Farooqui *et al.*
Egyptian Journal of Medical Human Genetics (2024) 25:23
https://doi.org/10.1186/s43042-024-00491-9

Egyptian Journal of Medical Human Genetics

RESEARCH

Open Access

System biology approach to delineate expressional difference in the blood mononuclear cells between healthy and Turner syndrome individuals

Anam Farooqui[1], Naaila Tamkeen[2], Safia Tazyeen[1], Sher Ali[3] and Romana Ishrat[1*]

Abstract

Background Turner syndrome (TS) is a rare disorder associated either with complete or partial loss of one X chromosome in women. The information on the genotype–phenotype relationship in TS is inadequate. Comparing the healthy and Turner syndrome patients may help elucidate the mechanisms involved in TS pathophysiology. Gene expression differences between healthy and individuals with Turner syndrome were characterized using the systems-biology approach of weighted gene coexpression network analysis (WGCNA) on 182 microarray peripheral mononuclear blood samples (PBMC).

Results The coexpression networks of healthy and TS had scale-free topology that ensures network robustness. In the process, five modules were preserved between healthy and TS, which carry several genes common in each module. Two of them, SMCHD1 and PGK1, have already been reported to be involved in TS. Previously reported genes of TS, specifically, PTPN22, RPS4X, CSF2RA, and TIMP1, were missing in their respective modules. Dysfunction, differential expression, or absence of these genes could lead to a progressive disruption of molecular pathways leading to the pathophysiology of TS. Indeed, we observed a significant difference in the functions of these modules when compared within and across the healthy and TS samples. We identified four clusters in the PPI network constructed from the top 15 KME enriched in significant functions.

Conclusion Overall, our work highlights the potential molecular functions, pathways, and molecular targets of TS that can be exploited therapeutically in the human healthcare system.

Keywords Coexpression network, Scale-free, Turner Syndrome, WGCNA

Background

This era of genomic technologies provides an immense wealth of molecular information offering a remarkable opportunity for new target discovery [1]. The microarray gene expression data provides an understanding of the underlying system-level view of the disconcerted networks of the disease cells. This, in turn, allows recognition of the potentially responsible key genes or pathways that can be targeted therapeutically [2]. However, the identification of key molecular targets remains a challenge for syndromes whose pathophysiology is mainly due to global genomic imbalance rather than defect in a single gene. Turner syndrome (TS) is a rare disorder associated either with complete or partial loss of one X chromosome in women. Often the women with TS have mosaic karyotypes [3]. Comorbidities associated with

*Correspondence:
Romana Ishrat
romanaishrat@gmail.com
[1] Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi 110025, India
[2] Department of Biosciences, Jamia Millia Islamia, New Delhi 110025, India
[3] VC Office, Era University, Lucknow 226003, India

© The Author(s) 2024. **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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/by/4.0/.

TS include short stature, absent or incomplete central puberty, ovarian dysgenesis, infertility, congenital malformations of the heart, endocrine disorders such as type 1 and type 2 diabetes mellitus, osteoporosis, recurrent miscarriages, and autoimmune disorders [4]. A multidisciplinary approach to care is required because Turner syndrome can affect multiple organs at any stage of life [5].

Previously, we demonstrated that the TS networks, constructed from manually curated genes from literature searches and microarray gene expression data, carry scale-free hierarchical fractal attributes [6, 7]. This organization showed that the TS networks were dynamic and heterogeneous, suggesting this to be a compact and self-organized and had efficient information processing potential. We identified few key regulators and essential genes (interologs) from these TS networks that we envisage having potential as therapeutic targets.

Weighted gene coexpression network analysis (WGCNA) is another powerful approach that lets us identify the higher-order relationships among the genes built on their coexpression relationships, demarcating highly similar coexpressed genes into modules, and allowing a robust view of transcriptome organization [8]. Exploring TS through WGCNA may explain a whole new aspect of this syndrome, and this may shed light on the physiology and gene network perturbations that occur in TS.

Our present work highlights the potential molecular functions, pathways, and molecular targets when genomic data from TS and healthy women are compared with each other and analyzed at the level of gene coexpression modules. Integrating this type of data, in which a comparative analysis of modules of healthy and TS samples can be studied in response to relevant molecular perturbations, may further facilitate the identification and validation of novel molecular targets.

Methods

Data set acquisition

The raw gene expression data were downloaded from the GEO data repository (9) (http://www.ncbi.nlm. nih.gov/geo/). As we aimed to compile a wide set of comparable data, we collected as many relevant data sets as we could find. We then filtered out all barring a core collection of data sets that were similar enough to be used for bioinformatics comparison. We considered four inclusion criteria. First, we considered only those data sets that were run on the Affymetrix platform HG-U133 Plus 2. Second, we included only those samples in each data set that were healthy (control) from human peripheral blood mononuclear Cells (PBMC). Third, only female samples were considered. And fourth, we made the correlations between genes more comparable across the studies by omitting all the data sets with fewer than five samples. Qualifying these inclusion criteria, a total of 12 data sets were considered for further analysis (Table 1). These expression data were combined, and prior to the analysis, batch corrected. The genes detected by multiple probes were represented by their mean value. These expression data were then subjected to WGCNA analysis.

Data preprocessing

The datasets that we collected individually had a small sample size. Performing the statistical analysis individually on each series may not give robust results [19]. Thus, we performed a combined analysis of the collected gene expression data from different series (Table 1). However,

Table 1 List of microarray series used in the study

S.n	Series	Platform	Sample types	Total number of samples	Samples used in the study	References
1	GSE90763	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	45	8	[9]
2	GSE46687	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	36	10	[4]
3	GSE26554	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	91	15	[10]
4	GSE34205	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	101	7	[11]
5	GSE22255	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	40	10	[12]
6	GSE21942	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	29	15	[13]
7	GSE17114	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	29	7	[12]
8	GSE19314	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	66	15	[14]
9	GSE27567	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	162	31	[15]
10	GSE22356	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	38	10	[16]
11	GSE13501	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	195	34	[17]
12	GSE42057	**GPL570** [HG-U133_Plus_2]	Peripheral blood mononuclear cell	136	20	[18]

Total number of Samples = 182

combining these datasets can lead to batch effects. So, we removed the batch effects between these datasets through the Empirical Bayes method (ComBat) [20] using the GeneSpring Software 14.9.1 (Agilent Technologies, Inc., Santa Clara, CA, USA). This algorithm adjusts the batch effect in microarray expression data. We assumed that the measured expression value of gene i in sample j of batch g can be expressed in a general form as follows:

$$x_{ij} = x_{ij}^{'} + b_{ij}^{x} + \varepsilon_{ij}^{x} \quad (1)$$

Then, to remove the batch effect, we standardize it:

$$\hat{x}_{ij} = \frac{x_{ij} - \overline{x}_{ij}}{\sigma_{x_i}} \quad (2)$$

Here, x_i is the average value of the expression value from gene i in all the samples; $\sigma\, x_i$ is the standard deviation for x_i of the expression value from gene i in all the samples.

Coexpression network construction: a system biology approach

Analyzing the coexpression patterns of the genes can provide an understanding of the underlying cellular processes because the coordinated coexpression of the genes encodes interacting proteins [21]. Therefore, we used the collected data sets (Table 1) to construct the coexpression network following the protocols of WGCNA [8, 22] to create a coexpression network of healthy humans (females) from PBMC samples. Separately, we again constructed a coexpression network of Turner syndrome patients from PBMC samples (GSE46687). We did this to get an overview of the comparison between healthy and TS patients from the perspective of a coexpression network.

To construct the gene coexpression network and identify its modules, we used the WGCNA (weighted gene coexpression network analysis) implemented in the R software package (http://www.r-project.org/). We used the Pearson correlation coefficient as a coexpression measure for cluster analysis. In these networks, a node represents the gene expression profile of a given gene. Thus, nodes are connected only if they have a significant pairwise expression profile cutoff across the samples. In our study, for all the pairwise comparisons of gene expression, the absolute value of the Pearson correlation coefficient was calculated across all microarray samples, and the appropriate soft threshold power was chosen. Then, the Pearson correlation matrix was transformed into an adjacency matrix. An adjacency matrix is a matrix of connection strengths through a power function. Thus, the connection strength (adjacency) a_{ij} between gene expressions x_i and x_j is defined as:

$$a_{ij} = |cor(x_i, x_j)|^{\beta} \quad (3)$$

As known, in general, all types of biological networks (gene expression networks) have been found to exhibit an approximate scale-free topology (2–6). Scale-free networks follow the power law and decay as $P(k) \sim k^{-g}$, where g is some exponent. Depending on the graph, we used the scale-free topology criterion (β). It is expected that the weighted gene coexpression network has an advantage over the unweighted coexpression network as the former one is more robust and the continuous nature of the gene coexpression information is preserved in it. For weighted gene coexpression network, whole network connectivity k_i of the ith gene expression profile x_i is the sum of the connection strengths with all other genes in the network, i.e.,

$$k_i = \sum_{j \in N, j \neq i} |cor(x_i, x_j)|^{\beta} \quad (4)$$

where N refers to the set of network genes.

Then, modules were identified following the dynamic tree cut method. The modules are the groups of genes with high topological overlap (1, 7). Through this method, the spurious or isolated connections were excluded during module formation.

The topological overlap for weighted networks (7) is given by,

$$t_{ij} = \frac{\sum_u a_{iu} a_{uj} + a_{ij}}{\min\left(\sum_u a_{iu}, \sum_u a_{ju}\right) + 1 - a_{ij}} \quad (5)$$

where an adjacency matrix is represented by,

$$A = [a_{ij}], 0 \leq a_{ij} \leq 1 \quad (6)$$

We used the topological overlap dissimilarity measure (1—topological overlap matrix (TOM)) in an average linkage hierarchical clustering to define the modules as branches of the resulting dendrogram. The genes of the modules were then assigned the eigenvalues based on which the significant module eigengenes (ME) were identified in each module.

Functional and pathway enrichment

The functional enrichment analysis was conducted on the genes in the identified modules. GO biological process term and Kyoto encyclopedia of genes and genomes (KEGG) pathway analyses [23] were conducted using DAVID online server [24]. Functional enrichment analysis was based on the cutoff value of Benjamini P-value < 0.05.

