The Pgbd5 DNA transposase is required for mouse cerebral cortex development through DNA double-strand breaks formation

Authors: Alessandro Simi[1], Federico Ansaloni[1], Devid Damiani[1], Azzurra Codino[1], Damiano Mangoni[1], Pierre Lau[1], Diego Vozzi[1], Luca Pandolfini[1], Remo Sanges[1,2*], Stefano Gustincich[1*]

Affiliations:

[1] Central RNA Laboratory, Istituto Italiano di Tecnologia (IIT); Genova, 16152, Italy

[2] Area of Neuroscience, International School for Advanced Studies (SISSA); Trieste, 34136, Italy

* Corresponding authors:

Stefano Gustincich, PhD; stefano.gustincich@iit.it

Remo Sanges, PhD; remo.sanges@sissa.it

Abstract: Transposons are conserved mobile genetic elements and one of the major sources of genetic variability during organisms' evolution. PiggyBac Transposable Element Derived 5 (*Pgbd5*), an evolutionarily conserved vertebrate DNA transposase-coding gene, is highly expressed in the central nervous system during development and adult life, suggesting it may play a role in neuronal differentiation and maintenance.

Here, we show that *Pgbd5* controls the expression of pro-neuronal and proliferative genes in a cell autonomous manner, leading to a decreased generation of neurons with altered migration. This activity depends on the induction of endogenous DNA double-strand breaks (DSBs).

These results position *Pgbd5* as a crucial regulator of brain development through its endonuclease activity.

One-Sentence Summary: *Pgbd5* regulates cerebral cortex neurogenesis through DNA double-strand breaks modulation.

Main Text:

Transposable elements (TEs) are mobile genetic elements constituting a large portion of eukaryotic genomes. TEs mobilization along the genome is one of the major sources of genetic variability during organisms' evolution. TEs are subdivided in two main classes: retrotransposons (class 1) and DNA transposons (class 2). While the first replicates using RNA intermediates and reverse transcription, DNA transposons do not replicate but use the so called "cut and paste" mechanisms (*1–3*), whereby a transposon-encoded protein, typically a transposase, catalyzes the excision of the transposon from the host genome and its subsequent reintegration at a different genomic site. The last wave of DNA transposon mobilization within the human genome appears to have occurred between ~85 and ~40 million years ago (Mya), during early evolution of the primate lineage (*4*). The transposase enzymes encoded by TEs are able to interact with nucleic acids, or other proteins, and their biochemical activities may be exploited by the host cell upon domestication. A striking example of protein-coding genes derived from domesticated TEs, are the Transib DNA Transposon-derived RAG1/2 recombinases involved in the V(D)J somatic recombination in lymphocytes, a crucial mechanism for adaptive immune response (*5*).

Pgbd5 gene is related to piggyBac DNA transposons. *Pgbd5* was domesticated ~500 Mya (*6*, *7*) and it is almost exclusively expressed in the brain as well as in embryonic solid tumors (*7–9*). It encodes for an endonuclease and it is capable to induce stereotypical cut-and-paste DNA transposition in cells (*8*). In childhood solid tumors, its over-expression leads to widespread genomic rearrangements (*9*). The physiological role of *Pgbd5* in neurons is currently unknown.

The mammalian cerebral cortex is generated during embryonic development. In the mouse, starting from E11.5, neural progenitors cells (NPC) e.g. radial glia (RG) cells, residing in the ventricular zone (VZ) and intermediate progenitors (IP) in the subventricular zone (SVZ), undergo different rounds of proliferation producing glutamatergic projection neurons (*10*, *11*). Newborn neurons migrate through the cortex following an inside-first outside-last pattern crossing the intermediate zone (IZ) by RG-guided migration and finally populating the forming cortical plate (CP). The first neuronal layers to be generated are the deep layers (DL) constituted by layer 6 (L6) and 5 (L5), which will project to subcortical targets (striatum, thalamus, and spinal cord). Later on, late-born neurons will populate the upper layers (UL) 4 and 2/3 (L4 and L2/3) and establish mainly cortico-cortical callosal connections (*12*, *13*).

Given the almost exclusive expression of *Pgbd5* in the brain and the high degree of conservation among organisms, we hypothesized that *Pgbd5* may play a key role in cerebral cortex development.

Results

Pgbd5 is expressed in embryonic mouse neocortex and controls early neuronal commitment

We followed *Pgbd5 in-vivo* expression in the embryonic mouse neocortex by quantitative PCR (RT-qPCR) performed on wild-type (WT) mouse cortices from E12.5 to E18.5 (Fig. 1a). *Pgbd5* was expressed at all stages up to 2-fold at E18.5 compared to E12.5, showing an expression drop at E14.5. To investigate *Pgbd5* function in brain development *in-vivo*, we knocked down (KD) *Pgbd5* RNA by means of *in utero* electroporation (*iue*) (Fig. 1b). We injected a short hairpin RNA (shRNA)-containing plasmid targeting the mouse *Pgbd5* RNA (shPgbd5) at E12.5, right after cerebral cortex neurogenesis starts. As control, we injected a shRNA sequence taken from a

freely available plasmid (Addgene plasmid #69040), targeting no known sequences in human or mouse transcriptomes (shCtrl). We have searched for possible targets of the shCtrl and shPgbd5 against the whole mouse transcriptome using BLASTN in Ensembl release 109 (*14*). The only perfect match spanning the entire sequence (nucleotide 1 to 21) of the stem region resulted from the shPgbd5 on the *Pgbd5* transcript suggesting the absence of possible off-targets. In order to visualize electroporated cells, a pCAGGS-IRES-GFP vector was co-electroporated with a 1:1 molar ratio. In these experimental conditions, a reduction of ~75% in *Pgbd5* RNA expression was obtained as measured by RT-qPCR on FAC-sorted cells expressing GFP (GFP^+) at E14.5 (Fig. 1c) compared to shCtrl injected cortices. To study the acute impact of *Pgbd5* KD on NPCs and immature neurons generation, E12.5 electroporated brains were harvested 1 day post-injection (*dpi*) and processed for immunofluorescence (IF) analysis (Fig.1 d-g). While the number of Pax6-expressing RG cells and Tbr2-expressing IP cells remained unaltered, the number of cells expressing the proneuronal gene marker Neurog2 and the neuronal differentiation gene marker Neurod1 were significantly decreased upon shPgbd5 injection compared to shCtrl-treated cortices (Fig.1h). To further validate a change in the number of newborn neurons, we carried out an acute cell-cycle exit analysis by injecting EdU at E12.5, right after *iue*, and stained for the proliferation marker Ki67 24 h later (Fig.1i). A significant decrease of $Ki67^-$ EdU^+ GFP^+ cells exiting the cell cycle (i.e. neuronal committed cells) was observed in shPgbd5-injected cortices (Fig.1j). These data suggest that *Pgbd5* expression in the developing neocortex is important to regulate the early steps of neuronal commitment.

Pgbd5 modulates neocortical neurons differentiation

To study the impact of *Pgbd5* repression on newborn cortical neurons, we silenced *Pgbd5* at E12.5 and collected GFP^+ cortices 2 *dpi*. At this stage, immature neurons migrate through the IZ and start to populate the DL of the developing CP. By IF (Fig. 2a-e) we analyzed the expression of cerebral cortex markers. When compared to the total GFP^+ cell population, we observed a significant increase in the amount of GFP^+ cells expressing apical and basal NPCs markers (Pax6 and Tbr2 respectively) compared to shCtrl-injected cortices (Fig. 2a'-2b'). Furthermore, the percentage of cells expressing Neurod1, at this stage expressed in the proliferative zone and IZ, was significantly decreased when compared to shCtrl treated cells. These data suggest that *Pgbd5* KD increases the amount of proliferating progenitor cells at the expenses of neuronal differentiation.

