Genome Medicine

RESEARCH

Open Access

Contribution of rare coding variants to microcephaly in individuals with neurodevelopmental disorders

Jihoon G. Yoon[1,2†], Hyunsoo Jang[3,4†], Seungbok Lee[1,5], Se Song Jang[5], Soojin Park[5], Jaeso Cho[1,5,6], Minji Kim[3,4], Jiye Han[3,4], Hyounji Yun[3,4], Man Jin Kim[1], Soo Yeon Kim[1,5], Woo Joong Kim[5], Anna Cho[6], Jin Sook Lee[7], Murim Choi[8], Alberto Fernandez-Jaen[9], Sebastian Silva[10,11], Reinaldo Uribe-San-Martín[12,13], Christian Cantillano[13,14], Noriko Miyake[15], Byung Chan Lim[5], Jung Min Ko[5], Ki Joong Kim[5], Ki-Jun Yoon[3*†] and Jong-Hee Chae[1,5*†]

Abstract

Background Microcephaly, characterized by an abnormally small head size, frequently co-occurs with neurodevelopmental disorders (NDDs). While the genetic basis of NDDs has been widely investigated, the contribution of rare coding variants to microcephaly remains poorly understood.

Methods We investigated the relationships between head circumference and rare coding variants in 418 individuals with microcephaly, analyzing data from 1050 exomes (312 trios and 106 proband-only samples). Participants were classified into primary microcephaly (PM) and secondary microcephaly (SM) groups, and their clinical and genetic characteristics were systematically assessed. The functional impact of high-priority candidate genes, *RTF1* and *ASAP2*, was further validated using neural progenitor cells (NPCs) and human forebrain organoid models.

Results Exome sequencing revealed 142 causative and 12 candidate genes associated with microcephaly. Pathway analyses indicated that PM genes are linked to early phases of brain development, whereas SM genes are more associated with later stages of neuronal maturation. In addition, the PM group had a significantly higher proportion of autosomal recessive disorders and exhibited more severe microcephaly than the SM group. Notably, females displayed greater microcephaly severity than males, primarily attributable to differences in the origin of the allele and inheritance patterns on the X chromosome. Functional experiments using CRISPR-Cas9 knockout in NPCs and brain organoids demonstrated reduced NPC proliferation, supporting the essential role of *RTF1* and *ASAP2* in brain development.

Conclusions This study sheds light on the complex genetic architecture of microcephaly, emphasizing the impact of rare coding variants on brain development and delineating distinct clinical and molecular profiles underlying PM and SM.

Keywords Microcephaly, Head circumference, Neurodevelopmental disorder, Exome sequencing, Brain organoids

†Jihoon G. Yoon and Hyunsoo Jang contributed equally to this work.

†Jong-Hee Chae and Ki-Jun Yoon jointly directed this work.

*Correspondence:
Ki-Jun Yoon
kijunyoon@kaist.ac.kr
Jong-Hee Chae
chaeped1@snu.ac.kr
Full list of author information is available at the end of the article

Background

The brain is the most sophisticated organ in the body, having undergone significant evolution throughout our species' history. This evolutionary process has enabled *Homo sapiens* to develop a uniquely high level of intelligence [1]. However, there is considerable variation in brain size within and between human populations. Microcephaly, or "small head," represents the extreme end of this variation, with affected individuals generally having a head circumference (HC) more than two standard deviations below the average for their matched age and sex.

Microcephaly has long been of interest to researchers as a disease model for studying brain development [2], offering valuable insights into the genetic and molecular pathways that govern brain growth. This condition can be categorized into two types: primary microcephaly (PM), when present at birth, and secondary microcephaly (SM), when it develops postnatally [3]. Recent advancements have significantly expanded our understanding of the genes and molecular mechanisms involved in microcephaly [2]. PM is frequently associated with autosomal recessive (AR) inheritance, where defects in mitotic spindle formation or cell division play a pivotal role in the disease's pathogenesis [4]. In contrast, SM is often linked to autosomal dominant (AD) inheritance and disruptions in later developmental processes, such as myelination, synapse formation, or gene expression regulation in the forebrain and hindbrain [5, 6].

To date, 1340 genes have been associated with microcephaly (HP:0000252) in the human phenotype ontology (HPO) database (v2.0.4) [7], indicating complex genetic regulation during brain development. Genetic studies have shown that microcephaly can arise from a wide range of genetic conditions, many of which impair the function of key proteins essential for brain development [8]. These proteins are involved in a variety of molecular pathways that control the proliferation and differentiation of neural stem cells as well as the migration and maturation of neurons and glial cells. The major molecular mechanisms responsible for microcephaly include DNA repair, centriole biogenesis, mitotic spindle formation, chromatin condensation, transcriptional regulation, Wnt/β-catenin signaling, and nuclear envelope pathways [9].

Recent advancements in sequencing technologies have accelerated the identification of the genetic architecture underlying numerous neurodevelopmental disorders (NDDs), including developmental disorders (DD), intellectual disability (ID), autism spectrum disorder (ASD), craniosynostosis, and microcephaly [4, 5, 10–15]. Nonetheless, the genetic determinants of small head sizes remain largely unexplored. Here, we investigated the association between rare variants and head sizes in 418 individuals with NDDs by measuring HCs, conducting bioinformatic analyses, performing thorough genotype–phenotype matchings, and demonstrating the functional roles of high-priority candidate genes through assays in human neural progenitor cells and brain organoids.

Methods

Study cohort

The study participants, who visited outpatient clinics at the Rare Disease Center of Seoul National University Children's Hospital and Seoul National University Bundang Hospital located in Seoul and Seongnam, Republic of Korea, between April 2019 and November 2022, were retrospectively enrolled (Additional file 2: Table S1). Inclusion criteria consisted of (1) individuals with rare diseases who underwent exome sequencing for diagnostic evaluation, (2) individuals with NDDs under 18 years of age and an occipitofrontal circumference lower than 3 percentiles for sex- and age-matched references at birth or later [16], and (3) probands with microcephaly not resulting from acquired causes, such as perinatal infections (e.g., Zika virus infection) or hypoxic brain damage. Individuals with positive results from previous genetic evaluations, such as chromosomal microarray (CMA) or *MECP2* sequencing, were excluded. During the study period, 2053 individuals with rare diseases, mostly suspected to have neurogenetic conditions, visited our clinics and underwent exome sequencing for genetic evaluation. Of these, 418 unrelated Korean participants were enrolled who met our inclusion criteria. Exome sequencing was conducted on these participants and their parents, with trio sequencing available for 75.6% of cases. In addition, we used ancestry-matched Korean population controls, namely the Korean Variant Archive 2 (KOVA2) dataset, for genetic analyses, consisting of 5,305 unrelated and healthy individuals [17]. The diagnostic yield of exome sequencing, along with the associations between HCs and various clinical and genomic factors—including mode of inheritance, variant types, and frequently altered genes—were evaluated in this cohort.

Exome and genome sequencing

Whole blood was obtained from the probands and their parents if available. Genomic DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen) following the manufacturer's protocol. A SureSelect XT Human All Exon V5/V6 Kit (Agilent Technologies Inc.) was used for hybridization. Genomic DNA samples were sequenced using the NovaSeq 6000 system (Illumina) with a targeted mean depth of 50X or above. Sequenced reads were aligned to the human reference genome GRCh38 (hg38)

using the Burrow-Wheeler Aligner mem (0.7.15). We adopted the WDL Analysis Research Pipelines (WARP, https://broadinstitute.github.io/warp/; Broad Institute) for reproducible analysis using the Exome Germline Single Sample pipeline (v3.0.4) [18]. Variant calling and joint genotyping were performed using Deepvariant and GLnexus [19]. We only included variants classified into the PASS category for further analysis. Sampling error checks and quality control for sex, relatedness, and ancestry were performed using *Peddy* (Additional file 1: Fig. S1) [20]. For genotyping, we requested quality control criteria consisting of (1) a minimum depth (DP) ≥ 8, (2) Phred-scaled genotype quality (GQ) ≥ 20, and (3) allelic balance between 0.2 and 0.8 for heterozygote calls and above 0.9 for homozygote or hemizygote calls. We utilized Ensembl Variant Effect Predictor (VEP) [21] for variant annotation with various databases: dbNSFP v4.2a [22], gnomAD exome v4.0 [23], KOVA2 [17], Online Mendelian Inheritance in Man (OMIM) [24], InterVar [25], and SpliceAI [26]. Genome sequencing was utilized for the individual (MCPH171) to determine the precise breakpoint of the *SMPD4* homozygous deletion (Additional file 1: Fig. S4). Libraries were prepared using the TruSeq DNA PCR-Free High Throughput Library Prep Kit (Illumina) following the manufacturer's instructions. Bioinformatics analysis was conducted with the Whole Genome Germline Single Sample pipeline (v2.3.3) provided by WARP and was utilized following the Functional Equivalence specifications [27].

Variant interpretation

Variant interpretation and genotype–phenotype correlations were discussed during weekly multidisciplinary board meetings, which included pediatricians, geneticists, laboratory specialists, and bioinformaticians. We screened approximately 7000 Mendelian forms of genetic disorders in the OMIM database [24]. For AD disorders (Additional file 2: Table S2), candidate variants were selected if they occurred de novo or were reported to be pathogenic (P) or likely pathogenic (LP) in the ClinVar database [28]. For AR disorders (Additional file 2: Table S3), the *trans*-conformation of candidate variants was confirmed in trios or proband-only samples using Sanger sequencing with additional parental sampling. A similar approach was applied to X-linked dominant (XD) and recessive (XR) disorders (Additional file 2: Table S4). Consequently, P/LP variants, classified in the ClinVar database or by the American College of Medical Genetics and Genomics (ACMG)/Association for Molecular Pathology (AMP) criteria [29], were considered. Following the identification of causative variants, we conducted phenotype matching—including reverse phenotyping—to achieve final diagnoses, aligning clinical presentations with the established clinical profiles of candidate genes. To maximize diagnostic yield, we further investigated potential phenotypic expansions for genes or disorders of particular interest through targeted literature reviews. These reviews involved systematic searches in PubMed and Google Scholar using gene names, variant types, and associated phenotypic terms. We prioritized recent peer-reviewed case reports, cohort studies, and functional studies published within the past five years. Articles were evaluated for evidence of phenotypic expansion, novel genotype–phenotype correlations, or previously unreported clinical presentations. All findings from the literature were discussed during multidisciplinary board meetings to assess whether the candidate variant could plausibly explain the observed phenotype in our cohort. Final interpretations were made based on a consensus of literature evidence and comprehensive clinical assessment.

De novo and copy-number variation (CNV) analysis

De novo variant analysis was conducted using denovolyzeR for 316 trios [30], with stringent criteria to ensure the identification of true de novo events (Additional file 1: Fig. S2, Additional file 2: Table S5). Initially, candidate de novo variants were filtered to be heterozygous in probands, reference genotypes in both parents, and have a population allele frequency (AF) < 0.01% in both the gnomAD and KOVA2 databases. Only high-confidence sites were included, defined by a DP ≥ 10, GQ ≥ 20, and an allelic balance between 0.3 and 0.7 in probands, and allelic balance < 0.03 in both parents. Sanger sequencing was used to validate suspected de novo variants identified in proband-only samples, with additional parental samples when available. For CNV detection, we employed two complementary methods: HMZDelFinder for small homozygous and hemizygous deletions [31], and CoNIFER for heterozygous deletions or duplications [32]. The functional impact of identified CNVs was predicted using AnnotSV [33], focusing on variants classified as P/LP variants by the ACMG/ClinGen guidelines [34]. Clinically significant pathogenic CNVs were validated using additional methods, including CMA, methylation-specific multiplex ligation-dependent probe amplification, quantitative PCR, or whole genome sequencing (Additional file 2: Table S6).

Gene dosage PCR for *SMPD4* deletion

To assess the copy number status of the *SMPD4* deletion in the MCPH171 family (Additional file 1: Fig. S4), we employed a gene dosage PCR method with minor modifications based on previously established protocols [35]. We designed two sets of primers targeting exon 9 and exon 19 of *SMPD4* (Additional file 2: Table S7). The PCR

reactions were conducted under the following cycling conditions: initial denaturation and activation at 94 ℃ for 5 min, followed by 18 cycles of amplification at 94 ℃ for 30 s, 60 ℃ for 30 s, and 72 ℃ for 1 min. The PCR concluded with a final extension at 72 ℃ for 5 min.

