SUPPLEMENTARY INFORMATION

Eye movement defects in KO zebrafish reveals *SRPK3* as a causative gene for an X-linked intellectual disability

Yu-Ri Lee[1,2,16], Mervyn G. Thomas[3,16], Arkaprava Roychaudhury[1,16], Cindy Skinner[4], Gail Maconachie[3,5], Moira Crosier[6], Holli Horak[7], Cris S. Constantinescu[8,9], Tae-Ik Choi[1], Jae-Jun Kyung[1], Tao Wang[10], Bonsu Ku[11], Bernard N Chodirker[12], Michael F. Hammer[13], Irene Gottlob[3,9], William H. J. Norton[14], Albert E. Chudley[12], Charles E. Schwartz[4,15,17], and Cheol-Hee Kim[1,17]

MATERIAL AND METHODS

X-chromosome exome sequencing for Family-1

X-chromosome exome sequencing for mutation identification in the probands of XLID families were conducted as described [1].Briefly, sequence libraries were prepared using a TruSeqTM Genomic DNA Library Preparation kit (Illumina, San Diego CA, USA), enriched for the human X chromosome Exome using a SureSelect Target Enrichment kit (Agilent, Santa Clara, CA, USA), and sequenced using the 75bp pair-end sequence module on the HiSeq2000 (Illumina, San Diego CA, USA). Alignment of the fastq reads, base recalibration, and variant calling were completed using Bowtie2 and Unified Genotyper (GATK).

Disease-causing mutations were enriched using the nonclinical portion of the dbSNP, the male-restricted portion of the 1000 Genomes, Exome Variant Server datasets, and an affected sib-pair/cross-cohort-based algorithm [1]. Sanger sequencing was used for variant validation, segregation analysis, and polymorphism studies of each mutation using the BigDye Terminator v3.1 Cycle Sequencing Kit on an ABI3100 automatic DNA analyzer (Applied Biosystems, Massachusetts, USA) following manufacturer's instruction. Variant analysis was completed on standard sequence alignment software (CodonCode and MacVector) followed by manual investigations of the chromatograms.

A total of 3 missense variants involving highly conserved amino acids residues in SRPK3 were identified in 7 affected males from three X-linked intellectual disability families recruited at Greenwood Genetic Center, Greenwood, SC, USA. Among these variants, c.475C>G; p.H159D was identified in Family 1 by X-chromosome exome sequencing [1].; c.1373C>A; p.Thr458Asn was identified in Family 2 at NHGRI; and c.1585G>A; p.E529K was identified in Family 3.

Whole exome sequencing for Family-2

Whole exome sequencing was performed twice, once by array capture of 50 Mb of exome target sequence using the Agilent SureSelectXT Human All Exon V5 enrichment kit and again using the 64 Mb Illumina TruSeq enrichment kit for all four samples in the family. The four libraries from

each kit were then pooled, and each set was sequenced on one lane of an Illumina HiSeq 2000 using a 2x100 bp paired-end protocol. Sequences were trimmed with trimmomatic, v.0.32 [2] and then aligned to the human genome (GRCh37) using Burrows-Wheeler Aligner, v.0.6.1 [3]. PCR duplicates were removed with Picard, v.1.70. Base quality recalibration, indel realignment, and variant calling were performed using the Genome Analysis Toolkit, v.1.4-37, as previously described [4]. Variants were annotated using SnpEff v. 3.4 [5] with gene annotations made against Ensembl release 73. Previously known variants were annotated with their allele frequencies (using all ethnic affiliations) from the 1000 Genomes Project (www.1000genomes.org/data) and the NHLBI GO Exome Sequencing Project (ESP) 6,500 samples release (http://evs.gs.washington.edu/evs/). Nonsynonymous variants were annotated with Polyphen2 [6] and SIFT scores [7]..

Candidate X-linked variants that were predicted to alter protein function (i.e., nonsynonymous, stop-gain, stop-loss and splice-junction mutations) were identified when the male proband is hemizygous on variant for which mother is a carrier and father has a reference nucleotide on the allele. In order to winnow our list of candidates to a set of true X-linked variants, we implemented a prioritizing system that hierarchically placed each candidate into one of 6 ‘levels’ according to sequence or genome architecture features that ranged from lowest (level 1) to highest (level 6) probability of being a false positive.

The following level criteria has been used:

- **Level 6** Probable genotype calling error (significant allelic disbalance in child and/or in parents for the corresponding genotype call) (2 variants found)
- **Level 5** Variant is within 5bp of an indel
- **Level 4** Variant is in gene with more than one Mendelian inheritance error (5 variants found)
- **Level 3** Variant coverage (read depth) is less than minimal coverage threshold (16 reads) below which sensitivity and zygocity of calls is questionable
- **Level 2** Variant is found in 6500 exomes and/or 1000 genomes/dbSNP public genomic variant databases with alternative allele frequency lower than 0.03, but above 0.0 (3 variants found)
- **Level 1** Highest priority (1 variant found)

All candidate X-linked variants within level 1 were submitted for validation testing by Sanger sequencing. We also screened for inherited variants with a frequency <3% in the ESP, both in cases of homozygous recessive or compound heterozygous state (i.e., when the same variant was present as two copies in the proband but only in one copy in both parents, and when two different variants were present in the same gene in the proband and were inherited from different parents, respectively). Candidate de novo mutations in proband present neither in the mother nor in the father and never previously observed in the 1000 Genomes or ESP datasets were also examined.