We then carried out a migration analysis: the neocortex was subdivided into three main zones (CP, IZ and VZ+SVZ) (Fig.2f) and the distribution of GFP^+ cells in each zone was calculated. In *Pgbd5*-silenced cortices, GFP^+ cells located mostly in the VZ/SVZ and the amount of migrating neurons reaching the CP was significantly reduced (Fig. 2f') compared to control conditions. Moreover, when considering the total GFP^+ population, we observed a strong reduction of GFP^+ cells expressing the DL markers Tbr1 (L6 marker) and Ctip2 (L5 marker) (Fig. 2d-2e, 2d'-2e') in *Pgbd5*-silenced cells compared to shCtrl treated ones further supporting previous evidences. All together, these data suggest that *Pgbd5* is required in NPCs to generate neurons able to migrate through the neocortical thickness.

Pgbd5 knockdown impairs cell metabolism and neurogenesis

To identify the molecular pathways affected by *Pgbd5* KD we performed RNA sequencing (RNA-Seq) on FAC-sorted E14.5 GFP$^+$ cells from E12.5 shPgbd5- and shCtrl-injected cortices (Fig.3). *Pgbd5* silencing down-regulated 2879 genes and up-regulated 2334 genes (Table S1). The most up-regulated or down-regulated 50 differentially expressed genes (DE-genes), ranked by FDR, are shown in Fig.3a. Gene Ontology (GO) analysis revealed that the majority of up-regulated genes included components of the cell cycle (e.g. *Ccnd* genes, *Ccng* genes, *Cdk* genes) and nervous system development (e.g. *Lhx* genes, *Wnt* genes) (Fig.3c-d and Table S2). Among the up-regulated genes normally involved in the modulation of neocortex neurogenesis, we found NPCs markers *Pax6*, *Tbr2*, *Nestin* and the targets of the Notch pathway *Hes1* and *Hes5* (Fig.3d and Table S1). Remarkably, the up-regulation of the Notch pathway targets is accompanied by a strong down-regulation of the proneural genes of the *Neurod* family *(Neurod1*, *Neurod2* and *Neurod6*) supporting the observed neurogenesis impairment. Other down-regulated genes are involved in cell metabolism being part of mitochondrial oxidative processes and respiration (e.g. *Cox* genes, *Nduf* subunits, *Atp5* subunits), translation and mitochondrial ribosome structure (e.g. *Rpl* and *Mrpl* genes, *Rps* genes) (Fig.3b and Table S2). Among the significantly down-regulated genes we found *Ctip2*, *Fezf2*, *Satb2* which are neocortical layer markers and control neurons specification and laminar position. Moreover, genes controlling migration (*Dcx*, *Rnd2*, *Rac1*, *Rac3*, *RhoA*, *RhoB*) and neuronal cell morphology-polarity (*Sox11* and *Cdc42*) (Fig.3d and Table S1) resulted down-regulated further supporting the observed impaired migration. These data confirmed, at transcriptional level, what we observed from IF analysis suggesting that *Pgbd5* is a key player in the balance between cell proliferation/renewal and neuron differentiation during mouse neocortex development.

Expression of Pgbd5 regulates radial migration of neocortical neurons

To further characterize the impaired cell localization observed at E14.5, we analyzed E12.5 shPgbd5-injected brains at E18.5, a stage at which the final layered structure of the cerebral cortex is almost at completion and the majority of DL neurons reached their final location. By IF analysis (Fig. 4a-b) we observed that the number of *Pgbd5*-silenced GFP$^+$ cells expressing the DL markers Tbr1 and Ctip2 were significantly decreased compared to control conditions (Fig.4c). We therefore ran a migration analysis to study the behavior of GFP$^+$ *Pgbd5*-silenced cells. Upon subdividing the CP in 10 bins (Fig. 4d), we calculated the distribution of GFP$^+$ cells for each bin. After shPgbd5 injection, we had a significant accumulation of electroporated cells in the DL of the cortex compared to controls (Fig. 4e). During image acquisition, it was evident that in *Pgbd5* KD cortices, a fraction of GFP$^+$ cells was not able to complete migration and located below the CP (Fig.4d). By subdividing the cerebral cortex into three main areas including the white matter (WM) (UL, DL, WM) (Fig.4d) we could confirm that ~20% of cells localized just underneath the CP with less cells reaching the DL and UL (Fig. 4f). Taken together these data suggest that neurons depends on *Pgbd5* in order to migrate properly through the cerebral cortex thickness and reach their final position.

Pgbd5 dowregulation blocks physiological occurrence of DNA double-strand breaks in neocortical progenitors

An important feature characterizing VZ and SVZ NPCs is the presence of a conspicuous amount of DNA breaking points (*15*, *16*). Considering that Pgbd5 presents endonuclease activity (*8*), we tested if *Pgbd5* silencing was altering physiological DSBs in NPCs. By IF staining for the DNA

DSBs marker γH2AX, we confirmed a strong signal in neocortical VZ and SVZ NPCs in WT mice (Fig.5a). In shCtrl-injected embryos, the γH2AX signal pattern was equal to the WT conditions, excluding any possible surgery-dependent artifacts (Fig. 5b left). On the other hand, γH2AX signal was significantly reduced upon *Pgbd5* silencing, with a clear gap between electroporated and not-electroporated areas (Fig. 5b right). The quantification of γH2AX signal in GFP^+ cells population confirmed that *Pgbd5* repression strongly reduced the amount of DSBs in the VZ and SVZ (Fig.5c). Taken together these data show that *Pgbd5* is necessary for the induction of physiological DSBs occurring in neocortical NPCs during development.

Discussion

Neurogenesis, the process by which dividing NPCs produce post-mitotic neurons, involves stereotyped molecular pathways conserved among different organisms. The control of the balance between proliferation and differentiation is the key for a proper development of cerebral cortex layered structure. The neocortex cytoarchitecture is heavily controlled during embryonic development and refined during postnatal periods by genetics/epigenetics factors and activity (*17–19*). In this framework, neuronal delamination, migration and positioning in the developing cortex is crucial, leading to brain disorders when altered (*20*).

Our study shows for the first time that *Pgbd5*, the most evolutionarily conserved transposable element-derived nuclease gene in vertebrates (*7*), is involved in cerebral cortex formation.

In-vivo acute cell-autonomous *Pgbd5* knockdown during the early stages of neocortex development resulted in neurogenesis impairment. Both immunohistochemistry and transcriptomic analysis clearly show a massive shutdown of genes involved in the determination and maintenance of neuronal cells in the neocortex (e.g. *Neurod* genes, *Dcx*) in favor of genes associated to cell proliferation/renewal (e.g. *Hes* genes, *Nestin*, *Cyclin* genes) typical of NPCs (*21*).

Furthermore, *Pgbd5*-silenced NPCs, already at 2 *dpi*, gave rise to neurons showing an altered migration. This was confirmed by both migration analysis and RNA-seq showing a significant downregulation of neocortical migration-controlling genes such as small *GTPases* (*22*). The defective migration phenotype become more severe at perinatal stages with a conspicuous amount of misplaced cells below the cortical plate.

