

**Spring 2019 – Epigenetics and Systems Biology**  
**Discussion Session (Epigenetics and Development)**  
**Michael K. Skinner – Biol 476/576**  
**Week 9 (March 7)**

**Epigenetics of Cell and Developmental Biology**

Primary Papers

1. Schworer, et al., (2016) Nature 540:428
2. Lister, et al., (2013) Science. 341:e1237905
3. Lyko F, et al., (2010) PLoS Biol. 2;8(11):e1000506

**Discussion**

Student 25 – Ref #1 above

- What is the epigenetic aging effect observed?
- What stem cell effect was observed?
- How do epigenetics and genetics cooperate in this process?

Student 26 – Ref #2 above

- What was the experimental design to investigate brain development?
- What technology was used to examine epigenetics?
- What observations regarding neuronal DNA methylation were made?

Student 27 – Ref #3 above

- What are the cast systems in the bee?
- How does epigenetics influence the development of the bee?
- What is the environmental factor that alters the epigenetic programming?

# Epigenetic stress responses induce muscle stem-cell ageing by *Hoxa9* developmental signals

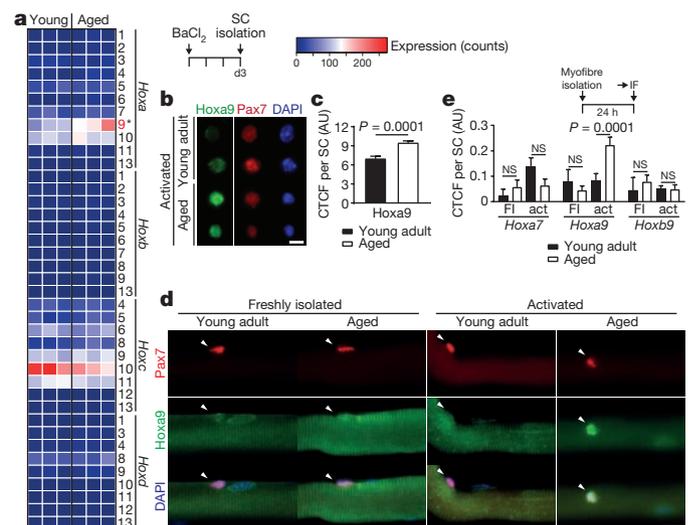
Simon Schwörer<sup>1</sup>, Friedrich Becker<sup>1</sup>, Christian Feller<sup>2</sup>, Ali H. Baig<sup>1</sup>, Ute Köber<sup>1</sup>, Henriette Henze<sup>1</sup>, Johann M. Kraus<sup>3</sup>, Beibei Xin<sup>4</sup>, André Lechel<sup>5</sup>, Daniel B. Lipka<sup>6</sup>, Christy S. Varghese<sup>1</sup>, Manuel Schmidt<sup>1</sup>, Remo Rohs<sup>4</sup>, Ruedi Aebersold<sup>2,7</sup>, Kay L. Medina<sup>8</sup>, Hans A. Kestler<sup>1,3</sup>, Francesco Neri<sup>1</sup>, Julia von Maltzahn<sup>1,§</sup>, Stefan Tümpel<sup>1,§</sup> & K. Lenhard Rudolph<sup>1,9</sup>

The functionality of stem cells declines during ageing, and this decline contributes to ageing-associated impairments in tissue regeneration and function<sup>1</sup>. Alterations in developmental pathways have been associated with declines in stem-cell function during ageing<sup>2–6</sup>, but the nature of this process remains poorly understood. Hox genes are key regulators of stem cells and tissue patterning during embryogenesis with an unknown role in ageing<sup>7,8</sup>. Here we show that the epigenetic stress response in muscle stem cells (also known as satellite cells) differs between aged and young mice. The alteration includes aberrant global and site-specific induction of active chromatin marks in activated satellite cells from aged mice, resulting in the specific induction of *Hoxa9* but not other Hox genes. *Hoxa9* in turn activates several developmental pathways and represents a decisive factor that separates satellite cell gene expression in aged mice from that in young mice. The activated pathways include most of the currently known inhibitors of satellite cell function in ageing muscle, including Wnt, TGF $\beta$ , JAK/STAT and senescence signalling<sup>2–4,6</sup>. Inhibition of aberrant chromatin activation or deletion of *Hoxa9* improves satellite cell function and muscle regeneration in aged mice, whereas overexpression of *Hoxa9* mimics ageing-associated defects in satellite cells from young mice, which can be rescued by the inhibition of *Hoxa9*-targeted developmental pathways. Together, these data delineate an altered epigenetic stress response in activated satellite cells from aged mice, which limits satellite cell function and muscle regeneration by *Hoxa9*-dependent activation of developmental pathways.

Age-dependent declines in the number and function of Pax7<sup>+</sup> satellite cells (SCs) impair the regenerative capacity of skeletal muscle<sup>2,4,9</sup>. Genes and pathways that contribute to this process<sup>2–6</sup> often also have a role in regulating embryonic development<sup>10–13</sup>. Despite these parallels, the function of the master regulators of development, Hox genes, has not been determined in SC ageing. An analysis of freshly isolated, *in vivo* activated SCs from young adult and aged mice (Extended Data Fig. 1a–e) revealed a specific upregulation of *Hoxa9* in SCs from aged mice, both at the mRNA (Fig. 1a, Extended Data Fig. 2a, b) and protein level (Fig. 1b, c). Similar results were obtained by immunofluorescence staining of SCs (Extended Data Fig. 2c) and myofibre-associated SCs (Fig. 1d, e, Extended Data Fig. 2d) that were activated in culture (Extended Data Fig. 1f, g).

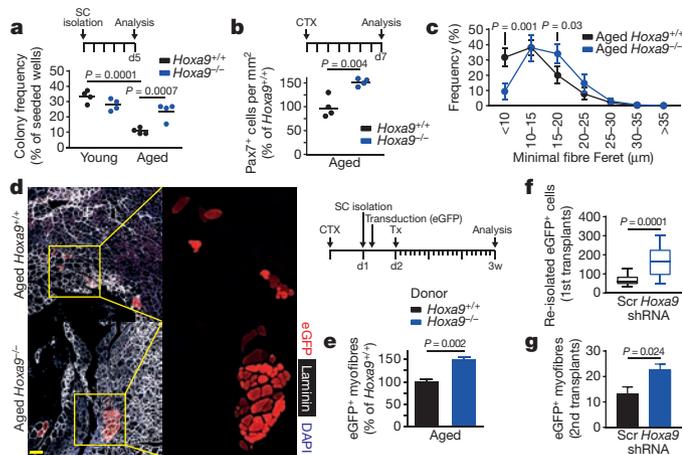
Ageing reduces the proliferative and self-renewal capacity of SCs in wild-type mice<sup>2,9,14,15</sup> (*Hoxa9*<sup>+/+</sup>; Extended Data Fig. 3). Homozygous deletion of *Hoxa9* (*Hoxa9*<sup>-/-</sup>) did not affect the colony-forming capacity of SCs from young adult mice but ameliorated ageing-associated impairment in colony formation of single-cell-sorted SCs in culture (Fig. 2a). *Hoxa9* deletion also increased the self-renewal of myofibre-associated SCs from aged mice in culture but had no effect on SCs

from young adult mice under these conditions (Extended Data Fig. 4a–c). Similar results were obtained by short interfering RNA (siRNA)-mediated knockdown of *Hoxa9* in myofibre-associated SC cultures derived from aged mice (Extended Data Fig. 4d–h). The number of SCs decreases in resting tibialis anterior muscle of ageing wild-type mice<sup>2,4,9</sup>; this phenotype was not affected by *Hoxa9* gene status (Extended Data Fig. 5a). However, homozygous deletion or siRNA-mediated knockdown of *Hoxa9* increased the total number of Pax7<sup>+</sup> SCs (Fig. 2b, Extended Data Fig. 5b–e) and improved myofibre regeneration



**Figure 1 | Upregulation of *Hoxa9* in aged activated SCs.** **a–c**, Analysis of freshly isolated, *in vivo* activated SCs (3 days after muscle injury with BaCl<sub>2</sub>) from young adult and aged mice. **a**, Heatmap showing the mRNA expression of all Hox genes as determined by RNA-sequencing analysis. **b**, Representative immunofluorescence staining for Hoxa9 and Pax7. Nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI). **c**, Corrected total cell fluorescence (CTCF) for Hoxa9 per SC as shown in **b**. AU, arbitrary units. **d**, **e**, Immunofluorescence (IF) staining for Hoxa9 and Pax7 in myofibre-associated SCs that were quiescent (freshly isolated (FI) myofibres) or activated (act; 24 h culture of myofibres). **d**, Representative images with arrowheads denoting Pax7<sup>+</sup> cells. **e**, CTCF for indicated Hox genes. Note the specific induction of *Hoxa9* in activated SCs isolated from aged mice. Scale bars, 5  $\mu$ m (**b**) and 20  $\mu$ m (**d**). *P* values were calculated by two-sided Mann–Whitney *U*-test (**c**) or two-way analysis of variance (ANOVA) (**e**). NS, not significant.  $n = 3$  mice in **a**;  $n = 134$  nuclei (young),  $n = 181$  nuclei (aged) from 3 mice in **c**;  $n = 12/13/17/56$  nuclei (*Hoxa7*),  $n = 9/42/102/62$  nuclei (*Hoxa9*),  $n = 7/35/34/25$  nuclei (*Hoxb9*) from 2 young and 4 aged mice in **e**.

<sup>1</sup>Leibniz-Institute on Aging – Fritz Lipmann Institute (FLI), Beutenbergstrasse 11, 07745 Jena, Germany. <sup>2</sup>Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, Auguste-Piccard-Hof 1, 8093 Zürich, Switzerland. <sup>3</sup>Institute of Medical Systems Biology, Ulm University, James-Frank-Ring, 89081 Ulm, Germany. <sup>4</sup>Molecular and Computational Biology Program, University of Southern California, 1050 Childs Way, Los Angeles, California 90089, USA. <sup>5</sup>Department of Internal Medicine I, Ulm University, Albert-Einstein-Allee 23, 89081 Ulm, Germany. <sup>6</sup>Division of Epigenomics and Cancer Risk Factors, DKFZ, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. <sup>7</sup>Faculty of Science, University of Zürich, Zürich, Switzerland. <sup>8</sup>Department of Immunology, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, USA. <sup>9</sup>Faculty of Medicine, Friedrich-Schiller-University, Jena, Germany. §These authors jointly supervised this work.

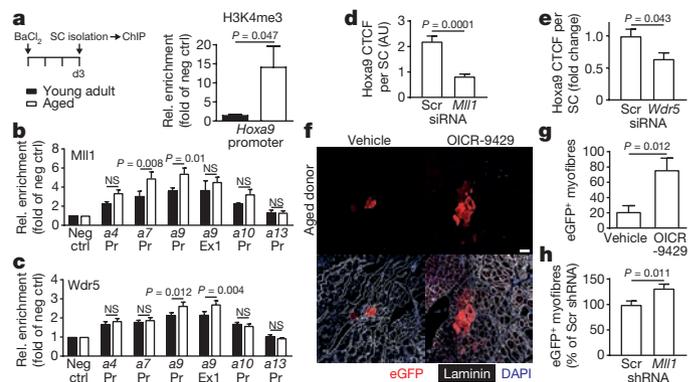


**Figure 2 | *Hoxa9* deficiency improves muscle regeneration in aged mice.**

**a**, Frequency of myogenic colonies derived from single-cell-sorted SCs from young adult or aged *Hoxa9*<sup>+/+</sup> and *Hoxa9*<sup>-/-</sup> mice after 5 days (d5) of culture. **b**, **c**, Quantification of Pax7<sup>+</sup> cells per area (**b**) and frequency distribution of minimal Feret's diameter (**c**) of tibialis anterior muscle fibres from aged *Hoxa9*<sup>+/+</sup> and *Hoxa9*<sup>-/-</sup> mice, 7 days after muscle injury with cardiotoxin (CTX). **d**, **e**, Transplantation (Tx) of enhanced green fluorescent protein (eGFP)-labelled SCs from aged *Hoxa9*<sup>+/+</sup> and *Hoxa9*<sup>-/-</sup> mice. **d**, Representative immunofluorescence staining for eGFP, laminin and DAPI in engrafted tibialis anterior muscles. Scale bar, 50  $\mu$ m. **e**, Quantification of donor-derived (eGFP<sup>+</sup>) myofibres in **d**. **f**, Quantification of donor-derived (eGFP<sup>+</sup>) SCs re-isolated from primary recipients. Scr, scrambled control shRNA. **g**, Quantification of donor-derived (eGFP<sup>+</sup>) myofibres from secondary recipients. Data in **f** represent median with 50% confidence interval box and 95% confidence interval whiskers. *P* values were calculated by two-way ANOVA (**a**, **c**), two-sided Student's *t*-test (**b**, **e**, **g**), or two-sided Mann–Whitney *U*-test (**f**). *n* = 4 mice in **a**; *n* = 4 mice in **b**, **c**; *n* = 8 recipient mice in **e**; *n* = 20 recipient mice in **f**; *n* = 5 recipient mice in **g**.

in injured muscle of aged mice almost to the levels in young adult mice (Fig. 2c, Extended Data Fig. 5f), albeit without affecting overall SC proliferation rates seven days after muscle injury (Extended Data Fig. 5g, h). *Hoxa9* gene deletion also improved the cell-autonomous, *in vivo* regenerative capacity of transplanted SCs derived from aged donor mice but did not affect the capacity of SCs derived from young adult donors (Fig. 2d, e, Extended Data Fig. 6a). Similarly, *Hoxa9* downregulation by short hairpin RNA (shRNA) infection rescued the regenerative capacity and the engraftment of transplanted SCs derived from aged mice almost to the level of SCs from young adult mice (Extended Data Fig. 6b–h). When transduced at similar infection efficiency (Extended Data Fig. 6i), *Hoxa9* shRNA compared to scrambled shRNA improved the self-renewal of serially transplanted SCs from aged mice in primary recipients (Fig. 2f, Extended Data Fig. 6j) as well as the regenerative capacity of 500 re-isolated SCs from primary donors that were transplanted for a second round into the injured tibialis anterior muscle of secondary recipients (Fig. 2g, Extended Data Fig. 6k). Together, these results demonstrate that the induction of *Hoxa9* limits SC self-renewal and muscle regeneration in aged mice, and that the deletion of *Hoxa9* is sufficient to revert these ageing-associated deficiencies.

The expression of *Hoxa9* in development and leukaemia is actively maintained by Mll1-dependent tri-methylation at lysine 4 of histone 3 (H3K4me3)<sup>16–18</sup>. Chromatin immunoprecipitation (ChIP) revealed that H3K4me3 is strongly enriched at the promoter and first exon of *Hoxa9* in activated SCs from aged compared to young adult mice, which was not detected to the same extent for other *Hoxa* genes (Fig. 3a, Extended Data Fig. 7a). ChIP analyses for Mll1 and Wdr5 (a scaffold protein of the Mll1 complex) revealed increased recruitment of these factors to the *Hoxa* cluster with Wdr5 enrichment being confined to the *Hoxa9* locus (Fig. 3b, c). Although no changes were

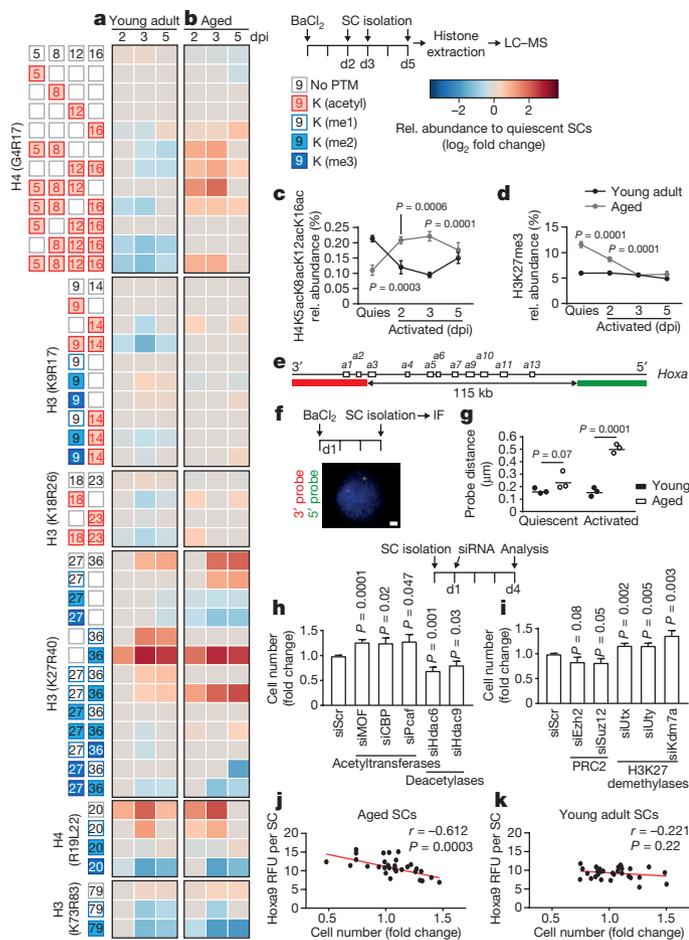


**Figure 3 | Mll1 complex-dependent chromatin modification induces *Hoxa9* and limits muscle regeneration in aged mice.**

**a–c**, ChIP-quantitative PCR (qPCR) analysis of the indicated *Hox* promoters (Pr) and exons (Ex) in activated SCs from young adult and aged mice using antibodies against H3K4me3 (**a**), Mll1 (**b**), or Wdr5 (**c**). **d**, **e**, CTFC for *Hoxa9* per SC after *Mll1* siRNA (**d**) or *Wdr5* siRNA transfection (**e**) of freshly isolated myofibre-associated SCs from aged mice. **f–h**, Transplantation of eGFP-labelled SCs from aged mice. **f**, Representative immunofluorescence staining for eGFP, laminin and DAPI in engrafted tibialis anterior muscles after transplantation of OICR-9429 treated SCs. Scale bar, 50  $\mu$ m. **g**, **h**, Quantification of donor-derived (eGFP<sup>+</sup>) myofibres after transplantation of OICR-9429-treated (**g**) or shRNA-treated (**h**) SCs. *P* values were calculated by two-way ANOVA (**b**, **c**), two-sided Student's *t*-test (**a**, **g**, **h**) or two-sided Mann–Whitney *U*-test (**d**, **e**). *n* = 6 mice in **a**; *n* = 7 mice (young), *n* = 10 mice (aged) in **b**, **c**; *n* = 109 nuclei (Scr siRNA), *n* = 110 nuclei (*Mll1* siRNA) from 3 mice in **d**; *n* = 116 nuclei (Scr siRNA), *n* = 65 nuclei (*Wdr5* siRNA) from 3 mice in **e**; *n* = 5 recipient mice in **g**; *n* = 6 recipient mice in **h**.

observed for Mll1, both H3K4me3 and Wdr5 showed significantly increased levels in nuclei of myofibre-associated SCs from aged versus young adult mice upon activation (Extended Data Fig. 7b–e). Of note, knockdown of either *Mll1* (also known as *Kmt2a*) or *Wdr5* reduced H3K4me3 levels as well as Mll1 recruitment to the *Hoxa9* locus and ameliorated *Hoxa9* induction in activated myofibre-associated SCs from aged mice (Fig. 3d, e, Extended Data Fig. 7f–i). Similar results were obtained by treatment of aged myofibre-associated SCs with OICR-9429, an inhibitor of the Mll1–Wdr5 interaction<sup>19</sup> (Extended Data Fig. 7j, k). Moreover, both *Mll1* knockdown and OICR-9429 treatment increased the self-renewal and lowered the myogenic commitment of myofibre-associated SCs from aged mice (Extended Data Fig. 7l–q), resulting in increased SC numbers in cultures of purified SCs or myofibre-associated SCs derived from aged mice (Extended Data Fig. 7r, s). Notably, Mll1 inhibition by either stable shRNA knockdown (Extended Data Fig. 7t) or OICR-9429 treatment improved the regenerative capacity of SCs from aged mice when transplanted into injured muscle of recipient mice (Fig. 3f–h). Taken together, these experiments demonstrate that the Mll1 complex contributes to *Hoxa9* induction in activated SCs from aged mice, resulting in impairment in SC function and muscle regeneration. Pax7 expression was downregulated in activated SCs of aged mice (Extended Data Fig. 7u–w) and did not correlate with *Hoxa9* expression (Extended Data Fig. 7x, y), indicating that Mll1-dependent regulation of Pax7 target genes<sup>20</sup> was not involved in the Mll1-dependent induction of *Hoxa9* in activated SCs from aged mice.

Next, a global analysis of histone post-translational modifications was carried out on freshly isolated SCs obtained before muscle injury (quiescent state) or two, three and five days after *in vivo* SC activation mediated by muscle injury (Fig. 4a, b, Extended Data Fig. 8a). Using a recently developed mass-spectrometry-based proteomic strategy<sup>21</sup>, 46 histone H3 and H4 lysine acetylation and methylation motifs were quantified. Quiescent SCs from aged mice compared to young adult mice showed increased levels of repressive marks (H3K9me2 and



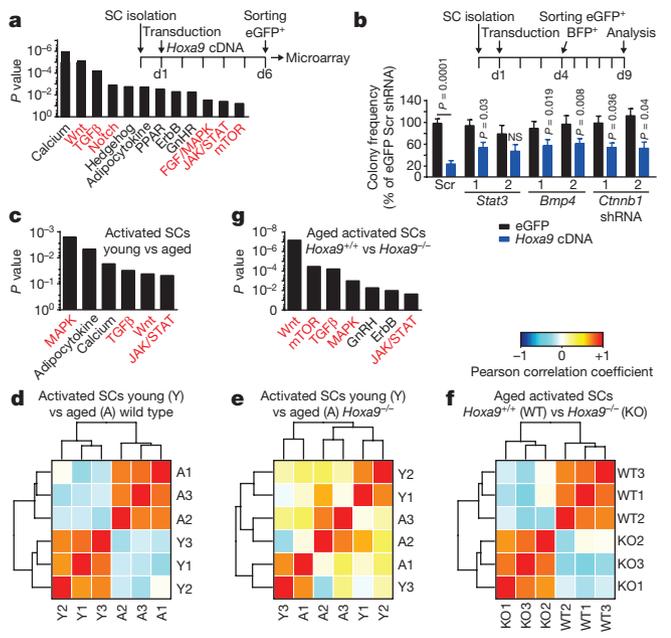
**Figure 4 | Altered epigenetic stress response in aged SCs.** **a, b**, Heatmap of mass spectrometry (LC-MS) analysis displaying significant ( $P < 0.05$ ) relative changes in abundance of the indicated histone modifications (measured at the indicated peptides) at the indicated days post injury (dpi). **c, d**, Trajectory time-course plots showing relative abundance of H4K5acK8acK12acK16ac (**c**) or H3K27me3 (**d**) in freshly isolated quiescent (quies) or *in vivo* activated SCs purified at indicated time points post muscle injury. **e–g**, Fluorescence *in situ* hybridization of freshly isolated quiescent or *in vivo* activated SCs with the indicated probes spanning the *Hoxa* cluster (**e**); an exemplary image (**f**); and the average probe distance (**g**). Scale bar, 1  $\mu\text{m}$ . **h, i**, Relative changes in SC number 4 days after transfection of freshly isolated SCs from aged mice with the indicated siRNAs. **j, k**, Pearson correlation of relative cell number and *Hoxa9* immunofluorescence signal of SCs from young adult and aged mice 4 days after transfection with a selection of siRNAs targeting different classes of chromatin modifiers. RFU, relative fluorescence units.  $P$  values were calculated by two-way ANOVA (**c, d, g**), two-sided Student's  $t$ -test (**a, b, h, i**), or Pearson correlation (**j, k**).  $n = 4$  mice in **a–d**;  $n = 3$  mice with 50 nuclei per replicate in **g**;  $n = 7$  mice (*Ezh2* siRNA), 8 mice (all others) in **h, i**;  $n = 6$  mice (aged),  $n = 3$  mice (young) in **j, k**.

H3K27me3; Extended Data Fig. 8a; consistent with ref. 22), and lower amounts of histone modifications typically enriched on active genes (for example, various H4 acetylation motifs, H3K14ac, H3K18ac and H3K36me2; Extended Data Fig. 8a). A time-dependent shift towards a heterochromatic state occurred during SC activation in young adult mice, whereas activation in aged SCs generated the opposite response (Fig. 4a, b). Although selective active marks such as H3 and H4 acetylation motifs declined in SCs from young adult mice during activation, there was a substantial increase in these marks in aged SCs (Fig. 4a–c). Conversely, repressive marks (for example, H3K27me3) decreased in SCs from aged mice but remained stable in SCs from young adult mice during activation (Fig. 4a, b, d). The observed shift of the chromatin towards a more permissive state after SC activation appeared to also

affect the *Hoxa* cluster as this locus displayed an increased chromatin decompaction after SC activation in aged mice but not in young adult mice (Fig. 4e–g).

To analyse the functional contribution of different types of chromatin modifications in activated SCs from aged mice, a set of genetic and pharmacological experiments was conducted. The expression of key enzymes involved in chromatin modifications detected by RNA-sequencing analysis was similar in activated SCs from young adult and aged mice (Extended Data Fig. 8b). However, knockdown of the acetyltransferases *MOF* (also known as *Kat8*), *CBP* (*Crebbp*) or *Pcaf* (*Kat2b*) improved the proliferative capacity of SCs from aged mice in bulk culture, whereas knockdown of histone deacetylases led to a reduction (Fig. 4h). Furthermore, knockdown of the H3K27 demethylases *Utx* (also known as *Kdm6a*), *Uty* or *Kdm7a* promoted the proliferation of aged SCs, which was instead inhibited by knockdown of *Suz12* and *Ezh2* (Fig. 4i), members of the PRC2 protein complex responsible for H3K27me3. Multi-acetylation motifs, as observed in activated SC from aged mice (Fig. 4b, c), are preferred binding sites for bromodomain-containing proteins<sup>23</sup>. Eight out of eleven non-toxic bromodomain inhibitors available from the Structural Genomics Consortium exhibited positive effects on the proliferative capacity of SCs from aged mice (Extended Data Fig. 8c, d,  $P = 4.2 \times 10^{-4}$ ). Targeting major classes of chromatin modifiers by a selected set of siRNAs (Supplementary Table 1) revealed a significant inverse correlation ( $r = -0.612$ ) between siRNA-mediated changes in *Hoxa9* protein expression and the proliferative capacity of SCs from aged mice, with no such effects observed in SCs from young adult mice (Fig. 4j, k). Similarly, siRNAs against *MOF* and *Utx* as well as bromodomain inhibitors led to significant decreases in the *Hoxa9* protein level in activated myofibre-associated SCs from aged mice (Extended Data Fig. 8e–g). In summary, activated SCs from aged mice exhibit site-specific and global aberrations in the epigenetic stress response, resulting in *Hoxa9* activation and profound negative effects on SC function, which are ameliorated by targeting the respective enzymes underlying these alterations.

By analysing the downstream effects of *Hoxa9* induction through lentiviral-mediated *Hoxa9* overexpression, we found a strong reduction in the colony forming and proliferative capacity of SCs from young adult mice (Extended Data Fig. 9a–c). The overexpression of other *Hox* genes exerted similar effects (Extended Data Fig. 9d) but the *Hoxa9* results are probably most relevant for physiological ageing because only *Hoxa9* was upregulated in activated SCs from aged mice (Fig. 1). The impaired myogenic capacity of SCs in response to *Hoxa9* overexpression was associated with increased rates of apoptosis and decreased cell proliferation (Extended Data Fig. 9e–h). Furthermore, *Hoxa9* induction associated with the suppression of several cell cycle regulators and induction of cell cycle inhibitors and senescence-inducing genes (Extended Data Fig. 9i) as well as with increased staining for senescence-associated  $\beta$ -galactosidase (Extended Data Fig. 9j, k). Microarray expression analysis of *Hoxa9*-overexpressing SCs compared to controls revealed that among the top 12 pathways regulated by *Hoxa9* were several major developmental pathways that have previously been shown to impair SC function and muscle regeneration in the context of ageing<sup>2,3,5,6,9,24,25</sup> (Fig. 5a, Extended Data Fig. 9l–o). ChIP analysis of putative *Hoxa9*-binding sites (Supplementary Table 1) in *Hoxa9*-overexpressing primary myoblasts indicated that a high number of these genes are probably direct targets of *Hoxa9* (Extended Data Fig. 9p; cumulative  $P$  value over tested genes:  $P = 1 \times 10^{-7}$ ). *Hoxa9* strongly induced downstream targets of the Wnt, TGF $\beta$  and JAK/STAT pathways, but targeted activation of each one of these pathways alone only led to slight changes in the expression of target genes of the other two pathways (Extended Data Fig. 9q–s), suggesting that *Hoxa9* acts as a central hub required for the parallel induction of these pathways in aged SCs. Of note, the inhibition of *Stat3*, *Bmp4* or *Cttnb1* (encoding  $\beta$ -catenin) by shRNAs as well as pharmacological inhibition of the Wnt, TGF $\beta$  or JAK/STAT pathway was sufficient to improve the myogenic colony forming capacity of SCs overexpressing *Hoxa9* (Fig. 5b,



**Figure 5 | Activation of *Hoxa9* induces developmental pathways.** **a**, KEGG pathway analysis of differentially expressed genes (DEGs) of SCs overexpressing *Hoxa9* compared to eGFP. Red-highlighted pathways were previously shown to impair the function of SCs in aged mice. **b**, Colony formation of single-cell-sorted SCs derived from young adult mice that were co-infected with a *Hoxa9* cDNA and the indicated shRNAs; comparison to *Hoxa9*/scrambled shRNA co-infected cells. **c, g**, KEGG analysis of DEGs from indicated transcriptomes. **d-f**, Heatmaps displaying Pearson correlation analysis of indicated transcriptomes. *P* values were calculated by two-way ANOVA (**b**). *n* = 4 pools of 3 mice in **a**; *n* = 6 mice (*Stat3* shRNA1/2), *n* = 7 mice (all others) in **b**; *n* = 3 mice per group in **c-f**.

Extended Data Fig. 10a, b). In line with previous results, knockdown of *Stat3* also increased the total number and lowered early differentiation of myofibre-associated SCs from aged mice, and in addition, increased the regenerative capacity of transplanted SCs from aged mice to a similar extent as *Hoxa9* knockdown (Extended Data Fig. 10c-g).

Differentially expressed genes were determined using RNA-sequencing data of freshly isolated, *in vivo* activated SCs from young adult and aged wild-type mice as well as from aged *Hoxa9*<sup>-/-</sup> mice. There was a highly significant overlap between genes induced by *Hoxa9* overexpression in SCs from young adult mice with those genes that were dysregulated in *in vivo* activated SCs from aged compared to young adult mice (*P* = 2.2 × 10<sup>-19</sup>; Extended Data Fig. 10h). Pathways that are currently known to be associated with SC ageing were again among the highest ranked pathways differentially expressed in activated SCs from aged compared to young adult mice including MAPK, TGFβ, Wnt and JAK/STAT signalling (Fig. 5c). Of note, *Hoxa9* deletion abrogated the separate clustering of gene expression profiles of activated SCs from aged compared to young adult mice (Fig. 5d, e). Comparing transcriptomes of activated SC from aged *Hoxa9*<sup>+/+</sup> to aged *Hoxa9*<sup>-/-</sup> mice re-established the separate clustering (Fig. 5f) characterized by enrichment of the same set of developmental pathways that associate with SC ageing in wild-type mice (Fig. 5g, compare to Fig. 5c).

Taken together, the current study provides experimental evidence that an aberrant epigenetic stress response impairs the functionality of SCs from aged mice by *Hoxa9*-dependent activation of developmental signals (Extended Data Fig. 10i). Notably, a proof of concept is provided that key enzymes that promote global and site-specific alterations in the epigenetic stress response of aged SCs are druggable, and that the inhibition of these targets leads to improvement in SC function and muscle regeneration during ageing. These findings provide experimental support for the recent hypothesis that a 'shadowed' dysregulation

of developmental pathways represents a driving force of stem-cell and tissue ageing<sup>26,27</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 October 2015; accepted 3 November 2016.

Published online 30 November 2016.

