

Identification of a unique, *de novo* MYCBP2 variant in an individual with highly superior autobiographical memory

Andreas Papassotiropoulos^{1,2,3*}, Jana Petrovska^{1,2+}, Andreas Arnold^{1,2+}, Aurora K.R. LePort⁴, Pavlina Mastrandreas^{1,2}, Melanie Neutzner^{1,2}, Virginie Freytag^{1,2,3}, Dmytro Nesterenko⁵, Vaibhav Gharat^{1,2}, Nathalie Schick Tanz^{2,6}, Vanja Vukojevic^{1,2}, David Coynel^{2,6}, Attila Stetak^{1,2,3}, Noëlle Burri^{1,2}, Navid Ghaffari⁷, Claudia Riva⁸, Janani Durairaj^{9,10}, Torsten Schwede^{9,10}, Oliver Bieri^{11,12}, Johannes Gräff¹³, Efthimios M.C. Skoulakis¹⁴, Katharina Henke¹⁵, Sven Cichon^{8,16,17}, Verdon Taylor⁸, Craig E.L. Stark¹⁸, James L. McGaugh¹⁸, Camin Dean⁵, Dominique J.-F. de Quervain^{2,3,6*}

¹ Division of Molecular Neuroscience, Department of Biomedicine, University of Basel, CH-4055 Basel, Switzerland

² Research Cluster Molecular and Cognitive Neurosciences, Department of Biomedicine, University of Basel, CH-4055 Basel, Switzerland

³ Psychiatric University Clinics, University of Basel, CH-4055 Basel, Switzerland

⁴ Windfall Data Inc., San Francisco, CA 94133, USA

⁵ German Center for Neurodegenerative Diseases, Charité University of Medicine, 10117 Berlin, Germany

⁶ Division of Cognitive Neuroscience, Department of Biomedicine, University of Basel, CH-4055 Basel, Switzerland

⁷ Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA

⁸ Department of Biomedicine, University of Basel, CH-4058 Basel, Switzerland

⁹ Department Biozentrum, University of Basel, CH-4056 Basel, Switzerland

¹⁰ SIB Swiss Institute of Bioinformatics, Department Biozentrum, University of Basel, CH-4056 Basel, Switzerland

¹¹ Department of Biomedical Engineering, University of Basel, CH-4123 Allschwil, Switzerland

¹² Division of Radiological Physics, Department of Radiology, University Hospital Basel, CH-4031 Basel, Switzerland

¹³ Laboratory of Neuroepigenetics, Brain Mind Institute, School of Life Sciences, EPFL, CH-1015 Lausanne, Switzerland

¹⁴ Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center "Alexander Fleming", 16672 Athens, Greece

¹⁵ Institute of Psychology, University of Bern, CH-3012 Bern, Switzerland

¹⁶ Institute of Medical Genetics and Pathology, University Hospital Basel, CH-4031 Basel, Switzerland

¹⁷ Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, 52428 Juelich, Germany

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

¹⁸ Department of Neurobiology and Behavior, University of California Irvine, Irvine, CA 92697-4550, USA

+ These authors contributed equally to the study

* Corresponding authors:

Andreas Papassotiropoulos: Division of Molecular Neuroscience, Department of Biomedicine, University of Basel, Birmannsgasse 8, CH-4055 Basel, Switzerland,
andreas.papas@unibas.ch

Dominique J.-F. de Quervain: Division of Cognitive Neuroscience, Department of Biomedicine, University of Basel, Birmannsgasse 8, CH-4055 Basel, Switzerland,
dominique.dequervain@unibas.ch

Abstract

Highly Superior Autobiographical Memory (HSAM) is an extremely rare condition characterized by an individual's unparalleled ability to recall personal past events with exceptional detail and accuracy, including exact dates and days of the week, spanning many decades¹⁻³. The molecular underpinnings of HSAM are unknown. Here, we investigated an individual with HSAM through neuropsychological testing, structural brain imaging, and genetic analyses. HSAM was confirmed as an isolated exceptional cognitive ability, with brain imaging revealing exceptionally large volumes of regions within the hippocampal formation, which have been previously linked to autobiographical memory. Using whole exome sequencing of the HSAM individual and their unaffected parents, we identified a unique *de novo* missense variant in *MYCBP2*, which encodes an E3 ubiquitin-protein ligase^{4,5}. To explore the potential behavioral consequences of this variant, we introduced the homologous variant into *C. elegans*, which resulted in reduced forgetting and increased membrane-bound glutamate receptor in relevant neuronal cells. These findings show that the studied HSAM individual carries a unique, *de novo* missense variant in *MYCBP2*, which reduces forgetting in a model organism. The identification of functionally relevant genetic variants in individuals with superior memory traits has the potential to inform future research into memory-modulating therapies.

Introduction

Highly Superior Autobiographical Memory (HSAM) is an extremely rare condition, with approximately 100 individuals identified over more than a decade of ongoing recruitment by the research group that initially reported this phenomenon^{3,6}. Notwithstanding reports dating back to the 19th century⁷, HSAM first gained attention in 2006 through the case study of “AJ,” a woman who could remember autobiographical events of almost every day of her life since adolescence. That report showcased a form of memory performance vastly different from typical or even other exceptional memory types, such as those observed in savants¹.

Studies show that HSAM and non-HSAM individuals recall a similar number of autobiographical details from the preceding week, indicating comparable encoding of events⁸. However, while recall over extended intervals – spanning the past month, year, and decade – severely declines in people without HSAM, it shows only minimal decay over time in HSAM individuals⁸. This unique trait suggests that HSAM may be driven less by an enhanced capacity to learn and remember and more by a reduced ability to forget^{3,9,10}.

Exceptional autobiographical memory is shared among all HSAM individuals, however, there is heterogeneity in their overall cognitive profiles. Not all exhibit the same memory strength across different types or contents of memory, and variations exist in their cognitive, clinical, and neural profiles^{9,10}. This diversity suggests that multiple molecular and biological mechanisms might underlie HSAM. Notably, molecular and biological mechanisms underpinning HSAM remain unexplored.

Here, we performed whole exome sequencing in a family trio of a male HSAM individual and his unaffected, non-consanguineous parents, hypothesizing that the HSAM individual carries

a functionally relevant *de novo* variant. Further investigations included structural brain imaging, neuropsychological assessments, and experiments in *Caenorhabditis elegans* (*C. elegans*) to investigate the potential behavioral consequences of the identified variant in this model organism.

Results

Neuropsychological assessment of the HSAM participant

The male proband, hereinafter referred to by the fictional initials “JN” for the purpose of anonymization, had participated in previous studies that assessed autobiographical memory of HSAM individuals^{8,9}. We invited JN, now in his sixties, to our research institute to obtain a detailed cognitive profile and a structural brain scan. The study protocol was approved by the Ethics Committee of Northwest and Central Switzerland.

We validated JN’s HSAM status through the administration of the 10 Dates Quiz², which assesses autobiographical memory by asking participants to recall the day of the week, a verifiable public event that occurred within \pm one month of the generated date, and a personal event for ten randomly selected dates, ranging from age fifteen to the present. JN achieved a score of 29 out of a possible 30 points. A score of 20 points or higher qualifies an individual for HSAM, whereas non-HSAM controls typically score below 3 points².

Subsequently, a neurocognitive profile for JN was established using standardized cognitive assessment tools. These included the Wechsler Adult Intelligence Scale (WAIS-IV, age-matched), the Wechsler Memory Scale IV (WMS-IV Older Adult Battery, age-matched), and the Stroop Color and Word Test (age- and education-matched). Testing was conducted across

verbal and non-verbal memory domains, attention, and processing speed (see Supplementary Table 1 for detailed results).

The WAIS-IV was administered to evaluate overall cognitive functioning, encompassing four composite scores: verbal comprehension, perceptual reasoning, working memory, and processing speed. JN achieved an overall Full-Scale Intelligence Quotient (FSIQ) score of 104, which is classified within the average range (percentile rank = 61, 95% CI = 100-108). The breakdown of JN's composite scores was as follows: verbal comprehension: 107 (average, percentile rank = 68, 95% CI = 101-112); perceptual reasoning: 98 (average, percentile rank = 45, 95% CI = 92-104); working memory: 114 (high average, percentile rank = 82, 95% CI = 106-120); processing speed: 97 (average, percentile rank = 42, 95% CI = 89-106).

The WMS-IV was utilized to measure auditory and visual memory, alongside immediate and delayed memory capabilities. JN's scores were: auditory memory: 109 (average, percentile rank = 73, 95% CI = 102-115); visual memory: 104 (average, percentile rank = 61, 95% CI = 98-109); immediate memory: 107 (average, percentile rank = 68, 95% CI = 100-113); delayed memory: 108 (average, percentile rank = 70, 95% CI = 101-114).

The Stroop test, comprising word, color, and color-word subtests, assessed JN's cognitive flexibility and processing speed. His results were: T-score of 55 (average) for word subtest; T-score of 34 (below average) for color subtest; T-score of 45 (average) for color-word subtest; T-score of 52 (average) for the Stroop interference score.

In summary, JN exhibited average cognitive functioning across most domains (Fig. 1a). In the working memory domain JN performed in the top 18% of age-matched peers, indicating above-average capabilities. Performance on the Stroop Color subtest was below average, indicating potential difficulties with some tasks performed under conditions of interference.

In summary, JN's autobiographical memory represents an isolated exceptional cognitive capability which does not extend across other cognitive domains.

Structural brain imaging

We investigated if JN displays distinctive structural brain features compared to a normative dataset. This comparison utilized the UKBiobank database to select a sample of matched controls—specifically men aged 58 to 68—totaling 7755 participants (see Methods). Among these controls, any participants with brain measurement values that indicated measurement artifacts (i.e., values deviating more than 10 standard deviations from the population mean) were excluded. We employed the UKBiobank's acquisition protocol and pipeline for structural imaging (T1 and T2 data), yielding image-derived phenotypes (IDPs) related to cortical volume, thickness, and surface area, as well as volumes of subcortical regions. All volumetric measurements were adjusted for total intracranial volume, and thickness and surface area values were scaled by the respective population averages (see Methods).

To quantitatively assess JN's brain structure, his ranking within the distribution of each IDP among the control group was calculated through standardization of his raw measurements (Z-score). The analysis focused on identifying whether any of JN's brain IDPs exceeded $Z=3$, indicating metrics that can be empirically considered as extremely large values¹¹ (i.e., lying in the outer 0.13% of the respective population distribution) (see Supplementary Table 2 for complete results).

Brain volumetry based on Freesurfer's automatic segmentation¹²⁻¹⁵ showed that the volumes of the right parasubiculum ($Z=3.52$, corresponding to top 0.022% of the reference

population), right head of the granule cell and molecular layers of the dentate gyrus ($Z=3.35$, i.e., top 0.04%), right head of the CA4 region ($Z=3.40$, i.e., top 0.034%), and right head of the CA3 region ($Z=3.20$, i.e., top 0.07%) exceeded $Z=3$, indicating extreme large volumes of these regions of the hippocampal formation in JN compared to the reference population (Fig. 1b,c). No other subcortical IDPs exhibited volumes exceeding $Z=3$ (Supplementary Table 2). Similarly, none of the 327 IDPs for cortical and subcortical volume, the 278 IDPs for cortical thickness, or the 344 IDPs for cortical surface area exceeded $Z=3$ (Supplementary Table 2). In summary, exclusively regions within the hippocampal formation - areas previously linked to autobiographical memory¹⁶⁻²⁰ - exhibited exceptionally large volumes in this individual with HSAM. Of note, among the 7755 UKBiobank participants, none demonstrated volumes exceeding a Z-score of 3 in all four sub-regions of the hippocampal formation, a feature uniquely present in JN.

