The American Journal of Human Genetics, Volume 99

# Supplemental Data

# Loss-of-Function Mutations in ELMO2 Cause

# Intraosseous Vascular Malformation

# by Impeding RAC1 Signaling

Arda Cetinkaya, Jingwei Rachel Xiong, \_ Ibrahim Vargel, Kemal Kösemehmetoğlu, Halil İbrahim Canter, Ömer Faruk Gerdan, Nicola Longo, Ahmad Alzahrani, Mireia Perez Camps, Ekim Zihni Taskiran, Simone Laupheimer, Lorenzo D. Botto, Eeswari Paramalingam, Zeliha Gormez, Elif Uz, Bayram Yuksel, Şevket<br>Ruacan, Mahmut Samil Sağıroğlu, Tokiharu Takahashi, Bruno **Mahmut** Sağıroğlu, Takahashi, Reversade, and Nurten Ayse Akarsu

# **Supplemental Figures**



## **Figure S1. Resolving the Complex Rearrangement**

(A) IGV images of the aligned paired-ends obtained via targeted sequencing in affected individual D-II:2. Proximal segment and Distal segment represent sequences in the vicinity of either end of the deletion breakpoints, whereas Segment A and Segment B are the sequences inserted in the complex rearrangement. For interpretation of the colors please refer to the IGV user guide on the IGV website (interpreting color by pair orientation). Briefly; grey paired-ends: both ends are facing each other (normal orientation); blue pairedends: both ends are in left orientation; aqua paired-ends: both ends are in right orientation; green paired-ends: both ends are in opposite orientation.

#### **Figure S1. Resolving the Complex Rearrangement (continued)**

(B) Schematic representation of the paired-ends that are discordant and/or long-insert (>1000 bp) as visualized in IGV. Detailed examination in family D using IGV showed 6 different types of discordant and/or long-insert (>1000 bp) paired-end reads around the deleted sequence, shown in (A). The upper panel shows 6 types of abnormal paired-end reads in the targeted sequencing data. Rearrangement of the genomic sequence, as shown in the lower panel, resolves discordant and/or long-insert paired-ends and predicts the complex rearrangement sequence. A: Segment A; P: Proximal segment; B: Segment B; D: Distal Segment.

(C) Schematic representation of the events that lead to the complex rearrangement. Circled numbers 1, 2, and 3 indicate the joining events during formation of the complex rearrangement, as follows, Event 1: Joining of Proximal segment and inverted Segment A; Event 2: Joining of inverted Segments A and B via a 13 bp joining segment; Event 3: Joining of inverted Segment B and Distal Segment.

(D) The 3-primer PCR designed to detect wild-type and rearranged alleles of *ELMO2*. Schematic representation of the PCR primer binding positions on the genome shows that the wild-type allele produces a 542 bp long PCR product, whereas deletion due to the complex rearrangement produces of a 697 bp long product using a different reverse primer (shown on the left). The right panel shows the agarose gel electrophoresis of the PCR products for affected individual D-II:2, who is homozygous, and her father (D-I:1), who is heterozygous for the complex rearrangement. An unrelated control (UC) is homozygous for the wild-type allele. M: 100bp DNA ladder (Promega).