Prioritization of candidate genes

To uncover novel gene-disease associations within our cohort, we implemented a modified strategy based on a previous report [36]. We focused on ultra-rare coding variants with minor allele frequencies < 0.01% in both the gnomAD and KOVA2 databases, prioritizing variants that were de novo or inherited from both parents for AD or AR disorders, respectively. To enhance the likelihood of identifying functionally relevant variants, we focused on high-impact coding changes, including frameshift, splice junction (± 2 bp), nonsense, and damaging missense variants. Importantly, the criteria used for prioritization were identical to those applied for variant interpretation, ensuring consistency across analytic steps. High-impact variants were further refined using evolutionary and functional prediction scores: variants with GERP scores > 4 and CADD scores > 20 were retained. For missense variants, we additionally applied the AlphaMissense score (> 0.564), indicative of likely pathogenicity [37–39]. For AD disorders, genes with high constraint metrics (probability of being loss-of-function intolerant [pLI] score > 0.9 or loss-of-function observed/expected upper bound fraction [LOEUF] score < 0.6) were selected, as they are less likely to harbor loss-of-function (LOF) variants in the general population [23]. For AR disorders, we selected genes in which affected trios carried two predicted deleterious alleles in trans. Additional cases with *RTF1* variants were identified and recruited via GeneMatcher [40].

sgRNA design and plasmid cloning

Single guide RNAs (sgRNAs) targeting *RTF1* and *ASAP2* genes were designed using the CRISPick (https://portals.broadinstitute.org/gppx/crispick/public) (Additional file 2: Table S7). For each sgRNA, two complementary oligonucleotides containing the sgRNA sequence with overhangs for BbsI restriction enzyme recognition sites were synthesized. The complementary oligonucleotides were annealed and subsequently cloned into the BbsI (NEB, #R3539)-digested pSpCas9(BB)−2A-GFP (PX458; Addgene, #48138) using a Golden Gate cloning method. To generate *RTF1* overexpression plasmids, full-length cDNA sequences were obtained from Ensembl (http://www.ensembl.org) and synthesized (Twist Bioscience). Sequence-specific degenerate codons were introduced at the sgRNA PAM site to confer resistance to sgRNA targeting, and patient-specific *RTF1* variants (p.Arg608Gly and p.Ser558Pro) were incorporated (Additional file 1: Fig. S7F). The synthesized cDNAs were then cloned into the piggyBac transposon-based vector, PBCAG-eGFP (Addgene, #40973), by replacing the EGFP sequence via Gibson Assembly (New England Biolabs, E2611S), according to the manufacturer's protocol.

Maintenance of human ESCs

H9 human embryonic stem cells (ESCs; WA09 cells) obtained from the Wisconsin International Stem Cell Bank (WiCell Research Institute) were used. The cells were cultured in a feeder-dependent culture system using Dulbecco's modified Eagle medium (DMEM)/F12 (Invitrogen) medium supplemented with 20% Knockout™ Serum Replacement (Gibco), 1X Non-essential Amino Acids (Invitrogen), 1X Penicillin/Streptomycin (Invitrogen), 1X 2-Mercaptoethanol (Millipore), 1X Glutamax (Invitrogen), and 10 ng/ml FGF-2 (Peprotech), as previously described [41]. The culture medium was refreshed daily, and cells were passaged every 6 days onto new plates pre-seeded with CF1 mouse embryonic fibroblasts pre-treated with mitomycin C (AG Scientific). For passaging, ESC colonies were fragmented using an insulin syringe and detached from the plate by incubating with 1 mg/ml Collagenase Type IV (Invitrogen) for 5 min.

Generation of forebrain organoids and organoid-derived NPCs

Forebrain-patterned organoids were generated following previously established protocols (Additional file 1: Fig. S7) [42, 43]. Briefly, human ESC colonies were detached 6 days post-passage using 1 mg/ml of Collagenase Type IV and washed with fresh stem cell medium. The detached ESC colonies were then transferred to an Ultra-Low attachment 6-well plate (Corning Costar) containing 3–4 ml of stem cell medium without FGF-2, supplemented with 2 µM each of Dorsomorphine and A83-01 (STEMCELL Technologies). On days 5–6, half of the medium was replaced with neural induction medium, consisting of DMEM/F12, 1X N2 Supplement (Invitrogen), 1X Penicillin/Streptomycin, 1X Non-essential Amino Acids, 1X Glutamax, 1 µM each of CHIR99021 and SB-431542 (STEMCELL Technologies). On day 7, organoids were embedded in hESC-qualified Matrigel (Corning), and neural induction media were refreshed every two days for an additional 6 days. On day 14, the embedded organoids were mechanically dissociated from the Matrigel by gentle pipetting using a 10 ml serological pipette. For long-term culture, organoids were maintained in a differentiation medium containing DMEM/F12, 1X N2 Supplement, 1X B27 Supplement minus vitamin A (Gibco), 1X Penicillin/Streptomycin, 1X Non-essential Amino

Acids, 1X Glutamax, and 2.5 μg/ml Insulin (Sigma), and kept on an orbital shaker at 120 rpm in the incubator.

To generate organoid-derived neural progenitor cells (NPCs), forebrain organoids were dissociated from Matrigel on day 14 and further dissociated into single cells using Accutase (Sigma) for 5 min. The organoids were then mechanically dissociated by pipetting and washed with an induction medium. The dissociated cells were cultured on Matrigel-coated culture dishes in a neural maintenance medium, consisting of DMEM/F12 supplemented with 1X N2 Supplement, 1X B27 Supplement minus vitamin A, 1X Glutamax, 1X Penicillin/Streptomycin, 5 μg/ml of Heparin (Sigma), 20 ng/ml FGF-2, and 20 ng/ml EGF (Peprotech). NPCs were passaged every 5 days onto fresh Matrigel-coated dishes by dissociation into single cells with Accutase.

NPC and brain organoid electroporation

For NPC electroporation, NPCs at 80–90% confluency were detached by incubating with Accutase for 5 min, followed by washing with neural maintenance media and resuspension in Opti-MEM (Thermo Fisher). A total of 6×10^5 cells, counted using the EVE-Plus (NanoEntek inc., Seoul, Korea), were mixed with a combination of the piggyBac donor plasmid PBCAG-eGFP, the helper plasmid CAG-PBase, and px458 containing the sgRNA of interest, with the final volume adjusted to 100 μl. For rescue experiments, sgRNA-resistant cDNA expression plasmids were included in the transfection mixture. The cell and plasmid DNA mixture was then transferred into Nepa electroporation cuvettes with a 2 mm gap (Nepa Gene), and three pulses (150 V, 5 ms duration with 450 ms intervals) were delivered using a square wave electroporator (CUY21SC, Nepa Gene). The electroporated cells were subsequently plated into a Matrigel-coated culture dish in neural maintenance media supplemented with ROCK inhibitor Y-27632 (Selleckchem). The following day, the media was changed to neural maintenance media without the ROCK inhibitor and refreshed daily thereafter.

For brain organoid electroporation, 2-month-old brain organoids were placed in a 100 mm cell culture dish containing differentiation media. A mixture of 0.05% Fast Green and plasmid DNAs was injected into the lumen structures of the organoids using a calibrated micropipette [43]. Electroporation was performed with electrode paddles featuring a 5 mm gap (CUY650-5, Nepa Gene), positioning the positive electrode near the DNA-injected region within the organoid. Five pulses (80 V, 85 ms duration with a 950 ms interval) were applied using square wave electroporation. Afterward, the transfected organoids were transferred back to the orbital shaker in differentiation media for continued culture.

Gene editing efficiency analysis

One day after NPC electroporation, cells were dissociated into single cells by incubating with Accutase for 5 min. The dissociated cells were resuspended in FACS buffer (Enzynomics) and GFP-labeled cells were sorted using a flow cytometer sorter (Sony SH800). Genomic DNA was extracted using DirectPCR Lysis Reagent (Viagen Biotech) and proteinase K (Enzynomics). The target loci of the gene of interest were amplified by PCR (Additional file 2: Table S7), and the resulting amplicons were subjected to sequencing using the MiniSeq system (Illumina). Indel frequencies were analyzed using Cas-Analyzer (http://www.rgenome.net/cas-analyzer/), and gene editing efficiencies were subsequently evaluated.

Immunohistology, confocal imaging, image quantification

For immunocytochemistry of NPC cultures, cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 15 min at room temperature (RT). Fixed samples were blocked and permeabilized using a blocking solution (3% bovine serum albumin, 5% normal donkey serum, and 0.1% Triton X-100 in PBS) for 1 h at RT. Primary antibodies were diluted in the blocking solution and applied for 1 h at RT, followed by washing and incubation with secondary antibodies, also diluted in the blocking solution, for 1 h at RT. Nuclei were counterstained with Hoechst 33,342 solution (Thermo Fisher) for 10 min at RT. For immunohistochemistry of brain organoids, the organoids were fixed with 4% paraformaldehyde in PBS overnight at 4 °C, cryoprotected in 30% sucrose in PBS, and embedded in optimal cutting temperature compound. Sections of 20 μm thickness were obtained using a Leica CM1860 cryostat. Organoid sections were stained following the same immunocytochemistry protocol, with the exception that primary antibodies were applied overnight at 4 °C. A complete list of antibodies is available in Additional file 2: Table S8. Stained sections were mounted using an in-house mounting medium and visualized on a Nikon A1 confocal microscope. Quantification of marker-expressing cells was performed using ImageJ (v.1.54) software.

For image analysis, brain organoids containing well-formed neural rosettes with a distinct progenitor layer ($PAX6^+$, 50–150 μm in diameter) and neuronal layer ($CTIP2^+$) were selected. Among these, two to three organoids per condition were analyzed for quantification, ensuring the presence of GFP^+ cells within the neural rosettes. For each organoid, at least two different neural rosettes were analyzed. The proportion of $Ki\text{-}67^+$ cells among GFP^+ cells was determined by counting the number of $Ki\text{-}67^+GFP^+$ cells and dividing it by the total

number of GFP^+ cells within each rosette. Final quantification was expressed as the average percentage of $Ki67^+$ cells among GFP^+ cells.

Cell cycle analysis

For cell cycle analysis, NPCs were pulsed with 10 μM EdU (Thermo Fisher) for 1 h, dissociated by incubating with Accutase for 5 min, and fixed with 4% paraformaldehyde in PBS for 15 min at RT. The fixed cells were stained using the Click-iT EdU Alexa 647 Flow Cytometry Kit (Thermo Fisher), following the manufacturer's protocol. Cells were then stained with Hoechst 33342 solution to assess DNA content. Samples were analyzed using a BD LSRFortessa™ flow cytometer (BD Biosciences), with GFP^+ cells gated for further analysis. The DNA content and percentages of EdU^+ populations were quantified.

Western blot

For western blot analysis, NPCs were lysed in RIPA buffer (LPS Solution, CBR002) supplemented with protease inhibitor cocktail (Millipore, 535,140-1SET). Lysates were kept on ice for 10 min and centrifuged at 16,000 g for 30 min at 4 °C. The supernatant was collected, mixed with SDS-PAGE loading buffer (LPS Solution, CBS002), and boiled for 5 min. Proteins were separated by SDS-PAGE and transferred to polyvinylidene difluoride membranes. Immunoblotting was performed using appropriate primary antibodies and HRP-conjugated secondary antibodies (Additional file 2: Table S8).

Statistical analysis

Statistical analysis was performed using GraphPad Prism (v9.5.0) and *R* (v4.1.2) software. A P-value < 0.05 was considered significant. All data are presented as *n* (%) or median (range). Categorical variables were compared using a two-tailed Fisher's exact test, and continuous variables were compared using the Mann–Whitney *U* test. Correction for multiple comparisons was carried out using the Bonferroni method with the false discovery rate cutoff < 0.05. HC data were transformed to *Z*-scores based on sex- and age-matched references [16]. HC was compared among the available participants, and missing data were excluded from the analysis. For multiple measurements in an individual during the follow-up period, the lowest *Z*-score was used for statistical analyses. Gene-set enrichment with the gene ontology (GO) database and protein interaction network analysis were performed and visualized using Metascape (https://metascape.org) and Cytoscape (https://cytoscape.org/) [44, 45].