- Rando, T. A. Stem cells, ageing and the quest for immortality. *Nature* **441**, 1080–1086 (2006).
- Brack, A. S. *et al.* Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. *Science* **317**, 807–810 (2007).
- Carlson, M. E. *et al.* Relative roles of TGF-β1 and Wnt in the systemic regulation and aging of satellite cell responses. *Aging Cell* **8**, 676–689 (2009).
- Sousa-Victor, P. *et al.* Geriatric muscle stem cells switch reversible quiescence into senescence. *Nature* **506**, 316–321 (2014).
- Conboy, I. M., Conboy, M. J., Smythe, G. M. & Rando, T. A. Notch-mediated restoration of regenerative potential to aged muscle. *Science* **302**, 1575–1577 (2003).
- Price, F. D. *et al.* Inhibition of JAK-STAT signaling stimulates adult satellite cell function. *Nat. Med.* **20**, 1174–1181 (2014).
- Krumlauf, R. Hox genes in vertebrate development. *Cell* **78**, 191–201 (1994).
- Lawrence, H. J., Sauvageau, G., Humphries, R. K. & Largman, C. The role of HOX homeobox genes in normal and leukemic hematopoiesis. *Stem Cells* **14**, 281–291 (1996).
- Cosgrove, B. D. *et al.* Rejuvenation of the muscle stem cell population restores strength to injured aged muscles. *Nat. Med.* **20**, 255–264 (2014).
- Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
- Lyons, K. M., Pelton, R. W. & Hogan, B. L. Organogenesis and pattern formation in the mouse: RNA distribution patterns suggest a role for bone morphogenetic protein-2A (BMP-2A). *Development* **109**, 833–844 (1990).
- Muñoz-Espín, D. *et al.* Programmed cell senescence during mammalian embryonic development. *Cell* **155**, 1104–1118 (2013).
- Sehgal, P. B., Levy, D. E. & Hirano, T. *Signal Transducers And Activators Of Transcription (STATs): Activation And Biology* (Kluwer Academic, 2003).
- Sinha, M. *et al.* Restoring systemic GDF11 levels reverses age-related dysfunction in mouse skeletal muscle. *Science* **344**, 649–652 (2014).
- Bernet, J. D. *et al.* p38 MAPK signaling underlies a cell-autonomous loss of stem cell self-renewal in skeletal muscle of aged mice. *Nat. Med.* **20**, 265–271 (2014).
- Soshnikova, N. & Duboule, D. Epigenetic temporal control of mouse Hox genes *in vivo*. *Science* **324**, 1320–1323 (2009).
- Ayton, P. M. & Cleary, M. L. Transformation of myeloid progenitors by MLL oncoproteins is dependent on *Hoxa7* and *Hoxa9*. *Genes Dev.* **17**, 2298–2307 (2003).
- Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. & Korsmeyer, S. J. Altered Hox expression and segmental identity in *Mll*-mutant mice. *Nature* **378**, 505–508 (1995).
- Grebien, F. *et al.* Pharmacological targeting of the Wdr5-MLL interaction in C/EBPα N-terminal leukemia. *Nat. Chem. Biol.* **11**, 571–578 (2015).
- McKinnell, I. W. *et al.* Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. *Nat. Cell Biol.* **10**, 77–84 (2008).
- Feller, C., Forné, I., Imhof, A. & Becker, P. B. Global and specific responses of the histone acetylome to systematic perturbation. *Mol. Cell* **57**, 559–571 (2015).
- Liu, L. *et al.* Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. *Cell Reports* **4**, 189–204 (2013).
- Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214–231 (2012).
- Tierney, M. T. *et al.* STAT3 signaling controls satellite cell expansion and skeletal muscle repair. *Nat. Med.* **20**, 1182–1186 (2014).
- Chakkalakal, J. V., Jones, K. M., Basson, M. A. & Brack, A. S. The aged niche disrupts muscle stem cell quiescence. *Nature* **490**, 355–360 (2012).
- Blagosklonny, M. V. Aging is not programmed: genetic pseudo-program is a shadow of developmental growth. *Cell Cycle* **12**, 3736–3742 (2013).
- Martin, N., Beach, D. & Gil, J. Ageing as developmental decay: insights from p16<sup>INK4a</sup>. *Trends Mol. Med.* **20**, 667–674 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank Y. Morita and A. Illing for providing guidance regarding FACS analysis. We are thankful to the FLI Core Facilities Functional Genomics (T. Kroll, A. Ploubidou) and DNA Sequencing (M. Groth) for their services. We express our thanks to M. Burkhalter, T. Sperka and A. Illing for discussions and suggestions. We are grateful to B. Wollscheid and S. Goetze for providing support for proteomic measurements. We thank V. Sakk and M. Kettering for mouse husbandry as well as S. Eichwald, K. Tramm and A. Abou Seif for experimental assistance. We are grateful to M. Kessel, M. Kyba, G. Sauvageau and D. Wellik for sharing plasmids with *Hox* cDNAs. We thank the Structural Genomics Consortium and S. Ackloo for providing access to the epigenetic probe library. We further thank M. Cerletti for providing protocols on SC isolation and E. Perdiguer for advice on infection of SCs before

transplantation. Work on this project in K.L.R.'s laboratory was supported by the DGF (RU-745/10, RU-745/12), the ERC (2012-AdG 323136), the state of Thuringia, and intramural funds from the Leibniz association. J.V.M. was supported by a grant from the DFG (MA-3975/2-1). C.F. acknowledges support by the DFG (FE-1544/1-1) and EMBO (long-term postdoctoral fellowship ALTF 55-2015). R.A. was supported by the ERC (AdvGr 670821 (Proteomics 4D)). The funding for the *Hoxa9*<sup>-/-</sup> mice to K.L.M. was provided by a grant of the NIH (HL096108). R.R. was supported by a grant from the NIH (R01GM106056). This work was further supported by grants to H.A.K. from the DFG (SFB 1074 project Z1), the BMBF (Gerontosys II, Forschungskern SyStaR, project ID 0315894A), and the European Community's Seventh Framework Programme 390 (FP7/2007-2013, grant agreement 602783).

**Author Contributions** S.S. designed and performed most experiments, analysed data, interpreted results and wrote the manuscript. F.B. designed and performed RNAi, ChIP and FISH experiments on isolated SCs, analysed data, interpreted results and wrote the manuscript. C.F. and R.A. designed and performed LC-MS experiments, analysed data, interpreted results and wrote the manuscript. A.H.B., U.K., H.H., C.S.V. and M.S. performed individual experiments

and analysed data. A.L. performed microarray experiments. D.B.L. provided support and suggestions for ChIP experiments. K.L.M. provided *Hoxa9*<sup>-/-</sup> mice. J.M.K. and H.A.K. performed microarray and pathway analysis, analysed putative *Hoxa9*-binding sites and provided support for statistical analysis. B.X. and R.R. conducted analysis of putative *Hoxa9*-binding sites. F.N. analysed RNA-sequencing data and performed correlation analysis. J.V.M. and S.T. conceived the project, designed and performed individual experiments, interpreted results and wrote the manuscript. K.L.R. conceived the project, designed experiments, interpreted results and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.V.M. ([julia.vonmaltzahn@leibniz-fli.de](mailto:julia.vonmaltzahn@leibniz-fli.de)), S.T. ([stefan.tuempel@leibniz-fli.de](mailto:stefan.tuempel@leibniz-fli.de)) or K.L.R. ([lenhard.rudolph@leibniz-fli.de](mailto:lenhard.rudolph@leibniz-fli.de)).

**Reviewer Information** *Nature* thanks J. Gil and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Data reporting.** No statistical methods were used to estimate sample size. No randomization was used. No animals were excluded. The evaluator was blinded to the identity of the specific sample as much as the nature of the experiment allowed it.

**Mice.** We purchased female young adult C57/BL6j mice (3–4 months) and aged C57/BL6j mice (22–28 months) from Janvier (wild-type mice). Female and male *Hoxa9*<sup>-/-</sup> mice have been described<sup>28</sup> and were obtained together with age- and gender-matched littermate controls from K. L. Medina. Mice were housed in a pathogen-free environment and fed with a standard diet *ad libitum*. Animal experiments were approved by the Thüringer Landesamt für Verbraucherschutz (Germany) under Reg.-Nr. 03-006/13, 03-012/13 and 03-007/15 and by the Regierungspräsidentium Tübingen (Germany) under Reg.-Nr. 35/9185.81-3/919.

**Muscle injury.** Mice were anaesthetized using isoflurane in air and oxygen through a nose cone. For SC activation, muscles were injured by injecting a total volume of 50  $\mu$ l of 1.2% BaCl<sub>2</sub> (Sigma) into approximately 20 sites in the hindlimb muscles. For regeneration and transplantation experiments, tibialis anterior muscle of the right leg was injected with 50  $\mu$ l cardiotoxin (CTX, 10  $\mu$ M, Sigma).

**SC isolation and FACS.** Muscles from hindlimbs from young adult or aged mice were dissected and collected in PBS on ice. Muscles were rinsed with PBS, minced with scissors and incubated in DMEM with Collagenase (0.2%, Biochrom) for 90 min at 37 °C and 70 r.p.m. Digested muscles were washed with 10% FBS in PBS, triturated and incubated in Collagenase (0.0125%) and Dispase (0.4%, Life Technologies) for 30 min at 37 °C and 100 r.p.m. The muscle slurry was diluted with 10% FBS in PBS, filtered through 100- $\mu$ m cell strainers and spun down at 500g for 5 min. Cell pellets were resuspended in FACS buffer (2% FBS in HBSS) and filtered through 40- $\mu$ m cell strainers and pelleted at 500g for 5 min. Pellets were resuspended in FACS buffer and stained with anti-mouse CD45 PE conjugate (30-F11, eBioscience), anti-mouse CD11b PE conjugate (M1/70, eBioscience), anti-mouse Sca-1 PE conjugate (D7, BioLegend), anti-mouse CD31 PE/Cy7 conjugate (390, BioLegend) and anti-mouse  $\alpha$ 7-integrin Alexa Fluor 647 conjugate (R2F2, AbLab) for 20 min at 4 °C on a rotating wheel. Cells were washed with FACS buffer. Live cells were identified as calcein blue positive (1:1,000, Invitrogen) and propidium iodide negative (PI, 1  $\mu$ g ml<sup>-1</sup>, BD Biosciences). SCs were identified as CD45<sup>-</sup>Sca-1<sup>-</sup>CD11b<sup>-</sup>CD31<sup>-</sup> $\alpha$ 7-integrin<sup>+</sup>. Cell sorting was performed on a FACSARIAIII with Diva Software (BD).

**Culture of SCs.** SCs and SC-derived primary myoblasts were cultured at 37 °C, 5% CO<sub>2</sub>, 3% O<sub>2</sub> and 95% humidity in growth medium on collagen/laminin-coated tissue culture plates for the indicated time periods. Growth medium was comprised of F10 (Life Technologies) with 20% horse serum (GE), 1% penicillin/streptomycin (Life Technologies) and 5 ng ml<sup>-1</sup> bFGF (Sigma). For coating, tissue culture plates were incubated with 1 mg ml<sup>-1</sup> collagen (Sigma) and 10 mg ml<sup>-1</sup> laminin (Life Technologies) in ddH<sub>2</sub>O for at least 1 h at 37 °C and allowed to air-dry. For passaging or FACS analysis, cultured cells were incubated with 0.5% trypsin in PBS for 3 min at 37 °C and collected in FACS buffer. Treatment of SCs with noggin (Preprotech) or DKK1 (Preprotech) was done at 100 ng ml<sup>-1</sup> concentration. SCs and SC-derived primary myoblasts were treated with 1  $\mu$ M of chemical probes provided by the Structural Genomics Consortium (SGC, <http://www.thesgc.org/chemical-probes/epigenetics>)<sup>29,30</sup>. OICR-9429 and bromodomain inhibitors were described previously<sup>19,31–39</sup>.

**Clonal myogenesis assay.** Freshly isolated SCs from young adult and aged mice were sorted in growth medium in 96-well plates using the automated cell deposition unit of the FACSARIAIII. After 5 days, wells containing myogenic colonies were counted by brightfield microscopy. For clonal analysis of lentivirus-transduced SCs, infected (eGFP<sup>+</sup> and/or BFP<sup>+</sup>) live (DAPI<sup>-</sup>) cells were sorted as one cell per well in growth medium and wells containing myogenic colonies were counted by fluorescence microscopy (Axio Observer, Zeiss) after 5 days. A colony was defined by the presence of at least two cells.

**Alamar blue assay.** SCs or SC-derived primary myoblasts were seeded at 500 cells per well in growth medium into 96-well plates. After 4 days of culture, the viability was measured by adding Alamar Blue (Life Technologies) as 10% of the sample volume. Cells were incubated for 2 h at 37 °C and fluorescence intensity was measured at an excitation/emission wavelength of 560/590 nm.

**BrdU assay.** SCs were incubated with 5  $\mu$ M BrdU (Sigma) in growth medium for 2 h. Cells were fixed with 4% PFA, permeabilized with 0.5% Triton X-100 and incubated with 2 N HCl/PBS for 30 min at room temperature. Incorporated BrdU was detected using anti-BrdU (347580, BD Biosciences) and Alexa-594 fluorochrome (Life Technologies) for 1 h at room temperature. Nuclei were counterstained with DAPI in PBS.

**TUNEL assay.** TUNEL assay was performed using the *In situ* Cell Death Detection Kit, Texas Red (Roche) according to the manufacturer's instructions.

**Senescence-associated  $\beta$ -galactosidase assay.** SCs were fixed in 4% PFA and stained with staining solution (5 mM potassium ferricyanide, 5 mM potassium

ferricyanide, 2 mM MgCl<sub>2</sub>, 150 mM NaCl, 1 mg ml<sup>-1</sup> X-Gal) in citrate/sodium-phosphate buffer (pH 6) overnight at 37 °C. Staining solution was removed by rinsing several times with PBS.

**Myofibre isolation and culture.** Individual myofibres were isolated from the extensor digitorum longus muscle as described previously<sup>40,41</sup>. Isolated myofibres were cultured in DMEM containing 20% FBS and 1% chicken embryo extract (Biomol) in dishes coated with horse serum. Freshly isolated fibres or fibres cultured for 24–34 h and 72 h were fixed with 2% PFA and subjected to immunofluorescence analysis. Clusters of SCs were counted on at least 10–15 fibres per replicate. A cluster was defined by the presence of at least three adjacent cells. For quantification of immunofluorescence staining of myofibre-associated quiescent and activated SCs, at least 20 fibres were analysed per replicate. Treatment of myofibre-associated SCs with chemical probes provided by the Structural Genomics Consortium (SGC) was done 4 h after isolation at 1  $\mu$ M concentration.

**siRNA transfection.** Transfection of SCs was performed in a reverse manner: SCs were seeded in growth medium into individual wells of a 384-well plate pre-filled with transfection mix. For floating cultures of single myofibres, transfections were performed 4 h after isolation in myofibre culture medium. Transfections were done using Lipofectamin RNAiMAX (Life Technologies) according to manufacturer's instructions. For gene knockdown either Silencer Select siRNAs (Life Technologies) or ON-TARGETplus siRNA SMART-pools (Dharmacon) were used. Respective Silencer Select or ON-TARGETplus SMART-pool non-targeting siRNAs were used as negative control. siRNA sequences are listed in Supplementary Table 1. Transfection efficiency was monitored using a Cy3-labelled control siRNA (Life Technologies). After transfection, FACS-sorted SCs or myofibre-associated SCs were cultured for the indicated time periods and fixed in 2% PFA in PBS. *In vivo* knockdown experiments were performed as described earlier<sup>41</sup>. siRNA sequences were modified to the Accell self-delivering format (Dharmacon) and 100  $\mu$ g Accell siRNA were injected into tibialis anterior muscle 2 days after CTX injury. *In vivo* knockdown was evaluated from SCs isolated from injected tibialis anterior muscle 3 days after transfection. Transfected muscles were collected 5 days after siRNA injection, frozen in 10% sucrose/OCT in liquid nitrogen and stored at –80 °C.

**Lentivirus production and transduction.** Lentivirus was produced in Lenti-X cells (Clontech) after co-transfection of 15  $\mu$ g shRNA or cDNA plasmid, 10  $\mu$ g psPAX2 helper plasmid and 5  $\mu$ g pMD2.G according to standard procedures<sup>42</sup>. Virus was concentrated by centrifugation for 2.5 h at 106,800g and 4 °C, and virus pellet was resuspended in sterile PBS. Lentiviral transduction was carried out in growth medium supplemented with 8  $\mu$ g ml<sup>-1</sup> polybrene (Sigma).

**Plasmids.** cDNA was inserted into the SF-LV-cDNA-eGFP plasmid<sup>43</sup>. Primers used for cloning of individual *Hox* cDNAs are listed in Supplementary Table 1. shRNA was inserted into the SF-LV-shRNA-eGFP plasmid using mir30 primers (Supplementary Table 1). shRNA sequences are listed in Supplementary Table 1.

**SC transplantation.** SCs were FACS purified and transduced with a lentivirus on Retrofectin (Takara) coated 48-well plates<sup>4</sup>. After 8–10 h, SCs were obtained by resuspension and washed several times with FACS buffer. For each engraftment, 10,000 SCs were resuspended in 0.9% NaCl and immediately transplanted into tibialis anterior muscles of adult immunosuppressed mice that had been injured with CTX 2 days before. Immunosuppression with FK506 (5 mg kg<sup>-1</sup> body weight, Sigma) was started at the day of injury using osmotic pumps (model 2004, Alzet) and maintained throughout the entire time of engraftment. Engrafted muscles were collected 3 weeks after transplantation and fixed in 4% PFA for 30 min at room temperature followed by incubation in 30% sucrose/PBS overnight at 4 °C. Fixed muscles were frozen in 10% sucrose/OCT in liquid nitrogen and stored at –80 °C.

**Immunohistochemistry.** Cryosections of 10  $\mu$ m were cut from frozen muscle using the Microm HM 550. Cryosections were rinsed once with PBS and fixed in 2% PFA in PBS for 5 min at room temperature. Sections were rinsed three times for 5 min with PBS, permeabilized with 0.5% Triton X-100/0.1 M glycine in PBS for 5 min at room temperature followed again by rinsing them three times with PBS. Sections were blocked in PBS supplemented with 5% horse serum and 1:40 mouse on mouse blocking reagent (Vector labs) for 1 h at room temperature. Incubation with primary antibodies was carried out overnight at 4 °C. The next day, sections were rinsed three times with PBS followed by incubation with secondary antibodies for 1 h at room temperature. Sections were rinsed again with PBS and nuclei were counterstained with 1:1,000 DAPI in PBS before mounting with Permafluor (Thermo Scientific). Slides were stored at 4 °C until analysis. The following primary antibodies were used: 1:1,000 chicken anti-GFP (ab6556, AbCam), 1:1,000 rabbit anti-laminin (L9393, Sigma), 1:200 rabbit anti-Ki67 (ab15580, AbCam), undiluted mouse anti-Pax7 (DSHB). The following secondary antibodies were used at 1:1,000: anti-chicken IgG Alexa-Fluor 488, anti-rabbit IgG Alexa-Fluor 488, anti-mouse IgG1 Alexa-Fluor 594 (Life Technologies).

**Immunofluorescence.** Freshly isolated SCs were allowed to settle on poly-L-lysine-coated diagnostic microscope slides for 30 min at room temperature. All cells and myofibres were fixed with 2% PFA, permeabilized with 0.5% Triton X-100 and blocked with 10% horse serum in PBS for 1 h at room temperature. Cells and fibres were stained with primary antibodies in blocking solution overnight at 4°C. Samples were washed three times with PBS and incubated with secondary antibodies for 1 h at room temperature. Nuclei were counterstained with DAPI. Cultured cells were kept in PBS; freshly isolated SCs and myofibres were mounted with Permafluor. The following primary antibodies were used: undiluted mouse anti-Pax7 (DSHB), 1:300 rabbit anti-Hoxa9 (07-178, Millipore), 1:500 mouse anti-Mll1 (05-765, Millipore), 1:500 rabbit anti-Wdr5 (A302-429A, Bethyl Laboratories), 1:300 rabbit anti-H3K4me3 (C15410003-50, Diagenode), 1:200 rabbit anti-MyoD (sc-304, Santa Cruz). The following secondary antibodies were used at 1:1,000: anti-rabbit IgG Alexa-Fluor 488, anti-mouse IgG Alexa-Fluor 594, anti-mouse IgG1 Alexa-Fluor 594 (Life Technologies).

**Fluorescence *in situ* hybridization (FISH).** Chromatin compaction FISH was done as described previously<sup>44</sup>. DNA of the 3' and 5' probe (Fosmid clones WIBR1-1312N03 and WIBR1-2209G09, CHORI) was labelled with digoxigenin or biotin by nick-translation (Roche). 100 ng of probe DNA was used per slide, together with 5 µg mouse CotI DNA (Life Technologies) and 5 µg single-stranded DNA (Ambion). Approximately 5,000 freshly sorted SCs were allowed to settle on poly-L-lysine-coated diagnostic microscope slides for 30 min at room temperature and were fixed with 2% PFA for 5 min. After washing three times with PBS, slides were incubated with 0.1 M HCl for 5 min and permeabilized with 0.5% Triton X-100 in 0.5% saponin for 10 min before freeze-thaw in 20% glycerol in PBS. Denaturation was performed in 50% formamide, 1% Tween-20 and 10% dextran sulfate/2× SSC for 5 min at 75°C before applying the hybridization cocktail. Probes were hybridized overnight at 37°C in a humidified chamber. Slides were rinsed three times with 2× SSC, blocked with 2% BSA in 0.1% Tween-20 in PBS for 1 h at room temperature, and hybridized probes were visualized with anti-digoxigenin-rhodamine (S7165, Millipore) and Streptavidin-Cy2 (016-220-084, IR USA) for 30 min at room temperature. Nuclei were counterstained with DAPI.

**Digital image acquisition and processing.** Immunofluorescence images of muscle sections, myofibres and freshly isolated SCs were acquired using the upright microscope Axio Imager (Zeiss) with 10×, 20× and 100× objectives and a monochrome camera. Brightfield and immunofluorescence images of cultured SCs were captured using the microscope Axio Observer (Zeiss) with 5×, 10× and 20× objectives and a monochrome camera. Image acquisition and processing was performed using the ZEN 2012 software (Zeiss). Brightness and contrast adjustments were applied to the entire image before the region of interest was selected. For the analysis of muscle sections, several images covering the whole area of the section were acquired in a rasterized manner and assembled in Photoshop CS6 (Adobe) to obtain an image of the entire section. Images were analysed using ImageJ software. The number of Pax7<sup>+</sup> cells in regeneration experiments was normalized to the area of the entire muscle section. CTCF was determined for each SC using the calculation: integrated density – (area of selected cell × mean fluorescence of background readings) (ref. 45).

**RNA isolation and reverse transcription.** Total RNA was isolated from freshly FACS-isolated or cultured SCs by using the MagMAX 96 total RNA Isolation Kit (Ambion) according to the manufacturer's protocol. The GoScript Reverse Transcription System (Promega) was used for cDNA synthesis from total RNA according to manufacturer's instructions.

**ChIP.** 5 × 10<sup>4</sup>–1 × 10<sup>5</sup> cells were crosslinked in 1% formaldehyde (Thermo Scientific) for 10 min. Crosslinking was quenched with glycine and cells were washed two times with ice-cold PBS. For ChIP of H3K4me3, cells were lysed in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1, 1× Roche cComplete Protease Inhibitor) and chromatin was sonicated in Snap Cap microTUBEs using a Covaris M220 sonicator to a fragment size of 150–300 bp. Chromatin was cleared for 10 min at 17,000g, and one-tenth of the chromatin was removed as input fraction. Chromatin was immunoprecipitated overnight with 20 µl Protein A/G bead mix (1:1, Dynabeads, Invitrogen) pre-coupled with 1 µg antibody (C15410003-50, Diagenode) in ChIP-dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl, 16.7 mM Tris-HCl pH 8.1, 1× Roche cComplete Protease Inhibitor). Beads were washed three times with low-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, Tris-HCl, pH 8.1) and three times with LiCl buffer (350 mM LiCl, 1% IPEGAL CA630, 1% deoxycholic acid, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1). For ChIP of Mll1, Wdr5 or haemagglutinin (HA)-tagged Hoxa9 cells were resuspended in sonication buffer (0.1% SDS, 1% Triton X-100, 0.1% Na-deoxycholate, 1 mM EDTA, 140 mM NaCl, 50 mM HEPES, pH 7.9), incubated on ice for 10 min and sonicated to a fragment size of 300–600 bp as described above. Chromatin was cleared for 10 min at 17,000g and unspecific binding was absorbed with 5 µl of Protein G beads for 1 h. One-tenth (Mll1/Wdr5) or one-twentieth

(HA-tag) of the chromatin was removed as input fraction. Chromatin was immunoprecipitated overnight with 2 µg of antibody (Mll1: A300-086A, Wdr5: A302-429A, Bethyl Laboratories; HA-tag: ab9110, Abcam). Chromatin-antibody complexes were captured with 20 µl Protein A/G bead mix (1:1, Dynabeads, Invitrogen) for 2 h. Beads were washed twice with sonication buffer, twice with NaCl buffer (0.1% SDS, 1% Triton X-100, 0.1% Na-deoxycholate, 1 mM EDTA, 500 mM NaCl, 50 mM HEPES, pH 7.9), twice with LiCl buffer and once with TE buffer. Decrosslinking and elution was performed in 50 µl decrosslinking buffer (1% SDS, 100 mM NaHCO<sub>3</sub>, 250 mM NaCl) for 4 h at 65°C with continuous shaking and subsequent Proteinase K treatment for 1 h at 45°C. DNA was purified using Agencourt AMPure XP beads (Beckman Coulter) with a beads:sample ratio of 1.8:1 or MinElute PCR Purification Kit according to manufacturer's protocols. **Quantitative PCR.** Quantitative PCR (qPCR) was performed with an ABI 7500 Real-Time PCR System (Applied Biosystems) in technical duplicates from the indicated number of biological replicates. The qPCR was carried out in a volume of 12 µl using the Absolute qPCR Rox Mix (Thermo Scientific) and the Universal Probe Library (Roche). Primer and probe sets for the detection of single genes are listed in Supplementary Table 1. *Gapdh* was detected with rodent *Gapdh* control reagents (Applied Biosystems). Relative expression values were calculated using the  $\Delta C_t$  method.

$$\Delta C_t = C_t[\text{gene of interest}] - C_t[\text{Gapdh}]$$

$$\text{Relative expression} = 2^{(-\Delta C_t)}$$

qPCR analysis of ChIP samples was performed using SYBR Green Supermix (Biorad) in a final reaction volume of 10 µl and 0.75 µM final primer concentration. Primers are listed in Supplementary Table 1. HA-tag ChIP signals were calculated as percentage of the input fraction. The  $\Delta\Delta C_t$  method was used to calculate fold enrichment of a genomic locus over the ChIP specific background control (*Actb* intergenic region for H3K4me3 or gene desert for *Mll1* and *Wdr5*), both normalized to the signal in the input fraction:

$$\Delta C_t[\text{normalized to input}] = (C_t[\text{ChIP}] - (C_t[\text{input}] - \log_2(\text{input dilution factor})))$$

$$\Delta\Delta C_t = \Delta C_t[\text{region of choice normalized to input}] - \Delta C_t[\text{control region normalized to input}]$$

$$\text{Fold enrichment} = 2^{(-\Delta\Delta C_t)}$$

**Nanostring analysis.** Pellets of freshly isolated SCs were lysed with 3 µl RLT buffer (QIAGEN) and subjected to Nanostring analysis according to manufacturer's instructions using a custom-made Hox gene nCounter Elements TagSet (Nanostring Technologies). Relative expression to the housekeeping genes *Gapdh*, *Hmbs* and *Polr2a* was calculated using nSolver Software (v2.0) after background correction and normalization to hybridized probe signals.

**Proteomic analysis of histone modifications.** Preparation of histones for mass spectrometry, data acquisition and analysis were essentially performed as described previously<sup>21</sup> with modifications described below. In brief, histones were isolated by acid extraction, derivatised by d6-acetic anhydride (CD<sub>3</sub>CO, Aldrich) and digested with sequencing-grade trypsin (Promega) overnight at a trypsin:protein ratio of 1:20. To acetylate free peptide N termini, trypsinised histones were derivatised again for 45 min at 37°C using 1:20 (v/v) d6-acetic anhydride (CD<sub>3</sub>CO, Aldrich) in 50 mM ammonium bicarbonate buffered to pH 8 by ammonium hydroxide solution. After derivatization, peptides were evaporated in a speed-vac at 37°C to near dryness, resuspended in 50 µl of 0.1% formic acid and purified by a StageTip protocol using two discs of C18 followed by one disc of activated carbon (3 M Empore). After StageTip purification, the samples were evaporated in a speed-vac to near dryness, resuspended in 20 µl of 0.1% formic acid and stored at –20°C until mass spectrometry acquisition. The histone samples were separated on a reversed-phase liquid chromatography column (75-µm, New Objective) that was packed in-house with a 15-cm stationary phase (ReproSil-Pur C18-AQ, 1.9 µm). The column was connected to a nano-flow HPLC (EASY-nLC 1000; Thermo Scientific) and peptides were electrosprayed in a Q Exactive mass spectrometer (Thermo Fisher Scientific). Buffer A was composed of 0.1% formic acid in HPLC-grade water and buffer B was 0.1% formic acid in ACN. Peptides were eluted in a linear gradient with a flow rate of 300 nl per minute, starting at 3% B and ramping to 35% in 52 min, followed by an increase to 50% B in 4 min, followed by an increase to 98% in 4 min and then holding at 98% B for another 6 min. Mass spectrometry was operated in a combined shotgun-PRM mode targeting positional isomers. Ion chromatograms were extracted with Thermo Xcalibur and Skyline and data summarization and statistical analysis was performed in Excel and R. Relative abundances were calculated from the raw signal reads, according to the formulas described previously<sup>21</sup> without further normalizations.

**Microarray and bioinformatics analysis.** Gene expression analysis was performed using the Mouse GE 8x60K Microarray Kit (Agilent Technologies, Design ID 028005). 100 ng total RNA isolated from SCs were used for the labelling. Samples were labelled with the Low Input Quick Amp Labelling Kit (Agilent Technologies) according to the manufacturer's instructions. Slides were scanned using a microarray scanner (Agilent Technologies). Expression data were extracted using the Feature Extraction software (Agilent Technologies). Preprocessing of expression data was performed according to Agilent's standard workflow. Using five quality flags (gIsPosAndSignif, gIsFeatNonUnifOL, gIsWellAboveBG, gIsSaturated, and gIsFeatPopnOL) from the Feature Extraction software output, probes were labelled as detected, not detected, or compromised. Gene expression levels were background corrected, and signals for duplicated probes were summarized by geometric mean of non-compromised probes. After  $\log_2$  transformation, a percentile shift normalization at the 75% level and a baseline shift to the median baseline of all probes was performed. All computations were performed using the R statistical software framework (<http://www.R-project.org>). Differentially expressed genes were calculated by the shrinkage T-statistic<sup>46</sup> and controlled for multiple testing by maintaining a false discovery rate (FDR) < 0.05 (ref. 47).

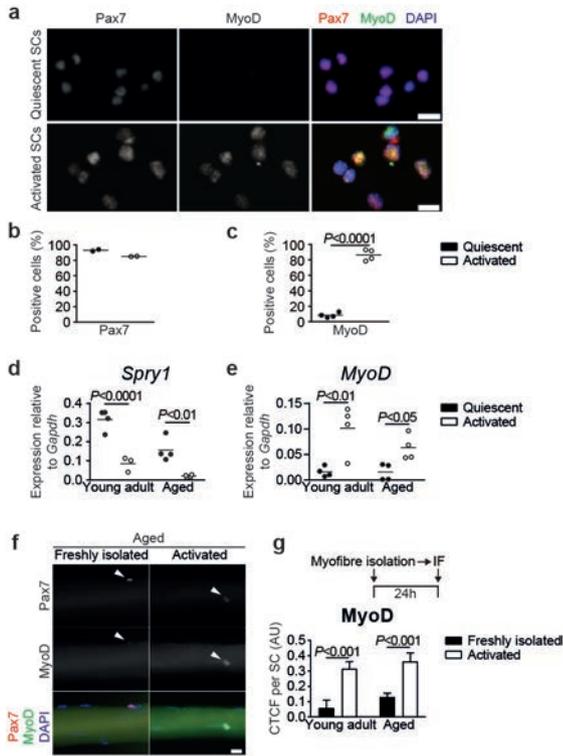
**RNA-sequencing analysis.** Sequencing reads were filtered out for low quality sequences and trimmed of low quality bases by using FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Mapping to mm9 genome was performed by using TopHat software<sup>48</sup>. Gene quantification was performed by using HT-Seq and differentially expressed genes (DEGs) were estimated by using DESEQ2 (refs 49, 50) within the R statistical software framework (<http://www.R-project.org>) with  $P < 0.01$ . Pearson correlation heatmaps were generated by using custom R scripts by selecting genes having more than 10 read counts in all the samples of at least one condition and an interquartile range (IQR) > 0.5. Significance of overlapping DEGs was calculated by normal approximation of hypergeometric probability.

