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 Cacciarelli**[1,2], **Andreas Gnirke**[1], **Franziska Michor**[1,2,4,5,#], and **Alexander Meissner**[1,2,7,#]

[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

Abstract

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 analyzed 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 orthologously methylated across most human cancer types[9]. Our

data suggest that the evolutionary innovation of extraembryonic tissues may have required cooption of DNA methylation-based suppression as an alternative to the embryonically utilized Polycomb group proteins, which coordinate germlayer formation in response to extraembryonic cues[10]. Moreover, we establish that this decision is made deterministically downstream of promiscuously utilized, and frequently oncogenic, signaling 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 subsequently evolve during early mammalian development, we generated whole genome bisulfite sequencing (WGBS) and RNA-seq datasets from mouse precompacted 8-cell stage embryos, Inner Cell Mass (ICM) and Trophectoderm from E3.5 blastocysts, as well as Epiblast and Extraembryonic Ectoderm (E×E) from E6.5 conceptuses, the latest stage where these major developmental 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, where ~80% of the genome becomes differentially methylated (Extended Data Fig. 2a). Specifically, the extraembryonic lineage lacks canonical bimodality: most CpGs are moderately less methylated than in epiblast, while 1.36% are specifically methylated (Extended Data Fig. 2b). E×E-specific hypo or hypermethylation segregate into distinct genomic compartments by CpG density and location, with *de novo* methylation preferentially enriched for CGIs near transcription start sites (TSSs) and 5′ exons (Fig. 1c, Extended Data Fig. 2c-f). Once established, these alternative landscapes are largely preserved across embryonic tissues and placenta, respectively[11,12] (Extended Data Fig. 2g).

Intriguingly, E×E-methylated CGIs (E×E Hyper CGIs) frequently overlap with PRC2-regulated genes, including master transcription factors that direct germlayer and body-axis formation (Fig. 1e, Extended Data Fig. 3a,b, Supplementary Table 1). Although the majority are not yet expressed in the epiblast, E×E-specific promoter methylation is generally associated with repression, including of many pluripotency-specific regulators, as well as concurrent loss of chromatin accessibility (Fig. 1e, Extended Data Fig. 3, 4). The lower global methylation in E×E also results in a more abrupt relationship between promoter-methylation and gene-repression that is apparent at levels as low as 0.1 (Extended Data Fig. 4c). DNA methylation surrounding these promoters is largely dispersive, with flanking regions less methylated in E×E than Epiblast, but a maximal increase specifically at the TSS (Fig. 1d). Moreover, while *de novo* CGI methylation only reaches ~0.25, methylated CpGs are distributed across 80% of the unique sequencing reads that fall within them, with a median per read methylation status of 0.25 (Extended Data Fig. 4d). The consistency between per molecule and aggregate methylation measurements can only 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 E×E-targeted regions invariably leads 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, E×E expresses alternative Wnt proteins, suppresses Fgf loci 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 signaling pathways, which are promiscuously utilized 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 TE and can be cultured independently of FGF, 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 (PD), 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 Fig. 6, Methods). Those cultured in FGF4+CHIR progressively diverged into two separate, morphologically distinguishable interior and exterior tissues that were independently isolated.

In combination, PD 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 nonphysiologically high levels (Fig. 2d). Surprisingly, when coupled with FGF, WNT agonism effectively blocks genome remethylation but redirects CGI-level methylation to a broader subset of extraembryonic targets (Fig. 2e, f). Targeting is specific to the FGF+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* methylated CGIs and E×E 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 2i condition, those methylated in E×E and FGF+CHIR, but not in FGF alone, generally encompass neuroectodermal regulators, and E×E-exclusive islands are often endodermal and induced by dual FGF and WNT activity (Extended Data Fig. 5c). Seemingly, E×E-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 programs.