PPI network of TS and MODULE identification

To further explore the significant genes of each module in terms of protein–protein interactions, the top 15 K_{ME} genes in each module were extracted. These genes were then entered into STRING database [25] to form the PPI network. The Cytoscape plug-in molecular complex detection (MCODE) [26, 27] was used to analyze the most notable clustering module. Parameters for MCODE were set with Degree Cutoff as 2, Node Score Cutoff as 0.2, K-Core as 2, and Max. Depth from Seed as 100. Thus, system biology approach may be exploited in human health care system including complex polygenic diseases.

Results

The workflow of the work presented here is illustrated in Fig. 1.

Data set acquisition and pre-processing

By filtering the datasets with our inclusion criteria (mentioned in the methodology), we collected a total of 12 microarray series which gave us 182 healthy PBMC female samples (see Table 1).

Module formation and characterization

Coexpression networks were generated using the weighted gene coexpression network (WGCNA) package, in which all the coexpressed genes were connected with varying correlation strengths. This was accomplished using soft thresholding, thereby preserving the continuous nature of the data set and eliminating the need to set an arbitrary correlation score cutoff. The soft threshold power β is a value used to power the correlation of the genes to that threshold. It is assumed that raising the correlation to a power (β) will reduce the noise of the correlations in the adjacency matrix. To pick up one threshold, we used the pickSoftThreshold function, which calculates if the network resembles a scale-free graph, for each power. The lowest power that resulted in a scale-free topology is the one we used. In a scale-free network, small-degree nodes are the most abundant, but the frequency of high-degree nodes decreases relatively slowly. Thus, nodes that have degrees much higher than average, so-called hubs, exist. Because of the heterogeneity of scale-free networks, random node disruptions do not lead to a major loss of connectivity, but the loss of the hubs causes the breakdown of the network into isolated clusters [28]. The biological networks are considered scale-free networks. This is the reason, in case of WGCNA, we selected that β value, which gives scale-free topology. The genes that had highly similar expression levels, clustered together [29]. Such highly correlated genes may shed light on the shared biological processes or shared regulatory mechanisms that could be targeted. Therefore, we performed combined global gene expression profiling on 182 patient samples from PBMCs Set 1, $n=182$ (combined 12 datasets in Table 1) and Set 2, $n=26$ (only Turner syndrome samples in GSE46687)) and separately constructed a weighted gene coexpression network of Set 1 and Set 2 based on pairwise Pearson correlations between the expression profiles for the identification of gene modules. A gene module is a group of highly coexpressed genes. These modules were detected through unsupervised hierarchical clustering [30]. In a coexpression network, for fulfilling the property of scale-free topology, the parameter β is considered very crucial. Most likely, the biological networks which are based on gene expression data are scale-free [31]. Therefore, $\beta=6$ (for Set 1) and $\beta=4$ (for Set 2) were considered to obtain scale-free topology by the fit index greater than 0.87 and 0.84, respectively. Figures 2 and 3 show the result of several powers for finding a network with scale-free topology properties of Sets 1 and 2, respectively. We found that the coexpression networks had a scale-free topology (Figs. 2C and 3C), which is believed to ensure network robustness and thus resisting random node attacks [32].

After identifying modules of coexpressed genes, each module in effect becomes a new network, and a new measure of connectivity (intramodular connectivity, or kin) is defined as the sum of a gene's connection strengths with all other genes in its module. Intramodular

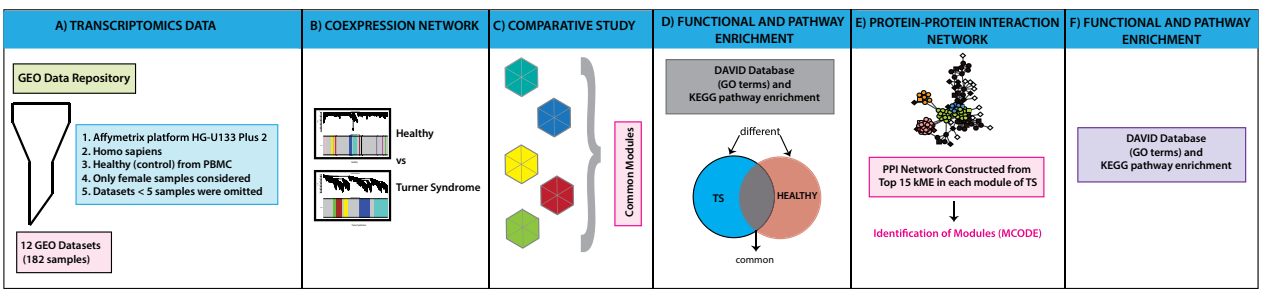

Fig. 1 Schematic illustration of the workflow

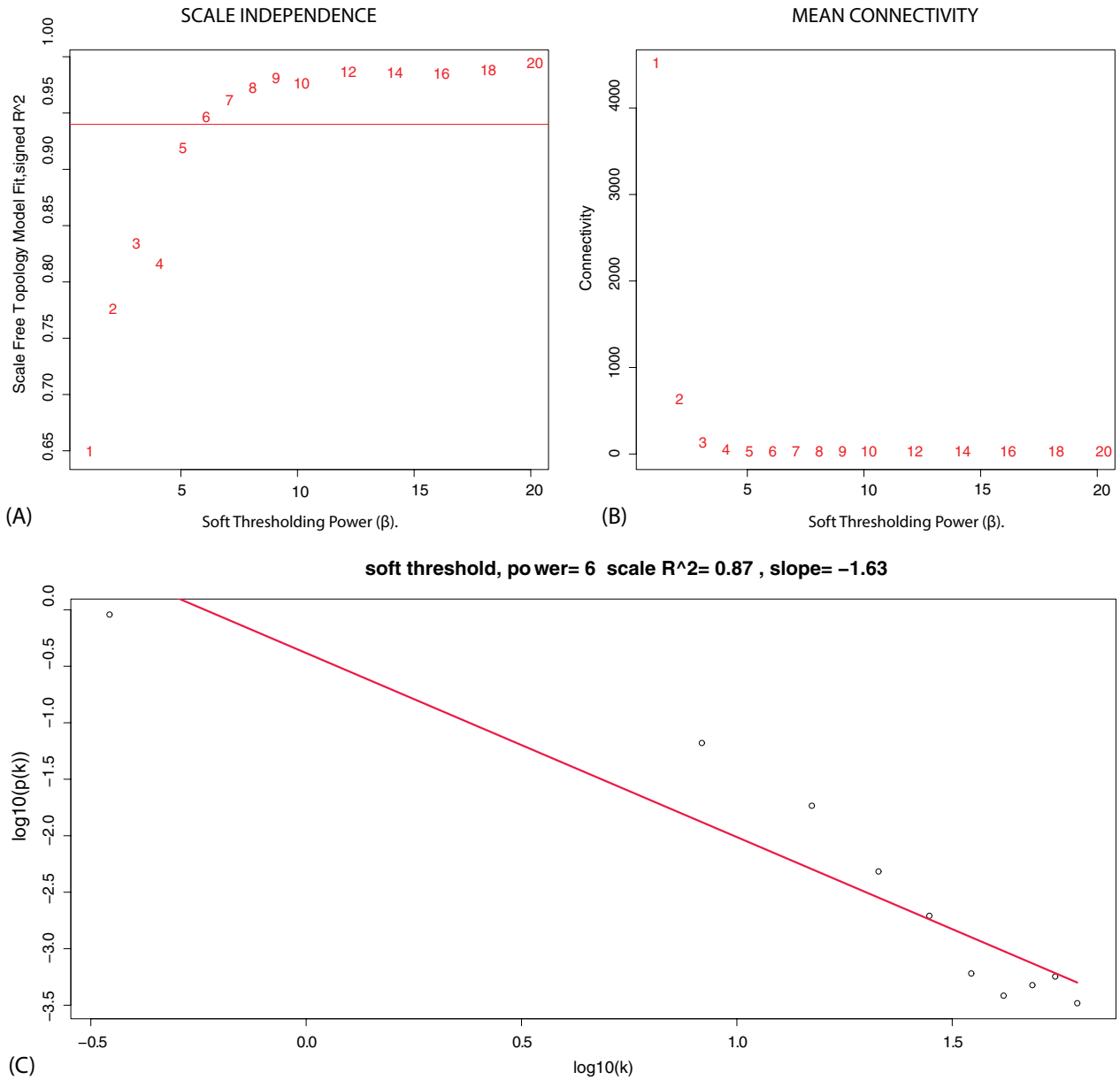

Fig. 2 Determination of soft-thresholding powers (*β*) in the weighted gene coexpression network analysis (WGCNA) for **Set 1**. **A** Analysis of the scale-free fit index for various soft-thresholding powers (β). **B** Analysis of the mean connectivity for various soft-thresholding powers (β). **C** Log–log plot of whole network connectivity distribution. The x-axis shows the logarithm of whole network connectivity, y-axis the logarithm of the corresponding frequency distribution. On this plot, the distribution approximately follows a straight line, which is referred to as approximately scale-free topology when $\beta = 6$

connectivity measures how connected, or coexpressed, a given gene is with respect to the genes of a particular module. The module network dendrogram, constructed by clustering ME distances, showed that modules with high Kin were positioned at the tip of the branches since they exhibit the highest interconnectedness with the rest of the module. After highly similar modules were merged, eleven gene expression modules were determined in Set 1 (Fig. 4A), and five gene expression modules were determined in Set 2 (Fig. 4B). The colors are assigned based on module size. Turquoise (others refer to it as cyan) colors the largest module, next comes blue, next brown, next yellow, then green, and so on. A gray color module is reserved for unassigned genes. Eigengenes, the first principal component of a cluster, is thought to be a representative of a cluster's expression profile [22].

