Title: Haploinsufficiency of the intellectual disability-gene *SETD5* disturbs developmental gene expression and cognition

Authors: Elena Deliu[1†], Niccolò Arecco[2,3†], Jasmin Morandell[1†], Christoph P. Dotter[1†], Ximena Contreras[1], Charles Girardot[2], Eva-Lotta Käsper[2,3], Alena Kozlova[1], Kasumi Kishi[1], Ilaria Chiaradia[1], Kyung-Min Noh[2@], Gaia Novarino[1@]

Affiliations:

[1]Institute of Science and Technology (IST) Austria, Klosterneuburg, Austria

[2]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany

[3]Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences

[†]These authors contributed equally to this work

[@]Correspondence to: noh@embl.de and gnovarino@ist.ac.at

Abstract

SETD5 gene mutations have been identified as a frequent cause of idiopathic intellectual disability. Here we show that *Setd5* haploinsufficient mice present developmental defects such as abnormal brain to body weight ratio and neural crest defect associated phenotypes. Furthermore, *Setd5* mutant mice show impairments in cognitive tasks, enhanced long-term potentiation, delayed ontogenetic profile of ultrasonic vocalisation and behavioural inflexibility. Behavioural issues are accompanied by abnormal expression of postsynaptic density proteins previously associated with cognition. Our data additionally indicate that Setd5 regulates RNA polymerase II dynamics and gene transcription via its interaction with the Hdac3 and

Paf1 complexes, findings potentially explaining the gene expression defects observed in *Setd5* haploinsufficient mice. Our results emphasize the decisive role of Setd5 in a biological pathway found to be disrupted in intellectual disability and autism spectrum disorder patients.

Main

Intellectual disability (ID), characterized by substantial limitation of cognitive functions and adaptive behaviours, affects 1.5-2 % of the Western population[1]. ID often co-exists with core symptoms of autism spectrum disorders (ASD), such as impairment in social interaction, communication and presence of repetitive behaviours[1]. ID and ASD have a strong genetic component and several ID- and ASD-genes have been identified[2,3].

The *SET-domain containing 5* (*SETD5*) gene belongs to the SET-domain containing gene family, encoding histone-modifying proteins[4]. *De novo* mutations in the *SETD5* gene have been recently identified as a cause of ID and ASD with up to 0.7% of patients with idiopathic ID carrying heterozygous mutations in the *SETD5* gene[5,6]. While the number of reported cases increases, allowing the delineation of a *"SETD5* syndrome"[7,8]*,* SETD5's molecular function and role in cognition remain elusive.

Here, we show that *Setd5* haploinsufficiency in mice leads to embryonic development defects and behavioural abnormalities. Developmental and behavioural issues correlate with abnormal control of gene expression in embryonic and adult mutant animals. Particularly, we found that disruption of *Setd5* impairs the proper expression of a number of post-synaptic density proteins implicated in synaptic plasticity and learning. At the molecular level Setd5 interacts with the histone

deacetylase 3 (Hdac3) and polymerase-associated factor 1 (Paf1) complexes to regulate gene transcription. Intriguingly, several components of the HDAC3 and PAF1 complexes have been previously involved in ID and ASD.

Results

Setd5 haploinsufficiency leads to early developmental defects.

Setd5 is expressed in several tissues, including the brain, during development and in adulthood[6,9]. To examine the effect of *Setd5* haploinsufficiency *in vivo*, we studied *Setd5*$^{+/-}$ mice (Supplementary Fig. 1). While full deletion of *Setd5* leads to lethality at early embryonic stages[9], *Setd5*$^{+/-}$ mice are viable but born at a non-Mendelian rate (Supplementary Fig. 2a) and have lower survival probability (Supplementary Fig. 2b). Interestingly, newborn and adult mutant mice present with increased brain to body weight ratio (Fig. 1a and Supplementary Fig. 2c) suggesting developmental defects. Many *Setd5*$^{+/-}$ mice also display tooth displacement, eye problems, white spotting of the belly and craniofacial abnormalities (Fig. 1b-c and Supplementary Fig. 2d-e), issues linked to aberrant head and trunk neural crest proliferation and cell fate determination[10,11]. These features, in part also observed in patients with *SETD5* mutations[5-7], suggest that *Setd5* haploinsufficiency may impact developmental processes at very early stages.

Following these observations, we performed RNA sequencing (RNA-seq) of 9.5 day-old (E9.5) control and *Setd5* mutant embryos. We detected 487 differentially expressed genes (DEGs) between *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$ mice (Fig. 1d and Supplementary Table 2), revealing an approximately equal number of up and down regulated genes. Gene ontology (GO) enrichment analysis, however, highlighted that genes upregulated in *Setd5*$^{+/-}$ embryos are associated with head and brain

development, while downregulated genes are involved in the formation of other embryonic parts including the neural crest, heart, limbs and skeleton (Fig. 1e, Supplementary Table 3).

Among pathways disrupted by *Setd5* haploinsufficiency there is the Wnt signalling cascade (adjusted (adj.) p-value 2e-2, Supplementary Table 3). Accordingly, downregulated genes are enriched for β-catenin and transcription factor 3 (TCF3) targets (adj. p-value: 1e-5 and 2e-10, respectively). Furthermore, DEGs are enriched for genes previously associated with dysmorphic features observed in *Setd5* mutant mice and patients[6] (Fig. 1b-c; Supplementary Fig. 2d-e), for example genes linked to eye developmental defects such as microphthalmia (e.g. *Otx2*, *Aldh1a3*, *Stra6* and *Sox2*, adj. p-value 4e-5).

To substantiate our analysis and study how *Setd5* haploinsufficiency affects gene expression in homogenous cell populations, we generated *Setd5*$^{+/-}$ mouse embryonic stem cells (ESCs) (Supplementary Fig. 3a) and differentiated them to embryoid bodies (EBs) and neural progenitor cells (NPCs) (Supplementary Fig. 3b). RNA-seq followed by differential expression analysis from *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$ (Supplementary Fig. 3c) ESCs, EBs and NPCs (Supplementary Table 2) indicated a significant overlap with the DEGs of the E9.5 embryos (67 genes; p-value: 6e-15, Supplementary Fig. 3d). GO-term analysis revealed that *Setd5*$^{+/-}$ ESCs and EBs display upregulation of genes related to epithelial and neuronal differentiation (Supplementary Fig. 3e, Supplementary Table 3) and downregulation of endoderm and mesoderm-related genes, indicating premature and excessive expression of neuroectodermal genes and suppression of genes associated with other germ layers. Similar to our findings in E9.5 embryos, downregulated DEGs in mutant EBs and NPCs are enriched for Wnt signalling (adj. p-value: 3e-5 and 6e-4, respectively)

and TCF3 targets (adj. p-value: 1.8e-7 and 8e-20, respectively). These results indicate that at early developmental stages *Setd5* haploinsufficiency favours the expression of neuronal-related genes at the expense of other cell lineages, thus reflecting the abnormalities apparent in *Setd5*$^{+/-}$ mice.

Setd5$^{+/-}$ mice exhibit overall normal brain morphology.

In the central nervous system Setd5 is found throughout development and adulthood (Supplementary Fig. 4a-c), in multiple brain regions (Supplementary Fig. 4c) and cell types, including inhibitory and excitatory neurons (Supplementary Fig. 4d-e). Thus, we investigated whether haploinsufficiency of *Setd5* affects brain morphology. Initially, we assessed NPC proliferation by 2-hour bromodeoxyuridine (BrdU) incorporation at embryonic day 12.5 (E12.5). In *Setd5*$^{+/-}$ embryos the number of BrdU labelled cells appears normal, indicating that NPC proliferation is not affected by *Setd5* mutations (Supplementary Fig. 5a,d). Furthermore, *Setd5*$^{+/-}$ embryos present no alterations in cell survival, as indicated by the very low number of cleaved Caspase 3 positive cells (Supplementary Fig. 5b,e) and normal numbers of Pax6 and Tuj1 positive cells at E12.5 (Supplementary Fig. 5c,f). This data is in agreement with the observation that while the brain/body weight ratio is increased in *Setd5*$^{+/-}$ mice compared to *Setd5*$^{+/+}$, the absolute brain size does not vary between the two genotypes. We also found no anomalies in brain morphology of *Setd5*$^{+/-}$ adult animals when assessed by Nissl staining (Supplementary Fig. 5g). Normal cortical lamination of *Setd5*$^{+/-}$ adult brain was confirmed in stainings for Cux1 and Ctip2 neurons. The whole cortex, individual layers, and the white matter in coronal brain sections displayed expected thicknesses (Supplementary Fig. 5h-i). In conclusion, *Setd5*$^{+/-}$ gross brain morphology is normal.

Setd5$^{+/-}$ mice show behavioural abnormalities, cognitive defects and enhanced long-term potentiation.

Next, we investigated whether *Setd5* haploinsufficiency affects mouse behaviour employing tests relevant for ASD- and ID-related mouse phenotypes[12]. Locomotion is not affected in *Setd5*$^{+/-}$ mice (Supplementary Fig 6a). Mutant mice also perform similarly to controls in tasks evaluating social interaction and social novelty, repetitive behaviours (i.e. marble burying and rotations in the open field) and anxiety (Supplementary Fig 6b-d), although *Setd5*$^{+/-}$ females show a slight reduction of anxiety-like behaviour in the elevated plus maze (Supplementary Fig 6d).

Setd5$^{+/-}$ females, but not males, fail to build proper nests (Fig. 2a and Supplementary Fig. 6e), indicating deficient reproductive[13] or maternal care abilities[14]. In addition, a delayed ontogenetic profile of ultrasonic vocalisation is apparent in *Setd5*$^{+/-}$ mouse pups, as they reach peak ultrasonic vocalisations approximately four days after their age- and sex-matched control littermates (Fig. 2b and Supplementary Fig. 6f), a delay comparable to deficits observed in ASD mouse models[15]. Because *SETD5* patients display ID, we next assessed *Setd5*$^{+/-}$ mice with learning tasks. First, we examined behaviour of group-caged mice in the Intellicage, where animals get access to water by nose poking to doors located in front of bottles placed at the four corners of the cage. During learning trials, each mouse is randomly assigned to an “incorrect” corner, where nose poking would trigger an aversive air puff instead of access to water (Fig. 2c). On average, *Setd5*$^{+/-}$ mice perform more nose pokes per corner visit (Fig. 2c), even when they are not seeking for water (no licks) (Supplementary Fig.7a). As opposed to marble burying, in which digging[16], but not stereotypic behaviour[17] may play a role, nose poke repetitions represent a cleaner, higher order repetitive behaviour, previously reported in other

ASD mouse models[18,19]. Moreover, mutants perform fewer visits with nose pokes but without licks (Supplementary Fig. 7a), a behaviour previously linked to the hippocampus[20]. During the place avoidance trial, both *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$ animals visit the incorrect corner less frequently than any correct corner (Fig. 2c). After roughly 24h, *Setd5*$^{+/+}$ animals learn the task, as they significantly reduce the number of nose pokes per visit in the incorrect corner. In contrast, *Setd5*$^{+/-}$ mice underperform, in that they continue to nose poke similarly frequent in both correct and incorrect corners (Fig. 2c-d), suggesting deficits in adaptive behaviour.

Next, we assessed context fear memory acquisition and consolidation. Mutant and control mice are indistinguishable during acquisition of contextual fear, indicating similar sensory acuity (Supplementary Fig. 7b). However, one day after acquisition, when exposed to the context, female *Setd5*$^{+/-}$ mice freeze significantly more than their wild type littermates, denoting enhanced fear memory retention (Fig. 2e). This difference was not observed in males trained with the same strong protocol (Supplementary Fig. 7c) but it became obvious upon a slightly weaker training (Supplementary Fig. 7d). In all circumstances, both female and male *Setd5*$^{+/-}$ animals further fail to extinguish the repellent context-shock association and form a new, neutral memory, since they do not significantly diminish freezing when the context becomes safe (Fig. 2e and Supplementary Fig. 7c-d). Thus, the contextual fear conditioning (CFC) tests suggest that *Setd5* haploinsufficiency is associated with a better memory retention but a worse capacity to extinguish memories, a feature already observed in mouse mutants for intellectual disability genes[21].

Although, *Setd5*$^{+/-}$ mice do not show increased anxiety-like behaviour, we substantiated our CFC analysis by performing the novel object location memory test, a hippocampus-dependent, aversive stimulus-independent, long-term memory test.

Mutant mice, when tested 24 h after receiving a subthreshold (3 minutes) training, show a significantly increased ability to discriminate between the old and new location (Fig. 2f, Supplementary Fig. 7e). However, this type of training typically produces no memory trace (defined as discrimination index > 30) in wild-type animals[22], as is the case with our control littermates.

To better understand the basis of the observed cognitive abnormalities, we tested long-term potentiation (LTP) of the synaptic transmission at the CA3-CA1 synapses in hippocampal slices from *Setd5*$^{+/-}$ and *Setd5*$^{+/+}$ littermates. In agreement with the observed increased memory retention, both early (Fig. 3a) and late (Fig. 3b and Supplementary Fig. 7f-g) LTP of local field potentials are enhanced in *Setd5*$^{+/-}$. This phenotype, already observed in an ASD and ID mouse model[23], suggests detrimental cellular and/or molecular synaptic plasticity mechanisms associated with abnormal learning and memory.

Cognitive defects in *Setd5*$^{+/-}$ mice are accompanied by abnormal dynamics of postsynaptic gene expression.

Because *Setd5* is part of a histone modifier gene family we hypothesized that *Setd5* loss of function mutations may prompt faulty regulation of gene expression during learning. Thus, we performed RNA-seq followed by differential expression analysis from hippocampal samples of naïve (homecage) and context fear conditioned mice (Supplementary Table 4).

Compared to homecage, control samples obtained one hour after conditioning display a marked regulation of gene expression (280 DEGs FDR < 0.05, Fig. 4a, left). Among these genes we observed a significant upregulation of activity-dependent immediate early genes, such as *Fos* and *Egr2*[24] (Fig. 4a, left), and enrichment for cAMP responsive element binding protein (CREB)-dependent genes

(adj. p-value 5.2e-5, Supplementary Fig. 8a, top). Three hours post-conditioning, however, in agreement with previous studies[25], gene expression levels were mostly back to baseline and showed a pattern similar to homecage samples (63 DEGs, FDR < 0.05, Fig. 4a, left).

Setd5$^{+/-}$ samples, while showing the expected upregulation of CREB-dependent and activity–dependent immediate early genes one hour post-conditioning (Fig. 4a, right and Supplementary Fig. 8a, bottom), differ markedly in their gene expression compared to samples obtained from control mice one hour and three hours post conditioning (180 and 212 DEGs respectively, FDR < 0.05). In contrast, *Setd5*$^{+/-}$ and control samples obtained from naïve homecage mice are virtually indistinguishable (4 DEGs, FDR < 0.05) indicating a dramatically different transcriptional response to training.