Results

Clinical characteristics

In total, 1050 exomes from 418 individuals (316 trios and 102 proband-only samples) were analyzed (Fig. 1A). The cohort comprised 52.6% female and 47.4% male participants. Quality control of the exome dataset using *Peddy* confirmed that all participants were unrelated, of East Asian ancestry, and showed no sample anomalies or relatedness violations (Additional file 1: Fig. S1). These participants were classified into three groups: 143 with PM (34.2%), 258 with SM (61.7%), and 17 with unknown (4.1%) classification. While the sex distribution did not differ significantly between groups, individuals in the PM group were diagnosed at an earlier age and exhibited more severe microcephaly. Additionally, they had a higher frequency of preterm birth (< 37 weeks) and lower birth weight (< 2500 g) compared to those in the SM group, as summarized in Table 1. Notably, the PM group displayed a higher incidence of comorbidities compared to the SM group, except for ASD. Specifically, the PM group had a higher prevalence of short stature/failure to thrive (59.4% vs. 48.4%, $P = 0.037$) and brain anomaly (59.4% vs 38.4%, $P < 0.001$). In contrast, ASD was more prevalent in the SM group (11.9% vs 22.5%, $P < 0.001$), as depicted in Fig. 1B.

Genomic landscape of microcephaly

In our cohort, we identified a total of 142 genes associated with microcephaly, which showed the genetic heterogeneity (Fig. 2A). Of these, 113 genes had previously been linked to microcephaly (HP:0000252), while 29 genes have not yet been associated with the condition in the HPO database. A detailed review of these genes revealed that there were clinical reports exhibiting microcephaly harboring variants in 20 genes, whereas the remaining 9 had no prior association with the microcephaly phenotype (Additional file 2: Table S9). Specifically, 51, 104, and 7 genes were identified in the PM, SM, and unknown groups, respectively. Among these, 14 genes were shared between the PM and SM subtypes. Exome sequencing provided an overall diagnostic yield of 56.7%, with 237 out of 418 participants obtaining a genetic diagnosis (Fig. 2B). Diagnostic yields were comparable between the PM and SM groups, with 53.8% and 59.7% of PM and SM cases achieving a diagnosis, respectively. Our results aligned well with the previous report [5], showing that the PM group exhibited a higher prevalence of AR inheritance (24.7 vs. 5.2%), while the SM group more frequently had AD inheritance (37.7 vs. 68.2%; Fig. 2C).

Causative variants were found predominantly in AD (33.0%), followed by X-linked (XL; 9.6%) and AR

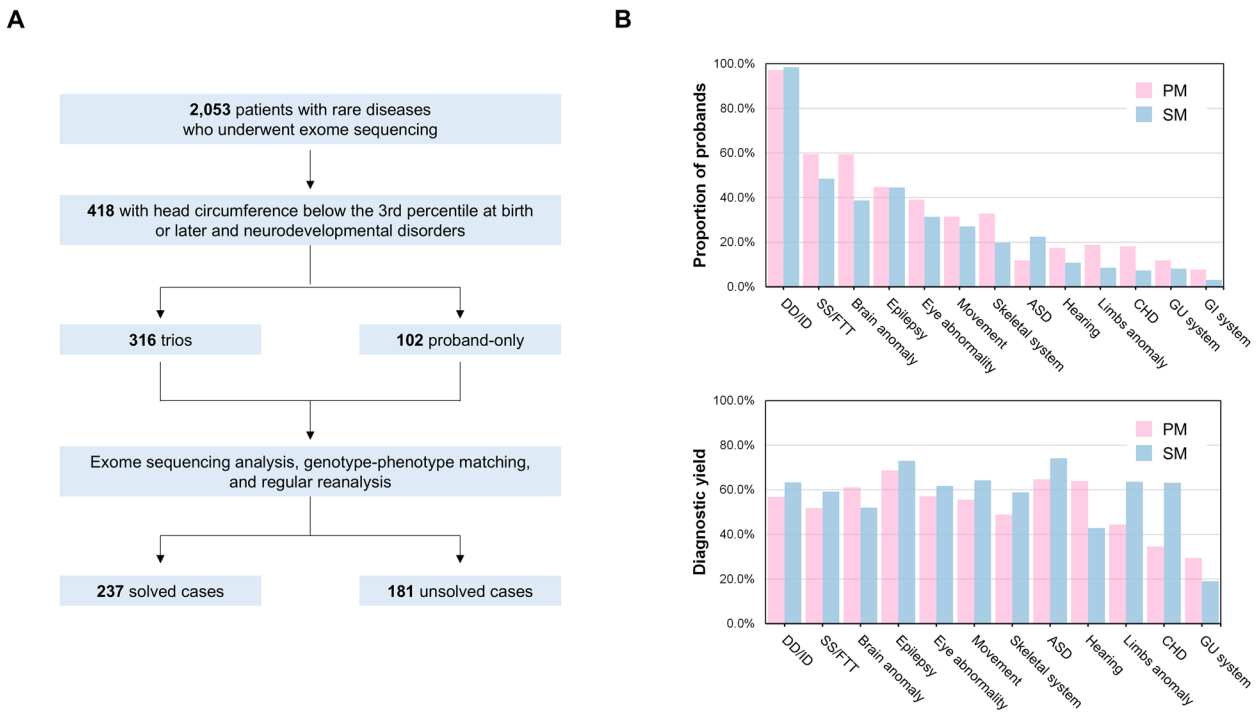

Fig. 1 Study cohort and clinical spectrum. **A** Overview of the study cohort and analytical flowchart. **B** Clinical spectrum and diagnostic yield in the study cohort. The upper panel depicts the proportion of clinical features observed within 418 individuals with microcephaly, stratified by primary microcephaly (PM, $n=143$) and secondary microcephaly (SM, $n=258$) groups. The lower panel presents the diagnostic yield associated with each clinical feature, represented as the percentage of resolved cases in each category. Abbreviations: DD, developmental delay; ID, intellectual disability; SS, short stature; FTT, Failure to thrive; ASD, autism spectrum disorder; CHD, congenital heart disorder; GU, genitourinary; GI, gastrointestinal

disorders (6.5%). For AD disorders, most cases were attributed to de novo variants, with *CTNNB1* as the most frequent gene ($n=10$), all of which were LOF variants. Other commonly affected genes included *FOXG1*, *KAT6A*, *GRIN1*, and *KMT2A* ($n=4$ each, Fig. 2D). De novo variant analysis in 316 trios identified 524 de novo coding variants, including 468 single-nucleotide variants (SNVs), 55 small indels, and one multi-nucleotide variant, with an average of 1.66 de novo coding variants per proband. The transition-to-transversion ratio for the detected de novo SNVs was 2.30. The distribution of de novo variant counts per proband in coding regions aligned with the expected *Poisson* distribution. Our findings indicate a high burden of de novo variants among probands, particularly for LOF variants (Additional file 1: Fig. S2).

For AR disorders, *PMM2* was the most frequent gene, having three cases. Several recessive variants were predominantly observed in Korean or East Asian populations (Additional file 2: Table S10), including two *PMM2* alleles: p.Arg103Ter (Korean AF 0.1%) and p.Arg147Pro (Korean AF 0.03%) variants. Notably, we identified previously unreported homozygous variants in *THOC6* (p.Leu241Profs*73) and *SPART* (p.Gln76Ter), absent from both the ClinVar and gnomAD databases. For XL disorders, *ATRX* (OMIM #309580) was the most frequently implicated gene ($n=8$), followed by *CASK* (OMIM #300749) and *DDX3X* (OMIM #300958).

Furthermore, pathogenic CNVs were detected in 20 individuals (5.3%), and multi-locus pathogenic variations (MPVs) were also found in 10 individuals (2.4%) (Additional file 1: Fig. S3). Recurrent regions with two or more pathogenic CNVs included 15q11.2 (*UBE3A*, $n=3$), Xp11.4 (*CASK*; $n=2$), and 17q24.2 (*BPTF*; $n=2$). Notably, pathogenic CNVs overlapping with *CASK* ($n=2$) and *FOXG1* ($n=1$) were also implicated by SNVs. Additionally, we identified an individual (MCPH171) with a homozygous deletion in *SMPD4* (OMIM #610457) through CNV analysis (Additional file 1: Fig. S4). The proband's unrelated parents had a history of recurrent infant losses associated with pachygyria and a thin corpus callosum. Notably, this deletion was undetectable by conventional CMA and remains unreported in public databases. This deletion was confirmed by genome sequencing, which revealed a 66.7 kb deletion and approximate breakpoints. Gene dosage PCR verified the

Table 1 Clinical characteristics of individuals with microcephaly

Characteristics		Total		PM		SM		*P* value
		n	%	*n*	%	*n*	%	
No. of patients		418	100.0%	143	34.2%	258	61.7%	
Sex	Female	220	52.6%	79	55.2%	131	50.8%	0.405
	Male	198	47.4%	64	44.8%	127	49.2%	
Head circumference, median (*Z*-score)		−3.31		−4.32		−2.78		< 0.0001
Age at diagnosis, median (months)		11.8		5.4		17.5		< 0.0001
Family history		11	2.6%	3	2.1%	8	3.1%	0.753
Post-term (≥ 42^0 weeks)		1	0.2%	0	0.0%	1	0.4%	< 0.0001
Term	Full term (39^0–41^6 weeks)	239	57.2%	58	40.6%	168	65.1%	
	Early term (37^0–38^6 weeks)	106	25.4%	44	30.8%	58	22.5%	
Pre-term	Late preterm (34^0–36^6 weeks)	40	9.6%	25	17.5%	15	5.8%	
	Moderate preterm (32^0–33^6 weeks)	6	1.4%	6	4.2%	0	0.0%	
	Very preterm (28^0–31^6 weeks)	8	1.9%	6	4.2%	2	0.8%	
	Extremely preterm (< 28^0 weeks)	4	1.0%	3	2.1%	1	0.4%	
	Not available	14	3.3%	1	0.7%	13	5.0%	
Birth weight	Normal (≥ 2500 g)	211	50.5%	58	40.6%	143	55.4%	< 0.0001
	Low (1500–2499 g)	84	20.1%	54	37.8%	29	11.2%	
	Very low (1000–1499 g)	7	1.7%	6	4.2%	1	0.4%	
	Extremely low (< 1000 g)	5	1.2%	3	2.1%	2	0.8%	
	Not available	111	26.6%	22	15.4%	83	32.2%	

PM, primary microcephaly; *SM*, secondary microcephaly

copy number status across three generations in this family, facilitating genetic counseling and enabling prenatal genetic diagnosis to support family planning. Finally, while we previously identified an individual with microcephaly caused by short tandem repeats in *ATN1*, screening for repeat expansions using exomes in known genes yielded no additional cases [46].

Pathway and network analysis of PM and SM

Using the identified genes, we performed gene-set enrichment analysis and protein–protein interaction analysis. These analyses revealed several key pathways significantly implicated in microcephaly development, including chromatin remodeling, neuron projection development, mitotic cell cycle, and regulation of protein-containing complex assembly (Fig. 3A). To further elucidate the distinct mechanisms underlying PM and SM, we next conducted subgroup analyses. These analyses revealed that damaged DNA binding, regulation of microtubule cytoskeleton organization, and GTPase regulator activity are specifically important in PM. Conversely, actin filament-based processes, regulation of stem cell differentiation, regulation of growth, and the Wnt signaling pathway, which are more related to the late stage of neuronal maturation, were distinct from SM (Fig. 3B). Additionally, protein–protein interaction network analyses revealed several key processes in PM and SM. In PM, protein networks were significantly enriched in cell division and DNA damage response, which may be more crucial during the early phase of brain development. In contrast, protein networks in SM were more prominent for the mid to late phase of brain development, such as modulation of synaptic transmission and actin filament-based processes (Fig. 3C). Collectively, these findings align well with previous literature, substantiating the validity of our results [2–6].

