

UC San Diego

UC San Diego Previously Published Works

Title

Comprehensive multi-omic profiling of somatic mutations in malformations of cortical development.

Permalink

<https://escholarship.org/uc/item/6qk17474>

Journal

Nature genetics, 55(2)

ISSN

1061-4036

Authors

Chung, Changuk
Yang, Xiaoxu
Bae, Taejeong
et al.

Publication Date

2023-02-01

DOI

10.1038/s41588-022-01276-9

Supplemental Material

<https://escholarship.org/uc/item/6qk17474#supplemental>

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, availalbe at
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

1 **Comprehensive multi-omic profiling of somatic mutations in**
2 **malformations of cortical development**

4 Changuk Chung ^{1,2,#}, Xiaoxu Yang ^{1,2,#}, Taejeong Bae ³, Keng Ioi Vong ^{1,2}, Swapnil Mittal ^{1,2},
5 Catharina Donkels ⁴, H. Westley Phillips ⁵, Zhen Li ^{1,2}, Ashley P. L. Marsh ^{1,2}, Martin W. Breuss
6 ^{1,2,6}, Laurel L. Ball ^{1,2}, Camila Araújo Bernardino Garcia ⁷, Renee D. George ^{1,2}, Jing Gu ^{1,2},
7 Mingchu Xu ^{1,2}, Chelsea Barrows ^{1,2}, Kiely N. James ^{1,2}, Valentina Stanley ^{1,2}, Anna S. Nidhiry
8 ^{1,2}, Sami Khoury ^{1,2}, Gabrielle Howe ^{1,2}, Emily Riley ^{1,2}, Xin Xu ^{1,2}, Brett Copeland ^{1,2}, Yifan
9 Wang ³, Se Hoon Kim ⁸, Hoon-Chul Kang ⁹, Andreas Schulze-Bonhage ¹⁰, Carola A. Haas ^{4,10},
10 Horst Urbach ¹¹, Marco Prinz ^{10,12,13}, David D. Limbrick, Jr. ¹⁴, Christina A. Gurnett ¹⁴, Matthew
11 D. Smyth ¹⁵, Shifteh Sattar ¹⁶, Mark Nespeca ¹⁶, David D. Gonda ¹⁶, Katsumi Imai ¹⁷, Yukitoshi
12 Takahashi ¹⁷, Hsin-Hung Chen ¹⁸, Jin-Wu Tsai ¹⁹, Valerio Conti ²⁰, Renzo Guerrini ²⁰, Orrin
13 Devinsky ²¹, Wilson A. Silva Jr ²², Helio R. Machado ⁷, Gary W. Mathern ⁵, Alexej Abyzov ³,
14 Sara Baldassari ²³, Stéphanie Baulac ²³, Focal Cortical Dysplasia Neurogenetics Consortium [&],
15 Brain Somatic Mosaicism Network ^{*} and Joseph G. Gleeson^{1,2,*}

16
17 ¹Department of Neurosciences, University of California San Diego, La Jolla, CA, 92037, USA

18 ²Rady Children's Institute for Genomic Medicine, San Diego, CA, 92123, USA

19 ³Department of Quantitative Health Sciences, Center for Individualized Medicine, Mayo Clinic,
20 Rochester, MN, 55905, USA

21 ⁴Department of Neurosurgery, Experimental Epilepsy Research, Medical Center-University of Freiburg,
22 Faculty of Medicine, University of Freiburg, Freiburg, 79106, Germany

23 ⁵Department of Neurosurgery, University of California at Los Angeles, CA, 90095, USA

24 ⁶Department of Pediatrics, Section of Clinical Genetics and Metabolism, University of Colorado Aurora,
25 CO, 80045, USA

26 ⁷Laboratory of Pediatric Neurosurgery and Developmental Neuropathology, Department of Surgery and
27 Anatomy, University of São Paulo (USP), Ribeirão Preto, 14000-000, Brazil

28 ⁸Department of Pathology, Severance Hospital, Yonsei University College of Medicine, Seoul, 03722,
29 South Korea

30 ⁹Division of Pediatric Neurology, Department of Pediatrics, Severance Children's Hospital, Yonsei
31 University College of Medicine, Seoul, 03722, South Korea

32 ¹⁰Center for Basics in NeuroModulation, Faculty of Medicine, University of Freiburg, Freiburg, 79106,
33 Germany

34 ¹¹Department of Neuroradiology, Medical Center-University of Freiburg, Faculty of Medicine, Freiburg,
35 79106, Germany

36 ¹²Institute of Neuropathology, Medical Center-University of Freiburg, Faculty of Medicine, Freiburg,
37 79106, Germany

38 ¹³Signalling Research Centres BIOSS and CIBSS, University of Freiburg, Freiburg, 79106, Germany.

39 ¹⁴Department of Neurology, St. Louis Children's Hospital, Washington University St Louis, MO, 63110,
40 USA

41 ¹⁵Department of Neurosurgery, St. Louis Children's Hospital, Washington University St Louis, MO,
42 63110, USA

43 ¹⁶Epilepsy Center, Rady Children's Hospital, San Diego, CA, 92123, USA

44 ¹⁷National Epilepsy Center, Shizuoka Institute of Epilepsy and Neurological Disorders, Shizuoka 420-
45 8688, Japan

46 ¹⁸Division of Pediatric Neurosurgery, The Neurological Institute, Taipei Veterans General Hospital,
47 Taipei City, 11217, Taiwan

48 ¹⁹Institute of Brain Science, Brain Research Center, National Yang Ming Chiao Tung University, Taipei,
49 11221, Taiwan

50 ²⁰Pediatric Neurology Unit and Laboratories, A. Meyer Children's Hospital, University of Florence,
51 Firenze, 50139, Italy

52 ²¹Comprehensive Epilepsy Center, Department of Neurology, New York University Langone Health, NY
53 10016, USA

54 ²²Department of Genetics, Center for Cell-Based Therapy, Center for Integrative Systems Biology,
55 University of São Paulo (USP), Ribeirão Preto, 14000, Brazil

56 ²³Sorbonne Université Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, Hôpital de la
57 Pitié Salpêtrière, Paris, 75006, France

58 [#]These authors contributed equally

59 *Correspondence to: jogleeson@health.ucsd.edu

60 [&]Full membership of the FCD Neurogenetic Consortium is listed at the end of this file

61 ^{*}Full membership of the Brain Somatic Mosaicism Network is listed at the end of this file

62

63

64 **Abstract: 138 words, Manuscript: 3965 words, Figures: 6. Extended Data items: 10. Suppl
65 Tables: 9.**

66 **Keywords: epilepsy, focal cortical dysplasia, brain mosaicism, mTOR, single-cell
67 sequencing, whole-exome sequencing, somatic, migration, excitation**

68

69 **Abstract**

70 Malformations of cortical development (MCD) are neurological conditions displaying focal
71 disruption of cortical architecture and cellular organization arising during embryogenesis, largely
72 from somatic mosaic mutations, and causing intractable epilepsy. Identifying the genetic causes
73 of MCD has been a challenge, as mutations remain at low allelic fractions in brain tissue resected
74 to treat condition-related epilepsy. Here, we report a genetic landscape from 283 brain resections,
75 identifying 69 mutated genes through intensive profiling of somatic mutations, combining
76 whole-exome and targeted-amplicon sequencing with functional validation including *in utero*
77 electroporation of mice and single-nucleus RNA sequencing. Genotype-phenotype correlation
78 analysis elucidated specific MCD gene sets associating distinct pathophysiological and clinical
79 phenotypes. The unique single-cell level spatiotemporal expression patterns of mutated genes in
80 control and patient brains implicate critical roles in excitatory neurogenic pools during brain
81 development, and in promoting neuronal hyperexcitability after birth.

82

83 **Introduction**
84 MCDs are heterogeneous groups of neurodevelopmental disorders with localized malformation
85 of cortical structures, often presenting with intractable epilepsy¹. Major MCD subtypes include
86 different classes of focal cortical dysplasia (FCD), hemimegalencephaly (HME), and tuberous
87 sclerosis complex (TSC)². The International League Against Epilepsy (ILAE) has classified FCD
88 subtypes based on neuropathological features and cell types³. Fifty percent of patients with
89 epileptic surgery due to refractory epilepsy show cortical dysplasia, and 50~75% of MCD
90 patients become seizure-free after surgical resection, which has led to remarkable clinical
91 benefits⁴⁻⁶. The abnormal histology of resected regions includes loss of cortical lamination,
92 enlarged dysplastic neurons, or balloon cells, sometimes accompanied by other brain
93 abnormalities. Similar to brain tumors, it can be difficult to predict pathology before surgery.

94 Again, like with brain tumors, genetic studies may offer insights into mechanisms.
95 Somatic mTOR pathway gene mutations are frequently detected in HME and type II FCD foci^{7,8}.
96 Recently, small- or medium-size cohort studies (<130 cases) confirmed these results and
97 correlated defects in neuronal migration, cell size, and neurophysiology⁹⁻¹¹. Still, 30~70% of
98 MCD cases remain genetically unsolved⁹⁻¹¹, suggesting other genes are yet to be discovered.

99 Detecting mutant alleles in resected bulk tissue from MCD patients is challenging
100 because unlike in brain tumors, the mutant cells in MCD are probably not hyperproliferative, and
101 thus variant allelic fraction (VAF) are low, often <5%, diluted by genomes of surrounding non-
102 mutated cells¹². Fortunately, new computational algorithms can help reduce false-positive and
103 false-negative signals, even when no ‘normal’ paired sample is available for comparison¹³⁻¹⁵.
104 Recent deep-learning technologies and state-of-the-art image-based artificial intelligence
105 software such as DeepMosaic, trained on non-cancer mosaic variants have significantly
106 improved accuracy¹⁴. The NIH-supported Brain Somatic Mosaicism Network (BSMN)
107 established a ‘Common pipeline’, incorporating a ‘best practice’ workflow to reliably and
108 reproducibly identify low VAF somatic variants¹⁶. With these advances, we assessed for mosaic
109 variants that might point to gene networks beyond mTOR in MCD lesions. Our results offer
110 insights into potentially druggable pathways in cases of incomplete resection or drug-resistant
111 forms of MCD.

112

113 **Results**

114 **Sequencing approach to identify causes of MCD**

115 To perform a thorough genetic screening for somatic mutations in resected epileptic tissue, we
116 formed the FCD Neurogenetics Consortium and from 22 separate international centers, enrolled
117 293 cases that met clinical and pathological criteria for FCD or HME. Our cohort included 30
118 HME cases, 80 type I-, 128 type II-, 31 type III-, and 12 unclassified-FCD cases. We included
119 acute resected brains from 10 neurotypicals and 2 TSC cases for comparison (Fig. 1a,
120 Supplementary Table 1). Patients with environmental causes, syndromic presentations such as
121 TSC, inherited mutations, multifocal lesions, or tumors were excluded (Methods).

122 We used a three-phase genetic screening, each followed by filtering for likely causative
123 mutations using published methods^{17,18}, and each followed by orthogonal targeted amplicon
124 sequencing (TASeq) for intra-case validation and VAF quantification, normalized with controls
125 (~5000 X, TASeq) (Fig. 1b, Extended Data Fig. 1). In Phase 1, we performed amplicon
126 sequencing (AmpliSeq, ~1000 X) profiling the entire open reading frame of 87 genes ('MCD
127 panel v1', Supplementary Table 2a) previously detected in FCD/HMEs or known PI3K-AKT3-
128 mTOR interactors (Supplementary Table 2b). In Phase 2, 80 cases (75 unsolved cases plus 5

129 solved cases) from Phase 1 and additionally collected 54 cases, we performed unbiased deep
130 whole-exome sequencing (WES, ~300 X) on paired samples, where available, or unpaired
131 samples (i.e. brain plus blood/saliva vs. brain only). In Phase 3, from an additional subcohort of
132 96 new cases (86 MCD plus 10 neurotypical), we designed the ‘MCD panel v2’ (Supplementary
133 Table 2c) including known and novel genes detected in Phases 1 and 2 (Methods). We also re-
134 sequenced unsolved cases from Phase 2 (30 cases), expecting that the higher read depth afforded
135 by panel sequencing could provide greater sensitivity to detect low VAF mutations potentially
136 missed by WES. Phase 2 used BSMN best practice guidelines for mapping and variant calling¹⁶
137 (Extended Data Fig. 1b,c).

138 From Phases 1 to 3, 1181 candidate somatic single nucleotide variant (sSNV) calls were
139 identified (Supplementary Table 3a). Of these, 627 were excluded based on gnomAD allele
140 frequencies, dinucleotide repeats, homopolymers, and additional BSMN established criteria
141 (Methods)^{19,20}. This yielded 554 candidate sSNVs, each assessed by TASEq, yielding 108
142 validated sSNVs (19.4% validation rate, Fig. 1c, Supplementary Table 3b), comparing similarly
143 to other BSMN effort validation rates in WGS^{16,21}. The validation rate of candidate sSNV calls in
144 each phase was 12.1% (15/124), 20.9% (67/320), and 23.6% (26/110), respectively. The
145 measured square-root transformed VAFs between the AmpliSeq/WES and TASEq were
146 correlated as expected (Spearman $\rho = 0.7725$) (Fig. 1d). Of the 69 genes mutated in 76 patients,
147 60 were not previously implicated in MCD. Eight were recurrently mutated, including 6 known
148 MCD genes (*MTOR*, *PIK3CA*, *SLC35A2*, *TSC2*, *AKT3*, *BRAF*) as well as 2 novel candidates
149 (*ATP2A1*, *PPFIA4*) (Fig. 1e, Extended data Fig. 2a,b). There were also several genes mutated
150 that were recently linked to epilepsy-associated developmental lesions (*FGFR2*, *NIPBL*,
151 *NPRL3*)²², one gene recently identified in FCD (*RHEB*)²³, and 57 genes we found mutated in a
152 single brain sample.

153 We estimate only ~7% of mutations identified are likely attributable to false discovery
154 during variant calling, based upon the background mutation rate in 75 BSMN neurotypical brain
155 samples, and published experience from the BSMN^{16,24}, processed with the same workflow (see
156 Methods). Thus, 93% of our candidate and known MCD mutations would not have been
157 identified in a size-matched neurotypical control cohort. We estimate the false negative rate of
158 phase 2 was 1.67%, assuming VAF rates comparable to what was detected (Methods). We also
159 calculated the probability of identifying the same gene mutated in two separate patients by
160 chance (Methods), taking into consideration the mutation rate, cohort size of each phase, gene
161 length, and panel size. *ATP2A1* and *PPFIA4*, the two novel recurrently mutated candidates, both
162 reached significance from our permutation analysis ($p = 0.000127$ for *ATP2A1* and $p = 0.000258$
163 for *PPFIA4*, Methods).

164 Most patients (80.52%, 62 cases) showed a single somatic mutation, but some showed two
165 somatic mutations (14.29%, 11 cases), and a few showed more than two mutations (5.19%, 4
166 cases). Interestingly, HME-4144 showed 11 different somatic mutations, all of which were
167 validated with TASEq. Although there are several possible explanations for HME-4144, we
168 suspect this reflects clonal expansion from a driver mutation, with the detection of multiple
169 passenger mutations, as reported in brain tumors²⁵.

170 Single-base mutational signatures (SBS) can describe potential mutational mechanisms in
171 human disease²⁶. We found 60.2% of mutations were C>T, likely arising from mutation of the
172 methylated CpG dinucleotide DNA epigenetic mark²⁷ (Extended Data Fig. 3). Enrichment of
173 SBS1 and SBS5, clock-like mutational signatures suggest endogenous mutations arising during
174 corticogenesis DNA replication.

175

176 **Functional dissection of the MCD genes**

177 Interestingly, most validated genes were non-recurrently mutated in our cohort, suggesting
 178 substantial genetic heterogeneity in MCD. This nevertheless provided an opportunity to study
 179 converging functional gene networks. Thus, we performed Markov clustering with a STRING
 180 network generated from the putative MCD genes²⁸, as well as recently reported novel MCD
 181 candidates (*NAV2*, *EEF2*, *CASK*, *NFI*, *KRAS*, *PTPN11*)^{22,29} (Fig. 2a). We identified four
 182 clusters, with cluster 1 (“mTOR pathway”) showing the highest term enrichment to the
 183 mTOR/MAP kinase signaling, supporting prior results for Type II MCDs. Because MCD panel
 184 v1 included many mTOR pathway genes not identified as mutated in patients from WES, we
 185 repeated the analysis by excluding genes not identified *a priori* from WES and recovered the
 186 same clusters (Extended Data Fig. 4). Cluster 1 also highlighted newly identified genes *FGFR2*,
 187 *KLHL22*, *RRAGA*, *PPP2R5D*, *PIK3R3*, *EEF2*, *EIF4G1*, and *MAPK9*. Cluster 2 identified
 188 “Calcium Dynamics” and included genes *ATP2A1*, *RYR2*, *RYR3*, *PSEN2*, *TTN*, and *UTRN*.
 189 Cluster 3 was labeled “Synaptic Functions” and included genes *CASK*, *GRIN2C*, and *PPFIA4*.
 190 Cluster 4 was labeled “Gene Expression” and included intellectual disability genes, mostly
 191 involved in nuclear function, including *NUP214*, *PRR14*, *PCNT*, *NIPBL*, *SRCAP*, *ASH1L*,
 192 *TRIP12*, and *MED13* (Fig. 2b, Methods). We further performed ClueGO analysis and found
 193 enrichment in mTOR signaling, focal adhesion assembly, cardiac muscle cell contraction, and
 194 artery morphogenesis (Extended Data Fig. 5). ClueGO also displayed isolated gene ontology
 195 (GO) term clusters such as ‘calcium ion import’ and ‘protein localization to synapse’. There
 196 were recurrently mutated genes in all four clusters, and while several of these clusters were not
 197 previously reported in MCDs, they were previously implicated in epilepsy, neurodevelopmental
 198 and neurodegenerative disease^{30,31}, suggesting functional overlap with MCDs.

199

200 **Functional validation of selected genes in mouse brain**

201 To investigate the roles of novel MCD genes and modules, we first revisited previous
 202 publications characterizing our novel MCD genes in the context of cortical development. In the
 203 ‘gene expression’ group in Fig. 2, *NIPBL*³², *ZNF335*³³, and *ZNF423*³⁴ were already reported as
 204 key regulators of neuronal migration during murine cortical development, but not in somatic
 205 mosaicism or MCD contexts. To expand functional validation, we selected two potential
 206 canonical mTOR pathway mutations (*RRAGA* p.H226R, *KLHL22* p.R38Q), and two genes not
 207 known as mTOR interactors (*GRIN2C* p.T529M, *RHOA* p.P75S), discovered in FCD-7967,
 208 3560, 5157, and 3876 respectively. *RRAGA* encodes Ras-related GTP binding A (RAGA), a
 209 GTPase sensing amino acid and activating mTOR signaling. The mosaic p.H226R mutation
 210 occurs within the C-terminal ‘roadblock’ domain (CRD), which binds to the RAGB protein and
 211 is conserved throughout vertebrate evolution (Extended Data Fig. 2c) and thus could change
 212 binding affinity³⁵. *KLHL22* encodes a CUL3 adaptor, determining E3 ubiquitin ligase specificity,
 213 and mediating degradation of DEPDC5, required for mTORC1 activation³⁶. The *KLHL22*
 214 p.R38Q variant in FCD-3560 is near the BTB (Broad-Complex, Tramtrack, and Bric- à-brac)
 215 domain that interacts with CUL3 (Extended Data Fig. 2d), suggesting the variant could enhance
 216 mTORC1 activity. *GRIN2C* encodes a subunit of the NMDA receptor regulating synaptic
 217 plasticity, memory, and cognition^{37,38}, the dysfunction of which is implicated in neurocognitive
 218 diseases^{39,40}. *GRIN2C* p.T529M mutation is located in the S1 glutamate ligand-binding domain
 219 (S1 LBD) (Extended Data Fig. 2e). *GRIN2A* p.T531M mutation, an analog mutation of *GRIN2C*

220 p.T529M in our cohort, was previously reported in epilepsy-aphasia spectrum disorders, where it
221 increased NMDA receptors ‘open-state’ probability⁴⁰. This suggests that the p.T529M mutation
222 activates the channel, likely in an mTOR-independent fashion. *RHOA* encodes RHOA protein, a
223 small GTPase, regulating cytoskeletal dynamics, cell migration, and cell cycle. RHOA p.P75S
224 mutation is located in the interdomain region between the second GTP/GDP binding domain and
225 Rho insert domain (Extended Data Fig. 2f). This mutation is implicated in skin cancer multiple
226 times in the Catalogue Of Somatic Mutations In Cancer (COSMIC) database (DB)⁴¹. Thus, all
227 mutations assessed here are likely gain-of-function and exert functional impact on cells in which
228 they are expressed.