**Identification of Hoxa9-binding sites.** Transcription start and end sites of putative Hoxa9 target genes were collected from the UCSC Genome Browser<sup>51</sup> with mm8 track. Sequences in gene body regions (from transcription start to end sites), promoter regions ( $-2/+1$  kb relative to transcription start sites), and distal intergenic regions ( $-50/+50$  kb relative to transcription start sites) of 26 genes were prepared for identification of Hoxa9 binding sites. These sequences were aligned based on the previously reported consensus motifs for Hoxa9-Meis1-Pbx1 (ATGATTTATGGC)<sup>52</sup> and Meis1 (TGTC)<sup>53</sup>. Putative Hoxa9-binding sites were aligned when they contained either no mismatch or one mismatch, and Meis1 motifs were aligned with no mismatch allowed. Hoxa9-binding sites with at least one Meis1-binding site within 300 bp on the same DNA strand were selected for further analysis. Identified Hoxa9-binding sites are listed in Supplementary Table 1.

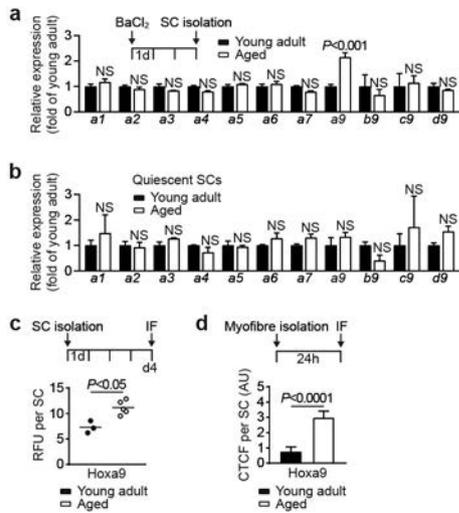
**Statistics.** If not stated otherwise, results are presented as mean and s.e.m. from the number of samples indicated in the figure legends. Two groups were compared by two-sided Student's *t*-test or two-sided Mann-Whitney *U*-test. For multiple comparisons a two-way ANOVA was performed using a FDR < 0.5 to correct for multiple comparisons. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ . Statistical analysis was done using GraphPad Prism 6 software and R (v3.3.1).

**Data availability statement.** Microarray and RNA-sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) with the accession code GSE87812. Further data that support the findings of this study are available from the corresponding authors upon reasonable request. Source data for the Figures and Extended Data Figures are provided with the paper.

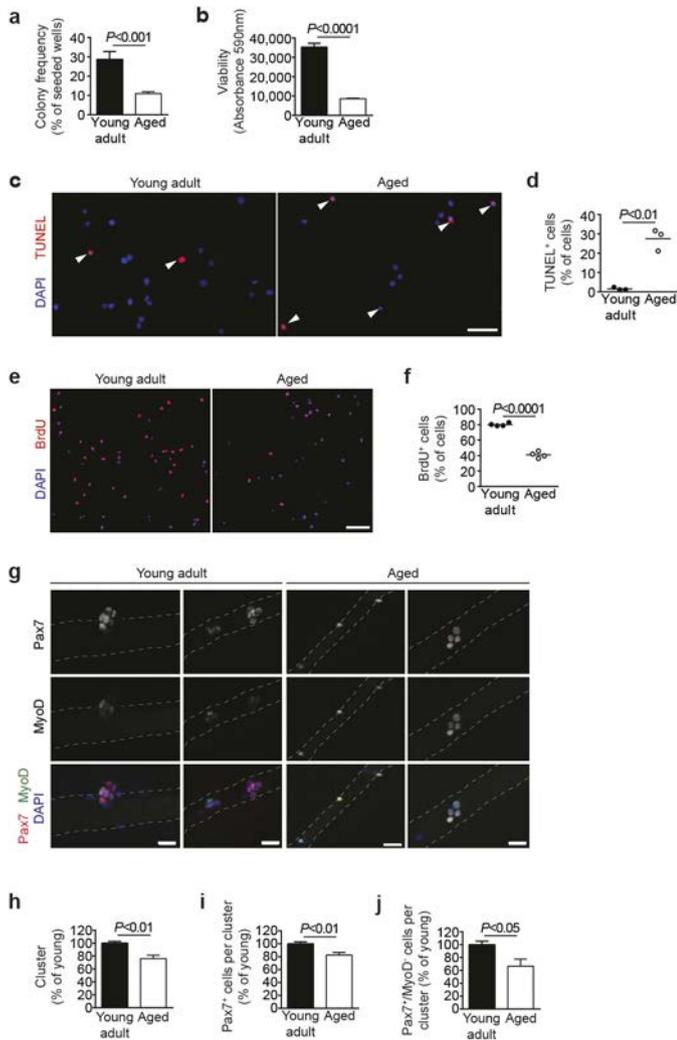
28. Lawrence, H. J. *et al.* Mice bearing a targeted interruption of the homeobox gene *HoxA9* have defects in myeloid, erythroid, and lymphoid hematopoiesis. *Blood* **89**, 1922–1930 (1997).
29. Brown, P. J. & Müller, S. Open access chemical probes for epigenetic targets. *Future Med. Chem.* **7**, 1901–1917 (2015).
30. Barsyte-Lovejoy, D. *et al.* Chemical biology approaches for characterization of epigenetic regulators. *Methods Enzymol.* **574**, 79–103 (2016).
31. Theodoulou, N. H. *et al.* Discovery of I-BRD9, a selective cell active chemical probe for bromodomain containing protein 9 inhibition. *J. Med. Chem.* **59**, 1425–1439 (2016).
32. Picaud, S. *et al.* Generation of a selective small molecule inhibitor of the CBP/p300 bromodomain for leukemia therapy. *Cancer Res.* **75**, 5106–5119 (2015).
33. Picaud, S. *et al.* PFI-1, a highly selective protein interaction inhibitor, targeting BET bromodomains. *Cancer Res.* **73**, 3336–3346 (2013).
34. Martin, L. J. *et al.* Structure-based design of an *in vivo* active selective BRD9 inhibitor. *J. Med. Chem.* **59**, 4462–4475 (2016).
35. Hay, D. A. *et al.* Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains. *J. Am. Chem. Soc.* **136**, 9308–9319 (2014).
36. Drouin, L. *et al.* Structure enabled design of BAZ2-ICR, a chemical probe targeting the bromodomains of BAZ2A and BAZ2B. *J. Med. Chem.* **58**, 2553–2559 (2015).
37. Clark, P. G. *et al.* LP99: discovery and synthesis of the first selective BRD7/9 bromodomain inhibitor. *Angew. Chem.* **127**, 6315–6319 (2015).
38. Chen, P. *et al.* Discovery and characterization of GSK2801, a selective chemical probe for the bromodomains BAZ2A and BAZ2B. *J. Med. Chem.* **59**, 1410–1424 (2016).
39. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
40. Pasut, A., Jones, A. E. & Rudnicki, M. A. Isolation and culture of individual myofibers and their satellite cells from adult skeletal muscle. *J. Vis. Exp.* **73**, 50074 (2013).
41. Bentzinger, C. F. *et al.* Fibronectin regulates Wnt7a signaling and satellite cell expansion. *Cell Stem Cell* **12**, 75–87 (2013).
42. Schambach, A. *et al.* Lentiviral vectors pseudotyped with murine ecotropic envelope: increased biosafety and convenience in preclinical research. *Exp. Hematol.* **34**, 588–592 (2006).
43. Wang, J. *et al.* A differentiation checkpoint limits hematopoietic stem cell self-renewal in response to DNA damage. *Cell* **148**, 1001–1014 (2012).
44. Chambeyron, S. & Bickmore, W. A. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130 (2004).
45. Burgess, A. *et al.* Loss of human Greatwall results in G2 arrest and multiple mitotic defects due to deregulation of the cyclin B-Cdc2/PP2A balance. *Proc. Natl Acad. Sci. USA* **107**, 12564–12569 (2010).
46. Opgen-Rhein, R. & Strimmer, K. Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Stat. Appl. Genet. Mol. Biol.* **6**, <http://dx.doi.org/10.2202/1544-6115.1252> (2007).
47. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 12 (1995).
48. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
49. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
51. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
52. Shen, W. F. *et al.* HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells. *Mol. Cell. Biol.* **19**, 3051–3061 (1999).
53. Huang, Y. *et al.* Identification and characterization of Hoxa9 binding sites in hematopoietic cells. *Blood* **119**, 388–398 (2012).



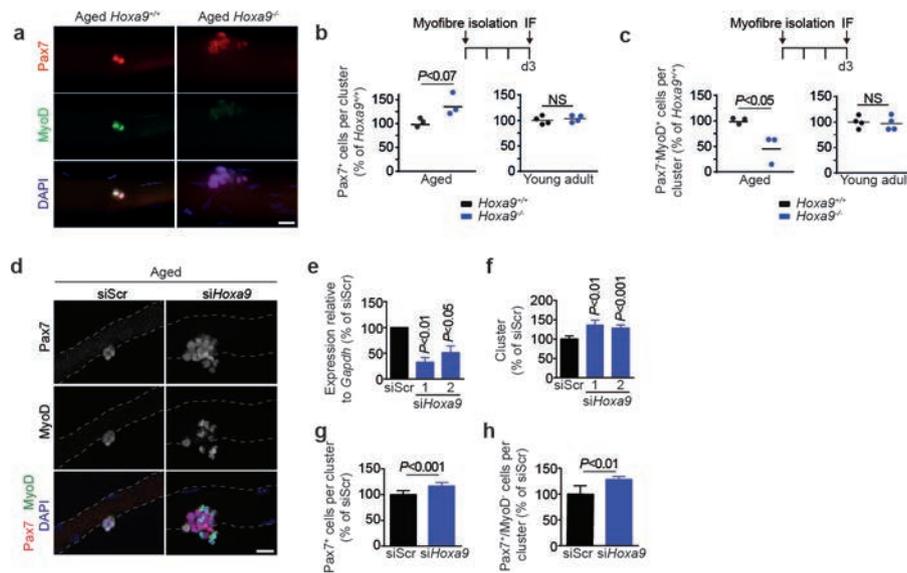
**Extended Data Figure 1 | SC activation.** **a**, Immunofluorescence staining for Pax7 and MyoD of freshly isolated SCs from injured (activated SCs) and uninjured muscles (quiescent SCs) from young adult mice. Nuclei were counterstained with DAPI (blue). **b**, **c**, Quantification of Pax7<sup>+</sup> cells (**b**) and MyoD<sup>+</sup> cells (**c**) in **a**. **d**, **e**, qPCR analysis of *Spry1* (**d**) and *MyoD* (**e**) expression in freshly isolated quiescent and *in vivo* activated SCs of young adult and aged mice. **f**, Immunofluorescence staining for Pax7 and MyoD on freshly isolated and 24-h cultured myofibre-associated SCs from aged mice. Nuclei were counterstained with DAPI (blue). **g**, Corrected total cell fluorescence (CTCF) for MyoD per SC as in **f**. Scale bars, 10  $\mu$ m (**a**) and 20  $\mu$ m (**f**). *P* values were calculated by two-sided Student's *t*-test (**b**, **c**) or two-way ANOVA (**d**, **e**, **g**). *n* = 2 mice in **b**; *n* = 4 mice in **c**; *n* = 3 mice (young activated), *n* = 4 mice (all others) in **d**; *n* = 4 mice in **e**; *n* = 33/24 nuclei (young), *n* = 35/20 nuclei (aged) from 3 mice in **g**.



**Extended Data Figure 2 | Expression of Hox genes in SCs.** **a, b**, Nanostring analysis of mRNA expression of *Hoxa* genes and *Hoxa9* paralogues (*b9-c9-d9*) in *in vivo* activated (**a**) and quiescent (**b**) freshly isolated SCs from young adult and aged mice. **c**, Relative fluorescence units (RFU) for *Hoxa9* per SC in 4-day cultured SCs from young adult and aged mice. **d**, Corrected total cell fluorescence (CTFC) for *Hoxa9* per activated SC on 24-h cultured myofibres as in Fig. 1d. *P* values were calculated by two-way ANOVA (**a, b**) or two-sided Mann–Whitney *U*-test (**c, d**). *n* = 3 mice in **a, b**; *n* = 3 mice (young), *n* = 5 mice (aged) in **c**; *n* = 34 nuclei (young), *n* = 32 nuclei (aged) from 4 mice in **d**.

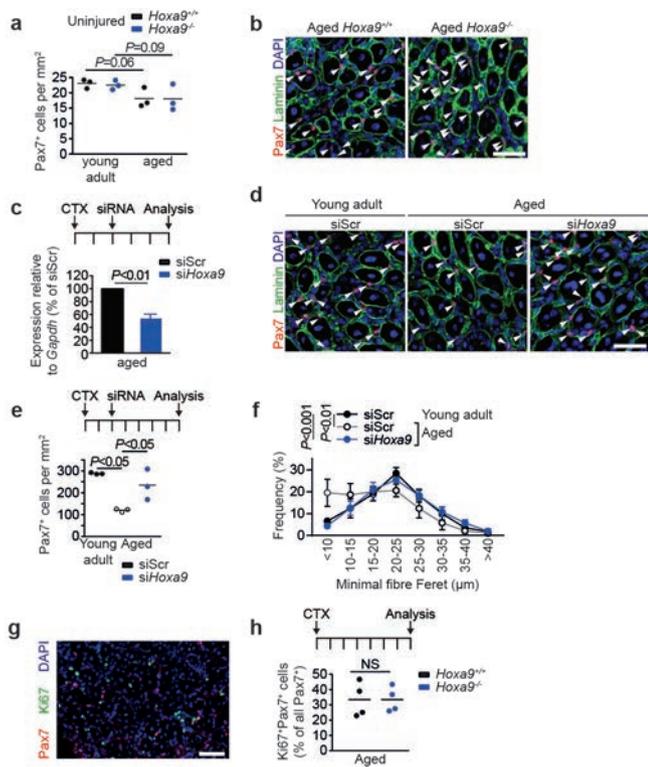


**Extended Data Figure 3 | Functional decline in aged SCs.** **a**, SCs from young adult and aged mice were sorted as single cells. After 5 days, the frequency of myogenic colonies was assessed. **b**, Equal numbers of FACS-isolated SCs from young adult and aged mice were cultured for 4 days and Alamar Blue assay was performed. **c**, TUNEL staining of SCs isolated from young adult or aged mice after 4 days of culture. Nuclei were counterstained with DAPI (blue). **d**, Quantification of apoptosis based on TUNEL staining in **c**. **e**, BrdU staining of SCs isolated from young adult or aged mice after 4 days of culture. Nuclei were counterstained with DAPI (blue). **f**, Quantification of proliferation based on BrdU staining in **e**. **g**, Immunofluorescence staining for Pax7 and MyoD on myofibers isolated from young adult and aged mice after 72 h in culture. Nuclei were counterstained with DAPI (blue). **h–j**, Quantification of the number of SC-derived clusters with at least 3 adjacent cells (**h**), average number of all Pax7<sup>+</sup> cells (**i**), or proportion of Pax7<sup>+</sup>/MyoD<sup>-</sup> cells (**j**) within clusters as in **g**. Scale bars, 20  $\mu\text{m}$  (**c**, **g**) and 50  $\mu\text{m}$  (**e**). *P* values were calculated by two-sided Student's *t*-test.  $n = 8$  mice (young),  $n = 10$  mice (aged) in **a**;  $n = 7$  mice (young),  $n = 5$  mice (aged) in **b**;  $n = 3$  mice in **d**;  $n = 4$  mice in **f**;  $n = 4$  mice (aged) in **j**,  $n = 5$  mice (all others) in **h–j**.



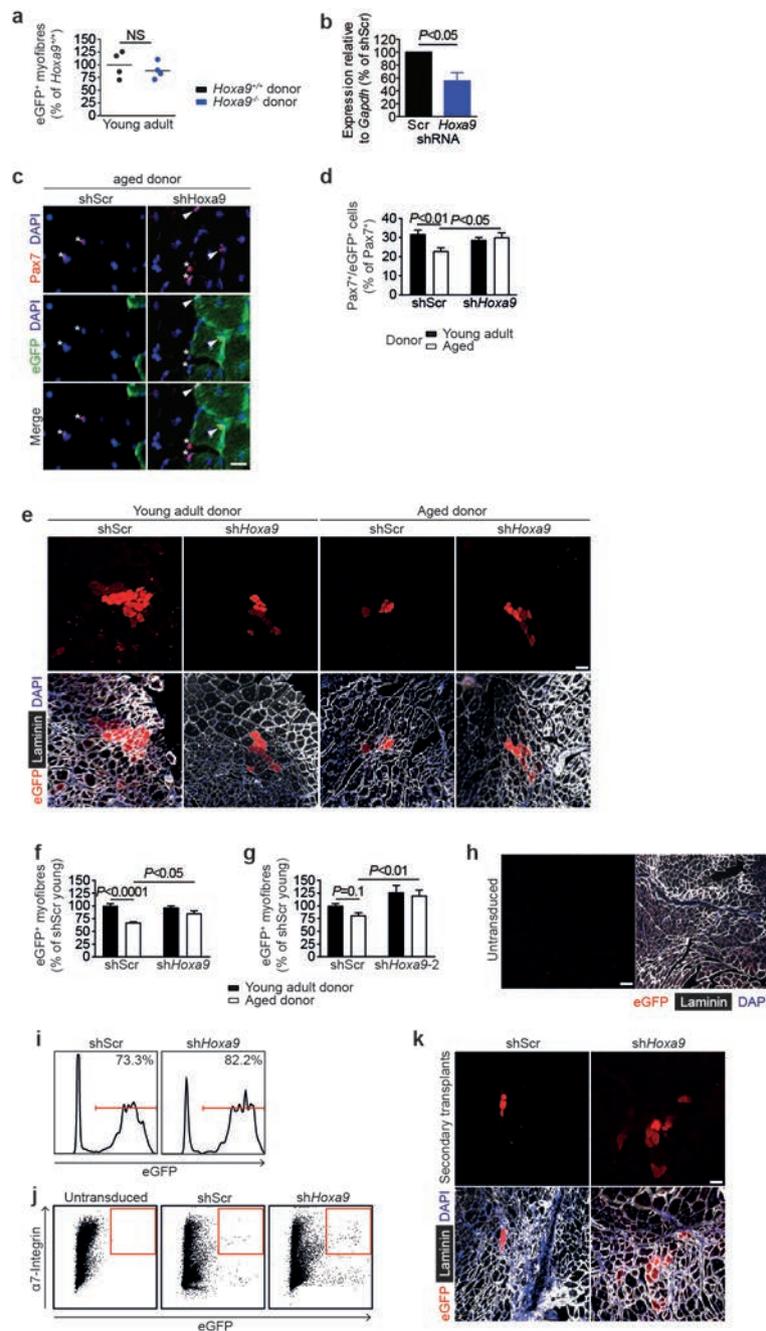
**Extended Data Figure 4 | Deletion or knockdown of *Hoxa9* improves SC function in myofibre cultures.** **a**, Immunofluorescence staining for Pax7 and MyoD on 72 h cultured myofibre-associated SCs from aged *Hoxa9*<sup>+/+</sup> and *Hoxa9*<sup>-/-</sup> mice. **b**, **c**, Average number of all Pax7<sup>+</sup> cells (**b**) or Pax7<sup>+</sup>/MyoD<sup>+</sup> cells (**c**) within clusters from aged or young adult *Hoxa9*<sup>+/+</sup> and *Hoxa9*<sup>-/-</sup> mice as shown in **a**. **d**, Immunofluorescence staining for Pax7 and MyoD on 72-h cultured myofibres isolated from aged mice transfected with *Hoxa9* or scrambled (Scr) siRNAs. Nuclei were counterstained with DAPI (blue). **e**, qPCR analysis of *Hoxa9* expression

in SCs transfected with *Hoxa9* siRNA or scrambled control. Two *Hoxa9* siRNAs with different target sequences (Supplementary Table 1) were used. **f–h**, Analysis of 72-h cultured myofibre-associated SCs from **d**. Quantification of the number of SC-derived clusters with at least 3 adjacent cells (**f**), average number of all Pax7<sup>+</sup> cells (**g**), or proportion of Pax7<sup>+</sup>/MyoD<sup>+</sup> cells (**h**) within clusters. Scale bars, 20  $\mu$ m (**a**, **d**). Dashed lines outline myofibres. *P* values were calculated by two-sided Student's *t*-test. *n* = 3 mice (aged), *n* = 4 mice (young) in **b**, **c**; *n* = 3 mice in **e**; *n* = 5 mice in **f–h**.



### Extended Data Figure 5 | Inhibition of *Hoxa9* improves muscle regeneration in aged mice.

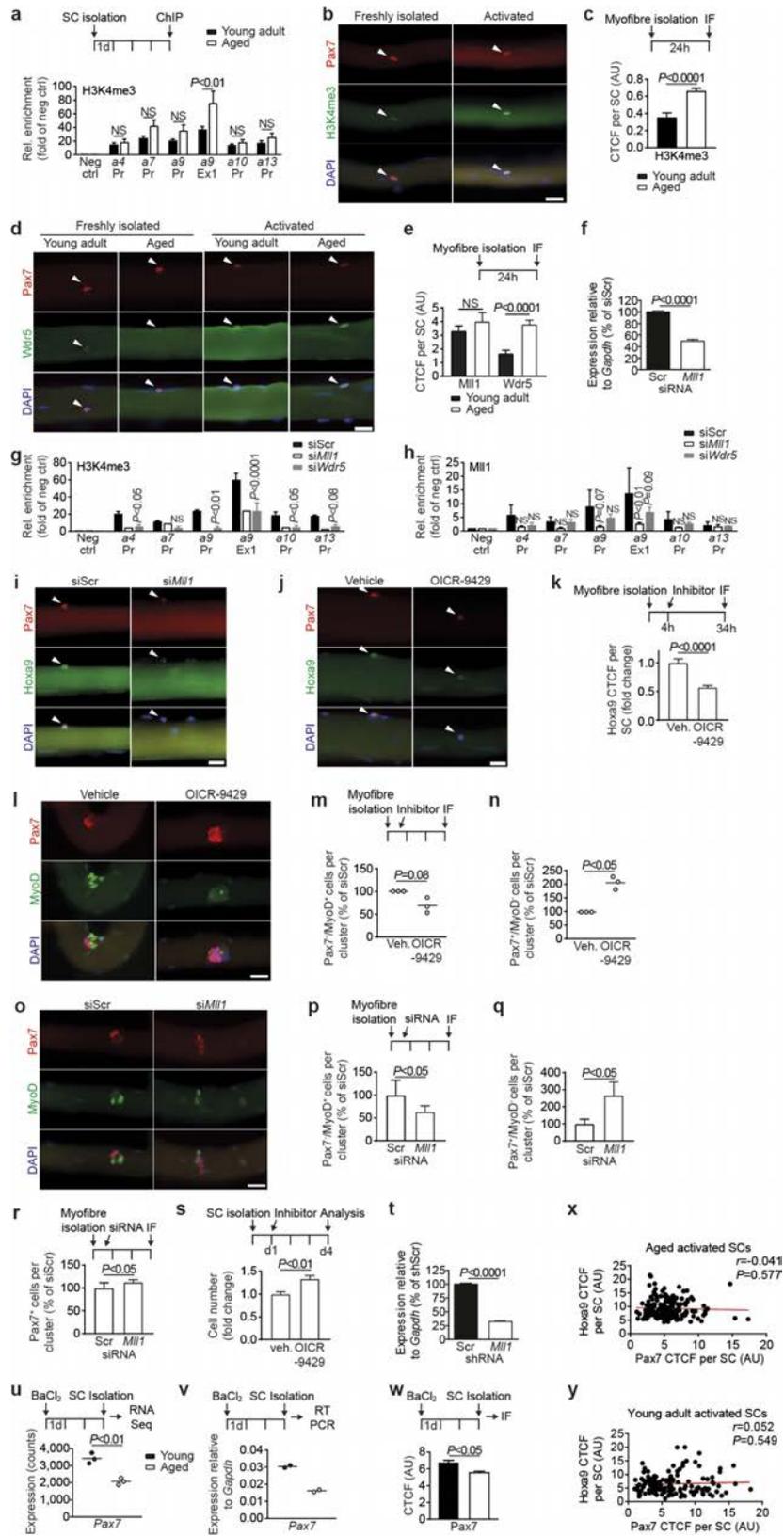
**a**, Quantification of Pax7<sup>+</sup> cells per area in uninjured tibialis anterior muscles from young adult and aged *Hoxa9*<sup>+/+</sup> and *Hoxa9*<sup>-/-</sup> mice. **b**, Representative immunofluorescence staining for Pax7 and laminin on tibialis anterior muscles from aged *Hoxa9*<sup>+/+</sup> and *Hoxa9*<sup>-/-</sup> mice that were collected 7 days after cardiotoxin (CTX) injury. **c**, qPCR analysis of *Hoxa9* expression in SCs isolated from tibialis anterior muscles injected with a self-delivering *Hoxa9* or scrambled siRNA and collected 5 days after muscle injury. **d**, Representative immunofluorescence staining for Pax7 and laminin of injured tibialis anterior muscles from young adult and aged mice that were injected with a self-delivery siRNA and collected 7 days after muscle injury. Nuclei were counterstained with DAPI (blue). Arrowheads denote Pax7<sup>+</sup> cells. **e**, Quantification of Pax7<sup>+</sup> cells from **d** per area. **f**, Frequency distribution minimal Feret's diameter of muscle fibres from **d**. **g**, Exemplary immunofluorescence staining for Pax7 and Ki67 on tibialis anterior muscles from aged *Hoxa9*<sup>+/+</sup> and *Hoxa9*<sup>-/-</sup> mice collected 7 days after muscle injury. Nuclei were counterstained with DAPI (blue). **h**, Quantification of proliferating SCs (Ki67<sup>+</sup>/Pax7<sup>+</sup>) as depicted in **g**. Scale bars, 50  $\mu$ m. *P* values were calculated by two-sided Student's *t*-test (**c**, **h**) or two-way ANOVA (**a**, **e**, **f**). *n* = 3 mice in **a**; *n* = 3 mice in **c**; *n* = 3 mice in **e**, **f**; *n* = 4 mice in **h**.



### Extended Data Figure 6 | Inhibition of *Hoxa9* improves regenerative capacity of aged SCs.

**a**, Quantification of donor-derived (eGFP<sup>+</sup>) myofibres from transplantation of SCs from young adult *Hoxa9*<sup>+/+</sup> and *Hoxa9*<sup>-/-</sup> mice. **b**, qPCR analysis of *Hoxa9* expression in SCs transduced with scrambled control or *Hoxa9* shRNA encoding lentivirus. **c–g**, Transplantation of eGFP-labelled SCs from young adult and aged mice that were targeted with shRNAs against *Hoxa9* or a scrambled control. **c**, Representative immunofluorescence staining for Pax7 and eGFP of transplanted muscle sections. Nuclei were counterstained with DAPI (blue). Arrowheads denote Pax7<sup>+</sup>/eGFP<sup>+</sup> cells, asterisks label Pax7<sup>+</sup>/eGFP<sup>-</sup> cells. **d**, Quantification of donor-derived (eGFP<sup>+</sup>) Pax7<sup>+</sup> cells in **c**. **e**, Representative immunofluorescence staining for eGFP and laminin of transplanted muscle sections, nuclei were counterstained with DAPI (blue). **f, g**, Quantification of donor-derived (eGFP<sup>+</sup>) myofibres in **e** for two different *Hoxa9* shRNAs in two independent experiments.

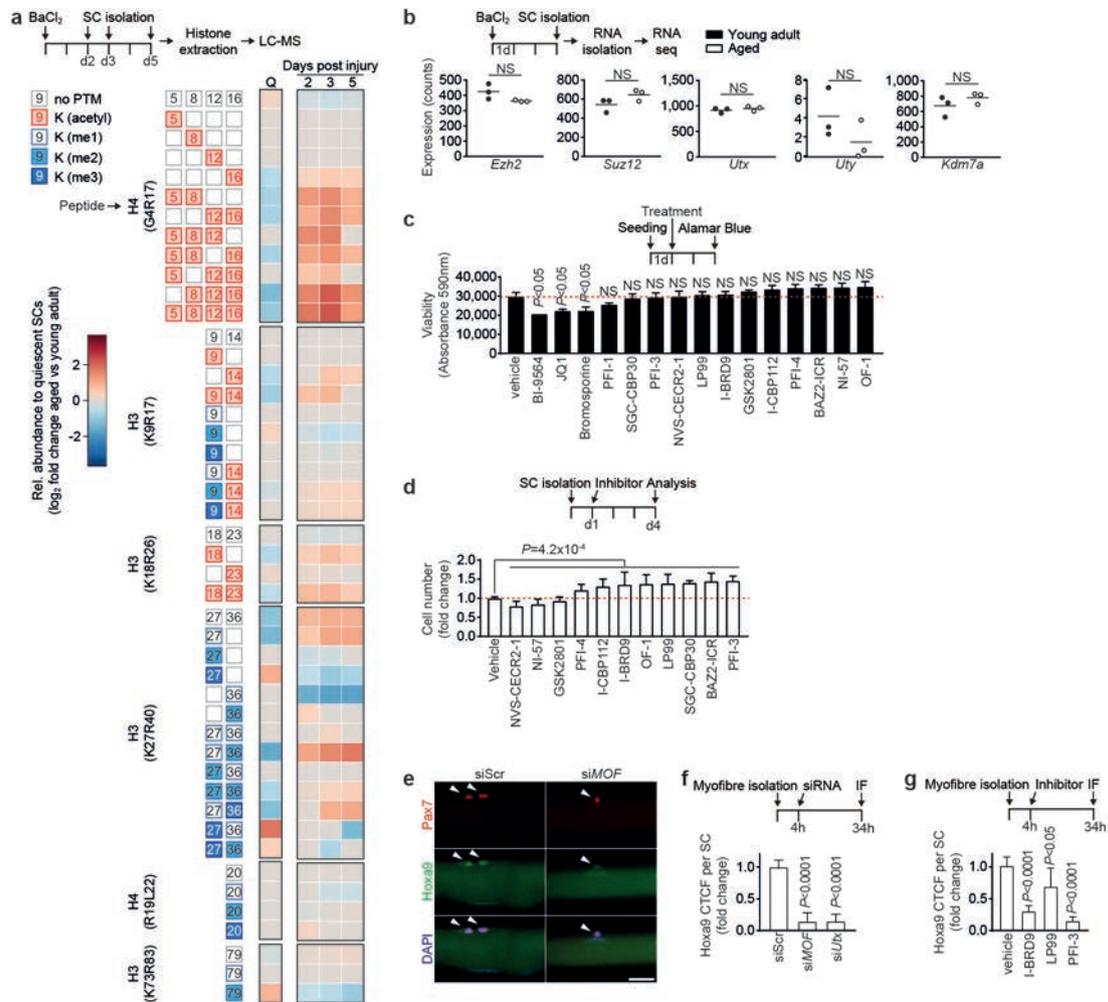
**h**, Exemplary immunofluorescence staining for eGFP and laminin in tibialis anterior muscles engrafted with untransduced aged SCs. Nuclei were counterstained with DAPI (blue). **i**, Flow cytometric analysis of transduction efficiency of donor SCs used for transplantation in primary recipients analysed in Fig. 2f. **j**, Representative flow cytometry plots for re-isolation of transplanted aged SCs that were untransduced as control or transduced with scrambled control or *Hoxa9* shRNA encoding lentivirus as quantified in Fig. 2f. **k**, Representative immunofluorescence staining for eGFP and laminin in engrafted tibialis anterior muscles from secondary recipients quantified in Fig. 2g. Nuclei were counterstained with DAPI (blue). Scale bars, 20 μm (**c**), 50 μm (**h**) and 100 μm (**e, k**). *P* values were calculated by two-sided Student's *t*-test (**a, b**) or two-way ANOVA (**d, f, g**). *n* = 4 recipient mice in **a**; *n* = 3 mice in **b**; *n* = 6 recipient mice (young donors), *n* = 4 recipient mice (aged donors) in **d, f**; *n* = 5 recipient mice in **g**.



Extended Data Figure 7 | See next page for caption.