Next, we clustered genes based on their transcriptional response at the different time points and compared their trajectories in mutant and control samples. While a number of genes, including early response genes, follow identical trajectories (Supplementary Fig. 8b, Supplementary Table 5) in the two genotypes, we identified 11 gene clusters displaying significantly different expression profiles between control and mutant animals (Fig. 4b, Supplementary Table 5). Thereafter, we performed *in-silico* analysis to identify biological pathways, cellular components and cell types potentially enriched in these clusters (Supplementary Table 6). Of note, the most significantly enriched set of genes belongs to the postsynaptic density (PSD) GO-term in cluster 6, consisting of genes displaying abnormally sustained expression three hours post-conditioning in *Setd5*$^{+/-}$ (Fig. 4b and Supplementary Table 6). Among the PSD protein encoding-genes belonging to this cluster, several, i.e. Shank1, CamKIIN1, CPEB1, FXR2P and Synaptopodin (Fig. 4c), have been

associated with memory retention, synaptic plasticity and ID[26-30]. Similarly, cluster 4 is significantly enriched for genes encoding for proteins regulating synapse structure and activity, including SynGAP1, LRRC4 and p140Cap, which are also predominantly expressed in the PSD[31-36] (Fig 4c). Finally, clusters 2, 5 and 7 (Fig. 4b), comprising genes significantly affected in *Setd5*$^{+/-}$, are enriched for histone H3 acetylation, lysine-acetylated histone binding proteins and histone H4-R3 methylation encoding genes, respectively (Supplementary Table 6), suggesting that the epigenetic response to CFC is partially affected in the mutants. Interestingly, the fine-tuning of histone H3 acetylation status has been previously associated with CFC dynamics and LTP of the CA3-CA1 hippocampus synapses[22].

SETD5 lacks methyltransferase activity and interacts with Hdac3 and Paf1 complexes.

Next, we investigated the molecular function of the SETD5 protein. Most SET domain-containing proteins are histone methyltransferases. Thus, we tested whether SETD5 exerts this enzymatic activity (Fig. 5a). Neither the SET domain (Fig. 5b) nor full-length SETD5 (Fig. 5c) showed methyltransferase activity *in vitro*. Likewise, histone H3 methylation status is comparable in control and *Setd5*$^{+/-}$ hippocampal samples (Supplementary Fig. 9a). Furthermore, neither in *Setd5*$^{+/-}$ ESCs nor in *Setd5* null (*Setd5*$^{-/-}$) ESCs (Supplementary Fig. 9b) are the histone methylation levels changed. These results are consistent with the fact that SETD5, like MLL5 (KMT2E), does not have a canonical substrate-binding domain, which is conserved in the other active SET family proteins (Supplementary Fig. 9c). Thus, in agreement with previous reports[9,37], we concluded that the SET domain of SETD5, similar to MLL5 (KMT2E)[38] and the Drosophila SET protein UpSET[39], lacks histone modifying enzymatic activity.

SETD5 might exert a function via interactions with other proteins. To address this point in a tractable system, and due to a lack of reliable Setd5 antibodies, we generated mouse ESCs expressing an endogenously tagged *Setd5* (Setd5-HA-FLAG) (Supplementary Fig. 9d-f). Immunoprecipitation coupled with quantitative mass spectrometry in ESCs and ESC-derived NPCs revealed that endogenous Setd5 is bound to two distinct protein complexes, Hdac3 and Paf1 (Fig. 5d,e, Supplementary Table 7 Supplementary Fig. 9g). The Hdac3 complex, containing Ncor1/2, Tbl1x/1xr1 and Hdac3 (Fig. 5e), regulates gene expression by modulating histone acetylation[40,41], while the Paf1 complex, containing Paf1, Leo1, Cdc73, Wdr61 and Ctr9 (Fig. 5e), is associated with controlling RNA Polymerase II dependent transcription[42].

These results suggest that Setd5 acts with at least two different protein complexes in regulating chromatin environment and gene expression. Intriguingly, Tbl1x and Tblxr1 have been previously linked to the regulation of Wnt signalling[43] a pathway appearing disrupted in *Setd5* mutants. Furthermore, all the components of the HDAC3 complex (https://gene.sfari.org/database/human-gene/) and a part of the PAF1 complex (e.g., LEO1)[44] (Supplementary Fig. 9h) have reported mutations associated with autism, suggesting a tight link between Setd5-associated protein complexes and neurodevelopment or brain function.

SETD5 patient mutations disrupt SETD5-HDAC3/-PAF1 interaction.

Several different *SETD5* mutations have been reported in intellectually disabled individuals. If the interaction between SETD5 and its complexes plays a critical role in the brain, *SETD5* mutations identified in patients (Supplementary Fig. 10a) would lead to disruption of the interactions. To test this hypothesis, we first investigated the effect of six human *SETD5* truncating mutations[5,45-47] (Fig. 6a) on

SETD5 protein expression and cellular localization. Ectopic expression of the six human mutant proteins in HEK cells revealed that one mutant (R308*) lost protein expression and another (K399*) failed at nuclear localization, but four others (T552*, E720*, R1001* and S1258*) expressed nuclear proteins (Fig. 6b). We then selected two mutants (E720* and S1258*) displaying nuclear protein expression and tested whether they interact with the Hdac3 and Paf1 complex. To avoid a skewed result due to overexpression, we expressed the same dosage of the full length WT and each of the two truncating mutants (E720* and S1258*) (Fig. 6a) from the Rosa26 locus in *Setd5*$^{-/-}$ ESCs (Supplementary Fig. 10b-d) and performed immunoprecipitation experiments. We found that only full-length wild type SETD5, but not the two truncated mutants, co-immunoprecipitated with Hdac3 and Tbl1x (Fig. 6c). These results imply that SETD5 is bound to the HDAC3 complex through its C-terminus (last 185 amino acids from 1258 to 1443) and that other mutations affecting these residues (e.g. T552* and R1001*) should result in loss of SETD5-HDAC3 interaction. Similarly, mutations truncating the C-terminal domain of SETD5 affect its interaction with Leo1, an ASD associated component of the Paf1 complex, although this interaction is only partially reduced (Fig. 6c). Altogether, these results highlight a potential role of SETD5 interacting chromatin proteins in the aetiology of the SETD5-associated disease.

Setd5 haploinsufficiency increases histone acetylation without affecting Hdac3 activity.

Considering the interaction between Setd5 and Hdac3, we studied whether Setd5 affects histone acetylation levels. Immunoblot analysis of control and *Setd5*$^{-/-}$ ESCs (Supplementary Fig. 11a,b) shows that deletion of *Setd5* results in increased histone H4 acetylation (Supplementary Fig. 11c). Likewise, the lysine 8 of histone 4

(H4K8) is more acetylated in hippocampal samples obtained from *Setd5*$^{+/-}$ mice than in controls (Supplementary Fig. 11d). Interestingly, H4K8 acetylation is also elevated three hours after CFC (Supplementary Fig. 11e) but is equivalent to control sample levels one hour post-conditioning (Supplementary Fig. 11f). Thus, H4K8 acetylation dynamics, observed after CFC training in wild type animals[48], are altered upon heterozygous loss of *Setd5* (Supplementary Fig. 11g). Despite the observed hyperacetylation, Hdac3 activity is not affected in *Setd5*$^{+/-}$ brain and ESC samples (Supplementary Fig. 11h,i), suggesting that the histone acetylation changes are independent of Hdac3 activity. They may thus be a consequence of other alterations in the chromatin environment for instance due to a defective interaction between Setd5 and the Paf1 complex.

Setd5 occupies transcription start sites (TSS) with Hdac3 and Pol II.

To elucidate the role of Setd5 in chromatin regulation, we examined the genomic distribution of Setd5 and its associated chromatin proteins (i.e., Hdac3 and Paf1 acting RNA polymerase II) in ESC using chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq). Read counts for tagged Setd5 at annotated genes revealed that Setd5 is significantly enriched at TSS (defined as a 2.5 kb proximal region centred on the gene start sites) and predicted ESC-specific enhancers, but no other regions, such as transcription end sites (TES) and inactive enhancers (Fig. 7a). Hdac3 signals were also enriched in TSS and ESC enhancers (Supplementary Fig. 12a), and correlated with Setd5 signals (Fig. 7b). RNA polymerase II (Pol II) ChIP-seq signals, detected with an antibody recognizing both Pol II recruited into the preinitiation complex (non-phosphorylated) and initiated Pol II (phosphorylated at serine 5), showed significant enrichment at TSS (Supplementary Fig. 12 b). Notably, the most pronounced signals of Setd5 were observed at TSS

occupied by both Hdac3 and Pol II (Fig. 7b), implying that our immunoprecipitation coupled with mass spectrometry result captured the Setd5-associated complexes at TSS.

Setd5 regulates Pol II occupancy at transcriptional start sites of neurodevelopmental genes.

To clarify SETD5's role at the TSS, we investigated whether Setd5 is required for Hdac3 or Pol II (via Paf1) recruitment. We analysed the genome-wide localization of Hdac3 and Pol II in control and *Setd5* mutant (i.e. *Setd5*$^{+/-}$ and *Setd5*$^{-/-}$) ESCs (Fig. 7c). We also examined the genomic distribution of acetylated histone H4 (pan-acetylated H4) and H3 (acetylated H3K27) as a potential outcome of Hdac3 activity. Global analysis revealed that Hdac3 occupies numerous promoters and enhancers across control and *Setd5* mutant ESCs (Supplementary Fig. 12a), with minimal changes between genotypes (21 differential peaks in *Setd5*$^{+/+}$ vs. *Setd5*$^{+/-}$, 34 differential peaks in *Setd5*$^{+/+}$ vs. *Setd5*$^{-/-}$, FDR < 0.05, Supplementary Fig. 12d). Our results suggest that Setd5 is not involved in Hdac3 recruitment to the chromatin, as opposed to previous hypotheses[9].

Interestingly, Pol II ChIP-seq profiles showed a broad redistribution in *Setd5*$^{+/-}$ and *Setd5*$^{-/-}$ cells. The ratio of promoter proximal to gene body Pol II density (i.e. pausing index) is decreased in 44% of genes (n=4144) and increased in 26% of genes (n=2487) containing Pol II at the TSS, in both *Setd5*$^{+/-}$ and *Setd5*$^{-/-}$ versus *Setd5*$^{+/+}$ ESCs (Fig. 7d). Moreover, when comparing *Setd5*$^{+/-}$ cells to controls, we observed significantly increased Pol II occupancy at 559 TSS (Fig. 7c,e and Supplementary Fig. 12e) and decreased occupancy at 182 TSS (Supplementary Fig. 12e,g, Supplementary Table 8). *Setd5*$^{-/-}$ cells also showed similar Pol II changes at those TSS (593 differential peaks in wild type vs. *Setd5*$^{-/-}$, FDR < 0.05,

Supplementary Fig. 12e). The TSS with increased Pol II are enriched for Hdac3 and depleted for both H3K27ac and H4ac signals over the inputs (Fig. 7e) implying a poised transcriptional state of these genes. Global distribution of H3K27 acetylation was unchanged in *Setd5*$^{-/-}$ ESCs at TSS as well as enhancers (Supplementary Fig. 12c,f), consistent with minimal changes in Hdac3. In agreement with the immunoblots (Supplementary Fig. 11c), average ChIP-seq profiles of acetylated H4 (pan-acetylated H4) revealed significantly increased acetylation at gene bodies in Setd5 mutant cells (Supplementary Fig. 12h), which might be linked to changes of Pol II rather than Hdac3 activity. All together, these results suggest that Setd5, through its interaction with chromatin-associated proteins (e.g. Paf1) functions to maintain proper levels of Pol II at TSS and gene bodies.

Next, we investigated whether chromatin alterations driven by *Setd5* haploinsufficiency are associated with changes in gene expression. We found that DEGs in *Setd5*$^{+/-}$ cells are enriched for genes showing both Hdac3 and Pol II signals, but no H3K27ac, at the TSS (adj. p-value: E9.5: 8.3e-22; ESCs: 4e-2; EBs: 1.6e-13; NPCs: 1.9e-14). This result suggests that TSS of DEGs show a poised chromatin state in ESC, implying that poised genes may be more susceptible to *Setd5* haploinsufficiency. Thus, we extended our analysis to the 559 genes featuring increased Pol II at the TSS in *Setd5*$^{+/-}$ ESCs and poised chromatin state (enriched for Hdac3 and depleted for both H3K27ac and H4ac signals), thereby also capturing genes possibly not expressed in control ESCs. GO term analysis of these genes revealed neuronal development and neuronal signalling pathways (Fig. 7f, Supplementary Table 8). Furthermore, genes showing a *Setd5* dosage dependent Pol II increase at the TSS (i.e. higher fold change in Pol II occupancy in *Setd5*$^{-/-}$ compared to *Setd5*$^{+/-}$ mutant cells) are significantly enriched for Wnt signalling-

related genes (adj. p-value: 8e-3), consistent with our differential expression analysis. In contrast, genes with decreased Pol II at the TSS were not enriched for any specific GO-term.

Finally, we examined the expression level of the 559 genes showing increased Pol II at the TSS. Average read counts were low in these genes, consistent with the low levels of H3K27ac and H4ac (Fig. 7e), indicating that in control conditions these genes are mostly not expressed in ESCs. Nevertheless, a subset (~20%, 124 genes) of the genes show increased expression in *Setd5*$^{+/-}$ ESCs (Fig. 7g,h, Supplementary Table 8), suggesting that these genes are more prone to transcription in *Setd5* haploinsufficiency in response to increased Pol II at TSS.

Collectively, these results indicate that Setd5 maintains proper Pol II levels at numerous TSS, most likely via its interaction with the Paf1 complex. When *Setd5* expression is reduced in ESCs, Pol II is globally redistributed and simultaneously increasingly positioned at the TSS of multiple neuronal-specific genes, possibly priming them for transcription. While Setd5 is not involved in global Hdac3 recruitment, our data suggests that Setd5 exerts its regulatory function preferentially at Hdac3 occupied TSS through Paf1-mediated Pol II recruitment and pausing.

Discussion

Mutations in *SETD5* are emerging as a relative frequent cause of ID and ASD, however the pathophysiological underpinnings remain uncovered.

We employed *in vivo* and *in vitro* mouse models to study how *Setd5* haploinsufficiency affects development and cognition. Our data suggest that during early development *Setd5* is critical in preventing expression of neuronal genes and maintaining correct levels of non-neuronal gene transcripts and cell types. Thus, the

variability of the clinical features of patients carrying different *SETD5* mutations and the phenotypic variability of *Setd5*$^{+/-}$ mice may reflect dosage-dependent effects on regulation of gene expression and/or stochastic variations in cell-fate determination early in development. Dosage dependent effects and changes in cell-fate determination were already suggested by previous studies, which analysed some effects of homozygous and transient reduction of *Setd5* in non-neuronal tissues[9,37]. Thus, our study, while focusing on the disease-relevant haploinsufficient model, supports some of the previous conclusions.

Most importantly, while *Setd5*$^{+/-}$ mice do not have obvious brain architecture abnormalities, our data show for the first time that *Setd5* haploinsufficiency leads to behavioural abnormalities and impairs the optimized transcription program associated with learning and memory. Interestingly, *Setd5* haploinsufficiency is associated with abnormal response of PSD gene expression. Thus, future studies analysing synaptic plasticity in *Setd5* mutants are warranted.

Our findings implicate Setd5 in the regulation of gene transcription through its interaction with the Hdac3 and Paf1 complexes. While we extend the observation of this interaction to neural precursors, this complex was ascertained in mouse ESC and HEK 293 cells[9], implying that SETD5-HDAC3-PAF1 interaction is conserved across different species and cell types. Interestingly, Hdac3 has been shown to act as a negative regulator of hippocampal long-term memory[22] and HDAC3 inhibitors described as memory enhancers. Similar to the *Setd5*$^{+/-}$ phenotype revealed here, focal deletion of *Hdac3* in the mouse dorsal hippocampus augmented novel object location memory[22], fear memory retention, and long-term potentiation[49]. However, the extinction of fear memory, which is substantially affected in the *Setd5*$^{+/-}$ animals, has not been examined in relation to *Hdac3* deficiency. This behavioural inflexibility

of *Setd5* mutants points to the possibility that chronic abnormal regulation of the HDAC3 complex may lead to impairment of cognition rather than its enhancement. In agreement, conditional homozygous *Hdac3* knockout in forebrain excitatory neurons was associated with impaired spatial learning[50], and our *Setd5* mutants present with learning inefficiency and/or impaired adaptive learning. Furthermore, Setd5 interacts with the Paf1 complex, which is also associated with ASD.