229 To test this hypothesis, we introduced episomal expression vectors carrying mutant or
230 wildtype (WT) genes co-expressing enhanced green fluorescent protein (EGFP) into the dorsal
231 subventricular zone via electroporation at mouse embryonic day 14 (E14), then harvested tissue
232 at either E18 to assess migration, or at postnatal day 21 (P21) to assess cell size and phospho-S6
233 as a reporter of mTOR activity⁴² (Fig. 3a). In E18 cortices, we found EGFP-positive cells
234 expressing mutant but not WT forms of *RRAGA* and *KLHL22* showed significant migration
235 defects of varying severity, whereas mutant *GRIN2C* showed no defect (Fig. 3b). These
236 migration defects in *RRAGA* and *KLHL22* mutant cells replicate major findings of MCD
237 disrupted cortical architecture. Notably, *RHOA* WT and mutant cells both showed significantly
238 disrupted neuronal migration and atypical cell clusters near the subventricular zone (SVZ)
239 (Extended Data Fig. 6a), likely due to the high level of expression. Nevertheless, some low GFP-
240 expressing cells, likely containing a small number of plasmid copies, showed evidence of
241 migration, whereas mutant cells, irrespective of the level of GFP, showed disrupted migration.
242 This indicates the *RHOA* mutation can contribute to the malformation of cortical development.

243 We next assessed cellular phenotype at P21 with samples available in both mice and the
244 corresponding patients and found enlarged cell body area in mutant forms of *KLHL22* and
245 *GRIN2C* compared with wildtype. In contrast, elevated levels of pS6 staining, described
246 previously in association with mTOR pathway mutations⁶, was found only in mutant *KLHL22*,
247 but not in mutant *GRIN2C* mice (Fig. 3c). Interestingly, the *RRAGA* mutant cells showed
248 increased pS6 level but not enlarged cell bodies compared to wildtype. Since this case (FCD-
249 7967, type 2B) has additional *MTOR* mutation, the increase in cell body size of dysplastic cells is
250 likely to be induced by *MTOR* but not *RRAGA* mutation (Extended Data Fig. 6b).

251 To assess correlation with human samples, we assessed available archived neuropathological
252 tissue sections for histology and pS6 activity. Similar to our mouse models, we found patient
253 FCD-3560 carrying *KLHL22* p.R38Q showed enlarged neurons that co-stained for excess pS6,
254 whereas FCD-5157 carrying *GRIN2C* p.T529M showed only a slight increase in cell body size
255 and no evidence of excessive pS6 (Fig. 3d). While this analysis does not take into account the
256 genotype of individual cells, it suggests *KLHL22* but not *GRIN2C* mutations impact mTOR
257 signaling.

258

259 Genotype-phenotype correlations in MCD patients

260 To assess the phenotypic contributions of the MCD genes we found, we focused on 76 of our
261 ‘genetically solved’ MCD cases, comparing detailed neuropathology, brain imaging, and clinical
262 course. We performed Pearson correlation followed by hierarchical clustering based upon ILAE
263 neuropathological diagnosis, compared with GO term-based curated genesets (Fig. 4,
264 Supplementary Table 3c) or with sSNVs in COSMIC DB because a subset of MCDs shows cell
265 over-proliferation, similar to cancer, in the lesion during cortical development. (Methods,

266 Supplementary Table 4). We found that FCD Type IIA and Type IIB, and HME were more
267 tightly clustered than FCD Type I or III (Fig. 4a), likely reflecting shared neuropathological
268 features that include large dysplastic neurons. As expected, FCD Type IIA, Type IIB, and HME
269 were positively associated with the mTOR pathway GO term and the presence of oncogenic
270 variants, FCD Type III, however, was associated with the MAPK pathway, consistent with
271 recent publications implicating *BRAF*, *FGFR2*, *NOD2*, and *MAPK9* in their etiology⁴³⁻⁴⁵. FCD
272 Type I showed strong positive correlations for glycosylation, consistent with recent findings of
273 somatic mutations in *SLC35A2* and *CANT1*^{46,47}.

274 We next investigated correlations between clinical phenotypes extracted from detailed
275 medical records including seizure type, neuropsychological examination, and positron emission
276 tomography (PET) metabolism, often used to help localize seizure focus^{48,49}. Seizure frequency,
277 early age of onset, Engel score, and history of infantile spasms drove clinical clustering, likely
278 reflecting shared clinical features in the most challenging patients. Focusing on the correlations,
279 PET hypometabolism around the resected region, correlated positively with COSMIC DB entry,
280 and negatively with MAPK and ubiquitination (Fig. 4b), suggesting divergent metabolic
281 mechanisms. Abnormal neurological examination correlated positively with COSMIC DB entry
282 and negatively with Type I histology, which may reflect the effects of mutations on baseline
283 neurological function.

284

285 **MCD genes enriched in the excitatory neuronal lineage**

286 To infer the cell type in which MCD genes function, we mapped the net expression of the MCD
287 genes (i.e. eigengene) onto a published single-cell transcriptome dataset from the 2nd-trimester
288 human telencephalon, at a time when mutations probably arose⁵⁰ (Fig. 5a,b). This showed a
289 strong positive correlation of the net expression levels of the MCD geneset with dividing radial
290 glial cells (Pearson $r = 0.3655$, $p = 7.915 \times 10^{-121}$) and a moderate correlation in dividing
291 intermediate progenitor cells (IPCs; Pearson $r = 0.1527$, $p = 2.448 \times 10^{-21}$) and mature excitatory
292 neuron cells (Pearson $r = 0.1780$, $p = 1.557 \times 10^{-28}$). We found a lack of positive correlation with
293 inhibitory neuronal lineages including medial and central ganglionic eminences (MGE, CGE)
294 and mature interneuron clusters (Fig. 5c). We next performed deconvolution of the MCD geneset
295 into four module eigengenes, utilizing weighted co-expression network analysis (WGCNA),
296 which identified cell types classified as mature excitatory neurons (turquoise and blue),
297 microglia (brown), and unassigned (grey) (Fig. 5d). Quantification supported enrichment in
298 dividing radial glia, excitatory neurons, and microglia, the latter likely driven by MCD candidate
299 genes *IRF8* and *VSIG4* (Fig. 5e). Taken together, the expression of MCD genes is more enriched
300 in dorsal cortex neurogenic pools and implicated in the maturation of excitatory rather than
301 inhibitory neurogenic pools, as well as microglia.

302

303 **MCD gene expression is enriched in dysplastic cells**

304 We next performed single-nucleus RNA sequencing (snRNAseq) analysis in resected MCD
305 brain tissue. We reasoned that single-nucleus transcriptomes would be more revealing than bulk
306 transcriptomes, but the average VAF of ~6% in our MCD cohort meant that the vast majority of
307 sequenced cells would be genetically wild-type. We thus decided to focus snRNAseq on resected
308 cortex from patients with shared pathological MCD hallmarks across a range of VAFs. We
309 selected five resected brain samples, one from a patient with FCD (FCD-4512 *SLC24A2*
310 p.V631I, 0.6% VAF and *SRCAP* c.2630+1G>A, 1.64% VAF), two from patients with HME

311 (HME-4688 *PIK3CA* p.E545K, 25.1% VAF and HME-6593 *PIK3CA* p.H1047R, 13.1% VAF),
312 and two from patients with TSC meeting full diagnostic criteria. We also included brains from
313 four neurotypical cases for comparison. We sequenced a total of 33,206 nuclei (see Methods).

314 While the FCD brain snRNAseq data showed substantial overlapping cell clusters with
315 controls using UMAP clustering, HME and TSC brains showed distinct cell cluster distributions
316 (Fig. 6a,b). We found that very few HME cells matched expression patterns for typical brain
317 cells, even after standard normalization and scaling (Fig. 6b,c, see Methods). We characterized
318 transcriptomic profiles of each cluster based upon established marker gene expression (Fig. 6d)
319 as well as differentially expressed gene (DEG) analysis (Extended Data Fig. 7a, Supplementary
320 Table 5a) and weighted gene coexpression network analysis (WGCNA, Extended Data Fig. 7b,
321 Supplementary Table 5b) to assign likely cell types. We labeled these clusters according to their
322 closest relatives: ‘Excitatory neuron-like (ExN-L)’, ‘astrocyte-like (Ast-L)’, ‘oligodendrocyte
323 progenitor cell-like (OPC-L) or ‘oligodendrocyte-like (OD-L)’. Even with these categories, some
324 clusters remained undefined (U) (Fig. 6a, Extended Data Fig. 7c).

325 We noted that several of the HME clusters showed excessive expression of growth factor
326 receptor (GFR) gene families, specifically *FGFR1* in cluster OPC-L1/2 in HME, *FGFR2* in
327 cluster Ast-L1/3 and OD-L, *FGFR3* in Ast-L1/3, *EGFR* in Ast-L1/3 and OPC-L1/2, and
328 *PDGFRA* in cluster OPC-L1/2 (Fig. 6d). To identify the cell types expressing these genes, we
329 performed RNA in situ hybridization in HME brain sections followed by hematoxylin-eosin
330 staining. We found co-localization of these same *FGFR* family, *EGFR*, and *PDGFRA* transcripts
331 with dysplastic cells (Extended Data Fig. 8). Previous experiments indicate that it is most often
332 the dysplastic cells within HME and MCD that carry mosaic mutations⁹, suggesting an effect of
333 these mutations on growth factor receptor expressions that correlates with dysplasia.

334 Next, we investigated the net expression patterns of genes mutated in MCD within this
335 snRNAseq dataset. The net expression of the MCD geneset was enriched in Ast-L1/3 and OD-L,
336 which were labeled as dysplastic cells (Fig. 6e). Interestingly, the individual mutated genes
337 displayed converging expression patterns resulting in four different eigengenes (Fig. 6f, grey,
338 turquoise, blue and brown), with gene members for each eigengene (Extended Data Fig. 9).
339 These show distinct enrichment patterns across cell types, implying that members of each
340 eigengene may be associated with the pathophysiology of the corresponding dysplastic cell type
341 in HMEs.

342 We next performed a pseudo-bulk DEG analysis by aggregating all cell types within the
343 same disease condition, comparing HME, TSC, and FCD with CTRL (Extended Data Fig. 10a,
344 Supplementary Table 5c). and detected 1735, 1017, and 981 differentially expressed genes in
345 HME, TSC, and FCD, respectively. Intriguingly, 21.33% (16/75) or 13.33% (10/75) of MCD
346 mutated genes in our list overlapped with DEGs of HME or TSC, but less overlapped with FCD
347 (7/75, 9.33%) likely due to low VAF. Permutation tests suggest that the overlaps in HME and
348 TSC are unlikely to have arisen by chance (see Methods). This suggests that many MCD-
349 mutated genes are misregulated in MCD-specific cell types.

350 Next, we performed cell-type specific DEG analysis comparing disease with CTRL group
351 to investigate pathophysiology within Ast, ExN, OPC, and OD lineages determined by single-
352 cell analysis (Extended Data Fig. 10b, Supplementary Table 5c). Interestingly, in the HME brain,
353 the Ast lineage, which has the largest upregulated DEGs across cell types, showed increased
354 expression of genes related to lipid biosynthesis or energy metabolism and postsynaptic
355 membrane. These same genes were also enriched in mutated mTOR-expressing sorted cell
356 populations in a rodent FCD model⁵¹. In the TSC Ast lineage, the genes related to calcium

357 dynamics and synaptic function were strongly upregulated compared to CTRL Ast, which was
358 also found in the ExN lineage (Supplementary Table 5d). In FCD ExN lineage compared to
359 CTRL ExNs, terms relevant to mTOR downstream pathways (ribosome complex-related terms),
360 calcium dynamics (calcium channel complex), and synapse (postsynapse-related terms) were
361 enriched. Moreover, at the individual gene level, many histone remodeling enzymes or
362 transcription factors participating in gene expression regulation were encountered within the
363 DEGs of FCD ExN (Supplementary Table 5c), consistent with the four enriched terms in Fig. 2.
364 Taken together, the predicted pathways and functions in MCD based on the gene ontology of 75
365 MCD genes are altered in Ast or ExN lineage in MCD, suggesting underlying pathogenic
366 mechanisms of MCDs.

368 Discussion

369 Our multi-omics study of the genetic landscape of MCD confirmed the important role of
370 mTOR/MAP kinase and glycosylation pathways, seen in about 60.5% of those with mutations.
371 Moreover, our results also suggest a role for biological processes including gene expression,
372 synaptic function, and calcium dynamics, which made up the other 39.5% of those with
373 mutations. While these pathways may be independent in mediating MCD pathophysiology, it is
374 possible that they may have some crosstalk⁵²⁻⁵⁴.

375 Only 76 of 283 patients showed one or more putative somatic mutations as a likely cause
376 of MCD. There could be numerous causes for the relatively low solve rate in MCD, including the
377 potential to miss very low VAF (<0.5%) mutations and the contribution of complex mutations
378 that could be missed by our pipeline. Finally, although patients with environmental causes,
379 syndromic, or inherited causes were excluded from our cohort, these factors could still contribute
380 to MCD.

381 With our approach, we identified two recurrently-mutated genes not previously
382 implicated in MCD and validated 3 other genes found mutated in single patients from prior
383 studies. Confirming the remaining candidate and identifying further MCD candidate genes will
384 require larger MCD cohorts. Including novel MCD candidate genes emerging from 300X WES
385 into the 1000X Phase 3 AmpliSeq allowed both confirmation of mutations, a more accurate
386 estimate of VAF, and identification of additional patients with these genes mutated that would
387 have been perhaps missed with 300X WES. *In vivo* functional validation by modeling mutations
388 in developing mouse brains suggests many identified genes likely contribute to disease. Like
389 with de novo germline mutations discovered in autism, we suggest that there could be dozens if
390 not hundreds of additional MCD genes, based on the low number of recurrently mutated genes
391⁵⁵.

392 The four gene networks, mTOR/MAP kinase, calcium dynamics, synapse, and gene
393 expression, play important roles both during neurogenesis and neuronal migration, as well as in
394 establishing neuronal excitability. For instance, calcium dynamics regulates cytoskeletal activity
395 and excitability^{56,57}. Genotypic information also showed correlations with phenotype, for
396 instance, PET brain hypometabolism and abnormality in the neurological examination correlated
397 with the presence of a likely oncogenic variant within COSMIC DB.

398 We also characterized the expression patterns of MCD mutated genes in neurotypical and
399 MCD brains at single-cell resolution. The cell types most strongly expressing mutated genes
400 include dorsal forebrain radial glial progenitors and their daughter excitatory neurons, consistent
401 with the likely site of origin of somatic brain mutations⁵⁸. The fact that mutated genes also
402 showed the strongest enrichment with these same clusters suggests that mutated genes can drive

403 gene expression in convergent pathways. The prior studies indicating that MCD dysplastic cells
404 express markers for both glia and neurons⁵⁹ are consistent with our findings, with mutations
405 driving critical roles predominantly in dividing radial glia, with profound effects on lineage and
406 cellular dysplasia. The MCD genes in patient brains found in our study demonstrated critical
407 roles during cortical development, significantly correlating with patient phenotypes. These
408 findings could lead to new molecular classifications and diagnostic strategies for MCD, and
409 ultimately to personalized therapies for epilepsy.

410

411 **Acknowledgments**

412 AmpliSeq, TASEq, and snRNAseq were supported by NIH P30CA023100 and S10OD026929 at
413 the UCSD IGM Genomics Center. Rady Children's Institute for Genomic Medicine, Broad
414 Institute (U54HG003067, UM1HG008900), the Yale Center for Mendelian Disorders
415 (U54HG006504), and the New York Genome Center provided whole-exome sequencing. UCSD
416 Microscopy core (NINDS P30NS047101) provided imaging support. UCSD Tissue Technology
417 Shared Resources Team (National Cancer Institute Cancer Center Support Grant, P30CA23100)
418 supported paraffin sectioning and H&E staining. This study is supported by the 2021 NARSAD
419 Young Investigator Grant from Brain & Behavior Research Foundation (30598 to CC), NIH
420 (NIMH U01MH108898 and R01MH124890 to JGG and GWM, and NIA R21AG070462,
421 NINDS R01NS083823 to JGG), the Regione Toscana under the Call for Health 2018
422 (DECODE-EE to RG) and Fondazione Cassa di Risparmio di Firenze (to RG). Fig. 1b and Fig.
423 3a are created with BioRender.com. The funders had no role in the study design, data collection,
424 analysis, decision to publish, or preparation of the manuscript.

