

SCIENTIFIC REPORTS



OPEN

Assessment of the key regulatory genes and their Interologs for Turner Syndrome employing network approach

Anam Farooqui, Safia Tazyeen, Mohd. Murshad Ahmed, Aftab Alam, Shahnawaz Ali , Md. Zubbair Malik, Sher Ali & Romana Ishrat

Turner Syndrome (TS) is a condition where several genes are affected but the molecular mechanism remains unknown. Identifying the genes that regulate the TS network is one of the main challenges in understanding its aetiology. Here, we studied the regulatory network from manually curated genes reported in the literature and identified essential proteins involved in TS. The power-law distribution analysis showed that TS network carries scale-free hierarchical fractal attributes. This organization of the network maintained the self-ruled constitution of nodes at various levels without having centrality-lethality control systems. Out of twenty-seven genes culminating into leading hubs in the network, we identified two key regulators (KRs) i.e. KDM6A and BDNF. These KRs serve as the backbone for all the network activities. Removal of KRs does not cause its breakdown, rather a change in the topological properties was observed. Since essential proteins are evolutionarily conserved, the orthologs of selected interacting proteins in *C. elegans*, cat and macaque monkey (lower to higher level organisms) were identified. We deciphered three important interologs i.e. KDM6A-WDR5, KDM6A-ASH2L and WDR5-ASH2L that form a triangular motif. In conclusion, these KRs and identified interologs are expected to regulate the TS network signifying their biological importance.

Search of disease related genes has gained momentum during the past one decade. Perhaps this is due to ever growing newer diseases on the horizon or perhaps a drastic change in the lifestyle of the people or both. Finding the cure of the disease requires its identification and diagnosis well on time. Though an increase in discovering disease-associated genes have been observed with time, there is still a large fraction of diseases without a known molecular basis. Network-based analysis of proteins has attained ample attention in recent years. Network-based approaches serve as potent prognostic tools that have been successfully applied in the characterization of genes in complex diseases such as cancer, ataxia, multiple sclerosis etc.¹⁻⁴. Understanding gene regulatory network advances our knowledge regarding initiation and progression of disease. Such studies will augment system biology research enhancing the efficacy of different therapeutic approaches.

Turner Syndrome (TS) is one such condition where a partial knowledge of its molecular basis is known but a large proportion of their associated candidate genes are still unknown. It is a rare chromosomal disorder affecting females where an X chromosome of a female is partly or completely missing due to sporadic chromosomal non-disjunction⁵. This results in XO condition. Besides the cases of monosomy X (45, X) being the most common in TS, several cases have also been reported with mosaicism, where 45, X cell line is accompanied by one or more other cell lines having a complete or structurally abnormal sex chromosomes (X or Y)⁶. Signs and symptoms of TS are highly variable differing dramatically from one person to another. Thus, the association between genotype and phenotype also remains a challenge. It is known that mosaic cases show up milder phenotypic anomalies compared to those with 45, X karyotype. Of all the observed symptoms, short stature and gonadal dysgenesis remains the most consistent one⁶. The other phenotypes of TS can be associated to the features that are less frequent such as cardiovascular congenital defects, aorta anomalies, renal alterations, cognitive inability that includes selective non-verbal deficiencies etc. Mental deficiency is not a characteristic of TS. It is believed that additional,

Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, 110025, India. Correspondence and requests for materials should be addressed to R.I. (email: romana05@gmail.com)

SN	Gene Name	Description	Location	References
1)	SHOX	Short Stature Homeobox	Xp22.33 and Yp11.2	5,35
2)	SRY	Sex-determining Region Y	Yp11.2	36,37
3)	KDM6A	Lysine Demethylase 6 A	Xp11.3	12,38
4)	TSPY1	Testis specific protein, Y-linked 1	Yp11.2	39
5)	RPS4X	Ribosomal protein S4, X-linked	Xq13.1	40,41
6)	RPS4Y1	Ribosomal protein S4, Y-linked	Yp11.2	42
7)	CSF2RA	Colony Stimulating Factor 2 Receptor Alpha	Xp22.33 and Yp11.2	43,44
8)	PRKX	Protein Kinase, X linked	Xp22.33	12,45
9)	ZFYVE9	Zinc finger FYVE domain-containing protein 9	1p32.3	12
10)	TIMP1	TIMP metalloproteinase inhibitor 1	Xp11.3	12,46
11)	IGF1	Insulin-like growth factor 1	12q23.2	47
12)	STS	Steroid Sulphate	Xp22.31	43,48
13)	NLGN4X	Neuroigin 4, X-Linked	Xp22.32-p22.31	48,49
14)	MTHFR	Methylenetetrahydrofolate reductase	1p36.22	50
15)	GHR	Growth Hormone Receptor	5p13.1-p12	51
16)	BDNF	Brain derived Neurotrophic Factor	11p14.1	13
17)	VDR	Vitamin D (1,25- dihydroxyvitamin D3) receptor	12q13.11	52
18)	AR	Androgen Receptor	Xq12	53
19)	FOXP3	Forkhead box P3	Xp11.23	54
20)	KCNH2	Potassium voltage-gated channel subfamily H member 2	7q36.1	55
21)	SCN5A	Sodium voltage-gated channel alpha subunit 5	3p22.2	55
22)	IGFBP3	Insulin like growth factor binding protein 3	7p12.3	56
23)	PTPN22	Protein Tyrosine Phosphatase, non-receptor type 22	1p13.2	57
24)	XIAP	X-Linked Inhibitor of Apoptosis	Xq25	58
25)	AMH	Anti-Müllerian Hormone	19p13.3	59
26)	PTPN1	Protein Tyrosine Phosphatase, Non-Receptor Type 1	20q13.13	60
27)	DAZ1	Deleted in azoospermia 1	Yq11.223	36
28)	USP9X	Ubiquitin Specific Peptidase 9, X-Linked	Xp11.4	48,61
29)	TMEM27	Transmembrane protein 27	Xp22.2	48,62
30)	EFHC2	EF-Hand Domain Containing 2	Xp11.3	63
31)	SOCS2	Suppressor Of Cytokine Signalling 2	12q22	56

Table 1. List of manually curated genes involved in TS. *The dosage sensitive X linked genes are highlighted in bold.

as-yet-unidentified genes on the X chromosome and some autosomal genes may play a role in the development of these co morbidities of TS. Identification of these factors is still lacking and needs to be evaluated.

The current approach for the prioritization of disease genes is mainly centred on the ‘guilt-by-association’ assumption, which means that the physically and functionally related genes are likely to be involved in the same biological pathways having comparable effects on the phenotypes⁷. Network theory is an important approach to understand topological properties and the dynamics of complex systems to co-relate with their functional modules. Most of the existing networks may be categorised into one of them, namely, scale-free, small world, random and hierarchical network. Amongst them, hierarchical type of network is of special interest to biologists as it includes the appearance of modules and sparsely distributed hubs regulate the network.

Based on this understanding, the key regulators of the TS were identified by integrating protein-protein interaction (PPI) network in the present study. Our current paradigm for studying TS revolves around the identification of the key regulators of TS among manually curated genes by combining protein interactions, functions, disease networks and orthologs. We also aim to understand its topological properties to predict important key regulators among which some are of fundamental importance for their activities and regulating mechanism.

Results

Data mining and curation of genes related to Turner Syndrome. Through literature search, a list of thirty-one genes reportedly involved in TS and its related comorbidities was obtained (Table 1). These genes will be used for network construction and to study their biological significance.