Expression studies for protein and mRNA

Paraffin sections were dewaxed and rinsed in two changes in absolute ethanol. Endogenous peroxidase activity was blocked with methanol peroxide solution for 10 min at room temperature. Sections were then rinsed in Tris Buffered Saline (TBS) and incubated with 10% normal serum in TBS for 10 min at room temperature. Primary SRPK3 antibody (R&D Systems cat MAB72301) was added in normal serum blocking solution in TBS at a 1:50 dilution and incubated overnight at

4°C. Sections were washed and incubated with horse anti-mouse secondary antibody in normal serum in TBS at 1/500 dilution for 30 min at RT. Sections were then incubated with tertiary complex utilizing standard VECTASTAIN Elite ABC kit PK6100 in TBS for 30 min at RT and developed with Sigma DAB kit (D-4293) for 10 min at RT. Sections were then counterstained with haematoxylin (10s) then washed, dehydrated, cleared, and mounted for light microscopy. Human embryos with normal karyotype and morphology were studied at Carnegie stage (CS) 15 (~33 days post ovulation (dpo)), CS19 (~48 dpo) and CS23 (~56 dpo). The human embryonic and fetal material was provided by the Joint MRC/Wellcome Trust (Grants MR/006237/1, MR/X008304/1 and 226202/Z/22/Z) Human Developmental Biology Resource.

Zebrafish husbandry

Adult zebrafish were reared at 28.5℃ under standard conditions of 14:10 hours of light and dark cycle. Male and female zebrafish were mated and the eggs were harvested and kept in egg water (5mM NaCl, 0.17mM KCl, 0.33mM CaCl 2, 0.33mM MgSO 4,10-5% Methylene Blue) at 28.5°C.

Whole mount *in situ* hybridization

RT-PCR was performed with larval zebrafish cDNA by using Superscript Ⅲ (Invitrogen, Waltham, Massachusetts, USA) reverse transcriptase and primers according to manufacturer's protocol. The desired amplicon was cloned into pGEM-T easy vector (Promega, Madison, Wisconsin, USA) and digested using *Not I* (NEB, Ipswich, Massachusetts, USA) restriction enzyme. *In vitro* transcription was carried using mMESSAGE mMACHINE T7 Transcription Kit (Ambion, Waltham, Massachusetts, USA) according to manufacturer's protocol. 1 µg of linearized DNA was used in synthesizing RNA probe which was dissolved in HYB+ at 50 ng/µl concentration and stored in -20ºC till experiment was conducted. Antisense RNA probes were synthesized using Dioxygenin (Roche, Basel, Switzerland). Different stages of zebrafish embryos, treated with 0.03% Phenylthiourea (PTU) (Sigma Aldrich, St. Louis, Missouri, USA), were fixed overnight using 4% Paraformaldehyde (PFA) (Sigma Aldrich, St. Louis, Missouri, USA) and dehydrated with 100% MeOH and stored at -20°C. On the day of the experiment, embryos were rehydrated using with 0.1% DEPC-PBST (Diethyl Pyro-carbonate-Phosphate Buffer Solution-Tween 20). Embryos were permeabilized using 0.5% Proteinase K and refixed using 4% PFA for 20 mins. The larvae were washed with DEPC-PBST and incubated in prehybridized HYB solution (50% formamide, 5X saline sodium citrate, 50µg/ml heparin, 500µg/ml torula RNA, 46mM citric acid pH 6.0, 0.1% Tween20) for 1 hour at 70°C. Antisense DIG labelled RNA probes were added in HYB solution and incubated overnight at 70°C. Larvae were washed in a preheated mixture of 50% saline sodium citrate containing 0.1% Tween-20 and 50% hybridization solution at 70ºC and subsequently washed several times with PBST. Then the larvae were treated with blocking solution 0.5% Horse serum with PBST for 1 hour at room temperature. Then the larvae were incubated in with 1/4000 concentration of anti-DIG Fab fragment conjugated with alkaline phosphatase (Roche, Basel, Switzerland) overnight at 4ºC. The larvae were washed in PBST at room temperature and incubated in staining solution included Nitro-blue tetrazolium chloride/5-bromo-4-chloro-3-indoyl phosphate (NBT/BCIP) (Roche, Basel, Switzerland) as an alkaline phosphatase substrate in the dark until sufficient staining appeared. Larvae were mounted in 90% glycerol in PBST and used for imaging.

Generation of *srpk3* KO zebrafish model

Zebrafish *srpk3* gene sequences were acquired from the NCBI database (*srpk3*: XM_689881.9). Primers for *in vitro* transcription of sgRNA were designed for *srpk3* target sites. Primers contained 5' T7 promoter and 15 nucleotides at 3' complement to the universal primer. Templates for in vitro transcription of sgRNAs to knock-out *srpk3* targets were produced by PCR with universal primer

