

1 **Single-cell transcriptomic resolution of stem cells and their developmental trajectories**
2 **in the hippocampus reveals epigenetic control of cell state perseverance**

3 Adrián Salas-Bastos^{1,2}, Martin Treppner^{3,4,2}, Josip S. Herman^{2,5,6,7}, Dimitrios Koutsogiannis²,
4 Harald Binder^{3,4}, Michael B. Stadler^{8,9,10}, Dominic Grün^{5,7,11}, Tanja Vogel^{1,12,13,*}

5 ¹ Institute of Anatomy and Cell Biology, Department of Molecular Embryology, Medical
6 Faculty, Albert-Ludwigs-University Freiburg, 79104 Freiburg, Germany

7 ² Faculty of Biology, Albert-Ludwigs-University Freiburg, Freiburg, Germany

8 ³ Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center,
9 Albert-Ludwigs-University Freiburg , 79104 Freiburg, Germany

10 ⁴ Freiburg Center for Data Analysis and Modeling, Albert-Ludwigs-University Freiburg ,
11 79104 Freiburg, Germany

12 ⁵ Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany

13 ⁶ International Max Planck Research School for Molecular and Cellular Biology (IMPRS-
14 MCB), Freiburg, Germany

15 ⁷ Würzburg Institute of Systems Immunology, Julius-Maximilians-University Würzburg,
16 97078 Würzburg, Germany

17 ⁸ Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

18 ⁹ SIB Swiss Institute of Bioinformatics, Basel, Switzerland

19 ¹⁰ Faculty of Science, University of Basel, Basel, Switzerland

20 ¹¹ Centre for Integrative Biological Signaling Studies, Albert-Ludwigs-University Freiburg,
21 79104 Freiburg, Germany

22 ¹² Center for Basics in NeuroModulation (NeuroModul Basics), Medical Faculty, Albert-
23 Ludwigs-University Freiburg, 79104 Freiburg, Germany

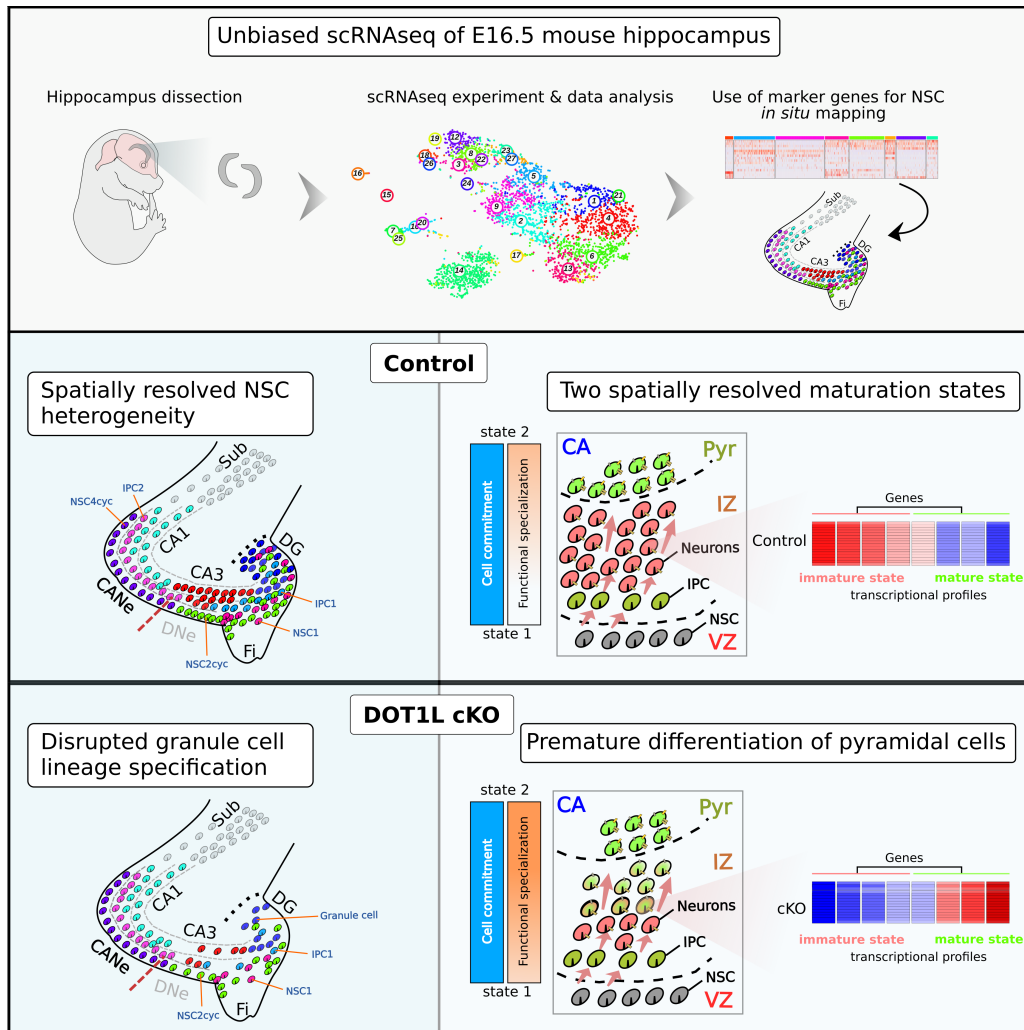
24 ¹³ Freiburg Institute for Advanced Studies (FRIAS), Albert-Ludwigs-University Freiburg,
25 79104 Freiburg, Germany

26 * corresponding author

27 **Abstract**

28 Despite conceptual research on hippocampus development and the application of single-
29 cell-resolved technologies, the nature and maturation of its diverse progenitor populations
30 are unexplored. The chromatin modifier DOT1L balances progenitor proliferation and
31 differentiation, and conditional loss-of-function mice featured impaired hippocampus
32 development. We applied single-cell RNA sequencing on DOT1L-mutant mice and
33 explored cell trajectories in the E16.5 hippocampus. We resolved in our data five distinct
34 neural stem cell populations with the developmental repertoire to specifically generate the
35 cornu ammonis (CA) 1 field and the dentate gyrus (DG). Within the two developing CA1-
36 and CA3-fields, we identified two distinct maturation states and we thus propose CA1- and
37 CA3-differentiation along the radial axis. In the developing hippocampus, DOT1L is
38 primarily involved in the proper development of CA3 and the DG, and it serves as a state-
39 preserving epigenetic factor that orchestrates the expression of several important
40 transcription factors that impact neuronal differentiation and maturation.

41 **Graphical Abstract**



- 42 • The developing hippocampus contains distinct and spatially separated NSC
 43 populations that differ in expression of a specific set of firstly described marker
 44 genes.
 45 • CA pyramidal neurons mature along the radial axis and pass through distinct
 46 maturation states.
 47 • DOT1L preserves the dentate granule cell lineage in the developing hippocampus
 48 and limits maturation in the CA1- and CA3-fields development.
 49 • DOT1L gates cell maturation as upstream regulator of transcription factor
 50 expression that confer instrumental roles in hippocampus development.

51 **Introduction**

52 The hippocampus develops from the medial pallium of the dorsal telencephalon¹. As part of
53 the limbic system, the hippocampal neuronal networks are important for learning, memory,
54 emotions and other functions². It contains three neuronal layers, and is a conserved
55 structure among vertebrates³. But surprisingly, up to date, the hippocampus, that is heavily
56 studied to understand such specific function as spatial navigation and orientation, is poorly
57 understood in its developmental origins.

58 The mature hippocampus is composed of the cornu ammonis (CA) fields and the dentate
59 gyrus (DG). The development of both regions starts in embryonic and continues during
60 postnatal stages⁴⁻⁶. CA and DG neurons originate from neural stem cells (NSCs) that locate
61 to the ventricular zone (VZ)^{4,7} and generate glutamatergic pyramidal cells (PC) or granule
62 cells (GC). Cell migration of committed hippocampal progenitors and differentiating
63 neurons along the radial and longitudinal axes represent crucial processes during embryonic
64 development of the CA-fields^{8,9} and DG¹⁰.

65 Several studies report on transcription factors (TFs) impacting hippocampus development¹¹⁻
66 ²¹ and they give valuable insights into how this process is controlled transcriptionally. It is
67 still unclear, but highly likely, that upstream signals coordinate the expression of instructing
68 TFs, which probably act in organised networks. Epigenetic mechanisms that converge on
69 concerted enhancer activation or on histone modifications to balance activation/repression
70 of TF expression in a spatio-temporal manner, could be in place to orchestrate
71 transcriptional programs driven by individual TFs. We here present the histone
72 methyltransferase *disruptor-of-telomeric-silencing-1-like*, DOT1L, as upstream regulator of
73 TFs that highly impacts hippocampus development by balancing progenitor proliferation
74 and differentiation.

75 DOT1L conditional mouse mutants (cKO) present a disrupted cytoarchitecture of the
76 hippocampus in which the DG is seemingly lost²². DOT1L impacts development of various
77 parts of the central nervous system (CNS)^{23,24}. In all brain regions studied, DOT1L
78 deficiency accelerates neuronal differentiation and migration. Thus, the alterations in the
79 hippocampus in DOT1L-cKO suggested that studying the loss-of-function (LOF) of
80 DOT1L will provide novel insights into general mechanisms of neuronal differentiation
81 during hippocampus development. Using single-cell transcriptome (scRNAseq) analysis of

82 control and mutant E16.5 hippocampus we resolved spatially distinct progenitor
83 populations exerting region-specific transcriptional programs. We reconstructed neuronal
84 differentiation along the radial dimension and different maturation trajectories which were
85 characterised by expression of a distinct set of marker genes. We showed that DOT1L
86 preserved hippocampus identity and repelled expansion of the subiculum. DOT1L
87 prevented premature differentiation of precursors and maturation of committed neurons
88 through concerted transcriptional activation of key TFs. We provide an integrated high-
89 resolution view of the lineage trajectories in place during hippocampus development,
90 identify the central genetic regulators and establish that the histone modifier DOT1L acts as
91 state-preserving epigenetic factor preventing untimely neuronal differentiation and
92 maturation.

93 **Results**

94 ***DOT1L deficiency changes proportions of NSCs, IPCs, CA3 PCs and DG GCs***

95 *Emx1-cre;Dot1l^{fl/fl}* mice (DOT1L-cKO) presented with a disorganised CA and DG region
96 (**Figure 1A**), similar to the phenotype observed using *Foxg1-cre* (**Figure S1A**). We
97 performed scRNAseq²⁵ on E16.5 hippocampus from one control and two DOT1L-cKO
98 (**Figure 1B**), using 4135 cells. This sample size was estimated to be suitable for subsequent
99 clustering analyses²⁶ (**Figure S2**). 3701 cells remained after pre-processing, and RaceID²⁷
100 analysis resolved 27 clusters (**Figure 1C**), most of which were annotated using known
101 marker genes^{6,10,15,28-30} (**Figure 1D**). All expected cell-types were identified, including five
102 clusters of NSCs. We observed a separation of CA1 from CA3 PCs and dentate GCs, and
103 that the latter two related transcriptionally closer to each other than to CA1 (**Figure 1D**).
104 We did not observe any major influence of the sex nor the genotype on the cell distribution
105 (**Figure S3**).

106 Although all clusters contained both control and mutant cells (**Figure 1C, inset**), an in-
107 depth analysis of the distribution of cells per cluster for each animal showed a significant
108 decrease of the proportion of DG granule, NSC1, CA3 pyramidal and IPC1 clusters upon
109 DOT1L-LOF (adjusted $P < 0.05$; Fisher's test) (**Figure 1E**). The subiculum (Subic1, Subic2)
110 and interneuron (INs) clusters showed significantly higher cell proportions for each of the
111 two DOT1L-cKO samples compared to controls (adjusted $P < 0.05$; Fisher's test) (**Figure**

112 **1E**). Thus, DOT1L-cKO quantitatively altered the cell-type composition in the developing
113 hippocampus in a cell-type specific manner.

114 We studied the enrichment of DOT1L-cKO cells with expression of subiculum marker
115 genes and the boundary to the adjacent CA1-field. ZBTB20-expressing cells localized in
116 the intermediate zone (IZ), the CA PC layer and the DG, but they were largely absent in the
117 subiculum (**Figure 1F**). The ZBTB20 expression domain in DOT1L-cKO seemed smaller
118 and localised ventro-laterally with fewer positive cells in the DG region compared to
119 controls (**Figures 1F, S4A**). The proportions of CA1 PCs did not decrease significantly in
120 the scRNAseq data suggesting a spatial re-arrangement of CA1 and subiculum, whereby
121 the subiculum was shifted ventro-laterally upon DOT1L-LOF. ISH for *Id2* as
122 cortical/subiculum marker²⁰ showed an expanded subiculum expression domain reaching
123 the prospective CA1-field ventrally in DOT1L-cKOs (**Figure 1G**). SATB2-positive cells,
124 present in the subiculum but mostly absent from the CA1-field³¹, also extended ventrally in
125 DOT1L-cKOs (**Figure 1H**). LHX2-positive cells, marking the hippocampal/cortical
126 boundary¹³, were almost completely absent in the DOT1L-cKOs in the hippocampal VZ,
127 CA-fields and DG, compared to their presence in controls (**Figure 1H**). Foxg1-cre
128 DOT1L-cKO (**Figures S1B-D**) displayed similar alterations, which strongly indicated that
129 DOT1L preserved hippocampal identity.

130 ISH for *Grik4*⁶ confirmed the decreased representation of CA3 PCs upon DOT1L-LOF
131 compared to controls (**Figure 1I**). Further, we confirmed the impaired DG in the DOT1L-
132 cKO, because we observed fewer PROX1- and EOMES-positive cells compared to controls
133 along the dentate migratory stream (DMS) and in the developing DG in DOT1L-cKOs
134 (**Figures 1J, 4B, S1E,F**). PAX6-expressing GC progenitors of the DG presented a similar
135 pattern as PROX1-positive cells in the DOT1L-cKOs (**Figure S4C**), underlining reduced
136 presence of DG granule precursors and neurons upon DOT1L-LOF. Summarising, DOT1L
137 prevented expansion of subiculum identity into the hippocampus and regulated cell
138 numbers that settle in the CA3-field and DG.