Turner Syndrome network follows hierarchical scale free features. The candidate genes listed in Table 1 were used to construct their regulatory network of which only twenty-seven genes participated as leading hubs. The main constructed network consisted of 3294 nodes and 97361 edges. The topological properties used to characterise the structural and organizational topographies of the TS network are probability of degree distribution $P(k)$, clustering co-efficient $C(k)$ and neighborhood connectivity $C_N(k)$. These properties could perhaps relate to the functional and self-similar (fractal) constituents of the network. It was observed that these

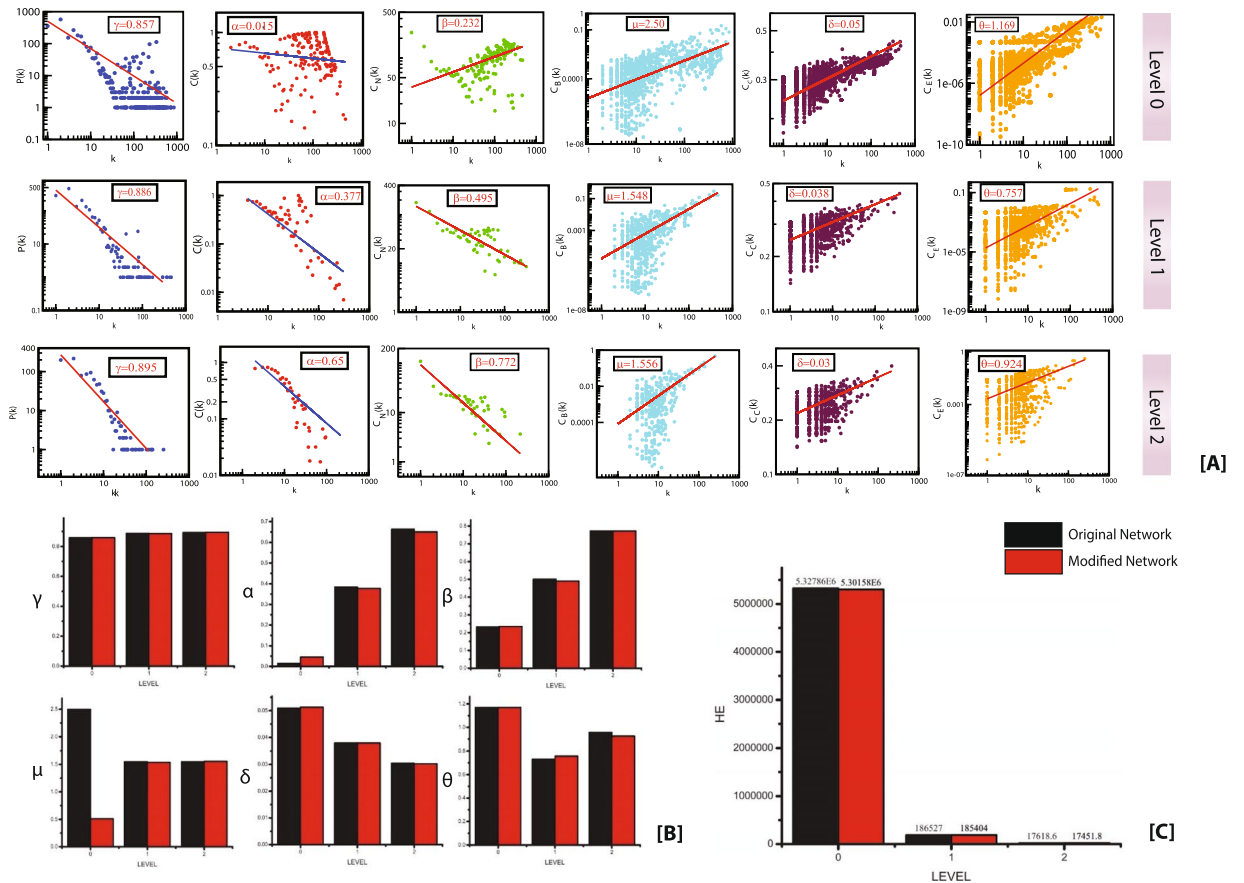


Figure 1. (A) The behaviours of degree distributions ($P(k)$), clustering co-efficient ($C(k)$), neighborhood connectivity ($C_N(k)$), betweenness ($C_B(k)$), closeness ($C_C(k)$) and eigen-vector ($C_E(k)$) measurements as a function of degree k for original and BDNF knockout network at different levels of organization. (B) The changes in the exponents of the six topological parameters due to BDNF knock-out experiment. (C) Changes in the Energy distribution in the network quantified by Hamiltonian calculation as a function of network levels in original and BDNF knockout network.

topological properties obey power law behaviour as a function of degree k (Figs 1(A) and 2(A) first row against Level 0). The power law fits on the data sets of the topological variables of the network are done and verified following a standard statistical fitting procedure proposed by Clauset *et al.*⁸. Here, all statistical p-values for all data sets, calculated against 2500 random samplings, are found to be larger than a critical value 0.1, and goodness of fits is found to be less than and equal to 0.33. The values of the exponents are obtained from the power law fittings. The results for the complete network are summarized as follows,

$$\begin{bmatrix} P \\ C \\ C_N \end{bmatrix} \sim \begin{bmatrix} k^{-\gamma} \\ k^{-\alpha} \\ k^{+\beta} \end{bmatrix} \rightarrow \begin{bmatrix} 0.857 \\ 0.015 \\ 0.232 \end{bmatrix} \quad (1)$$

The negative values of α and γ suggest that the TS network follows hierarchical nature. The value of $\gamma = \ln(2.6)/\ln(3)$ which is 0.857, means that number of nodes increase with the advancement of disease as a power of 2.6 while links as a power of 3 thus, giving us the idea of being hierarchical as it shows presence of modules in our clustering experiment. The positive value of β indicates that the network carries the assortive mixing specifying that a large cluster of degree nodes (formation of rich club) regulates the TS network.

The centrality measurements, namely betweenness centrality $C_B(k)$ and closeness centrality $C_C(k)$ represent the flow of information in the network and predict the influential candidates in the network. The well connectedness of nodes in a network is characterised by eigenvector centrality $C_E(k)$. It measures the efficacy of the spreading (receiving) power of information of nodes from the network. These properties obey power law behaviours as follows,

$$\begin{bmatrix} C_B \\ C_C \\ C_E \end{bmatrix} \sim \begin{bmatrix} k^\mu \\ k^\delta \\ k^\theta \end{bmatrix} \rightarrow \begin{bmatrix} 2.50 \\ 0.051 \\ 1.16 \end{bmatrix} \quad (2)$$

