LETTER

doi:10.1038/nature23891

Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer

Zachary D. Smith[1,2,3]*, Jiantao Shi[4,5]*, Hongcang Gu[1], Julie Donaghey[1,2], Kendell Clement[1,2,6], Davide Cacchiarelli[1,2], Andreas Gnirke[1], Franziska Michor[1,2,4,5]§ & Alexander Meissner[1,2]†§

In mammals, the canonical somatic DNA methylation landscape is established upon specification of the embryo proper and subsequently disrupted within many cancer types[1–4]. However, the underlying mechanisms that direct this genome-scale transformation remain elusive, with no clear model for its systematic acquisition or potential developmental utility[5,6]. Here, we analysed global remethylation from the mouse preimplantation embryo into the early epiblast and extraembryonic ectoderm. We show that these two states acquire highly divergent genomic distributions with substantial disruption of bimodal, CpG density-dependent methylation in the placental progenitor[7,8]. The extraembryonic epigenome includes specific *de novo* methylation at hundreds of embryonically protected CpG island promoters, particularly those that are associated with key developmental regulators and are orthologously methylated across most human cancer types[9]. Our data suggest that the evolutionary innovation of extraembryonic tissues may have required co-option of DNA methylation-based suppression as an alternative to regulation by Polycomb-group proteins, which coordinate embryonic germ-layer formation in response to extraembryonic cues[10]. Moreover, we establish that this decision is made deterministically, downstream of promiscuously used—and frequently oncogenic—signalling pathways, via a novel combination of epigenetic cofactors. Methylation of developmental gene promoters during tumorigenesis may therefore reflect the misappropriation of an innate trajectory and the spontaneous reacquisition of a latent, developmentally encoded epigenetic landscape.

To compare how epigenetic landscapes evolve during early mammalian development, we generated whole-genome bisulfite sequencing (WGBS) and RNA sequencing (RNA-seq) datasets from mouse precompacted 8-cell stage embryos, inner cell mass (ICM) and trophectoderm from embryonic day (E)3.5 blastocysts, as well as epiblast and extraembryonic ectoderm (ExE) from E6.5 conceptuses, the latest stage at which these progenitors remain largely homogeneous and undifferentiated (Fig. 1a, Extended Data Fig. 1, Supplementary Tables 1 and 2). Holistically, our time series captures the expected transition through the indistinguishably hypomethylated—but transcriptionally distinct—blastocyst-stage tissues, followed by a considerable departure at implantation, at which approximately 80% of the genome becomes differentially methylated (Extended Data Fig. 2a). Specifically, the extraembryonic lineage lacks canonical bimodality: most CpGs are incompletely methylated in comparison to the epiblast and 1.36% are methylated in the ExE (Fig. 1b, Extended Data Fig. 2b). ExE-specific hypo- or hypermethylated CpGs segregate into distinct genomic compartments by CpG density and location, with *de novo* methylation preferentially enriched for CpG islands (CGIs) near transcription start sites (TSSs) and 5′ exons (Fig. 1c, Extended Data Fig. 2c–f). Once established, these two alternative landscapes are largely preserved across embryonic tissues or in the midgestation placenta, respectively[11,12] (Extended Data Fig. 2g).

Notably, ExE-methylated CGIs (ExE hyper CGIs) frequently overlap with Polycomb repressive complex 2 (PRC2)-regulated genes, including master transcription factors that direct germ-layer and body-axis formation (Extended Data Fig. 3a, b, Supplementary Table 1). Although the majority of targeted genes are not yet expressed in the epiblast, ExE-specific promoter methylation is generally associated with repression, including of many pluripotency-specific regulators, as well as concurrent loss of chromatin accessibility (Extended Data Figs 3, 4). Moreover, the global relationship between promoter methylation and gene repression is more pronounced in the ExE than in the epiblast (Extended Data Fig. 4c). DNA methylation surrounding these promoters is largely dispersive, with flanking regions less methylated in the ExE than in the epiblast, but with a maximal increase specifically at the TSS (Fig. 1d, e). ExE hyper CGIs only reach methylation levels of ~0.25, but methylated CpGs are distributed across 80% of the sequencing reads that fall within them and have a median per-read methylation status that matches the unphased measurement (Extended Data Fig. 4d). The consistency between per-molecule and aggregate methylation is most likely to be explained by population-wide recruitment of *de novo* methyltransferases, followed by stochastic gains at individual CpGs in phase, similar to a variety of cancer systems[13,14]. Importantly, the higher CpG density of ExE-targeted regions leads invariably to a higher local methylation density, even though the per-CpG methylation status is intermediate (Fig. 1e).

Suppression overlaps with WNT pathway effectors that are induced in the proximal epiblast to promote primitive streak formation (Fig. 2a). However, the ExE expresses alternative WNT proteins, suppresses fibroblast growth factor (*Fgf*) promoters by *de novo* methylation, and specifically expresses receptors for epiblast-secreted factors (Fig. 2b, Extended Data Fig. 5a). The extraembryonic landscape may proceed deterministically from these two major signalling pathways, which are used promiscuously in many downstream developmental processes and frequently misregulated in cancers. To investigate this hypothesis, we selected the ICM as a model because it is indistinguishably hypomethylated from the trophectoderm and can be cultured independently of FGFs, whereas extraembryonic development rapidly attenuates[15]. ICMs were cultured in four conditions using combinations of FGF4, the mitogen-activated protein kinase kinase (MAPKK or MEK) inhibitor PD0325901, and the GSK3β inhibitor, WNT agonist CHIR99021 (CHIR) (Fig. 2c, Extended Data Fig. 5b). Isolated outgrowths were dually assayed by a combined RNA-seq and reduced representation bisulfite sequencing (RRBS) approach (Extended Data

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [2]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA. [3]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA. [4]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [5]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA. [6]Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts, USA.
†Present address: Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany.
*These authors contributed equally to this work.
§These authors jointly supervised this work.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

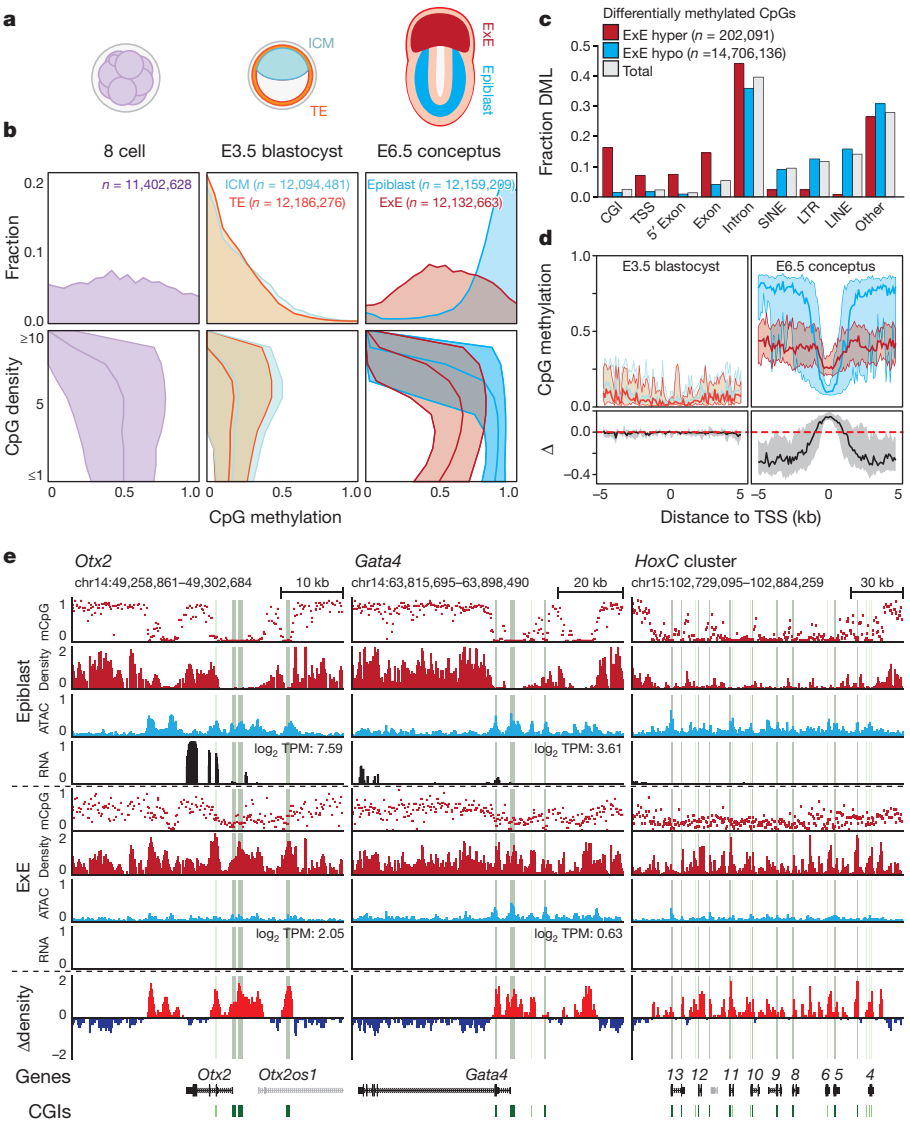


Figure 1 | Divergent postimplanation DNA methylation landscapes. a, Early developmental time series collected for this study, including precompacted 8-cell stage embryos (2.25 days post fertilization; E2.25), trophectoderm (TE) and inner cell mass (ICM) of the E3.5 blastocyst, and extraembryonic ectoderm (ExE) and epiblast of the E6.5 conceptus ($n = 2$ WGBS libraries per sample, see Methods). **b**, CpG methylation distribution for 100-bp tiles (top); median 100 bp tile methylation as a function of local CpG density (bottom), where the shaded area represents the 25th and 75th percentiles. **c**, Feature-level enrichment for differentially methylated CpGs compared to genomic background. ExE-hypomethylated CpGs are predominantly found in non-genic sequences, whereas ExE-hypermethylated CpGs localize to CpG islands (CGIs), transcription start sites (TSSs) and 5′ exons. Here, TSS refers to the 1 kb upstream of an annotated TSS only, whereas 5′ exon and exons represent non-overlapping sets. **d**, Median methylation architecture flanking ExE-hypermethylated TSSs within embryonic and extraembryonic tissues, as well as the relative methylation difference (Δ), which diverges considerably upon implantation. The shaded area represents the 25th and 75th percentiles per 100-bp bin. **e**, Genome browser tracks for WGBS, assay for transposase accessible chromatin with high-throughput sequencing (ATAC–seq) and RNA-seq data capturing three emblematic loci. Density refers to the projected number of methylated CpGs per 100 bp of primary sequence and highlights the extensive epigenetic signal present over these regions within ExE (Δdensity refers to the difference compared to the epiblast). For *Otx2* and *Gata4*, ExE-specific methylation and repression are concurrent, whereas the *HoxC* cluster is expressed later in embryonic development. CGIs are highlighted in green. TPM, transcripts per million.

Fig. 6, Methods). Those cultured in FGF4 plus CHIR progressively diverged into two separate, morphologically distinguishable interior and exterior tissues that were independently isolated.

In combination, PD0325901 and CHIR comprise the '2i' condition, an FGF-impeded, WNT-activated state that maintains preimplantation-like global hypomethylation[16]. Alternatively, exogenous FGF is sufficient to drive genome and CGI methylation to higher than physiological levels (Fig. 2d). Surprisingly, when coupled with FGF, WNT agonism effectively blocks genome remethylation but redirects CGI-level methylation to a greater subset of extraembryonic targets (Fig. 2e, f). CGI-targeting is specific to the FGF plus CHIR outgrowth exterior, which establishes an asymmetric Fgfr2 and Fgf4 expression pattern with the interior, similar to what occurs *in vivo* (Supplementary Tables 3, 4). The specific overlap between *in vitro* and ExE-methylated CGI promoters appears to reflect progressive restriction of potential targets over early development: those shared across conditions have early developmental functions and are often expressed in the ICM and the 2i condition such as *Prdm14*; those methylated in the ExE and in FGF plus CHIR, but not in FGF alone, generally encompass neuroectodermal regulators such as *Otx2* and *Pax6*; and ExE-exclusive targets are often endodermal and induced by dual FGF and WNT activity such as *FoxA2* and *Sox17* (Extended Data Fig. 5c). Seemingly, ExE-like global hypomethylation and CGI methylation can be recapitulated *in vitro* by WNT and FGF, but target specificity can be modulated to include multiple discrete developmental programmes.

We next sought to investigate the configuration of epigenetic regulators that specifically execute this transition. Whereas Dnmt1 and Dnmt3b are expressed in both tissues, Dnmt3l and Dnmt3a isoform 2 are reciprocally expressed in either the ExE or the epiblast and regulated by *de novo* promoter methylation in the other (Extended Data Fig. 7a–d). A truncated, non-catalytic isoform of the histone 3 lysine 36 (H3K36) demethylase Kdm2b is expressed during preimplantation and within the ExE, whereas a longer Jumonji demethylase domain-containing isoform is specifically induced in the epiblast[17] (Extended Data Fig. 7e). Otherwise, epigenetic regulator expression appears relatively stable between the two tissues at this time, such that their specific integration could explain the assembly of such profoundly different landscapes. To compare their capacity to direct both global and CGI methylation, we acutely disrupted *Dnmt1*, *Dnmt3a*, *Dnmt3b*, and *Dnmt3l*, the essential PRC2 component *Eed*, and *Kdm2b* by zygotic CRISPR–Cas9 injection (Supplementary Tables 5, 6, Methods). We found that *Dnmt1*, *Dnmt3b*, and *Dnmt3l* ablation substantially disrupt the ExE methylome, including at CGI targets, but show no obvious specificity for these regions or corresponding changes in expression (Fig. 3a, b, Extended Data Fig. 7f–h). The near complete loss of methylation in *Dnmt1*-null ExE compared to sample-matched epiblast indicates diminished *de novo* activity, and greater reliance

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

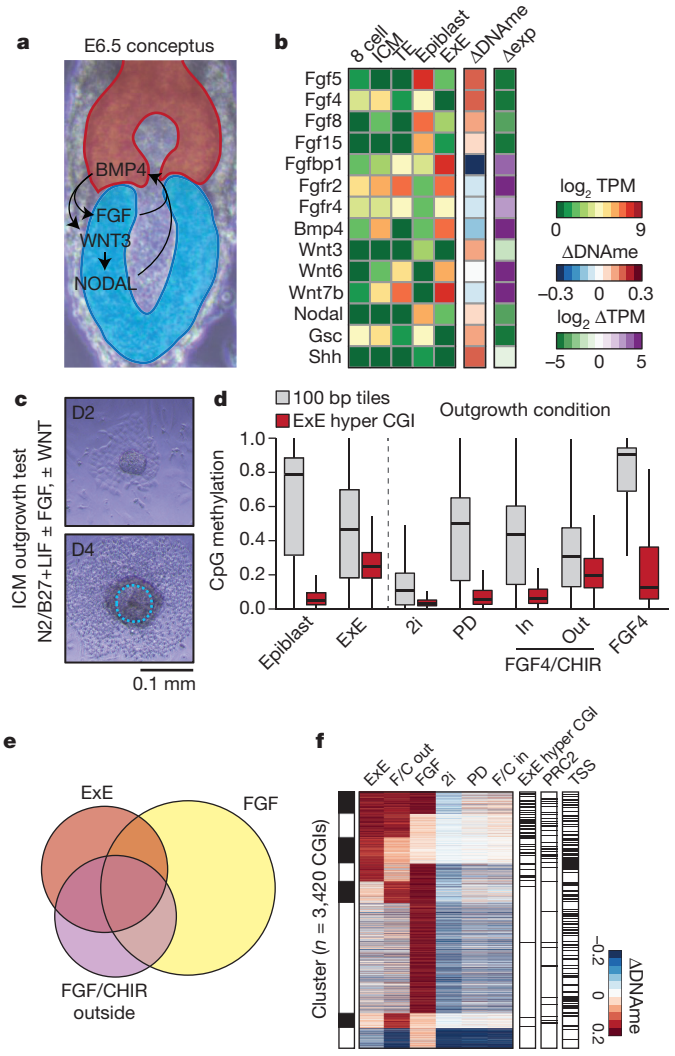


Figure 2 | *De novo* methylation of early developmental gene promoters can be modulated by external conditions. a, Schematic of signalling pathway interactions between the epiblast (blue) and the ExE (red). Epiblast-produced fibroblast growth factors (FGFs) promote ExE development, which expresses bone morphogenic protein 4 (BMP4) to induce WNT proteins in the epiblast. Epiblast secreted pro-Nodal is processed by the ExE to establish a proximal–distal gradient and the primitive streak[10]. **b**, Differential expression and promoter methylation of key signalling components between the ExE and epiblast. Many Fgfs and associated receptors exhibit reciprocal expression and promoter methylation. Wnt3 induction is apparent in the epiblast, whereas Wnt6 and 7b are highly expressed in both the trophectoderm and the ExE. Differential promoter methylation refers to the annotated TSS (±1 kb) with the greatest absolute difference (Supplementary Table 2). **c**, ICM outgrowths are cultured for four days under disparate growth factor or small molecule conditions intended to either stimulate or repress FGF and WNT activity. The outline highlights the purified component (Methods). **d**, Methylation boxplots for the conditions described in **c**, including all RRBS-captured 100 bp tiles and ExE-targeted CGIs (ExE hyper CGIs). Edges refer to the 25th and 75th percentiles, whiskers the 2.5th and 97.5th percentiles, respectively. **e**, The ExE, FGF/CHIR exterior, and FGF outgrowth all display substantial CGI methylation. Shown is the intersection of methylated CGIs with ≥0.1 increase in comparison to the epiblast ($n = 3,420$). The FGF4 condition has the highest number of methylated CGIs, but fewer intersect with ExE than when CHIR is also present: 25% of ExE hyper CGIs overlap with both conditions, whereas 51% overlap with the FGF/CHIR exterior outgrowth (outside). **f**, Clustering of differentially methylated CGIs from **e**, with methylation status in the ExE, embryonic regulation by PRC2, and TSS proximity (± 2 kb) included. F/C in and out refer to the interior and exterior FGF/CHIR outgrowth conditions, respectively.