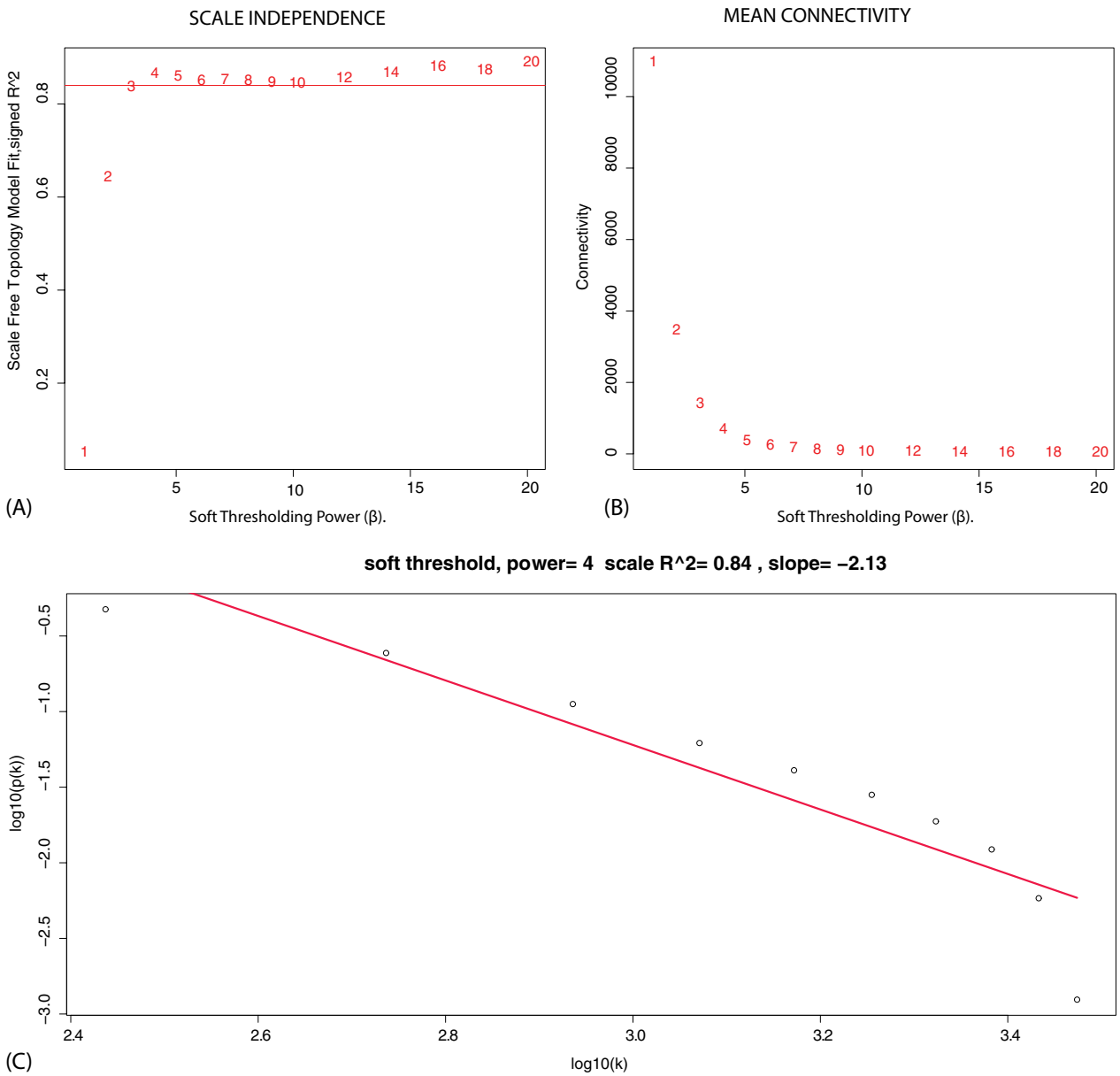

Fig. 3 Determination of soft-thresholding power (*β*) in the weighted gene coexpression network analysis (WGCNA) for **Set 2. A** Analysis of the scale-free fit index for various soft-thresholding powers (β). **B** Analysis of the mean connectivity for various soft-thresholding powers (*β*). **C** Log–log plot of whole network connectivity distribution. The x-axis shows the logarithm of whole network connectivity, and y-axis the logarithm of the corresponding frequency distribution. On this plot, the distribution approximately follows a straight line, which is referred to as approximately scale-free topology when $\beta = 4$

To get a sense of how related the modules are, one can summarize each module by its first eigengene (referred to as principal components), and then correlate these module eigengenes with each other based on K_{ME} value (Fig. 4C and D). The module eigengene-based intramodular connectivity measure kME roughly approximates the standard intramodular connectivity kIN. This measure is determined by correlating the expression profile of a gene i with the module eigengene of its resident module: kMEi = |cor(x i, ME)|. This gives top 15 most connected genes of each module.

The multiple dimensional scaling (MDS) plot provides an alternate visualization of the module structure. Figure 5 represents multidimensional scaling plots (MDS) of coexpressed modules in Set 1 and Set 2. Modules tend to form separate 'fingers' in this plot. The color denotes

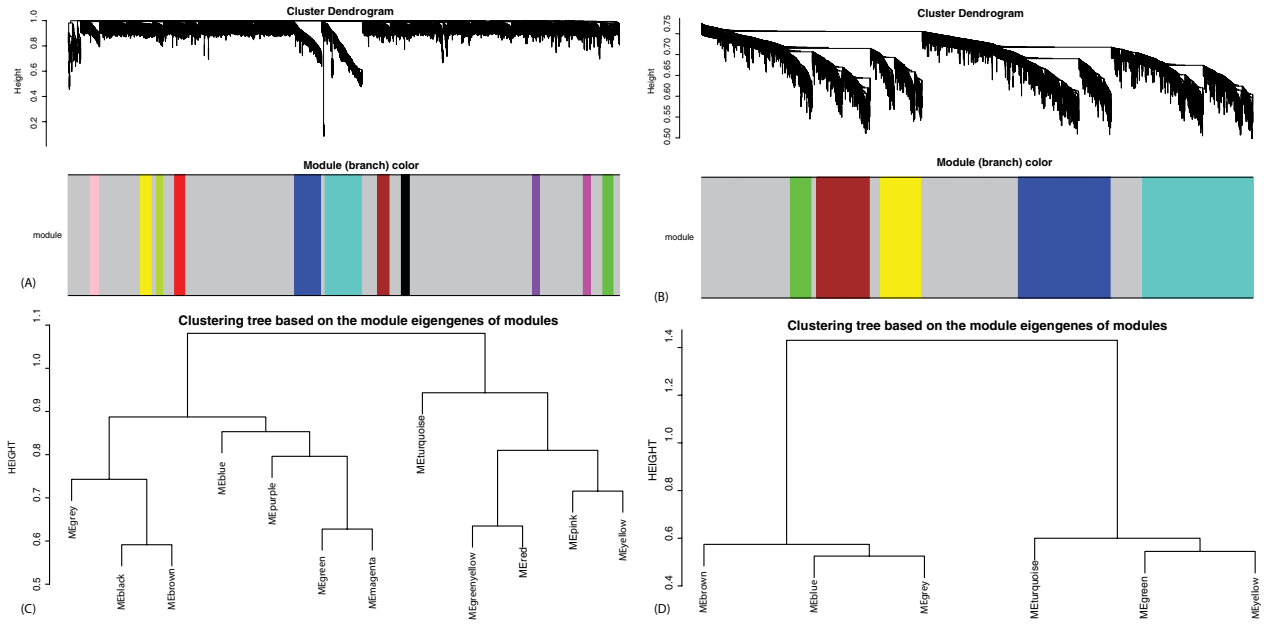

Fig. 4 WGCNA network and module detection **A** and **B** (Upper) Cluster dendrograms in the human Sets 1 (A) and 2 (B). The y-axes correspond to distance (1 − TO). (Lower) Dynamic tree cutting was used to determine modules, generally by dividing the dendrogram at significant branch points. Modules with significant overlap were assigned the same labels. Colors in the horizontal bar represent the modules. **C** and **D.** The module network dendrogram was constructed by clustering module eigengene distances in Sets 1 **C** and 2 **D**

genes of that module. Intramodular hub genes are located at the finger tips. The MDS plots are shown in Fig. 5. All the plots are color-coded according to the modules concerning respective Sets 1 or 2. The points with the same color are almost always clustered together which means that the relative position of the points is well-preserved. However, the spatial distributions of the points vary to a large extent when we compare MDS plots of healthy vs TS. This is because the expression pattern of genes is disrupted due to TS.

The detailed list of genes of each module in each set is listed in electronic supplementary material, (Additional file 1: Table S1. and S2).

Comparison of healthy and TS patients modules provide insights into TS

A comparison of coexpression networks between healthy and TS human samples could provide valuable insight into the syndrome. We found that five modules were common between Set 1 (Healthy) and Set 2 (Turner Syndrome). The genes that were common between these modules are listed in Table 2. It was identified that the genes highlighted in bold were differentially expressed in TS. Two of them, namely *SMCHD1* [33] and *PGK1* [34], have already been reported to be involved in TS. Thus, the functions performed by these genes (highlighted in bold) in their respective modules in healthy individuals are perturbed in TS patients due to their differential expression. This may contribute to the pathophysiology of TS. Therefore, as studied previously, it can also be concluded that the TS phenotype is caused due to the additive effect of genes from different loci and due to global genomic imbalance [7].

Earlier, we reported seed genes, few key regulatory genes, and signature genes involved in Turner syndrome from manually curated genes of TS from the literature [6] and microarray analysis [7] employing the network approach. Table 3 contains the list of those seed genes that are present in the gene expression modules of the healthy and TS coexpression network. We found that many of these genes were present in the different modules of Set 1 and Set 2. Since these reported genes are expected to play an important role in the pathophysiology of TS, we can see that many important genes (present in the modules of healthy samples coexpression network) are missing in their respective modules of TS samples (Table 3). This shows that due to the TS state, the important genes start to function in a different way resulting in disturbed gene functioning. The genes that coexpressed to share some biological process or regulatory mechanisms in healthy females changed their pattern of functioning, thus, causing the perturbation in the normal functioning of an individual with TS.