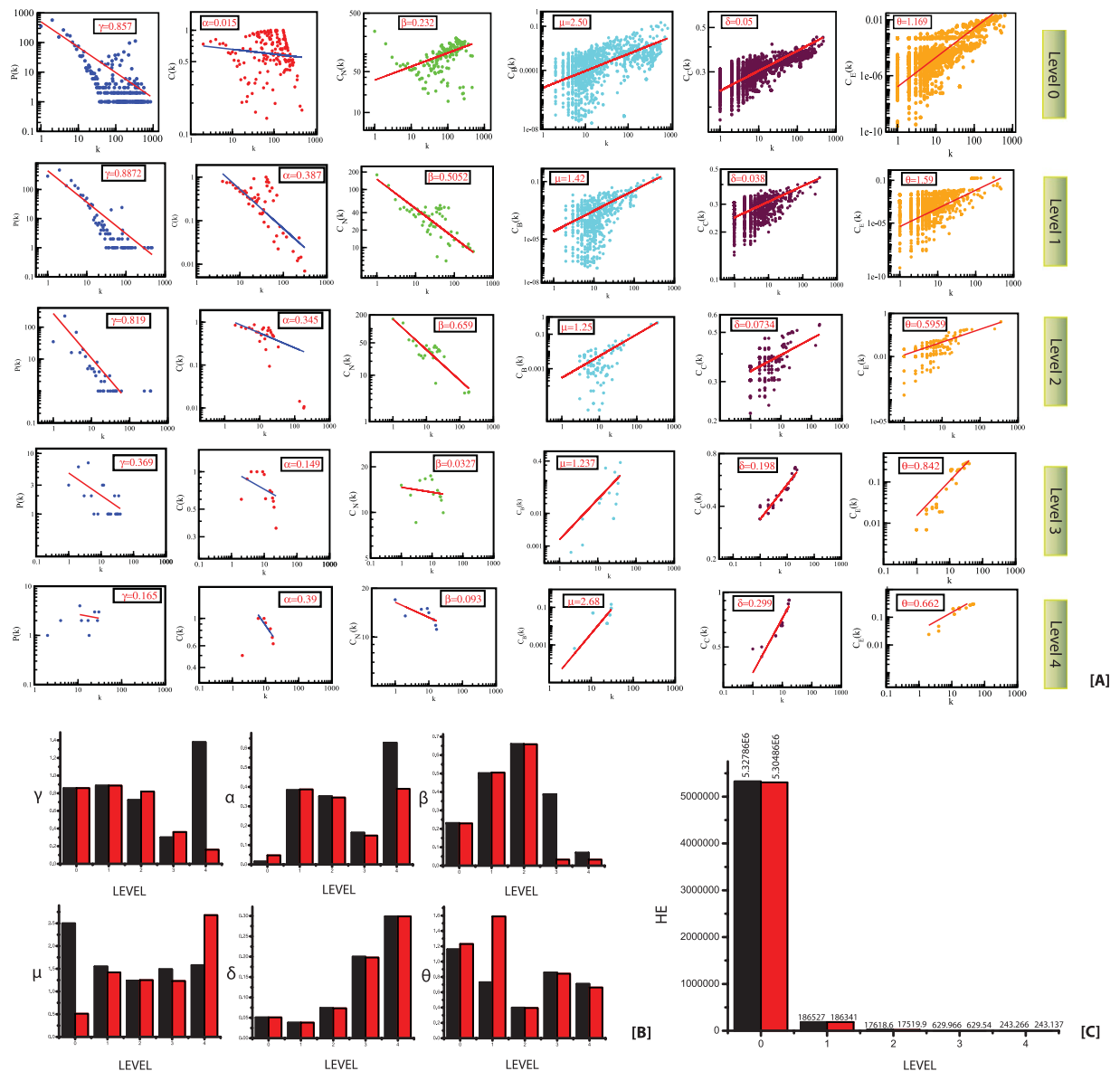
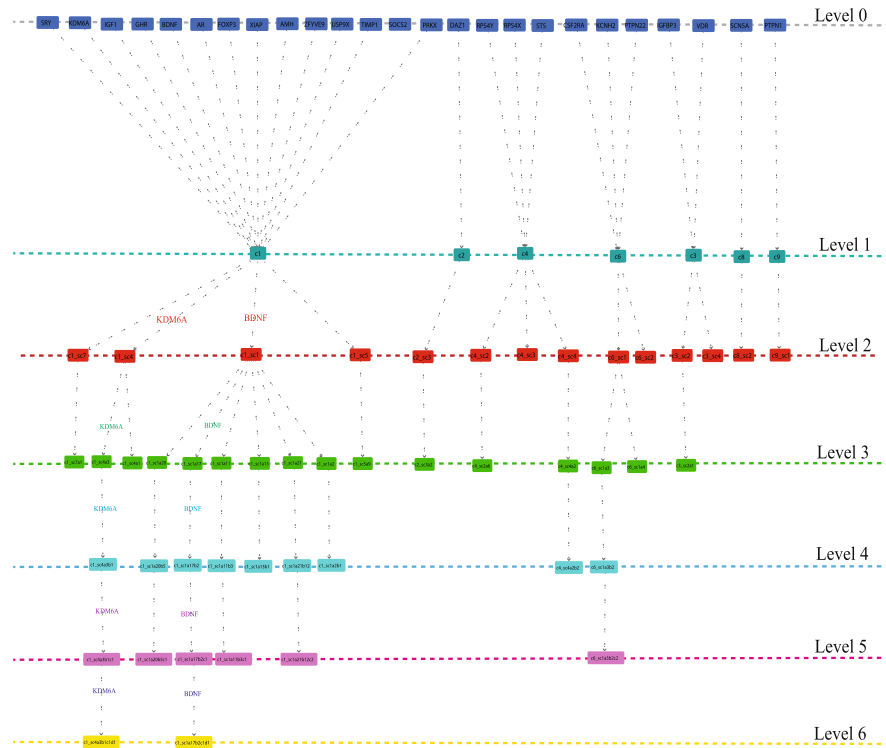


Figure 2. (A) The behaviours of degree distributions ($P(k)$), clustering co-efficient ($C(k)$), neighborhood connectivity ($C_N(k)$), betweenness ($C_B(k)$), closeness ($C_C(k)$) and eigen-vector ($C_E(k)$) measurements as a function of degree k for original and KDM6A knockout network at different levels of organization. (B) The changes in the exponents of the six topological parameters due to KDM6A knock-out experiment. (C) Changes in the Energy distribution in the network quantified by Hamiltonian calculation as a function of network levels in original and KDM6A knockout network.

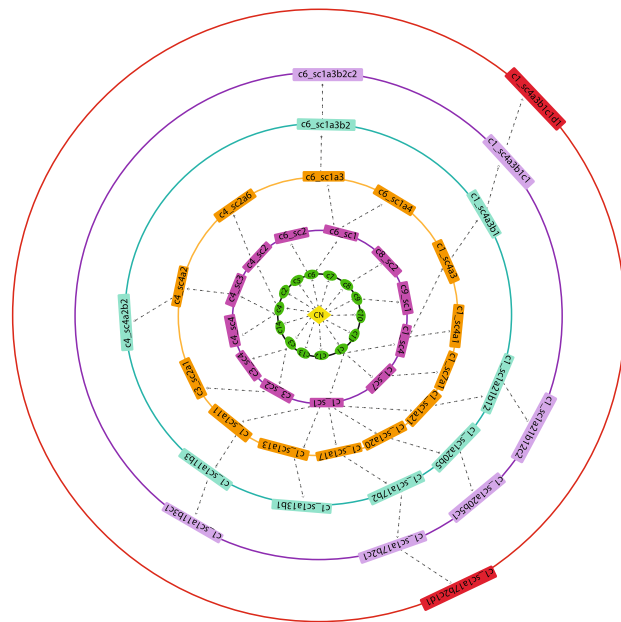
The power law natures of the three centrality measurements are again verified and confirmed using the procedure of Clauset *et al.*⁸ of statistical power law fitting. Here p values are found to be more than 0.1 and goodness of fit larger than 3.5. It was observed that only few numbers of higher degree nodes have large centrality values (for all three centrality measurements). The number of most influencing hubs, which can control the network, is few. Therefore, the TS network is dominated by the low degree nodes (genes/proteins) and these low degree nodes regulate the functioning and organization of the network. Few of the sparsely distributed leading hubs might, however, take important participation in regulating as well as maintaining the network stability. The positive values of these centrality measurements show that the network exhibits hierarchical scale free or fractal features.

Thus, the overall topological properties of the TS network indicate that the same self-organise into a scale free fractal state and have hierarchical organization, i.e. they are composed of successive interconnected or inter-nested communities.

Identification of key regulators and properties. The modular structure and their arrangement at various levels of organization are done following Newman and Girvan's standard community finding algorithm⁹. Using this algorithm, the TS network is found to be hierarchically organised through six different levels



[A]



[B]

Figure 3. (A) Tracing of KRs through different levels in the network. (B) Organization of the modules/sub-modules of the network.

(Fig. 3(B)). The corresponding modularity Q_N and Hamiltonian Energy (HE) as a function of levels of organization are found to be decreased as one goes from top to down organisation (Fig. 4(A,B), respectively).

Here we put forward the idea of *key regulators* (KRs) as the genes/proteins which are deeply rooted from top to bottom of the network and vice versa, which serve as the backbone of the network organization. It is not essential for these KRs to be the large leading hubs in the network, but they rather change their popularities randomly at various levels of organization (Fig. 5). Since the network qualifies hierarchical characteristics, the removal of the leading hubs will not cause its breakdown. However, the removal of KRs from the network may cause maximum local and global perturbations, especially at a deeper level of organization. These perturbations will propagate through various levels of organization's bottom to top or top to bottom causing topological change in the network. Thus, these KRs could be the possible key target genes of the TS network.