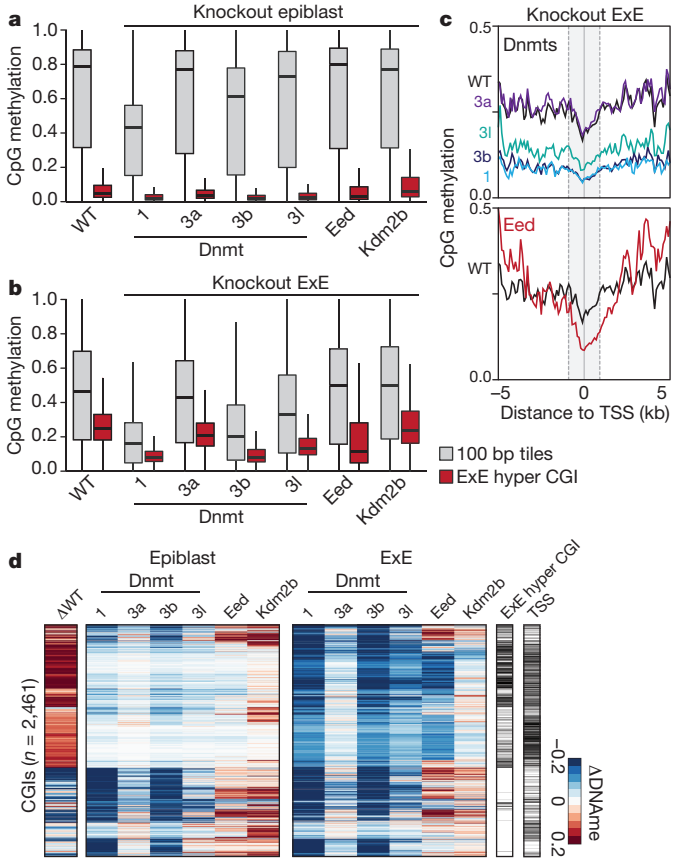


Figure 3 | A novel configuration of epigenetic regulators contributes to the extraembryonic methylation landscape. a, Boxplots for E6.5 epiblast tissue for wild-type (WT) and CRISPR–Cas9 disrupted samples, including for 100 bp tiles and ExE hyper CGIs, as measured by RRBS. Edges refer to the 25th and 75th percentiles, whiskers the 2.5th and 97.5th percentiles, respectively. **b**, Boxplots as in **a** for sample-matched ExE. In comparison to the Dnmt3a- and Dnmt3b-positive epiblast, *Dnmt1* or *Dnmt3b* disruption have a far greater effect on global methylation levels and result in a highly depleted genome. **c**, Composite plots of ExE hyper CGIs by knockout status. CGI methylation is disrupted in *Eed*-null ExE, particularly within +1 kb of the TSS, without affecting global levels. The black line represents the wild-type median. Composite plots map the median of 200-bp windows over 50-bp intervals from RRBS data. Grey box indicates ±1 kb of the TSS. **d**, Heat map of the differential CGI methylation (≥0.1) between the CRISPR–Cas9-targeted epiblast or ExE compared to their wild-type counterparts ($n = 2,461$). Differential ExE methylation status in comparison to epiblast and TSS proximity (±2 kb) are included for reference.

on epigenetic maintenance, despite prolonged Dnmt3l expression (Fig. 3a, b). Alternatively, *Eed*-null ExE disrupts CGI methylation without affecting global levels, suggesting that PRC2 may specifically coordinate repression upstream of DNMT3B as part of a novel developmental pathway (Fig. 3b–d, Extended Data Fig. 7f, g). Consistently, *Eed*-null ExE fails to suppress associated genes, which are induced to similar levels to those of the sample-matched epiblast (Extended Data Fig. 7h).

Our data indicate a point in early development at which sensitivity to promiscuously used growth factors instructs a distinct epigenome that is not observed during downstream ontogeny. However, *de novo* CGI methylation is also a general feature of tissue culture, cancer cell lines, and primary tumours, indicating a latent vulnerability in somatic cells[5,18] (Fig. 4a, Extended Data Figs 8, 9). To investigate a possible link with the subsequent re-emergence of this landscape in cancer, we mapped orthologous CGIs to compare patient-matched DNA methylation profiles from The Cancer Genome Atlas (TCGA) project, an age-matched chronic lymphocytic leukaemia (CLL) cohort, as well as

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

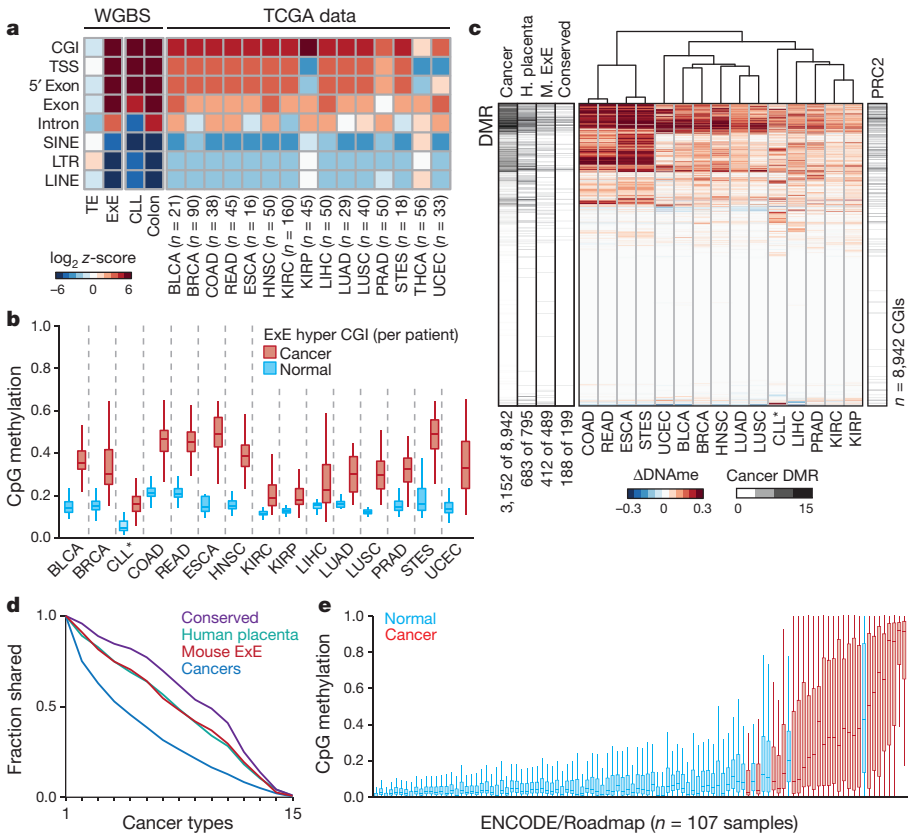


Figure 4 | Extraembryonically-targeted CpG islands are pervasively methylated across human cancer types. a, Disruption of global methylation creates similar biases for CGIs and promoters between ExE/epiblast or patient- or age-matched normal/cancer tissue comparisons. Heat map shows the log *z*-score enrichment for features by the binomial test for the above comparisons, as well as for trophectoderm/ICM (TE). Of these 16 cancer types, only THCA does not display a notably dysregulated methylome. *n* values refer to the number of matched cancer/normal tissue isolates for each type. The Cancer Genome Atlas (TCGA) samples include bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), colorectal adenocarcinoma (READ), oesophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillar cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), stomach and oesophageal carcinoma (STES), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC). Here, chronic lymphocytic leukaemia (CLL) to B lymphocyte comparison is between age-matched samples measured by WGBS. **b**, Feature level boxplots of 489 ExE hyper CGIs that preserve their status in humans, calculated as a feature per cancer or normal tissue for the 15 cancer types in which CGI methylation is generally apparent. Asterisk: CLL samples were measured by RRBS ($n = 119$) and represent a comparison between age-matched healthy B lymphocytes ($n = 24$). Edges refer to the 25th and 75th percentiles, whiskers the 2.5th and 97.5th percentiles, respectively. **c**, Differential methylation heat map for 8,942 orthologous CGIs measured in TCGA or by RRBS and clustered by Euclidean distance. The DMR bar includes the cumulative number of cancer types where a given CGI is called as hypermethylated, as well as the DMR status in either human placenta compared to human embryonic stem (ES) cells (H. placenta), mouse ExE compared to epiblast (M. ExE), or shared between both comparisons (Conserved). PRC2 denotes regulation by Polycomb in human ES cells. The numbers reflect the proportion of each set that is differentially methylated in at least one cancer type. **d**, Intersection analysis for DMR status across TCGA and CLL samples. Both placenta and ExE DMRs are similarly enriched for methylation in at least one human cancer type (86% and 84%, respectively, compared to 35% for all CGIs) and are more frequently methylated across them. Enrichment for conserved DMRs is greater than for extraembryonic DMRs from each individual species, and 94% are methylated in at least one cancer type. **e**, Boxplots of orthologous ExE hyper CGIs across 107 ENCODE/Roadmap Epigenomics Project samples, ranked by mean methylation and with cancer or cancer cell line assignment highlighted (red). 'Normal' assigned samples that sort with cancer include the trophoblast cell line HTR8svn, primary colon and colonic mucosa, placenta, and $CD8^+$ T lymphocytes, in descending order. Extended Data Fig. 9a and Supplementary Table 7 include additional sample characteristics.

data from the Encyclopedia of DNA Elements (ENCODE) and the Roadmap Epigenomics Project[14,19–21]. Of the 16 cancer types with sufficient normal biopsied samples, 15 significantly methylate ExE hyper CGIs (Fig. 4a, b). The signal is surprisingly robust and segregates cancer and normal tissue when measured as a feature across patients or when examining CGI-level changes (Fig. 4b, Extended Data Fig. 8). 84% of ExE hyper CGIs are methylated in at least one cancer type, and they are more frequently shared as conserved, pan-cancer targets (Fig. 4c, d, Extended Data Fig. 10a, b). We find some direct and indirect evidence that CGI methylation can be influenced by FGF sensing. For example, matched mutational and methylation analyses of the entire TCGA dataset ($n = 10{,}629$ cancers) show a 19.3% increase in the average methylation of ExE hyper CGIs when any FGF pathway member is mutated (from 0.275 to 0.328, Extended Data Fig. 10c). Similarly, statistical assessment of the connectivity between our ExE hyper CGIs and the 10 most mutated pathways in cancer reveals a notable enrichment for FGFR signalling in disease (enrichment *z*-score = 3.88, Extended Data Fig. 10d, e). Over the more expansive, but less internally controlled, ENCODE and Roadmap data, cancers and immortalized cell lines are clearly separated from primary tissues by their ExE hyper CGI methylation status (Fig. 4e, Supplementary Table 7). Notably, mature adaptive immune cells and endodermal lineages are generally more susceptible to low-level methylation within these regions, suggesting a pre-existing heterogeneity even in normal populations.

We present the developmental acquisition of an epigenetic landscape that partitions extraembryonic tissues within the embryo

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

and resembles a frequent, global departure in genome regulation in human cancers. This landscape co-occurs with the establishment of the first major signalling axes, can be partially directed from the hypomethylated ICM *in vitro*, and appears to be determined by disparate regulation of the DNMTs and associated cofactors. Notably, *de novo* methylation of CGIs in the ExE requires PRC2, which indicates a transient, biochemical interaction with DNMT3B or an upstream role in either determining the ExE state or priming CGIs for suppression. The coordination of this alternative, and presumably more permanent, repressive mechanism warrants further investigation and shares features with the somatic transition to cancer. Most obviously, FGF sensing passes through RAS/MAPK/ERK signalling, which has extensive oncogenic potential and putative roles in the establishment of the cancer methylome[22–24]. Similarly, the ExE displays attenuated *de novo* methylation activity directed wholly by DNMT3B, broadly resembling the high frequency of somatic *DNMT3A* mutations in acute myeloid leukaemia and myelodysplastic syndrome or DNMT3B-directed CGI methylation during colorectal transformation[25–28]. Transgenic mouse cancer models confirm conserved ExE hyper CGI methylation in similar contexts (Extended Data Fig. 10f). The extraembryonic landscape depends on extrinsic cues with numerous downstream developmental functions, which may provide a latent opportunity for spontaneous state transition without genetic perturbation in later development. If so, the likelihood of such a transition may relate to how closely a given regulatory network resembles the one governing extraembryonic specification. Whether or not additional morphological and molecular features of placental development that appear analogous to cancer hallmarks[29,30]—such as immunosuppression, tissue invasion, and angiogenesis—proceed as part or downstream of this primary epigenetic switch remains unexplored, but would provide a parsimonious developmental foundation for their systematic emergence during transformation.

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 7 July 2016; accepted 3 August 2017.
Published online 20 September 2017.

1. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–220 (2013).
2. Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.* **39,** 237–242 (2007).
3. Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nat. Genet.* **39,** 232–236 (2007).
4. Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet.* **39,** 157–158 (2007).
5. Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* **7,** 21–33 (2006).
6. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357,** eaal2380 (2017).
7. Schroeder, D. I. *et al.* The human placenta methylome. *Proc. Natl Acad. Sci. USA* **110,** 6037–6042 (2013).
8. Branco, M. R. *et al.* Maternal DNA methylation regulates early trophoblast development. *Dev. Cell* **36,** 152–163 (2016).
9. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–1022 (2011).
10. Arnold, S. J. & Robertson, E. J. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10,** 91–103 (2009).
11. Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* **45,** 1198–1206 (2013).
12. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500,** 477–481 (2013).
13. Landan, G. *et al.* Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* **44,** 1207–1214 (2012).
14. Landau, D. A. *et al.* Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* **26,** 813–825 (2014).
15. Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J. K. & Lonai, P. Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proc. Natl Acad. Sci. USA* **95,** 5082–5087 (1998).
16. Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20,** 311–316 (2013).
17. Boulard, M., Edwards, J. R. & Bestor, T. H. Abnormal X chromosome inactivation and sex-specific gene dysregulation after ablation of FBXL10. *Epigenet. Chromatin* **9,** 22 (2016).
18. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).
19. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
20. Hoadley, K. A. *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158,** 929–944 (2014).
21. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).
22. MacLeod, A. R., Rouleau, J. & Szyf, M. Regulation of DNA methylation by the Ras signaling pathway. *J. Biol. Chem.* **270,** 11327–11337 (1995).
23. Lu, C. W. *et al.* Ras-MAPK signaling promotes trophectoderm formation from embryonic stem cells and mouse embryos. *Nat. Genet.* **40,** 921–926 (2008).
24. Serra, R. W., Fang, M., Park, S. M., Hutchinson, L. & Green, M. R. A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. *eLife* **3,** e02313 (2014).
25. Ley, T. J. *et al. DNMT3A* mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363,** 2424–2433 (2010).
26. Walter, M. J. *et al.* Recurrent *DNMT3A* mutations in patients with myelodysplastic syndromes. *Leukemia* **25,** 1153–1158 (2011).
27. Rhee, I. *et al.* DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416,** 552–556 (2002).
28. Lin, H. *et al.* Suppression of intestinal neoplasia by deletion of *Dnmt3b*. *Mol. Cell. Biol.* **26,** 2976–2983 (2006).
29. Novakovic, B. & Saffery, R. Placental pseudo-malignancy from a DNA methylation perspective: unanswered questions and future directions. *Front. Genet.* **4,** 285 (2013).
30. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).

