Supplementary information, Data S1

Materials and Methods

Diagnosis of pectus excavatum (PE)

In this study, all participants underwent an extensive, standardized physical examination. The study protocols were approved by the Institutional Review Board and the Ethics Committee of Shenzhen Second People's Hospital and informed consents were obtained from all the participants. PE most frequently affects the lower sternum and cartilages but the degrees of rotation and asymmetry can be varied greatly. The diagnosis for PE in this study required the involvement of the lower sternum and cartilages. All patients showed a well-defined PE phenotype which is characterized by the obviously deep depression of the lower end of sternum as well as the costal cartilages through 4 to7.

Targeted capture and exome sequencing

Genomic DNA was extracted from whole blood in EDTA, using QIAGEN QIAamp DNA Blood Mini Kit. 12 µg of genomic DNA from each of the five selected individuals (II:5, III:4, III:6, IV:1 and IV:2) were sheared into ~90 bp DNA fragments by sonication. Exome capture was performed to collect the protein coding regions of human genome DNA using a NimbleGen Ez V2.0 (44.1 Mb) according to the manufacturer's instructions (Roche NimbleGen, Inc., Madison, WI, USA). The array covers more than 30,000 genes in the human genome and the gene or exon

annotations for this array come from the RefSeq (Jan. 2010), CCDS (Sep. 2009) and miRBase (v.14, Sep. 2009) databases. The exon-enriched DNA libraries were then subjected to a second library construction in preparation for Illumina exome sequencing and were sequenced using the Illumina Hiseq 2000 platform, following the manufacturer's instructions (Illumina, San Diego, USA).

Read mapping and variant analysis

The human reference genome, together with its gene annotations, was downloaded from the UCSC database (http://genome.ucsc.edu/, version hg19 (build37)). The sequencing reads from the five individuals was aligned using SOAPaligner after we removed the duplicated reads and the genetic variants were called using SOAPsnp with the default parameters [1]. The other thresholds for calling genetic variants included the following: 1) the number of unique mapped reads supporting a variant should be ≥ 1 and ≤ 200; and 2) the consensus quality score should be ≥ 20. The candidate disease-causing variants were obtained by filtering the shared variants of the four patients against the list of variants from the normal individual. All candidate disease-causing variants were further filtered against the total variants data available in the 1000 Genomes Project (November 23, 2010 releases for SNPs, ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/), and then against the SNP set in the dbSNP132. SIFT (http://sift.jcvi.org) was used to predict whether the identified mutations affect the protein sequences.

Mutation validation

All potentially pathogenic variants were amplified by Bigdye terminator v3.1 cycle sequencing kits (ABI, Foster City, CA, USA) and analyzed on an ABI 3730XL Genetic Analyzer.

Expression cloning

Wild-type GAL3ST4 coding sequences was cloned into the pPyCAGIP expression vector. GAL3ST4 p.R11W mutant was generated with the PCR primers used for plasmid construction and site-directed mutagenesis (Supplementary information, Table S3).

Hela cell culture

HeLa cells were grown in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum, 100 U/ml penicillin and 100 μg/ml streptomycin at 37 °C, 95% humidity and 5% CO2. One day before transfection, cells were split and seeded at 50% confluency directly into 24-well plates in medium without antibiotics. Then, 0.8 μg plasmid DNA was used per transfection per well (24-well plate) using Lipofectamine 2000 (Invitrogen) reagent (2.5 μl Lipofectamine per μg DNA).

Immunofluorescence

24 h post-transfection, cells were split and seeded at 10% confluency directly into

24-well plates in medium without antibiotics. 24 h later, cells were fixed in 4% paraformaldehyde for 15 min, washed in PBS, permeabilized in 0.25% TritonX-100, then washed in PBS, followed by block in 1% BSA for 30 min and incubation with rabbit anti-GAL3ST4 antibody (1:100, Abcam) overnight. After washing in PBS, cells were incubated with anti-rabbit-Cy3 (1:200, Invitrogen). The nuclei were counter-stained with DAPI (1:10000, Invitrogen) for 5 min and rinsed again in PBS. All steps were performed at the room temperature.

Reference

1 Li R, Yu C, Li Y, *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009; **25**:1966-1967.