In summary, our results are consistent with a model in which Setd5 regulates Pol II occupancy at TSS of neuronal-related genes via its interaction with the Hdac3 and Paf1 complexes. Whether Setd5 interacts with Hdac3 and Pol II simultaneously remains to be determined and warrants a structural study of Setd5 within its protein complexes. In addition, while providing evidences that Setd5 regulates gene transcription through its interaction with Hdac3 and Paf1 complexes, it remains unclear whether anomalies in these interactions underlie all of the mutants' phenotypes. In fact, although our data dismiss methyltransferase activity of SETD5, there remains the possibility that Setd5, potentially through interaction with other proteins, contributes to brain development and function also in other ways.

Our study, comprehensively combining molecular, circuit and behavioral analyses, is the first to analyse the consequences of *Setd5* haploinsufficiency in neural cells, brain and behaviour, advancing the understanding of the function of epigenetic factors for development, learning and cognition. Furthermore, our data indicate that *SETD5* mutations affect a biological pathway that is dysfunctional in other ID and ASD cases, implying that this work is relevant for a number of related disorders.

Accession codes

Mouse RNA-seq data are deposited at GEO with accession number GSE119498. RNA-seq and ChIP-seq data from in vitro samples is deposited at ENA under the following accession numbers: PRJEB28477, PRJEB28476, PRJEB28475, PRJEB28474, PRJEB28473 and PRJEB28472.

Acknowledgements

We thank V. Stein and M. Döngi for guidance in setting up LTP experiments, P. Jonas for consultancy on LTP data analysis, M Kutzer for help in statistical analysis of nonparametric data sets, S Deixler and A Lepold for management of animal colony, T Asenov and his team in the Machine Shop, F Marr, H.C. Önal, A. Coll, as well as M. Schunn and the Preclinical Facility team for technical assistance. We thank the EMBL proteomics, flow cytometry, and genomics core facilities for the sample preparation, data generation, and analysis support. We thank M. van Gerven for helping with the SETD5 human mutation study. This work was supported by the Simons Foundation Autism Research Initiative (grant 401299) to G.N. and the DFG (SPP1738 grant NO 1249) to K-M.N.

Author contributions: E.D., N.A., J.M. and C.D. designed and performed experiments, analysed data and prepared figures. X.C. and E-L.K. performed experiments and data analysis. C.G. supervised data analysis. A.K., K.K. and I.C. performed experiments under the supervision of E.D. and/or J.M.. G.N. and K-M.N. conceived and supervised the study. G.N. wrote the paper together with K-M.N, E.D., N.A., J.M. and C.D. All authors read and approved the final version of the manuscript.

Competing interests: The authors declare no competing financial interests.

476 **References**
477

478 1 Mefford, H. C., Batshaw, M. L. & Hoffman, E. P. Genomics, intellectual
479 disability, and autism. *N Engl J Med* **366**, 733-743,
480 doi:10.1056/NEJMra1114194 (2012).
481 2 Rauch, A. *et al.* Range of genetic mutations associated with severe non-
482 syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*
483 **380**, 1674-1682, doi:10.1016/S0140-6736(12)61480-9
484 S0140-6736(12)61480-9 [pii] (2012).
485 3 Najmabadi, H. *et al.* Deep sequencing reveals 50 novel genes for recessive
486 cognitive disorders. *Nature* **478**, 57-63, doi:10.1038/nature10423
487 nature10423 [pii] (2011).
488 4 Dillon, S. C., Zhang, X., Trievel, R. C. & Cheng, X. The SET-domain protein
489 superfamily: protein lysine methyltransferases. *Genome biology* **6**, 227,
490 doi:10.1186/gb-2005-6-8-227 (2005).
491 5 Grozeva, D. *et al.* De novo loss-of-function mutations in SETD5, encoding a
492 methyltransferase in a 3p25 microdeletion syndrome critical region, cause
493 intellectual disability. *American journal of human genetics* **94**, 618-624,
494 doi:10.1016/j.ajhg.2014.03.006 (2014).
495 6 Kuechler, A. *et al.* Loss-of-function variants of SETD5 cause intellectual
496 disability and the core phenotype of microdeletion 3p25.3 syndrome. *Eur J*
497 *Hum Genet* **23**, 753-760, doi:10.1038/ejhg.2014.165
498 ejhg2014165 [pii] (2015).
499 7 Powis, Z. *et al.* Expansion and further delineation of the SETD5 phenotype
500 leading to global developmental delay, variable dysmorphic features, and
501 reduced penetrance. *Clin Genet*, doi:10.1111/cge.13132 (2017).

8 Green, C., Willoughby, J., Study, D. D. D. & Balasubramanian, M. De novo SETD5 loss-of-function variant as a cause for intellectual disability in a 10-year old boy with an aberrant blind ending bronchus. *Am J Med Genet A* **173**, 3165-3171, doi:10.1002/ajmg.a.38461 (2017).

9 Osipovich, A. B., Gangula, R., Vianna, P. G. & Magnuson, M. A. Setd5 is essential for mammalian development and the co-transcriptional regulation of histone acetylation. *Development* **143**, 4595-4607, doi:10.1242/dev.141465 (2016).

10 Baxter, L. L., Hou, L., Loftus, S. K. & Pavan, W. J. Spotlight on spotted mice: a review of white spotting mouse mutants and associated human pigmentation disorders. *Pigment Cell Res* **17**, 215-224, doi:10.1111/j.1600-0749.2004.00147.x (2004).

11 Cordero, D. R. *et al.* Cranial neural crest cells on the move: their roles in craniofacial development. *Am J Med Genet A* **155A**, 270-279, doi:10.1002/ajmg.a.33702 (2011).

12 Santos, A. R., Kanellopoulos, A. K. & Bagni, C. Learning and behavioral deficits associated with the absence of the fragile X mental retardation protein: what a fly and mouse model can teach us. *Learn Mem* **21**, 543-555, doi:10.1101/lm.035956.114 (2014).

13 Deacon, R. M. Assessing nest building in mice. *Nature protocols* **1**, 1117-1119, doi:10.1038/nprot.2006.170 (2006).

14 Moretti, P., Bouwknecht, J. A., Teague, R., Paylor, R. & Zoghbi, H. Y. Abnormalities of social interactions and home-cage behavior in a mouse model of Rett syndrome. *Hum Mol Genet* **14**, 205-220, doi:10.1093/hmg/ddi016 (2005).

15 Wohr, M. Ultrasonic vocalizations in Shank mouse models for autism spectrum disorders: detailed spectrographic analyses and developmental profiles. *Neuroscience and biobehavioral reviews* **43**, 199-212, doi:10.1016/j.neubiorev.2014.03.021 (2014).

16 Thomas, A. *et al.* Marble burying reflects a repetitive and perseverative behavior more than novelty-induced anxiety. *Psychopharmacology (Berl)* **204**, 361-373, doi:10.1007/s00213-009-1466-y (2009).

17 Wolmarans de, W., Stein, D. J. & Harvey, B. H. Of mice and marbles: Novel perspectives on burying behavior as a screening test for psychiatric illness. *Cogn Affect Behav Neurosci* **16**, 551-560, doi:10.3758/s13415-016-0413-8 (2016).

18 Moy, S. S. *et al.* Development of a mouse test for repetitive, restricted behaviors: relevance to autism. *Behavioural brain research* **188**, 178-194, doi:10.1016/j.bbr.2007.10.029 (2008).

19 Nolan, S. O. *et al.* Deletion of Fmr1 results in sex-specific changes in behavior. *Brain Behav* **7**, e00800, doi:10.1002/brb3.800 (2017).

20 Voikar, V. *et al.* Automated dissection of permanent effects of hippocampal or prefrontal lesions on performance at spatial, working memory and circadian timing tasks of C57BL/6 mice in IntelliCage. *Behav Brain Res*, doi:S0166-4328(17)30962-2 [pii]
10.1016/j.bbr.2017.08.048 (2017).

21 Balemans, M. C. *et al.* Hippocampal dysfunction in the Euchromatin histone methyltransferase 1 heterozygous knockout mouse model for Kleefstra syndrome. *Hum Mol Genet* **22**, 852-866, doi:10.1093/hmg/dds490 (2013).

22 McQuown, S. C. *et al.* HDAC3 is a critical negative regulator of long-term memory formation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **31**, 764-774, doi:10.1523/JNEUROSCI.5052-10.2011 (2011).

23 Nuytens, K. *et al.* Haploinsufficiency of the autism candidate gene Neurobeachin induces autism-like behaviors and affects cellular and molecular processes of synaptic plasticity in mice. *Neurobiol Dis* **51**, 144-151, doi:10.1016/j.nbd.2012.11.004 (2013).

24 Mews, P. *et al.* Acetyl-CoA synthetase regulates histone acetylation and hippocampal memory. *Nature* **546**, 381-386, doi:10.1038/nature22405 (2017).

25 Peixoto, L. L. *et al.* Memory acquisition and retrieval impact different epigenetic processes that regulate gene expression. *BMC Genomics* **16 Suppl 5**, S5, doi:10.1186/1471-2164-16-S5-S5 (2015).

26 Alarcon, J. M. *et al.* Selective modulation of some forms of schaffer collateral-CA1 synaptic plasticity in mice with a disruption of the CPEB-1 gene. *Learn Mem* **11**, 318-327, doi:10.1101/lm.72704 (2004).

27 Bontekoe, C. J. *et al.* Knockout mouse model for Fxr2: a model for mental retardation. *Hum Mol Genet* **11**, 487-498 (2002).

28 Deller, T. *et al.* Synaptopodin-deficient mice lack a spine apparatus and show deficits in synaptic plasticity. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10494-10499, doi:10.1073/pnas.1832384100 (2003).

29 Vigil, F. A., Mizuno, K., Lucchesi, W., Valls-Comamala, V. & Giese, K. P. Prevention of long-term memory loss after retrieval by an endogenous CaMKII inhibitor. *Sci Rep* **7**, 4040, doi:10.1038/s41598-017-04355-8 (2017).

30 Zhang, J., Hou, L., Klann, E. & Nelson, D. L. Altered hippocampal synaptic plasticity in the FMR1 gene family knockout mouse models. *J Neurophysiol* **101**, 2572-2580, doi:10.1152/jn.90558.2008 (2009).

31 Wilkinson, B., Li, J. & Coba, M. P. Synaptic GAP and GEF Complexes Cluster Proteins Essential for GTP Signaling. *Sci Rep* **7**, 5272, doi:10.1038/s41598-017-05588-3 (2017).

32 Zeng, M. *et al.* Phase Transition in Postsynaptic Densities Underlies Formation of Synaptic Complexes and Synaptic Plasticity. *Cell* **166**, 1163-1175 e1112, doi:10.1016/j.cell.2016.07.008 (2016).

33 Kim, S. *et al.* NGL family PSD-95-interacting adhesion molecules regulate excitatory synapse formation. *Nat Neurosci* **9**, 1294-1301, doi:10.1038/nn1763 (2006).

34 DeNardo, L. A., de Wit, J., Otto-Hitt, S. & Ghosh, A. NGL-2 regulates input-specific synapse development in CA1 pyramidal neurons. *Neuron* **76**, 762-775, doi:10.1016/j.neuron.2012.10.013 (2012).

35 Alfieri, A. *et al.* Synaptic Interactome Mining Reveals p140Cap as a New Hub for PSD Proteins Involved in Psychiatric and Neurological Disorders. *Front Mol Neurosci* **10**, 212, doi:10.3389/fnmol.2017.00212 (2017).

36 Repetto, D. *et al.* p140Cap regulates memory and synaptic plasticity through Src-mediated and citron-N-mediated actin reorganization. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 1542-1553, doi:10.1523/JNEUROSCI.2341-13.2014 (2014).

37 Yu, S. E. *et al.* SET domain-containing protein 5 is required for expression of primordial germ cell specification-associated genes in murine embryonic stem cells. *Cell Biochem Funct* **35**, 247-253, doi:10.1002/cbf.3269 (2017).

38 Mas, Y. M. S. *et al.* The Human Mixed Lineage Leukemia 5 (MLL5), a Sequentially and Structurally Divergent SET Domain-Containing Protein with No Intrinsic Catalytic Activity. *PloS one* **11**, e0165139, doi:10.1371/journal.pone.0165139 (2016).

39 Rincon-Arano, H., Halow, J., Delrow, J. J., Parkhurst, S. M. & Groudine, M. UpSET recruits HDAC complexes and restricts chromatin accessibility and acetylation at promoter regions. *Cell* **151**, 1214-1228, doi:10.1016/j.cell.2012.11.009 (2012).

40 Perissi, V. *et al.* TBL1 and TBLR1 phosphorylation on regulated gene promoters overcomes dual CtBP and NCoR/SMRT transcriptional repression checkpoints. *Mol Cell* **29**, 755-766, doi:10.1016/j.molcel.2008.01.020 (2008).

41 Gallagher, D. *et al.* Ankrd11 is a chromatin regulator involved in autism that is essential for neural development. *Dev Cell* **32**, 31-42, doi:10.1016/j.devcel.2014.11.031 S1534-5807(14)00770-9 [pii] (2015).

42 Chen, F. X. *et al.* PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell* **162**, 1003-1015, doi:10.1016/j.cell.2015.07.042 (2015).

43 Li, J. & Wang, C. Y. TBL1-TBLR1 and beta-catenin recruit each other to Wnt target-gene promoter for transcription activation and oncogenesis. *Nat Cell Biol* **10**, 160-169, doi:10.1038/ncb1684 (2008).

44 Brandler, W. M. *et al.* Paternally inherited cis-regulatory structural variants are associated with autism. *Science* **360**, 327-331, doi:10.1126/science.aan2261 (2018).

45 Grozeva, D. *et al.* Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Hum Mutat* **36**, 1197-1204, doi:10.1002/humu.22901 (2015).

46 Farwell Hagman, K. D. *et al.* Candidate-gene criteria for clinical reporting: diagnostic exome sequencing identifies altered candidate genes among 8% of patients with undiagnosed diseases. *Genet Med* **19**, 224-235, doi:10.1038/gim.2016.95 (2017).

47 Halvardson, J. *et al.* Mutations in HECW2 are associated with intellectual disability and epilepsy. *J Med Genet* **53**, 697-704, doi:10.1136/jmedgenet-2016-103814 jmedgenet-2016-103814 [pii] (2016).

48 Peleg, S. *et al.* Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science* **328**, 753-756, doi:10.1126/science.1186088 (2010).

49 Kwapis, J. L. *et al.* Context and Auditory Fear are Differentially Regulated by HDAC3 Activity in the Lateral and Basal Subnuclei of the Amygdala. *Neuropsychopharmacology* **42**, 1284-1294, doi:10.1038/npp.2016.274 npp2016274 [pii] (2017).

50 Nott, A. *et al.* Histone deacetylase 3 associates with MeCP2 to regulate FOXO and social behavior. *Nat Neurosci* **19**, 1497-1505, doi:10.1038/nn.4347 (2016).