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

Acknowledgements We thank members of the Meissner and Michor laboratories for discussions and advice, in particular R. Karnik for help with data processing and alignment, as well as B. E. Bernstein and R. P. Koche for their expertise. F.M. and J.S. gratefully acknowledge support from the Dana-Farber Cancer Institute Physical Sciences-Oncology Center (NIH U54CA193461). The work was funded by the New York Stem Cell Foundation, the Broad-ISF Partnership for Cell Circuit Research, the Starr Foundation, NIH grants (1P50HG006193, P01GM099117 and R01DA036898) and the Max Planck Society. A.M. is a New York Stem Cell Foundation Robertson Investigator.

Author Contributions Z.D.S., J.S., F.M. and A.M. designed and conceived the study and prepared the manuscript. Z.D.S. performed all experiments and assisted in data analysis as performed by J.S. J.D. made the ATAC–seq libraries, D.C. made RNA-seq libraries, and H.G. made the dual RRBS and RNA-seq libraries with supervision from A.G. and alignment by K.C. F.M. and A.M. jointly supervised the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to F.M (michor@jimmy.harvard.edu) or A.M. (meissner@molgen.mpg.de).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Sample isolation and library preparation. Preparation of preimplantation and postimplantation samples was performed as described in ref. 31. In brief, B6D2F1 hybrid females between 5 and 8 weeks old (Charles River) were serially primed with 5 IU pregnant mare gonadotropin (Sigma) followed by 5 IU human chorionic gonadotropin (Millipore) after 46 h, and subsequently mated with B6D2F1 male mice ≤6 months old. For preimplantation time points, zygotes from mated females were isolated from the oviduct the following morning (E0.5) and cultured in KSOM media (Millipore) droplets under mineral oil until E2.25. The 8-cell sample was collected by careful monitoring of 4-cell embryos from ~E2 onwards, and emergent 8-cell embryos were swapped into KSOM supplemented with 1 μg ml^{-1} aphidicolin (Sigma) to ensure synchronization and minimal entry into the fourth replication cycle. 8-cell embryos were collected within 4 h of the first apparent embryo of this stage. Prior to collection, embryos were serially transferred through Acidic Tyrode's solution (Sigma) to remove the zona pellucida and carefully pipetted with a drawn glass capillary through 0.25% Trypsin-EDTA (Life Technologies) to remove maternal polar bodies. E3.5 blastocysts were also treated with Acidic Tyrode's solution to remove the zona, and the ICM and trophectoderm of matched samples were dissected using standard micromanipulation equipment (Eppendorf) and a Hamilton Thorne XYClone laser with 300 μs pulsing at 100% intensity. Isolation of postimplantation tissues was performed as described[32]. The deciduae of mated female mice were isolated on the morning of E6.5 and the conceptuses removed. Then, under a stereomicroscope, the embryo was carefully bisected along the extraembryonic–embryonic axis, removing the ectoplacental cone from the extraembryonic ectoderm when apparent. After separation, the epiblast and the ExE were incubated for 15 min at 4 °C in 0.5% trypsin, 2.5% pancreatin dissolved in PBS and allowed to rest for 5–10 min in KSOM at room temperature. Finally, the visceral endoderm was removed by drawing the embryo through a narrow, flame-drawn glass capillary and only samples with no apparent contamination were collected. On average, matched ExE and epiblast or ICM and trophectoderm samples from 5–10 embryos or from 20 or more 8-cell embryos were collected per assay.

DNA for whole-genome bisulfite sequencing was isolated as described previously[33], and libraries were prepared using the Accel-NGS Bisulfite DNA library kit (Swift Biosciences) according to the manufacturer's protocol. Final libraries were generated from 10–12 PCR cycles. RNA was purified using the RNeasy Micro Kit (Qiagen) and RNA-seq libraries were generated using the SMRT-seq v4 Ultra Low Input Kit (Clontech) according to the manufacturer's protocol with 10–11 long-distance PCR cycles. Libraries were generated from 150 pg of the subsequent cDNA using the Nextera XT DNA library preparation kit (Illumina) and 12 PCR cycles. ATAC–seq libraries were generated as described previously[34] using a 10 μl reaction and incubation with the TN5 transposase mixture (Nextera DNA library preparation kit, Illumina) for 45 min. The reaction was stopped according to the protocol described previously[35] and purified using silane beads (Thermo Fisher). Tagmented DNA was amplified for 12–14 cycles to generate the library. WGBS libraries were sequenced as a pool using the HiSeq X Ten platform (Illumina), and RNA-seq and ATAC–seq data were sequenced using the HiSeq 2500 (Illumina).

Outgrowth experiments. To generate controlled outgrowth data, ICM were immunosurgically isolated from BDF1×129S1/SvImJ strain blastocysts at 96 h post-fertilization as described[31]. In brief, oocytes were isolated by hormone priming from B6D2F1 females 12–14 h after administration of human chorionic gonadotropin and fertilized by intracytoplasmic sperm injection using piezo-actuated injection of 129S1/SvImJ strain sperm[36]. At 96 h post-fertilization, blastocysts were stripped of their zona pellucida by brief incubation in Acidic Tyrode's solution and incubated for 30 min in 1:10 diluted whole mouse antisera (Sigma) in CO_2-equilibrated KSOM, followed by destruction of the trophectoderm by culture in 1:10 diluted guinea pig complement sera (Sigma). After 15 min at 37 °C, the ICM separates from the complement-lysed trophectoderm and could be cleanly isolated by brief pulsing through a narrow glass capillary. ICM were isolated in batches of ~12 per drop. Once isolated, ICM were then plated into basal N2/B27 media supplemented with 1,000 U ml^{-1} leukaemia inhibitory factor (made in house) and one of the following conditions; '2i' supplemented with 1 μM PD0325901 and 3 μM CHIR99021(Reagents Direct)[37]; 'PD0325901' supplemented with 1 μM PD0325901 and 10 ng ml^{-1} BMP4 to promote outgrowth expansion (Peprotech)[38]; 'FGF plus CHIR' supplemented with 25 ng ml^{-1} mouse recombinant FGF4 (R&D systems) and 3 μM CHIR99021; and 'FGF' supplemented with 25 ng ml^{-1} FGF4 only. FGF4 was selected because it is the most highly expressed FGF family member in the preimplantation embryo and we sought to direct specific remethylation changes as is observed *in vivo*. ICM were placed into gelatin-treated tissue culture dishes plated with irradiated CF-1 strain embryonic fibroblasts to promote attachment. The primary outgrowth from the ICM, characterized as a centrally expanding, three-dimensional mass, was isolated after four days of culture. In all cases but the 2i condition, an outer layer of differentiated cells was apparent and removed using an identical strategy to that of removal of the visceral endoderm from E6.5 samples described above. However, under the FGF plus CHIR condition, the 'outer layer' was often of the same size or larger than the internal outgrowth, and only became defined during the latter portion of culture (see Extended Data Fig. 5b). As such, we collected both interior and exterior portions as they could clearly be distinguished as mutually ICM-derived. After incubation and either isolation or removal of external cells, outgrowths were serially washed through several KSOM drops under mineral oil before being snap-frozen in minimal volume for RNA-seq and RRBS profiling.

Generation of knockout embryos by zygotic CRISPR–Cas9 injection. Zygotic injection was performed essentially as described[39]. To improve the efficiency with which null alleles were generated, three separate single-guide RNA (sgRNA) sequences were designed per target, prioritizing highly scored protospacer sequences with no high scoring off-target sites using the CHOPCHOP web tool[40] and as 5′ as possible given these constraints to disrupt the coding frame Protospacer sequences were input into the following oligonucleotide primer pair and used to amplify off of the pX300 plasmid (Addgene): forward primer, AGTCAGTTAATACGACTCACTATAGN19GTTTTAGAGCTAGAAATAGC AAG; reverse primer, AAAAAAAGCACCGACTCGGTGCCAC. Protospacer sequences that did not begin with a G to initiate T7 transcription were inserted and an additional 5′ G was added. 200 ng of gel-purified, T7 promoter-containing sgRNA templates were used to generate sgRNAs by *in vitro* transcription using the MEGAshortscript T7 transcription kit (Thermo Fisher), followed by purification with phenol:chloroform and ethanol precipitation. Translation-competent spCas9 RNA was *in vitro* transcribed from a similarly designed, T7 promoter-driven template using the mMESSAGE mMACHINE T7 Ultra kit (Thermo Fisher) and purified using the RNA Clean and Concentrator Kit (Zymo Research). RNA was resuspended in an injection buffer comprising 5 mM Tris-HCl and 0.1 mM EDTA at pH 7.4. Zygotes were isolated from hormone-primed B6D2F1 females mated with B6D2F1 males as described above. Shortly after the formation of visible pronuclei (pronuclear stage 3), zygotes were cytoplasmically injected with 100 ng μl^{-1} of all three targeted sgRNAs pooled 1:1:1 and 200 ng μl^{-1} *Cas9* mRNA. At E3.5, cavitated blastocysts were transferred in clutches of 10–15 into one uterine horn of pseudopregnant CD-1 strain mice (Charles River) that had been mated with vasectomized male Swiss–Weber strain mice (Taconic) two days previously. To account for the ~1 day offset in developmental progression that results from uterine transfer, appropriately E6.5 stage conceptuses were isolated 4 days after uterine transfer and epiblast and extraembryonic ectoderm tissue were isolated as described above before snap-freezing in minimal volume. Each replicate consisted of at least 4 embryos and all experimental series include replicates generated from at least 2 rounds of zygotic injection. Care was taken to ensure epiblast and extraembryonic ectoderm tissue from matched embryos were included for each replicate set, and RRBS data in which both fractions did not cover >1 million CpGs at ≥5× coverage each were excluded from further analysis. Disruption of the target allele was confirmed by PCR amplification from the primary cDNA using primers that flank all three protospacer sequences to capture multiple simultaneous perturbations or truncations in phase.

Dual RNA-seq and RRBS profiling. Genomic DNA and mRNA purifications from low input samples were performed as described previously with modifications[41]. In brief, the cells were mixed with 15 μl of RLT plus buffer (Qiagen) containing 1 U μl^{-1} of SUPERase·In RNase inhibitor (ThermoFisher), 1% β-mercaptoethanol (Sigma), and were then transferred to 1 well in a 96-well DNA LoBind plate (Eppendorf). After adding 10 μl of M-280 streptavidin bead-conjugated reverse transcription primer to each sample, the reaction was incubated at 72 °C for 3 min in a thermocycler followed by incubation at room temperature for 25 min with gentle rotation. The genomic DNA and mRNA were separated in a DynaMag-96 Side Magnet (Thermo Fisher). The bead-tagged mRNA was subjected to reverse transcription as described previously[41] and the genomic DNA in the supernatant was transferred to a fresh 96-well DNA LoBind plate. After reverse transcription, the cDNA was PCR amplified and the RNA-seq library was generated according to the Smart-seq2 protocol[42]. Indexed RNA-seq libraries were pooled and sequenced in an Illumina Hiseq2500 sequencer.

Genomic DNA was isolated using 1× Agencourt AMPure beads (Beckman Coulter) and was eluted with 15 μl of low Tris–EDTA buffer. The RRBS library was generated as reported previously with modifications[43]. We used the CutSmart buffer (New England Biolabs) for all three enzymatic reactions including MspI digestion, end-repair/A-tailing and T4 DNA ligation. To minimize DNA loss, the DNA purification step was eliminated after each enzymatic reaction. In brief, the genomic DNA was digested by 16 units of MspI (New England Biolabs) for

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

80 min at 37 °C, and followed by heat inactivation at 65 °C for 15 min. The digested DNA fragments were end-repaired and A-tailed by adding 4 units of Klenow fragment (3′→5′ exo-) (New England Biolabs), 0.03 mM dCTP, 0.03 mM dGTP and 0.3 mM dATP; the reaction was carried out at 30 °C for 25 min and 37 °C for 25 min, followed by incubation at 70 °C for 10 min to inactive the enzyme. We then ligated the A-tailed DNA fragments with indexed adapters overnight at 16 °C, by adding 2,000 U of T4 DNA ligase, 0.75 mM ATP and 7 nM of the adapters. The T4 ligase was heat-inactivated at 65 °C for 15 min before pooling libraries together. To remove adaptor dimers, the library pool was cleaned up using 1.8× AMPure beads and the adaptor-tagged DNA fragments were eluted to 30 μl of low Tris-EDTA buffer. The bisulfite conversion of the adaptor-tagged DNA fragments was conducted using a Qiagen EpiTect Fast Bisulfite Conversion Kit following the manufacturer's instructions with a minor modification. We extended the bisulfite conversion time from 2 cycles of 10 min to 2 cycles of 20 min to achieve bisulfite conversion rates >99%. The bisulfite-converted DNA fragments were PCR amplified according to the following thermocycler settings: 98 °C for 45 s, 6 cycles of 98 °C for 20 s, 58 °C for 30 s, 72 °C for 1 min, and then 8–10 cycles of 98 °C for 20 s, 65 °C for 30 s, 72 °C for 1 min, followed by a final extension cycle of 5 min at 72 °C. The PCR-amplified library DNA was cleaned up using 1.3× AMPure beads and the RRBS libraries were paired-end sequenced for 2 × 100 cycles. Only instances in which the matched pool of Epiblast and ExE from a given replicate both had >1 million CpGs covered at ≥5× were included for downstream analysis.

For each sample, 10 μl of M-280 streptavidin beads (Thermo Fisher) were prepared as per the manufacturer's recommendations. Specifically, after washing with Solution A (0.1 M NaOH, 0.05 M NaCl) and B (0.1 M NaCl) sequentially, the beads were resuspended in 10 μl of 2× binding and washing buffer (10 mM Tris-HCl, 1 mM EDTA, 2 M NaCl) and then mixed with an equal volume of 2 μM of reverse transcription primer[41]. The mixture was incubated for 15 min at room temperature with gentle rotation. The bead-bound reverse transcription primer was collected using a magnet and was subsequently resuspended in 10 μl of binding buffer (10 mM Tris-HCl (pH 8.0), 167 mM NaCl, 0.05% Tween-20).

Estimating methylation levels. The methylation level of each sampled cytosine was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T. Single CpG methylation levels were limited to those CpGs that had at least fivefold coverage. For 100 bp tiles, reads for all the CpGs that were covered more than fivefold within the tile were pooled and used to estimate the methylation level as described for single CpGs. The CpG density for a given single CpG is the number of CpGs 50 bp up- and downstream of that CpG. The CpG density for a 100 bp tile is the number of CpGs in the tile. The methylation level reported for a sample is the average methylation by pooling all reads across replicates.