Detection of de novo variants by whole exome sequencing

Given that the HSAM participant's non-consanguineous parents have no reported HSAM, we hypothesized that this individual's HSAM may result from a heterozygous *de novo* variant. We therefore performed whole exome sequencing on the participant and his unaffected parents to search for *de novo* variants that might be linked to highly superior autobiographical memory. We predefined the following criteria for considering a *de novo* variant as potentially linked to HSAM and warranting further investigation: 1) The variant must be identifiable as *de novo* by two distinct algorithms (i.e., GATK refinement workflow and VarScan2); 2) Given the extreme scarcity of HSAM, we looked for variants that were not reported in the Genome Aggregation Database (gnomAD) encompassing over 125,000 individual exomes; 3) The variant must exhibit a high *in silico* probability of functional impact

(i.e., CADD score >20); 4) The respective gene must be intolerant to missense or loss-of-function variants, as indicated by the gnomAD constraint metrics.

Four variants were identified as *de novo* by both the GATK refinement workflow and VarScan2 (Supplementary Table 3). Of these, only one variant (13-77825365-C-A, GRCh37) met the additional criteria of uniqueness, potential functionality, and non-tolerability to missense or loss-of-function variants of the respective gene. Sanger sequencing confirmed the presence of the variant in the HSAM individual and its absence in his parents. The variant predicts a non-conservative valine to phenylalanine substitution in *MYCBP2*, which encodes an E3 ubiquitin-protein ligase and is member of the PHR (Phr1/MYCBP2, highwire and RPM-1) family of proteins²¹ (Fig. 1d,e). *MYCBP2* is involved in axon guidance and synapse formation and is highly conserved across species²¹⁻²³. The identified variant affects the valine residue at codon 768 of the human reference sequence in the RCC1-like domain (RLD) located in the N-terminal part of the protein. This residue is conserved across vertebrates (e.g., non-human primates, *Mus musculus*, *Canis familiaris*, *Xenopus tropicalis*, *Danio rerio*) and invertebrates (e.g., *Drosophila melanogaster*, *Caenorhabditis elegans*) (Fig. 1e). The three additional *de novo* variants of the HSAM proband affected *DIS3L* (encoding DIS3-like exosome 3'-5' exoribonuclease), *FPGT* (encoding fucose-1-phosphate guanylyltransferase), and *LMCD1* (encoding LIM and cysteine rich domains 1), but did not fulfil the criteria that would have warranted further investigation of these variants (Supplementary Table 3).

CRISPR-edited MYCBP2 variant effects on human induced pluripotent stem cell (hiPSC)-derived neuronal development

MYCBP2 plays significant roles in the development of the nervous system across species²¹. Recently, *MYCBP2* variants with C-terminal orientation have been linked to a neurodevelopmental syndrome that includes developmental delay, intellectual disability, and autistic features²⁴. We assessed the impact of JN's variant on the expression of genes essential for neuronal differentiation from early to late stages of neuronal maturation by introducing the *MYCBP2* variant V768F into hiPSCs using CRISPR-Cas9 homology-directed repair (HDR). To assess the genomic integrity of the edited iPSC lines, we performed a qPCR based genetic analysis that detects karyotypic abnormalities frequently reported in iPSC cells. Our analysis revealed no detectable abnormalities. Furthermore, we assessed the functional pluripotency of our CRISPR/Cas9 edited iPSCs by evaluating their ability to differentiate into each of the three germ layers. All iPSCs tested could differentiate into each of the three germ layers, as shown by positive immunostaining for the ectoderm marker PAX6, the mesoderm marker SMA, and the endoderm marker SOX17. We then performed RNA sequencing in an edited heterozygous and a wild-type (WT) clone at days 19, 35, 55, 70 and 90 of neuronal differentiation (see Methods). Principal component analysis (PCA) and analysis of sample clustering of the transcriptomic data showed clear separation of early and late stages of neuronal differentiation of both edited heterozygous and WT cells (Fig. 2a,b). Both clones exhibited the expected temporal gene expression profile necessary for neuronal differentiation and maturation (Fig. 2c-e), with no alterations in these expression profiles observed as a result of the variant. Likewise, the variant had no effect on *MYCBP2* expression on any of the assessed days of neuronal differentiation (comprehensive gene expression data are provided in Supplementary Table 4).

Effect of CRISPR-edited MYCBP2/rpm-1 variant on forgetting in C. elegans

To investigate the potential behavioral consequences of the *MYCBP2* variant, we utilized *C. elegans* as an *in vivo* genetic model system and conducted learning and long-term memory tests. *C. elegans* harbors a single highly conserved orthologue of *MYCBP2*, known as *rpm-1*, which plays a crucial role in regulating axonal development and synaptic strength²¹. Hence, we used CRISPR/Cas9 genome editing to introduce JN's *de novo* variant into the corresponding conserved residue of RPM-1 in *C. elegans*.

We generated animals in which endogenous *rpm-1* harbors the variant (*rpm-1(V682F)*) analogous to JN's *de novo* *MYCBP2* variant V768F and confirmed the CRISPR edit by sequencing. The variant is located in the N-terminal RCC1-like domain (RLD) of *rpm-1* (Fig. 3a). Visualization of the RLD structure based on homology modeling (see Methods) indicated that the V682 residue forms the tip of a blade of the 7-bladed β -propeller structure of RLD with a hydrophobic interaction to a proline from the next blade (Fig. 3a). The position of V682 on the surface of the RLD domain suggests that the bulkier phenylalanine variant may sterically interfere with binding partners of this domain. The homology modeling analysis also showed that the mutated residue is adjacent to an intrinsically disordered region of the protein, predicted for residues 662-678 (Fig. 3a). Such regions, lacking a stable tertiary structure, are particularly sensitive to changes in their environment, including temperature, which affect the dynamics, conformation and interaction potential of these regions²⁵⁻²⁷. We therefore conducted experiments under two different temperature conditions.

Mutants were tested in an olfactory learning assay, in which worms are trained to avoid diacetyl (DA), which they are normally attracted to since it signifies food, by pairing it with an aversive condition of food deprivation. Because olfactory conditioning relies on normal detection of volatile attractants, we first tested the chemotaxis of *rpm-1(V682F)* animals toward different odorants. The chemotaxis of *rpm-1(V682F)* mutants to two different volatile

attractants (DA, isoamyl alcohol (IAA)) and a repellent (octanol) was comparable to the response of the wild-type (WT) N2 strain, regardless of the growth temperature (Supplementary Fig. 1a). Furthermore, both WT and *rpm-1(V682F)* mutants had normal locomotor behavior (Fig. 3b) and responded similarly to food (Supplementary Fig. 1b), indicating that *rpm-1(V682F)* mutants have no obvious sensory or motor defects. In addition, WT animals and *rpm-1(V682F)* mutants exhibited similar body lengths (Fig. 3c), suggesting that the variant does not impact normal growth of these worms. In contrast, *rpm-1* loss-of-function mutants (*rpm-1(lf)*) and animals harboring variants that severely affect RPM-1 ubiquitin ligase activity (*rpm-1(LD)*, i.e., ligase-dead)²⁸ showed decreased locomotor activity and reduced body length (Fig. 3b,c).

In the olfactory learning assay, both unconditioned WT and *rpm-1(V682F)* mutant animals demonstrated robust chemotaxis towards DA (Fig. 3d). Moreover, following a 1-hour period of food deprivation in the presence of DA (aversive conditioning), both WT and *rpm-1(V682F)* mutants exhibited normal associative learning, evidenced by a significantly diminished attraction to DA to a similar extent (Fig. 3d). Both strains exhibited similar attraction to DA after food deprivation and after exposure to DA alone (in the presence of ample food) (Supplementary Fig. 1b).

To assess the impact of the *V682F* variant on long-term associative memory (LTAM), we evaluated animals' retention of conditioned behavior over time as described previously²⁹. We observed significant differences in LTAM retention between *rpm-1(V682F)* mutant animals and WT worms following a 24-hour delay period at 25°C ($P=7.8 \times 10^{-8}$; Fig. 3d). After 24 hours, WT animals demonstrated a complete loss of LTAM retention, as their chemotaxis towards DA was almost identical to that of unconditioned animals of the same age (Fig. 3d, last condition, i.e., 24h after food deprivation for 1h in the absence of DA). In contrast, the

chemotaxis towards DA of *rpm-1(V682F)* mutants 24h after conditioning was significantly lower than that of unconditioned *rpm-1(V682F)* mutants of the same age ($P=1.8*10^{-7}$; Fig. 3d, last condition, i.e., 24h after food deprivation for 1h in the absence of DA). No difference in LTAM between conditioned strains was observed following a 24-hour delay period at 20°C (Supplementary Fig. 1c). In summary, the V682F variant inhibits the physiological loss of LTAM in a temperature-dependent manner.

Abundance and mobility of the GLR-1 receptor in C. elegans CRISPR-edited MYCBP2/rpm-1 mutants

Previous studies in nematodes have demonstrated the importance of glutamatergic neurotransmission in learning and memory processes, including forgetting. Of the many glutamate receptor types in *C. elegans*, GLR-1 is the best investigated and is necessary for associative memory²⁹⁻³¹. GLR-1 encodes a receptor subunit of a non-NMDA excitatory ionotropic iGluR subtype and shows high homology with human α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors GRIA1-4. Importantly, total and membrane surface-bound GLR-1 receptor levels correlate with associative memory performance³². We used a transgenic strain expressing functional GLR-1 protein fused to both superecliptic pHluorin (SEP, a pH-sensitive form of GFP) and mCherry at the extracellular N-terminal domain (SEP::mCherry::GLR-1) specifically in the AVA interneuron³³, which is a key regulator of LTAM in *C. elegans*³⁴. CRISPR/Cas9 genome editing was used to introduce JN's *de novo* variant in this strain to evaluate the influence of the variant on GLR-1 abundance and dynamics, which were assessed by fluorescence recovery after photobleaching (FRAP). We investigated the recovery of mCherry (representing total GLR-1) and SEP (representing membrane-bound GLR-1) fluorescence after photobleaching to

measure the amount and mobility of total and plasma membrane surface-bound fractions of GLR-1 in the proximal region of the AVA axon. After photobleaching, the signal of mCherry as well as membrane surface-bound SEP are bleached, whereas SEP fluorescence of non-membrane-bound GLR-1 remains undetectable due to the quenched state of the fluorophore in acidic endosomes^{33,35}.

We found significantly higher levels of both total and membrane surface-bound GLR-1 in the AVA interneuron of *rpm-1(V682F)* mutants as compared to WT animals (Fig. 3e).