425

426 **Author contributions**

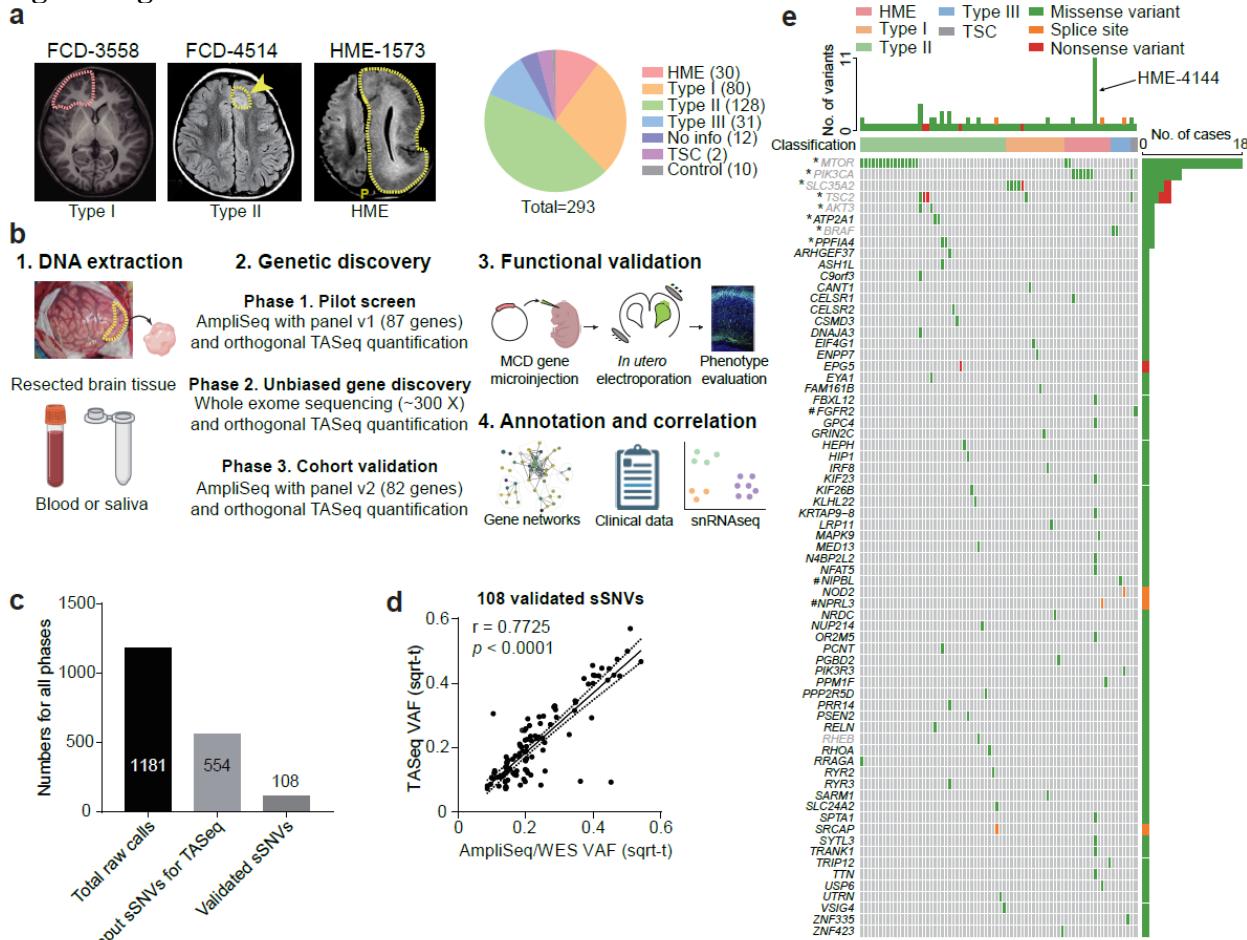
427 C.C., X.Y., Sa.B., St.B., and J.G.G. designed the study. C.C., S.M., and S.K. conducted
428 functional validation. C.B., V.S., A.S.N., E.R., C.C., and G.H. coordinated the clinical database.
429 X.Y., C.C., M.W.B., L.L.B., R.D.G., J.G., M.X., A.P.L.M., and K.N.J. organized, handled, and
430 sequenced human samples. X.Y., C.C., T.B., Y.W., A.A., X.X., Z.L., and B.C. performed
431 bioinformatics and data analysis. C.C. and K.I.V. performed the RNAscope experiment. C.D.,
432 H.W.P., C.A.B.G., S.H.K., H.K., H.U., M.P., A.S., C.A.H., D.D.L., C.A.G., M.D.S., S.S., M.N.,
433 D.D.G., K.I., Y.T., H.C., J.T., V.C., R.G., O.D., W.A.S., H.R.M., and G.W.M. provided resected
434 brain tissues and clinical data from FCD patients. C.C., X.Y., and J.G.G. wrote the manuscript.
435 All authors read and commented on the manuscript before submission.

436

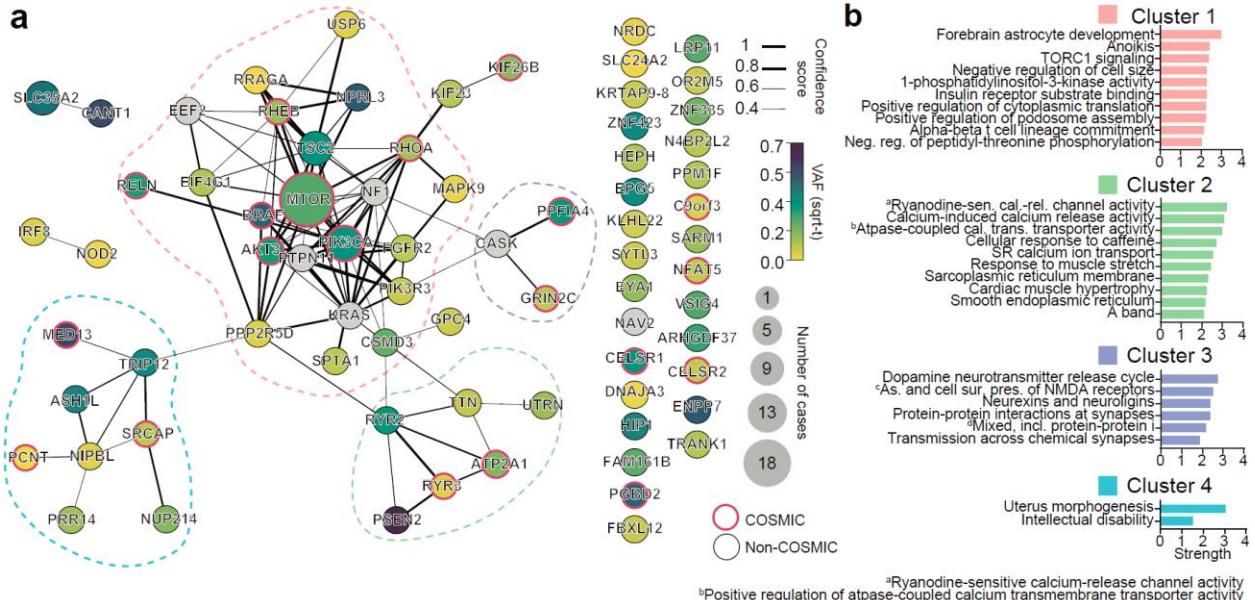
437 **Competing Interests Statement**

438 The authors declare no competing interests.

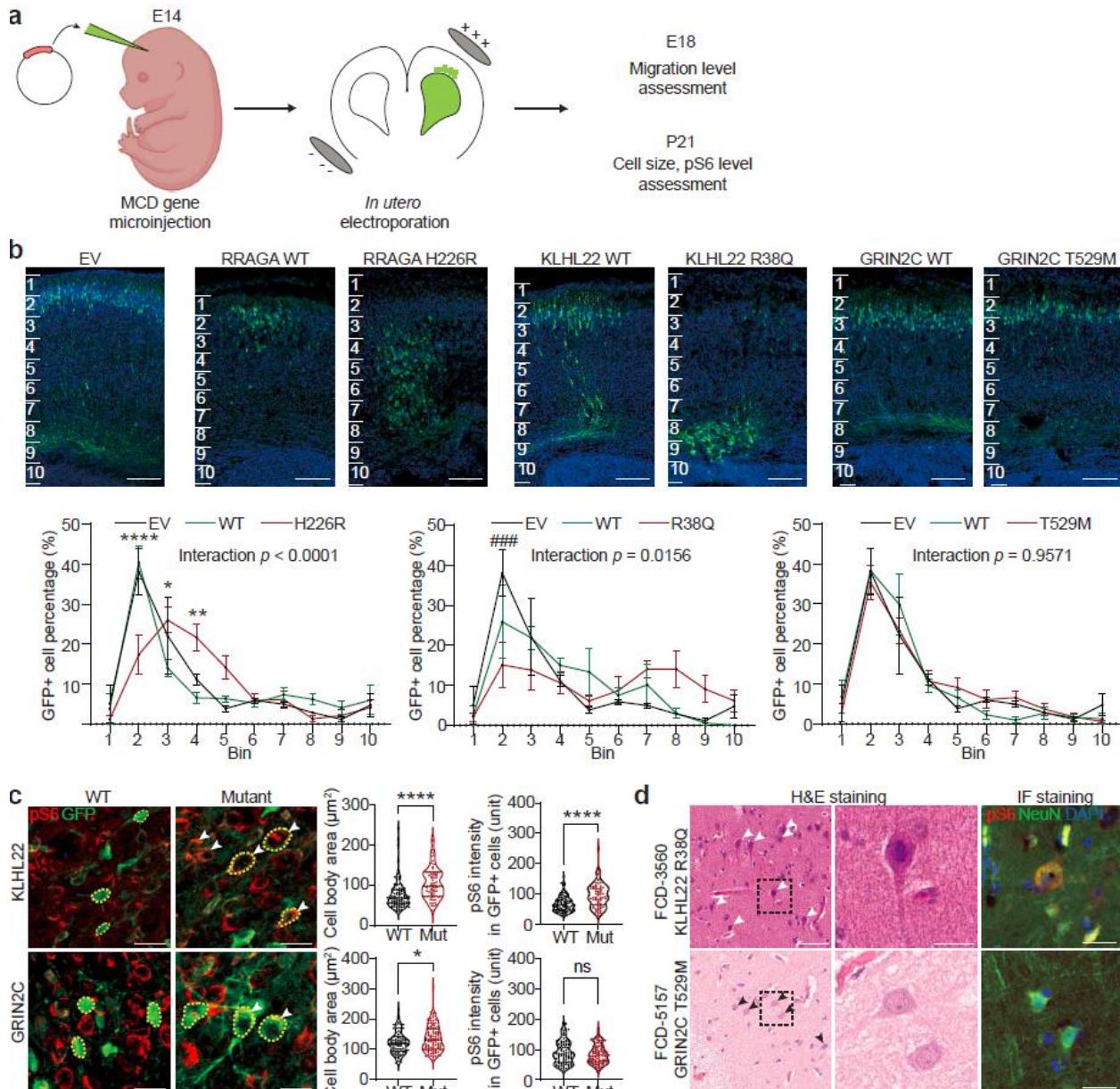
439

440 **Figure Legends**

441 **Figure 1. Comprehensive genetic profiling and validation of somatic variants in 283 MCD**
442 **patients.** (a) Representative MRI image of FCD-3558 (FCD type I), FCD-4514 (FCD type II),
443 HME-1573, and composition MCD cohort. Yellow arrow and dash: affected brain regions. FCD-
444 3558, MRI was non-lesional, the epileptic focus was mapped to the right frontal lobe (red dashed
445 line) and resected. (b) Three-phase comprehensive genetic MCD profiling workflow, followed
446 by quantification/validation of each variant with target amplicon sequencing (TASeq). Phase 1] 1000 X pilot
447 screening of DNA with an 87-gene mTOR-related panel. Phase 2] 300 X whole-exome sequencing (WES)
448 with best-practice somatic variant discovery for novel candidate genes. Phase 3] Cohort-level validation
449 with an updated, high-confidence TASeq gene set arising from Phases 1 and 2. A subset of somatic mutations
450 was functionally validated in mice.
451 Annotation and correlation with gene networks, clinical data, and single-single-nucleus RNAseq.
452 (c) 1181 sSNV calls were detected from all three phases, yielding 108 validated sSNVs. (d)
453 Correlation between square-root transformed (sqrt-t) AmpliSeq/WES variant allele fraction
454 (VAF) and TASeq VAF. Solid line: Simple linear regression (goodness of fit). Dashed lines:
455 95% confidence band. Spearman correlation coefficient ρ and corresponding two-tailed t -test p -value.
456 (e) Oncoplot of 69 genes from 108 validated sSNVs. Top: most patients had one gene
457 mutated, a few patients had more than one gene mutated, and patient HME-4144 had 11 different
458 validated genes mutated. Grey: genes recurrently mutated in previous HME/FCD cohorts. *:
459 genes recurrently mutated in our cohort. #: genes non-recurrently mutated but complementary to
460 other cohorts of epilepsy-associated developmental lesions.

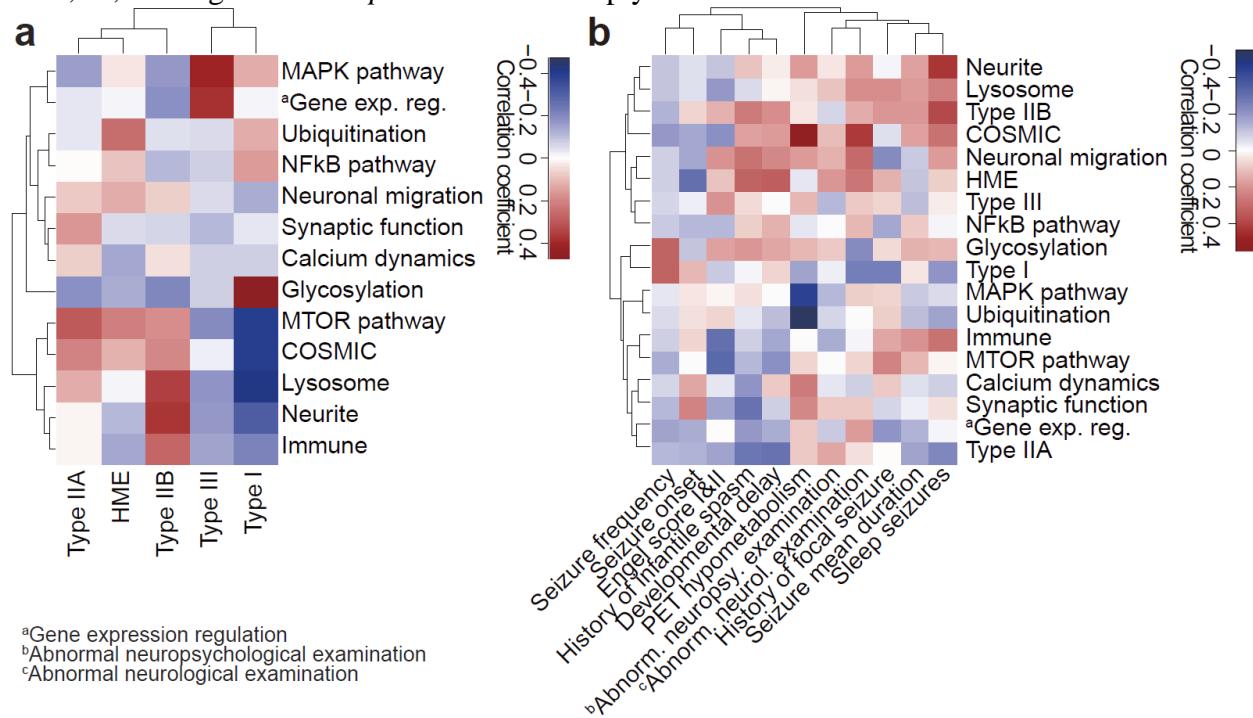


462
463 **Figure 2. Genes mutated in MCD highlight four major gene networks.** (a) STRING DB
464 pathway analysis of the 69 MCD discovered genes and six novel genes (a total of 75 genes) from
465 recent publications identify MTOR/MAP kinase pathway (pink, Cluster 1), Calcium dynamics
466 (green, Cluster 2), Synapse (purple, Cluster 3), Gene expression (blue, Cluster 4). Edge
467 thickness: STRING confidence score. Node size and color: square root transformed (sqrt-t)
468 number of patients carrying a given mutation and average VAF across all patients, respectively.
469 Non-clustered orphan genes were listed at right. Red border: variant reported in COSMIC DB.
470 (b) Gene Ontology (GO) analysis results confirm the functions of compositions in each network.
471 Top GO terms or KEGG pathways based on strength. Strength calculated by STRING.

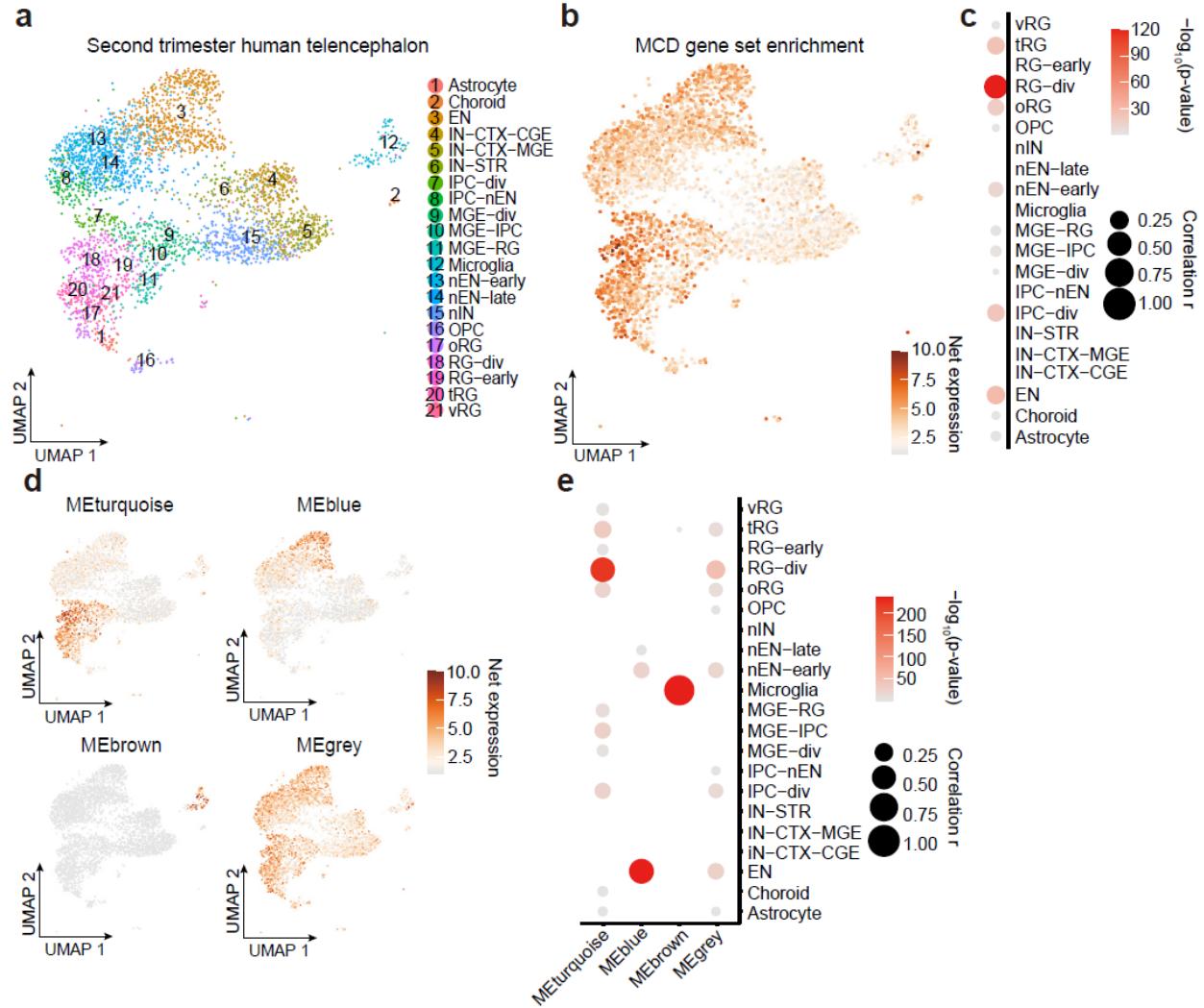


472
473 **Figure 3. Selected novel MCD somatic variants show functional defects in embryonic**
474 **mouse brain and patient samples.** (a) Validation of candidate mosaic variants in mice. (b) Two
475 different mutations in novel FCD type II genes, *RRAGA* H226R and *KLHL22* R38Q, but not a
476 novel FCD type I gene, *GRIN2C*, disrupt cellular radial migration from the subventricular zone
477 (SVZ). Below: two-way ANOVA and Sidak multiple comparisons with *p*-values of interaction
478 between genotype and bin factor. * or # indicates a *p*-value in comparison between WT and
479 mutant group, or EV and mutant group respectively. Ten bins from the surface of the cortex (top)
480 to SVZ (bottom). Scale bar: 100 μ m. Mean \pm SEM (standard error mean). n=3, 3, 6, 3, 6, 4, 3
481 biologically independent animals for EV, RRAGA WT, RRAGA H226R, KLHL22 WT,
482 KLHL22 R38Q, GRIN2C WT, and GRIN2C T529M, respectively. (c) Immunofluorescence in
483 postnatal day 21 mouse cortices for KLHL22 and GRIN2C wild-type (WT) or mutant isoform.
484 Neurons expressing mutant KLHL22 and GRIN2C recapitulate histological phenotypes shown in
485 (d), with enlarged cell bodies (white arrow) compared to WT isoforms (WT control), whereas

486 only neurons expressing KLHL22 but not GRIN2C mutant isoform display increased pS6 levels
 487 compared to control. Yellow dashed lines: examples of cell body size quantification. Two-tailed
 488 Mann-Whitney *U*-test. Dashed lines and dotted lines in the violin plots indicate median and
 489 quartiles, respectively. Scale bar: 20 μ m. n=105 cells (3 mice), 70(3), 95(3), 107(3) for KLHL22
 490 WT, KLHL22 R38Q, GRIN2C WT, and GRIN2C T529M, respectively. (d) H&E and phospho-
 491 S6 (pS6) staining of the resected brain from FCD-3560 and FCD-5157. One representative H&E
 492 or IF staining is shown for each patient case. The box area is zoomed in the middle image.
 493 Arrows: dysplastic cells. Right: Immunofluorescence (IF) for pS6 and NeuN. Note that
 494 dysplastic pS6-positive neurons with increased pS6 levels are present in FCD-3560 but not in
 495 FCD-5157. Scale bar: 60 μ m on the left, 20 μ m on the middle and right. *** $p < 0.0001$; * $p <$
 496 0.05; ns, non-significant. ## $p < 0.001$. EV: empty vector.