5’ –TAATACGACTCACTATAGGGAGGAAGGACTGGCACACGTTTTAGAGCTAGAA – 3’ (4). *In vitro* transcription was carried out using MaxiScript T7 Kit (Ambion). Cas9 expression vector pT3TS-nCas9n (Addgene plasmid # 46757), linearized with *Xba I* (NEB) and purified using DNA purification kit (ELPIS, Korea). Cas9 mRNA was synthesized using mMESSAGE mMACHINE T3 Kit (Ambion, Waltham, Massachusetts, USA) and tailed poly (A) with E. coli Poly(A) Polymerase (NEB, Ipswich, Massachusetts, USA). sgRNAs of *srpk3* were injected with Cas9 mRNAs into one-cell stage embryos with 300 ng/µl Cas9 mRNA and 150 ng/µl sgRNA. Mutations were validated using T7E1 Assay. Founder (F0) were raised to adulthood and outcrossed with wild type zebrafish to generate germ line mutations. Once the founder fishes were confirmed, F1 generation were raised to adulthood and similar genotypes were incrossed resulting in stable *srpk3* KO zebrafish line (F2). For the morphological analysis, adult male and female *srpk3* heterozygous KO zebrafish were crossed and embryos were collected and kept at 28.5°C until imaging was done. 5 dpf zebrafish larvae were anaesthetized using MS222 (Sigma Aldrich, St. Louis, Missouri, USA). Then the larvae were mounted on 3% methyl cellulose (Sigma Aldrich, St. Louis, Missouri, USA) for imaging.

Analysis of spontaneous eye movements in KO zebrafish at 5 dpf

Adult male and female *srpk3* heterozygous KO zebrafish were crossed and embryos were collected and kept at 28.5°C until the experiment was conducted. 5 dpf zebrafishes were mounted with 2% low melting agarose in a 6 cm petri dish (SPL Biosciences, Korea) and video record was taken for 3 minutes using MZ16FA Brightfield microscope (Leica Microsystems, Germany). Analysis of eye movements and tracking was done using MATLAB (Mathworks, USA) .

Social interaction assay in adult KO zebrafish

Social interaction assay was conducted as described previously. Behavior tank (24 x 12.5 x 12.5 cm), covered on three sides with non-transparent material and open on side, was filled with over three liters of system water. The tank was divides into two sections with a transparent acrylic partition. One section of the chamber was designated for the cue fish whereas the other part of the chamber was assigned for the tester fish. The section was divided into four equal chambers as Zone 1,2,3 and 4. Behavioral experiment was conducted between 13:00 hrs. and 17:00 hrs. During the experiment, three adult zebrafish were added as cue fish and single *srpk3* sibling were added as the tester fish. 15 minutes of video was taken from the lateral side using a SONY HDR-CX405 camera. Videos were analyzed using Ethovision XT (Noldus, the Netherlands).

Histological analysis

The brain was sectioned and stored at -20°C. Sections were stained with H&E for evaluation of brain structure. Selected sections were used to prepare cryosections for histological studies described below.

Adult zebrafish were fixed using 4% Paraformaldehyde (Sigma Aldrich, St. Louis, Missouri, USA) in 1X PBS overnight. After fixation, the samples were dehydrated and stored in 70% EtOH at -20°C. Paraffin embedding and microtome were performed to obtain tissue sections for H&E staining. 5 dpf zebrafish larvae and tissue imaging of H&E staining were done under Brightfield stereo microscope MZ16 Brightfield microscope (Leica Biosystems, Wetzlar, Germany).

Evaluation of apoptosis

Evaluation of cell death was determined via a vital fluorescent dye acridine orange method that often used as a marker of apoptotic cells in zebrafish. Acridine orange staining is a rapid method

and has been shown to be highly selective for apoptotic cells where determination of apoptosis was made by measuring relative acridine orange fluorescence in zebrafish embryos. Zebrafish larvae were placed in the egg water containing 4 µg/ml of acridine orange (Sigma Aldrich, St. Louis, Missouri, USA) for 20 min. Afterwards, the live zebrafish larvae were washed with the egg water 5 times for 5 min each, anesthetized with tricaine and mounted in 3% methylcellulose, prior examined by stereomicroscopy and fluorescent microscopy.

RESULTS

Table S1 Clinical features of seven XLID patients from three families having point mutation in the *SRPK3* gene. Most common features are learning difficulties and agenesis of the corpus callosum in patients. Poor attention span, abnormal smooth pursuit in eye movement, and lazy eye movement were clinically reported in four patients.