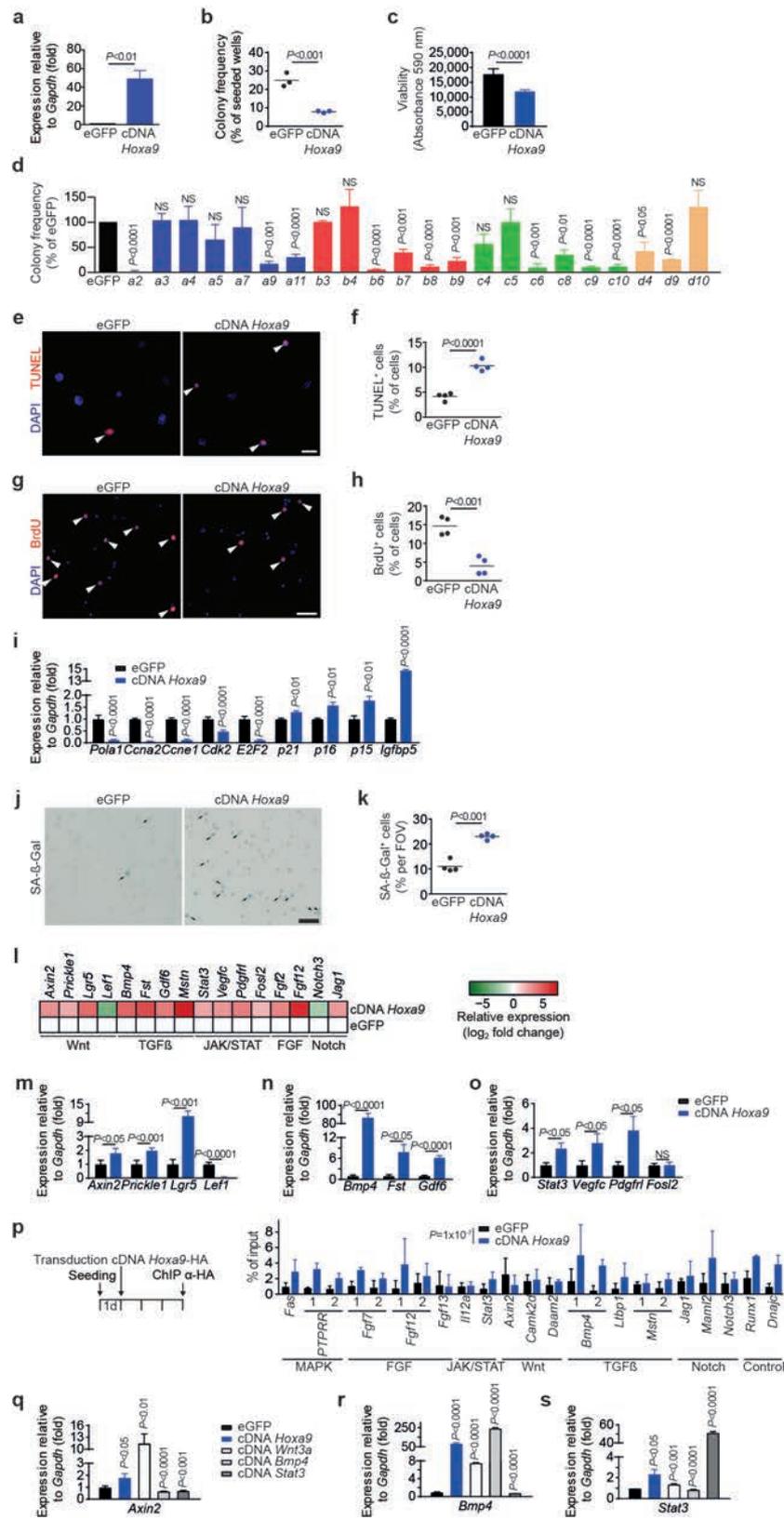
**Extended Data Figure 7 | Inhibition of Mll1 rescues H3K4me3 induction, Hoxa9 overexpression, and functional impairment of activated SCs from aged mice.** **a**, ChIP for H3K4me3 at promoters or exons of indicated Hox genes in activated SCs (4 day culture) from young adult and aged mice. **b**, Representative immunofluorescence staining for Pax7 and H3K4me3 on myofibre-associated SCs from aged mice that were freshly isolated or activated by 24-h culture of myofibres. **c**, Corrected total cell fluorescence (CTCF) for H3K4me3 on activated SCs shown in **b**. **d**, Representative immunofluorescence staining for Pax7 and Wdr5 on myofibre-associated SCs from young adult and aged mice that were freshly isolated or activated by 24-h culture of myofibres. **e**, CTCF for Mll1 and Wdr5 per activated SC as shown in **d**. **f**, qPCR analysis of *Mll1* in SCs transfected with *Mll1* siRNA or scrambled control. **g, h**, ChIPs for H3K4me3 (**g**) and Mll1 (**h**) in primary myoblasts 3 days after transfection with the indicated siRNAs. **i, j**, Immunofluorescence staining for Pax7 and Hoxa9 in myofibres from aged mice after transfection with *Mll1* siRNA or scrambled control (**i**, quantification in Fig. 3d) or after treatment with OICR-9429 or vehicle (**j**). **k**, CTCF for Hoxa9 per SC as shown in **j**. **l**, Representative immunofluorescence staining for Pax7 and MyoD on OICR-9429 treated myofibre-associated SCs from aged mice after 72 h culture. Nuclei were counterstained with DAPI (blue). **m, n**, Average number of Pax7<sup>-</sup>/MyoD<sup>+</sup> cells (**m**) or Pax7<sup>+</sup>/MyoD<sup>-</sup> cells (**n**) within clusters as shown in **l**. **o**, Representative immunofluorescence staining for Pax7 and MyoD on siRNA-treated myofibre-associated SCs from aged

mice after 72-h culture. Nuclei were counterstained with DAPI (blue). **p–r**, Average number of Pax7<sup>-</sup>/MyoD<sup>+</sup> cells (**p**), Pax7<sup>+</sup>/MyoD<sup>-</sup> cells (**q**) or Pax7<sup>+</sup> cells (**r**) within clusters in **o**. **s**, Relative changes in cell number of aged SCs after treatment with OICR-9429 and 4 days of culture, compared to vehicle control. **t**, qPCR analysis of *Mll1* in SCs transduced with *Mll1* shRNA or scrambled control. **u–w**, Analysis of Pax7 expression in *in vivo* activated SCs from young adult and aged mice by RNA-sequencing (**u**), qPCR (**v**), or immunofluorescence as depicted in Fig. 1b (**w**). **x, y**, Pearson correlation comparing the Hoxa9 immunofluorescence signal (quantification in Fig. 1c) and the Pax7 immunofluorescence signal (quantification in **w**) of activated SCs from aged (**x**) and young adult (**y**) mice. Note, there is no correlation between Hoxa9 expression level and Pax7 expression level in activated SCs from aged mice. Scale bars, 20 μm (**b, d, i, j, l, o**). *P* values were calculated by two-way ANOVA (**a, g, h**), two-sided Student's *t*-test (**f, m, n, p–v**), two-sided Mann–Whitney *U*-test (**c, e, k, w**) or Pearson correlation (**x, y**). *n* = 4 mice (young), *n* = 7 mice (aged) in **a**; *n* = 27 nuclei from 2 mice (young), *n* = 27 nuclei from 4 mice (aged) in **c**; *n* = 40/52 nuclei (Mll1), *n* = 44/99 nuclei (Wdr5) from 3 young/aged mice in **e**; *n* = 3 mice in **f**; *n* = 3 biological replicates (*Wdr5* siRNA), *n* = 2 biological replicates (*Mll1* siRNA) in **g**; *n* = 3 biological replicates in **h**; *n* = 173 nuclei (DMSO), *n* = 324 nuclei (OICR-9429) from 4 mice in **k**; *n* = 3 mice in **m, n**; *n* = 7 mice in **p–r**; *n* = 6 mice in **s**; *n* = 3 mice in **t**; *n* = 3 mice in **u**; *n* = 2 mice in **v**; *n* = 134 nuclei (young), *n* = 181 nuclei (aged) from 3 mice in **w–y**.



**Extended Data Figure 8 | Alterations in the epigenetic stress response of activated SCs from aged mice.** **a**, Heatmap displaying relative changes in abundance of different histone modifications (measured at the indicated peptides) in freshly isolated SCs from aged compared to young adult mice. SCs were analysed in quiescence (Q, derived from uninjured muscle) or at the indicated time points after activation mediated by muscle injury. Relative abundances at indicated days after injury are first normalized to quiescent SCs, and then compared between SCs isolated from aged and young adult mice and log<sub>2</sub> scaled. Only significant changes are shown ( $P < 0.05$ ). **b**, Expression analysis of the indicated genes in freshly isolated *in vivo* activated SCs from young adult and aged mice based on RNA-sequencing. **c**, Viability of primary myoblasts after 48-h treatment with bromodomain inhibitors (1  $\mu$ M) from the Structural Genomics Consortium probe set, measured by Alamar Blue assay. **d**, Relative changes in cell number of aged SCs after treatment with non-toxic bromodomain

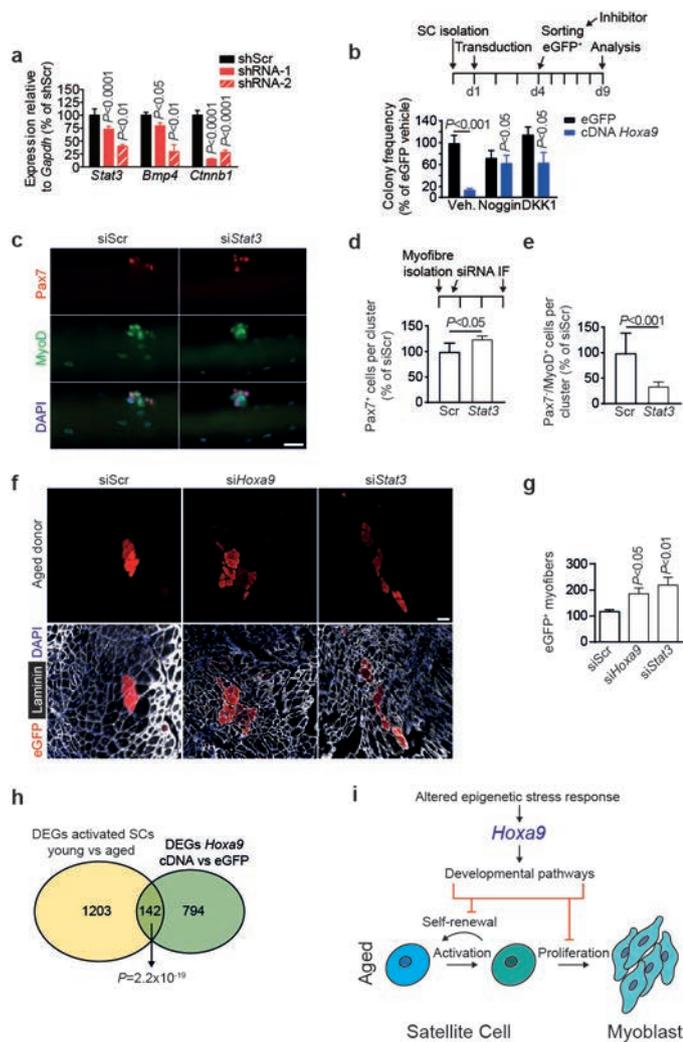
inhibitors (1  $\mu$ M) from **c** and 4 days of culture, compared to vehicle control. A Wilcoxon rank-sum test on the ratio of all cell counts being equal to 1 was performed to test the hypothesis of a general effect of the inhibitors on cell number. **e**, Representative immunofluorescence staining for Pax7 and Hoxa9 in siRNA-treated myofibre-associated SCs from aged mice. Scale bar, 20  $\mu$ m. **f**, CTCF for Hoxa9 per SC as shown in **e**. **g**, Quantification of immunofluorescence staining for Hoxa9 in Pax7<sup>+</sup> cells on myofibre-associated SCs from aged mice treated with bromodomain inhibitors.  $P$  values were calculated by two-sided Student's  $t$ -test (**a–c**), Wilcoxon rank-sum test (**d**) or two-sided Mann–Whitney  $U$ -test (**f, g**).  $n = 4$  mice in **a**;  $n = 3$  mice in **b**;  $n = 4$  biological replicates in **c**;  $n = 6$  mice in **d**;  $n = 71$  nuclei (scrambled siRNA),  $n = 48$  nuclei (*MOF* siRNA),  $n = 98$  nuclei (*Utx* siRNA) from 3 mice in **f**;  $n = 60$  nuclei (vehicle),  $n = 59$  nuclei (I-BRD9),  $n = 38$  nuclei (LP99),  $n = 62$  nuclei (PFI-3) from 3 mice in **g**.



Extended Data Figure 9 | See next page for caption.

**Extended Data Figure 9 | Overexpression of *Hox* genes inhibits SC function.** **a**, Expression of *Hoxa9* in SCs transduced with *Hoxa9* cDNA or eGFP as control. **b, c**, FACS-isolated SCs from young adult mice were transduced with a lentivirus either containing both eGFP and *Hoxa9* cDNA or only eGFP. Infected (eGFP<sup>+</sup>) cells were isolated after 3 days. **b**, Frequency of myogenic colonies from single-cell-sorted SCs. **c**, Quantification of cell number based on Alamar Blue assay of bulk cultures. **d**, Frequency of myogenic colonies of SCs overexpressing the indicated *Hox* genes. **e, g**, TUNEL (**e**) or BrdU (**g**) staining of SCs overexpressing *Hoxa9* or eGFP. Infected (eGFP<sup>+</sup>) cells were isolated 3 days after transduction and analysed 3 days later. Nuclei were counterstained with DAPI (blue). Arrowheads mark TUNEL- or BrdU-positive cells. **f, h**, Quantification of apoptosis (**f**) or proliferation (**h**) based on TUNEL or BrdU staining as in **e** or **g**. **i**, qPCR-based expression analysis of various cell-cycle and senescence markers in SCs overexpressing *Hoxa9* compared to eGFP-infected controls, 5 days after infection. **j**, Senescence-associated- $\beta$ -galactosidase (SA- $\beta$ -Gal) staining of SCs overexpressing *Hoxa9* or eGFP at day 5 after infection. Arrowheads mark SA- $\beta$ -Gal-positive cells. **k**, Quantification of senescence per field of view (FOV) based on SA- $\beta$ -Gal staining in **j**. **l**, Heatmap displaying log<sub>2</sub> fold changes

of expression of selected genes from microarray analysis in Fig. 5a. **m–o**, qPCR validation of differentially expressed genes annotated to Wnt (**m**), TGF $\beta$  (**n**) and JAK/STAT pathways (**o**) as in **l**. **p**, Identification of *Hoxa9*-binding sites by anti-HA ChIP of primary myoblasts overexpressing HA-tagged *Hoxa9* cDNA or eGFP as control. Shown is the qPCR for 1 or 2 putative *Hoxa9*-binding sites at the indicated loci. *Hoxa9*-binding sites at target genes were identified as described in the Methods and are listed in Supplementary Table 1. A two-sided block bootstrap test on the difference of the percentage of bound DNA for all binding sites being equal to 0 was performed to test the hypothesis of a generally increased binding of *Hoxa9*. **q–s**, SCs were infected with lentiviruses expressing *Hoxa9*, *Wnt3a*, *Bmp4* or *Stat3* cDNAs or eGFP. qPCR analysis of expression of the indicated target genes at 5 days after infection: *Axin2* (**q**), *Bmp4* (**r**) and *Stat3* (**s**). Scale bars, 20  $\mu$ m (**e, g**) and 50  $\mu$ m (**j**). *P* values were calculated by two-sided Student's *t*-test (**a–d, f, h, k, q–s**) or two-way ANOVA (**i, m–o**). *n* = 4 mice in **a**; *n* = 3 mice in **b**; *n* = 7 mice in **c**; *n* = 3 mice in **d**; *n* = 4 mice in **f, h, k**; *n* = 3 mice (p15, p21), *n* = 6 mice (p16), *n* = 4 mice (all others) in **i**; *n* = 4 pools of 3 mice in **l**; *n* = 4 mice in **m–o**; *n* = 3 biological replicates for **p**; *n* = 3 mice (*Wnt3a, Bmp4, Stat3*), *n* = 4 mice (eGFP, *Hoxa9*) in **q–s**.



### Extended Data Figure 10 | Validation of *Hoxa9* downstream targets.

**a**, Knockdown efficiency of two shRNAs (red bars) for *Stat3*, *Bmp4* and *Cttnb1*. **b**, SCs from young adult mice were transduced with an *Hoxa9* and *eGFP*-encoding lentivirus. *eGFP*<sup>+</sup> cells were sorted as single cells and cultured in the presence of noggin, DKK1 or 0.1% BSA in PBS as vehicle. Colony frequency was assessed after 5 days and is compared to *Hoxa9* cDNA expressing cells treated with vehicle control. **c**, Representative immunofluorescence staining for Pax7 and MyoD on siRNA-transfected myofibers from aged mice after 72 h of culture. Nuclei were counterstained with DAPI (blue). **d**, **e**, Average number of Pax7<sup>+</sup> cells (**d**) or Pax7<sup>+</sup>/MyoD<sup>+</sup> cells (**e**) within clusters in **c**. **f**, Representative immunofluorescence staining for eGFP and laminin in tibialis anterior muscles engrafted with siRNA-transfected SCs isolated from eGFP transgenic aged mice. Nuclei were counterstained with DAPI (blue). **g**, Quantification of donor-derived (*eGFP*<sup>+</sup>) myofibers in **f**. **h**, Area-proportional Venn diagram of differentially expressed genes from indicated transcriptomes. **i**, Model for the *Hoxa9*-mediated impairment of SC function during ageing: quiescent SCs become activated upon muscle injury and proliferate as myoblasts to repair damaged muscle tissue. After activation, aged SCs display global and locus-specific alterations in the epigenetic stress response resulting in overexpression of *Hoxa9*, which in turn induces developmental pathways inhibiting SC function and muscle regeneration in aged mice. Scale bars, 20 μm (**c**), and 100 μm (**f**). *P* values were calculated by two-way ANOVA (**a**, **b**) or two-sided Student's *t*-test (**d**, **e**, **g**). *n* = 3 mice in **a**; *n* = 4 mice in **b**; *n* = 5 mice in **d**, **e**; *n* = 5 recipient mice in **g**; *n* = 3 mice per group (activated SCs), *n* = 4 pools of 3 mice (*Hoxa9* overexpression) in **h**.

# Global Epigenomic Reconfiguration During Mammalian Brain Development

Ryan Lister,<sup>1,2\*†</sup> Eran A. Mukamel,<sup>3\*</sup> Joseph R. Nery,<sup>1</sup> Mark Urich,<sup>1</sup> Clare A. Puddifoot,<sup>3</sup> Nicholas D. Johnson,<sup>3</sup> Jacinta Lucero,<sup>3</sup> Yun Huang,<sup>4</sup> Andrew J. Dwork,<sup>5,6</sup> Matthew D. Schultz,<sup>1,7</sup> Miao Yu,<sup>8</sup> Julian Tonti-Filippini,<sup>2</sup> Holger Heyn,<sup>9</sup> Shijun Hu,<sup>10</sup> Joseph C. Wu,<sup>10</sup> Anjana Rao,<sup>4</sup> Manel Esteller,<sup>9,11</sup> Chuan He,<sup>8</sup> Fatemeh G. Haghghi,<sup>5</sup> Terrence J. Sejnowski,<sup>3,12,13</sup> M. Margarita Behrens,<sup>3†</sup> Joseph R. Ecker<sup>1,13†</sup>

DNA methylation is implicated in mammalian brain development and plasticity underlying learning and memory. We report the genome-wide composition, patterning, cell specificity, and dynamics of DNA methylation at single-base resolution in human and mouse frontal cortex throughout their lifespan. Widespread methylome reconfiguration occurs during fetal to young adult development, coincident with synaptogenesis. During this period, highly conserved non-CG methylation (mCH) accumulates in neurons, but not glia, to become the dominant form of methylation in the human neuronal genome. Moreover, we found an mCH signature that identifies genes escaping X-chromosome inactivation. Last, whole-genome single-base resolution 5-hydroxymethylcytosine (hmC) maps revealed that hmC marks fetal brain cell genomes at putative regulatory regions that are CG-demethylated and activated in the adult brain and that CG demethylation at these hmC-poised loci depends on Tet2 activity.

**D**ynamic epigenetic changes have been observed during brain development, maturation, and learning (1–6). DNA methylation (mC) is a stable covalent modification that persists in postmitotic cells throughout their lifetime, defining their cellular identity. However, the methylation status at each of the ~1 billion cytosines in the genome is potentially an information-rich and flexible substrate for epigenetic modification that can be altered by cellular activity (7, 8). Changes in DNA methylation were implicated in learning and memory (9, 10), as well as in age-related

cognitive decline (11). Mice with a postnatal deletion of DNA methyltransferases *Dnmt1* and *Dnmt3a* in forebrain excitatory neurons, or with a global deletion of methyl-CpG-binding protein 2 (*MeCP2*), show abnormal long-term neural plasticity and cognitive deficits (2, 12).

DNA methylation composition and dynamics in the mammalian brain are highly distinct. A modification of mC catalyzed by the Tet family of mC hydroxylase proteins, 5-hydroxymethylcytosine (hmC), accumulates in the adult brain (13–15) along with its more highly oxidized derivatives 5-formylcytosine and 5-carboxylcytosine. These modifications of mC were implicated as intermediates in an active DNA demethylation pathway (16–19). In addition, methylation in the non-CG context (mCH, where H = A, C, or T) is also present in the adult mouse and human brains (20, 21) but is rare or absent in other differentiated cell types (22, 23). Little is known about cell type-specific patterning of DNA methylation and its dynamics during mammalian brain development. Here, we provide integrated empirical data and analysis of DNA methylation at single-base resolution, across entire genomes, with cell-type and developmental specificity. These results extend our knowledge of the unique role of DNA methylation in brain development and function and offer a new framework for testing the role of the epigenome in healthy function and in pathological disruptions of neural circuits.

## Accumulation of Non-CG DNA Methylation During Brain Development

To identify the composition and dynamics of transcription and methylation during mammalian brain development, we performed transcriptome profiling (mRNA-Seq) and whole-genome bi-

sulfite sequencing [MethylC-Seq (24)] to comprehensively identify sites of cytosine DNA methylation (mC and hmC) and mRNA abundance at single-base resolution throughout the genomes of mouse and human frontal cortex (table S1). DNA methylation in embryonic stem (ES) cells occurs in both the CG (mCG) and non-CG (mCH) contexts, but mCH is largely lost upon cell differentiation (22, 23, 25, 26). We found that although mCH levels are negligible in fetal cortex, abundant mCH occurs in adult frontal cortex (Fig. 1A). mCH has previously been identified throughout the genome of the adult mouse brain (20) and at several hundred genomic positions in the human adult brain (21). Supporting previous studies, we found that mammalian brain mCH is typically depleted in expressed genes, with genic mCH level inversely proportional to the abundance of the associated transcript (Fig. 1, A and B) (20). This pattern is the opposite of that observed in ES cells (22) and suggests that genic mCH in the brain may inhibit transcription. The absence of mCH in fetal brain suggests that this signature for gene repression is added to the genome at a later developmental stage.

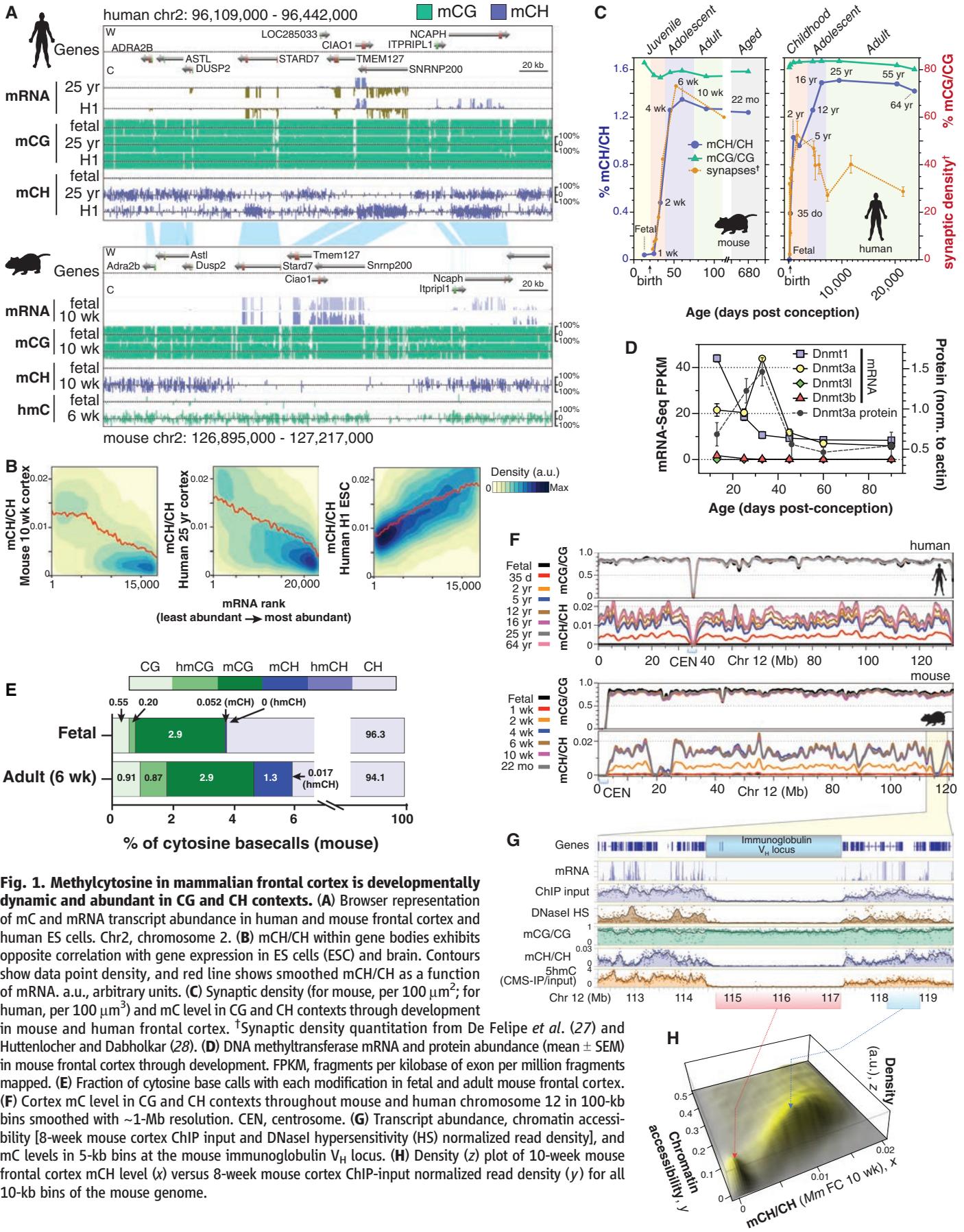
We performed MethylC-Seq on mouse and human frontal cortex during early postnatal, juvenile, adolescent, and adult stages (Fig. 1C). CH methylation level, defined as the fraction of all base calls at CH genome reference positions that were methylated (denoted mCH/CH), accumulates in mouse and human brain during early postnatal development to a maximum of 1.3 to 1.5% genome-wide at the end of adolescence before diminishing slightly during aging. mCH increases most rapidly during the primary phase of synaptogenesis in the developing postnatal brain, from 2 to 4 weeks in mouse (27) and in the first 2 years in humans (28), followed by slower accumulation of mCH during later adolescence. mCH accumulation initially parallels the increase in synapse density within human middle frontal gyrus (synaptogenesis lasts from birth to 5 years), but it subsequently continues to increase during the period of adolescent synaptic pruning, which in humans occurs between 5 and 16 years of age (Fig. 1C). Notably, the accumulation of mCH in mice from 1 to 4 weeks after birth coincides with a transient increase in abundance of the *de novo* methyltransferase *Dnmt3a* mRNA and protein (Fig. 1D). Analysis of the context of mCH sites showed that it is mainly present in the CA context (fig. S1, A to F), as previously reported for mCH (20, 22, 23, 26).

Overall, genomes in the frontal cortex are highly methylated. Whereas CG partially methylated domains (PMDs) account for about a third of the genome of various differentiated human cells (22, 25), human brain genomes have negligible CG PMDs, resembling pluripotent cell methylomes (25) (fig. S1, G and H). Given the high spatial concordance of CG PMDs and nuclear lamina-associated domains reported previously (29), the

<sup>1</sup>Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. <sup>2</sup>Plant Energy Biology [Australian Research Council Center of Excellence (CoE)] and Computational Systems Biology (Western Australia CoE), School of Chemistry and Biochemistry, The University of Western Australia, Perth, WA 6009, Australia. <sup>3</sup>Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. <sup>4</sup>La Jolla Institute for Allergy and Immunology and Sanford Consortium for Regenerative Medicine, La Jolla, CA 92037, USA. <sup>5</sup>Department of Psychiatry, Columbia University and The New York State Psychiatric Institute, New York, NY 10032, USA. <sup>6</sup>Department of Pathology and Cell Biology, Columbia University, New York, NY 10032, USA. <sup>7</sup>Bioinformatics Program, University of California at San Diego, La Jolla, CA 92093, USA. <sup>8</sup>Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA. <sup>9</sup>Cancer Epigenetics Group, Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona 08907, Spain. <sup>10</sup>Department of Medicine, Division of Cardiology, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>11</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain. <sup>12</sup>Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92037, USA. <sup>13</sup>Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA.

\*These authors contributed equally to this work.

†Corresponding author. E-mail: ryan.lister@uwa.edu.au (R.L.); mbehrens@salk.edu (M.M.B.); ecker@salk.edu (J.R.E.)



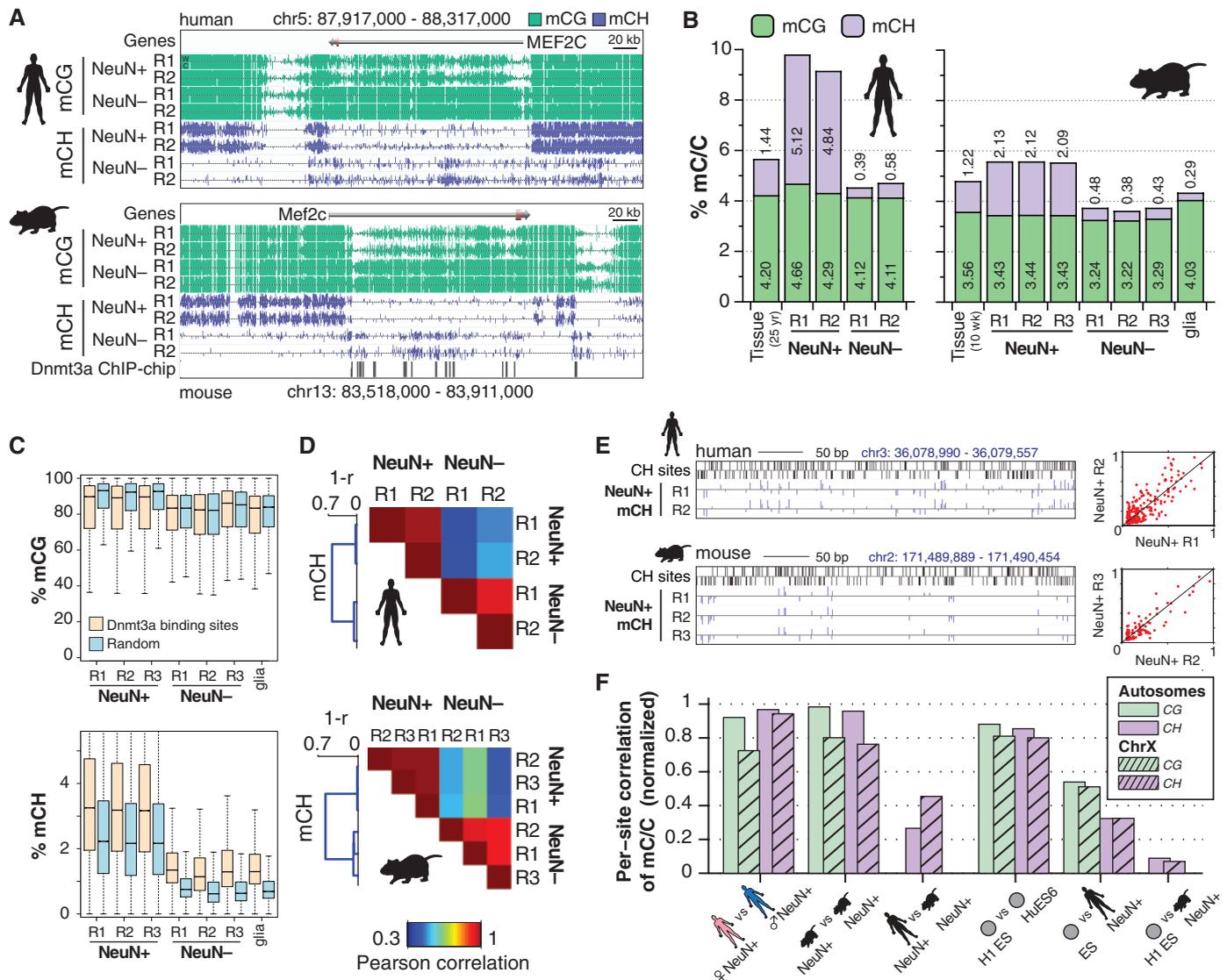
**Fig. 1. Methylation in mammalian frontal cortex is developmentally dynamic and abundant in CG and CH contexts.** (A) Browser representation of mC and mRNA transcript abundance in human and mouse frontal cortex and human ES cells. Chr2, chromosome 2. (B) mCH/CH within gene bodies exhibits opposite correlation with gene expression in ES cells (ESC) and brain. Contours show data point density, and red line shows smoothed mCH/CH as a function of mRNA. a.u., arbitrary units. (C) Synaptic density (for mouse, per  $100 \mu\text{m}^2$ ; for human, per  $100 \mu\text{m}^3$ ) and mC level in CG and CH contexts through development in mouse and human frontal cortex. <sup>†</sup>Synaptic density quantitation from De Felipe *et al.* (27) and Huttenlocher and Dabholkar (28). (D) DNA methyltransferase mRNA and protein abundance (mean  $\pm$  SEM) in mouse frontal cortex through development. FPKM, fragments per kilobase of exon per million fragments mapped. (E) Fraction of cytosine base calls with each modification in fetal and adult mouse frontal cortex. (F) Cortex mC level in CG and CH contexts throughout mouse and human chromosome 12 in 100-kb bins smoothed with  $\sim 1$ -Mb resolution. CEN, centrosome. (G) Transcript abundance, chromatin accessibility [8-week mouse cortex ChIP input and DNaseI hypersensitivity (HS) normalized read density], and mC levels in 5-kb bins at the mouse immunoglobulin V<sub>H</sub> locus. (H) Density (z) plot of 10-week mouse frontal cortex mCH level (x) versus 8-week mouse cortex ChIP-input normalized read density (y) for all 10-kb bins of the mouse genome.

paucity of CG PMDs in these brain methylomes could indicate that lamina-associated domains are altered or much less frequent in the brain.