497
 498 **Figure 4. Clinical phenotypic outcomes correlate with genotype-based classifications in**
 499 **MCD.** (a) Correlation heatmap for classification based on genetic information (y-axis) vs.
 500 International League Against Epilepsy (ILAE) classification based on histology (x-axis) using
 501 Pearson correlation. Shade: the value of the Phi coefficient. Note that Type IIA and HME are
 502 enriched with mTOR and Ubiquitination genes, while Type I is enriched in Glycosylation and
 503 depleted in MTOR and COSMIC genes. HME: hemimegalencephaly. (b) Correlation between
 504 classification based on genetic information and various clinical phenotypes. Shade: the value of
 505 Phi (binary data) or Pearson (continuous) correlation coefficient. For example, positron emission
 506 tomography (PET) hypometabolism is enriched in COSMIC genes and depleted in the MAPK
 507 pathway, whereas abnormal neurological examination is enriched in COSMIC genes. Raw data
 508 are in Supplementary Table 4.



509
510 **Figure 5. Single-nucleus transcriptomes reveal MCD gene enrichment in radial glia and**
511 **excitatory neurons in the developing human cortex.** (a) Uniform Manifold Approximation and
512 Projection (UMAP) for single-nucleus transcriptome in 2nd-trimester fetal human telencephalon
513 from a public dataset⁵⁰. (b) UMAP enrichment patterns of an eigengene using MCD genes. Note
514 enrichment for excitatory neurons and radial glia (dark brown). vRG: vertical radial glia, tRG:
515 truncated radial glia, RG-div: dividing radial glia, oRG: outer radial glia, EN: excitatory neuron,
516 nEN: newborn excitatory neuron, IPC: intermediate progenitor cell, STR: striatum, IN:
517 interneuron, CTX: cortex, MGE: medial ganglionic eminence, CGE: central ganglionic
518 eminence. (c) Quantification of enrichment of (b) based on cell types, showing enrichment for
519 RG-div. (d) Four eigengenes decomposed from (b). (e) Quantification of enrichment of (d) based
520 on cell types showing enrichment in dividing radial glia, microglia, and inhibitory cortical
521 neurons from the medial ganglionic eminence (MGE). Net expression: Relative and scaled net
522 expression level of a given eigengene generated by a defined gene list. The size and color of the
523 dot plots in (c) and (e) indicate Pearson correlation coefficient r and corresponding non-adjusted
524 asymptotic p-value derived from a two-sided Student's t-test, respectively.

525

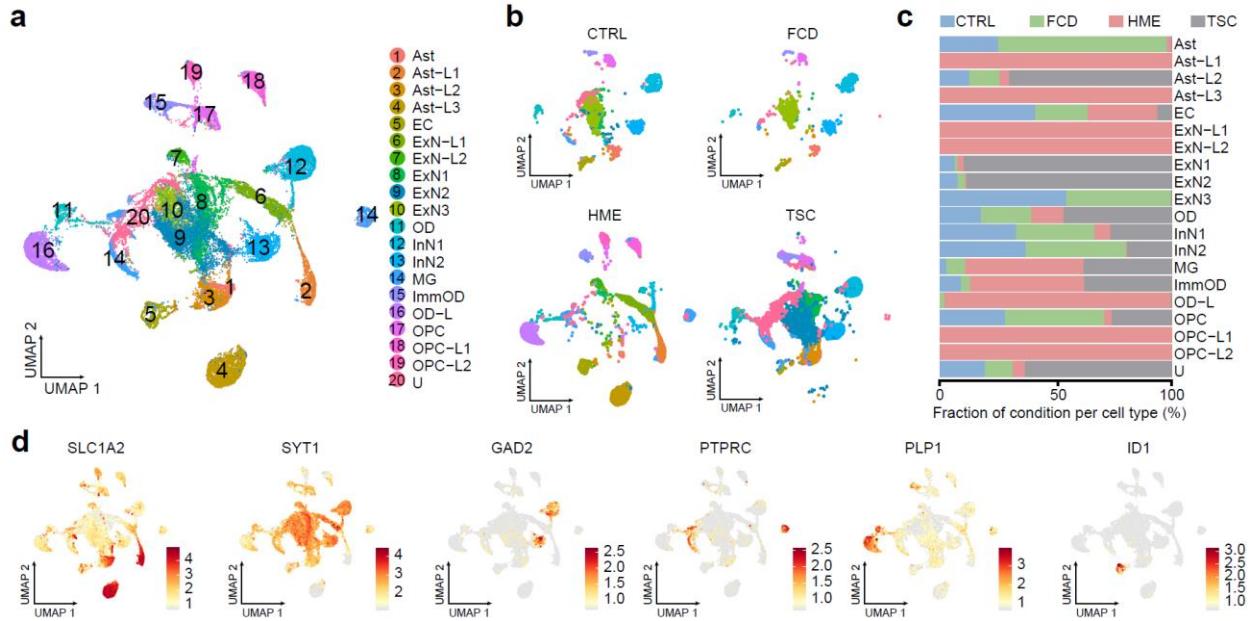


Figure 6. Single-nucleus transcriptomes showed MCD gene expression enriched in MCD-specific cell types. (a) UMAP for the 33,206 single-nucleus transcriptomes from cortical resections, revealing disrupted cell clusters, especially for HME and TSC brain, but only mild disruption in FCD brain. Cell type classification. Ast: astrocyte, EC: endothelial cells, ExN: excitatory neuron, ImmOD: immature oligodendrocyte, InN: inhibitory neuron, MG: microglia, OD: oligodendrocyte, OPC: Oligodendrocyte precursor cells, U: unidentified. (b) UMAP is split by disease condition. (c) The proportion of disease conditions for each cell type. (d) Expression pattern of selected marker genes in normal human cortex vs. several atypical markers related to brain cancer or Alzheimer's disease. Color scale: Average expression level. (e) Eigengene with all 75 MCD genes and quantification of enrichment based on cell types. (f) four genes decomposed from (e) and quantification of results. Net expression: Relative and scaled net expression level of a given eigengene generated by a defined gene list. The size and color of the dot plots in (e) and (f) are Pearson correlation coefficient r and corresponding non-adjusted asymptotic p -value derived from a two-sided Student's t -test, respectively.

526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541

542 **References**

- 544 1. Leventer, R.J., Guerrini, R. & Dobyns, W.B. Malformations of cortical development and epilepsy. *Dialogues*
545 *Clin Neurosci* 10, 47-62 (2008).
- 546 2. Barkovich, A.J., Dobyns, W.B. & Guerrini, R. Malformations of cortical development and epilepsy. *Cold*
547 *Spring Harb Perspect Med* 5, a022392 (2015).
- 548 3. Blumcke, I. et al. The clinicopathologic spectrum of focal cortical dysplasias: a consensus classification
549 proposed by an ad hoc Task Force of the ILAE Diagnostic Methods Commission. *Epilepsia* 52, 158-74 (2011).
- 550 4. Choi, S.A. & Kim, K.J. The Surgical and Cognitive Outcomes of Focal Cortical Dysplasia. *J Korean*
551 *Neurosurg Soc* 62, 321-327 (2019).
- 552 5. Krsek, P. et al. Different features of histopathological subtypes of pediatric focal cortical dysplasia. *Ann*
553 *Neurol* 63, 758-69 (2008).
- 554 6. Bast, T., Ramantani, G., Seitz, A. & Rating, D. Focal cortical dysplasia: prevalence, clinical presentation and
555 epilepsy in children and adults. *Acta Neurol Scand* 113, 72-81 (2006).
- 556 7. Poduri, A. et al. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron*
557 74, 41-8 (2012).
- 558 8. Lee, J.H. et al. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause
559 hemimegalencephaly. *Nat Genet* 44, 941-5 (2012).
- 560 9. Baldassari, S. et al. Dissecting the genetic basis of focal cortical dysplasia: a large cohort study. *Acta*
561 *Neuropathol* 138, 885-900 (2019).
- 562 10. Lai, D. et al. Somatic variants in diverse genes leads to a spectrum of focal cortical malformations. *Brain*
563 (2022).
- 564 11. Bedrosian, T.A. et al. Detection of brain somatic variation in epilepsy-associated developmental lesions.
565 *Epilepsia* (2022).
- 566 12. Sim, N.S. et al. Precise detection of low-level somatic mutation in resected epilepsy brain tissue. *Acta*
567 *Neuropathol* 138, 901-912 (2019).
- 568 13. Dou, Y. et al. Accurate detection of mosaic variants in sequencing data without matched controls. *Nat*
569 *Biotechnol* 38, 314-319 (2020).
- 570 14. Yang, X. et al. Control-independent mosaic single nucleotide variant detection with DeepMosaic. *Nat*
571 *Biotechnol* (2022). doi.org/10.1038/s41587-022-01559-w
- 572 15. Huang, A.Y. et al. MosaicHunter: accurate detection of postzygotic single-nucleotide mosaicism through
573 next-generation sequencing of unpaired, trio, and paired samples. *Nucleic Acids Res* 45, e76 (2017).
- 574 16. Wang, Y. et al. Comprehensive identification of somatic nucleotide variants in human brain tissue. *Genome*
575 *Biol* 22, 92 (2021).
- 576 17. Breuss, M.W. et al. Autism risk in offspring can be assessed through quantification of male sperm mosaicism.
577 *Nat Med* 26, 143-150 (2020).
- 578 18. Yang, X. et al. Developmental and temporal characteristics of clonal sperm mosaicism. *Cell* 184, 4772-4783
579 e15 (2021).
- 580 19. Garcia, C.A.B. et al. mTOR pathway somatic variants and the molecular pathogenesis of
581 hemimegalencephaly. *Epilepsia Open* 5, 97-106 (2020).
- 582 20. Pelorosso, C. et al. Somatic double-hit in MTOR and RPS6 in hemimegalencephaly with intractable epilepsy.
583 *Hum Mol Genet* 28, 3755-3765 (2019).
- 584 21. Breuss, M.W. et al. Somatic mosaicism in the mature brain reveals clonal cellular distributions during cortical
585 development. *Nature* (2022).
- 586 22. Lai, D. et al. Somatic mutation involving diverse genes leads to a spectrum of focal cortical malformations.
587 *medRxiv* (2021).
- 588 23. Lee, W.S. et al. Gradient of brain mosaic RHEB variants causes a continuum of cortical dysplasia. *Ann Clin*
589 *Transl Neurol* 8, 485-490 (2021).
- 590 24. Bae, T. et al. Somatic mutations reveal hypermutable brains and are associated with neuropsychiatric
591 disorders. *medRxiv* (2022).
- 592 25. Bozic, I. et al. Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad*
593 *Sci U S A* 107, 18545-50 (2010).
- 594 26. Alexandrov, L.B. et al. The repertoire of mutational signatures in human cancer. *Nature* 578, 94-101 (2020).
- 595 27. Kim, M. & Costello, J. DNA methylation: an epigenetic mark of cellular memory. *Exp Mol Med* 49, e322
596 (2017).

- 597 28. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting
598 functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47, D607-D613 (2019).
- 599 29. Bedrosian, T.A. et al. Detection of brain somatic variation in epilepsy-associated developmental lesions.
600 medRxiv (2021).
- 601 30. Tarabeux, J. et al. Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders
602 and schizophrenia. *Transl Psychiatry* 1, e55 (2011).
- 603 31. Bezprozvanny, I. Calcium signaling and neurodegenerative diseases. *Trends Mol Med* 15, 89-100 (2009).
- 604 32. van den Berg, D.L.C. et al. Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical
605 Neuron Migration. *Neuron* 93, 348-361 (2017).
- 606 33. Yang, Y.J. et al. Microcephaly gene links trithorax and REST/NRSF to control neural stem cell proliferation
607 and differentiation. *Cell* 151, 1097-112 (2012).
- 608 34. Massimino, L. et al. TBR2 antagonizes retinoic acid dependent neuronal differentiation by repressing Zfp423
609 during corticogenesis. *Dev Biol* 434, 231-248 (2018).
- 610 35. Su, M.Y. et al. Hybrid Structure of the RagA/C-Ragulator mTORC1 Activation Complex. *Mol Cell* 68, 835-
611 846 e3 (2017).
- 612 36. Chen, J. et al. KLHL22 activates amino-acid-dependent mTORC1 signalling to promote tumorigenesis and
613 ageing. *Nature* 557, 585-589 (2018).
- 614 37. Behar, T.N. et al. Glutamate acting at NMDA receptors stimulates embryonic cortical neuronal migration. *J
615 Neurosci* 19, 4449-61 (1999).
- 616 38. Paoletti, P., Bellone, C. & Zhou, Q. NMDA receptor subunit diversity: impact on receptor properties, synaptic
617 plasticity and disease. *Nat Rev Neurosci* 14, 383-400 (2013).
- 618 39. Strehlow, V. et al. GRIN2A-related disorders: genotype and functional consequence predict phenotype. *Brain*
619 142, 80-92 (2019).
- 620 40. Prickett, T.D. & Samuels, Y. Molecular pathways: dysregulated glutamatergic signaling pathways in cancer.
621 *Clin Cancer Res* 18, 4240-6 (2012).
- 622 41. Tate, J.G. et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 47, D941-D947
623 (2019).
- 624 42. Ruvinsky, I. & Meyuhas, O. Ribosomal protein S6 phosphorylation: from protein synthesis to cell size.
625 *Trends Biochem Sci* 31, 342-8 (2006).
- 626 43. Sumimoto, H., Imabayashi, F., Iwata, T. & Kawakami, Y. The BRAF-MAPK signaling pathway is essential
627 for cancer-immune evasion in human melanoma cells. *J Exp Med* 203, 1651-6 (2006).
- 628 44. Ornitz, D.M. & Itoh, N. The Fibroblast Growth Factor signaling pathway. *Wiley Interdiscip Rev Dev Biol* 4,
629 215-66 (2015).
- 630 45. Chen, X. et al. TNF-alpha-Induced NOD2 and RIP2 Contribute to the Up-Regulation of Cytokines Induced
631 by MDP in Monocytic THP-1 Cells. *J Cell Biochem* 119, 5072-5081 (2018).
- 632 46. Yates, T.M. et al. SLC35A2-related congenital disorder of glycosylation: Defining the phenotype. *Eur J
633 Paediatr Neurol* 22, 1095-1102 (2018).
- 634 47. Paganini, C. et al. Calcium activated nucleotidase 1 (CANT1) is critical for glycosaminoglycan biosynthesis
635 in cartilage and endochondral ossification. *Matrix Biol* 81, 70-90 (2019).
- 636 48. Lee, N. et al. Neuronal migration disorders: positron emission tomography correlations. *Ann Neurol* 35, 290-
637 7 (1994).
- 638 49. Kim, Y.H. et al. Neuroimaging in identifying focal cortical dysplasia and prognostic factors in pediatric and
639 adolescent epilepsy surgery. *Epilepsia* 52, 722-7 (2011).
- 640 50. Nowakowski, T.J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the
641 human cortex. *Science* 358, 1318-1323 (2017).
- 642 51. Kim, J.K. et al. Brain somatic mutations in MTOR reveal translational dysregulations underlying intractable
643 focal epilepsy. *J Clin Invest* 129, 4207-4223 (2019).
- 644 52. Li, R.J. et al. Regulation of mTORC1 by lysosomal calcium and calmodulin. *Elife* 5(2016).
- 645 53. Hoeffer, C.A. & Klann, E. mTOR signaling: at the crossroads of plasticity, memory and disease. *Trends
646 Neurosci* 33, 67-75 (2010).
- 647 54. Laplante, M. & Sabatini, D.M. Regulation of mTORC1 and its impact on gene expression at a glance. *J Cell
648 Sci* 126, 1713-9 (2013).
- 649 55. Coe, B.P. et al. Neurodevelopmental disease genes implicated by de novo mutation and copy number
650 variation morbidity. *Nat Genet* 51, 106-116 (2019).
- 651 56. Ridley, A.J. et al. Cell migration: integrating signals from front to back. *Science* 302, 1704-9 (2003).
- 652 57. Brini, M., Cali, T., Ottolini, D. & Carafoli, E. Neuronal calcium signaling: function and dysfunction. *Cell*

- 653 Mol Life Sci 71, 2787-814 (2014).
- 654 58. Lamparello, P. et al. Developmental lineage of cell types in cortical dysplasia with balloon cells. Brain 130,
655 2267-76 (2007).
- 656 59. Englund, C., Folkerth, R.D., Born, D., Lacy, J.M. & Hevner, R.F. Aberrant neuronal-glial differentiation in
657 Taylor-type focal cortical dysplasia (type IIA/B). Acta Neuropathol 109, 519-33 (2005).

658 **Methods**

659
660 The study protocol was approved by the UC San Diego IRB (#140028). Informed consent was
661 obtained from all participants or their legal guardians at the time of enrollment. All work with
662 mice was performed in accordance with UCSD IACUC protocol S15113.

663
664 **Overview of the FCD cohort**

665 This study is a multi-center international collaboration. We recruited a cohort of 283 individuals
666 from the ‘FCD Neurogenetics Consortium’ (see the member list). These individuals were
667 diagnosed with FCD type I, II, III, HME, or TSC and underwent surgical resection to treat drug-
668 resistant epilepsy between 2013 and 2021. Any cases that underwent surgical resection due to
669 environmental factors, for example, stroke, or acute trauma, were excluded. For each individual,
670 resected brain tissue was collected, along with paired blood or saliva samples and parental
671 samples, where available. Clinical history, pre- and post-operative brain imaging,
672 histopathology, ILAE classification according to the surgical tissue pathology report, and Engel
673 surgical outcome score (at least two years after surgery) were collected, when available.

674
675 **DNA extraction**

676 Pulverized cortical samples (~0.3 g) were homogenized with a Pellet Pestle Motor (Kimble,
677 #749540-0000) or Handheld Homogenizer Motor (Fisherbrand, #150) depending on the size of
678 the tissue, and resuspended with 450 µL RLT buffer (Qiagen, #40724) in a 1.5 ml
679 microcentrifuge tube (USA Scientific, #1615-5500). Homogenates were then vortexed for 1
680 minute and incubated at 70°C for 30 minutes. 50 µL Bond-Breaker TCEP solution (Thermo
681 Scientific, #77720) and 120 mg stainless steel beads with 0.2 mm diameter (Next Advance,
682 #SSB02) were added, and cellular disruption was performed for 5 minutes on a DisruptorGenie
683 (Scientific industries). The supernatant was transferred to a DNA Mini Column from an AllPrep
684 DNA/RNA Mini Kit (Qiagen, #80204) and centrifuged at 8500 xg for 30 seconds. The column
685 was then washed with Buffer AW1 (kit-supplied), centrifuged at 8500 xg for 30 seconds and
686 washed again with Buffer AW2 (kit-supplied), and then centrifuged at full speed for 2 minutes.
687 The DNA was eluted two times with 50 µL of pre-heated (70°C) EB (kit-supplied) through
688 centrifugation at 8,500 xg for 1 minute.