139 ***Lineage trajectory analyses resolve different progenitor populations in the developing***
140 ***hippocampus***

141 Studies of hippocampus development incompletely resolved transcriptional programs
142 underlying spatio-temporal cell fate trajectories of progenitors, but rely on broadly
143 expressed NSC and IPC markers, including *Pax6*, *Sox9*, *Eomes* or Notch signalling
144 pathway members¹⁰. Due to the high resolution inherent to the scRNAseq technology we
145 resolved five different NSC and two IPC clusters (**Figures 1C,D**) in the E16.5
146 hippocampus. RNA velocity analysis³² using scvelo³³ resolved cell trajectories (**Figure 2A**),
147 in which NSCs clustered together and tended to move towards IPC1, followed by IPC2.
148 IPC2 could thus represent a mature intermediate state towards neuronal differentiation
149 compared to IPC1. IPC2 connected to the CA1 PCs, followed by CA3 PCs and DG GCs.
150 Notably, the velocity vectors for the subiculum clusters converged with both CA PC
151 populations, but they lacked direct connection either to NSCs or IPCs identified in our
152 scRNAseq data set (**Figure 2A**).

153 To resolve the connections between the different NSC clusters in detail, we applied
154 diffusion maps (DM)³⁴ and diffusion pseudotime (DPT)³⁵ analyses, after filtering out cells
155 unrelated to the hippocampal lineage (see methods). DM and DPT analyses showed an
156 increased pseudotime as cells moved from the NSC (left) towards the more committed cell
157 populations (right). CA3 PCs appeared as the most mature population at E16.5 (**Figures**
158 **2B,C**). NSC2cyc and NSC4cyc localized closest to the trajectory start. Due to
159 transcriptional similarities within stem cells, they seemingly intermingled with each other
160 and with the remaining NSC1,3,5 and cortical hem (CH, CortHem) clusters (**Figure 2B**).
161 The cell arrangement based on the first three diffusion components indicated that NSCs
162 conserved partly their cluster structure (**Figure 2D**), with differences between the
163 respective NSC populations. DMs corroborated sequential development from the NSCs to
164 the IPC1 and IPC2 states, followed by CA1 PCs, DG GC and CA3 PCs (**Figures 2 B,D**).

165 We investigated the main gene sets and corresponding biological processes that changed
166 along the trajectory by combining the cell order provided by the DPT with self-organizing
167 maps (**Figure 2E**). Our analysis retrieved a series of nodes representing groups of genes
168 highly expressed at different stages along the pseudotime axis. NSCs expressed strongly
169 genes related to metabolic processes (**Figure 2E**, Node 26), and both NSCs and IPCs

170 expressed gene sets associated with translation and mRNA processing (**Figure 2E**, Nodes
171 25 and 28). Genes included in these nodes decreased in expression with advanced
172 differentiation. As cells matured towards IPCs and neuronal states, gene sets involved in
173 synaptic function increased in transcription (**Figure 2E**, Node 2). Noticeable, genes
174 affecting chromatin organization and histone modifications increased in expression from
175 NSCs to IPCs, had a maximum expression in IPCs, and decreased as differentiation moved
176 ahead on the diffusion pseudotime axis (**Figure 2E**, Node 9). IPC1 preceded IPC2 cells in
177 this node, corroborating different maturation states in the intermediate progenitors.
178 DM analysis indicated seemingly hierarchical ordered NSC populations, but could not
179 delineate a clear distinction of the different NSC clusters. We thus defined, based on visual
180 inspection of the expression pattern of the DEGs between populations (adjusted $P < 0.05$) in
181 the tSNE representation, marker genes of the respective NSC and IPC subtypes and used
182 single-molecule FISH (smFISH) to resolve the spatial locations of the distinct cell
183 populations (**Figure 2F**). *Sox21* transcription delineated the CH, NSC1 and NSC2cyc (low
184 *Sox21*) from NSC3, 4cyc, and 5 (high *Sox21*). *Sox21*-signals were strongest in the cortical
185 VZ and CA neuroepithelium (CANe), not extending into the dentate neuroepithelium
186 (DNe), origin of the DG (**Figure 2G**). *Cybrd1* expression marked NSC4cyc and CH, and
187 smFISH showed that NSC4cyc cells located basally in the CANe, compared to apically
188 located *Sox21*-expressing cells, which confined NSC3 and 5 within the VZ of the CANe.
189 Consistent with its expression in CH cells, a defined *Cybrd1*-expressing cell population
190 localized near the fimbria (**Figure 2H**).
191 *Sned1*, *Wnt8b* and *Adamts19* expression characterised NSC1 (expression of all three
192 markers, *Sned1* higher) and NSC2cyc (*Adamts19* high; *Sned1*, *Wnt8b* low) (**Figures 2I-K**).
193 According to *Sned1* expression, NSC1 localised in the VZ of the DNe extending towards
194 the DG. NSC2cyc localised to the same area. Although NSC1 and NSC2cyc had similar
195 expression domains, expression of *Sox21* in the NSC2cyc (**Figure 2F**) probably localises
196 these cells apically in the DNe. scRNAseq and smFISH together resolved two spatially
197 separated progenitor regions: NSC3, NSC4cyc and NSC5 precursors comprise the CANe,
198 whereas NSC1 and NSC2cyc demarcated the DNe.
199 The lineage trajectories suggested that IPC1 and 2 are common, probably consecutive states
200 during differentiation, originating from both progenitor populations (CANe, DNe). We

201 selected marker genes for discriminating both IPC populations and characterizing their
202 spatial location: all NSCs, CH and IPC1 expressed strongly *Hes5*, but had low expression
203 in IPC2; *Tac2* was almost exclusively expressed in IPC2 with the exception of some few
204 IPC1 (**Figure 2F**). *Hes5/Tac2* co-staining showed that *Hes5*-expressing cells mainly
205 localized along the entire VZ of the CANe and DNe (**Figure 2L**). *Tac2*-expressing cells
206 localized in the SVZ along the CANe and DNe, extending into the DG (**Figure 2L**). The
207 expression domain of both markers seemed well separated between VZ and SVZ in the
208 CANe but positive cells appeared intermingled in the DG region. IPC1 showed higher
209 levels of *Adamts19* compared to IPC2 (**Figure 2F**), and *Adamts19* expression was restricted
210 in the DNe and DG (**Figure 2K**). Thus, IPC2 are basal progenitors in the SVZ of the
211 CANe, and *Hes5/Adamts19*-positive IPC1 localised to the DNe and DG. Together with the
212 bioinformatics prediction that IPC1 seemingly preceded IPC2, we interpreted that IPC1 and
213 IPC2 are two spatially separated progenitor fractions that were seemingly depicted in an
214 earlier (IPC1) and advanced (IPC2) maturation state. Their distribution reflected probably
215 the slightly advanced development of the CA1 PC compared to DG GC.

216 ***NSCs and IPCs are biased towards CA1 PC and DG GC fate at E16.5***

217 Our data suggested two separate germinative regions to generate the CA-fields and DG. To
218 investigate the connections between the specific NSC/IPC populations to the mature cell
219 lineages we performed CellRank³⁶ and FateID³⁷ analyses. As we knew the cell populations
220 that represented the final cell states, we provided this information to the algorithms (**Figure**
221 **3A**). CellRank (**Figure 3B**) and FateID (**Figure 3C**) results highlighted distinct fractions of
222 the NSCs with high probabilities to generate either CA1 PC or DG GC. Specifically,
223 FateID analysis highlighted that CA1 PC fate connected to the NSC3, NSC4cyc, and NSC5
224 clusters (**Figure 3C**). This observation was in line with the NSC spatial expression domains
225 we identified *in vivo* (**Figures 2G-H**) and confirmed the separation of the CANe and DNe.
226 Both algorithms indicated that IPC1 had a fate bias towards DG GC, whereas IPC2 seemed
227 to generate CA1 PCs at E16.5. This observation supported that IPC2 connected to the
228 advanced development of CA1 as suggested by their localisation *in vivo*. The different NSC
229 and IPC populations identified at E16.5, however, had very low fate probabilities towards
230 the CA3 PC and subiculum cell lineages. CellRank analysis additionally indicated a

231 gradually declining differentiation potential, which was highest in the NSCs, moderate in
232 IPC, and lowest in neurons. Within the IPCs, IPC2 had a lower differentiation potential
233 compared to IPC1, confirming that the former represented a slightly advanced intermediate
234 state (**Figure 3B**).

235 ***DOT1L preserves the DNe and prevents premature differentiation***

236 We used the newly identified marker gene expressions in the stem cell compartment of the
237 developing hippocampus to study by smFISH the specific NSCs/IPCs affected in DOT1L-
238 cKOs. *Cybrd1* expression in the CANe (NSC4cyc) appeared similar in both genotypes
239 (**Figure 4A**). *Sned1*-positive NSC1 within the DNe/DG region reduced largely upon
240 DOT1L-cKO (**Figure 4B**). DNe/DG-located *Hes5*-expressing IPC1 decreased strongly in
241 the DOT1L-cKOs compared to the CANe, where *Hes5*-expression was unaffected (**Figure**
242 **4C**). *Tac2*-positive IPC2 in the SVZ were kept in the DOT1L-cKO CANe, but reduced in
243 the DNe/DG. These observations confirmed strong reduction of the stem cell populations
244 residing in the DNe/DG (NSC1, IPC1) upon DOT1L-cKO, and seemingly normal CANe
245 appearance. The finding of fewer DG stem cells explained reduced numbers of PROX1-
246 positive DG GC in DOT1L-cKO (**Figure 1J**).

247 NSC1 and IPC1 in the DNe/DG region seemed to be particular sensitive to DOT1L
248 presence. We thus explored expression changes upon DOT1L-cKO in stem cells, which
249 revealed that NSC1 separated from the other populations (**Figure 4D**). Seven genes altered
250 significantly their expression level specifically in NSC1 but not in other stem cells
251 (adjusted $P < 0.05$). The increased expression of *Pak3*³⁸ and *Neurod2*³⁹⁻⁴¹ indicated that
252 NSC1 were prone towards premature differentiation upon DOT1L-cKO. Among the IPCs,
253 IPC1 had 19 unique DEG upon DOT1L-cKO compared to controls. Here, the decreased
254 expression of *Nr2f1*⁴² and *Dmrta2/Dmrt5*⁴³ correlated with precocious neuronal
255 differentiation, similar to increased expression of *Cyfp2*⁴⁴, *Neurod2*³⁹⁻⁴¹ and *Nrn1*⁴⁵ (**Figure**
256 **4E**). Thus, expression changes upon DOT1L-cKO suggested that premature differentiation
257 caused the impaired development of the DG region.

258 ***DOT1L prevents maturation and activates TFs involved in hippocampus development***

259 We analysed DEGs within the entire data set to retrieve mechanistic insight into the
260 observed phenotypic changes. 21 clusters contained cells from both experimental
261 conditions and had DEGs (adjusted $P < 0.05$), of which DG GC, CA1 and CA3 PCs had the
262 highest numbers of DEGs (**Figure 5A**). Hierarchical clustering based on the transcriptional
263 alterations upon DOT1L-cKO divided the cell populations into two main groups. Mature
264 hippocampus cells (IPC2, DG GC, CA1 and CA3 PC) had evident changes in gene
265 expression (**Figure 5B, group 2**). GO-term analysis of genes with decreased expression
266 enriched in processes linked to brain development and morphogenesis. Genes with
267 increased expression upon DOT1L-LOF retrieved significant enrichment for biological
268 processes indicative of differentiation, including neuronal differentiation, synaptic and
269 cognitive function terms ($q\text{-value} < 0.05$) (**Figure 5C**). From this finding we concluded that
270 DOT1L impacted hippocampus development beyond the stem cells.

271 We focussed further analyses on TFs, because they have essential roles in hippocampus
272 development. Expression levels of 49 TFs changed significantly upon DOT1L-cKO
273 (adjusted $P < 0.05$) (**Figure 5D**) and NSCs shared a decreased expression of known key
274 regulators (**Figure 5D, cluster 4**). TFs both de- and increased in expression in mature cell
275 populations. Similar TF expression grouped DG GC and CA3 PCs in cluster 1, and IPC2
276 and CA1 PCs in cluster 2. Noticeable, *Lhx2*, *Tcf4*, *Nfix*, and *Nfib* decreased in all, and
277 *Nr2f1*, *Zbtb20* and *Nfia* in most represented populations. The genomic regions coding for
278 these different TFs showed decreased H3K79me2 in neuronal progenitor cells (NPCs)
279 treated with the DOT1L inhibitor EPZ5676, compared to DMSO control (**Figure 5E**),
280 rendering this set of TFs as potential targets of DOT1L. All TFs also showed lower
281 expression levels in the DOT1L-cKO along the pseudotime axis when compared to
282 controls. *Lhx2*, *Nfix* and *Nfib* increased in expression from NSCs to IPCs, and decreased as
283 the cells differentiated (**Figure 5F**). These results suggested a main role of the mentioned
284 TFs in IPCs, in which chromatin remodelling gained particular relevance (**Figure 2E**, Node
285 9). A decreased expression on the protein levels of ZBTB20, LHX2 and NFIB (**Figures**
286 **1F,1H, 5G**) confirmed that DOT1L regulated these TFs involved in hippocampus
287 development. Deletion of all the individual TFs impaired hippocampus development in a

288 similar way as we observed^{12,19,46}. This suggested a strong mechanistic link between altered
289 TFs expression and the observed phenotype upon DOT1L-cKO.

290 **CA PCs are in two maturation states at E16.5**

291 As CA PCs maturation lacks detailed understanding, and our data indicated that DOT1L-
292 LOF impacted both cell populations transcriptionally, we characterized the differentiation
293 trajectories of CA1 and CA3 PCs at single-cell resolution. We used the expression of
294 *Pou3f1* and *Grik4*⁶, to annotate both PC clusters. Visual inspection of the data showed that
295 neither *Pou3f1* nor *Grik4* was expressed homogeneously in all cells of the respective
296 cluster. In contrast, CA1 and CA3 PCs contained cells with either high or low expression of
297 *Pou3f1* or *Grik4*, which was indicative of unresolved heterogeneity within these two cell
298 populations (**Figures 6A,G**). We performed independent sub-clustering analysis in each
299 CA population (see methods) and resolved different maturation states. For both CA
300 populations, one of the obtained sub-clusters enriched with cells expressing noticeable
301 levels of either *Pou3f1* or *Grik4* (**Figures 6B,H**). GO-term analysis for significantly
302 increased genes (adjusted $P < 0.05$) in the CA1 and CA3 sub-clusters, enriched in either
303 *Pou3f1* or *Grik4* expression, retrieved terms indicative for neuronal differentiation (**Figures**
304 **6C,I**). scvelo analysis corroborated that sub-cluster 1 cells (low *Pou3f1* (CA1) or *Grik4*
305 (CA3)) moved towards the sub-cluster 2 (high *Pou3f1* or *Grik4*), and supported the
306 existence of two maturation states in CA1 and CA3 PCs at E16.5 (**Figures 6D,J**).