Figure Legends

Fig. 1. Early developmental defects and altered gene expression at E9.5 in *Setd5*$^{+/-}$ mice. a, Representative images of brain and body from >P30 *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$ females indicating similar brain size, but abnormally small body in the mutant (left). Quantitative comparison of the normalised brain/body weight and kidney/body weight ratios in P1 pups (*n*=15) and adult females (*n*=12, bar graphs on the right) (values normalised to *Setd5*$^{+/+}$), *P<0.05 (two-tailed Mann Whitney U test), ***P<0.001 (one-way ANOVA/two-tailed t test), n.s. not significant; data presented as means and SEM, as well as scatter plot (exact P values and detailed statistics provided in Supplementary Table 1**;** also refer to **Supplementary Fig. 2c**). **b,** Representative picture of the white belly spot observed in 31% of *Setd5*$^{+/-}$ animals as compared to only 2% in *Setd5*$^{+/+}$ animals (total *n* is 222 *Setd5*$^{+/+}$ and 157 *Setd5*$^{+/-}$, both females and males were included). **c**, Characteristic photographs of *Setd5*$^{+/-}$ eye abnormalities found in 15% of *Setd5*$^{+/-}$ animals (bottom picture: from our own cohort; top picture: taken with permission from the International Mouse Phenotyping Consortium http://www.mousephenotype.org/data/genes/MGI:1920145); doughnut charts show observations made in our own cohorts (total *n* is 222 *Setd5*$^{+/+}$ and 157 *Setd5*$^{+/-}$); eye abnormalities in *Setd5*$^{+/-}$ mice include corectopia, mydriasis and microphthalmia; **d**, Volcano plot for differential expression analysis of genes in embryonic day (E) 9.5 *Setd5*$^{+/-}$ mouse embryo samples (*n*=3 embryos per genotype). Y-axis: negative decimal logarithm of adj. P value; X-axis: binary logarithm of fold change; colour code: genes significantly (sign.) (FDR<0.05, EdgeR, likelihood ratio test) down- (blue) and upregulated (red) in *Setd5*$^{+/-}$ embryos; not significant expression changes (grey); **e**, Selected biological processes (GO) enriched in differentially expressed genes (DEGs) in E9.5 animals; Inner circle: score for enrichment of either upregulated (red) or downregulated DEGs (blue). Numbers show maximum enrichment scores and gridlines indicate maximum and half maximum values. Colour intensity scales with p-value. Outer circle: log2(fold change) (y-axis) for up- and downregulated DEGs per term. Scale is indicated at the top of the plot, gridlines divide plot into four equal parts, grey line represents zero.

Fig. 2. Behavioural abnormalities and cognitive defects as consequence of *Setd5* haploinsufficiency in mice. a, Representative images of the nests built by controls and mutants (left) and nest scores (right) showing impaired nesting abilities

in *Setd5*$^{+/-}$ females (*n*=10 animals per genotype, *P<0.05, two-tailed Mann-Whitney U test). **b,** Isolation-induced ultrasonic vocalization profiles of *Setd5*$^{+/+}$ (grey) and *Setd5*$^{+/-}$ (red) female mouse pups displaying delayed peaking in the mutants (*n*=21 P2, 19 P4, 19 P6, 18 P8, 18 P10, 17 P12 female mouse pups per genotype, *P>0.05, **P>0.01, one-way ANOVA/two-tailed t test or Mann-Whitney U test – detailed statistics presented in Supplementary Table 1). **c-d**, Place avoidance task in the Intellicage: cartoon showing that nose poking in the correct corner allows access to water (green), while incorrect nose pokes (red corner) trigger an aversive air puff (black) (**c**, left); the two genotypes (*n*=14 females per genotype in both **c** and **d**) visit incorrect and correct corners similarly during the 48h long trial (**c**, middle), but *Setd5*$^{+/-}$ make more nose pokes per visit (**c**, right), at both correct and incorrect corners; the bar graph (**c**, right) and the kinetics of nose pokes/visit ratio (**d**) indicate that *Setd5*$^{+/+}$ females learn the task after 24h and nose poke significantly less at the incorrect corner during the second peak of activity, while the *Setd5*$^{+/-}$ females continue to nose poke similarly often (**d**); peaks of the fitted curves in **d** have amplitudes (mean and SEM) of 7.25 ± 0.62 nose pokes per incorrect corner (control, grey) and 8.72 ± 0.83 (mutant, red), during the first 24h (learning phase), and of 4.63 ± 0.39 (control) and 8.75 ± 0.7 (mutant) during the last 24h (test phase). **e,** Contextual fear-conditioned memory retention and extinction scored as percent freezing during a 3-min exposure to the context, *n*=18 female mice per genotype; **f,** Subthreshold training (3 min) in novel object location memory test induces abnormal memory retention in mutants (*n*=13 females per genotype). ***P<0.001, **P<0.01, *P<0.05, n.s., not significant; data presented either as means and SEM, or as medians, as well as scatter plot. For detailed statistics, please refer to Supplementary Table 1.

Fig. 3. Increased LTP in *Setd5*$^{+/-}$ mice. a, Comparison of the post-tetanic potentiation and early-LTP (end of first hour after induction) of the CA3-CA1 synapses in the stratum radiatum of acute dorsal hippocampal slices obtained from *n*=10 male mice (P21-P24) of either genotype; one slice per mouse; stimuli were applied every 30s and results are means ± SEM for all mice; a stable baseline was recorded for 30-60 min (shown here – 10 min) and LTP was induced at time 0 (arrow) by high frequency stimulation (HFS, 4 x 100 stimuli of 0.2 ms at 100 Hz, every 5s). Right panel shows representative traces of fEPSPs obtained at baseline

(top), 1h after HFS (middle, early-LTP) and 5h after HFS (bottom, late-LTP) from either genotype (also refer to **Supplementary Fig. 7f** containing the means ± SEM for entire recordings). **b,** Quantifications of the percent potentiation at various moments after LTP induction, in dorsal hippocampal slices from control and mutant animals, showing elevated early LTP (1h, *n*=10) and late LTP (3h *n*=10, 4h *n*=8, 5h *n*=6) in *Setd5*$^{+/-}$ mice; *P<0.05. Boxplots indicate medians (middle line), 25%-75% interquartile range (box) and minimum and maximum data point (whiskers). For detailed statistics, please refer to Supplementary Table 1.

Fig. 4. Gene expression dynamics are altered in the *Setd5*$^{+/-}$ hippocampus after CFC training. a, Volcano plots displaying gene expression changes 1h or 3h after conditioning in wild type (*n*=4 animals) and mutant (*n*=5 animals): 280, 229, 169 and 273 significant DEGs, respectively (FDR <=0.05, EdgeR, likelihood ratio test). The y-axis shows the negative decimal logarithm of the adjusted P value and the binary logarithm of the fold change is shown on the x-axis; blue/red, genes down- and upregulated at the later time point, respectively; green, differentially expressed immediate-early genes (see reference 22); grey, not significant. **b,** Genes with significantly different responses (*n*=286) to the CFC clustered by their expression trajectories. Top row (left to right): 15, 43, 18 and 23 genes, middle row: 44, 24, 33 and 21 genes, bottom row: 37, 16 and 12 genes, respectively. Thick lines, average of all genes per cluster; thin lines, single genes; grey, *Setd5*$^{+/+}$; red, *Setd5*$^{+/-}$; homecage (HC), *n*=5; 1h, *n*=4; 3h, *n*=5 animals per genotype; **c,** Schematic illustration of a synapse highlighting post-synaptic genes exhibiting different responses in *Setd5*$^{+/-}$ animals upon CFC. Colouring indicates association of genes to clusters in **b**.

Fig. 5. SETD5 lacks methyltransferase activity and interacts with PAF1 and HDAC3 protein complexes. a, Schematic representation of the SET domain methylation assay. The GST-tagged SET domain of SETD5 is incubated with a ^{3}H-labeled S-Adenosyl methionine (SAM) substrate and mono- or oligonucleosomes purified from HeLa cells, followed by SDS-PAGE. The methylation activity signal is detected by autoradiography. **b,** Methylation assay results for the GST-tagged SET domain. Left: radioactive films for the SET domain; Right: PRC2 complex used as a positive control for the assay. Arrowheads indicate background levels of histone H3

methylation caused by contaminants bound to nucleosomes (Nuc.). The experiment was repeated three times with similar results. **c,** Methylation assay results for full length FLAG-tagged SETD5 overexpressed in HEK293T cells. Left: radioactive film results; Right: coomassie staining of SDS-PAGE; Bottom-right: Western blot of immunoprecipitated FLAG-tagged SETD5. Arrowheads indicate FLAG-beads immunoglobulins (Ig); b, FLAG-beads; the experiment was performed once. Refer to **Supplementary Fig. 14** for full scans of films, blots and gel in **b**-**c; d,** Correlation plot of the Setd5 interacting proteins in ESCs (x-axis) and NPCs (y-axis) identified by quantitative TMT mass spectrometry. Immunoprecipitation samples of the endogenously tagged Setd5-HA are compared with non-tagged wild type samples. Colour code: proteins significantly enriched in both ESCs and NPCs (red); significant only in ESCs (yellow) or in NPCs (purple); not significant (grey). Adjusted P value < 0.05 and $\log_2$ (Fold Change) > 0, limma statistical analysis; *n*=2 biologically independent samples. **e,** Interaction network of proteins found to significantly interact with Setd5 both in ESCs and NPCs from both TMT and label free mass spectrometry. Black edge thickness indicates interaction confidence from data generated in this study. Thin edge colour indicates String interaction score (blue, low; red, high). Orange, Setd5 bait; blue, Hdac3-Ncor-Tbl1x complex; green, Paf1 complex; grey, other proteins significantly interacting with Setd5.

Fig. 6. SETD5 patient mutations disrupt SETD5-HDAC3/-PAF1 interactions. a, Schematics of SETD5 rescue constructs representing selected human patient mutations. Open reading frame carrying the disease-associated truncating mutations or the endogenous stop codon is fused in frame with a P2A sequence and a GFP reporter. All constructs carry a N-terminal FLAG-HA tag. **b,** Western blot of wild type (WT) and truncated mutant SETD5 proteins overexpressed in HEK cells. Replicate lanes come from two independent cell transfections; samples were separated into cytoplasmic and nuclear fractions. Neg., empty vector; black arrows indicate expected bands for each construct. The experiment was repeated twice with similar results. **c,** Western blots of FLAG immunoprecipitation of SETD5 WT, SETD5 1258*, and SETD5 E720* in *Setd5*$^{-/-}$ ESC rescue lines. The nuclear protein Oct4 served as negative control. IN, input: nuclear extract before immunoprecipitation. FT, flow through, unbound to beads; IP, immunoprecipitation eluate with triple-FLAG peptide; 1% of IN and FT and 50% of IP eluate were loaded. The experiment was repeated

three times with similar results. All immunoblots shown are cropped; see **Supplementary Fig. 14** for full-length blots in **b**-**c**.

Fig. 7. *Setd5* haploinsufficiency in ESCs results in increased Pol II occupancy at the TSS of a subset of neurodevelopmental genes. a, Boxplots of ChIP-seq signal for 5 different genomic features: transcription start sites (TSS) bound (Pol II TSS, *n*=16342) or not bound (Inactive TSS, *n*=17817) by RNA-Pol II, transcription end sites (TES, *n*=34159), intergenic ESCs specific enhancers (*n*=4099), and other tissue intergenic enhancers (*n*=43657). Setd5 ChIP-seq was performed using endogenously tagged (Setd5-HA) compared to non-tagged ($Setd5^{+/+}$) ESCs for antibody background. One-sided Mann–Whitney U test; p-values are in the figure. **b,** Boxplots of Setd5 ChIP-seq signal at 4 genomic regions: TSS with either or both Hdac3 and Pol II (+Hdac3 +Pol II, *n*=13314; +Hdac3 –Pol II, *n*=701; –Hdac3 +Pol II, *n*=2890), and intergenic ESCs specific enhancers with Hdac3 peaks (+Hdac3 ESCs Enhancer, *n*=688); one-sided Mann–Whitney U test. **c,** Representative ChIP-seq tracks of Setd5-HA and Pol II in $Setd5^{+/HA}$ and *Setd5* mutant ESCs at the Meis2 gene locus. Signals are input-subtracted, merged replicates Reads per Genomic Content (RPGC); Refer also to **Supplementary Fig. 14** for complete tracks; experiments performed with two independent clonal lines with similar results. **d,** Density plot of genes with reduced (upper panel, *n*=4144) or increase (lower panel, *n*=2487) Pol II pausing index in both $Setd5^{+/-}$ and $Setd5^{-/-}$; X-axis is in log10 scale; one-sided Kolmogorov–Smirnov test, as compared to $Setd5^{+/+}$. **e,** Boxplots showing Setd5-HA, Pol II, Hdac3, H3K27ac and H4 pan acetylation ChIP-seq signals in $Setd5^{+/HA}$ and *Setd5* mutants with significantly increased Pol II occupancy at TSS in $Setd5^{+/-}$ (FDR < 0.05, EdgeR statistical analysis, **Supplementary Fig. 12e**), *n*=559 genes. **f,** Gene ontology enrichment analysis for genes presented in (**e**); GOstats statistical analysis. **g,** Boxplot showing RNA-seq counts for a subset of genes with increased Pol II at the TSS. Genes in (**e**) were clustered (k-means) and a cluster (*n*=124 genes) with more reads in $Setd5^{+/-}$ ESCs was presented. $Setd5^{+/+}$ and CRISPR/Cas9 clonal control lines ($Setd5^{+/+;\ tr.}$) were merged. One-sided Mann–Whitney *U* test; Y-axis in log10 scale. **h**, Heatmap of 35 representative genes used in (**g**). Boxplots in **a**, **b** and **e** represent means of 2 biological replicates from two independent clonal lines, normalized by input RPKM values (dashed horizontal line represents input baseline). Data points**,** less than 5 or more than 95 percentile of mean RPKM, were removed in

a, **b,** and **e**. All data points are presented in **g.** The upper, centre, and lower line of the boxplot indicates 75%, 50%, and 25% quantile, respectively. Whiskers extend to the most extreme datapoint within 1.5-times the interquartile range. Outliers are not shown.

METHODS

Ethical approval

All animal protocols complied with directive 2010/63/EU of the European Parliament and Council, the 3R (reduce, replace, refine) principle (Russel and Burch) and were approved by the Institutional Animal Care and Use and Ethical Committee at IST Austria.

Mice

Setd5$^{tm1a(EUCOMM)Wtsi}$ (International Mouse Phenotyping Consortium) were mated with homozygous Flip mice (Jackson Laboratories) to remove the NeoStop cassette. The resultant heterozygous *Setd5*$^{+/fl}$ females (Supplementary Fig. 1a) were crossed with CMVCre (B6.C-Tg(CMV-cre)1Cgn/J) males to obtain *Setd5*$^{+/-}$ mice (Supplementary Fig. 1a). *Setd5*$^{+/-}$ mice were backcrossed to the N10 generation in C57BL/6J mice and used for our experiments. The Setd5^{GFP} mouse has been described elsewhere[51] and was kindly provided by Prof M.A. Magnuson. Animals were housed in groups of 3-4 animals per cage and kept on a 12 h light/dark cycle (lights on at 7:00 am), with food and water available *ad libitum*, unless otherwise specified.

Generation of *Setd5* knock in and knock out cell lines

The *Setd5* knock in cell line was created by adapting a published method[52]. A donor plasmid was generated from the pFETCh_Donor plasmid (addgene ID #63934). We exchanged the Neomycin/Kanamycin resistance with Emerald GFP, introduced an HA tag upstream of the triple FLAG tag, and cloned 1Kb of Setd5 homology arms upstream and downstream of the construct. ESCs were electroporated with a CRISPR/Cas9 plasmid targeting the unique transcription end site of *Setd5* together with the modified pFETCH_Donor plasmid. To increase homologous-directed recombination efficacy, cells were treated with 10 μM of SCR7 (Xcessbio Biosciences) for 24h before the electroporation. Successfully electroporated cells were selected by single cell fluorescence-activated cell sorting (FACS) for GFP positive events. Correct insertion was confirmed by genotyping PCR and protein expression was tested using western blot (WB).

