The American Journal of Human Genetics, Volume *100*

Supplemental Data

Mutations in *NKX6-2* Cause Progressive Spastic Ataxia and Hypomyelination

Viorica Chelban, Nisha Patel, Jana Vandrovcova, M. Natalia Zanetti, David S. Lynch, Mina Ryten, Juan A. Botía, Oscar Bello, Eloise Tribollet, Stephanie Efthymiou, Indran Davagnanam, SYNAPSE Study Group, Fahad A. Bashiri, Nicholas W. Wood, James E. Rothman, Fowzan S. Alkuraya, and Henry Houlden

Figure S1

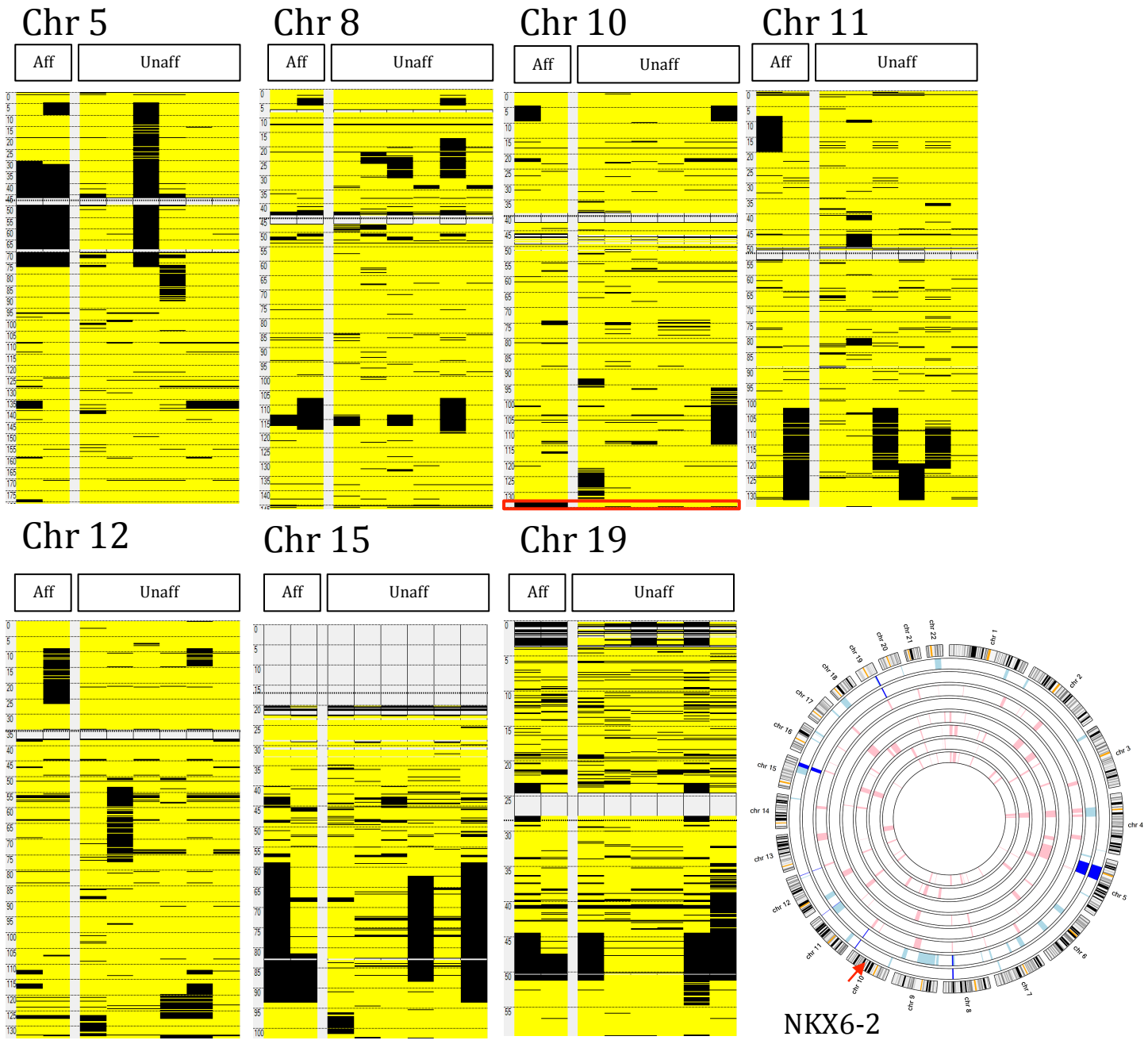
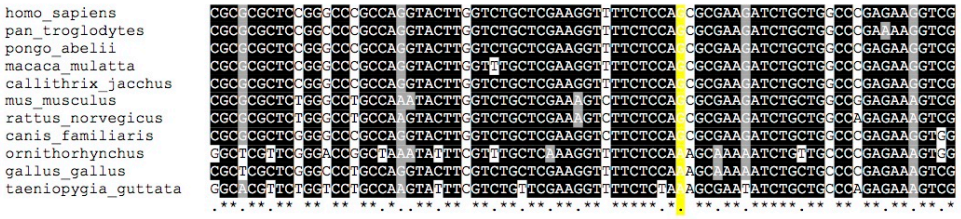

