SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Transcript activity of Cl3 promoter of *LRRC37* genes
RT-PCR analysis on the cDNA isolated from the total RNA extracted from Human *H. sapiens* by using brain, fetal brain, testis, liver, kidney, heart, lung, skeletal muscle, thymus, and cerebellum tissues using the primers designed to exon 1 and 4 starting from the Cl3 transcription start point **(A, left)**, *DND1* first two exons **(B, left)**, and first two exons Cl8 derived transcript **(C, left)**, respectively. *UBE1* is used as negative control. Testis consists of only two exons including only the first and fourth exons of *LRRC37A1* **(A, left)**. The figure shows the relative fold expression of Cl3 promoter activity **(A, right)**, *DND1* promoter activity **(B,right)** and Cl8 promoter activity **(C, right)**, on the cDNA isolated from the total RNA extracted from Human *H. sapiens* by using brain, fetal brain, testis, liver, kidney, heart, lung, skeletal muscle, thymus, and cerebellum tissues. Expression data were first normalized against housekeeping gene *UBE1* and then cross-compared using the heart tissue as reference.

Figure S2. Phylogenetic reconstruction of *DND1* promoter duplication within New World monkeys
Phylogenetic reconstruction of *DND1* derived Cl3 promoter in New World monkeys (NWM) leading to human by using the NJmethod (Saitou and Nei 1987). Species names are indicated as: Mmu Rhesus macaque (*Macaca mulatta*), Ppy Orangutan (*Pongo pygmaeus*), Hga Gibbon *(Hylobates gabriellae)*, Ggo Gorilla *(Gorilla gorilla)*, Ptr Chimpanzee (*Pan troglodytes*), and Hs Human (*Homo sapiens*). Different human individuals are indicated as NA18555, GM12236, NA19129, NA12878, NA18507, GM15510, Hs 1f (Cl3 promoter H1 Haplotype Forward), Hs 1r (Cl3 promoter H1 Haplotype Reverse), Hs 2f (Cl3 promoter H2 Haplotype Forward), Hs 2r (Cl3 promoter H2 Haplotype Reverse), and Hs *DND1* (Human original *DND1* promoter). Lower case next to the species name indicates name of the respective individual. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (pairwise deletion option). Phylogenetic analyses were conducted in MEGA4 (Tamura et al. 2007).

Figure S3. Evolutionary relationships of the *DND1*-derived Cl3 promoter
Phylogenetic reconstruction of *DND1* derived Cl3 promoter in different primates, cow, horse, and mouse species using the NJ method (Saitou and Nei 1987). Species names are indicated as:Bt Bovine (*Bostaurus*), Ef Horse (*Equusferus*), Mmd Mouse (*Musmusculusdomesticus*), Mmu Rhesus macaque (*Macaca mulatta*), Ppy Orangutan (*Pongopygmaeus*), Ptr Chimpanzee (*Pan troglodytes*), and Hs Human (*Homo sapiens*). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed

using the Jukes-Cantor method and are in the units of the number of base substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option). Detection of transcription factors were determined by publically available database (Alibaba2) and phylogenetic footprinting analysis. Phylogenetic footprinting analysis is performed using a program (Footprinter 2.0) (Blanchette and Tompa 2003). The results were embedded on the phylogenetic tree (see Methods). Arrow indicates the fixation of mutations in respective transcription factor binding motif on phylogenetic tree leading to human.Phylogenetic analyses were conducted in MEGA4 (Tamura et al. 2007).

Figure S4. Alignment of *DND1* and Cl3 regions

The alignment of the core duplicon region from *DND1* and Cl3 promoter including first exon. The species names are indicated as Bt Bovine (*Bos taurus*), Ef Horse (*Equus ferus*), Mmd Mouse (*Mus musculus domesticus*), Mmu Rhesus macaque (*Macaca mulatta*), Ppy Orangutan (*Pongo pygmaeus*), Ptr Chimpanzee (*Pan troglodytes*), and Hs Human (*Homo sapiens*). Red highlighted sequence indicates the position of transcription start site. Numbers next to the species name indicate the position of the extracted sequence from the respective species. Alignment was performed using ClustalW (Ebi) with default options and highlighted using Boxshade server version 3.21.

