

## Regulation of CHD2 expression by the *Chaserr* long noncoding RNA is essential for viability

Aviv Rom<sup>1</sup>, Liliya Melamed<sup>1</sup>, Micah Jonathan Goldrich<sup>1</sup>, Rotem Kadir<sup>2</sup>, Matan Golan<sup>3</sup>, Inbal Biton<sup>4</sup>, Rotem Ben-Tov Perry<sup>1</sup>, Igor Ulitsky\*

<sup>1</sup> Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel

<sup>2</sup> National Institute for Biotechnology in the Negev and Department of Microbiology, Immunology and Genetics, Ben-Gurion University of the Negev, Beer-Sheva, Israel

<sup>3</sup> Department of Molecular Cell Biology, <sup>4</sup> Department of Veterinary Resources, Weizmann Institute of Science, Rehovot, Israel

\* - Corresponding author, [igor.ulitsky@weizmann.ac.il](mailto:igor.ulitsky@weizmann.ac.il)

### Abstract

Genomic loci adjacent to genes encoding for transcription factors and chromatin remodelers are enriched for long non-coding RNAs (lncRNAs), but the functional importance of this enrichment is largely unclear. Chromodomain helicase DNA binding protein 2 (*Chd2*) is a chromatin remodeller with various reported functions in cell differentiation and DNA damage response. Heterozygous mutations in human *CHD2* have been implicated in epilepsy, neurodevelopmental delay, and intellectual disability. Here we show that *Chaserr*, a highly conserved lncRNA transcribed from a region near the transcription start site of *Chd2* and on the same strand, acts in concert with the CHD2 protein to maintain proper *Chd2* expression levels. Loss of *Chaserr* in mice leads to early postnatal lethality in homozygous mice, and severe growth retardation in heterozygotes. Mechanistically, loss of *Chaserr* leads to substantially increased *Chd2* mRNA and protein levels, which in turn lead to increased transcriptional interference by inhibiting promoters found downstream of highly expressed genes. We further show that *Chaserr* production represses *Chd2* expression solely in *cis*, and that the phenotypic consequences of *Chaserr* loss are rescued when *Chd2* is perturbed as well. Targeting *Chaserr* is thus a potentially viable strategy for increasing CHD2 levels in haploinsufficient individuals.

## Introduction

The mammalian transcriptome is highly complex, and contains tens of thousands of non-coding RNA genes<sup>1-4</sup>. A significant subset of these, referred to as long noncoding RNAs (lncRNAs), are at least 200 nucleotides (nt) in length, 3' polyadenylated, and 5' capped, and are therefore structurally similar to mRNAs, but lack protein coding potential<sup>5</sup>. Only a small portion of lncRNAs have been functionally characterized, and only very few of these have been studied in the context of organismal development<sup>6</sup>. Multiple lines of evidence point to a strong link between lncRNA functions and those of chromatin modifying complexes<sup>7,8</sup>. Numerous chromatin modifiers have been reported to interact with lncRNAs<sup>8</sup>. In addition, lncRNAs in vertebrate genomes are enriched in the vicinity of genes that encode for transcription-related factors<sup>2,9</sup>, including numerous chromatin-associated proteins, but the functions of the vast majority of these lncRNAs remain unknown.

We hypothesized that the proximity of some lncRNA genes to genes involved in chromatin biology may imply a functional connection. To explore the biology of such interactions, we focused on one of the most conserved lncRNAs in vertebrates, found in close proximity to *Chromodomain Helicase DNA Binding Protein 2 (Chd2)*. CHD2 is an ATP-dependent chromatin-remodeling enzyme, which together with CHD1 belongs to subfamily I of the chromodomain helicase DNA-binding (CHD) protein family. Members of this subfamily are characterized by two chromodomains located in the N-terminal region and a centrally located SNF2-like ATPase domain<sup>10</sup>, and facilitate disassembly, eviction, sliding, and spacing of nucleosomes<sup>11</sup>. There are conflicting reports on the genomic occupancy of CHD2. According to one report, based on ChIP-seq data obtained with antibodies against CHD1 and CHD2, both proteins bind predominantly in the proximity of gene promoters and share up to 60% of their DNA binding sites in human cell lines<sup>12</sup>. Another study used MNase-ChIP-seq of endogenously tagged *Chd1* and *Chd2* in mouse embryonic stem cells (mESCs)<sup>13</sup>, and reported that the two proteins have different binding patterns – CHD1 binds predominantly to promoter regions, whereas CHD2 is associated with gene bodies of actively transcribed genes. CHD2 has also been linked to the deposition and incorporation of the H3.3 histone variant at transcriptionally activated genes<sup>12,14,15</sup> and at DNA damage foci, with the latter activity promoting repair of double strand breaks<sup>16</sup>.

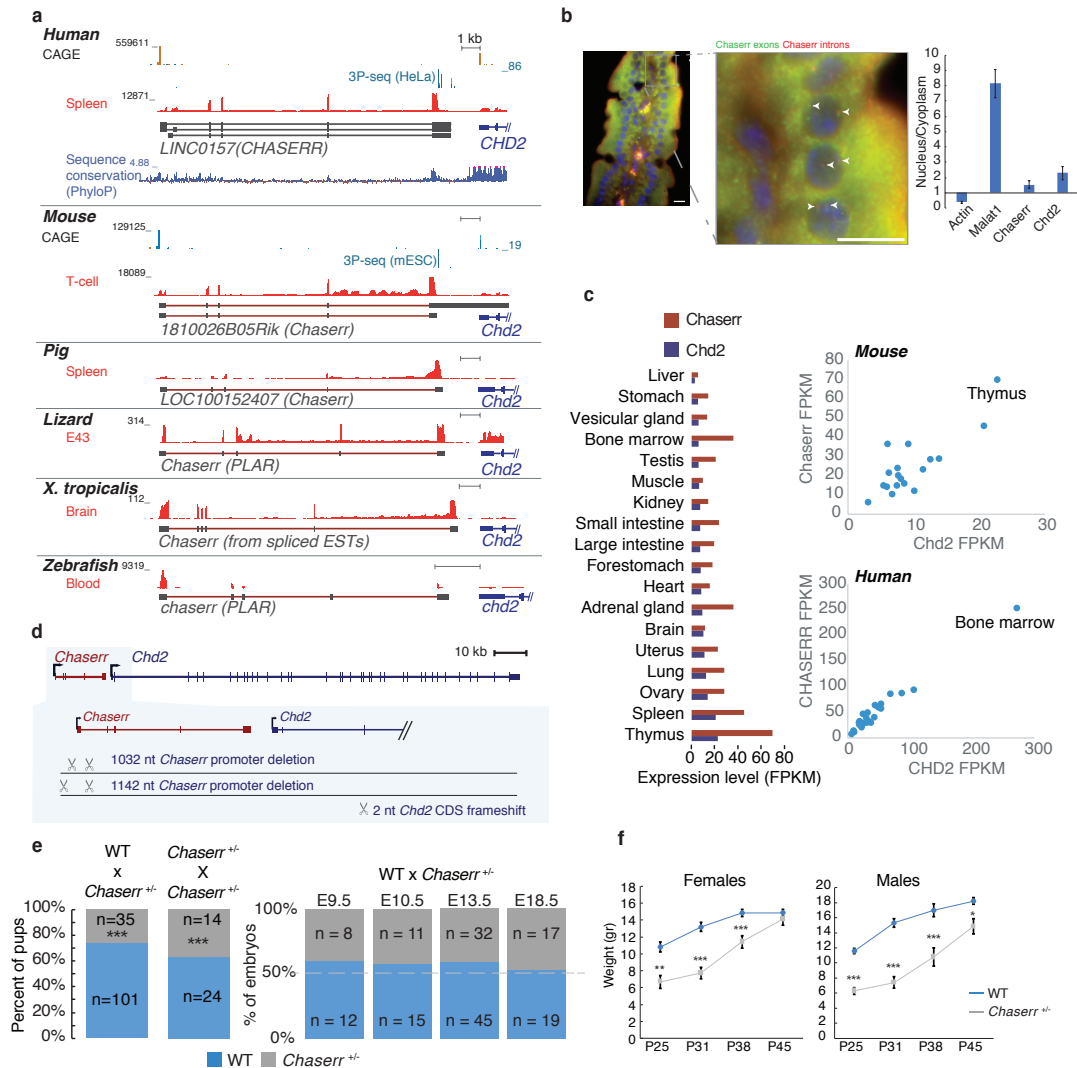
Mice homozygous for a gene-trap stop cassette in intron 27 of *Chd2* survive until E18.5 with a marked growth retardation, and no viable offspring of these mice can be recovered<sup>17</sup>. Heterozygotes with this mutation show increased mortality at postnatal days 1–4, and in the long-term they exhibit growth retardation, shorter life spans, and altered morphology in various organs. However, a dominant negative effect of the truncated protein could not be

excluded in this model. A different model for *Chd2* loss-of-function was recently created by the International Mouse Phenotyping Consortium, where exon 3 was replaced by a lacZ cassette and a stop signal<sup>18</sup>. No significant changes in mortality and aging were reported for these mice, but they exhibit slightly decreased body weight and length, skeletal abnormalities, an abnormal bone structure, decreased fat amount and bone mineral density, and abnormalities in blood composition, such as decreased erythrocyte cell number, hemoglobin content, and mean platelet volume (<http://www.mousephenotype.org/>).

In humans, *CHD2* haploinsufficiency is associated with neurodevelopmental delay, intellectual disability, epilepsy, and behavioral problems (reviewed in<sup>19</sup>). Studies in mouse models and cell lines also implicate *Chd2* in neuronal dysfunction: perturbations of *Chd2* affect neurogenesis in the mouse developing cerebral cortex<sup>20</sup> and in human stem cells differentiated to neurons<sup>21</sup>, and loss of a single *Chd2* copy leads to deficits in neuron proliferation and a shift in neuronal excitability<sup>22</sup>. Approaches for increasing CHD2 levels may thus have therapeutic relevance.