Patient	F1:II-1	F1:III-3	F1:III-4	F1:III-5	F1:III-7	F2	F3:III-1
Mutation	c.475C>G; p.H159D	c.475C>G; p.H159D	c.475C>G; p.H159D	c.475C>G; p.H159D	c.475C>G; p.H159D	c.1373C>A; p.T458N	c.1585G>A; p.E529K
De novo/Maternal	Maternal	Maternal	Maternal	Maternal	Maternal	Maternal	Maternal
Age (last exam)	69 yrs	35 yrs	27 yrs	35 yrs	32 yrs	28 yrs	15 yrs
Global Developmental Delay	Yes	Yes	Yes	Yes	Yes	n/a	Yes
Intellectual Disability (IQ)	Severe	Moderate (IQ 60-70); delayed language skills	Moderate (IQ 60-70); delayed language skills	Moderate (IQ 60-70); delayed language skills	Moderate (IQ 60-70); delayed language skills	Learning problems; borderline (IQ low normal)	Learning difficulties (IQ unknown)
Dysmorphic Feature	Prominent forehead; macrocephaly	None	None	None	None	None	None
Neurological Feature	n/a	n/a	n/a	n/a	Poor balance a nd postural reaction with p oor integration of vestibular stimulation	Ataxia and vertigo	Ataxia; proximal muscle weakness
Psychiatric Feature	n/a	n/a	Poor attention span	Poor attention span; emotional instability	n/a	n/a	n/a
Vision/hearing	None	n/a	None	None	n/a	Lazy eye at birth; strabismus; hyperacute	Abnormal smooth pursuit in eye movement; convergence insufficiency
Verbal	Yes	Yes	Yes	Delayed speech	Yes	Speech disturbance	n/a
Heart/Kidney/Genitalia	None	n/a	n/a	n/a	n/a	None	n/a
Musculo-skeletal	Contractures of digits	Delayed motor	Delayed motor	Delayed motor	Delayed motor	n/a	Mild muscle weakness
Neuroimaging (X ray, MRI, CT PET)	Enlargement of the lateral ventricles and the 3rd and 4th ventricles; partial absence of the corpus callosum	Irregular and asymmetrically dilated lateral ventricles; white matter loss; partial agenesis of the corpus callosum and colpocephaly	Irregular and asymmetrically dilated lateral ventricles; white matter loss; agenesis of the corpus callosum and colpocephaly	n/a	Irregular and asymmetrically dilated lateral ventricles; white matter loss; partial agenesis of the corpus callosum	CT PET - slightly decreased cerebellar uptake	Loss of volume within posterior body of the corpus callosum

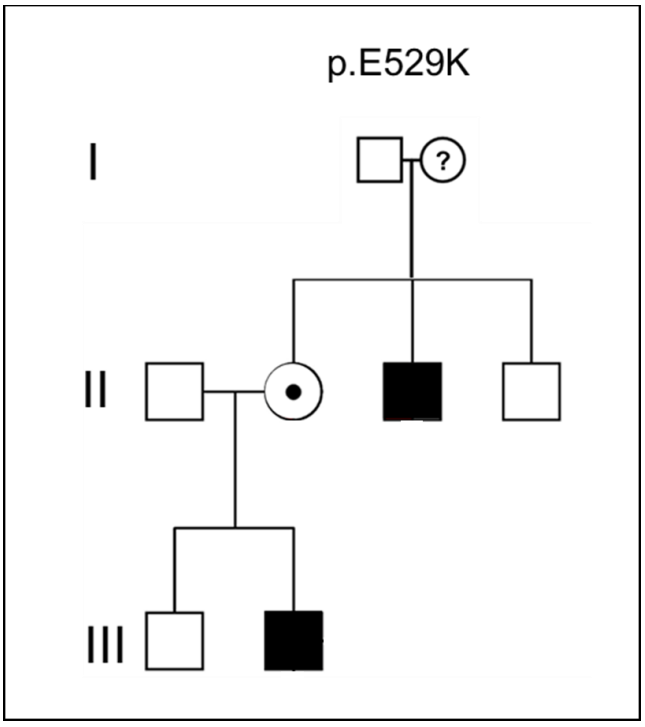


Fig S1 Pedigree of Family 3. The proband presented with ID, muscle hypotonia, and ataxia. His maternal uncle also had similar phenotypical characteristics. Whole exome sequencing was performed on four samples from the family. The third mutation in the *SRPK3* gene (c.1585G>A; p.E529K) was identified from this family. The proband's mother was heterozygous for the *SRPK3* variant and asymptomatic.