Setd5 knock out cell lines were generated using CRISPR/Cas9. For the generation of heterozygous *Setd5* line 1kb homology arms matching intron 2 and intron 15 of the *Setd5* locus were cloned inside the HR110PA-1 plasmid (System Biosciences) carrying a red fluorescent protein (RFP) and puromycin resistance cassette (PuroR). ESCs were electroporated with two CRISPR/Cas9 plasmids (pSpCas9(BB)-2A-GFP (PX458), addgene ID #48138) targeting intron 2 and intron 15 of *Setd5* and the HR100PA-1 donor plasmid carrying the RFP/PuroR cassette. Positive clones were selected by 1μg/ml puromycin treatment for 1 week. Validation of the positive cassette insertion in the *Setd5* locus was carried out by genotyping PCR. Biological replicates come from different clones.

The generation of ΔSET *Setd5* heterozygous or homozygous cell lines was performed by transfecting ESCs with two CRISPR/Cas9 plasmids with guides targeting intronic regions 20-to-70bp upstream of exon 7 or downstream of exon 8. Single cells were FACS sorted and the resulting clones were genotyped by PCR. Biological replicates are different clonal lines carrying the same mutation. CRISPR control lines are single cell-derived clones that were not edited in the targeted site or

in the top 10 potential off target sites. CRISPR/Cas9 guides were designed using Optimized CRISPR Design tool (http://crispr.mit.edu/). We choose guides with predicted off target sites in intergenic regions. For each guide we selected the top 10 off target sites and designed PCR primers spanning 300bp upstream and downstream from the predicted sites. PCR bands were resolved on 1% agarose gel. None of the guides showed a band shift in the clonal lines compared to wild type (WT) controls (data not shown). To check for small DNA indels the bands were cut out from the gel and sent for Sanger sequencing. Sequences were aligned against the reference mouse genome and no guide showed signs of off target activity in the predicted site.

Methylation assay

The human GST-tagged SET domain of SETD5 (E200-E566) was purified from BL21(DE3) E.Coli. For each reaction 2 μg of recombinant SET or 2 μg of recombinant PRC2 complex (Active Motif #31387) were incubated with 10 μg of HeLa cells purified mononucleosomes or 2 μg of di/tri- or oligonucleosomes together with 5 μCi of ^{3}H-S-Adenosylmethionine. The final reaction volume was brought to 30 μL with methylation assay buffer (20 mM Tris-HCl pH 8, 10% Glycerol, 5 mM Mg_2Cl, 60 mM KCl, and 2 mM DTT). Samples were incubated for 2h at 30°C. The reaction was blocked by adding SDS-loading buffer and boiling the samples at 95°C for 10 minutes. The remaining ^{3}H-S-Adenosylmethionine was separated from the rest of the proteins by SDS-PAGE. Coomassie stained gels were incubated with Amersham Amplify Fluorographic Reagent (GE NAMP100V) for 30 minutes followed by gel dehydration for 2h at 60°C. Radioactive signal was detected by exposing the gel to autoradiography film from 4h up to 1 month at -80°C.

Messenger RNA-sequencing (mRNA-seq)

ESC, EB and NPC mRNA extraction was performed as described in the Supplementary note. RNA extracts that passed quality control were used for library preparation. Libraries were prepared using the oligo-dT capture kit (NEB) and sequenced on Illumina HiSeq 2000 at EMBL, Heidelberg Genomics Core facility. Fastq files that passed quality control were aligned to the mm10 reference genome using bowtie2 v0.3[53], the generated bam files were used for counting reads using the HTseq tool v0.6.1[54]. Coherence between samples, time points and replicates was verified by principal component analysis (PCA). Differential expression analysis was performed using the R package DESeq2 v1.16[55]. Gene Ontology enrichment analysis was done using the Bioconductor package GOstats version 2.46.0[56].

Mouse tissues for mRNAseq were rapidly dissected on ice and snap frozen in liquid nitrogen. For embryonic day (E) 9.5 embryos genotyping was carried out on placental tissues. Whole embryos or the left hippocampal CA region were used for RNA extraction and library preparation.

Sequenced libraries raw reads were trimmed before alignment to mm10 reference genome using STAR version 2.5.4[57]. Read counts per gene were derived using STAR (Option -quantMode GeneCounts). Coherence of biological replicates was assessed using PCA, and sample distance clustering; one sample (WT, 1h) was excluded due to its clustering away from all other samples. Differential expression analysis of CA and mouse embryos was performed employing RUVSeq as described[58] and EdgeR v3.22.2[59] including the RUV variables in the design. Genes with FDR<=0.05 were deemed significantly differentially expressed. Averaged log transformed CPM values were normalized by calculating gene-wise z-scores. K-

means clustering was performed using Pearson correlation as distance measure and selecting a suitable cluster number (k=11) with minimal Davies Bouldin index and a minimum cluster size above 10 genes. To identify clusters with similar trajectories between genotypes, genes differentially expressed in at least one comparison were clustered in the same way (k=22). Gene Ontology enrichment analysis was done using GOstats v2.46.0[56] with a P value cut off of 0.001 and conditional testing enabled. For the E9.5 data random sampling was used to exclude terms enriched in random gene sets to control for enrichment in overall expressed genes. E9.5 mouse embryos GO enrichment results were visualized using a custom script and the GOplot package (1.0.2)[60]. Overlap between E9.5 data and *in vitro* datasets was visualized using a customized script based on the circlize package (0.4.4)[61]. Enrichment for cell type specific expression was performed using the EWCE tool (0.99.2)[62] using the scRNA-seq data from Zeisel et al[63]. Enrichment analysis for disease genes (OMIM), transcription factor targets as well as pathway related genes was done via the Enrichr web tool[64,65]. Other enrichment analyses were all done using a one-sided Fisher test in R. Hdac3 target genes for comparison with CA data were derived from published ChIP-seq data from mouse hippocampus[66], E9.5 data was compared to ChIP-seq data generated in this study. Refer to the Supplementary note for more details.

Setd5 immunoprecipitation (IP) followed by liquid chromatography-mass spectrometry (LC-MS)

For the IP of endogenously tagged Setd5 we used ESCs and NPCs. Briefly, 200 million cells were harvested and snap frozen in liquid nitrogen. Cells were lysed as described[67]. Nuclear enriched pellets were resuspended in a 420 mM salt buffer, incubated 15 min at 4°C and spun down to remove insoluble material. The soluble fraction was diluted to 150 mM salt concentration. Nuclear enriched samples were incubated with 15 µL of anti-FLAG M2 beads (Sigma Aldrich) overnight at 4°C. After the incubation, beads were washed 3 times with a 300 mM wash buffer, followed by a two-times 1h elution with 40 µL of FLAG elution buffer containing 0.25 mg/mL triple FLAG peptide at 4°C. The resulting immunoprecipitated material was analyzed by WB to verify a successful IP and submitted for sample preparation at the EMBL, Heidelberg Proteomics facility. Samples were prepared with the SP3 protocol[68] and digested peptides analyzed by liquid chromatography followed by label free or tandem mass tag (TMT) labeled mass spectrometry. *Setd5* knock in samples were compared against a matching background WT negative control lacking the endogenous Setd5 FLAG-HA. Each sample at each time point was processed in duplicates. For data analysis, label free peptides were identified and mapped using Isobarquant software[69], while the TMT labeled peptides were mapped and quantified by MaxQuant software[70] v1.5.6.5. For both datasets the raw peptide counts data were analyzed by applying a variance stabilization normalization method with the R package vsn (v 3.44)[71]. Possible batch effects were removed by fitting a linear model to the data that explains the variance between the replicates, using R package limma v3.32[72]. Coherence between samples, time points and replicates was verified by PCA. Differential protein IP analysis was performed using R package limma; proteins with an adjusted P value < 0.05 and $\log_2$(Fold Change) > 0 were considered to have a significantly different binding between the endogenous tagged Setd5 and the matching WT control. The correlation plot between the ESC stage and the NPC stage was obtained by comparing the log2 fold change of the 2 time points. Protein-protein interaction network analysis was performed with Cytoscape[73] using STRING

database v10[74] information channels and the data generated in this study. Protein-protein Interactions from STRING database with a combined score lower than 300 were discarded . Both label free and TMT labeled significant proteins were used to generate the network. To quantify the protein-protein interaction between Setd5 and the other proteins (black edge thickness) the $\log_2$(Fold Change) and the -$\log_{10}$(adjusted P value) of significantly enriched proteins from both label free and TMT mass spectrometry experiments were scaled from 0 to 1 and then summed together to generate a unified score value. Proteins with scaled fold change lower than 0.048 and scaled adjusted P value lower than 0.04 were discard from the pool of proteins used to build the networks. The $\log_2$(Fold Change) and the -$\log_{10}$(adjusted P value) of proteins identified both in ESCs and NPCs were summed together to present a unique interaction value during neurodevelopment.

Chromatin IP coupled with sequencing (ChIP-seq)

For Hdac3, Pol II and Setd5 ChIP each replicate of 100 million ESCs were harvested and cross linked in 1.5 mM EGS in PBS for 1h followed by 10 min cross linking in 1.5% paraformaldehyde (PFA) in PBS at RT. For H3K27ac and H4 pan acetylation each replicate of 20 million ESCs were harvested and cross linked in 1% PFA. Cross-linking solution was quenched with 125 mM Glycine for 5 min. Fixed cells were spun down at 3000 rpm for 10 min at 4 °C and washed twice with ice cold PBS freshly supplemented with complete protease inhibitor cocktail. Samples were snap frozen in liquid nitrogen and stored at -80 °C. Briefly, chromatin was prepared by re-suspending cross linked cells in hypotonic buffer on ice for 10 min, followed by 35 strokes of douncing with a tissue homogenizer. The cytoplasmic fraction was removed by gentle centrifugation and nuclear pellets were re-suspended in MNase digestion buffer. Nuclei were digested with MNase (Worthington) at a ratio 15 μL per 50 million cells (MNase conc: 25U/μL) for 5 min at 37 °C. MNase enzymatic activity was quenched by adding EDTA/EGTA based quenching buffer. Digested chromatin was sonicated with 3 cycles of Bioruptor Pico sonicator (Diagenode). To separate insoluble chromatin, samples were spun down at 15,000*g* for 10 min at 4 °C. Soluble fragmented chromatin was used as ChIP input. Both soluble input and precipitated pellet were resolved on agarose DNA gel to check the digestion pattern across replicates and 5% of each chromatin input were set aside before the IP. Each sample input was incubated with 70 μL of Invitrogen M-280 beads (Invitrogen) pre-coupled for 6h at 4°C with the specific antibody (see Life Sciences Reporting Summary for details). Samples were incubated overnight at 4°C and the following day beads were washed four times with LB3-100 buffer (10 mM TRIS, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% Sarcosine) followed by one wash with LB3-500 buffer (500 mM NaCl), followed by four washes with RIPA-250 buffer (10 mM Hepes, 250 mM LiCl, 1 mM EDTA, 0.7% Na-Deoxycholate, 1% Igepal-630), finally beads were quickly rinsed with 10 mM Tris/1 mM EDTA buffer. Elution of the IP chromatin from the beads was performed in elution buffer (1% SDS, 50 mM Tris, 10 mM EDTA) for 30 min at 65 °C. Eluted chromatin and ChIP inputs were reverse cross linked and proteins were digested with proteinase K overnight at 65 °C. DNA was purified using a PCR purification kit (QIAgen). Libraries were prepared with NEBNext Ultra II kit. For each ChIP two biological replicates per condition (WT or ΔSET *Setd5* heterozygous or homozygous) and one input per *Setd5* genotype, or Setd5 WT and Setd5-HA in the case of Setd5 ChIP, were pooled together and sequenced on an Illumina HiSeq 2000 at EMBL, Heidelberg Gene Core facility.

ChIP-seq data analysis

Analysis was performed in a local installation of Galaxy[75] maintained by the EMBL Genome Biology Computational Support using mouse genome version MM10 and in R version 3.5.0 (Bioconductor version 3.7). Data visualisations were performed using custom-made scripts in R. The 50bp single-end reads were aligned to the mouse genome with Bowtie version 2[53] using standard options (Galaxy Tool version 0.2, sensitive preset). Reads failing to be mapped or mapping at several locations (as identified by the XS tag set by bowtie2) were removed. Read duplicates were identified and removed using Picard's Mark Duplicates (http://broadinstitute.github.io/picard). Sequencing data quality was assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and the Deeptools2 package[76]. Peak calling was performed using MACS version 2[77]. See Life Sciences Reporting Summary for peak calling details. For all experiments MACS2 defined peaks from all conditions (WT and two mutants) were merged into a unique non-overlapping set of 25234 peaks for Hdac3 (filtered for Qvalue < 0.001) 129203 peaks for Pol II (filtered for Qvalue < 0.001), 45842 peaks for H3K27ac (filtered for Q-value < 0.00025) and 53778 for H4 pan acetylation (filtered for Qvalue < 0.1) using bedtools[78]. Enhancers from Shen et al[79] were adjusted to 2Kb (centered on their middle point as defined in downloaded files) and lifted from mm9 to mm10 using the UCSC liftover tool. For ChIP-seq analysis at transcription start sites, TSSs were defined using ENSEMBL version 91 considering gene start position for genes longer than 1250bp (total number of genes 34159). Peaks falling within 2.5Kb regions centered on TSS were considered for differential binding analysis. Peak intersection with defined TSS resulted in 14015 peaks for Hdac3, 16152 peaks for Pol II, and 8804 for H3K27ac. Differential analysis between mutant and WT samples was performed using the DiffBind[80] and edgeR[59] packages. Differentially bound peaks with corrected P values lower than 0.05 were selected for further analysis. RPKM signal were calculate using the Deeptool2 "multiBamSummary" tool from duplicates removed bam files, the count tables were than normalized to genomic region length and sample library size, replicates were averaged and all samples were background normalized by dividing them by the merged input RPKM. Signal files (bigwig format) were generated using the Deeptools2 "bam-Coverage" tool using the "Normalize coverage to 1x" option (to correct for library size). Replicates were averaged and input subtracted (IP minus input) using the Deeptools2 "bigwig-Compare" tool. Pol II pausing index definition was adapted from[81]. For each Setd5 merged replicate, input subtracted bigwig sample the average signal at TSS (-100 /+300bp) was divided by the gene body signal downstream of the TSS (+301 to +3301bp). Genes shorter than 3.5Kb were not included in the analysis, genes with signal lower than 1 at TSS and at the 3Kb gene body were removed from the analysis, in total 9490 were included in the pausing index calculation. The correlation between biological replicates was verified (Pearson's r ~ 0.98 Pol II at TSS and Gene body; r ~ 0.89 Hdac3 at TSS) and samples were merged for data visualisation.