689
690 **AmpliSeq and WES sequencing for somatic mutation candidates**

691 AmpliSeq and whole-exome sequencing (WES) were used at different phases to perform the
692 genetic screening within available samples from the cohort. Customized AmpliSeq DNA panels
693 for Illumina (Illumina, #20020495) were used for Massive Parallel Amplicon Sequencing²¹. 87
694 or 82 genes related to the mTOR pathway or curated based on the results of Phase 1 and 2,
695 respectively, were subjected to the AmpliSeq design system; a list of designed genes is provided
696 in Supplementary Table 2a-c. Two pools were designed for tiling the capture region. Genomic
697 DNA from extracted tissue was diluted to 5 ng/uL in low TE provided in AmpliSeq Library
698 PLUS (384 Reactions) kit (Illumina, #20019103). AmpliSeq was carried out following the
699 manufacturer's protocol (document #1000000036408v07). For amplification, 14 cycles each with
700 8 minutes were used. After amplification and FUPA treatment, libraries were barcoded with
701 AmpliSeq CD Indexes (Illumina, #20031676) and pooled with similar molecular numbers based
702 on measurements made with a Qubit dsDNA High Sensitivity kit (Thermo Fisher Scientific,

703 #Q32854) and a plate reader (Eppendorf, PlateReader AF2200). The pooled libraries were
704 subjected to Illumina NovaSeq 6000 platform for PE150 sequencing. The AmpliSeq design in
705 ‘Phase 1’ is under design ID IAA7610, and the AmpliSeq design in ‘Phase 3’ is under design ID
706 IAA26010.

707 Genomic DNA (~ 1.0 µg) was prepared for whole-exome sequencing, and libraries were
708 captured using the Agilent SureSelect XT Human All Exon v.5 or Nextera DNA Exome kits.
709 Then, 100, 125, or 150 bp paired-end reads (median insert size ~ 210 bp) were generated using
710 the Illumina HiSeq X 2500 platform. The sequencing experiments were designed to yield three
711 datasets of ~ 100X coverage on each sample, with a coverage goal of 300X from the brain and
712 100X from blood/saliva.

713

714 **Somatic variant calling from AmpliSeq and WES**

715 Reads were aligned to GRCh37 using BWA (version 3.7.16a), sorted per each read group with
716 SAMtools (version 1.7 for WES, version 1.9 for AmpliSeq), and merged into a single BAM file
717 with Sambamba (version 0.6.7). The merged BAM files were marked for duplicate reads using
718 Picard (v2.12.1 for WES, v2.18.27 for AmpliSeq), duplicated reads were not removed for
719 AmpliSeq because of the nature of the method. Then, we performed indel realignment and base
720 quality recalibration using GATK (v3.7.0), resulting in the final uniformed processed BAM files.

721 Both tissue-specific and tissue-shared mosaic variants were called from the AmpliSeq
722 and WES sequencing data. AmpliSeq and WES variants were called according to the availability
723 of the control tissue. Brain- and blood/saliva-specific variants were called using MuTect2
724 (GATK v3.8.1 for WES, v4.0.4 for AmpliSeq) paired mode and Strelka2 (v2.9.2) somatic
725 mode⁶⁰; the BAM files from the brain sample (combined and non-combined from independent
726 sequencing libraries) and blood/saliva samples were treated as “tumor-normal” and “normal-
727 tumor” pairs separately and cross-compared between each other. Variants called by both callers
728 were listed. Mosaic variants shared between the brain and blood/saliva samples were called using
729 the single mode of MosaicHunter¹⁵ by either combining all brain replicates or calling each
730 separate sample. Variants that passed all the MosaicHunter filters also were listed. Somatic
731 variants from WES data were further called by GATK (v3.7.0) haplotypewriter with ploidy
732 parameter set to 50, followed by a series of heuristic filters described as the best-practice by the
733 Brain somatic mosaicism network¹⁶, tissue-shared variants were called by the combination of
734 MuTect2⁶¹ (GATK v3.8.1) single-mode and DeepMosaic¹⁴.

735 A union of different pipelines was selected to get maximum sensitivity. Mosaic
736 candidates from the combined lists were refined using the following criteria: (i) the variant had
737 more than 3 reads for the alternative allele; (ii) the variant was not present in UCSC repeat
738 masker or segmental duplications; (iii) the variant was at least 2 bp away from a homopolymeric
739 tract; and (iv) the variant exhibited a gnomAD allele frequency lower than 0.001. Variants that
740 exist in the 1000 genome project (phase 3) also were excluded from the analysis. Variants from
741 both exome data sources were tested and a combination of tissue-specific mosaic variants and
742 tissue-shared mosaic variants were collected and the credible interval of VAFs was calculated
743 using a Bayesian-based method described previously⁶². To filter for candidate MCD disease-
744 causing variants, we further filtered out synonymous variants in coding regions, variants with
745 CADD Phred score < 25, and candidates that fell out of coding regions and were not predicted to
746 affect splicing, annotated by ANNOVAR and BEDtools (version 2.27.1), the annotation scripts
747 were provided on GitHub (https://github.com/shishenyxx/MCD_mosaic).

748 **False discovery estimation**

749 To calculate the false discovery of random variants detected in normal samples, we incorporated
750 75 normal control samples (71 brains and 4 other organs) previously sequenced with 250-300X
751 WGS, which should provide similar sensitivity as our exomes, the deep WGS were generated by
752 efforts from the NIMH Brain Somatic Mosaic Consortium¹⁶, from controls²¹, and from our
753 recent mutation detection pipeline²⁴. Variants were filtered based on the identical criteria as
754 described in the above data analysis part, with >0.005 VAF, all on exonic regions defined by
755 NCBI, and CADD score >25. While 13 variants remain positive from this pipeline from the 75
756 samples (0.17 per control), 306 candidate variants were determined in our 134 MCD exomes
757 (2.28 per MCD case), which lead to an estimated 7.59% per sample false discovery rate
758 (Supplementary Table 6).

759 **False negative rate estimation for phase 2 relative to phase 1**

760 Only FCD type II cases fulfilling the clinical criteria were considered for this calculation to
761 control the comparison conditions between phase I and phase II. Of the 67 type II patients who
762 underwent Phase 1 AmpliSeq with the 87 mTOR genes, 9 cases were positively validated with
763 mTOR mutations (9/67, 13.43%). Of the 17 novel type IIA/B patients in WES that did not
764 undergo Phase 1 AmpliSeq, 2 were detected with positively validate mTOR mutations on the 87
765 genes (2/17, 11.76%). Thus, we conclude that our false-negative rate for Phase 2 WES relative to
766 the Phase 1 AmpliSeq is 13.43% - 11.76% = 1.67%.

767 **Estimation of probability of observing recurrency in mutated genes in MCD**

768 Based on the previously established estimation^{63,64}, we simulated the number of detected sSNVs
769 and their recurrence. In phases 2 and 3 of our genetic discovery, we positively validated 67 and
770 26 sSNVs from 134 and 126 brains, respectively. Assuming the same mosaic mutation rate and
771 cohort size, we would be expecting the same number (67 + 26) of positively detected sSNVs. We
772 permuted 10,000 times for 67 deleterious sSNVs from the 19909 human coding genes, plus 26
773 deleterious sSNVs from the 59 genes in phase 3, and estimated the distribution of the same gene
774 being hit more than once. After correcting for gene length (average length of human coding
775 genes: 66645.9, the average length of the 59 genes in phase 3: 158929.7, correlated relative to
776 the average length⁶⁵), from the permutation analysis, the probability of observing more than hits
777 is $p = 0.000127$ for *ATP2A1* and $p = 0.000258$ for *PPFIA4*. The probability of observing more
778 than one hit on each of the 59 genes is provided, respectively in Supplementary Table 7.

779 **Orthogonal validation and quantification of mosaic mutations with targeted amplicon
780 sequencing**

781 Targeted amplicon sequencing (TASeq) with Illumina TruSeq was performed with a coverage
782 goal of >1000X for 554 candidate variants detected by computational pipelines described above
783 for both AmpliSeq and WES, to experimentally validate the mosaic candidates before functional
784 assessment. PCR products for sequencing were designed with a target length of 160-190 bp with
785 primers being at least 60 bp away from the base of interest. Primers were designed using the
786 command-line tool of Primer3^{66,67} with a Python (v3.7.3) wrapper^{17,18} (Supplementary Table 8).
787 PCR was performed according to standard procedures using GoTaq Colorless Master Mix
788 (Promega, M7832) on sperm, blood, and an unrelated control. Amplicons were enzymatically

789 cleaned with ExoI (NEB, M0293S) and SAP (NEB, M0371S) treatment. Following
790 normalization with the Qubit HS Kit (ThermFisher Scientific, Q33231), amplification products
791 were processed according to the manufacturer's protocol with AMPure XP beads (Beckman
792 Coulter, A63882) at a ratio of 1.2x. Library preparation was performed according to the
793 manufacturer's protocol using a Kapa Hyper Prep Kit (Kapa Biosystems, KK8501) and barcoded
794 independently with unique dual indexes (IDT for Illumina, 20022370). The libraries were
795 sequenced on Illumina HiSeq 4000 or NovaSeq 6000 platform with 100 bp paired-end reads.
796 Reads from TASEq were aligned to GRCh37 with BWA (version 3.7.16a), sorted, realigned, and
797 recalibrated with SAMtools (version 1.9), Picard (version 2.18.27), and GATK v3.8.1. Candidate
798 variants were annotated with the same ANNOVAR and BEDtools (version 2.27.1) scripts also
799 provided on GitHub (https://github.com/shishenyxx/MCD_mosaic) and exact binomial
800 confidence intervals were calculated for the same variants in the target sample as well as normal
801 controls. Variants detected from AmpliSeq and/or WES are considered to be positively validated
802 in a given tissue by TASEq if 1) the 95% lower binomial confidence interval is higher than
803 0.5% ; 2) the 95% higher binomial confidence interval is lower than 40%; 3) the 95% lower
804 binomial confidence interval of the negative control is below 0.5%.

805

806 **Oncoplot generation**

807 Oncoplot in Fig. 1e was generated using maftools (v2.6.05) R library.

808

809 **Mutational signature analysis**

810 Mutational signature analysis was performed using a web-based somatic mutation analysis
811 toolkit (Mutalisk)⁶⁸. PCAWG SigProfiler full screening model was used.

812

813 **STRING analysis**

814 STRING analysis was performed by STRING v11²⁸. A total of 75 MCD genes (69 novel and
815 known genes from our cohort and 6 novel genes from two other recent MCD cohort studies)
816 were loaded as input and MCL clustering was performed. The terms in Gene Ontology (GO),
817 KEGG pathways, and the top 10 terms GO or KEGG pathways were shown in Fig. 2b. If there
818 are less than 10 terms for those terms (such as clusters 3 and 4 in Fig. 2), we included all the
819 terms in GO or KEGG pathways, Local network cluster (STRING), Reactome pathways, and
820 Disease-gene associations (DISEASES) to show the enriched terms. Visualization was
821 performed by Cytoscape v3.9.

822

823 **ClueGO analysis**

824 Visualization of the functionally grouped biological terms was performed by ClueGO v2.5⁶⁹, a
825 Cytoscape plug-in. A total of 75 MCD genes from Fig. 2 were loaded and GO terms in the
826 'Biological Process' category were used for visualization. Terms with a p < 0.01, a minimum
827 count of 3, and an enrichment factor > 1.5, are grouped into clusters based on membership
828 similarities.

829

830 **Animals**

831 Pregnant Crl:CD1(ICR) mice (E14) for mouse modeling were purchased from Charles River
832 Laboratory. All mice used were maintained under standard group housing laboratory conditions

833 with temperatures of 18~23°C, 40~60% humidity, 12 hours of light/dark cycle, and free access to
834 food and water. The age and number of mice used for each experiment are detailed in the figure
835 legends. The sex of the embryos or subject mice used was not tested.

836

837 **DNA constructs**

838 *RRAGA*, *KLHL22*, and *RHOA* ORF regions were amplified from the hORFeome library and
839 inserted into the pCIG2 (pCAG-IRES-GFP) vector. *GRIN2C* ORF region was purchased from
840 DNASU Plasmid Repository at Arizona State University Biodesign Institute. Gibson Assembly
841 Cloning Kit (E5510S, New England Biolabs) was used for joining point mutation-carrying gene
842 fragments (amplified by primers below) and linearized pCIG2 vector (digested by XhoI and
843 XmaI). The mutation was confirmed by Sanger sequencing.

844 The sequence information of primers used to amplify mutation-carrying DNA fragments is in
845 Supplementary Table 8.

846

847 **In utero electroporation**

848 In utero electroporation was performed as described⁷⁰ with endotoxin-free plasmids (0.5–1 µg)
849 plus 0.1% Fast Green (Sigma, catalog no. 7252) injected into a single lateral ventricle in E14.5
850 embryos then electroporated with BTX ECM830 instrument using pulses of 45 V for 50 ms with
851 455-ms intervals were used.

852

853 **Mouse brain sectioning**

854 The brain was fixed by submersion or perfusion with 4% paraformaldehyde (PFA) for E18.5 or
855 P21 mice respectively, cryoprotected in 30% sucrose for 48 hrs, embedded in Tissue-Tek A,
856 sectioned at 20 um (CryoStar NX70, Thermo Fisher Scientific), mounted onto SuperFrost Plus
857 slides, and dried on a 50 °C heating block before staining.

858

859 **Immunofluorescence staining and imaging**

860 A section was rehydrated and washed by 1X PBS for 10 min 3 times, permeabilized in PBST
861 (0.3% Triton X-100 in 1X PBS) for 10 min, and blocked by blocking solution (5% normal BSA
862 in 1X PBS) for 2 hrs in room temperature. Sections were stained with diluted primary antibodies
863 in the blocking solution overnight at 4 °C. The next day, the sections were washed with PBST for
864 5 min three times and stained with secondary antibodies in blocking solution for 2 hrs in RT.
865 Blocking solution was dropped off from the slides and nuclei staining with DAPI solution
866 (0.1ug/ml of DAPI in PBST) was performed for 15 min. The slides were mounted with DAKO
867 fluorescent mount solution (catalog no. S3023). Zeiss 880 Airyscan Confocal is used for imaging
868 according to the manufacturer's instructions.

869

870 **Antibodies**

871 phospho-S6 (1:800 dilution, catalog no. 5364S ;Cell Signaling, AB_10694233), NeuN (1:100,
872 MAB377X; Sigma-Aldrich, AB_2149209), GFP (1:500, catalog no. GFP-1020, Aves Labs,
873 AB_10000240), Alexa Fluor Goat 488 chicken IgY (H+L) (1:1,000 dilution, catalog no. A-
874 11039, AB_2534096), Alexa Fluor 594 donkey anti-rabbit IgG (H+L) (1:1,000, catalog no.
875 R37119, AB_2556547).

876

877 **Genotype-phenotype association**

878 The functional modules to be tested were selected based on the enriched GO terms. A given
 879 known and candidate MCD gene was assigned as a member to one or multiple modules based on
 880 GO terms related to the given gene (results summarized in Supplementary Table 3c).

881 Subsequently, a given patient became a member of one (or multiple) functional module(s) if the
 882 genes detected in that patient were assigned to that (those) functional module(s). To associate
 883 likely oncogenic sSNVs with clinical phenotypes, the cases carrying a (or multiple) sSNV(s)
 884 listed in COSMIC DB were labeled as ‘COSMIC’. All available clinical information on the
 885 patient was collected and harmonized using ILAE terms (summarized in Supplementary Table
 886 4). Pearson correlation coefficients were calculated by cor.test() function in R. The value of
 887 correlation coefficients was displayed as colors in the heatmap of Fig. 4 using r-gplots (v3.1.1)
 888 package. If two groups with binary values were used for calculation, Phi coefficient was used.

889

890 **Single-nucleus RNA sequencing**

891 A fresh-frozen brain tissue (~50 mg) was placed into a glass dounce homogenizer containing 1
 892 ml cold lysis buffer (0.05% (v/v) NP-40, 10 mM Tris (pH 7.4), 3 mM MgCl₂, 10 mM NaCl) and
 893 dounce 10 times with a loose pestle and following 10 times with a tight pestle. The homogenate
 894 was incubated for 10 min in RT. 9 ml of wash buffer (1% BSA in 1X PBS) was added to the
 895 homogenate and filtered by a 30 um cell strainer. The strained homogenate was spun down in
 896 500 g to remove the supernatant. The pellet was resuspended with 5 ml of wash buffer. Straining
 897 and spinning down steps were performed once more, and the pellet was resuspended into 500 ul
 898 of wash buffer. 10 ul of nuclei resuspension was mixed with counting solution (0.02% Tween 20,
 899 0.1ug/ml DAPI, 1% BSA in 1X PBS) and nuclei density was measured by manual nuclei
 900 counting using DAPI signal. The resuspension was diluted by wash buffer to make the desired
 901 concentration (800~1000 nuclei/ul). Maximum 2 samples were pooled together targeting 10000
 902 nuclei per reaction. Gel beads emulsion (GEM) generation, cDNA, and sequencing library
 903 constructions were performed in accordance with instructions in the Chromium Single Cell 3'
 904 Reagent Kits User Guide (v3.1). A library pool was sequenced with 800 million read pairs using
 905 NovaSeq 6000. Age and sex information in specimens used in Fig. 6: 5-year-old (yo) male for
 906 CTRL-8352, 3 yo female for CTRL-8353, 4 month-old female for HME-4688, 3 yo male for
 907 HME-6593 and 1 yo female for TSC-4258. The information for FCD-4512 is in Supplementary
 908 Table 4.

909

910 **Single-nucleus RNAseq bioinformatics pipeline**

911 Fastq files from single-nucleus libraries were processed through Cell Ranger (v6.0.2) analysis
 912 pipeline with –include-introns option and hg19 reference genome. Pooled library was
 913 demultiplexed and singlets were taken by demuxlet (v1.0). Seurat (v4.0.5) package was used to
 914 handle single nuclei data objects. Nuclei passed a control filter (number of genes > 500, number
 915 of reads >1000, percentage of mitochondrial gene < 10%) was used for downstream analysis.
 916 Protein coding genes were used for further downstream analysis. Data were normalized and
 917 scaled with the most variable 3000 features using the ‘SCTtransform’ functions. Dimensionality
 918 reduction by PCA and UMAP embedding was performed using runPCA and runUMAP
 919 functions. Clustering was performed by FindNeighbors and FindClusters functions. Cell type
 920 identification was performed using known cell type markers expressed in the brain including

921 excitatory (RORB, CUX2, SATB2), inhibitory neuron (*GAD1*, *GAD2*), astrocyte (*SLC1A2*,
922 *SLC1A3*), oligodendrocyte (*MOBP*, *PLP1*), immature oligodendrocyte (*BCAS1*),
923 oligodendrocyte precursor cell (*PDGFRA*), microglia (*PTPRC*), and endothelial cell markers
924 (*CLDN5*, *ID1*) as well as using positive markers found by FindAllMarkers function with 3000
925 most variable features in scaled data. DEG analysis was performed by ‘FindMarkers’ function in
926 Seurat v4.0 with all genes available in the assay. The genes with adjusted p-value < 0.01 were
927 taken and listed in Supplementary Table 5c. The final visualization of various snRNAseq data
928 was performed by ggplot2 (v3.3.5) and matplotlib (v3.5.0).

