

# Comprehensive multiomic profiling of somatic mutations in malformations of cortical development

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

<sup>1</sup>Department of Neurosciences, University of California San Diego, La Jolla, CA, 92037, USA

<sup>2</sup>Rady Children's Institute for Genomic Medicine, San Diego, CA, 92123, USA

<sup>3</sup>Department of Quantitative Health Sciences, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, 55905, USA

<sup>4</sup>Department of Neurosurgery, Experimental Epilepsy Research, Medical Center-University of Freiburg, Faculty of Medicine, 79106 Freiburg, Germany

<sup>5</sup>Department of Neurosurgery, University of California at Los Angeles, CA, 90095, USA

<sup>6</sup>Department of Pediatrics, Section of Clinical Genetics and Metabolism, University of Colorado Aurora, CO, 80045, USA

<sup>7</sup>Laboratory of Pediatric Neurosurgery and Developmental Neuropathology, Dept. of Surgery and Anatomy, University of São Paulo (USP), Ribeirão Preto, 14000-000, Brazil

<sup>8</sup>Div Pediatric Neurology, Dept of Pediatrics, Severance Children's Hospital, Yonsei U College of Medicine, Seoul, Korea

<sup>9</sup>Center for Basics in NeuroModulation, Faculty of Medicine, University of Freiburg, 79106 Freiburg, Germany

<sup>10</sup>Department of Neuroradiology, Medical Center-University of Freiburg, Faculty of Medicine, 79106 Freiburg, Germany

<sup>11</sup>Institute of Neuropathology, Medical Center-University of Freiburg, Faculty of Medicine, 79106 Freiburg, Germany

<sup>12</sup>St. Louis Children's Hospital, Washington University St Louis, MO, 63110, USA

<sup>13</sup>Epilepsy Center, Rady Children's Hospital, San Diego, CA, 92123, USA

<sup>14</sup>National Epilepsy Center, Shizuoka Institute of Epilepsy and Neurological Disorders, Shizuoka 420-8688, Japan

<sup>15</sup>Institute of Brain Science, National Yang-Ming University, Beitou 112, Taipei, Taiwan

<sup>16</sup>Paediatric Neurology Unit and Laboratories, A. Meyer Children's Hospital, University of Florence, Italy

<sup>17</sup>Epilepsy Service, Dept. Neurology, New York University, NY, 10016, USA

<sup>18</sup>Department of Genetics, Center for Cell-Based Therapy, Center for Integrative Systems Biology, University of São Paulo (USP), Ribeirão Preto, 14000-000, Brazil

<sup>19</sup>Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, Hôpital de la Pitié Salpêtrière, Paris, France

<sup>#</sup>These authors contributed equally

<sup>\*</sup>Correspondence to: [jogleeson@health.ucsd.edu](mailto:jogleeson@health.ucsd.edu)

48 &Full membership of the FCD Neurogenetic Consortium is listed in the Supplement  
49 \*Full membership of the Brain Somatic Mosaicism Network is listed in the Supplement  
50 Number of figures: 6, Number of extended data items: 7

51  
52 Emails: [chchung@health.ucsd.edu](mailto:chchung@health.ucsd.edu), [xiy010@health.ucsd.edu](mailto:xiy010@health.ucsd.edu), [bae.taejeong@mayo.edu](mailto:bae.taejeong@mayo.edu),  
53 [kivong@health.ucsd.edu](mailto:kivong@health.ucsd.edu), [swmittal@ucsd.edu](mailto:swmittal@ucsd.edu), [catharina.donkels@uniklinik-freiburg.de](mailto:catharina.donkels@uniklinik-freiburg.de),  
54 [carola.haas@uniklinik-freiburg.de](mailto:carola.haas@uniklinik-freiburg.de), [hphillips@mednet.ucla.edu](mailto:hphillips@mednet.ucla.edu), [amarsh@health.ucsd.edu](mailto:amarsh@health.ucsd.edu),  
55 [martin.breuss@cuanschutz.edu](mailto:martin.breuss@cuanschutz.edu), [llball@health.ucsd.edu](mailto:llball@health.ucsd.edu), [camila.neurociencias@gmail.com](mailto:camila.neurociencias@gmail.com),  
56 [reneegeorge@gmail.com](mailto:reneegeorge@gmail.com), [j3gu@ucsd.edu](mailto:j3gu@ucsd.edu), [mxu.china@gmail.com](mailto:mxu.china@gmail.com), [cbarrows@ucsd.edu](mailto:cbarrows@ucsd.edu),  
57 [kiely.n.james@gmail.com](mailto:kiely.n.james@gmail.com), [vstanley146@gmail.com](mailto:vstanley146@gmail.com), [anidhiry@gmail.com](mailto:anidhiry@gmail.com),  
58 [samikhouri619@gmail.com](mailto:samikhouri619@gmail.com), [ghowe@ucsd.edu](mailto:ghowe@ucsd.edu), [eariley@ucsd.edu](mailto:eariley@ucsd.edu), [virginiaxuxin@gmail.com](mailto:virginiaxuxin@gmail.com),  
59 [brcopeland@gmail.com](mailto:brcopeland@gmail.com), [Yifan.Wang@mayo.edu](mailto:Yifan.Wang@mayo.edu), [paxco@yuhs.ac](mailto:paxco@yuhs.ac), [hipo0207@yuhs.ac](mailto:hipo0207@yuhs.ac),  
60 [andreas.schulze-bonhage@uniklinik-freiburg.de](mailto:andreas.schulze-bonhage@uniklinik-freiburg.de), [carola.haas@uniklinik-freiburg.de](mailto:carola.haas@uniklinik-freiburg.de),  
61 [horst.urbach@uniklinik-freiburg.de](mailto:horst.urbach@uniklinik-freiburg.de), [marco.prinz@uniklinik-freiburg.de](mailto:marco.prinz@uniklinik-freiburg.de), [gardnerc@wustl.edu](mailto:gardnerc@wustl.edu),  
62 [gurnettc@wustl.edu](mailto:gurnettc@wustl.edu), [ssattar@ucsd.edu](mailto:ssattar@ucsd.edu), [mnespeca@ucsd.edu](mailto:mnespeca@ucsd.edu), [dgonda@rchsd.org](mailto:dgonda@rchsd.org), [takahashi-ped@umin.ac.jp](mailto:takahashi-ped@umin.ac.jp), [roberthhchen3@gmail.com](mailto:roberthhchen3@gmail.com), [jinwu.tsai@gmail.com](mailto:jinwu.tsai@gmail.com), [valerio.conti@meyer.it](mailto:valerio.conti@meyer.it),  
63 [renzo.guerrini@meyer.it](mailto:renzo.guerrini@meyer.it), [wilsonjr@usp.br](mailto:wilsonjr@usp.br), [od4@nyu.edu](mailto:od4@nyu.edu), [hrmachad@fmrp.usp.br](mailto:hrmachad@fmrp.usp.br),  
64 [gmathern@ucla.edu](mailto:gmathern@ucla.edu), [stephanie.baulac@icm-institute.org](mailto:stephanie.baulac@icm-institute.org), [sara.baldassari@icm.institute.org](mailto:sara.baldassari@icm.institute.org),  
65 [jogleeson@health.ucsd.edu](mailto:jogleeson@health.ucsd.edu)  
66

67  
68 **Abstract: 129 words, Manuscript: 3101 words, Figures: 6, Extended Data items: 7, Suppl**  
69 **Tables: 3**  
70 **Keywords: epilepsy, focal cortical dysplasia, brain mosaicism, mTOR, single-cell**  
71 **sequencing, whole-exome sequencing**

### 72 **Abstract**

73 Malformations of cortical development (MCD) are neurological conditions displaying focal  
74 disruption of cortical architecture and cellular organization arising during embryogenesis, largely  
75 from somatic mosaic mutations. Identifying the genetic causes of MCD has been a challenge, as  
76 mutations remain at low allelic fractions in brain tissue resected to treat epilepsy. Here, we report  
77 a genetic atlas from 317 brain resections, identifying 69 mutated genes through intensive  
78 profiling of somatic mutations, combining whole-exome and targeted-amplicon sequencing with  
79 functional validation and single-cell sequencing. Genotype-phenotype correlation analysis  
80 elucidated specific MCD gene sets associating distinct pathophysiological and clinical  
81 phenotypes. The unique spatiotemporal expression patterns identified by comparing single-  
82 nucleus transcriptional sequences of mutated genes in control and patient brains implicate critical  
83 roles in excitatory neurogenic pools during brain development, and in promoting neuronal  
84 hyperexcitability after birth.

85

### 86 **Introduction**

87 MCDs are heterogeneous groups of neurodevelopmental disorders with localized malformation  
88 of cortical structures, often presenting with intractable epilepsy<sup>1</sup>. Major MCD subtypes include  
89 different classes of focal cortical dysplasia (FCD), hemimegalencephaly (HME), and tuberous  
90 sclerosis complex (TSC)<sup>2</sup>. The International League Against Epilepsy (ILAE) has classified FCD

92 subtypes based on neuropathological features and cell types<sup>3</sup>. MCD patients often undergo  
93 surgical resection of the lesion to treat drug-refractory epilepsy, which has led to remarkable  
94 clinical benefits in published series<sup>4</sup>. The abnormal histology of resected regions includes loss of  
95 lamination of cortical layers, enlarged dysplastic neurons, or balloon cells, sometimes  
96 accompanied by other brain abnormalities. But similar to brain tumors, it can be difficult to  
97 predict pathology prior to surgery.

98 Again, like with brain tumors, genetic studies may offer insights into mechanisms.

99 Somatic mTOR pathway gene mutations are frequently detected in HME and type II FCD foci<sup>5,6</sup>.  
100 Recently, small- or medium-size cohort studies (<100 cases) have confirmed these results and  
101 have correlated defects in neuronal migration, cell size, and neurophysiology<sup>7</sup>. Still, the vast  
102 majority of MCD cases still remain genetically unsolved, suggesting other genes or modules  
103 contribute to MCD.

104 Detecting mutant alleles in bulk resected foci from MCD patients is challenging because  
105 unlike in brain tumors, the mutant cells in MCD are probably not hyperproliferative, and thus  
106 variant allelic fraction (VAF) are often <5%, diluted by genomes of surrounding non-mutated  
107 cells<sup>8</sup>. Fortunately, new computational algorithms have helped reduce false-positive and false-  
108 negative signals, even when no ‘normal’ paired sample is available for comparison<sup>9-11</sup>. The NIH-  
109 supported Brain Somatic Mosaicism Network established the ‘BSMN common pipeline’,  
110 incorporating a ‘best practice’ workflow to reliably and reproducibly identify somatic variants  
111 contributed by members of the Network<sup>12</sup>. With these advances, we thus assessed the possibility  
112 of gene networks beyond mTOR that could underlie MCDs. This new gene discovery may give  
113 insights into novel druggable pathways in cases of incomplete resection due to regional  
114 importance or drug-resistant forms of MCD.

115

## 116 **Results**

### 117 **The genetic landscape of MCD from targeted and unbiased sequencing**

118 To perform a thorough genetic screening of somatic mutations in resected epileptic tissue, we  
119 formed the FCD Neurogenetics Consortium and enrolled 327 samples that met clinical and  
120 pathological criteria for FCD or HME. We excluded TSC from our enrollment criteria because  
121 genes are already well known. Our cohort included 31 HME cases, 98 type I-, 142 type II-, 32  
122 type III-, and 12 unclassified-FCD cases. We included acute resected brains from 10  
123 neurotypicals and 2 TSC cases for comparison (Fig. 1a, supplementary table 1). Patients with  
124 environmental causes, syndromic presentations, inherited mutations, multifocal lesions, or  
125 tumors were excluded (Methods).

126 We used a three-phase genetic screening, each followed by filtering for likely causative  
127 mutations using published methods<sup>13,14</sup>, and each followed by orthogonal targeted amplicon  
128 sequencing (TASeq) intra-case validation and VAF quantification compared with controls  
129 (~5000 X, TASeq)(Fig. 1b). In Phase 1, we performed amplicon sequencing (AmpliSeq, ~1000  
130 X) profiling the entire open reading frame of 87 genes previously detected in FCD/HMEs or  
131 known PI3K-AKT3-mTOR interactors (‘MCD panel v1’, Supplementary Table 2a). In Phase 2,  
132 for 75 unsolved cases from Phase 1 and additionally collected 54 cases, we performed unbiased  
133 deep whole-exome sequencing (WES, ~300 X) on paired samples, where available, or on  
134 unpaired samples (i.e. brain plus blood/saliva vs. brain only). In Phase 3, from an additional  
135 subcohort of 132 new cases, we designed the ‘MCD panel v2’ (Supplementary Table 2b)  
136 including known and novel genes detected in Phases 1 and 2 (Extended Data Fig. 1, Methods).

137 We re-sequenced unsolved cases from Phase 2, expecting that the higher read depth afforded by  
138 panel sequencing could provide greater sensitivity to detect low VAF mutations, and used  
139 BSMN best practice guidelines for mapping and variant calling<sup>12</sup>.

140 From Phases 1 to 3, 1181 candidate somatic SNVs were identified. Of these, 628 were  
141 excluded based on gnomAD allele frequencies, dinucleotide repeats, homopolymers, and  
142 additional BSMN established criteria (Methods)<sup>15,16</sup>. This yielded 554 candidate somatic SNV  
143 that were further assessed by TASEq, yielding 108 validated somatic SNV calls (19.4%  
144 validation rate, Fig. 1c, Supplementary Table 3), compared to other BSMN effort validation rates  
145 in WGS<sup>12,17</sup>. In detail, 15, 67, and 26 validated somatic SNV calls were derived from phase 1, 2,  
146 and phase 3, respectively. The measured VAFs between the AmpliSeq/WES and TASEq were  
147 correlated as expected ( $R^2 = 0.7243$ ) (Fig. 1d). Of the 69 candidate MCD genes mutated in 76  
148 patients, 8 were recurrently mutated, including known mTOR pathway genes as well as several  
149 novel candidates (Fig. 1e).

150 We estimate only ~7% of mutations identified are likely attributable to false discovery  
151 during variant calling, based upon background mutation rate in 75 BSMN neurotypical brain  
152 samples, and published experience from the BSMN<sup>12,18</sup>, processed with the same workflow (see  
153 Methods). Thus, 93% of our candidate MCD mutations would not have been identified in a size-  
154 matched neurotypical control cohort.

155 Most patients (80.52%, 62 cases) showed a single somatic mutation, but some showed two  
156 somatic mutations (14.29%, 11 cases), and some showed more than two mutations (5.19%, 4  
157 cases). Interestingly, HME-4144 showed 11 different somatic mutations, all of which were  
158 validated with TASEq. Although there are several possible explanations for HME-4144, we  
159 expect this reflects clonal expansion from a driver mutation, with detection of multiple passenger  
160 mutations, as reported in brain tumors<sup>19</sup>.

161 Single-base mutational signatures (SBS) were developed to describe potential mutational  
162 mechanisms in human disease<sup>20</sup>. We found 60.2% of mutations were C>T, likely arising from  
163 DNA epigenetic marks<sup>21</sup> (Extended Data Fig. 2). Enrichment of SBS1 and SBS5, clock-like  
164 mutational signatures suggest endogenous mutations arising during corticogenesis DNA  
165 replication.

166

## 167 **Functional dissection of the MCD genes**

168 Interestingly, most validated genes were non-recurrently mutated (88.4%, 61 of 69) in our  
169 cohort, suggesting substantial genetic heterogeneity in MCD. This nevertheless provided an  
170 opportunity to study converging functional gene networks. Thus, we performed Markov  
171 clustering with a STRING network generated from the putative MCD genes<sup>22</sup>, as well as recently  
172 reported novel MCD candidates (*NAV2*, *EEF2*, *CASK*, *NF1*, *KRAS*, *PTPN11*)<sup>23,24</sup> (Fig. 2a). We  
173 identified four clusters, with cluster 1 (“mTOR pathway”) showing the highest term enrichment  
174 to the mTOR/MAP kinase signaling, supporting prior results for Type II MCDs. Cluster 1 also  
175 highlighted newly identified genes *FGFR2*, *KLHL22*, *RRAGA*, *PPP2R5D*, *PIK3R3*, *EEF2*,  
176 *EIF4G1*, and *MAPK9*. Cluster 2 identified “Calcium Dynamics” and included genes *ATP2A1*,  
177 *RYR2*, *RYR3*, *PSEN2*, *TTN*, *UTRN*. Cluster 3 was labeled “Synaptic Functions” and included  
178 genes *CASK*, *GRIN2C*, and *PPFIA4*. Cluster 4 was labeled “Gene Expression” and included  
179 intellectual disability genes, mostly involved in nuclear function, including *NUP214*, *PRR14*,  
180 *PCNT*, *NIPBL*, *SRCAP*, *ASH1L*, *TRIP12*, and *MED13* (Fig. 2b).