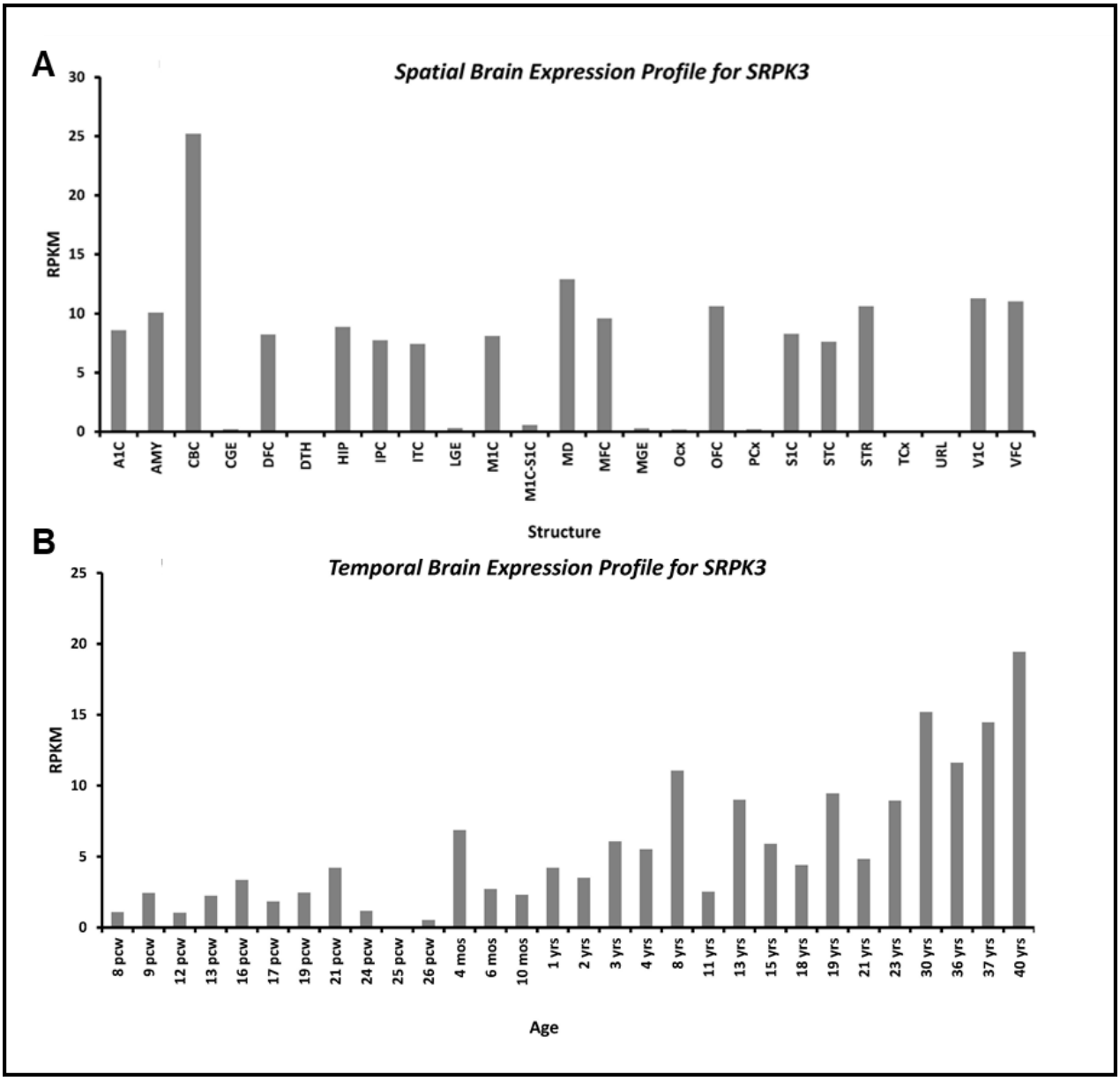


Fig S2 Transcriptomic analysis identifies spatiotemporal expression of *SRPK3* transcripts in human. **(A)** Spatial mRNA expression profile of *SRPK3* showing widespread brain expression with highest level of expression in the cerebellum. Expression RPKM values are cumulative from 8 weeks post conception to 40 years. RPKM = reads per kilobase of exon model per million mapped reads; A1C = primary auditory cortex; AMY = amygdaloid complex; CBC = cerebellar cortex; CGE = caudal ganglionic eminence; DFC = dorsolateral prefrontal cortex; DTH = dorsal thalamus; HIP = hippocampus; IPC = posteroventral parietal cortex; ITC = inferolateral temporal cortex; LGE = lateral ganglionic eminence; M1C = primary motor cortex; M1C-S1C = primary motor-sensory cortex; MD = mediodorsal nucleus of thalamus; MFC = anterior cingulate cortex; MGE = medial ganglionic eminence; Ocx = occipital neocortex; OFC = orbital frontal cortex; PCx = parietal neocortex; S1C = primary somatosensory cortex; STC = posterior superior temporal cortex; STR = striatum; TCx = temporal neocortex; URL = upper rhombic lip; V1C = primary visual cortex; VFC = ventrolateral prefrontal cortex. **(B)** Temporal mRNA expression profile of *SRPK3* showing increasing expression levels with age.