929

930 **Weighted gene co-expression network analysis**

931 ‘r-wgcna’ package (v1.69) was used for WGCNA according to instructions⁷¹. Briefly, a
932 similarity matrix was generated based on Pearson’s correlation coefficient value among the top
933 3000 variable features in single-nucleus transcriptome data, which was used to calculate the
934 subsequently signed type of network adjacency matrix. Next, the topological overlap matrix
935 (TOM) and the corresponding dissimilarity (1-TOM) value were generated from the adjacency
936 matrix. Finally gene modules were generated by ‘cutreeDynamic’ function with ‘tree’ method,
937 minAbsSplitHeight = 0.9 and minClusterSize = 30 option. Similar gene modules were merged by
938 ‘mergeCloseModules’ function with cutHeight = 0.25.

939

940 **Correlation between cell type and eigengene expression**

941 Pearson correlation coefficient r (>0 , negative values were not presented) between a given cell
942 type and net expression levels of a given geneset and Student asymptotic p-value for the
943 correlation were plotted as the colour and size of dots in dot plots, respectively.

944

945 **RNAscope**

946 We used published methods and purchased target probes (FGFR1 (catalog no. 310071-C2),
947 FGFR2 (311171), FGFR3 (310791), EGFR (310061-C4), PDGFRA(604481-C2)) for genes of
948 interest containing an 18-25 base region complementary to the target, as spacer sequencing, and
949 a 14 base Z-tail sequence⁷², including RNA pol III positive control and random sequence
950 negative control, following the manufacturer recommendations (Advanced Cell Diagnostics,
951 Hayward, CA). Images were acquired on a Leica STED Sp8 with a Falcon microscope.

952

953 **Permutation analysis for the enrichment of MCD genes**

954 To test the enrichment of differentially expressed MCD genes in RNA sequencing against a
955 random distribution, we designed a permutation analysis. All human genes used in the single-cell
956 RNA-seq analysis (n=19909) were randomly shuffled 10,000 times and the same number of
957 genes as described in the differential expression analysis (Extended Data Fig. 10a) was selected
958 for each shuffle. The number of overlaps between each shuffle and the MCD genes was
959 compared and the number of overlaps was used as the outcome and a null distribution was
960 generated from the 10,000 shuffles. All 75 positively validated MCD genes are confirmed to be
961 existing in the initial gene list. After 10,000 permutations, the permutation p-value was
962 calculated with numbers observed to overlap.

963

964 **Statistics and Reproducibility**

965 Statistical analyses were performed by R or Prism 8 (GraphPad Software). Two-way ANOVA
966 and Sidak multiple comparisons were performed in Fig. 3b and Extended Data Fig. 6a with *p*-
967 values of interaction between genotype and bin factor. Mann-Whitney *U*-test was performed for
968 Fig. 3c and Extended Data Fig. 6b. The images used in Fig. 3d and Extended Data Fig. 8 were
969 taken from patient-derived tissue slices which are unique and not biologically reproducible.
970 Detailed statistical information is described in Supplementary Table 9.

971

972 **Data availability**

973 WES and AmpliSeq data are deployed on NIMH Data Archive under study number 1484
974 “Comprehensive multi-omic profiling of somatic mutations in malformations of cortical
975 development” and SRA under accession number PRJNA821916: “Comprehensive multi-omic
976 profiling of somatic mutations in malformations of cortical development”. The BSMN
977 neurotypical brain data are available at NIMH Data Archive (NDA
978 study 644, 792 and 919, <https://nda.nih.gov/study.html?tab=result&id=644>,
979 <https://nda.nih.gov/study.html?tab=result&id=792> and
980 <https://nda.nih.gov/study.html?tab=result&id=919>) and SRA: PRJNA736951. The raw and
981 processed snRNAseq dataset was deposited in Gene Expression Omnibus (GEO) under
982 accession number GSE218022.
983 gnomAD: <https://gnomad.broadinstitute.org/>
984 COSMIC: <https://cancer.sanger.ac.uk/cosmic/>
985 STRING: <https://string-db.org/>
986 Single-cell RNA-seq data from developing cortex (Nowakowski et al., 2017):
987 <https://cells.ucsc.edu/?ds=cortex-dev>
988 GRCh37 genome accession number: PRJNA31257

989

990 **Code availability**

991 Code to generate the figures and analyze the data are publically available on GitHub⁷³
992 (https://github.com/shishenyxx/MCD_mosaic).
993

994 **Method-only references**

- 995 60. Kim, S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* 15, 591-594
996 (2018).
- 997 61. Benjamin, D. et al. Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* (2019).
- 998 62. Yang, X. et al. Genomic mosaicism in paternal sperm and multiple parental tissues in a Dravet syndrome
999 cohort. *Sci Rep* 7, 15677 (2017).
- 1000 63. O'Roak, B.J. et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum
1001 disorders. *Science* 338, 1619-22 (2012).
- 1002 64. Krumm, N. et al. Excess of rare, inherited truncating mutations in autism. *Nat Genet* 47, 582-8 (2015).
- 1003 65. Piovesan, A. et al. Human protein-coding genes and gene feature statistics in 2019. *BMC Res Notes* 12, 315
1004 (2019).
- 1005 66. Untergasser, A. et al. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35, W71-4
1006 (2007).
- 1007 67. Untergasser, A. et al. Primer3--new capabilities and interfaces. *Nucleic Acids Res* 40, e115 (2012).
- 1008 68. Lee, J. et al. Mutualisk: a web-based somatic MUTation AnaLyIS toolKit for genomic, transcriptional and
1009 epigenomic signatures. *Nucleic Acids Res* 46, W102-W108 (2018).
- 1010 69. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway
1011 annotation networks. *Bioinformatics* 25, 1091-3 (2009).
- 1012 70. Koizumi, H., Tanaka, T. & Gleeson, J.G. Doublecortin-like kinase functions with doublecortin to mediate
1013 fiber tract decussation and neuronal migration. *Neuron* 49, 55-66 (2006).
- 1014 71. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC
1015 Bioinformatics* 9, 559 (2008).
- 1016 72. Wang, F. et al. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded
1017 tissues. *J Mol Diagn* 14, 22-9 (2012).
- 1018 73. Yang, X. & Chung, C. shishenyxx/MCD_mosaic: MCD mosaic codes v1.0.0. (Zenodo, 2022).
1019 doi:10.5281/ZENODO.7196403.

1020 **Focal Cortical Dysplasia Neurogenetics Consortium**

1021
1022 Joseph G. Gleeson¹, Marilyn Jones¹, Diane Masser-Frye¹, Shifteh Sattar¹, Mark Nespeca¹, David
1023 D. Gonda¹, Katsumi Imai², Yukitoshi Takahashi², Hsin-Hung Chen³, Jin-Wu Tsai⁴, Valerio
1024 Conti⁵, Renzo Guerrini⁵, Orrin Devinsky⁶, Helio R. Machado⁷, Camila Araújo Bernardino
1025 Garcia⁷, Wilson A. Silva Jr.⁷, Se Hoon Kim⁸, Hoon-Chul Kang⁸, Yasemin Alanay⁹, Seema
1026 Kapoor¹⁰, Carola A. Haas¹¹, Georgia Ramantani¹², Thomas Feuerstein¹², Ingmar Blumcke¹³,
1027 Robyn Busch¹³, Zhong Ying¹³, Vadym Biloshytsky¹⁴, Kostiantyn Kostiuk¹⁴, Eugene
1028 Pedachenko¹⁴, Gary W. Matherne¹⁵, Christina A. Gurnett¹⁶, Matthew D. Smyth¹⁶, Ingo Helbig¹⁷,
1029 Benjamin C. Kennedy¹⁷, Judy Liu¹⁸, Felix Chan¹⁸, Darcy Krueger¹⁹, Richard Frye²⁰, Angus
1030 Wilfong²⁰, David Adelson²⁰, William Gaillard²¹, Chima Oluigbo²¹, Anne Anderson²²

1031
1032 ¹Rady Children's Hospital, San Diego, CA, USA

1033 ²Shizuoka Institute of Epilepsy and Neurological Disorders, Shizuoka, Japan

1034 ³Taipei Veterans General Hospital, Taipei City, Taiwan

1035 ⁴National Yang Ming Chiao Tung University, Taipei City, Taiwan

1036 ⁵A. Meyer Children's Hospital, University of Florence, Firenze, Italy

1037 ⁶New York University Langone Health, New York, NY, USA

1038 ⁷University of São Paulo (USP), Ribeirão Preto, Brazil

1039 ⁸Severance Hospital, Yonsei University College of Medicine, Seoul, South Korea

1040 ⁹Acibadem Hospital, Istanbul, Turkey

1041 ¹⁰Lok Nayak Hospital & Maulana Azad Medical Center, New Delhi, India

1042 ¹¹University of Freiburg, Freiburg, Germany

1043 ¹²Albert-Ludwigs University, Freiburg, Germany

1044 ¹³University Hospital Erlangen, Erlangen, Germany

1045 ¹⁴Romodanov Institute of Neurosurgery, Kyiv, Ukraine

1046 ¹⁵University of California at Los Angeles, CA, USA

1047 ¹⁶St. Louis Children's Hospital, Washington University St Louis, MO, USA

1048 ¹⁷Children's Hospital Philadelphia, PA, USA

1049 ¹⁸Brown University, Providence, RI, USA

1050 ¹⁹Cincinnati Children's Hospital, Cincinnati, OH, USA

1051 ²⁰Barrow Neurological Institute at Phoenix Children's Hospital, University Arizona College of Medicine,
1052 Phoenix, AZ, USA

1053 ²¹Children's National Hospital, Washington DC, USA

1054 ²²Baylor College of Medicine, Texas Children's Hospital, Houston, TX, USA

1055

1056 **Brain Somatic Mosaicism Network**

1057
1058 Alice Lee¹, August Yue Huang¹, Alissa D'Gama¹, Caroline Dias¹, Christopher A. Walsh¹,
1059 Eduardo Maury¹, Javier Ganz¹, Michael Lodato¹, Michael Miller¹, Pengpeng Li¹, Rachel Rodin¹,
1060 Rebeca Borges-Monroy¹, Robert Hill¹, Sara Bizzotto¹, Sattar Khoshkhoo¹, Sonia Kim¹, Zinan
1061 Zhou¹, Alice Lee², Alison Barton², Alon Galor², Chong Chu², Craig Bohrson², Doga Gulhan²,
1062 Eduardo Maury², Elaine Lim², Euncheon Lim², Giorgio Melloni², Isidro Cortes², Jake Lee², Joe
1063 Luquette², Lixing Yang², Maxwell Sherman², Michael Coulter², Minseok Kwon², Peter J. Park²,
1064 Rebeca Borges-Monroy², Semin Lee², Sonia Kim², Soo Lee², Vinary Viswanadham², Yanmei
1065 Dou², Andrew J. Chess³, Attila Jones³, Chaggai Rosenbluh³, Schahram Akbarian³, Ben
1066 Langmead⁴, Jeremy Thorpe⁴, Sean Cho⁴, Andrew Jaffe⁵, Apua Paquola⁵, Daniel Weinberger⁵,
1067 Jennifer Erwin⁵, Jooheon Shin⁵, Michael McConnell⁵, Richard Straub⁵, Rujuta Narurkar⁵,
1068 Alexej Abyzov⁶, Taejeong Bae⁶, Yeongjun Jang⁶, Yifan Wang⁶, Anjene Addington⁷, Geetha
1069 Senthil⁷, Cindy Molitor⁸, Mette Peters⁸, Fred H. Gage⁹, Meiyang Wang⁹, Patrick Reed⁹, Sara
1070 Linker⁹, Alexander Urban¹⁰, Bo Zhou¹⁰, Reenal Pattni¹⁰, Xiaowei Zhu¹⁰, Aitor Serres Amero¹¹,
1071 David Juan¹¹, Inna Povolotskaya¹¹, Irene Lobon¹¹, Manuel Solis Moruno¹¹, Raquel Garcia
1072 Perez¹¹, Tomas Marques-Bonet¹¹, Eduardo Soriano¹², Gary Mathern¹³, Danny Antaki¹⁴, Dan
1073 Averbuj¹⁴, Eric Courchesne¹⁴, Joseph G. Gleeson¹⁴, Laurel L. Ball¹⁴, Martin W. Breuss¹⁴,
1074 Subhojit Roy¹⁴, Xiaoxu Yang¹⁴, Changuk Chung¹⁴, Chen Sun¹⁵, Diane A. Flasch¹⁵, Trenton J.
1075 Frisbie Trenton¹⁵, Huira C. Kopera¹⁵, Jeffrey M. Kidd¹⁵, John B. Moldovan¹⁵, John V. Moran¹⁵,
1076 Kenneth Y. Kwan¹⁵, Ryan E. Mills¹⁵, Sarah B. Emery¹⁵, Weichen Zhou¹⁵, Xuefang Zhao¹⁵,
1077 Aakrosh Ratan¹⁶, Adriana Cherskov¹⁷, Alexandre Jourdon¹⁷, Flora M. Vaccarino¹⁷, Liana
1078 Fasching¹⁷, Nenad Sestan¹⁷, Sirisha Pochareddy¹⁷, Soraya Scuder¹⁷

1079
1080 ¹Boston Children's Hospital, Boston, MA, USA

1081 ²Harvard University, Boston, MA, USA

1082 ³Icahn School of Medicine at Mount Sinai, New York, NY, USA

1083 ⁴Kennedy Krieger Institute, Baltimore, MD, USA

1084 ⁵Lieber Institute for Brain Development, Baltimore, MD, USA

1085 ⁶Mayo Clinic, Rochester, MN, USA

1086 ⁷National Institute of Mental Health (NIMH), Bethesda, MD, USA

1087 ⁸Sage Bionetworks Seattle, WA, USA

1088 ⁹Salk Institute for Biological Studies, La Jolla, CA, USA

1089 ¹⁰Stanford University CA, USA

1090 ¹¹Universitat Pompeu Fabra, Barcelona, Spain

1091 ¹²University of Barcelona, Barcelona, Spain

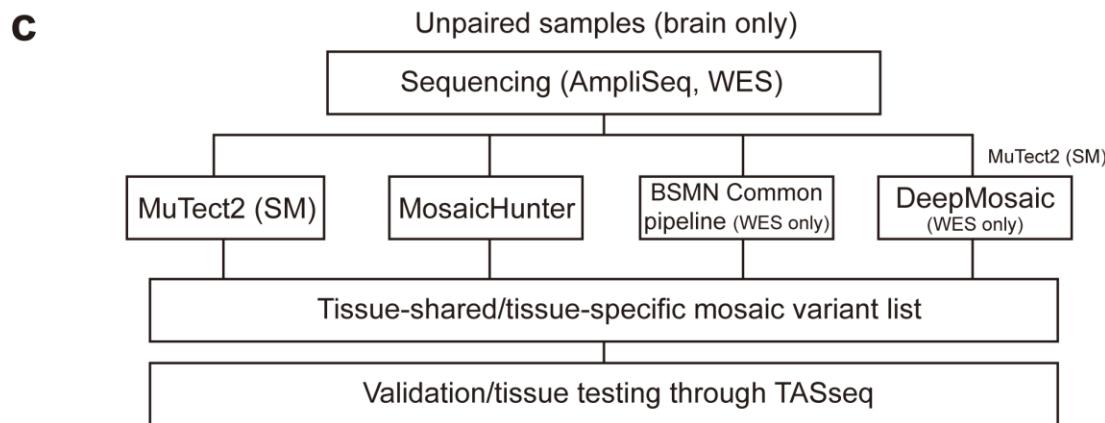
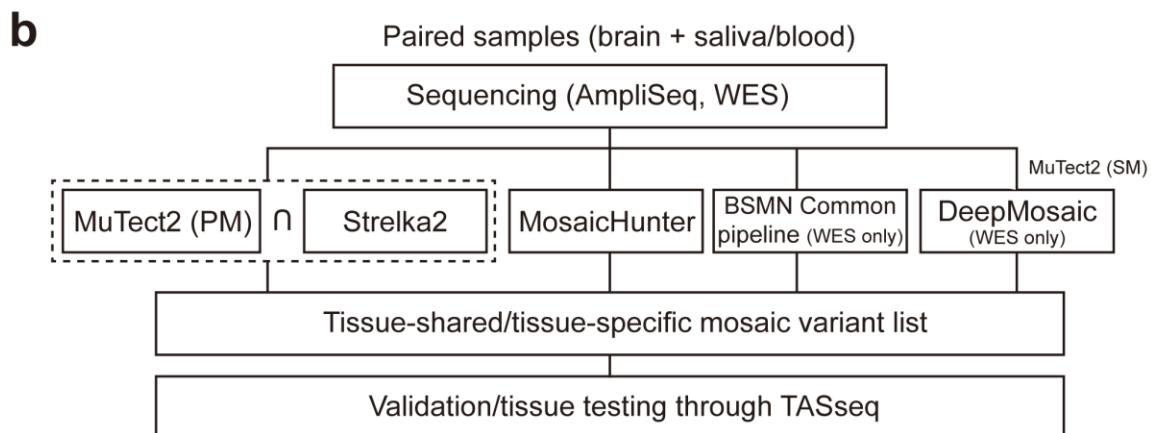
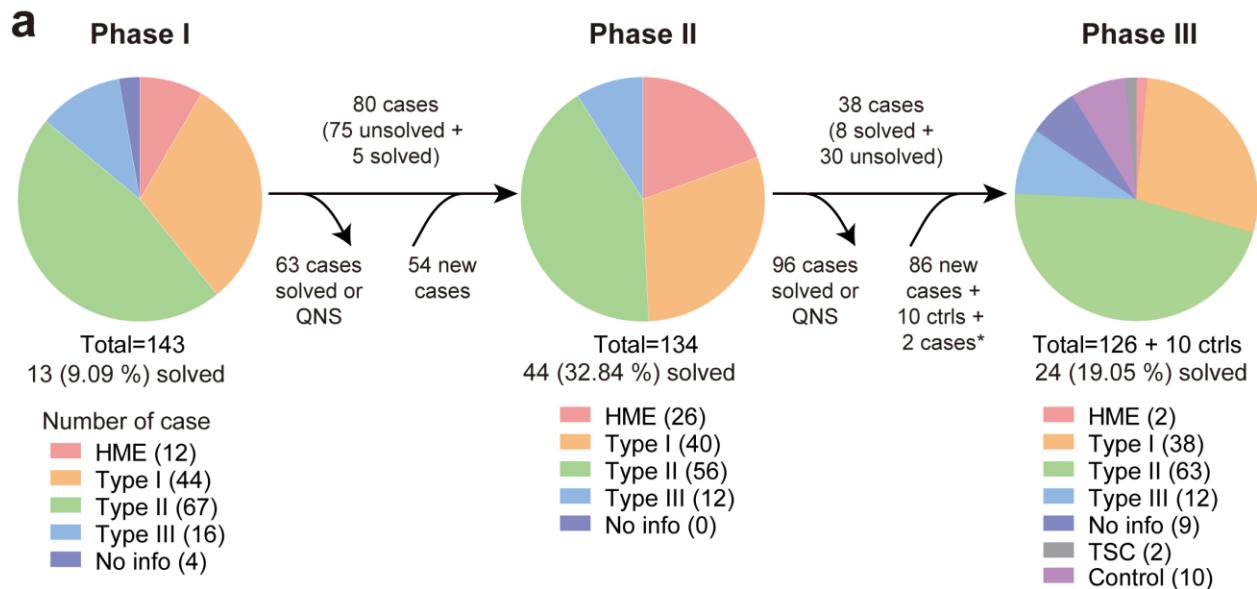
1092 ¹³University of California, Los Angeles, CA, USA