181 Notably, *ATP2A1*, *PPFIA4*, and *NIPBL* were recurrently mutated, either within our  
182 cohort or with a recent report<sup>24</sup> (Extended Data Fig. 3a-b), occurring within the latter 3 clusters.  
183 While these clusters were not previously reported in MCDs, they were previously implicated in  
184 epilepsy, neurodevelopmental and neurodegenerative disease<sup>25,26</sup>, suggesting functional overlap  
185 with MCDs. We further performed ClueGO analysis and found enrichment in mTOR signaling,  
186 focal adhesion assembly, cardiac muscle cell contraction, and artery morphogenesis (Extended  
187 Data Fig. 4). ClueGO also displayed isolated gene ontology (GO) term clusters such as ‘calcium  
188 ion import’ and ‘protein localization to synapse’.

189

## 190 **Functional validation of selected module genes in embryonic mouse brain**

191 To investigate the roles of novel MCD genes and modules, we selected two potential mTOR  
192 pathway mutations (*RRAGA* p.H226R, *KLHL22* p.R38Q), and non-mTOR gene mutation  
193 (*GRIN2C* p.T529M), discovered in FCD-7967, 3560, and 5157, respectively. *RRAGA* encodes  
194 Ras-related GTP binding A (RAGA), a GTPase sensing amino acid and activating mTOR  
195 signaling, with two functional domains: GTPase domain and C-terminal ‘roadblock’ domain  
196 (CRD)<sup>27</sup>. The mosaic p.H226R mutation occurs within the CRD, which binds to the RAGB  
197 protein and is conserved throughout vertebrate evolution (Extended Data Fig. 3c) and thus could  
198 change binding affinity. *KLHL22* encodes a CUL3 adaptor, determining E3 ubiquitin ligase  
199 specificity. The CUL3-*KLHL22* complex mediates the degradation of DEPDC5, required for  
200 mTORC1 activation<sup>28</sup>. The *KLHL22* p.R38Q variant in FCD-3560 is near the BTB (Broad-  
201 Complex, Tramtrack, and Bric-à-brac) domain that interacts with CUL3 (Extended Data Fig. 3d),  
202 suggesting the variant could enhance mTORC1 activity. *GRIN2C* encodes a subunit of the  
203 NMDA receptor regulating synaptic plasticity, memory, and cognition<sup>29,30</sup>, dysfunction of which  
204 is implicated in many neurocognitive diseases including epilepsy, neurodevelopment, and  
205 tumors<sup>31,32</sup>. *GRIN2C* p.T529M mutation is located in the S1 glutamate ligand-binding domain  
206 (S1 LBD) (Extended Data Fig. 3e). *GRIN2A* p.T531M mutation, an analog mutation of *GRIN2C*  
207 p.T529M in our cohort, was previously reported in epilepsy-aphasia spectrum disorders, where it  
208 increased NMDA receptors ‘open-state’ probability<sup>32</sup>. This suggests that the p.T529M mutation  
209 activates the channel, likely in an mTOR independent fashion. Thus, all mutations assessed here  
210 are likely gain-of-function and exert functional impact on cells in which they are expressed.

211 To test this hypothesis, we introduced mutant or wildtype (WT) genes co-expressing  
212 enhanced green fluorescent protein (EGFP) into the dorsal subventricular zone via  
213 electroporation at mouse embryonic day 14 (E14), then harvested tissue at either E18 to assess  
214 migration, or at postnatal day 21 (P21) to assess cell size and phospho-S6 as a reporter of mTOR  
215 activity<sup>33</sup> (Fig. 3a). In E18 cortices, we found EGFP-positive cells expressing mutant but not WT  
216 forms of *RRAGA* and *KLHL22* showed significant migration defects of varying severity, whereas  
217 mutant *GRIN2C* showed no defect (Fig. 3b). These migration defects in *RRAGA* and *KLHL22*  
218 mutant cells replicate major findings of MCD disrupted cortical architecture.

219 We next assessed cellular phenotype at P21 with samples available in both mice and the  
220 corresponding patients and found enlarged cell body area in both mutant forms of *KLHL22* and  
221 *GRIN2C* compared to according wildtype. In contrast, the elevated levels of pS6 staining,  
222 described previously in association with mTOR pathway mutations<sup>6</sup>, was found only in mutant  
223 *KLHL22*, but not in mutant *GRIN2C* mice (Fig. 3c).

224 To assess correlation with human samples, we assessed archived neuropathological tissue  
225 sections for histology and pS6 activity. Similar to our mouse models, we found patient FCD-

226 3560 carrying *KLHL22* p.R38Q showed enlarged neurons that co-stained for excess pS6 staining,  
227 whereas FCD-5157 carrying *GRIN2C* p.T529M showed only a slight increase in cell body size  
228 and no evidence of excessive pS6 staining (Fig. 3d). While this analysis does not take into  
229 account the genotype of individual cells, it suggests *KLHL22* but not *GRIN2C* mutations impact  
230 mTOR signaling.

231

### 232 **Genotype-phenotype correlations in MCD patients**

233 To assess the phenotypic contributions of the MCD genes we found, we focused on 76 of our  
234 ‘genetically solved’ MCD cases, comparing detailed neuropathology, brain imaging, and clinical  
235 course. We performed Pearson correlation followed by hierarchical clustering based upon ILAE  
236 neuropathological diagnosis, compared with GO term-based curated genesets and whether the  
237 genetic variant was present in COSMIC DB (Methods, Supplementary Table 3,4, Fig. 4). We  
238 found that FCD Type IIA and Type IIB, and HME were more tightly clustered than FCD Type I  
239 or III (Fig. 4a), likely reflecting shared neuropathological features that include large dysplastic  
240 neurons. As expected, FCD Type IIA, Type IIB, and HME were positively associated with the  
241 mTOR pathway GO term and COSMIC DB entry, FCD Type III, however, was associated with  
242 the MAPK pathway, consistent with recent publications implicating *BRAF*, *FGFR2*, *NOD2*, and  
243 *MAPK9* in their etiology<sup>34-36</sup>. FCD Type I showed few strong positive correlations for  
244 glycosylation, consistent with recent findings of somatic mutations in *SLC35A2* and *CANT1*<sup>37,38</sup>.

245 We next investigated correlations between clinical phenotypes extracted from detailed  
246 medical records including seizure type, neuropsychological examination, and positron emission  
247 tomography (PET) metabolism, often used to help localize seizure focus<sup>39,40</sup>. Seizure frequency,  
248 early age of onset, Engel score, and history of infantile spasms drove clinical clustering, likely  
249 reflecting shared clinical features in the most challenging patients. Focusing on the correlations,  
250 PET hypometabolism correlated positively with COSMIC DB entry, and negatively with MAPK  
251 and Ubiquitination (Fig. 4b), suggesting divergent metabolic mechanisms. Abnormal  
252 neurological examination correlated positively with COSMIC DB entry and negatively with  
253 Type I histology, which may reflect the effects of mutations on baseline neurological function.

254

### 255 **MCD genes enriched in the excitatory neuronal lineage**

256 To infer the cell type in which MCD genes function, we accessed a published single-cell  
257 transcriptome dataset from the 2nd-trimester human telencephalon, at a time when these  
258 mutations probably arose<sup>41</sup> (Fig. 5a). We generated an eigengene, by mapping the average  
259 expression of our MCD genes against the UMAP plot (Fig. 5b). This showed a strong positive  
260 correlation with dividing radial glial cells, and a moderate correlation in dividing intermediate  
261 progenitor cells (IPCs) and mature excitatory neuron cells. We found a negative correlation with  
262 inhibitory neuronal lineages including medial and central ganglionic eminences (MGE, CGE)  
263 and mature interneuron clusters (Fig. 5c). We next performed deconvolution into four major  
264 module eigengene (MEs), which revealed cell types classified as mature excitatory neurons  
265 (turquoise and blue), microglia (brown), and unassigned (grey) (Fig. 5d). Quantification  
266 supported enrichment in dividing radial glia, excitatory neurons, and microglia, the latter likely  
267 driven by MCD candidate genes *IRF8* and *VSIG4* (Fig. 5e). Taken together, the expression of  
268 MCD genes is more enriched in dorsal cortex neurogenic pools and implicated in the maturation  
269 of excitatory rather than inhibitory neurogenic pools, as well as microglia.

270

## 271 MCD gene expression is enriched in dysplastic cells

272 We next performed differentially expressed gene (DEG) analysis in the MCD brain. We  
273 reasoned that single-nucleus transcriptomes would be more revealing than bulk transcriptomes,  
274 but the average VAF of ~6% in our MCD cohort meant that the vast majority of sequenced cells  
275 would be genetically wild-type. We thus decided to focus snRNAseq on resected cortex from  
276 patients with shared pathological MCD hallmarks but higher VAFs. We selected four resected  
277 brain samples, two from patients with HME (HME-4688 *PIK3CA* p.E545K, 25.1% VAF and  
278 HME-6593 *PIK3CA* p.H1047R, 13.1% VAF), and two from patients with TSC meet full  
279 diagnostic criteria. We also included brains from four neurotypical cases as a comparison and  
280 sequenced a total of 22,067 nuclei (see Methods).

281 While the TSC brain single nucleus transcriptomes showed substantial overlapping pools  
282 with controls, HME brains showed a distinct UMAP distribution, located at the edges of the plot  
283 (Fig. 6a). We found that very few HME cells matched expression patterns for typical brain cells,  
284 even after standard normalization and scaling (Fig. 6b, Extended Data Fig. 5a, see Methods). We  
285 thus labeled these clusters according to their closest relatives based upon established marker  
286 gene expression in the control brain, labeled as ‘astrocyte-like (Ast-L)’ or ‘oligodendrocyte-like  
287 (OD-L)’. Even with these categories, some clusters remained undefined (U) (Extended Data Fig.  
288 5b,c). Interestingly, there was no single cell cluster that matched the VAF in the brain,  
289 suggesting the mutant cells, as well as surrounding non-mutant cells, have dramatically disrupted  
290 transcriptomes.

291 We noted that several of the HME clusters showed excessive expression of fibroblast  
292 growth factor receptor (FGFR) gene families, specifically *FGFR1* in cluster U1/2 in HME,  
293 *FGFR2/3* in cluster Ast-L1/3 and OD-L, *EGFR* in Ast-L1/3 and U1/2, and *PDGFRA* in cluster  
294 U1/2 (Extended Data Fig. 5b,c). To identify the cell types expressing these genes, we performed  
295 RNA in situ hybridization in HME brain sections followed by hematoxylin-eosin staining. We  
296 found co-localization of these same *FGFR* family, *EGFR*, and *PDGFRA* transcripts with  
297 dysplastic cells (Extended Data Fig. 6). Previous experiments indicate that it is most often the  
298 dysplastic cells within HME and MCD that carry disease mutations<sup>7</sup>, suggesting an effect of  
299 these mutations on growth factor receptor expressions that correlates with dysplasia.

300 Next, we investigated the expression patterns of MCD genes in this HME/TSC  
301 snRNAseq dataset. An eigengene representing expression patterns of MCD genes was enriched  
302 in Ast-L1/3 and OD-L, which were labeled as dysplastic cells (Fig. 6c). Interestingly, the  
303 individual MCD genes displayed converging expression patterns resulting in six different  
304 eigengenes (Fig. 6d, gene members for each eigengene are described in Extended Data Fig. 7)  
305 which show distinct enrichment patterns across cell types (Fig. 6e), implying that membership of  
306 each eigengene may be associated with the pathophysiology of the corresponding dysplastic cell  
307 type in HMEs. We performed a pseudo-bulk DEG analysis comparing HME with CTRL and  
308 detected 590 up-regulated genes and 1096 down-regulated genes. Intriguingly, 20% (15/75) of  
309 MCD mutated genes in our list overlapped with DEGs of HME. Permutation testing suggests  
310 that this overlap is unlikely to have arisen by chance (Fig. 6f, see Methods). Taken together,  
311 many MCD genes are misregulated in MCD-specific cell types, suggesting that our MCD genes  
312 may play important roles in the pathogenesis of dysplastic cells in MCDs.

313

## 314 Discussion

315 In this study, we use a multiomics approach to study the genetic landscape of MCD in the largest  
316 reported cohort to date. We confirmed the important role of mTOR/MAP kinase and  
317 glycosylation pathways, seen in about 60.5% of those with mutations. Moreover, our results also  
318 linked novel biological processes including gene expression, synaptic function, and calcium  
319 dynamics, which made up the other 39.5% of mutations. Nevertheless, only 76 of 317 patients  
320 showed one or more putative somatic mutations as a likely cause of MCD. There could be  
321 numerous causes for the relatively low solve rate in MCD, including the potential to miss very  
322 low VAF mutations and the contribution of complex mutations like structural variants or short  
323 tandem repeats polymorphism. Finally, although patients with environmental causes, syndromic,  
324 or inherited causes were excluded from our cohort, these factors could still contribute to MCD.

325 With our approach, we identified several recurrently-mutated genes not previously  
326 implicated in MCD. Confirming the remaining candidate and identifying further MCD candidate  
327 genes will require larger MCD cohorts. Including novel MCD candidate genes emerging from  
328 300X WES into the 1000X Phase 3 AmpliSeq allowed both confirmation of mutations, a more  
329 accurate estimate of VAF, and identification of additional patients with these genes mutated that  
330 would have been perhaps missed with 300X WES. Functional validation by modeling mutations  
331 in embryonic mouse brains suggests that most candidate genes we identified are likely to  
332 contribute to disease. Perhaps it is not surprising that there are so many MCD genes, because  
333 such mutations may avoid embryonic lethality due to their expression in just a small subset of  
334 cells. Like with de novo germline mutations discovered in autism, we suggest that there could be  
335 dozens if not hundreds of additional MCD genes, based in part upon the low number of  
336 recurrently mutated genes<sup>42</sup>.

337 The four gene networks, mTOR/MAP kinase, calcium dynamics, synapse, and gene  
338 expression, are intriguing, as they should play important roles for these genes both during brain  
339 development and homeostasis. All four pathways are critical both for corticogenesis during  
340 neurogenesis and neuronal migration, as well as neuronal excitability. For instance, calcium  
341 dynamics is shown to regulate cytoskeletal activity and excitability<sup>43,44</sup>. The genotypic  
342 information also showed correlations with clinical features, for instance, PET brain  
343 hypometabolism and abnormality in the neurological examination correlated with COSMIC DB  
344 variants, opening the possibility to predict genotype based on phenotype.

345 We also characterized the expression patterns of MCD genes in the developmentally  
346 normal and MCD brains at single-cell resolution. The cell types most strongly expressing  
347 candidate MCD genes include dorsal forebrain radial glial progenitors and their daughter  
348 excitatory neurons, as well as brain microglia, fitting well with the likely site of origin of somatic  
349 brain mutations<sup>45</sup>. Surprisingly, the dramatic gene dysregulation seen in the HME brain skewed  
350 the UMAP plots in ways that could not be accounted for simply by the VAF. The fact that the  
351 MCD genes also showed the strongest enrichment with these same clusters suggests that the  
352 MCD genes are very likely to have pivotal roles in the HME condition. Prior studies on MCD  
353 indicated that dysplastic cells express markers for both glia and neurons<sup>46</sup>. Our findings,  
354 however, suggest that MCD mutations drive critical roles predominantly in dividing radial glia,  
355 with profound effects on lineage and cellular dysplasia. To conclude, the MCD genes in patient  
356 brains found in our study demonstrated critical roles during cortical development, significantly  
357 correlate with patient phenotypes, and could open doors to novel treatments for MCDs.

