



OPEN

## Transcriptome analysis of *MBD5*-associated neurodevelopmental disorder (MAND) neural progenitor cells reveals dysregulation of autism-associated genes

Sureni V. Mullegama<sup>1,8</sup>, Steven D. Klein<sup>2,9,10</sup>, Stephen R. Williams<sup>3</sup>, Jeffrey W. Innis<sup>4</sup>, Frank J. Probst<sup>1</sup>, Chad Haldeman-Englert<sup>5</sup>, Julian A. Martinez-Agosto<sup>2</sup>, Ying Yang<sup>6</sup>, Yuchen Tian<sup>7</sup>, Sarah H. Elsea<sup>1</sup>✉ & Toshihiko Ezashi<sup>7</sup>✉

*MBD5*-associated neurodevelopmental disorder (MAND) is an autism spectrum disorder (ASD) characterized by intellectual disability, motor delay, speech impairment and behavioral problems; however, the biological role of methyl-CpG-binding domain 5, *MBD5*, in neurodevelopment and ASD remains largely undefined. Hence, we created neural progenitor cells (NPC) derived from individuals with chromosome 2q23.1 deletion and conducted RNA-seq to identify differentially expressed genes (DEGs) and the biological processes and pathways altered in MAND. Primary skin fibroblasts from three unrelated individuals with MAND and four unrelated controls were converted into induced pluripotent stem cell (iPSC) lines, followed by directed differentiation of iPSC to NPC. Transcriptome analysis of MAND NPC revealed 468 DEGs ( $q < 0.05$ ), including 20 ASD-associated genes. Comparison of DEGs in MAND with SFARI syndromic autism genes revealed a striking significant overlap in biological processes commonly altered in neurodevelopmental phenotypes, with TGF $\beta$ , Hippo signaling, DNA replication, and cell cycle among the top enriched pathways. Overall, these transcriptome deviations provide potential connections to the overlapping neurocognitive and neuropsychiatric phenotypes associated with key high-risk ASD genes, including chromatin modifiers and epigenetic modulators, that play significant roles in these disease states.

Autism spectrum disorder (ASD) is an encompassing term that describes neurodevelopmental disorders that have common deficits in communication and social interactions, accompanied by stereotyped behaviors and restricted interest. The exact molecular mechanism(s) that leads to this clinical entity is unknown. Studies suggest that between 20 and 45% of all ASDs are associated with rare pathogenic variants in specific genes<sup>1,2</sup>. The use of chromosomal microarray (CMA) is advocated as the first tier genetic test for individuals with ASD<sup>3,4</sup>.

<sup>1</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. <sup>2</sup>Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. <sup>3</sup>10X Genomics, San Francisco, CA, USA. <sup>4</sup>Departments of Human Genetics, Pediatrics and Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA. <sup>5</sup>Mission Fullerton Genetics Center, Asheville, NC 28803, USA. <sup>6</sup>Department of Molecular Pharmacology and Physiology, University of South Florida, Tampa, FL 33612, USA. <sup>7</sup>Division of Animal Sciences and Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA. <sup>8</sup>Present address: Department of Molecular and Cellular Biology, College of Osteopathic Medicine, Sam Houston State University, Conroe, TX 77304, USA. <sup>9</sup>Present address: Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA. <sup>10</sup>Present address: Department of Medical Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA. ✉email: elsea@bcm.edu; ezashit@missouri.edu

The utilization of CMA led to the identification of *MBD5*-associated neurodevelopmental disorder (MAND, MIM 156200), which is an ASD characterized by intellectual disability, motor delay, severe speech impairment, and behavioral problems<sup>5</sup>. MAND serves as an umbrella term that encompasses 2q23.1 deletion syndrome, *MBD5* variant cases, and 2q23.1 duplications<sup>5</sup>. The conditions in this group of disorders have copy number variations or pathogenic sequence variants in the methyl-binding domain 5 gene (*MBD5*, MIM 611472), which is responsible for the majority of the phenotypes present in MAND<sup>6,7</sup>. *MBD5* variants have been associated with ASD, epilepsy, and schizophrenia in multiple cohorts<sup>7,8</sup>. The pathological mechanisms and pathways underlying the primarily neurological, neurodevelopmental, and neurobehavioral aspects of the MAND phenotype are not well delineated. Deficiency of *Mbd5* in mice resulted in reduced body weight motor deficits, abnormal social behavior, impaired learning, and iron metabolism abnormalities<sup>9,10</sup>. However, the importance of the genes associated with MAND in specific neurodevelopmental functions remains largely undefined.

Utilization of animal models to understand autism and other human neurodevelopmental or psychiatric disorders proves to be difficult<sup>11</sup>. Further, the absence of the “perfect” animal model leads to the inability to truly understand the human pathological mechanism which can lead to potential therapeutic targets<sup>12,13</sup>. Recent studies utilizing stem cell technology to derive neurons has served as the popular method to model neurodevelopmental disorders and has revealed deeper understanding of the phenotype of these disorders<sup>13–15</sup>. In addition, next-generation RNA-sequencing (RNA-seq) has been utilized in numerous studies and is a valid tool to identify therapeutic targets, biomarkers, pathways, and genes associated with ASD modeled by stem cell-derived human neurons<sup>16–19</sup>.

In this study, we explore the underpinnings of the neurodevelopmental phenotype of MAND by creating neural progenitor cells (NPC) derived from patients with 2q23.1 deletion syndrome, followed by RNA-seq to identify the contributing molecular pathways. Data show that haploinsufficiency of one or more genes the 2q23.1 region leads to *FOXG1* upregulation in NPC, accompanied with differentially expressed genes implicated in forebrain regionalization in early neurogenesis. Altered genes significantly overlap with ASD-associated genes, providing further insight into mechanisms underlying ASD by identifying shared network perturbances caused by the genetically penetrant MAND-associated 2q23.1 deletion.

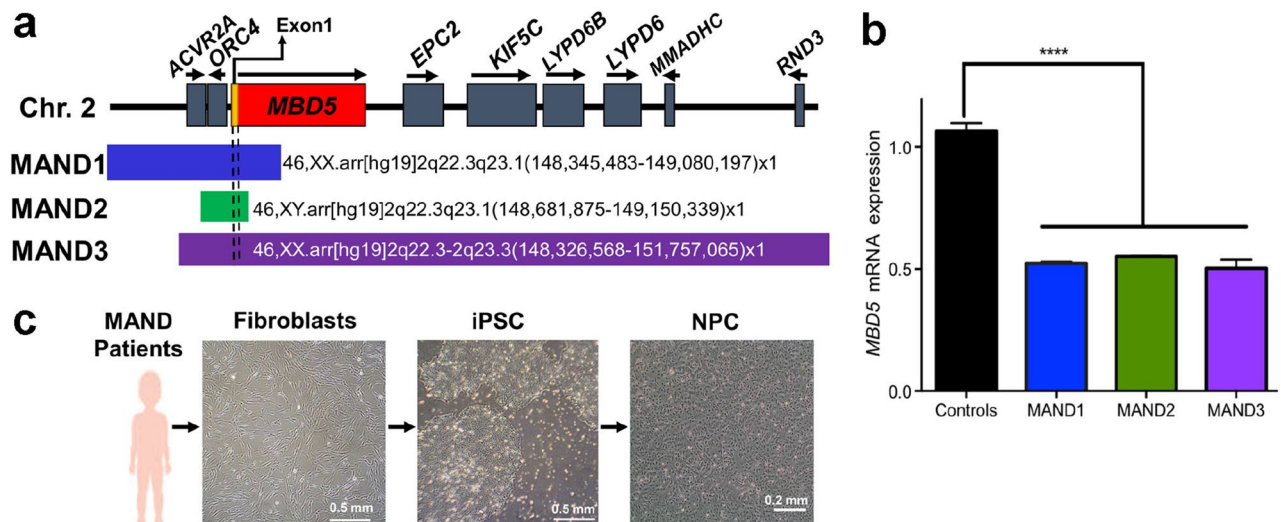
## Materials and methods

This study was approved by the Baylor College of Medicine Institutional Review Board. All samples and information were collected after obtaining written informed consent. Whole blood and skin samples were provided by each participant. Participant recruitment, sample collection, and all experiments were performed in accordance with relevant guidelines and regulations by the Baylor College of Medicine Institutional Review Board and associated regulatory committees.

