The American Journal of Human Genetics, Volume *104*

Supplemental Data

Missense Variants in the Histone Acetyltransferase Complex Component Gene *TRRAP* Cause Autism and Syndromic Intellectual Disability

Benjamin Cogné, Sophie Ehresmann, Eliane Beauregard-Lacroix, Justine Rousseau, Thomas Besnard, Thomas Garcia, Slavé Petrovski, Shiri Avni, Kirsty McWalter, Patrick R. Blackburn, Stephan J. Sanders, Kévin Uguen, Jacqueline Harris, Julie S. Cohen, Moira Blyth, Anna Lehman, Jonathan Berg, Mindy H. Li, Usha Kini, Shelagh Joss, Charlotte von der Lippe, Christopher T. Gordon, Jennifer B. Humberson, Laurie Robak, Daryl A. Scott, Vernon R. Sutton, Cara M. Skraban, Jennifer J. Johnston, Annapurna Poduri, Magnus Nordenskjöld, Vandana Shashi, Erica H. Gerkes, Ernie M.H.F. Bongers, Christian Gilissen, Yuri A. Zarate, Malin Kvarnung, Kevin P. Lally, Peggy A. Kulch, Brina Daniels, Andres Hernandez-Garcia, Nicholas Stong, Julie McGaughran, Kyle Retterer, Kristian Tveten, Jennifer Sullivan, Madeleine R. Geisheker, Asbjorg Stray-Pedersen, Jennifer M. Tarpinian, Eric W. Klee, Julie C. Sapp, Jacob Zyskind, Øystein L. Holla, Emma Bedoukian, Francesca Filippini, Anne Guimier, Arnaud Picard, Øyvind L. Busk, Jaya Punetha, Rolph Pfundt, Anna Lindstrand, Ann Nordgren, Fayth Kalb, Megha Desai, Ashley Harmon Ebanks, Shalini N. Jhangiani, Tammie Dewan, Zeynep H. Coban Akdemir, Aida Telegrafi, Elaine H. Zackai, Amber Begtrup, Xiaofei Song, Annick Toutain, Ingrid M. Wentzensen, Sylvie Odent, Dominique Bonneau, Xénia Latypova, Wallid Deb, CAUSES Study, Sylvia Redon, Frédéric Bilan, Marine Legendre, Caitlin Troyer, Kerri Whitlock, Oana Caluseriu, Marine I. Murphree, Pavel N. Pichurin, Katherine Agre, Ralitza Gavrilova, Tuula Rinne, Meredith Park, Catherine Shain, Erin L. Heinzen, Rui Xiao, Jeanne Amiel, Stanislas Lyonnet, Bertrand Isidor, Leslie G. Biesecker, Dan Lowenstein, Jennifer E. Posey, Anne-Sophie Denommé-Pichon, Deciphering Developmental Disorders study, Claude Férec, Xiang-Jiao Yang, Jill A. Rosenfeld, Brigitte Gilbert-Dussardier, Séverine Audebert-Bellanger, Richard Redon, Holly A.F. Stessman, Christoffer Nellaker, Yaping Yang, James R. Lupski, David B. Goldstein, Evan E. Eichler, Francois Bolduc, Stéphane Bézieau, Sébastien Küry, and Philippe M. Campeau