358

359 **Online Methods**

360

361 **Overview of the FCD cohort**

362 This study is a multi-center international collaboration. We recruited a cohort of 317 individuals  
363 from the 'FCD Neurogenetics Consortium' (see the member list). These individuals were  
364 diagnosed with FCD type I, II, III, HME, or TSC and underwent surgical resection to treat drug-  
365 resistant epilepsy between 2013 and 2021. Any cases that underwent surgical resection due to  
366 environmental factors, for example, stroke, or acute trauma, were excluded. For each individual,  
367 resected brain tissue was collected, along with paired blood or saliva samples and parental  
368 samples, where available. Clinical history, pre- and post-operative brain imaging,  
369 histopathology, ILAE classification according to the surgical tissue pathology report, and Engel  
370 surgical outcome score (at least two years after surgery) were collected, when available.

371

372 **Informed consent and study approval**

373 The study protocol was approved by the UC San Diego IRB (#140028). Informed consent was  
374 obtained from all participants or their legal guardians at the time of enrollment.

375

376 **DNA extraction**

377 Pulverized cortical samples (~0.3 g) were homogenized with a Pellet Pestle Motor (Kimble,  
378 #749540-0000) or Handheld Homogenizer Motor (Fisherbrand, #150) depending on the size of  
379 the tissue, and resuspended with 450 µL RLT buffer (Qiagen, #40724) in a 1.5 ml  
380 microcentrifuge tube (USA Scientific, #1615-5500). Homogenates were then vortexed for 1  
381 minute and incubated at 70°C for 30 minutes. 50 µl Bond-Breaker TCEP solution (Thermo  
382 Scientific, #77720) and 120 mg stainless steel beads with 0.2 mm diameter (Next Advance,  
383 #SSB02) were added, and cellular disruption was performed for 5 minutes on a DisruptorGenie  
384 (Scientific industries). The supernatant was transferred to a DNA Mini Column from an AllPrep  
385 DNA/RNA Mini Kit (Qiagen, #80204) and centrifuged at 8500 xg for 30 seconds. The column  
386 was then washed with Buffer AW1 (kit-supplied), centrifuged at 8500 xg for 30 seconds and  
387 washed again with Buffer AW2 (kit-supplied), and then centrifuged at full speed for 2 minutes.  
388 The DNA was eluted two times with 50 µl of pre-heated (70°C) EB (kit-supplied) through  
389 centrifugation at 8,500 xg for 1 minute.

390

391 **MPAS and WES sequencing for somatic mutation candidates**

392 Massive parallel amplicon sequencing (MPAS) and whole-exome sequencing (WES) were used  
393 at different phases to perform the genetic screening within available samples from the cohort.  
394 Customized AmpliSeq DNA panels for Illumina ( Illumina, #20020495) were used for Massive  
395 Parallel Amplicon Sequencing<sup>17</sup>. 87 or 82 genes related to the mTOR pathway or curated based  
396 on the results of Phase 1 and 2, respectively, were subjected to the AmpliSeq design system; a  
397 list of designed genes is provided in Supplementary Table 2a-b. Two pools were designed for  
398 tiling the capture region. Genomic DNA from extracted tissue was diluted to 5 ng/uL in low TE  
399 provided in AmpliSeq Library PLUS (384 Reactions) kit (Illumina, #20019103). AmpliSeq was  
400 carried out following the manufacturer's protocol (document #1000000036408v07). For  
401 amplification, 14 cycles each with 8 minutes were used. After amplification and FUPA  
402 treatment, libraries were barcoded with AmpliSeq CD Indexes (Illumina, #20031676) and pooled

403 with similar molecular numbers based on measurements made with a Qubit dsDNA High  
404 Sensitivity kit (Thermo Fisher Scientific, #Q32854) and a plate reader (Eppendorf, PlateReader  
405 AF2200). The pooled libraries were subjected to Illumina NovaSeq 6000 platform for PE150  
406 sequencing. The AmpliSeq design in the ‘Phase 1’ is under the design ID IAA7610, and the  
407 AmpliSeq design in ‘Phase 3’ is under the design ID IAA26010.

408 Genomic DNA (~ 1.0 µg) was prepared for whole-exome sequencing, and libraries were  
409 captured using the Agilent SureSelect XT Human All Exon v.5 or Nextera DNA Exome kits.  
410 Then, 100, 125, or 150 bp paired-end reads (median insert size ~ 210 bp) were generated using  
411 the Illumina HiSeq X 2500 platform. The sequencing experiments were designed to yield three  
412 datasets of ~ 100X coverage on each sample, with a coverage goal of 300X from the brain and  
413 100X from blood/saliva.

414

#### 415 **Somatic variant calling from MPAS and WES**

416 Reads were aligned to GRCh37 using BWA (version 3.7.16a), sorted per each read group, and  
417 merged into a single BAM file with Sambamba (version 0.6.7). The merged BAM files were  
418 marked for duplicate reads using PICARD (v2.12.1), duplicated reads were not removed for  
419 MPAS because of the nature of the method. Then, we performed indel realignment and base  
420 quality recalibration using GATK (v3.7–0), resulting in the final uniformed processed BAM  
421 files.

422 Both tissue-specific and tissue-shared mosaic variants were called from the MPAS and  
423 WES sequencing data. MPAS and WES variants were called according to the availability of the  
424 control tissue. Brain- and blood/saliva-specific variants were called using MuTect2 (GATK3.8)  
425 paired mode and Strelka2 somatic mode<sup>47</sup>; the BAM files from the brain sample (combined and  
426 non-combined from independent sequencing libraries) and blood/saliva samples were treated as  
427 “tumor-normal” and “normal-tumor” pairs separately and cross-compared between each other.  
428 Variants called by both callers were listed. Mosaic variants shared between the brain and  
429 fibroblast samples were called using the single mode of MosaicHunter<sup>11</sup> by either combining all  
430 brain replicates or calling each separate sample. Variants that passed all the MosaicHunter filters  
431 also were listed. Somatic variants from WES data were further called by GATK (v3.7–0)  
432 haplotypecaller with ploidy parameter set to 50, followed by a series of heuristic filters described  
433 as the best-practice by the Brain somatic mosaicism network<sup>12</sup>, tissue-shared variants were called  
434 by the combination of MuTect2<sup>48</sup> (GATK 3.8) single-mode and DeepMosaic<sup>10</sup>.

435 A union of different pipelines was selected to get maximum sensitivity. Mosaic  
436 candidates from the combined lists were further filtered using the following criteria: (i) the  
437 variant had more than 3 reads for the alternative allele; (ii) the variant was not present in UCSC  
438 repeat masker or segmental duplications; (iii) the variant was at least 2 bp away from a  
439 homopolymeric tract; and (iv) the variant exhibited a gnomAD allele frequency lower than  
440 0.001. Variants that exist in the 1000 genome project (phase 3) also were excluded from the  
441 analysis. Variants from both exome data sources were tested and a combination of tissue-specific  
442 mosaic variants and tissue-shared mosaic variants were collected and the credible interval of  
443 VAFs was calculated using a Bayesian-based method described previously<sup>49</sup>. To filter for  
444 candidate disease-causing variants for FCD, we further filtered out synonymous variants in  
445 coding regions, variants with CADD Phred score < 25, and candidates that fell out of coding  
446 regions and were not predicted to affect splicing by ANNOVAR.

447 **False discovery estimation**

448 To calculate the false discovery of random variants detected in normal samples, we incorporated  
449 75 normal control samples (71 brains and 4 other organs) previously sequenced with 250-300X  
450 WGS, which should provide similar sensitivity as our exomes, the deep WGS were generated by  
451 efforts from the NIMH Brain Somatic Mosaicism Consortium<sup>12</sup>, from controls<sup>17</sup>, and from our  
452 recent mutation detection pipeline<sup>18</sup>. Variants were filtered based on the identical criteria as  
453 described in the above data analysis part, with >0.01 VAF, all on exonic regions defined by  
454 NCBI, and CADD score >25. While 13 variants remain positive from this pipeline from the 75  
455 samples (0.17 per control), 306 candidate variants were determined in our 134 MCD exomes  
456 (2.28 per MCD case), which lead to an estimated 7.59% per sample false discovery rate  
457 (Supplementary Table 5).

458 **Orthogonal validation and quantification of mosaic mutations with targeted amplicon  
459 sequencing**

460 Targeted amplicon sequencing (TASeq) with Illumina TruSeq was performed with a coverage  
461 goal of >1000X for 554 candidate variants detected by computational pipelines described above  
462 for both MPAS and WES, to experimentally validate the mosaic candidates before functional  
463 assessment. PCR products for sequencing were designed with a target length of 160-190 bp with  
464 primers being at least 60 bp away from the base of interest. Primers were designed using the  
465 command-line tool of Primer3<sup>50,51</sup> with a Python (v3.7.3) wrapper<sup>13,14</sup>. PCR was performed  
466 according to standard procedures using GoTaq Colorless Master Mix (Promega, M7832) on  
467 sperm, blood, and an unrelated control. Amplicons were enzymatically cleaned with ExoI (NEB,  
468 M0293S) and SAP (NEB, M0371S) treatment. Following normalization with the Qubit HS Kit  
469 (ThermFisher Scientific, Q33231), amplification products were processed according to the  
470 manufacturer's protocol with AMPure XP beads (Beckman Coulter, A63882) at a ratio of 1.2x.  
471 Library preparation was performed according to the manufacturer's protocol using a Kapa Hyper  
472 Prep Kit (Kapa Biosystems, KK8501) and barcoded independently with unique dual indexes  
473 (IDT for Illumina, 20022370). The libraries were sequenced on Illumina HiSeq 4000 or NovaSeq  
474 6000 platform with 100 bp paired-end reads.

475

476 **Mutational signature analysis**

477 Mutational signature analysis was performed using a web-based somatic mutation analysis  
478 toolkit (Mutalisk)<sup>52</sup>. PCAWG SigProfiler full screening model was used.

479

480 **STRING analysis**

481 STRING analysis was performed by STRING v11<sup>22</sup>. A total of 75 MCD genes were loaded as  
482 input and MCL clustering was performed. The terms in Gene Ontology (GO), KEGG pathways,  
483 and Top 10 terms GO or KEGG pathways were shown in Fig. 2b. If there are less than 10 terms  
484 for those terms (such as clusters 3 and 4 in Fig. 2), we included all the terms in GO or KEGG  
485 pathways, Local network cluster (STRING), Reactome pathways, and Disease-gene associations  
486 (DISEASES) to show the enriched terms. Visualization was performed by Cytoscape v3.9.

487

488 **ClueGO analysis**

489 Visualization of the functionally grouped biological terms was performed by ClueGO v2.5<sup>53</sup>, a  
490 Cytoscape plug-in. A total of 75 MCD genes from Fig. 2 were loaded and GO terms in the  
491 'Biological Process' category were used for visualization. Terms with a  $p < 0.01$ , a minimum  
492 count of 3, and an enrichment factor  $> 1.5$ , are grouped into clusters based on membership  
493 similarities.

494

## 495 **Animals**

496 Pregnant Crl: CD1(ICR) mice for mouse modeling were purchased from Charles River  
497 Laboratory. All mice used were maintained under standard group housing laboratory conditions  
498 with 12 hours light/dark cycle and free access to food and water. The age and number of mice  
499 used for each experiment are detailed in the figure legends. The sex of the embryos used was not  
500 tested. All work with mice was performed in accordance with UCSD IACUC protocol S15113.

501

## 502 **DNA constructs**

503 *RRAGA*, *KLHL22*, and *RHOA* ORF regions were amplified from the hORFeome library and  
504 inserted into the pCIG2 (pCAG-IRES-GFP) vector. *GRIN2C* ORF region was purchased from  
505 DNASU Plasmid Repository in Arizona State University Biodesign Institute. All sequences of  
506 clones were confirmed by sanger sequencing.

507

## 508 **In utero electroporation**

509 In utero electroporation was performed as described previously<sup>54</sup> with modifications as follows.  
510 Endotoxin-free plasmids (0.5–1  $\mu$ g) plus 0.1% Fast Green (Sigma, catalog no. 7252) was injected  
511 into one lateral ventricle of E14.5 embryos. Electroporation was performed by placing the anode  
512 on the side of the DNA injection and the cathode on the other side of the head to target cortical  
513 progenitors. Four pulses of 45 V for 50 ms with 455-ms intervals were used.

514

## 515 **Mouse brain section preparation**

516 An E18 mouse brain is fixed in 4% paraformaldehyde (PFA) for 2 hrs. For the P21 mouse brain,  
517 a mouse was anesthetized by isoflurane and perfused by cold 1X PBS for 8 min and following  
518 4% cold PFA for 8 min. The brains were dehydrated in 30% sucrose in 1x PBS for 48 hrs and  
519 embedded in Tissue-Tek optimal cutting temperature compound and frozen on dry ice. A frozen  
520 block was sectioned with 20  $\mu$ m thickness in a cryostat (CryoStar NX70, Thermo Fisher  
521 Scientific) and placed on sliding glass. The attached sections were dried on a 50 °C heating block  
522 for 3 hrs.

523

## 524 **Immunofluorescence staining and imaging**

525 A section was rehydrated and washed by 1X PBS for 10 min 3 times, permeabilized in PBST  
526 (0.3% Triton X-100 in 1X PBS) for 10 min, and blocked by blocking solution (5% normal BSA  
527 in 1X PBS) for 2 hrs in room temperature. Sections were stained with diluted primary antibodies  
528 in the blocking solution overnight at 4 °C. The next day, the sections were washed with PBST for  
529 5 min three times and stained with secondary antibodies in blocking solution for 2 hrs in RT.  
530 Blocking solution was dropped off from the slides and nuclei staining with DAPI solution  
531 (0.1  $\mu$ g/ml of DAPI in PBST) was performed for 15 min. The slides were mounted with DAKO

532 fluorescent mount solution (catalog no. S3023). Zeiss 880 Airyscan Confocal is used for imaging  
533 according to the manufacturer's instructions.

534

### 535 **Antibodies**

536 phospho-S6 (1:800 dilution, catalog no. 5364S ;Cell Signaling, AB\_10694233), NeuN (1:100,  
537 MAB377X; Sigma-Aldrich, AB\_2149209), GFP (1:500, catalog no. GFP-1020, Aves Labs,  
538 AB\_10000240), Alexa Fluor Goat 488 chicken IgY (H+L) (1:1,000 dilution, catalog no. A-  
539 11039, AB\_2534096), Alexa Fluor 594 donkey anti-rabbit IgG (H+L) (1:1,000, catalog no.  
540 R37119, AB\_2556547).

541

### 542 **Genotype-phenotype association**

543 The functional modules to be tested were selected based on the enriched GO terms (Fig. 2 and  
544 Extended Data Fig. 4). A given candidate MCD gene was assigned as a member to one or  
545 multiple modules based on GO terms related to the given gene (results summarized in  
546 Supplementary Table 3c). Subsequently, a given patient became a member of one (or multiple)  
547 functional module(s) if the genes detected in that patient were assigned to that (those) functional  
548 module(s). All available clinical information on the patient was collected and harmonized using  
549 ILAE terms (summarized in Supplementary Table 4). Pearson correlation coefficients were  
550 calculated by cor.test() function in R. The value of correlation coefficients were displayed as  
551 colors in the heatmap of Fig. 4. If two groups with binary values were used for calculation, Phi  
552 coefficient was used.

553

### 554 **Single-nucleus RNA sequencing**

555 A fresh-frozen brain tissue (~50 mg) was placed into a glass dounce homogenizer containing 1  
556 ml cold lysis buffer (0.05 % (v/v) NP-40, 10 mM Tris (pH 7.4), 3 mM MgCl<sub>2</sub>, 10 mM NaCl) and  
557 dounce 10 times with a loose pestle and following 10 times with a tight pestle. The homogenate  
558 was incubated for 10 min in RT. 9 ml of wash buffer (1% BSA in 1X PBS) was added to the  
559 homogenate and filtered by a 30 um cell strainer. The strained homogenate was spun down in  
560 500 g to remove the supernatant. The pellet was resuspended by 5 ml of wash buffer. Straining,  
561 spinning down steps was performed once more, and the pellet was resuspended into 500 ul of  
562 wash buffer. 10 ul of nuclei resuspension was mixed with counting solution (0.02 % Tween 20,  
563 0.1ug/ml DAPI, 1% BSA in 1X PBS) and nuclei density was measured by manual nuclei  
564 counting using DAPI signal. The resuspension was diluted by wash buffer to make the desired  
565 concentration (800~1000 nuclei/ul). 1~4 samples were pooled together targeting 10000 nuclei  
566 per reaction. Gel beads emulsion (GEM) generation, cDNA, and sequencing library  
567 constructions were performed in accordance with instructions in the Chromium Single Cell 3'  
568 Reagent Kits User Guide (v3.1). A library pool was sequenced with 800 million read pairs using  
569 NovaSeq 6000.