Quantification of the GLR-1::mCherry and the GLR-1::SEP FRAP signal dynamics demonstrated that the membrane surface-bound GLR-1 mobile fraction was significantly higher in mutant animals compared to WT controls (Fig. 3f). Total GLR-1 was also higher in mutants compared to WT, but this difference did not reach statistical significance. In summary, these experiments suggest that the variant leads to increased GLR-1 receptor abundance and membrane surface-bound mobility.

Axonal development in CRISPR-edited MYCBP2/rpm-1 C. elegans mutants

In addition to regulating synaptic strength, MYCBP2 plays a significant role in axonogenesis²¹. In nematodes, disruption of RPM-1 function leads to axon termination defects, characterized by the overgrowth of sensory and motor axons^{36,37}. This phenotype is likely attributable to altered ubiquitination and subsequent impaired autophagosome formation at axon termination sites^{24,28}. In *C. elegans*, C-terminal mutations compromise RPM-1 ubiquitin ligase activity and result in axonal abnormalities²⁴.

We tested the effect of JN's N-terminal variant on axon termination in the anterior lateral mechanosensory (ALM) and posterior lateral mechanosensory (PLM) neurons utilizing a strain that expresses GFP in these neurons (*muls2* [*mec-7p::GFP*])³⁸. Conventional *rpm-1*

loss-of-function mutants (*rpm-1(lf)*) and animals harboring variants that severely affect RPM-1 ubiquitin ligase activity (*rpm-1(LD)*) displayed severe axon termination defects (Supplementary Fig. 1d). In contrast, axon termination in *rpm-1(V682F)* mutants was not impaired under physiological growing conditions (Supplementary Fig. 1d) and was comparable to that of WT animals under conditions known to affect axon development (Supplementary Fig. 1d). These results indicate that JN's *de novo* variant does not alter *in vivo* axon development linked to ubiquitin ligase activity.

Discussion

We report that JN carries a unique, *de novo* missense variant in *MYCBP2*, which reduces forgetting in *C. elegans*. JN's HSAM was confirmed through the 10 Dates Quiz, where he achieved a score of 29 out of 30 points. In all other standard cognitive tasks assessing intelligence and cognitive capacities across verbal and non-verbal memory domains, attention, and processing speed, JN scored within the expected performance range (Fig. 1a). This indicates that his HSAM represents an isolated exceptional cognitive capability, not extending to other cognitive domains. The subsequent quantification of JN's brain volumetric data was conducted to identify, in an unbiased manner, any structural brain features that might be linked to his extreme autobiographical memory capacity. Using a sample of 7755 male participants from the UKBiobank aged 58 to 68 years as a reference distribution, we observed atypically high volumes in several sub-regions of the hippocampus (Fig. 1b,c). While confidence in the precise localization of the structural changes must always be tempered by the limitations of current MRI techniques³⁹ and automatic segmentation algorithms, these increases appear to derive from the right parasubiculum, right head of the

granule cell and molecular layers of the dentate gyrus, right head of the CA4 region, and right head of the CA3 region. This observation did not apply to any other structural brain features measured. Of note, among the 7755 UKBiobank participants, none demonstrated extreme volumes in all four sub-regions of the hippocampal formation, a feature uniquely present in JN. Importantly, previous research has shown that the volumes of these sub-regions correlate positively with autobiographical memory in average- and lower-performing individuals¹⁶⁻²⁰. For example, bilateral CA2/3 volume is positively associated with autobiographical memory recall performance in healthy individuals with rather low memory recall performance¹⁷. In young, healthy individuals, high-resolution MRI showed that the volumes of the dentate gyrus, CA2/3 regions, and the subiculum were positively correlated with episodic autobiographical memory¹⁹. Positive correlations between the amount of autobiographical memory internal details produced over time and the volumes of the left pre/parasubiculum and the right CA3/2 were also reported in healthy young subjects¹⁶. Thus, the unbiased, brain-wide comparison of JN's volumetric data with a large reference population suggests a connection between his extraordinary cognitive ability and hippocampal regions known to be critical for autobiographical memory. Supporting the role of the hippocampus in HSAM, an fMRI study in eight individuals with this condition revealed that their ability to access autobiographical memory was associated with enhanced functional connectivity between the prefrontal cortex and the hippocampus, compared to controls⁴⁰.

MYCBP2 encodes MYC binding protein 2, a member of the highly conserved PHR (Phr1/MYCBP2 in vertebrates, Highwire in *Drosophila*, RPM-1 in *C. elegans*) family of proteins. PHR proteins consist of several conserved domains (RCC1-like GEF domain (RLD),

PHR family specific domains (PHR), RAE-1 binding domain (RBD), FSN-1 binding domain 1 (FBD1), Myc binding domain (MBD) and RING-H2 ubiquitin ligase domain (RING)) (Fig. 1e), represent intracellular signaling hubs with positive and negative impact on multiple downstream signaling pathways (reviewed in^{21,41}), and have critical roles in axonogenesis, neuronal growth, and synaptic function^{21-23,41}. While C-terminal domains (FBD1, RING) are critical for the ubiquitin E3 ligase function of PHR proteins, the downstream events regulated by the N-terminal domain harboring JN's *de novo* variant are not well understood.

Deletion of *MYCBP2* and its homologues in various animal models leads to neurodevelopmental abnormalities. For example, loss of *rpm-1* function in *C. elegans* results in developmental alterations of neuronal morphology and function^{22,36}, and *rpm-1* mutants exhibit defects in spontaneous locomotion⁴². *highwire* *Drosophila* mutants exhibit increased axon branch length and walking defects²³. In mice, constitutive loss of *Phr1* function results in severe neurodevelopmental defects and death shortly after birth^{43,44}. Interestingly, loss of function of *MYCBP2* and its homologues can also enhance memory capacity and neuroprotection in animal models: *highwire* mutants exhibit facilitated long-term memory formation⁴⁵, and conditional loss of *Mycbp2* function in mice results in prolonged survival of injured axons in both the peripheral and central nervous systems⁴⁶. These dual outcomes highlight the gene's complex role in neural circuit formation and cognitive function, suggesting that specific alterations in human *MYCBP2* could modify synaptic efficiency and stability, thereby impacting memory retention.

The observed reduction of forgetting in *C. elegans* carrying the human variant aligns with JN's HSAM phenotype, suggesting a potential evolutionary conservation of the gene's role in

memory processes. Notably, the variant induces an increase in GLR-1 in *C. elegans*, implicating enhanced glutamatergic signaling. A substantial body of research has demonstrated the critical role of glutamatergic signaling in learning, memory, and forgetting (reviewed, for example, in⁴⁷⁻⁴⁹). Located in the N-terminal RCC1-like domain (RLD) of *MYCBP2*, the variant does not induce developmental phenotypes in *C. elegans*, suggesting that it does not interfere with the E3 ligase function of *MYCBP2*, or does so in a manner insufficient to trigger developmental abnormalities. Moreover, the variant did not influence the expression of genes essential for neuronal differentiation and maturation in hiPSC-derived neuronal cultures. This specificity may explain why the variant impacts memory without broader developmental consequences. Of note, JN's variant inhibited the physiological loss of LTAM in a temperature-dependent manner. This implies that, in addition to the variant, specific permissive environmental conditions may be required to elicit the very rare HSAM phenotype.

At this stage, it is important to emphasize that the identified variant is confined to a single individual with HSAM and does not inform about other cases of HSAM. Additionally, the specificity of the variant's impact on autobiographical memory warrants further research on the mechanisms by which the *MYCBP2* variant selectively influences this type of human memory without affecting other cognitive domains. Nevertheless, the identification of functionally relevant genetic variants in individuals with superior memory traits opens a new field with the potential to inform future research into memory-modulating therapies.

Methods

Brain imaging

UK Biobank participants

Data for the control population were obtained from the UK Biobank through research application number 101531. Informed consent was obtained by the UK Biobank for all participants⁵⁰. The data release included 502'413 participants. To match the age and sex of our HSAM participant, we restricted the sample to participants fulfilling the following criteria:

- participated to baseline visit as well as to the imaging visit (instance 2)
- self-reported overall health rating "Excellent", "Good" or "Fair" (data-field 2178)
- sex "Male" (data-field 31)
- age at imaging visit greater than 58 and lower than 68 years old (data-field 21003)

This corresponded to 9'356 participants.

HSAM participant: MRI acquisition and pre-processing

Structural MRI data was acquired on a 3T Siemens Magnetom Prisma scanner with a 20-channel RF receive head-neck coil, following the UK Biobank imaging protocol⁵¹. Recent evidence confirms high inter-site reliability of the Skyra-based UK Biobank imaging protocols on the Prisma platform, particularly for grey matter phenotypes such as volume and cortical thickness, which are central to this study. Protocols were adapted to minimize variability between Skyra and Prisma scanner data⁵².

The T1 structural image was a 1mm isotropic resolution three-dimensional magnetization-prepared rapid-acquisition gradient-echo (3-D MPRAGE) acquisition with the following parameters: field of view = 256x256 mm; time of acquisition = 4min53s; repetition time = 2000ms; inversion time = 880ms; echo time = 2.03ms; flip angle = 8°; 208 sagittal slices. The T2-weighted image consisted of a fluid-attenuated inversion recovery (FLAIR) image, with the 3D SPACE optimized readout and the following parameters: field of view = 256x256 mm; time of acquisition = 5min52s; repetition time = 5000ms; inversion time = 1800ms; echo time = 397ms; flip angle = 120°; slice thickness = 1.05mm; 192 sagittal slices.

Data were pre-processed using the pipeline for brain imaging processing of the UK Biobank⁵³ (https://git.fmrib.ox.ac.uk/falmagro/uk_biobank_pipeline_v_1.5), via an adapted Docker container (<https://git.fmrib.ox.ac.uk/paulmc/fbp-dockerfile>, <https://github.com/dcoynel/fbp-dockerfile>). Correction of gradient distortion was performed as a first step of the T1 pipeline, using the scanner's own gradient coefficients (*coeff.grad*)⁵³.

Image-derived phenotypes

Of interest for this project were the image-derived phenotypes (IDPs) provided by the cortical and subcortical pipeline from FreeSurfer v6.0.0¹²⁻¹⁴. Both the T1 and T2 images were used as input for FreeSurfer¹⁵. The resulting IDPs consisted of cortical and subcortical volumes, cortical thickness and cortical surface area. Concerning the cortical data, we included data from all cortical parcellations available: the Desikan atlas, the Desikan-Killiany–Tourville atlas (denoted DKT), the pial surface parcellation the using Desikan-Killiany atlas (denoted pial) and the Destrieux atlas (denoted a2009s). The following subcortical structures were further sub-divided: amygdala (nuclei volumes), hippocampus (subfield volumes), thalamus (nuclei volumes), brainstem (substructure volumes). Subjects without estimated

total intracranial volume were excluded (data-field 26521; N=22). The following IDPs were excluded from further analyses because more than 5% of the UK Biobank participants had a value of 0: volume of 5th-Ventricle (data-field 26525) and volume of non-WM-hypointensities (data-field 26529). For a given IDP, a subject was excluded if he had a value that was outside the population's mean \pm 10 standard deviations. This was the case 138 times, distributed over 61 subjects and 70 IDPs.