Interestingly, the GO analysis on the DE-genes revealed a strong impact of *Pgbd5* downregulation on cell metabolism: numerous genes encoding for the multisubunits complexes of the mitochondrial oxidative phosphorylation (OXPHOS) system (*Ndufs*, *Ndufa*, *Cox* and *Atp5* gene families) (*23*) were significantly decreased upon *Pgbd5* silencing. During differentiation of NPCs to neurons, cells rely less on glycolysis and become progressively more dependent on OXPHOS to generate energy to be spent for neuronal processes rather than proliferation (*24*, *25*). It remains to be determined whether the aforementioned metabolic switch is under direct control by *Pgbd5* or a consequence of an altered balance between proliferation and differentiation.

Strikingly, we show that absence of Pgbd5 almost abolishes occurrence of DSBs in the mouse embryonic cortex, linking physiological mechanisms of genomic plasticity with developmental processes as neurogenesis and radial migration. This is intriguing given that recurrent DSBs clusters have been shown to control the expression of neurodevelopmental genes in NPCs in a cell autonomous manner (*26*). Moreover, genes involved in non-homologous recombination repair upon DSB formation have been shown to be required for a proper brain development (*27*).

These results suggest a novel mechanism in which a domesticated transposon is required for the generation of endogenous DSBs in NPCs, and that this process is crucial for corticogenesis. Our study opens up to the intriguing hypothesis that DSBs may be non-random and not exclusively due to topological stress, but tightly regulated for the establishment of a concerted gene expression signature.

References and Notes:

1. D. Rodriguez-Terrones, M. E. Torres-Padilla, Nimble and Ready to Mingle: Transposon Outbursts of Early Development. *Trends Genet.* **34**, 806–820 (2018).
2. J. A. Erwin, M. C. Marchetto, F. H. Gage, Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat. Rev. Neurosci.* **15**, 497–506 (2014).
3. E. Chesnokova, A. Beletskiy, P. Kolosov, The Role of Transposable Elements of the Human Genome in Neuronal Function and Pathology. *Int. J. Mol. Sci*. **23**, 5847 (2022).
4. J. K. Pace, C. Feschotte, The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res.* **17**, 422–432 (2007).
5. V. V. Kapitonov, J. Jurka, RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol*. **3**, 0998–1011 (2005).
6. A. Sarkar, C. Sim, Y. S. Hong, J. R. Hogan, M. J. Fraser, H. M. Robertson, F. H. Collins, Molecular evolutionary analysis of the widespread piggyBac transposon family and related “domesticated” sequences. *Mol. Genet. Genomics*. **270**, 173–180 (2003).
7. T. Pavelitz, L. T. Gray, S. L. Padilla, A. D. Bailey, A. M. Weiner, PGBD5: A neural-specific intron-containing piggyBac transposase domesticated over 500 million years ago and conserved from cephalochordates to humans. *Mob. DNA.* **4**, 1–17 (2013).
8. A. G. Henssen, E. Henaff, E. Jiang, A. R. Eisenberg, J. R. Carson, C. M. Villasante, M. Ray, E. Still, M. Burns, J. Gandara, C. Feschotte, C. E. Mason, A. Kentsis, Genomic DNA transposition induced by human PGBD5. *Elife*. **4**, 1–20 (2015).
9. A. G. Henssen, R. Koche, J. Zhuang, E. Jiang, C. Reed, A. Eisenberg, E. Still, I. C. Macarthur, E. Rodríguez-Fos, S. Gonzalez, M. Puiggròs, A. N. Blackford, C. E. Mason, E. De Stanchina, M. Gönen, A. K. Emde, M. Shah, K. Arora, C. Reeves, N. D. Socci, E. Perlman, C. R. Antonescu, C. W. M. Roberts, H. Steen, E. Mullen, S. P. Jackson, D. Torrents, Z. Weng, S. A. Armstrong, A. Kentsis, PGBD5 promotes site-specific oncogenic mutations in human tumors. *Nat. Genet.* **49**, 1005–1014 (2017).
10. B. J. Molyneaux, P. Arlotta, J. R. L. Menezes, J. D. Macklis, Neuronal subtype specification in the cerebral cortex. *Nat. Rev. Neurosci.* **8**, 427–437 (2007).
11. P. Rakic, Evolution of the neocortex: A perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).
12. D. Jabaudon, Fate and freedom in developing neocortical circuits. *Nat. Commun.* **8**, 16042 (2017).
13. A. M. Oproescu, S. Han, C. Schuurmans, New Insights Into the Intricacies of Proneural Gene Regulation in the Embryonic and Adult Cerebral Cortex. *Front. Mol. Neurosci.* **14**, 1–24 (2021).

14. F. J. Martin, M. R. Amode, A. Aneja, O. Austine-Orimoloye, A. G. Azov, I. Barnes, A. Becker, R. Bennett, A. Berry, J. Bhai, S. K. Bhurji, A. Bignell, S. Boddu, P. R. Branco Lins, L. Brooks, S. B. Ramaraju, M. Charkhchi, A. Cockburn, L. Da Rin Fiorretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. G. Giron, T. Genez, G. S. Ghattaoraya, J. G. Martinez, C. Guijarro, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, M. Kay, V. Kaykala, T. Le, D. Lemos, D. Marques-Coelho, J. C. Marugán, G. A. Merino, L. P. Mirabueno, A. Mushtaq, S. N. Hossain, D. N. Ogeh, M. P. Sakthivel, A. Parker, M. Perry, I. Piližota, I. Prosovetskaia, J. G. Pérez-Silva, A. I. A. Salam, N. Saraiva-Agostinho, H. Schuilenburg, D. Sheppard, S. Sinha, B. Sipos, W. Stark, E. Steed, R. Sukumaran, D. Sumathipala, M. M. Suner, L. Surapaneni, K. Sutinen, M. Szpak, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, B. Walts, E. Wass, N. Willhoft, J. Allen, J. Alvarez-Jarreta, M. Chakiachvili, B. Flint, S. Giorgetti, L. Haggerty, G. R. Ilsley, J. E. Loveland, B. Moore, J. M. Mudge, J. Tate, D. Thybert, S. J. Trevanion, A. Winterbottom, A. Frankish, S. E. Hunt, M. Ruffier, F. Cunningham, S. Dyer, R. D. Finn, K. L. Howe, P. W. Harrison, A. D. Yates, P. Flicek, Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).

15. K. Onishi, A. Uyeda, M. Shida, T. Hirayama, T. Yagi, N. Yamamoto, N. Sugo, Genome stability by DNA polymerase β in neural progenitors contributes to neuronal differentiation in cortical development. *J. Neurosci.* **37**, 8444–8458 (2017).

16. N. Gionchiglia, A. Granato, L. Lossi, Brain from Embryonic Life to Old Age (2021), doi:10.3390/molecules26237198.

17. C. R. Cadwell, A. Bhaduri, M. A. Mostajo-Radji, M. G. Keefe, T. J. Nowakowski, Development and Arealization of the Cerebral Cortex. *Neuron.* **103**, 980–1004 (2019).

18. A. Simi, M. Studer, Developmental genetic programs and activity-dependent mechanisms instruct neocortical area mapping. *Curr. Opin. Neurobiol.* **53**, 96–102 (2018).