570

### 571 **Single-nucleus RNAseq bioinformatics pipeline**

572 Fastq files from single-nucleus libraries were processed through Cell Ranger (v6.0.2) analysis  
573 pipeline with –include-introns option and hg19 reference genome. Pooled library was  
574 demultiplexed and singlets were taken by demuxlet. Seurat (v4) package was used to handle

575 single nuclei data objects. Protein coding genes were used for further downstream analysis.  
576 Nuclei passed a control filter (number of genes > 500, number of reads >1000, percentage of  
577 mitochondrial gene < 10%) was used for downstream analysis. Data were normalized and scaled  
578 with the most variable 5000 features using the ‘NormalizeData’ and ‘ScaleData’ functions.  
579 Dimensionality reduction by PCA and UMAP embedding was performed using runPCA and  
580 runUMAP function. Clustering was performed by FindNeighbors and FindClusters function. Cell  
581 type identification was performed using known cell type markers expressed in the brain  
582 including excitatory/inhibitory neuron, astrocyte, oligodendrocyte, microglia, and endothelial  
583 cell markers as well as using positive markers found by FindAllMarkers function with 3000 most  
584 variable features in scaled data.

585

#### 586 **Weighted gene co-expression network analysis**

587 ‘r-wgcn’ package (v1.69) was used for WGCNA according to instructions (PMID: 19114008).  
588 Briefly, a similarity matrix was generated based on Pearson’s correlation coefficient value  
589 among the top 3000 variable features in single-nucleus transcriptome data, which was used to  
590 calculate the subsequently signed type of network adjacency matrix. Next, the topological  
591 overlap matrix (TOM) and the corresponding dissimilarity (1-TOM) value were generated from  
592 the adjacency matrix. Finally gene modules were generated by ‘cutreeDynamic’ function with  
593 ‘tree’ method, minAbsSplitHeight = 0.9 and minClusterSize = 30 option. Similar gene modules  
594 were merged by ‘mergeCloseModules’ function with cutHeight = 0.25. String analysis was  
595 performed using each gene module for the identification of the given module’s functional  
596 characteristics.

597

#### 598 **RNAscope**

599 We used published methods and purchased target probes for genes of interest containing an 18-  
600 25 base region complementary to the target, as spacer sequencing, and a 14 base Z-tail  
601 sequence<sup>55</sup>, including RNA pol III positive control and random sequence negative control,  
602 following the manufacturer recommendations (Advanced Cell Diagnostics, Hayward, CA).  
603 Images were acquired on a Leica STED Sp8 with Falcon microscope.

604

#### 605 **Permutation analysis for the enrichment of MCD genes**

606 To test the enrichment of differentially expressed MCD genes in RNA sequencing against a  
607 random distribution, we designed a permutation analysis. All human genes used in the single-cell  
608 RNA-seq analysis (n=19909) were randomly shuffled 10,000 times and the same number of  
609 genes as described in the differential expression analysis (n=1686) was selected for each shuffle.  
610 The number of overlaps between each shuffle and the MCD candidates was compared and the  
611 number of overlaps was used as the outcome and a null distribution was generated from the  
612 10,000 shuffles. All 75 positively validated MCD genes are confirmed to be existing in the initial  
613 gene list. After 10,000 permutations, the permutation p-value was calculated with numbers >=  
614 observed overlap (p=0.0017 for the data shown in the main text).

615

#### 616 **Statistical analyses**

617 Statistical analyses were performed by R or Prism 8 (GraphPad Software). Two-way ANOVA  
618 and Sidak multiple comparisons were performed in Fig 3b with p-values of interaction between  
619 genotype and bin factor. \*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ .

620

## 621 **Code and data availability**

622 Code to generate the figures and analyze the data are publically available on GitHub  
623 ([https://github.com/shishenyxx/MCD\\_mosaic](https://github.com/shishenyxx/MCD_mosaic)). WES and AmpliSeq data are deployed on NIMH  
624 Data Archive under study number 1484 “Comprehensive multiomic profiling of somatic  
625 mutations in malformations of cortical development” and SRA under accession number  
626 PRJNA821916: “Comprehensive multiomic profiling of somatic mutations in malformations of  
627 cortical development”. The snRNAseq R object was deposited in Single Cell Portal  
628 ([https://singlecell.broadinstitute.org/single\\_cell/study/SCP1815/comprehensive-multiomic-  
629 profiling-of-somatic-mutations-in-malformations-of-cortical-development#study-download](https://singlecell.broadinstitute.org/single_cell/study/SCP1815/comprehensive-multiomic-profiling-of-somatic-mutations-in-malformations-of-cortical-development#study-download)).

630

## 631 **Acknowledgments**

632 AmpliSeq, TASEq, and snRNAseq were supported by NIH P30CA023100 and S10OD026929 at  
633 the UCSD IGM Genomics Center. Rady Children’s Institute for Genomic Medicine, Broad  
634 Institute (U54HG003067, UM1HG008900), the Yale Center for Mendelian Disorders  
635 (U54HG006504), and the New York Genome Center provided whole-exome sequencing. UCSD  
636 Microscopy core (NINDS P30NS047101) provided imaging support. CC was supported by a  
637 2021 Brain & Behavior Research Foundation Young Investigator Grant. This study was  
638 supported by the NIH (NIMH U01MH108898 and R01MH124890 to JGG and GWM, and NIA  
639 R21AG070462, NINDS R01NS083823 to JGG). We thank Stéphanie Baulac and Sara  
640 Baldassari for sharing unpublished exome data.

641

## 642 **Author contributions**

643 C.C., X.Y., and J.G.G. designed the study. C.C., S.M., and S.K. conducted functional validation.  
644 C.B., V.S., A.N., E.R., C.C., and G.H. coordinated the clinical database. X.Y., C.C., M.W.B.,  
645 L.L.B., R.D.G., J.G., M.X., A.P.L.M., and K.N.J. organized, handled, and sequenced human  
646 samples. X.Y., C.C., T.B., X.X., and B.C. performed bioinformatics and data analysis. C.C. and  
647 K.I.V. performed the RNAscope experiment. C.D., H.W.P., C.A.B.G., S.H.K., H.K., A.S.,  
648 C.A.H., C.G., C.A.G., S.S., M.N., D.D.G., K.I., Y.T., R.C., J.T., V.C., R.G., O.D., W.A.S.,  
649 H.R.M., and G.W.M. provided resected brain tissues and clinical data from FCD patients. C.C.,  
650 X.Y., and J.G.G. wrote the manuscript. All authors read and commented on the manuscript  
651 before submission.

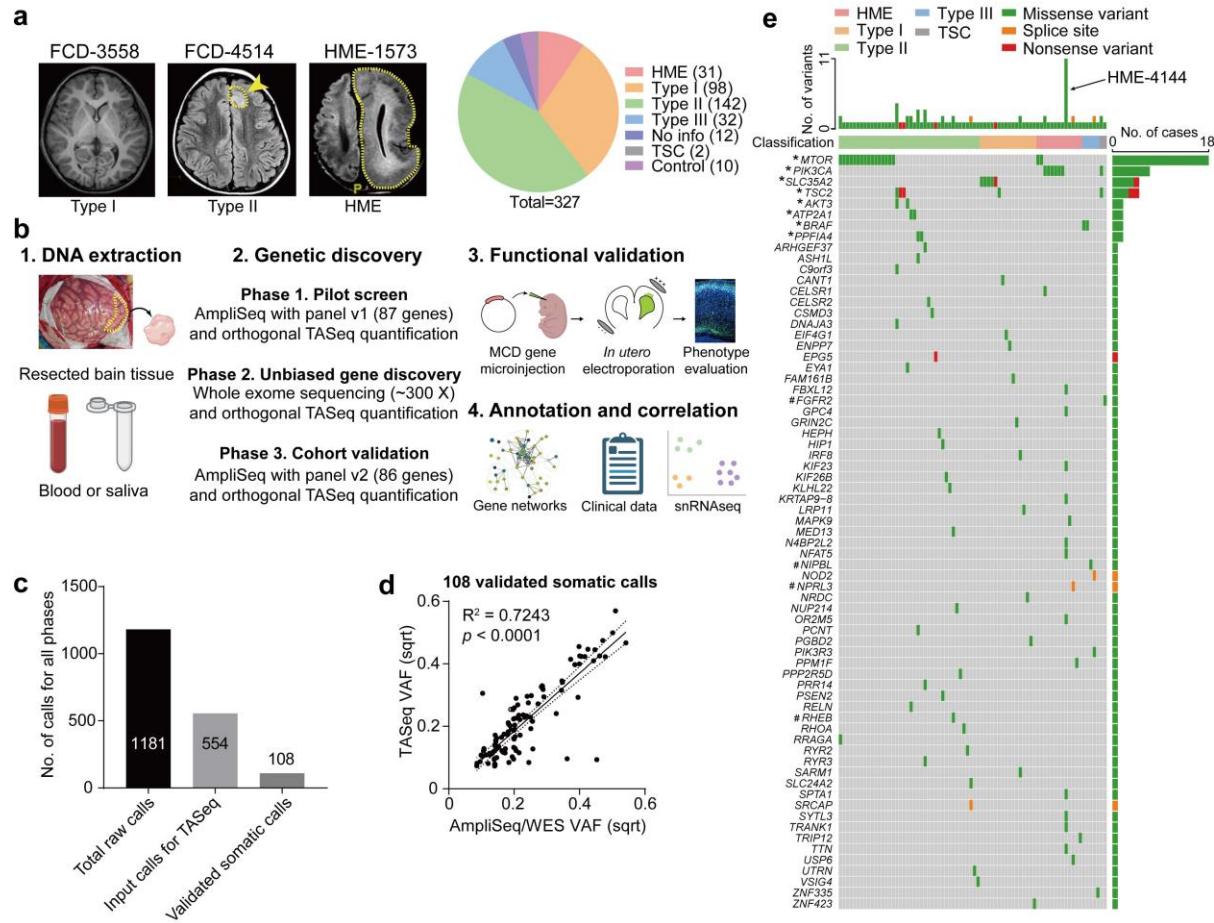
652

## 653 **Competing Interests Statement**

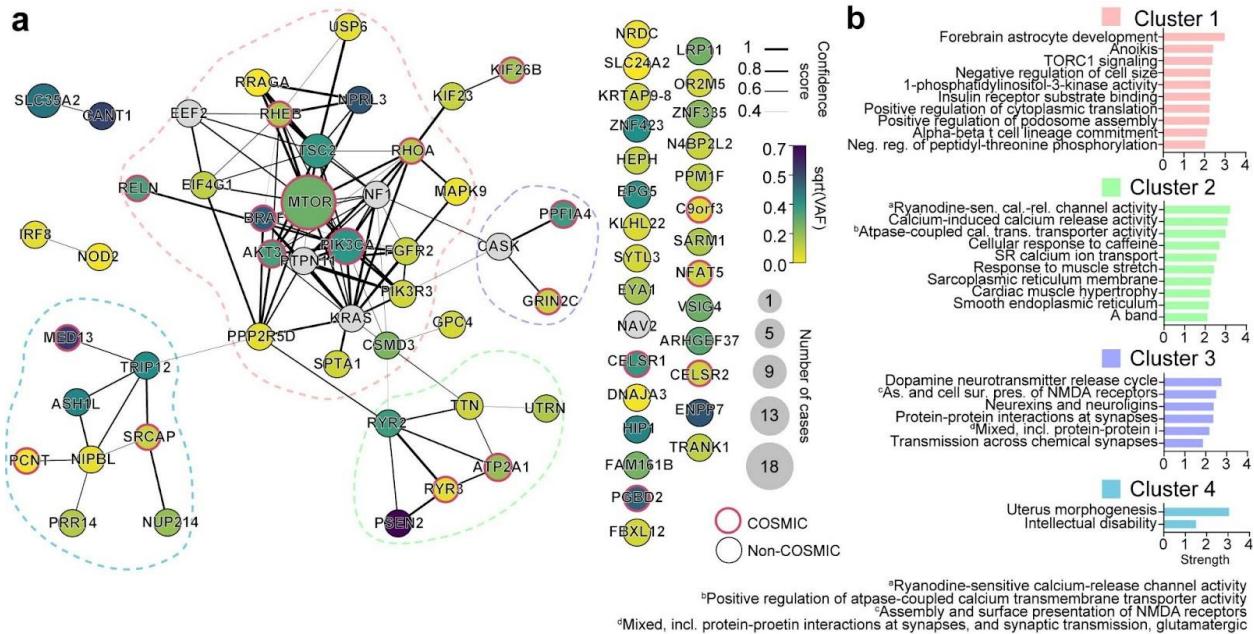
654 The authors declare no competing interests.

655

656



657 **Figure 1. Comprehensive genetic profiling and validation of somatic variants in 327 MCD**  
658 **patients.** (a) Representative MRI image of FCD-3588 (FCD type I), FCD-4514 (FCD type II),  
659 HME-1573, and a pie chart for the composition of our MCD cohort. Yellow arrow and dash:  
660 affected brain regions. (b) Workflow for comprehensive genetic profiling of MCD, using a three-  
661 phase approach from patient DNA. Each phase was followed by quantification/validation of each  
662 variant with target amplicon sequencing (TASEq). Phase 1] 1000  $\times$  pilot screening of DNA with  
663 an 87-gene mTOR-related panel. Phase 2] 300  $\times$  whole-exome sequencing (WES) with best-  
664 practice somatic variant discovery for novel candidate discovery. Phase 3] Cohort-level  
665 validation with an updated, high-confidence TASEq gene set based on knowledge from Phase 1  
666 and 2. A subset of the somatic mutations was further functionally validated by mouse modeling.  
667 Candidate genes were annotated and correlated with external datasets such as STRING DB,  
668 clinical phenotype dataset, and newly generated single-nucleus RNAseq dataset from MCD  
669 brain. (c) Somatic variant calls were detected from all three phases of genetic discovery, yielding  
670 108 validated somatic calls. (d) Correlation between square-root transformed (sqrt)  
671 AmpliSeq/WES variant allele fraction (VAF) and TASEq VAF. Solid line: best-fit line linear  
672 regression. Dotted lines: 95 % confidence band of the best-fit line. (e) Oncoplot with all 69  
673 validated somatic SNVs from this study. Top: most patients had one gene mutated, a few patients  
674 had more than one gene mutated, and patient HME-4144 had 11 different validated gene  
675 mutations. Color: type of variant. \* and #: recurrent genes in our cohort, and non-recurrent in our  
676 cohort but recurrent in other studies, respectively.



678  
679 **Figure 2. Four major gene networks were discovered from the comprehensive MCD gene**  
680 **profiling.** (a) STRING DB pathway analysis of the 69 MCD discovered genes and six novel

681 genes from recent publications identifies MTOR/MAP kinase pathway (pink, Cluster 1),

682 Calcium dynamics (green, Cluster 2), Synapse (purple, Cluster 3), Gene expression (blue,

683 Cluster 4). Edge thickness: confidence score calculated by STRING. Size and color of a node:

684 square root transformed (sqrt) number of patients carrying a given mutation and average sqrt

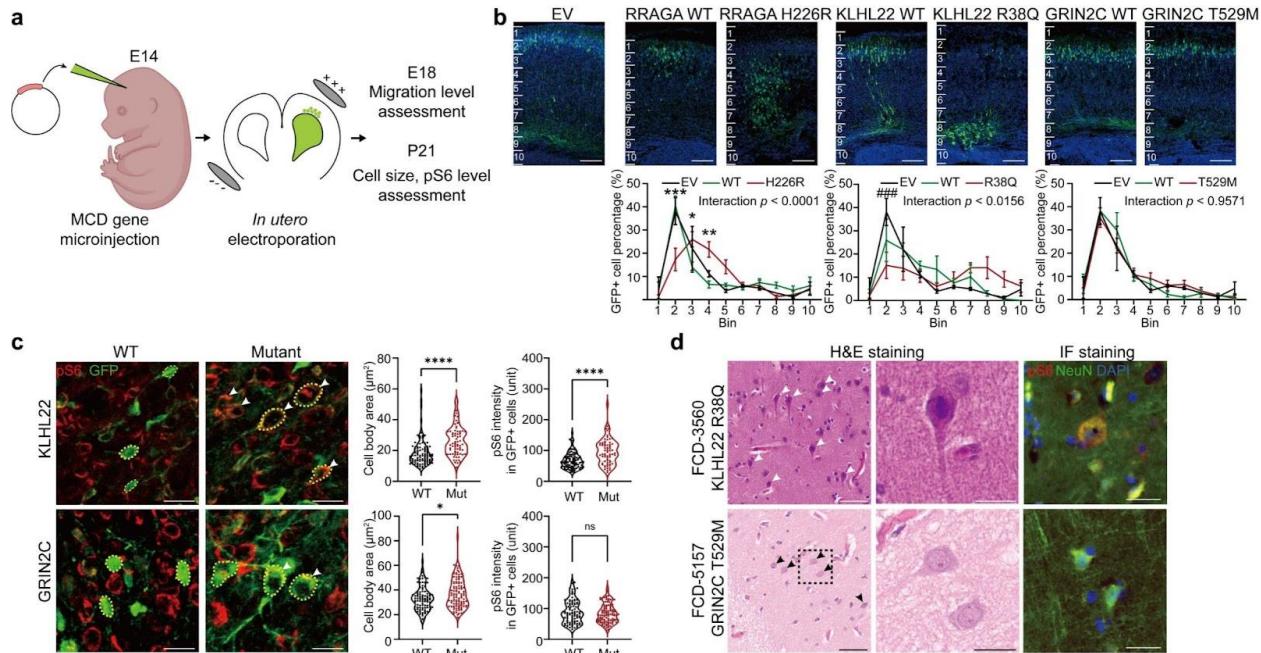
685 VAF across all patients, respectively. Non-clustered orphan genes are listed on the right.

