The American Journal of Human Genetics, Volume *104*

Supplemental Data

Expanding the Spectrum of BAF-Related Disorders: *De Novo* Variants in *SMARCC2* Cause a Syndrome with Intellectual Disability and Developmental Delay

Keren Machol, Justine Rousseau, Sophie Ehresmann, Thomas Garcia, Thi Tuyet Mai Nguyen, Rebecca C. Spillmann, Jennifer A. Sullivan, Vandana Shashi, Yong-hui Jiang, Nicholas Stong, Elise Fiala, Marcia Willing, Rolph Pfundt, Tjitske Kleefstra, Megan T. Cho, Heather McLaughlin, Monica Rosello Piera, Carmen Orellana, Francisco Martínez, Alfonso Caro-Llopis, Sandra Monfort, Tony Roscioli, Cheng Yee Nixon, Michael F. Buckley, Anne Turner, Wendy D. Jones, Peter M. van Hasselt, Floris C. Hofstede, Koen L.I. van Gassen, Alice S. Brooks, Marjon A. van Slegtenhorst, Katherine Lachlan, Jessica Sebastian, Suneeta Madan-Khetarpal, Desai Sonal, Naidu Sakkubai, Julien Thevenon, Laurence Faivre, Alice Maurel, Slavé Petrovski, Ian D. Krantz, Jennifer M. Tarpinian, Jill A. Rosenfeld, Brendan H. Lee, Undiagnosed Diseases Network, and Philippe M. Campeau

Supplemental Data

Supplemental Figures

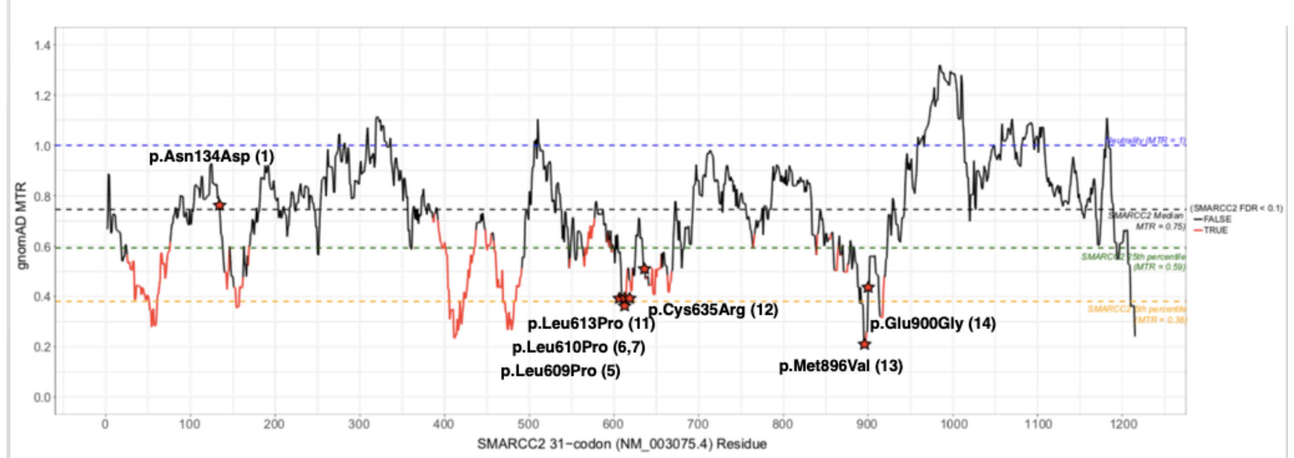
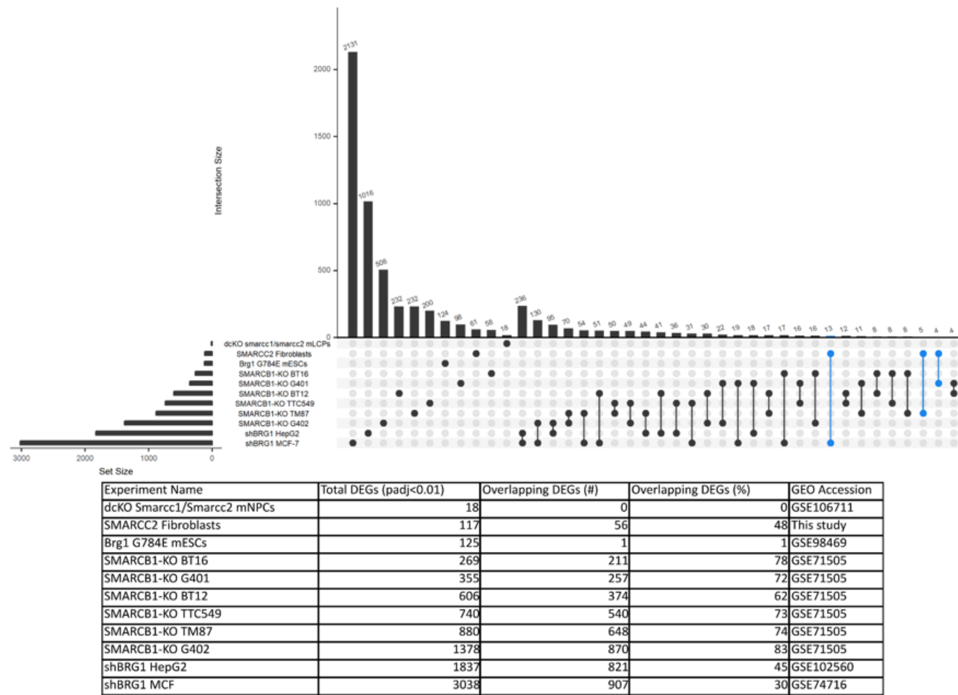