## Results

LncRNAs are spread throughout vertebrate genomes, but have a notable enrichment in proximity to chromatin- and transcription-related genes<sup>2,9</sup>. One such lncRNA, annotated as *1810026B05Rik* in mouse (which we denote as *Chaserr*, for *CHD2* adjacent, suppressive regulatory RNA) and *LINC01578/LOC100507217* in human (*CHASERR*), is an almost completely uncharacterized lncRNA, found upstream of and transcribed from the same strand as *Chd2* (**Fig. 1a**). *Chaserr* has five exons, is polyadenylated, and is a bona fide lncRNA according to PhyloCSF<sup>23</sup> (**Extended Data Fig. 1a**), CPAT<sup>24</sup> (coding probability 0.296), and CPC<sup>25</sup> (coding potential score -1.23). According to FANTOM5 transcription start site (TSS) annotations, 3P-seq poly(A) site mapping, and RNA-seq data, *Chaserr* transcript is independent of *Chd2* (**Fig. 1a**). The ‘tandem’ organization with *Chd2*, *Chaserr* exon-intron structure, and parts of *Chaserr* sequence are conserved throughout vertebrates (**Fig. 1a**), which makes it one of the most conserved mammalian lncRNAs<sup>9,26</sup>. According to RefSeq annotation, the last exon of *Chaserr* in mouse overlaps *Chd2*; however, according to RNA-seq and 3P-seq data from various tissues, the predominant *Chaserr* isoform ends ~500 bp after its last 3' splice site, ~2.2 Kb upstream of the *Chd2* TSS (as in GENCODE transcript ENSMUST00000184554), and we therefore considered this isoform in further studies. The RNA product of *Chaserr* is localized in the nucleus, in proximity to the *Chd2* site of transcription (**Fig. 1b**). This nuclear enrichment is due at least in part to nonsense mediated decay (NMD) that acts on *Chaserr*, likely triggered by a 117 codon non-conserved ORF that ends in the 2<sup>nd</sup> exon (**Extended Data Fig. 1b-c**).



**Figure 1 | *Chaserr* is essential for postnatal viability.** **a**, Genomic locus of *Chaserr* in the indicated vertebrate genomes. CAGE read coverage is from the FANTOM5 project<sup>64</sup>. 3P-seq data are from<sup>65</sup>. RNA-seq data are from HPA<sup>66</sup>, ENCODE (mouse), FAANG (pig), SRP009831 (lizard), SRP039546 (xenopus), SRP024369 (zebrafish). **b**, Left: Single molecule FISH with probes targeting *Chaserr* introns (red) and exons (green) in mouse duodenum. Co-localization of introns and exons signal marked by white arrowhead. Scale bar shows 10 $\mu$ m. Right: qRT-PCR comparing cytoplasm and nuclear fractions in mESCs. **c**, Left: *Chaserr* and *Chd2* RNA expression across different tissues profiled in the mouse BodyMap<sup>67</sup>. Right: Scatter plots for *Chaserr* and *Chd2* expression in mouse BodyMap and human HPA<sup>66</sup> projects. **d**, Schematic of the positions of regions targeted by gRNAs used to generate *Chaserr*<sup>-/-</sup> and *Chd2*<sup>m/m</sup> mice. **e**, Observed rates of *Chaserr*<sup>+/-</sup> and *Chaserr*<sup>-/-</sup> neonates. **f**, Weights of neonates at the indicated age (males n=7–15, females n=6–18 per group). Error bars show S.E.M. \*P<0.05 \*\*P<0.01 and \*\*\*P<0.001 (two-sided t-test).

*Chaserr* and *Chd2* mRNA are tightly co-expressed across a panel of human and mouse tissues, and during mouse development according to ENCODE and FANTOM5 data (Fig. 1c and Extended Data Fig. 1d-e). Interestingly, both genes are expressed at appreciable levels in all studied samples, with particularly high expression in lymphocytes. The ratio between the expression levels of the two genes is also similar across the tissues, with notable

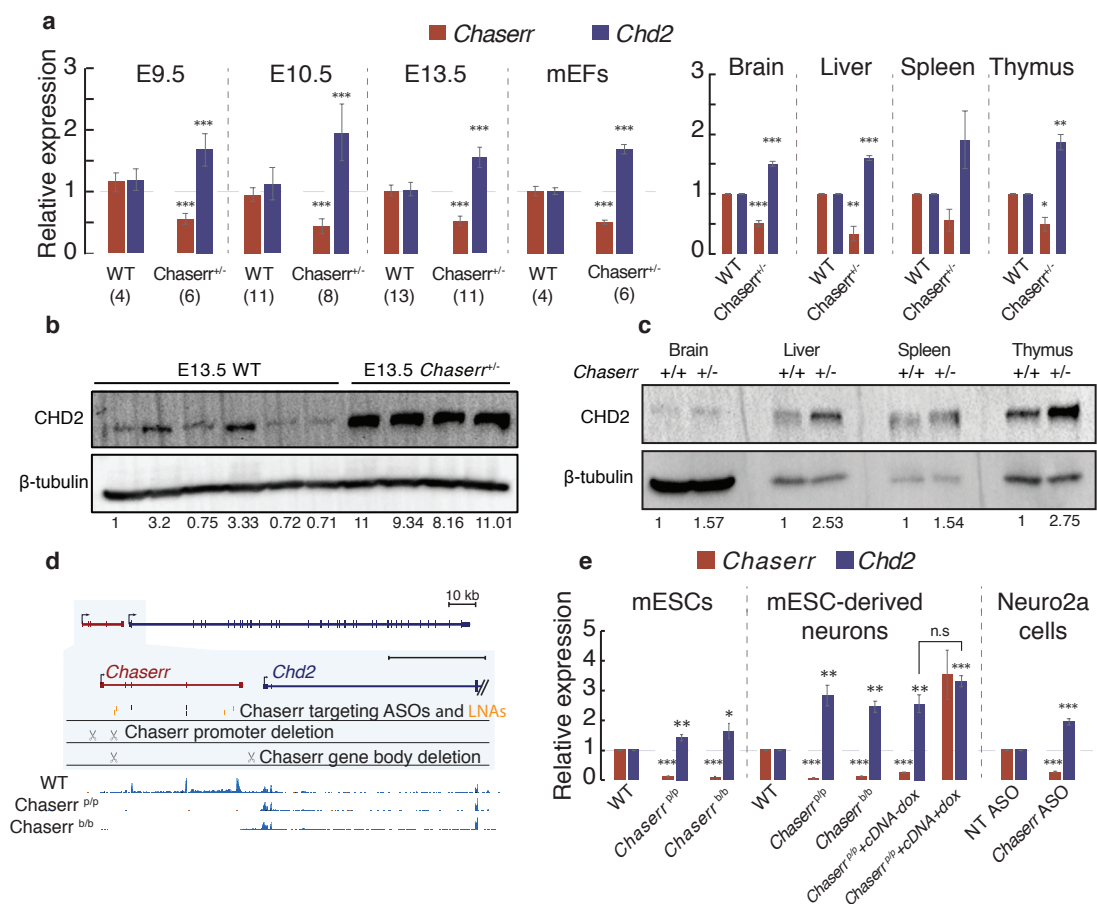
exceptions of neuronal cells and fibroblasts, where *Chaserr* levels are relatively high (**Fig. 1c and Extended Data Fig. 1d-e**).

In order to study its function, we generated *Chaserr* null alleles in mice by injection of CRISPR/Cas9 with a pair of gRNAs targeting sequences flanking the promoter and the first exon of *Chaserr* (**Fig. 1d**). This resulted in two different *Chaserr*<sup>+/-</sup> mouse strains that carry deletions of 1032 and 1142 bps (**Extended Data Fig. 1f**), which were sufficient to abolish expression of mature *Chaserr* (see below), and the two lines were phenotypically indistinguishable from each other, and so we used them interchangeably in subsequent experiments. Strikingly, out of 38 pups born following crosses between *Chaserr*<sup>+/-</sup> mice, we observed no *Chaserr*<sup>-/-</sup> pups. The numbers of *Chaserr*<sup>+/-</sup> pups at weaning also deviated from expected Mendelian ratios (~37%,  $P < 10^{-8}$ , **Fig. 1e**), and the surviving mice exhibited substantial growth retardation, occasional malocclusion, and neonatal lethality (**Fig. 1f and Extended Data Fig. 1g-i**). Pathological analysis of *Chaserr*<sup>+/-</sup> mice revealed a wide range of abnormalities, including fat depletion, kyphosis, and thymic depletion, neither of which was highly penetrant. *Chaserr*<sup>+/-</sup> females very rarely became pregnant, therefore we focused on crosses between *Chaserr*<sup>+/+</sup> females and *Chaserr*<sup>+/-</sup> males for most of this study. Out of 136 pups born from such crosses, 35 were *Chaserr*<sup>+/-</sup> (~25%), significantly less than the expected 50% ( $P < 10^{-5}$ , **Fig. 1e**), suggesting two copies of *Chaserr* are required for proper postnatal survival. In contrast, embryos at four different developmental stages up to E18.5 were recovered with the expected ~50% *Chaserr*<sup>+/-</sup> vs. *Chaserr*<sup>+/+</sup> ratios, suggesting that a single copy of *Chaserr* is sufficient for viable embryonic development (**Fig. 1e**).

Due to the close proximity and co-expression of *Chaserr* and *Chd2*, we tested whether *Chd2* expression is affected in *Chaserr*<sup>+/-</sup> embryos. *Chd2* mRNA was significantly upregulated by ~1.5-fold during embryonic development at the examined time points (E9.5, E10.5, and E13.5), in mouse embryonic fibroblasts (mEFs) derived from E13.5 embryos, and in four adult tissues (**Fig. 2a**). Western blot (WB) in E13.5 embryos and in the adult tissues also demonstrated strong up-regulation of CHD2 protein (**Fig. 2b-c**).