The adult mammalian brain contains the highest levels of hmC that have been observed (15), accounting for about 40% of methylated CG sites in cerebellar Purkinje cells (30). hmC accumulates during early postnatal brain development in mice (31, 32), becoming enriched in highly expressed genes (33). Given the evidence that hmC can be an intermediate in an active DNA demethylation pathway (16, 17), high-resolution analysis of the genomic distribution of hmC is needed to understand its role in the control of DNA methylation dynamics through brain development. Standard bisulfite-sequencing data does not distin-

guish between methylated and hydroxymethylated sites, so methylcytosines identified by MethylC-Seq analysis represent the sum of these two contributions. Therefore, we used Tet-assisted bisulfite sequencing [TAB-Seq (34)], a base-resolution technique that distinguishes hmC from C and mC genome-wide, to profile hmC in mouse fetal and adult frontal cortex (Fig. 1A). Integration of the genome-wide profiles of mC and hmC enabled a detailed breakdown of the methylated subset of the genome at these distinct developmental stages (Fig. 1E). hmC constitutes 0.20% of total cytosine base calls in fetal cortex and increases to 0.87% in adult cortex. This modification appears to be restricted to the CG context, as also observed in human and mouse ES cells (34); after correction

for false detection, we estimated that 0.017% of cytosine base calls were hmCH genome-wide (99% confidence interval: 0 to 0.059%), and significant hmCH was detected at few individual sites (fig. S2, A and B). The overwhelming presence of hmC in the CG context (99.98%) in mouse adult and fetal frontal cortex is consistent with recent findings in human ES cells, where 99.89% of hmC is in the CG context (34). hmC was present at many highly methylated CG sites (fig. S2C). Therefore, although only a small fraction of all cytosines throughout the genome are methylated (mCG = 2.9%, mCH = 1.3%, hmC = 0.87%), mCH and hmC constitute major, and nonoverlapping, components of the methylated fraction of the genome in adult frontal cortex (mCG = 57.2%,



**Fig. 2. mCH is positionally conserved and is the dominant form of DNA methylation in human neurons.** (A) Browser representation of mCG and mCH in NeuN+ and NeuN- cells. Human NeuN+/NeuN- samples: R1, 53-year-old male; R2, 55-year-old male. Mouse NeuN+/NeuN- samples: R1, 7-week male; R2, 6-week female; R3, 12-month females (not shown). (B) Percentage of methylated base calls in each sequence context throughout the genome. (C) Box and whisker plot of mCG and mCH level in neurons and glia at genomic

regions bound by Dnmt3a versus a random set. Whiskers indicate 1.5 times the interquartile range. (D) mCH correlation between NeuN+ and NeuN- cells in mouse and human, measured in 10-kb bins. (E) Browser representation of mCH sites in neurons. Scatter plots (right) show consistent mCH/CH at all single sites in a 20-kb window overlapping the example region (left). (F) Correlation analysis of methylation state at single sites between neurons and ES cells in human and mouse. Correlation values are normalized by a simulation (62).

mCH = 25.6%, hmC = 17.2%). These data suggest that the steady-state population of hmC in the adult brain is not an intermediate stage in the demethylation of mCH. However, these steady-state measurements do not preclude the possibility that hmCH could be rapidly turned over after conversion from mCH, leading to negligible detected hmCH despite Tet-mediated demethylation at CH sites.

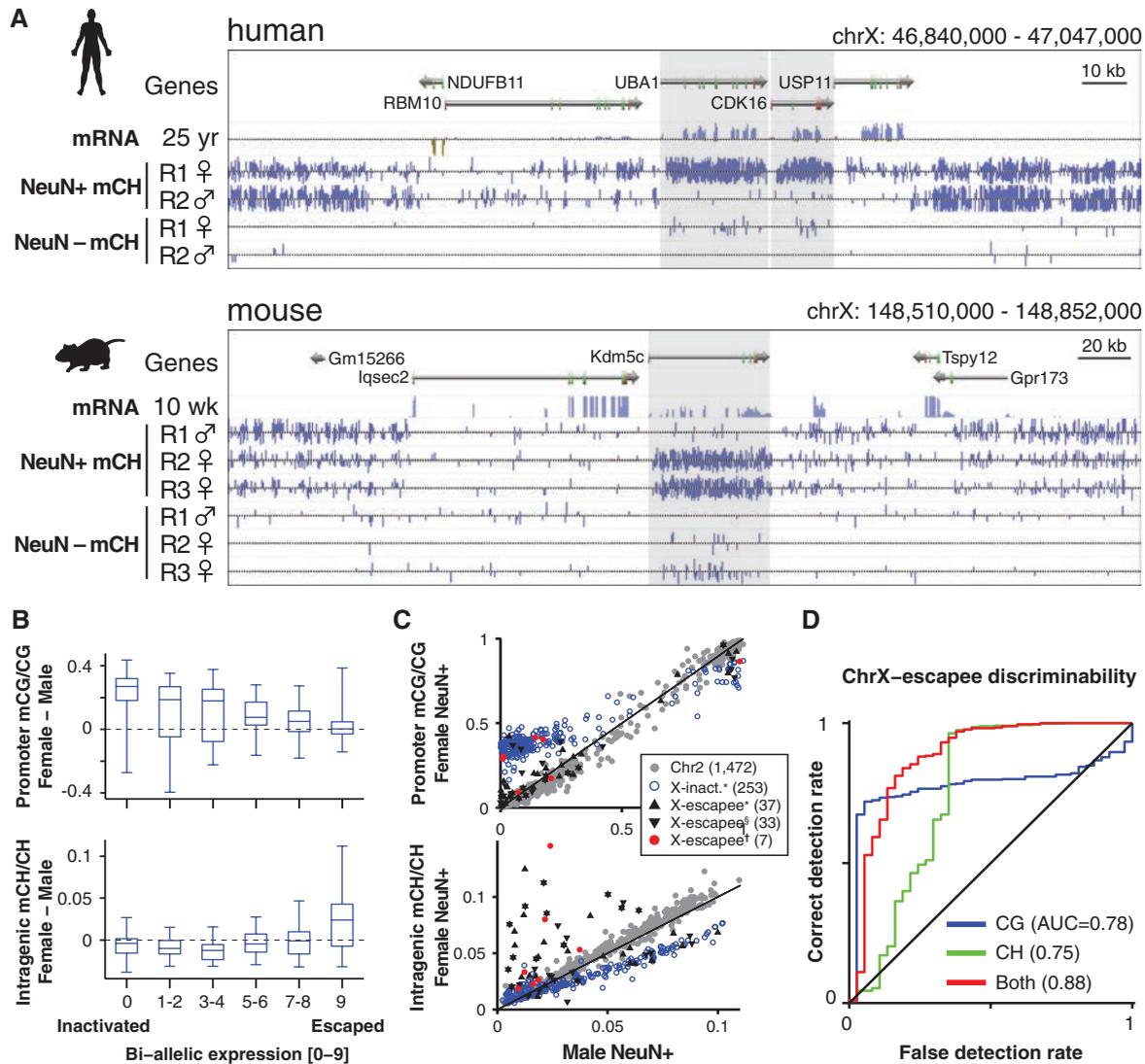
### Protection of Inaccessible Genomic Regions from de Novo Methylation

mCH accumulates in parallel across most of the genome (Fig. 1F). However, we found numerous (36 in human, 34 in mouse) noncentromeric, megabase-sized regions that do not accumulate mCH. These regions, which we termed mCH deserts, are enriched for large gene clusters that encode proteins involved in immunity and recep-

tors required for sensory neuron function (table S2). One mCH desert spans the immunoglobulin  $V_H$  locus, which encodes variable domains of the immunoglobulin heavy chain that rearrange in B lymphocytes. The  $V_H$  locus is transcriptionally quiescent in the frontal cortex of 10-week-old mice, and the chromatin state is highly inaccessible, as inferred from deoxyribonuclease I (DNaseI) hypersensitivity profiling (35) and chromatin immunoprecipitation (ChIP) input sequence read density data (36, 37) (Fig. 1G). In contrast, mCG is not depleted in mCH deserts.

Genome-wide detection of hmC by cytosine 5-methylsulphonate immunoprecipitation (CMS-IP) (38, 39) revealed that hmC is also strongly depleted in the  $V_H$  locus. mCH deserts are observed at other loci in the genome, including olfactory receptor gene clusters that form heterochromatic aggregates required for mono-

allelic receptor expression in olfactory sensory neurons (40, 41). Genome-wide comparison of mCH/CH with chromatin accessibility, as inferred from ChIP input read density (36, 37), for all 10-kb windows of the mouse genome revealed two discrete groups of genomic regions (Fig. 1H). Low-accessibility regions tend to contain minimal mCH, whereas more-accessible regions of the genome show a proportional relationship between genome accessibility and mCH levels. Thus, although mCG is unaffected in these regions, lower chromatin accessibility appears to be highly inhibitory to deposition of mCH and hmC, potentially via inaccessibility to de novo methyltransferases and Tet mC hydroxylases. Furthermore, this indicates that accumulation of mCH and hmC during mammalian brain development occurs via processes that are at least partly independent from methylation at CG dinucleotides.



**Fig. 3. mCH is enriched in genes that escape X inactivation.** (A) Browser representation showing mCH-hypermethylated female human and mouse genes that escape X inactivation (shaded regions). (B) Box and whisker plots of gender differences in promoter mCG and intragenic mCH in inactivated and escapee genes on human chrX. (C) Scatter plot of gender differences in mCG and mCH

in human chrX genes. Reported X inactivated and escapee genes: \*Carrel and Willard (49); †Sharp *et al.* (50); ‡predicted escapee genes, and autosomal (Chr2) genes are indicated. (D) Discriminability analysis of genes that escape female X inactivation using mC data, showing correct versus false detection rate mapped for all possible mC/C thresholds.



synaptic plasticity (43), and regulation of synapse number and function (44). Genome-wide surveys identified 174 mouse genes in which glia were hypermethylated relative to neurons in the CH context (table S3). Unbiased gene ontology analysis revealed that these glial hyper-mCH genes are highly enriched for roles in neuronal and synaptic development and function (table S3). These genes also overlapped significantly with a set of 461 genes expressed at higher levels in neurons than in astrocytes (13-fold higher overlap than chance,  $P < 10^{-30}$ , Fisher exact test, FET) (45) and 233 developmentally up-regulated genes (7.5-fold,  $P < 10^{-7}$ , FET). These genes show hypomethylation of CG and CH in neurons and hypermethylation of CH in glia (fig. S3A), consistent with a potential role of mCH in transcriptional repression of neuronal genes in the glial genome. Furthermore, genes associated with oligodendrocyte or epithelial function accumulate mCH through development (fig. S3B), with oligodendrocyte up-regulated genes showing intragenic mCH hypermethylation in neurons and hypomethylation in glia, whereas epithelial genes

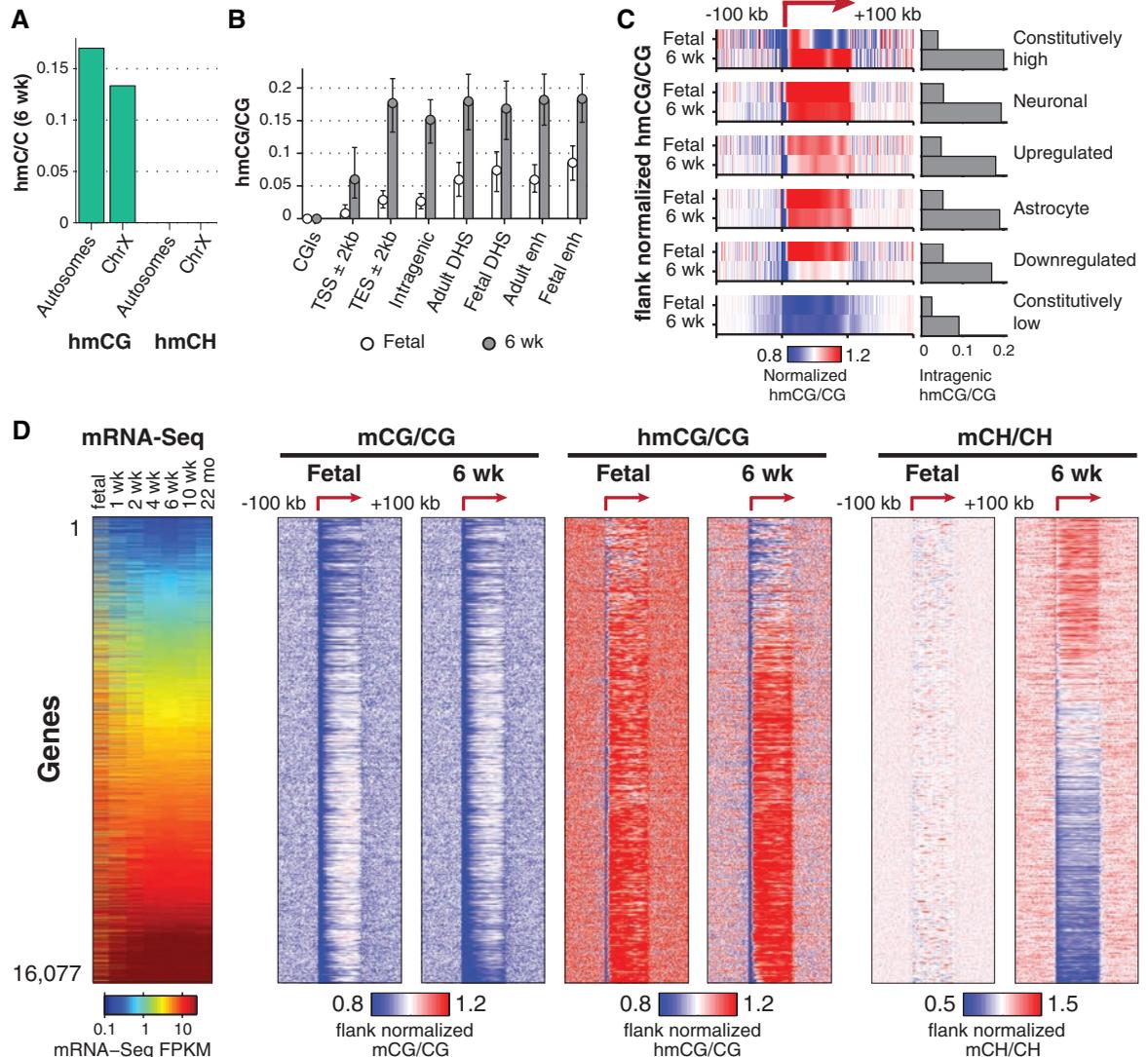
display mCH hypermethylation in both neuronal and glial populations. Consistent with CH methylation requiring Dnmt3a, glial hyper-mCH genes frequently intersect areas of the genome bound by Dnmt3a in mouse postnatal neural stem cells (46). Dnmt3a-binding regions are greatly enriched for mCH, particularly in glia, whereas mCG is not enriched in Dnmt3a-binding regions in glia and is depleted in neurons (Fig. 2C). Thus, there is an association with Dnmt3a binding sites specific to mCH and not mCG, suggesting partial independence between these two marks.

### mCH Position Is Highly Conserved

We examined whether the position of DNA methylation is stochastic or precisely controlled at different genomic scales. The level of mCH in 10-kb windows throughout the genome was highly reproducible between independent samples of the same cell type, with lower, but substantial, correlation between cell types (Fig. 2D). Closer inspection revealed consistency between the methylation level at individual mCH sites in neurons from different individuals in both mice and hu-

mans (Fig. 2E). At single-base resolution (fig. S4), perfect correlation between individuals would not be observed even if the true methylation level were identical at each site because of the stochastic effect of a finite number of sequenced reads. To correct for this, we normalized the observed correlation by that of simulated data sets with the same coverage per site as each of our experimental samples but with identical methylation levels (Fig. 2F and fig. S4). To assess statistical significance, we used a permutation test, which compared the data correlation with the correlation after randomly shuffling the relative positions of CH sites in each sample (fig. S4). This revealed that autosomal CG and CH sites have nearly identical methylation levels in neuronal populations isolated from different individuals of the same species. Observed differences could be explained by stochastic sampling rather than true individual variation. Unexpectedly, normalized per-site correlation is higher for mCH than mCG between neuronal populations isolated from the frontal cortices of different human individuals, and mouse neuronal mCG and mCH per-site

**Fig. 5. hmCG is enriched within active genomic regions in fetal and adult mouse brain.** (A) hmCG level in 6-week mouse frontal cortex for autosomes and ChrX. (B) Median hmCG level within genomic features (error bars 32nd to 68th percentile). enh, enhancer. (C) Median normalized hmCG throughout gene body and flanking 100 kb for indicated gene sets. Bars show absolute hmCG/CG levels within gene bodies for each class. (D) mC and hmC throughout gene body and flanking 100 kb for each. Transcript abundance (mRNA-Seq FPKM) during mouse development is also shown (left).



correlations are equivalent. Per-site correlation between two human ES cell lines (H1 and HUES6) is also high (>0.8) for both mCG and mCH.

The high interindividual correlation of mCH at the kilobase and single-site scales indicates that methylation of CH positions, particularly in mammalian neurons, is a highly controlled process. It is not consistent with a stochastic event that takes place at any available CH position in a particular genomic region that accumulates mCH. Comparison of mCH between human and mouse neurons at conserved exonic CH positions revealed a low but significant interspecies correlation (Fig. 2F;  $P < 0.005$ , shuffle test), possibly indicating conservation of the cellular processes that precisely target or restrict mCH at these positions. Last, per-site mCG and mCH correlation between human ES cells and neurons is significantly lower, likely because of differences in the processes governing methylation of particular genomic features in the distinct cell types, for example, enrichment and depletion of mCH in highly transcribed genes in ES cells and neurons, respectively (Fig. 1, A and B).

The precise conservation of mCH position may be partly caused by the physical configuration of DNA within nucleosomes. Consistent with this, neuronal mCH patterns contain robust periodic components at the scale of nucleosome spacing [~170 base pairs (bp), fig. S5A] and the DNA helix coil length (~10.5 bp, fig. S5B). Such periodic components may arise from sequence-dependent constraints on mCH position, which would be the same in every neuronal cell. Alternatively, epigenetic heterogeneity within the population of NeuN+ nuclei in our sample may lead to stronger correlation for CH sites located on the same physical chromosome, compared with the correlation between the same locations on chromosomes from different cells. To test this, we measured the cross-correlation within individual reads, revealing a contribution of within-chromosome correlation to the periodic methylation pattern (fig. S5C).

### Gender-Specific DNA Methylation Patterns on the X Chromosome

Interindividual correlation of mCG and mCH on chromosome X (ChrX) is frequently lower than on autosomes (Fig. 2F), prompting a closer analysis of ChrX mC patterns. ChrX mCG and mCH levels were generally lower in females compared with males, presumably because of the effect of ChrX inactivation (fig. S5, F and G) (47, 48). However, a subset of genes in both humans and mice have significantly greater intragenic mCH levels in females compared with males (Fig. 3A). Inspection of these genes revealed that most were previously found to escape inactivation in human females (X-escapees), displaying biallelic expression (49) and a reduction in promoter mCG hypermethylation, a DNA methylation signature of inactivated alleles (50). Quantification of human gender differences in neuronal DNA methylation for ChrX genes previously characterized as show-

ing biallelic expression (49) revealed that females have reduced promoter mCG and a large increase in intragenic mCH but not intragenic mCG (Fig. 3, B and C, and fig. S5, D and E). The sequence composition of mCH is very similar in the whole genome, within autosomal gene bodies, and within X-chromosome inactivated and escapee gene bodies (fig. S5H). Analysis of gender-specific methylation in additional human cell types revealed that female promoter mCG hypomethylation is observed at X-escapee genes in glia and human embryonic stem cells (fig. S6). Intragenic mCH hypermethylation of X-escapees was also observed in female glia, albeit to a lesser extent than in neurons, but was not present in ES cells. Thus, X-escapee mCH hypermethylation may be a feature that is specific to neural cell types. Although both promoter CG hypomethylation and intragenic CH hypermethylation provide significant information for discriminating X-escapees [Fig. 3D, discriminability index (area under the curve, AUC) = 0.75 and 0.78, respectively], combining both mCG and mCH measurements boosts discriminability (AUC = 0.88). By using this intragenic mCH hypermethylation signature, we identified seven new putative X-escapee genes (table S4). On the basis of these data, we hypothesize that intragenic CH hypermethylation in neurons may play a compensatory role in genes that fail to acquire repressive CG hypermethylation in the promoter region, restoring equal gene expression between male and female cells (51).

### Distinct Genic DNA Methylation States Demarcate Functionally Relevant Gene Clusters

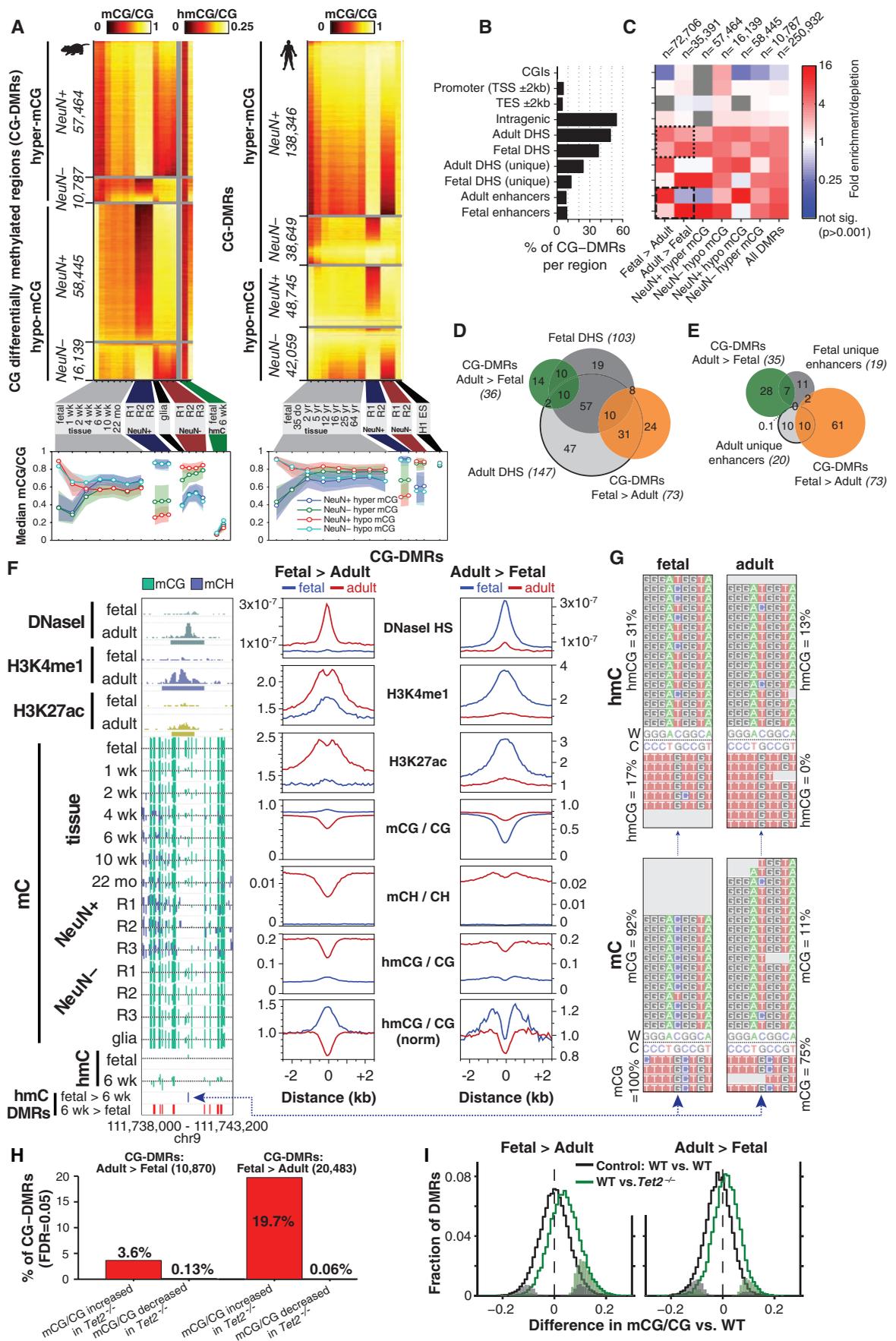
DNA methylation within promoter regions and in gene bodies is implicated in regulation of gene expression (22, 52), suggesting that the precisely conserved, cell type-specific DNA methylation patterns may be related to specific neuronal and glial cellular processes. We therefore used an unbiased approach to classify patterns of mCG and mCH within each annotated gene body and in flanking regions extending 100 kb up- or downstream. After normalizing the methylation pattern around each autosomal gene by the local baseline mCG or mCH level in each adult neuronal or glial sample, we combined these features into a large data matrix containing 4200 individual DNA methylation measurements for each gene [seven samples, two contexts (CG and CH), 300 1-kb bins within and around each gene]. Using principal component (PC) analysis, we extracted five methylation features (PCs) that together account for 46% of the total data set variance (fig. S7A). Gene sets with specific neuronal or astrocytic expression, as well as ChrX genes, segregate within PC space (fig. S6B). We then used *k*-means clustering to classify all genes into 15 clusters on the basis of their mCG and mCH patterns (Fig. 4A and fig. S8). Several dominant patterns of DNA methylation and transcript abundance and dynamics between developmental and cellular states are evident. A cluster of genes that progressively loses gene-body mCG and mCH through development

contains constitutively highly expressed genes that are strongly enriched for neuronal function and depleted for astrocyte-specific roles (Fig. 4A, box 1). These genes show intragenic mCG enrichment in glia and depletion in neurons (box 2), indicating that glial gene body mCG resembles that of the neural precursor cells that predominate the fetal brain. This indicates that the loss of mCG in brain tissue during development is due to CG hypomethylation in mature neurons. These constitutively highly expressed genes enriched for neuronal function also show extensive intragenic mCH hypomethylation in neurons in contrast to glia (box 3), and they are enriched for hmCG (box 4) as previously described (32, 33). Genes that are not as highly transcribed, but that are associated with neuronal function and are developmentally up-regulated, also show intragenic mCG and mCH hypomethylation in neurons but not glia (box 5). For these gene sets, mCG and mCH enrichment or depletion is precisely localized to transcribed regions, suggesting that this modification of genic mC is tightly coupled to transcription. Notably, the bodies of constitutively high genes that are not enriched for neuronal function (box 6) do not show marked fetal/glial mCG enrichment or neuronal mCH depletion, indicating that this differential methylation is specific for genes enriched for neuronal function and not simply an association with particular levels of transcriptional activity. Genes associated with astrocyte function show an opposite pattern to genes associated with neuronal function: a progressive increase in intragenic mCG and mCH in frontal cortex tissue over development, neuronal mCG and mCH hypermethylation, and glial mCG and mCH hypomethylation (Fig. 4A, boxes 7 to 9). Last, genes with constitutively low expression do not show developmental or cell type-specific DNA methylation patterns (box 10), demonstrating that dynamic DNA methylation in genes is highly associated with differential transcriptional activity in mammalian brain development and neural cell specialization.

Each of the gene clusters identified in our unbiased analysis was significantly enriched or depleted for cell type-specific function [neuronal or astrocytic genes (45)] or particular expression patterns (constitutively high or low expression, developmentally up- or down-regulated) (Fig. 4B). Profiling the median mCG and mCH of genes within each of these categories allows direct comparison of developmental and cell type-specific DNA methylation in mouse (Fig. 4C) and human (fig. S3C). This analysis recapitulates many of the conclusions of the unbiased clustering (Fig. 4A).

The inverse relationship observed between genic mCH level and transcriptional activity is consistent with a model whereby intragenic accumulation of mCH impedes transcriptional activity. Alternatively, the process of transcription could interfere with mCH de novo methylation or induce active mCH demethylation, although these are not consistent with the DNMT3A-dependent intragenic mCH in human embryonic stem cells that is positively correlated with gene expression

**Fig. 6. Developmental and cell type-specific differential mCG.** (A) Heat map of absolute mCG level in CG-DMRs identified between neurons and glia and over development in mouse (left) and human (right). (B) Fraction of all CG-DMRs located in distinct genomic features in mouse. (C) Enrichment or depletion of distinct cell type-specific and developmental CG-DMR sets within genomic features from (B). (D and E) Intersection of developmentally dynamic CG-DMRs and (D) DNaseI hypersensitive sites or (E) enhancers in mouse brain in thousands. (F) Browser representation of mouse developmentally dynamic CG-DMRs and quantification of local enrichment of chromatin modifications, genome accessibility, and mC. (G) TAB-Seq reads showing fetal-specific hmCG in the Fetal > Adult CG-DMR in mouse. (H) Proportion of mouse developmental CG-DMRs where mCG/CG is significantly increased or decreased in *Tet2* knockout mice. (I) Distribution of mCG level difference between wild-type (WT) and *Tet2* mutant at mouse CG-DMRs. Significantly different DMRs are indicated by coloration.



(22, 23). Overall, glial mCG and mCH patterns closely resemble those of the fetal and the early postnatal brain, indicating that DNA methylation in early mammalian brain developmental stages may be a default state that largely persists through to maturity in glial cells, whereas neuronal differentiation and maturation involve extensive re-configuration of the DNA methylome that is highly associated with cell type-specific changes in transcriptional activity.

### hmCG Is Enriched Within Active Genomic Regions in Fetal and Adult Mouse Brain

Our base-resolution analysis of hmC using TAB-Seq revealed that intragenic and global hmCG levels are largely equivalent between chromosomes, whereas hmCG/CG is 22% lower on the male ChrX, consistent with previous reports from enrichment based detection of hmC (32, 33) (Fig. 5A). Analysis of hmCG levels in different genomic regions revealed that, although adult hmCG/CG is similar across transcriptional end sites and intragenic, DNaseI-hypersensitive (DHS), and enhancer regions, the fetal frontal cortex shows a relative enrichment of hmCG in DHS regions and enhancers, in particular enhancer regions that are unique to the fetal developmental stage (Fig. 5B). The inverse pattern can be observed for adult mCG levels, which are lower in DHS regions and enhancers (fig. S9, A and B), suggesting that regions of relatively high hmCG levels in the fetal brain show relatively low mCG levels in the adult brain. Analysis of intragenic hmCG enrichment relative to flanking genomic regions, for cell type-specific or developmentally dynamic gene sets (Fig. 5C), showed that neuronal and astrocyte gene bodies that are highly enriched with hmCG in adult are also highly enriched at the fetal stage.

Thus, despite lower absolute levels of intragenic hmCG in the fetal stage, the adult patterns of hmCG enrichment at these cell type-specific genes are already forming in utero. Constitutively lowly expressed genes show intragenic depletion of hmCG, in contrast to constitutively highly transcribed genes, which show localized enrichment of hmCG throughout part or all of the gene body. Developmentally down-regulated genes show enrichment of hmCG in the fetal frontal cortex but not in adults, indicating that reduced transcription is accompanied by a loss of hmCG enrichment. Overall, transcriptional activity is associated with intragenic hmCG enrichment, as reported (33), with in utero establishment of adult hmCG patterns for cell type-specific genes and loss of hmC enrichment associated with developmentally coupled transcriptional down-regulation.

