation induces neurite outgrowth and differentiation in pc12 cells: role for caveolin-1 in the signal transduction pathway. *J. Neurochem.*, **110**, 194–207.

Prusiner,S.B. (1981) Disorders of glutamate metabolism and neurological dysfunction. *Annu. Rev. Med.*, **32**, 521–542.

Prusiner,S.B. (1982) Novel proteinaceous infectious particles cause scrapie. *Science*, **216**, 136–144.

Radovanovic,I. *et al.* (2005) Truncated prion protein and doppel are myelinotoxic in the absence of oligodendrocytic PrPC. *J. Neurosci.*, **25**, 4879–4888.

Roffé,M. *et al.* (2010) Prion protein interaction with stress-inducible protein 1 enhances neuronal protein synthesis via mTOR. *Proc. Natl Acad. Sci. USA*, **107**, 13147–13152.

Ruepp,A. *et al.* (2010) Corum: the comprehensive resource of mammalian protein complexes–2009. *Nucleic Acids Res.*, **38**, 497–501.

Rutishauser,D. *et al.* (2009) The comprehensive native interactome of a fully functional tagged prion protein. *PLoS One*, **4**.

Salès,N. *et al.* (2002) Developmental expression of the cellular prion protein in elongating axons. *Eur. J. Neurosci.*, **15**, 1163–1177.

Sandberg,M.K. and Löw,P. (2005) Altered interaction and expression of proteins involved in neurosecretion in scrapie-infected gt1-1 cells. *J. Biol. Chem.*, **280**, 1264–1271.

Santuccione,A. *et al.* (2005) Prion protein recruits its neuronal receptor NCAM to lipid rafts to activate p59fyn and to enhance neurite outgrowth. *J. Cell Biol.*, **169**, 341–354.

Si,K. *et al.* (2003a) A neuronal isoform of CPEB regulates local protein synthesis and stabilizes synapse-specific long-term facilitation in aplysia. *Cell*, **115**, 893–904.

Si,K. *et al.* (2003b) A neuronal isoform of the aplysia CPEB has prion-like properties. *Cell*, **115**, 879–891.

Storey,J.D. (2002) A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B*, **64**, 479–498.

Taraboulos,A. *et al.* (1995) Cholesterol depletion and modification of COOH-terminal targeting sequence of the prion protein inhibit formation of the scrapie isoform. *J. Cell. Biol.*, **129**, 121–132.

Taylor,D.R. and Hooper,N.M. (2007) The low-density lipoprotein receptor-related protein 1 (lrp1) mediates the endocytosis of the cellular prion protein. *Biochem. J.*, **402**, 17–23.

Trevitt,C.R. and Collinge,J. (2006) A systematic review of prion therapeutics in experimental models. *Brain*, **129**(Pt 9), 2241–2265.

Wang,K. *et al.* (2009) Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat. Biotechnol.*, **27**, 829–839.

Watts,J.C. *et al.* (2009) Interactome analyses identify ties of PRP and its mammalian paralogs to oligomannosidic n-glycans and endoplasmic reticulum-derived chaperones. *PLoS Pathog.*, **5**.

Yehiely,F. *et al.* (1997)Identification of candidate proteins binding to prion protein. *Neurobiol. Dis.*, **3**, 339–355.

Zampieri,M. *et al.* (2008) Origin of co-expression patterns in E. coli and S. cerevisiae emerging from reverse engineering algorithms. *PLoS One*, **3**.