Specifically, four of the genes, namely *PTPN22* [35], *RPS4X* [36, 37], *CSF2RA* [38, 39], and *TIMP1* [40, 41] that we identified in our previous studies [6] as the key

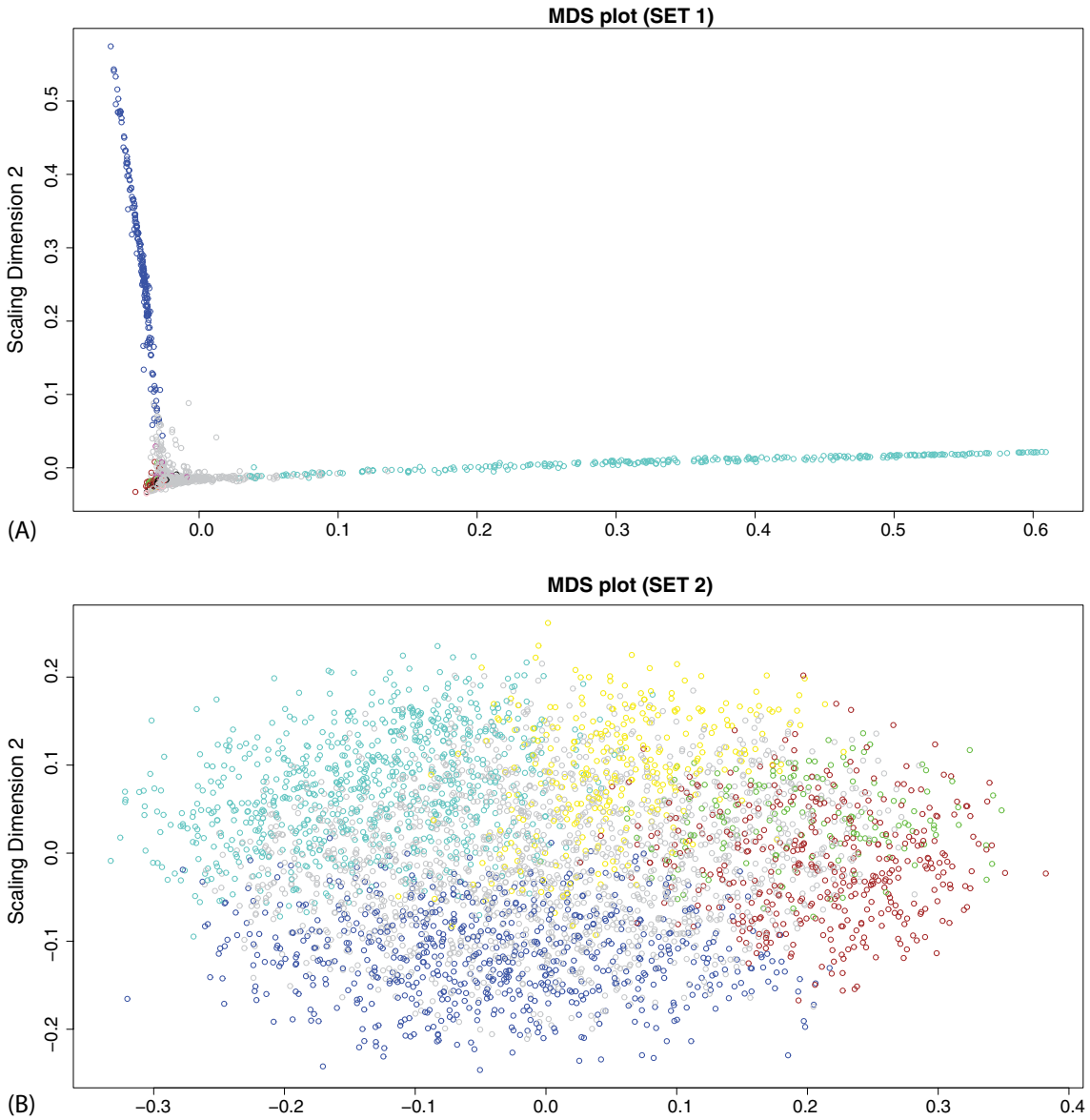

Fig. 5 A multidimensional scaling plots (MDS) of coexpressed modules in Set 1 and Set 2. Modules tend to form separate 'fingers' in this plot. The color denotes genes of that module. Intramodular hub genes are located at the finger tips. **A** An MDS plot displaying expression data of genes in different modules in Set 1. **B** An MDS plot displaying expression data of genes in different modules in Set 2

regulatory genes of the TS network were found to be missing in their respective modules of TS. Thus, dysfunction, differential expression, or absence of these genes could then lead to a progressive disruption of the molecular pathways in a predictable manner leading to the pathophysiology of TS. This also suggests a high level of genetical, organizational and expressional heterogeneity among the TS patients.

Comparison of functional and pathway enrichment of different modules of healthy vs Turner Syndrome

To compare the potential biological functions of critical modules between healthy and TS samples, we performed the functional enrichment analysis of the GO and KEGG categories. There was a significant difference in the functions that different modules were enriched in when

Table 2 Common genes between the common modules of Set 1 (Healthy) and Set 2 (TS)

S.n	Common module	Common Gene
1	Turquoise	*GDI2, RPS5, UBB, RPS27, RPLP0, LAPTM5, BTF3, RPL22, DLAT,* ***SMCHD1****, GLYR1,* ***CDC42SE2***
2	Blue	–
3	Brown	*RHOA,* ***PGK1****, SDHC,* ***ARPC4***
4	Yellow	***RUNX1-IT1****,* ***ZBTB20****,* ***USP34***
5	Green	*NAGA, AGTRAP, ATP6V0B, GRN, BLVRB, GNG5, SLC43A3, SLC15A3,* ***EMILIN2****,* ***GLIPR2***

Genes highlighted in bold are differentially expressed in TS based on the analysis by [7]. *The genes that are bold underlined have reports that they are associated with TS*

compared within and across the healthy and TS samples (Table 4). However, few biological functions were consistent within and across the healthy and TS samples.

We identified five modules common between healthy and TS coexpression networks. In the **turquoise module** of both the healthy and TS coexpression network, the common molecular functions identified were protein binding, RNA binding, and poly(A) RNA binding. The common enriched cellular components were found to be the cytosol, extracellular exosome, and cytoplasm and none of the biological functions were found to be common in turquoise module. In the **blue module** of both, the healthy and TS coexpression network, the common molecular function identified was protein binding, while none of the biological function and cellular component was found to be common. In the **green module,** extracellular exosome while in the **yellow module,** cytosol was found to be the common enriched cellular component of both the healthy and TS coexpression network. None of the functions were common in the **brown module** of both the healthy and TS coexpression network. None of the KEGG pathways was found to be common in the common modules of healthy and TS coexpression networks.

Though there were few enriched functions and cellular components that were common, there was a significant difference in the enriched functions and pathways among these five common modules. The functional enrichment and KEGG pathway of different modules of healthy and TS coexpression networks are listed in Tables 4 and 5.

It can be explained that the similar enriched functions though similar but the set of genes that perform these functions are slightly different (in each module) in TS samples as compared to healthy ones. This means that the functions being similar are somehow being performed by a different set of genes, and due to differential gene expression in the TS state, these functions may not be performed in synchronized and a proper way resulting in the pathophysiology of Turner syndrome. And obviously, there are many enriched molecular functions, biological functions, cellular components, and KEGG pathways that are significantly different among the five common modules between healthy and TS coexpression networks.

While many genes (Table 2) are common between common modules of the healthy and TS patients, most of the genes are different in these modules between healthy and TS. This clearly makes sense that in a

Table 3 List of previously reported seed genes present in the gene expression modules of the healthy and TS coexpression network

S.n	Modules of healthy	Previously reported Seed genes/ Key genes present in Modules of Healthy	Modules of TS	Previously reported Seed genes/ Key genes present in Modules of TS
1	Turquoise	*TOB2, RPL31,* ***PTPN22****,* ***RPS4X***	Turquoise	*YME1L1, UBE3B, TCAIM, SYNCRIP, STAT1, SAMD4B, PEX3, PCCB, PAPOLA, OCIAD1, NNT, N4BP2L1, MAPK1 FOXN3, DOT1L, CELF1, ATXN7L1*
2	Blue	*TSPAN33, TPM1, THBS1, GUCY1A3, CXCL5*	Blue	*SYNCRIP, SLC16A7, PPP6R2, POGK, OCIAD1, NPC1, KMT2A, DOT1L, DGKZ, BTBD7, ANKRD44*
3	Brown	*PDLIM5, APLP2,* ***CSF2RA***	Brown	*SLC16A7, MAST4, KRR1, KMT2A, GOPC, GGCX, FAHD2A, EML4, BOD1L1*
4	Yellow	–	Yellow	*MAPK1, DCAF8, AKAP10*
5	Green	*LGALS3, ALOX5,* ***TIMP1***	Green	*PCNX2, KMT2A, KIAA2026, APLP2*
6	Black	–	–	–
7	Red	*SLC16A7, MTF2, KRAS, KMT2A*	–	–
8	Greenyellow	*POGK, KMT2A, HIPK1*	–	–
9	Magenta	–	–	–
10	Pink	*OCIAD1*	–	–
11	Purple	–	–	–

Genes highlighted in bold are differentially expressed in TS based on the analysis by [7]. *The genes that are underlined have reports that they are associated with TS* [7]

healthy state, these common genes function in synchronization and combination of the genes listed in electronic supplementary material, (Additional file 1: Table S1). These genes get replaced by a new set of genes in the TS state (electronic supplementary material, (Additional file 1: Table S2)), and the complete functional and biological process is performed with a whole new set of genes. Many of these genes are differentially expressed in TS. Due to this differential expression, the genes perform the biological and molecular functions in a perturbed way. This explains the mechanism of Turner syndrome and pathophysiology.