686 Red border: variant reported in the COSMIC database. (b) Gene Ontology (GO) analysis results

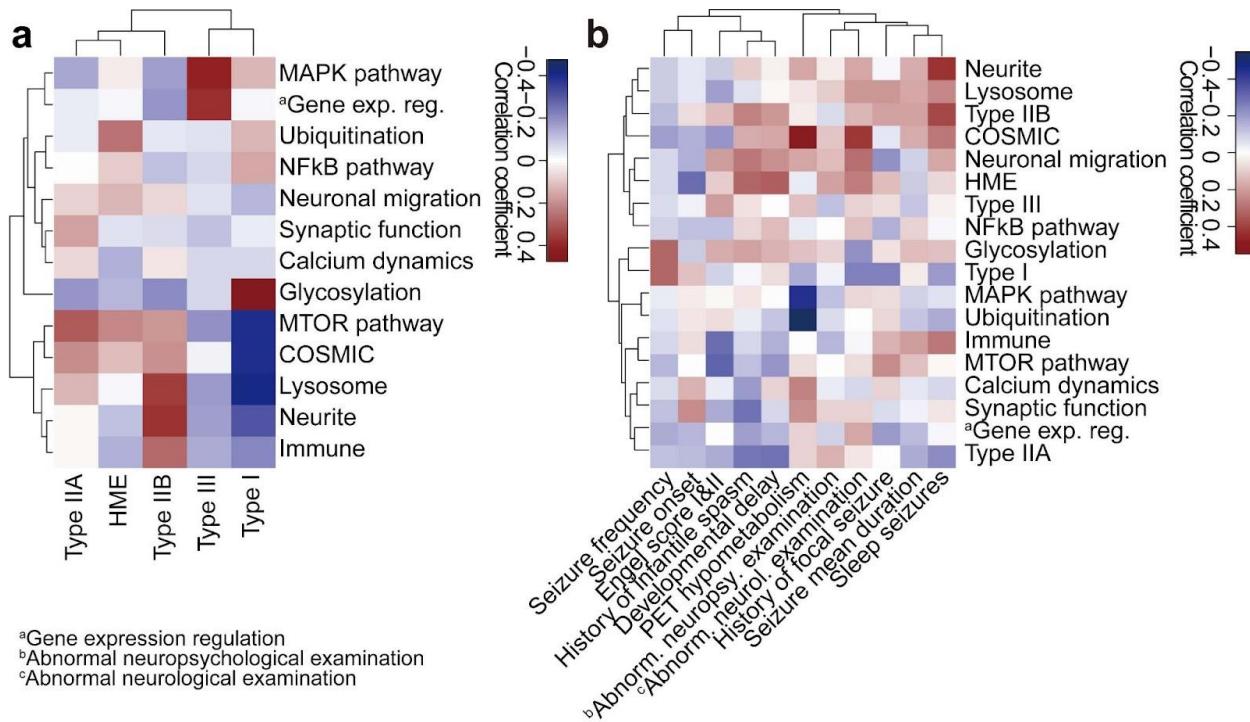
687 confirmed the functions of compositions in each network. Top GO terms or KEGG pathways.

688 Strength calculated by STRING.

689

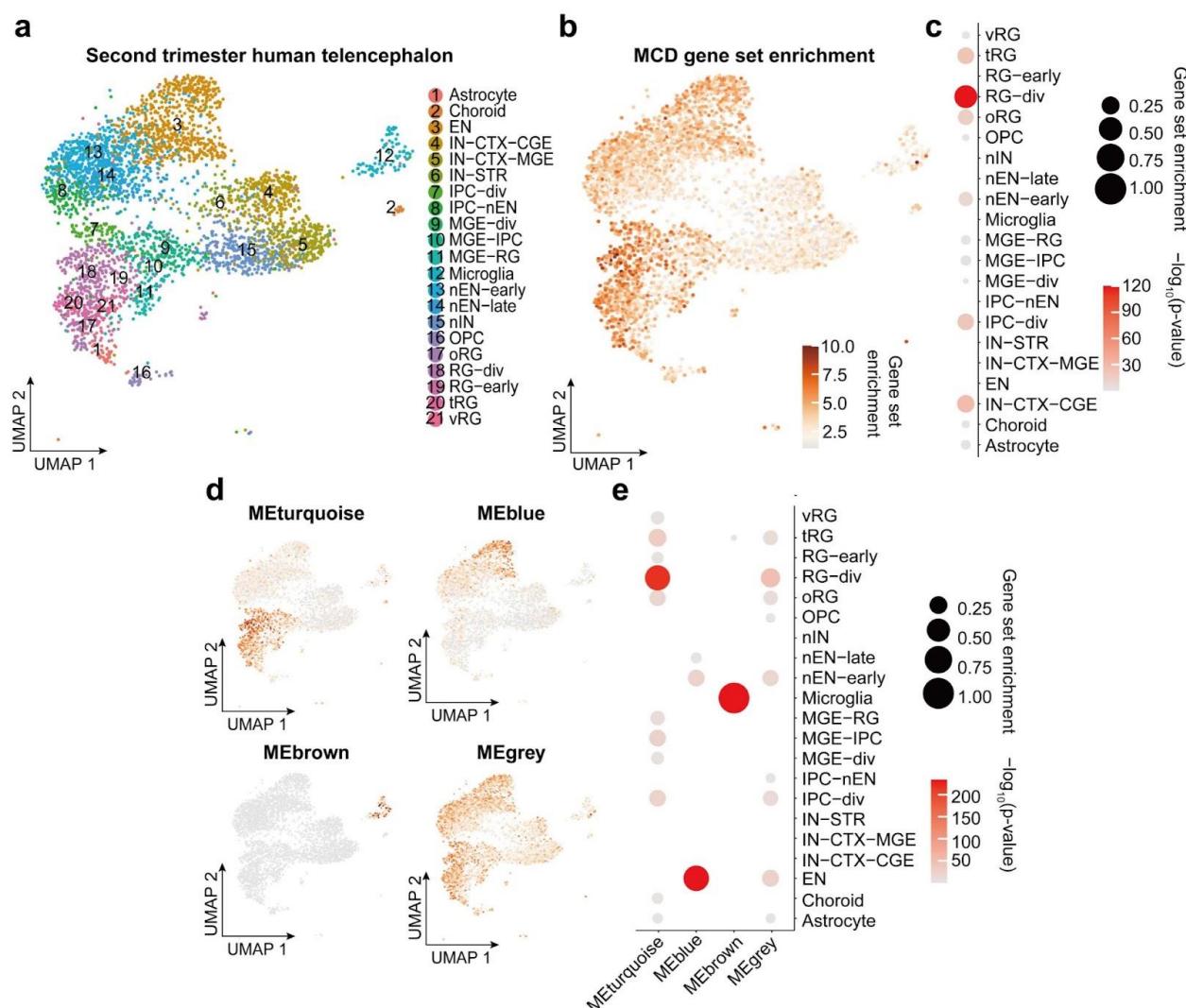


690  
691 **Figure 3. Selected novel MCD somatic variants show functional defects in embryonic**  
692 **mouse brain and patient samples.** (a) Workflow for functional validation of candidate mosaic  
693 variants in mice. (b) Two different mutations in novel FCD type II genes, *RRAGA* H226R and  
694 *KLHL22* R38Q, but not a novel FCD type I gene, *GRIN2C*, disrupt cellular radial migration from  
695 the subventricular zone (SVZ). Below: two-way ANOVA and Sidak multiple comparisons with  
696 *p*-values of interaction between genotype and bin factor. \* or # indicates a *p*-value in comparison  
697 between WT and mutant group, or EV and mutant group respectively. Ten bins from the surface  
698 of the cortex (top) to SVZ (bottom). Scale bar: 100  $\mu\text{m}$ . Error bar:  $\pm\text{SE}$ . (c) Immunofluorescence  
699 in postnatal day 21 mouse cortices for *KLHL22* and *GRIN2C* wild-type (WT) or mutant isoform.  
700 Neurons expressing mutant *KLHL22* and *GRIN2C* recapitulate histological phenotypes shown in  
701 (d), with enlarged cell bodies (white arrow) compared to WT isoforms (WT control), whereas  
702 only neurons expressing *KLHL22* but not *GRIN2C* mutant isoform display increased pS6 levels  
703 compared to control. Dotted lines: examples of cell body size quantification. Two-sided  
704 Student's *t*-test. Scale bar: 20  $\mu\text{m}$ . (d) H&E and phospho-S6 (pS6) staining of the resected brain  
705 from FCD-3560 and FCD-5157. Box area is zoomed in the middle image. Arrows: dysplastic  
706 cells. Right: Immunofluorescence (IF) for pS6 and NeuN. Note dysplastic pS6-positive neurons  
707 with increased pS6 levels are present in FCD-3560 but not in FCD-5157. Scale bar: 60  $\mu\text{m}$  on the  
708 left, 20  $\mu\text{m}$  on the middle and right. \*\*\*p < 0.0001; \*p < 0.05; ns, non-significant. ###p <  
709 0.001. EV: empty vector.



710  
711 **Figure 4. Clinical phenotypic outcomes correlate with genotype-based classifications in**  
712 **MCD. (a)** Correlation heatmap for classification based on genetic information (y-axis) vs.  
713 International League Against Epilepsy (ILAE) classification based on histology (x-axis) using  
714 Pearson correlation. Shade: the value of Phi coefficient. Note Type IIA and HME are enriched  
715 with mTOR and Ubiquitination genes, while Type I is enriched in Glycosylation and depleted in  
716 MTOR and COSMIC genes. HME: hemimegalencephaly. **(b)** Correlation between classification  
717 based on genetic information and various clinical phenotypes. Shade: the value of Phi (binary  
718 data) or Pearson (continuous) correlation coefficient. For example, positron emission  
719 tomography (PET) hypometabolism is enriched in COSMIC genes and depleted in the MAPK  
720 pathway, whereas abnormal neurological examination is enriched in COSMIC genes. The whole  
721 dataset is in Supplementary Table 4.

722



723

724

725

726

727

728

729

730

731

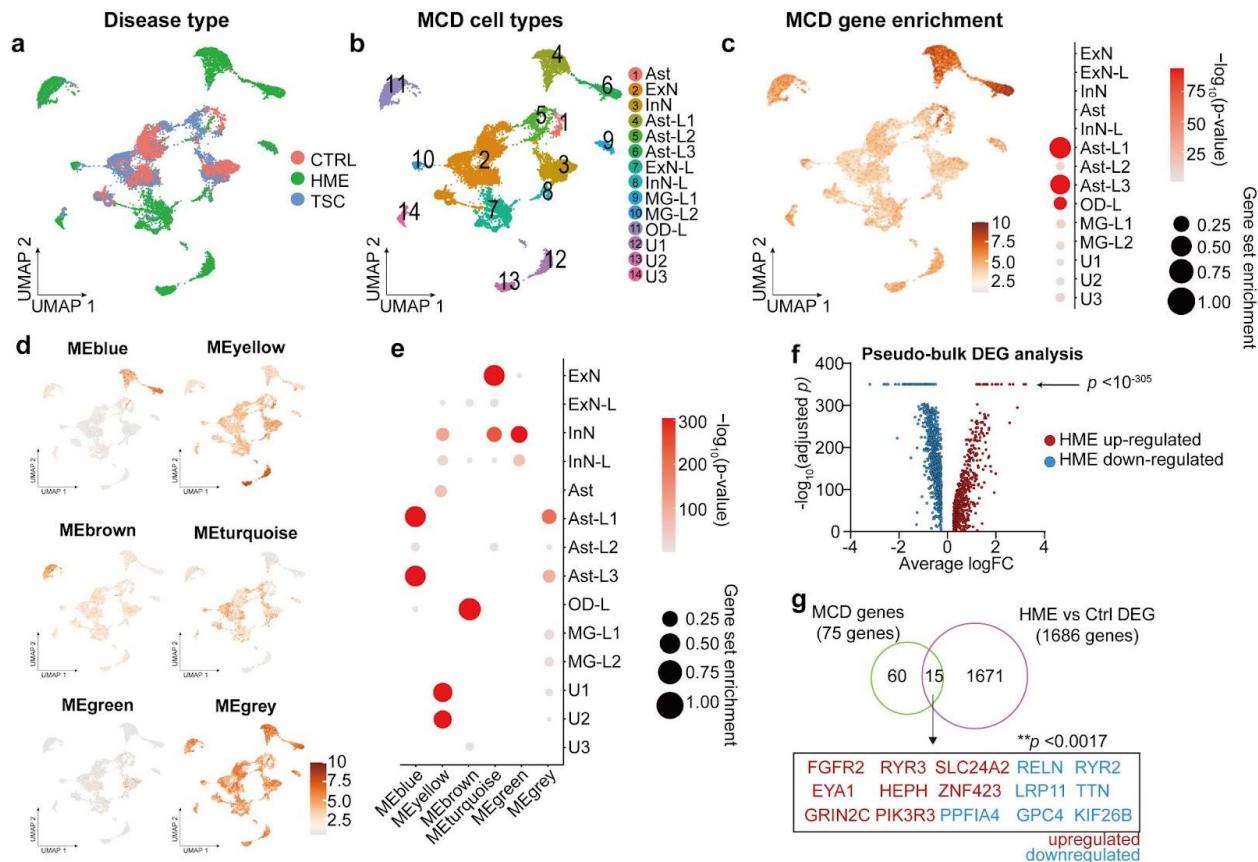
732

733

734

735

**Figure 5. Single-nucleus transcriptomes reveal MCD gene enrichment in radial glia and excitatory neurons in the developing human cortex.** (a) Uniform Manifold Approximation and Projection (UMAP) for single-nucleus transcriptome in 2nd-trimester fetal human telencephalon from a public dataset<sup>41</sup>. (b) UMAP enrichment patterns of an eigengene using MCD genes. Note enrichment for excitatory neurons and radial glia (dark brown). vRG: vertical radial glia, tRG: truncated radial glia, RG-div: dividing radial glia, oRG: outer radial glia, EN: excitatory neuron, nEN: newborn excitatory neuron, IPC: intermediate progenitor cell, STR: striatum, IN: interneuron, CTX: cortex, MGE: medial ganglionic eminence, CGE: central ganglionic eminence. (c) Quantification of enrichment of (b) based on cell types, showing enrichment for RG-div. (d) Four eigengenes decomposed from (b). (e) Quantification of enrichment of (d) based on cell types showing enrichment in dividing radial glia, microglia, and inhibitory cortical neurons from the medial ganglionic eminence (MGE).