307 Based on different expression levels in the respective sub-clusters, we selected *Tac2*, *Luzp2*
308 and *Pcp4* as markers for less mature and *Pou3f1* for mature CA1 PCs; *Grp* for immature
309 and *Cpne4* for mature CA3 PCs, respectively (**Figures 6E,K**). Immature CA1 markers
310 were expressed in the CANe, within the NSC3, 4cyc, and 5 expression domain, but their
311 expression extended also into cortical areas including the subiculum. *Tac2* and *Pcp4*
312 expression comprised the DNe/DG (**Figures 6F, S5**). *Pou3f1* expression extended in the IZ,
313 where its expression was lower than in the overlying PCs, where mature cells expressed
314 strongly *Pou3f1* (**Figure 6F**). *Grp* (immature CA3) was expressed in the CA3 region with a
315 slightly higher level in the IZ than in PCs (**Figure 6L**). *Cpne4*, marking mature CA3 cells,
316 was restricted to the PCs (**Figure 6L**). Together, we resolved two different maturation

317 states during embryonic CA-field development that sequentially follow each other along
318 the radial axis.

319 ***DOT1L gates CA pyramidal cell maturation***

320 DOT1L gates differentiation of progenitor cells, seemingly in a cell-type specific manner,
321 e.g. with the CANe being less affected than the DNe. The transcriptional changes upon
322 DOT1L-cKO in the mature CA1 and CA3 PCs (**Figures 5A-C**) and the two different
323 maturation states in both CA-fields, suggested that DOT1L gates also maturation of
324 differentiated cells. Indeed, DOT1L-LOF favoured the transition from less to more mature
325 PC states, because more cells belonged to mature CA1 or CA3 sub-clusters compared to
326 controls, which contained a significantly higher proportion of cells in less mature sub-
327 clusters (adjusted $P < 0.05$, Fisher's test) (**Figures 7A,G**). In support and indicative of
328 premature differentiation, mutants showed a stronger *Pou3f1* signal at the level of the SVZ
329 and IZ compared to controls (**Figure 7B**). The CA3 marker *Grp* was nearly completely lost
330 in DOT1L-cKO, corroborating a drastic depletion of the less mature cells (**Figure 7H**).

331 DEG analysis revealed discrete gene sets, expression of which either decreased or increased
332 significantly in the less mature cell states upon DOT1L-cKO compared to controls
333 (adjusted $P < 0.05$) (**Figures 7C,I**). In both CA-fields significantly decreased gene sets
334 enriched for GO-terms related to transcriptional control ($q\text{-value} < 0.05$), which indicated
335 that TFs represented again important targets affected upon DOT1L-cKO (**Figures 7C,I**).
336 We determined, independent on the DOT1L-cKO, TFs that changed significantly in
337 expression in mature compared to less mature CA cells (**Figures 7D,J**). Both, CA1 and
338 CA3 maturation correlated with decreased expression of *Lhx2*, *Nfia*, *Nfib*, *Nfix*, and *Nr2f1*
339 (adjusted $P < 0.05$) (**Figures 7D,J**). The significant higher expression of the former TFs in
340 the less mature cell state and their role as transcriptional activators indicated that they are
341 favouring a less differentiative cell state. We subsequently identified those TFs that might
342 mediate the increased maturation upon DOT1L-cKO by intersecting the TFs changing upon
343 CA maturation with TFs changing significantly upon DOT1L-cKO in less mature CA cells.
344 13 TFs for CA1, and 8 for CA3 were differentially expressed (adjusted $P < 0.05$) (**Figures**
345 **7E,K**), of which *Nfia*, *Nfix*, *Nr2f1* and *Lhx2* decreased, and *Bcl11b* increased in expression
346 upon DOT1L-LOF in both CA-fields (**Figures 7F,L**). *Lhx2* was transcriptionally most

347 affected and its expression also decreased on the protein level in the CANe, IZ and the
348 pyramidal layers upon DOT1L-cKO (**Figures 7M, 1H**).

349 **Discussion**

350 The presented data provide novel insights into cell lineage trajectories and fates in the
351 developing hippocampus that are under control of the histone methyltransferase DOT1L.
352 DOT1L functions in the CNS to gate and control adaption of subsequent cell fates^{22-24,47} and
353 LOF mice had an impaired hippocampal structure²². We used the DOT1L-cKO to dissect
354 cell lineage differentiation in the hippocampus that is in this regard remarkably
355 understudied. Based on high-resolution scRNAseq, we demonstrated different stem cell
356 populations in the developing hippocampus. Other recent work that applied scRNAseq
357 technology for the study of the developing hippocampus represented NSCs as one
358 homogeneous population, and did not discriminate either between NSCs and IPCs
359 committed to specific hippocampal lineages (i.e., PCs and GCs) or between possible
360 distinct NSC states^{5,32}. We resolved heterogeneity in the stem cell populations of the
361 hippocampus. The subdivision of NSCs and IPCs into discrete clusters pointed either
362 towards transitory cell states during proliferation and differentiation, or to discrete
363 progenitor classes with the potential to generate specific cell lineages. These alternatives
364 were not exclusive, but were both true. We defined a set of novel marker genes to
365 discriminate the two germinative regions, CANe and DNe, that were proposed by others,
366 but were molecularly ill-defined due to lack of specific markers for the respective cell
367 populations. From recently reported markers, used to dissect DG development on a protein
368 level (SOX9, HES1/5, NOTCH1/2, HEY2 for NSCs; EOMES, DLL3 for IPCs¹⁰), only *Dll3*
369 enriched in one single specific cell population in our data set, i.e. IPC1. All other markers
370 were transcribed at similar levels in all identified NSC clusters in our study (**table 1**), and,
371 moreover, some of them in Cajal-Retzius cells, oligodendrocyte precursors, CH or
372 pericytes. *Eomes* was expressed by NSC1 and 3, by Cajal-Retzius cells and in both IPC
373 clusters. Thus, our data is of major relevance for high resolution studies of lineage
374 relationships in the developing hippocampus. We identified NSC3, NSC4cyc, and NSC5 as
375 CANe, expressing *Sox21* and *Cybrd1*, the NSC2cyc as DNe, expressing *Sned1* and
376 *Adams19*, and NSC1 feeding the DG trajectory. Given that derailed development of the

377 hippocampus is underlying diseases like autism and schizophrenia^{48,49}, it is of paramount
378 relevance to deepen our knowledge on the basic cellular relationships in this part of the
379 limbic system.

380 Our data advance prevailing views on hippocampus development. Firstly, our data analysis
381 suggested that subiculum cells lacked connection to NSCs and IPCs present in the E16.5
382 hippocampus. This points towards an extra-hippocampal origin of subiculum cells, which
383 argues against recent findings that subiculum and CA1 PCs share a common developmental
384 path³². The discrepancy between both studies could be attributed to the different
385 developmental stages analyzed (E16.5 in our work, P0 and P5 in La Manno et al³²).
386 Moreover, the bioinformatic tools (scvelo vs Velocyto) and their specific parameterizations
387 (for example selection of variable genes) employed to predict lineage relationships might
388 also affect the predictions⁵⁰.

389 Secondly, we observed that the CA1-lineage was affected less severely compared to CA3
390 upon DOT1L-cKO. A possible reason for this could be that both populations originate
391 from, however so far not defined, different progenitor pools of which only one is DOT1L-
392 sensitive. Our data set could suggest a trajectory from NSC1 via IPC1 to CA3 and DG,
393 because these clusters decreased in cell numbers upon DOT1L-cKO. A common CA3/DG
394 progenitor could be postulated as different concentrations of WNT3A differentiated human
395 induced pluripotent stem cell-derived hippocampal progenitors into either CA3 PC or DG
396 GC⁵¹. Moreover, *Prox1*-LOF in postmitotic neurons induced transcriptional changes
397 indicative of a switch from DG GC to CA3 PC identity⁵², highlighting an inter-conversion
398 plasticity between both lineages. Despite the quantitative changes observed in our data set
399 and the transcriptional similarity among CA3 PCs and DG GCs, the algorithms we used to
400 predict this lineage trajectory in the scRNAseq data set and our visualisations in tissue
401 sections could not find support for CA3 PCs deriving from the DNe. Because, surprisingly,
402 our in-depth analysis of differentiation trajectories into CA1 and CA3 PCs, DG GCs or
403 subiculum cell fates did not reveal convincing differentiation probability from any
404 NSC/IPC clusters towards CA3. The lack of identification of CA3 progenitors could be
405 caused by technical limitations, but our data set derived from three individuals. The

406 likelihood that we did not capture CA3 progenitors in three different cell isolation attempts
407 seems low. This raises the hypothesis that CA3 progenitors are not represented at E16.5,
408 implying that they occur earlier in development. This is possible, because CA-field
409 development is controlled temporally, with CA3 arising earlier compared to CA1⁸. Our data
410 supports this observation, as at E16.5 CA3 PCs were the most differentiated cell
411 population. Thus, CA3 progenitors should differentiate earlier than CA1 progenitors, if they
412 share a common ancestor. Or, equally possible, CA3 progenitors differentiate earlier but
413 from a shared ancestor of the DG granule lineage. Precise definition of lineage trajectories
414 and spatio-temporally separated progenitor populations for the CA1- and CA3-fields
415 requires time resolved scRNAseq experiments. Using DOT1L-cKO with relative normal
416 numbers of CA1 but loss of CA3, seems an adequate approach to provide such in-depth
417 insights.

418 Thirdly, previous studies proposed maturation of the CA PCs along the longitudinal axis in
419 a pole-inward pattern, starting at embryonic and prolonging into postnatal stages⁶. We
420 refine this view by showing two radially separated maturation states in the CA PC lineages.
421 *Grp* and *Cpne4* are highly specific novel markers that highlight this radial maturation
422 trajectory of CA3 PCs. In conjunction are altered expression levels of TFs, most of which
423 decreased with maturation. DOT1L-LOF decreased a TF-set that is silenced during normal
424 developmental CA-field maturation, which might be responsible for the advanced
425 maturation observed in both CA-fields. Additionally, a set of TFs is differently regulated
426 during the maturation of CA1 or CA3 PCs: *Zfp462*, *Pou3f3*, *Insm1*, *Id2*, *Foxg1* and *Bcl11b*
427 increased during CA1-maturation, and they decreased during CA3-maturation (**Figure**
428 **6D,J**). Thus, this TF-set might be involved in cell-specific maturation programs, and it
429 might confer the loss of CA3 and relative normal numbers of CA1 PCs upon DOT1L-LOF.
430 In this regard DOT1L-cKO increased *Bcl11b* expression in CA3 PCs , thus in opposing
431 direction compared to the normal maturation trajectory. *Bcl11* homologues might play
432 important roles in conferring lineage identity between CA-fields because whereas *Bcl11a*
433 and *Bcl11b* are co-expressed in CA1/2, *Bcl11b* is not expressed in CA3, but *Bcl11a*
434 expression extends into this region⁵³. Increased levels of *Bcl11b* upon DOT1L-cKO

435 specifically in the CA3 PCs might interfere at least in part with adapting of a
436 phenotypically distinct CA3-identity.

437 Lastly, our data establish DOT1L as first histone methyltransferase to act as an upstream
438 regulator of instructive TFs that maintain progenitor states, but also drive maturation of CA
439 PCs or DG GCs. The histone deacetylases, HDAC1, 2 and 3⁵⁴⁻⁵⁶ are other histone modifiers
440 impacting hippocampus development. Thus, epigenetic mechanisms play pivotal roles in
441 hippocampus development, in cell fate and maturation processes. Our data suggest that
442 chromatin plasticity and histone modifications are important mechanisms for generating
443 and resolving transient developmental stages in the hippocampus.

444 **Methods**

445 **Generation of DOT1L-cKO mice, dissection and genotyping**

446 All animal experiments were approved by the animal welfare committees of the University
447 of Freiburg and local authorities (G16/11). *Emx1-cre*⁵⁷ and *Foxg1-cre*⁵⁸ mice were kept on a
448 C57BL6/J background and used to breed *Emx1-Dot1l* and *Foxg1-Dot1l-cKO* mice,
449 respectively. Mutant offspring correspond to both *Emx1*^{cre/+};*Dot1l*^{fl/fl} and *Foxg1*^{cre/+};*Dot1l*^{fl/fl},
450 while WT offspring correspond to *Emx1*^{+/+};*Dot1l*^{fl/fl}, *Emx1*^{+/+};*Dot1l*^{fl/+}, *Foxg1*^{+/+};*Dot1l*^{fl/fl} and
451 *Foxg1*^{+/+};*Dot1l*^{fl/+}.

452 Adult pregnant animals were sacrificed by cervical dislocation for embryo collection at
453 embryonic day (E) E16.5 or E18.5. The embryos were sacrificed by decapitation. The
454 brains isolated at defined embryonic stages were washed in cold phosphate-buffered saline
455 (PBS) and then fixed in 4% PFA overnight at 4° C. For cryosections, fixed brains were
456 incubated in 30% sucrose solution at 4° C until, embedded in tissue freezing medium
457 (Leica Biosystems, Germany) and frozen at -80 °C. Forebrains were cut coronally into
458 16µm sections and mounted on Superfrost Plus Microscope Slides (Thermo Fisher, USA).