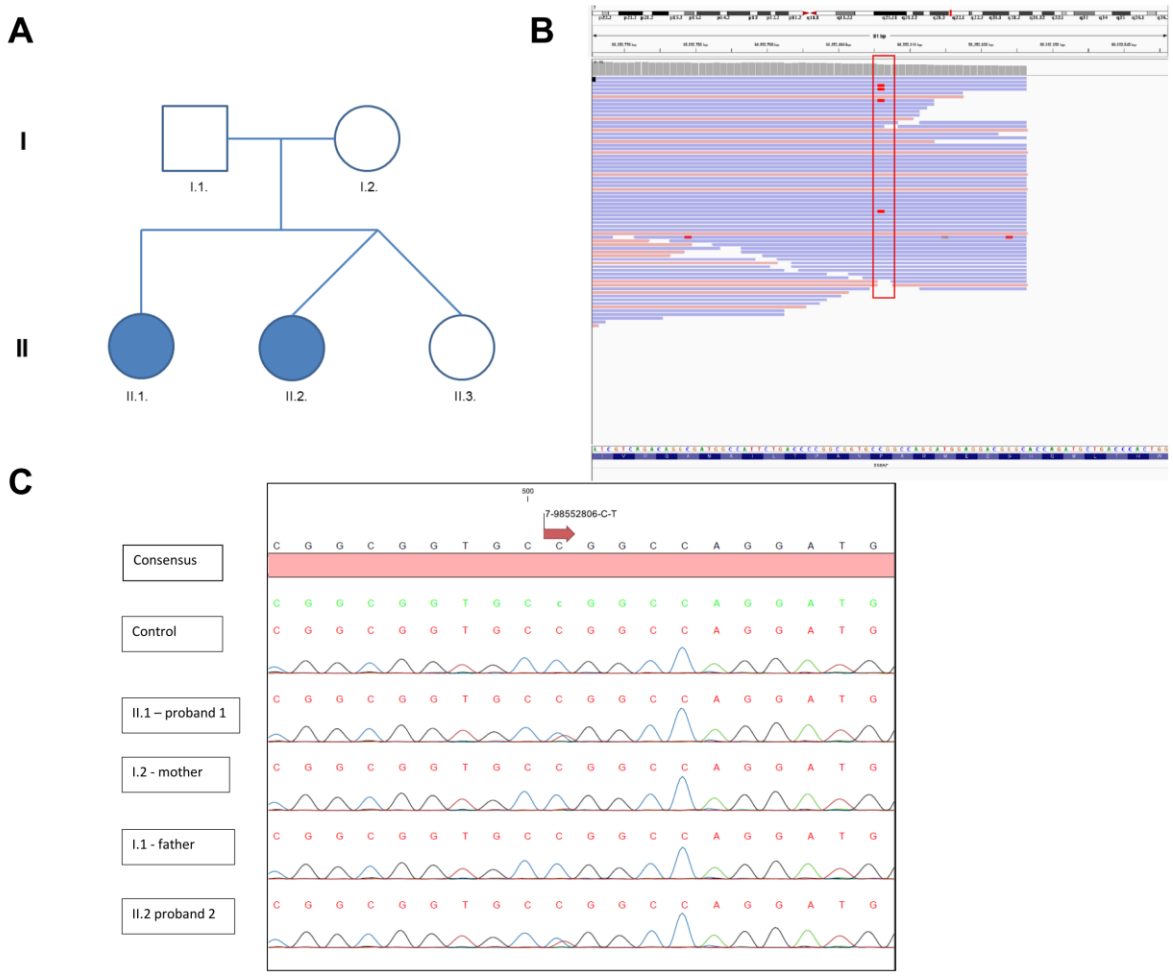

Figure S1. Maternal mosaicism of variant c.5795C>T (p.Pro1932Leu) identified in individuals 22 and 23. (A) Pedigree of the family. (B) Integrative Genome Viewer snapshot of BAM emitted by GATK haplotype caller in the mother (individual I.2.). Reads are shown aligned to the genome in blue for forward strand and red for reverse strand. The C to T variant is shown in red on 4/55 reads. This low 7% representation is significantly deviated from the expected 50% and is indicative of mosaicism. (C) Segregation analysis of the variant by Sanger sequencing confirms the heterozygous variant in the siblings and the mosaicism in the mother.