Genomic features. LINE, LTR and SINE annotations were downloaded from the UCSC (University of California, Santa Cruz) browser (mm9) RepeatMasker tracks. CGI annotations were downloaded from the UCSC browser (mm9) CpG Islands track. Gene annotations (exon, 5′ exon, intron) were downloaded from the UCSC browser (mm9) RefSeq track. Promoters (TSSs) are defined as ±2 kb of the RefSeq annotation. Corresponding human annotations were downloaded from the UCSC browser for hg19. In each case, the methylation level of an individual feature is estimated by averaging methylation for all CpGs within the feature that are covered greater than fivefold. Assignment of CGIs to a given TSS (CGI promoters) included annotated CGIs that fell within this boundary. Methylation was estimated for 'core TSS' sequences defined as ±1 kb of the RefSeq annotation and only included CpGs measured at ≥5× in both samples (WGBS) or pooled samples (RRBS). For Fig. 2b and Extended Data Figs 3f and 5c, promoters for all isoforms are included and the maximally different alternative TSS was reported. Within the Supplementary Tables, the methylation levels of all annotated TSSs were calculated and reported in this manner, with the mean transcripts per million (TPM) estimate for the gene reported for all associated TSSs.

Identification of differentially methylated loci and regions. For WGBS data, identification of differentially methylated loci was performed using the DSS package, which uses biological replicates and information from CpG sites across the genome to stabilize the estimation of the dispersion parameters[44]. Only CpGs that were covered at least fivefold across all samples were considered for a given comparison. A false discovery rate (FDR) cut-off of 5% was used to identify differentially methylated CpGs. A CGI was called as differentially methylated if it was covered by at least 5 CpGs and 80% of them were significantly hyper/hypo methylated. For TCGA Illumina Infinium HumanMethylation450K BeadChip data, given that most cancer types have more than 20 cancer and normal samples, Wilcoxon rank-sum test was used to identify differentially methylated CpGs, with a FDR cut-off of 5%. All statistical tests throughout this study are two-sided. A CGI was called as differentially methylated if 80% of covered CpGs were significantly hyper/hypo methylated. For RRBS data, a simple cut-off of 10% difference in CGI-level methylation was used to call differential methylation.

Gene expression analysis. Alignment was performed using TopHat2 against mouse genome assembly mm9 with default settings. Isoform-level expression was quantified by kallisto, which performs pseudoalignment of reads against cDNA sequence of transcripts. Gene-level expression was estimated as the sum of expression of associated isoforms. Refseq mRNA sequences were downloaded from the UCSC genome browser. Expression levels were reported as transcripts per million (TPM).

Pathway enrichment. Pathway enrichment was performed by a hypergeometric test using the GSEA online tool. The *P* value was adjusted for multiple hypothesis testing according to Benjamini and Hochberg, with 5% as a cut-off. Regulation by PRC2 in human ES cells taken from ref. 45.

Connectivity analysis. We used GRAIL (gene relationships across implicated loci)[46] to test whether a query gene is functionally related to a set of seed genes. GRAIL uses text-mining to quantify the relatedness between two genes in the genome, by which a global gene network is built. It has been demonstrated that genes that function in the same pathway tend to distribute in a coherent subnetwork. In this study, we built a subnetwork using ExE hyper CGI-associated genes, which were significantly enriched in several pathways. To predict whether a query gene is functionally related to the ExE hyper subnetwork, we project this gene to the global network, and test whether connection of this gene to the subnetwork is random or statistically significant.

ATAC–seq data processing. Reads were aligned to mouse genome mm9 using BWA with default parameters. Duplicates were removed by the function MarkDuplicates from the Picard tool kit. Reads with low mapping quality (<10) or in the mitochondrial chromosome were removed. NucleoATAC was used to generate insert density, which was normalized by the total number of insertions in each sample[47].

Orthology mapping between human and mouse. Mouse mm9 CGIs were mapped to human hg19 segments using liftOver with chain file mm9ToHg19.over.chain. Then human orthologous CGIs were defined as the nearest CGIs to the mapped segments.

Data availability. All datasets have been deposited in the Gene Expression Omnibus and are accessible under GSE84236. Additional data include: Roadmap and ENCODE samples from RnBeads Methylome Resource (http://rnbeads.mpi-inf.mpg.de/methylomes.php), mouse adult tissues from GSE42836, and CLL and normal B lymphocytes from GSE58889.

31. Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nature* **511,** 611–615 (2014).
32. Chenoweth, J. G. & Tesar, P. J. Isolation and maintenance of mouse epiblast stem cells. *Methods Mol. Biol.* **636,** 25–44 (2010).
33. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484,** 339–344 (2012).
34. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).
35. Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345,** 943–949 (2014).
36. Yoshida, N. & Perry, A. C. Piezo-actuated mouse intracytoplasmic sperm injection (ICSI). *Nat. Protocols* **2,** 296–304 (2007).
37. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
38. Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115,** 281–292 (2003).
39. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153,** 910–918 (2013).
40. Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **44,** W272–W276 (2016).
41. Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12,** 519–522 (2015).
42. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* **9,** 171–181 (2014).
43. Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protocols* **6,** 468–481 (2011).
44. Wu, H. *et al.* Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res.* **43,** e141 (2015).
45. Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* **40,** 499–507 (2008).
46. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5,** e1000534 (2009).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

47. Schep, A. N. *et al.* Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* **25,** 1757–1770 (2015).
48. Ciruna, B. G. & Rossant, J. Expression of the T-box gene *Eomesodermin* during early mouse development. *Mech. Dev.* **81,** 199–203 (1999).
49. Ralston, A. & Rossant, J. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev. Biol.* **313,** 614–629 (2008).
50. Savory, J. G. *et al.* Cdx2 regulation of posterior development through non-Hox targets. *Development* **136,** 4099–4110 (2009).
51. Donnison, M. *et al.* Loss of the extraembryonic ectoderm in *Elf5* mutants leads to defects in embryonic patterning. *Development* **132,** 2299–2308 (2005).
52. Goldin, S. N. & Papaioannou, V. E. Paracrine action of FGF4 during periimplantation development maintains trophectoderm and primitive endoderm. *Genesis* **36,** 40–47 (2003).
53. Kang, M., Piliszek, A., Artus, J. & Hadjantonakis, A. K. FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. *Development* **140,** 267–279 (2013).
54. Nichols, J., Silva, J., Roode, M. & Smith, A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136,** 3215–3222 (2009).
55. Auclair, G., Guibert, S., Bender, A. & Weber, M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol.* **15,** 545 (2014).
56. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* **43,** 811–814 (2011).
57. Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448,** 714–717 (2007).
58. He, J. *et al.* Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat. Cell Biol.* **15,** 373–384 (2013).
59. Wu, X., Johansen, J. V. & Helin, K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol. Cell* **49,** 1134–1146 (2013).
60. Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157,** 1445–1459 (2014).
61. Boulard, M., Edwards, J. R. & Bestor, T. H. FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat. Genet.* **47,** 479–485 (2015).
62. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41,** 178–186 (2009).
63. Steine, E. J. *et al.* Genes methylated by DNA methyltransferase 3b are similar in mouse intestine and human colon cancer. *J. Clin. Invest.* **121,** 1748–1752 (2011).
64. Schulze, I. *et al.* Increased DNA methylation of Dnmt3b targets impairs leukemogenesis. *Blood* **127,** 1575–1586 (2016).
65. Yang, L. *et al.* DNMT3A loss drives enhancer hypomethylation in FLT3-ITD-associated leukemias. *Cancer Cell* **29,** 922–934 (2016); erratum **30**, 363–365, (2016).
66. Mayle, A. *et al. Dnmt3a* loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood* **125,** 629–638 (2015).
67. Haney, S. L. *et al.* Promoter hypomethylation and expression is conserved in mouse chronic lymphocytic leukemia induced by decreased or inactivated Dnmt3a. *Cell Rep.* **15,** 1190–1201 (2016).

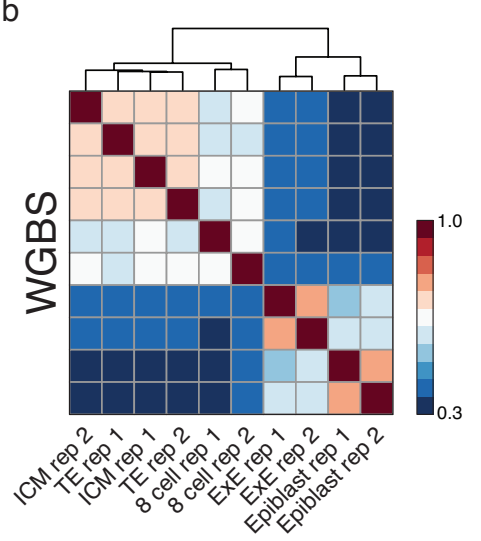
© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a

WGBS data

	Total CpGs			Conversion Rate		5x covered CpGs				100 bp tiles	
Sample	n	5x	10x	Rep 1	Rep 2	Median	Mean	Euclidean	Pearson	Euclidean	Pearson
8 cell	2	19,145,511	10,442,879	0.9915	0.9921	0.46	0.46	0.25	0.66	0.20	0.70
ICM	2	20,673,920	17,417,328	0.9948	0.9950	0.17	0.21	0.19	0.68	0.12	0.75
TE	2	20,968,642	19,366,304	0.9948	0.9948	0.14	0.18	0.17	0.69	0.10	0.76
Epiblast	2	20,865,403	20,058,090	0.9873	0.9870	0.86	0.78	0.14	0.76	0.07	0.88
ExE	2	20,786,316	19,824,067	0.9931	0.9928	0.54	0.53	0.18	0.77	0.09	0.89

b

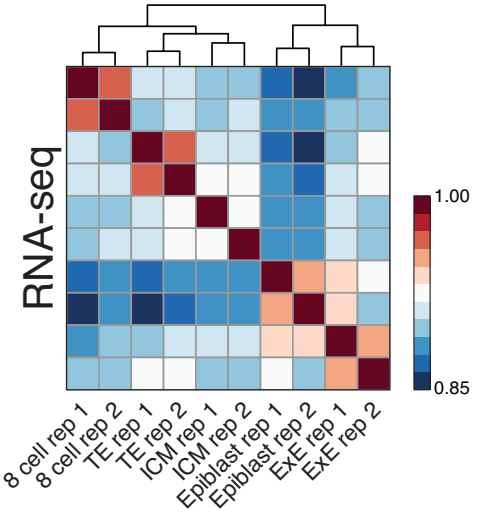


c

RNA-seq data

	Mapped reads		Mean Per Base Cov.		
Sample	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Pearson
8 Cell	108,036,120	101,901,332	108.07	97.71	0.96
ICM	96,481,760	106,439,920	91.35	104.79	0.92
TE	110,221,714	90,140,072	128.91	102.00	0.96
Epiblast	91,896,290	86,300,074	78.05	64.69	0.95
ExE	98,585,124	89,507,686	98.21	103.19	0.95

d

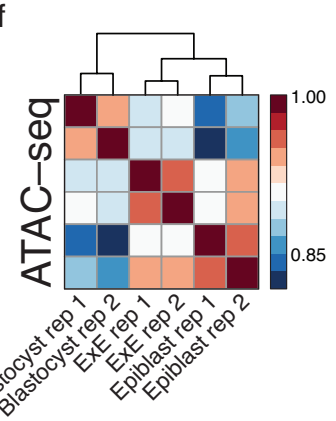


e

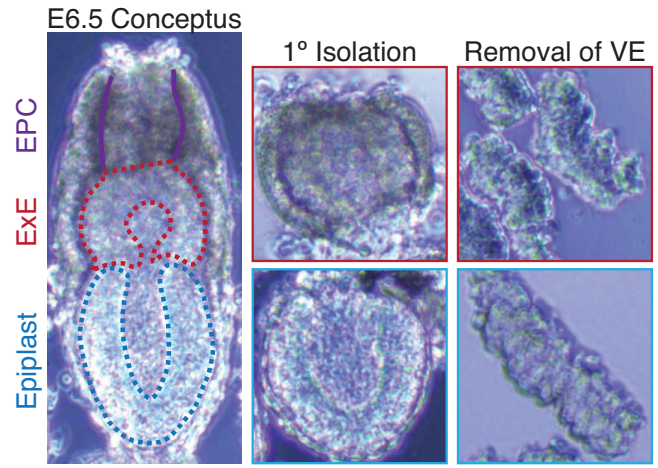
ATAC–seq data

	Mapped reads		Proper Pair (%)		
Sample	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Pearson
Blastocyst	25,941,379	15,997,038	97.94%	98.68%	0.94
Epiblast	36,489,734	112,497,843	98.16%	99.14%	0.95
ExE	35,258,834	73,997,976	98.75%	98.41%	0.96

f

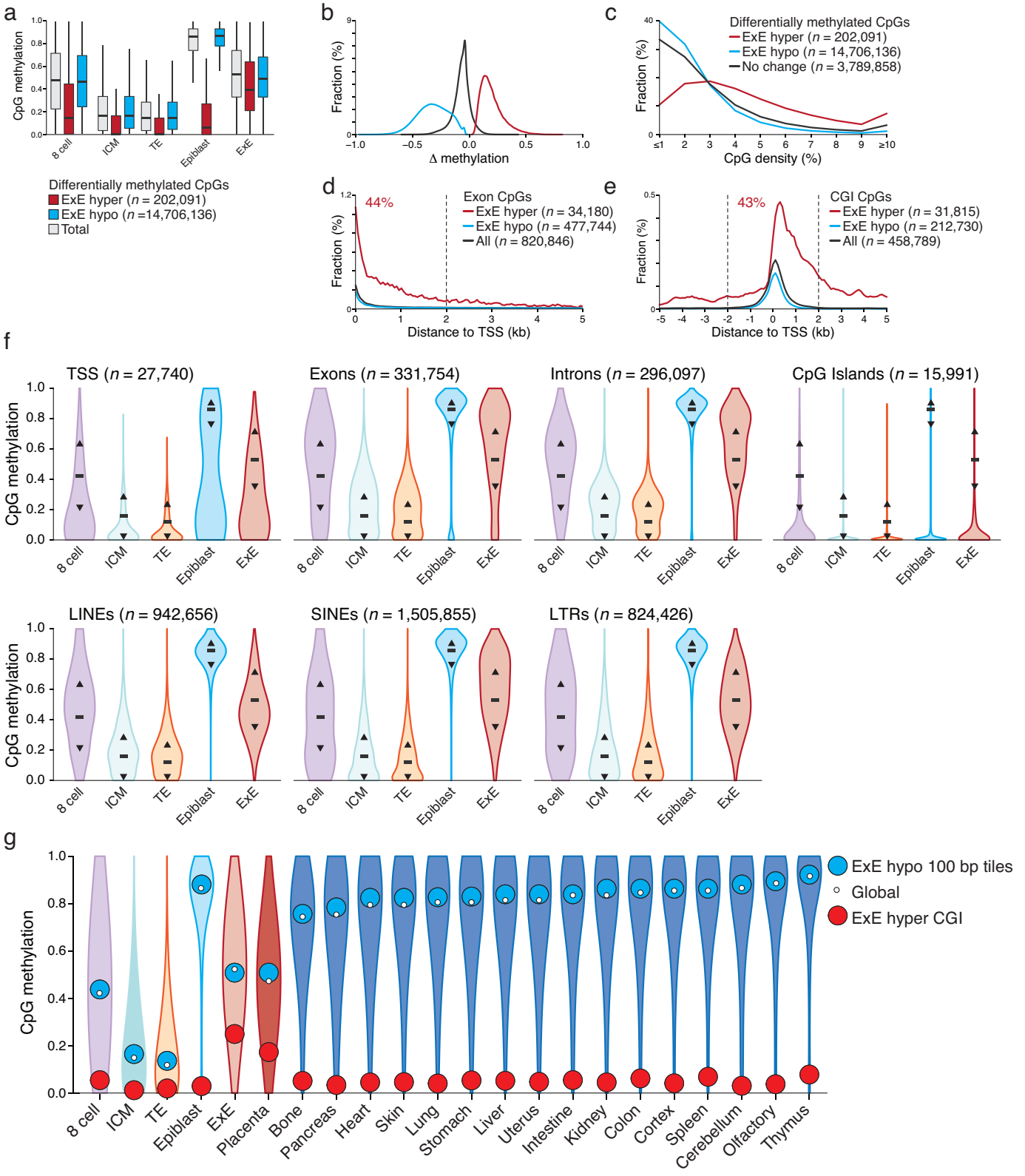


g



Extended Data Figure 1 | Tracking divergence in DNA methylation landscapes during mouse implantation. a–f, Sequencing metrics and coverage information for WGBS, RNA-seq, and ATAC–seq data including hierarchical clustering and Pearson correlation for CpGs, genes, and gene promoters, respectively. WGBS data also includes Euclidean distance, which can be beneficial for examining sample similarity in globally hypomethylated samples, as well as similarity scores for 100 bp tiles, which locally merge the intrinsically higher variance of intermediately methylated CpGs to reduce noise. For RNA-seq and ATAC–seq data, biological replicates cluster together, as do 8-cell and postimplantation WGBS data, whereas tissues of the E3.5 blastocyst cluster together but not as discrete inner cell mass (ICM) and trophectoderm (TE) compartments. In general, there is minimal variation between the methylation status of the ICM and trophectoderm, with only slight deviations around the minimal global value that is reached during this developmental period. **g**, Isolation of the epiblast and ExE from the E6.5 post-implantation embryo. The conceptus is first removed from maternal decidual tissue and portioned into epiblast and ExE fractions, taking care to remove the apical ectoplacental cone (EPC). Then, outer visceral endoderm (VE) and trophoblast cells are enzymatically digested and mechanically removed using a thin glass capillary.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

