The American Journal of Human Genetics, Volume *107*

Supplemental Data

NCKAP1 Disruptive Variants Lead to a Neurodevelopmental Disorder with Core Features of Autism

Hui Guo, Qiumeng Zhang, Rujia Dai, Bin Yu, Kendra Hoekzema, Jieqiong Tan, Senwei Tan, Xiangbin Jia, Wendy K. Chung, Rebecca Hernan, Fowzan S. Alkuraya, Ahood Alsulaiman, Mohammad A. Al-Muhaizea, Gaetan Lesca, Linda Pons, Audrey Labalme, Linda Laux, Emily Bryant, Natasha J. Brown, Elena Savva, Samantha Ayres, Dhamidhu Eratne, Hilde Peeters, Frédéric Bilan, Lucile Letienne-Cejudo, Brigitte Gilbert-Dussardier, Inge-Lore Ruiz-Arana, Jenny Meylan Merlini, Alexia Boizot, Lucia Bartoloni, Federico Santoni, Danielle Karlowicz, Marie McDonald, Huidan Wu, Zhengmao Hu, Guodong Chen, Jianjun Ou, Charlotte Brasch-Andersen, Christina R. Fagerberg, Inken Dreyer, Anne chun-hui Tsai, Valerie Slegesky, Rose B. McGee, Brina Daniels, Elizabeth A. Sellars, Lori A. Carpenter, Bradley Schaefer, Maria J. Guillen Sacoto, Amber Begtrup, Rhonda E. Schnur, Sumit Punj, Ingrid M. Wentzensen, Lindsay Rhodes, Qian Pan, Raphael A. Bernier, Chao Chen, Evan E. Eichler, and Kun Xia

Supplementary Figures

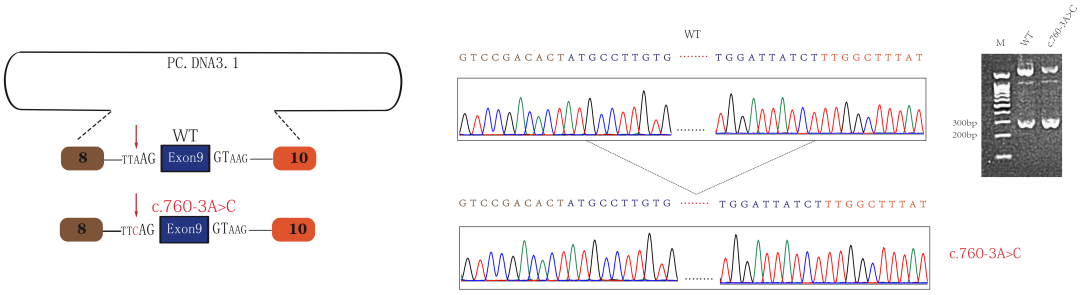

Figure S1 No abnormal splicing was observed in minigene assay experiment for *de novo* intronic variants c.760-3A>C.

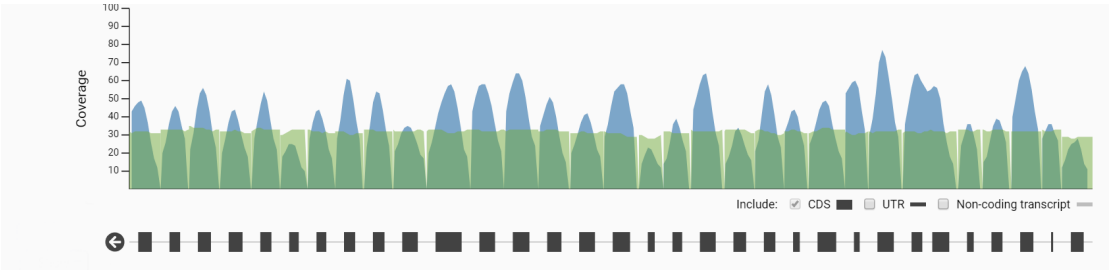

Figure S2 Mean coverage of the coding regions of *NCKAP1* in gnomAD WES/WGS data. Majority of the coding nucleotide sites has a mean coverage over 30×.