## >out\_236 size=4247 cov=147.71

Proximal Segment

(45.027.684) GGAGGCACAGAGGAGTTAAGTAGAAGGCTCAGTGTCACATAGCTAGCAAGTGGCACA GCCAGAATTTGATACCAGGTTGTCTGGTGTTGCTGGAAAGCCATGAGAAAGCAGGTGGCTCTGCACCTGA CCCGCAGCAGGTGCTTTGTGGAGATTTCTGGTTTGTGTATCTGAAACCAAAGCACACTGAGGCTAAGAAA CCTTCACAGCAAGAACTCACCATTGTAACAGAAGACAGCTCTCCACTGTCTACAGGTGGGCCACACTCCC TCGTGTGCTATTTGGTCCTTTACAATCTAGCCATCCTGGACTACGGGCTATTTCTATAAGGCAGGCCATG TTCTTTCAATTCTTCTGTACCTTTGCACATTCTCTTCCTCCTACTGACTGTCAGAAGAGCTCCTATTCAA GCATCAGAGCCTAGTTCGGATGTTACCTCCTCCTTGTACTAGTCTTAGCTTTCTCTACTTTCCCTGGCAA AAGTAAATAGGACAGGGGCCTCTGTGTTCTCACAGCATCCTGTTCATACTTCCGTTAATGAGTACGATTT TGTCATACTCTATTACAGTTACATATTTACACATCTCCACTAGCTCCTCCTCCAGAACAGGTATAACTCT GCAGTCTCATAACCTAGCACAGCACTTGGTGTCCAGTGAAGGCTCAGTAAGCACCTGTGTCTGCTTAAAC CCAATGGCTCTGATTTCTTCCACTCCTGTGAACCAGGCTGGCTCCAAACAGACAGGGACAGGCATGAACT GTGCAGGCTGTCTGGTGTTCCCTCTGCGTCCCAATACCCTCTAACTACACAGTATTTCCATTGTTTAACA GTTTGTTTTGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAATGCTTTGGGAGGTCAAGGCAGGAGGAT CACTTCAGGCCAGGAGTTCGAGACATTTCTCTACAAAAAATAAAAATTAGCTGGGCATGGTGGCGTGTGC CTGTAGTCCCTAGCTACTCAGGAGGCTGAGGCAGGAGAGTTGCTTGAGCCCAGGAGTTCGAGGCTGCAGT GAACTATGATCATGTCACTGAACTCTAGCCTGGGTGACAAAGTGAGATCCTGTCTCTATTAAAAAATAAA AATAAAATAAAAAACTGTTGTTTTCATATCAAACATTAGAATAAGTCTTACTGAAAAATCAGGGTTAGCA CCCCTAACTCTAGTTGAAAGAAAGAGGAAAACCTCCCTCTGCTTTTCTCCCTATCGAAAAGAAACATCAG GAGATAGGAGACTAGAAATAATTCTGTTCTCCAGATAATATTCCCAGGATAAAGGATCTCCTTATAGGGA ACCCTTCCAGCACGTGTTTGCTTACTCCCTTCATTTTCATTCTAGAAAGGAAGTAGAAATTAGGTCAAAG GAATTCAGTAACATTGCAAAATTCTCAGGGTTATAAATAGCCTCTTGCTCCGCCACTCAGCAGTACTGTA GCTCAATTGCAAACCACCAGCATAAGGGAGCTGTAGTTCCATATCCAGGCCCAGAGGGCCCCTTCCGGAT TAAACAACTCCCCACCCTTCTACCTTCCACTCCTCTTTAAGCATTACAGATATAAGGATGATTTGGAAAA CTCAAAATGCTACAAAACAGGAGACAGGCACACACTGACCTGATGACAATTTTACTATAAGTCATAAGAA TGAAAGGCCAGTTTTTCAAACACAGAGCTGGGAAAAGTAGAGAAAAATAGGAACATTAGTGTCTCCAATG GAGCAGTCTCTAAGTGGCTTCTAGAAGGTGTGCTTATAAAATCAGCACTGTTACAGAAGCCCCATGGAGC TTTATCAATCACCCATGCAGTGACTGCTGCTTCACGGAGCTGGATTCTGAGAGCAGGATTGGAAACCGAA TCTTAAGGAATGAAGTGAAAGATGATCTCCTCCTATTAGTAAGGAATTTGGGATGAAGAATGCCAGGTTT CCAGCTCCCCTCCTTCCCCCTATTCCTTTAAGGATAAGAACGGATGTGTACTTTTTGTTCAGTTATTTGG CTTTTGATTTGGTGAAGGAGGCATGGGTAAGTTTGGTTTATGTTTCCTGCAACAGAATGCAACCTCATCA AGAAGGCAAGTTGGAGTCACGCCCTTTCTAATCGAGAACACAATCTCTATAAAGGCTTCCCCCTCTGAAA TCTTCTCATCTGAGAAGTGATTCAACTTCCCTTAGTTCAGTCCACTGAATGACGATCAGACTCTGAGGTA CAGAAAAGGGAGGATTTGACTGGCAGAGATTAAGGTTCTTATGTTTACTGCTGGAAGCCCTGGGAAGGAC CACCATACCAACAGCAGCTGAGGTTTCCCCTTTTTGTTGAAGTATGTTAGACACAGGACTCATTCACACA TGGTAAAAAGCCACAAACTTACATAGAGGCAAAGGTGTGGTGAGAAGGGGTGTGGGGGCAGGGGAACTAT TTGGGTTTCTTCATCTTGATTCCTGCTGTAAGAATGGCCTGGGTCTTCTCTTATCCTGCACAGACACCTT CTGTAGTGTTCCTATACTTAAAGAATAGTGCTTAAGTAAGCTCCTTAGCAGAGCATTCAACTTACTCCAG TTCCACCGTGGCTCTGTGGTGGAACTGGCTTTCAGATTCGCTTTTGCAGATGCCATGTTCTCACCAGACC ACCTGTTCCCTGAAAAACCATGCTCTTCTGAGTCTGTATGTTTGCTTGTGCTATTATTTCCTCTTCCTTG TCCAGCTGGCAAACTTCAACTCAACCTTCAAGGCCCAGCTAAGATAATCACTCTTCTATGTGGCTTTCCT AATCCTCTCACCTCAGGGAAAATTAATTGCAGCCTCATCTGTGATCCCAGTCTTCTGTACATACCTTCCT GATCCTACTGATTTATAATTCTCTATGTGCAGTCTCTTCCCTGGCTACGATAGGTGTGATGTGCATACCA GCAGCGGCAGTATCCCTGGGAGCCCAGACCCACCCAATCAGAATCTGCCTTCAAACAAGATCCTGAGGTG GTTCAAGTGCATGTTCAAGTCTGGGACATAGGCTCTAGATCAGATAAGGTCCGAGAAACAGTGGTTCTCA ACTCTGGTTGTGCACCATCAACACACACCCCAGGTCATTTATAAACTCTCAGATGCTCAGGGTTGACCCA CAGCTCCAGGAGAGTTGACAACCTGAGGGCAAGGACTCATGAAGCCTAAGTCTATTCCCCCATCTACTCT TCCCCCAGTTTGCAGCACACTGCCTGCCACATAGCTGTGTCCACAGAATTCCACAGGAAAATTCCCTCCA GGGTGCCTCCAGACCTAAATGCCTTCAGAAGAGAGTGAAAACACTACTGACTGAGGATTATTAATTGCCT AGAACTTTCCAAATCCCTAACCACAATGACTGAAGAGATGAATTTACAAAAAGTCAGTGTTTACAAACCA CTTACCTCAGTTTTTGGAAATGAGAGGAAACAGAAACTGAGATGGAAAATACAGATGAACTTCAGAATTA AACTTTAGAAAACCGTGATT(45,031,190)