Table 4 Comparative analysis of Functional Enrichment of coexpression modules of Healthy with Turner Syndrome

S.n	Category	Term	Count		Bemjamini *P*-Value	
			Healthy	TS	Healthy	TS
TURQUOISE						
1	GOTERM_MF	GO:0005515 ~ protein binding	120	417	2.30E-06	4.03E-25
2	GOTERM_MF	GO:0044822 ~ poly(A) RNA binding	88	110	2.11E-56	2.84E-22
3	GOTERM_MF	GO:0003735 ~ structural constituent of ribosome	77	–	5.39E-100	–
4	GOTERM_MF	GO:0003723 ~ RNA binding	49	46	6.77E-31	5.41E-06
5	GOTERM_MF	GO:0098641 ~ cadherin binding involved in cell–cell adhesion	15	–	2.82E-05	–
6	GOTERM_CC	GO:0005829 ~ cytosol	109	207	5.75E-38	3.88E-20
7	GOTERM_CC	GO:0070062 ~ extracellular exosome	98	140	8.60E-35	1.66E-05
8	GOTERM_CC	GO:0005737 ~ cytoplasm	92	225	7.94E-11	7.38E-05
9	GOTERM_CC	GO:0016020 ~ membrane	85	–	5.21E-32	–
10	GOTERM_CC	GO:0005634 ~ nucleus	78	260	9.81E-05	6.42E-11
11	GOTERM_BP	GO:0006413 ~ translational initiation	81	–	1.06E-129	–
12	GOTERM_BP	GO:0006412 ~ translation	79	–	1.99E-98	–
13	GOTERM_BP	GO:0000184 ~ nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	77	–	2.83E-127	–
14	GOTERM_BP	GO:0006614 ~ SRP-dependent co-translational protein targeting to membrane	76	–	1.32E-137	–
15	GOTERM_BP	GO:0019083 ~ viral transcription	75	–	1.73E-125	–
16	KEGG_PATHWAY	hsa03010: Ribosome	77	–	1.00E-100	–
17	GOTERM_MF	GO:0005524 ~ ATP binding	–	73	–	0.043912
18	GOTERM_MF	GO:0000166 ~ nucleotide binding	–	26	–	0.020279
19	GOTERM_BP	GO:0015031 ~ protein transport	–	33	–	0.00205
20	GOTERM_BP	GO:0000398 ~ mRNA splicing, via spliceosome	–	32	–	7.47E-08
21	GOTERM_BP	GO:0016032 ~ viral process	–	27	–	0.003497
22	GOTERM_BP	GO:0050852 ~ T cell receptor signaling pathway	–	25	–	2.87E-07
23	GOTERM_BP	GO:0043161 ~ proteasome-mediated ubiquitin-dependent protein catabolic process	–	23	–	8.11E-04

Table 4 (continued)

S.n	Category	Term	Count		Bemjamini *P*-Value	
			Healthy	TS	Healthy	TS
24	GOTERM_CC	GO:0005654 ~ nucleoplasm	–	194	–	6.26E-24
25	KEGG_PATHWAY	hsa03040: Spliceosome	–	21	–	8.47E-05
BLUE						
1	GOTERM_MF	GO:0005515 ~ protein binding	103	251	0.008723	1.33E-08
2	GOTERM_MF	GO:0046982 ~ protein heterodimerization activity	34	–	5.69E-18	–
3	GOTERM_MF	GO:0098641 ~ cadherin binding involved in cell–cell adhesion	17	–	9.14E-07	–
4	GOTERM_MF	GO:0042393 ~ histone binding	11	–	1.21E-05	–
5	GOTERM_MF	GO:0003779 ~ actin binding	11	–	0.008723	–
6	GOTERM_CC	GO:0070062 ~ extracellular exosome	66	–	8.95E-12	–
7	GOTERM_CC	GO:0005886 ~ plasma membrane	56	–	0.017636	–
8	GOTERM_CC	GO:0005576 ~ extracellular region	30	–	0.004611	–
9	GOTERM_CC	GO:0005615 ~ extracellular space	27	–	0.003648	–
10	GOTERM_CC	GO:0000786 ~ nucleosome	24	–	1.70E-24	–
11	GOTERM_BP	GO:0007596 ~ blood coagulation	20	–	1.45E-12	–
12	GOTERM_BP	GO:0002576 ~ platelet degranulation	18	–	2.00E-14	–
13	GOTERM_BP	GO:0098609 ~ cell–cell adhesion	16	–	1.97E-06	–
14	GOTERM_BP	GO:0006334 ~ nucleosome assembly	15	–	6.74E-10	–
15	GOTERM_BP	GO:0044267 ~ cellular protein metabolic process	11	–	8.50E-06	–
16	KEGG_PATHWAY	hsa05202: Transcriptional misregulation in cancer	15	–	4.44E-06	–
17	KEGG_PATHWAY	hsa04611: Platelet activation	9	–	0.018321	–
18	KEGG_PATHWAY	hsa04270: Vascular smooth muscle contraction	8	–	0.038538	–
19	GOTERM_MF	GO:0044822 ~ poly(A) RNA binding	–	63	–	3.71E-09
20	GOTERM_CC	GO:0005634 ~ nucleus	–	174	–	3.12E-08
21	GOTERM_CC	GO:0005654 ~ nucleoplasm	–	120	–	2.12E-12
22	GOTERM_CC	GO:0005829 ~ cytosol	–	118	–	4.45E-07
23	GOTERM_CC	GO:0016020 ~ membrane	–	76	–	0.001716
24	GOTERM_CC	GO:0043234 ~ protein complex	–	21	–	0.039779
BROWN						
1	GOTERM_CC	GO:0070062 ~ extracellular exosome	34	–	1.94E-05	–
2	GOTERM_CC	GO:0016020 ~ membrane	32	–	1.61E-06	–
3	GOTERM_CC	GO:0005829 ~ cytosol	32	–	0.004395	–
4	GOTERM_CC	GO:0005925 ~ focal adhesion	10	–	0.004824	–

Table 4 (continued)

S.n	Category	Term	Count		Bemjamini *P*-Value	
			Healthy	TS	Healthy	TS
5	GOTERM_MF	GO:0005515 ~ protein binding	–	145	–	0.02457
6	GOTERM_MF	GO:0044822 ~ poly(A) RNA binding	–	49	–	1.99E-10
7	GOTERM_MF	GO:0003676 ~ nucleic acid binding	–	31	–	0.003619
8	GOTERM_MF	GO:0008270 ~ zinc ion binding	–	31	–	0.029941
9	GOTERM_MF	GO:0003682 ~ chromatin binding	–	15	–	0.039254
10	GOTERM_BP	GO:0006351 ~ transcription, DNA-templated	–	48	–	0.010354
11	GOTERM_BP	GO:0000398 ~ mRNA splicing, via spliceosome	–	13	–	0.010354
12	GOTERM_BP	GO:0008380 ~ RNA splicing	–	12	–	0.008027
13	GOTERM_BP	GO:0006306 ~ DNA methylation	–	6	–	0.008027
14	GOTERM_CC	GO:0005634 ~ nucleus	–	101	–	0.002754
15	GOTERM_CC	GO:0005654 ~ nucleoplasm	–	83	–	2.07E-11
16	GOTERM_CC	GO:0005694 ~ chromosome	–	8	–	0.036564
GREEN						
1	GOTERM_MF	GO:0016176 ~ superoxide-generating NADPH oxidase activator activity	4	–	8.84E-04	–
2	GOTERM_CC	GO:0070062 ~ extracellular exosome	30	–	2.15E-06	–
3	GOTERM_CC	GO:0005576 ~ extracellular region	15	–	0.04831	–
4	GOTERM_CC	GO:0043020 ~ NADPH oxidase complex	4	–	4.89E-04	–
5	KEGG_PATHWAY	hsa04380: Osteoclast differentiation	10	–	7.53E-06	–
6	GOTERM_MF	GO:0005515 ~ protein binding	–	94	–	0.005757
7	GOTERM_MF	GO:0004872 ~ receptor activity	–	9	–	0.024421
8	GOTERM_MF	GO:0031996 ~ thioesterase binding	–	4	–	0.018449
9	GOTERM_MF	GO:0031997 ~ N-terminal myristoylation domain binding	–	3	–	0.018449
10	GOTERM_BP	GO:0006954 ~ inflammatory response	–	12	–	0.03351
11	GOTERM_BP	GO:0045087 ~ innate immune response	–	12	–	0.039301
12	GOTERM_BP	GO:0002576 ~ platelet degranulation	–	8	–	0.016525
13	GOTERM_BP	GO:0045454 ~ cell redox homeostasis	–	6	–	0.037236
14	GOTERM_BP	GO:0000302 ~ response to reactive oxygen species	–	5	–	0.03351
15	GOTERM_CC	GO:0070062 ~ extracellular exosome	–	63	–	8.56E-13
16	GOTERM_CC	GO:0016020 ~ membrane	–	38	–	4.20E-04

Table 4 (continued)

S.n	Category	Term	Count		Bemjamini *P*-Value	
			Healthy	TS	Healthy	TS
17	GOTERM_CC	GO:0005789 ~ endoplasmic reticulum membrane	–	21	–	8.12E-04
18	GOTERM_CC	GO:0005794 ~ Golgi apparatus	–	18	–	0.011764
19	GOTERM_CC	GO:0043231 ~ intracellular membrane-bounded organelle	–	15	–	0.005962
20	KEGG_PATHWAY	hsa04142: Lysosome	–	11	–	7.95E-05
21	KEGG_PATHWAY	hsa04145: Phagosome	–	11	–	1.98E-04
YELLOW						
1	GOTERM_CC	GO:0005829 ~ cytosol	9	63	0.038311	0.002252
2	GOTERM_MF	GO:0005515 ~ protein binding	–	146	–	1.14E-05
3	GOTERM_BP	GO:0016485 ~ protein processing	–	9	–	0.003794
4	GOTERM_CC	GO:0070062 ~ extracellular exosome	–	68	–	8.00E-07
5	GOTERM_CC	GO:0016020 ~ membrane	–	55	–	9.69E-06
6	GOTERM_CC	GO:0005739 ~ mitochondrion	–	35	–	6.58E-04
7	GOTERM_CC	GO:0005794 ~ Golgi apparatus	–	28	–	3.24E-04
8	KEGG_PATHWAY	hsa00190: Oxidative phosphorylation	–	12	–	3.53E-04
9	KEGG_PATHWAY	hsa04932: Non-alcoholic fatty liver disease (NAFLD)	–	11	–	0.00346
BLACK						
1	GOTERM_MF	GO:0005515 ~ protein binding	53	–	2.85E-06	–
2	GOTERM_MF	GO:0044822 ~ poly(A) RNA binding	17	–	1.29E-04	–
3	GOTERM_CC	GO:0005737 ~ cytoplasm	30	–	0.026891	–
4	GOTERM_CC	GO:0070062 ~ extracellular exosome	25	–	7.16E–04	–
5	GOTERM_CC	GO:0005654 ~ nucleoplasm	22	–	0.006953	–
6	GOTERM_CC	GO:0016020 ~ membrane	18	–	0.017508	–
7	GOTERM_CC	GO:0005925 ~ focal adhesion	8	–	0.011224	–
GREENYELLOW						
1	GOTERM_MF	GO:0003677 ~ DNA binding	12	–	0.038074	–
2	GOTERM_MF	GO:0042800 ~ histone methyltransferase activity (H3-K4 specific)	3	–	0.038074	–
3	GOTERM_CC	GO:0005634 ~ nucleus	22	–	0.041888	–
4	GOTERM_CC	GO:0005654 ~ nucleoplasm	18	–	0.001011	–
5	GOTERM_CC	GO:0016607 ~ nuclear speck	7	–	3.66E-04	–
RED						
1	GOTERM_MF	GO:0005515 ~ protein binding	51	–	0.005499	–
2	GOTERM_MF	GO:0044822 ~ poly(A) RNA binding	20	–	6.04E-06	–
3	GOTERM_MF	GO:0008270 ~ zinc ion binding	15	–	0.008017	–