Supplementary Methods and Notes

Family 1 and Family 14

Whole-exome sequencing: Details of WES are described elsewhere (PMID: 25363768). Briefly, SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen) reagents were used with a custom barcoding protocol that enabled simultaneous exome enrichment. All exome sequencing was performed using paired-end 100-base pair (bp) reads on Illumina HiSeq 2000 platform. BWA was used to align sequence reads to the hg19 reference genome, and both Picard (http://broadinstitute.github.io/picard/) and GATK were used for marking PCR duplicates, family-based sequence realignment and quality score recalibration. A multinomial model-based family genotyper was used to generate candidate SNV and indels. Candidate variants were subjected to experimental validation. Gene-specific primers were designed for PCR amplification of candidate single nucleotide variables and indels, and amplicons were pooled and sequenced on an Illumina MiSeq. WES Mean Depth of Coverage 89.2x.

No other positive findings on WES besides the *NCKAP1* variants in the probands.

Family 2

Whole-exome sequencing: Details of WES are described elsewhere (PMID: 25363768). In brief, Whole blood-derived genomic DNA was enriched for exonic sequences using SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen) reagents. Samples were barcoded and each pool of four samples was sequenced using 75-bp paired-end reads on single lanes of the Illumina HiSeq 2000 instrument. Reads were aligned to hg19 with BWA, and SAMtools was used for marking PCR duplicates and genotyping. In-house scripts were used for family-based assessment of de novo mutations and annotation against genes and the exome variant server (http://varianttools. sourceforge.net/Annotation/EVS). Primers for candidate variants were designed for PCR amplification of candidate SNVs and indels from all family members, and amplicons were sent for Sanger sequencing. WES Mean Depth of Coverage 89.2x

No other positive findings on WES besides the *NCKAP1* variants in the probands.

Family 3

Whole-exome sequencing (same as family 18 and 19): Using genomic DNA from the proband and parents, the exonic regions and flanking splice junctions of the genome were captured using the IDT xGen Exome Research Panel v1.0. Massively parallel (NextGen) sequencing was done on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19, and analyzed for sequence variants using a custom-developed analysis tool. Additional sequencing technology and variant interpretation protocol has been previously described (Cite Retterer et al., PMID 26633542). The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page (http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/)".

WES Mean Depth of Coverage 92x, 98.6% covered at 10x

No other positive findings on WES besides the *NCKAP1* variant in the proband.

Previous genetic testing for this patient: Clinical Fragile X analysis and an array CGH. Both are normal.

Family 4 and Family 15

Whole-exome sequencing: Details of WES are described elsewhere (PMID: 31130284). Briefly, Ion AmpliSeq Exome RDY Kit (Thermo Fisher, Carlsbad, CA, USA) was used for exome enrichment followed by exome sequencing on Ion Proton instrument (Thermo Fisher, Carlsbad, CA, USA) for an avg. depth: 208, average coverage at 1x: 99.7%, average coverage at 20X: 97.5%.

Other genetic testing: The patients also had CMA that was normal (Affymetrix Cytogenetics Whole-Genome 2.7m Array, analyzed using Affymetrix Chromosome Analysis Suite ChAS version 1.2).

Patient 4 had a dystonia NGS panel (Exons and intron/exon junctions [±5 nucleotides on either side of exon] are screened within the limitations of the NGS protocol. Clinical information supplied determines the extent of capillary sequencing performed) containing the following genes:

ACTB (102630), ADAR (601059), AFG3L2 (604581), AP1S2 (300629), APTX (606350), ARSA (607574), ARX (300382), ATM (607585), ATP13A2 (610513),