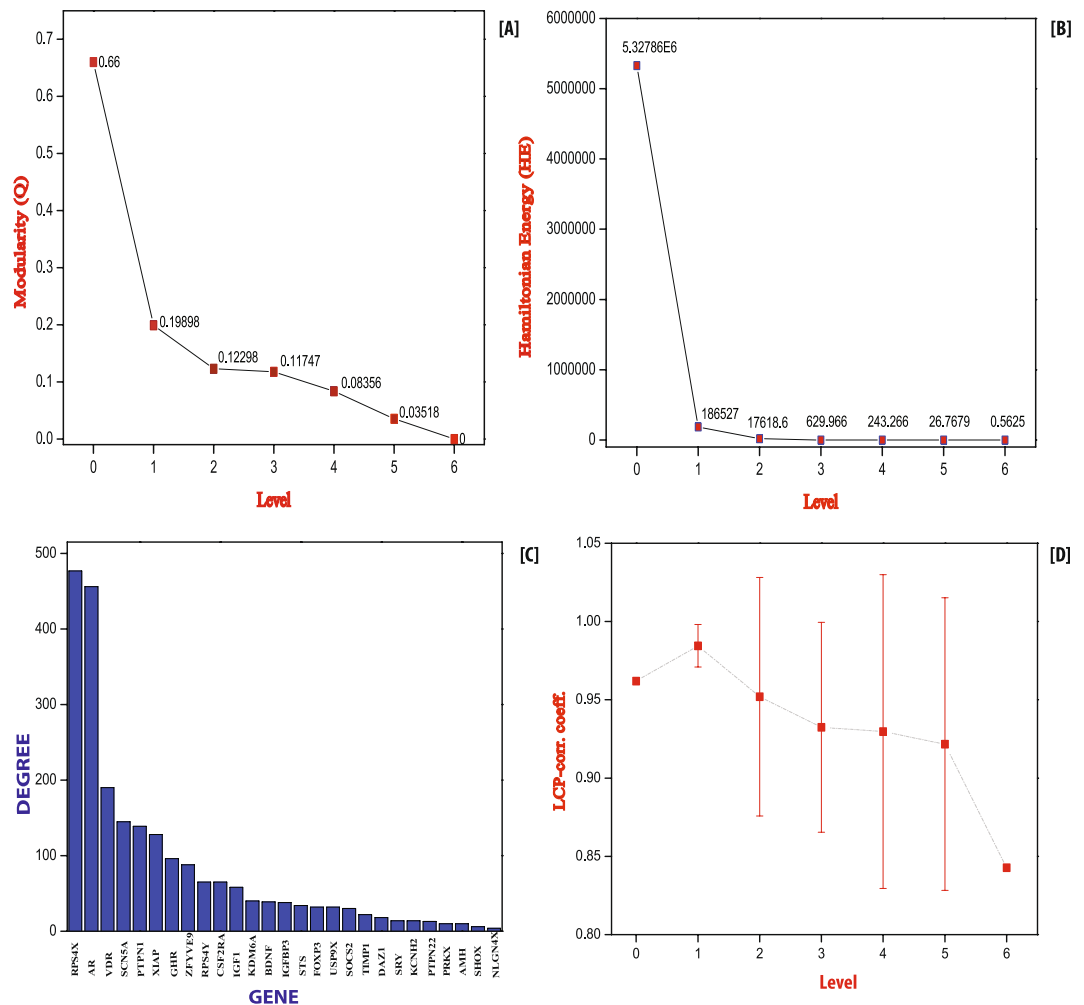


Figure 4. (A) Corresponding modularity Q_N as a function of levels of organization. (B) Corresponding Hamiltonian Energy (HE) as a function of levels of organization. (C) Characterization of twenty-seven leading hubs of the network by degree (D). Variation in the calculated average LCP-corr for TS network as a function of network level.

Following the definition of KR, we identified two KRs, namely, KDM6A and BDNF (Fig. 5), which are key regulators/organizers of the TS network. Unexpectedly, the first eleven leading hubs are not found to be KRs since they fail to reach till the deepest level of organization (Figs 4(C) and 5). These two KRs maintain a low profile/popularity thereby regulating the network till the bottom level of organization. These KRs separate from each other after level one and then move separately till the motif level (Fig. 3(A)). These KRs may act as signal propagators from top to bottom and vice versa to maintain network stability and inherent properties. The proteins SOCS2, GHR, CSF2RA and PRKX move till fifth level but fail to reach the motif level (Fig. 5).

To understand the regulating ability of each of the 2 KRs, we calculated the Probability $P_x(y^l)$ of KR (Fig. 5).

$$P_x(y^l) = \frac{y^l}{E^l} \quad (3)$$

Where x is number of edges y^l at level l and E^l is total number of edges of the network or modules and sub-modules. The calculated Probability $P_x(y^l)$ of all the key genes show an increase in P_x as one goes top to bottom direction as level l increases. This means the regulating ability of each key gene becomes more important at deeper level of organization.

Evidence of self-organization: local-community-paradigm (LCP) approach. The LCP architecture assists not only the rapid delivery of information across the various network modules, but also through the local processing. We analysed the TS network to check the maintenance of its self-organization at various levels of network organization using LCP technique. The LCP-corr of all the modules/sub-modules at various levels was calculated. The average values of LCP-corr at each level (modules having zero LCP-corr are not taken in average) are greater than 0.85 and the values do not change with error bar (Fig. 4(D)). This indicates that the network

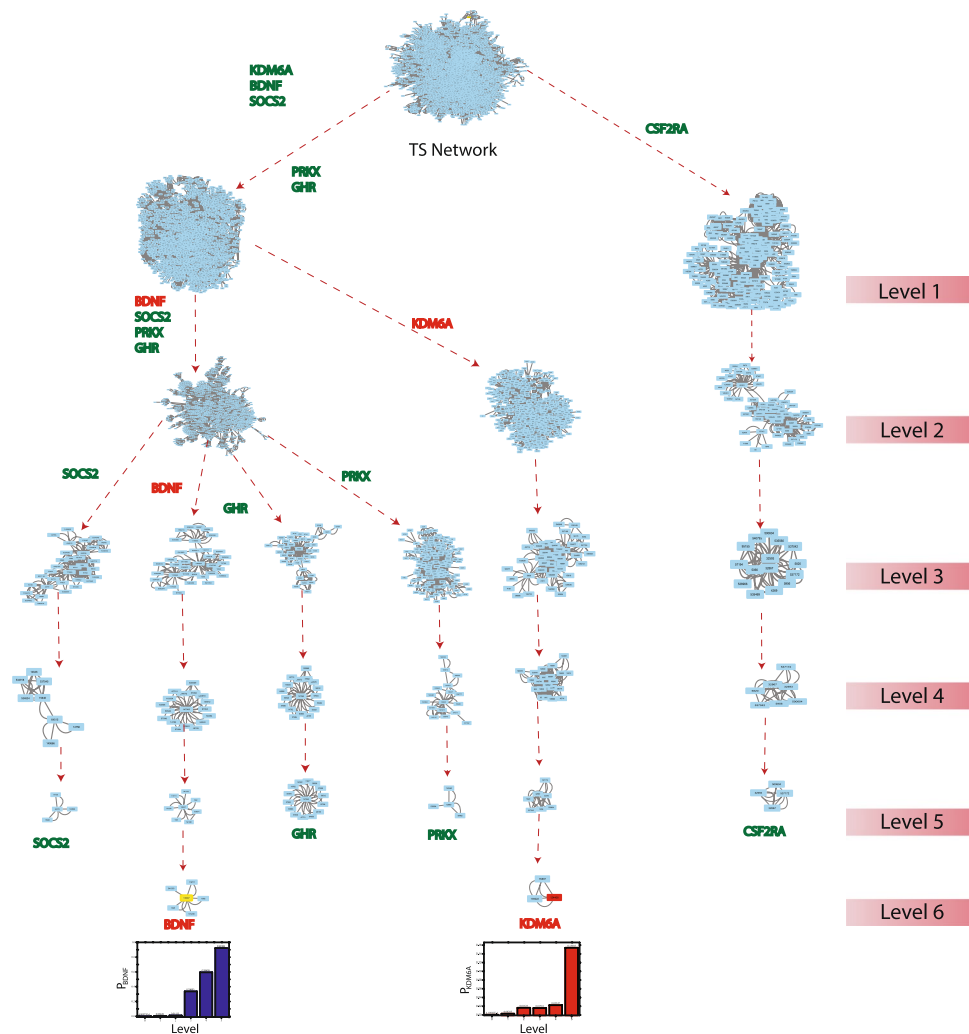


Figure 5. Network/modules/sub-modules at different network levels which accommodate leading hubs and key regulators. The probability distribution of the KRs as a function of level.

maintains self-organisation and is compact and has efficient information processing. It represents strong LCP networks that are dynamic and heterogeneous, which facilitate network evolution and reorganization.