Table 4 (continued)

S.n	Category	Term	Count		Bemjamini *P*-Value	
			Healthy	TS	Healthy	TS
4	GOTERM_MF	GO:0001046 ~ core promoter sequence-specific DNA binding	4	–	0.027245	–
5	GOTERM_CC	GO:0005634 ~ nucleus	42	–	3.70E-06	–
6	GOTERM_CC	GO:0005654 ~ nucleoplasm	38	–	5.63E-12	–
7	GOTERM_CC	GO:0005737 ~ cytoplasm	33	–	0.022071	–
8	GOTERM_CC	GO:0016607 ~ nuclear speck	10	–	3.13E-06	–
9	GOTERM_BP	GO:0016569 ~ covalent chromatin modification	7	–	0.00304	–
MAGENTA						
1	GOTERM_MF	GO:0030246 ~ carbohydrate binding	7	–	9.31E-04	–
2	GOTERM_MF	GO:0004872 ~ receptor activity	6	–	0.011029	–
3	GOTERM_CC	GO:0070062 ~ extracellular exosome	19	–	0.003651	–
4	GOTERM_CC	GO:0043202 ~ lysosomal lumen	5	–	0.003651	–
5	KEGG_PATHWAY	hsa04145: Phagosome	7	–	8.56E-04	–
6	KEGG_PATHWAY	hsa04142: Lysosome	6	–	0.002174	–
PURPLE						
1	GOTERM_CC	GO:0005925 ~ focal adhesion	9	–	8.64E-04	–
PINK						
1	GOTERM_MF	–	–	–	–	–
2	GOTERM_BP	–	–	–	–	–
3	KEGG_PATHWAY	–	–	–	–	–

Table 5 Top 15 K_{ME} in each module of TS samples

S.n	MODULES				
	Turquoise	Blue	Brown	Yellow	Green
1	GOLGA7	NDUFS8	ANKRD36B	PCMT1	MS4A6A
2	XRCC5	OCIAD1	RP9	TMED10	PSAP
3	MAGT1	SRRM2	NUTM2B-AS1	COPZ1	MS4A6A
4	TMEM230	NLN	CCAR1	SH3GLB1	MNDA
5	CD164	PRR11	NEK1	SDHC	CAPNS1
6	FYTTD1	FBXW12	GON4L	VAMP3	MS4A6A
7	ATG5	FGFR1	PNISR	C4orf3	APLP2
8	UBE2D3	PDE4C	KMT2A	MAPRE1	NAGA
9	CAPRIN1	EWSR1	PRRC2C	MAPK1	RAC1
10	PCNP	MGC12488	ATRX	ARF1	APLP2
11	WTAP	PRR11	JPX	CHMP1B	TNFSF13B
12	PTPN11	POLR1B	RNPC3	TMEM59	BLVRB
13	SMAD2	UBXN2A	EIF5B	DOCK2	MEGF9
14	UBE2G1	ATP8B1	EIF5B	RPN2	NAGA
15	PCBP1	MALAT1	PPIG	ATP6AP2	TM9SF2

PPI network of TS and identification of its clusters

The eigengenes are the principal component of a module, and they can be a representative of a cluster's expression profile. To understand how these modules interact with one another, we selected the top 15 K_{ME} genes of each module. These K_{ME} genes of each module are listed in Table 5. Being the eigengenes of each module, the genes listed in Table 5 may have a potential role in TS. These genes were then entered into the STRING database to construct the protein–protein interaction (PPI) network. The confidence score was set at 0.7, and the genes in the first shell was set at 100. The main network constructed had a clustering coefficient of 0.750 with 129 nodes and 990 edges. In total, we considered the top four sub-networks (modules 1, 2, 3, and 4) with scores 34, 12.5, 7, and 7, respectively, that were detected by MCODE (Fig. 6). Each cluster had the highest scoring node called a seed. It is the node from which the cluster was derived. The genes *NDUFV1, EFTUD2, POLR1C, and STT3A* are the highest scoring node, i.e., seed genes of clusters 1, 2, 3, and 4, respectively. They act as the eigengene of those cluster. The list of genes of each cluster is listed in electronic supplementary S3. The functional and pathway enrichment of modules of the PPI network of TS was analyzed, and it was found that cluster 1 is mainly enriched in the processes and reactions related to mitochondrial

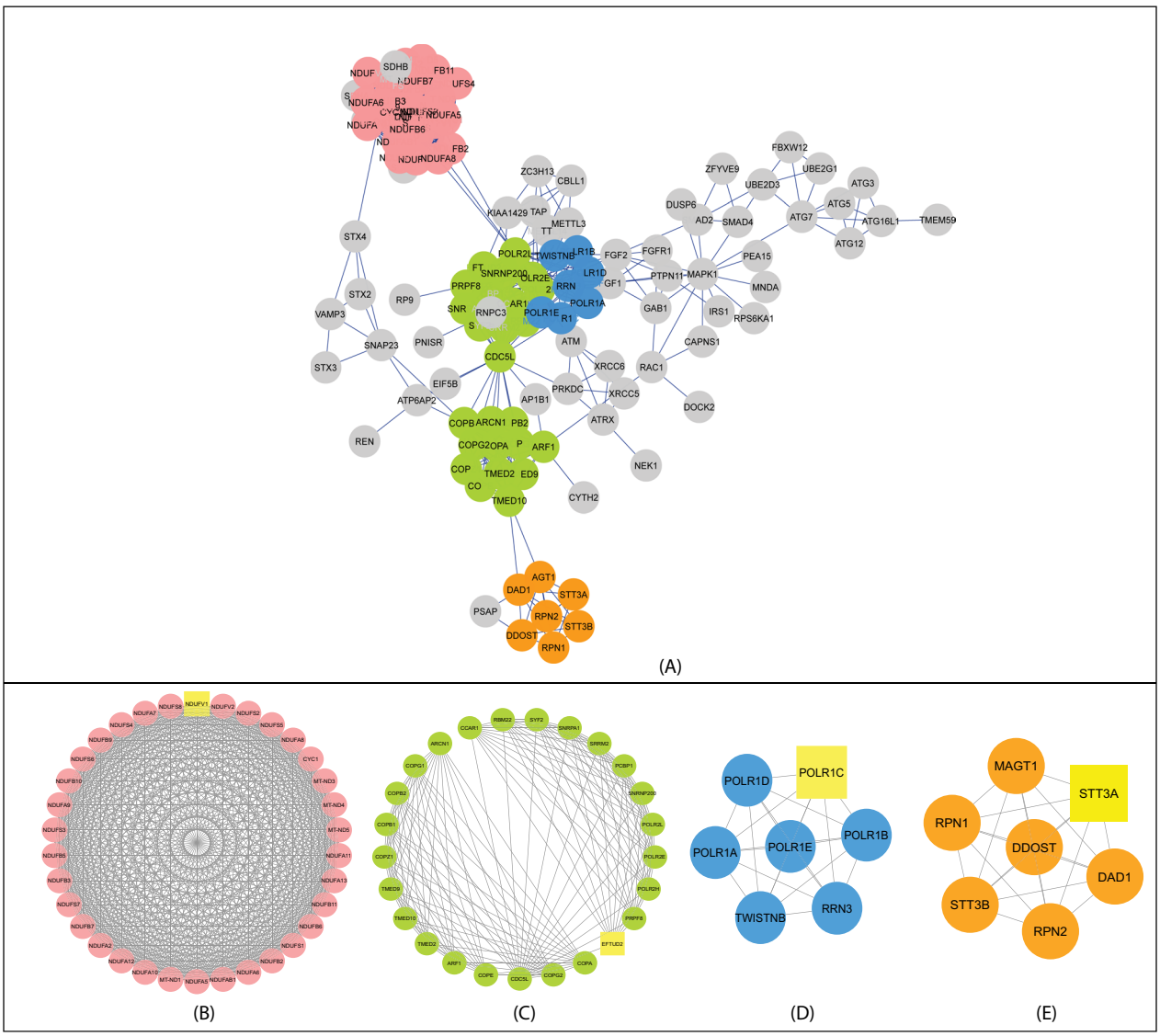

Fig. 6 **A** Protein–protein interaction network of top 15 K_{ME} genes of each module of TS coexpression network**. B-E** Modules identified by MCODE where **B** is cluster 1, **C** is cluster 2, **D** is cluster 3, and **E** is cluster 4. The highest scoring node in the cluster is called a seed. It is the node from which the cluster was derived and is represented by a square shape (see square yellow in **B**, **C**, **D** and **E**)

phosphorylation, cluster 2 is mainly enriched in golgi-mediated protein transport and reactions related to it, cluster 3 is mainly enriched in processes related to transcription, and cluster 4 is mainly enriched in post-translational modification processes like glycosylation. The list of different molecular functions, biological functions, cellular components, and KEGG pathways in which these four clusters are enriched is listed in the electronic supplementary material (Additional file 1: Table S4).