**Figure 6. Single-nucleus transcriptomes showed MCD gene expression enriched in MCD-specific cell types.** (a) UMAP for the single-nucleus transcriptome of 22067 nuclei from the cortical lesions of control (CTRL), hemimegalencephaly (HME), and tuberous sclerosis complex (TSC) brain. (b) Cell type classification. Ast: astrocyte, ExN: excitatory neuron, InN: inhibitory neuron, MG: microglia, OD: oligodendrocyte, U: unidentified. (c) The expression pattern of an eigengene made with all MCD genes and the quantification of enrichment based on cell types. (d) Six eigengenes decomposed from (c). (e) Quantification of the cell-type-specific enrichment in (d). (f) A volcano plot from DEG list of HME versus CTRL pseudo-bulk data. The genes having adjusted  $p < 10^{-305}$  were pointed by the arrow. (g) The MCD genes overlap with DEGs of HME in contrast to controls. A permutation test (10,000 times) shows a very rare chance ( $p < 0.0017$ ) to show this overlap in a random sampling of 1686 genes from 19909 protein-coding genes used in these DEGs. Red or blue coloring of gene names indicates upregulated or downregulated DEGs in HMEs compared to CTRLs, respectively.

754 **Supplementary Table Descriptions**

755 **Supplementary Table 1. The cohort list and corresponding sequencing methods.** The 327  
756 cases are listed in each row and corresponding sequencing methods used for a given sample were  
757 described.

758

759 **Supplementary Table 2. AmpliSeq primer pool designs** (a) Ampliseq primer pool design used  
760 in phase 1. (b) Ampliseq primer pool design used in phase 3.

761

762 **Supplementary Table 3. The summary of SNV calls across the three phases of genetic**  
763 **discovery.** (a) 1811 raw calls derived from the combination of variant callers described in  
764 Extended Data Fig. 1. (b) 554 input SNV calls participated in TASEq quantification. (c)  
765 Validated brain somatic SNV calls from (b). (d) Annotation table of the genes listed in (c) based  
766 on GO terms.

767

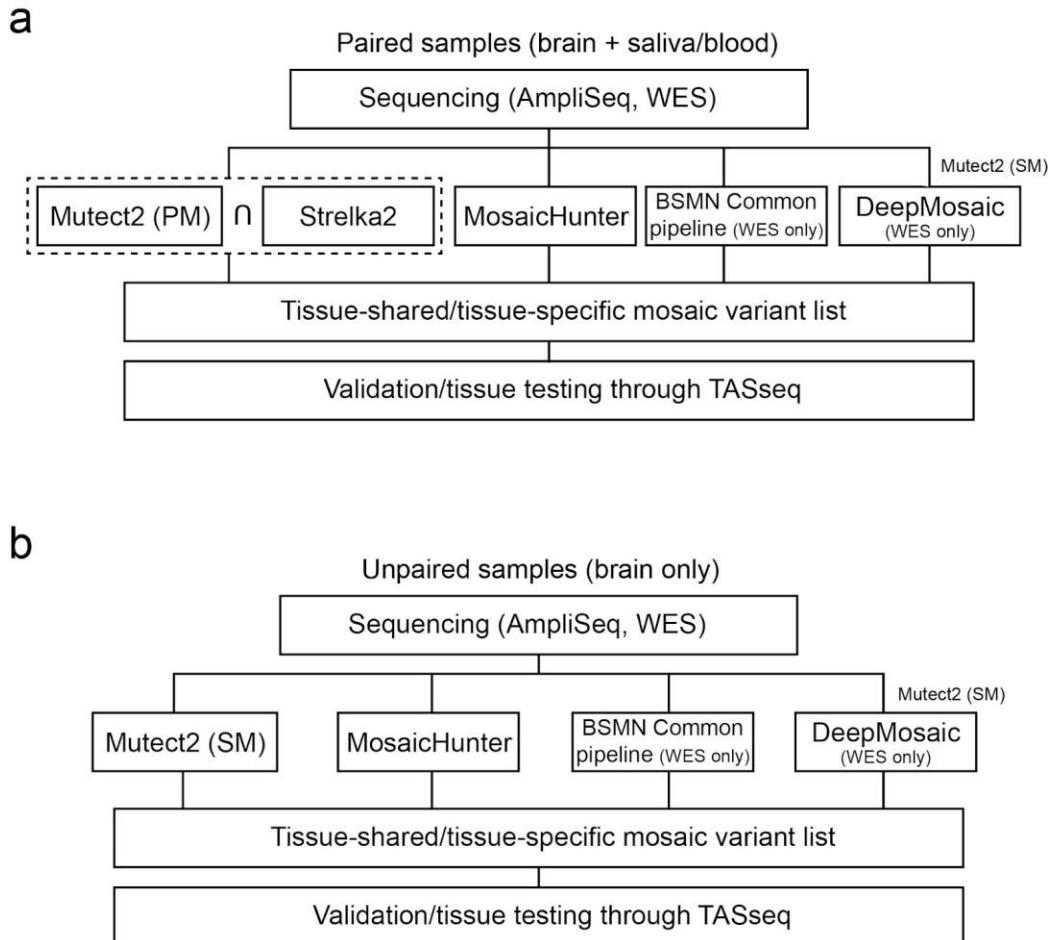
768 **Supplementary Table 4. The summary of phenotype and genotype information for the**  
769 **‘genetically solved’ cases.**

770

771 **Supplementary Table 5. The summary table used for false discovery estimation.**

772

773



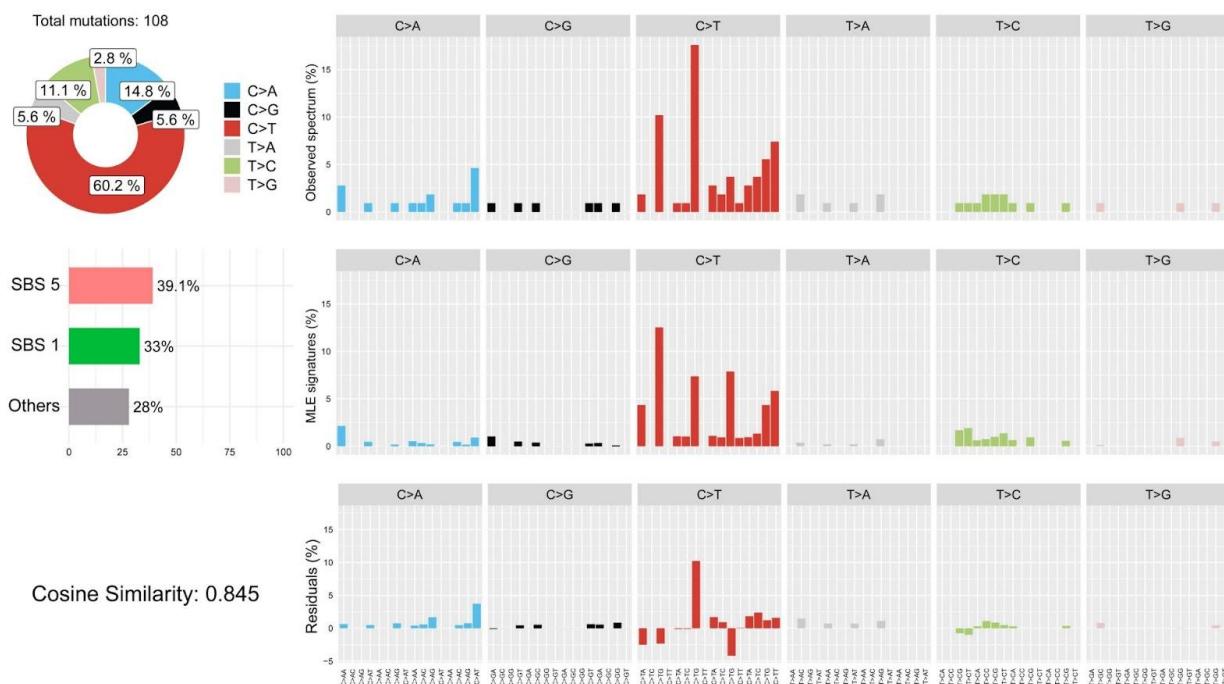
774

775

776 **Extended Data Fig. 1 Bioinformatic pipeline to detect somatic SNVs in the MCD cohort. (a)**  
777 The pipeline for paired samples. Notably, the dashed square indicates that the sharing variants  
778 between MuTect2 paired mode and Strelka2 were used for the downstream analysis. BSMN  
779 common pipeline and DeepMosaic were used only for WES datasets. The DeepMosaic input  
780 variants were generated by MuTect2 single mode. (b) The pipeline for unpaired samples. The  
781 pipeline is similar except that MuTect2 single mode without Strelka2 is used. PM: paired mode,  
782 SM: single mode.

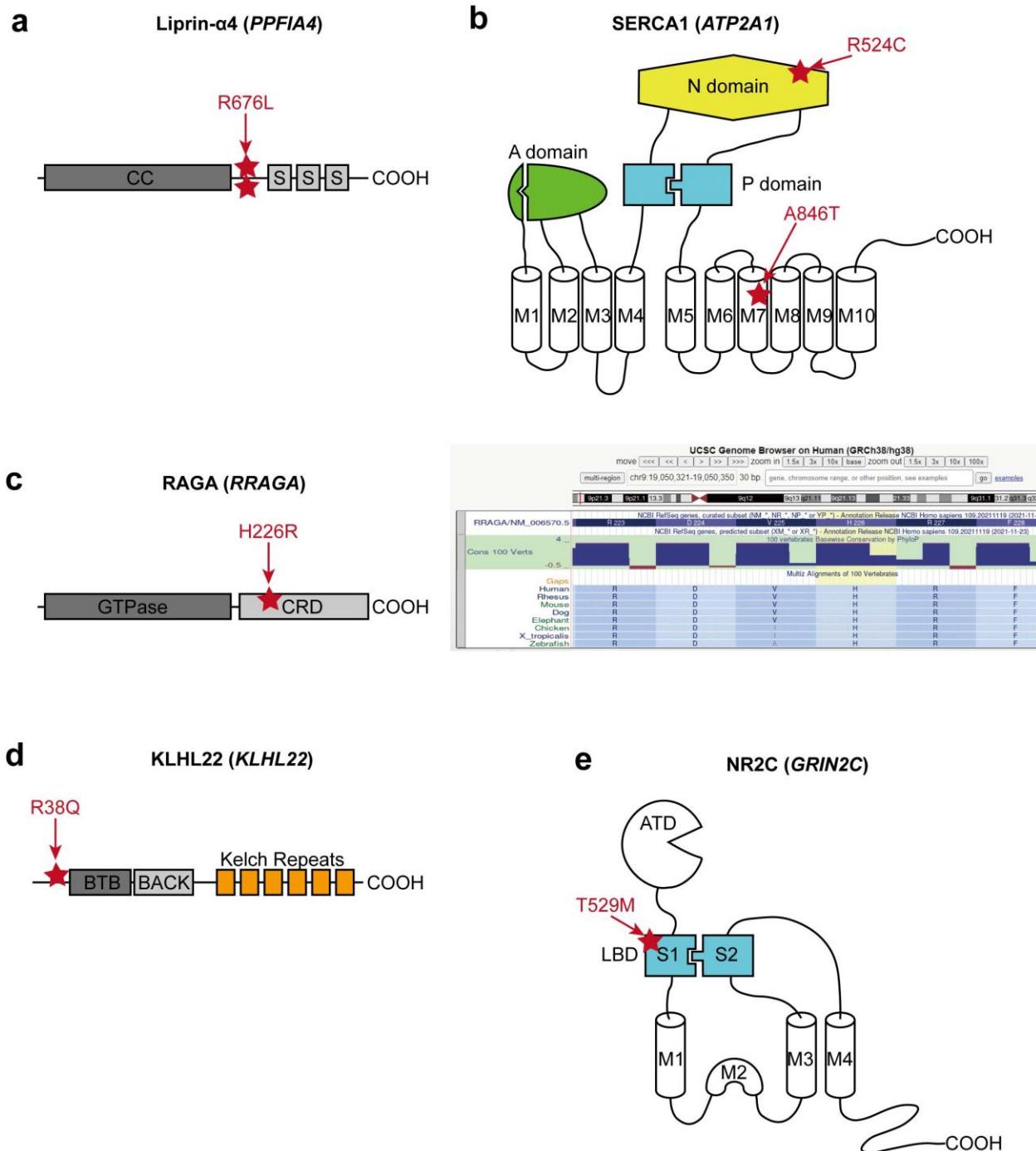
783

784



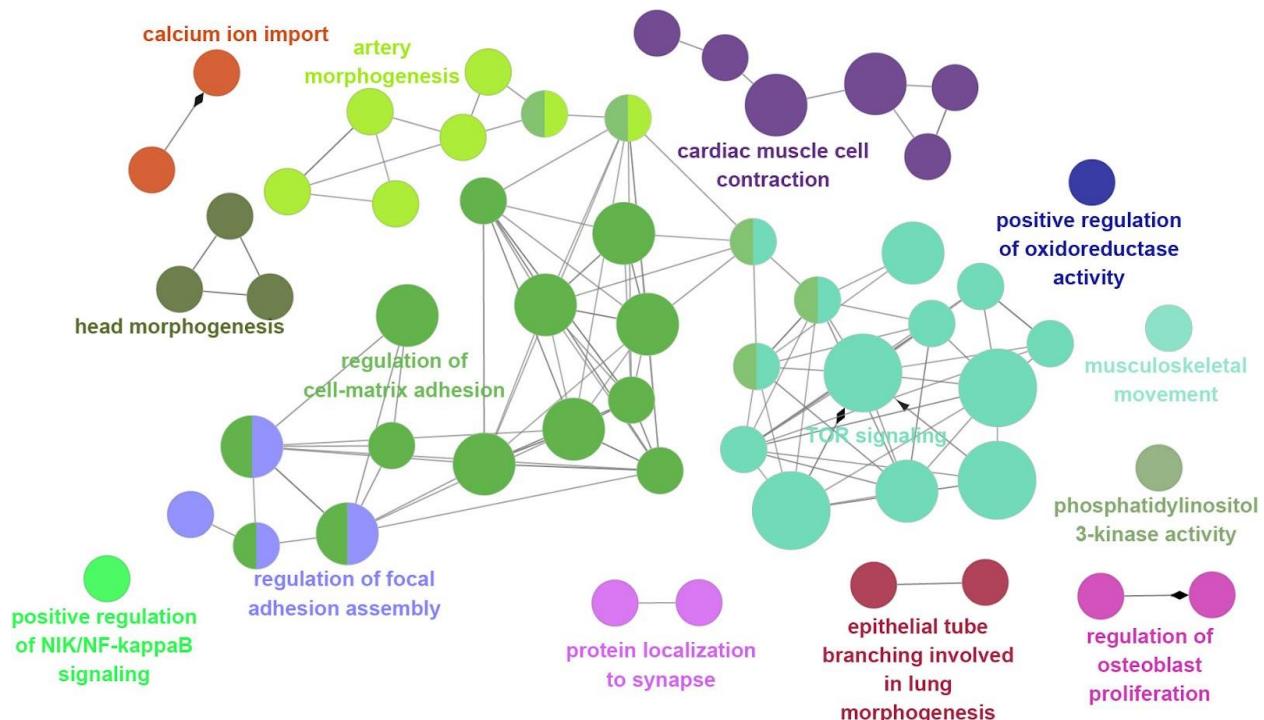
785  
786 **Extended Data Fig. 2 Mutational signature analysis through Mutalisk shows cell-division-  
787 related clock-like signatures in the MCD cohort.** SBS5 (39.1%) and SBS1 (33%) are clock-  
788 like mutational signatures. SBS1 especially correlates with cell division and mitosis of stem  
789 cells.