19. L. Telley, G. Agirman, J. Prados, N. Amberg, S. Fièvre, P. Oberst, G. Bartolini, I. Vitali, C. Cadilhac, S. Hippenmeyer, L. Nguyen, A. Dayer, D. Jabaudon, Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science (80-.).* **364** (2019), doi:10.1126/science.aav2522.

20. G. Sokpor, B. Brand-Saberi, H. P. Nguyen, T. Tuoc, Regulation of Cell Delamination During Cortical Neurodevelopment and Implication for Brain Disorders. *Front. Neurosci.* **16**, 1–21 (2022).

21. N. Bertrand, D. S. Castro, F. Guillemot, PRONEURAL GENES AND THE SPECIFICATION OF NEURAL CELL TYPES. **3** (2002), doi:10.1038/nrn874.

22. R. Azzarelli, F. Guillemot, E. Pacary, Function and regulation of Rnd proteins in cortical projection neuron migration. *Front. Neurosci.* **9**, 1–13 (2015).

23. W. J. H. Koopman, F. Distelmaier, J. A. M. Smeitink, P. H. G. M. Willems, OXPHOS mutations and neurodegeneration. *EMBO J.* **32**, 9–29 (2013).

24. M. Khacho, A. Clark, D. S. Svoboda, J. Azzi, J. G. MacLaurin, C. Meghaizel, H. Sesaki, D. C. Lagace, M. Germain, M. E. Harper, D. S. Park, R. S. Slack, Mitochondrial Dynamics Impacts Stem Cell Identity and Fate Decisions by Regulating a Nuclear Transcriptional Program. *Cell Stem Cell.* **19**, 232–247 (2016).

25. M. Khacho, R. Harris, R. S. Slack, Mitochondria as central regulators of neural stem cell

fate and cognitive function. *Nat. Rev. Neurosci.* **20**, 34–48 (2019).

26. P. C. Wei, A. N. Chang, J. Kao, Z. Du, R. M. Meyers, F. W. Alt, B. Schwer, Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell*. **164**, 644–655 (2016).

27. Y. Gao, Y. Sun, K. M. Frank, P. Dikkes, Y. Fujiwara, K. J. Seidl, J. M. Sekiguchi, G. A. Rathbun, W. Swat, J. Wang, R. T. Bronson, B. A. Malynn, M. Bryans, C. Zhu, J. Chaudhuri, L. Davidson, R. Ferrini, T. Stamato, S. H. Orkin, M. E. Greenberg, F. W. Alt, A Critical Role for DNA End-Joining Proteins in Both Lymphogenesis and Neurogenesis. **95**, 891–902 (1998).

Acknowledgments:

We are indebted to all the members of the SG, RS and RNA initiative labs for thought-provoking discussions. A special thanks to Andrea Contestabile for Neurod1 antibody. We are grateful to technical and administrative staff of IIT. We acknowledge that the research activity herein was carried out using the IIT HPC infrastructure.

Funding:

IIT intramural funding (SG)

Marie Curie MINDED grant 754490 (AS)

Author contributions:

Conceptualization: AS, DD, RS, SG

Methodology: AS, FA, DD, AC, DV, LP

Investigation: AS, FA, DD, AC, DM, PL, DV, LP

Visualization: AS

Funding acquisition: SG, RS

Project administration: SG, RS, AS

Supervision: SG, RS

Writing – original draft: AS, FA, SG, RS

Writing – review & editing: All authors

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials.

Supplementary Materials

Materials and Methods
Table S1-S5

Figure 1

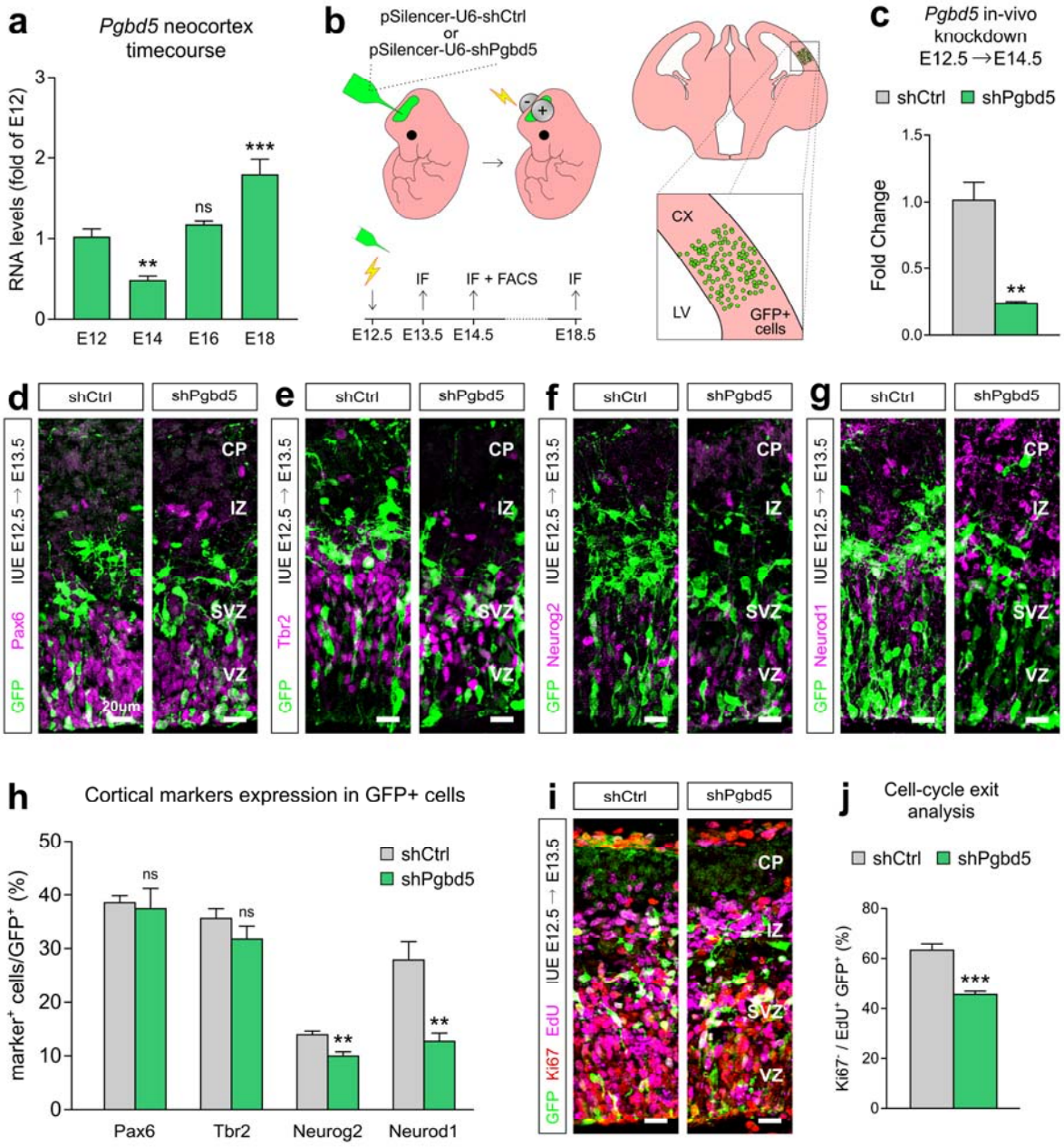

Fig.1 *Pgbd5* repression alters neurogenesis

a, Expression levels of *Pgbd5* in the developing mouse cerebral cortex. Data are mean ± s.e.m.; *n* = 6 independent biological samples. One-way ANOVA with Dunnett's multiple comparisons test. **b**, Schematic of *in utero* electroporation (*iue*) procedure (top left), experimental timeline (bottom left), electroporated GFP$^+$ cells in a coronal brain slice (top right) and electroporated cerebral cortex close-up (bottom right). **c**, Expression of *Pgbd5* RNA in GFP+ cells from E14.5 cortices injected with shCtrl or shPgbd5 at E12.5. RNA levels are normalized on shCtrl group. n = 3 technical replicates. Unpaired Student's t-test. **d-g**, Immunofluorescence (IF) staining for RG and IP cells (Pax6, Tbr2) and pro-neuronal (Neurog2, Neurod1) markers at E13.5, shRNAs *iue* at