1093 ¹⁴University of California, San Diego, La Jolla, CA, USA

1094 ¹⁵University of Michigan, Ann Arbor, MI, USA

1095 ¹⁶University of Virginia, Charlottesville, VA, USA

1096 ¹⁷Yale University, New Haven, CT, USA

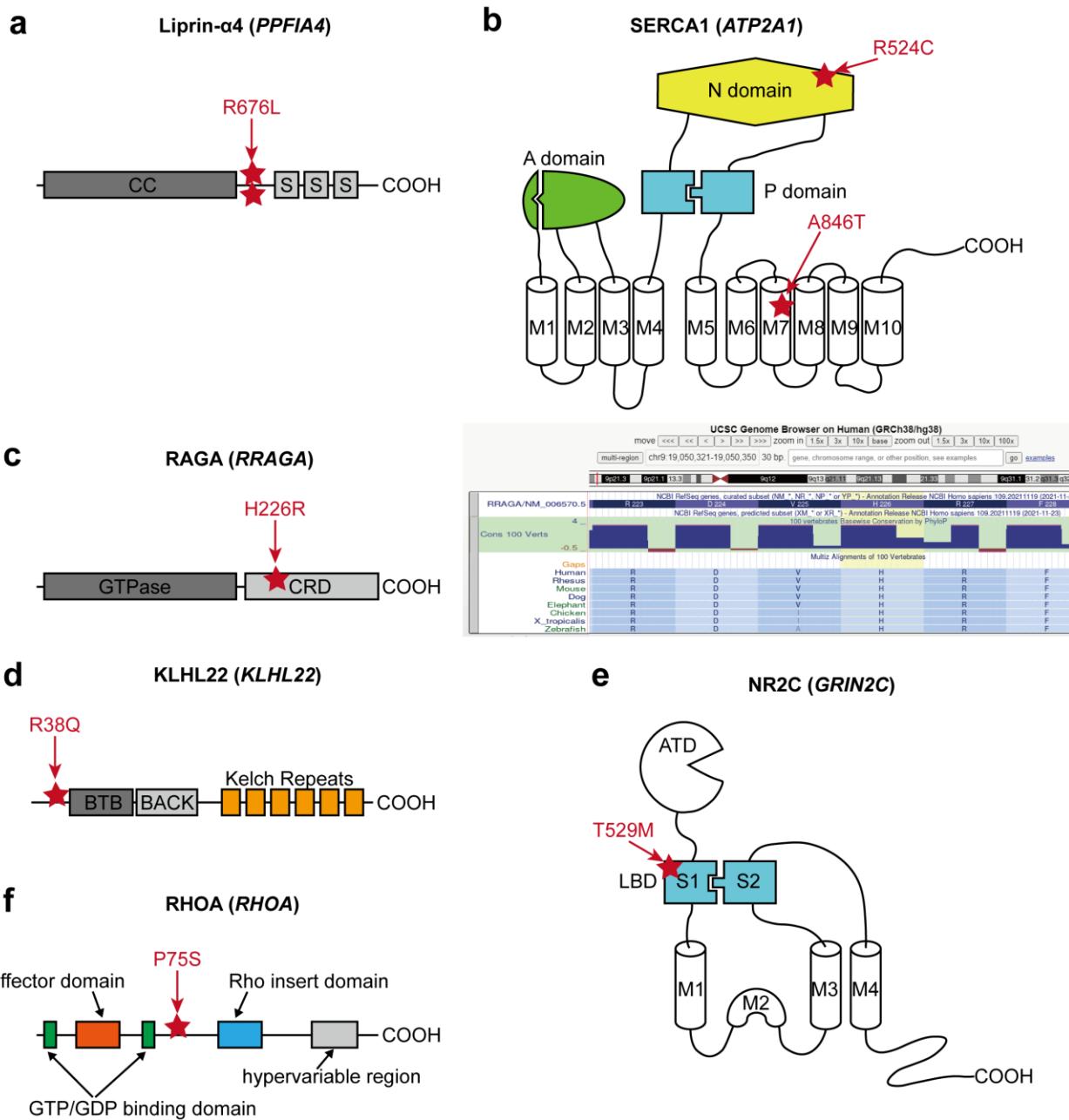


1098

1099 Extended Data Fig. 1 Workflow of genetic discovery and bioinformatic pipeline to detect sSNVs in the
 1100 MCD cohort. (a) Workflow chart describing the flow of cases for each phase of genetic discovery.
 1101 QNS: quantity not sufficient. 2 cases labeled by a star are sequenced in phase 1 but not phase 2. (b) The
 1102 pipeline for paired samples. Notably, the dashed square indicates that the sharing variants between
 1103 MuTect2 paired mode and Strelka2 were used for the downstream analysis. BSMN common pipeline and

1104 DeepMosaic were used only for WES datasets. The DeepMosaic input variants were generated by
1105 MuTect2 single mode. (c) The pipeline for unpaired samples. The pipeline is similar except that MuTect2
1106 single mode without Strelka2 is used. PM: paired mode, SM: single mode.

1107



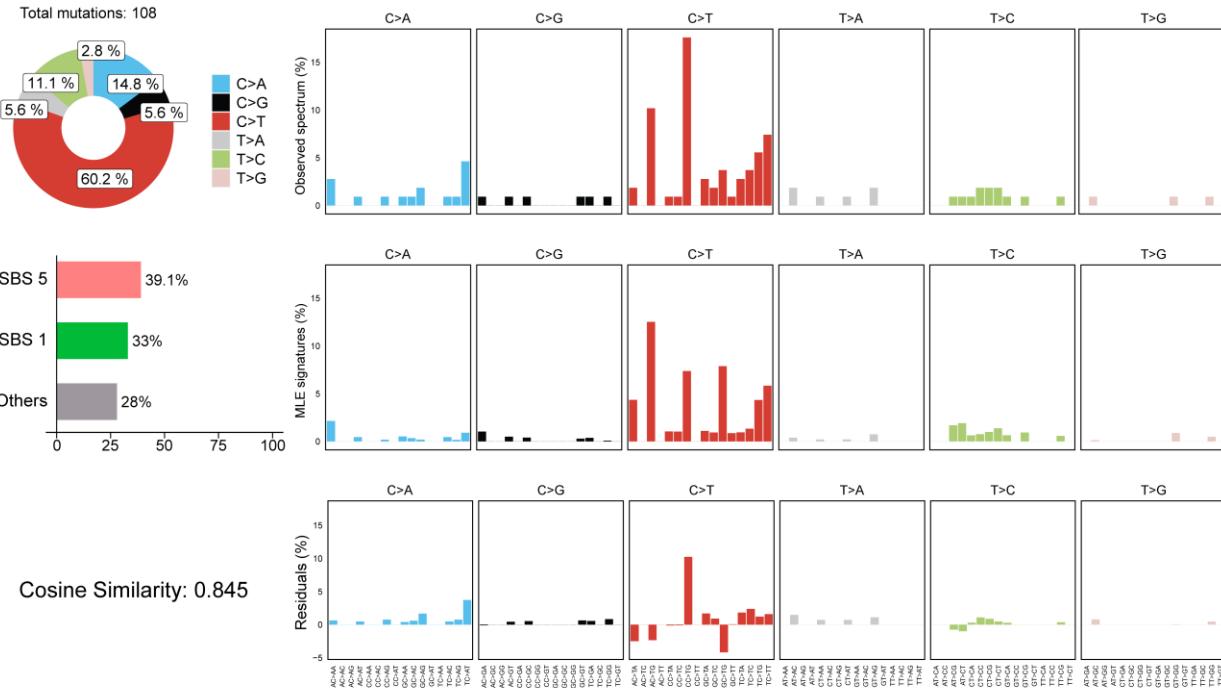
1110 Extended Data Fig. 2 The locations of the selected MCD variants. (a) The location of two recurrent
 1111 SNV calls is at the same position between the coiled-coil domain (CC) and the first SAM domain (S) of
 1112 Liprin- α 4. (b) Two different variants in SERCA1. p.R524C mutation is at the nucleotide ATP-binding (N)
 1113 domain, whereas the p.A846T variant is in the 7th transmembrane (M7) domain. A: Actuator domain, P:
 1114 Phosphorylation domain, M: Transmembrane domain. (c) Left: The location of p.H226R variant in
 1115 RAGA protein. GTPase: GTPase domain, CRD: C-terminal roadblock domain. Right: UCSC genome
 1116 browser screenshot describing that p.H226 is a conserved site across all vertebrates. (d) The location of
 1117 the p.R38Q variant in the N-terminal region before the BTB (Broad-Complex, Tramtrack, and Bric-à-
 1118 brac) domain of KLHL22. (e) A variant in the S1 domain of NR2C. S1 and S2 together make the ligand-

1119 binding domain (LBD), the target of glutamate. ATD: Amino-terminal domain. (f) RHOA p.P75S variant
1120 in the interdomain space between the second GTP/GDP binding domain and Rho insert domain.

1121

1122

1123



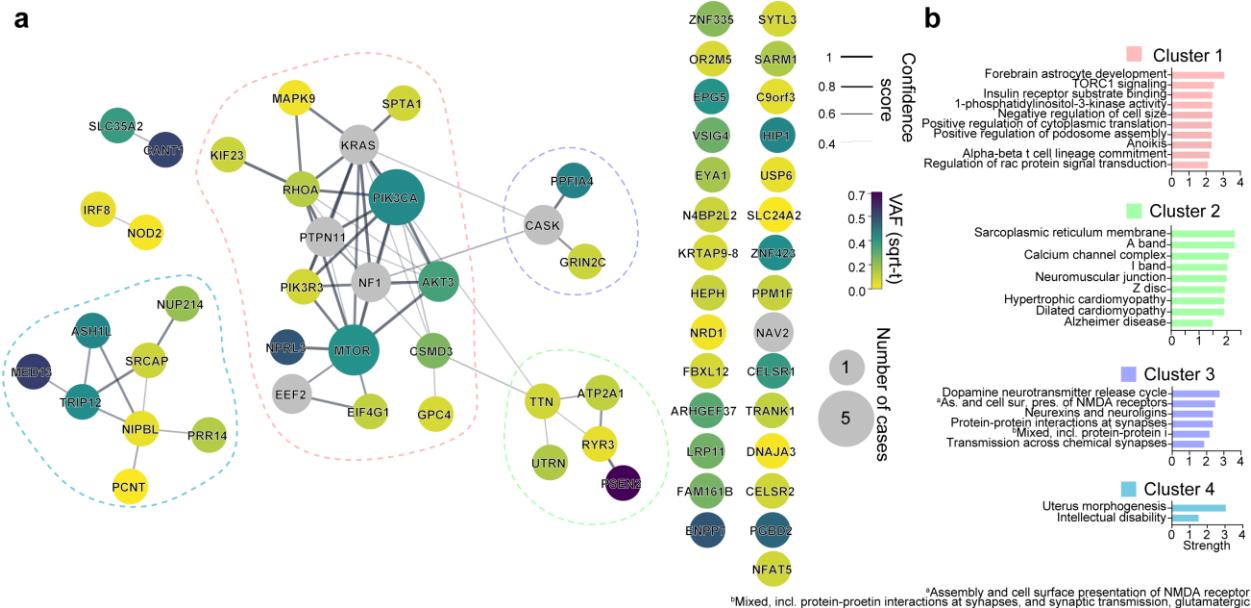
1124

1125 Extended Data Fig. 3 Mutational signature analysis shows cell-division-related clock-like signatures in
1126 the MCD cohort. SBS5 (39.1%) and SBS1 (33%) revealed by Mutalisk are clock-like mutational
1127 signatures. SBS1 especially correlates with stem cell division and mitosis.

1128

1129

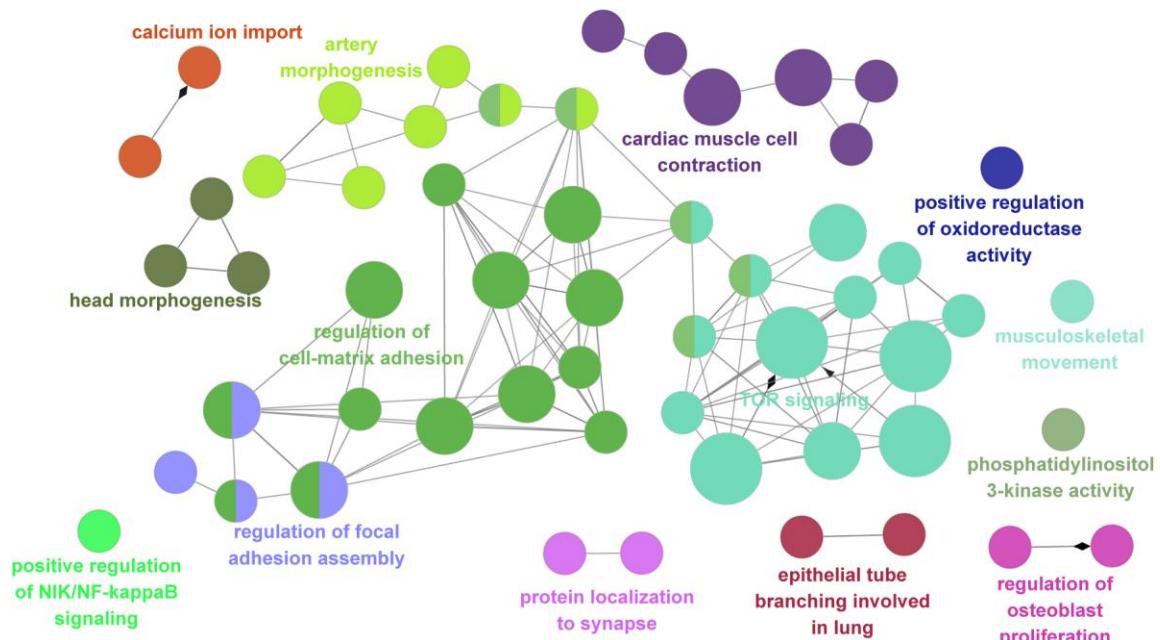
1130



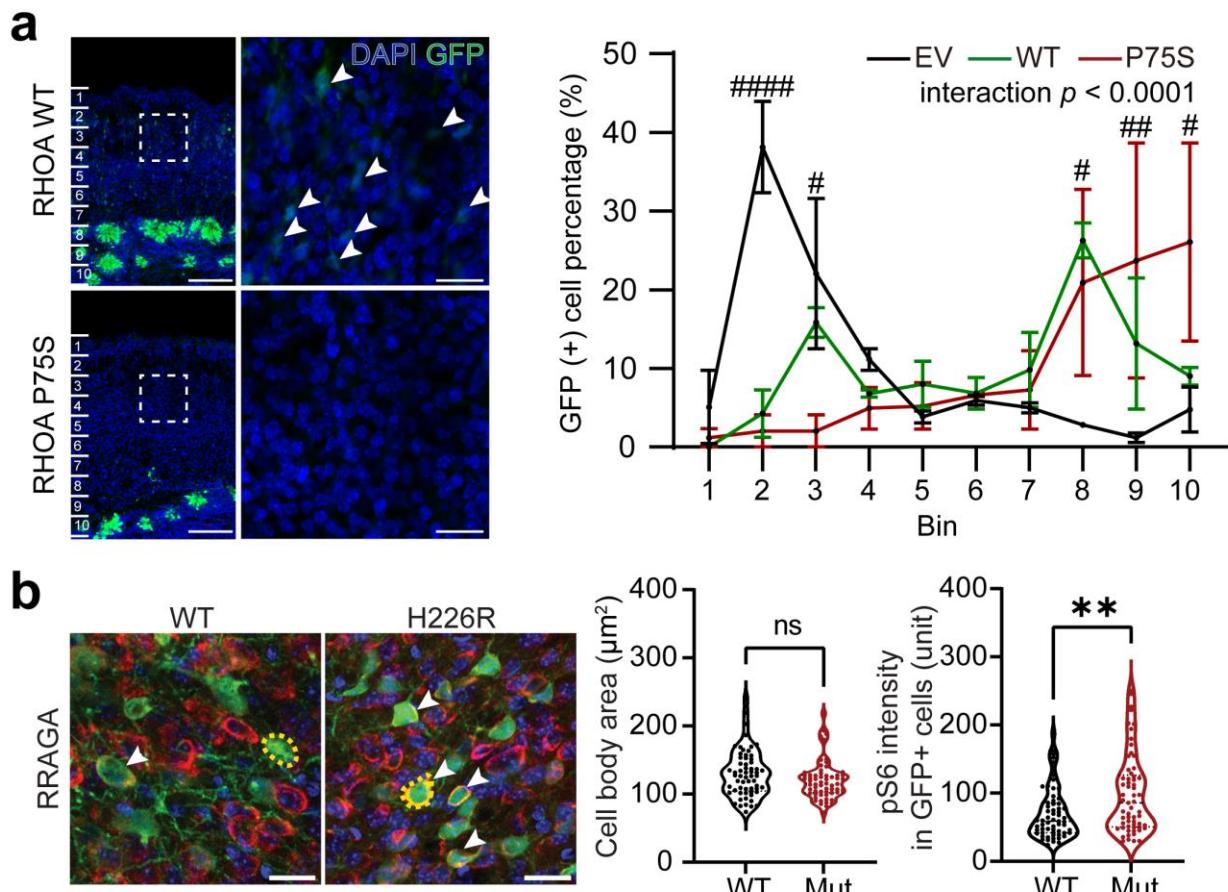
1131

1132 Extended Data Fig. 4 Four major gene networks were reconstructed from the WES dataset. (a)
 1133 STRING DB pathway analysis of the 59 MCD discovered genes and six novel genes (a total of 65 genes)
 1134 from recent publications identifies MTOR/MAP kinase pathway (pink, Cluster 1), Calcium dynamics
 1135 (green, Cluster 2), Synapse (purple, Cluster 3), Gene expression (blue, Cluster 4). Edge thickness:
 1136 confidence score calculated by STRING. Size and color of a node: square root transformed (sqrt-t)
 1137 number of patients carrying a given mutation and average VAF across all patients, respectively. Non-
 1138 clustered orphan genes are listed on the right. (b) Gene Ontology (GO) analysis results confirmed the
 1139 functions of compositions in each network. Top GO terms or KEGG pathways. Strength calculation and
 1140 cluster generation were performed by STRING.