The final data used for statistical analyses consisted of 1070 IDPs for 7755 UK Biobank subjects.

MRI statistical analyses

The data were further corrected for total brain volume and z-transformed. For correcting for total brain volume, we considered for each IDP the residuals of a linear model including only the intercept and the total intracranial volume (TIV). Those values were then Z-transformed using the mean and standard deviation of the UK Biobank participants only. The resulting Z-scores then represent the amount of standard deviation a given subject is from the TIV-corrected control population's mean. The analysis focused on identifying whether any of JN's brain IDPs exceeded $Z=3$, indicating metrics that can be empirically considered as extremely large values¹¹ (i.e., lying in the outer 0.13% of the respective population distribution).

Supplementary Table 2 reports all Z values (positive and negative).

Human genetics methods

Whole exome sequencing (WES)

WES for the HSAM participant and his unaffected parents was performed together with additional 253 non-HSAM healthy young adults, from an unrelated cohort with European

ancestry. Blood sampling, DNA isolation and library preparation was done as described in ⁵⁴. WES was done on the Illumina HiSeq platform (paired-end reads). Samples were processed in a double-randomized manner. Samples were randomly pooled and each pool was sequenced in multiple lanes. For each sample, over 9 Gb of sequence were generated. The SureSelect XT Human All Exon V6 + UTR kit (Agilent) was used for targeting regions. Each sample's sequence was mapped to the hg19 human reference genome, using BWA (Burrow-Wheeler Alignment) v0.7.12 ⁵⁵. Duplicates were flagged with Picard v1.135. Local realignment and base quality score recalibration were performed with the Genome Analysis Toolkit (GATK) ⁵⁶ v3.4-0, following the standard GATK protocol (<https://gatk.broadinstitute.org>). Haplotype Caller was used for variant calling and recalibration. We chose 99% sensitivity for a variant to be 'true' based on an adaptive error model and filtered out false-positive variants using this threshold. Variants with any missing values were also filtered out.

De novo variant identification

De novo variant identification based on WES data is highly susceptible to false positive findings. Therefore, we took the following steps to nominate potential *de novo* variants with high confidence. First, we ran the Genotype Refinement Pipeline (<https://gatk.broadinstitute.org/hc/en-us/articles/360035531432-Genotype-Refinement-workflow-for-germline-short-variants>), GATK 3.4-0, in order to improve the accuracy of genotype calls and to filter calls that are not reliable enough for downstream analysis. This workflow is recommended as particularly useful for analysis concerning transmission or *de novo* origin of variants in a family. Specifically, the "CalculateGenotypePosteriors" tool was used to calculate the posterior probabilities for each sample at each variant site, including

family priors, i.e., pedigree data, for the HSAM trio. Next, the “VariantFiltration” tool was used to identify genotypes with $GQ \geq 20$ based on the posteriors, indicating that there is a 99% chance that the genotype in question is correct. Next, “VariantAnnotator” was used to tag possible *de novo* variants. High confidence *de novo* sites were defined as having $GQs \geq 20$.

To further refine the *de novo* variants search, we also applied another *de novo* variants filtering pipeline, independently from the GATK pipeline. We used SAMtools 1.9 for variant calling and VarScan v2.4.1⁵⁷ for identification of *de novo* variants. For comparison of the two pipelines, see^{58,59}. Variants identified as high-confidence *de novo* by both GATK and VarScan2 were considered for further analyses.

Annotation of genetic variants

Genetic variants were annotated based on their prevalence in the human population, functional consequences and deleteriousness. Considering that the HSAM phenomenon is extremely rare, we only kept variants which did not appear in the Genome Aggregation Database (gnomAD), v2.1.1⁶⁰. The gnomAD v2.1.1 database consists of 125748 exomes and 15708 genomes worldwide. Furthermore, we filtered out variants with any carriers in 253 unrelated non-HSAM individuals sequenced together with the HSAM trio, or any carriers in unrelated non-HSAM individuals of European ancestry from six additional WES batches of our exome studies ($N_{batch1}=94$; $N_{batch2}=94$; $N_{batch3}=376$; $N_{batch4}=282$; $N_{batch5}=187$; $N_{batch6}=285$), processed in the same manner as the WES batch including the HSAM participant. We took this step to exclude false-positive variants due to potential sequencing or data processing noise, specific to our data and processing pipeline.

The impact of variants on genes was determined using the Ensembl Variant Effect Predictor (VEP) ⁶¹. Variants with high impact (splice acceptor variant, splice donor variant, stop gained, frameshift variant, stop lost, start lost) or moderate impact (inframe insertion, inframe deletion, missense variant, protein altering variant) on protein coding genes, as defined by VEP (https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html), were considered of interest. In cases where one variant was assigned to multiple genes, the functional consequences for each gene were considered separately. If a variant had multiple functional consequences for a single gene, the functional consequences with highest impact were taken for annotation.

Sanger sequencing of the locus harboring the de novo MYCBP2 variant in the trio

The *MYCBP2* locus was amplified using the primer pairs 5'-TCATTGTCTCTTACCCTCCGG-3'/5'-TGGGAGGCTTTCTTTGGTCT-3' and 5'-TGCAAAGAAATGTGTTCTGCA-3'/5'-AGCAATGGATGAAGACCTGGA-3' via polymerase chain reaction (PCR) employing the KAPA HiFi HotStart PCR Kit (standard Protocol Mix, KK2501). The PCR conditions were as follows: initial denaturation at 95°C for 3 minutes, followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 1 minute, with a final extension at 72°C for 5 minutes and a hold at 4°C. Following amplification, the PCR products were excised from the gel and purified using the Zymoclean Gel DNA Recovery Kit (D4001). The purified gel fragments were subsequently cloned into the pGEM[®]-T Easy Vector System (A1360) using the standard 10 µl ligation protocol. The ligation mix was used to transform JM109 High Efficiency Competent Cells under selection conditions (LB medium supplemented with ampicillin, IPTG, and X-Gal). Colonies presenting with white coloration on X-Gal plates were selected and cultured overnight in 5 ml of LB medium containing

ampicillin. Plasmid DNA was then extracted from these cultures using the ZymoPURE Plasmid Miniprep Kit (D4208T). Finally, the extracted plasmids were Sanger sequenced using the standard T7 primer (TAATACGACTCACTATAGGG). Sequencing was conducted at Microsynth.

C. elegans experiments

Strains

The *C. elegans* Bristol strain, variety N2, was used as the WT control. The following strains with integrated transgenes were used: *muls32[mec-7p::GFP]* and *akls201[rig-3pr::SEP:GFP::mCherry::glr-1]*. The following CRISPR-engineered strains were used: *rpm-1(utr86[V682F])*, *rpm-1(utr93[lf])* and *rpm-1(utr112[C3535A; H3537A; H3540A])*. Due to very close proximity of the *rpm-1* locus and the integration site of *akls201*, the point variant for the V682F substitution had to be introduced *de novo* in *rpm-1* in the *akls201* background, *rpm-1(utr103[V682F]); akls201*. A list of all the strains used in this study is shown in Supplementary Table 5. All strains were outcrossed at least four times. Standard methods were used for maintaining and manipulating *C. elegans*⁶². Animals were grown at either 20°C or 25°C. The necessity to shift animals to 25°C was due to the apparent temperature-sensitive phenotype caused by the V682F substitution.

CRISPR/Cas9 engineering

A DNA-based (plasmid/oligo) co-CRISPR strategy was used to obtain the *rpm-1* modifications of interest and *dpy-10* served as the co-CRISPR marker⁶³. Cas9 was expressed from pDD162⁶⁴. The Q5 Site-directed mutagenesis kit (New England Biolabs) was used to clone

the sgRNA targeting sequence of interest into a version of pDD162 lacking the Cas9 sequence⁶⁴. In this way the sgRNA concentration could be adapted independently of Cas9. Plasmids were purified using the Midiprep kit from Qiagen and PAGE-purified repair oligos were ordered from Microsynth AG (www.microsynth.com). pDD162 and sgRNA plasmids were injected at 25-50 ng/ μ l and DNA repair oligos at 20-40 ng/ μ l. A micropipette puller (Sutter Instrument) was used together with Borosilicate Glass Capillaries (World Precision Instruments, Inc.) to prepare injection needles. The injection mix was prepared freshly and 2 μ l were transferred into the needle using Microloader tips (Eppendorf). The injection setup consisted of the FemtoJet 4i pump (Eppendorf), the DMi8 microscope (Leica) and a micromanipulator (Narishige). Injected worms were transferred to 6 cm plates, stored at 15°C for 2 hours and then shifted to 25°C. Three days later, animals showing the co-CRISPR phenotype (rollers) were singled and genotyped the day after for the presence of the modification of interest. The list of oligos and plasmids used to obtain and genotype the different CRISPR/Cas9 modifications is shown in Supplementary Table 5.

Measuring body length and motility

A mixed population of animals was grown at 25°C. Five L4-stage larvae were singled onto seeded 6 cm NGM plates for each strain. Sixteen hours later the animals were imaged (Leica M50 with FlexaCam C1) and the body length was measured in Fiji⁶⁵. To quantify motility, the same animals were “activated” by picking them gently to the center of the plate using an eyelash. After two minutes of recovery the body bends were counted for one minute. The experiment was repeated three times.

Behavioral assays

Chemotaxis towards different chemicals, negative olfactory conditioning, as well as long-term associative memory (LTAM) testing were performed as previously described^{29,66,67} with the following adaptations: 1) Animals usually went through only one round of conditioning instead of the regular two rounds for LTAM, as this was sufficient to observe the memory retaining phenotype of *rpm-1(V682F)*. 2) Animals were shifted to 25°C for the 24 h recovery phase after conditioning and only remained at the regular 20°C during recovery for specific control experiments.

Analysis of axon termination defects

ALM and PLM axon termination defects were analyzed using the *muls32[mec-7p::GFP]* transgenic line³⁸. Staged animals were grown at either 20°C or 25°C. Day-1 adults were picked into a drop of M9 / 0.2% sodium azide on a 3% agarose pad on a microscopy slide and covered with a cover slip. ALM and PLM axon termination defects were scored as normal or defective under a fluorescent dissecting microscope (Leica M165 FC). ALM axon termination was considered defective, if a short or long hook was observed⁶⁸. PLM axon termination was considered defective, if it extended beyond the ALM cell body or ended in a hook⁶⁸. One hundred animals per genotype and temperature were scored.

*Quantifying *glr-1* abundance and dynamics*

L4 larvae were picked and shifted to 25°C the day before imaging. Microscopy on the SEP and mCherry-tagged *glr-1* line⁶⁹ and subsequent analysis were carried out as described previously³².

Structural modeling

Three structural models of the RCC1-like domain (RLD) from the RPM-1 protein were generated using the SWISS-MODEL automated modeling workflow^{70,71}, based on template Protein Data Bank (PDB) structures 8gqe, 7vvg, and 1a12. The resulting models exhibited QMEANDisCo⁷² global scores of 0.59, 0.58, and 0.54, respectively, and local confidence scores of 0.67, 0.69, and 0.58 for the V682 residue. The V382F mutation was introduced for visualisation using FoldX⁷³.