As *Chaserr*<sup>-/-</sup> mice could not be generated, in order to study regulation of *Chd2* in *Chaserr*<sup>+/-</sup> cells and efficiently compare different *Chaserr* perturbations we used CRISPR/Cas9 in mouse embryonic stem cell (mESCs) to engineer clones with either homozygous loss of *Chaserr* promoter and first exon (*Chaserr*<sup>ΔP</sup>), or deletion of the rest of the *Chaserr* gene body from the first intron to just downstream of *Chaserr* polyadenylation site (*Chaserr*<sup>b/b</sup>) (**Fig. 2d**). Importantly, the *Chaserr*<sup>b/b</sup> line has an intact *Chaserr* promoter, and thus helps address the possibility of a competition between the *Chaserr* and *Chd2* promoters. *Chaserr* expression was abolished in both mutant lines, as evident in RNA-seq data (**Fig. 2d**), leading to significant *Chd2* mRNA upregulation (**Fig. 2e**). This up-regulation is consistent with a

previous study that examined *Chaserr* as part of a panel of lncRNAs (*Chaserr* was referred to as ‘linc2025’ in that study) and also observed an increase in *Chd2* upon deletion of *Chaserr* promoter<sup>27</sup>. *Chd2* was also significantly upregulated in neurons derived from *Chaserr*<sup>p/p</sup> and *Chaserr*<sup>b/b</sup> mESCs (Fig. 2e). Similarly, targeting of *Chaserr* in Neuro2a cells using antisense oligonucleotides (ASOs) or LNA Gapmers reduced *Chaserr* expression by ~75%, and led to a significant increase in *Chd2* mRNA and protein levels (Fig. 2e and Extended Data Fig. 2a-b). We note that other perturbation methods, like insertion of polyA sites or ribozymes<sup>28</sup> are not effective for perturbing *Chaserr* (<sup>27</sup> and data not shown). We conclude that an intact *Chaserr* represses *Chd2* expression in all the systems we studied.



**Figure 2 | *Chaserr* represses *Chd2* expression.** **a**, qPCR of the indicated genes in whole embryos from the indicated developmental stage, mEFs, or adult tissues with the indicated background. n for each group of embryos or mEFs is indicated in parentheses. Normalized to *Actb*. n=3 for the tissues. **b**, Western blot for the indicated protein in individual E13.5 whole embryos from indicated background. **c**, as in b for the adult tissues. **d**, Top: Scheme of regions targeted by gRNAs used to generate *Chaserr*<sup>p/p</sup>, *Chaserr*<sup>b/b</sup> mESCs, or the antisense reagents; Bottom: RNA-seq read coverage in mESCs from the indicated background. **e**, qRT-PCR of the indicated genes in the indicated backgrounds or treatments. Normalized to *Actb*. n≥3. Error bars show S.E.M. \*P<0.05 \*\*P<0.01 and \*\*\*P<0.001 (two-sided t-test).

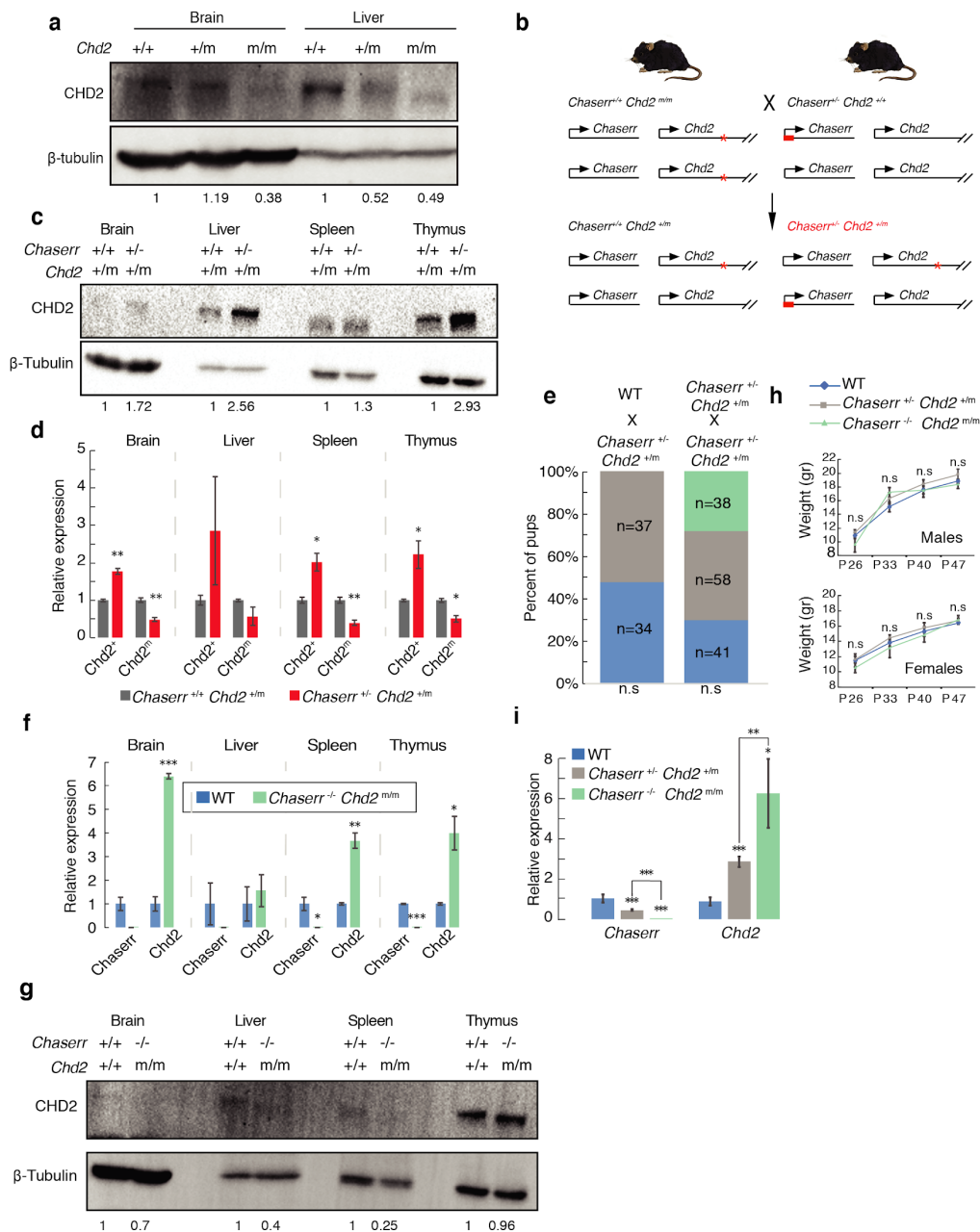
In order to test whether the mature *Chaserr* RNA is sufficient for repression of *Chd2*, we infected the *Chaserr*<sup>p/p</sup> mESCs with a lentivirus carrying a doxycycline (dox)-inducible *Chaserr* cDNA. *Chaserr* was overexpressed relative to the WT levels by ~3.5-fold following Dox addition, but *Chd2* mRNA expression was not affected compared to no Dox control (**Fig. 2e**). We conclude that either the RNA product of *Chaserr*, or transcription from its endogenous locus (which might be affected by the ASO/Gapmer cleavage of the nascent transcript), are required for repression of *Chd2* mRNA.

To evaluate if the broader genomic context is important for *Chaserr* function, we reconstituted the *Chaserr* locus on a plasmid, which contains a fragment of the mouse genome from the *Chaserr* promoter to the start codon of *Chd2*, excluding *Chaserr* introns, cloned immediately upstream of a firefly luciferase coding sequence (**Extended Data Fig. 2c**). Deletion of *Chaserr* promoter in this system led to a ~50% increase in luciferase expression (**Extended Data Fig. 2c**), consistent with an increase we observed in the endogenous setting. This suggests that the information required for at least part of *Chaserr* activity is encoded in the tested locus.

We next examined whether the early lethality of mice with *Chaserr* loss-of-function is mediated by increased *Chd2* levels. We used CRISPR/Cas9 with a single gRNA targeted to the 3<sup>rd</sup> exon of *Chd2* (**Fig. 1d** and **Extended Data Fig. 2d**) to generate *Chd2*<sup>m/m</sup> mice that had a two bp deletion in the coding sequence of *Chd2*. This mutation had a mild effect on *Chd2* mRNA expression, but dramatically reduced CHD2 protein levels, as expected from a frameshift in the main coding frame, with the residual protein possibly emanating from an alternative start codon (**Fig. 3a**, note that the monoclonal antibody we used recognizes a peptide in the C-terminal part of CHD2). *Chd2*<sup>m/m</sup> mice were born at expected Mendelian ratios (**Extended Data Fig. 2e**) and had no gross developmental phenotypes, similarly to the *Chd2*<sup>tm1b(EUCOMM)Hmgu</sup> mice generated by insertion of a LacZ cassette into the 2<sup>nd</sup> intron and deletion of 3<sup>rd</sup> exon of *Chd2*<sup>29</sup>. This enabled us to breed *Chaserr*<sup>+/-</sup> mice with *Chd2*<sup>m/m</sup> mice (**Fig. 3b**). We intercrossed the resulting *Chaserr*<sup>+/-</sup> *Chd2*<sup>m/+</sup> offspring, and out of 44 pups born from such crosses, only 14 (~32%) were *Chaserr*<sup>+/-</sup> (**Extended Data Fig. 2f**), suggesting that one hypomorphic allele of *Chd2* is not sufficient for compensating for loss of one *Chaserr* allele. Indeed, CHD2 protein levels were substantially higher in *Chaserr*<sup>+/-</sup> *Chd2*<sup>m/+</sup> mice when compared to their *Chaserr*<sup>+/+</sup> *Chd2*<sup>m/+</sup> littermates (**Fig. 3c**), potentially explaining the neonatal lethality and growth retardation (**Extended Data Fig. 2g**).

The *Chaserr*<sup>+/-</sup> *Chd2*<sup>m/+</sup> offspring that survived allowed us to test with allele-specific qRT-PCR whether *Chaserr* affected *Chd2* expression from the *cis* allele *in vivo* (**Extended Data Fig. 2h**). Only *Chd2* from the *Chaserr*<sup>-</sup> allele was up-regulated in *Chaserr*<sup>+/-</sup> *Chd2*<sup>m/+</sup> mice when compared to *Chaserr*<sup>+/+</sup> *Chd2*<sup>m/+</sup> littermates (**Fig. 3d**), suggesting that *Chaserr* represses

*Chd2* through *cis*-acting regulation. Interestingly, the *Chd2* from the *Chaserr*<sup>+</sup> allele was repressed compared to WT levels, hinting at a possible feedback regulation resulting from excess CHD2 expression (see Discussion).