459 Genotyping was carried out by polymerase chain reaction (PCR) from tail DNA. Briefly,
460 tail samples were lysed in QuickExtract DNA Extra Solution 2.0 (Lucigen, USA) and PCR
461 reactions were carried out with GoTaq DNA polymerase and with the Primers: *Emx1-Cre*
462 forward: 5'-ATGCTTCTGTCCGTTTGCCG-3', *Emx1-Cre* reverse: 5'-
463 CCTGTTTTGCACGTTACCG-3', *DOT1L* forward: 5'-
464 GCCTACAGCCTTCATCATT-3', *DOT1L* reverse: 5'-
465 CCCATACAGTACTCACCGGAT-3', *Bf1-F25* forward: 5'-
466 GCCGCCCCCGACGCCTGGGTGATG-3', *Bf1-R159* reverse: 5'-
467 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG-3' and *Bf1-Rcre222* reverse: 5'-
468 ATAATCGCGAACATCTTCAGGTTCTGCGGG-3'. PCR products were analyzed by
469 agarose gel electrophoresis.

470 **Histological analyses**

471 *Cresyl Violet (Nissl) Staining*

472 Frozen embryonic brains were sectioned at 16µm with a Leica cryostat (CM3050S) and
473 stained with cresyl violet, following standard protocol. The sections were mounted on

474 Superfrost Plus Microscope Slides (Thermo Fisher, USA) with Eukitt mounting medium
475 (O. Kindler, Germany).

476 *Immunofluorescence staining (IF)*

477 Embryonic mouse forebrain sections fixed on Superfrost Plus Microscope Slides (Thermo
478 Fisher, USA) were surrounded with a hydrophobic pen and permeabilized and blocked with
479 10% normal donkey serum (NDS; Bio-Rad, USA)/ 0.1% Triton X-100 (Carl Roth,
480 Germany)/ PBS for 1 h at room temperature (RT). Sections were incubated with primary
481 antibodies diluted in the blocking solution over night at 4° C. After 3 washing steps with
482 PBS/ 0.1% Triton X-100 the secondary antibodies (dilution 1:500) were applied in blocking
483 solution and incubated for 1 h at room temperature. Sections were washed again 3 times
484 with PBS and nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI, dilution
485 1:1000). After 3 final washing steps sections were mounted with Dako fluorescent
486 mounting medium (Agilent Technologies, USA) and sealed with nail polish. The respective
487 antibodies and the dilutions used are listed in **table 2**.

488 *In situ hybridization (ISH)*

489 Forebrain slices at the different developmental stages were hybridized with digoxigenin-
490 labelled riboprobes in hybridization buffer (12.7 mM Tris base, 184.4 mM NaCl, 5.9 mM
491 NaH₂ PO₄, 6.27 mM Na₂ HPO₄, 5 mM EDTA pH 8.0, 0.5x Denhardt's solution, 1 mg/ml
492 Yeast RNA, 10% Dextran sulfate, 50% v/v Formamide) at 68°C overnight. Sections were
493 washed 3 times in a solution containing 50% formamide, 0.1% Tween-20 and 5% saline
494 sodium citrate at 68°C in a water bath. They were then transferred to an incubation
495 chamber and washed twice with maleic acid buffer and Tween-20 (MABT) for 30 min at
496 RT. After blocking in a MABT solution containing 20% lamb serum, sections were
497 incubated with an alkaline phosphatase-conjugated anti-digoxigenin antibody (1:1500 in
498 blocking solution; Roche, Switzerland) overnight at RT. After four washing steps in MABT
499 (10 min, 3x 20 min) and three washing steps (7 min each) in pre-staining buffer, the
500 reaction product was developed using NBT/BCIP solution diluted in pre-staining buffer
501 (1:100; Roche, Switzerland) overnight at RT. Stained sections were washed 4 times in PBS
502 and then embedded using Aquatex (Merck Millipore, USA).

503 *Single molecule FISH (smFISH)*

504 Brain tissue cryosections with a thickness of 16µm were processed following the
505 RNAScope kit protocol, with some few modifications. Firstly, the tissue sections were
506 incubated at 40°C for 1 h in the HybEZ hybridization oven (Advanced Cell Diagnostics,
507 USA) to ensure adhesion of the tissue on the slides. The sections were then washed 5 min
508 in 1X PBS to remove the remaining tissue freezing medium. Afterwards tissue slides were
509 treated with RNAScope Hydrogen Peroxide for 10 min at RT and the sections were
510 immediately washed in deionized water. Secondly, the slides were put in a slide holder and
511 incubated for 5 min inside a beaker filled with 1X RNAScope Target Retrieval Reagent
512 preheated at approximately 95°C. At the end of the incubation time the slides were washed
513 in deionized water, incubated in 100% Ethanol for 2 min, and then dried at RT. Afterwards,
514 an hydrophobic pen was used to delineate the area around the sections and RNAScope
515 ProteaseIII was added on top of the sections followed by incubation 30 min at 40°C in the
516 HybEZ hybridization oven. Thirdly, the slides were washed with deionized water and
517 subjected to the corresponding hybridization and amplification steps in the HybEZ
518 hybridization oven at 40°C. At this step the sections were incubated with the specific
519 probes for 2 h. For the amplification, the sections were incubated with RNAScope
520 Multiplexv2 Amp1 for 30 min, RNAScope Multiplexv2 Amp2 for 30 min and RNAScope
521 Multiplexv2 Amp3 for 15 min, each step followed by two 2 min washing step on Washing
522 Buffer at RT. Fluorescent signal was revealed by incubation of the sections with RNAScope
523 Multiplex FLv2 HRP-C specific reagents (C1, C2 or C3 depending on the probes channel)
524 for 15 min, the respective fluorophore for 30 min and then RNAScope Multiplex FL v2
525 HRP Blocker for 15 minutes, each step followed by two 2 min washing step on Washing
526 Buffer at RT. In the case of double stainings the slides were incubated with the respective
527 RNAScope Multiplex FL v2 HRP-C reagent targeting the second probe and the respective
528 fluorophore following the same steps as above. All the incubation steps at 40°C were
529 carried out inside an aluminium chamber containing the slides arranged on an ACD EZ-
530 Batch Slide Rack and a deionized water moistened tissue. Finally, the slides were
531 counterstained with DAPI for 30 sec at RT and mounted using Dako fluorescent mounting
532 medium (Agilent Technologies, USA). The dyes used for signal detection were Opal Dye

533 520 and Opal Dye 570 (Akoya Biosciences, USA). The respective probes as well as the
534 fluorophore dilutions used in the different stainings are listed in **table 3**.

535 *Image acquisition and processing*

536 Images were acquired with Axio Imager M2 (Zeiss, Germany). High resolution images
537 were downsampled using GIMP (version 2.10.24) and figure panels were arranged using
538 Inkscape (version 1.0.2: 394de47547, 2021-03-26).

539 **scRNAseq experiment**

540 ***Sample size estimation***

541 To determine the number of cells that we needed to sequence for stable clustering, we used
542 data from the cortex, hippocampus and subventricular zone of two E18 mice
543 (https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.3.0/1M_neurons).

544 The data set was filtered for hippocampal cell populations based on the expression of
545 marker genes, namely *Zbtb20*, *Satb2*, *Etv1*, *Man1a*, *Pou3f1*, *Prox1*, *Pax6* and *Mef2c*. The
546 obtained data set served as a pilot data set to determine the number of cells for the planned
547 experiment. We used cells from the original (filtered) dataset to determine cluster stability
548 with varying numbers of cells (250, 500, 1000, 2000, and 3000). To determine the
549 clustering stability for a higher number of cells, exceeding the number of cells in the pilot
550 data, we trained the single-cell deep Boltzmann machine (scDBM)²⁶ on the pilot data set
551 and generated synthetic cells which we combined with the initial data.

552 The number of cells at which the clustering reached stability, meaning that there were no or
553 few outlier clustering solutions, was chosen as the optimal number of cells.

554 First, we subsampled 250, 500, 1000, 2000, and 3000 cells from the pilot data for 30 times
555 each and ran k-means clustering on each of these subsamples. From previous analyses, we
556 saw that 16 clusters were appropriate for the pilot data set, and therefore we defined $k = 16$
557 for the clustering. Next, we assessed clustering stability for each subsample using the
558 Davies-Bouldin Index (DBI). Additionally, we trained a scDBM on the pilot data. We
559 generated 1000 and 2000 (30 times each) additional cells, after which we combined the
560 synthetic cells with the pilot data and performed k-means clustering. A stable clustering

561 occurred at a cell count of 4000 cells, indicated by a generally low DBI and few outliers
562 after clustering the 4000 cells for 30 times.

563 ***Cell dissociation, sorting and sample processing for scRNAseq***

564 Three samples corresponding to 1 control and 2 cKOs at stage E16.5 were collected and
565 processed for scRNAseq analysis. The hippocampi of E16.5 embryos were dissected on
566 cold HBSS (Hank's Balance Salt Solution) buffer and transferred to 1.5 ml tubes
567 containing the dissociation solution (Papain + DNase). The dissociation procedure was
568 carried out with the Papain Dissociation System (Worthington-Biochem, USA) following
569 the manufacturer's instructions with some few modifications. Briefly, all the reagents were
570 reconstituted in the respective solutions and equilibrated with 95% O₂ : 5% CO₂ if needed.
571 The dissociation solution (Papain + DNase) was pre-incubated at 37°C and aliquoted in 1.5
572 ml tubes for sample collection. Tissue samples were disrupted inside the tubes using a p200
573 pipette tip precoated with normal goat serum and then incubated at 37°C in a water bath for
574 approximately 30 min with manual inversion every 5 min. After the incubation time
575 samples were pipetted up and down, as mentioned above, for obtaining a single cells
576 suspension. The tubes were left undisturbed for 60 sec and the supernatants were
577 transferred to another tube and pelleted down by centrifugation at 300 rcf, 4°C, 5 min.
578 Supernatants were discarded and the cells were resuspended with resuspension solution
579 (DNase + Ovomuroid inhibitor) using p200 pipette tips precoated with serum, and
580 immediately and carefully transferred on top of another set of tubes containing Ovomuroid
581 inhibitor solution. The tubes were then centrifuged at 76 rcf, 4°C, 6 min, the supernatants
582 were removed, and the cell pellets resuspended in 200µl HBSS buffer and kept on ice.

583 The cells were stained with Zombie Green™ Fixable Viability Kit (BioLegend, USA) in
584 PBS for 15 min at RT and in the dark using a dilution of 1:500. After that the cells were
585 washed with 1ml 1x MojoSort™ Buffer (BioLegend, USA), centrifuged at 300 rcf for 5
586 min, resuspended in 2 - 3 ml 1x MojoSort™ Buffer and passed through a 30 µm filter
587 (Sysmex, CellTrics). Finally live cells were sorted on a MoFlo XDP Cell Sorter (Beckman
588 Coulter, USA) into 384-well plates containing prepared lysis buffer and processed with the
589 CEL-Seq2 modified protocol (Sagar et al., 2018). Briefly, the single cells were sorted in
590 384 well plates containing RT (reverse transcription) primers (anchored polyT primers

591 having a 6 bp cell barcode, 6 bp unique molecular identifiers (UMIs), a part of 5' Illumina
592 adapter and a T7 promoter), dNTPs, Triton X-100 and Vapor-Lock (Qiagen). The MoFlo
593 XDP Cell Sorter was calibrated for dispensing the single cells in the centre of each well
594 prior to sorting and the machine was run using trigger pulse width to exclude doublets. The
595 384 well plates containing the sorted cells were centrifuged at maximum speed and stored
596 at -80°C until library preparation as described in Sagar et al., 2018. Libraries were
597 sequenced on a HiSeq3000 Illumina sequencing machine and the demultiplexing of the raw
598 data was performed by running bcl2fastq (version 2.17.1.14.).

599 ***scRNAseq data analysis***

600 *Quantification of transcript abundance*

601 BWA (version 0.6.2-r126) was used to align paired-end reads to the transcriptome using
602 default parameters (H. Li & Durbin, 2010). The transcriptome was based on all gene
603 models from the mouse ENCODE VM9 release from the UCSC genome browser, which
604 contained 57207 isoforms with 57114 isoforms mapping to fully annotated chromosomes
605 (1-19, X, Y, M). Gene isoforms were merged per gene to a single gene locus and gene loci
606 were grouped to gene groups, if there was >75% overlap, resulting in 34111 gene groups.
607 The right mate of the read pair was mapped to all gene groups and multimapping reads
608 were excluded. The left mate carried the barcode information with 6bp corresponding to the
609 UMI and the following 6bp to the cell barcode. The remaining sequence contained a
610 poly(T) and adjacent gene information but was not used for quantification. Transcript
611 counts were obtained by aggregating the number of UMI per gene locus and for every cell
612 barcode⁵⁹.

613 *Filtering, normalization, dimensionality reduction and clustering*

614 The output count matrices generated were combined using R (versions 3.6.3 and 3.5.1) and
615 Rstudio (version 1.2.5001, build 93 (7b3fe265, 2019-09-18)) and the data filtering,
616 normalization, dimensionality reduction and clustering analyses were performed with the
617 RaceID package (version 0.1.3). As pre-processing steps, the genes with names starting
618 with “mt”, “Gm” or “Rik” were filtered out from the expression matrix, and cells with less
619 than 500 total counts were removed. For the filtering step we ran RaceID filterdata function

620 with default parameters after adjusting *minnumber*=1, *CGenes*= *c*(“*Mki67*”, “*Pcna*”) and
621 *Fgenes*=*c*(“*Kcnq1ot1*”). A total of 3701 cells were kept after filtering (control = 1287;
622 cKO1= 1102; cKO2= 1312) and used for further analysis. For the computation of the
623 distance between cells we used the default RaceID settings and for the clustering we set the
624 *clustexp* function parameter *cln*=16, which represents the number of clusters at which the
625 change in the log of within cluster dispersion became small. Additionally, outlier estimation
626 was performed using the default parameters.

627 *Sub-clustering analyses*

628 For sub-clustering analyses the count matrix was subset using the information of the cell
629 names belonging to the specific clusters. The RaceID pipeline was run using as input the
630 subset data and following the steps described above, setting *cln*=2 and the *metric*=
631 “*logpearson*” as the metric used for computing the distance between cells. The selection of
632 2 as the expected number of clusters was based on the observed gene expression patterns on
633 the t-SNE representations. We proposed that the 2 expected clusters should represent: less
634 and more mature cells. On the other hand, we defined log-Pearson as the distance parameter
635 for the clustering step. This selection was based on the idea that a putative unresolved
636 heterogeneity could be driven by lowly expressed transcripts that were not considered for
637 the clustering step in our initial analysis.