#### Segment A inverted

(45,023,171) GCAGCCGTGTCTGTGTTTTTGTCTCGCAGAATTAGAGCCCATTGGGAACGATGCCAC CACCGTCAGACATTGTCAAAGTGGCCATTGAGTGGCCAGGTGCTAACGCCCAGCTCCTTGAAATCGACCA GGTATGCTCCTGAAGTGAGAAGCAGTGGTTCAAGGAAAGGCACCTGGGGAGTGCATGGCAGAGGACATCT TGAGGGATGGGGACCACCGGCATCAAGAGTAAGAACGAGCAACAGGAAGGCTAAGCTTTGGGC (45,022,912)

# Joiningsegment<br><mark>TTACAA</mark>CTTAAGC

#### Segment B inverted (45,031,203)

(45,031,259) TTAGTTGTAACCCAGGACTCTTCCCAGCACACCACACTCCCTCCCTACCCAAGCTCC

#### DistalSegment

(45,037,129) CAGTGGGCAACTAGGTTCAGGTGCCGTGTTTTCAGGCTGGAATCAAATAGGAGGAGT ACCTGCAGCTGAATGTCATGATTACACAAATGCAATTGTGTCTGGCACGTGGGAGAGGATAGTAGAGATT CTGAGCTTTGTAACCTTGCAGACCTTTGTGCTATTTACCAGCTGTTTGTCCCTGGGTAAGTTACTAGATC TCTCTGAGCCTGTCTCCTCATGTGTACAATGGGGAGAATAACAGTACCAGCCTTAATGGATTGTTGTGAG AAATGAGATCACTGTAAATATACAACAGTTAGCAAAATGTCTGCATTTAAGTGCTCAATAAATTGTAAAT TATAAAATTTTAAAATTATAGTAGGCCTTCCATAAGTACTTATTGTGTGAGTGAATTATCTACCCCTCTC AGT(45.037.538)