**Figure 3 | *Chaserr* regulates *Chd2* expression in *cis*.** **a**, CHD2 western blot in the indicated tissues from mice with the indicated genotypes. **b**, Scheme of the cross between *Chaserr*<sup>+/-</sup> and *Chd2*<sup>m/m</sup> mice. **c**, Western blot in the indicated tissues from mice with the indicated genotypes. **d**, *Chd2* allele-specific expression in the indicated backgrounds in four different mouse tissues. Normalized to *Actb*. n=3 for the indicated tissues. **e**, Neonate survival rates for the *Chaserr*<sup>+/-</sup> *Chd2*<sup>m/m</sup> intercross. **f**, qRT-PCR in the indicated adult tissues and backgrounds. Normalized to *Actb*. n=3 for the indicated tissues. **g**, CHD2 western blot from the indicated tissues and backgrounds. **h**, Pup weights for the *Chaserr*<sup>+/-</sup> *Chd2*<sup>m/m</sup> intercross. n=5–19 pups per genotype per time point. **i**, qRT-PCR for mEFs with the indicated background. Normalized to *Actb*. WT n=5, *Chaserr*<sup>+/-</sup> *Chd2*<sup>+/-</sup> n=20, *Chaserr*<sup>+/-</sup> *Chd2*<sup>m/m</sup> n=5. Error bars show S.E.M. \*P<0.05 \*\*P<0.01 and \*\*\*P<0.001 (two-sided t-test).

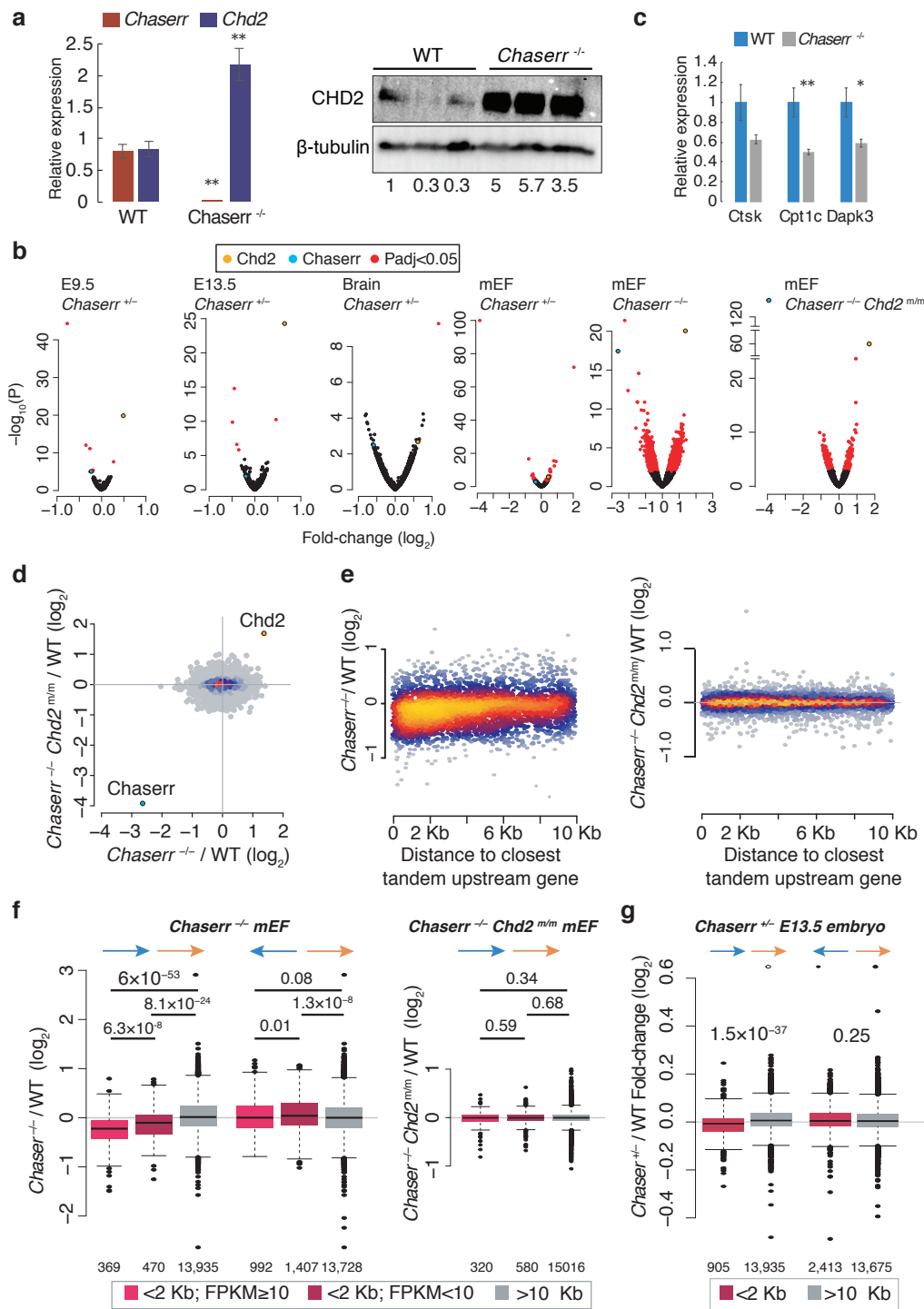


In order to directly test whether the severe *Chaserr* loss-of-function phenotype is mediated by CHD2 over-expression, we used CRISPR/Cas9 to delete the *Chaserr* promoter and first exon in a *Chd2<sup>m/m</sup>* mouse (**Extended Data Fig. 3a**), and thus generated a model in which loss of *Chaserr* increases the expression of a hypomorphic allele of *Chd2*. Out of 137 pups born from intercrosses of *Chaserr<sup>+/-</sup> Chd2<sup>+/m</sup>* mice, 38 were *Chaserr<sup>-/-</sup> Chd2<sup>m/m</sup>* (~27%), 58 *Chaserr<sup>+/-</sup> Chd2<sup>+/m</sup>* (~42%), and 41 WT (~30%), which did not deviate significantly from normal Mendelian ratios ( $P = 0.2$ ; **Fig. 3e** and **Extended Data Fig. 3b,c**), despite up to 6-fold up-regulation of *Chd2* mRNA in different tissues of *Chaserr<sup>-/-</sup> Chd2<sup>m/m</sup>* mice (**Fig. 3f**). Despite the mRNA up-regulation, CHD2 mutant protein levels in tissues from *Chaserr<sup>-/-</sup> Chd2<sup>m/m</sup>* mice were not higher than the WT protein levels in WT mice (**Fig. 3g**). *Chaserr<sup>-/-</sup> Chd2<sup>m/m</sup>* and *Chaserr<sup>+/-</sup> Chd2<sup>+/m</sup>* mice showed no significant differences in weight (**Fig. 3h** and **Extended Data Fig. 3d**), and no obvious phenotypes, in stark contrast to the common and pleiotropic phenotypes observed in *Chaserr<sup>+/-</sup>* neonates on *Chd2<sup>+/+</sup>* background. A hypomorphic allele of *Chd2* can thus rescue the lethality caused by loss of *Chaserr*, when the two occur on the same allele.

To further characterize the changes in levels of *Chd2* mRNA in the double mutants, we isolated mEFs from E13.5 embryos from *Chaserr<sup>+/+</sup>*, *Chaserr<sup>+/-</sup>*, and *Chaserr<sup>-/-</sup>* genotypes on *Chd2<sup>m/m</sup>* background. qRT-PCR analysis showed that *Chaserr* expression is completely abolished in *Chaserr<sup>-/-</sup> Chd2<sup>m/m</sup>* mice, and that *Chd2* mRNA was significantly overexpressed in a manner that was inversely dependent on *Chaserr* dosage (**Fig. 3i**).

Having established that the severe phenotype resulting from *Chaserr* loss is mediated by CHD2, we were next interested to understand the consequences on gene expression of *Chaserr* loss and CHD2 up-regulation. We isolated *Chaserr<sup>-/-</sup>* mEFs, which showed a >2-fold increase in *Chd2* mRNA and protein levels (**Fig. 4a**), and used RNA-seq to characterize gene expression in *Chaserr<sup>+/+</sup>* and *Chaserr<sup>+/-</sup>* E9.5 and E13.5 embryos, adult brains, as well as *Chaserr<sup>+/+</sup>*, *Chaserr<sup>+/-</sup>*, *Chaserr<sup>-/-</sup>*, and *Chaserr<sup>-/-</sup> Chd2<sup>m/m</sup>* mEFs (**Fig. 4b**). As expected from the difference in the severity of the phenotype, we observed limited differential expression in *Chaserr<sup>+/-</sup>* embryos or mEFs compared to WT, whereas substantial numbers of differentially expressed genes were found in *Chaserr<sup>-/-</sup>* mEFs. The 1,493 significantly down-regulated genes (down by at least 25%, adjusted  $P < 0.05$ ) were enriched for GO categories such as nervous and connective tissue development and transcriptional regulation, and were very significantly enriched for genes whose loss-of-function in mice is associated with phenotypes such as “decreased length of long bones”, “decreased body weight”, “respiratory distress”, and “premature death” (**Supplementary Table 1**), closely related to the phenotypes observed in our *Chaserr* loss-of-function model. Specifically, some of the most down-regulated genes,

such as *Ctsk*, *Cpt1c*, and *Dapk3*, which were also validated by qRT-PCR (Fig. 4c), are known to be required for proper embryonic development<sup>30</sup>.