638 *Analysis of the effect of sex and genotype on clustering*

639 To ensure that there were no profound differences between mice of different sex that could
640 potentially confound our analyses, we visualized the biological variation in the data via a
641 denoised principal components analysis (PCA)⁶⁰ using the *scrn* package (version 1.16.0).
642 Here, a threshold value for the number of principal components is calculated based on the
643 proportion of variance explained by the biological components in the data. We separated
644 male and female mice using the expression of *Xist* gene (2 males: 1 control and 1 cKO; and
645 1 cKO female). By visual examination of the first five principal components, we could not
646 detect any substantial differences between the sex of the mice and thus ruled out the sex as
647 confounding factor. A similar analysis performed using the genotype (control versus cKO)
648 as grouping factor, which showed the absence of major effects due to the genotype.

649 *Differential gene expression analysis*

650 Analyses of the differentially expressed genes were performed with RaceID built in
651 function. Briefly, a background model for the expected transcript count variability is
652 internally computed by RaceID algorithm, and used for inferring negative binomial
653 distributions reflecting the gene expression variability within each group. A P-value for the
654 observed difference in transcript counts between the two compared groups is calculated
655 based on the inferred distributions, and subsequently adjusted for multiple testing using the
656 Benjamini–Hochberg method. Expression of the *Xist* gene was removed from the DEGs for
657 the distinct plots generated.

658 *Comparative analysis of cell proportions*

659 Comparison of cell population proportions between control and cKOs per cluster was
660 carried out in a 1 versus 1 basis using the Fischer’s exact test, by means of the `fisher.test`
661 function included in R. Additionally, in order to correct for multiple comparisons a post hoc
662 adjustment of the obtained P-values was performed using the `p.adjust` function with
663 Bonferroni as the correction method. We considered a change in proportions between
664 conditions for a specific cluster to be significant only if both cKO samples showed an
665 adjusted P-value less than 0.05 when individually compared to the control sample.

666 **RNA velocity analysis**

667 *Single cell expression data analysis*

668 Reference and Quantification:

669 In order to estimate levels of both spliced transcripts and introns for use in RNA velocity
670 analysis, a reference for pseudo-alignment was constructed similar to the "alevin_sep_gtr"
671 method described⁵⁰.

672 Mouse genome sequence and transcript annotation were obtained from Ensemble release 98
673 (`fasta/mus_musculus/dna/Mus_musculus.GRCm38.dna.primary_assembly.fa` and
674 `gtf/mus_musculus/Mus_musculus.GRCm38.98.gtf` files from
675 <http://ftp.ensembl.org/pub/release-98/>). The sequences of transcripts and introns were
676 extracted using R (<https://r-project.org>, version 3.6) and the `eisaR` package (version 0.8
677 available from

678 <https://github.com/fmicompbio/eisaR/tree/051496c6f90a7a4320821a8f3c0518fb4cf4a85a>
679 using `extractTxSeqs` with `type = "spliced"` and `extractIntronSeqs` with `type = "separate"`,
680 `flanklength = 50`, `joinOverlappingIntrons = FALSE`. A transcript-to-gene table was created
681 which links transcripts to genes, and introns to distinct genes for simultaneous
682 quantification (see below). The whole procedure was then performed as described in [http://](http://bioconductor.org/packages/release/bioc/vignettes/eisaR/inst/doc/rna-velocity.html)
683 bioconductor.org/packages/release/bioc/vignettes/eisaR/inst/doc/rna-velocity.html.
684 Extracted sequences were indexed with Salmon (version 1.1.0)⁶¹ with arguments `-k 23 --`
685 `type puff --gencode` and using the genome as a decoy. Reads were quantified using
686 Salmon/Alevin (version 1.1.0)⁶² with parameters `--celseq2` and the transcript-to-gene map
687 created above.

688 *Quality control and filtering:*

689 Technical quality of single cell experiments, cell barcode identification and quantification
690 were assessed using the Bioconductor package `alevinQC` (version 1.2.0,
691 <https://doi.org/doi:10.18129/B9.bioc.alevinQC>). Salmon/Alevin counts from all samples
692 were imported into R using the `tximeta` package (version 1.4.3)⁶³ and stored in a
693 `SingleCellExperiment` container for downstream analysis (55421 genes, 5678 cells). The
694 number of detected genes per cell and the fraction of counts in genes encoded on the
695 mitochondrion (`chrM`) were calculated using the `addPerCellQC` function from the `scater`
696 package (version 1.14.6)⁶⁴, and cells with more than 1000 detected genes and less than 25%
697 mitochondrial counts were retained (3972 cells). Cell size factors for normalization of raw
698 counts were calculated using the `scran` package (version 1.14.6)⁶⁰, by first clustering the
699 cells using the `quickCluster` function with `method="igraph"` and then `computeSumFactors`
700 with `min.mean=0.5`, which implements the deconvolution strategy described⁶⁵. Log-
701 transformed normalized counts were then calculated using the `logNormCounts` function
702 from the `scater` package.

703 *Velocity estimation*

704 RNA velocity analysis was performed with the `scvelo` package (version 0.1.25)³³. In order
705 to match the original cell annotations, the matrices generated with the Alevin pipeline were

706 filtered using the cell IDs that were kept in the initial analysis with RaceID algorithm. The
707 filtered matrices were then used as input for the scvelo pipeline.

708 **Diffusion maps and diffusion pseudotime analyses**

709 Diffusion maps and diffusion pseudotime were performed with the destiny package
710 (version 2.14.0)⁶⁶. Firstly, an object of the single cell experiment class was created with the
711 SingleCellExperiment package (version 1.6.0) using as input the raw expression matrix
712 obtained after filtering out the clusters corresponding to the INs, red blood cells, OPCs,
713 pericytes, choroid plexus, microglia, Cajal-Retzius, subiculum and unknown clusters. The
714 diffusion map was run with the DiffusionMap function using as values for the sigmas and k
715 parameters the outputs of running the functions find_sigmas and find_dm_k, respectively.
716 The diffusion pseudotime was subsequently computed using as input the diffusion maps
717 object into the DPT function with default arguments.

718 **Cell fate bias estimation**

719 Analyses of the cell fate bias probabilities were performed by applying a RNA velocity
720 independent, FateID (version 01.9)³⁷, and a RNA velocity dependent, CellRank (version
721 1.2.0)³⁶, algorithm. Both analyses were run after filtering out the clusters corresponding to
722 the INs, red blood cells, OPCs, pericytes, choroid plexus, microglia, Cajal-Retzius, and
723 unknown clusters. Cell fate probabilities were computed after presetting CA1 and CA3
724 pyramidal, DG granule, Subic1 and Subic2 cell populations as the final states. The original
725 tSNE projections computed by RaceID were provided as the dimensionally reduced
726 coordinates for data visualization.

727 **Functional annotation analysis**

728 To identify the main processes changing along the differentiation trajectory, functional
729 over-representation analysis of gene ontology biological processes was performed with the
730 R package clusterProfiler (version 3.12.0)⁶⁷ using as input the genes represented in the
731 respective nodes. For identification of the main biological processes related to the DEGs
732 between control and cKOs in each of the individual clusters genes increased and decreased
733 in expression were considered independently. Analyses of the main processes represented

734 by the DEGs increasing in expression between the more mature and the less mature CA cell
735 states, as well as in the case of the processes represented by the DEGs increasing or
736 decreasing in expression among control and cKOs in the less mature CA cell state, were
737 performed selecting the gene ontology for molecular functions. In all the independent
738 analyses a significance *q-value* cutoff of 0.05 was selected.

739 **Transcription factors analysis**

740 The exploratory analyses of the transcription factors changing in expression upon DOT1L
741 depletion and differentially expressed among CA maturation states were performed by
742 intersection of the respective DEGs with a list of *mus musculus* reported transcription
743 factors obtained from the TRRUST (version 2) database (available at:
744 <https://www.grnpedia.org/trrust/>), the database of transcription co-factors and transcription
745 factor interactions (TcoF-DB, version 2) (available at:
746 <https://tools.sschmeier.com/tcof/doc/>), and the AnimalTFDB (version 3.0) database
747 (available at: <http://bioinfo.life.hust.edu.cn/AnimalTFDB/#!/download>).

748 **Analysis of the H3K79me2 profiles**

749 The H3K79me2 ChIPseq data sets for NPC treated with DMSO or EPZ for 48h have been
750 previously published⁴⁷ and the files are accessible at the Gene Expression Omnibus
751 database under the accession number GSE135318 (GSM4005219:
752 NPC48h_DMSO_H3K79me2_rep2; GSM4005235: NPC48h_EPZ_H3K79me2_rep2).
753 H3K79me2 profiles for the selected genes were plotted using the pyGenomeTracks
754 package (version 3.6)⁶⁸ and assembled in inkscape.

755 **Data accessibility, interoperability and reproducibility**

756 The data files derived from the scRNAseq experiment discussed in this publication have
757 been deposited in NCBI's Gene Expression Omnibus⁶⁹ and are accessible through GEO
758 Series accession number GSE178105 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE178105>). All bioinformatics analyses steps and settings are deposited under the
759 link https://github.com/adsalas/DOT1L_hippocampus_development.
760

761 **Acknowledgments**

762 ASB, MT, TV, HB, DG are members of the Deutsche Forschungsgemeinschaft (DFG,
763 German Research Foundation)–funded 322977937/GRK2344 and thank for the support.
764 DG was also supported by the Max Planck Society and the DFG (GR4980/3-1). The
765 authors want to thank Dr. Mehmet Tekman (Department of Computer Science,
766 Bioinformatics Group, University Freiburg) for providing suggestions regarding data
767 analysis, and Dr. Sagar (Department of Medicine II, University Medical Center Freiburg)
768 for his kind support during the libraries preparation for the scRNAseq experiment.
769 Moreover, the authors thank Dr. Alejandro Villarreal, Dr. Henriette Franz, Dr. Nicole
770 Hellbach, Stefanie Heidrich, Ute Baur (Institute of Anatomy and Cell Biology, Department
771 of Molecular Embryology, University Freiburg) and HaiQin Zhang (RISE-Fellow) for
772 performing some initial staining experiments and for help with animal experimentation.
773 Further, the authors thank the Freiburg Galaxy Server team and all members of the Vogel
774 group for discussion and their support.

775 **Author contributions**

776 TV conceptualised the study. ASB, MT, HB and TV designed the experimental strategy.
777 ASB performed the experiments. ASB and TV analyzed and interpreted data and results.
778 JSH collaborated with the execution of the scRNAseq experiment and provided input for
779 data analysis. DK contributed with smFISH experiments. MT analyzed data and prepared
780 figures. MBS performed the quantification of exonic and intronic reads, and the scvelo
781 analysis. HB, and DG supervised and provided input for the data analysis. ASB and TV
782 wrote the manuscript and prepared the figures, with input from all co-authors. The
783 manuscript was read and approved by all co-authors.

784 **Competing interests**

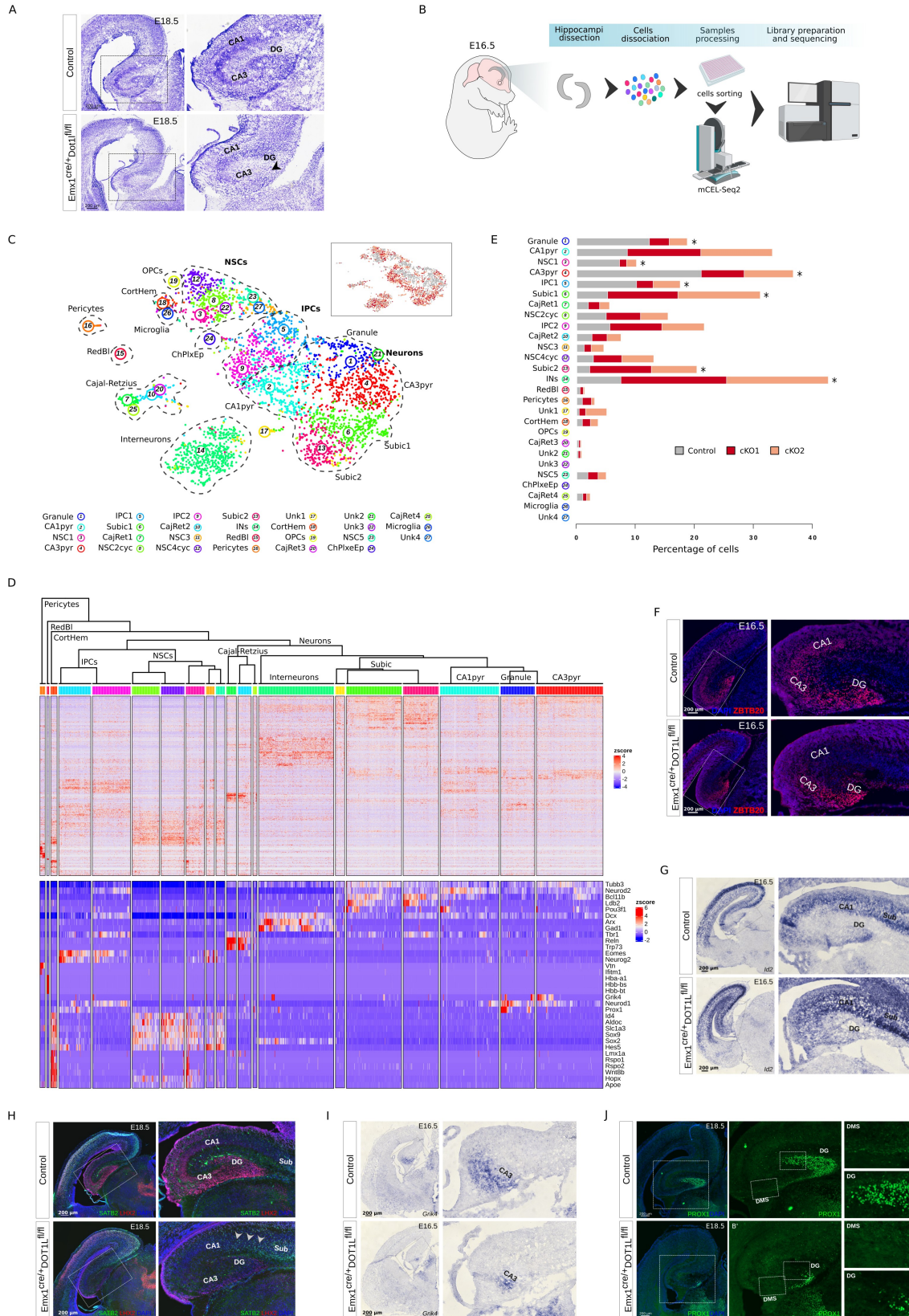
785 The authors declare no competing financial interests.