We next sought to investigate the configuration of epigenetic regulators that specifically E×Ecute this transition. While *Dnmt1* and *Dnmt3b* are expressed in both tissues, *Dnmt3l* and *Dnmt3a* isoform 2 (*Dnmt3a2*) are reciprocally expressed in either the E×E or epiblast and regulated by *de novo* methylation in the alternate (Extended Data Fig. 7a-d). A truncated, non-catalytic isoform of the H3K36 demethylase *Kdm2b* is expressed over preimplantation and within E×E, while a longer Jumonji (JMJ) demethylase domain-containing isoform is specifically induced in epiblast[17] (Extended Data Fig. 7e). Otherwise, epigenetic regulator expression is 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*, *3a*, *3b*, and *3l*, the essential PRC2 component *Eed*, and *Kdm2b* by CRISPR/Cas9 injection into zygotes (Supplementary Tables 5 and 6, Methods). We find that

Dnmt1, 3b and 3l ablation substantially disrupt the E×E methylome, including at CGI targets, but show no specificity for these regions or corresponding changes to expression (Fig. 3a, b, Extended Data Fig. 7f). The near complete loss of methylation in *Dnmt1-null* E×E compared to epiblast indicates diminished *de novo* activity, and greater reliance on epigenetic maintenance, despite prolonged Dnmt3l expression (Fig. 3a, b). Alternatively, *Eed-null* E×E disrupts CGI methylation without affecting global levels, suggesting that PRC2 specifically coordinates repression upstream of DNMT3B as part of a novel developmental pathway (Fig. 3c, d, Extended Data Fig. 7g). Consistently, *Eed-null* E×E fails to suppress associated genes, which are induced to levels similar to those of sample-matched epiblast (Extended Data Fig. 7h).

Our data indicate a point in early development where sensitivity to promiscuously utilized growth factors instructs a novel epigenome that is not observed during ontogeny. However, *de novo* CGI methylation is also a general feature of tissue culture, cancer cell lines, and primary tumors, suggesting that somatic cells remain vulnerable[5,18] (Fig. 4a, Extended Data Figs. 8-10). To investigate a possible link with the subsequent reemergence of this landscape in cancer, we mapped orthologous CGIs to directly compare extraembryonically-methylated CGIs across patient-matched DNA methylation profiles from The Cancer Genome Atlas (TCGA) project, an age-matched Chronic Lymphocytic Leukemia (CLL) cohort, as well as data from ENCODE and the Roadmap Epigenomics Project[14,19-21]. Of the 16 tumor types with sufficient normal biopsied samples, 15 significantly methylate E×E Hyper CGIs (Fig. 4a, b). The signal is surprisingly robust and segregates tumor and normal tissue when measured as a feature across patients or when examining CGI-level changes (Fig. 4b, c, Extended Data Fig. 8). 84% of E×E Hyper CGIs are methylated in at least one cancer type, and they are more frequently shared as conserved, pan-cancer targets (Fig. 4d). 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 data set (n=10,629 tumors) shows a 5.3 percentage point increase in the average methylation of E×E Hyper CGIs when any FGF pathway member is mutated (Extended Data Fig. 10). Similarly, statistical assessment of the connectivity between our E×E Hyper CGIs and the ten most mutated pathways in cancer reveals a striking enrichment for FGFR signaling in disease (Enrichment Z-score=3.88, Extended Data Fig. 10). Over the more expansive, but less internally controlled, ENCODE and ROADMAP data, cancers and immortalized cell lines clearly separate from primary tissues by E×E Hyper CGI methylation status (Fig. 4e, Supplementary Table 7). Intriguingly, mature adaptive immune cells and endodermal lineages are generally more susceptible to low level methylation within these regions, suggesting a preexisting heterogeneity even in normal populations.

