Appendix S1

SUPPLEMENTARY SUBJECTS AND METHODS

Subjects

A family of Chinese Han origin with typical features of TOC was identified in Liaoning Province, China. This family consists of 65 individuals, 12 of who (5 females and 7 males) exhibited manifestations of TOC. A written informed consent was obtained from all subjects. This study was approved by the Institutional Review Board and the Ethics Committee of The First Hospital of China Medical University and conducted according to the Declaration of Helsinki Principles. Plantar skin and whole layer of the esophageal mucosa samples were collected from 3 TOC family members with tylotic skin and esophageal cancer (III9, III11 and III18), and 3 healthy family members (III10, III12 and III19). Biopsies of the whole layer of mucosa were also obtained from 3 sporadic squamous cell esophageal cancer patients. Two experienced dermatologists carefully examined all the individuals. Two well-trained pathologists read all tissue sections. Peripheral blood samples were collected from members of this family, and 100 unrelated healthy control individuals of Chinese Han origin. Genomic DNA was isolated from these samples.

Library preparation and sequencing

Exome capture was carried out for 3 affected and one unaffected family members. The exome sequences were efficiently enriched from 0.4 μg genomic DNA using Agilent liquid capture system (Agilent SureSelect Human All Exon V6) according to the manufacturer's protocol. Firstly, qualified genomic DNA was randomly fragmented to a mean size of 180–280 bp by Covaris S220 sonicator. Remaining overhangs were converted into blunt ends via exonuclease polymerase activities. Secondly, DNA fragments were end repaired and phosphorylated, followed by A-tailing and ligation at the 3'ends with paired-end adaptors (Illumina). DNA fragments with ligated adapter molecules on both ends were selectively enriched in a PCR reaction. After the PCR reaction, libraries hybridize with liquid phase with biotin-labeled probe, then use magnetic beads with streptomycin to capture the exons of genes. Captured libraries were enriched in a PCR reaction to add index tags in order to prepare for sequencing. Products were purified using AMPure XP system (Beckman Coulter, Beverly, USA) and quantified using the Agilent high sensitivity DNA assay on the Agilent Bioanalyzer 2100 system. At last, DNA library were sequenced on Illumina for paired-end 150 bp reads.

Reads mapping to reference sequence

Valid sequencing data was mapped to the reference genome (GRCh37/hg19) by Burrows-Wheeler Aligner (BWA) software to get the original mapping result in BAM format. Subsequently, Samtools and Sambamba were respectively utilized in both the steps of sorting bam files and doing duplicate-marking to generate final bam file. If one or one pair read(s) has multiple mapping positions, the strategy adopted by BWA is to select the best one. If there are multiple best mapping positions, we randomly pick one. Mapping step is very difficult due to mismatches, including true mutation and sequencing error, and duplicates resulted from PCR amplification. These duplicated reads are uninformative and should not be considered as evidence for the variants. Picard was employed to mark these duplicates so that we would ignore them in the following analysis.

Variant calling

In this step, reads that aligned to the exon regions were collected for mutation identification and subsequent analysis. Samtools mpileup and bcftools were used to do variant calling and identify SNP, indels. Since CNV detection is not so accurate, we employed the reliable user-friendly computational pipeline-CoNIFER to discover disruptive genic CNVs in human genetic studies of disease, which might be missed by standard approaches.

Functional annotation

Functional annotation is very important because the link between genetic variation and disease can be found in this step. ANNOVAR was performed to do annotation for Variant Call Format file obtained in the previous step. The variant position, variant type, conservative prediction and other information were obtained in this step through a variety of databases, such as dbSNP, 1000 Genome, GnomAD, CADD and HGMD. Since we are interested in the exonic variants, gene transcript annotation databases, such as Consensus CDS, RefSeq, Ensemble and UCSC, were also applied for annotation to determine the amino acid alternation. The variants were shared among 3 cases, but were neither present in the public database nor in the control individuals of the family. These variants were considered to be the candidate variants.

Sanger sequencing validation

All exons of *RHBDF2*, *ITGB4* and *CDK3* genes, including intron–exon boundaries, were amplified by PCR, using oligonucleotide primers. DNA sequencing was performed by ABI 3700 DNA Sequencer (Applied Biosystems, Foster, USA). The DNA sequences of the affected individuals, unaffected individuals and 100 unrelated normal controls were compared with the *RHBDF2*, *ITGB4* and *CDK3* gene sequences in GenBank.

Hematoxylin-eosin staining and Immunohistochemistry

Specimens from normal and affected plantar skin, normal esophagus tissues, sporadic and tylotic squamous cell tumors tissues were fixed in 10% formalin, paraffin-embeded, cut into 5-μm serial sections and mounted on slides for histological examinations with hematoxylin-eosin staining. RHBDF2 protein expression was studied by immunohistochemistry in formalin-fixed and paraffin-embeded material as described previously (13). Tissue sections were incubated with the rabbit anti-human RHBDF2 antibody (1:50, Sigma, Saint Louis, USA) overnight at 4°C, and further developed with biotinylated goat anti-rabbit IgG antibody (1:100, ZSGB-Bio, Beijing, China). The samples were dehydrated and clarified, using a conventional method, and prepared for examination under a light microscope.

Immunofluorescence

Skin slices were fixed in acetone for 10 min, incubated with rabbit anti-human RHBDF2 antibody (1:50, Sigma) for 60 min at 37°C, and rinsed. R-phycoerythrin-conjugated goat anti-rabbit IgG antibody (1:80, ZSGB-Bio) was applied for 60 min to label RHBDF2. Control is the substitution of primary anti-human RHBDF2 antibody with isotype mouse IgG (1:50, ZSGB-Bio).