E12.5. **h**, Quantification of marker^{+}/GFP^{+} cells at E13.5 in shPgbd5- vs shCtrl-injected brains. Data are mean ± s.e.m.; n = 4 each. Unpaired Student's t-test. **i**, IF staining at E13.5 for proliferation marker Ki67 and EdU injected at E12.5 right after *iue*. **j**, Quantification of Ki67^{-} EdU^{+} GFP^{+} differentiating cells 24 h after *iue* and EdU injection. Data are mean ± s.e.m.; n = 4 each. Unpaired Student's t-test. LV, lateral ventricle; CX, cerebral cortex; VZ, ventricular zone; SVZ, subventricular zone; IZ, Intermediate zone; CP, cortical plate. $^{**}P < 0.01$, $^{***}P < 0.001$, ns = not significant.

Figure 2

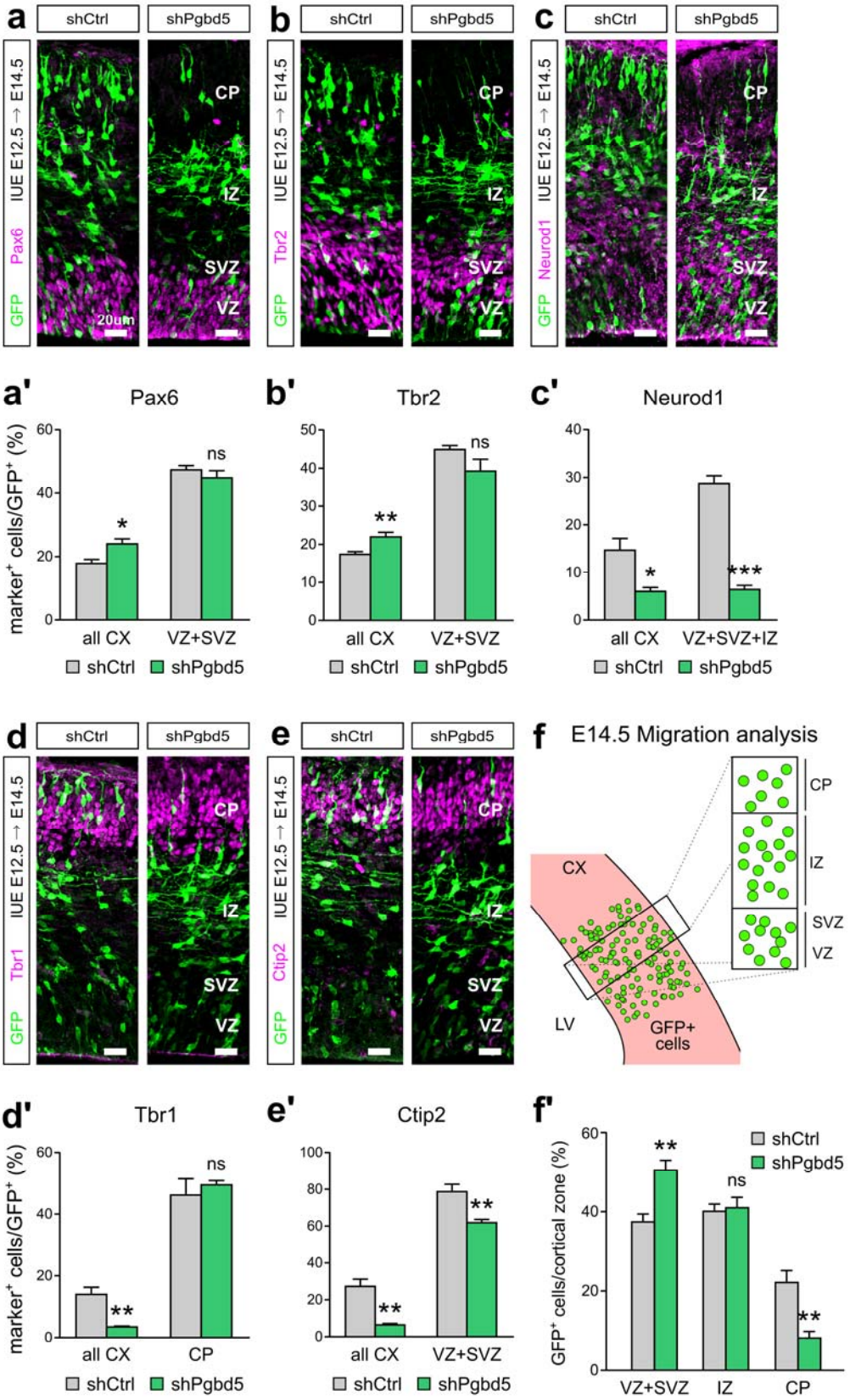
a
shCtrl
shPgbd5
IUE E12.5 → E14.5
GFP Pax6
CP
IZ
SVZ
VZ
20um
b
GFP Tbr2
c
GFP Neurod1
a'
Pax6
marker+ cells/GFP+ (%)
all CX
VZ+SVZ
ns
*
b'
Tbr2
**
c'
Neurod1
VZ+SVZ+IZ
***
d
GFP Tbr1
e
GFP Ctip2
f
E14.5 Migration analysis
CX
LV
GFP+ cells
d'
Tbr1
CP
e'
Ctip2
f'
GFP+ cells/cortical zone (%)
VZ+SVZ
IZ

Fig.2 *Pgbd5* is essential for proper neuronal differentiation and migration

a-e, IF staining for NPCs (Pax6, Tbr2), immature neuronal cells (Neurod1) and DL neurons (Tbr1, Ctip2) markers at E14.5, shRNAs *iue* at E12.5. **a'-e'**, Quantification of marker$^+$/GFP$^+$ cells at E14.5. Data are mean ± s.e.m.; n = 4 each. Unpaired Student's t-test. **f**, Schematic of migration analysis. **f'**, E14.5 GFP$^+$ cells distribution in VZ+SVZ, IZ and CP of E12.5 shPgbd5- vs shCtrl-injected brains. Data are mean ± s.e.m.; n = 4 each. Unpaired Student's t-test. VZ, ventricular zone; SVZ, subventricular zone; IZ, Intermediate zone; CP, cortical plate; LV, lateral ventricle; CX, cerebral cortex. *P < 0.05, **P < 0.01, ***P < 0.001, ns = not significant.