786 **References**

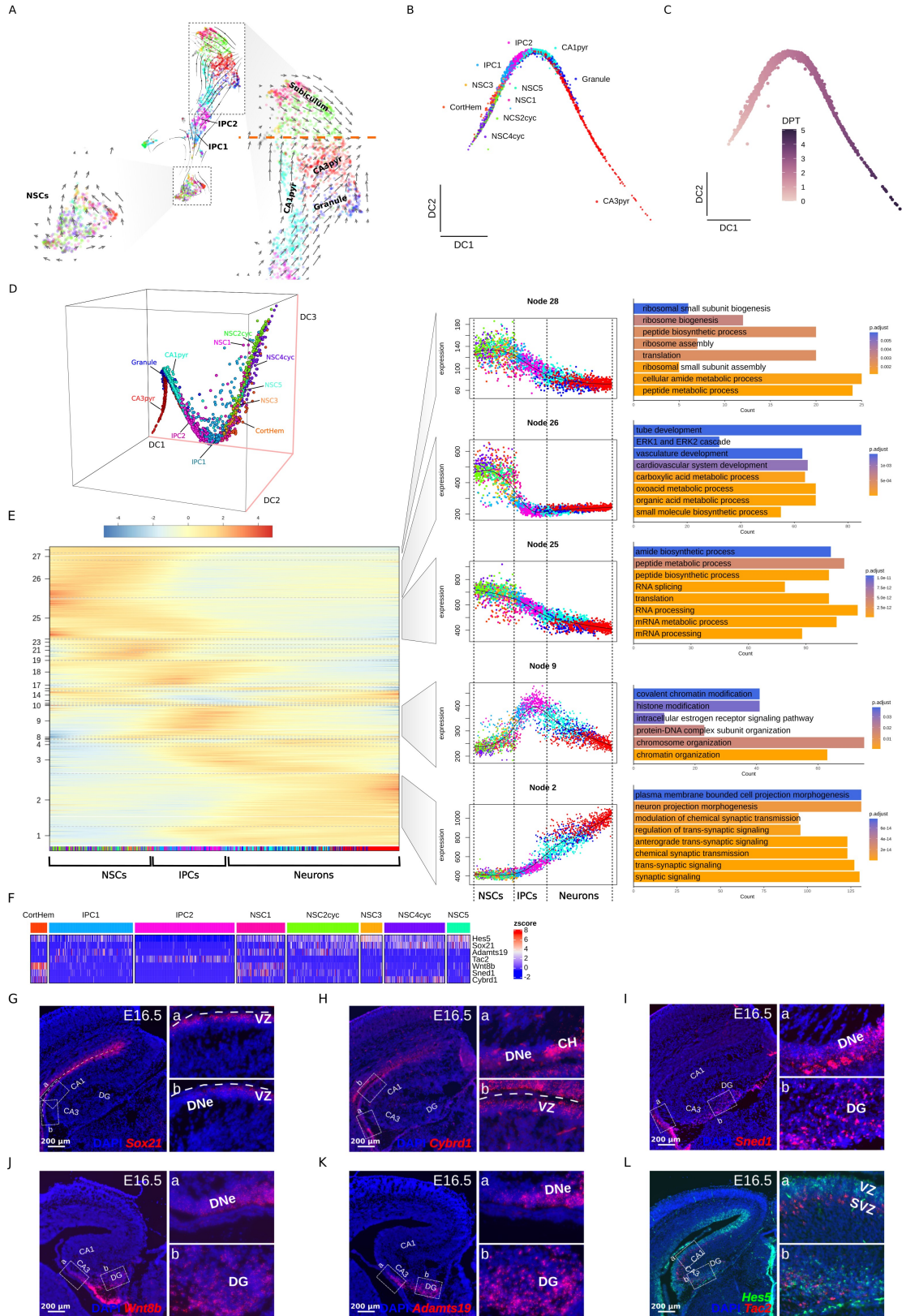
- 787 1. Kumamoto, T. & Hanashima, C. in *Neuroscience Research*, Vol. 86 37-49 (Elsevier
788 Ireland Ltd and Japan Neuroscience Society, 2014).
- 789 2. Fanselow, M.S. & Dong, H.W. in *Neuron*, Vol. 65 7-19 (Elsevier Inc., 2010).
- 790 3. Amaral, D. & Lavenex, P. in *The Hippocampus Book*, Edn. Vol. 1. (eds. P.
791 Andersen, R. Morris, D. Amaral, T. Bliss & J. O'Keefe) 37-110 (Oxford
792 Neuroscience, Oxford; 2007).
- 793 4. Berg, D.A. *et al.* in *Cell*, Vol. 177 654-668.e615 (Elsevier Inc., 2019).
- 794 5. Hochgerner, H., Zeisel, A., Lönnerberg, P. & Linnarsson, S. in *Nature*
795 *Neuroscience*, Vol. 21 290-299 (2018).
- 796 6. Tole, S., Christian, C. & Grove, E.A. in *Developmental Biology*, Vol. 4970 4959-
797 4970 (1997).
- 798 7. Grove, E.A., Kirkwood, T.B.L. & Price, J. in *Neuron*, Vol. 8 217-229 (1992).
- 799 8. Altman, J. & Bayer, S.A. in *Journal of Comparative Neurology*, Vol. 301 343-364
800 (1990).
- 801 9. Kitazawa, A. *et al.* in *Journal of Neuroscience*, Vol. 34 1115-1126 (2014).
- 802 10. Nelson, B.R. *et al.* in *eLife*, Vol. 9 1-30 (2020).
- 803 11. Arnold, S.J. *et al.* in *Genes and Development*, Vol. 22 2479-2484 (2008).
- 804 12. Barry, G. *et al.* in *Journal of Neuroscience*, Vol. 28 12328-12340 (2008).
- 805 13. Bulchand, S., Grove, E.A., Porter, F.D. & Tole, S. in *Mechanisms of Development*,
806 Vol. 100 165-175 (2001).
- 807 14. Galceran, J., Miyashita-Lin, E.M., Devaney, E., Rubenstein, J.L. & Grosschedl, R.
808 in *Development* (Cambridge, England), Vol. 127 469-482 (2000).
- 809 15. Galichet, C., Guillemot, F. & Parras, C.M. in *Development*, Vol. 135 2031-2041
810 (2008).
- 811 16. Miquelajauregui, A. *et al.* in *Proceedings of the National Academy of Sciences of*
812 *the United States of America*, Vol. 104 12919-12924 (2007).
- 813 17. Nielsen, J.V., Nielsen, F.H., Ismail, R., Noraberg, J. & Jensen, N.A. in
814 *Development*, Vol. 134 1133-1140 (2007).
- 815 18. Porter, F.D. *et al.* in *Development*, Vol. 124 2935-2944 (1997).
- 816 19. Rosenthal, E.H., Tonchev, A.B., Stoykova, A. & Chowdhury, K. in *Hippocampus*,
817 Vol. 22 2144-2156 (2012).
- 818 20. Xie, Z. *et al.* in *Proceedings of the National Academy of Sciences*, Vol. 107 6510-
819 6515 (2010).
- 820 21. Zhao, Y. *et al.* in *Science*, Vol. 284 1155-1158 (1999).
- 821 22. Franz, H. *et al.* in *Nucleic Acids Research*, Vol. 47 168-183 (Oxford University
822 Press, 2019).
- 823 23. Bovio, P.P. *et al.* in *Molecular Neurobiology*, Vol. 56 4273-4287 (*Molecular*
824 *Neurobiology*, 2019).
- 825 24. Gray De Cristoforis, A., Ferrari, F., Clotman, F. & Vogel, T. in *Molecular Brain*,
826 Vol. 13 1-20 (*Molecular Brain*, 2020).
- 827 25. Sagar, Herman, J.S., Pospisilik, J.A. & Grün, D. in *Methods in Molecular Biology*,
828 Vol. 1766 257-283 (2018).
- 829 26. Treppner, M. *et al.* in *Scientific Reports* 1-11 (Nature Publishing Group UK, 2021).
- 830 27. Grün, D. *et al.* in *Nature*, Vol. 525 251-255 (2015).
- 831 28. Artegiani, B. *et al.* in *Cell Reports*, Vol. 21 3271-3284 (ElsevierCompany., 2017).

- 832 29. Fotaki, V. *et al.* in *Developmental Dynamics*, Vol. 239 284-296 (2010).
833 30. Hatami, M., Conrad, S., Naghsh, P., Alvarez-Bolado, G. & Skutella, T. in *Frontiers*
834 in *Cellular Neuroscience*, Vol. 12 1-18 (2018).
835 31. Zhang, L. *et al.* in *Cell Death & Differentiation*, Vol. 27 1604-1617 (Springer US,
836 2020).
837 32. La Manno, G. *et al.* in *Nature*, Vol. 560 494-498 (2018).
838 33. Bergen, V., Lange, M., Peidli, S., Wolf, F.A. & Theis, F.J. in *Nature Biotechnology*,
839 Vol. 38 1408-1414 (Springer US, 2020).
840 34. Haghverdi, L., Buettner, F. & Theis, F.J. in *Bioinformatics*, Vol. 31 2989-2998
841 (2015).
842 35. Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F. & Theis, F.J. in *Nature*
843 *Methods*, Vol. 13 845-848 (2016).
844 36. Lange, M. *et al.* in *bioRxiv* (2020).
845 37. Herman, J.S., Sagar & Grün, D. in *Nature Methods*, Vol. 15 379-386 (2018).
846 38. Souopgui, J., Sölter, M. & Pieler, T. in *EMBO Journal*, Vol. 21 6429-6439 (2002).
847 39. Pleasure, S.J., Collins, A.E. & Lowenstein, D.H. in *Journal of Neuroscience*, Vol.
848 20 6095-6105 (2000).
849 40. Ravanpay, A.C., Hansen, S.J. & Olson, J.M. in *Molecular and Cellular*
850 *Neuroscience*, Vol. 44 178-189 (Elsevier Inc., 2010).
851 41. Sugimoto, Y., Furuno, T. & Nakanishi, M. in *Cell Biology International*, Vol. 33
852 174-179 (International Federation for Cell Biology, 2009).
853 42. Parisot, J., Flore, G., Bertacchi, M. & Studer, M. in *Development*, Vol. 144 2045-
854 2058 (2017).
855 43. De Clercq, S. *et al.* in *Cerebral Cortex*, Vol. 28 493-509 (2018).
856 44. Lee, Y. *et al.* in *NeuroReport*, Vol. 28 749-754 (2017).
857 45. Telley, L. *et al.* in *Science*, Vol. 351 1443-1446 (2016).
858 46. Chou, S.J., Perez-Garcia, C.G., Kroll, T.T. & O'Leary, D.D.M. in *Nature*
859 *Neuroscience*, Vol. 12 1381-1389 (2009).
860 47. Ferrari, F. *et al.* in *Nature Communications*, Vol. 11 1-13 (Springer US, 2020).
861 48. Bacon, C. *et al.* in *Molecular Psychiatry*, Vol. 20 632-639 (Nature Publishing
862 Group, 2015).
863 49. Cachia, A. *et al.* in *Translational Psychiatry*, Vol. 10 (Springer US, 2020).
864 50. Sonesson, C., Srivastava, A., Patro, R. & Stadler, M.B. in *PLoS Computational*
865 *Biology*, Vol. 17 (2021).
866 51. Sarkar, A. *et al.* in *Cell Stem Cell*, Vol. 22 684-697.e689 (Elsevier Inc., 2018).
867 52. Iwano, T., Masuda, A., Kiyonari, H., Enomoto, H. & Matsuzaki, F. in
868 *Development*, Vol. 139 3051-3062 (2012).
869 53. Simon, R., Wiegreffe, C. & Britsch, S. in *Frontiers in Molecular Neuroscience*, Vol.
870 13 1-10 (2020).
871 54. Li, L., Jin, J. & Yang, X.J. in *iScience*, Vol. 20 148-167 (Elsevier Inc., 2019).
872 55. MacDonald, J.L. & Roskams, A.J. in *Developmental Dynamics*, Vol. 237 2256-
873 2267 (2008).
874 56. Montgomery, R.L., Hsieh, J., Barbosa, A.C., Richardson, J.A. & Olson, E.N. in
875 *Proceedings of the National Academy of Sciences of the United States of America*,
876 Vol. 106 7876-7881 (2009).
877 57. Gorski, J.A. *et al.*, Vol. 22 6309-6314 (2002).