Cell culture experiments

Human induced pluripotent stem cell culture

Human induced pluripotent stem cell (hiPSC) line NAS2⁷⁴ was cultured and expanded on TC-treated plates coated with Matrigel® hESC-Qualified Matrix (Corning, Cat. #354277) in mTeSR™ Plus medium (STEMCELL Technologies, Cat. #100-0274). hiPSCs cells were maintained at 37 °C, 95% humidity, 5% CO₂ with medium changes every two days, and split when colonies were big with dense centers (within 3–5 days of seeding). Cells were passaged with ReLeSR™ (STEMCELL Technologies, Cat. #100-0483) to obtain small aggregates of colonies and plated in the presence of 10 μM Y27632 ROCK inhibitor (Miltenyi Biotec, Cat. #130-104-169).

Generation of MYCBP2 mutant hiPSCs

To introduce the variant of interest (GTT to TTT) in the NAS2 cell line, a single-guide RNA was designed using CRISPOR (<http://crispor.gi.ucsc.edu/>). Selection of the guide RNA was based

on the proximity of the cleavage site to the desired gene editing site. Additional criteria included the highest chance of successful targeting and the least chance of off-target editing reported by the software. The guide RNA sequence (5'-TGCATGGTTTGCCTGTCTG-3') was synthesized as a single-guide RNA (sgRNA) using a commercial synthesis service (ThermoFisher Scientific, Cat. #A35534). The CRISPR-associated protein 9 (Cas9) was procured as a recombinant protein (ThermoFisher Scientific, TrueCut™ Cas9 Protein v2, Cat. #A36498). The HDR template (5'-
AGATGAAAAGATGAGAAGTCTATGATGTGCCCTCCAGGCATGCACAAATGGAAGCTGGAGCAGTGT
ATGTTTTGTACAGTGTGTGGAGACTGTACAGGTTATGGAGCCAGCTGTGTCAGTAGTGGACGGCCAGA
CAGAGT-3') was designed by using the Integrated DNA Technologies (IDT) Alt-R™ CRISPR HDR Design Tool (<https://www.idtdna.com/pages/tools/alt-r-crispr-hdr-design-tool>) and synthesized as a single-stranded oligodeoxynucleotide (ssODN) by IDT . The HDR donor template included the desired variant, four silent variants to prevent re-cutting by Cas9 to enhance HDR efficiency, along with flanking homology arms (~70 bp) identical to the sequences adjacent to the targeted site. Codons with similar usage in human to the wild-type codons were chosen to insert the silent variants.(ssODN) by IDT(Integrated DNA Technologies). To enhance HDR efficiency, the donor template included silent variants to prevent re-cutting by Cas9.

NAS2 iPSCs were cultured in 6-well plates (Corning, Cat. #3516) for 3 days to reach a confluence of 50–60% (approx. 2×10^6 cells/well). The ribonucleoprotein (RNP) complex was prepared by incubating 20 pmol of custom sgRNA and 3 µg of TrueCut™ Cas9 Protein v2 at room temperature for at least 20 min, together with 60 pmols of ssODN donor for the repair condition, in a final volume of 10 µL of P3 Primary Cell Nucleofector® Solution. During RNP

complex incubation, cells were washed with DBPS (Gibco, Cat. #14040-117) and dissociated using StemPro™ Accutase™ Cell Dissociation Reagent (Gibco, Cat. #A11105-01) for 4–5 min. For each transfection condition, 300 000 cells were used and resuspended in 10 µL of P3 Primary Cell Nucleofector® Solution (Lonza, V4XP-3032). Cells were added to the mix with the RNP complex and the 20 µL reaction was transferred to a 16-well nucleocuvette strip (Lonza, Cat. #V4XP3032) and nucleofected using LONZA 4D-Nucleofector® with the CA137 pulse program. Post-transfection, cells were immediately transferred to alaminin 521-coated (BioLamina, Cat. #LN521) 24-well plate in mTeSR1™ (STEMCELL Technologies, Cat. #85850) medium supplemented with 10 µM Y-27632 to enhance cell survival. After 24 hours, the medium was replaced with fresh mTeSR1™ medium.

Three days after electroporation, around 100.000 cells from each condition were collected for genotyping and the remaining cells were re-plated, expanded and frozen. Genomic DNA (gDNA) was extracted by resuspending cell pellets in 50 µL of lysis buffer (50mM KCl (ThermoScientific Chemicals, Cat. #424090010), 10mM TRIS Base (Sigma-Aldrich®, Cat. #T1503), 0.1% Triton X-100 (PanReac AppliChem, Cat. #A4975,0500), 10mg/ml proteinase K (CarlRoth, Cat. #7528.2)) and incubating them in a thermo cycler at 68°C for 15 min, followed by a 15 min incubation at 95°C. gDNA concentration was measured using a NanoDrop™ 2000. 100 ng human iPSC gDNA was used as a template for genotyping PCRs. gPCR primers were designed to amplify a 305 bp product flanking the CRISPR/Cas9 cut site, with the 3' nucleotide of the forward primers corresponding either to the WT or the mutant nucleotide. Primers were as follows: primer pair for detection of WT allele: F1 5'-CTGGAGCAGTGCATGG-3' with R1 5'-GGCAACCTTAGAGATGAC-3' and primer pair for detection of edited allele: F2 5'-GCTGGAGCAGTGTATGT -3' and R1 5'-GGCAACCTTAGAGATGAC-3'. The PCRs were performed using FIREPol® DNA Polymerase (Solis

Biodyne, Cat. #01-02-00500). The PCR annealing temperature for the amplification of the WT and the edited allele was 62°C. PCR fragments were run on 2% agarose gels to confirm single bands. Following this initial screening, edited cells were plated onto laminin 521-coated 3.5 cm plates at a density of 2000 cells/plate and allowed to expand until single colonies emerged. Single colonies were then picked onto laminin 521-coated 96-well plates and their genotypes were assessed using PCR. Sanger sequencing was performed on positive clones in order to discriminate heterozygous and homozygous variants. The hPSC Genetic Analysis Kit (STEMCELL Technologies, Cat. # 07550) was used to test for possible karyotypic abnormalities of the clones.

To assess potential off-target effects, regions of homology to the sgRNA sequence were predicted using CRISPOR. We selected the top 9 potential off-target regions and amplified each one by targeted PCR of gDNA from edited and unedited clones and further analyzed them by Sanger Sequencing. Each amplicon was purified using the QIAquick PCR Purification Kit (Qiagen, Cat. #28104) and the concentration was determined using a Nanodrop 2000 Spectrophotometer. All off-target regions and primer sequences can be found in Supplementary Table 6.

Pluripotency and Three Germ Layer Differentiation Test

The pluripotency of WT and MYCBP2 mutant NAS2 cells was verified using staining for expression of the conventional surface markers NANOG and OCT3/4, and through differentiation into the three germ layers. To differentiate iPSCs into each of the three germ layers, cells were passaged as single cells using StemPro™ Accutase™ Cell Dissociation Reagent (Gibco, Cat. #A11105-01) and seeded on Matrigel-coated ibidi chambers (ibidi, Cat. #80826) with the STEMdiff™ Trilineage Differentiation Kit (according to the manufacturer's

instructions, STEMCELL Technologies, Cat. #05230). Cells were fixed on the indicated day and germ lineage specific markers were assessed using immunofluorescence. Markers used were the following: PAX6 for ectoderm, Smooth Muscle Actin (SMA) for mesoderm and SOX-17 for endoderm.

Directed differentiation of human iPSC cells to neural progenitor cells.

iPSC colonies were treated with StemPro™ Accutase™ Cell Dissociation Reagent (Gibco, Cat. #A11105-01) to obtain single cells and seeded at a density of 2×10^6 cells/well in a 6 well plate in the presence of 10 μ M Y27632. Neural induction was initiated 24h later when cells formed a confluent monolayer by changing the culture medium to a 1:1 mixture of N2- and B27-containing media (referred to N2/B27). N2 medium consisted of DMEM/F-12 Glutamax (Gibco, Cat. #31331-028), N2 supplement (Gibco, Cat. #17502048), 100 μ M non-essential amino acids (Gibco, Cat. #11140-050) and 100 μ M 2-mercaptoethanol (Gibco, Cat. #21985-023). B27 medium consisted of Neurobasal medium (Gibco, Cat. #21103-049), B27 supplement with vitamin A (Gibco, Cat. #17504-044) and 200 mM glutamine (Gibco, Cat. #35050-038). N2/B27 medium was supplemented with 10 μ M SB431542 (Miltenyi Biotec, Cat. #130-105-336) to inhibit TGF β signaling, and 100 nM LDN-193189 (Miltenyi Biotec, Cat. #130-103-925) to inhibit BMP signaling for the first 10 days of differentiation during which the medium was exchanged daily⁷⁵. Neuroepithelial cells were collected on day 10 by dissociation with Accutase and replated in N2/B27 medium in the presence of 10 μ M Y27632 on Poly-L-lysine hydrobromide (Sigma-Aldrich™, Cat. #P9155) and laminin-coated (Sigma, Cat. #11243217001) tissue culture treated plates (Corning). Cells were cultured with daily medium changes in N2/B27 medium till day 19. At this stage cells were fixed with 4% paraformaldehyde and stained for PAX6 and SOX2 to show their identity as neural progenitor

cells or harvested with Accutase for RNA preparation or frozen in CryoStor® CS10 (Stemcell Technologies, Cat. #100-1061). To further differentiate NPCs to cortical neurons, day 19 cells were seeded in N2/B27 medium at a density of 50000 cells/cm² on Poly-L-lysine hydrobromide and laminin-coated plates in the presence of 10 ng/ml BDNF (Miltenyi Biotec, Cat. #130-096-286) and 10ng/ml GDNF (Miltenyi Biotec, Cat. #130-098-449). At day 24 of the differentiation protocol cells were again passaged on Poly-L-lysine hydrobromide and laminin-coated plates in N2/B27 medium supplemented with 10 ng/ml BDNF and 10 ng/ml GDNF at a density of 25000 cells/cm². For passages at days 19 and 25 10 μM Y27632 was added. Cells were cultured in N2/B27 medium supplemented with BDNF and GDNF up till day 90 during which the medium was replaced every other day. At days 35, 55, 70 and 90 cells were harvested with Accutase for RNA preparation.

Immunofluorescence

Cells were washed using PBS and fixed with 4% (v/v) paraformaldehyde for 20 min at room temperature. Fixed cells were washed three times with 1x PBS, permeabilized by 0.15% (v/v) Triton X-100 in PBS for 15 min followed by washing with PBS twice. Cells were blocked with 3% (w/v) BSA (Sigma-Aldrich, Cat. #A9647) in PBS for 1h and then incubated with primary antibodies diluted in the same blocking buffer at 4°C overnight. Subsequently, cells were thoroughly washed with PBS three times, and incubated with appropriate secondary antibodies conjugated with fluorescence dyes diluted in blocking buffer for 2h at room temperature, followed by three washes with PBS. Nuclei were stained during the first PBS wash after secondary antibody incubation using Hoechst 33342 (1:1000, Thermofisher Scientific, Cat. #62249) for 15 mins. Cells were imaged using either the spinning disk Nikon Ti CSU-W1 confocal or the Zeiss point scanning LSM880 inverted microscope and images were

acquired using either a 20x, 25x or 40x objective. Primary antibodies used were: PAX6 (1:500, BioLegend, Cat. #901301); SOX1 (10ug/ml, R&D SYSTEMS, Cat. #AF3369), Nestin (1:1000, ThermoFisher Scientific, Cat. #MA1-110) and MAP2 (1:500, Sigma-Aldrich™, Cat. #M4403). Secondary antibodies used for primary antibody detection were species-specific, Alexa-dye conjugates (Jackson ImmunoResearch).