Local perturbations driven by key regulators. The knock-out experiment of the KRs from the TS network highlights the local perturbations driven and their consequence on global network properties. The knockout experiment for both the KRs was performed separately. In both the cases a significant change in the topological properties of the network is observed (Figs 1(A) and 2(A)). It was seen that α change significantly at level 0, whereas β and γ change only slightly (Figs 1(B) and 2(B)). Similarly, the changes in the exponents of centrality measurements (δ , μ and θ) at level 0 also show significant change (Figs 1(B) and 2(B)). The values of δ and θ changes slightly whereas a significant change is observed in the value of μ . It is evident from the changes in the exponents of topological parameters that as one goes to deeper level i.e. top to down direction the network perturbation increases (Figs 1(B) and 2(B)). In case of KDM6A, after the fourth level, its removal almost breaks down the sub-modules present in the remaining deeper levels. Whereas in case of BDNF, after the second level, its removal almost breaks down the sub-modules present in the remaining deeper levels. This indicates that local perturbation is maximum at deeper levels starting from bottom to top.

We then calculate Hamiltonian energy of the respective complete network and modules/sub-modules in the KR knockout experiment to understand change in energy distributions in the respective network. A slight decrease in the Hamiltonian energy is observed at each level due to knockout of the KRs (Figs 1(C) and 2(C)). This indicates that the removal of KRs cause enormous loss of wiring/rewiring energy which is propagated throughout the levels of network organization.

Centrality-Lethality in Turner Syndrome network. The TS network is close to hierarchical network and hence the modules/sub-modules emerged are compact at upper levels of organization. The knockout of KRs (KDM6A and BDNF) from TS network does not cause the network breakdown. As observed in case of KDM6A

Hub Gene	Interacting Partner	Location of Interacting partner
SRY	HDAC3 (Histone Deacetylase 3)	5q31.3
RPS4Y1	RPS3 (40S ribosomal protein S3)	11q13.4
KDM6A	WDR5 (WD repeat-containing protein 5)	9q34.2
	ASH2L (Ash2 histone methyltransferase complex subunit ASH2)	8p11.23
BDNF	MBTP1 (Membrane-bound transcription factor site-1 protease)	16q23.3-q24.1
	CAPS2 (Calcium-dependent secretion activator 2)	12q21.1-q21.2
	CPE (Carboxypeptidase E)	4q32.3
	NOS3 (Nitric oxide synthase 3, endothelial)	7q36.1

Table 2. Essential PPI interactions in TS network.

and BDNF, their knockout almost breaks down the sub-modules from fourth and second level respectively, whereas other modules/sub-modules remain stable to preserve the network properties. Hence, TS network rules out centrality-lethality rule¹⁰. However, the identified KR's have important regulating activities in the network which is reflected in the changes in the topological properties and other parameters of the network and its associated communities at various levels of organization.

Predicting essential protein interactions: Interologs. Following the centrality measurements-based methodology (see in Methodology), we examined the top thirty genes each one identified by its centrality and degree measurements. We got 120 genes from all measurements (Supplementary Table S1). Among these 120 central genes, 2 genes (HDAC3 and RPS3) were found to be the interacting partners of the hub genes, SRY and RPS4Y respectively in TS network.

Analysing the interacting partners of the disease associated genes play an important role in the prediction of genotype-phenotype associations and helps in identifying new disease genes candidates (i.e. the genes coding for the interacting proteins are putative disease-causing genes). It is thus expected that the interacting partners of the key regulators may also be associated with TS. Built on this assumption 8 PPI were selected for further analysis listed in Table 2.

One of the most important features of essential proteins is their conservative property. Many studies have shown that essential proteins evolve much slower than the other proteins. They are more evolutionarily conserved than non-essential proteins¹¹. By considering the facts that essential proteins depend not only on the interactions between proteins but also their orthologous properties, we find orthologs of the proteins listed in Table 2 in different species namely, *C. elegans*, cat and macaque monkey (Supplementary Table S2). If two proteins physically interact in one species and they have orthologous counterparts in another species, it is likely that their orthologs interact in that species too. If such conserved interactions exist, they are called interologs.

The network of *C. elegans*, cat, macaque monkey, and human were constructed from these 8 essential interacting proteins as hub genes for the analysis of the interologs (Fig. 6). Though it is expected that all these interactions play important role in TS network regulation, it was observed that only two protein-protein interactions (highlighted in bold in Table 2) remained conserved at each level of organism. Also, it is noteworthy that both these interactions involve the key regulator KDM6A and these proteins form a triangular motif (Fig. 6). These predicted interologs might play major role in the pathophysiology of TS.

Discussion

Understanding the regulation of a disease network is of great relevance towards the development of treatments for various diseases in the field of pharmacogenomics. One of the ways is to identify the key drivers that regulate the complete network. We have attempted to construct TS network focusing on genes that are regulated through this network.

The TS network constructed from manually curated set of genes show hierarchical features, which means that the network has system level organization involving modules/communities which are interrelated. Since the network is hierarchical, individual gene activities are not of much significance, rather their synchronisation exhibits various important functional regulation of the network. Significant genes (leading hubs) were recognized as key regulators of the network by influencing motifs and module regulation, indicating their biological significance. The leading hubs have significantly important functions. They integrate the lower degree nodes for organizing and regulating activities like inter and intra cross-talk among various other essential genes, maintaining network properties and stability, and optimizing the network signal processing. However, out of these leading hubs, two are key regulators, playing important roles in keeping network in order. In TS network, out of twenty-seven leading hubs, we identified two such KR's which are KDM6A and BDNF. KDM6A is an X-linked gene that plays a central role in coding the histones. It escapes X-chromosome inactivation suggesting that it could be a potential candidate gene of TS¹². Also, it has been reported that this gene may be involved in premature ovarian failure. BDNF gene is localised on chromosome 11 and is a member of the neurotrophin family of growth factors. It has been reported that TS patients have higher BDNF levels than healthy ones¹³. These KR's serve as the backbone for any network activities and their regulations and could be a possible target gene for disease control. Surprisingly, these KR's do not fall in first few largest hubs (eleven hubs) and thus keep a low profile in the network. Since the network possesses hierarchical properties, removal of KDM6A and BDNF does not cause network breakdown, instead the network adapt itself functionally.