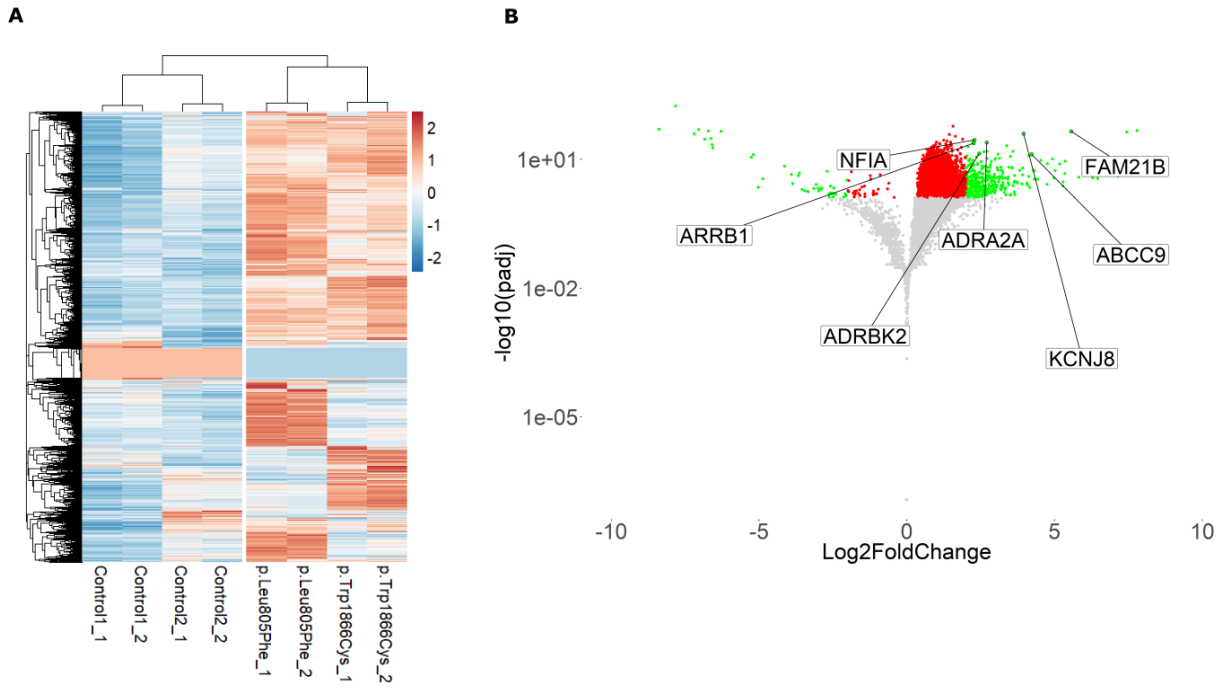

Figure S2. Fibroblasts harboring the p.(Leu805Phe) or p.(Trp1866Cys) variants have differential gene expression patterns.

RNAseq was performed on two healthy controls and two individual fibroblasts to assay gene expression with technical duplicates. (A) Heatmap of genes expressed with at least 10 counts in at least one condition. Normalized count values from "DESeq2" were plotted and scaled row-wise using the "pheatmap" R package. (B) Log_2 Fold Change (Log_2FC) and p-values adjusted for 10% False Discovery Rate (padj) were calculated for the two affected individuals and two controls respectively pooled together using the "DESeq2" R package then plotted using the "ggplot2" R package as a volcano plot. Green dots represent genes with a padj lower than 0.01, and an absolute Log_2FC higher than 2 (padj < 0.01, abs(Log_2FC) > 2) and are referred to as Differentially Expressed Genes (DEGs). Red dots represent genes with a padj lower than 0.01 but an absolute Log_2FC lower than 2 (padj < 0.01, abs(Log_2FC) < 2).

Supplementary methods

RNAseq

Human primary fibroblasts for patient were cultured in DMEM (ThermoFisher cat# 11995-065), 10% FBS, 1mM GlutaMax (ThermoFisher cat# 35050-061) and antibiotics-antimycotics (ThermoFisher cat# 15240-062). Fibroblasts were plated at 1 million cells per 150 mm dish and allowed to grow until they reached 80% confluency. Cells were washed twice with D-PBS, resuspended in QIAzol (Qiagen cat# 79306) and stored at -80°C until all samples were ready for RNA extraction. RNA isolation was performed using the RNeasy mini kit (Qiagen cat# 74104), according to the manufacturer's protocol. Samples were treated with the Turbo DNA free kit (ThermoFisher cat# AM1907) and quality was assessed using the Agilent 2100 Bioanalyzer. Sequencing was performed at the CHU Sainte-Justine and Génome Québec Integrated Clinical Genomic Centre in Pediatrics (CIGCP). mRNA Libraries were prepared using the TruSeq Stranded mRNA kit for 48 samples (Illumina), according to the manufacturer's instructions. Samples were run on the Illumina HiSeq 4000 PE100 with 7 samples per lane. Output files were analyzed using the MUGQIC RNAseq pipeline (MUGQIC) steps 1 through 14 on the Guillimin Génome Québec HPC. In summary, BAM files were converted to FASTQ using Picard (BROAD Institute), sequences were trimmed using Trimmomatic (Bolger, 2014) then aligned to the GRCh37 genome using STAR (Dobin, 2013); duplicate and misaligned reads were discarded using Picard, and raw counts were called using HTseq (Anders S, 2015). Differential expression analysis was performed using the DESeq2 R package (Love, 2014), with default parameters. GO annotation analysis was performed using the GOrilla web application (Eden, 2009). Control and patient cell lines of the same type were respectively pooled together into two groups for differential gene expression analysis, so as to find genes which were significantly differentially expressed in all patients compared to all controls. Significantly expressed genes were selected with an adjusted p value (10% False Discovery Rate, padj)