### **Figure S2.** *De novo* **Assembly of the Sequence Around Complex Rearrangement**

### **Figure S2.** *De novo* **Assembly of the Sequence Around Complex Rearrangement (continued)**

Presented here is the continuous sequence of the complex rearrangement assembled using paired-end reads from targeted massively parallel sequencing of D-II:2 that align between Chr20:45,027,684 and Chr20:45,037,538 on the reference sequence. The segments of the complex rearrangement are named as shown in Figure 2. The starting and ending positions of each segment on the reference genome are indicated in red for Proximal Segment, Segment A inverted, Segment B inverted, and Distal Segment, whereas no position is indicated for the 13 bp long joining segment sequence that joins inverted Segments A and B. Sequences highlighted in yellow show microhomology segments at breakpoint junctions of Proximal Segment-Segment A inverted and Segment B inverted-Distal Segment. Note that the joining segment is not a random sequence, but consists of 2 microhomology segments near the boundaries of Segments A and B highlighted in blue and green. The sequence highlighted in blue is a direct repetition of 6 nucleotides near the boundary of Segment A inverted with an additional thymidine. The sequence highlighted in green is an inverted repetition of 6 nucleotides near the boundary of Segment B inverted. Presence of  $>1$ microhomologies, several basepairs apart from breakpoints suggests that >1 DNA synthesis and trimming event may have occurred during event 2 (Figure S1C).



## **Figure S3. iPSCs Derived from Human Dermal Fibroblasts Are Pluripotent**

Immunofluorescence (IF) images of control (ELMO2-C9) and affected individual (ELMO2-9) iPSCs stained for various markers of pluripotency, along with alkaline phosphatase (AP) staining. Scale bars on IF and AP staining at  $10\times$  magnification represent 200  $\mu$ m. Scale bar on AP staining at 4 x represents 500 µm. AF: Alexa Fluor; DAPI: 4',6-diamidino-2-phenylindole. To generate human iPSCs, retroviral vectors encoding the human cDNAs of KLF4, SOX2, OCT4, and C-MYC (Addgene) were used and cells were grown in KnockOut DMEM containing 20% KnockOut serum replacement, 2 mM L-glutamine, 1% NEAA, 0.1 mM beta-mercaptoethanol, 1% NEAA, 0.1% penicillin and streptomycin, and 4 ng/ml bFGF on irradiated fibroblast feeders. iPSC colonies were manually picked after 3–4 weeks and subsequently adapted to Matrigel-coated (Corning) dishes.



# **Figure S4. The Effect of the c.1065+1G>A Mutation in** *ELMO2* **on the RNA Level**

**(A)** Expression of *ELMO2* detected via qRT-PCR in the non-carrier sibling control (A-II:2), the heterozygous father (A-I:1), and the affected individual (A-II:3), respectively. Briefly, total RNA for qRT-PCR was isolated from primary fibroblasts grown to 70% confluence using TRI Reagent (Sigma-Aldrich). cDNA was generated via reverse transcription using a QuantiTect Reverse Transcription Kit (QIAGEN). Quantitative real-time PCR (qRT-PCR) was performed using Jump Start SYBR Green Mix (Sigma-Aldrich) and the housekeeping gene *ACTB* to normalize gene expression. Data are shown as mean ± SEM, unpaired two-tailed t-test. \*p<0.05; \*\*p<0.01. **(B)** Sanger sequences with amino acid annotations of the 4 aberrantlyspliced transcripts in affected individual A-II:3, as compared to the canonical splice variant.



## **Figure S5. Generation of Zebrafish** *ELMO2* **Allelic Series**

Schematic diagram showing nucleotide deletions on zebrafish genomic loci in the 3 *elmo2* knockout alleles, along with amino acid annotations. Yellow box represents guide RNA 1 target loci.



## **Figure S6. Phylogenetic Analysis of ELMO Protein Family**

A molecular phylogenetic tree showing the relationships between vertebrate and amphioxus ELMO proteins based on neighbour-joining analysis. Phylogenetic analysis of the ELMO family was performed using the NCBI reference sequences of ELMO1, ELMO2, and ELMO3 proteins of the following species: human (*Homo sapiens*; GenBank: NP\_055615, NP\_573403, NP\_078988), mouse (*Mus musculus*; GenBank: NP\_525027, NP\_525026, NP\_766348), chicken (*Gallus gallus*; GenBank: NP\_001026165, XP\_417479, NP\_001243730), frog (*Xenopus tropicalis*; GenBank: XP\_002939462, NP\_001008123, XP\_002931693), zebrafish (*Danio rerio*; GenBank: NP\_998256, NP\_001186987, XP\_009296233), and elephant shark (*Callorhinchus milii*; GenBank: XP\_007887897, XP\_007883848, XP\_007904034).