Clinical and genetic correlates of head circumference

Next, we assessed the association between clinical and genetic factors and the degree of microcephaly. HCs were transformed into *Z*-scores based on age and sex references. The age at measurement did not appear to affect the microcephaly severity (Additional file 1: Fig. S5). Subgroup analyses revealed that the PM group exhibited more pronounced microcephaly than the SM group, with a median *Z*-score of -4.32 and -2.78, respectively ($P < 0.0001$; Fig. 4A). Analysis of 12 clinical comorbidities showed no significant association with the severity of microcephaly in our cohort (Fig. 4B). However, we found a significant association between inheritance patterns and microcephaly severity (Fig. 4C): individuals with AD disorders (median Z -3.01) had milder microcephaly than those with AR (median Z -3.66) or XL (median Z -3.51) disorders. This association may reflect the higher

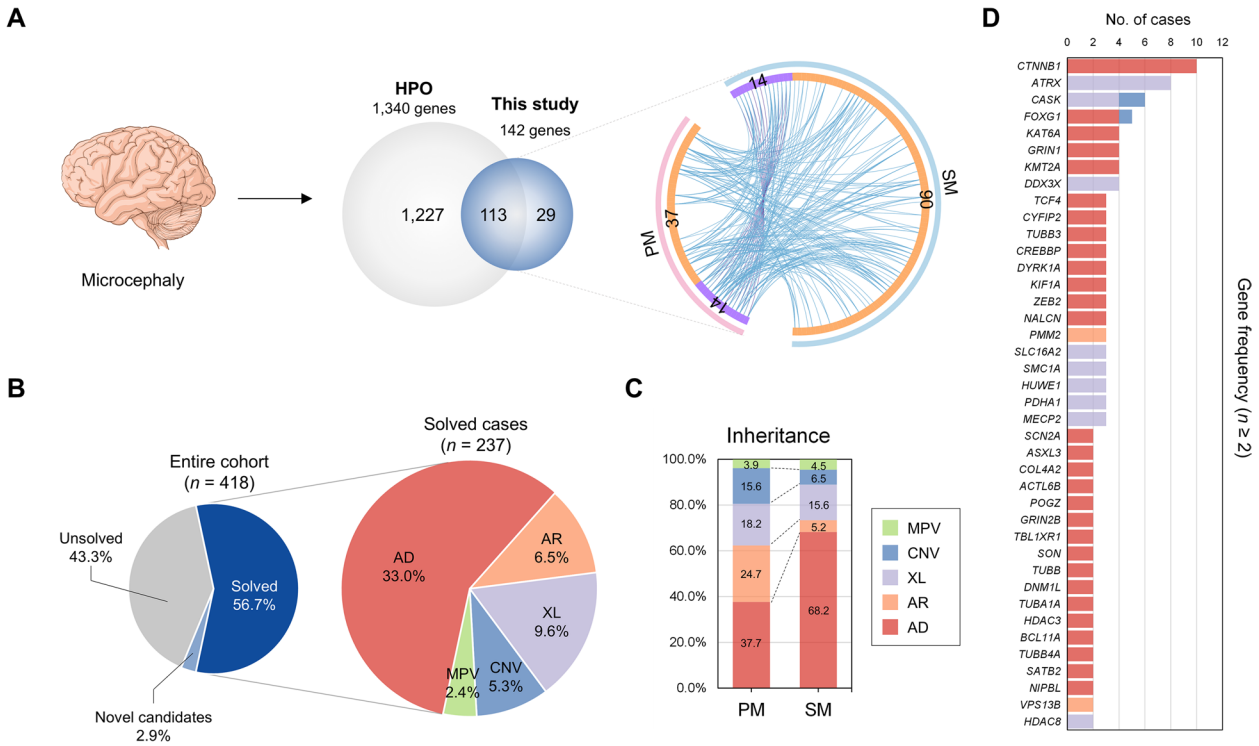

Fig. 2 Genomic landscape of genes associated with microcephaly. **A** Venn diagram analysis comparing genes identified in this study with those cataloged in the Human Phenotype Ontology (HPO) for microcephaly. A total of 142 genes were identified, with 113 overlapping with the HPO database and 29 not reported in the HPO. The Circos plot presents overlap and interactions between genes associated with primary microcephaly (PM, $n=51$) and secondary microcephaly (SM; $n=104$), with an overlap of 14 genes between the two subtypes. **B** Diagnostic outcomes of exome sequencing for the study cohort. The left pie chart depicts the proportion of solved and unsolved cases, while the right pie chart breaks down the solved cases ($n=237$) by inheritance patterns, highlighting autosomal dominant (AD), autosomal recessive (AR), X-linked (XL), copy number variants (CNV), and multi-locus pathogenic variations (MPV). **C** Comparison of inheritance patterns in solved cases by microcephaly subtypes. Notably, AR inheritance is prevalent in PM cases, whereas AD inheritance predominates in SM cases. **D** Genes that are frequently associated with two or more cases in the study cohort. The bar graph shows the frequency of alterations in specific genes, with colors corresponding to inheritance patterns. The most frequently affected gene is *CTNNB1* ($n=10$), followed by *ATRX* ($n=8$) and *CASK* ($n=6$), both located on the X chromosome

prevalence of AD inheritance in the SM group (Fig. 2C), which is linked to milder microcephaly. Additionally, we observed that 11 individuals (2.6%) reported a family history of neurodevelopmental issues, with 6 families having other affected family members with microcephaly (Additional file 2: Table S11). Genetic causes were identified in 7 individuals: 4 with AD and 3 with AR disorders, while the cause remained unclear in 4 families. Among the cases of microcephaly observed in other family members, 3 were associated with AR disorders, and 3 remained undiagnosed. Notably, none of the carrier parents of AD disorders exhibited microcephaly, although they did present with mild neurodevelopmental issues or mosaicism. These findings suggest that AR disorders may have a stronger association with familial microcephaly than AD disorders, aligning with the observation that AR disorders have a greater impact on microcephaly severity (Fig. 4C). However, further studies involving larger cohorts are warranted to validate these associations.

Furthermore, we identified that individuals with de novo variants on the X chromosome had more severe microcephaly (median Z-4.25) than those with de novo variants on autosomes (median Z-3.08; $P=0.039$; Fig. 4D). Sex differences in microcephaly severity were also notable, with females showing more severities than males, particularly in cases involving XL conditions (Fig. 4E). Further analysis revealed distinct inheritance patterns and genes involved between sexes in XL disorders: males primarily had maternally inherited variants in XR disorders, while females mostly had de novo variants in XD disorders. These differences suggest that the more severe microcephaly observed in females may result from the involvement of different genes and the greater impacts of de novo variants compared to maternally inherited variants. Among frequently altered genes with four or more cases (Fig. 4F), we noticed that individuals with *CASK* variants had significantly reduced HCs compared to others (median Z-5.06, $P=0.036$).

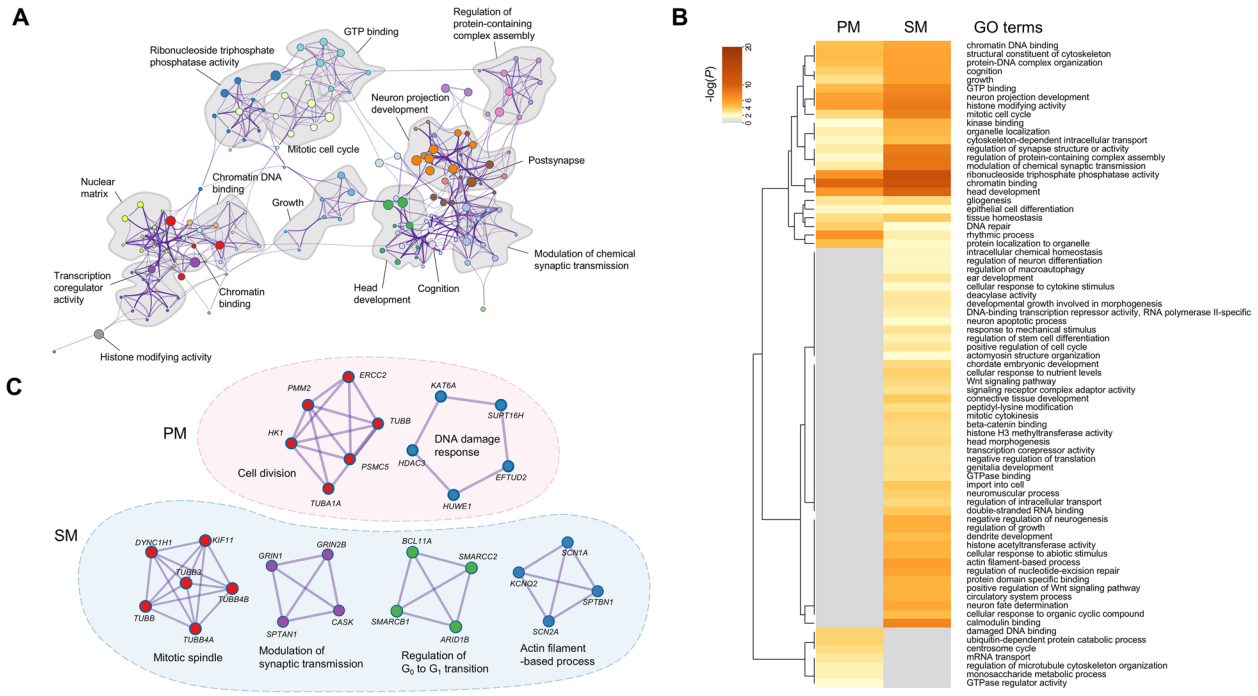

Fig. 3 Gene-set enrichment and protein interaction network analysis reveal shared and distinct mechanisms in primary and secondary microcephaly. **A** Network analysis of enriched Gene Ontology (GO) terms across all microcephaly-associated genes, with nodes connected by edges indicating a similarity score > 0.3. Node colors represent specific functional pathways, including head development, chromatin binding, neuron projection development, and mitotic cell cycle. **B** Heatmap presentation of enriched GO terms for genes associated with primary microcephaly (PM) and secondary microcephaly (SM), illustrating both shared and distinct biological processes linked to each condition. **C** Representative protein–protein interaction networks enriched in PM- and SM-associated genes, highlighting key processes such as cell division and DNA damage response in PM, and mitotic cell cycle, modulation of synaptic transmission, regulation of G_0 to G_1 transition, and actin filament-based processes in SM

Six individuals, including four females and two males, were identified with two large deletions and four SNVs in *CASK* (Additional file 1: Fig. S6). Although limited in cases, our data replicated the established association between *CASK* and severe microcephaly due to pontine and cerebellar hypoplasia (OMIM #300749).

Novel genotype–phenotype associations

In unsolved cases, we sought to identify novel candidate genes associated with disease, employing a strategy similar to that described in previous reports (see "Prioritization of candidate genes" in the "Methods" section) [36]. This approach allowed us to report a novel disease association between de novo variants in *HDAC3* and variable NDDs [47]. Additionally, we identified novel candidate genes in 12 cases (2.9%; Table 2), including 9 with AD and 3 with AR inheritance, which may represent uncharacterized genes for NDDs or microcephaly. These genes contain de novo or compound heterozygous variants with predicted deleterious effects, as indicated by in silico predictors and high conservation scores. We found that *RTF1* and *ASAP2* have been previously implicated in variable neurodevelopmental and neuropsychiatric conditions, including DD, ASD, bipolar disorder (BD), and schizophrenia (SCZ; Additional file 2: Table S12) [10–13, 48–51], with reported variants suggesting a LOF mechanism (Fig. 5A). Furthermore, our cases with *RTF1* and *ASAP2* variants exhibited PM, suggesting them suitable for brain organoid modeling [52]. Unsolved cases with identical inheritances involving these genes were also identified in the UK Deciphering Developmental Delay (DDD) study, further supporting their relevance and prioritization [11].

First, we identified eight individuals with *RTF1* (ENST00000389629) de novo variants previously reported in the literature associated with DD, ASD, and BD [11, 12, 48, 49]. All were missense variants except for one splicing-site variant, which was predicted to have a donor loss effect with a spliceAI score of 0.86. In our study, we recruited two individuals with *RTF1* variants (Additional file 2: Table S13). Individual 1 (MCPH363), with an *RTF1* (c.1822A > G, p.Arg608Gly) de novo