**Figure 4 | Transcriptional dysregulation following loss of *Chaserr*.** **a**, qRT-PCR (left, WT n=4, *Chaserr*<sup>-/-</sup> n=6) and Western blot (right) of *Chd2* expression in mEFs from the indicated background. n=3. qRT-PCR normalized to *Actb*. Western blot shows mEFs from three independent embryos for each background. **b**, Volcano plots for the indicated differential expression comparison using DESeq2<sup>53</sup>. **c**, qRT-PCR of the indicated genes in WT and *Chaserr*<sup>-/-</sup> mEFs. Normalized to *Actb*. WT n=4, *Chaserr*<sup>-/-</sup> n=6. **d**, Changes in gene expression in *Chaserr*<sup>-/-</sup> and *Chaserr*<sup>-/-</sup> *Chd2*<sup>m/m</sup> mEFs. **e**, Change in expression in *Chaserr*<sup>-/-</sup>

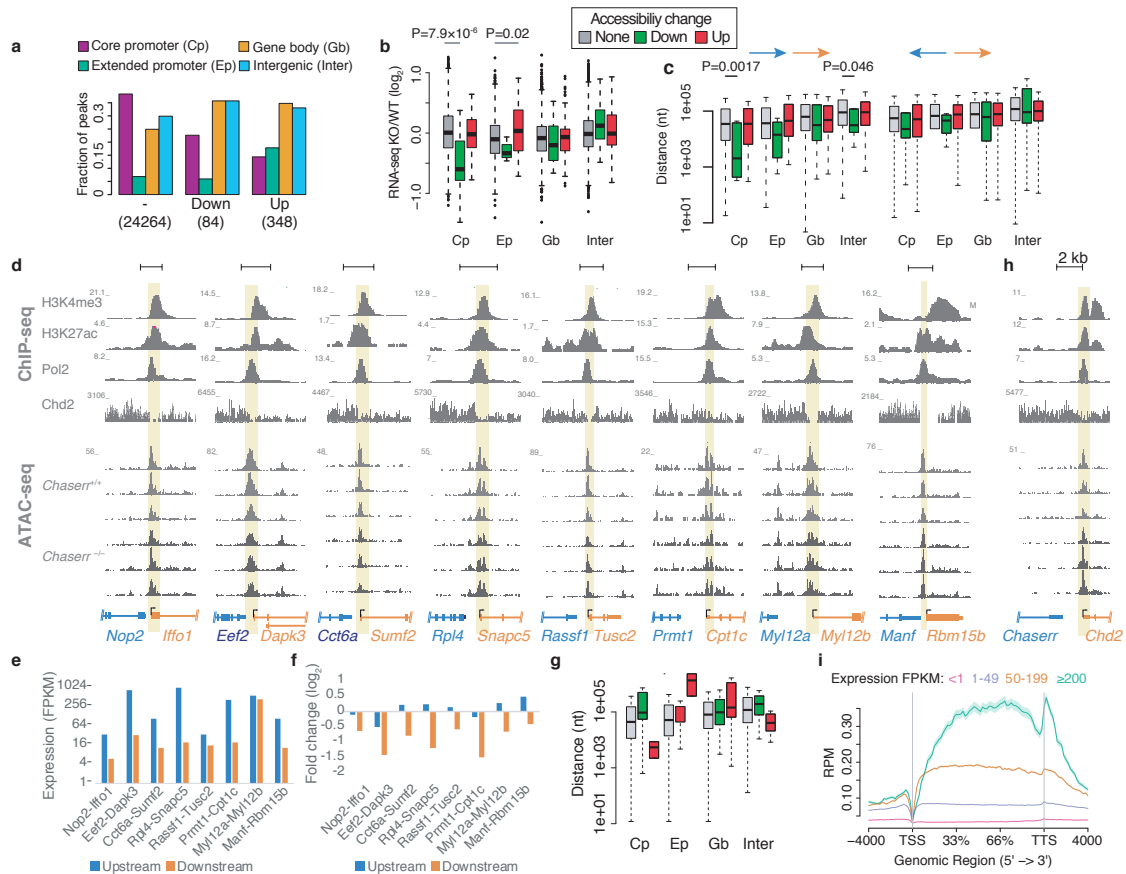
or *Chaserr*<sup>-/-</sup> *Chd2*<sup>m/m</sup> vs. WT mEFs and distance to the 3' end of the closest upstream gene transcribed from the same strand, for genes for which this distance is <10 Kb. **f**, Changes in gene expression in *Chaserr*<sup>-/-</sup> or *Chaserr*<sup>-/-</sup> *Chd2*<sup>m/m</sup> vs. WT mEFs for genes with the indicated distances from the closest upstream gene. Genes with distance <2 Kb were further subdivided based on the expression levels of the upstream gene. **g**, As in e, for the E13.5 whole embryo RNA-seq. The number of genes in each group is indicated below the distance threshold. Error bars show S.E.M. \*P<0.05, \*\*P<0.01 (two-sided t-test).

In contrast, the 616 significantly up-regulated genes (up by at least 25%, adjusted P<0.05) were enriched for various RNA-related GO categories (**Supplementary Table 1**). Notably, changes in E13.5 embryo were significantly anti-correlated with those in E13.5 medial ganglionic eminence (MGE) of *Chd2*<sup>+/-</sup> mice<sup>22</sup> (Spearman R=-0.27, P<10<sup>-100</sup>). Milder anti-correlation was observed between dysregulation in the *Chaserr*<sup>+/-</sup> adult brain and the *Chd2*<sup>+/-</sup> adult hippocampus<sup>22</sup> (R= -0.11, P= 8.8×10<sup>-47</sup>). Hundreds of genes were dysregulated in the *Chaserr*<sup>-/-</sup> *Chd2*<sup>m/m</sup> mEFs compared to WT mEFs from the same pregnancies, and these changes are likely due to *Chd2* loss-of-function, as there was no correlation between these changes and the changes in *Chaserr*<sup>-/-</sup> mEFs (Spearman R=-0.003 P=0.68, **Fig. 4d**). These observations further support the notion that the downstream effects of loss of *Chaserr* are driven by up-regulation of *Chd2*.

Inspection of the loci of some of the most affected genes, such as *Ctsk* and *Dapk3* (**Fig. 4c**), led us to suspect that promoters of down-regulated genes might be preferentially found in close proximity to transcription termination sites (TTSs) of genes transcribed on the same strand (and thus potentially susceptible to transcriptional interference), in a similar organization to the *Chaserr-Chd2* locus. Indeed, we found a mild yet highly significant correlation between changes in gene expression in *Chaserr*<sup>-/-</sup> mEFs and the distance to the TTS of the closest tandem upstream gene (Spearman R=0.2 P=2×10<sup>-151</sup>), with the effect observed mostly when the distance was shorter than ~6 Kb (**Fig. 4e**). In contrast, no such effect was observed in *Chaserr*<sup>-/-</sup> *Chd2*<sup>m/m</sup> mEFs (R=0.0055 P=0.45, **Fig. 4e**). Further, down-regulation was stronger when the expression of those upstream neighboring genes was higher (Spearman R=-0.17 P=3.4×10<sup>-7</sup> between the change in expression of the downstream gene and the absolute expression of the upstream gene for intergenic distances <2Kb). Genes with a close and abundant upstream neighbor were significantly repressed in *Chaserr*<sup>-/-</sup> mEFs, and in stark contrast, much smaller differences, and in opposite direction, were observed for neighboring genes transcribed on opposite strands, when considering distances between 5' ends (**Fig. 4f**). Notably, in most cases of repression of close tandem downstream gene, both the upstream and the downstream gene were protein-coding (190 of the 194 pairs with intergenic distance <2 Kb and a reduction of at least 25% in the expression of the downstream

gene in *Chaserr*<sup>-/-</sup> mEFs). Similar trends were also found in data from E13.5 embryos, with much smaller effect sizes (**Fig. 4g**).

To further characterize the regulatory dysregulation, we performed ATAC-seq<sup>31</sup> using WT and *Chaserr*<sup>-/-</sup> mEFs. Consistently with the increase in *Chd2* production, we observed a mild increase of ~7% in accessibility of *Chd2* promoter in *Chaserr*<sup>-/-</sup> mEFs compared to WT. In order to analyze the changes in accessibility throughout the genome following *Chaserr* depletion, we merged accessibility peaks from all samples, assigned them to the nearest gene, and classified the peaks as “Core promoter” (<300 nt from TSS), “Extended promoter” (>300nt and <2 kb from TSS), “Gene body”, or “Intergenic”, based on RefSeq annotations. Comparing WT and *Chaserr*<sup>-/-</sup> mEFs, we found 84 and 348 regions with reduced or increased accessibility, respectively (t-test P<0.05), with changes observed in gene bodies and in promoters (**Fig. 5a**). Promoter peaks whose accessibility was significantly decreased in *Chaserr*<sup>-/-</sup> mEFs were associated with concordant changes in gene expression (**Fig. 5b**). Strikingly, these peaks were predominantly found at promoters of genes separated by short intergenic regions from TTSSs of other genes with the same relative orientation (**Fig. 5c**, median distance of 2.6 Kb compared to 43.9 Kb for unaffected promoters). All nine genes with significantly reduced core promoter accessibility, a tandem upstream neighbor within <5 Kb, and robust expression in the RNA-seq data were repressed by at least 25% (all with adjusted P<0.05, **Fig. 5d,f**). Notably, in these cases, the upstream gene was expressed at dramatically higher levels than the affected downstream gene (15-fold on average, **Fig. 5e**), and the expression of the upstream genes was not substantially affected by loss of *Chaserr* (**Fig. 5f**).



**Figure 5 | Changes in chromatin accessibility following loss of *Chaserr*.** **a**, Distribution of ATAC-seq peaks that had reduced (‘Down’) or increased (‘Up’) accessibility in *Chaserr*<sup>-/-</sup> mEFs among the indicated four groups. n for each group indicated in parentheses. **b**, Changes in gene expression of genes assigned to peaks in the indicated group (from a). Group names are as in legend of a. P values computed using two sided Wilcoxon rank-sum test. **c**, Distance from the closest gene on the same (left) or opposite (right) strand for the genes in the indicated group (from a). **d**, Loci of eight genes with reduced accessibility of peaks in their core promoters in the *Chaserr*<sup>-/-</sup> background. Top: ChIP-seq coverage in mEFs from the ENCODE project. Middle: CHD2 MNase-ChIP-seq read coverage in mESCs<sup>13</sup>. Bottom - read coverage in ATAC-seq data from mEFs isolated from three embryos from each background. **e**, RNA-seq expression levels for the indicated pairs of genes in WT cells. **f**, Differential expression in *Chaserr*<sup>-/-</sup> mEFs compared to WT mEFs for the indicated gene pairs. **g**, Same as c left, but using data from *Chaserr*<sup>-/-</sup> *Chd2*<sup>m/m</sup> mEFs, compared to WT mEFs. **h**, As in d, for the *Chaserr/Chd2* locus. **i**, Metagene of CHD2 occupancy<sup>13</sup> on genes in mESCs, divided into four groups based on their expression FPKM.

To test if these effects stem from CHD2 over-expression, we used the same methodology to compare *Chaserr*<sup>-/-</sup> *Chd2*<sup>m/m</sup> mEFs with WT mEFs matched from the same pregnancies, and observed fewer changes in accessibility (67 and 26 peaks with decreased and increased accessibility, respectively, based on the same criteria as above), and no preferential association of decreased promoter accessibility with distance to tandem upstream genes (**Fig. 5g**), suggesting that the transcriptional interference in *Chaserr*<sup>-/-</sup> cells is driven by excess CHD2 levels.