RNA isolation

Total RNA was isolated using Qiagen RNeasy mini kit (Qiagen, Cat. #74104) with on column DNA digestion (DNase, Qiagen, Cat. #79254) according to manufacturer's specifications. RNA concentration was measured using a NanoDrop™ 2000 (Witec AG).

RNA sequencing and bioinformatic pipeline

RNA sequencing was performed at Novogene GmbH (www.novogene.com) on a NovaSeq X Plus Series (PE150) platform including RNA sample QC and mRNA library preparation (poly A enrichment). The library was checked with Qubit and real-time PCR for quantification and bioanalyzer for size distribution detection. Quantified libraries were pooled and sequenced according to effective library concentration and data amount.

Raw data (raw reads) of fastq format were firstly processed through fastp software. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing poly-N and low-quality reads from raw data. At the same time, Q20, Q30 and GC content the clean data were calculated. All downstream analyses were based on the clean data with high quality. Reference genome and gene model annotation files were downloaded from genome website directly. Index of the reference genome was built using Hisat2 v2.0.5 and paired-end clean reads were aligned to the reference genome using Hisat2

v2.0.5⁷⁶. featureCounts⁷⁷ v1.5.0-p3 was used to count the reads numbers mapped to each gene.

Differential expression analysis (three replicates for each clone across each time point) was performed using the DESeq2R package⁷⁸ (1.44.0). Genes with less than three samples having a count of at least 10 were filtered out, yielding a total of 20894 genes for analysis. Samples visualization (PCA and Heatmap) was obtained on variance stabilizing transformed counts. Differential expression analysis between time points was performed considering contrasts from an interaction model day x clone; log2 fold-changes were shrunk using the ashR method⁷⁹. Temporal profiles of markers of neuronal differentiation were obtained on DESeq2 normalized counts standardized across samples: genes showing significant differential expression at a given time point as compared to day 19 within WT or HTZ clones, were highlighted ($|\text{Log}_2 \text{FC}| > 1$ and Benjamini-Hochberg adjusted $P < 0.05$ across all 20894 genes for a given contrast).

The transcriptional profiles of the WT and edited heterozygous clones were compared to the global transcriptional changes of iPSC cell lines undergoing neuronal differentiation and maturation described in ⁸⁰. From the gene expression data in that study, we derived the first two principal components based on 128 RNA-seq samples obtained from an iPSC-based model of corticogenesis (self-renewal: days 2, 4, 6; accelerated dorsal: days 2, 4, 6, 9; NPC: day 15; Rosette: day 21; Neurons day 77; Neuron+rat astrocyte: days 49, 63, 77). The RNA-seq samples of our study were projected onto these two first principal components.

Acknowledgements

We are deeply thankful to JN and his family, without whose contribution and dedication to research made this work possible.

Funding

The study was funded through intramural funds of the University of Basel and by the Novartis Research Foundation (Novartis Forschungsstiftung) FreeNovation Award FN19-0000000028.

Author contributions

Conceptualization: A.P. and D.J.-F.d.Q. Methodology, formal analysis and investigation: A.P., J.P., A.A., A.K.R.L.P., P.M., M.N., V.F., V.G., N.S., V.V., D.C., A.S., N.B., C.R., J.D., T.S., O.B., K.H., V.T., C.E.L.S., J.L.Mc.G. and D.J.-F.d.Q. Resources: A.P., D.N., O.B., V.T., C.E.L.S., J.L.Mc.G. and D.J.-F.d.Q. Data curation: J.P., A.A., P.M., M.N., V.F., V.G., N.S., V.V. and D.C. Writing—original draft: A.P., J.P., A.A., P.M., M.N., N.S., D.C. and D.J.-F.d.Q. Writing—review, revision, and editing: A.P., J.P., A.A., A.K.R.L.P., P.M., M.N., V.F., V.G., N.S., V.V., D.C., A.S., N.B., N.G., C.R., J.D., T.S., O.B., J.G., E.M.C.S., K.H., S.C., V.T., C.E.L.S., J.L.Mc.G., C.D. and D.J.-F.d.Q. Supervision: A.P. and D.J.-F.d.Q.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Conccorrespondence and material requests should be addressed to A.P. or D.J.-F.d.Q.

Data availability

All data produced in the present study are available upon reasonable request to the authors.

References

- 1 Parker, E. S., Cahill, L. & McGaugh, J. L. A case of unusual autobiographical remembering. *Neurocase* **12**, 35-49, (2006).
- 2 LePort, A. K. *et al.* Behavioral and neuroanatomical investigation of Highly Superior Autobiographical Memory (HSAM). *Neurobiol Learn Mem* **98**, 78-92, (2012).
- 3 Talbot, J., Convertino, G., De Marco, M., Venneri, A. & Mazzoni, G. Highly Superior Autobiographical Memory (HSAM): A Systematic Review. *Neuropsychol Rev*, (2024).
- 4 Guo, Q., Xie, J., Dang, C. V., Liu, E. T. & Bishop, J. M. Identification of a large Myc-binding protein that contains RCC1-like repeats. *Proc Natl Acad Sci U S A* **95**, 9172-9177, (1998).
- 5 Pao, K. C. *et al.* Activity-based E3 ligase profiling uncovers an E3 ligase with esterification activity. *Nature* **556**, 381-385, (2018).
- 6 McGaugh, J. L. Highly Superior Autobiographical Memory, in *Learning and Memory: A Comprehensive Reference* (ed J. H. Byrne) 411-443 (Academic Press, 2017).
- 7 Henkle, W. D. Remarkable cases of memory. *The Journal of Speculative Philosophy* **5**, 6-26, (1871).
- 8 LePort, A. K., Stark, S. M., McGaugh, J. L. & Stark, C. E. Highly Superior Autobiographical Memory: Quality and Quantity of Retention Over Time. *Front Psychol* **6**, 2017, (2015).
- 9 LePort, A. K., Stark, S. M., McGaugh, J. L. & Stark, C. E. A cognitive assessment of highly superior autobiographical memory. *Memory* **25**, 276-288, (2017).
- 10 Palombo, D. J., Sheldon, S. & Levine, B. Individual Differences in Autobiographical Memory. *Trends Cogn Sci* **22**, 583-597, (2018).
- 11 Aggarwal, C. C. *Outlier Analysis*. 2nd edn, (Springer International Publishing, 2017).
- 12 Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341-355, (2002).
- 13 Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179-194, (1999).
- 14 Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195-207, (1999).
- 15 Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216, (2018).
- 16 Barry, D. N., Clark, I. A. & Maguire, E. A. The relationship between hippocampal subfield volumes and autobiographical memory persistence. *Hippocampus* **31**, 362-374, (2021).
- 17 Clark, I. A., Dalton, M. A. & Maguire, E. A. Posterior hippocampal CA2/3 volume is associated with autobiographical memory recall ability in lower performing individuals. *Sci Rep* **13**, 7924, (2023).
- 18 Clark, I. A. & Maguire, E. A. Release of cognitive and multimodal MRI data including real-world tasks and hippocampal subfield segmentations. *Sci Data* **10**, 540, (2023).
- 19 Palombo, D. J. *et al.* Episodic autobiographical memory is associated with variation in the size of hippocampal subregions. *Hippocampus* **28**, 69-75, (2018).
- 20 Bonnici, H. M., Chadwick, M. J. & Maguire, E. A. Representations of recent and remote autobiographical memories in hippocampal subfields. *Hippocampus* **23**, 849-854, (2013).

- 21 Grill, B., Murphey, R. K. & Borgen, M. A. The PHR proteins: intracellular signaling hubs in neuronal development and axon degeneration. *Neural Dev* **11**, 8, (2016).
- 22 Zhen, M., Huang, X., Bamber, B. & Jin, Y. Regulation of presynaptic terminal organization by *C. elegans* RPM-1, a putative guanine nucleotide exchanger with a RING-H2 finger domain. *Neuron* **26**, 331-343, (2000).
- 23 Wan, H. I. *et al.* Highwire regulates synaptic growth in *Drosophila*. *Neuron* **26**, 313-329, (2000).
- 24 AlAbdi, L. *et al.* Loss-of-function variants in MYCBP2 cause neurobehavioural phenotypes and corpus callosum defects. *Brain* **146**, 1373-1387, (2023).
- 25 Fuxreiter, M. *et al.* Disordered proteinaceous machines. *Chem Rev* **114**, 6806-6843, (2014).
- 26 Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing protein intrinsic disorder. *Chem Rev* **114**, 6561-6588, (2014).
- 27 Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature reviews. Molecular cell biology* **6**, 197-208, (2005).
- 28 Crawley, O. *et al.* Autophagy is inhibited by ubiquitin ligase activity in the nervous system. *Nature communications* **10**, 5017, (2019).
- 29 Vukojevic, V. *et al.* A role for alpha-adducin (ADD-1) in nematode and human memory. *EMBO J* **31**, 1453-1466, (2012).
- 30 Morrison, G. E. & van der Kooy, D. A mutation in the AMPA-type glutamate receptor, *glr-1*, blocks olfactory associative and nonassociative learning in *Caenorhabditis elegans*. *Behav Neurosci* **115**, 640-649, (2001).
- 31 Rose, J. K., Kaun, K. R., Chen, S. H. & Rankin, C. H. GLR-1, a non-NMDA glutamate receptor homolog, is critical for long-term memory in *Caenorhabditis elegans*. *J Neurosci* **23**, 9595-9599, (2003).
- 32 Gharat, V., Peter, F., de Quervain, D. J., Papassotiropoulos, A. & Stetak, A. Role of GLR-1 in Age-Dependent Short-Term Memory Decline. *eNeuro* **11**, (2024).
- 33 Hoerndli, F. J. *et al.* Neuronal Activity and CaMKII Regulate Kinesin-Mediated Transport of Synaptic AMPARs. *Neuron* **86**, 457-474, (2015).
- 34 Hadziselimovic, N. *et al.* Forgetting is regulated via Musashi-mediated translational control of the Arp2/3 complex. *Cell* **156**, 1153-1166, (2014).
- 35 Kennedy, M. J., Davison, I. G., Robinson, C. G. & Ehlers, M. D. Syntaxin-4 defines a domain for activity-dependent exocytosis in dendritic spines. *Cell* **141**, 524-535, (2010).
- 36 Schaefer, A. M., Hadwiger, G. D. & Nonet, M. L. *rpm-1*, a conserved neuronal gene that regulates targeting and synaptogenesis in *C. elegans*. *Neuron* **26**, 345-356, (2000).
- 37 Opperman, K. J. & Grill, B. RPM-1 is localized to distinct subcellular compartments and regulates axon length in GABAergic motor neurons. *Neural Dev* **9**, 10, (2014).
- 38 Ch'ng, Q. *et al.* Identification of genes that regulate a left-right asymmetric neuronal migration in *Caenorhabditis elegans*. *Genetics* **164**, 1355-1367, (2003).
- 39 Wisse, L. E. M. *et al.* Hippocampal subfield volumetry from structural isotropic 1 mm(3) MRI scans: A note of caution. *Hum Brain Mapp* **42**, 539-550, (2021).
- 40 Santangelo, V. *et al.* Enhanced brain activity associated with memory access in highly superior autobiographical memory. *Proc Natl Acad Sci U S A* **115**, 7795-7800, (2018).
- 41 Virdee, S. An atypical ubiquitin ligase at the heart of neural development and programmed axon degeneration. *Neural Regen Res* **17**, 2347-2350, (2022).