Measurement of mC and hmC in all genes in fetal and adult mouse frontal cortex indicated that both mCG and hmCG are depleted at promoters and in gene bodies of lowly expressed genes, whereas hmCG is enriched throughout the gene bodies of more highly transcribed genes (Fig. 5D). The most highly expressed genes in the adult frontal cortex show intragenic mCG hypomethylation (Figs. 4 and 5D) but still retain high intra-

genic hmCG. Ranking all genes by transcript abundance, it is evident that the highest mean intragenic hmCG levels, which occur in the most highly transcribed genes, correspond to hmCG/CG  $\sim 0.25$  and mCG/CG  $\sim 0.5$  (fig. S10D). Frontal cortex development is accompanied by increased enrichment of hmCG at intragenic regions that are already hyper-hydroxymethylated at the fetal stage (Fig. 5D and fig. S10), demonstrating that adult patterns of genic hmC are already evident in the immature fetal brain.

### CG Differentially Methylated Regions Enriched in Regulatory Regions

Because differences in genic mCG were observed over development and between neuronal and glial cell populations (Fig. 4), we scanned the human and mouse methylomes to comprehensively identify CG differentially methylated regions (CG-DMRs) throughout the genome. CG-DMRs were identified between fetal and adult frontal cortex, neurons and glia, and combined into four sets: neuronal and glial hyper- and hypo-methylated CG-DMRs. In total, 267,799 human and 142,835 mouse CG-DMRs were identified (median lengths: for mouse, 473 bp; human, 533 bp), revealing several predominant dynamics in mCG during brain development and cellular specialization (Fig. 6A and fig. S11). Neuronal CG-DMRs are the most numerous in both mice and humans, because of the very distinct mCG patterns that emerge during neuronal differentiation and maturation. At these sites, CG methylation in adult neurons is distinct compared with those in glial and/or fetal and early postnatal development frontal cortex tissue samples. Neuronal hypermethylated CG-DMRs also show mCH hypermethylation (fig. S11). In mouse, mCG/CG within neuronal hypomethylated CG-DMRs declines to a stable level by 1 week after birth. In contrast, neuronal hypermethylated CG-DMRs do not begin to change until 1 week after birth, after which they accumulate mCG until 2 to 4 weeks of age. These data indicate that increases in neuronal mCG occur during synaptogenesis after most decreases in neuronal mCG have already occurred. Furthermore, we found that hydroxymethylation in the adult cortex is highest in CG-DMRs that show neuronal hypermethylation and is depleted from CG-DMRs that display neuronal hypomethylation (Fig. 6A). This suggests that hmCG may be most abundant in neurons, rather than glial cells, in the frontal cortex.

Analysis of the genomic features in which CG-DMRs are located revealed that although half are found within gene bodies, they are not common within promoters and transcriptional start and end regions. Instead, they are disproportionately located at DHS regions and enhancers unique to fetal or adult brain (Fig. 6B). Closer inspection of the enrichment and depletion of these CG-DMRs revealed that fetal enhancers and DHS sites unique to the fetal brain are enriched for hypermethylation in adult brain but not in the fetal brain (Fig. 6C). In contrast, adult enhancers and unique adult DHS sites are highly associated with CG hyper-

methylation in fetal stages but are not associated with hypermethylated CG-DMRs in the adult brain and in neurons. Thus, developmentally dynamic enhancers and DHS sites in frontal cortex have dynamic CG methylation that is depleted where chromatin accessibility and regulatory element activity increase, consistent with a range of human cell lines (53).

To characterize gene functions associated with the CG-DMRs, we analyzed the association between proximal genes (transcriptional start site within 5 kb of the DMR) and cell type-specific or developmentally dynamic gene sets (fig. S12A). We observed an inverse relationship between methylation and gene function. Genes associated with neuronal function and up-regulation during development are enriched for promoter hypermethylation in glia and hypomethylation in neurons, whereas genes down-regulated during brain development and those related to astrocyte function are enriched for promoter hypermethylation in neurons and hypomethylation in glia. Genes that are constitutively expressed at either high or low levels are not associated with promoter/transcription start site CG-DMRs, indicating that dynamic CG methylation is highly associated with changes in transcriptional activity and cell type-specific transcriptional regulation.

Because the majority of all developmentally dynamic CG-DMRs are associated with DHS sites, we examined the directional relationships between dynamic mCG and DNA accessibility states over development (Fig. 6D). Notably, DHS sites unique to fetal frontal cortex overlap with 28% of CG-DMRs that gain methylation through development (Adult>Fetal). However, these sites only overlap 7.3% of CG-DMRs that lose mCG during development (Fetal>Adult). Similarly, DHS sites unique to adult frontal cortex rarely overlap Adult>Fetal CG-DMRs. A similar analysis of developmentally dynamic enhancers active in only one of the developmental stages (37) (Fig. 6E) showed that enhancer activation is associated with mCG hypomethylation of the enhancer, whereas enhancer inactivation is associated with enhancer mCG hypermethylation.

This inverse relationship between genome accessibility and mCG level at putative functional regions of the genome suggests that nuclear factors that bind the region and increase accessibility may cause localized reduction in mCG, as previously reported for a small number of DNA binding proteins (54). Alternatively, mCG hypermethylation may cause reduced genome accessibility by direct inhibition of DNA-protein interactions or induction of chromatin compaction, with loss of mCG enabling increased chromatin accessibility and genome interaction with DNA binding factors.

Discrete regions that show increased or decreased CG methylation through development are associated with specific local chromatin modifications. We found that CG-methylated regions of the fetal frontal cortex that become hypomethylated in the adult (Fig. 6F, Fetal>Adult) gain

localized histone modifications characteristic of active enhancers (H3K4me1 and H3K27ac) and increased DNaseI hypersensitivity in the adult. In addition, these regions have reduced accumulation of mCH (Fig. 6F), consistent with an overall decrease in mC linked to increased genome accessibility and enhancer activity. In contrast, genomic regions that gain mCG during development (Adult>Fetal) lose localized enrichment of H3K4me1, H3K27ac, and DNaseI hypersensitivity and show increased mCH. These changes indicate inactivation of these genomic regions through brain development and suggest that this inactivation is associated with increased local mCG.

### A Hydroxymethylation Signature of Developmentally Activated Regions

Adult>Fetal CG-DMRs show broad low-level hmCG enrichment flanking the CG-DMR and a localized depletion of hmC at the center in both fetal and adult genomes (Fig. 6F), and the absolute abundance of hmC is several-fold lower in fetal compared with adult frontal cortex (Fig. 1E). This suggests that although Adult>Fetal CG-DMRs gain mCG through development, they tend to be refractory to conversion to hmC, potentially because of lower accessibility to the Tet hydroxylases. In contrast, Fetal>Adult CG-DMRs have a local enrichment of both mCG and hmCG in the fetal cortex that becomes a local depletion in the adult. Two enrichment-based genome-wide hmC profiling techniques, CMS-IP (38) and biotin-glucosyl tagging (31), confirmed the localized enrichment of hmC at Fetal>Adult CG-DMRs (fig. S12B). The localized enrichment of hmC at these inaccessible and quiescent genomic regions, which lose mCG and hmCG later in development, indicates that they may be premodified with hmCG in the fetal stage to create a dormant state that is poised for subsequent demethylation and activation at a later developmental stage. Closer inspection of base-resolution hmC data revealed that 4% of the hmCG bases that have significantly higher hmC levels in fetal compared with adult [false discovery rate (FDR) 0.05] directly overlap with Fetal>Adult CG-DMRs, far exceeding the number expected by chance (0.5%). This indicates that despite lower global levels of hmC in the fetal brain, developmentally demethylated CG-DMRs are enriched for hmCG bases that are more highly hydroxymethylated in fetal than in adult brain (Fig. 6G). The localized enrichment of mCG at these CG-DMRs in the fetal cortex indicates that CG-demethylation has not yet taken place.

If fetal hmCG is poised at dormant genomic regions in order to facilitate active DNA demethylation at later developmental stages, then the Tet hydroxylase enzymes that catalyze conversion of mC to hmC should be necessary for mCG hypomethylation in the adult frontal cortex at these regions. To test this, we performed MethylC-Seq of genomic DNA from frontal cortex tissue of adult *Tet2*<sup>-/-</sup> mice. Adult>Fetal CG-DMRs, which gain mCG through development, are largely unaffected in *Tet2*<sup>-/-</sup> compared with wild-type adult mice

(Fig. 6, H and I; 3.6% hypermethylated, FET, FDR 0.05). By contrast, a substantial fraction of Fetal>Adult CG-DMRs are hypermethylated in *Tet2*<sup>-/-</sup> (19.7%) versus wild type. The mutant shows a small but significant increase in mCG at Fetal>Adult CG-DMRs (Fig. 6I and fig. S12C) (*Tet2*<sup>-/-</sup>: 7.9% ± 4.6%,  $P < 10^{-11}$ , Wilcoxon signed rank test). The partial effect of the mutation on CG methylation is not unexpected given that all three *Tet* genes are expressed in the brain (fig. S9C) and may exhibit some functional redundancy. Additionally, a genome-wide search identified 14,340 CG-DMRs hypermethylated in *Tet2*<sup>-/-</sup> relative to wild type (6 weeks, 10 weeks, and 22 months), >fourfold more numerous than hypomethylated CG-DMRs (3099). This further indicates a role for Tet2 in mediating mCG demethylation during brain development.

### Discussion

The essential role of frontal cortex in behavior and cognition requires the coordinated interaction, via electrical and chemical signaling, of multiple neuronal cell types and a diverse population of glial cells. Individual brain cells have unique roles within circuits that are defined by their location and pattern of connections as well as by their molecular identity. The development and maturation of the brain's physical structure and the refinement of the molecular identities of neurons and glial cells occur in parallel in a finely orchestrated process that starts early during the embryonic period and continues, in humans, well into the third decade of life (55, 56). An early postnatal burst of synaptogenesis is followed by activity-dependent pruning of excess synapses during adolescence (28, 57, 58). This process forms the basis for experience-dependent plasticity and learning in children and young adults (59), and its disruption leads to behavioral alterations and neuropsychiatric disorders (60). During this period, profound transcriptional changes lead to the appearance of adult electrophysiological characteristics in neocortical neurons.

Our study suggests a key role of DNA methylation in brain development and function. First, CH methylation accumulates significantly in neurons through early childhood and adolescence, becoming the dominant form of DNA methylation in mature human neurons. This shows that the period of synaptogenesis, during which the neural circuit matures, is accompanied by a parallel process of large-scale reconfiguration of the neuronal epigenome. Indeed, central nervous system deletion of *Dnmt3a* during late gestation induces motor deficits, and animals die prematurely (61). However, mice with a postnatal deletion restricted to the pyramidal cell population (complete recombination around 1 month old) do not show overt behavioral or transcriptional alterations (2). Our data suggest that expression of *Dnmt3a* specifically around the second postnatal week may be critical for establishing a normal brain DNA methylation profile and allowing healthy brain development.

Second, the precise positioning of mCG and mCH marks, which are conserved between individuals and across humans and mice, is consistent with a functional role. Whether this is the case, or whether the conserved patterns are instead a reflection of conserved nucleosome position or chromatin structure, requires further investigation. Third, the relationship between DNA methylation patterns and the function of neuron- or astrocyte-specific gene sets suggests a role for DNA methylation in distinguishing these two broad classes of cortical cells. DNA methylation could therefore play a key role in sculpting more-specific cellular identities. If this is the case, we expect that purified subpopulations will reveal high specificity of methylation at specific sites for particular cell types. Thus, the observation that most CH sites with nonzero methylation are methylated in ~20 to 25% of sampled cells (fig. S1H) could be explained by the heterogeneity of these brain circuits rather than by stochastic methylation within each cell. These conclusions obtained from our genome-wide, base-resolution, cell type-specific DNA methylomes for brain cells through key stages of development are the first steps toward unraveling the genetic program and experience-dependent epigenetic modifications leading to a fully differentiated nervous system.

### References and Notes

1. E. Borrelli, E. J. Nestler, C. D. Allis, P. Sassone-Corsi, Decoding the epigenetic language of neuronal plasticity. *Neuron* **60**, 961–974 (2008). doi: [10.1016/j.neuron.2008.10.012](https://doi.org/10.1016/j.neuron.2008.10.012); pmid: [19109904](https://pubmed.ncbi.nlm.nih.gov/19109904/)
2. J. Feng *et al.*, Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13**, 423–430 (2010). doi: [10.1038/nn.2514](https://doi.org/10.1038/nn.2514); pmid: [20228804](https://pubmed.ncbi.nlm.nih.gov/20228804/)
3. C. A. Miller, J. D. Sweatt, Covalent modification of DNA regulates memory formation. *Neuron* **53**, 857–869 (2007). doi: [10.1016/j.neuron.2007.02.022](https://doi.org/10.1016/j.neuron.2007.02.022); pmid: [17359920](https://pubmed.ncbi.nlm.nih.gov/17359920/)
4. S. Numata *et al.*, DNA methylation signatures in development and aging of the human prefrontal cortex. *Am. J. Hum. Genet.* **90**, 260–272 (2012). doi: [10.1016/j.ajhg.2011.12.020](https://doi.org/10.1016/j.ajhg.2011.12.020); pmid: [22305529](https://pubmed.ncbi.nlm.nih.gov/22305529/)
5. M. Suderman *et al.*, Conserved epigenetic sensitivity to early life experience in the rat and human hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* **109** (suppl. 2), 17266–17272 (2012). doi: [10.1073/pnas.1121260109](https://doi.org/10.1073/pnas.1121260109); pmid: [23045659](https://pubmed.ncbi.nlm.nih.gov/23045659/)
6. K. D. Siegmund *et al.*, DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS ONE* **2**, e895 (2007). doi: [10.1371/journal.pone.0000895](https://doi.org/10.1371/journal.pone.0000895); pmid: [17878930](https://pubmed.ncbi.nlm.nih.gov/17878930/)
7. K. Martinowich *et al.*, DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* **302**, 890–893 (2003). doi: [10.1126/science.1090842](https://doi.org/10.1126/science.1090842); pmid: [14593184](https://pubmed.ncbi.nlm.nih.gov/14593184/)
8. J. U. Guo *et al.*, Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat. Neurosci.* **14**, 1345–1351 (2011). doi: [10.1038/nn.2900](https://doi.org/10.1038/nn.2900); pmid: [21874013](https://pubmed.ncbi.nlm.nih.gov/21874013/)
9. I. B. Zovkic, M. C. Guzman-Karlsson, J. D. Sweatt, Epigenetic regulation of memory formation and maintenance. *Learn. Mem.* **20**, 61–74 (2013). doi: [10.1101/lm.026575.112](https://doi.org/10.1101/lm.026575.112); pmid: [23322554](https://pubmed.ncbi.nlm.nih.gov/23322554/)
10. J. M. Levenson *et al.*, Evidence that DNA (cytosine-5) methyltransferase regulates synaptic plasticity in the hippocampus. *J. Biol. Chem.* **281**, 15763–15773 (2006). doi: [10.1074/jbc.M511767200](https://doi.org/10.1074/jbc.M511767200); pmid: [16606618](https://pubmed.ncbi.nlm.nih.gov/16606618/)
11. A. M. M. Oliveira, T. J. Hemstedt, H. Bading, Rescue of aging-associated decline in *Dnmt3a2* expression restores

- cognitive abilities. *Nat. Neurosci.* **15**, 1111–1113 (2012). doi: [10.1038/nn.3151](https://doi.org/10.1038/nn.3151); pmid: [22751036](https://pubmed.ncbi.nlm.nih.gov/22751036/)
12. P. Moretti *et al.*, Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. *J. Neurosci.* **26**, 319–327 (2006). doi: [10.1523/JNEUROSCI.2623-05.2006](https://doi.org/10.1523/JNEUROSCI.2623-05.2006); pmid: [16399702](https://pubmed.ncbi.nlm.nih.gov/16399702/)
  13. N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* **126**, 781–790 (1972). pmid: [4538516](https://pubmed.ncbi.nlm.nih.gov/4538516/)
  14. N. W. Penn, Modification of brain deoxyribonucleic acid base content with maturation in normal and malnourished rats. *Biochem. J.* **155**, 709–712 (1976). pmid: [949331](https://pubmed.ncbi.nlm.nih.gov/949331/)
  15. D. Globisch *et al.*, Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010). doi: [10.1371/journal.pone.0015367](https://doi.org/10.1371/journal.pone.0015367); pmid: [21203455](https://pubmed.ncbi.nlm.nih.gov/21203455/)
  16. J. U. Guo, Y. Su, C. Zhong, G.-L. Ming, H. Song, Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011). doi: [10.1016/j.cell.2011.03.022](https://doi.org/10.1016/j.cell.2011.03.022); pmid: [21496894](https://pubmed.ncbi.nlm.nih.gov/21496894/)
  17. Y.-F. He *et al.*, Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011). doi: [10.1126/science.1210944](https://doi.org/10.1126/science.1210944); pmid: [21817016](https://pubmed.ncbi.nlm.nih.gov/21817016/)
  18. C.-X. Song *et al.*, Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013). doi: [10.1016/j.cell.2013.04.001](https://doi.org/10.1016/j.cell.2013.04.001); pmid: [23602153](https://pubmed.ncbi.nlm.nih.gov/23602153/)
  19. L. Shen *et al.*, Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706 (2013). doi: [10.1016/j.cell.2013.04.002](https://doi.org/10.1016/j.cell.2013.04.002); pmid: [23602152](https://pubmed.ncbi.nlm.nih.gov/23602152/)
  20. W. Xie *et al.*, Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816–831 (2012). doi: [10.1016/j.cell.2011.12.035](https://doi.org/10.1016/j.cell.2011.12.035); pmid: [22341451](https://pubmed.ncbi.nlm.nih.gov/22341451/)
  21. K. E. Varley *et al.*, Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **23**, 555–567 (2013). doi: [10.1101/gr.147942.112](https://doi.org/10.1101/gr.147942.112); pmid: [23325432](https://pubmed.ncbi.nlm.nih.gov/23325432/)
  22. R. Lister *et al.*, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009). doi: [10.1038/nature08514](https://doi.org/10.1038/nature08514); pmid: [19829295](https://pubmed.ncbi.nlm.nih.gov/19829295/)
  23. M. J. Ziller *et al.*, Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet.* **7**, e1002389 (2011). doi: [10.1371/journal.pgen.1002389](https://doi.org/10.1371/journal.pgen.1002389); pmid: [22174693](https://pubmed.ncbi.nlm.nih.gov/22174693/)
  24. R. Lister *et al.*, Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**, 523–536 (2008). doi: [10.1016/j.cell.2008.03.029](https://doi.org/10.1016/j.cell.2008.03.029); pmid: [18423832](https://pubmed.ncbi.nlm.nih.gov/18423832/)
  25. R. Lister *et al.*, Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68–73 (2011). doi: [10.1038/nature09798](https://doi.org/10.1038/nature09798); pmid: [21289626](https://pubmed.ncbi.nlm.nih.gov/21289626/)
  26. B. H. Ramsahoye *et al.*, Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5237–5242 (2000). doi: [10.1073/pnas.97.10.5237](https://doi.org/10.1073/pnas.97.10.5237); pmid: [10805783](https://pubmed.ncbi.nlm.nih.gov/10805783/)
  27. J. De Felipe, P. Marco, A. Fairén, E. G. Jones, Inhibitory synaptogenesis in mouse somatosensory cortex. *Cereb. Cortex* **7**, 619–634 (1997). doi: [10.1093/cercor/7.7.619](https://doi.org/10.1093/cercor/7.7.619); pmid: [9373018](https://pubmed.ncbi.nlm.nih.gov/9373018/)
  28. P. R. Huttenlocher, A. S. Dabholkar, Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178 (1997). doi: [10.1002/\(SICI\)1096-9861\(19971020\)387:2<167::AID-CNE1>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1096-9861(19971020)387:2<167::AID-CNE1>3.0.CO;2-Z); pmid: [9336221](https://pubmed.ncbi.nlm.nih.gov/9336221/)
  29. B. P. Berman *et al.*, Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* **44**, 40–46 (2012). doi: [10.1038/ng.969](https://doi.org/10.1038/ng.969); pmid: [22120008](https://pubmed.ncbi.nlm.nih.gov/22120008/)
  30. S. Kraucionis, N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009); [10.1126/science.1169786](https://doi.org/10.1126/science.1169786). doi: [10.1126/science.1169786](https://doi.org/10.1126/science.1169786); pmid: [19372393](https://pubmed.ncbi.nlm.nih.gov/19372393/)
  31. C.-X. Song *et al.*, Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68–72 (2011). doi: [10.1038/nbt.1732](https://doi.org/10.1038/nbt.1732); pmid: [21151123](https://pubmed.ncbi.nlm.nih.gov/21151123/)
  32. K. E. Szulwach *et al.*, 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14**, 1607–1616 (2011). doi: [10.1038/nn.2959](https://doi.org/10.1038/nn.2959); pmid: [22037496](https://pubmed.ncbi.nlm.nih.gov/22037496/)
  33. M. Mellén, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–1430 (2012). doi: [10.1016/j.cell.2012.11.022](https://doi.org/10.1016/j.cell.2012.11.022); pmid: [23260135](https://pubmed.ncbi.nlm.nih.gov/23260135/)
  34. M. Yu *et al.*, Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012). doi: [10.1016/j.cell.2012.04.027](https://doi.org/10.1016/j.cell.2012.04.027); pmid: [22608086](https://pubmed.ncbi.nlm.nih.gov/22608086/)
  35. J. A. Stamatoyannopoulos *et al.*, An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* **13**, 418 (2012). doi: [10.1186/gb-2012-13-8-418](https://doi.org/10.1186/gb-2012-13-8-418); pmid: [22889292](https://pubmed.ncbi.nlm.nih.gov/22889292/)
  36. R. K. Auerbach *et al.*, Mapping accessible chromatin regions using Sono-Seq. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14926–14931 (2009). doi: [10.1073/pnas.0905443106](https://doi.org/10.1073/pnas.0905443106); pmid: [19706456](https://pubmed.ncbi.nlm.nih.gov/19706456/)
  37. Y. Shen *et al.*, A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012). doi: [10.1038/nature11243](https://doi.org/10.1038/nature11243); pmid: [22763441](https://pubmed.ncbi.nlm.nih.gov/22763441/)
  38. W. A. Pastor *et al.*, Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011). doi: [10.1038/nature10102](https://doi.org/10.1038/nature10102); pmid: [21552279](https://pubmed.ncbi.nlm.nih.gov/21552279/)
  39. Y. Huang, W. A. Pastor, J. A. Zepeda-Martínez, A. Rao, The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine. *Nat. Protoc.* **7**, 1897–1908 (2012). doi: [10.1038/nprot.2012.103](https://doi.org/10.1038/nprot.2012.103); pmid: [23018193](https://pubmed.ncbi.nlm.nih.gov/23018193/)
  40. A. Magklara *et al.*, An epigenetic signature for monoallelic olfactory receptor expression. *Cell* **145**, 555–570 (2011). doi: [10.1016/j.cell.2011.03.040](https://doi.org/10.1016/j.cell.2011.03.040); pmid: [21529909](https://pubmed.ncbi.nlm.nih.gov/21529909/)
  41. E. J. Clowney *et al.*, Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell* **151**, 724–737 (2012). doi: [10.1016/j.cell.2012.09.043](https://doi.org/10.1016/j.cell.2012.09.043); pmid: [23141535](https://pubmed.ncbi.nlm.nih.gov/23141535/)
  42. H. Li *et al.*, Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9397–9402 (2008). doi: [10.1073/pnas.0802876105](https://doi.org/10.1073/pnas.0802876105); pmid: [18599437](https://pubmed.ncbi.nlm.nih.gov/18599437/)
  43. M. W. Akhtar *et al.*, In vivo analysis of MEF2 transcription factors in synapse regulation and neuronal survival. *PLoS ONE* **7**, e34863 (2012). doi: [10.1371/journal.pone.0034863](https://doi.org/10.1371/journal.pone.0034863); pmid: [22496871](https://pubmed.ncbi.nlm.nih.gov/22496871/)
  44. A. C. Barbosa *et al.*, MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9391–9396 (2008). doi: [10.1073/pnas.0802679105](https://doi.org/10.1073/pnas.0802679105); pmid: [18599438](https://pubmed.ncbi.nlm.nih.gov/18599438/)
  45. J. D. Cahoy *et al.*, A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008). doi: [10.1523/JNEUROSCI.4178-07.2008](https://doi.org/10.1523/JNEUROSCI.4178-07.2008); pmid: [18171944](https://pubmed.ncbi.nlm.nih.gov/18171944/)
  46. H. Wu *et al.*, Dnm3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444–448 (2010). doi: [10.1126/science.1190485](https://doi.org/10.1126/science.1190485); pmid: [20651149](https://pubmed.ncbi.nlm.nih.gov/20651149/)
  47. T. Mohandas, R. S. Sparkes, L. J. Shapiro, Reactivation of an inactive human X chromosome: Evidence for X inactivation by DNA methylation. *Science* **211**, 393–396 (1981). doi: [10.1126/science.6164095](https://doi.org/10.1126/science.6164095); pmid: [6164095](https://pubmed.ncbi.nlm.nih.gov/6164095/)
  48. K. Plath, S. Mlynarczyk-Evans, D. A. Nusinow, B. Panning, Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002). doi: [10.1146/annurev.genet.36.042902.092433](https://doi.org/10.1146/annurev.genet.36.042902.092433); pmid: [12429693](https://pubmed.ncbi.nlm.nih.gov/12429693/)
  49. L. Carrel, H. F. Willard, X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400–404 (2005). doi: [10.1038/nature03479](https://doi.org/10.1038/nature03479); pmid: [15772666](https://pubmed.ncbi.nlm.nih.gov/15772666/)
  50. A. J. Sharp *et al.*, DNA methylation profiles of human active and inactive X chromosomes. *Genome Res.* **21**, 1592–1600 (2011). doi: [10.1101/gr.112680.11](https://doi.org/10.1101/gr.112680.11); pmid: [21862626](https://pubmed.ncbi.nlm.nih.gov/21862626/)
  51. C. M. Johnston *et al.*, Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. *PLoS Genet.* **4**, e9 (2008). doi: [10.1371/journal.pgen.0040009](https://doi.org/10.1371/journal.pgen.0040009); pmid: [18208332](https://pubmed.ncbi.nlm.nih.gov/18208332/)
  52. A. K. Maunakea *et al.*, Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257 (2010). doi: [10.1038/nature09165](https://doi.org/10.1038/nature09165); pmid: [20613842](https://pubmed.ncbi.nlm.nih.gov/20613842/)
  53. R. E. Thurman *et al.*, The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012). doi: [10.1038/nature11232](https://doi.org/10.1038/nature11232); pmid: [22955617](https://pubmed.ncbi.nlm.nih.gov/22955617/)
  54. M. B. Stadler *et al.*, DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011). pmid: [22170606](https://pubmed.ncbi.nlm.nih.gov/22170606/)
  55. Z. Petanjek *et al.*, Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13281–13286 (2011). doi: [10.1073/pnas.1105108108](https://doi.org/10.1073/pnas.1105108108); pmid: [21788513](https://pubmed.ncbi.nlm.nih.gov/21788513/)
  56. B. Kolb *et al.*, Experience and the developing prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **109** (suppl. 2), 17186–17193 (2012). doi: [10.1073/pnas.1121251109](https://doi.org/10.1073/pnas.1121251109); pmid: [23045653](https://pubmed.ncbi.nlm.nih.gov/23045653/)
  57. H. T. Chugani, A critical period of brain development: Studies of cerebral glucose utilization with PET. *Prev. Med.* **27**, 184–188 (1998). doi: [10.1006/pmed.1998.0274](https://doi.org/10.1006/pmed.1998.0274); pmid: [9578992](https://pubmed.ncbi.nlm.nih.gov/9578992/)
  58. P. Levitt, Structural and functional maturation of the developing primate brain. *J. Pediatr.* **143** (suppl.), S35–S45 (2003). doi: [10.1067/S0022-3476\(03\)00400-1](https://doi.org/10.1067/S0022-3476(03)00400-1); pmid: [14597912](https://pubmed.ncbi.nlm.nih.gov/14597912/)
  59. M. V. Johnston, Plasticity in the developing brain: implications for rehabilitation. *Dev. Disabil. Res. Rev.* **15**, 94–101 (2009). doi: [10.1002/ddrr.64](https://doi.org/10.1002/ddrr.64); pmid: [19489084](https://pubmed.ncbi.nlm.nih.gov/19489084/)
  60. P. J. Uhlhaas, W. Singer, The development of neural synchrony and large-scale cortical networks during adolescence: Relevance for the pathophysiology of schizophrenia and neurodevelopmental hypothesis. *Schizophr. Bull.* **37**, 514–523 (2011). doi: [10.1093/schbul/sbr034](https://doi.org/10.1093/schbul/sbr034); pmid: [21505118](https://pubmed.ncbi.nlm.nih.gov/21505118/)
  61. S. Nguyen, K. Meletis, D. Fu, S. Jhaveri, R. Jaenisch, Ablation of de novo DNA methyltransferase Dnm3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Dev. Dyn.* **236**, 1663–1676 (2007). doi: [10.1002/dvdy.21176](https://doi.org/10.1002/dvdy.21176); pmid: [17477386](https://pubmed.ncbi.nlm.nih.gov/17477386/)
  62. Materials and methods are available as supplementary materials on Science Online.

**Acknowledgments:** We thank W. F. Loomis for critical reading of this manuscript; J. Chambers for technical assistance with animal breeding; D. Chambers for technical assistance with cell sorting; M. Lutz and T. Berggren for provision of the HUES6 cells; F. Yue, G. Hon, and B. Ren for providing mapped mouse ChIP-Seq data and assistance with TAB-Seq; and A. Nimmerjahn for providing the S100b-eGFP mice. Human brain tissue was obtained from the National Institute of Child Health and Human Development Brain and Tissue Bank for Developmental Disorders at the University of Maryland, Baltimore, Maryland, the IDIBELL Biobank, which is part of the BrainNet Europe Bank funded by the European Commission (LSHM-CT-2004-503039). We thank the Stamatoyannopoulos laboratory (University of Washington) and the Mouse ENCODE Consortium for generating and providing access to the DNase hypersensitivity data sets (35). This work was supported by a National Institutes of Mental Health grant to M.M.B. and J.R.E. (MH094670); the Howard Hughes Medical Institute to T.J.S. and J.R.E.; the Gordon and Betty Moore Foundation (GMBF3034) to J.R.E.; NIH grant HG006827 to C.H.; NIH RO1 grants AI44432, HD065812, and CA151535; grant RM-01729 from the California Institute of Regenerative Medicine; and Translational Research grant

TRP 6187-12 from the Leukemia and Lymphoma Society (to A.R.). R.L. was supported by an Australian Research Council Future Fellowship. Support was also provided by the Government of Western Australia through funding for the Western Australia CoE for Computational Systems Biology. E.A.M. was supported by the Center for Theoretical Biological Physics, University of California San Diego, and NIH (National Institute of Neurological Diseases and Stroke grant K99NS080911). W.A.P. was supported by a predoctoral graduate research

fellowship from the NSF. Analyzed data sets can be accessed at [http://neomorph.salk.edu/brain\\_methylomes](http://neomorph.salk.edu/brain_methylomes). Sequence data can be downloaded from National Center for Biotechnology Information GEO (GSE47966). Tet2 mutant mice are available under a material transfer agreement from the La Jolla Institute for Allergy and Immunology. Correspondence and requests for materials should be addressed to R.L. (ryan.lister@uwa.edu.au), M.M.B. (mbehrens@salk.edu), and J.R.E. (ecker@salk.edu).

#### Supplementary Materials

[www.sciencemag.org/content/341/6146/1237905/suppl/DC1](http://www.sciencemag.org/content/341/6146/1237905/suppl/DC1)  
Materials and Methods  
Figs. S1 to S12  
Tables S1 to S5  
References (63–78)

15 March 2013; accepted 7 June 2013  
10.1126/science.1237905

# The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers

Frank Lyko<sup>1,9</sup>, Sylvain Foret<sup>2,9</sup>, Robert Kucharski<sup>3</sup>, Stephan Wolf<sup>4</sup>, Cassandra Falckenhayn<sup>1</sup>, Ryszard Maleszka<sup>3\*</sup>

**1** Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg, Germany, **2** ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Australia, **3** Research School of Biology, the Australian National University, Canberra, Australia, **4** Genomics and Proteomics Core Facility, German Cancer Research Center, Heidelberg, Germany

## Abstract

In honey bees (*Apis mellifera*) the behaviorally and reproductively distinct queen and worker female castes derive from the same genome as a result of differential intake of royal jelly and are implemented in concert with DNA methylation. To determine if these very different diet-controlled phenotypes correlate with unique brain methylomes, we conducted a study to determine the methyl cytosine (mC) distribution in the brains of queens and workers at single-base-pair resolution using shotgun bisulfite sequencing technology. The whole-genome sequencing was validated by deep 454 sequencing of selected amplicons representing eight methylated genes. We found that nearly all mCs are located in CpG dinucleotides in the exons of 5,854 genes showing greater sequence conservation than non-methylated genes. Over 550 genes show significant methylation differences between queens and workers, revealing the intricate dynamics of methylation patterns. The distinctiveness of the differentially methylated genes is underscored by their intermediate CpG densities relative to drastically CpG-depleted methylated genes and to CpG-rich non-methylated genes. We find a strong correlation between methylation patterns and splicing sites including those that have the potential to generate alternative exons. We validate our genome-wide analyses by a detailed examination of two transcript variants encoded by one of the differentially methylated genes. The link between methylation and splicing is further supported by the differential methylation of genes belonging to the histone gene family. We propose that modulation of alternative splicing is one mechanism by which DNA methylation could be linked to gene regulation in the honey bee. Our study describes a level of molecular diversity previously unknown in honey bees that might be important for generating phenotypic flexibility not only during development but also in the adult post-mitotic brain.