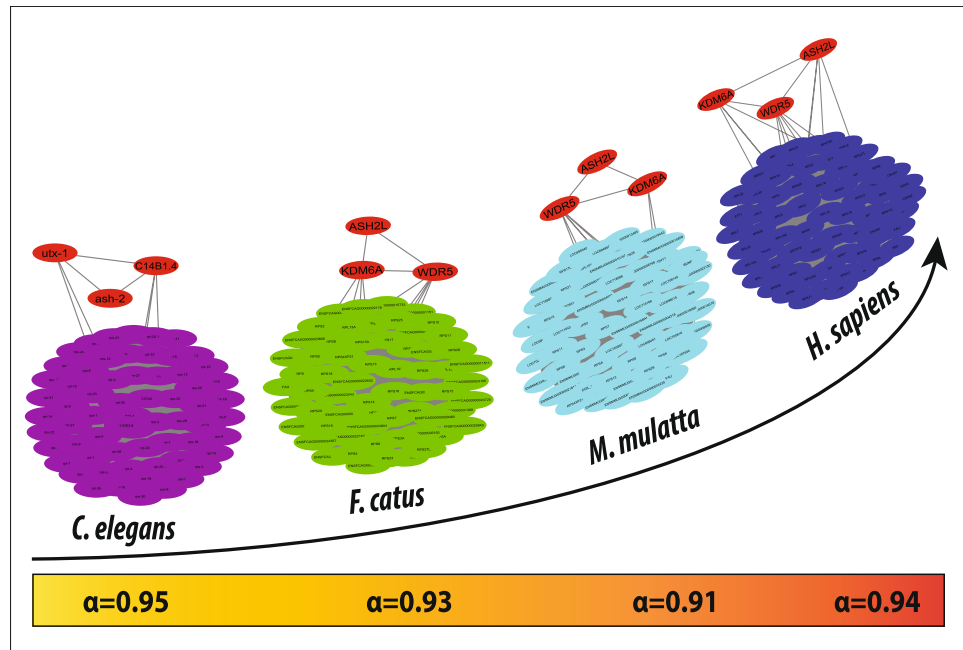


Figure 6. Interologs in the network from lower to higher organism. α is the clustering coefficient of the network.

The topological properties of the network exhibit power law behaviour indicating that the network obeys fractal nature. This could be a signature of self-organization in the network. It was also studied that TS represents strong LCP networks that are dynamic and heterogeneous, which facilitate its evolution and reorganization. This indicates that the network maintains self-organization and is compact and has efficient information processing.

The knock out experiment of the KR indicates that the change in the network properties do not cause significant change in its topology. This indicates that the system does not prefer a change due to perturbation imparted by KR knock out. The network reorganizes itself and adapt according to the changed topological properties. The ability to adapt for a better network organization without breakdown of the system is another signature of self-organization¹⁴. By considering the facts that essential proteins are more evolutionarily conserved than non-essential ones and essential proteins frequently bind each other, we predicted the essential proteins by integrating the orthology with PPI. We deciphered three important interologs (evolutionarily conserved protein-protein interactions) i.e. KDM6A-WDR5, KDM6A-ASH2L and WDR5-ASH2L, thus forming a triangular motif. KDM6A already has been found to be the key regulator of TS network, therefore these interologs are expected to play major role in TS.

It is likely that the above mentioned genotypic constitution is operative in normal female. However, the Table 1 shows presence of several Y-linked genes including SRY. This raises a question if this scenario is applicable to males also. Alternatively, this may be operative in both the sexes for normal function of both the genome. While this seems to be simple and more convincing proposition, it raises yet another question as we know no two Turner patients are alike. Thus, genotypically every Turner is unique with respect to its genotype. If we superimpose on different phenotype of Turner patient, will that eventually generate a consensus on regulatory protein and their interacting partner. While this may be a logical expectation, it still poses yet another challenge both logistical and experimental ones. It would be extraordinarily informative if every single gene listed in Table 1 is analysed with respect to its normal vs mutational status, copy number variation, expressional dynamics and interactomes involved therein. Such analysis would surely augment the understanding on the mechanism of formation of TS and perhaps identification of key regulatory genes. Information on these lines is envisaged to be useful for a better diagnosis and prognosis of TS.

Methodology

Data mining and curation of genes related to Turner Syndrome. The information of TS related genes and proteins were collected from the literatures. The genes expected to be involved in TS and related co-morbidities were manually curated from various sources like repositories (PIR)¹⁵, reviewed literatures (i.e. PubMed), OMIM¹⁶ etc.

Construction of Gene Regulatory Network. For the construction of the primary network of the expressed proteins, the curated genes were mapped to their respective UniProt IDs and their associated functional information were retrieved. The simple concept of one gene, one protein was used to build the gene regulatory network of TS.

The network was constructed with available large-scale protein interaction networks using Pathway Commons^{17,18} (a built-in cytoscape feature) and then visualised in cytoscape 3.4 version¹⁹. Pathway Commons is a metasearch platform, which, in addition to PPI data retrieval, also collects pathway data from multiple publicly available databases, including REACTOME, Systems Biology Centre New York (<http://sbcny.org/data.htm>), HumanCyc (<http://humancyc.org/>), The Cancer Cell Map (cancer.cellmap.org) and PID. Pathway Commons includes biochemical reactions, complex assembly, transport and catalytic events and physical interactions. It can easily be accessed either directly online, or through Cytoscape's built-in import "Network from Web Services" function.

Characterization of Topological Properties of Networks. The structural properties of complex networks are characterized through the behaviours of the topological parameters. The following topological properties of the networks (graph) were studied to learn the important behaviours of the network: Degree distribution, Neighborhood connectivity, clustering co-efficient, Betweenness centrality, Closeness centrality and Eigenvector centrality.

Degree distribution. In a network, the degree k is a centrality measure that represents the number of links the node connects with other nodes. For a network defined by a graph $G = (N, E)$, where N and E are number of nodes and edges respectively, the probability of degree distribution ($P(k)$) of the network is the ratio of the number of nodes having degree to the network size;

$$P(k) = \frac{n_k}{N} \quad (4)$$

Where, n_k is the number of nodes having degree k and N is the total number of nodes in the network. $P(k)$ indicates the importance of hubs or modules in the network. It obeys power law $P(k) \sim k^{-\gamma}$ in scale-free and hierarchical networks depending on the value of γ which specifies the importance of hubs or modules in the network²⁰.

Neighborhood connectivity. The number of neighbors of a node is considered as its connectivity. The neighborhood connectivity of a node n is defined as the average connectivity of all neighbors of n ²¹. In the network ($C_N(k)$) Neighborhood connectivity is given by,

$$C_N(k) = \sum_q q P\left(\frac{q}{k}\right) \quad (5)$$

where, $P\left(\frac{q}{k}\right)$ is the conditional probability that a link belonging to a node with connectivity k points to a node with connectivity q . The positive power dependence of $C_N(k)$ could be an indicator of assortivity in the network topology.

Clustering co-efficient. This topological parameter of a network represents the measure of the interconnection of a node with its neighborhood node and strength of its interconnection. It is the ratio of the number of its nearest neighborhood edges e_i to the total possible number of edges of degree k_i . For an undirected network, clustering co-efficient ($C(k_i)$) of i th node can be calculated by,

$$C(k_i) = \frac{2e_i}{k_i(k_i - 1)} \quad (6)$$

Betweenness centrality. Betweenness centrality C_B of a node represents the prominence of information flow in the network, and the extent to which the node has control over the other nodes in the network through communication^{22,23}. If $d_{ij}(v)$ indicates the number of geodesic paths from node i to node j passing through node v , and d_{ij} indicates number of geodesic paths from node i to j , then betweenness centrality ($C_B(v)$) of a node v can be calculated by,

$$C_B(v) = \sum_{i,j,i \neq j \neq k} \frac{d_{ij}(v)}{d_{ij}} \quad (7)$$

Closeness centrality. Closeness centrality (C_C) measures how fast information is spread from a node to other nodes accessible from it in the network²⁴. C_C of a node i is the reciprocal of the mean geodesic distance between the node and all other nodes connected to it in the network, and is given by,

$$C_C(i) = \frac{n}{\sum_j d_{ij}} \quad (8)$$

where d_{ij} represents the geodesic path length from nodes i to j , and n is the total number of vertices in the graph reachable from node i .

Eigenvector centrality. Eigenvector centrality of a node i ($C_E(i)$) in a network is proportional to the sum of i 's neighbor centralities²⁵, and it is given by,

$$C_E(i) = \frac{1}{\lambda} \sum_{j=nn(i)} v_j \quad (9)$$

where, $nn(i)$ indicates nearest neighbors of nodes i in the network. λ is eigenvalue of the eigenvector v_i is given by, $Av_i = \lambda v_i$ where, A is the adjacency matrix of the network (graph). The principal eigenvector of A , which corresponds to maximum eigenvalue λ_{max} , is taken to have positive eigenvector centrality score. Eigenvector centrality can be used as an indicator of node's spreading power in the network.