Figure 3

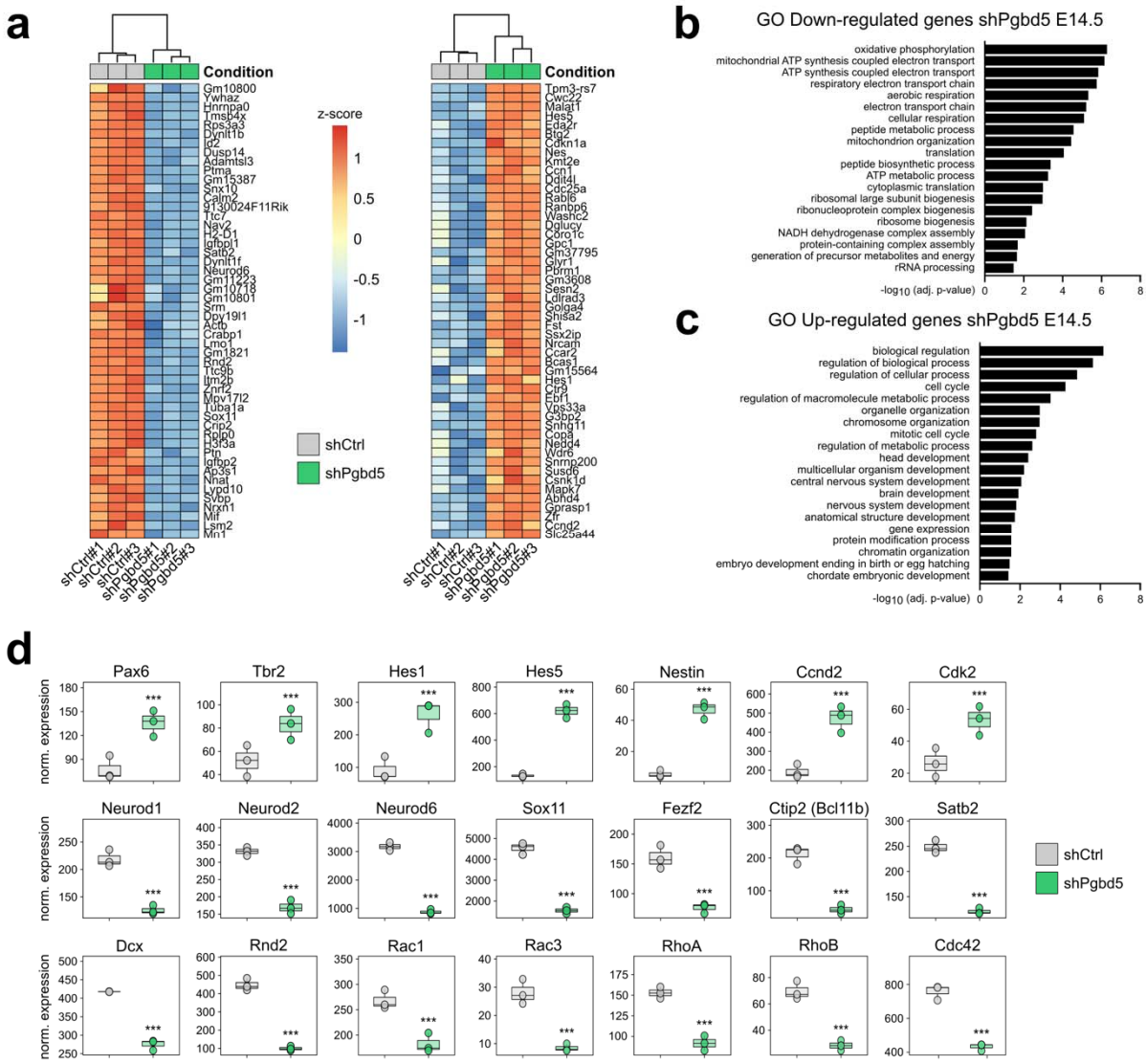

Fig.3 *Pgbd5* silencing impairs cerebral cortex transcriptome

a, Top 50 down and upregulated genes. Scaled normalized expression values (TMM) of the most significant (FDR) down- (left) and up-regulated genes (right) are reported in the heatmap. Gene names are reported in rows, whereas sample names in columns. Samples were clustered using the hclust "average" method. **b**,**c**, Top GO terms under the biological process category for down-regulated (b) and up-regulated (c) genes in mouse GFP^{+} cells after *iue* of shPgbd5 vs. shCtrl. **d**, Expression values of neural gene markers. Boxplots reporting the normalized expression levels (TMM) of specific NPCs and neuronal gene markers. Statistics is reported as FDR-corrected p-value from edgeR DE analysis (see methods). GO, Gene Ontology. ***$P < 0.001$.