**Citation:** Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, et al. (2010) The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. *PLoS Biol* 8(11): e1000506. doi:10.1371/journal.pbio.1000506

**Academic Editor:** Laurent Keller, University of Lausanne, Switzerland

**Received:** May 25, 2010; **Accepted:** August 24, 2010; **Published:** November 2, 2010

**Copyright:** © 2010 Lyko et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Work in FL's lab was supported by a grant from the Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg. Work in RM's lab was supported by the Australian Research Council grant DP1092706. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** DMG, differentially methylated gene; DNMT3, DNA methyltransferase 3; mCpG, methylated CpG; o/e, observed/expected

\* E-mail: maleszka@rsbs.anu.edu.au

<sup>9</sup> These authors contributed equally to this work.

## Introduction

Many animal species have evolved the capacity to generate organisms with contrasting morphological, reproductive, and behavioral phenotypes from the same genome. However, the question of how such strikingly different organismal outputs occur with no standard genetic changes remains one of the key unresolved issues in biology.

The nutritionally controlled queen/worker developmental divide in the social honey bee *Apis mellifera* is one of the best known examples of developmental flexibility in any phylum. Despite their identical nature at the DNA level, the queen bee and her workers are strongly differentiated by their anatomical and physiological characteristics and the longevity of the queen [1]. Furthermore, the behaviors of queens and workers are remarkably divergent, varying from the navigational proficiency of foragers to the colony-bound omnipresent chemical influences of the queen

which control many aspects of the colony's existence. A diet of royal jelly during larval development clearly influences the epigenetic status of the queen's cells without altering any of the hardwired characteristics of her genome. As a result, two contrasting organismal outputs, fertile queens and non-reproductive workers, are generated from the same genome.

Recently, we have shown that diet is not the only modulator of developmental trajectories in honey bees. By silencing the activity of DNA methyltransferase 3 (DNMT3), a key component of epigenetic machinery controlling global gene reprogramming, we were able to generate adult bees with queen characteristics [2]. This relatively simple perturbation of the DNA methylation system not only mimicked the dietary effect of royal jelly on phenotype but also changed the cytosine methylation pattern of an illustrative gene. Furthermore, analysis of gene expression in both queens and workers suggested that their alternative developmental pathways are associated with subtle transcriptional changes in a particular

## Author Summary

The queen honey bee and her worker sisters do not seem to have much in common. Workers are active and intelligent, skillfully navigating the outside world in search of food for the colony. They never reproduce; that task is left entirely to the much larger and longer-lived queen, who is permanently ensconced within the colony and uses a powerful chemical influence to exert control. Remarkably, these two female castes are generated from identical genomes. The key to each female's developmental destiny is her diet as a larva: future queens are raised on royal jelly. This specialized diet is thought to affect a particular chemical modification, methylation, of the bee's DNA, causing the same genome to be deployed differently. To document differences in this epigenomic setting and hypothesize about its effects on behavior, we performed high-resolution bisulphite sequencing of whole genomes from the brains of queen and worker honey bees. In contrast to the heavily methylated human genome, we found that only a small and specific fraction of the honey bee genome is methylated. Most methylation occurred within conserved genes that provide critical cellular functions. Over 550 genes showed significant methylation differences between the queen and the worker, which may contribute to the profound divergence in behavior. How DNA methylation works on these genes remains unclear, but it may change their accessibility to the cellular machinery that controls their expression. We found a tantalizing clue to a mechanism in the clustering of methylation within parts of genes where splicing occurs, suggesting that methylation could control which of several versions of a gene is expressed. Our study provides the first documentation of extensive molecular differences that may allow honey bees to generate different phenotypes from the same genome.

group of genes encoding conserved physio-metabolic proteins [2,3]. These findings prompted us to examine the hypothesis that significant behavioral differences between queens and workers are partly underpinned by differences between their brain epigenomes that have arisen from basically identical genomes during development. The choice of brain tissue is critical because it is a non-dividing, largely diploid tissue and is thus free of any complications that arise from differential genomic replication that may characterize polytene and endopolyploid tissues (nearly all adult tissues of insects are non-diploid). In the context of methylomes, the use of whole bodies, or abdomens, creates an unacceptable mixture of methylomic signatures that simply cannot be deconvoluted in regards to function in any biologically meaningful manner.

We used bisulfite converted brain DNA of both castes together with Solexa (Illumina GA) sequencing technology [4] to generate a DNA methylation map at single-nucleotide resolution across the *Apis* genome. This powerful approach has recently been used to compare DNA methylation profiles across a group of selected species, including DNA from a worker honey bee whole body [5]. The results confirm the antiquity of DNA methylation in eukaryotes [6,7] and provide more experimental evidence that this epigenomic modification is utilized in a lineage-specific manner [8–10].

Here we confirm that in contrast to heavily methylated mammalian genomes [11], only a small and specific fraction of the honey bee genome is methylated [5,10,12,13]. Furthermore, the methylated cytosines occur in a group of genes showing a

higher level of conservation than non-methylated genes. Nearly 600 of those genes show significant methylation differences in the brains of queens and workers, suggesting that their transcription might be epigenetically modulated in a context-dependent manner. Additional deep sequencing of selected genes in all three castes—queens, workers, and drones (haploid males)—suggests that brain methylation patterns are unique to each behavioral system. We discuss our findings in the context of epigenetic influences on global regulatory networks and their ability to generate contrasting phenotypic and behavioral outcomes from the same genome.

## Results

### Characterization of Brain Methylomes in Queens and Workers

The sequencing of bisulfite converted *Apis* DNA yielded a dataset of 131 million reads after filtration and quality checks, 68.5% of which were mapped to unique genomic regions. The total sequence output was 18.8 giga bases (10.2 Gb for the queen and 8.6 Gb for the worker) yielding a combined 20× coverage of the 260 Mb genome. Our reads also contained multiple coverage of thousands of unmethylated repeated elements (ALUs and mariners) giving false-positive rates of only 0.1% for the queen DNA and 0.2% for the worker DNA. Figure S1A shows the distribution of the coverage depth for all cytosines on both strands, whereas distribution of the CpG nucleotides is shown in Figure S1B. More than 90% of the 10,030,209 CpGs in the *Apis* genome were covered by at least two sequencing reads, allowing for the methylation status of individual sites to be determined with confidence.

The characteristics of the brain methylomes of queens and workers are shown in Tables 1 and 2. Three firm conclusions can be drawn. First, of the over 60 million cytosines that exist in the *Apis* genome, only approximately 70,000 are methylated. Second, nearly all the methylated cytosines occur in CpG dinucleotides. Third, the overriding majority of these methylated sites are in exons. Finally, the number of methylated cytosines in *Apis* is nearly three orders of magnitude lower than in the human genome [11]. This relatively small number of mCs overcomes the large technical hurdles that exist in both mammalian and plant genomes where the number of methylated sites that need to be examined in terms of their importance to biological phenomena is in the hundreds of millions.

As shown in Table 1 the quantities of methylated CpGs (mCpGs) in queen and worker brain DNA are very similar, 69,064 and 68,222, respectively, with 54,312 mCpGs in common. Similarly, the methylation levels of mCpG are almost identical in both castes (Figure S2). Methylation in honey bees appears to be restricted to cytosines associated with CpG dinucleotides, with no significant non-CpG or asymmetric methylation detected in either genomic or mitochondrial DNA (Table 1). Therefore, we conclude that methylation at non-CpG sites is either extremely rare or non-existent in the honey bee genome. In accord with previous analyses [2,5,12,13], methylated sites in *Apis* appear to be exclusively located in exons with only infrequent mCpGs detected in intronic regions (Table 2). Most importantly, the methylated exons reside in genomic regions with low CpG observed/expected (*o/e*) ratios (Figure 1), whereas non-methylated exons fall into the category with high CpG *o/e* ratios. This bimodal profile is consistent with previous predictions based on bioinformatics analyses [10,12,13] and reflects the propensity of methylated Cs to be converted over time to thymines, resulting in a lower than expected density of the CpGs in methylated genes. However, the

**Table 1.** Cytosine DNA methylation in queens and workers in CG, CHG, and CHH genomic contexts (H = A, T, or C).

	Total	Methylated in Queens	Methylated in Workers	Methylated in Both Castes
CG	10,030,209	69,064	68,222	54,312
CHG	8,673,113	14	130	0
CHH	45,072,611	561	3,019 <sup>a</sup>	0

The thresholds used for methylation calls are detailed in the Methylation Assessment section.

<sup>a</sup>Nearly all of the 3,019 CHH that were inferred to be methylated in worker brains on the basis of Solexa reads were found to be not methylated by an additional sequencing of selected amplicons using the 454 technology.

doi:10.1371/journal.pbio.1000506.t001

total number of methylated genes in *Apis* revealed by genome-wide bisulfite sequencing is 5,854 instead of the 4,000 predicted to be methylated on the basis of local CpG bias. One reason for this difference might be that some genes do not display significant CpG depletion as a result of evolutionary pressure to maintain a particular protein coding sequence.

The genome-wide profiling of mCpGs confirms that methylated genes in *Apis* encode proteins showing a higher degree of conservation than proteins encoded by non-methylated genes [10]. Figures S3, S4, S5 and Table S1 show the results of our cross-species comparisons for methylated and non-methylated genes (Figure S3), for high-CpG and low-CpG genes (Figure S4), and high-CpG methylated and non-methylated genes (Figure S5). Most of the highly conserved genes are expected to be utilized by most tissues. In contrast, less conserved genes expressed in specialized tissues, such as those encoding odorant-binding proteins or odorant receptors, are not methylated (not shown). The repeated elements, ALUs, and mariners that harbor most of the DNA methylation content in humans and plants are *not* methylated in the bee genome, certainly not in the brain (Figures S6). Similarly, the multi-gene families encoding rRNAs and tRNAs, mitochondrial DNA, and CpG islands show no evidence of methylation in the brain (Figure S6). Lastly, while methylation of sub-telomeric regions has been shown to be important for the control of telomere length and recombination [14], the honey bee telomeres are also not methylated (not shown). The lack of methylation in ALUs and transposons has also been reported in a recent study performed on DNA extracted from a worker's whole body [5]. Given the proposed role of cytosine methylation in defense against genomic parasites in plants and vertebrates [7], the lack of methylation in ALU repeats and mariner transposons suggests that these mobile elements do not significantly impact on genome stability in honey bees. Indeed the bee genome contains an unusually small percentage of common types of transposons and retrotransposons found in other insects, possibly as a result of a strong selective pressure against mobile elements in male bees

(drones) that develop from unfertilized eggs and carry a haploid set of chromosomes [15].

As in the human and *Arabidopsis* genomes [4,11], methylation in *Apis* shows evidence of periodicity, although due to a much lower density of modified CpGs in this species the periodicity of 10 nucleotides (one helical DNA turn) is not obvious. However, a 3-base periodic pattern is clearly detectable, reflecting a preferential methylation of CpGs occupying the first and second position of the arginine codons (autocorrelation data in Figure S7).

#### Detailed Analysis of Methylation Patterns in Selected Amplicons by Deep Bisulfite Sequencing

To validate our Solexa-based methylation results, we designed primers for selected regions of eight nuclear and four mitochondrial genes and re-sequenced the PCR-generated amplicons using 454 technology. As illustrated in Figure 2, the 454 sequencing profiles are essentially identical with the Solexa-based results. All nuclear genes show differential methylation in the brains of queens and workers, including those cases where the methylation is almost absent, such as GB18602 in queen brains (Figure 2). No methylation was detected by this approach in the four selected mitochondrial amplicons (not shown).

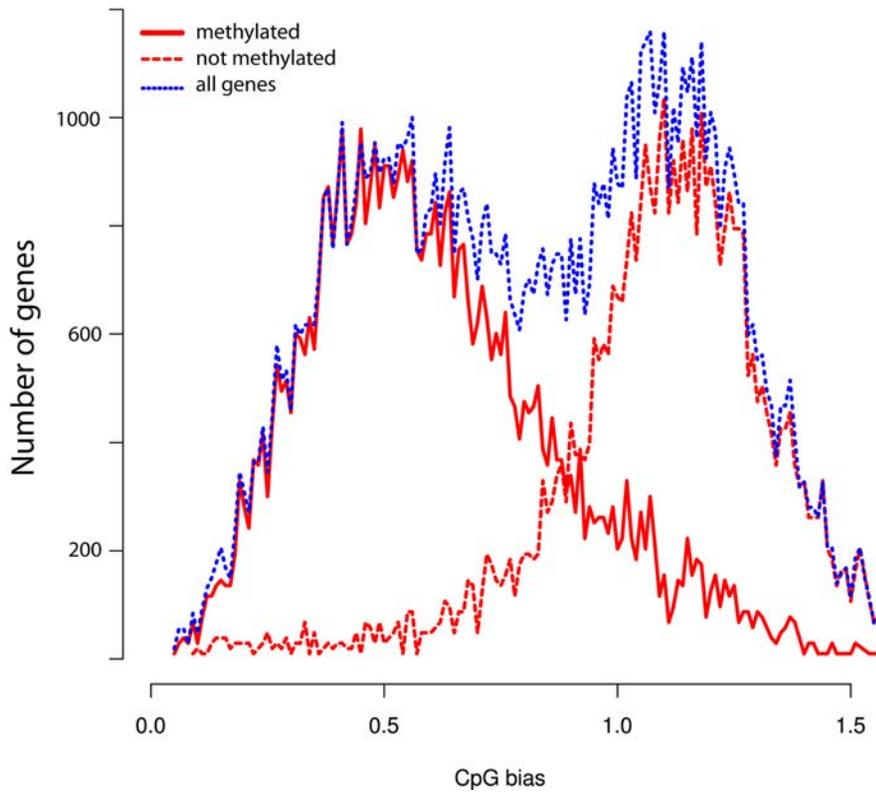
To further expand our analysis, we increased the 454 bisulfite sequencing coverage of the eight nuclear genes selected for validation and also included DNA from drone brains. We obtained several thousand high-quality reads for 24 amplicons (eight genes in three castes), with the total coverage ranging from 48 to 2,427×. The results shown in Figure 3 reveal both the dynamics and uniqueness of the methylation patterns in each cast. Out of the eight genes with differential worker/queen methylation, three show similar methylation patterns in workers and drones, but a distinct methylation pattern in queens (Figure 3A). Three additional genes show similar methylation patterns in queens and drones, but a distinct pattern in workers (Figure 3B). Two out of eight analyzed genes (GB11061 - seryl-tRNA synthetase and GB15356 - syd, chromosome segregation; Figure 3C) show distinct

**Table 2.** Cytosine DNA methylation in CG dinucleotides (mCG) in the exonic, intronic, and "intergenic" regions of queens and workers.

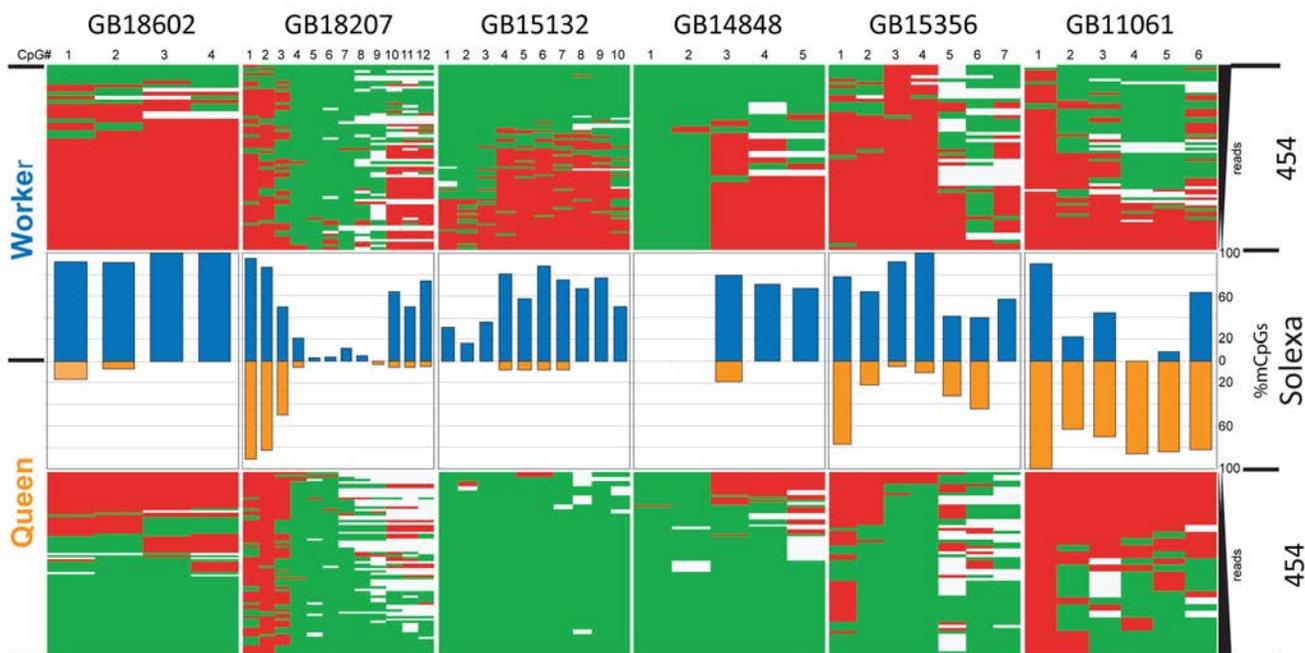
Genomic Location	Queens	% mCGs	% of All mCGs	Workers	% mCGs	% of all mCGs
Exons	54,378	8.6	78.74	51,658	8.16	75.72
Introns	5,992	0.2	8.68	6,720	0.22	9.85
Introns + exons	60,370	1.64	87.41	58,378	1.57	85.57
Intergenic regions <sup>a</sup>	8,694	0.16	12.59	9,844	0.17	14.43

<sup>a</sup>The annotation of the *Apis* transcriptome is largely limited to the coding regions, and it is likely that some of the intergenic regions may correspond to untranslated segments of mRNAs.

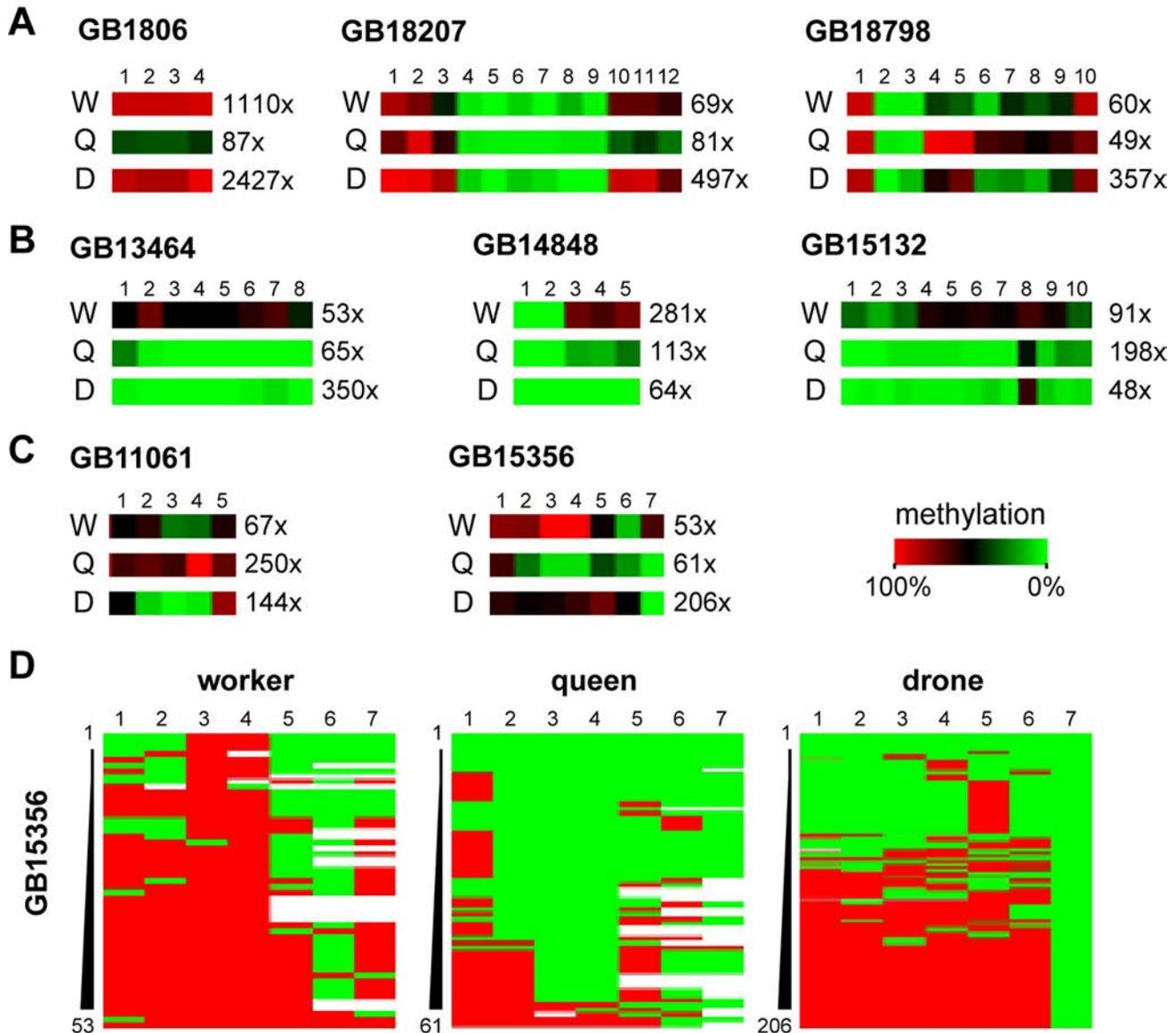
doi:10.1371/journal.pbio.1000506.t002



**Figure 1. CpG (o/e) bias of protein-coding regions in the honey bee genome.** Since the profiles for both queens and workers are virtually identical, only the queen profile is shown.  
doi:10.1371/journal.pbio.1000506.g001



**Figure 2. Comparison of CpG methylation profiles in differentially methylated genes generated by two technologies, Solexa genome-wide shotgun sequencing and 454 sequencing of PCR produced amplicons.** The "heat maps" represent the 454 sequencing of PCR amplified segments, whereas the bars illustrate the Solexa reads. The eight nuclear genes for this experiment were chosen from the list of DMGs shown in Tables 3 and S2, taking into account the availability of convenient CpG-containing regions for primer design. Six genes are shown in this figure and the others in Figure 3. Gene annotations: GB18602 - membrane protein; GB18207 - cadherin; GB15132 - TAP42 (TOR signaling); GB14848 - clathrin assembly protein; GB15356 - syd, chromosome segregation; GB11061 - seryl-tRNA synthetase.  
doi:10.1371/journal.pbio.1000506.g002



**Figure 3. Detailed analysis of deep sequencing of selected genes.** The bisulfite converted amplicons of selected genes were sequenced using 454 technology. The selection was based on differential methylation in brains of queens and workers, but DNA from male brains (drones) was also used in this experiment. The panels illustrate the uniqueness of brain methylation patterns in bees. 3A: Genes showing similar methylation patterns in workers and drones, but a distinct methylation pattern in queens. 3B: Genes with similar methylation patterns in queens and drones, but a distinct pattern in workers. 3C: Gene with distinct methylation patterns in all three castes. Panel 3D shows the full methylation heatmaps of GB15356. This result is discussed in the chapter “Detailed Analysis of Methylation Patterns in Selected Amplicons by Deep Bisulfite Sequencing.” Gene annotations: GB18798 - ubiquitin conjugation factor; GB13464 - RhoGAP93B. For other genes, see Figure 2. doi:10.1371/journal.pbio.1000506.g003

methylation patterns in all three castes. The latter finding was also confirmed by the analysis of full methylation heatmaps of GB15356 (Figure 3D). GB15356 is strongly methylated in workers, with many reads showing complete methylation in the 5'-half of the amplicon (Figure 3D). In queens, GB15356 methylation is strongly reduced and many reads show no methylation at all. Intriguingly, drones show a bimodal methylation pattern with approximately half of the reads methylated and the other half unmethylated (Figure 3D). These results further illustrate caste-specific differences in methylation patterns and suggest a complex role of DNA methylation in the regulation of caste-specific epigenomic differences in the brain.

### Identification of Differentially Methylated Genes

To determine if there is a link between DNA methylation patterns and the striking morphological and behavioral polymorphisms of queen bees and workers, we examined the levels of CpG methylation in all annotated transcription units in both brains using high stringency criteria (Supporting Information). This approach generated a list of 561 differentially methylated genes (DMGs, Tables 3 and S2) showing significant methylation differences between the two castes. With the exception of highly expressed genes encoding ribosomal proteins, DMGs in *Apis* are expressed at low or moderate levels across all analyzed tissues (Tables 3 and S2). In several cases their transcriptional activities

**Table 3.** Differentially methylated genes in brains of queens and workers.

Gene ID <sup>a</sup>	No. of CpGs	Antenna	Brain	Relative		Expression <sup>b</sup>			Gene Annotation
				HPG	Larva	Ovary	Thorax		
GB18602	30	1	1	1	1	1	1	Transmembrane protein YhhN	
GB18303	13	1	1	1	1	1	1	Activator protein of Rab-like small GTPases	
GB13368	9	2	1	2	10	1	3	3-hydroxyacyl-CoA dehydrogenase, NAD-binding	
GB13215	34	1	1	1	1	1	1	Glycine cleavage system P-protein,	
GB15588	9	1	1	1	1	1	1	Low-density lipoprotein receptor domain class A	
GB15132	24	1	1	1	1	1	1	TAP42 (regulates the TOR signaling pathway)	
GB12560	12	1	1	1	1	9	1	RNA-binding protein	
GB11648	13	1	1	1	1	2	1	Catalase	
GB19645	12	1	1	1	1	1	1	Phosphodiesterase 6	
GB12929	39	1	1	1	1	1	1	Paralytic - Na channel	
GB11421	31	1	1	1	1	1	1	Tight junction associated protein	
GB19503	33	1	1	1	1	1	1	Heat shock protein 8	
GB13740	24	1	1	1	1	1	1	Dysfusion, TF with PAS domain	
GB10394	8	1	1	1	1	1	1	TNF-receptor-associated factor 1	
GB16628	9	10	6	8	10	10	10	Ribosomal protein L6	

Only the top 15 genes are shown; see Table S2 for list of 561 genes that fall into this category. Based on microarray data from Foret et al. [10]. The genome assembly v.02 was used throughout this study.

<sup>a</sup>GB numbers refer to the proteins at BeeBase: [genomes.arc.georgetown.edu/drupal](http://genomes.arc.georgetown.edu/drupal).

<sup>b</sup>Genes were ranked into 10 bins based on their expression levels from low (1) to high (10).

doi:10.1371/journal.pbio.1000506.t003

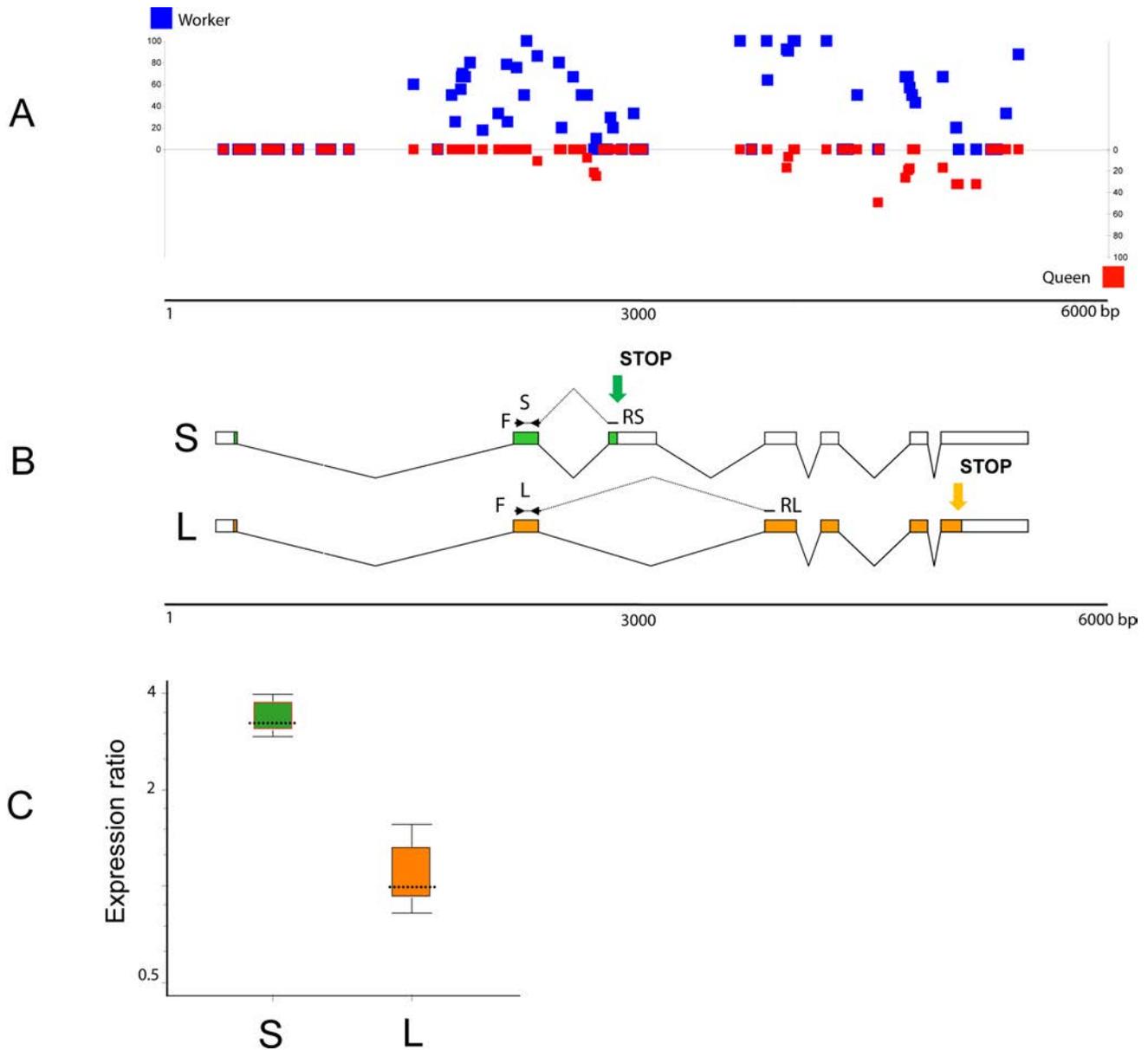
were found to be significantly up-regulated in some tissues relative to others. For example, the expression of 3-hydroxyl-CoA dehydrogenase (GB13368) is much higher in the larva than in the adult brain, and RNA-binding protein (GB12560) is significantly up-regulated in the ovaries relative to other tissues (Table 3). Almost all DMGs encode highly conserved, well-characterized proteins that have been implicated in core processes such as metabolism, RNA synthesis, nucleic acids binding, and signal transduction (Table S2). While a number of genes could not be clearly assigned to functional categories, their high level of conservation across phyla indicates that they are nevertheless likely to be involved in essential cellular processes (e.g. GB18943, GB13480, and GB18037). Several differently methylated genes encode proteins previously shown to be involved in either brain development or activity-dependent neural functions in both vertebrates and invertebrates. These include the Ephrin receptor GB1258516 [16], a nicotinic acetylcholine receptor GB19703, “no extended memory” GB16408 that is encoded by cytochrome B561 in *Drosophila*, two NMDA receptors GB19334 and GB15722, and a membrane channel GB12287 that mediates cell adhesion. When defective, GB12287 results in the “big brain” phenotype (Table S2). We note that Dynactin, used in our previous study [2] to illustrate the methylation differences between the two castes during larval growth in both royal jelly-fed and RNAi-treated individuals, does not show differential methylation in the brain. However, two genes, GB11197 and GB13866, encoding proteins associated with the large Dynein complex to which Dynactin also belongs are differentially methylated in the brain. Thus, the multi-protein Dynein complex appears to be epigenetically modulated during larval growth and in adult brains.