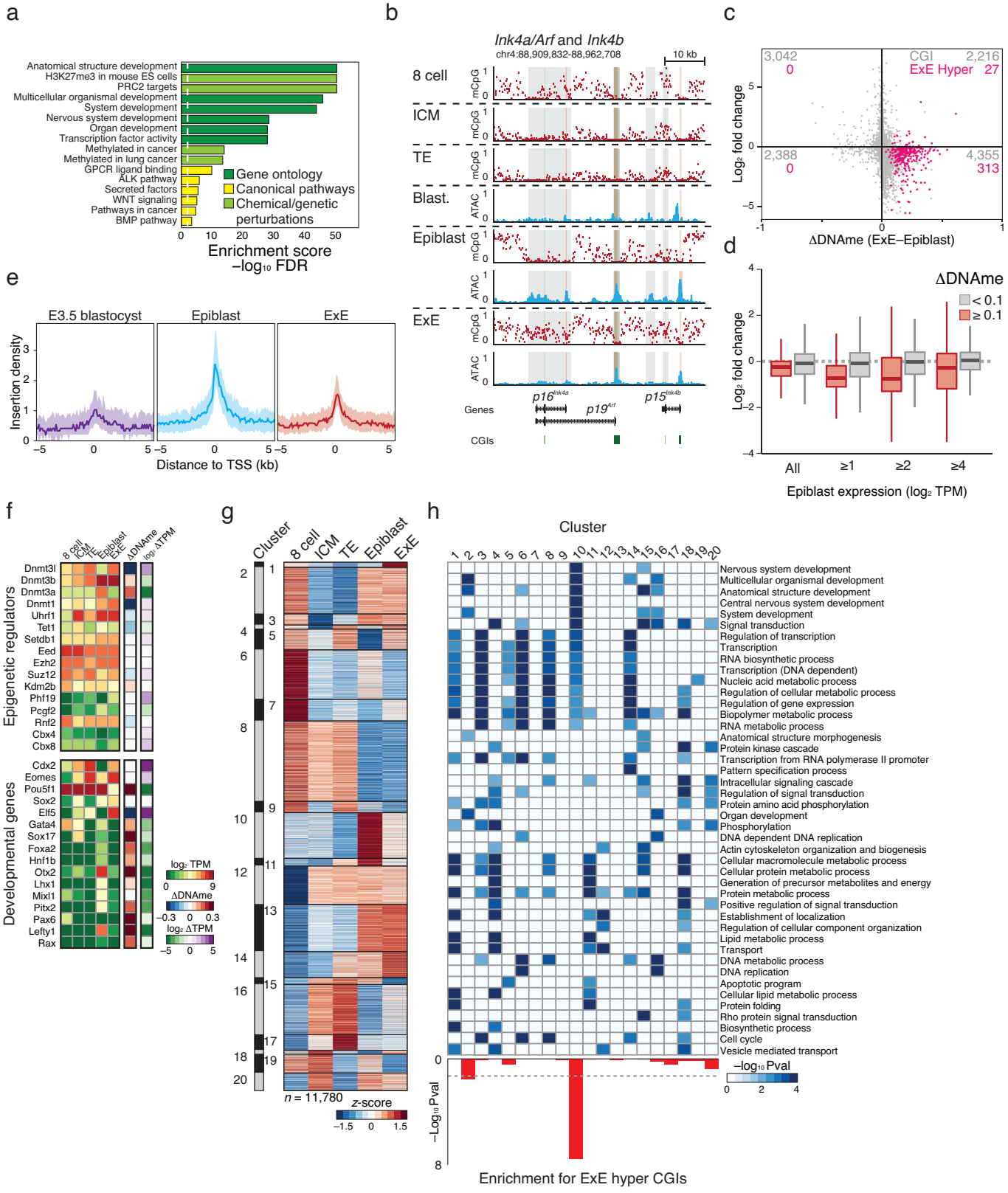


Extended Data Figure 2 | See next page for caption.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 2 | Unique features of the extraembryonic methylation landscape. a, CpG methylation boxplots for all covered CpGs as well as those that are significantly hyper- or hypomethylated within the ExE compared to epiblast (ExE hyper or ExE hypo, respectively). ExE hypo CpGs largely reflect differential remethylation compared to the epiblast across the genome. Alternatively, ExE hyper CpGs are mostly unmethylated in the ICM and trophectoderm and remain so in the epiblast, indicating an ExE-specific mechanism. Edges refer to the 25th and 75th percentiles, and whiskers the 2.5th and 97.5th percentiles, respectively. **b**, Differential methylation distribution for ExE hyper or hypo CpGs compared to epiblast. Hypomethylation appears to be a global feature of the ExE and deviates from a default hypermethylated state in the epiblast. Alternatively, increased DNA methylation appears to be directed focally and *de novo* at regions that are unmethylated within the epiblast and subsequent embryonic and adult somatic tissues. **c**, Alternate CpG density distributions for ExE hypo and hyper CpGs indicate differential enrichment within distinct genomic features. Whereas ExE hypo CpGs resemble the global average, ExE hyper CpGs occur within regions of higher CpG densities. **d**, The fraction of dynamically methylated CpGs that fall within annotated exons as a function of distance to their assigned TSS. 44% of exonal ExE hyper CpGs fall within 2 kb of their associated TSS. **e**, The fraction of dynamically methylated CpGs that fall within annotated CpG islands (CGIs) based upon their proximity to the nearest TSS. ExE hyper CpGs are generally TSS proximal and skew downstream of the TSS, with 43% falling within $\pm$2 kb. **f**, DNA methylation distribution for different genomic features including those associated with genic (TSS, exon, intron and CGI) and repetitive (LINE, SINE and LTR) sequences. For reference, black bar and arrows highlight the global median and the 25th and 75th percentiles, respectively. Globally, all features exhibit the expected passage through minimal DNA methylation values within the ICM and trophectoderm of the E3.5 blastocyst before remethylation at implantation. Compared to its global distribution, the ExE exhibits higher levels of *de novo* methylation within exons and introns, and lower than global levels within regions of LINE and LTR retrotransposon origin. The epiblast exhibits nearly complete hyper or hypomethylation depending on the genomic feature, and is bimodal at TSSs, which frequently contain CGIs. *n* values refer to the number of annotated features of a given type. **g**, Violin plots of 100 bp methylation data for early embryonic, placental, and fetal tissues demonstrate general epigenetic retention of either the somatic epiblast or extraembryonic architecture throughout subsequent development. The white dot highlights the global median, and blue and red reflect the median of ExE hypomethylated 100-bp tiles and ExE hyper CGIs, respectively. Notably, the placenta largely preserves the hypomethylated global landscape and targeted CGI methylation as they are established by E6.5 within the ExE. We show 100-bp tiles and CGIs for ExE-specific hypomethylation and hypermethylation, respectively, to restrict CpGs to a notable feature where they change as a group. WGBS data of somatic tissues and midgestation placenta taken from ref. 11.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

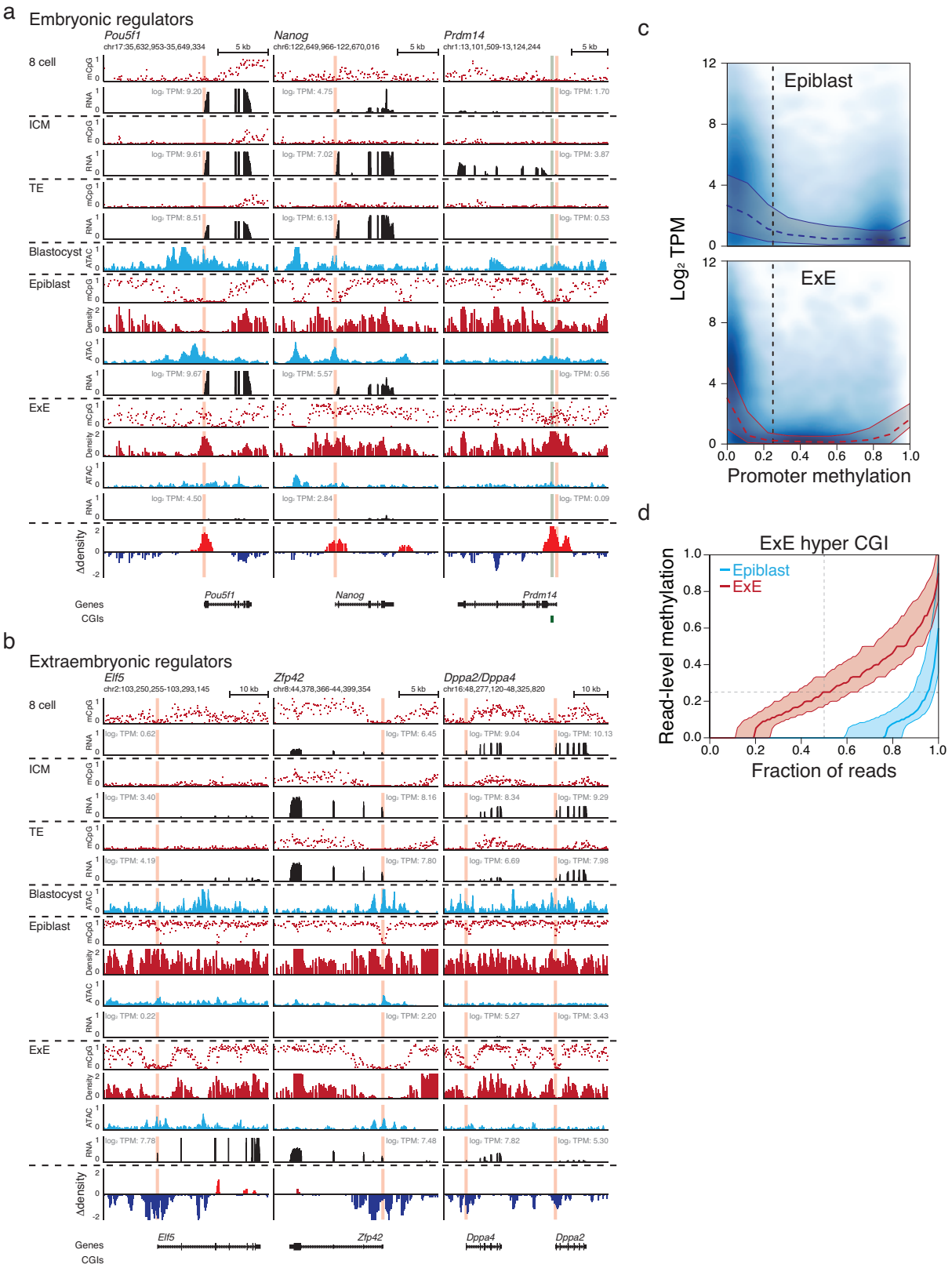


Extended Data Figure 3 | See next page for caption.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 3 | Transcriptional differences between epiblast and ExE are directed in part through DNA methylation. a, Select gene set enrichment analysis of ExE hypermethylated TSSs, including Gene Ontology, canonical pathways, and genetic and chemical perturbations, shows high enrichment for transcription factors and signalling pathways involved in patterning the early embryo. Moreover, these promoter CGI-containing genes are canonical targets of PRC2, which coordinates selective expression of key developmental regulators during gastrulation. **b**, DNA methylation and open chromatin dynamics for the tumour suppressors $p16^{Ink4a}$, $p19^{Arf}$ (both encoded by *Cdkn2a*) and $p15^{Ink4b}$ (encoded by *Cdkn2b*). Although these loci are either basally or non-transcribed during early development, three regions are dynamically methylated in the ExE (highlighted in grey), including a >10-kb region that encompasses the entirety of the $p16^{Ink4a}$ locus and is either wholly unmethylated in the epiblast or extensively methylated in the ExE. CGIs are highlighted in green, and the positions of included TSSs are highlighted in red. **c**, Scatterplot of $\log_2$ expression dynamics versus differential CGI methylation between the epiblast and the ExE. Although most dynamically methylated CGI promoter-containing genes have functions in later embryonic development and are not yet highly expressed, *de novo* methylation in the ExE is generally associated with transcriptional repression. ExE hyper CGIs are highlighted in pink. Promoter CGIs are assigned to the most proximal gene within a boundary of ±2 kb. **d**, Boxplots relating promoter methylation and expression in the restriction of extraembryonic and embryonic compartments. Promoters are defined as ±1 kb of an annotated TSS and scored as dynamically methylated in the ExE if the difference with the epiblast is ≥0.1. Expression changes between dynamically methylated and background promoter sets are provided over increasing thresholds according to their expression in the epiblast. Although many CGI promoter-containing genes are not dynamically expressed in either the epiblast or the ExE and are associated with downstream developmental functions, transcriptional repression is a consistent feature of promoter methylation, even at this low threshold. **e**, Median open chromatin signal as measured by ATAC–seq for ExE hyper CGI-associated TSSs in the transition from pre- to postimplantation. ExE hyper CGI-associated genes are heavily enriched for roles in patterning the embryo proper and are primarily not expressed until the onset of gastrulation. In the transition from blastocyst to epiblast, these promoters gain open chromatin signal, suggesting transcriptional priming or activation, which is not observed within the ExE, where they are *de novo* methylated. Shaded area reflects the 25th and 75th percentile per fixed 100-bp bin. **f**, Expression and differential promoter methylation of key epigenetic and master transcriptional regulators over early embryonic and extraembryonic development. Most epigenetic regulators exhibit minimal expression differences between the epiblast and the ExE, with the Dnmts being notable exceptions. Key isoforms of Dnmt3a and Dnmt3b are upregulated in the epiblast in conjunction with global remethylation, whereas the suppression of Dnmt3a in the ExE corresponds with *de novo* promoter methylation. Alternatively, the maintenance methyltransferase Dnmt1 and the non-catalytic cofactor Dnmt3l are induced within the blastocyst and maintained at higher levels in the ExE, with reciprocal methylation of the *Dnmt3l* promoter in the epiblast. The H3K36 demethylase Kdm2b displays differential expression of catalytically active and inactive isoforms within the epiblast and the ExE, respectively, with isoform switching seemingly imposed by *de novo* methylation around the somatically used CGI promoter. The ExE is characterized by persistent expression of the master regulators Cdx2, Eomes, and Elf5 (refs 48–51), whereas the still pluripotent epiblast remains Pou5f1 (also known as Oct4) positive. Many additional regulators of subsequent developmental stages are basally expressed within the epiblast and their promoters *de novo* methylated in the ExE. The difference in promoter methylation refers to the annotated TSS that exhibits the greatest absolute difference between ExE and epiblast. TPM, transcripts per million. Additional high-resolution genome browser tracks are displayed for select transcriptional and epigenetic regulators in Extended Data Figs 4 and 7, respectively. **g**, Unsupervised hierarchical clustering of 11,780 genes over late preimplantation and early post-implantation development, partitioned into 20 distinct dynamics ('clusters'). Cluster 10 includes genes that are specifically induced within the epiblast but not the ExE. Heat map intensity reflects the row-normalized *z*-score. **h**, Significant Gene Ontology enrichment for the 20 gene expression dynamics characterized in **f**, including those regulated by ExE-methylated CGI promoters, as calculated using the binomial test. Cluster 10 is enriched for both developmental functions and ExE promoter methylation.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