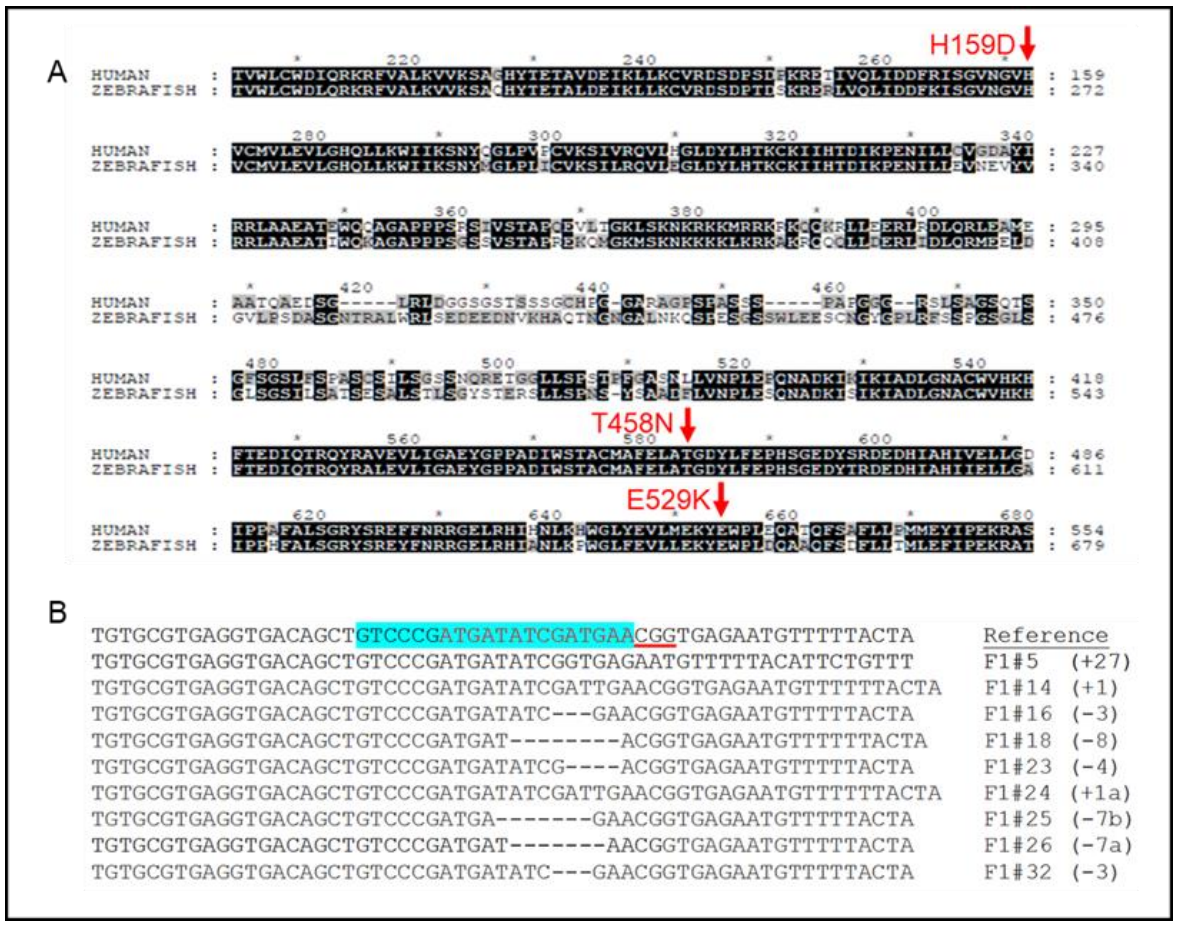
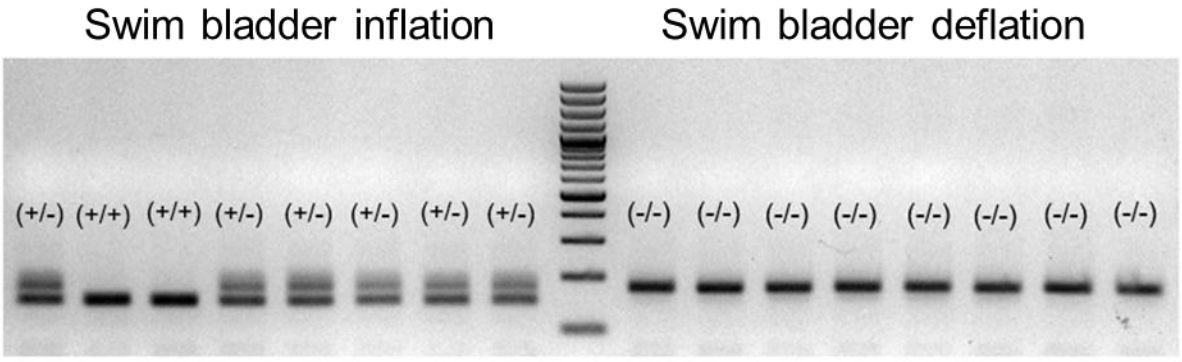


Fig S3 Conserved amino acid sequence of SRPK3 between human and zebrafish (A) Comparison of human and zebrafish SRPK3 proteins. Three variants identified from the XLID patients are indicated by red arrows. (B) Confirmation of induced mutations in *srpk3* KO zebrafish generated by CRISPR KO technology.

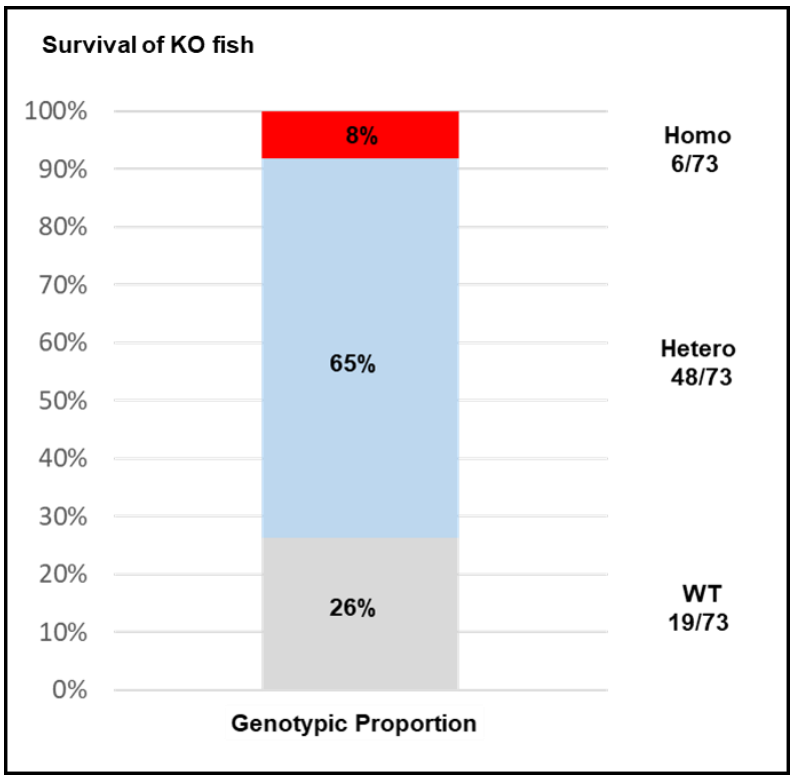
(A)