### CpG Bias and Epigenetic Modulation

Recently, Elango et al. [13] on the basis of bioinformatic analyses of a dataset of differentially expressed genes in brains of queens and workers proposed that “high-CpG genes in *A. mellifera* generally are more prone to epigenetic modulation than low-CpG genes.” We have tested this hypothesis using our new caste-specific brain methylome data. The results summarized in Table S3 suggest that (a) the methylation of a gene is a decreasing function of its CpG richness (Figure S8), (b) the “caste-specific genes” [13] that are methylated have a lower CpG content than the non-methylated genes (Table S3), and (c) DMGs are over-represented in the low CpG genes (Table S3). Therefore, our results do not support the hypothesis of Elango et al. [13]. However, it is noteworthy that although the DMGs are generally CpG-depleted, they tend to be less CpG-depleted than those genes that are not differentially methylated (Table S3). This intermediate CpG density observed in DMGs underscores the uniqueness of this class of genes and suggests that they might be methylated in a distinct manner from the rest of methylated genes. This class of genes showing differential patterns of methylation associated with phenotypic polymorphism is thus of special importance in the study of complex context-dependent phenotypes.

### Unraveling the Link between CpG Methylation and Splicing

To explore the relationship between differential methylation and expression patterns in queens and workers, we examined in more detail the first gene on the DMG list (GB18602) encoding a putative transmembrane protein with the YhhN domain conserved from bacteria to mammals. Figure 4 shows the distribution



**Figure 4. Expression profile of an alternatively spliced and differentially methylated gene GB18602 in queen and worker brains.** (A) The CpG methylation pattern indicating the level of methylation for individual CpGs (blue squares, workers; red squares, queens). (B) Gene model of GB18602 showing the two spliced variants S (short protein) and L (long protein) and the positions of PCR primers used for variant-specific amplifications. The green and orange arrows indicate the positions of two alternative Stop codons. (C) Relative expression of the two spliced variants in brains of queens and workers examined by real-time PCR. The level of transcript S (green) encoding the truncated protein is significantly up-regulated in the queen brain, whereas the L variant (orange) is expressed at the same level in both castes. The queen expression represents a combined set of data from three independent queen samples: 4 mo old (1 brain), 12 mo old (2 brains), and swarm queens of unknown age (2 brains). Workers were 8 d old (6 brains in 3 replicates). The reference gene was calmodulin [2]. Whisker-box plot of expression ratio values: dotted line, median value; box, inter-quartile range of values; whiskers, outer 50% of observations. For more details, see Table S4. doi:10.1371/journal.pbio.1000506.g004

of mCpGs against the GB18602 gene model (Figure 4A and 4B) and the relative expression of two spliced variants in both castes (Figure 4C). The L variant (L) encoding a long protein shows identical expression levels in both queens and workers, whereas the S variant (S) encoding a short protein is significantly up-regulated in queen relative to worker brain (Figure 4C). The majority of the differentially methylated sites in the GB18602 locus map to the region spanning the additional cassette-exon that contains a Stop codon for the short protein encoded by the S transcript, suggesting a correlation between methylation and the

outcome of alternative splicing of this gene in *Apis*. The increased level of methylation spanning the conditional splicing event (insertion or skipping of the cassette-exon) in the worker brain may impede the inclusion frequency of this exon into the mature transcript. Since the L variant is expressed at the same levels in both castes, the increased methylation in workers appears to be specifically affecting splicing, but not transcription. The observed differential pattern of expression of both transcripts in the brains of queens and workers (Figure 4C) supports this idea. Although the function of this gene is not known, the expression profiles of the

*Drosophila melanogaster* ortholog CG7582 suggest that it encodes a protein involved in fat and sugar metabolism [17]. In the fly, which has no CpG methylation, this gene is not alternatively spliced and shows the highest levels of expression in the nervous system (FlyAtlas.org). In contrast, the human ortholog of GB18602, designated TMEM86A, produces alternatively spliced variants, including one encoding a truncated protein similar to the honey bee variant S. In addition to GB18602 we found numerous other examples of methylated genes in *Apis* in which most or even all clusters of mCpGs show a non-random, highly significant tendency to be near differentially spliced exons (Figure S9). Another salient finding relevant to methylation of intron-containing genes is the differential methylation of the multi-gene histone family in *Apis*. As illustrated in Table 4 and Figure S10, all intron-containing histone genes are methylated, whereas intronless histone genes show no evidence of methylation. It is noteworthy that the methylated histone genes in *Apis* belong to a distinct class of histone variants. Unlike the canonical histones these variants are expressed constitutively and independently of replication and act as multifunctional regulators in a range of processes including DNA repair, transcription initiation and termination, meiotic recombination, etc. [18]. It is believed that they represent lineage specific innovation that is important for each organism's evolutionary specialization [18].

## Discussion

The discovery of a functional DNA methylation system in honey bees and other invertebrates [1,7–10,19] has brought a fresh perspective to the study of epigenetic regulation of development and behavior. It reinforced the view that this covalent modification of DNA is an ancient and widely utilized evolutionary mechanism that was present in the basal Metazoa and has been recruited to serve diverse functions in modern organisms, including regulation of gene expression, cell differentiation, and silencing of transposons [20–22]. However, the trajectories from methylation changes to complex phenotypes are indirect, multi-level, and virtually unknown. For example, the hundreds of millions of methylated cytosines in the human genome and their large variation in different cell types in vivo pose a major challenge to uncovering those changes causative to phenotype. By contrast, the honey bee *Apis mellifera* shares its basic methylation enzymology with humans, yet as shown in this and other studies [5,10,12,13] only a small and specific fraction of its genome is methylated. The present results show that honey bees utilize methyl tags to mark a core of mostly conserved and ubiquitously expressed critical genes whose activities cannot be switched off in most tissues. Recent data suggest that in spite of their permanent expression these genes might not be required at the same level

**Table 4.** Annotation of the histone gene family in *Apis mellifera*.

Class	Proposed Subclass	Type	Apis Histone Genes	NCBI RefSeq mRNA	Proposed Gene ID	Proposed Protein ID	Splicing Status	Methyl CpG	
<b>H1</b>	H1	canonical	GB12700 <sup>a</sup>	XM_001121111	H1.1	H1.A			
			GB12218	XM_001122184	H1.2	H1.B			
<b>H2A</b>	H2A	canonical	GB18806	XM_001120186	H2A.1	H2A.A			
			GB12818	XM_001120346	H2A.2	H2A.B			
			GB13800	XM_001119899	H2A.3	H2A.C			
			N/A <sup>c</sup>	XM_001120934	H2A.4	H2A.D			
		H2A.X	<b>variant</b>	GB18954	XM_624697	H2A.X	H2A.X	<b>Spliced</b>	<b>Yes</b>
		H2A.Z/H2AV	<b>variant</b>	GB12991	XM_624164	H2A.Z	H2A.Z	<b>Spliced</b>	<b>Yes</b>
	Pseudogene	Pseudogene		ψH2A					
<b>H2B</b>	H2B	canonical	GB12700 <sup>a</sup>	XM_001120238	H2B.1	H2B.A			
			GB13012	XM_001120889	H2B.2	H2B.B			
			N/A <sup>c</sup>	XM_001122218	H2B.3	H2B.C			
			GB12922 <sup>b</sup>	XM_001119846	H2B.4	H2B.D			
			GB11889	XM_001120014	H2B.5	H2B.D			
<b>H3</b>	H3	canonical	GB11223	XM_001120304	H3.1	H3			
			GB14620	XM_001121026	H3.2	H3			
			N/A <sup>c</sup>	XM_001120132	H3.3	H3			
		H3.3	<b>variant</b>	GB12948	XM_001120696	H3.3.1	H3.3	<b>Spliced</b>	<b>Yes</b>
				GB11228	XM_624496	H3.3.2	H3.3	<b>Spliced</b>	<b>Yes</b>
	<b>variant</b>	CENPA	GB18566	N/A	CENPA	CENPA	<b>Spliced</b>	<b>Yes</b>	
<b>H4</b>	H4	canonical	GB20104	XM_001120066	H4.1	H4			
			GB12644 <sup>a</sup>	XM_01119948	H4.2	H4			
			GB14107	XM_001120988	H4.3	H4			
			GB17789	NM_001011609	H4.4	H4			

See Figure S10 for additional details. These Bee Base protein entries are either incorrect or missing:

<sup>a</sup>chimeras,

<sup>b</sup>truncated,

<sup>c</sup>not available.

doi:10.1371/journal.pbio.1000506.t004

throughout development, or under changing environmental conditions [23–25].

In honey bees, feeding of newly hatched larvae destined to become queens with royal jelly leads to metabolic acceleration and increased growth driven by global but relatively subtle changes in the expression levels of a large number of ubiquitous genes [2,3,10]. These initial stages of larval development are later followed by the activation of more specific pathways to lay down caste-specific structures [3,10]. Interestingly, adult queen bees continue to be fed royal jelly, suggesting that this highly specialized diet is important for maintaining their reproductive as well as behavioral status. One possibility is that adult queens adjust their brain methylomes according to external instructions from their diet. One of the ingredients of royal jelly, phenyl butyrate [26], is a known histone deacetylase inhibitor and growth regulator that has been implicated in improving cognitive deficits in mice [27] and in life extension of *Drosophila* [28]. Although the significance of phenyl butyrate in royal jelly is not yet understood, it is conceivable that this complex diet evolved to provide two important functions for honey bees. It primarily serves as the source of nutrients for queen development but also as the regulator of epigenetic networks controlling gene expression in the brain. In addition to having different morphologies, reproductive capacities, and distinct behaviors, the genetically identical queen and worker honey bees also have different synaptic densities in their brains. In a recent study, Groh and Rössler [29] proposed that such developmental, diet-induced heterochrony results in fewer synapses in olfactory centers in queens, which may result in poorer performance on olfactory learning tasks compared to workers.

Recent studies using rodent models provided strong support for an idea that the nervous system has co-opted epigenetic mechanisms utilized during development for activity-dependent brain functions, including the generation and maintenance of long-term behavioral memories in adulthood [30,31]. Not surprisingly, DNA methylation has also been found to be involved in memory processing in honey bees [32], highlighting the significance of this epigenomic setting in conserved brain functions. These findings also provided evidence that DNA methylation, once believed to be an inert process after cellular differentiation, is dynamically regulated in the adult brain. Although both DNA methylation and chromatin remodeling have been implicated in these processes, the specific biological mechanisms underlying such adaptations remain largely unknown.

Our study provides experimental evidence that at least 560 differentially methylated ubiquitously expressed genes are involved in generating molecular brain diversity in female honey bees. Although it is still unclear how methylation might be linked to the gene regulatory networks, it has been proposed that DNA methylation together with changes in the histone profiles has the capacity to adjust DNA accessibility to cellular machinery by changing chromatin density [33–35]. Our findings support this notion and suggest that this mechanism provides an additional level of transcriptional control to fine tune the levels of messenger RNAs, including differentially spliced variants, encoded by the conserved genes. The association of mCpG clusters with alternatively spliced exons and genes containing introns in *Apis* is reminiscent of the distribution of mCpGs around the exon/intron junctions in human genes [36]. Epigenetic control of both splicing and mRNA levels might be utilized in different lineages, suggesting that a direct relationship between gene methylation and transcription is a widely spread phenomenon in both the animal and plant kingdoms [8,37].

Cytosine methylation may interact with other epigenetic features, such as distinctive histone modification signatures that

have been shown to correlate with the splicing outcome in a set of human genes [33–35]. The correlation between methylation and splicing is further highlighted by the differential methylation of two classes of histone genes in *Apis*. We find that only intron-containing histone variants are methylated, whereas intronless canonical histone genes are not methylated. Interestingly, histone variants have been implicated in multiple conserved roles in eukaryotes [18] and therefore are part of the cellular maintenance systems together with other ubiquitously expressed genes. In a broader context, methylated cytosines may specify information to set up, proliferate, and regulate splicing patterns during cellular processes such as development and differentiation.

Thus, rather than switching the genes on and off by promoter methylation, the intragenic methylation in *Apis* operates as a modulator of gene activities. As a result the entire topology of a complex brain network can be reprogrammed by subtle adjustments of many genes that act additively to produce a given phenotype [38]. Such adjustable DNA methylation levels generating variability in the transcriptional output of methylated genes could underlie genetically inherited propensity to phenotypic variability in accord with the recently proposed model of stochastic epigenetic variations as a heritable force of evolutionary change [39].

The technical advantages of the low number of methylated cytosines in the genome, together with diet-controlled phenotypes arising from the same genome, make the honey bee an extremely tractable, simplified *in vivo* system in which to examine fundamental principles underpinning transitions from methylomes to organismal plasticity. In particular, the absence of promoter methylation in honey bees brings into focus gene body methylation as an important mechanism controlling various aspects of transcription. The utility of honey bees for understanding the intricacies of this process in the behavioral context can now be experimentally tested.

## Materials and Methods

### Source of DNA

Total DNA was extracted from dissected gland-free brains of 50 age-matched egg-laying queens (2.5 wk old) and from fifty 8-d-old workers. These individuals represent early stages of the reproductive life of queen bees and mature young workers capable of performing foraging tasks [19].

### Sequencing of Bisulfite Converted DNA Libraries Using the Solexa GAIIx Platform (Illumina)

5 µg of high molecular weight DNA were used for fragmentation using the Covaris S2 AFA System in a total volume of 100 µl. Fragmentation-run parameters: Duty cycle 10%; Intensity: 5; Cycles/burst: 200; Time: 3 min; number of cycles: 3, resulting in a total fragmentation-time of 180 s. Fragmentation was confirmed with a 2100 Bioanalyzer (Agilent Technologies) using a DNA1000 chip. Fragment sizes were 140 bp on average for queen and worker DNAs, respectively. The fragmented DNAs were concentrated to a final volume of 75 µl using a DNA Speed Vac. End repair of fragmented DNA was carried out in a total volume of 100 µl using the Paired End DNA Sample Prep Kit (Illumina) as recommended by the manufacturer. For the ligation of the adaptors, the Illumina Early Access Methylation Adaptor Oligo Kit and the Paired End DNA Sample Prep Kit (Illumina) were used, as recommended by the manufacturer. For the size selection of the adaptor-ligated fragments, we used the E-Gel Electrophoresis System (Invitrogen) and a Size Select 2% precast agarose gel (Invitrogen). Each fragmented DNA was loaded on two lanes of

the E-gel. Electrophoresis was carried out using the “Size Select” program for 16 min. According to the standard loaded (50 bp DNA Ladder, Invitrogen), 240 bp fragments were extracted from the gel, pooled, and directly transferred to bisulfite treatment without further purification. For the bisulfite treatment we used the EZ-DNA Methylation Kit (Zymo) as recommended by the manufacturer with the exception of a modified thermal profile for the bisulfite conversion reaction. The conversion was carried out in a thermal cycler using the following thermal profile: 95°C for 15 s, 50°C for 1 h, repeat from step 1, 15×, 4°C for at least 10 min. The libraries were subsequently amplified, using the Fast Start High Fidelity PCR System (Roche) with buffer 2, and Illuminas PE1.1 and PE2.1 amplification primers. PCR thermal profile: 95°C for 2 min, 95°C for 30 s, 65°C for 20 s, 72°C for 30 s, then repeat from step 2, 11×, 72°C for 7min, hold at 4°C. PCR reactions were purified on PCR purification columns (MinElute, Qiagen) and eluted in 20 µl elution buffer (Qiagen).

### Validation of the Libraries

1 µl of the libraries were analyzed on a 2100 Bioanalyzer (Agilent Technologies) using a DNA1000 chip. The fragment sizes were 240 bp and 243 bp for the queen and worker libraries, respectively. The estimated concentrations of the libraries were 0.8 ng/µl for the queen library and 5.8 ng/µl for the worker library.

### Sequencing and Analysis

We used 8 pM of single stranded DNA per lane for Solexa sequencing. In total we sequenced 6 lanes. Worker: 1. single end - 36 bp - 10,187,567 reads (×2); 2. paired end - 76 bp - 7,960,842 reads (×2); 3. paired end - 76 bp - 7,444,938 reads (×2); 4. paired end - 76 bp - 11,642,135 reads (×2). Queen: 1. paired end - 76 bp - 16,752,247 reads (×2); 2. paired end - 76 bp - 16,778,784 reads (×2). For sequencing we used a Solexa Genoma Analyzer GAIIx with a v2 Paired End Cluster Generation Kit - GA II (Illumina) and v3 36 bp Cycle Sequencing Kits (Illumina). Extraction of sequences was done using Illumina Pipeline v1.4 software. Image analysis and basecalling was done using Illumina SCS v2.5 software.

### Mapping

Reads were mapped using BSMAP-1.0240 with minor modifications [40]. A number of trimming and mapping options were assessed, and the conditions yielding the highest genome coverage depth was used for further processing (-s 12 -v 5 -k 6, for word size, number of mismatches, and number of words). Only the reads mapping uniquely were used. Mapping was carried out on a Linux cluster running Debian 5.0 (lenny).

### Methylation Assessment

To increase the accuracy of methylation calls, only those cytosines fulfilling neighborhood quality standards NQS41 were counted [41]; namely, we only took into account bases of quality 20 or more, flanked by at least three perfectly matching bases of quality 15 or more. Deamination efficiency was assessed using the observation that the genomic repeats are not methylated in the honeybee (Figure S3). The deep coverage of these repeated sequences allowed us to estimate that the deamination rate is 99.76% for the queens and 99.71% for workers. The methylation status of each cytosine was then assessed by comparing the number of methylated and non-methylated reads to a binomial distribution with a probability of success equal to the deamination rate and a number of trials equal to the number of reads mapping to that

cytosine and adjusting the resulting  $p$  values for multiple testing with the method of Benjamini and Hochberg [42]. An adjusted  $p$  value of 0.05 was used as a threshold for methylation calls. All statistical computations were carried out using the R language ([www.r-project.org](http://www.r-project.org)).

Honeybee ESTs and predicted genes were loaded into a Mysql database and visualized with Gbrowse ([www.gmod.org](http://www.gmod.org)), where CpG methylation levels in queens and workers were added as separate tracks.

### Differential Methylation

Base-wise differences between queen and workers were estimated using Fisher exact tests. Gene-wise differences were assessed by generalized linear models of the binomial family, where methylation levels were modeled as functions of two categorical variables: caste and CpG position.  $p$  values were adjusted for multiple testing with the method of Benjamini and Hochberg [42].

### Amplicon Sequences Selection

Illumina sequencing and BSMAP mapping results were confirmed by 454 sequencing of a set of bisulfite amplicons. Amplicon sequences were selected using raw methylome data and the following criteria: minimum coverage - 5 mapped reads for each queen and worker sample; minimum 2 mCpGs within a maximum of ~600 bp of sequence showing at least 50% difference in methylation levels between the two samples. In addition, four regions of mtDNA were selected. All primers and other details are listed in Table S4.

### Other Protocols

All molecular protocols are described elsewhere [2,9,10,43].

### Supporting Information

**Figure S1 Coverage of all cytosines.** (A) Cumulative distribution of the coverage of all cytosines, on either strand of the genome, in workers and queens. On the  $x$ -axis, coverage refers to the coverage depth that is the number of reads uniquely mapped to a given cytosine. The  $y$ -axis is the cumulative distribution; for instance, approximately 50% of all cytosines are covered by less than 5 reads, and about 80% are covered by less than 10 reads. (B) Cumulative distribution of the coverage of all CpGs in the genome, in workers and queens. On the  $x$ -axis, coverage refers to the coverage depth that is the number of reads uniquely mapped to a given CpG dinucleotide. The  $y$ -axis is the cumulative distribution (for instance, approximately 50% of the CpGs are covered by less than 15 reads, and about 80% are covered by less than 25 reads).

Found at: [doi:10.1371/journal.pbio.1000506.s001](https://doi.org/10.1371/journal.pbio.1000506.s001) (0.41 MB PDF)

**Figure S2 Methylation levels of methylated CpGs.** Distribution of the methylation level of methylated CpGs. The methylation level is the proportion of methylated reads mapping to a given CpG. Over 30% of the CpGs are fully methylated.

Found at: [doi:10.1371/journal.pbio.1000506.s002](https://doi.org/10.1371/journal.pbio.1000506.s002) (0.16 MB PDF)

**Figure S3 Number of methylated and non-methylated Apis genes with BLAST hits to different species at various E-value thresholds.** The amino acid sequences of the genes were compared. Fisher exact tests were conducted to assess whether significantly more methylated genes have a BLAST hit than non-methylated genes. Statistically significant tests at the 5% level are denoted with a star, and non-significant tests are

shown with a dot. The details of this analysis can be found in Table S3.

Found at: doi:10.1371/journal.pbio.1000506.s003 (0.13 MB PDF)

**Figure S4 Number of high and low CpG honey bee genes with BLAST hits to different model species at various E-value thresholds.** The amino acid sequences of the genes were compared. Fisher exact tests were conducted to assess whether significantly more low CpG genes have a BLAST hit than high CpG genes. Statistically significant tests at the 5% level are denoted with a star, and nonsignificant tests are shown with a dot. The details of this analysis can be found in Table S3.

Found at: doi:10.1371/journal.pbio.1000506.s004 (0.13 MB PDF)

**Figure S5 Number of high CpG methylated and non-methylated honey bee genes with BLAST hits to different model species at various E-value thresholds.** The amino acid sequences of the genes were compared. Fisher exact tests were conducted to assess whether significantly more high CpG methylated genes have a BLAST hit than high CpG non-methylated genes. Statistically significant tests at the 5% level are denoted with a star, and non-significant tests are shown with a dot. The details of the analysis can be found in Table S3.

Found at: doi:10.1371/journal.pbio.1000506.s005 (0.13 MB PDF)

**Figure S6 Coverage (red) and methylation ratio (green) along various kinds of repetitive elements.** The methylation ratio is the proportion of the reads where a cytosine is either methylated or unmethylated. The  $y$ -axes are logarithmic in base 10 (the  $x$ -axis is truncated to the nearest multiple of 50, just like the  $y$ -axis is truncated to the nearest integer).

Found at: doi:10.1371/journal.pbio.1000506.s006 (0.63 MB PDF)

**Figure S7 Periodicity of methylation patterns.** (A) Autocorrelation of CpG methylation status over 1 kb. (B) Autocorrelation over 100 bp. Figures A and B show that the correlation of methylation status of neighboring CpGs increases sharply between 1 bp and 20 bp, then drops rapidly between 40 bp and 100 bp, and then slowly fades away. CpGs within a neighborhood of 2 bp to 100 bp are thus more likely to share the same methylation status than more distant CpGs. (C) Fourier transform of autocorrelation showing a clear periodicity peak at 33 cycles per 100 bp (every 3 bp). (D) Distribution of codon position of mCs, and distribution of methylation level depending on the position. These two panels indicate that the distance between methylated CpGs is often a multiple of three and that the methylated cytosine corresponds most frequently to the first nucleotide of an arginine codon.

Found at: doi:10.1371/journal.pbio.1000506.s007 (0.33 MB PDF)

**Figure S8 Correlation between CpG o/e and proportion of methylated CpGs.** Genes with a lower CpG content tend to have a higher proportion of methylated CpGs. The red line is a polynomial regression through the points. The Akaike Information Criterion for model selection and a (monotonously decreasing) polynomial of degree three was identified as the best model.

Found at: doi:10.1371/journal.pbio.1000506.s008 (0.71 MB PDF)

**Figure S9 Distribution of methylated CpGs relative to splicing sites.** For 169 genes, each containing a single well-defined alternative splicing event, the distance of all mCpGs to the centre of the alternatively spliced intron was computed, and the median of all these distances was calculated. A null distribution of this median distance was constructed using a randomization procedure (Manly, 2007): the methylation status of mCpGs of these genes were randomly shuffled 1,000 times, and the corresponding median distances computed. The observed value

(1,224) is smaller than the smallest of the null distribution (1,259); the probability of the methylated CpGs to be as close or closer to the alternatively spliced intron as in this dataset is thus less than 0.001.

Found at: doi:10.1371/journal.pbio.1000506.s009 (0.77 MB PDF)

**Figure S10 Annotation of the histone gene family in *Apis mellifera* showing the methylation profiles.** See Table 4 for details.

Found at: doi:10.1371/journal.pbio.1000506.s010 (0.34 MB PDF)

**Table S1 Sequence conservation of methylated and non-methylated genes.** (A) Number of high and low CpG *Apis* genes with blast hits to different species at various E-value thresholds. The amino acid sequences of the genes were compared. Fisher exact tests were conducted to assess whether significantly more low CpG genes have a blast hit than high CpG genes. (B) Number of methylated and non-methylated honey bee genes with blast hits to different model species at various E-value thresholds. The amino acid sequences of the genes were compared. Fisher exact tests were conducted to assess whether significantly more methylated genes have a blast hit than non-methylated genes. (C) Number of high CpG methylated and non-methylated honey bee genes with blast hits to different model species at various E-value thresholds. The amino acid sequences of the genes were compared. Fisher exact tests were conducted to assess whether significantly more high CpG methylated genes have a blast hit than high CpG non-methylated genes.

Found at: doi:10.1371/journal.pbio.1000506.s011 (0.19 MB DOC)

**Table S2 Differentially methylated genes in queens and worker brains.** A generalized linear model of the binomial family was used to identify genes that are differentially methylated between castes. The methylation level of each gene was modeled as a function of the caste and of each of its CpG dinucleotides. In the table, “Caste” indicates whether the caste is a statistically significant factor explaining differences in methylation levels, “CpG” represents the different dinucleotides of that gene, and “Caste \* CpG,” the interaction factor, indicates whether the CpG dinucleotides behave differently between castes. GB numbers refer to the proteins at BeeBase: genomes.arc.georgetown.edu/drupal. Genes were ranked into 10 bins based on their expression levels from low (1) to high (10). No value in the relative expression column indicates those genes that are not represented on the microarray. Based on microarray data from Foret et al. [10].

Found at: doi:10.1371/journal.pbio.1000506.s012 (1.13 MB DOC)

**Table S3 Evaluation of the Elango et al. hypothesis.** (A) CpG o/e in methylated genes. (B) Differential methylation and differential gene expression.

Found at: doi:10.1371/journal.pbio.1000506.s013 (0.04 MB DOC)

**Table S4 Details on genes used for deep 454 sequencing.** Found at: doi:10.1371/journal.pbio.1000506.s014 (0.03 MB XLS)

## Acknowledgments

We thank Berit Haldemann, Andre Leischwitz, and Matthias Schaefer for their help with various aspects of sequencing. We thank George Gabor L. Miklos for stimulating discussions and critical comments on the manuscript and Joanna Maleszka for providing the biological materials used in this study. We also thank Fiona Wilkes for editorial help, Ros Attenborough for drafting the non-technical summary and three anonymous reviewers for their constructive comments.

## Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: FL RM.

Performed the experiments: SF RK SW CF. Analyzed the data: FL SF RK. Contributed reagents/materials/analysis tools: FL RM. Wrote the paper: RM.

## References

- Maleszka R (2008) Epigenetic integration of environmental and genomic signals in honey bees: the critical interplay of nutritional, brain and reproductive networks. *Epigenetics* 3: 188–192.
- Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honey bees via DNA methylation. *Science* 319: 1827–1830.
- Barchuk AR, dos Santos Cristino A, Kucharski R, Simões ZLP, Maleszka R (2007) Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev Biol* 7: 70.
- Cokus SJ, Feng S, Zhang X, Chen Z, et al. (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452: 215–219.
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328: 916–919.
- Colot V, Rossignol JL (1999) Eukaryotic DNA methylation as an evolutionary device. *Bioessays* 21: 402–411.
- Göll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74: 481–514.
- Suzuki MM, Kerr AR, De Sousa D, Bird A (2007) CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* 17: 625–631.
- Wang Y, Jorda M, Jones PL, Maleszka R, et al. (2006) Functional CpG methylation system in a social insect. *Science* 314: 645–647.
- Foret S, Kucharski R, Pittelkow Y, Lockett GA, Maleszka R (2009) Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. *BMC Genomics* 10: 472.
- Lister R, Pelizzola M, Downen RH, Hawkins RD, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462: 315–322.
- Wang Y, Leung FC (2009) In silico prediction of two classes of honeybee genes with CpG deficiency or CpG enrichment and sorting according to gene ontology classes. *J Mol Evol* 68: 700–705.
- Elango N, Hunt BG, Goodisman MA, Yi SV (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci U S A* 106: 11206–11211.
- Gonzalo S, Jaco I, Fraga MF, Chen T, et al. (2006) DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nature Cell Biol* 8: 416–424.
- Honey Bee Genome Consortium (2006) Insights into social insects from the genome of the honey bee *Apis mellifera*. *Nature* 443: 931–949.
- Vidovic M, Nighorn A, Koblar S, Maleszka R (2007) Eph receptor and Ephrin signaling in developing and adult brain of the honeybee (*Apis mellifera*). *Dev Neurobiol* 67: 233–251.
- Zinke I, Schütz CS, Katzenberger JD, Bauer M, Pankratz MJ (2002) Nutrient control of gene expression in *Drosophila*: microarray analysis of starvation and sugar-dependent response. *EMBO J* 21: 6162–6173.
- Talbert PB, Henikoff S (2010) Histone variants—ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* 11: 264–275.
- Gabor Miklos GL, Maleszka R (2010) Epigenomic communication systems in humans and honey bees: from molecules to behaviour *Hormones & Behavior*; doi:10.1016/j.yhbeh.2010.05.016.
- Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. *Science* 187: 226–232.
- Riggs AD (1975) X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* 14: 9–25.
- Regev A, Lamb M, Jablonka E (1998) The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol Biol Evol* 15: 880–891.
- Butte AJ, Dzau VJ, Glueck SB (2001) Further defining housekeeping, or “maintenance,” genes: focus on “a compendium of gene expression in normal human tissues.” *Physiol Genomics* 7: 95–96.
- Sugden K, Pariente C, McGuffin P, Aitchison K, D’Souza U (2008) Housekeeping gene expression is affected by antidepressant treatment in a mouse fibroblast cell line. *J Psychopharmacol*; doi:10.1177/0269881108099690.
- Gibson G (2008) The environmental contribution to gene expression profiles. *Nat Rev Genet* 9: 575–581.
- Burzynski SR, Patil S, Ilkowska-Musial E, Chitur S, et al. (2008) Pathway analysis of the effect of chromatin remodeling agent phenylbutyrate on the brains of honeybees. Society for Neuroscience, Washington 2008, Abstract 494.2/UU42.
- Ricobaraza A, Cuadrado-Tejedor M, Pe’rez-Mediavilla A, Frechilla D, et al. (2009) Phenylbutyrate ameliorates cognitive deficit and reduces tau pathology in an Alzheimer’s disease mouse model. *Neuropsychopharmacology* 34: 1721–1732.
- Kang HL, Benzer S, Min KT (2002) Life extension in *Drosophila* by feeding a drug. *Proc Natl Acad Sci U S A* 99: 838–843.
- Groh C, Rössler W (2008) Caste-specific postembryonic development of primary and secondary olfactory centers in the female honeybee brain. *Arthropod Struct Dev* 37: 459–468.
- Miller CA, Sweatt JD (2007) Covalent modification of DNA regulates memory formation. *Neuron* 53: 857–869.
- Miller CA, Campbell SL, Sweatt JD (2008) DNA methylation and histone acetylation work in concert to regulate memory formation and synaptic plasticity. *Neurobiol Learn Mem* 89: 599–603.
- Lockett GA, Helliwell P, Maleszka R (2010) Involvement of DNA methylation in memory processing in the honey bee. *Neuroreport* 21: 812–816.
- Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nature Rev Genet* 10: 295–304.
- Ball MP, Li JB, Gao Y, Lee JH, et al. (2009) Cytosine methylation may interact with other epigenetic features, such as histone modifications. *Nature Biotech* 27: 361–368.
- Luco RF, Pan Q, Tominaga K, Blencowe BJ, et al. (2010) Regulation of alternative splicing by histone modifications. *Science* 327: 996–1000.
- Laurent L, Wong E, Li G, Huynh T, et al. (2010) Dynamic changes in the human methylome during differentiation. *Genome Res* 20: 1–12.
- Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* 39: 61–69.
- Wittkopp PJ (2007) Variable gene expression in eukaryotes: a network perspective. *J Exp Biol* 210: 1567–1575.
- Feinberg AP, Irizarry RA (2010) Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A* 107: 1757–1764.
- Xi Y, Li W (2009) BSMAP: whole genome Bisulfite Sequence MAPPING program. *BMC Bioinformatics* 10: 232.
- Altshuler D, Pollara V, Cowles C, Van Etten, et al. (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407: 513–516.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B* 57: 289–300.
- Kucharski R, Maleszka R (2005) Microarray and rtPCR analyses of gene expression in the honey bee brain following caffeine treatment. *J Mol Neurosci* 27: 269–276.