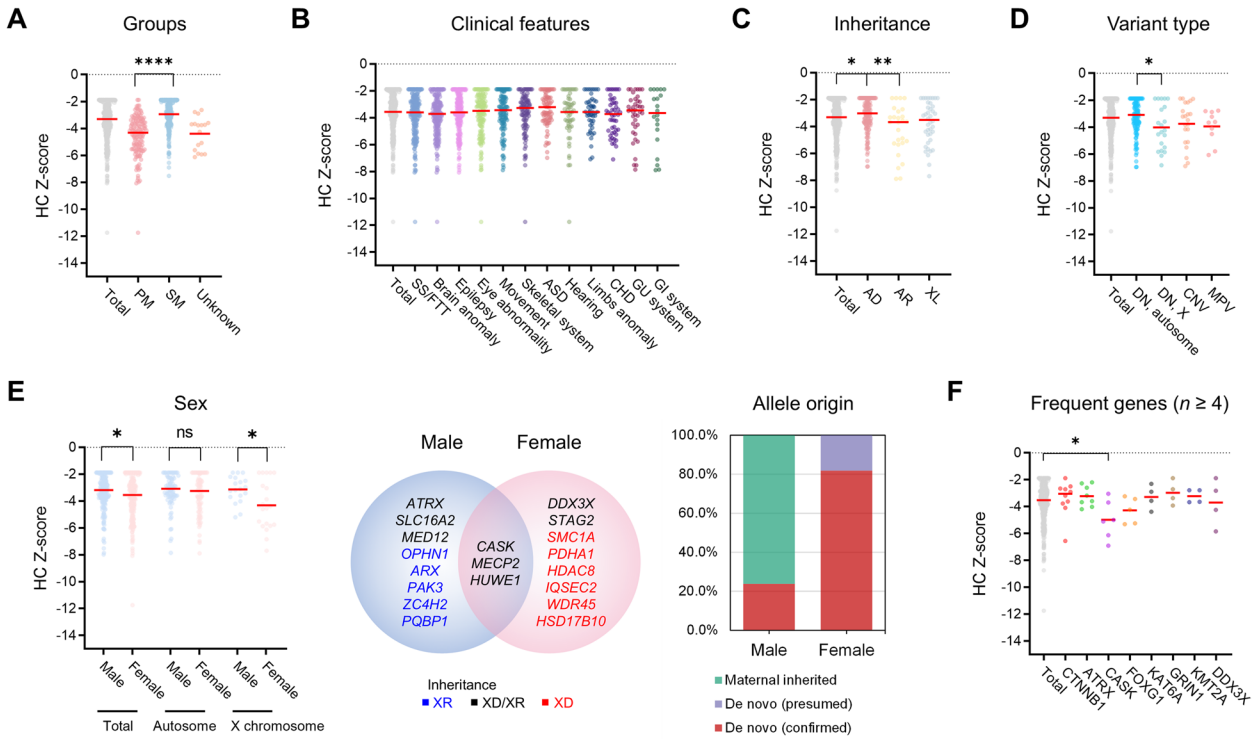

Fig. 4 Clinical and genetic factors associated with the severity of microcephaly. **A** Distribution of head circumference (HC) *Z*-scores across primary microcephaly (PM) and secondary microcephaly (SM) groups. The PM group shows significantly lower *Z*-scores than the SM group. **B** Correlation of HC *Z*-scores with clinical comorbidities. **C** HC *Z*-scores by inheritance pattern reveal that autosomal dominant (AD) conditions are associated with milder microcephaly compared to autosomal recessive (AR) and X-linked (XL) conditions. **D** Comparison of HC *Z*-scores by variant types, indicating that de novo (DN) variants on the X chromosome have a more pronounced impact on HC than those on autosomes. **E** Sex differences in microcephaly severity. Females exhibit more severe microcephaly than males, primarily driven by cases in XL conditions. Venn diagram illustrating the overlap and sex-specific distribution of genes associated with XL disorders. Genes such as *CASK*, *MECP2*, and *HUWE1* are present in both sexes, while others exhibit sex-specific inheritance patterns: X-linked recessive (XR) in males and X-linked dominant (XD) in females. Distinct allele origins are also observed, with most females having de novo variants in XD disorders, whereas males are predominantly affected by maternally inherited variants in XR disorders. **F** Gene-level analysis of HC *Z*-scores for genes with four or more cases. Individuals with variants in *CASK* exhibit significantly reduced HCs (median *Z*-5.06, $P = 0.36$). * $P < 0.05$, *** $P < 0.001$, **** $P < 0.0001$

variant, presented with DD and microcephaly. He was the second child of a dichorionic diamniotic twin. Unlike his twin brother, he exhibited multiple congenital anomalies, including an abdominal hemangioma, skull deformity, and chest wall abnormality. Prenatal ultrasonography detected microcephaly and inferior vermian hypoplasia, later confirmed by brain MRI, which also revealed delayed myelination (Fig. 5B). Neonatal complications included anemia and feeding difficulties, and he developed epilepsy at 4 months. He later underwent surgeries for bilateral hydrocele, congenital cataracts, and a preauricular skin tag. Follow-up MRI revealed dysplastic lens thinning and mild microphthalmia at 36 months, while a kidney biopsy confirmed congenital nephrotic syndrome with focal segmental glomerulosclerosis. Individual 2, who harbors an *RTF1* (c.1672 T > C, p.Ser558Pro) de novo variant, exhibited DD/ID and ASD, with comparatively milder phenotypes than Individual 1. His HC measured 52 cm in his 40s, within the normal range, and brain MRI revealed no anatomical abnormalities.

Similarly, *ASAP2* (ENST00000281419) variants have been associated with DD, ASD, and SCZ in the literature [10–13, 50, 51]. While most of these cases involve monoallelic de novo variants, our case and the one reported in the DDD study involved biallelic variants—either as homozygotes or compound heterozygotes (Additional file 2: Table S13). Individual 3 (MCPH052), who carried compound heterozygous *ASAP2* variants (c.1295G > A, p.Gly432Asp; c.1880A > G, p.Tyr627Cys), exhibited DD and arthrogryposis multiplex congenita, affecting all joints except the hips, as well as syndromic features with multi-organ anomalies in the brain, eyes, and skeletal system. Notably, the brain MRI revealed bilateral symmetric perisylvian polymicrogyria (Fig. 5C), and multiple eye abnormalities were observed, including microphthalmia, retinal degeneration, ectopic pupil with corneal opacity, cataract, and coloboma. Additional skeletal

Table 2 Candidate genes identified in individuals with neurodevelopmental disorders and microcephaly

MOI	Individual	Group	Gene (Transcript)	pLI[a]	LOEUF[a]	HGVS	Origin	GERP score	CADD score	AlphaMissense score	gnomAD AF[a]	Reported in DDD cohort[b]
AD	MCPH067	SM	*RPS5* (ENST00000196551)	1.00	0.31	c.437G>T, p.Arg146Leu	De novo	4.98	31	0.9984	NR	1 de novo
	MCPH176	SM	*DCAF5* (ENST00000341516)	1.00	0.19	c.1301C>A, p.Ser434Ter	De novo	5.68	40	NA	NR	1 de novo
	MCPH214	SM	*PPP2R2A* (ENST00000380737)	1.00	0.36	c.251T>C, p.Phe84Ser	De novo	5.98	29.9	0.9998	NR	1 de novo
	MCPH273	SM	*ATAD2B* (ENST00000238789)	1.00	0.38	c.2851C>T, p.Arg951Ter	De novo	5.59	38	NA	NR	3 de novo
	MCPH314	PM	*KPNB1* (ENST00000290158)	1.00	0.12	c.592C>T, p.Leu198Phe	De novo	6.02	25.7	0.9322	NR	No
	MCPH332	SM	*CEP170* (ENST00000366542)	1.00	0.26	c.1623_1624del, p.His541Glnfs*48	De novo	NA	NA	NA	NR	No
	MCPH344	PM	*DPYSL3* (ENST00000343218)	1.00	0.39	c.452T>A, p.Met151Lys	De novo	5.44	24.9	0.8965	NR	No
	MCPH363	PM	*RTF1* (ENST00000389629)	1.00	0.22	c.1822A>G, p.Arg608Gly	De novo	5.47	24.8	0.8614	NR	1 de novo
	MCPH365	PM	*IWS1* (ENST00000295321)	1.00	0.37	c.53C>T, p.Pro18Leu	De novo	5.8	29.7	0.9551	NR	No
AR	MCPH052	PM	*ASAP2* (ENST00000281419)	1.00	0.47	c.1295G>A, p.Gly432Asp	Paternal	5.24	26.8	0.9597	NR	1 homozygote
						c.1880A>G, p.Tyr627Cys	Maternal	5.67	30	0.8336	NR	
	MCPH218	PM	*PP2D1* (ENST00000389050)	0.00	1.22	c.762_763del, p.Phe255Ter	Paternal	NA	NA	NA	$1.4x10^{-5}$	1 de novo
						c.250C>T, p.Gln84Ter	Maternal	5.46	33	NA	$4.89x10^{-5}$	
	MCPH258	SM	*ABCC12* (ENST00000311303)	0.00	1.02	c.4058T>C, p.Leu1353Pro	Paternal	5.5	25	0.7635	NR	1 homozygote
						c.3161C>A, p.Ser1054Ter	Maternal	5.79	43	NA	$6.98x10^{-6}$	

Abbreviations: *MOI* mode of inheritance, *AD* autosomal dominant, *AR* autosomal recessive, *PM* primary microcephaly, *SM* secondary microcephaly, *pLI* probability of being loss-of-function intolerant, *LOEUF* Loss-of-function Observed/Expected Upper-bound Fraction, *HGVS* Human Genome Variation Society, *CADD* Combined Annotation Dependent Depletion, *GERP* the Genomic Evolutionary Rate Profiling, *NA* not available, *NR* not reported

[a] These metrics were obtained from gnomAD Exome v.4.0.0

[b] Unresolved cases with variants of unknown significance in the specified gene obtained from the UK Deciphering Developmental Delay (DDD) study[11]

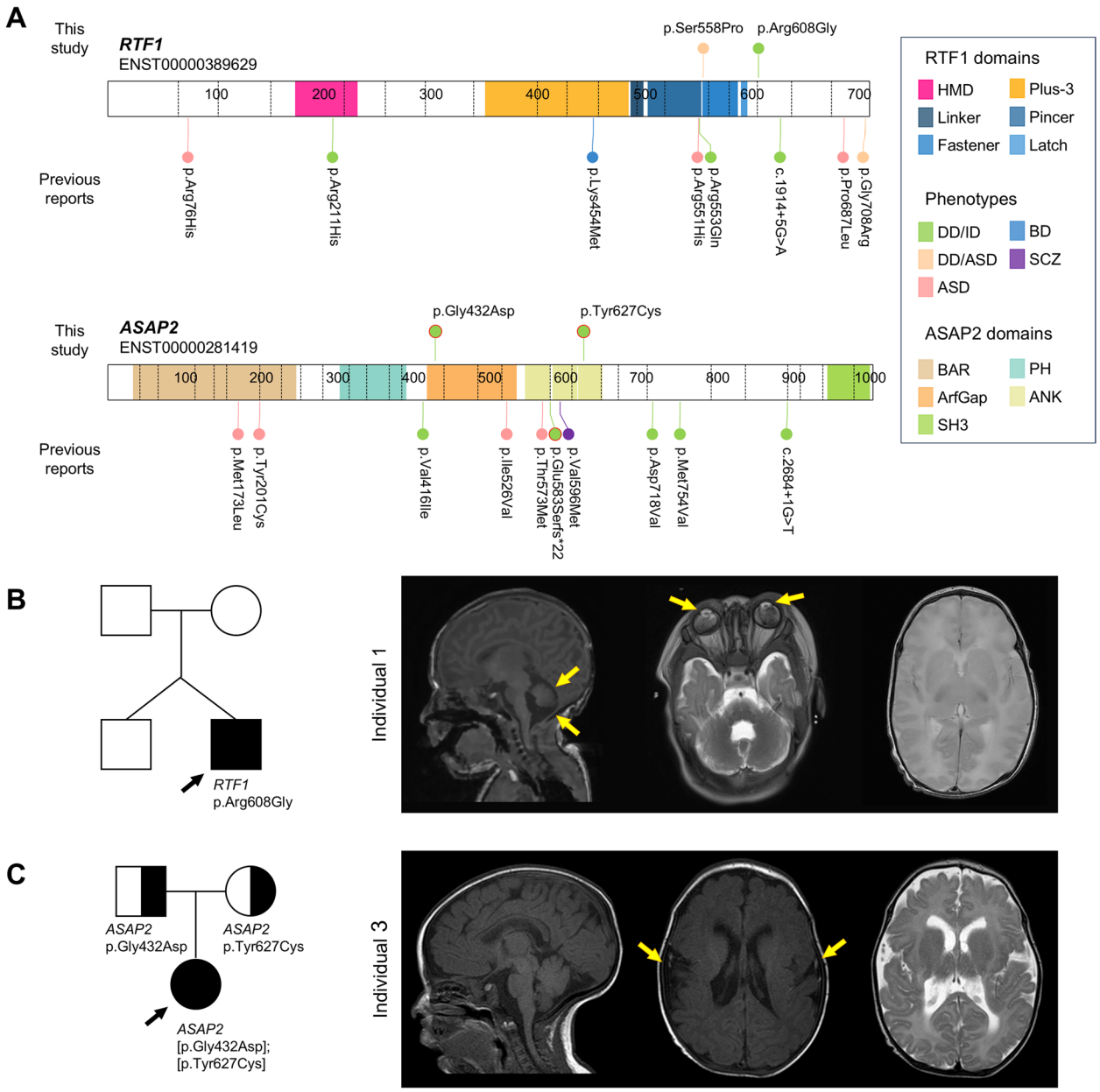

Fig. 5 Literature review and clinical findings uplift *RTF1* and *ASAP2* as high-priority candidate genes in neurodevelopmental disorders. **A** Schematic overview of both newly identified and previously reported variants in *RTF1* and *ASAP2*. De novo variants in these genes are implicated in various neurodevelopmental and neuropsychiatric conditions, including DD, ASD, BD, and SCZ. Notably, for *ASAP2*, both monoallelic and biallelic variants (compound heterozygotes or homozygotes; two independent cases highlighted with red circles) were observed. **B** Brain MRI of Individual 1 with an *RTF1* p.Arg608Gly de novo variant shows features of microcephaly, inferior vermian hypoplasia, delayed myelination, and dysplastic thinning of both lenses with mild microphthalmia, indicating broad brain anomalies. **C** Brain MRI of Individual 3 with compound heterozygous variants in *ASAP2* ([p.Gly432Asp]; [p.Tyr627Cys]) reveals bilateral symmetric perisylvian polymicrogyria and prominent primitive veins, indicating neuronal migration disorder. Abbreviations: HMD, histone modification domain; BAR, Bin/Amphiphysin/Rvs domain; PH, pleckstrin homology domain; ANK, ankyrin repeat domain; SH3, Src homology 3 domain; DD, developmental delay; ID, intellectual disability; ASD, autism spectrum disorder; BD, bipolar disorder; SCZ, schizophrenia

anomalies included dolichocephaly, thoracic deformities, bilateral clubfoot, and a high-arched palate. Collectively, evidence from the literature and our cohort suggests that dysfunction of *RTF1* and *ASAP2* may contribute to impaired brain development, potentially resulting in NDDs or microcephaly.