(B)

Clutch #	Swimbladder inflation	Swimbladder deflation
1	147/192	45/192 (23.4%)
2	211/268	57/268 (21.3%)
3	210/265	55/265 (20.1%)
4	300/375	75/375 (20%)
5	52/65	13/65 (20%)
6	235/290	55/290 (18.9%)
7	196/231	35/231 (15.2%)
8	201/232	31/232 (13.4%)
9	225/240	15/240 (6.3%)
10	200/209	9/209 (4.3%)
Total	1977/2367	390/2367 (16.5%)

Table S2 Correlation between swimbladder phenotype and genotype in *srpk3* KO zebrafish. (A) After selection of zebrafish larvae at 5 dpf on the basis of swimbladder phenotype (inflation vs. deflation), their genotype (for example, clutch #6) was confirmed by PCR analysis; wild type (+/+), heterozygous (+/-), and homozygous (-/-) KO siblings, respectively. **(B)** As a representative analysis, total 2367 larvae from 10 independent clutches were examined. The incidence of swim bladder defects was decreased from 23.4% to 4.3% in subsequent generations.



241

242 **Fig S4 Survival ratio of *srpk3* KO zebrafish to adulthood.** *srpk3* hetero in-cross produced three different
243 genotypes (Wild type, Heterozygous and Homozygous KO zebrafish) in 26%, 65% and 8%, respectively,
244 reaching adulthood. As a case study, total 73 adult zebrafish of siblings were genotyped.

245

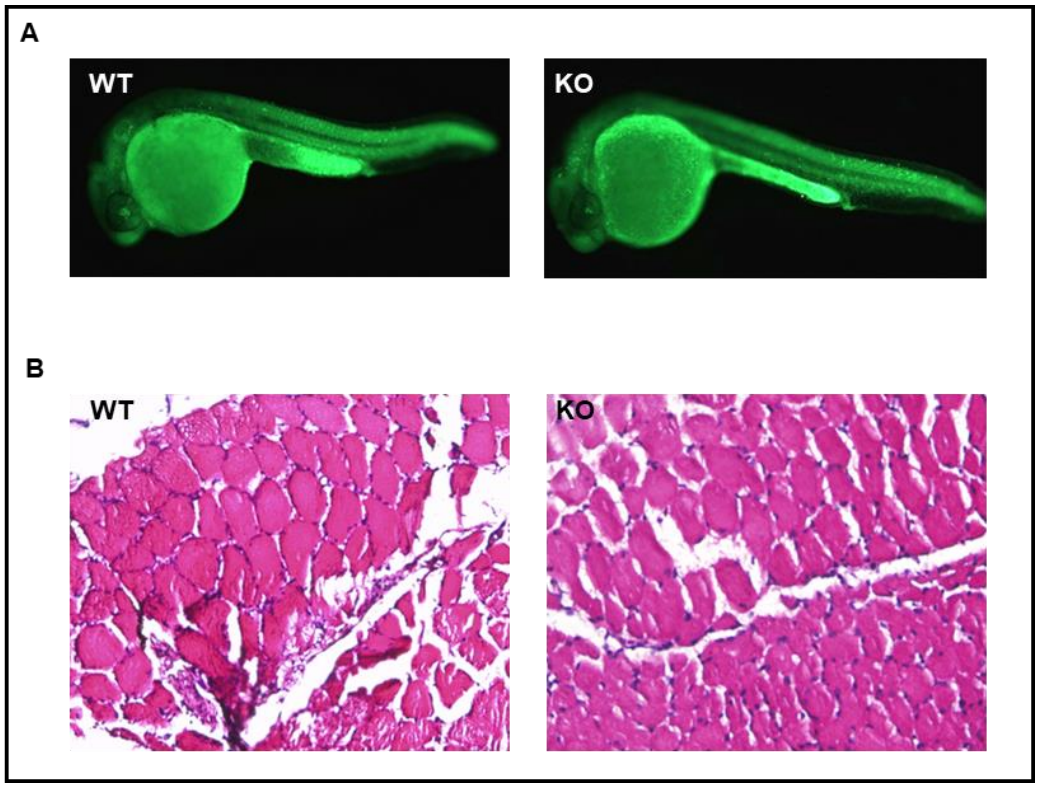


Fig S5 No muscle defect in *srpk3* KO zebrafish at early embryonic stage and in adult tissues. (A) Acridine orange staining for cell death in zebrafish embryos at 24 hpf. No significant change in cell death was observed in *srpk3* KO zebrafish. **(B)** Muscle histology (H&E staining) in adult zebrafish. No myopathy was observed in muscle tissue of adult *srpk3* KO zebrafish, compared to that of control WT fish.