lower than 0.01, and a log2 Fold Change (log2FC) higher than 2 or lower than -2; corresponding to an overall fold change of at least 4 or -4. Genes of interest were all significantly differentially expressed in individual patient analyses with DESeq2 compared to controls.

qPCR

Total RNA from human cell lines was extracted as described above. Affected individuals were compared to 5 controls for ABCC9 and 6 controls for KCNJ8. Among the controls used for ABCC9and KCNJ8, the following cell lines were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research: AGO8498A; GM00041 and GM07532. Reverse transcription was performed with 1 µg- of total RNA in a 20 µl reaction volume using qScript cDNA SuperMix (QuantaBio; CA101414-102) according to the manufacturer's specifications. For Real-Time PCR, cDNA samples were diluted 2-fold (*HMBS* and *KCNJ8*) or 35-fold (*GAPDH*, *ABCC9*), respectively. Gene-specific transcription level was measured in triplicates using the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific ; A25741) and the LightCycler® 96 System from Roche Life Science). Primers (amplified sequences 85 to 200 bp) were designed using PrimerBank (https://pga.mgh.harvard.edu/primerbank/) and Primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). *KCNJ8* (FW : GCTCTTCGCTATCATGTGGT; REV : GAAGAGAAAAGCAGAAGTGAAAGAC E=2.07), *ABCC9* (FW: TCGCTTCCTTTTGAGTCCTG; REV: ATGTCCTCTGTTTCTGCGG E= 2.15), and controls: *HMBS* (FW : GGCAATGCGGCTGCAA; REV : GGGTACCCACGCGAATCAC E= 1.93), *GABDH* (FW : AGCCACATCGCTCAGACA; GCCCAATACGACCAAATCC E=1.92). Melting curves for each primer sets showed a single symmetrical amplicon, and no primer-dimer peaks were observed in the no-reverse transcription-control (NoRT) reactions. PCR efficiency for each set of primers was calculated based on the slope of the standard curve from a 5-fold serial dilution and taken into account for the calculation of the relative quantification ratio (RQ) as described by (Pfaffl, 2001).

Statistics

Outliers were identify and deleted with the Grubbs' test, $\alpha = 0.05$ (Prism 6 – GraphPad). Results are expressed as the means ± standard errors of the means (SEMs). Unpaired parametric t-test (Prism 6 - GraphPad) was performed unless otherwise indicated.

Bibliography for methods section

1. Anders S, P. P. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics, 31(2), 166-9.
2. Bolger, A. M. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120.
3. BROAD Institute. (n.d.). Picard Tools. Retrieved 2017, from https://github.com/broadinstitute/picard
4. Dobin, A. D. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21.
5. Eden, E. N. (2009). GOrilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists. BMC Bioinformatics, 10(48).
6. Love, M. I. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biolog, 15(12), 550.
7. MUGQIC. (n.d.). MUGQIC pipelines RNAseq, 2.2.1-beta. Retrieved 2017, from https://bitbucket.org/mugqic/mugqic_pipelines/src/master/pipelines/rnaseq/
8. Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research, 29(9), 45e–45. https://doi.org/10.1093/nar/29.9.e45