The presence of a TTS closely upstream of the promoters most affected by loss of *Chaserr* (**Fig. 5d**) resembles the organization of the *Chaserr* and *Chd2* locus (**Fig. 5h**). We observed substantial occupancy of CHD2 protein (using tagged CHD2 MNase-ChIP-seq data from mESCs<sup>13</sup>) in the short intergenic regions between the upstream TTS and the promoters of the repressed genes, as between *Chaserr* TTS and *Chd2* (**Fig. 5d,h**), suggesting that the effect seen on the accessibility of these promoters and gene expression might be a direct result of increased CHD2 activity in the intergenic regions. Concordantly with these individual cases, a metagene analysis of tagged CHD2 occupancy across the gene body showed that CHD2 preferentially associates with ~2Kb regions immediately downstream of TTSs (**Fig. 5i**). CHD2 thus appears to function in regions downstream of TTSs in WT cells, and the CHD2 hyperactivity caused by loss of *Chaserr* leads to repression of promoters adjacent to those regions.

## Discussion

Decoding lncRNA functions is a formidable challenge, in particular because of the substantial heterogeneity in their biology and modes of action. A prominent group of vertebrate lncRNAs, including some of the most conserved ones, are found in regions flanking genes involved in transcription, including numerous chromatin modifiers. We hypothesized that such co-localization may imply a connection between the biology of the chromatin modifier and the mode of action of the lncRNA, potentially through a feedback loop. By focusing on one such pair that is particularly highly conserved in evolution, we uncovered a lncRNA-mediated circuit that regulates the expression of *Chd2*.

Negative autoregulatory feedback loops are prevalent in RNA biology, and help tune levels of factors involved in mRNA splicing, polyadenylation, editing, and modifications, and in processing of small RNAs<sup>32-38</sup>. There are also examples of such feedback loops regulating transcription-related genes<sup>39</sup>, but to the best of our knowledge, there are no known autoregulatory loops involving chromatin modifiers in mammals. We propose a model in which the *Chaserr-Chd2* tandem organization forms a negative feedback loop that tunes CHD2 levels. The genomic occupancy of CHD2 in regions immediately downstream of TTSs (**Fig. 5i**), as well as the enrichment of CHD2-mediated accessibility changes in gene bodies (**Fig. 5a**), suggest that CHD2 acts in TTS-proximal regions. Accordingly, we show that under conditions of excess levels of CHD2 protein, this TTS-proximal occupancy is associated with transcriptional interference on downstream neighbors of highly-expressed genes, especially when the intergenic regions between them are particularly short (**Fig. 4e-g, Fig. 5c-g**). Due to the short intergenic region between *Chaserr* and *Chd2*, we suggest that such transcriptional interference serves as the basis of a negative feedback loop that leads to repression of *Chd2*

production when CHD2 protein levels are high, thus maintaining a tight control on its expression. In cells with *Chaserr* loss-of-function, this loop is compromised, and CHD2 is upregulated. Supporting this model, in *Chaserr*<sup>+/-</sup> *Chd2*<sup>m/+</sup> tissues, increase in the CHD2 protein leads to repression of transcription of the *Chd2*<sup>m</sup> allele, which is found next to an intact *Chaserr* (**Fig. 3d**).

The exact mechanism by which *Chaserr* represses *Chd2* expression remains unknown. The simplest feedback model suggests that *Chaserr* transcription, or perhaps its termination and cleavage and polyadenylation, are important. Such a mechanism would resemble the activity of the SRG1 noncoding RNA in the yeast *S. cerevisiae*, that represses the transcription of SER3, which is found immediately downstream of SRG1 and on the same strand, by deposition of nucleosomes at the SER3 promoter that prevent binding of transcription activators<sup>40-44</sup>. The genomic arrangement of SRG1 and SER3 resembles that of *Chaserr* and *Chd2*, though the 3' end of SRG1 overlaps SER3 promoter<sup>41</sup>, and the distance between their promoters is much shorter (~500 bp between SRG1 TSS and SER3 TSS vs. ~17 kb between the TSS of *Chaserr* and TSS of *Chd2*). However, there is evidence that the RNA product of *Chaserr* is important for its function. The exon-intron architecture and sequence of *Chaserr* are highly conserved, suggesting that production of a particular RNA species, or splicing at particular locations, is important either because of the RNA product itself, or because they modulate the amount of transcription or Pol2 velocity upon encounter with the ~2 Kb intergenic region between *Chaserr* and *Chd2*. The similar effects we observed when deleting *Chaserr* promoter or gene body, or when targeting the RNA with ASOs or Gapmers, also point to the importance of the RNA product, though we cannot rule out that ASOs and Gapmers may affect transcription by cleaving the nascent transcript.

Both *Chaserr* and *Chd2* appear to be under the shared control of several enhancer regions found in the ~200 Kb gene desert upstream of *Chaserr* (**Extended Data Fig. 4**), and those enhancers potentially regulate both genes together, which can explain the tight co-expression between *Chaserr* and *Chd2* mRNA (**Fig. 1c**). We propose that shared regulation of both genes is potentially important for maintaining the integrity of the feedback mechanism, as we found that expression levels of the upstream gene appear to be consequential for transcriptional interference resulting from CHD2 hyperactivity (**Fig. 4f**).

Interestingly, in contrast to previous reports based on a gene trap inserted in intron 27 of *Chd2*<sup>17,45,46</sup>, nearly complete loss of CHD2 protein appears to be largely compatible with normal viability in mice in our hands, and in the similar mouse model generated by the EUCOMM consortium<sup>18</sup>. The previous observations of embryonic lethality of CHD2 loss-of-function can potentially be attributed to a dominant negative effect of the truncated protein product created by the gene trap. In contrast to the CHD2 loss models, we report here that

increase in CHD2 levels brought about by loss of even a single copy of *Chaserr* is toxic and leads to perinatal lethality. We note that this is one of the most severe phenotypes reported so far for loss of a lncRNA<sup>6,47</sup>. The severe, CHD2-mediated phenotype is consistent with the high conservation of *Chaserr-Chd2* genomic organization and sequence during 500 million years of vertebrate evolution. Of the lncRNAs annotated in human and mouse, only ~5% (~100 genes) have evidence of conservation in fish<sup>26,48</sup>, and we were able to identify homologs of *Chaserr* in every vertebrate species we examined (**Fig. 1a**).

Our results have important implications from the therapeutic perspective. Individuals that bear mutations in the *CHD2* gene exhibit epilepsy and neurodevelopmental disorders<sup>19</sup>. In all described cases, these individuals are haploinsufficient for *CHD2*, and so bear an intact WT copy of *CHD2*. Therefore, increase of *CHD2* expression through perturbation of *Chaserr*, e.g., using antisense oligonucleotides, might have a therapeutic benefit. Importantly, *Chaserr* is highly conserved between human and mouse (**Fig. 1a**), and targeting of the human *Chaserr* using Gapmers in human MCF7 and SH-SY5Y cells leads to an increase in *CHD2* mRNA and protein levels (**Extended Data Fig. 5**), suggesting that the results we observe in the mouse model are of direct relevance to human *Chaserr*.

Similarly to many other chromatin remodelling factors, that are increasingly implicated in human disease, the precise molecular function of *Chd2* and the direct consequences of its dysregulation are poorly understood. Here, we could leverage the similarity between the genomic arrangements of *Chaserr* and *Chd2* and those of the genes most affected by *Chaserr* perturbation to highlight the potential role of chromatin remodelling that CHD2 plays downstream of TTSs. Decoding lncRNA functions can therefore provide important insight into other layers of gene regulation. As chromatin modifiers are often flanked by lncRNAs, some as highly conserved as *Chaserr*, we expect that further research into the functions of these lncRNAs may uncover new paradigms in chromatin biology.



## Materials and Methods

### Animals

The study was conducted in accordance with the guidelines of the Weizmann Institutional Animal Care and Use Committee (IACUC). C57black6 Ola HSD mice were purchased from Harlan Laboratories (Rehovot, Israel). All other mouse strains were bred and maintained at the Veterinary Resources Department of the Weizmann Institute.

### Generation of KO mice

The CRISPR KO mice were generated as in (Wang et al., 2013). All KO mice were generated by standard procedures at the Weizmann transgenic core facility. To generate *Chaserr* KO mice we used four single guide RNAs (sgRNAs 1-4, see **Supplementary Table 2**), two targeting regions before the *Chaserr* transcription start site and two targeting regions in the first intron, which resulted in two founder lines. *Chaserr* mutant mice were genotyped using primers flanking the targeted region followed by Sanger sequencing. To generate *Chd2* mutant mice we used one sgRNA targeted to the third exon of *Chd2* mRNA. *Chd2* mutant mice were genotyped by amplicon sequencing which identified a founder with an ORF shift. Amplicon library preparation was as follows: *Chd2* mutant locus was amplified using *Chd2* genotype nesting primers (**Supplementary Table 2**). Thereafter 1µl PCR reaction was used as template for the addition of R1 and R2 Illumina adaptors (R1/R2\_Exon3\_Chd2, **Supplementary Table 2**), followed by fragment AMPure (Beckman Coulter, A63881) size selection and cleanup. A second PCR reaction was used in order to add sequencing barcodes to the amplicons, followed by AMPure size selection. Indexed amplicons were pooled and sequenced on NextSeq 500. To genotype *Chd2<sup>m</sup>* lines routinely, we used PCR with primers flanking the mutated region and digested the PCR product with the DdeI restriction enzyme whose recognition sequence is compromised in the mutated DNA. The fragment sizes were then analyzed on a 2% agarose gel. To generate *Chaserr<sup>-/-</sup> Chd2<sup>m/m</sup>* mice we used *Chd2<sup>m/m</sup>* background and continued as described above for generation of the *Chaserr<sup>-</sup>* allele, using only two sgRNAs (sgRNA2+3, **Supplementary Table 2**). sgRNA injection were done on CB6F1 Ola HSD mice which were later backcrossed with C57BL/6 Ola HSD. All the experiments were done on 4 to 15-week-old mice from F2–F5 generations and E9.5–E18.5 developmental time points.

### Tissue culture

mESCs were routinely cultured in mouse ES medium (mESM) consisting of: 500ml DMEM (Gibco, 11965-092), 15% ES-grade Fetal Calf Serum (Biological Industries), sodium pyruvate 1mM (Gibco, 11360-039), nonessential amino acids 1x (Gibco, 11140-035), 0.1mM

b-mercaptoethanol (Sigma, M6250-250ML), penicillin-streptomycin (Biological Industries), and 1000U/ml LIF (Weizmann Proteomics Unit).