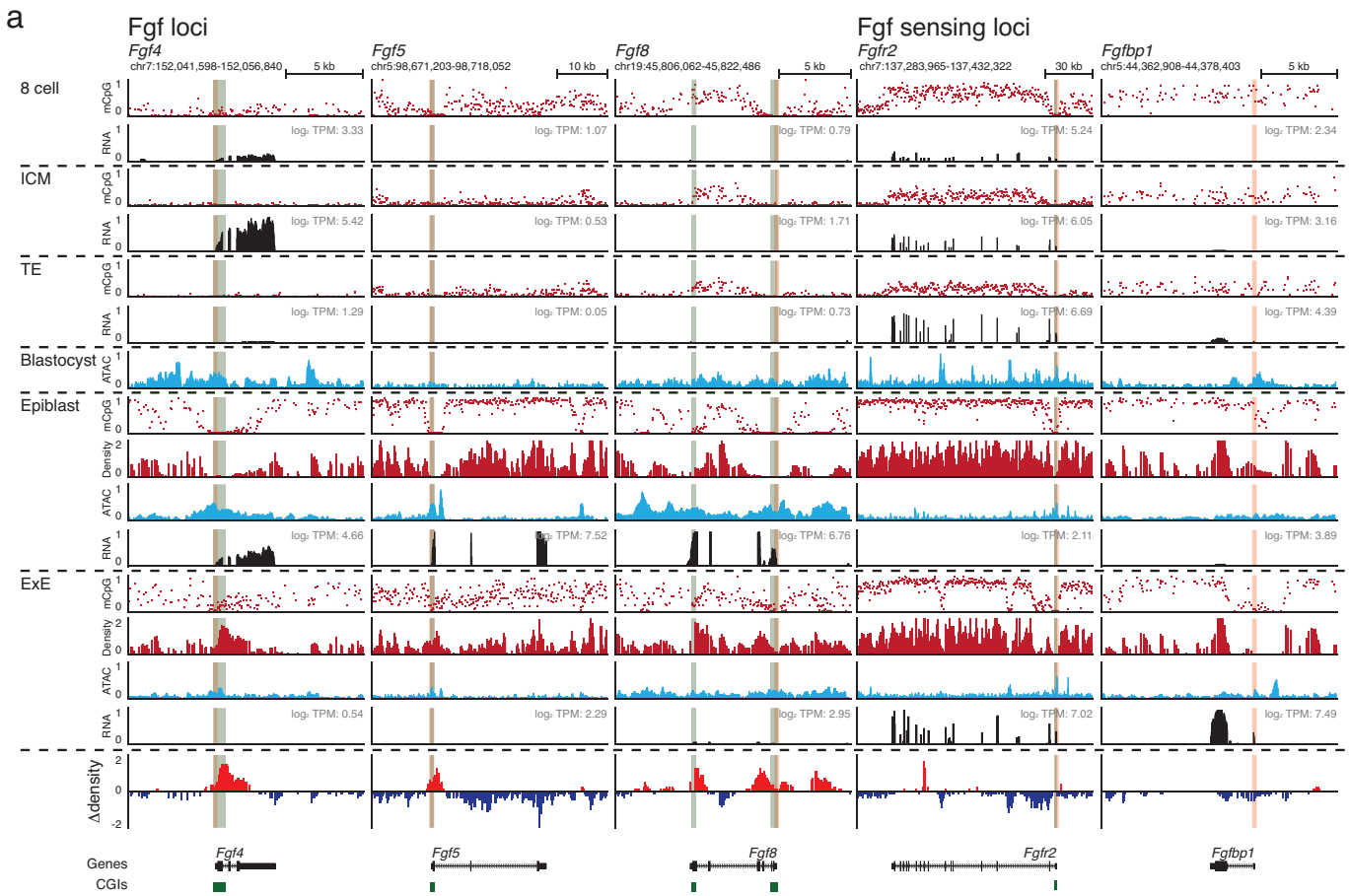


Extended Data Figure 4 | See next page for caption.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 4 | Unique bifurcation and epigenetic reinforcement of transcriptional regulators during postimplantation development. a, Genome browser tracks for WGBS, ATAC–seq and RNA-seq data for transcriptional regulators associated with embryonic or extraembryonic development. CGIs are highlighted in green, and the positions of included TSSs are highlighted in red. Embryonic regulators include Pou5f1, Nanog, and Pdrm14, which are progressively expressed over preimplantation and for which Pou5f1 and Nanog remain expressed in the epiblast. For these genes, repression in the ExE is accompanied by differential methylation of their TSSs, which is apparent as a local hypermethylation 'peak' at the *Pou5f1* locus within an ~5 kb region that is otherwise hypomethylated in the epiblast. At the *Nanog* locus, an upstream region remains hypomethylated in both tissues. Finally, *de novo* methylation of the *Prdm14* promoter is representative of ExE-specific CGI promoter methylation that occurs at hundreds of genes with downstream developmental functions. Density refers to the projected number of methylated CpGs per 100 bp of primary sequence and highlights the extensive epigenetic signal present over these regions within the ExE specifically (Δdensity refers to the difference compared to the epiblast). **b**, Extraembryonic development is in part directed by the master regulator Elf5, which is not induced until implantation and is reciprocally methylated at its TSS in the epiblast. Intriguingly, many transcriptional regulators associated with pluripotency and germline development persist within the ExE, including Zfp42 and the paralogues Dppa2 and Dppa4. As with *Elf5*, the promoters for these genes are differentially methylated in the epiblast and frequently characterized by broad kilobase-scale hypomethylation surrounding their TSSs in the ExE. **c**, Scatterplots for $\log_2$ TPM as a function of promoter methylation reveal a higher sensitivity to low methylation levels in the ExE in comparison to the epiblast. Median, 25th, and 75th percentiles for expression within 0.1 methylation bins are included for reference. The fraction of unmethylated promoters is very similar between each tissue and exhibit comparable expression values. Promoters are calculated as ±1 kb of an annotated TSS. Vertical dotted line indicates the median methylation value of ExE hyper CGIs. **d**, Read-level methylation of ExE hyper CGIs in the ExE and epiblast. The methylation status for every sequencing read within a given CGI was ranked and binned into percentiles. Plotted are the median and the 25th and 75th percentiles for these ranks across ExE hyper CGIs for both the ExE and the epiblast. In general, about 80% of reads falling within these regions are methylated in the ExE, with a median methylation value of 0.25. This value is very close to the average, unphased measurement for the CGI entirely, indicating that *de novo* methylation occurs in a high fraction of cells within the ExE and to a similar extent.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | See next page for caption.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 5 | Epigenetic restriction of FGF production and sensing to embryonic or extraembryonic compartments. **a**, Genome browser tracks for WGBS, ATAC–seq and RNA-seq data for select growth factors, receptors, and potentiators that are dynamically regulated during early post-implantation development. *Fgf* loci such as the ICM-expressed *Fgf4* and epiblast-expressed *Fgf5* and *Fgf8* are all regulated by CGI-containing promoters that are *de novo* methylated in the ExE. Alternatively, expression of FGF-sensing genes such as *Fgfr2* and the potentiating protein *Fgfbp1* becomes specific to the ExE and is characterized by broad kilobase-scale hypomethylated domains surrounding their respective TSSs in this tissue. Moreover, the asymmetric allocation of Fgfr2-expressing cells during the specification of the ICM indicates that this tissue is still sensitive to these growth factors before the epigenetic restriction that is imposed by DNA methylation during implantation[52,53]. CGIs are highlighted in green, and the positions of included TSSs are highlighted in red. Density refers to the projected number of methylated CpGs per 100 bp of primary sequence and highlights the extensive epigenetic signal present over these regions within ExE specifically (Δdensity refers to the difference compared to epiblast). **b**, Bright-field images of ICM outgrowths after two or four days under disparate growth factor or small molecule conditions. All ICMs were cultured on irradiated feeders in a basal N2/B27 media supplemented with leukaemia inhibitory factor (LIF). 2i refers to the canonical FGF-inhibited, WNT-active condition comprised of the MEK inhibitor PD0325901 and the GSK3β inhibitor CHIR99021, which functions as a WNT agonist[37]. PD refers to culture with PD0325901 alone and represents repressed FGF signalling in the absence of an additional WNT input[54]. FGF4/CHIR represents dual FGF and WNT activity by culture in recombinant FGF4 and CHIR99021 and includes notable interior and exterior tissue structures that emerged during culture and were independently isolated and profiled. Finally, ICMs were cultured in FGF4 alone. Outlines highlight the specific components of each outgrowth that were subsequently purified for analysis by dual RNA-seq and RRBS profiling (see Methods). Scale bar shown on the bottom right. **c**, Differential methylation of CGIs during *in vitro* culture differs from the ExE according to developmental trajectory. Shown are specific TSS-associated CGIs that are either methylated in the ExE and both conditions, ExE and FGF/CHIR, or ExE-only and the corresponding mean adjusted $\log_2$ fold change in gene expression. Shared targets include early developmental genes, such as *Prdm14*, that are repressed in each case, though often highly expressed in the FGF/CHIR interior. Notably, some of these genes, particularly those associated with the germline, can be *de novo* methylated later in embryonic development[55]. FGF differs from the ExE and FGF/CHIR conditions in the methylation of CGIs associated with either the epiblast or the neuroectoderm, including genes that are expressed in the FGF condition, such as *Otx2*, *Igfbp2*, and *Sfrp2*, though this set encompasses other neuroectodermal master regulators such as *Pax6* that are not yet expressed. Finally, ExE and FGF/CHIR diverge in the promoter methylation of endodermal master regulators, such as *Foxa2*, *Hnf1b*, *Gata4*, and *Sox17*, which are highly expressed in the transition from FGF/CHIR inside to outside. Notably, the bifurcation in CGI methylation corresponds to the expression of *Fgfr2* and repression of *Fgf4*, as is observed *in vivo*: Fgf4 is highly expressed within the interior and repressed in the exterior (32.0 to 3.5 TPM) while Fgfr2 is induced (2.3 to 13.5 TPM). PD and FGF/CHIR conditions are also uniquely positive for Dnmt3b and 3l expression, but ExE hyper CGI methylation is not observed with PD0325901 present (TPM = 30.2 and 60.9 for Dnmt3b and Dnmt3l in FGF/CHIR outside, and 61.0 and 41.3 for PD), indicating either the requirement for an additional cofactor or post-translational modification to redirect these enzymes to this feature set.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a RRBS data

	Sample	n	Total CpGs		100 bp tiles			
			5x	10x	Median	Mean	Pearson	Euclidean
Outgrowth	2i	4	2,543,848	2,277,000	0.11	0.14	0.74	0.12
Outgrowth	PD	6	2,708,588	2,448,519	0.50	0.43	0.88	0.18
Outgrowth	FGF	5	2,666,145	2,416,074	0.90	0.72	0.97	0.09
Outgrowth	FGF/CHIR inside	7	2,668,463	2,424,931	0.44	0.39	0.88	0.19
Outgrowth	FGF/CHIR outside	7	2,734,953	2,501,819	0.31	0.31	0.85	0.14
Epiblast	Wild-type	4	2,479,938	2,241,546	0.79	0.62	0.96	0.10
Epiblast	Dnmt1	5	2,585,520	2,341,277	0.43	0.38	0.81	0.17
Epiblast	Dnmt3a	5	2,774,163	2,542,382	0.77	0.60	0.97	0.11
Epiblast	Dnmt3b	6	2,765,347	2,539,808	0.61	0.51	0.94	0.15
Epiblast	Dnmt3l	5	2,490,838	2,242,846	0.73	0.58	0.86	0.22
Epiblast	Eed	4	2,506,687	2,246,193	0.80	0.62	0.92	0.16
Epiblast	Kdm2b	3	2,641,642	2,365,433	0.77	0.61	0.90	0.17
ExE	Wild-type	4	2,520,722	2,279,820	0.46	0.45	0.91	0.13
ExE	Dnmt1	5	2,520,255	2,274,229	0.16	0.19	0.62	0.16
ExE	Dnmt3a	4	2,602,481	2,356,566	0.43	0.42	0.85	0.17
ExE	Dnmt3b	6	2,543,780	2,302,780	0.20	0.25	0.82	0.14
ExE	Dnmt3l	4	2,745,180	2,479,070	0.33	0.36	0.84	0.17
ExE	Eed	4	2,460,919	2,178,707	0.50	0.45	0.88	0.16
ExE	Kdm2b	3	2,685,969	2,435,038	0.50	0.46	0.86	0.18

b RNA-seq data

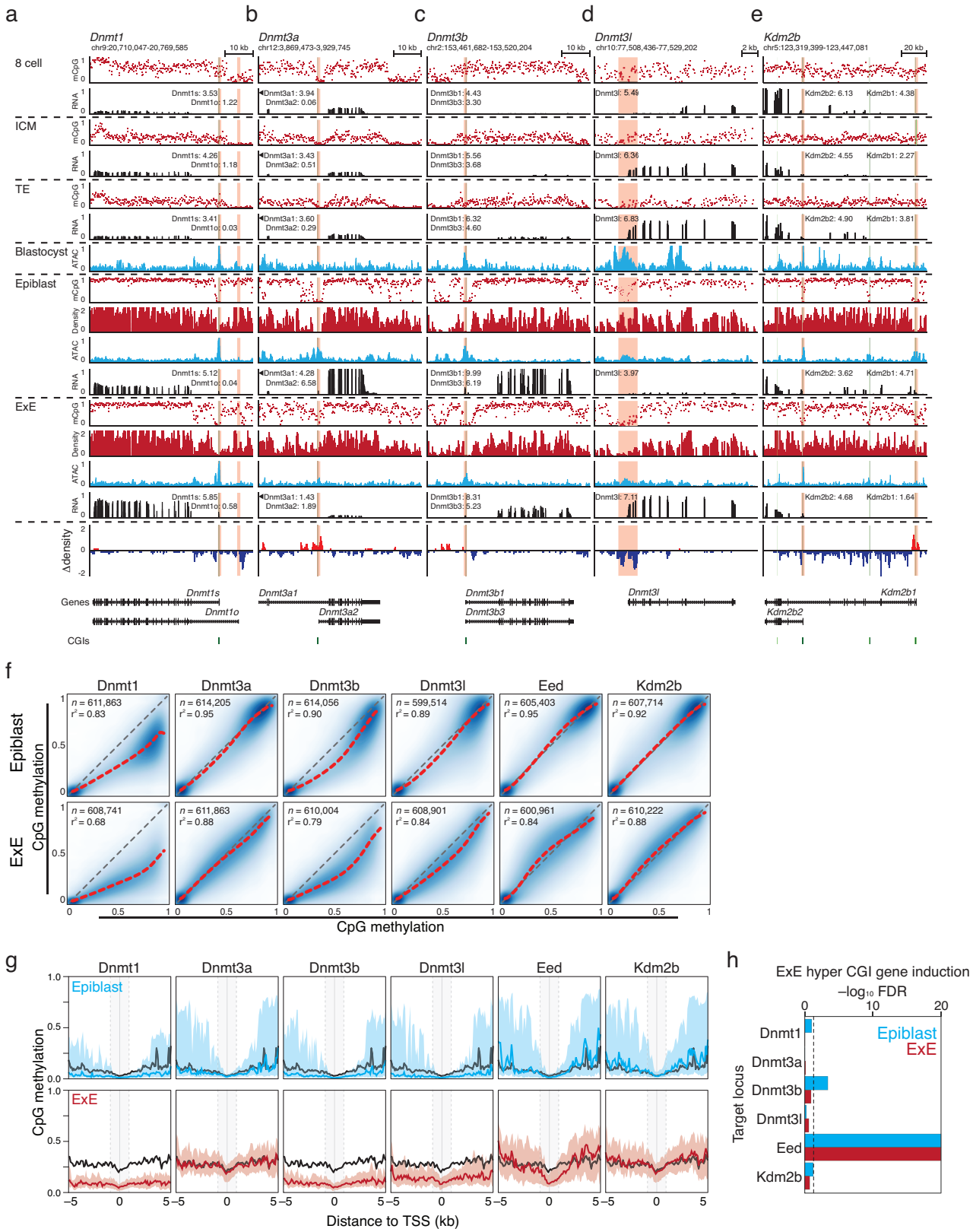
	Sample	n	Median of biological replicates			
			Mapped	Exonic rate	Mean per base cov.	Pearson
Outgrowth	2i	4	16,173,599	0.93	11.29	0.98
Outgrowth	PD	6	17,195,224	0.91	11.38	0.96
Outgrowth	FGF	5	19,735,502	0.92	13.40	0.98
Outgrowth	FGF/CHIR inside	3	25,130,944	0.91	17.51	0.99
Outgrowth	FGF/CHIR outside	3	45,754,712	0.92	33.50	0.99
Epiblast	Wild-type	4	13,611,109	0.92	9.25	0.95
Epiblast	Dnmt1	5	16,323,776	0.90	10.93	0.94
Epiblast	Dnmt3a	5	16,092,316	0.91	10.64	0.97
Epiblast	Dnmt3b	6	17,284,693	0.91	11.62	0.96
Epiblast	Dnmt3l	5	19,409,738	0.91	13.25	0.97
Epiblast	Eed	4	28,232,447	0.90	18.23	0.96
Epiblast	Kdm2b	3	26,003,962	0.93	18.14	0.98
ExE	Wild-type	4	11,537,966	0.93	8.25	0.98
ExE	Dnmt1	5	16,171,034	0.93	11.42	0.95
ExE	Dnmt3a	4	18,319,262	0.93	13.10	0.96
ExE	Dnmt3b	6	20,682,531	0.92	15.17	0.96
ExE	Dnmt3l	4	20,708,730	0.92	14.81	0.95
ExE	Eed	4	25,794,839	0.92	17.24	0.97
ExE	Kdm2b	3	26,295,000	0.93	18.60	0.91