Rescue experiment

SETD5 point mutations identified in patients were obtained from the Sfari Gene database and were selected to be equally spaced across the entire protein sequence. We chose six mutations at R308[82], K399[83], T552[84], E720[85], R1001 and S1258[82,83] all resulting in the gain of a stop codon. The WT and truncating mutations rescue constructs were designed to have an N-terminal FLAG tagged SETD5

followed by a P2A cleavage site and EmGFP marker. The WT human SETD5 open reading frame and the truncated versions were cloned from the pFN21A - SETD5-HaloTag construct (Promega). All six constructs, including WT, were cloned inside the pDonor MCS Rosa26 (addgene ID #37200) plasmid carrying 800bp homology arms of the Rosa26 gene locus. These donor plasmids were used to generate rescue clonal lines from ΔSET homozygous Setd5 cells. Cells were transfected with CRISPR/Cas9 and a single guide RNA targeting the first intron of the Rosa26 gene and the WT or truncated SETD5 rescue construct. Positive cells were FACS sorted selecting for GFP positive events. Correct insertion of the rescue gene was confirmed by genotyping PCRs and WBs. Different clonal cell lines carrying the same rescue construct were considered biological replicates and were derived from the same CRISPR/Cas9 transfection.

Behavioural studies

Behavioural studies were carried out during the light period. Two to 3.5 month old age- and sex-matched littermate mice were habituated to the test room for at least 1 h before each test. At least 1-week-long rest periods were given between tests. Behavior apparatuses were cleaned between trials with 70% ethanol. All behavioural tests were performed starting with the least aversive task first and ending with the most aversive, and either scored automatically or by an experimenter blind to the genotype. Both females and males were tested unless stated otherwise. When no inter-genotypic differences were observed the data for both sexes was pooled to avoid redundant graphs. However, in case of inter-genotypic differences, although both male and females show essentially the same phenotype, the data for the two sexes is reported separately. Extensive description of the behavioural studies can be found in the Supplementary note.

Histone acetylation

Histone extraction of the right CA hippocampal region from animals that underwent CFC was performed employing the EpiQuik Total Histone Extraction Kit (Epigentek, Cat. Nr. OP-0006-100), according to the manufacturer's indications. Detailed description of the protocol used here can be found in the Supplementary note.

Hdac3-IP and HDAC3 activity assay from adult mouse forebrain

Hdac3-IP and activity assays were performed according to a modified protocol, previously described[86]. Female 2 to 4 month old *Setd5*$^{+/-}$ and their WT littermates were sacrificed by cervical dislocation and the hippocampus and cortex (forebrain) of both hemispheres were rapidly dissected on ice. After washes in ice-cold 1x PBS, tissues were homogenized in a tissue grinder in ice-cold IP-RIPA-buffer (50 mM Tris pH 8.0, 120 mM NaCl, 0.5% NP-40 and 1 mM EDTA) with protease inhibitors (Roche, Ref. 04 693 159 0019). Tissue lysis was allowed for 30 min on ice, followed by 30 min centrifugation at 4°C at 13000 rpm. Supernatant was collected and protein concentration was determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher, Cat. no. 23225). Lysates (IP-input) were directly used for the HDAC3-IP followed by the HDAC3 activity assay (Sigma, EPI004).

2 mg of total protein lysate was brought to 400 µl with IP-RIPA and cleared under agitation for 1h on Protein A magnetic beads (Abcam, ab214286). Cleared lysates were collected and split into two 200 µl (1 mg total protein) aliquots. Both samples were incubated overnight (12-16h), at 4°C rotating, either with anti-HDAC3 conjugated protein A magnetic beads (IP+ samples, beads conjugated for 2h with 3.35 µg rabbit anti-HDAC3 primary antibody, Abcam, ab32369- for antibody details

refer to Life Sciences Reporting Summary) or with empty beads (IP- samples). The flow through was collected (FT samples), immune complexes washed and directly dissolved in HDAC3 assay buffer (provided with EPI004) and equilibrated for 1h on ice. The assay was performed according to the manufacturer's guidelines and the supernatant was used for flourometric measurement in polystyrene plates (Corning) at Em/Ex=380/500 using a Synergy H1 Hybrid Reader (Biotek). HDAC3 activity was normalized to total HDAC3 protein levels as determined by western blotting, see above, using mouse anti-HDAC3 (Santa Cruz, sc-376957, 1:200 in 5% milk, 1x TBST).

Hdac3-IP and HDAC activity assay from ESCs

For Hdac3-IP, 100 million *Setd5*$^{+/+}$ and *Setd5*$^{-/-}$ ESCs were prepared with the same protocol as for Setd5-IP. Nuclear extract were incubated over night with Dynabeads Protein G (Thermo Fischer, #1004D) pre-coupled with 20 μg of Hdac3 antibody (ab7030) per IP. Beads were washed 4 times with 300mM salt wash buffer. HDAC activity was tested using HDAC-GLOI/II assay kit (Promega, G6420, lot no 0000279650). Washed beads were resuspended in 100 μL Hdac assay buffer and 15 µL of resuspended beads were used for each reaction. All conditions were tested in triplicates. Trichostatin A (TSA) was used as enzymatic inhibitor at 50nM final concentration. Luminescence (400-750nm) was recorded on a 96-well plate reader (Tecan, inifinite M1000 pro) every 2 min for 40 cycles at 25°C. Reported data is from 45 minutes after reaction start.

Immunostaining, Nissl and fluorescence *in situ* hybridization (FISH) of adult and embryonic brains and BrdU pulse-labeling

For fluorescent immunostainings and Nissl stainings in adult mice *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$ animals were transcardially perfused, brains dissected, postfixed in 4% PFA, dehydrated and sliced at 40 μm. For bromodeoxyuridine (BrdU) pulse labeling pregnant females were injected with BrdU and sacrificed 2h later. Embryo heads were dissected, fixed overnight, dehydrated and sliced at 16 μm. For *FISH* experiments on adult wildtypes, brains of C57BL/6J animals were dissected, immediately frozen in O.C.T, sectioned at 20 μm and stored at -80°C. For detailed experimental protocols of imaging and quantifications refer to the Supplementary note, for details about primary antibodies used refer to the Life Sciences Reporting Summary.

Hippocampal slice preparation and electrophysiology

Transverse dorsal hippocampal slices (300 µm) were prepared from P21-P24 male littermates. Slices were recovered for 1h at 33 °C, and then at RT (21-23 °C) in carbogenated artificial cerebrospinal fluid (aCSF) containing (in mM): 130 NaCl, 2.75 KCl, 1.1 NaH_2PO_4, 28.82 $NaHCO_3$, 11 glucose, 1.43 $MgSO_4$, 2.5 $CaCl_2$, 5 Na ascorbate and 3 Na pyruvate (~320 mOsm, 7.2–7.4 pH). Filed potential in the CA1 stratum radiatum were generated and recorded using aCSF-filled electrodes. Signals were amplified by a Multiclamp 700B amplifier (Axon Instruments/Molecular Devices), and then digitized using Digidata 1550A (Molecular Devices). Extended protocol is described in the supplementary note.

Data analysis

All data analysis was performed blind to the genotype. Data exclusion criteria have always been pre-established and respected throughout data acquisition and

analysis. Examples of exclusion criteria: death of an animal during data acquisition lead to the exclusion of its littermate sibling; no LTP observable in the control mouse (experimental problems assumed) - both the control and the sibling mutant removed from data sets; one control-mutant pair was removed from the LTP data set after noticing that in the case of the mutant, the recording had been performed in a slice from the intermediate hippocampus, and not dorsal, like all other. We discarded all data resulting from experiments for protocol optimisation, because the working protocol was not respected. No outlier was removed from any data set with the exception of one sample in the RNA-sequencing analysis (1h, WT) that clustered clearly away from all other samples (clustering by euclidean sample distance and PCA) during the preliminary quality assessment.

Statistics

Statistical analyses were performed using R, Origin Software (Origin Inc) and GraphPad Prism. Shapiro–Wilk test was used to evaluate normal distribution and Levene's test for equal variance of all data sets. Parametric data were analyzed for significance using one-way or two-way ANOVA with Bonferroni post-hoc test, using $*P < 0.05$, $**P < 0.01$, $***P<0.001$, and presented as scatter plots and mean ± standard error of the mean (SEM), unless otherwise specified. Data sets with non-normal distributions and/or unequal variance were analyzed using the 2-tailed Mann-Whitney U test or Kruskal-Wallis test and presented as boxplots and individual data points. Adjustment for multiple comparisons were made using post-hoc tests Bonferroni (parametric data) or Dunn (non-parametric data). The data met the assumptions of the statistical test used and were generally evaluated in terms of distribution and variances, as seen in Supplementary Table 1. Although we report to have used two-way ANOVA with repeated measures and Kruskal-Wallis tests using corrections for multiple comparisons, our data sets were also evaluated in a generalized linear model in R with similar outcomes. Experiments were replicated at least three times (including behavioural experiments, which were performed on at least three different cohorts).

Sample size calculation: The sample size was predetermined using the tool provided here: http://biomath.info/power/ttest.htm. In addition, calculated sample size was compared against sizes known from previous experience to yield high power to detect specific effects for each experiment. Accordingly, in some of the tests we decided to use less animals (<30), since the generated numbers were ethically unacceptably high.

Randomization: For in vivo experiments, mice were chosen based on genotypes. Sex-matched animal pairs of control-mutant siblings from the same litters were compared to decrease variance due to age, environment and genetic background[87]. At least 5 litters were used in each behavioral test.

Life Sciences Reporting Summary: Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

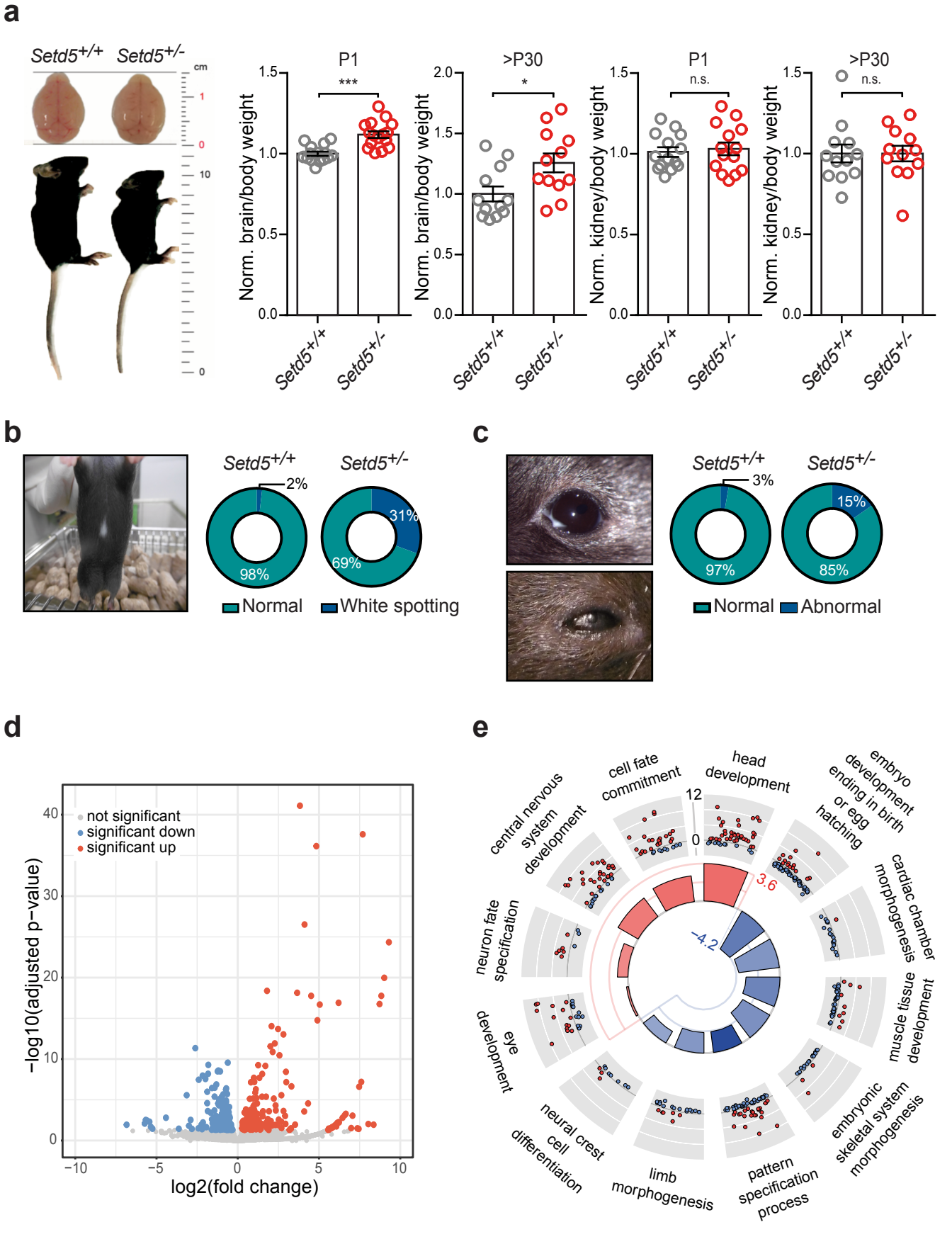
References

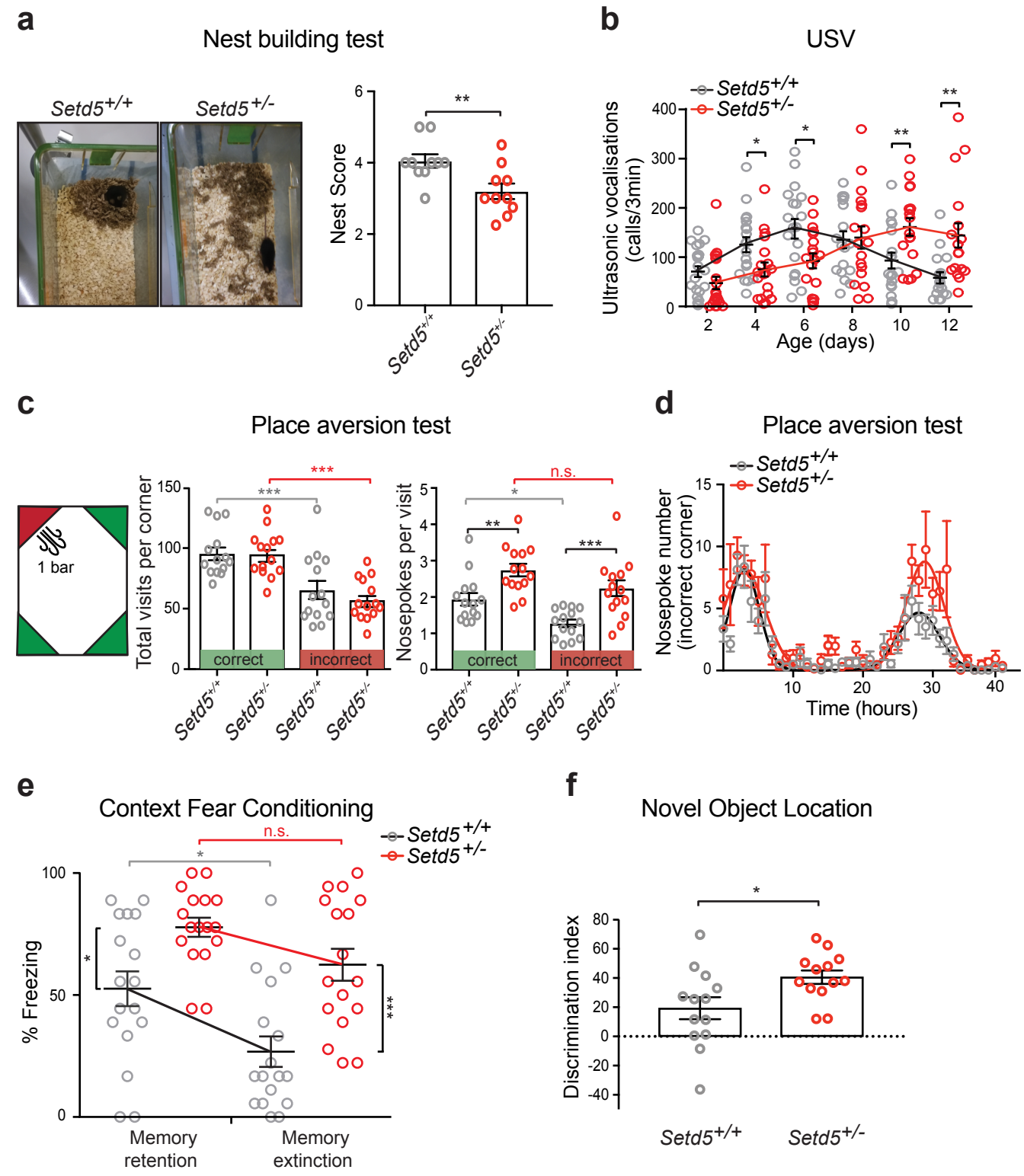
51 Osipovich, A. B., Gangula, R., Vianna, P. G. & Magnuson, M. A. Setd5 is essential for mammalian development and the co-transcriptional regulation of