ATP1A2 (182340), ATP1A3 (182350), ATP7B (606882), AUH (600529), C19orf12 (614298), CACNA1A (601011), CHMP2B(609512), CP (117700), DCAF17 (612515), DDC (107930), DLAT(608770), DRD5 (126453), EARS2 (612799), ERCC6 (133540), FA2H (611026), FASTKD2 (612322), FBXL4 (605654), FBXO7 (605648), FOXG1 (164874), FOXRED1 (613622), FTL (134790), GAMT (601240), GCDH (608801), GCH1 (600225), GJA1 (600309), HPRT1 (308000), HTT (613004), KCNMA1 (600150), KCNQ2 (602235), L2HGDH (609584), MAPT (157140), MARS2 (609728), MAT1A (610550), MCOLN1 (605248), MMADHC (611935), MPV17 (137960), NPC2 (601015), PANK2 (606157), PARK2 (602544), PDGFB (190040), PDGFRB (173410), PDHX (608769), PINK1 (608309), PLP1 (300401), PNKD (609023), PNPT1 (610316), PRKRA (603424), PRRT2 (614386), PSEN1 (104311), PTEN (601728), PTS (612719), QDPR (612676), RNASEH2A (606034), RNASEH2B (610326), RNASEH2C (610330), SAMHD1 (606754), SCP2 (184755), SDHAF1 (612848), SERAC1 (614725), SGCE (604149), SLC19A3 (606152), SLC20A2 (158378), SLC2A1 (138140), SLC46A1 (611672), SLC6A3 (126455), SPR (182125), SUCLA2 (603921), SUOX (606887), TAF1 (313650), TH (191290), THAP1 (609520), TIMM8A (300356), TOR1A (605204), TPK1 (606370), TREM2 (605086), TREX1(606609), TUBB4A (602662), UBQLN2 (300264), VPS37A (609927), WDR45 (300526), ZNF592 (606937).

This panel detected a likely pathogenic variant in GCDH (NM_000159.4: c.368A>G; p.Tyr123Cys). Although this homozygous variant has never been reported before or seen in any other Saudi patient with glutaric aciduria, it did satisfy PM1, PM2, PP2 and PP3 criteria, so it can be classified as likely pathogenic. This raised the possibility of a dual molecular diagnosis. Unfortunately, however, urine organic acids were requested at least five times and the results, while questionable on two occasions, were completely normal on three separate occasions.

Family 5 and Family 13

Whole-exome sequencing: Both families were studied by trio-based exome sequencing (Medexome Roche, sequenced on a NextSeq500 Illumina)

Patient 5: Mean Depth of Coverage 98.1x, 98.87% covered at 20x

Patient 13 : Mean Depth of Coverage 115.2x, 98.86% covered at 20x

Family 6

Whole-exome sequencing (same as family 11 and 20): Using genomic DNA from the proband and parents (if available), the exonic regions and flanking splice junctions of the genome were captured using the Clinical Research Exome kit (Agilent Technologies, Santa Clara, CA). Massively parallel (NextGen) sequencing was done on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19, and analyzed for sequence variants using a custom-developed analysis tool. Additional sequencing technology and variant interpretation protocol has been previously described (Cite Retterer et al., PMID 26633542). The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page (http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/)"

Mean Depth of Coverage 98x, 96.7% covered at 10x

No other positive findings on WES besides the *NCKAP1* variant in the proband.

Previous genetic testing for this patient: Infantile Epilepsy Gene Panel (GeneDx) with Reflex to Remainder of Comprehensive Epilepsy Gene Panel (GeneDx) 2013: Negative; SLC2A1 Gene Sequencing (Athena Diagnostic Laboratory) 2012: Negative; Chromosome Microarray Blood (ACL Laboratory) 2012: Normal male, 46 XY; Karyotype and Fragile X: Normal.

Family 7

Whole-exome sequencing: WES for the proband was performed by the Australian Genome Research Facility Ltd (AGRF) Agilent Sureselect XT CREv2 kit, Illumina instrument) with a targeted mean coverage of 100x and a minimum 90% of bases sequenced to at least 15x. Data was processed using Cpipe (Sadedin, SP., et al. (2015) Genome Medicine 7:68), in order to generate annotated variant calls within the target region (coding exons +/- 2bp), via alignment to the reference genome (GRCh37). Variants were annotated against all gene transcripts, with reporting of variants against the HGNC recommended transcript (according to HGVS nomenclature). Sample identity has been confirmed using a VCGS SNP-ID panel.

No other positive findings on WES besides the *NCKAP1* variant in the proband.

Previous genetic testing for this patient: Microarray analysis: Array type – Illumina Infinium GSA-24 v1.0; Resolution, 0.20Mb; Assembly hg19/GRCh37 (Feb 2009); Result: no clinically significant genomic imbalance detected. Fragile X PCR Result: Normal allele size, 37 CGG triplet repeats.