790  
791

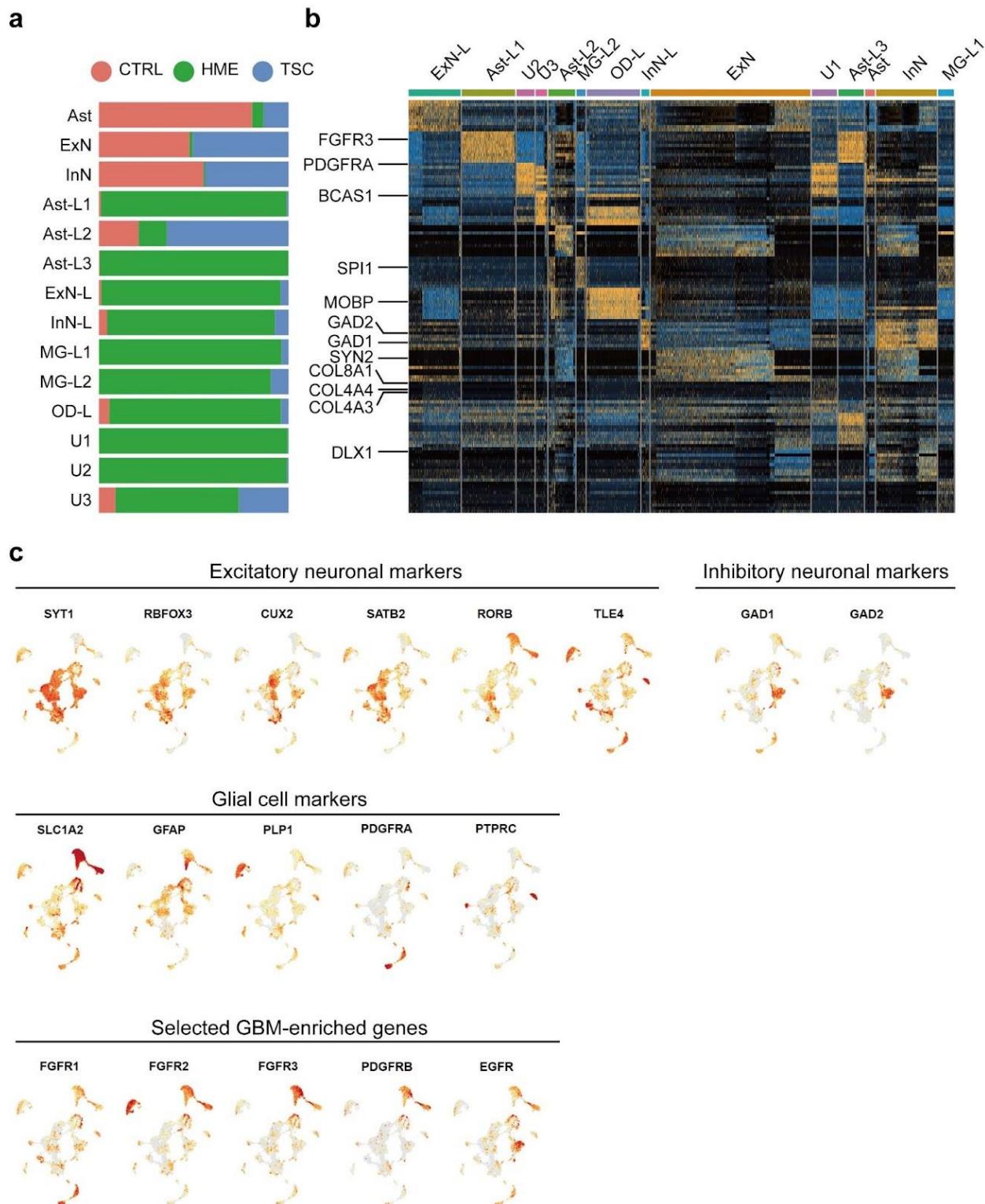


792  
 793 **Extended Data Fig. 3 The locations of the selected MCD variants.** (a) The location of two  
 794 recurrent variant calls is at the same position between the coiled-coil domain (CC) and the first  
 795 SAM domain (S) of Liprin- $\alpha$ 4. (b) Two different variants in SERCA1. p.R524C mutation is at  
 796 the nucleotide ATP-binding (N) domain, whereas the p.A846T variant is in the 7th  
 797 transmembrane (M7) domain. A: Actuator domain, P: Phosphorylation domain, M:  
 798 Transmembrane domain. (c) Left: The location of p.H226R variant in RAGA protein. GTPase:  
 799 GTPase domain, CRD: C-terminal roadblock domain. Right: UCSC genome browser screenshot

800 describing that p.H226 is a conserved site across all vertebrates. (d) The location of p.R38Q  
801 variant in the N-terminal region before BTB (Broad-Complex, Tramtrack, and Bric-à-brac)  
802 domain of KLHL22. (e) A variant in the S1 domain of NR2C. S1 and S2 together make the  
803 ligand-binding domain (LBD), the target of glutamate. ATD: Amino terminal domain.  
804

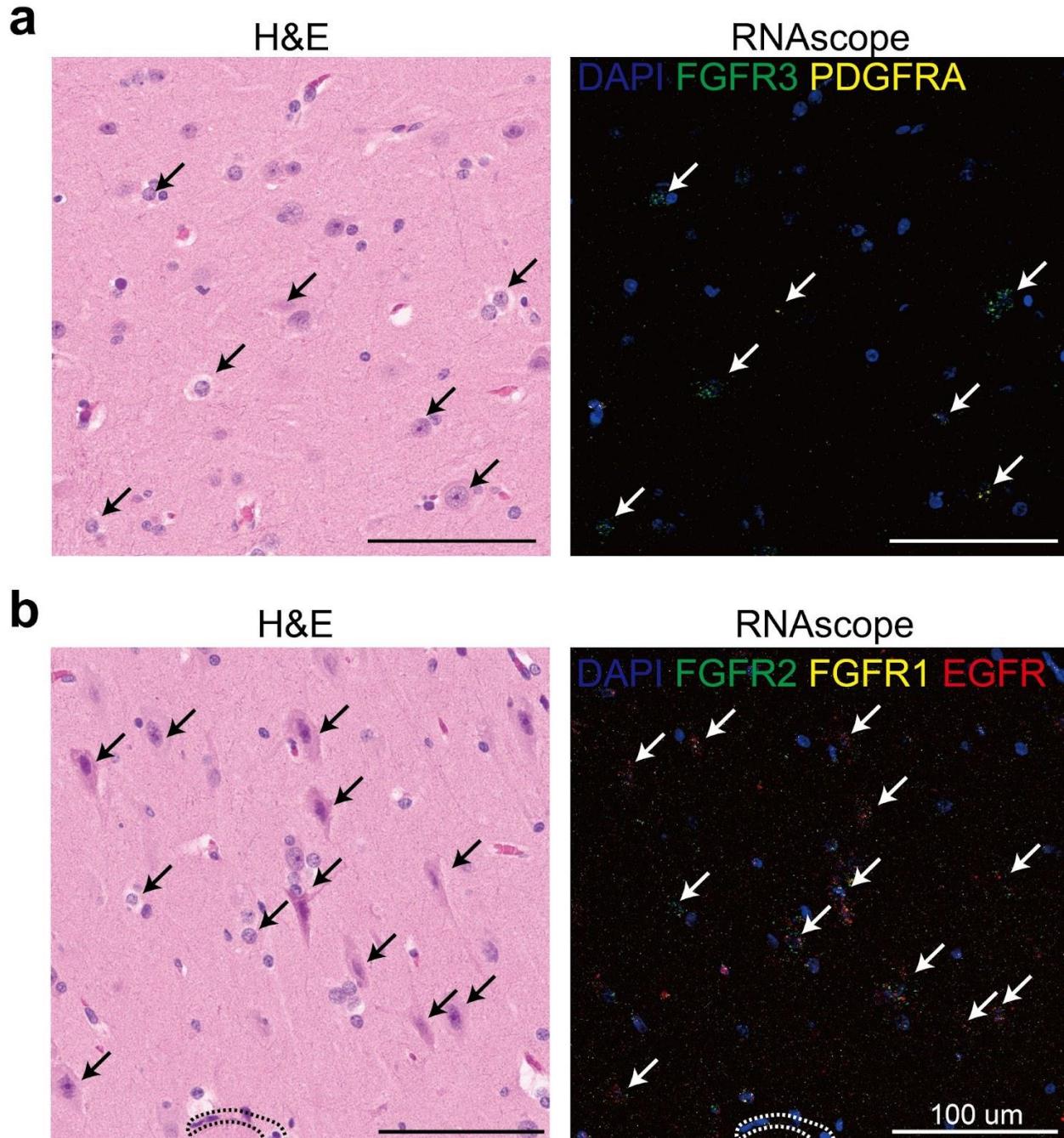


805  
806 **Extended Data Fig. 4 The ClueGO analysis using the MCD genes result identifies the**  
807 **biological processes and molecular pathways.** The main cluster is related to TOR signaling,  
808 regulation of cell-matrix adhesion, regulation of focal adhesion assembly, and artery  
809 morphogenesis. Notably, there are also isolated clusters that were not covered in previous  
810 studies, for example, cardiac muscle cell contraction, calcium ion import, and protein  
811 localization to the synapse. Term p-value with Bonferroni correction was reflected in node size  
812 (Large:  $p < 0.0005$ , medium:  $p < 0.05$ , small:  $p < 0.1$ ).



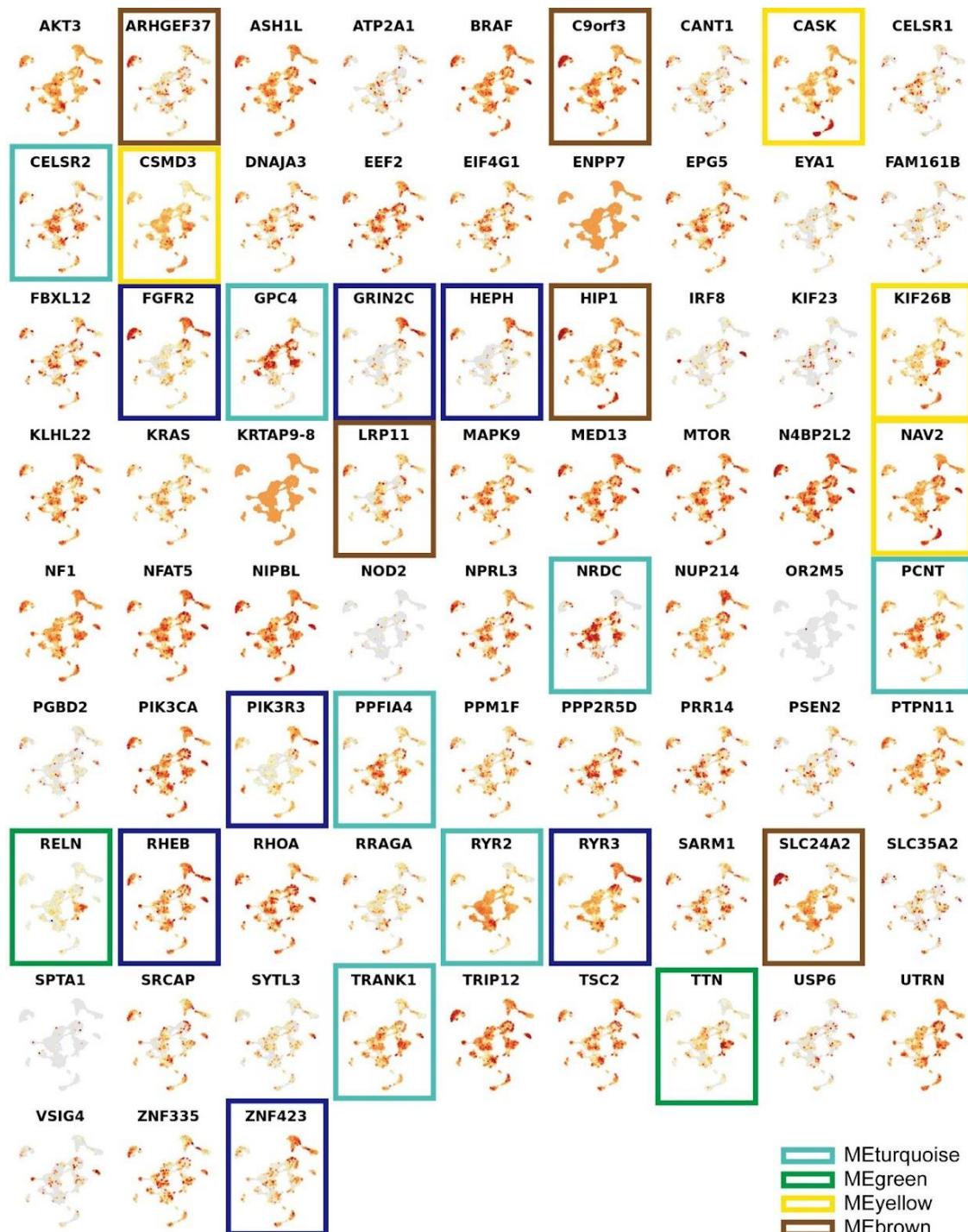
813  
814 **Extended Data Fig. 5 Cell-type identification by DEGs and known marker gene expression**  
815 **in the MCD snRNAseq dataset.** (a) MCD prefix was used for the clusters that have less than  
816 25 % of control origin. (b) DEG analysis using FindAllMarker function in Seurat v4 package.  
817 The top 10 genes for each cluster were presented. Several notable genes helping to define major

818 cell types were labeled on the left side. Note that *FGFR3* and *PDGFRA* are up-regulated in Ast-  
819 L1/3 and U1/2/3, respectively, implying that these genes can be the markers for MCD-dominant  
820 clusters. (c) Selected markers for major cell types in the human cortex. *CUX1*, *CUX2* for upper  
821 layer excitatory neuronal markers, *SATB2* for layer 4 excitatory neuronal marker, *RORB*,  
822 *FEZF2*, *BCL11B*, *FOXP2*, *ROBO2* for deep layer-specific markers, *GAD1*, *GAD2*, *DLX6*, *RELN*  
823 for inhibitory neuronal markers, *GFAP*, *SLC1A2*, *SLC1A3*, *MMD2* for astrocyte markers, *PTPRC*  
824 for the microglial marker, *OLIG1*, *OLIG2*, *MOBP*, *PLP1* for oligodendrocyte markers, *FGFR1*,  
825 *FGFR2*, *FGFR3*, *PDGFRB*, *EGFR* for the selected GBM-enriched genes covering subsets of  
826 MCD-enriched clusters.  
827



828  
829  
830  
831  
832  
833  
834  
835

**Extended Data Fig. 6 The validation of the snRNAseq result from HME-6593 shows MCD dominant clusters are highly correlated with dysplastic cells in MCD.** (a) H&E staining (left) and RNAscope (right) staining results in several MCD-dominant markers (*FGFR2*, *FGFR1*, *EGFR*) in the same formaldehyde-embedded-paraffin-fixed section. (b) H&E and RNAscope result in another section with different RNA probes (*FGFR3* and *PDGFR $\alpha$* ) enriched in MCD clusters. Dashed lines indicate blood vessels. White/black arrows are pointing to the dysplastic cells.



836  
837  
838  
839  
840

**Extended Data Fig. 7 Expression patterns of individual MCD genes in the MCD snRNAseq dataset.** The gene members of each eigen module shown in Fig. 6d were colored according to the name of a given eigengene.