Community detection/finding: Leading Eigen-vector method. The activities of the constructed network were defined at various levels of hierarchy to describe the modular nature, properties and the organizing principle of the hierarchical network. To detect the communities, the Leading Eigen Vector method (LEV)^{26,27} was used in R from package 'igraph'²⁸ in this study. The LEV method is the most promising one for community detection as it calculates the Eigen value for each link, exemplifying the significance of each link, not nodes. To obtain only motif, we detected modules from complete network and then sub-modules from the modules at each level of organization.

Modularity. Modularity is the measure of how fine a network is divided in communities⁹. Modularity (Q) is expressed as follows,

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j) \quad (10)$$

where m is the total number of edges in the community, A_{ij} is the adjacency matrix of size $i \times j$, k represents degrees, and the δ function yields 1 if nodes i and j are in the same community.

Tracing of the Genes and Knock out Experiment. One particular challenge is to identify the main drivers that control the regulation of TS network. This was done through tracing of genes. This tracing of genes was performed up to motif level in various modules/sub-modules obtained from clustering. Through tracing the most significant and influential nodes within the network constructed was identified that regulates the network.

Further, the change in the organization of the network in the absence of these significant nodes was observed through the knockout experiment. The identified key regulators were successively removed from the constructed complete network, and the topological properties of the modified network were calculated again to describe the perturbations caused within the network due to the absence of these key regulators. The knockout experiment was repeated at different level of network organization to comprehend the role of these key genes in the network. The topological properties of the network were calculated using Network Analyzer, a plug-in in Cytoscape version 3.4, whereas for eigen value calculation, we used CytoNCA²⁹, another plug-in in Cytoscape for topological properties.

Local-community-paradigm (LCP) approach: Compactness of the network. LCP-Decomposition Plot (LCP-DP) is an approach to represent topological self-organisation as a local-community-paradigm (LCP), and consequently is used to visualise and examine the effect of LCP on network topology. It is a function of common neighbors (CN) index of interacting nodes and local community links (LCL) of each pair of interacting nodes in the network. It provides information on number, size and compactness of the communities in a given network³⁰. The CN index between two nodes x and y can be calculated from the measure of overlapping between their sets of first-node-neighbors $S(x)$ and $S(y)$ given by, $CN = S(x) \cap S(y)$. If there is significant amount of overlapping between the sets $S(x)$ and $S(y)$ (large value of CN), the possible likelihood of interaction of these two nodes could happen and so an increase in CN represents the rise in compactness of the network, showing its faster information processing abilities. Further, the LCLs between the two nodes x and y , whose upper bound is defined by, $\max(LCL) = \frac{1}{2}CN(CN - 1)$, is the number of internal links in local-community (LC), which is strongly inter-linked group of nodes. Most probably these two nodes link together if CN of these two nodes are members of LC³⁰. LCP-DP has been found to have a linear dependence between CN and \sqrt{LCL} .

The LCP correlation (LCP-corr) is the Pearson correlation co-efficient between the variables CN and LCL defined by $LCP - corr = \frac{cov(CN, LCL)}{\sigma_{CN}\sigma_{LCL}}$ with $CN > 1$, where $cov(CN, LCL)$ is the covariance between CN and LCL, σ_{CN} and σ_{LCL} are standard deviations of CN and LCL, respectively.

Distribution of energy in the network: Hamiltonian energy calculation. At each level of the network, certain level of energy is maintained that helps organise the network at that level. This is measured by using Hamiltonian Energy (HE) of the network at that level/state within the formalism of Constant Potts Model^{31,32}. HE gives the energy distribution not only at the global level of a network, but also at modular level, which is in the self-organization of the system. HE of a network or module or sub-module can be calculated by,

$$H^{[c]} = -\sum_c [e_c - \gamma n_c^2] \quad (11)$$

Where e_c and n_c are the number of edges and nodes in a community 'c' and γ is the resolution parameter acting as an edge density threshold which is set to be 0.5.

Predicting essential interactions by integrating the Orthology with PPI. Study has shown that there is a positive correlation between essentialities (essential proteins) and topological properties (centralities) of the proteins in PPI networks. As a consequence, a series of centrality measures based on network topological features have been used for identifying essential proteins, such as Degree Centrality, Betweenness Centrality, Closeness Centrality and Eigenvector Centrality. The proteins in the TS network were ranked in terms of their centrality (top 30 in each centrality). Then the ranking scores of these proteins were used to judge whether a protein is essential. Further the interacting partners of the hub proteins that fall into this category were identified.

In view of the facts that essential proteins are more evolutionarily conserved than non-essential proteins and they frequently bind each other, the prediction of essential proteins was done by integrating the orthology with PPI networks. To measure the conservation of the selected interacting proteins, their orthologous proteins in four different species namely, *Caenorhabditis elegans*, *Felis catus* (domestic cat), *Macaca mulatta* (macaque monkey), and *Homo sapiens* were investigated (lower to higher level organisms). Information on orthologous proteins is taken from Version 8 of the InParanoid database³³ (an ortholog database) which contains a collection of pairwise comparisons between 100 whole genomes (99 eukaryotes and 1 prokaryote) constructed by the INPARANOID program. Further the network of *C. elegans*, cat, macaque monkey, and human were constructed considering these 8 essential interacting proteins as hub genes for the analysis of the conserved interactions from STRING Protein Database³⁴. STRING quantitatively integrates interaction data for a large number of organisms, and transfers information between these organisms where applicable.

References

- Lim, J. *et al.* A Protein–Protein Interaction Network for Human Inherited Ataxias and Disorders of Purkinje Cell Degeneration. *Cell* **125**, 801–814 (2006).
- Chuang, H.-Y., Lee, E., Liu, Y.-T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **3**, (2007).
- Pujana, M. A. *et al.* Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat. Genet.* **39**, 1338–1349 (2007).
- Baranzini, S. E. *et al.* Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum. Mol. Genet.* **18**, 2078–2090 (2009).
- Kesler, S. R. Turner Syndrome. *Child Adolesc. Psychiatr. Clin. N. Am.* **16**, 709–722 (2007).
- Muntaj, S., Feroze, A. G., Purva, S. V., Radhika, S. & Tilak P. Karyotypic Variables in Turner Syndrome: A Case Series. *Int. J. Sci. Study* **3**, 171–175 (2015).
- Lee, I., Blom, U. M., Wang, P. I., Shim, J. E. & Marcotte, E. M. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* **21**, 1109–1121 (2011).
- Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-Law Distributions in Empirical Data. *SIAM Rev.* **51**, 661–703 (2009).
- Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **69**, 026113 (2004).
- Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
- Peng, W. *et al.* Iteration method for predicting essential proteins based on orthology and protein-protein interaction networks. *BMC Syst. Biol.* **6**, 87 (2012).
- Trolle, C. *et al.* Widespread DNA hypomethylation and differential gene expression in Turner syndrome. *Sci. Rep.* **6**, 34220 (2016).
- Czyzyk, A. *et al.* Brain-derived neurotrophic factor plasma levels in patients with Turner syndrome. *Gynecol. Endocrinol.* **30**, 245–249 (2014).
- Ashby, W. R. Principles of the Self-Organizing System. in *Facets of Systems Science* 521–536, https://doi.org/10.1007/978-1-4899-0718-9_38 (Springer US, 1991).
- Wu, C. & Nebert, D. W. Update on genome completion and annotations: Protein Information Resource. *Hum. Genomics* **1**, 229–233 (2004).
- Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514–517 (2005).
- Cerami, E. G. *et al.* Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39**, D685–D690 (2011).
- Liu, H., Beck, T. N., Golemis, E. A. & Serebriiskii, I. G. Integrating *In Silico* Resources to Map a Signaling Network. In *Gene Function Analysis* (ed. Ochs, M. F.) **1101**, 197–245 (Humana Press, 2014).
- Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
- Maslov, S. Specificity and Stability in Topology of Protein Networks. *Science* **296**, 910–913 (2002).
- Ulrik Brandes, A. Faster Algorithm for Betweenness Centrality. *J. Math. Sociol.* **25**, 163–177 (2001).
- Mason, O. & Verwoerd, M. Graph theory and networks in Biology. *IET Syst. Biol.* **1**, 89–119 (2007).
- Canright, G. & Kenth, E. M. Roles in networks. *Sci. Comput. Program.* **53**, 195–214 (2004).
- Bonacich, P. Power and Centrality: A Family of Measures. *Am. J. Sociol.* **92**, 1170–1182 (1987).
- Newman, M. E. J. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **74**, (2006).
- Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* **103**, 8577–8582 (2006).
- Gabor, C. & Nepusz, T. The igraph software package for complex network research. *Inter J Comp Syst* **1695**, 1–9 (2006).
- Tang, Y., Li, M., Wang, J., Pan, Y. & Wu, F.-X. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* **127**, 67–72 (2015).
- Cannistraci, C. V., Alanis-Lobato, G. & Ravasi, T. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks. *Sci. Rep.* **3**, (2013).
- Traag, V. A., Van Dooren, P. & Nesterov, Y. Narrow scope for resolution-limit-free community detection. *Phys. Rev. E* **84**, (2011).
- Traag, V. A., Krings, G. & Van Dooren, P. Significant Scales in Community Structure. *Sci. Rep.* **3**, (2013).
- Sonnhammer, E. L. L. & Östlund, G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* **43**, D234–D239 (2015).
- von Mering, C. *et al.* STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* **31**, 258–261 (2003).
- Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A. & Marais, G. A. B. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc. Natl. Acad. Sci.* **109**, 5346–5351 (2012).
- Premi, S., Srivastava, J., Panneer, G. & Ali, S. Startling Mosaicism of the Y-Chromosome and Tandem Duplication of the SRY and DAZ Genes in Patients with Turner Syndrome. *PLoS ONE* **3**, e3796 (2008).
- Ali, S. & Hasnain, S. E. Molecular dissection of the human Y-chromosome. *Gene* **283**, 1–10 (2002).