**Primary cell culture.** Primary skin fibroblast cells were derived and cultured in DMEM (Thermo) containing 10% FBS, 1% non-essential amino acids, 2 mM glutamine, 0.1 mM 2-mercaptoethanol, and 4 ng/ml human recombinant FGF2 (in house)<sup>20</sup>. Fibroblasts and NPC lines utilized in this study were obtained from four controls (CTR1, CTR2<sup>21</sup>, CTR3, CTR4) and three MAND patients with confirmed molecular diagnosis of 2q23.1 deletion syndrome by chromosomal microarray (Fig. 1a, Supplementary Table 1)<sup>22–24</sup>. MAND1 (SMS388, 3 y old female) has a 0.73 Mb loss at 2q22.3q23.1, 46,XX,arr[hg19]2q22.3q23.1(148,345,483–149,080,197) × 1,15q11.1q11.2(20,262,223–22,285,757) × 1<sup>6,25</sup>. MAND2 (SMS447, 10 y male) has a 0.18 Mb 2q23.1 deletion, 46,XY,arr[hg19]2q23.1(148,687,657–148,872,174) × 1. MAND3 (SMS456, 8 y female) has a 3.43 Mb deletion at 2q23.1, 46,XX,arr[hg19]2q22.3–q23.3(148,326,568–151,757,065) × 1<sup>25</sup>. All MAND cases were confirmed to have haploinsufficiency of *MBD5* (Fig. 1b).

**Reprogramming skin fibroblasts into iPSC with episomal plasmids.** Early passages (between passages p3 and p5) of skin fibroblasts were cultured at low oxygen atmosphere of 4% O<sub>2</sub>/5% CO<sub>2</sub>/91% N<sub>2</sub> and maintained in the condition for the entire reprogramming period to enhance iPSC generation<sup>21,26</sup>. Episomal vectors carrying shRNA for p53 suppression and non-transforming *MYCL*, in addition to the usual reprogramming genes, *POU5F1*, *SOX2*, *KLF4*, and *LIN28* were employed to reprogram the fibroblasts<sup>27</sup>. Three micrograms of combined episomal plasmids (Addgene 27077, 27078 and 27080) were electroporated into 6 × 10<sup>5</sup> fibroblast cells by using a Nucleofector II device (Lonza) and an Amaxa NHDF Nucleofector kit (Lonza) according to the manufacturer’s instructions<sup>28</sup>. The electroporated cells were allowed to recover for 2 to 4 days by culturing in the above conditions. Cells (2 × 10<sup>5</sup>) were placed into 100 mm dishes previously coated with Matrigel (BD Bioscience). The following day the culture medium was switched to mTeSR1<sup>29</sup> (STEMCELL Technologies). Colonies emerged about 14 days post-transduction. Colonies were mechanically isolated around day 20 and expanded under feeder free conditions in mTeSR1 medium on a Matrigel coated substratum.

**Neural progenitor cell (NPC) differentiation.** NPC lines were generated from iPSCs through monolayer culture protocol using STEMdiff Neural Induction Medium (StemCell Technologies), following the manufacturer’s instructions. Briefly, 2 × 10<sup>6</sup> iPSCs were transferred onto Matrigel-coated d35 plate (day 0) with Neural Induction Medium. The medium was refreshed daily, and the cells were passaged when they were confluent (day 7–9). By that time, the cells showed robust PAX6 expression and extensive reductions of pluripotent markers (*POU5F1*, *NANOG*). After passage (p), PAX6 expression in the entire cell population was further improved and the cell doubling time was determined by cell counting between p1 and 2 (day 8–15). RNA preparations and cells for immunostaining were collected between p1 and p4 (day 12–22) to ensure that both control and MAND iPSC lines were equally and fully differentiated into NPC stage and avoided any undifferentiated cell to be included (Supplementary Table 1).



**Figure 1.** MAND patient 2q23.1 deletions include *MBD5* and exhibit reduced *MBD5* expression indicating haploinsufficiency. **(a)** Schematic diagram of the chromosome 2q23.1 deletion region. The coding region of *MBD5* is indicated by the red box; yellow region indicates non-coding exons. Chromosome 2q23.1 deletions present in the MAND patients in this study are summarized with bars. Additional details regarding MAND patients are provided in Supplementary Table 1. **(b)** qRT-PCR analysis showed ~50% reduction of *MBD5* mRNA expression in primary fibroblasts derived the three MAND patients compared to the controls. Expression of *MBD5* was calculated relative to that of the housekeeping gene, *GAPDH*. All expression values were calculated relative to control levels set at 1.0. Error bars represent the SEM. \*\*\*\* $p < 0.0001$ ,  $n = 3$  independent biological replicates performed in triplicate, GraphPad Prism 8.4.2 (graphpad.com). **(c)** Overview of iPSC and neural progenitor cell (NPC) line derivations in this study. Skin fibroblasts from MAND patients (bar, 0.5 mm) were converted into iPSC by transfection with episomal plasmids encoding reprogramming factors (bar, 0.5 mm). NPC were derived from iPSC with neuronal induction medium (bar, 0.2 mm).

**Immunostaining.** Cells were grown on Lab-Tek II 4-well chambered cover glass (Nunc) coated with Matrigel. After fixing the cells in 4% paraformaldehyde in PBS for 10 min and permeabilizing them in 1.0% Triton X-100/PBS for 30 min, coverslips were placed in 5% (v/v) goat serum/5% w/v BSA in PBS for 1 h. Cells were then incubated with appropriately diluted primary antibodies (Supplementary Table 2) overnight at 4 °C. Secondary antibody staining was performed with either Alexa Fluor 568-, 647- or 488-labeled detection reagents (goat anti-rabbit, goat anti-mouse antibodies; Molecular Probes) at 1:300 dilution (Supplementary Table 2). Images were captured under an Olympus IX70 inverted microscope with an ORCA-AG CCD camera (<http://www.biotech.missouri.edu/mcc/Olympus.html>).

**Quantitative RT-PCR.** This analysis was performed similar to previously reported<sup>6,30</sup>. Briefly, RNA was isolated from control and patient NPC lines using RNeasy Plus Mini Kits (QIAGEN) according to manufacturer's instructions. All samples were treated with DNase I. RNA was quantified using the NanoDrop ND-100 Spectrophotometer. First-strand cDNA synthesis was carried out with qScript cDNA SuperMix (Quanta Biosciences) according to manufacturer's instructions. For quantitative real-time PCR, predesigned Taqman MGB (minor groove binder) probes from Assays-on-Demand Gene Expression Products (Life Technologies) were used for all genes (see Supplementary Table 3). All samples were run in triplicate in 10  $\mu$ L reaction volumes. PCR conditions were the default setting of the ABI Prism 7900 HT Sequence Detection System (Life Technologies). The cycle threshold (Ct) was determined during the geometric phase of the PCR amplification plots, as recommended by the manufacturer. Relative differences in transcript levels were quantified with the  $\Delta\Delta$ Ct method with *GAPDH* as an endogenous control and were compared between control and MAND cell lines. Statistical analysis for gene expression differences by one-way ANOVA was performed with Prism 8.4.2 (GraphPad Software, graphpad.com).

**RNA-seq transcriptome analysis.** RNA was isolated from NPC using RNeasy Plus Mini Kits (QIAGEN). The quantitation and quality control of RNA from all samples were performed on a Fragment Analyzer (Advanced Analytical), and cDNA libraries were constructed by standard methods (Illumina TruSeq mRNA stranded kit) with index adapters (Illumina TruSeq indexes). All libraries were sequenced as paired-end, 50 base read length on an Illumina HiSeq 2500 with an average read count of 21–28 million reads per sample at University of Missouri DNA core (dnacore.missouri.edu/). Once reads were obtained, all reads were analyzed for quality using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), aligned to human genome GRCh38 using STAR<sup>31</sup> and counted using FeatureCounts<sup>32</sup>. The resulting gene-wise count matrices were used for downstream analyses. Differential gene expression analysis was conducted using DeSeq2<sup>33</sup> comparing MAND NPC lines with control NPC lines. Of the 59,074 transcripts analyzed, 19,801 passed all DeSeq2 quality filters

and were given a  $p$  value. Genes were considered differentially expressed if they reached a corrected  $p < 0.005$  using a Benjamini–Hochberg corrected false discovery rate (FDR)  $\alpha < 0.005$ . Data for these experiments are freely available at the Gene Expression Omnibus (GSE141835). The significantly differentially expressed gene set (498 genes shown in Supplementary Data 1) was cleaned and converted to HGNC approved gene names. The list was further edited for unmatched terms and synonyms resulting in a list of 468 genes (Supplementary Data 3, Tab 1).

**Gene expression network analysis: biological processes and enrichment.** To identify the prevalent biological pathways affected in MAND due to a 2q23.1 deletion, we used four independent pathway analysis databases (BIOCARTA, KEGG<sup>34</sup>, PANTHER<sup>35</sup> and REACTOM<sup>36</sup>) to incorporate and assess the 468 differentially expressed genes (significance cut off  $p = 0.05$ ). Since each of these approaches utilize different methodologies and algorithms, we conducted each analysis separately to provide the most comprehensive approach to pathway interrogation. We also utilized Gene Set Enrichment Analysis (GSEA)<sup>37</sup> with default parameters for Gene Ontology (GO) analysis of biological processes, molecular functions and pathways. To identify the enrichment of the 468 differentially expressed genes by chromosomal location, we utilized Enrichr<sup>38</sup>. This enrichment analysis relied on ranking the list of genes based on statistical significance with the default settings of GSEA<sup>39</sup>.