Chr 5
Aff
Unaff
Chr 8
Aff
Unaff
Chr 10
Aff
Unaff
Chr 11
Aff
Unaff
Chr 12
Aff
Unaff
Chr 15
Aff
Unaff
Chr 19
Aff
Unaff
NKX6-2

Figure S2

Sequence conservation based multiple alignments on 37 Eutherian mammals provided by Ensembl[1]. The figure shows a constrained region surrounding the p.Leu163Val mutation in NKX6-2 homeodomain.



Supplementary Methods

Whole Exome sequencing

Family 1 and 2

In family 1 exome Exome sequencing and alignments were carried out at Macrogen (Macrogen Inc.Seoul, South Korea). Exome capture was performed using SureSelect Human All Exon V4 enrichment kit (Agilent Technologies, Santa Clara, CA, USA) and sequenced on an Illumina HiSeq4000 instrument (Illumina, San Diego, CA,USA) generating 100bp paired-end reads. Reads were aligned using BWA [2] to the hg19 human genome reference. Base quality recalibration, realignment and variant calling was done using GATK HaplotypeCaller-based pipeline [3-6]. Called variants were annotated in-house using Annovar [7] and filtered for rare homozygous and compound heterozygous variants using custom R scripts. Only variants with MAF of < 0.01 in 1000 Genomes Project [8], NHLBI GO Exome Sequencing [9], and ExAC [10], were included. A mean coverage of 80x was obtained with 93.4% of based covered at more than 20x. Similarly, in family 2 Exome sequencing was performed using SureSelect Human All Exon enrichment kit (Agilent Technologies, Santa Clara, CA, USA). Reads were aligned using Novoalign (Novocraft, Selangor, Malaysia) to hg19 genome reference and processed with Dindel[11]. In family 3, Exome capture was performed on the index using TruSeq Exome Enrichment kit (Illumina) following the manufacturer's protocol. Sample was prepared as an Illumina sequencing library, and in the second step, the sequencing libraries were enriched for the desired target using the Illumina Exome Enrichment protocol. The captured libraries were sequenced using Illumina HiSeq 2000 Sequencer. The reads were mapped against UCSC hg19 (http://genome.ucsc.edu/) by BWA (http://bio-bwa.sourceforge.net/). The SNPs and Indels were detected by

SAMTOOLS (http://samtools.sourceforge.net/). Variants from WES were filtered such that only novel (or very low frequency <0.1%), coding/splicing, homozygous variants that are within the autozygome of the affected individual and are predicted to be pathogenic were considered as likely causal variants [12; 13].