Figure S1: Visualization of the eight *de novo* missense variants in *SMARCC2* using the Missense Tolerance Ratio (MTR) tool. Eight of the individuals in this cohort harbor one of seven *de novo* missense variants. Six of these missense variants preferentially affect one of the 25% most intolerant residues of SMARCC2 (p=0.0004). p.Asn134Asp is the only missense variant affecting a tolerant region of this gene. Numbers in brackets represent individual's number.



Experiment Name	Total DEGs (padj<0.01)	Overlapping DEGs (#)	Overlapping DEGs (%)	GEO Accession
dcKO Smarcc1/Smarcc2 mNPCs	18	0	0	GSE106711
SMARCC2 Fibroblasts	117	56	48	This study
Brg1 G784E mESCs	125	1	1	GSE98469
SMARCB1-KO BT16	269	211	78	GSE71505
SMARCB1-KO G401	355	257	72	GSE71505
SMARCB1-KO BT12	606	374	62	GSE71505
SMARCB1-KO TTC549	740	540	73	GSE71505
SMARCB1-KO TM87	880	648	74	GSE71505
SMARCB1-KO G402	1378	870	83	GSE71505
shBRG1 HepG2	1837	821	45	GSE102560
shBRG1 MCF	3038	907	30	GSE74716

Figure S2: Comparison of RNAseq data from various public datasets involving BAF-complex components. RNAseq results were obtained from the European Nucleotide Archive. Datasets were analyzed with a padj cutoff of 0.01 as some datasets did not have DEGs whose Log2FC was higher than 2. DEGs shared by at least 2 datasets were considered as overlapping. To compare DEGs between species, gene names were converted from mouse to human, then plotted using the UpSet R package (Conway J.R, 2017). Overlaps which contain genes found in *SMARCC2* Fibroblasts are colored in blue.

Supplemental Tables

Table S1: Elaborated table describing the clinical findings in 15 individuals harboring *SMARCC2* variants.

ADHD- attention deficit hyperactivity disorder; AG-aggressiveness; ASD- autism spectrum disorder; AX- anxiety; CNS – central nervous system; DD- developmental delay; DQ- developmental quotient; DSI- difficulties with social interactions (not qualify for ASD diagnosis); F-female; FTT – failure to thrive; HA- hyperactive; HP- hyperphagia; HS- hypersensitive; ID- intellectual disability; L – left; M- male; Mo- months; MRI- magnetic resonance imaging; n/a- not available information; n/ap- not applicable; n/r- not reported; OB- obsession; OFC- occipital frontal circumference; R- right; RG- rigidity/'routine driven'; SD- standard deviation; SI- self-injury; SLD- sleep disturbance; ST- stubborn; TAN- tantrums; Y- years;

Table S2: Differentially expressed genes in individuals 1 and 7 with *SMARCC2* variants p.Asn134Asp and p.Leu610Pro compared to control samples. Significant differentially expressed genes were selected with log_2 fold change of more than 2 or less than -2, and an adjusted p-value lower than 0.01. Due to sex differences between patients (males) and controls (females), genes on the X or Y chromosomes are not included.

Table S3: Differentially expressed genes in individual 1 with *SMARCC2* variant p.Asn134Asp compared to control samples. Significant differentially expressed genes were selected with log_2 fold change of more than 2 or less than -2, and an adjusted p-value lower than 0.01. Due to sex differences between patients (males) and controls (females), genes on the X or Y chromosomes are not included.

Table S4: Differentially expressed genes in individual 7 with *SMARCC2* variant p.Leu610Pro compared to control samples. Significant differentially expressed genes were selected with log_2 fold change of more than 2 or less than -2, and an adjusted p-value lower than 0.01. Due to sex differences between patients (males) and controls (females), genes on the X or Y chromosomes are not included.

Table S5: GOrilla annotation for differentially expressed genes in patients with *SMARCC2* variants p.Leu610Pro or p.Asn134Asp compared to control samples. Significant differentially expressed genes were selected with absolute Log_2FC higher than 2, and an adjusted p-value lower than 0.01. N is the total number of background genes, B is the total number of genes in the background list which correspond to the specific GO, n is the total number of genes with a differential expression in both the patients, and b is the number of genes in that target list with this GO annotation.