Functional impacts of *RTF1* and *ASAP2* knockout in NPCs and forebrain organoids

Based on the observed clinical findings, we investigated the roles of *RTF1* and *ASAP2* in brain development through functional experiments using CRISPR-Cas9 knockout models in NPCs derived from human forebrain organoids. To assess the impact of gene knockout on cell proliferation, two sgRNAs targeting each gene were designed and cloned into a plasmid co-expressing Cas9. These plasmids were then co-electroporated with a piggyBac transposon system to label transfected NPCs and their progeny with GFP, allowing for efficient tracking of electroporated cells [53]. Following electroporation, GFP-positive cells were sorted to validate knockout efficiency, which was within the acceptable range of 30% to 60% (Additional file 1: Fig. S7E). Cell cycle analysis using flow cytometry revealed that knockout of *RTF1* or *ASAP2* resulted in significant alterations in cell cycles, with a notable reduction in the percentage of cells in the S phase compared to scramble controls (Fig. 6A, B). These findings indicate that *RTF1* and *ASAP2* are essential for proper progression through the S phase of the cell cycle in NPCs, highlighting their critical roles in regulating NPC proliferation during human brain development.

To further verify whether these genes are essential for NPC proliferation in a physiologically relevant three-dimensional environment resembling the developing human cortex, we employed the same knockout strategy to forebrain organoids via electroporation [43, 54]. Electroporation was performed on 2-month-old forebrain organoids, followed by immunohistochemistry analysis at 2 and 5 days post-electroporation (Fig. 6C). The analysis revealed a significant reduction in the number of NPCs expressing Ki-67, a marker of proliferating cells, among PAX6 and GFP co-positive cells upon *RTF1* knockout, starting from two days post-electroporation (Fig. 6D). A similar reduction in Ki-67 positive cells was observed in *ASAP2* knockout cells by the five days post-electroporation (Fig. 6E). These findings align with results from two-dimensional NPC cultures, demonstrating that both *RTF1* and *ASAP2* are indeed essential for NPC proliferation during human brain development.

Lastly, we conducted rescue experiments using our established knockout system to assess the functional impact of patient-derived *RTF1* variants on NPC proliferation. In this approach, sgRNA-resistant cDNA constructs encoding either *RTF1* wild-type (WT) or the patient-specific variants p.Arg608Gly and p.Ser558Pro were introduced. Western blot analysis confirmed robust expression of HA-tagged cDNAs even in the presence of the sgRNA/Cas9 system (Fig. 6F). To ensure sufficient time for phenotypic recovery, a cell cycle assay was performed three days after electroporation. The results showed that expression of WT cDNA fully restored the reduced S-phase proportion observed in the knockout (GFP-only) condition, whereas cDNAs carrying patient-derived *RTF1* variants resulted in only partial rescue (Fig. 6G, Additional file 1: Fig. S7G). Overexpression of either the p.Arg608Gly or p.Ser558Pro variants led to a significantly lower S-phase proportion compared to the scramble control, with the p.Arg608Gly variant showing no significant difference from the knockout. These findings suggest that p.Arg608Gly is likely a complete LOF allele, whereas p.Ser558Pro exhibits a hypomorphic effect. This is consistent with the spectrum of disease severity observed between Individual 1 and 2. Collectively, our results demonstrate that patient-derived *RTF1* variants impair NPC proliferation, potentially contributing to abnormal brain development.

Discussion

To our knowledge, this study represents the largest investigation to date exploring the associations between HCs and clinical and genetic factors in individuals with microcephaly, based on exome sequencing data from a cohort of 418 individuals. We achieved a diagnostic yield of 56.7%, with novel candidate genes identified in 2.9% of the cohort (Fig. 2B). Overall, we found 142 genes associated with microcephaly. The analyses within our cohort revealed that *CTNNB1* is the most frequent gene, followed by two XL genes, *ATRX* and *CASK*. It is noteworthy that the molecular spectrum of microcephaly in Koreans differed from that reported in Europeans and Arabians [4, 5, 55]. Specifically, de novo variants in AD disorders emerged as the primary genetic drivers, while the number of solved cases within AR disorders was relatively low, which may be linked to no consanguineous marriages in this cohort. Therefore, the well-established microcephaly genes with AR inheritance such as *MCPH1*, *WDR62*, and *ASPM* did not play as major causative genes in Koreans. Nevertheless, we identified recurrent recessive variants frequently associated with Korean or East Asian populations, including rare variants in *PMM2* (Additional file 2: Table S10). Furthermore, we detected a homozygous deletion encompassing the entire *SMPD4* gene (OMIM #610457) using CNV detection tools in a non-consanguineous family with a history of recurrent baby loss. This deletion may represent a founder allele in the Korean population; therefore, similar cases should be suspected in Korean families with a recurrent history of fetal loss accompanied by multiple skeletal anomalies, pachygyria, and microcephaly (Additional file 1: Fig. S4).

The comparisons between PM ($n=51$) and SM-associated genes ($n=104$) revealed several key pathways involved in the pathogenesis of microcephaly (Fig. 3).

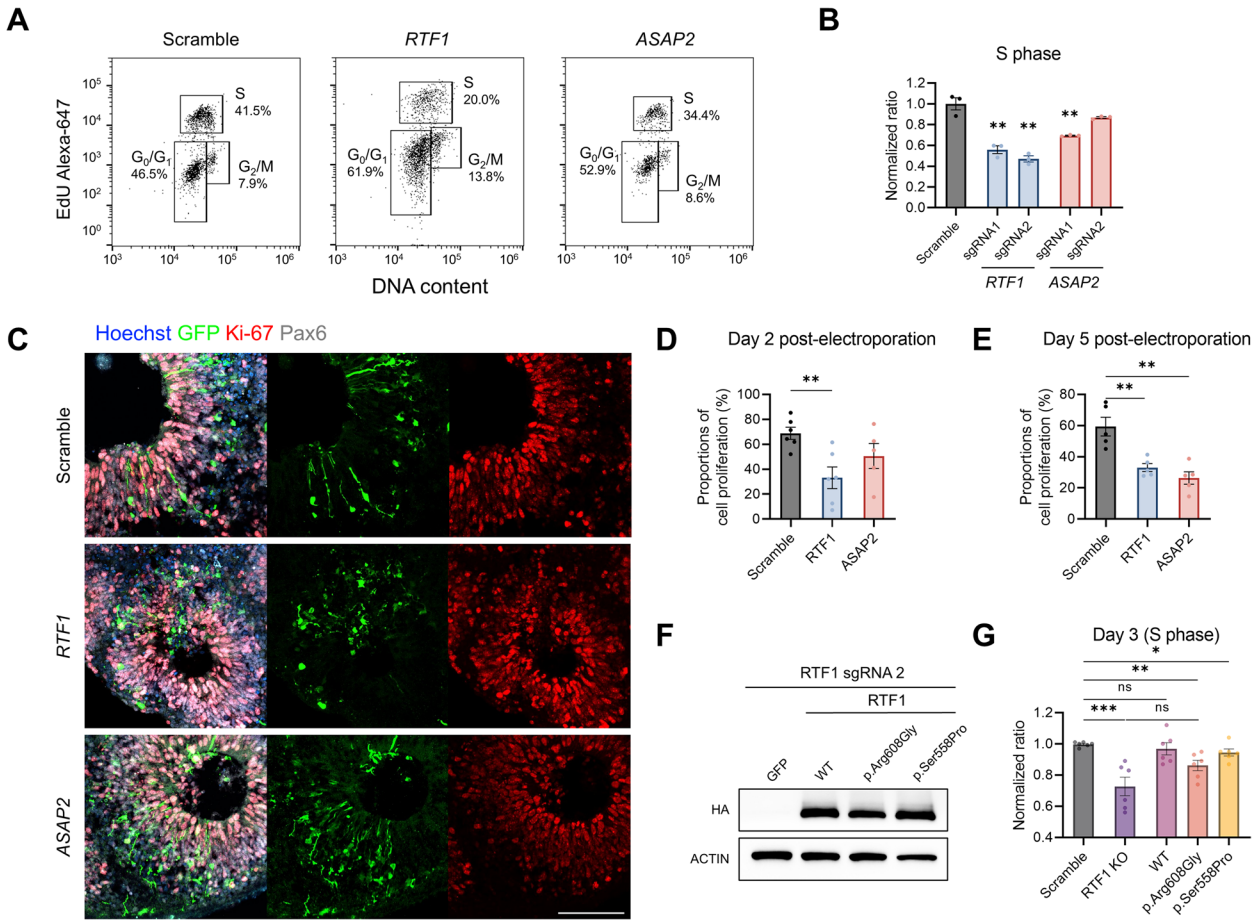

Fig. 6 Loss of RTF1 or ASAP2 impairs neural progenitor cell (NPC) proliferation in brain organoids. **A, B** Flow cytometry analysis of cell cycle progression in neural progenitor cells (NPCs) transfected with the piggyBac system for GFP labeling and a plasmid expressing Cas9 and sgRNA targeting *RTF1* and *ASAP2*. GFP-labeled cells were gated for analysis. **A** Representative flow cytometry plots show distinct distributions among cell populations in cell cycle phases in NPCs treated with sgRNA1 of *RTF1* and *ASAP2*, compared to the control (scramble). **B** Quantification of the percentage of cells in the S phase shows that knockout of *RTF1* and *ASAP2* in NPCs using the CRISPR-Cas9 system with sgRNAs significantly reduces the S phase cell population. Data represent mean ± SEM ($n = 3$). **C-E** Immunohistochemistry analysis of brain organoids at 2 and 5 days post-electroporation to deplete *RTF1* and *ASAP2* in NPCs. Data are shown as mean ± SEM from 5–6 sections across 2–3 independent organoids. **C** Confocal images of organoids stained with Hoechst, GFP, Ki-67, and Pax6. Hoechst stains nuclei, indicating all cells; GFP labels transfected cells; Ki-67 marks proliferating cells; and Pax6 identifies neural progenitor cells (scale bar, 100 µm). **D, E** Quantification of proliferating cells in forebrain organoid, which is measured by the proportion of $PAX6^+GFP^+Ki\text{-}67^+$ cells among the $PAX6^+GFP^+$ population at **D** 2 days and **E** 5 days post-electroporation. **F** Western blot analysis demonstrating robust expression of HA-tagged RTF1 constructs, including wild-type (WT) and patient-derived variants, following co-transfection with RTF1 sgRNA2 and Cas9. **G** Functional rescue assay following *RTF1* knockout. Overexpression of *RTF1* WT restored S-phase proportion, whereas the patient-derived variants (p.Arg608Gly and p.Ser558Pro) exhibited absent or reduced rescue efficiency. Data represent mean ± SEM ($n = 6$). ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Since individuals with SM were predominant in our cohort, the analysis yielded a greater number of enriched pathways associated with SM. Specifically, chromatin remodeling, regulation of the mitotic cell cycle, neuron projection, and gliogenesis were identified as common mechanisms shared by PM and SM. Conversely, pathways such as neurotransmitter regulation, membrane potential, apoptotic signaling, and the Wnt/β-catenin signaling pathway are specifically implicated in SM. These results aligned with our understanding that PM is primarily associated with defects in early brain development, whereas SM is characterized by abnormalities occurring during the mid to late phases of brain development [2–6]. It is notable that *CTNNB1*, which encodes the core protein of the Wnt/β-catenin signaling pathway, was the most frequently altered gene in our cohort, with all individuals harboring these variants presenting with SM. A recent study highlighted that genes associated with DD are mainly involved in early-phase cells, such as neuronal progenitors or immature cells, while those associated with ASD are mainly expressed in mature neurons [12]. Our findings are compatible with this finding, as the SM

group showed a higher prevalence of ASD compared to the PM group (22.5 vs. 11.9%, $P=0.011$; Fig. 1B).