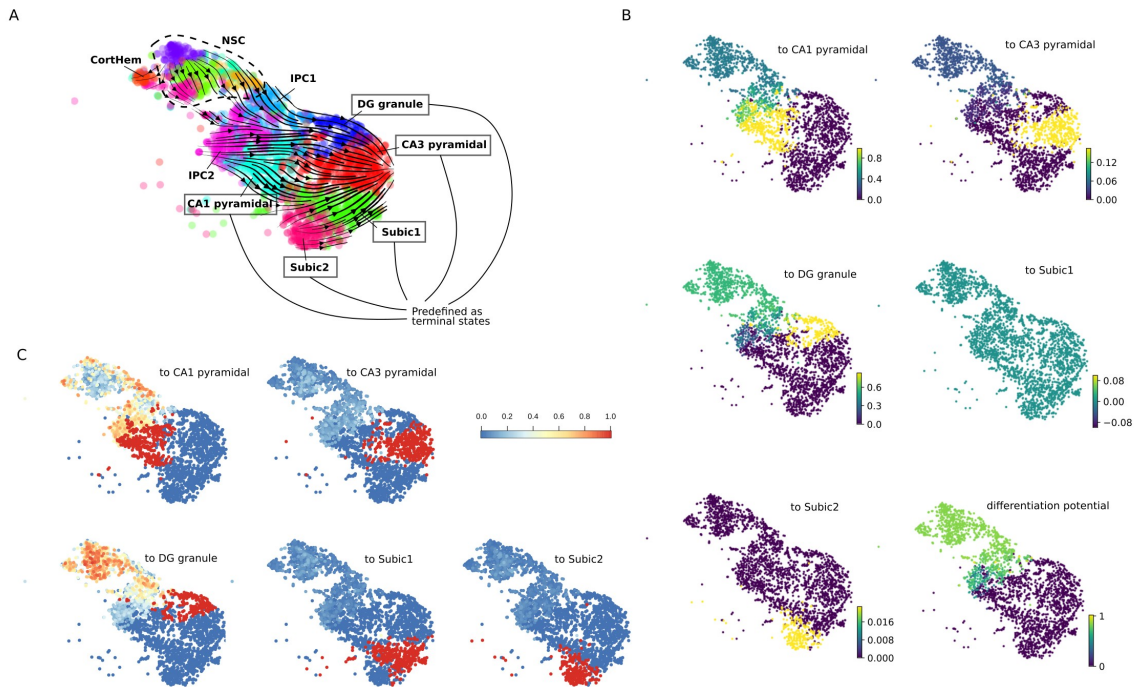
- 878 58. Hébert, J.M. & McConnell, S.K. in *Developmental Biology*, Vol. 222 296-306
879 (2000).
- 880 59. Grün, D., Kester, L. & Van Oudenaarden, A. in *Nature Methods*, Vol. 11 637-640
881 (2014).
- 882 60. Lun, A.T.L., McCarthy, D.J. & Marioni, J.C. in *F1000Research*, Vol. 5 1-69 (2016).
- 883 61. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. in *Nature*
884 *Methods*, Vol. 14 417-419 (Nature Publishing Group, 2017).
- 885 62. Srivastava, A., Malik, L., Smith, T., Sudbery, I. & Patro, R. in *bioRxiv* 1-16
886 (*Genome Biology*, 2018).
- 887 63. Love, M.I. *et al.* in *PLoS Computational Biology*, Vol. 16 1-13 (2020).
- 888 64. McCarthy, D.J., Campbell, K.R., Lun, A.T.L. & Wills, Q.F. in *Bioinformatics*, Vol.
889 33 1179-1186 (2017).
- 890 65. Lun, A.T.L., Bach, K. & Marioni, J.C. in *Genome Biology*, Vol. 17 1-14 (*Genome*
891 *Biology*, 2016).
- 892 66. Angerer, P. *et al.* in *Bioinformatics*, Vol. 32 1241-1243 (2016).
- 893 67. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. in *OMICS A Journal of Integrative*
894 *Biology*, Vol. 16 284-287 (2012).
- 895 68. Lopez-Delisle, L. *et al.* in *Bioinformatics*, Vol. 37 422-423 (2020).
- 896 69. Edgar, R., Domrachev, M. & Lash, A.E. in *Nucleic Acids Research*, Vol. 30 207-
897 210 (2002).



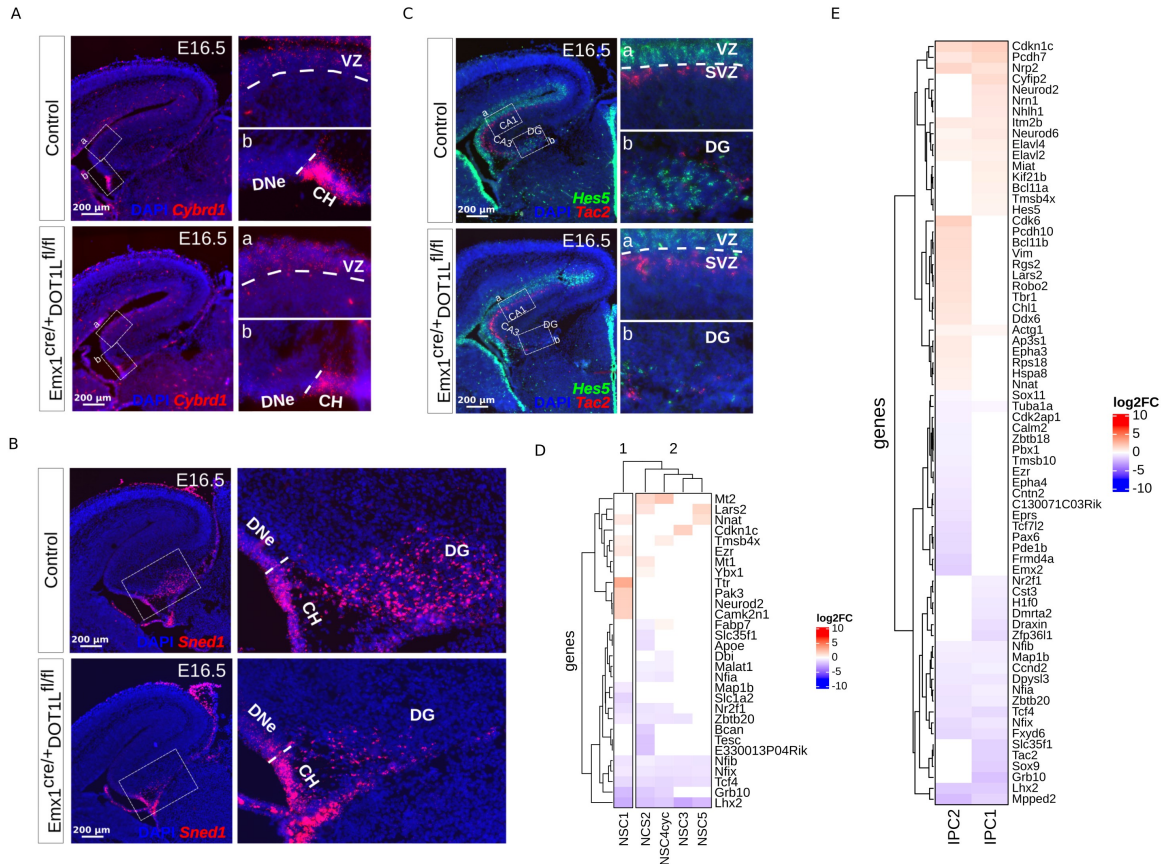
1 **Figure 1.** DOT1L-cKO alters hippocampal cytoarchitecture by affecting specific cell
2 populations. A) Nissl staining on E18.5 control and cKO brain sections (n = 3), showing
3 apparent loss of the DG region in the mutants (black arrowhead). B) Experimental design.
4 E16.5 hippocampi of control and cKO littermates were dissected, followed by cell
5 dissociation, single cells FACS sorting and libraries preparation using the mCEL-Seq2
6 protocol. C) tSNE plot showing the 27 clusters obtained after RaceID analysis. Inset shows
7 the distribution of control (grey) and cKO (red and salmon) cells in the tSNE
8 representation. D) Heatmaps showing the z-score normalized expression (red: high; blue:
9 low) of highly variable genes (top) and exemplary marker genes (bottom) used for cell
10 cluster annotation. Only clusters containing at least 10 cells are shown in the heatmaps.
11 Hierarchical clustering of cell populations was performed on highly variable genes based
12 on euclidean distance. E) Barplot showing the cell proportions in each cluster normalized
13 by sample. Clusters significantly changing in proportions based on Fisher's exact test
14 (adjusted $P < 0.05$) are indicated with asterisks. F) Immunofluorescence staining for
15 ZBTB20 on E16.5 control and mutant brain sections showing depletion of positive cells in
16 the cKOs (n = 3) in different areas of the hippocampal region. G) *In situ* hybridization for
17 *Id2* on E16.5 control and cKO brain sections indicating high expression at the level of the
18 putative CA1 in the mutants. H) Immunofluorescence co-staining for LHX2 and SATB2 on
19 E18.5 control and mutant brain sections showing depletion of LHX2 positive cells in the
20 cKOs (n = 3) in different areas of the hippocampal region and expansion of the SATB2
21 positive cells over the CA1 domain (white arrowheads). I) *In situ* hybridization for *Grik4*
22 on E16.5 control and cKO brain sections (n = 3). J) Immunofluorescence staining for
23 PROX1 on E18.5 control and mutant brain sections (n = 3) showing depletion of positive
24 cells in the cKOs in the DG region and along the DMS. Scale bars are indicated inside each
25 respective image. Sub: Subiculum; CA: cornu ammonis; DMS: dentate migratory stream;
26 DG: dentate gyrus; Pyr: pyramidal cell layer; IZ: intermediate zone.



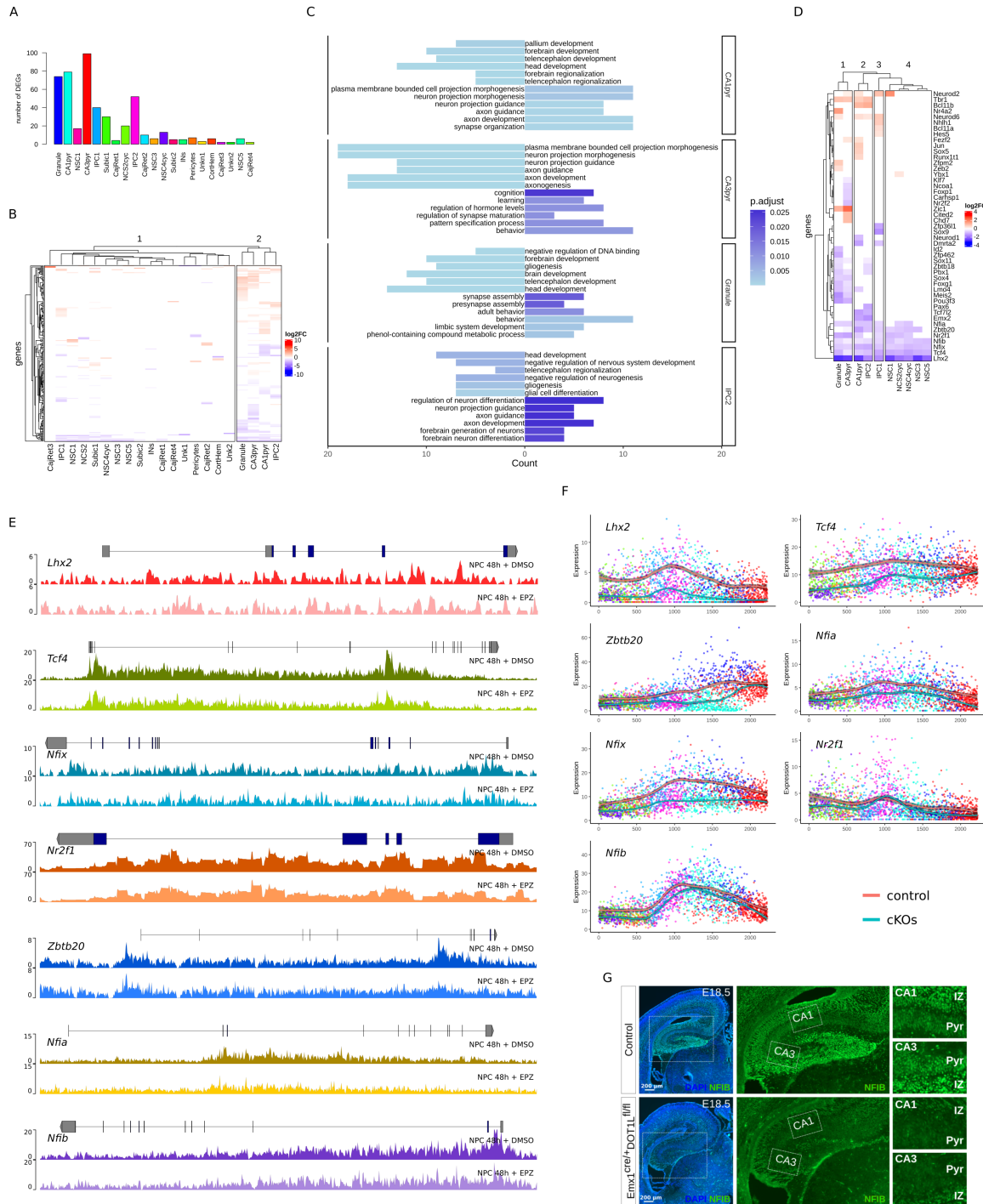
27 **Figure 2.** Lineage trajectory analysis identifies distinct and spatially separated progenitor
28 populations in the developing hippocampus. A) scvelo stream plot showing averaged
29 velocity vectors (arrows) indicative of the main transitions between the different cell
30 clusters identified at E16.5. Magnification panels for the NSCs and committed neuronal
31 populations are shown. Red dotted line highlights the point where the adjacent subiculum
32 and CA pyramidal cell populations meet each other. Clusters are color coded as in Figure
33 1C. B) Two-dimensional diffusion maps plot showing the lineage trajectory of the
34 hippocampal populations. Clusters are color coded as in Figure 1C. C) Two-dimensional
35 diffusion pseudotime plot showing increase in the pseudotime as cells progress from the
36 NSC to the committed cell states. D) Three-dimensional diffusion maps plot depicting the
37 separation of NSC4cyc and NSC2cyc populations, apparently intermingled in B. Clusters
38 are color coded as in Figure 1C. Red lines in the axes indicate the 3 distinct diffusion
39 components (DC). E) Self organizing map showing the expression levels of gene sets
40 (nodes) changing along the diffusion pseudotime axis (left). The running mean expression
41 for selected nodes is shown for the different cell clusters (middle). Barplots indicating the
42 main biological processes enriched ($q\text{-value} < 0.05$) in each highlighted node are shown
43 (right). F) Heatmap showing expression levels for putative marker genes for the NSCs and
44 IPCs clusters. G-L) smFISH stainings for selected NSCs and IPCs markers showing their
45 spatial expression domains on E16.5 control brain sections ($n = 3$). Scale bars are indicated
46 inside each image panel. DC: diffusion component; CA: cornu ammonis; DNe: dentate
47 neuroepithelium; DG: dentate gyrus; VZ: ventricular zone; SVZ: subventricular zone; CH:
48 cortical hem.