1141



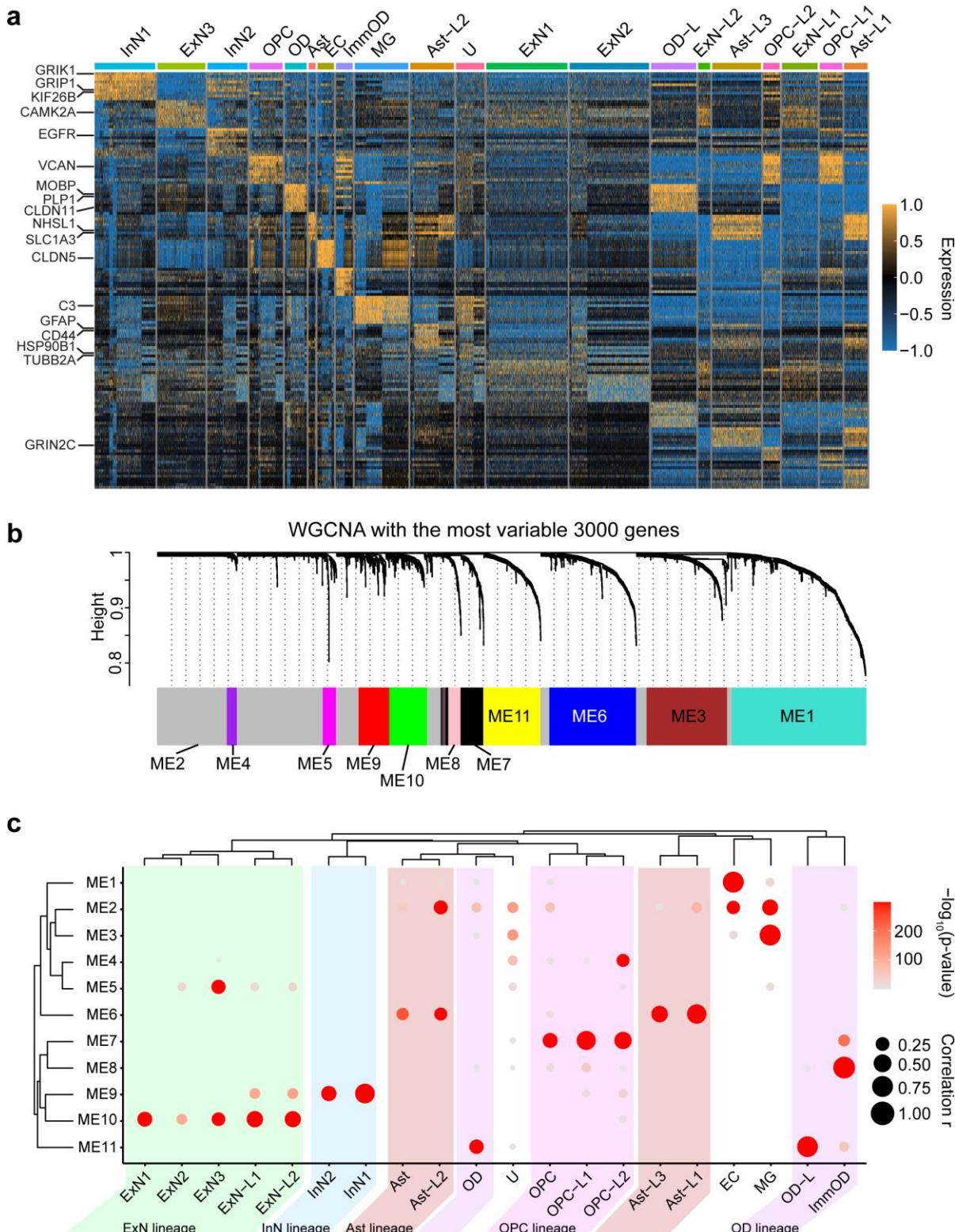
1144 Extended Data Fig. 5 ClueGO analysis using the MCD genes result identifies the biological processes
 1145 and molecular pathways. The main cluster is related to TOR signaling, regulation of cell-matrix
 1146 adhesion, regulation of focal adhesion assembly, and artery morphogenesis. Notably, there are also
 1147 isolated clusters that were not covered in previous studies, for example, cardiac muscle cell contraction,
 1148 calcium ion import, and protein localization to the synapse. Corrected p-value with Bonferroni step down
 1149 was reflected in node size (two-sided hypergeometric test, Large: $p < 0.0005$, medium: $p < 0.005$, small: p
 1150 < 0.05). All p-values are in Supplementary Table 9.



1152

1153 Extended Data Fig. 6 Additional functional analyses for new RRAGA and RHOA mutations. (a)
1154 Over-expression of RHOA WT and P75S mutant form in cortical neurogenic pool induce both significant
1155 defects in migration. Notably, some portion of WT form-expressing cells migrate to the superior cortical
1156 area (white arrow), whereas mutant form-expressing cells did not show any migrating cells at all. The
1157 dashed square area is magnified to the right side images. Scale bar: 100 μ m and 20 μ m for left and right
1158 images, respectively. Right, Quantification of the migration level. EV data was exported from Fig. 3b.
1159 Two-way ANOVA and Sidak multiple comparisons with p-values of interaction between genotype and
1160 bin factor. Ten bins from the surface of the cortex (top) to SVZ (bottom). n=3, 3, 2 biologically
1161 independent mice for EV, RHOA WT and RHOA P75S, respectively. Mean \pm SEM. (b)
1162 Immunofluorescence in postnatal day 21 mouse cortices for RRAGA wild-type (WT) or mutant isoform.
1163 Yellow dashed lines: examples of cell body size quantification. Dashed lines and dotted lines in the violin
1164 plots indicate median and quartiles, respectively. Two-tailed Mann-Whitney U-test. Scale bar: 20 μ m.
1165 n=61 cells (3 mice), 61 (2) for RRAGA WT and RRAGA H226R, respectively * or # indicates a p-value
1166 in comparison between WT and mutant group, or EV and mutant group respectively. #####p < 0.0001;
1167 ##p < 0.01; #p < 0.05; **p < 0.01; ns, non-significant.

1168



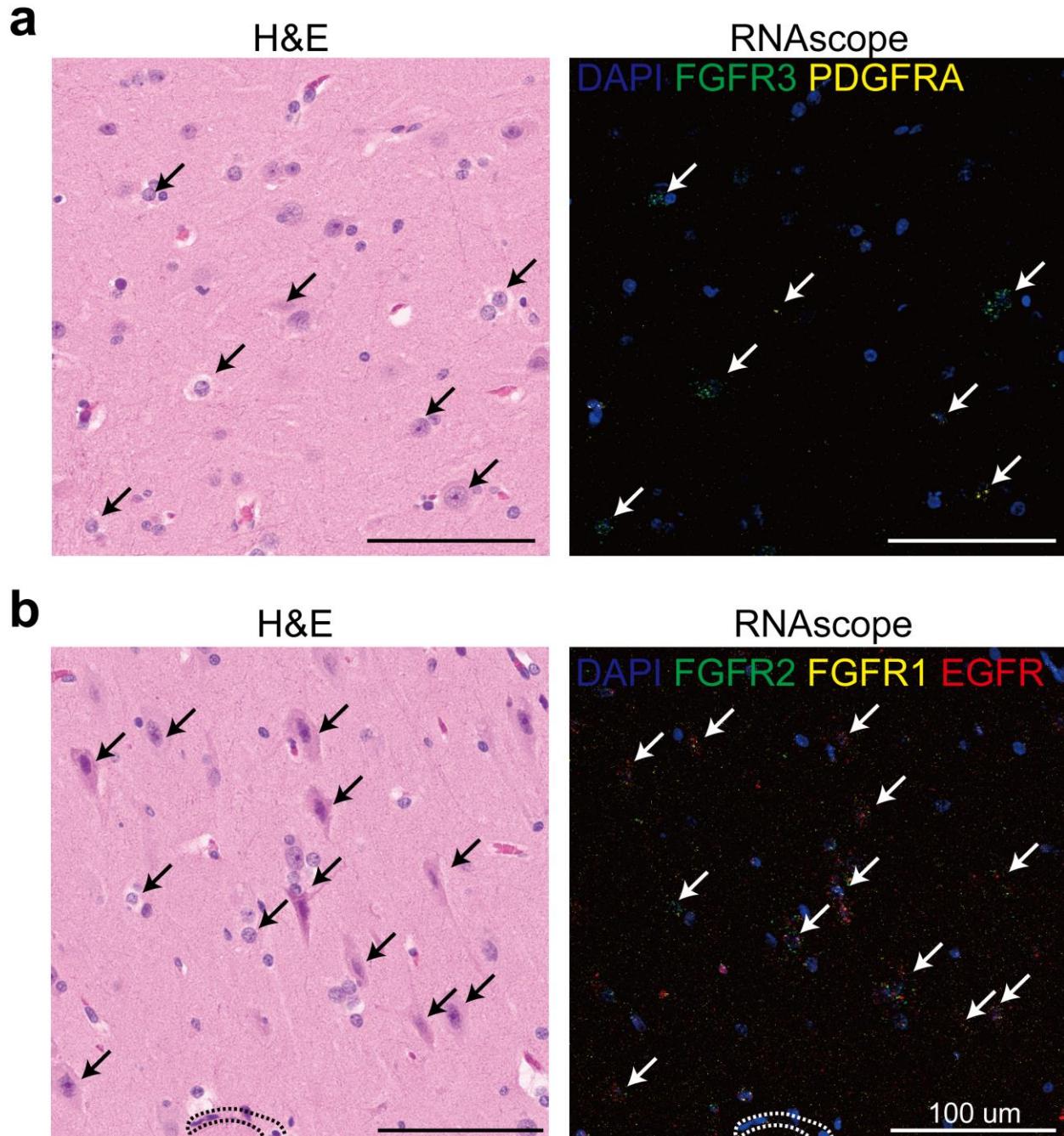
1169

1170 Extended Data Fig. 7 Cell-type identification by DEGs and WGCNA in the MCD snRNAseq dataset.

1171 (a) DEG analysis using FindAllMarker function in Seurat v4 package. The top 10 genes for each
1172 cluster were presented. Some notable marker genes are presented on the left side. Color: scaled gene
1173 expression level. (b) Description of WGCNA. The most variable 3000 genes were used for generating six

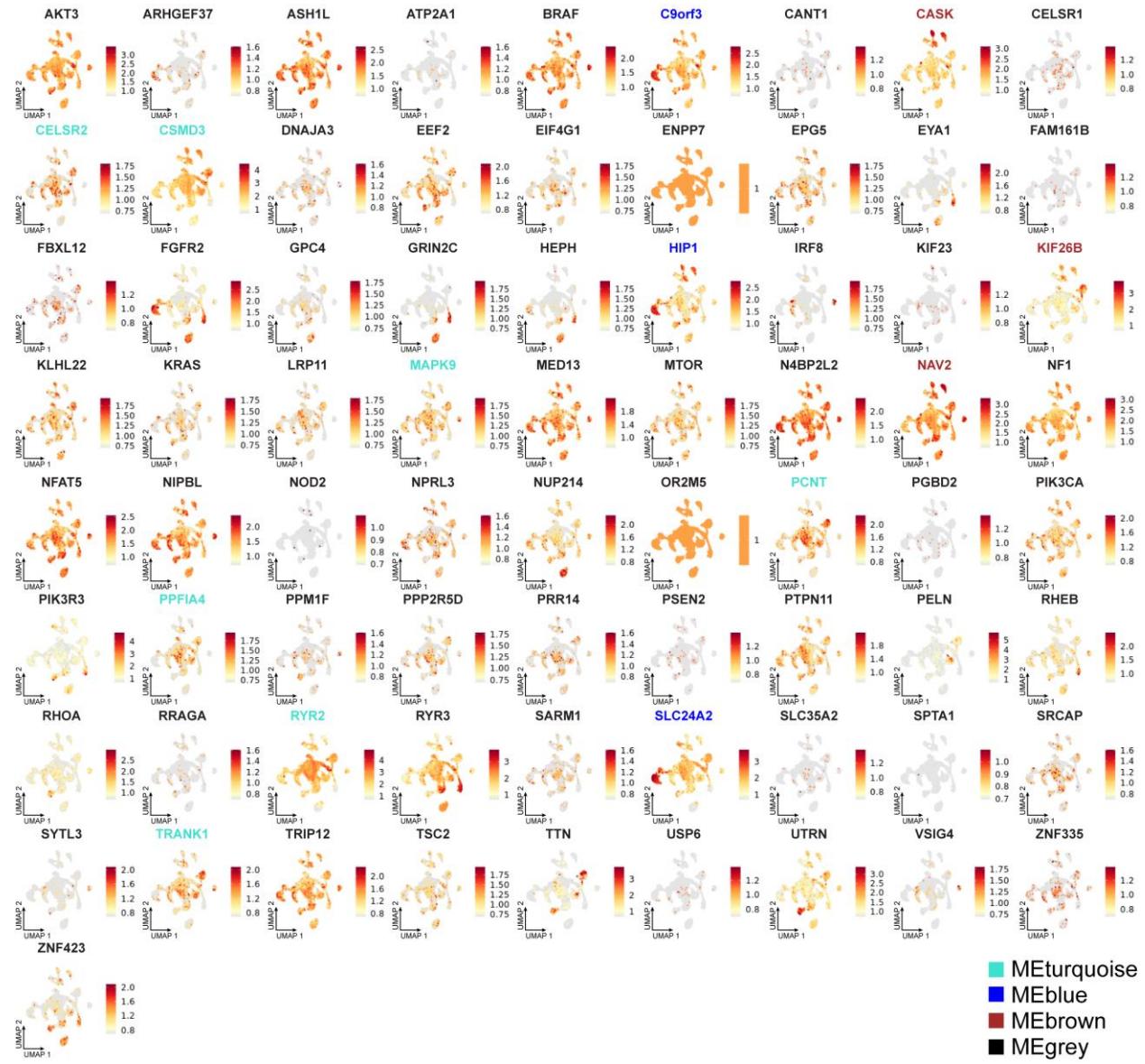
1174 co-expression module eigengenes (ME1 to ME11). The members of each ME are described in
1175 Supplementary Table 5b. (c) Enrichment of module eigengenes in cell type clusters. Atypical clusters
1176 showing similar patterns with a normal cell cluster were classified as the same lineage. We identified 5
1177 different lineages (Ast, OD, ExN, InN, OPC) coded as different colors. Notably, Ast-L1/2/3 and OPC-
1178 L1/2 show excessively increased expression of ME6 or ME7, Ast or OPC signature ME, respectively.
1179 OPC-L2 shows upregulation of ME4, related to the cell cycle, implying that HME has many over-
1180 proliferating OPC-L cells. Excitatory neuronal lineage typically expresses ME5 and ME10, but ExN-L1/2
1181 also shows increased expression of ME9, a signature of inhibitory neurons, compared to ExN1/2/3. OD-L
1182 cells are classified as OD lineage because they express excessive ME11, a signature to OD. U cluster,
1183 dominant in TSC, does not show a clear signature. The size and color of the dot plot are the Pearson
1184 correlation coefficient and corresponding non-adjusted asymptotic p-value derived from a two-sided
1185 Student's t-test, respectively.

1186



1188 Extended Data Fig. 8 The validation of the snRNAseq result from HME-6593 shows that MCD
 1189 dominant clusters are highly correlated with dysplastic cells in MCD. (a-b) H&E (left) and RNAscope
 1190 (right) for genes expressed highly in MCD brain (FGFR2, FGFR1, EGFR, top) or (FGFR3, PDGFRA
 1191 bottom). Dashed lines: blood vessels. White/black arrows: dysplastic cells. One representative section is
 1192 shown for each probe combination.

1193

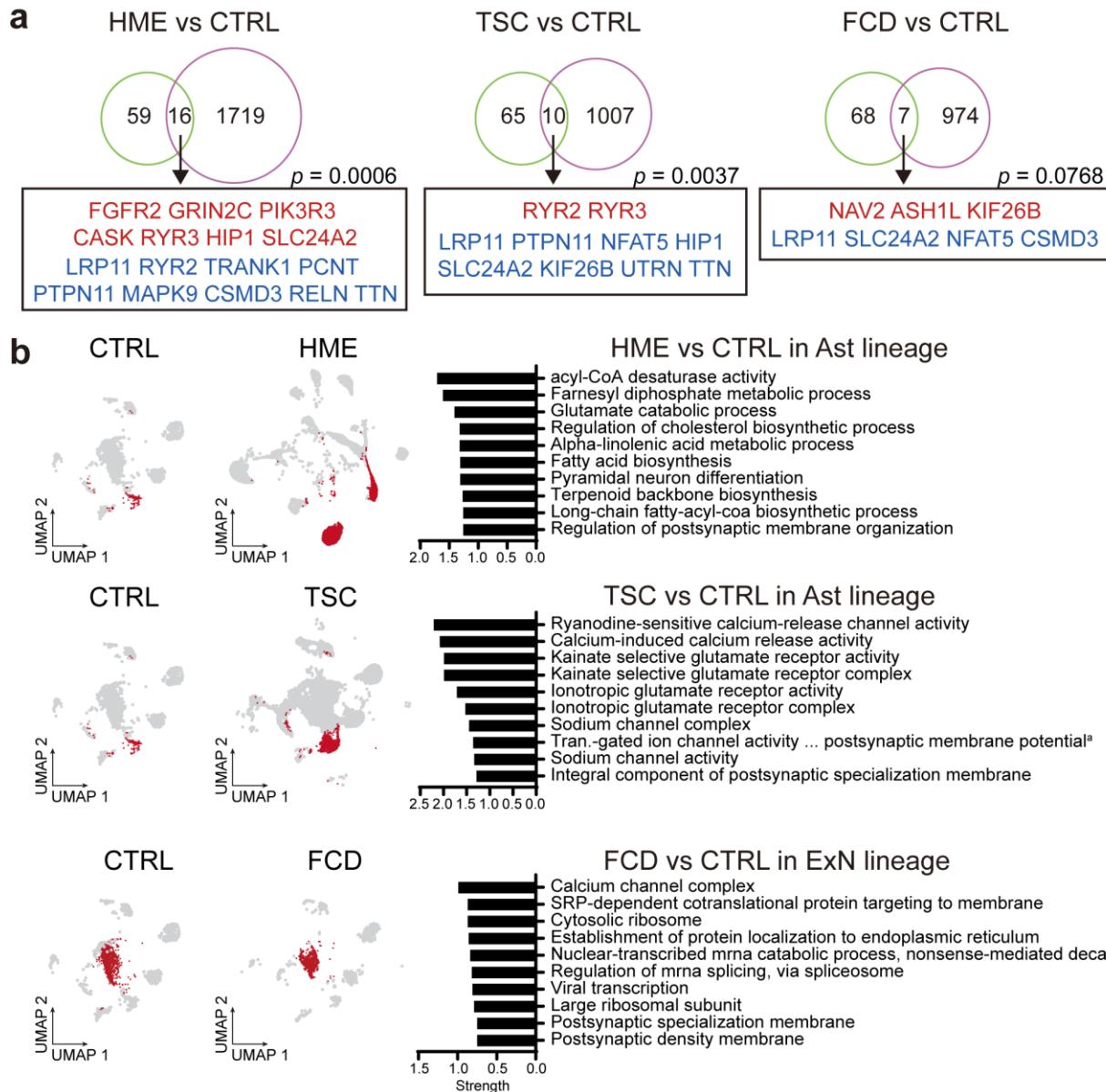


1194

Extended Data Fig. 9 Expression patterns of individual MCD genes in the MCD snRNAseq dataset.

1195 The gene members of each eigen module shown in Fig. 6d were colored according to the name of
1196 a given eigengene.
1197

1198



1199

^aTransmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential

1200 Extended Data Fig. 10 Functional implication of MCD genes in MCD snRNAseq dataset. (a) The
 1201 75 MCD genes overlap with DEGs of MCDs in contrast to controls. p-values derived from permutation
 1202 tests (10,000 times) show a chance to show this overlap in a random sampling of DEGs from 19909
 1203 protein-coding genes used in these DEGs. Red or blue coloring of gene names indicates upregulated or
 1204 downregulated DEGs in MCDs compared to CTRLs, respectively. HME and TSC show significant
 1205 overlap with the MCD genes whereas an FCD case with low VAF did not, which is probably because of
 1206 low VAF. One-sided permutation test. (b) Cell-lineage specific DEGs compared to the according to
 1207 normal cell lineage of CTRL represent alterations of mTOR downstream pathways, calcium dynamics,
 1208 and synaptic functions across. Red dots in UMAPs indicate the cells that participated in the comparison.
 1209 Top 10 enriched GO or KEGG terms representing lineage-specific DEGs.