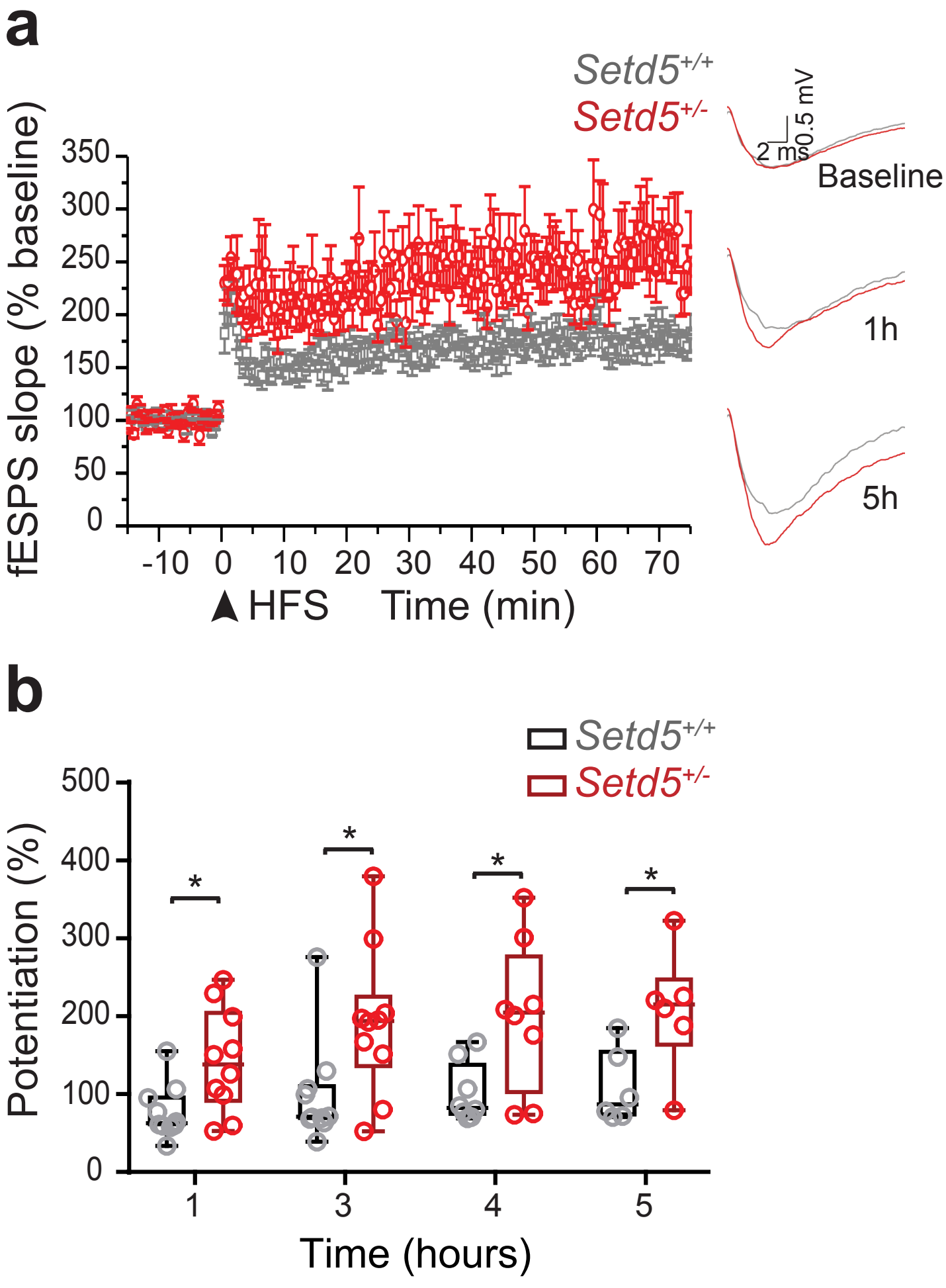
histone acetylation. *Development* **143**, 4595-4607, doi:10.1242/dev.141465 (2016).
52 Savic, D. *et al.* CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Res* **25**, 1581-1589, doi:10.1101/gr.193540.115 (2015).
53 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923
nmeth.1923 [pii] (2012).
54 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638
btu638 [pii] (2015).
55 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:s13059-014-0550-8 [pii]
10.1186/s13059-014-0550-8 (2014).
56 Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257-258, doi:btl567 [pii]
10.1093/bioinformatics/btl567 (2007).
57 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635
bts635 [pii] (2013).
58 Peixoto, L. *et al.* How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. *Nucleic acids research* **43**, 7664-7674, doi:10.1093/nar/gkv736 (2015).
59 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
60 Walter, W., Sanchez-Cabo, F. & Ricote, M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912-2914, doi:10.1093/bioinformatics/btv300 (2015).
61 Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812, doi:10.1093/bioinformatics/btu393 (2014).
62 Skene, N. G. & Grant, S. G. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Front Neurosci* **10**, 16, doi:10.3389/fnins.2016.00016 (2016).
63 Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142, doi:10.1126/science.aaa1934 (2015).
64 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).
65 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
66 Nott, A. *et al.* Histone deacetylase 3 associates with MeCP2 to regulate FOXO and social behavior. *Nat Neurosci* **19**, 1497-1505, doi:10.1038/nn.4347 (2016).

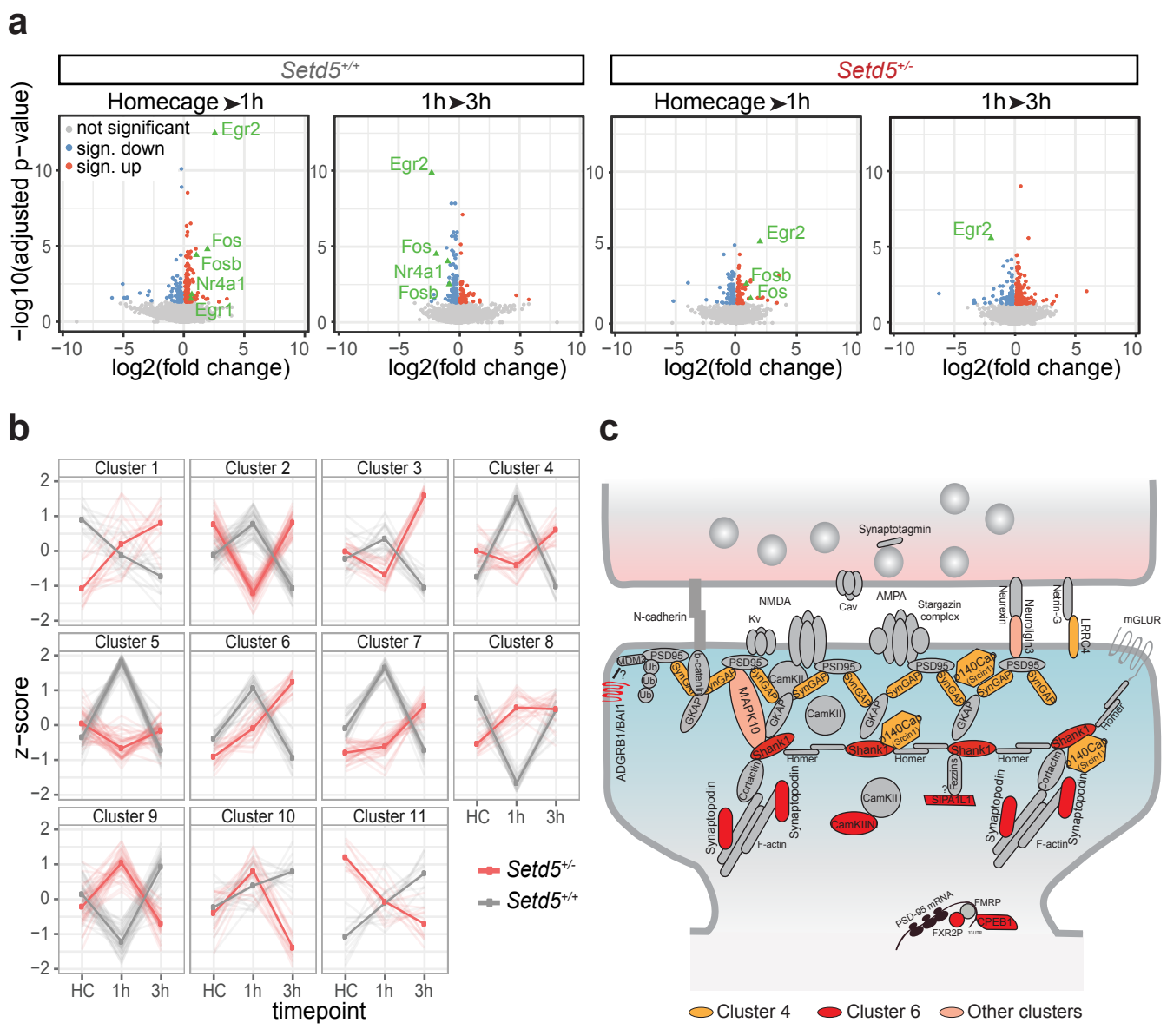
67 Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic acids research* **11**, 1475-1489 (1983).
68 Fraser, J. *et al.* Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol Syst Biol* **11**, 852, doi:10.15252/msb.20156492 (2015).
69 Liao, Z., Wan, Y., Thomas, S. N. & Yang, A. J. IsoQuant: a software tool for stable isotope labeling by amino acids in cell culture-based mass spectrometry quantitation. *Anal Chem* **84**, 4535-4543, doi:10.1021/ac300510t (2012).
70 Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319, doi:10.1038/nprot.2016.136
nprot.2016.136 [pii] (2016).
71 Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**, S96-104 (2002).
72 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007
gkv007 [pii] (2015).
73 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
74 Jensen, L. J. *et al.* STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**, D412-416, doi:10.1093/nar/gkn760
gkn760 [pii] (2009).
75 Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**, R86, doi:10.1186/gb-2010-11-8-r86
gb-2010-11-8-r86 [pii] (2010).
76 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191, doi:10.1093/nar/gku365
gku365 [pii] (2014).
77 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137
gb-2008-9-9-r137 [pii] (2008).
78 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11 12 11-34, doi:10.1002/0471250953.bi1112s47 (2014).
79 Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-120, doi:10.1038/nature11243 (2012).
80 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393, doi:10.1038/nature10730
nature10730 [pii] (2012).

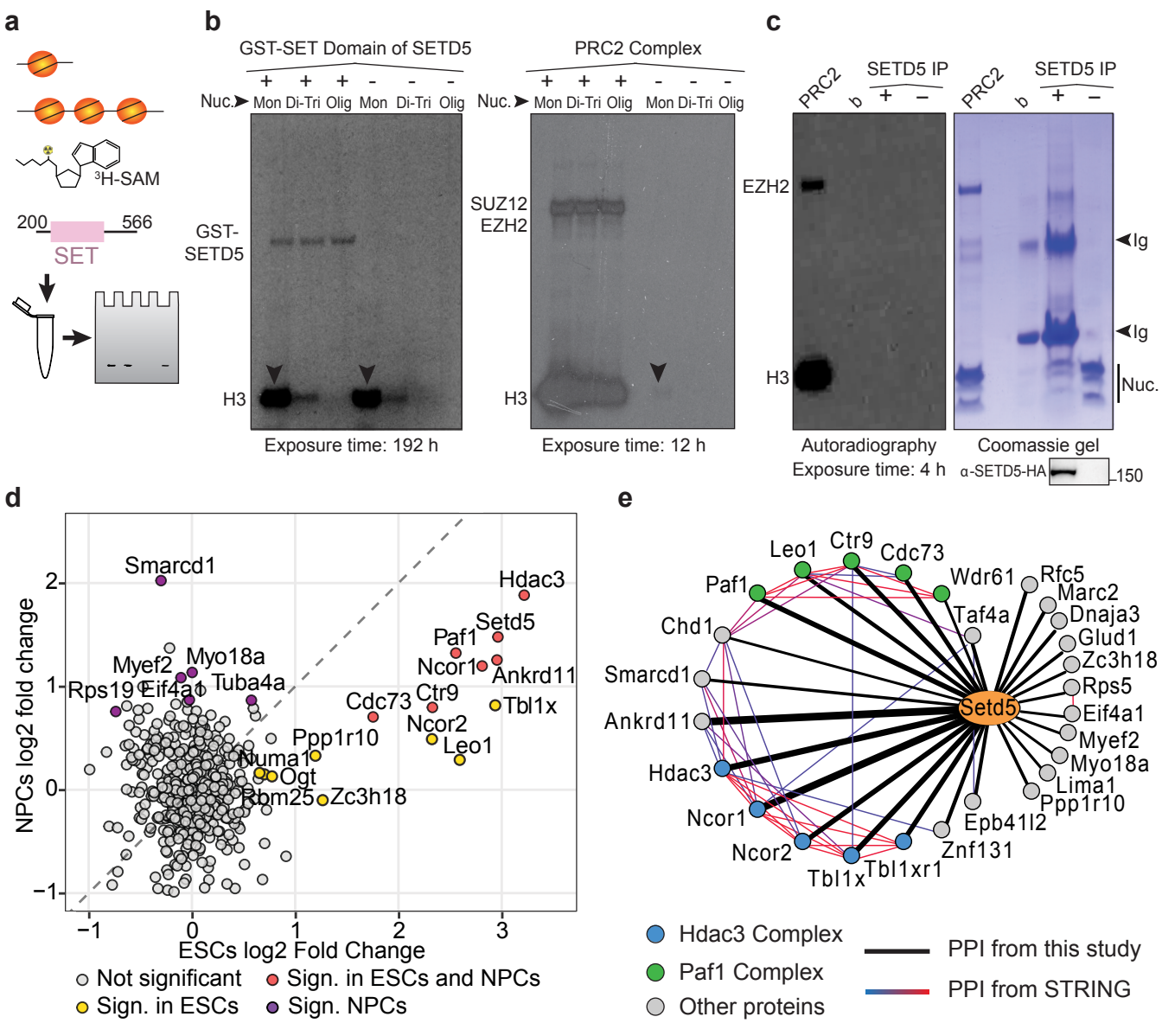
81 Chen, F. X. *et al.* PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell* **162**, 1003-1015, doi:10.1016/j.cell.2015.07.042 (2015).
82 De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-215, doi:10.1038/nature13772 nature13772 [pii] (2014).
83 Grozeva, D. *et al.* De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. *Am J Hum Genet* **94**, 618-624, doi:10.1016/j.ajhg.2014.03.006 S0002-9297(14)00108-6 [pii] (2014).
84 Farwell Hagman, K. D. *et al.* Candidate-gene criteria for clinical reporting: diagnostic exome sequencing identifies altered candidate genes among 8% of patients with undiagnosed diseases. *Genet Med* **19**, 224-235, doi:10.1038/gim.2016.95 gim201695 [pii] (2017).
85 Halvardson, J. *et al.* Mutations in HECW2 are associated with intellectual disability and epilepsy. *J Med Genet* **53**, 697-704, doi:10.1136/jmedgenet-2016-103814 jmedgenet-2016-103814 [pii] (2016).
86 Graff, J. *et al.* Epigenetic priming of memory updating during reconsolidation to attenuate remote fear memories. *Cell* **156**, 261-276, doi:10.1016/j.cell.2013.12.020 (2014).
87 Holmdahl, R. & Malissen, B. The need for littermate controls. *Eur J Immunol* **42**, 45-47, doi:10.1002/eji.201142048 (2012).