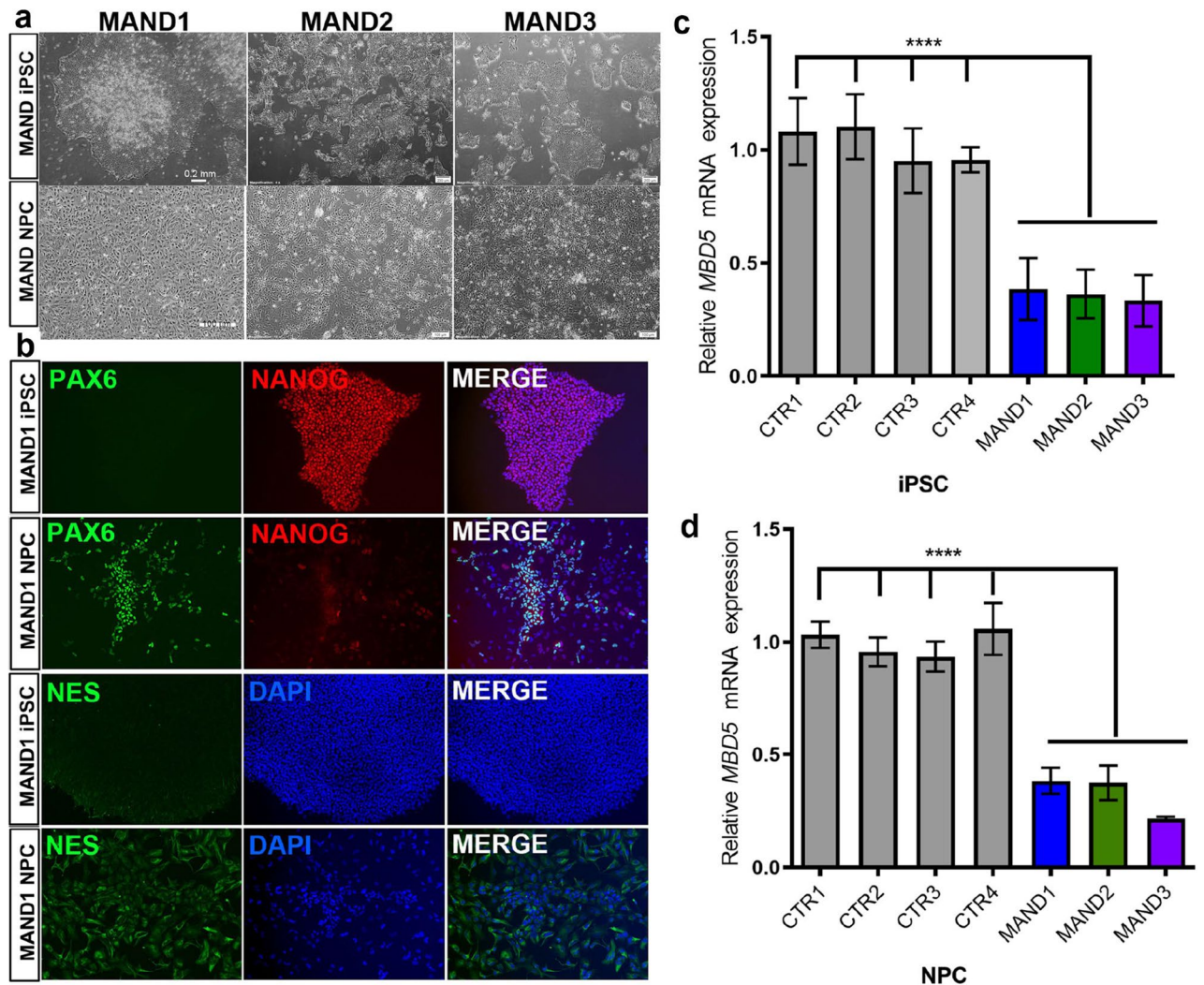
**Comparison of MAND and ASD biological processes and pathway analyses.** We first compared the MAND differentially expressed genes list (Supplementary Data 3, Tab 1) to the SFARI database gene list (Supplementary Data 3, Tab 2), which is a well-curated list of 913 genes, then we expanded this list to be inclusive for HGNC synonyms to a final list of 944 genes. We compared the two lists to identify the overlapping genes (Supplementary Data 3, Tab 3) and then sought to compare the biological processes which were represented by each gene list using the ToppFUN tool<sup>40</sup> (Supplementary Data 3, Tab 4). We limited our analysis to GO: Biological Process with a Bonferroni correction for multiple testing, a statistical  $p$  value cutoff of 0.05, and limited genes to  $1 < n < 1500$ . We captured the top 50 enriched biological processes and compared them between the two lists (additional tables in the Tab 4). Finally, we investigated whether our MAND differentially expressed genes were involved in known ASD associated pathways (Supplementary Data 3, Tab 5). Pathways were visualized utilizing the Reactome FIViz application in Cytoscape<sup>41</sup> (Fig. 4c).

## Results

**Identification of MAND (2q23.1 deletion syndrome) patients.** Three individuals with 2q23.1 deletion syndrome were identified clinically by chromosomal microarray analysis (CMA) (Fig. 1a, Supplementary Table 1). Despite the differently sized deletions (Fig. 1a), all deletions include *MBD5* exon 1, containing the transcription start site, as well as the *MBD5* translation start site (located in exon 6), and as a consequence, all three patients have ~50% *MBD5* mRNA expression ( $p < 0.0001$ ) resulting in haploinsufficiency (Fig. 1b). These three patients have phenotypes commonly observed in MAND, including developmental delay, intellectual disability, sleep disturbance, seizures, language delay, hypotonia, mild dysmorphic features, and ASD-associated behaviors (short attention span and stereotypic, repetitive behaviors), consistent with the most prevalent clinical features of 2q23.1 deletion syndrome<sup>5</sup>. Since the primary phenotypic concerns for patients with 2q23.1 deletion syndrome are neurodevelopmental, we took an approach to investigate the molecular mechanism(s) underlying these phenotypes by creating neural progenitor stem cells from fibroblast cell lines derived from skin biopsies from each patient (Fig. 1c).