- 42 Giles, A. C., Opperman, K. J., Rankin, C. H. & Grill, B. Developmental Function of the PHR Protein RPM-1 Is Required for Learning in *Caenorhabditis elegans*. *G3 (Bethesda)* **5**, 2745-2757, (2015).
- 43 Bloom, A. J., Miller, B. R., Sanes, J. R. & DiAntonio, A. The requirement for Phr1 in CNS axon tract formation reveals the corticostriatal boundary as a choice point for cortical axons. *Genes Dev* **21**, 2593-2606, (2007).
- 44 Burgess, R. W. *et al.* Evidence for a conserved function in synapse formation reveals Phr1 as a candidate gene for respiratory failure in newborn mice. *Mol Cell Biol* **24**, 1096-1105, (2004).
- 45 Huang, C. *et al.* A permissive role of mushroom body alpha/beta core neurons in long-term memory consolidation in *Drosophila*. *Curr Biol* **22**, 1981-1989, (2012).
- 46 Babetto, E., Beirowski, B., Russler, E. V., Milbrandt, J. & DiAntonio, A. The Phr1 ubiquitin ligase promotes injury-induced axon self-destruction. *Cell Rep* **3**, 1422-1429, (2013).
- 47 Diering, G. H. & Hugarir, R. L. The AMPA Receptor Code of Synaptic Plasticity. *Neuron* **100**, 314-329, (2018).
- 48 Hardt, O., Nader, K. & Wang, Y. T. GluA2-dependent AMPA receptor endocytosis and the decay of early and late long-term potentiation: possible mechanisms for forgetting of short- and long-term memories. *Philos Trans R Soc Lond B Biol Sci* **369**, 20130141, (2014).
- 49 Moreno, A. Molecular mechanisms of forgetting. *Eur J Neurosci* **54**, 6912-6932, (2021).
- 50 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779, (2015).
- 51 Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* **19**, 1523-1536, (2016).
- 52 Duff, E. *et al.* Reliability of multi-site UK Biobank MRI brain phenotypes for the assessment of neuropsychiatric complications of SARS-CoV-2 infection: The COVID-CNS travelling heads study. *PLoS One* **17**, e0273704, (2022).
- 53 Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400-424, (2018).
- 54 Heck, A. *et al.* Exome sequencing of healthy phenotypic extremes links TROVE2 to emotional memory and PTSD. *Nat Hum Behav* **1**, (2017).
- 55 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, (2009).
- 56 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, (2010).
- 57 Koboldt, D. C., Larson, D. E. & Wilson, R. K. Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. *Curr Protoc Bioinformatics* **44**, 15.14.11-17, (2013).
- 58 Chen, J., Li, X., Zhong, H., Meng, Y. & Du, H. Systematic comparison of germline variant calling pipelines across multiple next-generation sequencers. *Sci Rep* **9**, 9345, (2019).
- 59 Liang, Y. *et al.* Comparative Analysis for the Performance of Variant Calling Pipelines on Detecting the de novo Mutations in Humans. *Front Pharmacol* **10**, 358, (2019).

- 60 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation
in 141,456 humans. *Nature* **581**, 434-443, (2020).
- 61 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, (2016).
- 62 Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71-94, (1974).
- 63 Arribere, J. A. *et al.* Efficient marker-free recovery of custom genetic modifications
with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* **198**, 837-846, (2014).
- 64 Dickinson, D. J., Ward, J. D., Reiner, D. J. & Goldstein, B. Engineering the
Caenorhabditis elegans genome using Cas9-triggered homologous recombination.
Nat Methods **10**, 1028-1034, (2013).
- 65 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat*
Methods **9**, 676-682, (2012).
- 66 Bargmann, C. I., Hartwig, E. & Horvitz, H. R. Odorant-selective genes and neurons
mediate olfaction in *C. elegans*. *Cell* **74**, 515-527, (1993).
- 67 Nuttley, W. M., Atkinson-Leadbeater, K. P. & Van Der Kooy, D. Serotonin mediates
food-odor associative learning in the nematode *Caenorhabditiselegans*. *Proc Natl*
Acad Sci U S A **99**, 12449-12454, (2002).
- 68 Grill, B. *et al.* *C. elegans* RPM-1 regulates axon termination and synaptogenesis
through the Rab GEF GLO-4 and the Rab GTPase GLO-1. *Neuron* **55**, 587-601, (2007).
- 69 Hoerndli, F. J. *et al.* Kinesin-1 regulates synaptic strength by mediating the delivery,
removal, and redistribution of AMPA receptors. *Neuron* **80**, 1421-1437, (2013).
- 70 Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein
homology-modeling server. *Nucleic Acids Res* **31**, 3381-3385, (2003).
- 71 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and
complexes. *Nucleic Acids Res* **46**, W296-W303, (2018).
- 72 Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of
individual protein structure models. *Bioinformatics* **27**, 343-350, (2011).
- 73 Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res*
33, W382-388, (2005).
- 74 Devine, M. J. *et al.* Parkinson's disease induced pluripotent stem cells with
triplication of the alpha-synuclein locus. *Nature communications* **2**, 440, (2011).
- 75 Shi, Y., Kirwan, P., Smith, J., Robinson, H. P. & Livesey, F. J. Human cerebral cortex
development from pluripotent stem cells to functional excitatory synapses. *Nat*
Neurosci **15**, 477-486, S471, (2012).
- 76 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and
quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628,
(2008).
- 77 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program
for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930,
(2014).
- 78 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, (2014).
- 79 Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275-294, (2017).
- 80 Burke, E. E. *et al.* Dissecting transcriptomic signatures of neuronal differentiation and
maturation using iPSCs. *Nature communications* **11**, 462, (2020).

Figure 1

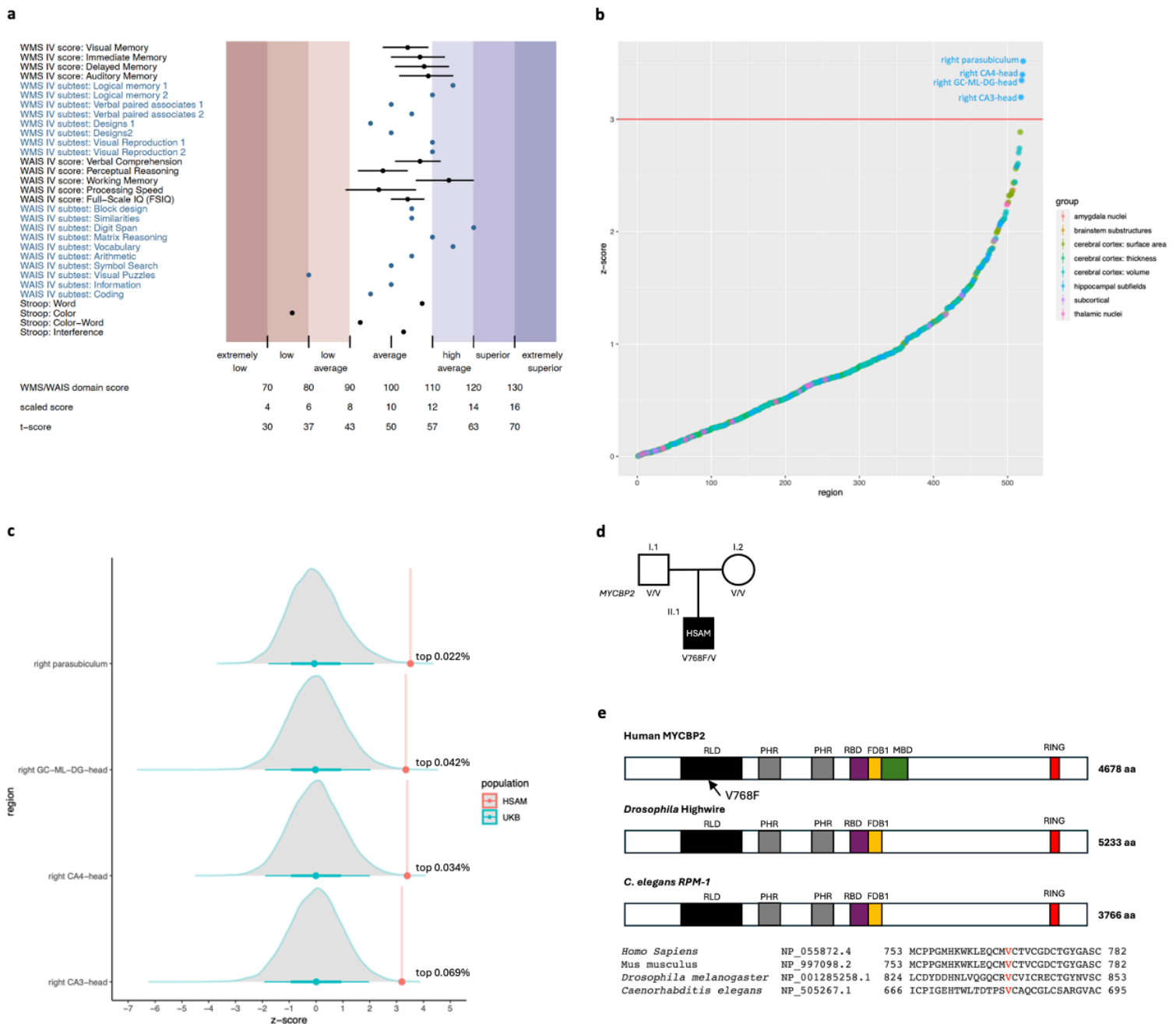


Fig. 1 | MYCBP2 variant in the individual with HSAM. a, JN's neurocognitive profile. **b**, Comparison of JN's image-derived phenotypes of cortical and subcortical brain regions to a normative dataset from the UKBiobank. Dots represent positive Z values in ascending order. Dots above the red horizontal line exceed $Z=3$, indicating extreme large values of the corresponding regions in JN as related to the reference population. GC-ML-DG: granule cell and molecular layers of the dentate gyrus. **c**, Distribution of the volumes of the four brain regions exceeding $Z=3$ in the reference population. Red dots and vertical lines denote JN's volumes in these regions. **d**, *MYCBP2* *de novo* variant in the family trio. **e**, Conservation of protein domains (RCC1-like GEF domain (RLD), PHR family specific domains (PHR), RAE-1 binding domain (RBD), FSN-1 binding domain 1 (FDB1), Myc binding domain (MBD) and RING-H2 ubiquitin ligase domain (RING)) and of the mutated residue.