All other cell lines were routinely cultured in DMEM containing 10% fetal bovine serum and 100 U penicillin/0.1 mg ml<sup>-1</sup> streptomycin at 37° C in a humidified incubator with 5% CO<sub>2</sub>. Cell lines were routinely tested for mycoplasma contamination.

Primary mEFs were isolated as previously described<sup>49</sup>.

### **Neuronal differentiation**

Neuronal differentiation was performed as previously described (Ying et al., 2003). mESCs were first grown in the absence of mEFs for two passages and then seeded on gelatin-coated plates at a density of 1.5x10<sup>5</sup> in N2B27 medium: 1:1 mixture of DMEM/F12 (Sigma) supplemented with N2 (Gibco), and Neurobasal medium (Gibco) supplemented with B27 (Gibco), 1X Glutamax (Gibco), 0.1 mM β-mercaptoethanol (Sigma), 100 U/ml penicillin, and 0.1 mg/ml streptomycin (Biological industries). After four days under these conditions, 3x10<sup>5</sup> cells were re-plated on Poly-D-Lysine (Sigma, P6407) and Laminin (Life, 23017-015) coated plates, in N2B27 medium supplemented with 20ng/ml FGF2 (Peprotech, 100-18B-50/100-18B-100). After 24 hr, FGF2 was removed and cells were cultured for three additional days.

### **Transfections**

mESCs were transfected with electroporation the Lonza protocol

([http://bio.lonza.com/fileadmin/groups/marketing/Downloads/Protocols/Generated/Optimized\\_Protocol\\_309.pdf](http://bio.lonza.com/fileadmin/groups/marketing/Downloads/Protocols/Generated/Optimized_Protocol_309.pdf)).

HEK293T cells were transfected with were performed using PolyEthylene Imine (PEI)<sup>50</sup> (PEI linear, Mr 25,000, Polyscience). Neuro2A transfection: 2×10<sup>5</sup> were seeded in 6 well-plate and transfected using Lipofectamine 3000 (Life, L3000-008) with LNA1, LNA2 or mix of LNA1-4 (LNA KD) or with ASO1, ASO2, ASO3 or mix of ASO1-3 (ASO KD) to final concentration of 50 nM. For transfection of MCF7 cells, 2×10<sup>5</sup> were seeded in a 6-well plate and transfected using PEI (as previously described) with LNA h1 and/or LNA h2 to final concentration of 50 nM. For luciferase experiments, 20×10<sup>3</sup> cells were seeded in a 24-well plate and co-transfected using PEI with pGL3-Chaserr-reporter vector (100 ng) and psiCHECK-1 (100 ng, Promega, C8011). Firefly and Renilla expression were measured using Dual-Luciferase Reporter Assay System (Promega, E1910) following manufacturer's protocol. For transfection of SH-SY5Y cells, 2×10<sup>5</sup> cells were seeded in a 6-well plate and transfected using DharmaFECT 4 Transfection Reagent (Dharmacon, T-2002-03) following manufacturer's protocol with LNA h1 and/or LNA h2 to final concentration of 50 nM. End time point for all KD experiments and luciferase assay was at 48 hr post transfection.

## Genome editing in mESCs

To generate mESC *Chaserr*<sup>p/p</sup> mESCs,  $2 \times 10^6$  mEF-depleted cells were transfected with sgRNAs 2+3 and pCas9\_GFP (a gift from Kiran Musunuru, Addgene #44719). To generate *Chaserr*<sup>b/b</sup> mESCs,  $2 \times 10^6$  cells were transfected with sgRNAs 2+5 and pCas9\_GFP. The next day fresh mESC medium was supplemented with 1 $\mu$ g/ml puromycin (Invivogen, ant-pr-1) for 72 hr, while replacing medium each 24 hr. Next,  $3\text{--}4 \times 10^3$  cells were seeded at low density on 10 cm plate until single colonies formed, thereafter colonies were picked, expanded, genomic deletion was verified by PCR, sequencing, and expression was tested by RT-qPCR.

## RNA and RT-qPCR

Total RNA was extracted from different cell lines and mouse tissues, using TRIAGENT (MRC) according to the manufacturer's protocol. cDNA was synthesized using qScript Flex cDNA synthesis kit (95049, Quanta). Fast SYBR Green master mix (4385614) was used for qPCR.

## RNA-seq

Strand-specific mRNA-seq libraries were prepared from 500–4000 ng total RNA using the TruSeq Stranded mRNA (Illumina) or SENSE mRNA-Seq (Lexogen) library preparation kits, according to the manufacturer protocols. Libraries were sequenced on a NextSeq 500 to obtain 38–50 nt paired-end reads. Coverage tracks for the UCSC genome browser were prepared by aligning reads to the mm9 genome assembly using STAR<sup>51</sup>. Gene expression levels were quantified using RSEM<sup>52</sup> and a RefSeq gene annotation database which was manually edited to correct the annotation of the last exon of *Chaserr*. Differential expression was computed using DESeq2 with default settings<sup>53</sup>. Genomic context was also analyzed using the RefSeq gene annotations. RNA-seq and ATAC-seq datasets are deposited in GEO database under the accession GSE124375 (reviewer token wtkleowgbhotjgn). RNA-seq data from previous studies were downloaded from the SRA database, and quantified using RSEM with the same annotation file.

## Western Blot

Total protein was extracted from tissues and cell lines by lysis with RIPA supplemented with protease inhibitors and DTT 1mM. Proteins were resolved on 8-10% SDS-PAGE gels and transferred to a Polyvinylidene difluoride (PVDF) membrane. After blocking with 5% non-fat milk in PBS with 0.1% Tween-20 (PBST), membranes were incubated with the primary antibody followed by the secondary antibody conjugated with horseradish peroxidase. Blots were quantified using Image Lab software. Primary antibodies as follows: anti-Chd2

(Millipore, MABE873), anti- $\beta$ -tubulin (Sigma, T4026). Secondary antibodies as follows: anti-rat (AP136P), anti-mouse (115-035-003).

### **ATAC-seq**

ATAC-seq was performed as previously described (Buenrostro et al., 2013).  $5 \times 10^4$  mEFs derived from WT, *Chaserr*<sup>-/-</sup>, or *Chaserr*<sup>-/-</sup> *Chd2*<sup>m/m</sup> backgrounds. Libraries were sequenced with paired-end sequencing on Illumina NextSeq 500. Reads were aligned to the mm9 genome assembly using Bowtie2<sup>54</sup> and peaks were called using MACS2<sup>55</sup>. Normalized read coverage files were computed by MACS2. Peaks from individual samples were merged using bedtools<sup>56</sup> and read coverage in the peaks was computed using bigWigAverageOver bed UCSC utility<sup>57</sup>. Differential peaks were then called using two-sided t-test. Peaks were assigned to the closest RefSeq gene and annotated as “Core promoter” if they fell within 300 nt of a TSS; “Extended promoter” if they fell within 2,000 nt of a TSS; “Gene body” if they overlapped a transcription unit; or “Intergenic” otherwise.

### **Single-molecule FISH**

Stellaris probe libraries targeting *Chaserr* introns or exons were designed using the Biosearch Technologies server (**Supplementary Table 3**) and ordered from Biosearch Technologies. Duodenum sections were hybridized according to a previously published protocol<sup>58</sup> with the exception of using 25% formamide and 6 $\mu$ M sections. DAPI (Sigma-Aldrich, D9542) and a FITC-conjugated antibody against E-Cadherin (BD Biosciences, 612131) were used to identify the nucleus and cell-membrane, respectively. Imaging was performed on a Nikon-Ti-E inverted fluorescence microscope with a 100 $\times$  oil-immersion objective and a Photometrics Pixis 1024 CCD camera using MetaMorph software as previously described<sup>59</sup>. The analysis was carried on ImageM, a custom Matlab program<sup>60</sup> to compute single-molecule mRNA concentrations in the nucleus or cytoplasm by segmenting each cell manually according to the cell borders and the nucleus.

### **Cycloheximide treatment**

mESCs and mEFs were treated with DMSO (vehicle) or CHX (Sigma #C7698) 100 $\mu$ g/mL and collected at the indicated time points for RNA/protein analysis.

### **Extraction of cytoplasmic and nuclear RNA**

Cells were washed twice with ice-cold PBS, then scraped with ice-cold buffer A (EGTA 15 $\mu$ M, EDTA 10 $\mu$ M, protease inhibitor cocktail (Sigma, P8340) and RNase inhibitor (ERX-E4210-01)) and centrifuged at 400g, 4°C for 5' min. Supernatant was discarded, fresh buffer A was added and the pellet was mechanically pipetted with 21G followed by 27G syringe. Cells were then centrifuged at 2000g, 4°C for 5 min and the syringe step was repeated. The

cells were then centrifuged at 2500g, 4°C for 5 min. The pellet was then kept as the nuclear fraction and the supernatant was centrifuged again at 6000g, 5 min followed by another supernatant collection (clean cytosolic fraction). Nuclear pellet was then washed three times with buffer A. RNA was extracted TRIAGENT (MRC).

### **Molecular cloning**

Guide RNAs were designed using CHOPCHOP<sup>61</sup>. Cloning of plasmids were done following [Zhang Lab General Protocol](http://www.genome-engineering.org/crispr/wp-content/uploads/2014/05/CRISPR-Reagent-Description-Rev20140509.pdf) (<http://www.genome-engineering.org/crispr/wp-content/uploads/2014/05/CRISPR-Reagent-Description-Rev20140509.pdf>) using pHU6-gRNA<sup>62</sup>, (a gift from Charles Gersbach, Addgene plasmid # 53188) or pKLV-U6gRNA(BbsI)-PGKpuro2ABFP<sup>63</sup> (a gift from Kosuke Yusa, Addgene plasmid # 50946). For luciferase experiments, a synthetic sequence containing *Chaserr* promoter, *Chaserr* cDNA, the intergenic region between *Chaserr* and *Chd2*, and *Chd2* promoter were synthesized by hylabs and cloned to into pGL3-basic backbone (Promega). To generate the plasmids with deleted fragments we used the following restriction enzymes followed by ligation: (i) *Chaserr* promoter deletion: XhoI+PvuII; (ii) *Chaserr* cDNA+intergenic deletion: PvuII+BglIII; (iii) *Chaserr* cDNA+intergenic+*Chd2* promoter: PvuII+HindIII; (iv) *Chaserr* promoter+*Chaserr* cDNA+intergenic: XhoI+BglIII; (v) *Chd2* promoter: BglIII+HindIII.