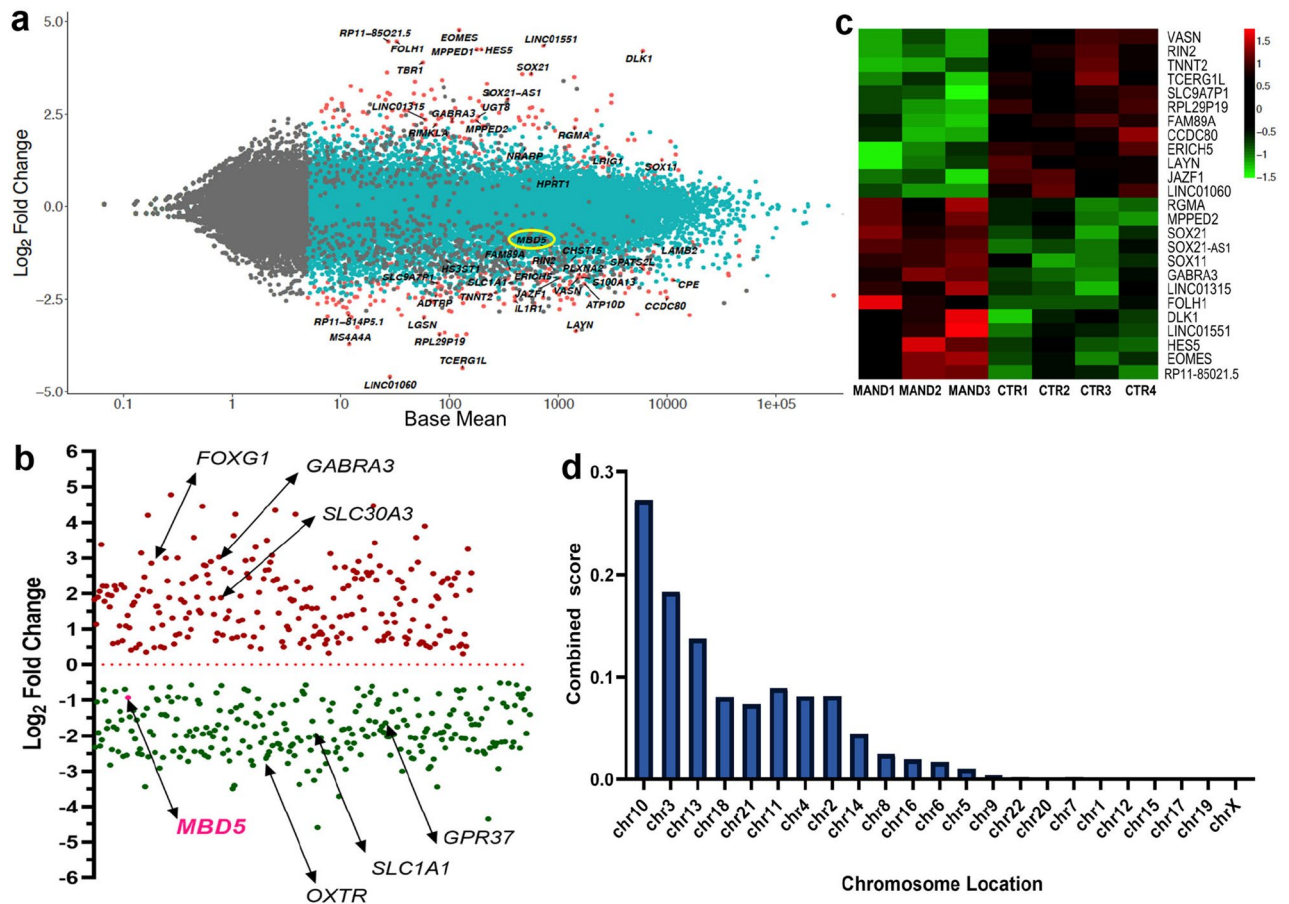
**Generation and characterization of iPSC derived from MAND patient fibroblasts.** Skin fibroblast cell lines from the three MAND patients and four normal controls were developed within one month from the time of the biopsy. The fibroblasts were expanded, stored, and then transduced to produce iPSCs in a feeder-free and defined culture system (Supplementary Fig. 1). The generated iPSC colonies expressed pluripotent markers, alkaline phosphatase, POU5F1 (also known as OCT4), NANOG, and TRA-1-60 (Supplementary Fig. 1). Programming efficiencies for MAND1-3 were 0.016, 0.4, and 0.022%, respectively. The successfully generated colonies were banked, and we utilized a subset of the iPSC lines to characterize MAND1-iPSC, MAND2-iPSC, MAND3-iPSC (Fig. 2a,b, Supplementary Fig. 2c), and CTR1-iPSC, CTR2-iPSC, CTR3-iPSC, and CTR4-iPSC (Supplementary Fig. 2a,b). Isolated DNAs from the iPSC lines were analyzed by chromosomal microarray (CMA) to confirm the molecular karyotype for each cell line was preserved and consistent with the original fibroblast line (Fig. 1a). Furthermore, we isolated RNA from these iPSCs and confirmed reduced *MBD5* mRNA expression in the MAND-iPSCs (Fig. 2c).

**Generation and characterization of NPCs derived from MAND-iPSCs.** Using an established neural induction monolayer culture protocols (STEMCELL Technologies)<sup>42,43</sup>, MAND and control iPSC lines were differentiated into PAX6 expressing NPCs (Fig. 2a,b & Supplementary Fig. 2). Expression of the pluripotent stem cell marker NANOG and POU5F1, which was initially expressed in 100% of the iPSCs before the differentiation, had decreased to undetectable level. Another neural stem cell marker NES was expressed (Fig. 2b). Finally, DNA and RNA were isolated for CMA and RT-qPCR studies, respectively, to confirm cells maintained their original integrity during differentiation into neural lineage cells. CMAs of each NPC line were consistent with fibroblast and iPSC analyses, and RT-qPCR studies confirmed significantly reduced *MBD5* mRNA expression in the MAND-NPC lines (Fig. 2d).



**Figure 2.** Generation of iPSC and NPC lines from individuals with MAND. (a) Bright-field images of undifferentiated iPSC lines from MAND1-3 (top row panel, l to r). Lower panel images are three NPC lines derived from MAND1-3 iPSC lines, respectively (l to r). After two weeks of neuronal induction, images were captured. Bars, 0.2 mm. (b) Immunostaining of MAND1 iPSC (1<sup>st</sup> row) and MAND1 NPC (2<sup>nd</sup> row) for neural progenitor cell marker PAX6 (1<sup>st</sup> column) and pluripotent cell marker NANOG (2<sup>nd</sup> column and 3<sup>rd</sup> for merged). Immunostaining of MAND1 iPSC and MAND1 NPC for NES (3<sup>rd</sup> and 4<sup>th</sup> rows, respectively). (c, d) qRT-PCR analysis confirmed ~50% reduction of *MBD5* mRNA expression in MAND iPSC (c) and NPC lines (d) when compared with control iPSC and NPC lines, respectively.  $p < 0.0001$ . Quantification of the test gene (*MBD5*) was calculated relative to that of the housekeeping gene, *GAPDH*. All expression values were calculated relative to control levels set at 1.0. Error bars represent the SEM. \*\*\*\* $p < 0.0001$ ,  $n = 3$  independent biological replicates performed in triplicate, GraphPad Prism 8.4.2 (graphpad.com).

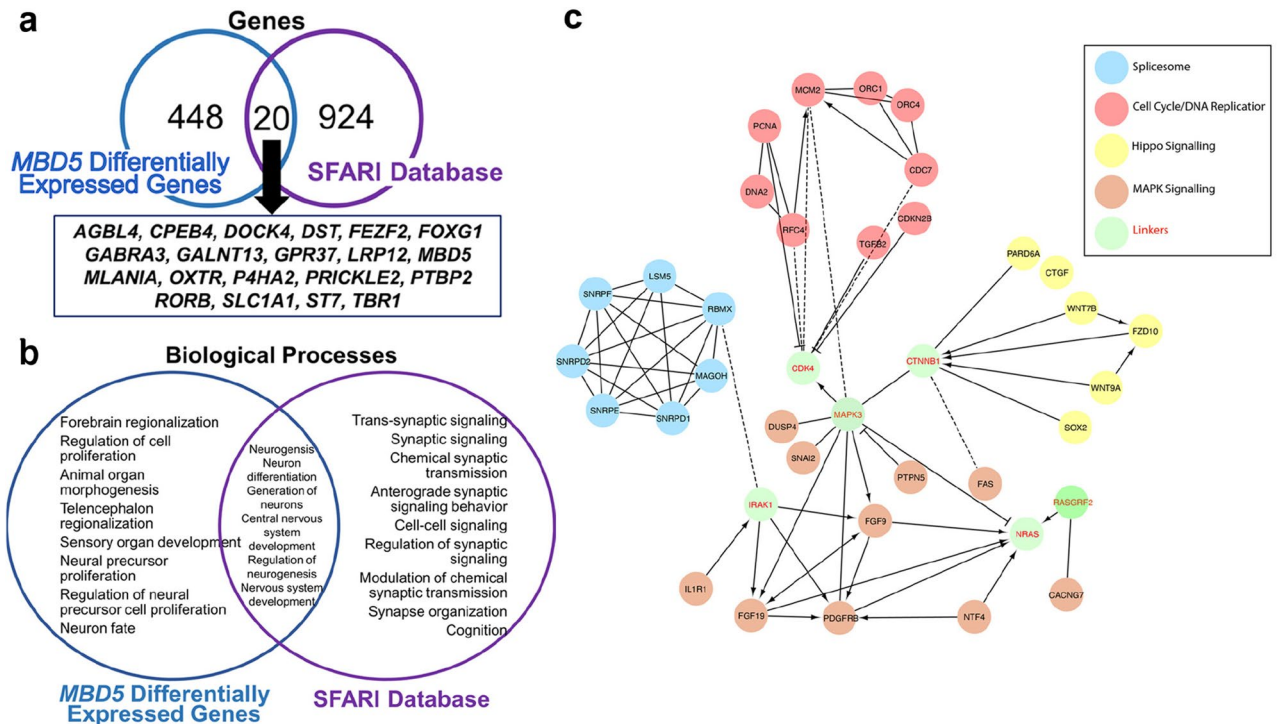
**Global transcriptome analysis of MAND NPC reveals genome-wide alterations.** In order to investigate the genes, pathways, and biological processes that are disrupted due to *MBD5* haploinsufficiency that might provide insight into the etiology of this complex neurodevelopmental condition, we performed RNA-seq analysis on these established NPC lines. RNA-seq analysis ( $n = 3$  MAND-NPCs,  $n = 4$  CTR-NPCs) initially identified 498 differentially expressed genes in MAND-NPCs (Fig. 3a). From these 498 genes, there was an even distribution of upregulated and downregulated genes, confirming that *MBD5* can act as a positive and negative regulator of gene expression (Fig. 3b). The top differentially expressed (up and down) genes in MAND-NPCs compared to the Control-NPCs are depicted in a heat map (Fig. 3c). The 228 genes up-regulated in MAND-NPCs (Fig. 3b, see Supplementary Data 2 for complete list) include those coding for a variety of transcription factors, including *FOXG1\** and other F-Box, T-Box, SOX, and homeobox genes; transmembrane proteins (transporters, receptors), including *GABRA3\** and *SLC30A3\**; ligands (*WNT7B*, *FGF9*); and non-coding RNAs (*VASN*, *RIN2*, *LINC01551*, *SOX21-AS1*). The 270 down-regulated genes (Fig. 3b, see Supplementary Data 1 for complete list) contained transcription regulators, including *MBD5\**, *FOXQ1*, *MSX2*, *GATA2*; transmembrane proteins: *SLC1A1\**, *GPR37\**, *OXTR\**, *FZD10*; and ligands: *NTF4*, *TGFB2*, *SMAD7*. Gene names with asterisk\* are in SFARI Gene (<https://www.sfari.org/resource/sfari-gene/>) [version of June 20, 2019 released]. The initial