Figure 4

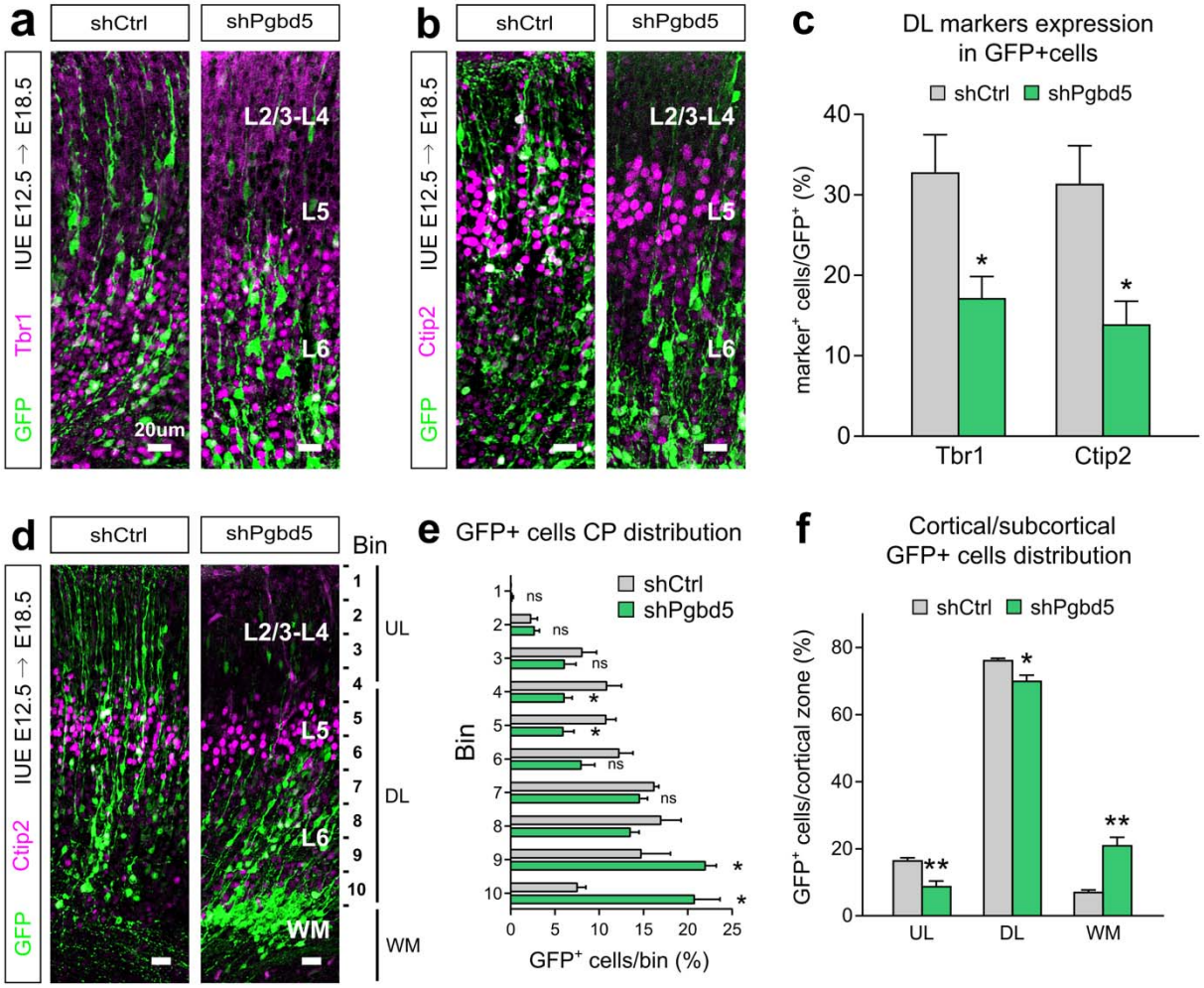

Fig.4 *Pgbd5* is required for neurons migration

a-b, IF staining for DL neuron markers (Tbr1, Ctip2) at E18.5, shRNAs *iue* at E12.5. **c**, Quantification of marker$^+$/GFP$^+$ cells at E18.5. Data are mean ± s.e.m.; n = 3 shCtrl, n = 4 shPgbd5. Unpaired Student's t-test. **d**, Representative image of cerebral cortex sub-division for migration analysis. **e**, E18.5 radial distribution of GFP$^+$ cells in the CP of E12.5 shRNAs-injected cortices. Data are mean ± s.e.m.; n = 3 shCtrl, n = 4 shPgbd5. Unpaired Student's t-test. **f**, Distribution of E18.5 GFP$^+$ cells in the three main cortical areas (UL, DL and WM). Data are mean ± s.e.m.; n = 3 shCtrl, n = 4 shPgbd5. Unpaired Student's t-test. L6, Layer 6; L5, Layer 5; L2/3-L4, Layers 2/3 and Layer 4; WM, white matter; DL, Deep Layers; UL, Upper Layers. *P < 0.05, **P < 0.01, ns = not significant.

Figure 5

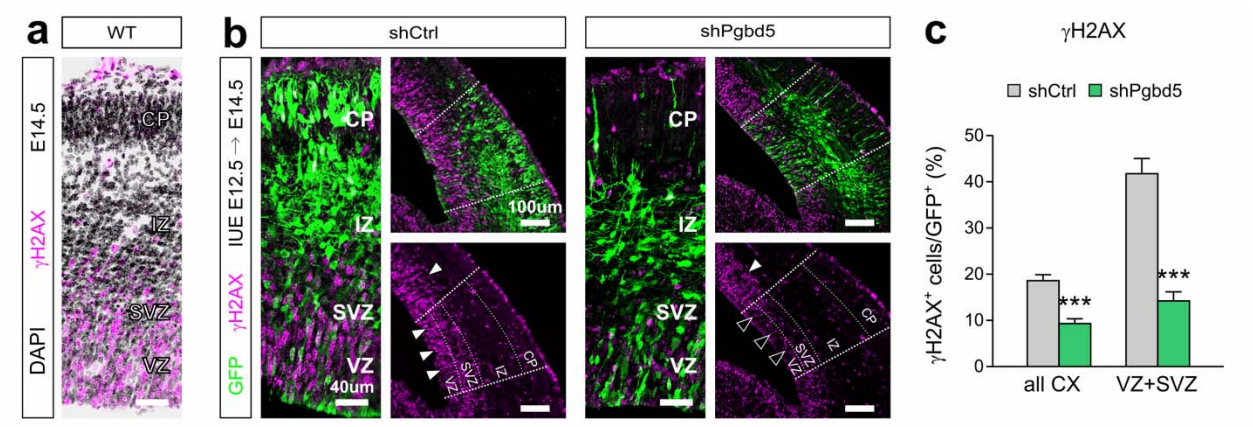

Fig.5 DNA DSBs in the cerebral cortex depend on *Pgbd5*

a, γH2AX IF staining in WT cerebral cortex at E14.5. **b**, γH2AX IF signal in shPgbd5- (right) vs. shCtrl-injected brains (left). Strong expression (white arrowheads) vs. weak expression (black arrowheads). **c**, Quantification of γH2AX$^+$/GFP$^+$ cells at E14.5. Data are mean ± s.e.m.; n = 4 shCtrl, n = 4 shPgbd5. Unpaired Student's t-test. VZ, ventricular zone; SVZ, subventricular zone; IZ, Intermediate zone; CP, cortical plate. ***P < 0.001.

Supplementary Materials for

The Pgbd5 DNA transposase is required for mouse cerebral cortex development through DNA double-strand breaks formation

Alessandro Simi[1], Federico Ansaloni[1], Devid Damiani[1], Azzurra Codino[1], Damiano Mangoni[1], Pierre Lau[1], Diego Vozzi[1], Luca Pandolfini[1], Remo Sanges[1,2*], Stefano Gustincich[1*]

Corresponding author: stefano.gustincich@iit.it, remo.sanges@sissa.it

The PDF file includes:

Materials and Methods
Table S1-S5

Materials and Methods

Animals

CD1 IGS mice for *iue* (Charles River) were housed at Istituto Italiano di Tecnologia (IIT) animal facility. All animal procedures were approved by IIT animal use committee and the Italian Ministry of Health and conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the European Community Council Directives. All mice were group-housed under a 12-hours light-dark cycle in a temperature and humidity-controlled environment with *ad libitum* access to food and water.

Plasmids

For in utero electroporation experiments, control shRNA (shCtrl) and Pgbd5 shRNA (shPgbd5) (Table S5) were cloned along the hU6 promoter in a pSilencer 3.0-H1 (ThermoFisher Scientific) plasmid using EcoRI and HindIII restriction sites and substituting the original H1 promoter. To visualize electroporated cells, pSilencer 3.0-U6-shCtrl and pSilencer 3.0-U6-shPgbd5 vectors were co-electroporated together with a pCAGGS-IRES-GFP vector with a 1:1 molar ratio.

In Utero Electroporation

Animal care and experimental procedures were performed in accordance with the IIT licensing and the Italian Ministry of Health license n° 693/2019-PR and 327/2021-PR. E12.5 timed-pregnant CD1 IGS mice were anesthetized with isoflurane and administered 2-5 mg/kg ketoprofene for analgesia. The uterine horns were exposed by laparotomy and the DNA plasmids (1 µg/µl) with 0.01% Fast Green dye (Sigma Aldrich) were injected in the lateral ventricle using a glass capillary (B100-58-10, Sutter Instrument). 3 mm diameter platinum tweezer electrodes (CUY650P3, NepaGene) were used to electroporate the cerebral cortex. Four electrical pulses (35 V, 50 msec duration, 950 msec interval) were delivered using a NEPA21 electroporator (NEPA21, NepaGene).