### ***Chaserr* cloning and lentiviral production**

cDNA from Source BioScience clone C130076G01 was amplified with PCR adding restriction sites for NheI at 5' end and AgeI at 3' end (**Supplementary Table 2**) and cloned into pLIX\_402 vector (a gift from David Root, Addgene #41394) using restriction ligation. To produce viruses, HEK293T ( $2.5 \times 10^6$ , 10cm plate) cells were transfected using PEI with psPAX2 (3.5µg, a gift from Didier Trono, Addgene #12260), pMD2.G (1.5µg, a gift from Didier Trono, Addgene, #12259) and pLIX\_402-*Chaserr* (5µg). Viruses were collected 48 hr post-transfection and filtered through 0.45µm sterile filters. Viruses were supplemented with polybrene (1:1000, Sigma, 107689-10G) upon cell infection.

### **Micro-CT scanning and analysis**

Prior to Micro-CT scanning the mice were anesthetized using IP injection of a mix of Xylazine (10mg/Kg) and Ketamine (100mg/Kg). Mice were scanned using a micro-CT device TomoScope® 30S Duo scanner equipped with two source-detector systems. The scanner uses two x-ray sources and a detector system that are mounted on a gantry that rotates around a bed holding the animal. The operation voltages of both tubes were 40kV. The integration time of protocols was 90ms (360 rotations) for 3cm length and axial images were obtained at an isotropic resolution of 80µm. Due to the maximum length limit, to cover the whole mouse body, imaging was performed in two-three parts with overlapping area and then all slices

merged to one dataset representing the entire ROI. The radiation dose range was 0.9Gy. All micro-CT scans were reconstructed using a filtered back-projection algorithm using scanner software. Then the reconstructed data sets for each mice were merged to one data set using ImageJ software. 3D volume rendering images were produced using Amira Software.

## References

1. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).
2. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
3. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
4. Ravasi, T. *et al.* Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.* **16**, 11–19 (2006).
5. Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
6. Perry, R. B.-T. & Ulitsky, I. The functions of long noncoding RNAs in development and stem cells. *Development* **143**, 3882–3894 (2016).
7. Han, P. & Chang, C.-P. Long non-coding RNA and chromatin remodeling. *RNA Biol.* **12**, 1094–1098 (2015).
8. Rutenberg-Schoenberg, M., Sexton, A. N. & Simon, M. D. The Properties of Long Noncoding RNAs That Regulate Chromatin. *Annu. Rev. Genomics Hum. Genet.* **17**, 69–94 (2016).
9. Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550 (2011).
10. Tajul-Arifin, K. *et al.* Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome. *Genome Res.* **13**, 1416–1429 (2003).
11. Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of

- ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490–503 (2013).
12. Siggins, L., Cordeddu, L., Rönnerblad, M., Lennartsson, A. & Ekwall, K. Transcription-coupled recruitment of human CHD1 and CHD2 influences chromatin accessibility and histone H3 and H3.3 occupancy at active chromatin regions. *Epigenetics Chromatin* **8**, 4 (2015).
  13. de Dieuleveult, M. *et al.* Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* **530**, 113–116 (2016).
  14. Harada, A. *et al.* Chd2 interacts with H3.3 to determine myogenic cell fate. *EMBO J.* **31**, 2994–3007 (2012).
  15. Semba, Y. *et al.* Chd2 regulates chromatin for proper gene expression toward differentiation in mouse embryonic stem cells. *Nucleic Acids Res.* **45**, 8758–8772 (2017).
  16. Luijsterburg, M. S. *et al.* PARP1 Links CHD2-Mediated Chromatin Expansion and H3.3 Deposition to DNA Repair by Non-homologous End-Joining. *Mol. Cell* **61**, 547–562 (2016).
  17. Marfella, C. G. *et al.* Mutation of the SNF2 family member Chd2 affects mouse development and survival. *J. Cell. Physiol.* **209**, 162–171 (2006).
  18. Dickinson, M. E. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508–514 (2016).
  19. Lamar, K.-M. J. & Carvill, G. L. Chromatin Remodeling Proteins in Epilepsy: Lessons From CHD2-Associated Epilepsy. *Front. Mol. Neurosci.* **11**, 208 (2018).
  20. Shen, T., Ji, F., Yuan, Z. & Jiao, J. CHD2 is Required for Embryonic Neurogenesis in the Developing Cerebral Cortex. *Stem Cells* **33**, 1794–1806 (2015).
  21. Meganathan, K. *et al.* Regulatory networks specifying cortical interneurons from human embryonic stem cells reveal roles for CHD2 in interneuron development. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11180–E11189 (2017).
  22. Kim, Y. J. *et al.* Chd2 Is Necessary for Neural Circuit Development and Long-Term Memory. *Neuron* (2018). doi:10.1016/j.neuron.2018.09.049
  23. Lin, M. F., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to

- distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–82 (2011).
24. Wang, L. *et al.* CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, e74 (2013).
  25. Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–9 (2007).
  26. Hezroni, H. *et al.* Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep.* (2015).  
doi:10.1016/j.celrep.2015.04.023
  27. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* (2016). doi:10.1038/nature20149
  28. Tuck, A. C. *et al.* Distinctive features of lincRNA gene expression suggest widespread RNA-independent functions. *Life Sci Alliance* **1**, e201800124 (2018).
  29. Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
  30. Saftig, P. *et al.* Impaired osteoclastic bone resorption leads to osteopetrosis in cathepsin-K-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13453–13458 (1998).
  31. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
  32. Lykke-Andersen, S., Ardal, B. K., Hollensen, A. K., Damgaard, C. K. & Jensen, T. H. Box C/D snoRNP Autoregulation by a cis-Acting snoRNA in the NOP56 Pre-mRNA. *Mol. Cell* **72**, 99–111.e5 (2018).
  33. Jangi, M., Boutz, P. L., Paul, P. & Sharp, P. A. Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes Dev.* **28**, 637–651 (2014).
  34. Triboulet, R. & Gregory, R. I. Autoregulatory Mechanisms Controlling the Microprocessor. in *Regulation of microRNAs* (ed. Großhans, H.) 56–66 (Springer US, 2010).



35. Luo, W. *et al.* The conserved intronic cleavage and polyadenylation site of CstF-77 gene imparts control of 3' end processing activity through feedback autoregulation and by U1 snRNP. *PLoS Genet.* **9**, e1003613 (2013).
36. Li, Y. *et al.* An intron with a constitutive transport element is retained in a Tap messenger RNA. *Nature* **443**, 234–237 (2006).
37. Pendleton, K. E. *et al.* The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **169**, 824–835.e14 (2017).
38. Feng, Y., Sansam, C. L., Singh, M. & Emeson, R. B. Altered RNA editing in mice lacking ADAR2 autoregulation. *Mol. Cell. Biol.* **26**, 480–488 (2006).
39. Crews, S. T. & Pearson, J. C. Transcriptional autoregulation in development. *Curr. Biol.* **19**, R241–6 (2009).
40. Pruneski, J. A., Hainer, S. J., Petrov, K. O. & Martens, J. A. The Paf1 complex represses SER3 transcription in *Saccharomyces cerevisiae* by facilitating intergenic transcription-dependent nucleosome occupancy of the SER3 promoter. *Eukaryot. Cell* **10**, 1283–1294 (2011).
41. Martens, J. A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* **429**, 571–574 (2004).
42. Martens, J. A., Wu, P.-Y. J. & Winston, F. Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. *Genes Dev.* **19**, 2695–2704 (2005).
43. Hainer, S. J., Pruneski, J. A., Mitchell, R. D., Monteverde, R. M. & Martens, J. A. Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev.* **25**, 29–40 (2011).
44. Thebault, P. *et al.* Transcription regulation by the noncoding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. *Mol. Cell. Biol.* **31**, 1288–1300 (2011).
45. Nagarajan, P. *et al.* Role of chromodomain helicase DNA-binding protein 2 in DNA damage response signaling and tumorigenesis. *Oncogene* **28**, 1053–1062 (2009).

46. Kulkarni, S. *et al.* Disruption of chromodomain helicase DNA binding protein 2 (CHD2) causes scoliosis. *Am. J. Med. Genet. A* **146A**, 1117–1127 (2008).
47. Feyder, M. & Goff, L. A. Investigating long noncoding RNAs using animal models. *J. Clin. Invest.* **126**, 2783–2791 (2016).
48. Ulitsky, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* (2016). doi:10.1038/nrg.2016.85
49. Jozefczuk, J., Drews, K. & Adjaye, J. Preparation of mouse embryonic fibroblast cells suitable for culturing human embryonic and induced pluripotent stem cells. *J. Vis. Exp.* (2012). doi:10.3791/3854
50. Durocher, Y., Perret, S. & Kamen, A. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res.* **30**, E9 (2002).
51. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
53. Love, M., Anders, S. & Huber, W. Differential analysis of count data--the DESeq2 package. *Genome Biol.* **15**, 550 (2014).
54. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
55. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
56. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
57. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
58. Itzkovitz, S. *et al.* Single-molecule transcript counting of stem-cell markers in the mouse

- intestine. *Nat. Cell Biol.* **14**, 106–114 (2011).
59. Bahar Halpern, K. & Itzkovitz, S. Single molecule approaches for quantifying transcription and degradation rates in intact mammalian tissues. *Methods* **98**, 134–142 (2016).
  60. Lyubimova, A. *et al.* Single-molecule mRNA detection and counting in mammalian tissue. *Nat. Protoc.* **8**, 1743–1758 (2013).
  61. Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–7 (2014).
  62. Kabadi, A. M., Ousterout, D. G., Hilton, I. B. & Gersbach, C. A. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res.* **42**, e147 (2014).
  63. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
  64. Consortium, Fantom *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
  65. Nam, J. W. *et al.* Global analyses of the effect of different cellular contexts on microRNA targeting. *Mol. Cell* **53**, 1031–1043 (2014).
  66. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014).
  67. Li, B. *et al.* A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci. Rep.* **7**, 4200 (2017).
  68. Weischenfeldt, J. *et al.* Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol.* **13**, R35 (2012).
  69. Li, T. *et al.* Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. *EMBO J.* **34**, 1630–1647 (2015).

70. Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16**, 152 (2015).
71. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99–101 (2016).