Figure S5. Evolutionary relationships of introns and Cl3 promoter sequences (linearized)

The evolutionary history of the intronic region between exons 3-4 and (-1)-(-2) from the beginning of ling coding exon and Cl3 promoters that is extracted from database and sequenced from clones. The species names are human Hs (*Homo sapiens*) and macaque Rh, Mmu (*Macaca mulatta*). The phylogenetic tree was reconstructed using the NJ method (Saitou and Nei 1987). Phylogenetic tree was linearized assuming equal evolutionary rates in all lineages. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tajima-Nei method and are in the units of the number of base substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option). Phylogenetic analyses were conducted in MEGA4 (Tamura et al. 2007).

Figure S6. Evolutionary relationships of Cl8 promoter region with ancestral *BPTF* promoter sequence

The evolutionary history was inferred using the NJ method (Saitou and Nei 1987). The species names are human Hs (*Homo sapiens)*, macaque Rh (*Macaca mulatta*), and mouse Mmd (*Mus musculus domesticus*). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tajima-Nei method and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of

208 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura et al. 2007).

Figure S7. Alignment of *BPTF* and Cl8 promoter regions
The alignment of duplication region from *BPTF* and Cl8 promoter region is shown. The species names are indicated as Mmd Mouse (*Mus musculus domesticus*), Rh Rhesus macaque (*Macaca mulatta*), and Hs Human (*Homo sapiens*). Numbers next to the species name indicate the position of the extracted sequence from respective chromosome of species. Alignment was performed using ClustalW (Ebi) with default options and highlighted using Boxshade server version 3.21.

Figure S8. Dot-plot comparative analysis of mouse and human *LRRC37* genes
A) Dot-plot matrix analysis of *LRRC37*A genes from human and mouse. The nucleotide sequence file for *LRRC37*A (NM_014834) and mouse *LRRC37*A extracted from the sequence database are compared. The dot-plot matrix analysis is performed using the program SEAVIEW (version 4.1) (Galtier et al. 1996) with the option window size 20 identity matches 13. **B)** Schematic representation of mouse and human LRRC37 protein structure is shown. The figure is scaled except for the exon 1 region. SS indicates the signal sequence presumably used for subcellular localization of the LRRC37 protein and RPT is unknown amino acid repeat sequence. Transmembrane domain (TMD) is highlighted with orange box. The data is obtained by using SMART (Simple Moduler Archtecture Research Tool)(http://smart.embl-heidelberg.de/)

Figure S9. Phylogeny of the *LRRC37* family
The evolutionary history of the *LRRC37*family was inferred using the neighbor-joining method (Saitou and Nei 1987). The intronic core duplicon region between exons 4 and 5 is used for alignment. Species names are indicated as Md Mouse (*Mus musculus domesticus*), Rn Rat (*Rattus norvegicus*), Cf Dog (*Canis familiaris*), Bt Bovine (*Bos taurus*), Ec Horse (*Equus caballus*), Cja Marmoset (*Callithrix jacchus*), Rh Rhesus macaque (*Macaca mulatta*), Ppy orangutan (*Pongo pygmaeus*), and Hs human (*Homo sapiens*). *LRRC37* copies detected to be active in expression analysis are colored in red. The copy of *LRRC37* (Rh6343), which is only active in cerebellum, is not included due to the deletion of core duplicon region including exons 4 and 5. *LRRC37* copies detected to be highly active in cerebellum (C), testis (T), or both tissues (C/T). The number next to the species name indicates the position of the duplicon within a genome of respective species. *LRRC37* core duplicon region $Hs_{27577498}$, $Ppy_{4014333}$, $Ppy_{28292972}$, $Ppy_{28296257}$, $Rh_{27312797}$, and $Rn_{92660724}$ are removed from the final dataset due to error formation in the phylogenetic tree reconstruction. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option). There were a total of 1583 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura et al. 2007).

Figure S10. Snapshot of *LRRC37* genes on *MAPT* locus

The figure depicts the depth-of-coverage for respective species within the region of MAPT locus. Species names are indicated as Mmu Rhesus macaque (*Macaca mulatta*), Ppy Orangutan (*Pongo pygmaeus*), Ptr Chimpanzee (*Pan troglodytes*), and Hs Human (*Homo sapiens*).