c sgRNA sequences

Target	sgRNA	Coordinates (mm10)	Sequence
Dnmt1	1	chr9:20936581-20936600:–	GAAACTGGAAGAGGTAACAG
Dnmt1	2	chr9:20929059-20929078:–	GACTCCCGAGGACAGAGACG
Dnmt1	3	chr9:20926483-20926502:+	AGTATCAAACCAGGTCGAGG
Dnmt3a	1	chr12:3895692-3895711:+	GCCAGTAGGAGGGGATGCTG
Dnmt3a	2	chr12:3896005-3896024:+	GGCATTGGAGAGCTGGTGTG
Dnmt3a	3	chr12:3896628-3896647:–	GCTTGTTGTAGGTGGCCTGG
Dnmt3b	1	chr2:153661463-153661482:+	AGAGGGTGCCAGCGGGTATG
Dnmt3b	2	chr2:153665300-153665319:+	GCCTCCCCCAGAATCACCCG
Dnmt3b	3	chr2:153667519-153667538:+	GGAATAGGTGACCTCGTGTG
Dnmt3l	1	chr10:78050661-78050680:–	GCCTGTCGGAGGCGAGGAGG
Dnmt3l	2	chr10:78052046-78052065:+	TGCACCATCTGCTGTTCCGG
Dnmt3l	3	chr10:78051869-78051888:+	CGGCACCCCTTGTTTGAGGG
Eed	1	chr7:89,977,027-89,977,046:–	GATGCTGTCAGTATTGAGAG
Eed	2	chr7:89,970,355-89,970,374:+	GATCCAGCAACTGCTAATAG
Eed	3	chr7:89,969,589-89,969,608:+	GAAGGTTTGGGTCTCGTGGG
Kdm2b	1	chr5:122888545-122888564:–	TGAAGCAGAGCTGCATCATG
Kdm2b	2	chr5:122888618-122888637:–	GCCTGCCTGCGGACGGAGTG
Kdm2b	3	chr5:122888692-122888711:+	GGCAGCCAACTTCACCGCTG

Extended Data Figure 6 | Generation of dual expression and methylation libraries from outgrowth and embryonic knockout data. **a**, **b**, Sequencing metrics and coverage information for dual RNA-seq and RRBS libraries generated for the evaluation of ICM outgrowths and CRISPR–Cas9 disrupted E6.5 embryos, including similarity metrics between replicates (Euclidean distance and Pearson correlation for RRBS and Pearson correlation for RNA-seq). Mean and median methylation of 100 bp tiles is also included for the RRBS samples. **c**, CRISPR–Cas9 disrupted embryos were generated by zygotic injection of three single guide RNA (sgRNA) sequences specific to early exons that are shared across different isoforms. The genomic coordinates and protospacer sequences are provided (see Methods).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

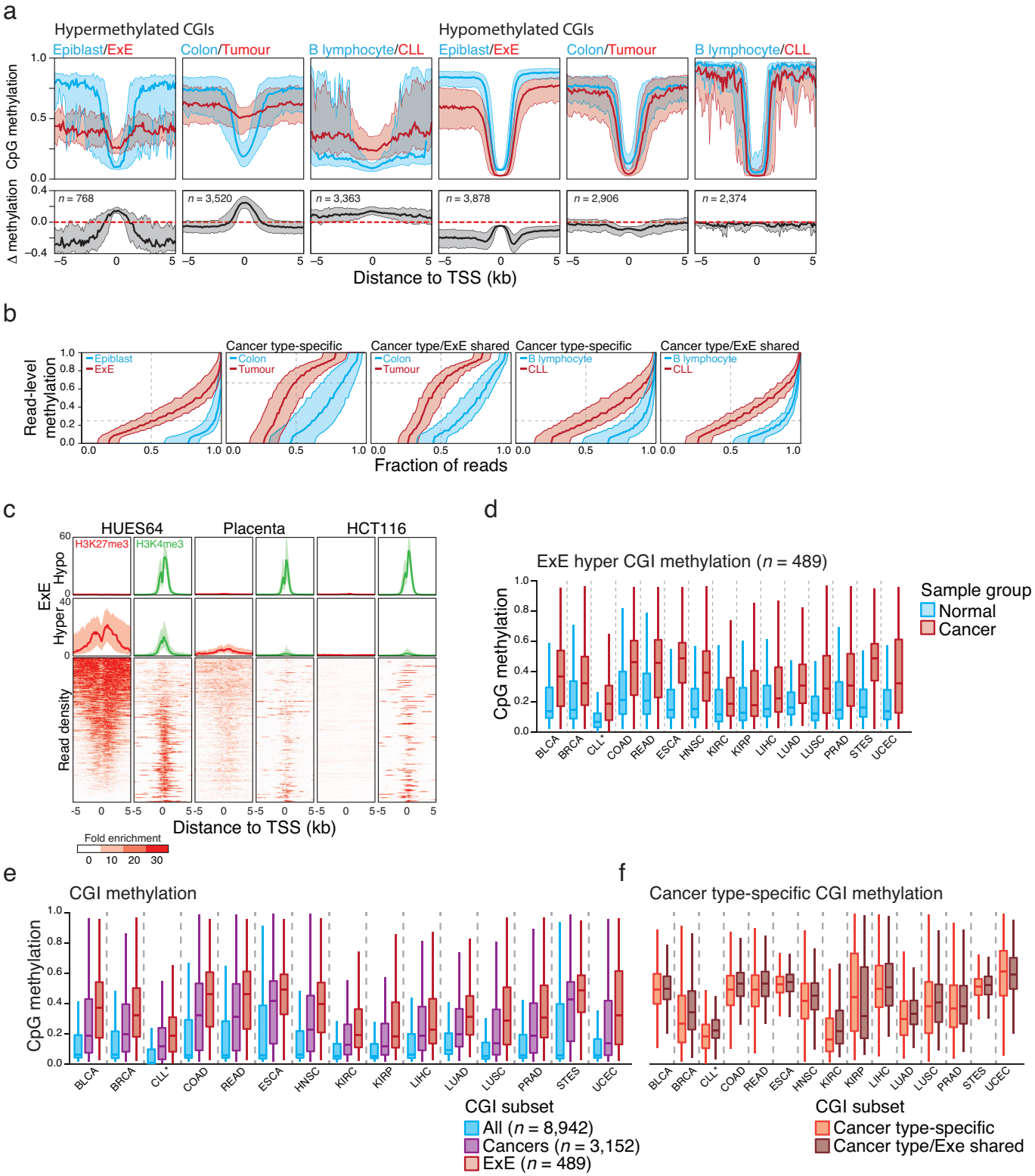


Extended Data Figure 7 | See next page for caption.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 7 | Dynamic behaviour of key epigenetic regulators during early implantation. Genome browser tracks for WGBS, ATAC–seq and RNA-seq data (log_2 TPM shown for selected isoforms). CGIs are highlighted in green, and the positions of included TSSs are highlighted in red. **a**, Dnmt1 is not appreciably expressed in early cleavage, in part owing to a transient maternal imprint over the somatically used TSS (*Dnmt1s*)[33,56], but shows moderate induction within the ICM. Then, at implantation, it is induced within both the epiblast and the ExE. Dnmt1 is expressed at higher levels within the ExE and displays persistent focal hypomethylation around the maternal-specific TSS (*Dnmt1o*) that is not observed in the epiblast, which resolves an area of preimplantation-specific hypomethylation to the hypermethylated genomic average. **b**, The short Dnmt3a2 isoform is induced to high levels in epiblast and is also expressed within embryonic stem (ES) cells. Alternatively, the CGI-containing promoter of *Dnmt3a2* is methylated in the ExE and its transcription is suppressed. **c**, Like *Dnmt1*, the *Dnmt3b* promoter contains a CGI that is maternally imprinted during preimplantation[33,56]. Induction is apparent within the blastocyst, but becomes asymmetrically abundant within the epiblast following implantation. **d**, DNMT3L is a non-catalytic cofactor that enhances the *de novo* activity of DNMT3A and B, with specific functions in the early embryo and germline[57]. During implantation, Dnmt3l is initially expressed in both the ICM and the trophectoderm, but it remains expressed in the ExE and is silenced by *de novo* promoter methylation in the epiblast. **e**, The H3K36 demethylase KDM2B has specific roles in establishing the boundary between promoters and actively transcribed gene bodies, as well as in PRC2 recruitment and the establishment of facultative heterochromatin[58–61]. A catalytically inactive isoform, *Kdm2b2*, initiates from an alternate TSS downstream of exons encoding the demethylating Jumonji domain of the catalytically active *Kdm2b1* (ref. 17). Kdm2b2 is the most prevalent isoform during preimplantation development and remains expressed in the ExE. Alternatively, Kdm2b1 is only induced during implantation within the Epiblast, whereas its CGI-containing promoter gains methylation in the ExE. Like *Dnmt1s* and *Dnmt3b*, the CGI promoter of *Kdm2b1* is a maternally methylated imprint that resolves to hypomethylation during implantation[33,56]. **f**, Extraembryonic genome remethylation is highly dependent on DNMT3B and DNMT1. Pairwise comparisons of 100-bp tiles as measured by RRBS for wild-type epiblast and ExE (*y* axis) versus matched CRISPR–Cas9-disrupted tissues (*x* axis). Extraembryonic methylation levels diminish genome-wide when *Dnmt1*, *Dnmt3b* and *Dnmt3l* are disrupted. The epiblast is only sensitive to *Dnmt1* and *Dnmt3b* disruption, both to a lesser extent than the ExE, presumably because of compensation from DNMT3A. Notably, the decrease in global methylation levels when *Dnmt1* is deleted is greater for ExE than epiblast, indicating a higher dependence on maintenance and less efficient *de novo* methyltransferase activity in this tissue. The identity line is included in grey and the best fit by LOESS regression in red. The number of 100 bp tiles used in each comparison and the r^2 values are included in the upper left of each plot. **g**, Composite plots of ExE hyper CGI-containing promoters in CRISPR–Cas9 targeted epiblast and ExE, respectively. In general, only limited effects are observed in the epiblast other than a slight increase in the peripheral methylation within the *Eed*-null sample. Alternatively, both TSS proximal and peripheral methylation is decreased in *Dnmt1*-, *Dnmt3b*-, and *Dnmt3l*-null ExE. The *Eed*-null ExE is unique in its specificity for diminished methylation at the TSS, particularly downstream within the first kilobase. In both the epiblast and the ExE, the wild-type median is included in black for comparison. Line represents the median and the shaded area the 25th and 75th percentiles, respectively. For RRBS data, composite plots are of the median for 200-bp windows, taken at intervals of 50 bp. **h**, Statistical test for the derepression of ExE hyper CGI associated genes demonstrates a comparable requirement for Eed in both the epiblast and the ExE. Gene expression of knockout samples were compared to matched wild-type samples using DESeq2 with raw counts as input. Enrichment for ExE hyper CGI associated genes were evaluated by Wilcoxon rank-sum test and represented as *z*-scores, which were converted to *P* values assuming a normal distribution. Bonferroni correction for multiple testing was applied to derive the FDR.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