Figure 2

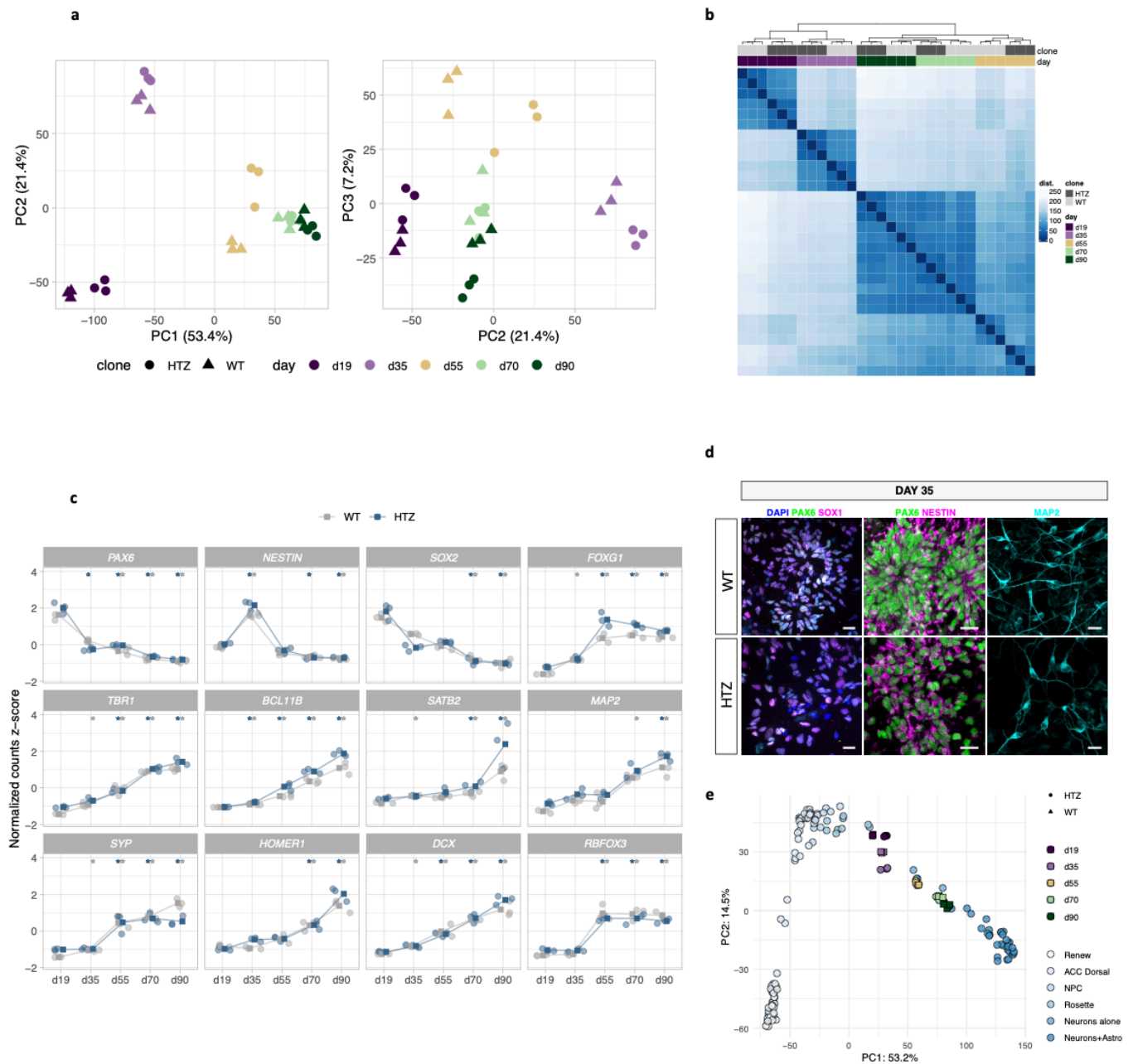


Fig. 2 | Gene expression in hiPSCs undergoing neuronal differentiation. a,b, Principal components (a) and sample clustering heatmap (b) of RNAseq data in WT and heterozygote samples across differentiation days. **c,** Temporal profiles of canonical markers for neuronal differentiation. Dots represent individual samples, per-day group means (WT, HTZ) are marked as squares. Asterisks denote differential expression at a given day vs. day 19, within the HTZ or WT group ($|\text{Log}_2 \text{FC}| > 1$; BH P value < 0.05). **d,** Representative images of day 35 neurons displaying a mix of neural progenitor cells and polarized neurons. Immunostaining of PAX6, SOX1, NESTIN and MAP2 of the WT and HTZ clone at day 35 of neural differentiation. Scale bar, 20 μm . **e,** Expression data of WT and HTZ clones across differentiation days projected onto the first two principal components of global transcriptional changes of iPSC cell lines undergoing neuronal differentiation and maturation as described in ⁸⁰. WT: Wild-type clone; HTZ: Heterozygote clone. DAPI: 4',6-diamidino-2-phenylindole.

Figure 3

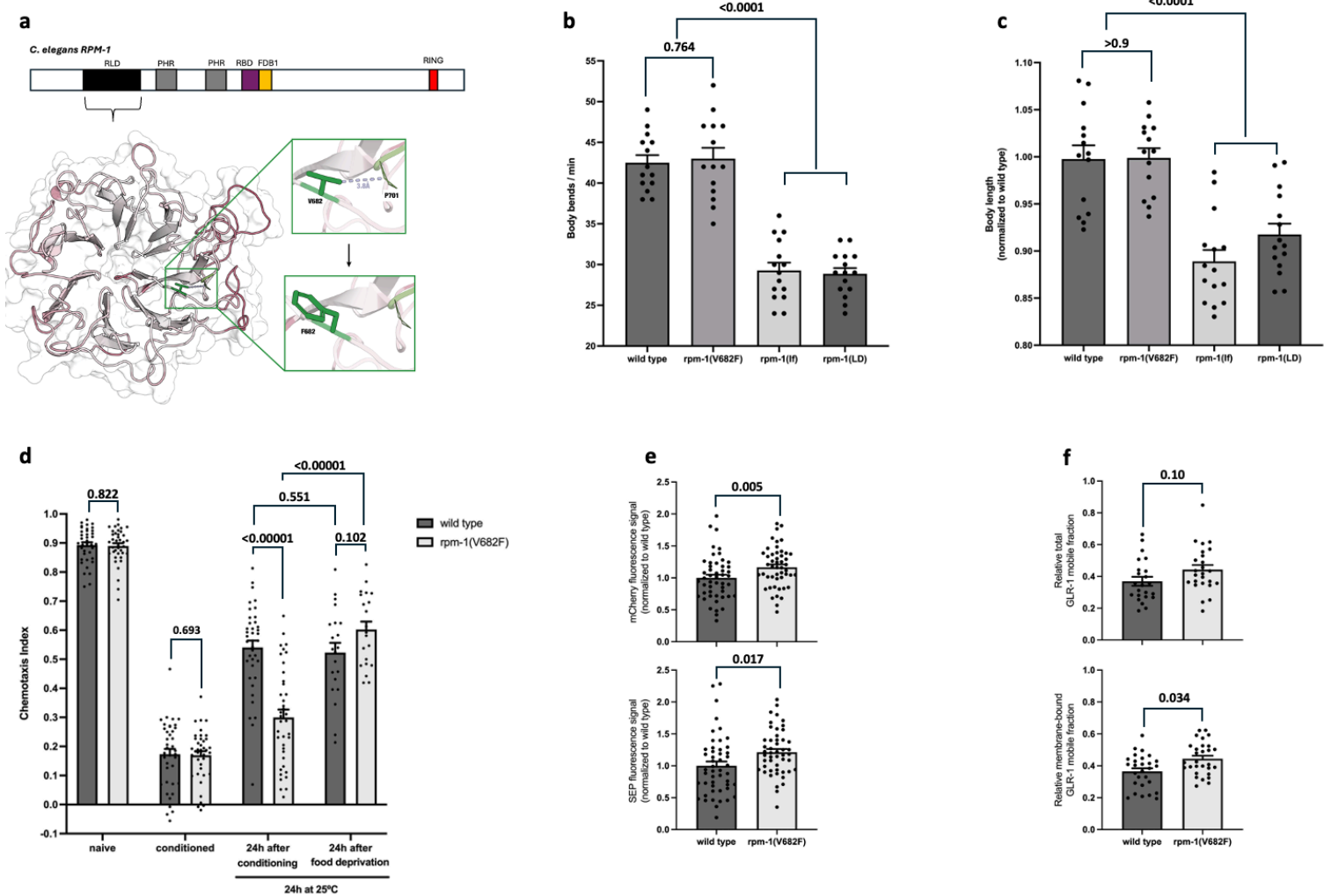


Fig. 3 | Structural modeling of the RLD, impact of the mutated residue on developmental, behavioral, and molecular measurements in *C. elegans*, and transcriptomic data in CRISPR-edited hiPSCs undergoing early corticogenesis. a, Schematic representation and domain structure of *C. elegans* RPM-1. Template PDB ID 8gqe was used for RPM-1 RLD domain structural modeling. Structure quality is depicted on a red to white scale using QMEAN scoring. Residue V682 and the mutated version F682 are colored green and depicted as sticks. The predicted hydrophobic interaction of V682 with P701 (green lines) is shown as a light blue dashed line. **b,c**, Locomotor behavior, i.e. body bends/min (**b**), and body length (**c**) of the adult N2 WT strain and three adult *rpm-1* mutant strains (*rpm-1(V682F)*), loss-of-function *rpm-1(lf)* (i.e., *rpm-1(utr93[3xstop+frameshift])*, see Methods), and an RPM-1 ligase-dead (*rpm-1(LD)*) site-directed mutant strain carrying variants within the RING domain (C3535A, H3537A, and H3540A) that impair catalytic activity²⁸. Dots represent single worms. **d**, Chemotaxis indices towards DA of WT and *rpm-1(V682F)* worms under four conditions: naïve (unconditioned), conditioned (immediately after 1h of food deprivation in the presence of DA), 24h after conditioning, 24h after 1h of food deprivation in the absence of DA. For the last two conditions, worms were kept at 25°C during 24h. Dots represent single plates. **e**, Fluorescence intensity of total (mCherry) and membrane-bound (SEP) GPCR receptors normalized to WT signal in the proximal part of the axon of the AVA neuron of WT and *rpm-1(V682F)* worms. Dots represent measurements in single worms. **f**, Quantity of the

mobile fraction of total (mCherry) and membrane-bound (SEP) GLR-1 in the proximal part of the axon of the AVA neuron of WT and *rpm-1(V682F)* worms. Dots represent measurements in single worms. For all bar plots in **b-f**, bars and error bars represent means and s.e.m., respectively. All *P* values in **b-f** are two-sided (non-parametric Wilcoxon-Mann-Whitney test).