Discussion

Turner syndrome (TS) is a rare disorder that is associated with either complete or partial loss of one X chromosome in women, with an approximate occurrence of 1 in 2500 females [42]. Surprisingly, our knowledge of genotype–phenotype relationships in TS is rather limited, with very few specific candidate genes linked to its clinical features. The genotype–phenotype relationship in TS refers to the relationship between karyotype and phenotypic features of TS. TS is defined by the presence of an abnormal X chromosome in the form of monosomy, mosaicism of a 45X cell line with another cell line, which could be 46XX, 46XY, or an abnormal sex chromosome rearrangement [43, 44]. In TS, there is a link between the exact cytogenetic appearance and the phenotype. Pure 45, X monosomy is the most common karyotype and is associated with the most abnormal phenotype. Although the clinical features of TS are well-defined, the severity of the phenotype in TS individuals varies according to the underlying chromosomal constitution [44]. To better characterize gene expression differences between healthy and TS, we took a systems-biology approach by using weighted gene coexpression network analysis on 182 microarray peripheral mononuclear blood samples (PBMC). By comparing the modules in both healthy and TS, we aim at finding more relations between both healthy and TS PBMC samples including both the similarities and the differences.

The coexpression networks of healthy and TS samples had a scale-free topology that ensured network robustness which means they can resist random node attacks. Five modules were preserved between healthy and Turner Syndrome that carry many genes that were common (listed in Table 2) in each module. This means that these five modules are TS specific and are related to TS progression, which implicates many genes that were previously known and many that were unknown that function in TS. We found that many genes were common between the modules of TS and healthy coexpressed networks. Many of these common genes were differentially expressed in TS modules. This signifies that most of these genes being differentially expressed fail to perform the molecular functions of their respective modules. Two of these genes, namely *SMCHD1* and *PGK1* that were common between turquoise and brown modules, respectively, of healthy and TS have already been reported to be involved in TS. *SMCHD1* (OMIM 614982) plays a role in X chromosome inactivation [45]. It was reported earlier in a 13-year-old girl with Turner syndrome-like clinical features in association with intellectual disability, facial dysmorphism, and the psychomotor developmental delay had chromosome deletion syndromes and also had *SMCHD1* polymorphism [33]. This *SMCHD1* may correlate with TS. In another case study of a 28-year-old Turner syndrome patient, it was found through the DNA studies that the locus controlling X-inactivation is proximal to *PGK1* suggesting the role of *PGK1*, as well, in TS [34]. There were few more common genes between healthy TS modules, namely *CDC42SE2*, *ARPC4, RUNX1-IT1, ZBTB20, USP34, EMILIN2, GLIPR2* that were differentially expressed in TS and may prove to be significant in the progression of TS pathophysiology.

Many genes were different among the common healthy and TS modules. Few more previously reported genes of TS, specifically, *PTPN22, RPS4X, CSF2RA*, and *TIMP1* were found to be missing in their respective modules in TS when compared with healthy ones. Dysfunction or absence of these genes (listed in Tables 1 and 2) could then lead to a progressive disruption of the molecular pathways in a predictable sequence leading to the pathophysiology of TS.

It is unsurprising that the presence of 45,X monosomy results in dysregulation of cellular growth and repair pathways, leading to significant detrimental effects during both embryonic development and later stages. The extended cell cycle observed may mark the initiation of Turner syndrome (TS) pathogenesis, giving rise to a cascade of phenotypic consequences throughout various life stages, including embryonic/fetal, neonatal, pediatric, adolescence, and adulthood phases. The X chromosome harbors transcription factors influencing cell cycle duration and haploinsufficiency may contribute to cell cycle delays. Moreover, the insufficient expression of genes associated with DNA replication or repair on the X chromosome could further prolong the cell cycle in 45,X cells, impacting their ability to respond to S replication signals. TS patients often experience embryonic lethality, growth retardation, short stature, osteopenia/osteoporosis, congenital heart disease, gonadal dysgenesis, impaired pancreatic β-cell function, and neurologic deficits, manifesting as compromised neuronal microstructural integrity (connectivity) in white matter pathways.

While there were few molecular, biological functions and cellular pathways common among these preserved modules of healthy and TS, we also observed a significant difference in the functions of these modules when compared within and across healthy and TS samples.

These genes mentioned in Tables 1 and 2 are likely to be involved in molecular and biological functions and pathways that are syndrome-related, thus providing a new window of an investigation into TS pathophysiology. The identified pathways in individuals with Turner syndrome (TS) offer valuable insights into the underlying molecular processes associated with this condition. The dysregulation observed in these pathways has specific implications for various aspects of TS pathology. The spliceosome pathway, crucial for RNA splicing, may influence gene expression patterns, contributing to the diverse clinical manifestations seen in TS. Transcriptional misregulation, involving errors in gene transcription control, further accentuates the complexity of TS molecular dysregulation and its potential impact on aberrant gene activity. Platelet activation and vascular smooth muscle contraction dysregulations hint at possible connections to cardiovascular complications observed in TS. Abnormalities in osteoclast differentiation suggest potential implications for bone health, possibly contributing to conditions like osteoporosis. Lysosome and phagosome dysregulation may influence cellular homeostasis and immune responses in TS. Oxidative phosphorylation, central to energy production, highlights potential metabolic perturbations in TS. The presence of the non-alcoholic fatty liver disease (NAFLD) pathway points to potential liver-related complications in individuals with TS. Collectively, the identification of these pathways not only provides a more elaborated understanding of TS at the molecular level, but also lays the groundwork for future research and therapeutic strategies targeting the specific biological dysregulations associated with this genetic condition.

Further, we identified four clusters in the PPI network constructed from the top 15 K_{ME} in each module of the TS coexpression network. These modules were enriched in different significant molecular and biological functions, cellular components, and KEGG pathways. Each cluster has the highest scoring node called a seed. It is the node from which the cluster was derived. The genes *NDUFV1, EFTUD2, POLR1C,* and *STT3A* are the highest scoring node representing the seed genes of cluster 1, 2, 3, and 4, respectively, as mentioned earlier. They act as the eigengene of that cluster.

This comprehensive network-based meta-analysis methodology has previously been implemented by Jeremy et al. on human and mouse brain transcriptome [19]. This methodology of WGCNA has many advantages over traditional transcriptional analyses, leading to more reliable results compared to previous studies. Although, through this approach we uncover many facets of TS, there are several limitations such as the limited size of TS samples. The currently available datasets of PBMCs samples of TS do not allow a more elaborated study at this moment. Thus, a larger sample size would provide more elaborate results. Overall, our work highlights the potential molecular functions, pathways, and molecular targets when genomic data of TS and healthy women are compared and analyzed at the level of gene coexpression modules that can be targeted therapeutically in the future.

This work also provides an opportunity to ascertain if these genes originating from different modules have lost their functional attributes owing to mutational load or aberrant signal transduction or defunct interactomes. Since the number of interactomes are expected to be more for all the genes, their own detailed characterization would prove to be a rewarding proposition. From these interactomes, regulatory genes may be identified using network-based approach. In any case, these observations are likely to fuel the thought along the line for precise identification of genes actually implicated in giving rise to TS. Another important aspect is that after the implantation of the blastocyst, at what stage TS sets in and how the molecular heterogeneity is accentuated and prevailed. If we know what compels such an events, we may use the information for better management of TS patients including that of prenatal diagnosis.

Conclusion

In this study, gene expression differences between healthy and individuals with Turner syndrome were characterized using the systems-biology approach of weighted gene coexpression network analysis (WGCNA) on 182 microarray peripheral mononuclear blood samples (PBMC). While many genes were found to be common between the modules of TS and healthy coexpressed networks, some of the genes were different among the common healthy and TS modules. These genes may prove to be significant in the progression of TS pathophysiology. Dysfunction or absence of these genes could lead to a progressive disruption of the molecular pathways in a predictable sequence leading to the pathophysiology of TS.

Overall, our scientific work holds potential benefits for various groups within society. Primarily, biotechnologists, geneticists, and doctors stand to gain valuable insights from our analysis, as it may contribute to early diagnosis and proactive intervention in individuals at risk of Turner Syndrome. By identifying specific genes associated with the syndrome, our research could aid geneticists in establishing reference panels for more accurate and efficient genetic testing. This has the potential to facilitate early detection and intervention before symptoms manifest or worsen.

Moreover, our findings could pave the way for advancements in personalized medicine. Pharmaceutical companies might leverage the identified genes to develop Turner Syndrome-specific drugs aimed at reducing the severity of symptoms. This personalized approach has the potential to significantly improve the quality of life for individuals with Turner syndrome by addressing the underlying genetic factors. In essence, our work carries implications for both diagnostic practices and the development of targeted therapeutic strategies, offering a comprehensive contribution to the fields of genetics, medicine, and biotechnology.

Abbreviations

TS	Turner Syndrome
WGCNA	Weighted Gene Coexpression Network Analysis
PBMC	Peripheral Mononuclear Blood Samples
PPI	Protein–Protein Interaction
GEO	Gene Expression Omnibus
MCODE	Molecular Complex Detection
MDS	Multiple Dimensional Scaling

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43042-024-00491-9.

Additional file 1. Table S1. List of genes of each module of Set 1 (Healthy). **Table S2.** List of genes of each module of Set 2 (Turner Syndrome). **Table S3.** The list of genes of each cluster identified through MCODE. **Table S4.** The list of different molecular functions, biological functions, cellular components, and KEGG pathways in which these four clusters (Benjamini < 0.05).

Author contributions

R.I., S.A., and A.F. conceived the model. AF did the numerical experiments. A.F., N.T., and S.T. analyzed and interpreted the results. A.F. wrote the manuscript. All the authors read, edited and approved the manuscript.

Funding

A.F. is financially supported by the Indian Council of Medical Research under SRF (Senior Research Fellowship) (Project No. ISRM/11(50)/2017).

Availability of data and materials

The datasets analyzed during the current study are available in the GEO data repository (https://www.ncbi.nlm.nih.gov/geo/).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing financial interests.