**Figure 3.** Global transcriptome analysis of MAND NPC reveals alterations of multiple genes. **(a)** Scatter plot of fold-change in gene expression in MAND NPC lines versus control NPC lines. The gene expression difference between MAND and CTR NPC lines are shown in log<sub>2</sub> fold changes in y-axis. Genes more highly expressed in MAND NPC than in CTR NPC have a positive fold-change, while the down-regulated genes are illustrated with negative fold change. X-axis represents mean normalized expression. Significantly differentially expressed genes are shown in red. Some representative differentially expressed genes are indicated, including *MBD5* (circled in yellow). **(b)** MAND-associated 2q23.1 deletion results in both positive and negative effects on gene expression. The 468 differentially expressed genes are plotted based on their log<sub>2</sub> fold changes in y-axis. Arrows point to *MBD5* and other genes that are associated with autism spectrum disorder (ASD). **(c)** Heatmap for top 25 differentially expressed genes; the 12 most downregulated and 13 most upregulated genes in the MAND NPC compared with control NPC lines are shown. **(d)** The 468 differentially expressed genes are plotted based on chromosome region. Based on the combined score (Supplementary Data 2, Tab 1), the greatest number of differentially expressed genes were on chromosomes 10, 3, 13, 18, and 21.

498 genes were assessed and converted to unique HGNC approved gene names, resulting in 468 unique genes, which were then used for all downstream analyses. Genome-wide expression differences as a consequence of the 2q23.1 deletion were examined using Enrichr, which sorted the 468 differentially expressed genes by their chromosome region and location. The chromosomes with the greatest number of differentially expressed genes were chromosomes 10, 3, 13, 18, 21, respectively (Fig. 3d, see Supplementary Data 2, Tab 1 for complete list of genes in the chromosome locations).

**MAND is associated with alterations in multiple biological processes and pathways.** To identify the biological processes that are affected by 2q23.1 deletions in MAND NPC, we utilized Gene Set Enrichment Analysis (GSEA), revealing high enrichment in these 5 categories: organ morphogenesis ( $p = 1.24e^{-18}$ ), regulation of cell population and proliferation ( $p = 3.07e^{-16}$ ), neurogenesis ( $p = 3.32e^{-16}$ ), neuron differentiation ( $p = 7.61e^{-15}$ ) and central nervous system development ( $p = 1.52e^{-14}$ ) (Supplementary Data 2, Tab 4). Performing GO analysis with the differentially expressed genes results showed that most genes dysregulated as a consequence of 2q23.1 deletion were involved in transcription regulatory region DNA binding (GO: 1990837 and 0044212) (Supplementary Data 2, Tab 3). To identify the key pathways impacted by the 468 differentially expressed genes, we ran our differential gene list through BIOCARTA and PANTHER pathway databases. Overall, our gene set was enriched in pathways involved in signal transduction, neuronal system, developmental biology, hemostasis, immune system, and disease (Supplementary Fig. 4). Key pathways that contained MAND differentially



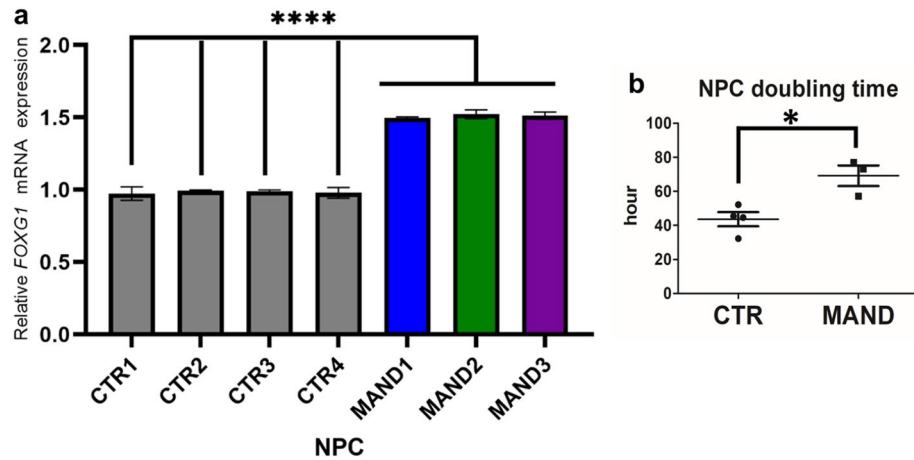
**Figure 4.** Altered cellular pathways and biological processes overlap between MAND and ASD. (a) Differentially expressed genes identified in MAND NPC RNA-seq data were compared to the syndromic and non-syndromic ASD genes in the SFARI database. The genes which are differentially expressed in MAND NPCs show overlap with 20 ASD associated genes and biological processes. In addition, the two gene lists also have their own unique biological processes. (b) Biological processes affected in MAND NPCs overlap those also enriched the SFARI dataset. Utilizing ToppGene, we see significant enrichment for multiple cellular and developmental processes, with overlap with central nervous system development, neurogenesis, neuron differentiation, generation of neurons, and nervous system development with the SFARI gene set. (c) Enrichment of ASD-associated pathways in MAND NPCs. Utilizing DAVID, significantly differentially expressed genes in MAND NPC are enriched in the following KEGG Pathways: spliceosome, MAPK signaling, cell division/DNA replication and WNT signaling. The solid and broken lines represent direct and indirect interactions, respectively. The arrow indicates activation, while a T head shows repression in the interaction.

expressed genes were TGF $\beta$  and Hippo signaling pathways, receptor signaling by FGF and tyrosine kinases, DNA replication, and cell cycle (Fig. 4c, Supplementary Data 3, Tab 5).

### Transcriptional impact of MAND-associated 2q23.1 deletion on expression of autism-associated genes.

We queried if there was overlap between the differentially expressed genes identified via RNA-seq in MAND NPCs and those genes known to be associated with autism. We compared MAND RNA-seq data to syndromic and non-syndromic ASD genes from SFARI (<https://www.sfari.org/resource/sfari-gene/>) (Fig. 4a, Supplementary Data 3, Tab 1–3). Genes differentially expressed in MAND NPCs show overlap with 4.3% of the ASD associated genes from SFARI ( $n = 20$ ). We then compared our list with the SFARI autism gene set utilizing ToppCluster<sup>40</sup>. This analysis revealed striking significant overlap between the two gene lists on biological processes, which are in keeping with the neurological phenotype and ASD, and include forebrain and telencephalon regionalization, neuron fate commitment, and forebrain development (Fig. 4b). Interestingly, processes unique to each gene list were also identified, which may be an area for future studies of pathogenesis for each genetic subgroup. Finally, we discovered that several MAND differentially expressed genes are also involved in overrepresented pathways commonly seen in ASD<sup>18,44,45</sup>, including Hippo (8 genes identified in the pathway, i.e.  $n = 8$ ), MAPK ( $n = 11$ ), DNA replication/cell cycle ( $n = 11$ ), and spliceosome formation pathways ( $n = 7$ ) (Fig. 4c, Supplementary Data 3, Tab 5), providing additional biological connections between MAND and ASD.