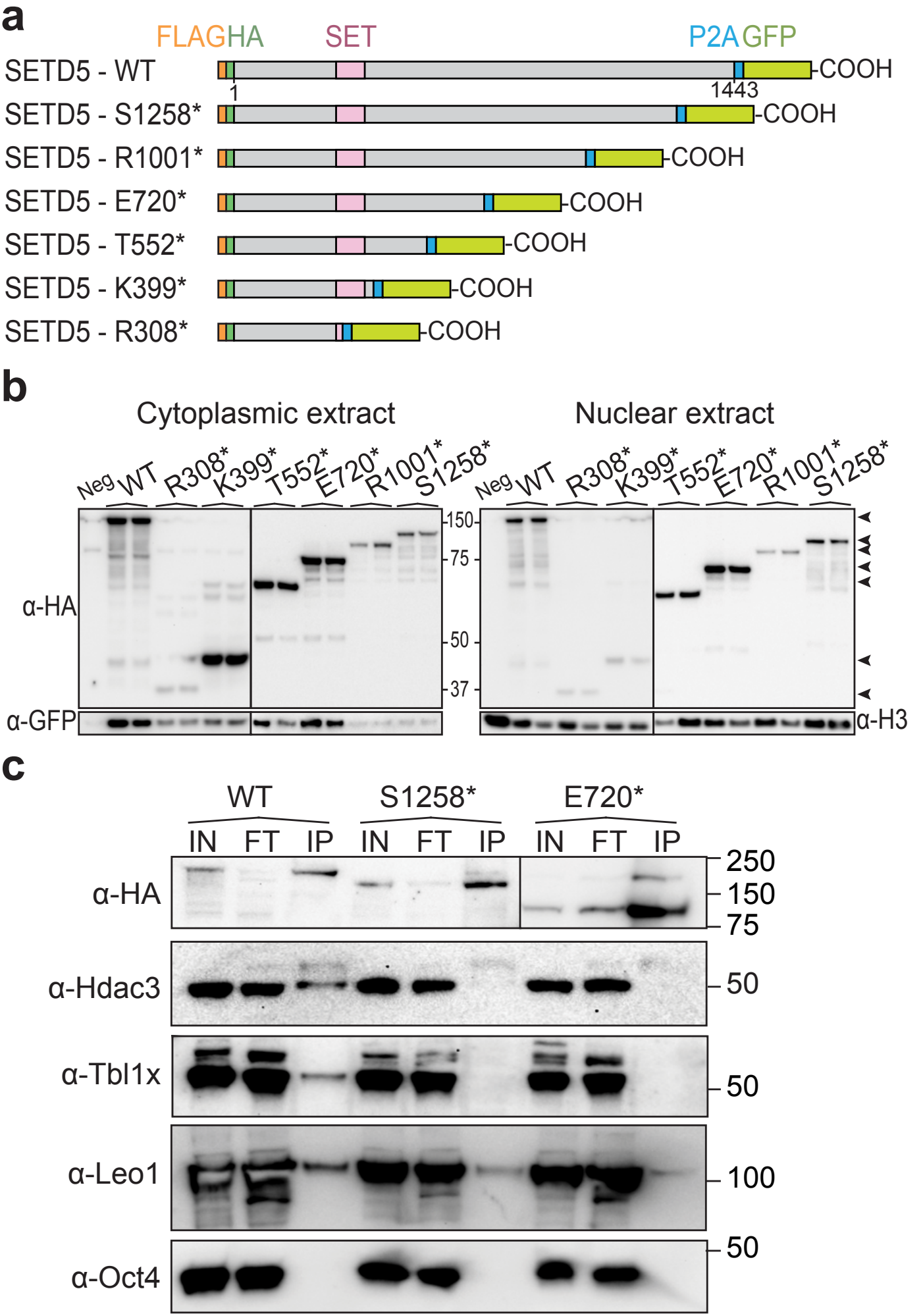

a
Setd5+/+
Setd5+/-
cm
1
0
10
0
P1
***
>P30
*
P1
n.s.
>P30
n.s.
Norm. brain/body weight
Norm. kidney/body weight
1.5
1.0
0.5
0.0
2.0
Setd5+/+
Setd5+/-
b
Setd5+/+
Setd5+/-
2%
98%
31%
69%
Normal
White spotting
c
Setd5+/+
Setd5+/-
3%
97%
15%
85%
Normal
Abnormal
d
not significant
significant down
significant up
−log10(adjusted p−value)
log2(fold change)
40
30
20
10
0
−10
−5
0
5
10
e
head development
embryo development ending in birth or egg hatching
cardiac chamber morphogenesis
muscle tissue development
embryonic skeletal system morphogenesis
pattern specification process
limb morphogenesis
neural crest cell differentiation
eye development
neuron fate specification
central nervous system development
cell fate commitment
12
0
3.6
−4.2

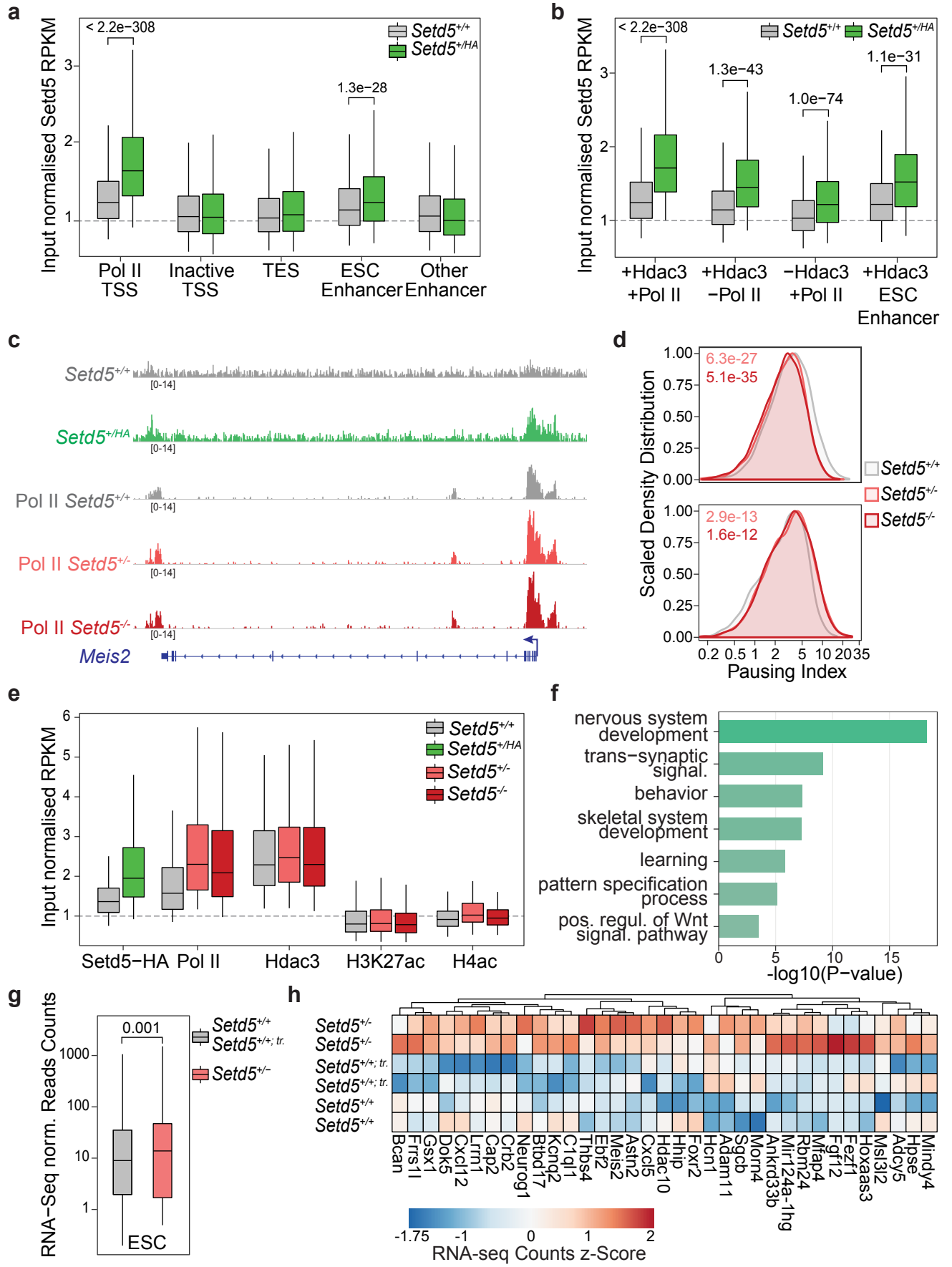

a
Nest building test
Setd5+/+
Setd5+/-
Nest Score
**
Setd5+/+
Setd5+/-
b
USV
Setd5+/+
Setd5+/-
Ultrasonic vocalisations (calls/3min)
Age (days)
c
Place aversion test
1 bar
Total visits per corner
Nosepokes per visit
correct
incorrect
n.s.
d
Place aversion test
Nosepoke number (incorrect corner)
Time (hours)
e
Context Fear Conditioning
% Freezing
Memory retention
Memory extinction
f
Novel Object Location
Discrimination index


a
Setd5+/+
Setd5+/-
fESPS slope (% baseline)
350
300
250
200
150
100
50
0
-10
0
10
20
30
40
50
60
70
HFS
Time (min)
2 ms
0.5 mV
Baseline
1h
5h
b
Setd5+/+
Setd5+/-
Potentiation (%)
500
400
300
200
100
0
*
*
*
*
1
3
4
5
Time (hours)


a
Setd5+/+
Setd5+/-
Homecage ➤1h
1h➤3h
not significant
sign. down
sign. up
Egr2
Fos
Fosb
Nr4a1
Egr1
Nr4a1
−log10(adjusted p−value)
log2(fold change)
b
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
z−score
HC
1h
3h
timepoint
Setd5+/-
Setd5+/+
c
Synaptotagmin
N-cadherin
NMDA
Kv
Cav
AMPA
Stargazin complex
Neurexin
Neuroligin3
Netrin-G
LRR64
mGLUR
ADGRB1/BAI1
MDM2
Ub
PSD95
δ-catenin
SynGAP
GKAP
MAPK10
CamKII
Shank1
Homer
14-3-3ζ/δ
Cortactin
Synaptopodin
F-actin
CamKIIN
SIPA1L1
Rezrins
PSD-95 mRNA
FMRP
CPEB1
FXR2P
Cluster 4
Cluster 6
Other clusters


a
3H-SAM
200
566
SET
b
GST-SET Domain of SETD5
PRC2 Complex
Nuc.
Mon Di-Tri Olig Mon Di-Tri Olig
Mon Di-Tri Olig Mon Di-Tri Olig
GST-SETD5
SUZ12
EZH2
H3
H3
Exposure time: 192 h
Exposure time: 12 h
c
PRC2
SETD5 IP
b
EZH2
H3
Ig
Ig
Nuc.
Autoradiography
Exposure time: 4 h
Coomassie gel
α-SETD5-HA
150
d
Smarcd1
Hdac3
Setd5
Paf1
Ncor1
Ankrd11
Myef2
Myo18a
Rps19
Eif4a1
Tuba4a
Cdc73
Ctr9
Tbl1x
Ncor2
Ppp1r10
Leo1
Numa1
Ogt
Rbm25
Zc3h18
NPCs log2 fold change
ESCs log2 Fold Change
Not significant
Sign. in ESCs and NPCs
Sign. in ESCs
Sign. NPCs
e
Leo1
Ctr9
Cdc73
Paf1
Wdr61
Rfc5
Marc2
Taf4a
Dnaja3
Chd1
Glud1
Smarcd1
Zc3h18
Rps5
Setd5
Ankrd11
Eif4a1
Myef2
Hdac3
Myo18a
Lima1
Ppp1r10
Ncor1
Epb41l2
Ncor2
Znf131
Tbl1x
Tbl1xr1
Hdac3 Complex
Paf1 Complex
Other proteins
PPI from this study
PPI from STRING


a
FLAG
HA
SET
P2A
GFP
SETD5 - WT
SETD5 - S1258*
SETD5 - R1001*
SETD5 - E720*
SETD5 - T552*
SETD5 - K399*
SETD5 - R308*
-COOH
1
1443
b
Cytoplasmic extract
Nuclear extract
Neg
WT
R308*
K399*
T552*
E720*
R1001*
S1258*
150
75
50
37
α-HA
α-GFP
α-H3
c
WT
S1258*
E720*
IN
FT
IP
250
150
75
50
100
α-HA
α-Hdac3
α-Tbl1x
α-Leo1
α-Oct4


a
Input normalised Setd5 RPKM
< 2.2e−308
1.3e−28
Setd5+/+
Setd5+/HA
Pol II TSS
Inactive TSS
TES
ESC Enhancer
Other Enhancer
b
Input normalised Setd5 RPKM
< 2.2e−308
1.3e−43
1.0e−74
1.1e−31
+Hdac3 +Pol II
+Hdac3 −Pol II
−Hdac3 +Pol II
+Hdac3 ESC Enhancer
c
Setd5+/+
Setd5+/HA
Pol II Setd5+/+
Pol II Setd5+/−
Pol II Setd5−/−
[0-14]
Meis2
d
Scaled Density Distribution
6.3e-27
5.1e-35
2.9e-13
1.6e-12
Setd5+/+
Setd5+/−
Setd5−/−
Pausing Index
e
Input normalised RPKM
Setd5+/+
Setd5+/HA
Setd5+/−
Setd5−/−
Setd5−HA
Pol II
Hdac3
H3K27ac
H4ac
f
nervous system development
trans−synaptic signal.
behavior
skeletal system development
learning
pattern specification process
pos. regul. of Wnt signal. pathway
-log10(P−value)
g
RNA−Seq norm. Reads Counts
0.001
Setd5+/+
Setd5+/+; tr.
Setd5+/−
ESC
h
Setd5+/−
Setd5+/−
Setd5+/+; tr.
Setd5+/+; tr.
Setd5+/+
Setd5+/+
Bcan
Frrs1l
Gsx1
Dok5
Cxcl12
Lrrn1
Cap2
Crb2
Neurog1
Btbd17
Kcnq2
C1ql1
Thbs4
Ebf2
Meis2
Astn2
Cxcl5
Hdac10
Hhip
Foxr2
Hcn1
Adam11
Sgcb
Mom4
Ankrd33b
Mir124a-1hg
Rbm24
Mtap4
Fgf12
Fezf1
Hoxaas3
Msl3l2
Adcy5
Hpse
Mindy4
-1.75
-1
0
1
2
RNA-seq Counts z-Score