Family 8

The *NCKAP1* variant in family 8 were detected using smMIPs by sequening 208 ID/DD/ASD candidate genes (PMID: 28191889). The smMIP sequencing and analysis was performed as previously described (PMID: 28191889). In summary, we targeted the coding portions as well as 5 bp into each exon-adjacent intron to capture variation at splice-donor/acceptor sites. Samples were barcoded and sequenced with an Illumina HiSeq 2000, and data analysis was performed with the MIPgen suite of tools. Variant calling of smMIP data was performed on each sequencing lane with FreeBayes v0.9.14 with default settings and the hg19 reference. Variants with sequencing depth (DP) of >8× and a quality score (QUAL) of >20 were annotated with the Ensembl Variant Effect Predictor tool for GRCh37 and with CADD scores. Candidate variants were validated by Sanger sequencing.

No other positive findings on the targeted genes besides the *NCKAP1* variant in the proband.

Other genetic testing: Illumina Human Omni2.5-8v1 array, negative.

Family 9

Whole-exome sequencing: Using genomic DNA from the proband, the exonic regions and their flanking splice junctions of the genome were captured using Sureselect Human All exon V7 (Agilent Technologies). Massively parallel (NextGen) sequencing was done on an Illumina system (NextSeq550) with 150bp paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19, and analyzed for sequence variants using a custom-developed analysis tool. WES quality parameters obtained from the proband were 127X read depth in average with 91.06% coverage (read depth strictly superior to 20X).

No other positive findings on WES besides the *NCKAP1* variant in the proband.

Other genetic testing: None.

Family 10

Whole-exome sequencing: Blood sample from our patient was analyzed by whole exome sequencing using Twist Human Core Exome Kit (TWIST Biosciences, San Francisco, CA, USA); sequencing was performed on an Illumina HiSeq4000 platform. The platform Varsome Clinical (Saphetor, Lausanne, Switzerland) powered with Sentieon DNAseq v.201911 (Sentieon, San Jose, CA, USA) was used for secondary analysis (mapping, alignment and variant detetection) of exome sequencing data. Sequenced reads were aligned to the GRCh37/hg19 reference human genome with an average coverage of 100x and coverage percentage of RefSeq coding region of 97.8% at 30x. Variants filtering was performed as described in previous studies (PMID: 25044680, PMID: 24389049).

No other putative causative variants in known genes for autism spectrum disorders, intellectual disability or neurological/psychiatric problems were detected.

Other genetic testing: unknown.

Family 11

Whole-exome sequencing (same as family 6 and 20): Using genomic DNA from the proband and parents (if available), the exonic regions and flanking splice junctions of the genome were captured using the Clinical Research Exome kit (Agilent Technologies, Santa Clara, CA). Massively parallel (NextGen) sequencing was done on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19, and analyzed for sequence variants using a custom-developed analysis tool. Additional sequencing technology and variant interpretation protocol has been previously described (Cite Retterer et al., PMID 26633542). The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page (http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/)"

WES Mean Depth of Coverage 98x, 96.7% covered at 10x

No other positive findings on WES besides the *NCKAP1* variant in the proband.

Other genetic testing: The patient also had a karyotype, array, and Prader-Willi/Angelman syndrome testing. The array and PWS/AS testing were normal. The karyotype revealed a de novo inversion [46,XX,inv(2)(p23q37.3)dn] with uncertain clinical significance.

Family 12

The *NCKAP1* variant in family 12 were detected using smMIPs. The smMIP sequencing and analysis was performed as previously described (PMID: 27824329). In summary, smMIPs were designed using MIPgen with an updated scoring algorithm. After amplification, libraries were sequenced using the Illumina HiSeq2000 platform. Sequences were aligned against GRCh37 using BWA-MEM (v.0.7.13) after removing incorrect read pairs and low-quality reads. SNVs/indels were called with FreeBayes (v.0.9.14). Variants exceeding tenfold sequence coverage and read quality over 20 (QUAL > 20) were annotated with ANNOVAR using reference GRCh37. Candidate variants were selected for validation using Sanger sequencing.

Other genetic testing: no.

Family 16

Array-CGH: Agilent 400K oligo array.

Result:

arr[hg19] 2q32.1(183,762,482-184,182,761)x1,Xp21.1(33,000,333-33,047,374)x1

The variant on 2q32 is *de novo* and the one with the *NCKAP1*-variant, the deletion on Xq21.1 is maternally inherited and includes exons 2 of the DMD gene, it is considered likely pathogenic. It has not been found in males in the family, the patient and her mother attends cardiological controls.