The severity of microcephaly was influenced by multiple clinical and genetic factors (Fig. 4). Overall, the PM group displayed a more severe phenotype compared to the SM group, which may be largely linked to the higher proportion of AR inheritance in the PM group (Fig. 2C) and more severe microcephaly in AR disorders (Fig. 4C). In addition, sex differences in microcephaly severity were noted. These differences were primarily driven by cases involving XL disorders (Fig. 4E), where females exhibited more severe microcephaly than males. Our analyses suggest a substantial impact of the X chromosome on brain development. The second and third most frequently affected genes in our cohort were both XL genes (*CASK* and *ATRX*). Moreover, individuals harboring variants in *CASK* presented with significantly reduced HCs, implying the strong influence of this gene (Fig. 4F). Additionally, de novo variants on the X chromosome result in more severe microcephaly than those on autosomes (Fig. 4D). Together, these findings emphasize the crucial impact of the X chromosome in brain development, compatible with the previous report [56], and may help elucidate the sex-based differences observed in microcephaly severity. Furthermore, MPVs were identified in 10 cases (2.4%), a proportion consistent with previous studies [57]. These included 3 cases of AD + AD, 2 cases each of AD + AR, AD + XL, AD + CNV, and 1 case of XL + CNV events (Additional file 1: Fig. S3). Most individuals with MPVs exhibited broad syndromic presentations, possibly due to phenotypic expansions or blended phenotypes [58, 59]. Although the cases are limited, microcephaly severity did not appear to be associated with the presence of MPVs (Fig. 4D).

Previous gene-based approaches highlighted the importance of large-scale data and de novo variants in novel gene discovery, particularly for AD disorders [36]. We adopted a similar strategy and utilized segregation information from trio sequencing, prioritizing 12 candidate genes with 9 AD and 3 AR disorders (Table 2). Among these candidates, we have validated the role of *RTF1* and *ASAP2* in brain development, and international collaborations enabled us to collect an additional case with an *RTF1* variant.

RTF1 (RNA polymerase-associated protein RTF1 homolog) is a critical component of the PAF1 complex, known for its regulatory influence on gene expression through interactions with chromatin-remodeling proteins, as well as for its function as a transcription elongation factor independent of the PAF1 complex [60]. In zebrafish, mutations in *Rtf1* and other PAF1 complex proteins lead to neural crest cell malformation, suggesting a significant role in neurodevelopmental processes [61]. Similarly, mutations in *Ctr9*, another core component of the PAF1 complex, were lethal at late embryonic or early larval stages in *Drosophila*, though partial rescue was possible with nervous system-specific expression [62]. Individuals with *RTF1* variants displayed a range of clinical features associated with brain development, including DD/ID, ASD, microcephaly, and other brain abnormalities (Additional file 2: Table S12) [11, 12, 48, 49]. These findings implied that *RTF1* dysfunction may lead to variable defects in brain development and may have pleiotropic effects [63]. *RTF1* has a pLI score of 1, indicating intolerance to LOF, and all identified variants in our cohort occurred de novo. Notably, several identified variants are localized at arginine residues (e.g., p.Arg551His, p.Arg553Gln, and p.Arg608Gly), which are important for helix stability and are situated near the "fastener" and "latch" helices (Fig. 5A). These helices (residues 486–710) are reported to be crucial for transcription elongation by facilitating RNA polymerase II translocation [64]. Given the pivotal role of this domain in transcriptional regulation, variants affecting these regions are likely to impair gene expression control, thereby contributing to neurodevelopmental disruption. Our functional rescue assays (Fig. 6G) further substantiate this model, delineating differences in disease severity between Individual 1 (p.Arg608Gly) and Individual 2 (p.Ser558Pro). However, we acknowledge that the limited number of enrolled cases constrains our ability to establish definitive genotype–phenotype correlations for *RTF1*. In particular, the multi-systemic involvement observed in Individual 1 raises the possibility of undetected MPVs [57]. Future studies employing comprehensive genomic analyses and expanded case ascertainment through collaborative efforts will be essential to confirm these novel gene–disease associations and to further elucidate the pleiotropic effects observed in this study.

ASAP2 (ArfGAP with SH3 domain, ANK repeat, and PH domain 2) is a member of the ArfGAP family, which is involved in endocytic trafficking and actin remodeling [65]. Currently, little is known about the role of ASAP2 in human diseases. Its paralog, ASAP1, is well-studied for its role in cell motility and cancer metastasis [66, 67]. Studies in animal models reveal that loss of ASAP1 function in mice results in perinatal lethality, growth retardation, and developmental defects in bone and adipose tissues [68]. Similarly, ASAP1-mutant zebrafish display impaired embryonic development and neutrophil migration [69]. Furthermore, AGAP1, another important protein in the ArfGAP family, has been linked to cerebral palsy and NDDs in humans, with AGAP1-morphant zebrafish exhibiting developmental abnormalities, neurological deficits, and reduced motility [70]. The functional overlap among the ArfGAP family members, together

with their expression in brain tissues, raises the possibility that ASAP2 dysfunction may result in similar phenotypic manifestations in humans. This hypothesis is supported by previous reports [10–13, 50, 51] and our case, where Individual 3 with *ASAP2* bi-allelic variants was associated with NDDs and arthrogryposis multiplex congenita. With a pLI score of 1, *ASAP2* presents evidence of being under constraint, suggesting it may function in a dosage-sensitive manner. It is plausible that monoallelic variants may result in milder phenotypes as reported in the literature, while biallelic variants are associated with more severe clinical presentations. However, considering the limited number of cases in our study, further investigations, including additional cases with ASAP2 variants and functional experiments, are required to clarify the role of ASAP2 in human disease and to better elucidate the underlying molecular mechanisms and genotype–phenotype correlations.

Functional investigations employing human brain organoids have demonstrated that both *RTF1* and *ASAP2* are essential for NPC proliferation. Brain organoids are three-dimensional brain models that recapitulate various aspects of the developing brain and have been a useful tool for investigating genes involved in NDDs such as microcephaly [52, 71]. Knockout of *RTF1* and *ASAP2* using CRISPR-Cas9 led to substantial defects in NPC proliferation (Fig. 6), indicating that these genes are essential for brain development. Indeed, the observed reduction in NPC populations during the S phase aligns with our pathway analyses, which emphasize the critical role of the regulation of G_0 to G_1 transition in microcephaly (Fig. 3C). While our study selected only two high-priority genes in the human brain organoid model, recent advancements indicate that high-throughput genetic screening can be effectively performed in human brain organoids [71]. This emerging platform offers the potential to accelerate the simultaneous investigation of multiple candidate genes involved in brain development and to more efficiently and comprehensively assess their contributions to microcephaly.

Conclusions

This study provides new insights into the diverse genomic landscape of microcephaly, emphasizing the role of rare coding variants in the complex genetic architecture linked to brain development. Our findings have important implications for advancing diagnostic approaches, therapeutic strategies, and genetic counseling for individuals affected by this debilitating condition. While our analysis concentrated on rare coding variants, the impact of non-coding variants remains to be explored. Future research involving larger, more diverse cohorts, along with functional studies to investigate the molecular pathways regulated by various candidate genes, is necessary to validate our results and deepen our understanding of the genetic mechanisms underlying microcephaly during neurodevelopment in humans.

Abbreviations

AD	Autosomal dominant
AR	Autosomal recessive
ASD	Autism spectrum disorder
BD	Bipolar disorder
CMA	Chromosomal microarray
CNV	Copy-number variation
DD	Developmental disorders
DDD	Deciphering Developmental Delay
ESC	Embryonic stem cell
GFP	Green fluorescent protein
gnomAD	Genome Aggregation Database
GO	Gene Ontology
HC	Head circumference
HPO	Human phenotype ontology
ID	Intellectual disability
LOF	Loss-of-function
KOVA2	Korean Variant Archive 2
LP	Likely pathogenic
MPV	Multi-locus pathogenic variation
NDD	Neurodevelopmental disorder
NPC	Neural progenitor cell
OMIM	Online Mendelian Inheritance in Man
P	Pathogenic
PM	Primary microcephaly
SCZ	Schizophrenia
sgRNAs	Single guide RNAs
SM	Secondary microcephaly
SNV	Single nucleotide variant
XD	X-linked dominant
XL	X-linked
XR	X-linked recessive

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-025-01513-w.

Additional file 1.

Additional file 2.

Acknowledgements

The authors thank all participants and their families for their support. We acknowledge NIAID NIH for the open brain image from the NIH BioArt collection (bioart.niaid.nih.gov/bioart/60).

Authors' contributions

Conceived and supervised the study, secured funding: J.G.Y., K-J.Y., J.H.C.; Recruited participants and/or provided samples, obtained ethical approval: S.Y.K., W.J.K., A.C., J.S.L., A.F-J., S.S., R.U-S-M., C.C., N.M., B.C.L., J.M.K., K.J.K, J.H.C.; Project administration: J.S.L.; Experiments: H.J., S.P., M.J.K., H.Y., J.H.; Data analysis: J.G.Y., H.J., S.L., S.S.J., S.P., J.C., M.C.; Visualization: J.G.Y., H.J.; Writing—original draft: J.G.Y., H.J., K-J.Y., J.H.C.; Writing—review & editing: all authors reviewed and approved the final manuscript.

Funding

This research was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT, grant number: 2022–0-00333) and SNUH Kun-hee Lee Child Cancer and Rare Disease Project Foundation, Republic of Korea (grant number: 22B-001–0100). Additional support was provided by the National

Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning (MSIP), under grant numbers 2019R1C1C1006600 and 2020M3A9E4039670, as well as RS-2024–00440778 (to K.-J.Y.). Further funding was supported by the Korea Health Technology R&D Project grant through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: 2025–02214700 to J.G.Y.). This work was also supported by a Young Investigator Grant from the Suh Kyungbae Foundation awarded to K.-J.Y. In addition, the study was supported by JSPS KAKENHI (grant numbers JP19H03621 and JP22H03047) and the NCGM Intramural Research Fund (grant numbers 21A1011 and 23A1014 to N.M.).

Data availability

All datasets generated or analyzed during this study are included in the main manuscript or in the additional files. Exome sequencing datasets from this study cannot be released publicly due to patient confidentiality and the restrictions of Institutional Review Board (IRB) approvals. Depending on specific IRB approvals and the optional consent provided by participants or their parents, access to subsets of genomic data may be granted upon request, subject to approval by the Institutional Review Board of Seoul National University Hospital. Requests for further information and data access should be directed to the lead contact, Dr. Jong-Hee Chae (chaeped1@snu.ac.kr).

Declarations

Ethics approval and consent to participate

This study was conducted in compliance with the Helsinki Declaration. All participants were enrolled according to a study protocol approved by the Institutional Review Board of Seoul National University Hospital (IRB No.1406–081-588). Written informed consent was obtained from all participants or their parents.

Consent for publication

All participants in this study provided written informed consent for the use and publication of anonymized clinical and genetic information. In multi-institutional cases, each participating institution obtained the required consent for publication from the relevant individuals or their legal representatives.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Genomic Medicine, Rare Disease Center, Seoul National University Children's Hospital and Seoul National University College of Medicine, Daehak-Ro 101, Jongno-Gu, Seoul 03080, Republic of Korea. [2]Department of Laboratory Medicine, Gangnam Severance Hospital and Yonsei University College of Medicine, Seoul, Republic of Korea. [3]Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea. [4]KAIST Stem Cell Center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. [5]Department of Pediatrics, Seoul National University Children's Hospital, Seoul, Republic of Korea. [6]Department of Pediatrics, Seoul National University Bundang Hospital, Seongnam, Republic of Korea. [7]Department of Pediatrics, Seoul National University Hospital Child Cancer and Rare Disease Administration, Seoul National University Children's Hospital, Seoul, Republic of Korea. [8]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea. [9]Department of Pediatric Neurology, Neurogenetic Section, Hospital Universitario Quironsalud, Universidad Europea, Madrid, Spain. [10]Child Neurology Service, Hospital de Puerto Montt, Puerto Montt, Chile. [11]Escuela de Medicina, Universidad San Sebastián, Sede Patagonia, Puerto Montt, Chile. [12]Departamento de Neurología, Pontificia Universidad Católica de Chile, Santiago, Chile. [13]National Reference Center for Epilepsy, Hospital Dr. Sótero del Río, Santiago, Chile. [14]Neurosurgery Department, Pontificia Universidad Católica de Chile, Santiago, Chile. [15]Department of Human Genetics, National Institute of Global Health and Medicine, Japan Institute for Health Security, Tokyo, Japan.

