Methods

Microarray Data

The microarray expression profiling dataset GSE46687, deposited by Bondy et al., was downloaded from the Gene Expression Omnibus. The dataset was based on the GPL570 Affymetrix Human Genome U133 Plus 2.0 Array platform. The experiment contained 36 samples consisting of 16 subjects with TS who were identified as having a maternally inherited X chromosome (45,Xm), 10 subjects with TS who were identified as having a paternally inherited X chromosome (45,Xp) and 10 subjects with the normal female karyotype (46,XX). Since it was public dataset, the information of age and health condition as well as the usage of the medication of the individuals was unavailable, which appears to be a potential limitation. The annotation file for GPL570 was also downloaded from the GEO.

Differential Expression Analysis

Differential expression analysis was performed using the online analysis tool GEO2R; the expression profiles of monosomy X TS patients and normal 46,XX females were compared to identify the DEGs. *P*-values and adjusted *P*-values were calculated using *t*-tests. Genes from each sample with the following criteria were retained: (1) a |log2 (fold-change)| >1 and (2) an adjusted $P < 0.05$. We selected the most significant genes when the DEGs were duplicated. We divided the TS patients into two groups depending on the parental origin of the existing X chromosome. Analyses were performed independently for the 45,Xm and 45,Xp TS samples, and the DEGs were determined by the intersection of the two datasets. A Venn diagram of DEGs was drawn using the online tool Venny 2.1, and the heatmap for the DEGs was created using Heml software.

Tissue-Specific Gene Expression Analysis

We used the online resource BioGPS to analyze the tissue-specific expression of the DEGs. Transcripts mapped to a single tissue with the following criteria were identified as highly tissue specific: the tissue-specific expression level was >10 times the median, and the second highest expression level was less than one-third as high as the highest level.

Functional Enrichment Analysis of DEGs

We used DAVID 6.8 to perform the functional enrichment analysis of DEGs; this analysis included the functional categories, Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The GO analysis included 3 categories, namely, biological process (BP), cellular component (CC) and molecular function (MF), which were used to predict protein functions. KEGG pathway analysis was used to assign sets of DEGs to specific pathways to enable the construction of the molecular interaction, reaction and relationship networks. The functional categories included COG_ONTOLOGY, UP_KEYWORDS and UP_SEQ_FEATURE. Benjamini-adjusted $P < 0.05$ and an enriched gene count >5 were chosen as the criteria for significance.

Protein-Protein Interaction (PPI) Network Analysis

The PPI network analysis was conducted using STRING, which is an online database of known and predicted protein-protein interactions. These interactions include physical and functional associations, and the data are mainly derived from computational predictions, high-throughput experiments, automated text mining and co-expression networks. We mapped the DEGs onto the PPI network and set an interaction score of >0.4 as the threshold

value. In addition, Cytoscape v3.6.0 software was used to visualize and construct the PPI network. Nodes with the greatest numbers of interactions with neighboring nodes were considered hub nodes.

To identify the key PPI network modules, the app ClusterOne from the Cytoscape software suite was used to perform the gene network clustering analysis. A $P < 0.05$ was set as the significance threshold for identifying key modules.

Results

Differentially Expressed Genes

We downloaded the microarray expression dataset GSE46687 from the GEO database and analyzed the DEGs between monosomy X TS patients and normal female 46,XX individuals using the online analysis tool GEO2R. In total, 42 upregulated and 91 downregulated genes were identified between Xm TS patients and normal individuals. In addition, 279 upregulated and 234 downregulated genes were identified between Xp TS patients and normal individuals. We identified the intersection of these two datasets and obtained a total of 25 upregulated and 60 downregulated genes. As shown in, 15 (17.6%, one upregulated and 14 downregulated) of the 85 DEGs were on the X chromosome. Most of these genes are involved in basic cellular activities, such as the structural maintenance of chromosomes and the mediation of transcription. Three of the downregulated genes (*AP1S2, CSF2RA*, and *CD99*) on the X chromosome were related to the immune system. The *X inactive specific transcript* (*XIST*) gene was downregulated, and the adjusted *P*-value was highly significant ($P < 0.0001$). The Venn diagram and

heatmap for the DEGs are presented in . As shown in , 25 upregulated and 60 downregulated genes were identified through the comparison of TS patients and normal individuals. LMF1, a protein-coding gene involved in the maturation and transport of lipoprotein lipase, was upregulated in Xp TS patients, whereas it was downregulated in Xm TS patients.