841 **Focal Cortical Dysplasia Neurogenetics Consortium (Additional Members)**

842

843 Dr. Yasemin Alanay, Division of Pediatric Genetics, Acibadem Hospital, Istanbul, Turkey

844 Dr. Seema Kapoor, Division of Genetics, Genetic & Metabolic Lab, Lok Nayak Hospital &  
845 Maulana Azad Medical Center, Pakistan

846 Dr. Georgia Ramantani, Dr. Thomas Feuerstein, Albert-Ludwigs University, Freiburg, Germany

847 Dr. Ingmar Blumcke, Dr. Robyn Busch, Dr. Zhong Ying, Department of Neuropathology,  
848 University Hospital Erlangen, Germany

849 Dr. Vadym Biloshytsky, Dr. Kostiantyn Kostiuk, Dr. Eugene Pedachenko, A. Romodanov  
850 Institute of Neurosurgery, Kyiv, Ukraine

851 Dr. Marilyn Jones, Diane Masser-Frye, Rady Children's Hospital, San Diego, CA

852 Dr. Ingo Helbig, Dr. Benjamin C. Kennedy, Division of Neurology, Children's Hospital  
853 Philadelphia, PA

854 Dr. Judy Liu, Dr. Felix Chan, Department of Molecular Biology, Cell Biology, and Biochemistry,  
855 Department of Neurology, Brown University, RI

856 Dr. Darcy Krueger, Department of Clinical Pediatrics and Neurology, Cincinnati Children's  
857 Hospital, OH

858 Dr. Richard Frye, Dr. Angus Wilfong, Dr. David Adelson, Barrow Neurological Institute at  
859 Phoenix Children's Hospital, U Arizona College of Medicine, Phoenix, AZ

860 Dr. William Gaillard, Dr. Chima Oluigbo, Children's National Hospital, Washington DC

861 Dr. Anne Anderson, Dept of Pediatrics, Baylor College of Medicine, Texas Children's Hospital,  
862 Houston, TX

863

864 [gmathern@ucla.edu](mailto:gmathern@ucla.edu), [jwchang@mednet.ucla.edu](mailto:jwchang@mednet.ucla.edu), [renzo.guerrini@meyer.it](mailto:renzo.guerrini@meyer.it), [yalanay@gmail.com](mailto:yalanay@gmail.com),  
865 [drseemakapoor@gmail.com](mailto:drseemakapoor@gmail.com), [Thomas.feuerstein@uniklinik-freiburg.de](mailto:Thomas.feuerstein@uniklinik-freiburg.de),

866 [georgia.ramantani@kispi.uzh.ch](mailto:georgia.ramantani@kispi.uzh.ch), [hipo0207@yuhs.ac](mailto:hipo0207@yuhs.ac), [carola.haas@uniklinik-freiburg.de](mailto:carola.haas@uniklinik-freiburg.de),

867 [catharina.donkels@uniklinik-freiburg.de](mailto:catharina.donkels@uniklinik-freiburg.de), [takahashi-ped@umin.ac.jp](mailto:takahashi-ped@umin.ac.jp), [hrmachad@gmail.com](mailto:hrmachad@gmail.com),

868 [camila.neurociencias@gmail.com](mailto:camila.neurociencias@gmail.com), [wilsonjr@usp.br](mailto:wilsonjr@usp.br), [gurnettc@neuro.wustl.edu](mailto:gurnettc@neuro.wustl.edu),

869 [gardnerc@wustl.edu](mailto:gardnerc@wustl.edu)", [ingmar.bluemcke@uk-erlangen.de](mailto:ingmar.bluemcke@uk-erlangen.de), [vabil@i.ua](mailto:vabil@i.ua), [ssattar@health.ucsd.edu](mailto:ssattar@health.ucsd.edu),

870 [dgonda@rchsd.org](mailto:dgonda@rchsd.org), [mnespeca@rchsd.org](mailto:mnespeca@rchsd.org), [mjone@rchsd.org](mailto:mjone@rchsd.org), [dmasser-frye@rchsd.org](mailto:dmasser-frye@rchsd.org),

871 [roberthhchen3@gmail.com](mailto:roberthhchen3@gmail.com), [jinwu.tsai@gmail.com](mailto:jinwu.tsai@gmail.com), [HELBIGI@chop.edu](mailto:HELBIGI@chop.edu),

872 [KENNEDYBC@chop.edu](mailto:KENNEDYBC@chop.edu), [judy\\_liu@brown.edu](mailto:judy_liu@brown.edu), [felix\\_chan@brown.edu](mailto:felix_chan@brown.edu),

873 [darcy.kreuger@cchmc.org](mailto:darcy.kreuger@cchmc.org), [r frye@phoenixchildrens.com](mailto:r frye@phoenixchildrens.com), [a wilfong@phoenixchildrens.com](mailto:a wilfong@phoenixchildrens.com),

874 [dadelson@phoenixchildrens.com](mailto:dadelson@phoenixchildrens.com), [WGAILLAR@childrensnational.org](mailto:WGAILLAR@childrensnational.org),

875 [COluigbo@childrensnational.org](mailto:COluigbo@childrensnational.org), [annea@bcm.edu](mailto:annea@bcm.edu),

876

877 **Brain Somatic Mosaicism Network**

878 Boston Children's Hospital: Alice Lee, August Yue Huang, Alissa D'Gama, Caroline Dias,  
879 Christopher A. Walsh, Eduardo Maury, Javier Ganz, Michael Lodato, Michael Miller, Pengpeng  
880 Li, Rachel Rodin, Rebeca Borges-Monroy, Robert Hill, Sara Bizzotto, Sattar Khoshkhoo, Sonia  
881 Kim, Zinan Zhou

882 Harvard University: Alice Lee, Alison Barton, Alon Galor, Chong Chu, Craig Bohrson, Doga  
883 Gulhan, Eduardo Maury, Elaine Lim, Euncheon Lim, Giorgio Melloni, Isidro Cortes, Jake Lee,  
884 Joe Luquette, Lixing Yang, Maxwell Sherman, Michael Coulter, Minseok Kwon, Peter J. Park,  
885 Rebeca Borges-Monroy, Semin Lee, Sonia Kim, Soo Lee, Vinary Viswanadham, Yanmei Dou

886 Icahn School of Medicine at Mt. Sinai: Andrew J. Chess, Attila Jones, Chaggai Rosenbluh,  
887 Schahram Akbarian

888 Kennedy Krieger Institute: Ben Langmead, Jeremy Thorpe, Sean Cho

889 Lieber Institute for Brain Development: Andrew Jaffe, Apua Paquola, Daniel Weinberger,  
890 Jennifer Erwin, Jooheon Shin, Michael McConnell, Richard Straub, Rujuta Narurkar

891 Mayo Clinic: Alexej Abyzov, Taejeong Bae, Yeongjun Jang, Yifan Wang

892 NIMH: Anjene Addington, Geetha Senthil

893 Sage Bionetworks: Cindy Molitor, Mette Peters

894 Salk Institute for Biological Studies: Fred H. Gage, Meiyang Wang, Patrick Reed, Sara Linker

895 Stanford University: Alexander Urban, Bo Zhou, Reenal Pattni, Xiaowei Zhu

896 Universitat Pompeu Fabra: Aitor Serres Amero, David Juan, Inna Povolotskaya, Irene Lobon,  
897 Manuel Solis Moruno, Raquel Garcia Perez, Tomas Marques-Bonet

898 University of Barcelona: Eduardo Soriano

899 University of California, Los Angeles: Gary Mathern

900 University of California, San Diego: Danny Antaki, Dan Averbuj, Eric Courchesne, Joseph G.  
901 Gleeson, Laurel L. Ball, Martin W. Breuss, Subhojit Roy, Xiaoxu Yang, Changuk Chung

902 University of Michigan: Chen Sun, Diane A. Flasch, Trenton J. Frisbie Trenton, Huiru C.  
903 Kopera, Jeffrey M. Kidd, John B. Moldovan, John V. Moran, Kenneth Y. Kwan, Ryan E. Mills,  
904 Sarah B. Emery, Weichen Zhou, Xuefang Zhao

905 University of Virginia: Aakrosh Ratan

906 Yale University: Adriana Cherskov, Alexandre Jourdon, Flora M. Vaccarino, Liana Fasching,  
907 Nenad Sestan, Sirisha Pochareddy, Soraya Scuder

908  
909  
910  
911  
912  
913  
914

[Christopher.Walsh@childrens.harvard.edu](mailto:Christopher.Walsh@childrens.harvard.edu), [peter\\_park@hms.harvard.edu](mailto:peter_park@hms.harvard.edu),  
[nenad.sestan@yale.edu](mailto:nenad.sestan@yale.edu), [gage@salk.edu](mailto:gage@salk.edu), [drweinberger@libd.org](mailto:drweinberger@libd.org), [moranj@umich.edu](mailto:moranj@umich.edu),  
[flora.vaccarino@yale.edu](mailto:flora.vaccarino@yale.edu), [abyzov.alexej@mayo.edu](mailto:abyzov.alexej@mayo.edu), [jogleeson@health.ucsd.edu](mailto:jogleeson@health.ucsd.edu),  
[gmathern@ucla.edu](mailto:gmathern@ucla.edu), [ecourchesne1949@gmail.com](mailto:ecourchesne1949@gmail.com), [s1roy@ucsd.edu](mailto:s1roy@ucsd.edu), [andrew.chess@mssm.edu](mailto:andrew.chess@mssm.edu),  
[schahram.akbarian@mssm.edu](mailto:schahram.akbarian@mssm.edu), [mette.peters@sagebase.org](mailto:mette.peters@sagebase.org), [cindy.molitor@sagebase.org](mailto:cindy.molitor@sagebase.org),

## 915 References

1. Leventer, R.J., Guerrini, R. & Dobyns, W.B. Malformations of cortical development and epilepsy. *Dialogues Clin Neurosci* **10**, 47-62 (2008).
2. Barkovich, A.J., Dobyns, W.B. & Guerrini, R. Malformations of cortical development and epilepsy. *Cold Spring Harb Perspect Med* **5**, a022392 (2015).
3. Blumcke, I. *et al.* The clinicopathologic spectrum of focal cortical dysplasias: a consensus classification proposed by an ad hoc Task Force of the ILAE Diagnostic Methods Commission. *Epilepsia* **52**, 158-74 (2011).
4. Choi, S.A. & Kim, K.J. The Surgical and Cognitive Outcomes of Focal Cortical Dysplasia. *J Korean Neurosurg Soc* **62**, 321-327 (2019).
5. Poduri, A. *et al.* Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron* **74**, 41-8 (2012).
6. Lee, J.H. *et al.* De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat Genet* **44**, 941-5 (2012).
7. Baldassari, S. *et al.* Dissecting the genetic basis of focal cortical dysplasia: a large cohort study. *Acta Neuropathol* **138**, 885-900 (2019).
8. Sim, N.S. *et al.* Precise detection of low-level somatic mutation in resected epilepsy brain tissue. *Acta Neuropathol* **138**, 901-912 (2019).
9. Dou, Y. *et al.* Accurate detection of mosaic variants in sequencing data without matched controls. *Nat Biotechnol* **38**, 314-319 (2020).
10. Yang, X. *et al.* DeepMosaic: Control-independent mosaic single nucleotide variant detection using deep convolutional neural networks. *bioRxiv* (2021).
11. Huang, A.Y. *et al.* MosaicHunter: accurate detection of postzygotic single-nucleotide mosaicism through next-generation sequencing of unpaired, trio, and paired samples. *Nucleic Acids Res* **45**, e76 (2017).
12. Wang, Y. *et al.* Comprehensive identification of somatic nucleotide variants in human brain tissue. *Genome Biol* **22**, 92 (2021).
13. Breuss, M.W. *et al.* Autism risk in offspring can be assessed through quantification of male sperm mosaicism. *Nat Med* **26**, 143-150 (2020).
14. Yang, X. *et al.* Developmental and temporal characteristics of clonal sperm mosaicism. *Cell* **184**, 4772-4783 e15 (2021).
15. Garcia, C.A.B. *et al.* mTOR pathway somatic variants and the molecular pathogenesis of hemimegalencephaly. *Epilepsia Open* **5**, 97-106 (2020).
16. Pelorosso, C. *et al.* Somatic double-hit in MTOR and RPS6 in hemimegalencephaly with intractable epilepsy. *Hum Mol Genet* **28**, 3755-3765 (2019).
17. Breuss, M.W. *et al.* Somatic mosaicism in the mature brain reveals clonal cellular distributions during cortical development. *Nature* (2022).
18. Bae, T. *et al.* Somatic mutations reveal hypermutable brains and are associated with neuropsychiatric disorders. *medRxiv* (2022).
19. Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A* **107**, 18545-50 (2010).
20. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).
21. Kim, M. & Costello, J. DNA methylation: an epigenetic mark of cellular memory. *Exp Mol Med* **49**, e322 (2017).
22. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613 (2019).

959 23. Bedrosian, T.A. *et al.* Detection of brain somatic variation in epilepsy-associated developmental lesions. *medRxiv* (2021).

960 24. Lai, D. *et al.* Somatic mutation involving diverse genes leads to a spectrum of focal cortical malformations. *medRxiv* (2021).

961 25. Tarabeux, J. *et al.* Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. *Transl Psychiatry* **1**, e55 (2011).

962 26. Bezprozvanny, I. Calcium signaling and neurodegenerative diseases. *Trends Mol Med* **15**, 89-100 (2009).

963 27. Su, M.Y. *et al.* Hybrid Structure of the RagA/C-Ragulator mTORC1 Activation Complex. *Mol Cell* **68**, 835-846 e3 (2017).

964 28. Chen, J. *et al.* KLHL22 activates amino-acid-dependent mTORC1 signalling to promote tumorigenesis and ageing. *Nature* **557**, 585-589 (2018).

965 29. Behar, T.N. *et al.* Glutamate acting at NMDA receptors stimulates embryonic cortical neuronal migration. *J Neurosci* **19**, 4449-61 (1999).

966 30. Paoletti, P., Bellone, C. & Zhou, Q. NMDA receptor subunit diversity: impact on receptor properties, synaptic plasticity and disease. *Nat Rev Neurosci* **14**, 383-400 (2013).

967 31. Strehlow, V. *et al.* GRIN2A-related disorders: genotype and functional consequence predict phenotype. *Brain* **142**, 80-92 (2019).

968 32. Prickett, T.D. & Samuels, Y. Molecular pathways: dysregulated glutamatergic signaling pathways in cancer. *Clin Cancer Res* **18**, 4240-6 (2012).

969 33. Ruvinsky, I. & Meyuhas, O. Ribosomal protein S6 phosphorylation: from protein synthesis to cell size. *Trends Biochem Sci* **31**, 342-8 (2006).

970 34. Sumimoto, H., Imabayashi, F., Iwata, T. & Kawakami, Y. The BRAF-MAPK signaling pathway is essential for cancer-immune evasion in human melanoma cells. *J Exp Med* **203**, 1651-6 (2006).

971 35. Ornitz, D.M. & Itoh, N. The Fibroblast Growth Factor signaling pathway. *Wiley Interdiscip Rev Dev Biol* **4**, 215-66 (2015).

972 36. Chen, X. *et al.* TNF-alpha-Induced NOD2 and RIP2 Contribute to the Up-Regulation of Cytokines Induced by MDP in Monocytic THP-1 Cells. *J Cell Biochem* **119**, 5072-5081 (2018).

973 37. Yates, T.M. *et al.* SLC35A2-related congenital disorder of glycosylation: Defining the phenotype. *Eur J Paediatr Neurol* **22**, 1095-1102 (2018).

974 38. Paganini, C. *et al.* Calcium activated nucleotidase 1 (CANT1) is critical for glycosaminoglycan biosynthesis in cartilage and endochondral ossification. *Matrix Biol* **81**, 70-90 (2019).

975 39. Lee, N. *et al.* Neuronal migration disorders: positron emission tomography correlations. *Ann Neurol* **35**, 290-7 (1994).

976 40. Kim, Y.H. *et al.* Neuroimaging in identifying focal cortical dysplasia and prognostic factors in pediatric and adolescent epilepsy surgery. *Epilepsia* **52**, 722-7 (2011).

977 41. Nowakowski, T.J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318-1323 (2017).

978 42. Coe, B.P. *et al.* Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat Genet* **51**, 106-116 (2019).

979 43. Ridley, A.J. *et al.* Cell migration: integrating signals from front to back. *Science* **302**, 1704-9 (2003).

980 44. Brini, M., Cali, T., Ottolini, D. & Carafoli, E. Neuronal calcium signaling: function and dysfunction. *Cell Mol Life Sci* **71**, 2787-814 (2014).

981 45. Lamparello, P. *et al.* Developmental lineage of cell types in cortical dysplasia with balloon cells. *Brain* **130**, 2267-76 (2007).

982 46. Englund, C., Folkerth, R.D., Born, D., Lacy, J.M. & Hevner, R.F. Aberrant neuronal-glial differentiation in Taylor-type focal cortical dysplasia (type IIA/B). *Acta Neuropathol* **109**, 519-33 (2005).

983 47. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* **15**, 591-594 (2018).

984 48. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* (2019).

985 49. Yang, X. *et al.* Genomic mosaicism in paternal sperm and multiple parental tissues in a Dravet syndrome cohort. *Sci Rep* **7**, 15677 (2017).

986 50. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**, W71-4 (2007).

987 51. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**, e115 (2012).

1013 52. Lee, J. *et al.* Mutualisk: a web-based somatic MUTation AnaLyIS toolKit for genomic, transcriptional and  
1014 epigenomic signatures. *Nucleic Acids Res* **46**, W102-W108 (2018).

1015 53. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway  
1016 annotation networks. *Bioinformatics* **25**, 1091-3 (2009).

1017 54. Koizumi, H., Tanaka, T. & Gleeson, J.G. Doublecortin-like kinase functions with doublecortin to mediate  
1018 fiber tract decussation and neuronal migration. *Neuron* **49**, 55-66 (2006).

1019 55. Wang, F. *et al.* RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded  
1020 tissues. *J Mol Diagn* **14**, 22-9 (2012).

1021