**FOXG1 expression is significantly increased in MAND NPC.** In this study, MAND NPC RNA-seq data revealed significantly increased *FOXG1* mRNA levels over the controls, which was confirmed by RT-qPCR (Fig. 5a). *FOXG1* is essential for the normal development of the telencephalon, which is one biological process that is highly enriched in the MAND NPC data set (Supplementary Fig. 3 and Fig. 4b)<sup>46,47</sup>. Pathogenic variants in *FOXG1* are associated with autosomal dominant Rett syndrome, congenital variant [MIM 613454], which is characterized by normal birth but development of progressive microcephaly, severe hypotonia, midline stereotypic activities typical of classic Rett syndrome, no spoken language, and electroencephalography (EEG) abnormalities. *MBD5* (2q23.1 deletion) and *FOXG1* have each been coupled with Angelman syndrome-like



**Figure 5.** MAND-associated 2q23.1 deletion including *MBD5* results in de-repression of *FOXG1*. **(a)** Altered expression of *FOXG1* in MAND NPCs. Quantitative real-time PCR was performed on MAND NPC and control NPC lines. Levels of *FOXG1* mRNA were calculated relative to that of the housekeeping gene *GAPDH*. All expression values were calculated relative to control levels set at 1.0. Error bars represent the SEM. \*\*\*\* $p < 0.0001$ ,  $n = 3$  independent biological replicates performed in triplicate. **(b)** Comparison of cell doubling time between MAND NPC and control NPC lines. Error bars represent the SEM. \* $p = 0.0152$ , GraphPad Prism 8.4.2 (graphpad.com).

(AS-L) cohorts and Rett syndrome-like cohorts (RS-L)<sup>48</sup>. Further, duplication of *FOXG1* is associated with West syndrome, a neurodevelopmental syndrome that includes epilepsy, developmental delays, and severe speech impairment<sup>46,49,50</sup>, similar to the phenotype observed in MAND. Increased *FOXG1* expression was also noted in ASD proband forebrain organoids over the control organoids, and down regulation of *FOXG1* in the organoid rescued the ASD-associated phenotype<sup>51</sup>. We compared the growth of MAND NPCs to CTR NPC and found that the doubling time of MAND NPC was significantly slower than CTR NPC (Fig. 5b). Overall, 2q23.1 deletion results in de-repression of *FOXG1* and other genes, including *SOX11*, *EZH2*, *EED*, *ETV5*, and *KDM1A*, which all function as chromatin modifiers and are implicated in neurogenesis (Supplementary Data 2, Tab 4), further connecting this network of epigenetic modifiers and their critical roles in neurodevelopment.

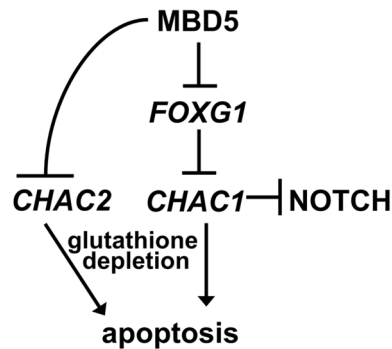
## Discussion

Investigation and understanding of neurodevelopmental disorders has been fueled by the development of iPSC from patients with these complex genetic conditions<sup>14</sup>. In this study, we generated and characterized iPSCs and NPCs that were derived from MAND (2q23.1 deletion syndrome) patients with overlapping phenotypes of ID and autism. By utilizing global transcriptome analysis, we were able to identify differentially expressed genes in early neuronal development that are associated with pathways and genes common to ASD. These findings support previous knowledge about the functional role of *MBD5* but also provide new insights into the consequences of chromosome 2q23.1 deletion in the regulation of numerous pathways that when altered contribute to the phenotype in 2q23.1 deletion syndrome.

In this study, cells were directly collected from patients with MAND and converted to iPSCs and NPCs. This approach allowed us to recapitulate the gene dosage effects in early neuronal developmental stages in human cells. There are advantages of our model compared to previous models to understand MAND. A previous study using an *Mbd5* gene-trap mouse model showed abnormal behavior and cognitive impairment<sup>9</sup>; however, this *Mbd5* gene-trap model is not representative of human 2q23.1 deletion syndrome. Another study utilized primary human neural precursor cells that were treated by a short hairpin RNA (shRNA) to suppress *MBD5* expression<sup>52</sup>; however, this methodology resulted only in a single cell line with reduced *MBD5* expression, limiting the analysis of data. In the most recent study, Seabra et al. (2020) generated CRISPR-derived neurons with mutation of *MBD5* and then conducted RNA-seq<sup>53</sup>. However, *MBD5* was not sufficiently dysregulated in the study; thus, the data presented are not truly representative of MAND<sup>53</sup>. The present study utilizing patient-derived NPC provides a representative disease model, consistent with other recent ASD studies that have shown the advantages of using iPSC technology for the study of genetic impact on human early neuronal developmental stage<sup>54</sup>.

Previous studies have shown that 2q23.1 deletions are critical for neurodevelopment and that *MBD5* haploinsufficiency is responsible for phenotypes, such as seizures, sleep, speech impairment, and autism<sup>5-7,30,55</sup>. These studies fell short in identifying the exact targets of *MBD5* and the perturbed biological processes and pathways consequent to *MBD5* haploinsufficiency that could explain these neurodevelopmental phenotypes. In our study, we recognize that the dysregulated genes and pathways identified may not necessarily be due to direct effects of *MBD5* haploinsufficiency alone and could be due to indirect, downstream effects of reduced *MBD5* expression or may be due to reduced expression of other deleted gene(s) in these patient samples. Additional molecular studies are required to understand the effect of *MBD5* haploinsufficiency specifically on these dysregulated genes. Differentially expressed genes in NPC of MAND patients included genes associated with other neurodevelopmental disorders (Figs. 3, 4). A study of pathogenic de novo mutations that link to intellectual disability identified





**Figure 6.** Schematic model of the possible role of MBD5 in regulation of glutathione, apoptosis, and Notch signaling pathway.

mutations in genes identified in our data set: *FOXG1*, *MBD5*, *TBRI*, *GABRB3* and others<sup>56</sup> (Supplementary Data 3, Tab 3). By the Gene Ontology (GO) enrichment analysis of the transcriptome data, biological functions of neuron development, brain development and cellular processes were altered in the cohort in alignment with the neurodevelopmental phenotype present in these individuals (Supplementary Fig. 3).

In this study, we took advantage of the ability to derive NPCs from MAND patient cells and RNA-seq analysis to investigate the role of MAND on ASD-associated genes and pathways (Figs. 3, 4). Studies of ASD patient-derived neuronal cells have demonstrated altered gene expression and dysregulation of pathways such as Hippo, MAPK, DNA replication/cell cycle, and spliceosome formation pathways<sup>57,58</sup>. While our study may have been hindered by small sample size, our analysis using the SFARI database identified 20 genes (4.3%) linked to ASD with altered gene expression due to 2q23.1 deletion, highlighting the commonalities between the MAND and ASD (Fig. 4a). This percentage is similar to the enrichment for SFARI genes in other key ASD RNA-seq studies<sup>44,57–59</sup>. These genes could be the key to understanding the connection between MAND and autism. Further, six key biological processes are shared between our data set and genes with known association to ASD (Fig. 4b), suggesting that these biological processes are key to understanding the overlapping phenotypes between MAND and ASD. Further study of these biological processes is needed to more clearly elucidate the physiological mechanisms that are altered.

The MAND-derived NPC RNA-seq data revealed differentially expressed genes previously linked to neurodevelopmental disorders. Some of the most interesting connections in phenotypically similar neurodevelopmental disorders include the *FOXG1* dosage-linked associations with Rett syndrome and West syndrome. Our findings in this study indicate that MBD5 either directly or indirectly regulates *FOXG1* expression in an early neurogenesis stage (Fig. 5). Mutation and/or altered expression of *FOXG1* have been shown to cause an autosomal dominant form of Rett syndrome<sup>60</sup>, associated with ASD<sup>59</sup>, West syndrome<sup>50</sup>, and schizophrenia<sup>61</sup>. Since *FOXG1* is a dosage sensitive gene, its overexpression would likely affect proper brain development and cognitive functions. Furthermore, this may represent a possible route for convergent phenotypes between MAND and other syndromes, including West syndrome associated with duplication of *FOXG1*<sup>50</sup>. The slow growth of MAND NPCs (Fig. 5b) may represent a link between *FOXG1* dysregulation and improper glutathione synthesis. *CHAC2* expression is also altered in our dataset (Supplementary Data 1). *CHAC2* is thought to be essential for glutathione maintenance<sup>62,63</sup>. Interestingly, *FOXG1* plays a role in the transcriptional repression of *CHAC1* which is in the same family as *CHAC2* and also plays a role in glutathione maintenance (Fig. 6). *CHAC1* is thought to be a pro-apoptotic factor and inhibitor of NOTCH signaling in glioblastoma cell lines<sup>64</sup>. Further, *FOXG1* is thought to play a role in glutathione maintenance and glutathione to regulate cell proliferation<sup>65,66</sup>. Future molecular studies to assess MBD5 regulation of *FOXG1* is necessary especially in relation to glutathione synthesis and Notch signaling (Fig. 6).