Haplotype analysis

Samples from family 1 and 2 were re-aligned using Novoalign (Novocraft, Selangor, Malaysia) to the hg19 genomic reference and variants were called using GATK HaplotypeCaller-based pipeline[3-5]. Shared regions of homozygosity were identified using HomozygosityMapper [14]

Comparative modeling

The Ensembl database [1] was used to identify homeodomain region with NKX6-2 and to obtain multispecies alignments. SWISS-MODEL [15] was used to identify known structures with a high sequence similarity to NKX6-2 and for PDB model generation. 3a01.2.A PDB model was selected as a template showing the highest sequence similarity (56.25%). DUET [16] was used to predict c.C487G mutation effect on protein stability.

Western blotting

Fibroblast samples from affected individuals and healthy control were obtained with consents. After reaching 90% confluence, the cultured cells were harvested and lysed in ice-cold lysis RIPA buffer (50 mM Tris-HCl (pH8.0), 150 mM NaCl, 0.5% DOC, 0.1% Triton x100, 0.1% Sodium Dodecyl Sulfate and SIGMAFAST Protease Inhibitor Cocktail) for 30 minutes. After that the lysate was centrifuged at 16000g at 4°C. The supernatant was transferred to a new tube and the protein concentration was measured using BCA protein assay kit (Pierce) according to manufacturer's protocol.

Equal amounts of proteins (12 ug per lane) from affected individuals and healthy controls were separated by SDS/PAGE using Bis-Tris gradient gels (4–12% NuPAGE; Invitrogen) and transferred by electrophoresis onto Immobilon-P transfer membranes (Millipore). After being blocked with fat-free milk, membranes were immunoblotted with the respective rabbit anti-NKX 6-2 (Abcam ab179532; 1:1000) antibodies at 4 °C overnight. Blots were then exposed to horseradish peroxidase-conjugated goat anti-rabbit IgG (17210, Bio-Rad Laboratories, 1:5,000) for 1h at room temperature. Blots were developed using ECL-Prime (GE Healthcare), visualized via a ChemiDoc Touch Imaging System, and analysed using Image Lab 5.2 software (Bio-Rad Laboratories). For the quantifications, the signal intensity of each of the Kv1 bands was normalized to the signal intensity of the corresponding reversible Ponceau staining as a reliable loading control[17]. The NKX6-2 level were expressed as a percentage of the control sample.

Gene Co-expression Network Analyses

We generated a Gene Co-expression Network (GCN) using expression data generated by the UK Brain Expression Consortium[18; 19] and assayed with the Affymetrix Exon 1.0 ST Array in intralobular white matter originating from 83 neuropathologically-normal individuals. In total, 19152 transcripts were used by the weighted gene co-expression network analysis (WGCNA) R package[20]. In brief, after outlier removal, a "signed" GCN was constructed by creating a signed Topological Overlap Measure (TOM) matrix based on Pearson correlation. Gene modules were created by hierarchical clustering based on a 1-TOM dissimilarity matrix. The results of the initial hierarchical clustering were post-processed using the k-means clustering search method with 30 iterations.

After this, gene modules were functionally annotated with gProfileR[21] R package using Gene Ontology (GO) database without Electronic Inferred Annotations (EIA) and accounting for multiple testing with gSCS. We consider all associations with corrected P-values of less than 0.05 as significant.

Given that *NKX6-2* is known to be a transcription factor we also used an Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context[22] (ARACNe-AP) to identify its most probable target genes in human white matter. The algorithm was used in its bootstrapping mode with 100 bootstraps. The resulting regulon (consisting of the predicted target genes) was functionally annotated with gProfileR R package using Gene Ontology (GO) database without Electronic Inferred Annotations (EIA) and accounting for multiple testing with gSCS.