49 **Figure 3.** NSCs and IPCs show cell fate biases towards CA1 pyramidal or DG granule cells
50 in the E16.5 hippocampus. A) Stream plot showing the average velocity vectors projected
51 on the tSNE representation of the hippocampal populations. The mature cell populations
52 preset as terminal states are indicated in the figure. Clusters are colour coded as in Figure
53 1C. B) tSNE plots showing the differentiation potential and the estimated fate probabilities
54 towards the distinct terminal states as computed by CellRank analysis. Scale bars indicate
55 the probability values. C) tSNE plots showing the estimated fate probabilities towards the
56 distinct terminal states as computed by FateID analysis. Scale bar indicate the probability
57 values.

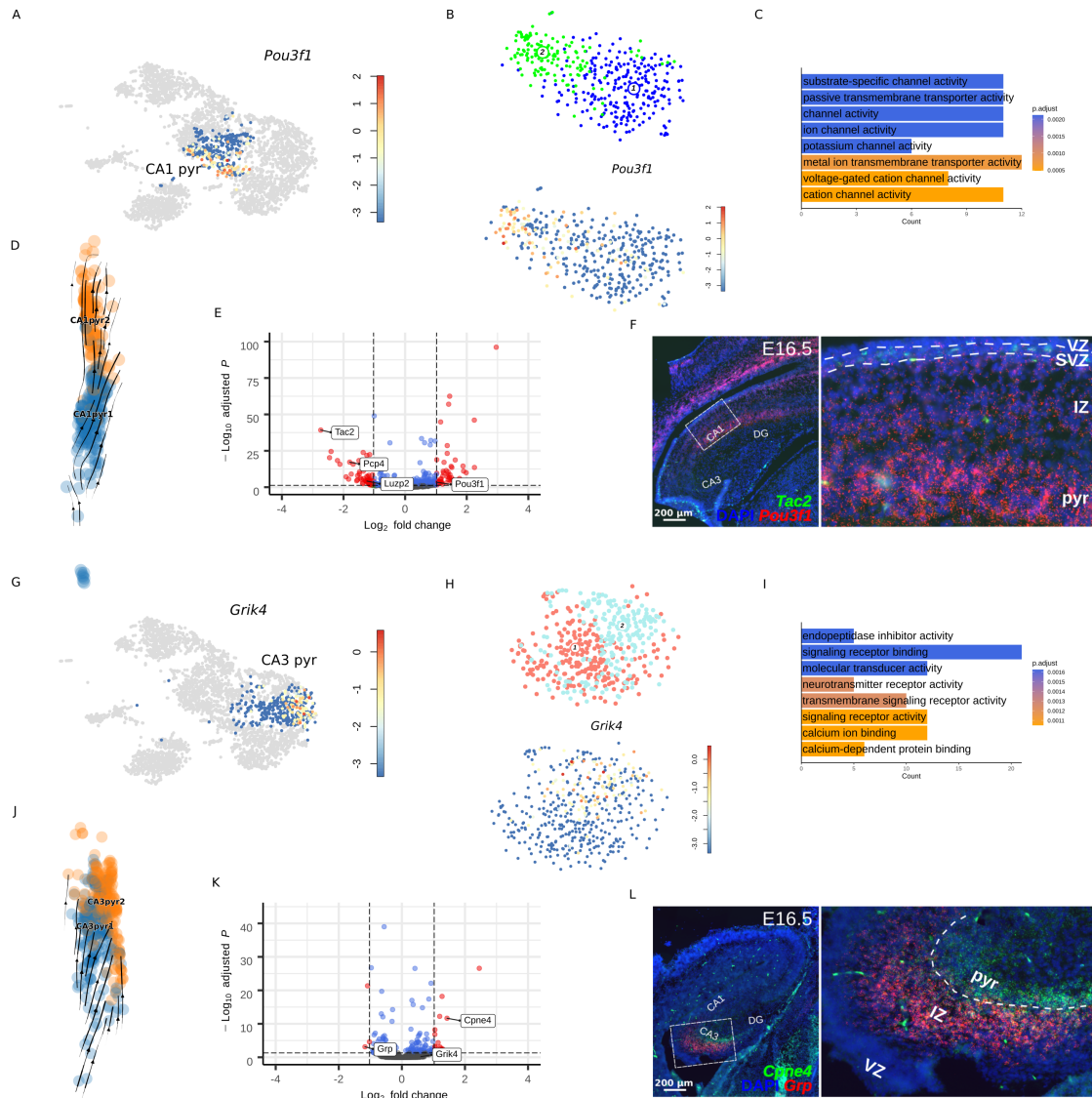


58 **Figure 4.** DOT1L is required for granule cells lineage generation and for controlling
 59 progression of IPCs towards differentiation. A-C) smFISH stainings for selected markers
 60 showing that NSC from the CANe are preserved (A) while granule cell lineage-related
 61 NSC (B) and IPC (C) are decreased upon DOT1L depletion (n = 3). Scale bars are
 62 indicated inside each image panel. DNe: dentate neuroepithelium; DG: dentate gyrus; VZ:
 63 ventricular zone; SVZ: subventricular zone; CH: cortical hem. D-E) Heatmaps of the DEGs
 64 (adjusted $P < 0.05$) between conditions for the NSC clusters (D) and IPC clusters (E). Scale
 65 corresponds to log₂ fold change between average expression for each condition within a
 66 cluster. Blue and red colors indicate significantly decreased or increased expression
 67 between DOT1L-cKO and control, respectively. Genes no significantly DE are depicted in
 68 white.



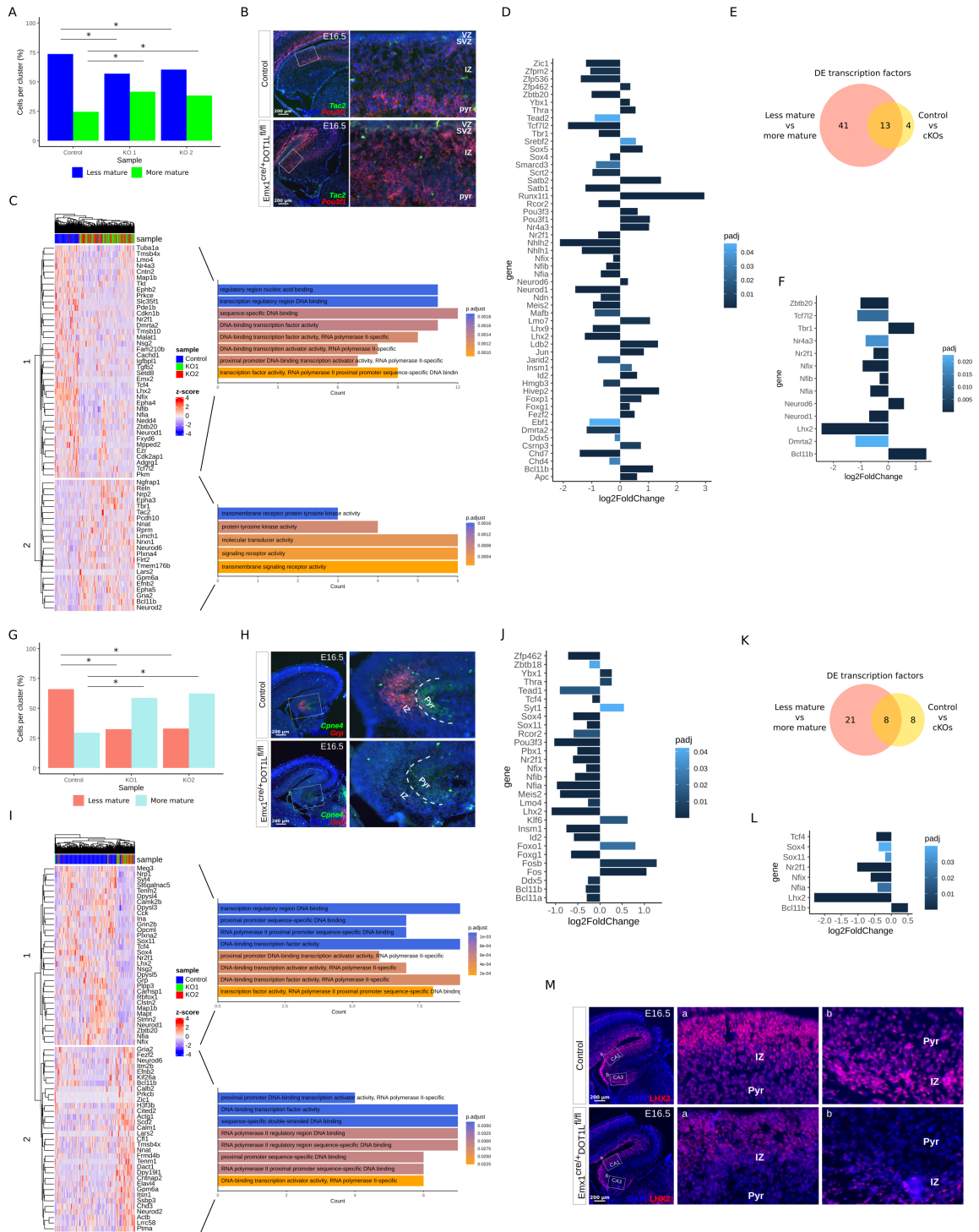
69 **Figure 5.** DOT1L prevents differentiation and activates master TFs involved in
 70 hippocampal development. A) Barplot showing the number of DEGs (adjusted P<0.05)
 71 obtained for the different cell clusters upon DOT1L deletion. B) Heatmap of the DEGs
 72 (adjusted P<0.05) between conditions for all the compared clusters. Scale corresponds to

73 log₂ fold change between average expression for each condition within a cluster. Blue and
74 red colors indicate significantly decreased or increased expression between DOT1L-cKO
75 and control, respectively. Genes no significantly DE are depicted in white color. C)
76 Barplots of the enriched biological processes (*q-value* < 0.05) obtained for the gene sets
77 significantly increased or decreased in expression in the respective clusters upon DOT1L-
78 cKO. D) Heatmap of the transcription factors significantly increasing or decreasing in
79 expression (adjusted *P*<0.05) in the DOT1L-cKOs for the clusters corresponding to the
80 hippocampal lineage. Scale corresponds to log₂ fold change between average expression
81 for each condition within a cluster. Colour scale as in B. E) H3K79me2 profiles for *Lhx2*,
82 *Tcf4*, *Nfix*, *Nr2f1*, *Zbtb20*, *Nfia* and *Nfib* in NPCs treated with DMSO (control) or EPZ for
83 48h showing decreased methylation signal in the gene bodies after EPZ treatment. Gene
84 structures are shown on the top of the first track for each gene. F) Expression profiles for
85 the genes in E along pseudotime axis. Red and light blue lines represent the smoothed
86 running mean expression for control and cKO cells, respectively. Grey shadow along each
87 line indicates a confidence interval of 0.95. Clusters are color coded as in Figure 1C. G)
88 Immunofluorescence staining for NFIB on E18.5 control and mutant brain sections (n = 3)
89 showing depletion of positive cells in the cKOs in different areas of the hippocampal
90 region. CA: cornu ammonis; Pyr: pyramidal cell layer; IZ: intermediate zone. Scale bars are
91 shown inside the image panels.



92 **Figure 6.** CA pyramidal cells are in two distinguishable maturation states at E16.5. A) and
 93 G) tSNE heatmaps showing non uniform expression of *Pou3f1* in CA1 pyramidal cells (A)
 94 and *Grik4* in CA1 pyramidal cells (G) (red: high; blue: low). Scale bars represent log2
 95 values. B) and H) tSNE plots showing the 2 sub-clusters obtained after subclustering
 96 analysis (top) and apparent enrichment of *Pou3f1* (B, bottom) or *Grik4* (H, bottom)
 97 expression in the sub-cluster 2. Scale bars represent log2 values. C) and I) Barplots of the
 98 enriched molecular functions (q -value < 0.05) obtained for the gene sets significantly
 99 increased in expression in sub-cluster 2 for each CA region. D) and G) scvelo stream plots

100 showing averaged velocity vectors (arrows) indicative of CA sub-cluster 1 cells movement
101 towards sub-cluster 2. E) and K) Volcano plots for the DEGs (adjusted $P < 0.05$) between
102 sub-cluster 2 and sub-cluster 1 highlighting putative markers for the more mature (right)
103 and less mature (left) cell states. F) and L) smFISH on E16.5 control brain sections ($n = 3$)
104 for selected cell state marker genes. Scale bars are shown inside the image panels. CA:
105 cornu ammonis; Pyr: pyramidal cell layer; SVZ: subventricular zone; VZ: ventricular zone;
106 IZ: intermediate zone.



107 **Figure 7.** Loss of DOT1L promotes maturation of CA pyramidal cells. A) and G) Barplots
 108 of cell proportions in each sub-cluster, normalized by sample. Significant changes in cell
 109 proportions based on Fisher's exact test (adjusted $P < 0.05$) are indicated with asterisks.

109 Colour codes are kept as in Figure 4 B and H (top). B) and H) smFISH on E16.5 control
110 and cKO brain sections (n = 3) for the selected cell state gene markers. C) and I) Heatmaps
111 (left) showing the DEGs (adjusted $P < 0.05$) between control and cKO samples in the less
112 mature CA sub-clusters. Genes are grouped based on hierarchical clustering and divided in
113 2 main clusters according to the direction of change. Barplots (right) indicating the main
114 biological processes enriched ($q\text{-value} < 0.05$) in each of the 2 gene clusters are shown. D)
115 and J) Barplots for the differentially expressed TF (adjusted $P < 0.05$) between the 2 CA
116 maturation states. Genes changing in expression in the more mature state relative to the less
117 mature state are shown. E) and K) Venn diagrams indicating the intersection of the TFs
118 differentially expressed between more and less mature sub-clusters with TFs differentially
119 expressed between conditions for the less mature cell state. F) and L) Barplots showing the
120 \log_2 fold change in expression for the differentially expressed TF (adjusted $P < 0.05$) in the
121 less mature cell state upon DOT1L depletion that are also differentially expressed between
122 maturation states. M) Immunofluorescence staining for LHX2 on E16.5 control and mutant
123 brain sections (n=3) showing depletion of positive cells in the IZ and Pyr upon DOT1L-
124 cKO. Scale bars are shown inside each image panel. CA: cornu ammonis; Pyr: pyramidal
125 cell layer; VZ: ventricular zone; SVZ: subventricular zone; IZ: intermediate zone.