Here, our study of MAND with human neural in vitro cultures through patient-derived NPC demonstrates *FOXG1* upregulation (Fig. 5), as well as perturbed expression of genes implicated in brain development, especially forebrain regionalization (Fig. 4b), and affected genes significantly overlap with ASD associated genes (Fig. 4a). These findings expand our knowledge about gene network differences and possible interactions between these related disease pathways (Figs. 4c, 6). Such findings may lead to potential connections between overlapping neurocognitive and neuropsychiatric phenotypes and associated risk factors, such as other chromatin modifiers and epigenetic factors that link to ASD. We expect these in vitro models will nurture hypotheses toward therapeutic interventions<sup>67,68</sup>.

### Web resources

The URLs for data present herein are as follows:

AutismKB, <http://autismkb.cbi.pku.edu.cn/>.

Cytoscape, <http://www.cytoscape.org/>.

DAVID Bioinformatics Resources 6.7, <http://david.abcc.ncifcrf.gov/>.

Enrichr, <http://amp.pharm.mssm.edu/Enrichr/>.

GSEA (Gene Set Enrichment Analysis), <https://www.gsea-msigdb.org/gsea/index.jsp>.

KEGG, <http://www.genome.jp/kegg/>.

Online Mendelian Inheritance in Man (OMIM), <http://www.omim.org/>.  
 PANTHER, <http://www.pantherdb.org>.  
 SFARI Human Gene Module (accessed February 18, 2020), [https://gene.sfari.org/autdb/HG\\_home.do](https://gene.sfari.org/autdb/HG_home.do).  
 ToppCluster <https://toppgene.cchmc.org/>.  
 UCS Genome Browser, <http://genome.ucsc.edu>.

## Data availability

The RNA-seq data are available in the Gene Expression Omnibus (GSE141835).

Received: 24 June 2020; Accepted: 20 April 2021

Published online: 28 May 2021

## References

- Devlin, B. & Scherer, S. W. Genetic architecture in autism spectrum disorder. *Curr. Opin. Genet. Dev.* **22**, 229–237. <https://doi.org/10.1016/j.gde.2012.03.002> (2012).
- Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221. <https://doi.org/10.1038/nature13908> (2014).
- Tammimies, K. *et al.* Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *JAMA* **314**, 895–903. <https://doi.org/10.1001/jama.2015.10078> (2015).
- Miller, D. T. *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* **86**, 749–764. <https://doi.org/10.1016/j.ajhg.2010.04.006> (2010).
- Mullegama, S. V. & Elsea, S. H. Clinical and molecular aspects of MBD5-associated neurodevelopmental disorder (MAND). *Eur. J. Hum. Genet.* **24**, 1376. <https://doi.org/10.1038/ejhg.2016.77> (2016).
- Mullegama, S. V. *et al.* Reciprocal deletion and duplication at 2q23.1 indicates a role for MBD5 in autism spectrum disorder. *Eur. J. Hum. Genet.* **22**, 57–63. <https://doi.org/10.1038/ejhg.2013.67> (2014).
- Talkowski, M. E. *et al.* Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. *Am. J. Hum. Genet.* **89**, 551–563. <https://doi.org/10.1016/j.ajhg.2011.09.011> (2011).
- Ishizuka, K. *et al.* Investigation of single-nucleotide variants in MBD5 associated with autism spectrum disorders and schizophrenia phenotypes. *Nagoya J. Med. Sci.* **78**, 465–474. <https://doi.org/10.18999/nagjms.78.4.465> (2016).
- Camarena, V. *et al.* Disruption of Mbd5 in mice causes neuronal functional deficits and neurobehavioral abnormalities consistent with 2q23.1 microdeletion syndrome. *EMBO Mol. Med.* **6**, 1003–1015. <https://doi.org/10.15252/emmm.201404044> (2014).
- Tao, Y. *et al.* MBD5 regulates iron metabolism via methylation-independent genomic targeting of Fth1 through KAT2A in mice. *Br. J. Haematol.* **166**, 279–291. <https://doi.org/10.1111/bjh.12863> (2014).
- Clowry, G., Molnar, Z. & Rakic, P. Renewed focus on the developing human neocortex. *J. Anat.* **217**, 276–288. <https://doi.org/10.1111/j.1469-7580.2010.01281.x> (2010).
- Dragunow, M. The adult human brain in preclinical drug development. *Nat. Rev. Drug Discov.* **7**, 659–666. <https://doi.org/10.1038/nrd2617> (2008).
- Dolmetsch, R. & Geschwind, D. H. The human brain in a dish: the promise of iPSC-derived neurons. *Cell* **145**, 831–834. <https://doi.org/10.1016/j.cell.2011.05.034> (2011).
- Fink, J. J. & Levine, E. S. Uncovering true cellular phenotypes: using induced pluripotent stem cell-derived neurons to study early insults in neurodevelopmental disorders. *Front. Neurol.* **9**, 237. <https://doi.org/10.3389/fneur.2018.00237> (2018).
- Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468**, 443–446. <https://doi.org/10.1038/nature09544> (2010).
- Huang, G. *et al.* Uncovering the functional link between SHANK3 deletions and deficiency in neurodevelopment using iPSC-derived human neurons. *Front. Neuroanat.* <https://doi.org/10.3389/fnana.2019.00023> (2019).
- Świtońska, K. *et al.* Identification of altered developmental pathways in human juvenile HD iPSC with 71Q and 109Q using transcriptome profiling. *Front. Cell. Neurosci.* <https://doi.org/10.3389/fncel.2018.00528> (2019).
- DeRosa, B. A. *et al.* Convergent pathways in idiopathic autism revealed by time course transcriptomic analysis of patient-derived neurons. *Sci. Rep.* **8**, 8423. <https://doi.org/10.1038/s41598-018-26495-1> (2018).
- Deneault, E. *et al.* CNTN5-/-or EHMT2-/-human iPSC-derived neurons from individuals with autism develop hyperactive neuronal networks. *Elife* **8**, e40092. <https://doi.org/10.7554/eLife.40092> (2019).
- Rapraeger, A. C., Guimond, S., Krufka, A. & Olwin, B. B. Regulation by heparan sulfate in fibroblast growth factor signaling. *Methods Enzymol.* **245**, 219–240 (1994).
- Sheridan, M. A. *et al.* Early onset preeclampsia in a model for human placental trophoblast. *Proc. Natl. Acad. Sci.* **116**, 4336–4345. <https://doi.org/10.1073/pnas.1816150116> (2019).
- Cheung, S. W. *et al.* Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **7**, 422–432. <https://doi.org/10.1097/01.gim.0000170992.63691.32> (2005).
- Boone, P. M. *et al.* Detection of clinically relevant exonic copy-number changes by array CGH. *Hum. Mutat.* **31**, 1326–1342. <https://doi.org/10.1002/humu.21360> (2010).
- Wiszniewska, J. *et al.* Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. *Eur. J. Hum. Genet.* **22**, 79–87. <https://doi.org/10.1038/ejhg.2013.77> (2014).
- Mullegama, S. V. & Elsea, S. H. Intragenic MBD5 familial deletion variant does not negatively impact MBD5 mRNA expression. *Mol. Cytogenet.* **7**, 80. <https://doi.org/10.1186/s13039-014-0080-9> (2014).
- Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T. & Yamanaka, S. Hypoxia enhances the generation of induced pluripotent stem cells. *Cell Stem Cell* **5**, 237–241. <https://doi.org/10.1016/j.stem.2009.08.001> (2009).
- Okita, K. *et al.* A more efficient method to generate integration-free human iPSC cells. *Nat. Methods* **8**, 409–412. <https://doi.org/10.1038/nmeth.1591> (2011).
- Lee, K. *et al.* Engraftment of human iPSC cells and allogeneic porcine cells into pigs with inactivated RAG2 and accompanying severe combined immunodeficiency. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7260–7265. <https://doi.org/10.1073/pnas.1406376111> (2014).
- Ludwig, T. E. *et al.* Feeder-independent culture of human embryonic stem cells. *Nat. Methods* **3**, 637–646. <https://doi.org/10.1038/nmeth902> (2006).
- Mullegama, S. V., Alaimo, J. T., Chen, L. & Elsea, S. H. Phenotypic and molecular convergence of 2q23.1 deletion syndrome with other neurodevelopmental syndromes associated with autism spectrum disorder. *Int. J. Mol. Sci.* **16**, 7627–7643. <https://doi.org/10.3390/ijms16047627> (2015).
- Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. <https://doi.org/10.1093/bioinformatics/bts635> (2013).

32. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. <https://doi.org/10.1093/bioinformatics/btt656> (2013).
33. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550–550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
34. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–114. <https://doi.org/10.1093/nar/gkr988> (2012).
35. Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33**, D284–288. <https://doi.org/10.1093/nar/gki078> (2005).
36. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–622. <https://doi.org/10.1093/nar/gkn863> (2009).
37. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* **23**, 3251–3253. <https://doi.org/10.1093/bioinformatics/btm369> (2007).
38. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **14**, 128. <https://doi.org/10.1186/1471-2105-14-128> (2013).
39. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550. <https://doi.org/10.1073/pnas.0506580102> (2005).
40. Kaimal, V., Bardes, E. E., Tabar, S. C., Jegga, A. G. & Aronow, B. J. ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res.* **38**, W96–W102. <https://doi.org/10.1093/nar/gkq418> (2010).
41. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
42. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPSC cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280. <https://doi.org/10.1038/nbt.1529> (2009).
43. Boyer, L. F., Campbell, B., Larkin, S., Mu, Y. & Gage, F. H. Dopaminergic differentiation of human pluripotent cells. *Curr. Protoc. Stem Cell Biol.* <https://doi.org/10.1002/9780470151808.sc01h06s22> (2012).
44. Griesi-Oliveira, K. *et al.* Transcriptome of iPSC-derived neuronal cells reveals a module of co-expressed genes consistently associated with autism spectrum disorder. *Mol. Psychiatry* <https://doi.org/10.1038/s41380-020-0669-9> (2020).
45. Kumar, S. *et al.* Impaired neurodevelopmental pathways in autism spectrum disorder: a review of signaling mechanisms and crosstalk. *J. Neurodev. Disord.* **11**, 10. <https://doi.org/10.1186/s11689-019-9268-y> (2019).
46. Hettige, N. C. & Ernst, C. FOXG1 dose in brain development. *Front. Pediatr.* <https://doi.org/10.3389/fped.2019.00482> (2019).
47. Hou, P.-S., hAilín, D. Ó., Vogel, T. & Hanashima, C. Transcription and beyond: delineating FOXG1 function in cortical development and disorders. *Front. Cell. Neurosci.* <https://doi.org/10.3389/fncel.2020.00035> (2020).
48. Tan, W. H., Bird, L. M., Thibert, R. L. & Williams, C. A. If not Angelman, what is it? A review of Angelman-like syndromes. *Am. J. Med. Genet. A* **164A**, 975–992 (2014).
49. Brunetti-Pierri, N. *et al.* Duplications of FOXG1 in 14q12 are associated with developmental epilepsy, mental retardation, and severe speech impairment. *Eur J Hum Genet* **19**, 102–107. <https://doi.org/10.1038/ejhg.2010.142> (2011).
50. Striano, P. *et al.* West syndrome associated with 14q12 duplications harboring FOXG1. *Neurology* **76**, 1600–1602. <https://doi.org/10.1212/WNL.0b013e3182194bbf> (2011).
51. Nolan, N. *The Role Of Foxg1 Overexpression In The Pathogenesis Of Severe, Macrocephalic Autism Spectrum Disorder* Medical Doctor (MD) thesis, Yale University, (2018).
52. Gigeck, C. O. *et al.* A molecular model for neurodevelopmental disorders. *Transl. Psychiatry* **5**, e565. <https://doi.org/10.1038/tp.2015.56> (2015).
53. Seabra, C. M. *et al.* Transcriptional consequences of MBD5 disruption in mouse brain and CRISPR-derived neurons. *Mol Autism* **11**, 45. <https://doi.org/10.1186/s13229-020-00354-1> (2020).
54. Gordon, A. & Geschwind, D. H. Human in vitro models for understanding mechanisms of autism spectrum disorder. *Mol. Autism* **11**, 26–26. <https://doi.org/10.1186/s13229-020-00332-7> (2020).
55. Williams, S. R. *et al.* Haploinsufficiency of MBD5 associated with a syndrome involving microcephaly, intellectual disabilities, severe speech impairment, and seizures. *Eur. J. Hum. Genet.* **18**, 436–441. <https://doi.org/10.1038/ejhg.2009.199> (2010).
56. Hamdan, F. F. *et al.* De novo mutations in moderate or severe intellectual disability. *PLoS Genet.* **10**, e1004772. <https://doi.org/10.1371/journal.pgen.1004772> (2014).
57. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021. <https://doi.org/10.1016/j.cell.2013.10.031> (2013).
58. Schafer, S. T. *et al.* Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. *Nat. Neurosci.* **22**, 243–255. <https://doi.org/10.1038/s41593-018-0295-x> (2019).
59. Mariani, J. *et al.* FOXG1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. *Cell* **162**, 375–390. <https://doi.org/10.1016/j.cell.2015.06.034> (2015).
60. Ariani, F. *et al.* FOXG1 is responsible for the congenital variant of Rett syndrome. *Am. J. Hum. Genet.* **83**, 89–93. <https://doi.org/10.1016/j.ajhg.2008.05.015> (2008).
61. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527. <https://doi.org/10.1038/nature19847> (2016).
62. Bachhawat, A. K. & Kaur, A. Glutathione degradation. *Antioxid Redox Signal* **27**, 1200–1216. <https://doi.org/10.1089/ars.2017.7136> (2017).
63. Wang, C. K. *et al.* CHAC2 is essential for self-renewal and glutathione maintenance in human embryonic stem cells. *Free Radic. Biol. Med.* **113**, 439–451. <https://doi.org/10.1016/j.freeradbiomed.2017.10.345> (2017).
64. Dali, R., Verginelli, F., Pramatarova, A., Sladek, R. & Stifani, S. Characterization of a FOXG1:TLE1 transcriptional network in glioblastoma-initiating cells. *Mol. Oncol.* **12**, 775–787. <https://doi.org/10.1002/1878-0261.12168> (2018).
65. Markovic, J. *et al.* The depletion of nuclear glutathione impairs cell proliferation in 3t3 fibroblasts. *PLoS ONE* **4**, e6413. <https://doi.org/10.1371/journal.pone.0006413> (2009).
66. Pallardó, F. V., Markovic, J., García, J. L. & Viña, J. Role of nuclear glutathione as a key regulator of cell proliferation. *Mol. Asp. Med.* **30**, 77–85. <https://doi.org/10.1016/j.mam.2009.01.001> (2009).
67. Fujimori, K. *et al.* Modeling sporadic ALS in iPSC-derived motor neurons identifies a potential therapeutic agent. *Nat. Med.* **24**, 1579–1589. <https://doi.org/10.1038/s41591-018-0140-5> (2018).
68. Ishizuka, K. *et al.* PM338 Investigation of the association of rare single nucleotide variants in methyl-CpG-binding domain protein 5 (MBD5) with phenotypes of autism spectrum disorders and schizophrenia. *Int. J. Neuropsychopharmacol.* **19**, 23–24. <https://doi.org/10.1093/ijnp/pyw041.338> (2016).

## Acknowledgements

We are grateful to the study participants and their families for their cooperation in this study. We thank the N.J. Bivens in the DNA core at University of Missouri, Columbia (UMC) for RNA sequencing and Dr. Janice Smith and her team for assistance with CMAs. The work was supported, in part, by the resources from University of

California, Los Angeles, Bond Life Sciences Center at UMC, Baylor College of Medicine, University of Michigan, and Jan and Duncan Neurological Research Institute at the Texas Children's Hospital. JWI acknowledges support from the Morton S. and Henrietta K. Sellner Professorship in Human Genetics. The PAX6 monoclonal antibody developed by Kawakami, A. with the N-terminal region, aa 1-223 was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242. The following cell lines were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research: GM05659 and GM24585. We thank Mr. Dennis Reith for editorial assistance.

### Author contributions

S.V.M., S.H.E., and T.E. were involved in the conception and design of this study. J.W.I., F.J.P., C.H.E. identified and referred the patients, clinical samples, and clinical reports. S.V.M. and S.H.E. were involved in collection of patient data and samples. S.V.M. was involved in the growing of cell lines and expression studies. T.E. generated the iPSCs and NPCs and Y.Y. helped with the immunostaining. RNA-seq studies were conducted by DNA core at University of Missouri, Columbia. RNA-seq raw data was analyzed by S.R.W. and S.V.M. Interpretation of the filtered RNA-seq data was done by S.V.M., S.D.K., and J.A.M.-A. Figures were designed by S.V.M., S.D.K. and T.E. S.V.M., S.D.K., S.H.E., T.E. were involved in the writing of the manuscript. S.V.M., S.H.E., T.E., S.R.W., S.D.K., J.A.M.-A., J.W.I., F.J.P., and C.H.E. all critically reviewed and edited the final version of the manuscript.

### Funding

This work was supported by grants from Patton Trust Research Grant program at the Kansas City Area Life Sciences Institute (TE), Fondation Jérôme Lejeune (SHE), March of Dimes (Grant #6-FY12-324, JAM-A), UCLA Children's Discovery Institute, UCLA CART (NIH/NICHD Grant# P50-HD-055784, JAM-A), NIH/NCATS UCLA CTSI (Grant # UL1TR000124, JAM-A), Autism Speaks Grant #9172 (SK) and the UCLA-Caltech MSTP NIH T32GM008042 (SK).

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-90798-z>.

**Correspondence** and requests for materials should be addressed to S.H.E. or T.E.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021