Supplemental Methods

Document S1: Materials and Methods

RT-qPCR for gene expression analysis:

Total RNA was isolated from fibroblasts using Qiagen RNeasy mini kit (Qiagen cat# 74104) and treated with DNase (Turbo DNA Free, ThermoFisher cat # AM1907) before the cDNA synthesis. cDNA was prepared from 2ug of total RNA using Superscript III Reverse Transcriptase and oligo(dT) primer (ThermoFisher). cDNA was quantified by using SsoAdvanced Universal SYBR

Green (Bio-rad) on a LightCycler 480 II (Roche) with a 10s denaturing step at 95ºC and 20s annealing step at 59ºC for 40 cycles.

Splicing PCR: The following primers generated from mRNA of SMACCR2 (NM_003075) were used. SMARCC2-F2 GGC TGC GCA CAG AC A TGT ACA CAA;

SMARCC2-R2 TAA CAG GGT TGC CCG ACT GAC TGA;

SMARCC2-R3 TCC GCC TTG CCT GTT ACT TTG GCT;

The PCR cycling condition for PCR is as following: 95°C 60seconds for denature, 60°C 30 second annealing, and 72°C

The PCR products were recovered from agarose gel and then sequenced.

Splicing RT-qPCR: Total RNA was isolated from LCLs using Qiagen RNeasy mini kit, and treated with DNase using the Turbo DNA Free kit. cDNA was prepared from 1 ug RNA using qScript cDNA SuperMix (VWR International cat # CA101414-104). cDNA was quantified by using PowerUp SYBR Green Master Mix (ThermoFisher cat# A25742) on a LightCycler 96 (Roche).

RNAseq Analysis

Human primary fibroblasts were sub-cultured in DMEM (ThermoFisher cat# 11995-065) 10% FBS, 1mM GlutaMax (ThermoFisher cat#35050-061) and 1X antibiotics-antimycotics (ThermoFisher cat# 15240-062). Fibroblasts were plated at 1 million cells per 150 mm dish and allowed to grow until they reached about 80% confluency. Cells were trypsinized, washed 2 times with 1X 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 McGill University and Genome Quebec Innovation Center (MUGQIC). mRNA Libraries were prepared using the TruSeq Stranded mRNA kit (Illumina) according to the manufacturer's instructions. Samples were run on the Illumina HiSeq 4000 PE100. Output files were analyzed using the MUGQIC RNAseq pipeline (MUGQIC, n.d.) steps 1 through 14 on the Guillimin Genome Quebec HPC. In summary, BAM files were converted to FASTQ using Picard (BROAD Institute, n.d.), sequences were trimmed using Trimmomatic (Bolger, 2014) then aligned to the GRCh37 genome using STAR (Dobin, 2013), 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 annotations were obtained using the GOrilla web application (Eden, 2009).

Supplementary Analysis of Overlaps (figure S2): To understand the low overlap between DEGs in the two assessed individuals, we compared our results to other public datasets involving SWI/SNF components. We looked at several datasets in both human and mouse cell lines and primary cells in culture (Table). We re-analyzed every dataset separately using only a padj cutoff of 0.01 as some datasets did not have DEGs whose Log2FC was higher than 2. To compare DEGs between species, gene names were converted from mouse to human, then plotted using the UpSet R package. There is a high discrepancy in the number of

significant DEGs for each dataset; mouse primary late cortical progenitors with a double SMARCC1/SMARCC2 knock-out (dcKO mLCPs) only have 18 DEGs, compared to 3038 DEGs for the BRG1 knockdown in the MCF-7 human breast adenocarcinoma cell-line. Additionally, overlaps between separate datasets are scarce; the highest overlap is 236 genes between the BRG1-KD in MCF-7 cells and the BRG1-KD in HepG2 hepatocellular carcinoma cells, corresponding to 8 and 12% overlap respectively.

Supplemental References

Anders S, P. P. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics, 31*(2), 166-9.

Bolger, A. M. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114–2120. doi:10.1093/bioinformatics/btu170

BROAD Institute. (n.d.). *Picard Tools*. Retrieved 2017, from https://github.com/broadinstitute/picard

Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938-2940

Dobin, A. D. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15–21. doi:10.1093/bioinformatics/bts635

Eden, E. N. (2009). GOrilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists. *BMC Bioinformatics, 10*(48).

Love, M. I. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biolog, 15*(12), 550. doi:10.1186/s13059-014-0550-8

MUGQIC. (n.d.). *MUGQIC pipelines RNAseq*, 2.2.1-beta. Retrieved 2017, from https://bitbucket.org/mugqic/mugqic_pipelines/src/master/pipelines/rnaseq/