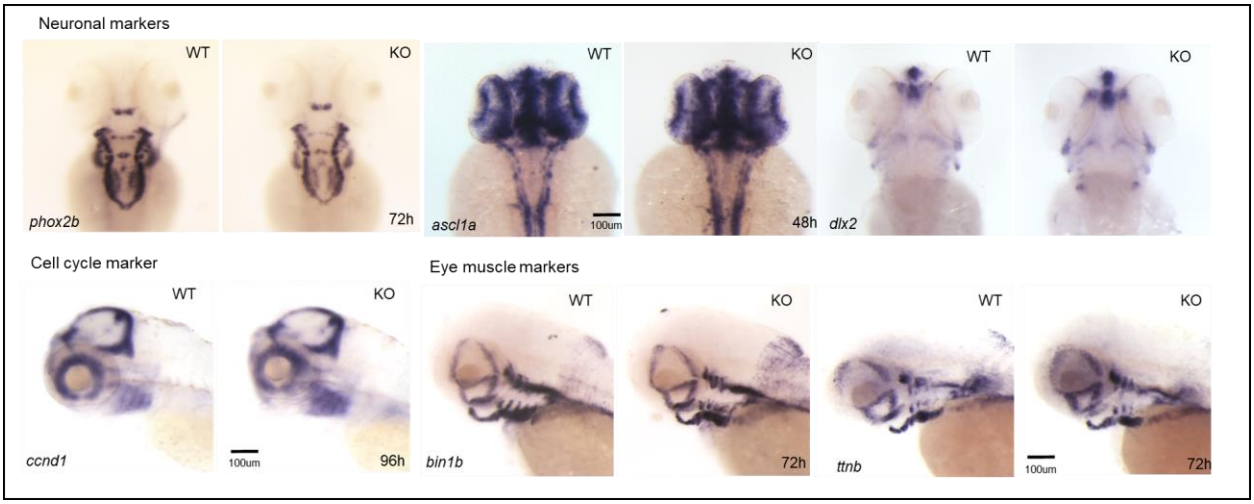


Fig S6 Expression of neuronal, cell cycle and eye muscle markers. Expressions of *paired-like homeobox* 2b *(phox2b), achaete-scute homolog 1a (ascl1a), distal-less homeobox 2* (*dlx2), cyclind1 (ccnd1), bridging integrator 1b (bin1b),* and *tandem duplicate 1 (ttnb)* were checked in early stages of zebrafish development but no significant changes were observed in wild type and its KO siblings.

Movie S1



Movie S2



Movie S1 Spontaneous eye movements in WT (+/+) siblings. Representative WT #2 larva (1:00 - 2:00 min), 2x speed.

Movie S2 Spontaneous eye movements in KO (-/-) siblings. Representative KO #1 larva (1:00 - 2:00 min), 2x speed.

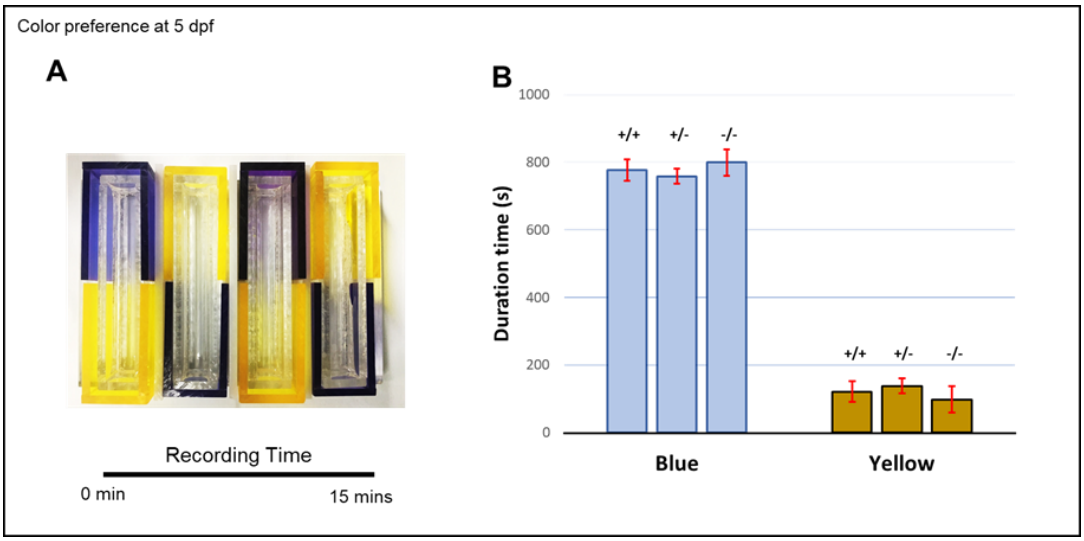


Fig S7 Color preference test of *srpk3* KO zebrafish at 5 dpf. Blue vs. yellow color preference assay has been conducted at 5 dpf in *srpk3* KO zebrafish, but no significant changes have been found in the wild type, heterozygous and homozygous KO zebrafish. Color preference test recording of representative control (+/+, n = 11), *srpk3* heterozygous knockout (+/-, n = 19) and *srpk3* homozygous knockout (-/-, n = 7) zebrafish for 15 minutes.

SUPPLEMENTAL RFERENCE

1. Niranjan TS, Skinner C, May M, Turner T, Rose R, Stevenson R, Schwartz CE, Wang T. Affected kindred analysis of human X chromosome exomes to identify novel X-linked intellectual disability genes. PLoS One. 2015 Feb 13;10(2):e0116454.
2. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20.
3. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010 Mar 1;26(5):589-95.
4. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010 Sep 1;20(9):1297-303.
5. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. fly. 2012 Apr 1;6(2):80-92.
6. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nature methods. 2010 Apr;7(4):248-9.
7. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009 Jul;4(7):1073-81.