Identification of phenotypically similar known Mendelian disorders

We used the web-resource Phenomizer (http://compbio.charite.de/phenomizer/ accessed on 10th. March 2017) to identify known Mendelian disorders with a high degree of phenotypic similarity to the affected individuals with homozygous deleterious variants in *NKX6-2*. This analysis was performed using a set of curated Human Phenotype Ontology terms, which were representative of the core clinical features of this patient group. The terms used were as follows: Abnormality of the cerebral white matter (HP:0002500); Brisk reflexes (HP:0001348); Dystonia (HP:0001332); CNS hypomyelination (HP:0003429); Cerebellar ataxia associated with quadrupedal gait (HP:0009878); Cerebellar atrophy (HP:0001272); Lower limb spasticity (HP:0002061) and Nystagmus (HP:0000639). Semantic similarity scoring was performed with no assumptions made regarding the mode of inheritance and using the Resnik (not symmetric) method. All Mendelian disorders with a corrected p-value of <0.001 (Benjamini Hochberg procedure) were retained.

SYNAPSE Study Group

A team of world-class researchers working towards a better understanding of the genetic basis in neurological disorders

Supported by the Welcome Trust

Collaborators

Prof Yamna Kriouile

Affiliation: Unit of Neuropediatrics, Children's Hospital of Rabat, University of Rabat, Rabat 6527, Morocco

Email: dr.kriouile@gmail.com

Prof. Mohamed El Khorassani

Affiliation: Children's Hospital of Rabat, University of Rabat, Rabat 6527, Morocco

Email: elkhorassani.mohamed@gmail.com

Prof. Mhammed Aguennouz

Affiliation: Department of Clinical and Experimental Medicine, University of Messina, Messina 98123, Italy

Email: aguenoz@unime.it

Prof Stanislav Groppa

Affiliation: Department of Neurology and Neurosurgery, Institute of Emergency Medicine, Chisinau, Republic of Moldova.

Email: sgroppa@gmail.com

Dr. Blagovesta Marinova Karashova

Affiliation: Department of Paediatrics, Medical University of Sofia, Sofia 1431, Bulgaria
Email: blagovestakarashova@gmail.com

Prof Lionel Van Maldergem

Affiliation: Centre of Human Genetics, University Hospital Liege, Liege 4000, Belgium

Email: lvanmaldergem@chu-besancon.fr

Dr. Wolfgang Nachbauer

Affiliation: Department of Neurology, Medical University Innsbruck, Anichstrasse 35, Innsbruck 6020, Austria

Email: Wolfgang.Nachbauer@i-med.ac.at

Prof. Sylvia Boesch

Affiliation: Department of Neurology, Medical University Innsbruck, Anichstrasse 35, Innsbruck 6020, Austria

Email: sylvia.boesch@i-med.ac.at

Dr. Larissa Arning

Affiliation: Department of Human Genetics, Ruhr-University Bochum, Bochum 44801, Germany

Email: Larissa.Arning@ruhr-uni-bochum.de

Prof. Dagmar Timmann

Affiliation: Braun Neurologische Universitätsklinik Universität *Essen,* Hufelandstr 55, *Essen* D-45122, Germany

Email: Dagmar.Timmann-Braun@uni-duisburg-essen.de

Prof. Bru Cormand

Affiliation: Department of Genetics, Universitat de Barcelona, Barcelona 08007, Spain

bcormand@ub.edu

Dr. Belen Pérez-Dueñas

Affiliation: Hospital Sant Joan de Deu, Esplugues de Llobregat 08950, Barcelona, Spain
bperez@sjdhospitalbarcelona.org

Dr Gabriella Di Rosa, MD, PhD

Affiliation: Department of Pediatrics, University of Messina, Messina 98123, Italy

Email: gdirosa@unime.it

Prof. Jatinder S. Goraya, MD, FRCP
Affiliation: Division of Paediatric Neurology, Dayanand Medical College & Hospital, Ludhiana, Punjab 141001, India

Email: gorayajs@gmail.com

Prof. Tipu Sultan

Affiliation: Division of Paediatric Neurology, Children's Hospital of Lahore, Lahore 381-D/2, Pakistan

Email: tipusultanmalik@hotmail.com

Prof Jun Mine

Affiliation: Department of Paediatrics, Shimane University, Faculty of Medicine, Izumo, 693-8501, Japan

Email: jmine@med.shimane-u.ac.jp

Prof. Daniela Avdjieva,

Affiliation: Department of Paediatrics, Medical University of Sofia, Sofia 1431, Bulgaria