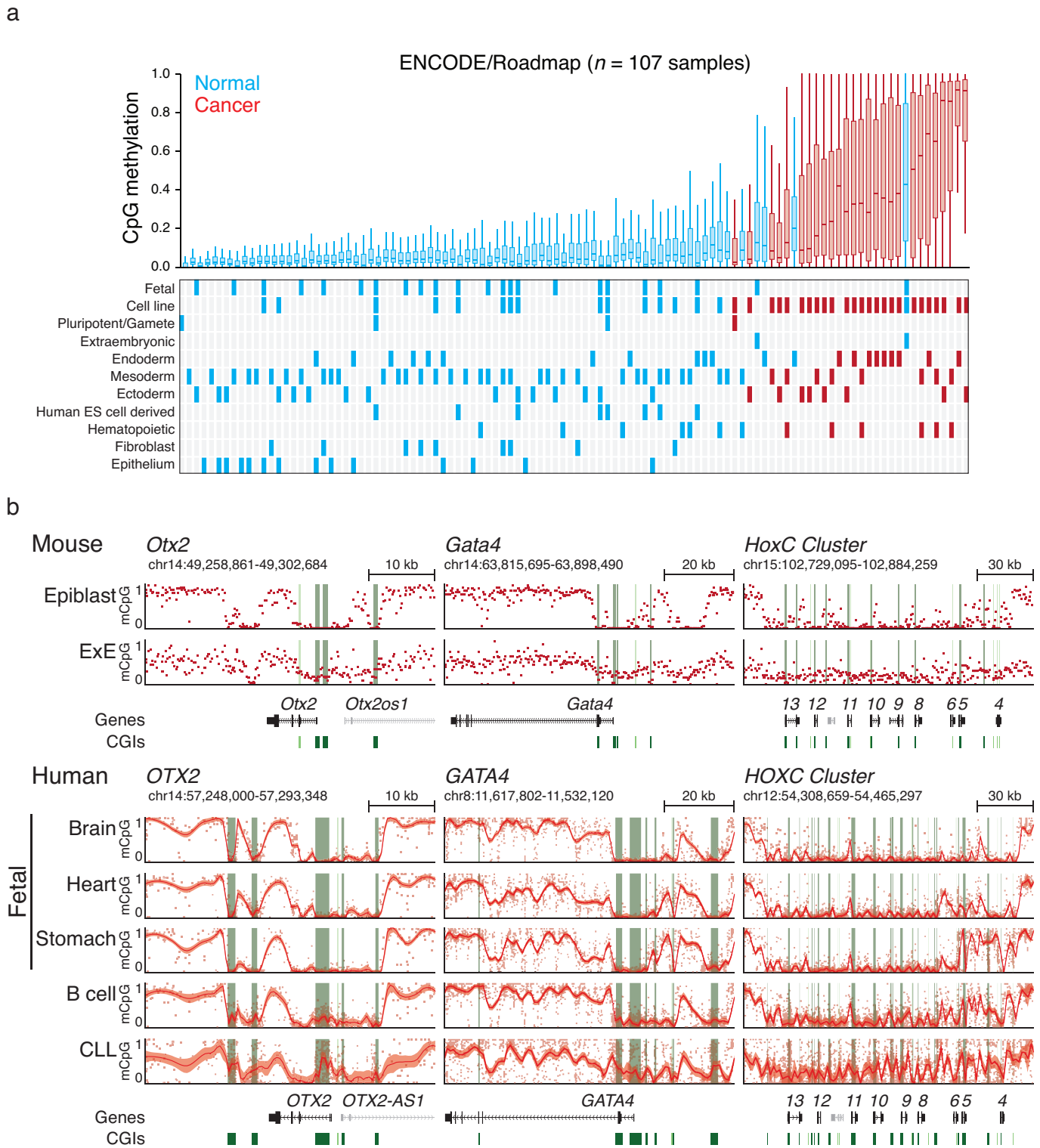


Extended Data Figure 8 | See next page for caption.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 8 | General features of the cancer methylome and of CGI DMRs. a, Median methylation of differentially regulated CGI-containing promoters in a primary colon tumour isolate and CLL compared to colon and B lymphocytes, respectively, as measured by WGBS. ExE hyper CGIs as identified in this study and shown in Fig. 1 are included for reference. The median methylation difference between extraembryonic or cancerous tissue compared to the epiblast or normal tissue is also included. The general features of both cancer methylomes are similar to those of the ExE, with a maximal increase in DNA methylation centred at the TSS that steadily diminishes within the periphery. Alternatively, hypomethylated CGIs in extraembryonic or tumorigenic contexts are maximally different a distance away from the TSS, within the boundary or 'CpG island shore', as previously reported for cancer[62]. Shaded area represents the 25th and 75th percentiles per 100-bp bin. **b**, Read-level methylation of hypermethylated CGIs in the ExE versus the epiblast, colon tumour versus colon, and CLL versus B lymphocyte, with those that share differential methylation status between the cancer and extraembryonic development included as a subset. The methylation status for every sequencing read within a given hypermethylated CGI was ranked and binned into percentiles. Plotted are the median and 25th and 75th percentiles for these ranks across CGIs called as hypermethylated in each pairwise comparison. The ExE/epiblast and CLL/B lymphocyte comparisons exhibit very similar distributions that indicate general discordance, meaning similar aggregate methylation across the feature as is observed in phase, which is most likely to be obtained by dispersive *de novo* methylation across the majority of alleles within the population. Colon tumour exhibits substantially higher read-level methylation, with a median of ~0.7. However, the per-read methylation level of the non-tumorous, matched colon tissue is also quite high, with >50% of reads exhibiting some methylation. This could indicate a transition in the epigenetic status of these loci within colon tissue that precedes tumorigenesis, as has been noted for several other tissues in Extended Data Fig. 9. The read-level methylation distribution is the same for cancer type-specific CGIs regardless of whether or not they are also ExE hyper CGIs. As such, the targeting to ExE hyper CGIs is a conserved feature of human cancer types, but the extent to which they are methylated can be specific to the system. **c**, Data taken from ENCODE samples that reflect embryonic and extraembryonic identities in human in comparison to the well-characterized human cancer cell line HCT116. The human ES cell line HUES64, a proxy for the pluripotent epiblast, displays notable enrichment for both repressive, PRC2-deposited H3K27me3 and activating H3K4me3 modifications at orthologous ExE hyper CGIs. Alternatively, human placenta exhibits diminished enrichment for both modifications at these regions, as does HCT116. Both systems display substantial methylation over ExE hyper CGIs as presented in Fig. 4, Extended Data Fig. 9 and Supplementary Table 7. As a control, 'ExE hypo' CGIs demonstrate uniformly high H3K4me3 levels. Enrichment density heat maps are provided for the full ExE hyper CGI set and are ranked across plots according to their enrichment for H3K27me3 in HUES64. Normalized enrichment represents the fold chromatin immunoprecipitation-enrichment against sample matched whole cell extract (WCE). **d**, Boxplots of mean methylation for 489 ExE-methylated, orthologous CGIs (ExE hyper CGIs) across the 14 tissue-matched TCGA cancer types that display disregulated DNA methylation landscapes and for CLL. Asterisk: CLL samples were measured by RRBS ($n = 119$) and represent a comparison between age-matched healthy B lymphocytes ($n = 24$). Edges refer to the 25th and 75th percentiles, whiskers the 2.5th and 97.5th percentiles, respectively. **e**, Boxplots for TCGA datasets and CLL for the absolute methylation values of all orthologously mapped CGIs, those methylated across cancer types, and those that are specifically methylated in mouse ExE. In all 15 cancer types that exhibit general global hypomethylation and CGI methylation as part of their departure from somatic cells, ExE hyper CGIs are specifically enriched, more so than for CGIs that are observed as hypermethylated in any given cancer type. **f**, Boxplots for the same data for cancer type-specific CGI DMRs and those that are also methylated in mouse ExE. Notably, the extent to which mouse ExE hyper CGIs are methylated reflects the cancer type, with some exhibiting higher absolute methylation values than others. However, in 14 out of 15 cases, the absolute methylation status of cancer type-specific CGI DMRs and those that are also methylated in the ExE are nearly identical, and often slightly greater. Absolute methylation values therefore appear to be determined by the specific cancer or cancer type, whereas targeting of extraembryonically methylated CGIs is a general feature.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

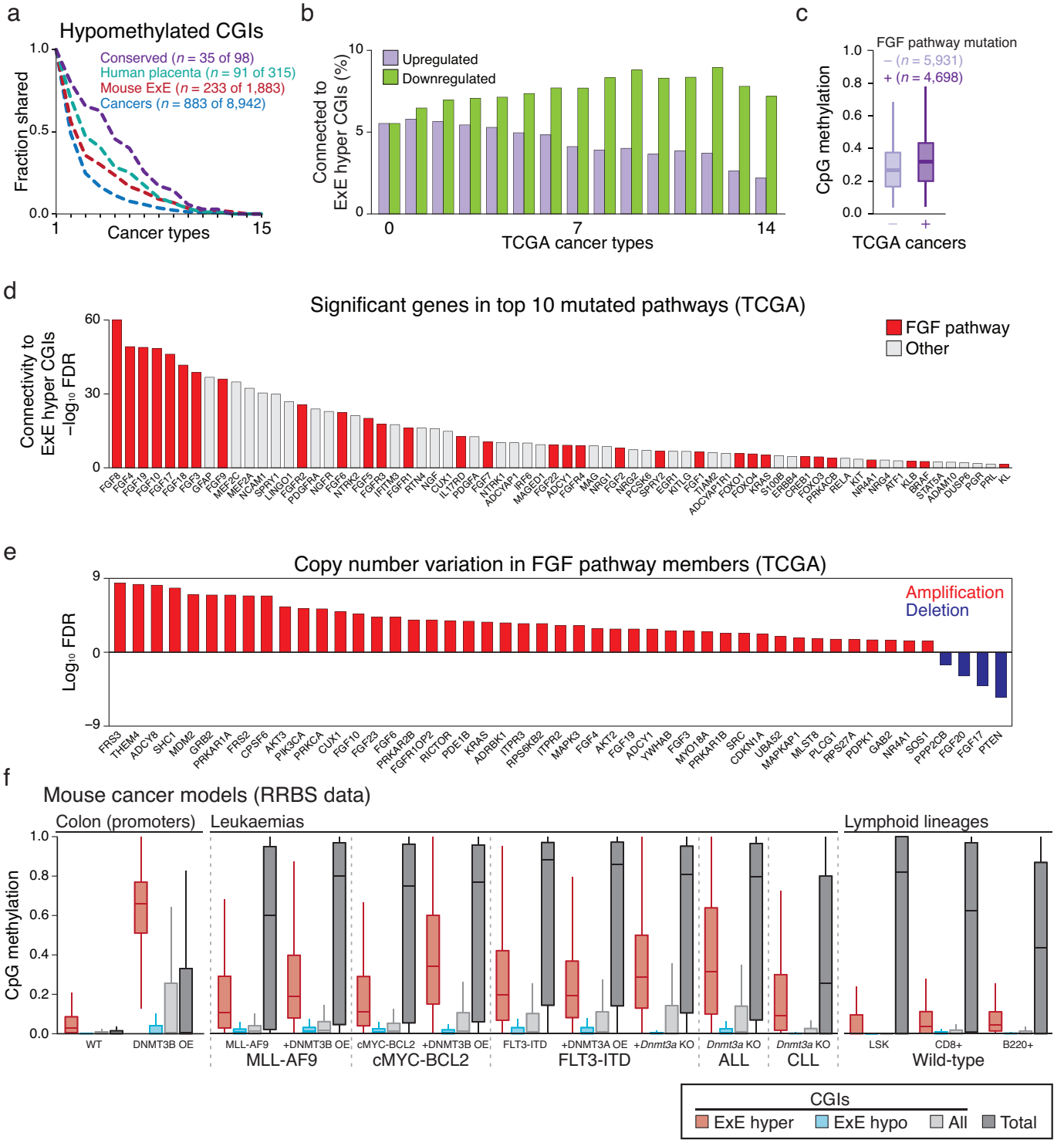


Extended Data Figure 9 | See next page for caption.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 9 | Broad conservation of extraembryonic methylation patterns across cancer types and cell lines. a, Boxplots of orthologous ExE hyper CGIs across 107 ENCODE/Roadmap Epigenomics Project samples as presented in Fig. 4, with notable additional features of each sample highlighted below. Human extraembryonic tissues, including a trophoblastic cell line and primary placenta, also share conserved CGI methylation with mouse. Normal tissues that appear to exhibit higher mean methylation of ExE hyper CGIs include numerous endodermal lineages, such as colonic mucosa, stomach and liver (mean methylation values of 0.225, 0.185 and 0.179, respectively) as well as mature cell types of the adaptive immune system, such as $CD8^+$ and $CD4^+$ T lymphocytes and B lymphocytes (mean methylation values of 0.199, 0.173 and 0.173, respectively). By contrast, ectodermal and epithelial cells are comparatively less methylated than other somatic tissues, although cancer cell lines and primary tumours derived from these tissues remain sensitive to hypermethylation. **b**, Genome browser tracks for orthologous loci as originally presented for mouse development in Fig. 1 for three human fetal tissues that represent each germ-layer (brain, ectoderm; heart, mesoderm; stomach, endoderm), primary human B lymphocytes, and a CLL sample. CGIs around these loci are preserved in a hypomethylated state during embryonic development, where the bimodal architecture of the DNA methylation landscape is clearly maintained. In B lymphocytes, some low-level, encroaching methylation is already apparent over developmentally hypomethylated regions, as is also observed in the Roadmap sample in **a**. However, in the transition to CLL, extensive methylation is observed across these CGIs although methylation values drop in the surrounding areas. Red line and shaded area reflect the local mean and standard deviation as calculated by local regression (LOESS) to compensate for the greater number of CpGs within the human orthologues versus mouse, which can complicate visual estimates of local methylation at these scales. CGIs are highlighted in green.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | See next page for caption.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 10 | Genetic features of ExE CGI methylation in cancers. a, Intersection analysis as presented in Fig. 4d for cancer-hypomethylated CGIs across the 14 TCGA cancer types and CLL that exhibit global loss of methylation in tandem with CGI hypermethylation. Generally, CGI hypomethylation is more specific, such that the intersection across cancers decays exponentially. Notably, even for hypomethylated CGIs, the intersection across cancer types remains higher for those that are also hypomethylated in mouse ExE, human placenta, or both (Conserved). **b**, Intersection analysis for cancer-dysregulated genes across TCGA cancer types. Of genes significantly dysregulated in at least *n* (0–14) TCGA cancer types, the fraction of genes that are functionally related to ExE hyper CGI-associated genes were predicted by GRAIL, using a global gene network built by text-mining (see Methods). An FDR of 5% was used as a cut-off. As the number of TCGA cancer types increases, the fraction of ExE hyper CGI-associated genes within the downregulated set generally increases, whereas those that are upregulated decrease substantially. **c**, Boxplots of the average methylation for the 489 orthologous ExE hyper CGIs across the 10,629 cancers available in TCGA with matched mutational and methylation data, segregated by mutational status of genes that function as part of the FGF signalling pathway. In aggregate, cancers with FGF pathway mutations have a median average ExE hyper CGI methylation level of 0.328 compared to 0.275 for those that do not ($P < 10^{-16}$, rank-sum test). Edges refer to the 25th and 75th percentiles, whiskers the 2.5th and 97.5th percentiles, respectively. **d**, Among 539 genes that are present in the top 10 recurrently mutated pathways in cancer, 68 are functionally related to ExE hyper CGI-associated genes (FDR < 5%), as predicted by GRAIL. Genes in the FGF signalling pathway are highlighted in red. In general, FGF signalling pathway genes have high connectivity scores to ExE hyper CGI-associated genes (enrichment *z*-score = 3.88 for FGF pathway members within the *P* value distribution for all 539 genes). **e**, Statistical enrichment for FGF pathway genes for either amplification or deletion within the TCGA database is notably skewed towards amplification, indicating a generally oncogenic nature for this pathway in tumorigenesis. **f**, Methylation status of ExE hyper CGIs across colonic and haematopoietic mouse cancer models in which *de novo* methyltransferase activity has been perturbed. All samples are measured by RRBS. Datasets include: primary colon tissue in which Dnmt3b has been overexpressed (promoter methylation status reported, ref. 63); genetic models of acute myeloid leukaemia (AML) including those transformed by the MLL–AF9 fusion (ref. 64), cMyc and BCL2 overexpression (ref. 64), and FLT3 internal tandem duplication (FLT3-IDT, ref. 65); and acute and chronic lymphoblastic leukaemia models driven by *Dnmt3a* knockout alone (refs 66 and 67). Methylation of ExE hyper CGIs is observed in both colonic Dnmt3b overexpression and haematopoietic *Dnmt3a* knockout. Additional oncogenic drivers appear to induce *de novo* methylation of these regions in the presence or absence of DNMT3 expression, indicating numerous potentials routes to accomplishing the same molecular phenotype. Wild-type haematopoietic tissues are included for reference and taken from refs 66 and 67. Edges refer to the 25th and 75th percentiles, whiskers the 2.5th and 97.5th percentiles, respectively.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

nature research

Corresponding author(s): Alexander Meissner
Franziska Michor

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   For WGBS and RRBS data, identification of differentially methylated loci was performed by DSS, which use biological replicates and information from CpG sites across the genome to stabilize the estimation of the dispersion parameters. Only CpGs that were covered at least fivefold across all samples were considered for a given comparison. An FDR cutoff of 5% was used to identify differentially methylated CpGs. A CGI was called DMR if it was covered by at least 5 CpGs and 80% of them were significantly hyper/hypo methylated. For TCGA 450K data, given that most cancer types have more than 20 tumor and normal samples, Wilcoxon rank-sum test was used to identify differentially methylated CpGs, with a FDR cutoff of 5%. A CGI was called DMR if 80% of covered CpGs were significantly hyper/hypo methylated.

   Alignment was performed using TopHat2 against mouse genome assembly mm9 with default settings. Isoform-level expression was quantified by kallisto, which performs pseudoalignment of reads against cDNA sequence of transcripts. Gene-level expression was estimated as the sum of expression of associated isoforms. Refseq mRNA sequences were downloaded from the UCSC genome browser. Expression levels were reported as TPM.

2. Data exclusions

   Describe any data exclusions.

   No samples were excluded from analysis at this time

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   Replicates and results are reported and clearly described in the Figures, Legends and Methods.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Embryos were randomized per collection but carried forward afterwards as "matched" ICM and TE or Epiblast and Extraembryonic samples.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   n/a

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Data was analyzed using standard computational pipelines that are highlighted in Methods when used.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	n/a

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	n/a
b. Describe the method of cell line authentication used.	n/a
c. Report whether the cell lines were tested for mycoplasma contamination.	n/a
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	n/a

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Embryos were generated from mating female B6D2F1 (BDF1) mice age 5-8 wks with males of the same strain ≤12 months of age. Resultant F2 embryos were either cultured in vitro or gestated in vitro to times of 2.25, 3.5, or 6.5 days post fertilization.

All studies described in this manuscript have been approved by the IACUC (protocol number 28-21).

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Provide all relevant information on human research participants, such as age, gender, genotypic information, past and current diagnosis and treatment categories, etc. OR state that the study did not involve human research participants.