### **Figure S6. Phylogenetic Analysis of ELMO Protein Family (continued)**

To identify the orthologous proteins of amphioxus and lampreys, the genome sequence databases of amphioxus (*Branchiostoma floridae*, JGI Genome Portal) and Japanese lamprey (*Lethenteron japonicum*, Japanese Lamprey Genome Project) were searched using BLASTP and TBLASTN. The single *Drosophila melanogaster* ced-12 protein (GenBank: NP\_609548) was used as an outgroup for phylogenetic analysis.

The ELMO protein or deduced amino acid sequences from the these species were aligned using ClustalX2. The sequences were then manually trimmed of all sites that were not unambiguously aligned. Phylogenetic analysis of amino acid sequences was performed using the neighbor-joining method implemented in ClustalX2, with outputs displayed using TreeView (Figure S6). Branch lengths are proportional to evolutionary distance corrected for multiple substitutions; the scale bar denotes 0.1 underlying amino acid substitutions per site. Figures on branches indicate robustness of each node, estimated from 1000 bootstrap replicates. Three ELMO proteins of jawed-vertebrates (ELMO1, ELMO2, ELMO3) form a monophyletic group generated by the two gene duplication events at the origin of vertebrates: one that gave rise to the Elmo3 and Elmo1/2 paralogous groups, and a second that gave rise to the Elmo1 and Elmo2 paralogous groups around the time of appearance of "jaw" and dermal bones in vertebrates.



# **Figure S7. Comparison of ELMO2 Proteins in Vertebrate Model Species**

Amino acid sequences from human (*Homo sapiens*, NP\_573403), mouse (*Mus musculus*,NP\_525026), chicken (*Gallus gallus*, XP\_417479), frog (*Xenopus tropicalis*, NP\_001008123) and zebrafish (*Danio rerio*, NP\_001186987) were aligned by ClustalX2. Black shades denote 100% conserved regions (including similarity groups); dark and light grey shades represent 80% and 60% conserved regions respectively. Critical domains interacting with other molecules such as the PH domain (cd13359, red underlines) with DOCK and Tyr-713 with Axl (red arrow) are highly conserved amongst different species .

# **Supplemental Tables**



## **Table S1. Filtering and Prioritization Scheme for Targeted Sequencing**

Mean coverage of exons and percentage of exons covered ≥4 times in the critical interval are shown for each individual. Variants from targeted sequencing for each individual were filtered for variants that had low quality, low coverage, and high frequency in the general population (1000 Genomes, Exome Sequencing Project, and IGBAM in-house exome database). Subsequently, potentially deleterious single nucleotide or small deletion/insertion variants (missense, stop gain/loss, splice site, insertion, deletion, and indel variants) that are homozygous were listed for each individual. Variants in *ELMO2* were common in affected individuals from Families A, B, and C; however, a complex rearrangement involving the first exon of *ELMO2* that was missed by variant calling was detected in Family D, indicated by \*. The 9 bp in-frame deletion in *NCOA3* is indicated by \*\*, which is homozygous in 10 of 498 additional individuals without vascular malformation in the IGBAM in-house database. GMAF: Global minor allele frequency, taken from the 1000 Genome Project; ESPMAF: Exome Sequencing Project minor allele frequency.



## **Table S2. Filtering and Prioritization Scheme for Exome Sequencing**

Mean coverage of the exome and percentage of exons covered ≥4 times are shown for each individual. Variants from exome sequencing for each individual were filtered for variants that had low quality, low coverage, and high frequency in the general population (1000 Genomes, Exome Sequencing Project, and IGBAM in-house exome database). Subsequently, potentially deleterious single nucleotide or small deletion/insertion variants (missense, stop gain/loss, splice site, insertion, deletion, and indel variants) that are consistent with Mendelian inheritance for autosomal recessive disorders and located in the critical region delineated by homozygosity mapping were listed for each individual. Variants in *ELMO2* were common in affected individuals from Families A, B, and C. GMAF: Global Minor Allele Frequency, taken from the 1000 Genome Project; ESPMAF: Exome Sequencing Project minor allele frequency.



# **Table S3. List of Primers**

All primer sequences used for DNA and cDNA sequencing, RT-PCR, and cloning, for both humans and zebrafish.