38. Greenfield, A. The UTX gene escapes X inactivation in mice and humans. *Hum. Mol. Genet.* **7**, 737–742 (1998).
39. de Marqui, A. B. T., da Silva-Grecco, R. L. & Balarin, M. A. S. Prevalence of Y-chromosome sequences and gonadoblastoma in Turner syndrome. *Rev. Paul. Pediatr. Engl. Ed.* **34**, 114–121 (2016).
40. Zinn, A. R. *et al.* Inactivation of the Rps4 gene on the mouse X chromosome. *Genomics* **11**, 1097–1101 (1991).
41. Fisher, E. M. *et al.* Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* **63**, 1205–1218 (1990).
42. Watanabe, M., Zinn, A. R., Page, D. C. & Nishimoto, T. Functional equivalence of human X- and Y-encoded isoforms of ribosomal protein S4 consistent with a role in Turner syndrome. *Nat. Genet.* **4**, 268–271 (1993).
43. Urbach, A. & Benvenisty, N. Studying Early Lethality of 45,XO (Turner's Syndrome) Embryos Using Human Embryonic Stem Cells. *PLoS ONE* **4**, e4175 (2009).
44. Berletch, J. B., Yang, F. & Distèche, C. M. Escape from X inactivation in mice and humans. *Genome Biol.* **11**, 213 (2010).
45. Zhang, Y. *et al.* Genes That Escape X-Inactivation in Humans Have High Intraspecific Variability in Expression, Are Associated with Mental Impairment but Are Not Slow Evolving. *Mol. Biol. Evol.* **30**, 2588–2601 (2013).
46. Anderson, C. L. & Brown, C. J. Polymorphic X-Chromosome Inactivation of the Human TIMP1 Gene. *Am. J. Hum. Genet.* **65**, 699–708 (1999).
47. Lebl, J., Průhová, S., Zapletalová, J. & Pechová, M. IGF-I resistance and Turner's syndrome. *J. Pediatr. Endocrinol. Metab. JPEM* **14**, 37–41 (2001).
48. Carrel, L. *et al.* Genomic Environment Predicts Expression Patterns on the Human Inactive X Chromosome. *PLoS Genet.* **2**, e151 (2006).
49. Zinn, A. R. *et al.* A Turner syndrome neurocognitive phenotype maps to Xp22.3. *Behav. Brain Funct.* **3**, 24 (2007).
50. Oliveira, K. C. De. *et al.* Prevalence of the polymorphism MTHFR A1298C and not MTHFR C677T is related to chromosomal aneuploidy in Brazilian Turner Syndrome patients. *Arq. Bras. Endocrinol. Metabol.* **52**, 1374–1381 (2008).
51. Binder, G., Baur, F., Schweizer, R. & Ranke, M. B. The d3-Growth Hormone (GH) Receptor Polymorphism Is Associated with Increased Responsiveness to GH in Turner Syndrome and Short Small-for-Gestational-Age Children. *J. Clin. Endocrinol. Metab.* **91**, 659–664 (2006).
52. Peralta López, M. *et al.* Vitamin D receptor genotypes are associated with bone mass in patients with Turner syndrome. *J. Pediatr. Endocrinol. Metab. JPEM* **24**, 307–312 (2011).
53. Kalkan, R., Özdağ, N., Bundak, R., Çirakoğlu, A. & Serakinci, N. A unique mosaic Turner syndrome patient with androgen receptor gene derived marker chromosome. *Syst. Biol. Reprod. Med.* **62**, 77–83 (2016).
54. Su, M. A. *et al.* The role of X-linked FOXP3 in the autoimmune susceptibility of Turner Syndrome patients. *Clin. Immunol.* **131**, 139–144 (2009).
55. Trolle, C. *et al.* Long QT Interval in Turner Syndrome – A High Prevalence of LQTS Gene Mutations. *PLoS ONE* **8**, e69614 (2013).
56. Braz, A. F. *et al.* Genetic Predictors of Long-Term Response to Growth Hormone (GH) Therapy in Children With GH Deficiency and Turner Syndrome: The Influence of a SOCS2 Polymorphism. *J. Clin. Endocrinol. Metab.* **99**, E1808–E1813 (2014).
57. Trovó de Marqui, A. B. Turner syndrome and genetic polymorphism: a systematic review. *Rev. Paul. Pediatr.* **33**, 363–370 (2015).
58. Jevalikar, G. S. *et al.* Turner syndrome patients with bicuspid aortic valves and renal malformations exhibit abnormal expression of X-linked inhibitor of apoptosis protein (XIAP). *J. Pediatr. Endocrinol. Metab. JPEM* **28**, 1203–1208 (2015).
59. Visser, J. A. *et al.* Anti-Müllerian hormone levels in girls and adolescents with Turner syndrome are related to karyotype, pubertal development and growth hormone treatment. *Hum. Reprod.* **28**, 1899–1907 (2013).
60. Clayton, P. *et al.* A pharmacogenomic approach to the treatment of children with GH deficiency or Turner syndrome. *Eur. J. Endocrinol.* **169**, 277–289 (2013).
61. Xu, J., Taya, S., Kaibuchi, K. & Arnold, A. P. Sexually dimorphic expression of Usp9x is related to sex chromosome complement in adult mouse brain. *Eur. J. Neurosci.* **21**, 3017–3022 (2005).
62. Pasquali, L. *et al.* Collectrin gene screening in Turner syndrome patients with kidney malformation. *J. Genet.* **88**, 105–108 (2009).
63. Startin, C. M., Fiorentini, C., de Haan, M. & Skuse, D. H. Variation in the X-Linked EFHC2 Gene Is Associated with Social Cognitive Abilities in Males. *PLOS ONE* **10**, e0131604 (2015).

Acknowledgements

A.F., S.T., M.M.A. and A.A. are financially supported by UGC Non-Net fellowship. M.Z.M. and S.A. are financially supported by Indian Council of Medical Research under SRF (Senior Research Fellowship).

Author Contributions

R.I. and A.F. conceived the model and did the numerical experiment. A.F. and S.T. prepared the figures of the numerical results. A.F. analysed and interpreted the results and wrote the manuscript. A.F., M.M.A., A.A., S.A., M.Z.M., S.A. and R.I. involved in the study and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-28375-0>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018