Other genetic testing: None.

Family 17

The proband in Family 17 underwent a karyotype that identified an apparently balanced inversion in chromosome 2. Parental studies confirmed *de novo* status of the inversion. SNP-chromosomal microarray, FMR1 CGG repeat studies, and expanded neurodevelopmental panel were all normal. At this point in the workup, the clinical team decided upon mate pair sequencing to better characterize the inversion rather than whole exome sequencing. Mate pair sequencing is a clinically available junction breakpoint sequencing test available through Mayo Clinic Laboratories. Exact details of methodology are not available to the clinical team, who is not affiliated with Mayo.

The patient has not had whole-exome sequencing. The expanded ASD/ID panel is run on an exome platform and was normal. All other genetic workup included normal SNP-chromosomal microarray, normal FMR1 CGG repeat studies.

Family 18

Whole-exome sequencing (same as family 3 and 19): Using genomic DNA from the proband and parents, the exonic regions and flanking splice junctions of the genome were captured using the IDT xGen Exome Research Panel v1.0. Massively parallel (NextGen) sequencing was done on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19, and analyzed for sequence variants using a custom-developed analysis tool. Additional sequencing technology and variant interpretation protocol has been previously described (Cite Retterer et al., PMID 26633542). The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page (http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/)"

Mean Depth of Coverage 72x, 98.6% covered at 10x

WES identified the *de novo* VUS in *NKCAP1* and paternal likely pathogenic SCN5A variant.

Previous genetic testing: Patient 18 had prior *RAG1* and *RAG2* panel testing which was homozygous for p.A444V mutation, consistent with his personal and family history of severe combined immunodeficiency (SCID). For his genetics evaluation with our team, he had a chromosome microarray (118k oligo probes and 66k SNP probes)

which was negative for copy number abnormality and positive for significant absence of heterozygosity (163 Mb, 5.66%) consistent with known parental consanguinity. Additional testing performed by the chromosome microarray testing lab ruled out chromosome 15 imprinting and copy number alterations associated with Prader-Willi and Angelman syndromes.

Family 19

Whole-exome sequencing (same as family 3 and 18): Using genomic DNA from the proband and parents, the exonic regions and flanking splice junctions of the genome were captured using the IDT xGen Exome Research Panel v1.0. Massively parallel (NextGen) sequencing was done on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19, and analyzed for sequence variants using a custom-developed analysis tool. Additional sequencing technology and variant interpretation protocol has been previously described (Cite Retterer et al., PMID 26633542). The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page (http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/)".

WES Mean Depth of Coverage 128x, 98.6% covered at 10x

Previous genetic testing: None.

Family 20

Whole-exome sequencing (same as family 6 and 11): Using genomic DNA from the proband and parents (if available), the exonic regions and flanking splice junctions of the genome were captured using the Clinical Research Exome kit (Agilent Technologies, Santa Clara, CA). Massively parallel (NextGen) sequencing was done on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19, and analyzed for sequence variants using a custom-developed analysis tool. Additional sequencing technology and variant interpretation protocol has been previously described (Cite Retterer et al., PMID 26633542). The general assertion criteria for variant classification are publicly

available on the GeneDx ClinVar submission page (http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/)"

WES Mean Depth of Coverage 124x, 96.8% covered at 10x

Besides variant in *NCKAP1*, WES identified a variant of uncertain significance in NLGN4X, c.2321T>C, maternally inherited, a variant of uncertain significance in DNM2, c.992A>G, maternally inherited, and a variant of uncertain significance in HERC2, c.2488G>A that was paternally inherited.

Previous genetic testing for this patient included a chromosomal microarray that utilized the Agilent 4-plex CGH+SNP oligonucleotide array with 108K probes and 60K SNP probes. This identified a duplication of 7q21.3, approximately 375.67kb in size and this included 4 genes (CASD1, SGCE, PEG10, and PPP1R9A). This is classified as a variant of uncertain significance. This patient also had Fragile X syndrome which was normal and a newborn hypotonia panel that included myotonic dystrophy type 1 triple nucleotide repeat analysis, methylation of the Prader Willi/Angelman syndrome region, and spinal muscular atrophy deletion analysis.