Received: 14 May 2023 Accepted: 1 February 2024
Published online: 25 February 2024

References

1. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, Laurance MF, Zhao W, Qi S, Chen Z et al (2006) Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proc Natl Acad Sci 103(46):17402–17407. https://doi.org/10.1073/pnas.0608396103
2. Hood L (2004) Systems biology and new technologies enable predictive and preventative medicine. Science 306(5696):640–643. https://doi.org/10.1126/science.1104635
3. Muntaj S, Ganaie FA, Purva SV, Radhika S, Tilak P (2015) Karyotypic variables in Turner Syndrome: a case series. Int J Sci Study. 3(4):171–175. https://doi.org/10.17354/ijss/2015/330
4. Bondy CA (2007) Care of girls and women with Turner Syndrome: a guideline of the Turner Syndrome study group. J Clin Endocrinol Metab 92(1):10–25. https://doi.org/10.1210/jc.2006-1374
5. Gravholt CH, Andersen NH, Conway GS, Dekkers OM, Geffner ME, Klein KO, Lin AE, Mauras N, Quigley CA, Rubin K et al (2017) Clinical practice guidelines for the care of girls and women with Turner Syndrome: proceedings from the 2016 Cincinnati International Turner Syndrome Meeting. Eur J Endocrinol 177(3):G1–G70. https://doi.org/10.1530/EJE-17-0430
6. Farooqui A, Tazyeen S, MohdM A, Alam A, Ali S, Malik MdZ, Ali S, Ishrat R (2018) Assessment of the key regulatory genes and their Interologs for Turner Syndrome employing network approach. Sci Rep 8(1):10091. https://doi.org/10.1038/s41598-018-28375-0
7. Farooqui A, Alhazmi A, Haque S, Tamkeen N, Mehmankhah M, Tazyeen S, Ali S, Ishrat R (2021) Network-based analysis of key regulatory genes implicated in Type 2 Diabetes Mellitus and Recurrent Miscarriages in Turner Syndrome. Sci Rep 11(1):10662. https://doi.org/10.1038/s41598-021-90171-0
8. Pei G, Chen L, Zhang W (2017) WGCNA Application to Proteomic and Metabolomic Data Analysis. Methods in Enzymology, vol 585. Elsevier, Amsterdam, pp 135–158
9. Bouchama A, Aziz MA, Mahri SA, Gabere MN, Dlamy MA, Mohammad S, Abbad MA, Hussein M (2017) A model of exposure to extreme environmental heat uncovers the human transcriptome to heat stress. Sci Rep 7(1):9429. https://doi.org/10.1038/s41598-017-09819-5
10. Thompson SD, Marion MC, Sudman M, Ryan M, Tsoras M, Howard TD, Barnes MG, Ramos PS, Thomson W, Hinks A et al (2012) Genome-wide association analysis of juvenile idiopathic arthritis identifies a new susceptibility locus at chromosomal region 3q13. Arthritis Rheum 64(8):2781–2791. https://doi.org/10.1002/art.34429
11. Ioannidis I, McNally B, Willette M, Peeples ME, Chaussabel D, Durbin JE, Ramilo O, Mejias A, Flaño E (2012) Plasticity and virus specificity of the airway epithelial cell immune response during respiratory virus infection. J Virol 86(10):5422–5436. https://doi.org/10.1128/JVI.06757-11
12. Krug T, Gabriel JP, Taipa R, Fonseca BV, Domingues-Montanari S, Fernandez-Cadenas I, Manso H, Gouveia LO, Sobral J, Albergaria I et al (2012) TTC7B emerges as a novel risk factor for ischemic stroke through the convergence of several genome-wide approaches. J Cerebral Blood Flow Metabolism 32(6):1061–1072. https://doi.org/10.1038/jcbfm.2012.24
13. Kemppinen AK, Kaprio J, Palotie A, Saarela J (2011) Systematic review of genome-wide expression studies in multiple sclerosis. BMJ Open 1(1):e000053. https://doi.org/10.1136/bmjopen-2011-000053
14. Su R, Li MM, Bhakta NR, Solberg OD, Darnell EPB, Ramstein J, Garudadri S, Ho M, Woodruff PG, Koth LL (2014) Longitudinal analysis of sarcoidosis blood transcriptomic signatures and disease outcomes. Eur Respir J 44(4):985–993. https://doi.org/10.1183/09031936.00039714
15. LaBreche HG, Nevins JR, Huang E (2011) Integrating factor analysis and a transgenic mouse model to reveal a peripheral blood predictor of breast tumors. BMC Med Genomics 4:61. https://doi.org/10.1186/1755-8794-4-61
16. Risbano MG, Meadows CA, Coldren CD, Jenkins TJ, Edwards MG, Collier D, Huber W, Mack DG, Fontenot AP, Geraci MW et al (2010) Altered immune phenotype in peripheral blood cells of patients with scleroderma-associated pulmonary hypertension. Clin Transl Sci 3(5):210–218. https://doi.org/10.1111/j.1752-8062.2010.00218.x
17. Barnes MG, Grom AA, Thompson SD, Griffin TA, Pavlidis P, Itert L, Fall N, Sowders DP, Hinze CH, Aronow BJ et al (2009) Subtype-specific

peripheral blood gene expression profiles in recent-onset juvenile idiopathic arthritis. Arthritis Rheum 60(7):2102–2112. https://doi.org/10.1002/art.24601
18. Bahr TM, Hughes GJ, Armstrong M, Reisdorph R, Coldren CD, Edwards MG, Schnell C, Kedl R, LaFlamme DJ, Reisdorph N et al (2013) Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol 49(2):316–323. https://doi.org/10.1165/rcmb.2012-0230OC
19. Miller JA, Horvath S, Geschwind DH (2010) Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc Natl Acad Sci 107(28):12698–12703. https://doi.org/10.1073/pnas.0914257107
20. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8(1):118–127. https://doi.org/10.1093/biostatistics/kxj037
21. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95(25):14863–14868. https://doi.org/10.1073/pnas.95.25.14863
22. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinf 9(1):559. https://doi.org/10.1186/1471-2105-9-559
23. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30
24. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: Database for annotation, visualization, and integrated discovery. Genome Biol 4(5):P3
25. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B (2003) STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 31(1):258–261
26. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4(1):2. https://doi.org/10.1186/1471-2105-4-2
27. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504. https://doi.org/10.1101/gr.1239303
28. Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47–97. https://doi.org/10.1103/RevModPhys.74.47
29. Saelens W, Cannoodt R, Saeys Y (2018) A comprehensive evaluation of module detection methods for gene expression data. Nat Commun 9(1):1090. https://doi.org/10.1038/s41467-018-03424-4
30. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genetics Mol Biol. https://doi.org/10.2202/1544-6115.1128
31. Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5(2):101–113. https://doi.org/10.1038/nrg1272
32. Wu J, Tan S-Y, Liu Z, Tan Y-J, Lu X (2017) Enhancing structural robustness of scale-free networks by information disturbance. Sci Rep 7(1):7559. https://doi.org/10.1038/s41598-017-07878-2
33. Chen C-P, Lin S-P, Chern S-R, Wu P-S, Chen S-W, Lai S-T, Chuang T-Y, Chen W-L, Wang W (2018) A 13-year-old girl with 18p deletion syndrome presenting Turner syndrome-like clinical features of short stature, short webbed neck, low posterior hair line, puffy eyelids and increased carrying angle of the elbows. Taiwan J Obstet Gynecol 57(4):583–587. https://doi.org/10.1016/j.tjog.2018.06.019
34. Pettigrew A, McCabe ER, Ledbetter DH, Elder FF (1991) Isodicentric X chromosome in a patient with Turner syndrome implications for localization X of the inactivation center. Hum Genet 87:498–502. https://doi.org/10.1007/BF00197176
35. Bianco B, Verreschi ITN, Oliveira KC, Guedes AD, Galera BB, Galera MF, Barbosa CP, Lipay MVN (2010) PTPN22 polymorphism is related to autoimmune disease risk in patients with Turner Syndrome: PTPN22 polymorphism in Turner Syndrome. Scand J Immunol 72(3):256–259. https://doi.org/10.1111/j.1365-3083.2010.02438.x
36. Omoe K, Endo A (1996) Relationship between the monosomy X phenotype and Y-linked ribosomal protein S4 (Rps4) in several species of mammals: a molecular evolutionary analysis ofRps4Homologs. Genomics 31(1):44–50. https://doi.org/10.1006/geno.1996.0007
37. Zinn AR, Bressler SL, Beer-Romero P, Adler DA, Chapman VM, Page DC, Disteche CM (1991) Inactivation of the Rps4 gene on the mouse X chromosome. Genomics 11(4):1097–1101
38. Urbach A, Benvenisty N (2009) Studying early lethality of 45, XO (Turner's Syndrome) embryos using human embryonic stem cells bridger JM, editor. PLoS ONE 4(1):e4175. https://doi.org/10.1371/journal.pone.0004175
39. Berletch JB, Yang F, Disteche CM (2010) Escape from X inactivation in mice and humans. Genome Biol 11(6):213. https://doi.org/10.1186/gb-2010-11-6-213
40. Anderson CL, Brown CJ (1999) Polymorphic X-chromosome inactivation of the human TIMP1 gene. Am J Human Gene 65(3):699–708. https://doi.org/10.1086/302556
41. Trolle C, Nielsen MM, Skakkebæk A, Lamy P, Vang S, Hedegaard J, Nordentoft I, Ørntoft TF, Pedersen JS, Gravholt CH (2016) Widespread DNA hypomethylation and differential gene expression in Turner syndrome. Sci Rep 6:34220. https://doi.org/10.1038/srep34220
42. Shen C-H (2019) Molecular Diagnosis of Chromosomal Disorders. Diagnostic Molecular Biology. Elsevier, Amsterdam, pp 331–358
43. Oliveira RMRD, Verreschi ITDN, Lipay MVN, Eça LP, Guedes AD, Bianco B (2009) chromosome in Turner syndrome: review of the literature. Sao Paulo Med J 127(6):373–378. https://doi.org/10.1590/S1516-31802009000600010
44. Sagi L, Zuckerman-Levin N, Gawlik A, Ghizzoni L, Buyukgebiz A, Rakover Y, Bistritzer T, Admoni O, Vottero A, Baruch O et al (2007) Clinical significance of the parental origin of the X chromosome in Turner Syndrome. J Clin Endocrinol Metab 92(3):846–852. https://doi.org/10.1210/jc.2006-0158
45. Blewitt ME, Gendrel A-V, Pang Z, Sparrow DB, Whitelaw N, Craig JM, Apedaile A, Hilton DJ, Dunwoodie SL, Brockdorff N et al (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat Genet 40(5):663–669. https://doi.org/10.1038/ng.142

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.