Received: 8 November 2024 Accepted: 10 July 2025
Published online: 06 August 2025

References

1. Gilbert SL, Dobyns WB, Lahn BT. Genetic links between brain development and brain evolution. Nat Rev Genet. 2005;6:581–90.
2. Jayaraman D, Bae B-I, Walsh CA. The Genetics of Primary Microcephaly. Annu Rev Genomics Hum Genet. 2018;19:177–200.
3. Woods CG, Parker A. Investigating microcephaly. Arch Dis Child. 2013;98:707–13.
4. Shaheen R, Maddirevula S, Ewida N, Alsahli S, Abdel-Salam GMH, Zaki MS, et al. Genomic and phenotypic delineation of congenital microcephaly. Genet Med. 2019;21:545–52.
5. Boonsawat P, Joset P, Steindl K, Oneda B, Gogoll L, Azzarello-Burri S, et al. Elucidation of the phenotypic spectrum and genetic landscape in primary and secondary microcephaly. Genet Med. 2019;21:2043–58.
6. Seltzer LE, Paciorkowski AR. Genetic disorders associated with postnatal microcephaly. Am J Med Genet C Semin Med Genet. 2014;166:140–55.
7. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J-P, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019;47:D1018–27.
8. Phan TP, Holland AJ. Time is of the essence: the molecular mechanisms of primary microcephaly. Genes Dev. 2021;35:1551–78.
9. Jean F, Stuart A, Tarailo-Graovac M. Dissecting the Genetic and Etiological Causes of Primary Microcephaly. Front Neurol. 2020;11: 570830.
10. Kosmicki JA, Samocha KE, Howrigan DP, Sanders SJ, Slowikowski K, Lek M, et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat Genet. 2017;49:504–10.
11. Kaplanis J, Samocha KE, Wiel L, Zhang Z, Arvai KJ, Eberhardt RY, et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. Nature. 2020;586:757–62.
12. Fu JM, Satterstrom FK, Peng M, Brand H, Collins RL, Dong S, et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat Genet. 2022;54:1320–31.
13. Zhou X, Feliciano P, Shu C, Wang T, Astrovskaya I, Hall JB, et al. Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. Nat Genet. 2022;54:1305–19.
14. Yoon JG, Hahn HM, Choi S, Kim SJ, Aum S, Yu JW, et al. Molecular Diagnosis of Craniosynostosis Using Targeted Next-Generation Sequencing. Neurosurg. 2020;87:294.
15. Yu JW, Yoon JG, Han C, Noh SH, Shin DM, Yang Y-M, et al. Digenic impairments of haploinsufficient genes in patients with craniosynostosis. JCI Insight. 2025;10: e176985.
16. Kim JH, Yun S, Hwang S-S, Shim JO, Chae HW, Lee YJ, et al. The 2017 Korean National Growth Charts for children and adolescents: development, improvement, and prospects. Korean J Pediatr. 2018;61:135–49.
17. Lee J, Lee J, Jeon S, Lee J, Jang I, Yang JO, et al. A database of 5305 healthy Korean individuals reveals genetic and clinical implications for an East Asian population. Exp Mol Med. 2022;54:1862–71.
18. Pipelines W, Awdeh A, Dingman W, Grant G, Khajouei F, Kiernan E, et al. WDL Analysis Research Pipelines: Cloud-Optimized Workflows for Biological Data Processing and Reproducible Analysis [Internet]. Preprints; 2024. Available from: https://www.preprints.org/manuscript/202401.2131/v1
19. Yun T, Li H, Chang P-C, Lin MF, Carroll A, McLean CY. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. Bioinformatics. 2021;36:5582–9.
20. Pedersen BS, Quinlan AR. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am J Hum Genet. 2017;100:406–13.
21. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:122.
22. Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. Genome Med. 2020;12:103.
23. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.
24. Amberger JS, Hamosh A. Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. Curr Protoc Bioinform. 2017;58:1.2.1–1.2.12.
25. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet. 2017;100:267–80.

26. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019;176:535-548.e24.
27. Regier AA, Farjoun Y, Larson DE, Krasheninina O, Kang HM, Howrigan DP, et al. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. Nat Commun. 2018;9:4038.
28. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. Nucleic Acids Res. 2020;48:D835–44.
29. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–24.
30. Ware JS, Samocha KE, Homsy J, Daly MJ. Interpreting de novo Variation in Human Disease Using denovolyzeR. Curr Protoc Hum Genet. 2015;87:7.25.1–7.25.15.
31. Gambin T, Akdemir ZC, Yuan B, Gu S, Chiang T, Carvalho CMB, et al. Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. Nucleic Acids Res. 2017;45:1633–48.
32. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012;22:1525–32.
33. Geoffroy V, Guignard T, Kress A, Gaillard J-B, Solli-Nowlan T, Schalk A, et al. AnnotSV and knotAnnotSV: a web server for human structural variations annotations, ranking and analysis. Nucleic Acids Res. 2021;49:W21–8.
34. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2020;22:245–57.
35. Blesa S, Olivares MD, Alic AS, Serrano A, Lendinez V, González-Albert V, et al. Easy One-Step Amplification and Labeling Procedure for Copy Number Variation Detection. Clin Chem. 2020;66:463–73.
36. Seaby EG, Smedley D, Taylor Tavares AL, Brittain H, van Jaarsveld RH, Baralle D, et al. A gene-to-patient approach uplifts novel disease gene discovery and identifies 18 putative novel disease genes. Genet Med. 2022;24:1697–707.
37. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.
38. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. PLOS Comput Biol. 2010;6: e1001025.
39. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023;381:eadg7492.
40. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. Hum Mutat. 2015;36:928–30.
41. Yoon K-J, Nguyen HN, Ursini G, Zhang F, Kim N-S, Wen Z, et al. Modeling a genetic risk for schizophrenia in iPSCs and mice reveals neural stem cell deficits associated with adherens junctions and polarity. Cell Stem Cell. 2014;15:79–91.
42. Qian X, Nguyen HN, Song MM, Hadiono C, Ogden SC, Hammack C, et al. Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. Cell. 2016;165:1238–54.
43. Yoon K-J, Ringeling FR, Vissers C, Jacob F, Pokrass M, Jimenez-Cyrus D, et al. Temporal Control of Mammalian Cortical Neurogenesis by m6A Methylation. Cell. 2017;171:877-889.e17.
44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003;13:2498–504.
45. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10:1523.
46. Yoon JG, Lee S, Cho J, Kim N, Kim S, Kim MJ, et al. Diagnostic uplift through the implementation of short tandem repeat analysis using exome sequencing. Eur J Hum Genet. 2024;32:584–7.
47. Yoon JG, Lim S-K, Seo H, Lee S, Cho J, Kim SY, et al. De novo missense variants in HDAC3 leading to epigenetic machinery dysfunction are associated with a variable neurodevelopmental disorder. Am J Hum Genet. 2024;111:1588–604.
48. Nishioka M, Kazuno A-A, Nakamura T, Sakai N, Hayama T, Fujii K, et al. Systematic analysis of exonic germline and postzygotic de novo mutations in bipolar disorder. Nat Commun. 2021;12:3750.
49. Callaghan DB, Rogic S, Tan PPC, Calli K, Qiao Y, Baldwin R, et al. Whole genome sequencing and variant discovery in the ASPIRE autism spectrum disorder cohort. Clin Genet. 2019;96:199–206.
50. Yuen RKC, Merico D, Cao H, Pellecchia G, Alipanahi B, Thiruvahindrapuram B, et al. Genome-wide characteristics of de novo mutations in autism. npj Genom Med. 2016;1:160271–1602710.
51. Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell. 2013;154:518–29.
52. Eichmüller OL, Knoblich JA. Human cerebral organoids - a new tool for clinical neurology research. Nat Rev Neurol. 2022;18:661–80.
53. Dady A, Davidson L, Loyer N, Sanders T, Januschke J, Storey KG. Engineering fluorescent reporters in human pluripotent cells and strategies for live imaging human neurogenesis. bioRxiv. 2024;2024.05.01.591467.
54. Yoon K-J, Song G, Qian X, Pan J, Xu D, Rho H-S, et al. Zika-Virus-Encoded NS2A Disrupts Mammalian Cortical Neurogenesis by Degrading Adherens Junction Proteins. Cell Stem Cell. 2017;21:349-358.e6.
55. Dawidziuk M, Gambin T, Bukowska-Olech E, Antczak-Marach D, Badura-Stronka M, Buda P, et al. Exome Sequencing Reveals Novel Variants and Expands the Genetic Landscape for Congenital Microcephaly. Genes. 2021;12:2014.
56. Mallard TT, Liu S, Seidlitz J, Ma Z, Moraczewski D, Thomas A, et al. X-chromosome influences on neuroanatomical variation in humans. Nat Neurosci. 2021;24:1216–24.
57. Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, et al. Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. N Engl J Med. 2017;376:21–31.
58. Karaca E, Posey JE, Coban Akdemir Z, Pehlivan D, Harel T, Jhangiani SN, et al. Phenotypic expansion illuminates multilocus pathogenic variation. Genet Med. 2018;20:1528–37.
59. Yoon JG, Yu JW, Shim KW, Kim YO, Lee MG. Multi-locus pathogenic variation identified in a patient with craniosynostosis. Am J Med Genet Part A. 2023;194: e63521.
60. Warner MH, Roinick KL, Arndt KM. Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. Mol Cell Biol. 2007;27:6103–15.
61. Jurynec MJ, Bai X, Bisgrove BW, Jackson H, Nechiporuk A, Palu RAS, et al. The Paf1 complex and P-TEFb have reciprocal and antagonist roles in maintaining multipotent neural crest progenitors. Development. 2019;146:dev180133.
62. Bahrampour S, Thor S. Ctr9, a Key Component of the Paf1 Complex, Affects Proliferation and Terminal Differentiation in the Developing Drosophila Nervous System. G3 (Bethesda). 2016;6:3229–39.
63. Lasser M, Sun N, Xu Y, Wang S, Drake S, Law K, et al. Pleiotropy of autism-associated chromatin regulators. Development. 2023;150:dev201515.
64. Vos SM, Farnung L, Linden A, Urlaub H, Cramer P. Structure of complete Pol II–DSIF–PAF–SPT6 transcription complex reveals RTF1 allosteric activation. Nat Struct Mol Biol. 2020;27:668–77.
65. Kahn RA, Bruford E, Inoue H, Logsdon JM Jr, Nie Z, Premont RT, et al. Consensus nomenclature for the human ArfGAP domain-containing proteins. J Cell Biol. 2008;182:1039–44.
66. Furman C, Short SM, Subramanian RR, Zetter BR, Roberts TM. DEF-1/ASAP1 Is a GTPase-activating Protein (GAP) for ARF1 That Enhances Cell Motility through a GAP-dependent Mechanism. J Biol Chem. 2002;277:7962–9.
67. Müller T, Stein U, Poletti A, Garzia L, Rothley M, Plaumann D, et al. ASAP1 promotes tumor cell motility and invasiveness, stimulates metastasis formation in vivo, and correlates with poor survival in colorectal cancer patients. Oncogene. 2010;29:2393–403.
68. Schreiber C, Saraswati S, Harkins S, Gruber A, Cremers N, Thiele W, et al. Loss of ASAP1 in mice impairs adipogenic and osteogenic differentiation of mesenchymal progenitor cells through dysregulation of FAK/Src and AKT signaling. PLOS Genet. 2019;15: e1008216.
69. Cui J, Wen D, Wang L, Meng C, Wang Y, Zhao Z, et al. CRISPR/Cas9-induced *asap1a* and *asap1b* co-knockout mutant zebrafish displayed abnormal embryonic development and impaired neutrophil migration. Gene Expr Patterns. 2023;49: 119331.

70. van Eyk CL, Corbett MA, Frank MSB, Webber DL, Newman M, Berry JG, et al. Targeted resequencing identifies genes with recurrent variation in cerebral palsy. npj Genom Med. 2019;4:1–11.
71. Esk C, Lindenhofer D, Haendeler S, Wester RA, Pflug F, Schroeder B, et al. A human tissue screen identifies a regulator of ER secretion as a brain-size determinant. Science. 2020;370:935–41.

Publisher's Note