Supplementary discussion on RNAseq analysis

Apart from common DEGs, each patient had a majority of DEGs which weren't shared. Of those, the p.Leu805Phe individual had upregulation of more than 80 cell cycle regulation genes (supplementary). GO analysis of DEGs in this patient showed a significant enrichment in chromosome segregation, chromosome organization, cell division, mitotic cell cycle process, and other related GOs. *TRRAP* has been shown to be necessary for cell-cycle regulation, and *Trrap* conditional knock-out in mouse brain leads to premature neural progenitor differentiation, microcephaly and disorganized neuronal layers[1]. Moreover, *Trrap* knockout mouse ESCs accumulate chromosomal aberrations because of chromosome fragmentation and lagging, and fail to arrest at the mitotic checkpoint[2], consistent with the genes upregulated in this individual (*NDE1, MIS18A, NEK2, CENPE, CENPF,* and others – Supplementary). Thus, it is possible that overexpressing *TRRAP*-related genes can lead to cell-cycle deregulation and abnormal neural progenitor differentiation, leading to the neural phenotype observed in *TRRAP* patients. The p.Trp1866Cys individual did not have the same enrichment in cell cycle regulation genes. Nevertheless, the DEGs with the lowest adjusted P values were associated mostly with neuronal function (*EMB*, *PDE11A*, *STMN2*, *DYSF*, *SGIP1*, *PLEKHG5,* and others - Supplementary). In addition, members of the *HOXC* cluster were significantly upregulated (*HOXC9*, *HOXC10*, *HOXC11*, *HOXC-AS1*, *HOXC-AS2*, *HOXC-AS3, HOTAIR*), as well as *HOXB9* and *HOXD13*, suggesting a role for TRRAP in regulating 5' Hox gene signaling. Homeobox containing transcription factors (*Hox* genes) have been implicated in embryonic morphogenesis and a multitude of disorders[3], and the spatio-temporal regulation of their expression has been thoroughly studied in a model of "temporal collinearity" [4; 5]. Each cluster of *Hox* genes is globally repressed, then progressively activated from the 3' to the 5' end during embryogenesis through the addition of active histone marks (H3K27Ac, H3K4m3) and the removal of repressive histone marks (H3K27m3) at the same time in all the clusters (*Hox* A, B,

C, and D). An irregular early expression of these *Hox* genes can cause abnormal embryonic development in mouse and drosophila[6; 7].

1. Tapias, A., Zhou, Z.W., Shi, Y., Chong, Z., Wang, P., Groth, M., Platzer, M., Huttner, W., Herceg, Z., Yang, Y.G., et al. (2014). Trrap-dependent histone acetylation specifically regulates cell-cycle gene transcription to control neural progenitor fate decisions. Cell stem cell 14, 632-643.
2. Herceg, Z., Hulla, W., Gell, D., Cuenin, C., Lleonart, M., Jackson, S., and Wang, Z.Q. (2001). Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression. Nature genetics 29, 206-211.
3. Quinonez, S.C., and Innis, J.W. (2014). Human HOX gene disorders. Molecular genetics and metabolism 111, 4-15.
4. Mallo, M., and Alonso, C.R. (2013). The regulation of Hox gene expression during animal development. Development 140, 3951-3963.
5. Montavon, T., and Duboule, D. (2013). Chromatin organization and global regulation of Hox gene clusters. Philosophical transactions of the Royal Society of London Series B, Biological sciences 368, 20120367.
6. Coiffier, D., Charroux, B., and Kerridge, S. (2008). Common functions of central and posterior Hox genes for the repression of head in the trunk of Drosophila. Development 135, 291-300.
7. Kmita, M., van Der Hoeven, F., Zakany, J., Krumlauf, R., and Duboule, D. (2000). Mechanisms of Hox gene colinearity: transposition of the anterior Hoxb1 gene into the posterior HoxD complex. Genes & development 14, 198-211.

Supplemental Acknowledgments

We would like to thank Frédérique Allaire from the Health Regional Agency of Poitou-Charentes for supporting this project. We acknowledge Léa Ferrand and Emilie Le Blanc's assistance for grant and data management. E.E.E. is an investigator of the Howard Hughes Medical Institute. CAUSES Study investigators include Shelin Adam, Christele Du Souich, Alison Elliott, Anna Lehman, Jill Mwenifumbo, Tanya Nelson, Clara Van Karnebeek, and Jan Friedman. We acknowledge also the DECIPHER Consortium which contributed to the exchange of genetic and clinical data between the teams.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.