Email: davdjieva@yahoo.com

Dr. Hadil Kathom,

Affiliation: Department of Pediatrics, Medical University of Sofia, Sofia 1431, Bulgaria

Email: hadilmk@gmail.com

Prof.Dr Radka Tincheva

Affiliation: Head of Department of Clinical Genetics, University Pediatric Hospital, Sofia 1431, Bulgaria

Email: radka.tincheva@gmail.com

Prof. Selina Banu

Affiliation: Neurosciences Unit, Institute of Child Health and Shishu Shastho Foundation Hospital, Mirpur, Dhaka 1216, Bangladesh

Email: selinabanu17@gmail.com

Prof. Mercedes Pineda-Marfa

Affiliation Servei de Neurologia Pediàtrica, l'Hospital Universitari Vall d'Hebron, Barcelona 08035, Spain

Email: pineda@hsjdbcn.org

Prof. Pierangelo Veggiotti

Affiliation: Unit of Infantile Neuropsychiatry Fondazione

Istituto Neurologico "C. Mondino" IRCCS, Via Mondino 2, Pavia 27100, Italy

Email: pierangelo.veggiotti@unipv.it

Prof. Michel D. Ferrari

Affiliation: Leiden University Medical Center, Albinusdreef 2, Leiden 2333, Netherlands

Email: M.D.Ferrari@lumc.nl

Prof Arn M J M van den Maagdenberg

Affiliation: Leiden University Medical Center, Albinusdreef 2, Leiden 2333, Netherlands

A.M.J.M.van_den_Maagdenberg@lumc.nl

Prof. Alberto Verrotti

Affiliation: University of L'Aquila, L'Aquila, Italy

Email: verrottidipianella@univaq.it

Prof Giangluigi Marseglia

Affiliation: Department of Pediatrics, University of Pavia, IRCCS Policlinico "San Matteo", Pavia 27100, Italy

Email: gl.marseglia@smatteo.pv.it

Dr. Salvatore Savasta

Affiliation: Division of Pediatric Neurology, Department of Pediatrics, University of Pavia, IRCCS Policlinico "San Matteo", Pavia 27100, Italy

Email: S.Savasta@smatteo.pv.it

Dr. Mayte García-Silva

Affiliation: Hospital Universitario 12 de Octubre, Madrid 28041, Spain

Email: mgarciasilva@salud.madrid.org

Dr. Alfons Macaya Ruiz

Affiliation: University Hospital Vall d'Hebron, Barcelona 08035, Spain

Email: amacaya@vhebron.net

Prof. Barbara Garavaglia

Affiliation: IRCCS Foundation, Neurological Institute “Carlo Besta”, Molecular Neurogenetics, 20126 Milan, Italy

Email: segr.neurogenetica@istituto-besta.it

Dr. Eugenia Borgione

Affiliation: Laboratorio di Neuropatologia Clinica, U.O.S. Malattie, Neuromuscolari Associazione OASI Maria SS. ONLUS – IRCCS, Via Conte Ruggero 73, 94018 Troina, Italy

Email: eborgione@oasi.en.it

Dr. Simona Portaro

Affiliation: IRCCS Centro Neurolesi "Bonino Pulejo", SS113, c.da Casazza, 98124 Messina, Italy

Email: simonaportaro@hotmail.it

Dr. Benigno Monteagudo Sanchez

Affiliation: Hospital Arquitecto Marcide, Avenida de la Residencia S/N, Ferrol (A Coruña), 15401 Spain

Email: benims@hotmail.com

Dr. Richard Boles

Affiliation: Courtagen Life Sciences, 12 Gill Street Suite 3700, Woburn, MA 01801 USA

Email: Richard.Boles@courtagen.com

Prof. Savvas Papacostas

Affiliation: Neurology Clinic B, The Cyprus Institute of Neurology and Genetics, 6 International Airport Road, 1683 Nicosia, Cyprus