Immunofluorescence and Imaging

Mice were sacrificed at indicated ages and the brains dissected and fixed in 4% paraformaldehyde (PFA) (Sigma Aldrich) at 4°C overnight. Brains were de-hydrated in 30% sucrose and sliced in 20 µm coronal cryosections using a CM 3050S cryostat (Leica). Slices were permeabilized and blocked in 1x Phosphate-Buffered Saline (PBS) containing 5% normal goat serum (NGS, Abcam) and 0.3% Triton X-100 (Sigma Aldrich). Brain sections were incubated overnight with primary antibodies (see Table S3) diluted in blocking solution at 4°C. After extensive washes in 1x PBS + 0.1% Triton X-100 (PBST) brain slices were incubated with fluorescent dye conjugated secondary antibodies (see Table S3) diluted in blocking solution for 1h at RT. Slices were counterstained with Hoechst 33342 (Thermo Fisher Scientific) 1:10.000 in PBST for 15 min and extensively washed in 1x PBS, mounted with Mowiol (Sigma Aldrich). Fluorescent images were acquired with Nikon A1 confocal microscope equipped with a 20x objective and processed with Nikon software version 4.11.0 (NIS Elements). Positive cells for the indicated marker were counted through the depth of the cortex in the electroporated area and the percentages normalized on the total number of GFP^+ cells as indicated in the figure legends. For cell number quantifications, all relevant sections containing GFP^+ electroporated cells from rostral to caudal were quantified upon shCtrl and shPgbd5 conditions by ImageJ version 1.53q (Wayne Rasband, National Institutes of Health, USA). In order to assess GFP^+ cells distribution

at E14.5, the electroporated cortex was subdivided radially in three bins using Tbr1 and Tbr2 markers to identify the CP, IZ or VZ+SVZ boundaries. For the E18.5 migration analysis, we divided the cerebral cortex in ten equal-sized bins from the pia to the upper edge of the white matter. In both analyses the number of GFP+ cells in each bin were quantified and reported as the percentage of total counted cells.

Cell cycle exit analysis

The nucleoside analog EdU (5-ethynyl 20-deoxyuridine) was injected intraperitoneally at 25 mg/kg in pregnant females at E13.5, 24h after the *iue* procedure. EdU was detected using EdU click-iT technology (Invitrogen, C10340). The amount of electroporated cells exiting the cell cycle was calculated by the ratio GFP+ Ki67- EdU+ / GFP+ EdU+.

FAC-sorting of GFP$^+$ cells from iue mouse cerebral cortex

Pregnant female CD1 IGS mice electroporated at E12.5 with shCtrl or shPgbd5, were sacrificed at E14.5 and brain cortices dissected under a stereomicroscope in ice-cold Hanks' Balanced Salt Solution (HBSS) and enzymatically dissociated using a Neural Tissue Dissociation Kit (Miltenyi Biotec) following manufacturer's protocol. After tissue digestion at 37°C, cells were manually dissociated by pipetting, filtered with a 40 µm cell strainer, centrifuged for 10 min at 300 rcf and resuspended in ice-cold HBSS. GFP$^+$ cells were FAC-sorted with a SH800S instrument (Sony). For each shCtrl and shPgbd5 sample, n = 3 pools of 250 GFP$^+$ cells were collected for direct RNA-Seq library preparation using the SMART-Seq HT PLUS kit (Takara), according to manufacturer's instructions.

RNA purification and RT-qPCR

Total RNA from mouse cerebral cortex was purified by TRIzol Reagent (Thermo Fisher Scientific) and DNA was removed by treatment with DNAse I (Sigma-Aldrich). cDNA was synthesized from 0.5 mg of RNA using an iScript cDNA Synthesis kit (Bio-Rad). Real time quantitative PCR was performed on a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad). For *Pgbd5* detection a SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) was used with Tbp as reference gene (Table S4) with the following cycling conditions: 95°C for 20 sec, followed 40 cycles of 95°C for 10 sec and 60°C for 1 min. No template and no RT controls were included. Expression levels were determined relative to the reference gene using the ΔΔCt method.

Gene expression analysis

RNA-Seq reads were mapped to the murine reference genome (version: GRCm39.vM28) using STAR with default parameters (version: 2.6.0a) (Dobin et al. 2013). Gene expression levels were estimated by the read counting module embedded within the STAR tool (–quantMode GeneCounts). DE-genes were identified by using edgeR (Robinson et al. 2009) using the TMM method to normalize raw read counts, whereas common, trended and tagwise dispersions were estimated by maximizing the negative binomial likelihood (default). Lowly expressed genes were removed, retaining only genes expressed >= 1 TMM in at least two of the six analyzed samples, and DE-genes were tested performing a quasi-likelihood F-tests (glmQLFit and glmQLTest). Genes were considered as DE when showing an FDR < 0.05.

Gene ontology enrichment analysis

Gene ontology (GO) enrichment analysis was performed by using gprofiler2 (Peterson et al. 2020). GO terms enriched in the set of up- and down-regulated genes were identified separately

providing a custom list of background genes (*i.e.*, genes expressed >= 1 TMM in at least two samples). gSCS p-value correction method was used. GO terms associate to at least 2 up- or down-regulated genes and showing FDR < 0.05 were considered as significant.

Statistical analysis

Statistical analyses were performed using GraphPad Prism (version 7, GraphPad Software) according to the number of replicates and group design. In all figures, significance of comparisons is denoted as follows: $*P<0.05$, $**P<0.01$, $***P<0.001$. In the figure legends, statistical tests used and the nature and numbers of samples analyzed (defined as *n*) are reported. Sample sizes were based on published experiments and previous experience in which differences were observed. No power calculations or statistical tests to pre-determine the sample size were used.

Supplementary tables:

Table S1 (separate file)

See file Table S1_DEgenes.xlsx

Table S2 (Separate file)

See file TableS2_GO.xlsx

Table S3

Antibodies table

Antibody	Host species	Company	Catalog number	Dilution
anti-GFP	chicken polyclonal	Abcam	ab13970	1:1000
anti-Ctip2	rat monoclonal	Abcam	ab18465	1:1000
anti-Tbr1	rabbit polyclonal	Abcam	ab31940	1:1000
anti-Tbr2	rabbit monoclonal	Abcam	ab183991	1:1000
anti-Pax6	rabbit polyclonal	Millipore	AB2237	1:1000
anti-Neurod1	goat polyclonal	Santa Cruz Biotech.	sc-1084	1:200
anti-Ki67	rabbit polyclonal	Abcam	ab15580	1:1000
anti-Neurog2	rabbit polyclonal	Cell Signaling	13144	1:500
anti-γH2AX	rabbit polyclonal	Cell Signaling	9718S	1:500
AlexaFluor-488 anti-Chicken	goat	Thermo Scientific	A11040	1:1000
AlexaFluor-546 anti-Rabbit	goat	Thermo Scientific	A-11035	1:1000
AlexaFluor-647 anti-Rabbit	goat	Thermo Scientific	A-21244	1:1000
AlexaFluor-647 anti-Rat	goat	Thermo Scientific	A-21247	1:1000
AlexaFluor-546 anti-Goat	donkey	Thermo Scientific	A-11056	1:1000
AlexaFluor-647 anti-Rabbit	donkey	Thermo Scientific	A-21447	1:1000

Table S4

RT-qPCR primer table

Oligo name	Forward primer sequence (5' - 3')	Reverse primer sequence (5' - 3')
Tbp	GTACCGCAGCTTCAAAATATTGTAT	AAATCAACGCAGTTGTCCGTG
Gapdh	TGTGTCCGTCGTGGATCTGA	CCTGCTTCACCACCTTCTTGA
Pgbd5	CTATATCATCTTTACAGGGCCCAGC	GATCAGTGACATGTTGCCCTTTGTC

Table S5

shRNA sequences

Oligo name	Oligo sequence (5' - 3')
shCtrl	CAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTG
shPgbd5	GGTACAGACCAACATGTATGCCTCGAGGCATACATGTTGGTCTGTACC