Email: savvas@cing.ac.cy

Dr. Michail Vikelis

Affiliation: Iatreio Kefalalgias Glyfadas, 8 Lazaraki str, 3rd floor, 16675, Athens, Greece

Email: mvikelis@headaches.gr

Prof James Rothman

Affiliation: Department of Cell Biology, Yale School of Medicine, New Haven, CT

Email: jrothman77@yahoo.com

Dr Paola Giunti

Affiliation: Department of Molecular Neuroscience, University College London, London, UK

Email: p.giunti@ucl.ac.uk

Prof Henry Houlden

Affiliation: Department of Molecular Neuroscience, University College London, London, UK

Email: h.houlden@ucl.ac.uk

Dr. Vincenzo Salpietro Damiano

Affiliation: Department of Molecular Neuroscience, University College London, London, UK

Email: v.salpietro@ucl.ac.uk

Dr Emer Oconnor

Affiliation: Department of Molecular Neuroscience, University College London, London, UK

Email: e.oconnor@ucl.ac.uk

Dr Viorica Chelban

Affiliation: Department of Molecular Neuroscience, University College London, London, UK

Email: v.chelban@ucl.ac.uk

Stephanie Efthymiou

Affiliation: Department of Molecular Neuroscience, University College London, London, UK

Email:s.efthymiou@ucl.ac.uk

Prof Dimitri Kullmann

Affiliation: University College London, London, UK

Email: d.kullmann@ucl.ac.uk

SUPPLEMENTARY MATERIALS LEGEND

Table S1. Haplotype analysis data in family 1 and 2.

Figure S1. Homozygozity mapping in family 3 data.

Figure S2. Sequence conservation of the homeobox domain.

Supplementary Methods

Synapse Study Group Collaborators

SUPPLEMENTAL REFERENCES

1. Aken, B.L., Achuthan, P., Akanni, W., Amode, M.R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., et al. (2017). Ensembl 2017. Nucleic Acids Res 45, D635-D642.
2. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589-595.
3. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303.
4. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43, 11 10 11-33.
5. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43, 491-498.
6. Danecek, P., and McCarthy, S.A. (2017). BCFtools/csq: Haplotype-aware variant consequences. Bioinformatics (Oxford, England).
7. Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc 10, 1556-1566.
8. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature 526, 68-74.
9. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. In. (
10. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.
11. Albers, C.A., Lunter, G., MacArthur, D.G., McVean, G., Ouwehand, W.H., and Durbin, R. (2011). Dindel: accurate indel calls from short-read data. Genome Res 21, 961-973.
12. Alkuraya, F.S. (2013). The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. Human genetics 132, 1197-1211.
13. Alkuraya, F.S. (2016). Discovery of mutations for Mendelian disorders. Human Genetics, 1-9.
14. Seelow, D., Schuelke, M., Hildebrandt, F., and Nurnberg, P. (2009). HomozygosityMapper--an interactive approach to homozygosity mapping. Nucleic Acids Res 37, W593-599.
15. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42, W252-258.
16. Pires, D.E., Ascher, D.B., and Blundell, T.L. (2014). DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res 42, W314-319.

17. Romero-Calvo, I., Ocon, B., Martinez-Moya, P., Suarez, M.D., Zarzuelo, A., Martinez-Augustin, O., and de Medina, F.S. (2010). Reversible Ponceau staining as a loading control alternative to actin in Western blots. Analytical biochemistry 401, 318-320.
18. Forabosco, P., Ramasamy, A., Trabzuni, D., Walker, R., Smith, C., Bras, J., Levine, A.P., Hardy, J., Pocock, J.M., Guerreiro, R., et al. (2013). Insights into TREM2 biology by network analysis of human brain gene expression data. Neurobiol Aging 34, 2699-2714.
19. Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., Coin, L., de Silva, R., Cookson, M.R., et al. (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. Nature neuroscience 17, 1418-1428.
20. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.
21. Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res 35, W193-200.
22. Lachmann, A., Giorgi, F.M., Lopez, G., and Califano, A. (2016). ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinformatics 32, 2233-2235.