We present the developmental acquisition of a novel epigenetic landscape that partitions extraembryonic tissues within the embryo and resembles a frequent, global departure in genome regulation in human cancers. This landscape cooccurs with the establishment of the first major signaling axes, can be partially directed from 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 E×E requires PRC2, which indicates either a transient, biochemical interaction with the DNMTs or an upstream, primary silencing role. The coordination of this alternative, and presumably more permanent, repressive mechanism

warrants further investigation and shares notable parallels to the somatic transition to cancer. Most obviously, FGF sensing passes through RAS/MAPK/ERK signaling, which has extensive oncogenic potential and putative roles in cancer methylome establishment[22-24]. Similarly, the E×E displays attenuated *de novo* methylation activity directed wholly by DNMT3B, broadly resembling the high frequency of somatic *DNMT3A* mutations in Acute Myeloid Leukemia (AML) and Myelodysplastic Syndrome (MDS) or DNMT3B-directed CGI methylation during colorectal transformation[25-28]. Transgenic mouse cancer models confirm conserved E×E Hyper CGI methylation in similar contexts (Extended Data Fig. 10). 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 for 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 to their systematic emergence during transformation.

Online Methods

Sample isolation and library preparation

Preparation of preimplantation and postimplantation samples was performed as described in Ref 31. Briefly, B6D2F1 hybrid females between 5-8 weeks old (Charles River) were serially primed with 5 IU Pregnant Mare Gonadotropin (Sigma) followed by 5 IU Human Chorionic Gonadotrophin (HCG, Millipore) after 46 hours, 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 aphidicolin (Sigma) to ensure synchronization and minimal entry into the fourth replication cycle. 8 cell embryos were collected within 4 hours of the first apparent embryo of this stage. Prior to collection, embryos were serially transferred through Acid 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 Acid Tyrode's solution to remove the Zona and the ICM and TE 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 decidua of mated female mice were isolated on the morning of E6.5 and the conceptus 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, Epiblast and Extraembryonic Ectoderm (E×E) 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 minutes in KSOM at room temperature. Finally, 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 E×E and Epiblast or ICM and TE 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 according to Ref 33 and libraries were prepared using the Accel-NGS™ Bisulfite DNA library kit (Swift Biosciences) according to the manufacturers protocol. Final libraries were generated from 10-12 PCR cycles. RNA was purified using the RNAeasy Micro Kit (Qiagen) and RNA-seq libraries generated using the SMRT-seq v4 Ultralow Input Kit (Clontech) according to the manufacturer's protocol with 10-11 LD 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 according to Ref 34 using a 10 μl reaction and incubation with the TN5 transposase mixture (Nextera DNA library preparation kit, Illumina) for 45 minutes. Reaction was stopped according to the protocol described in Ref 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 × ten platform (Illumina), while 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 96hpf as described[31]. Briefly, oocytes were isolated by hormone priming from B6D2F1 females 12-14 hours after administration of hCG and fertilized by Intracytoplasmic Sperm Injection using piezo actuated injection of 129S1/SvIMJ strain sperm[36]. At 96 hrs post-fertilization, blastocysts were stripped of their zona pellucida by brief incubation in Acid Tyrode's solution and incubated for 30 minutes 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 minutes at 37°C, the ICM separates from the complement-lysed TE 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 1000 U/mL LIF (made in house) and one of the following conditions; '2i' supplemented with 1 μM PD0325901 and 3 μM CHIR99021(Reagents Direct)[37]; 'PD' supplemented with 1 uM PD0325901 and 10 ng/mL BMP4 to promote outgrowth expansion (Peprotech)[38]; 'FGF+CHIR' supplemented with 25ng/mL mouse recombinant FGF4 (R and D) and 3 μM CHIR99021; and 'FGF' supplemented with 25ng/mL 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 plated in 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 4 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 removal of the extraembryonic endoderm from E6.5 samples described above. However, under the FGF +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 RRBS and RNA-seq profiling.

Generation of KO embryos by CRISPR/Cas9 and zygotic injection

Zygotic injection was performed essentially as described[39]. To improve the efficiency with which null alleles were generated, three separate guide RNA sequences were designed per target, prioritizing highly scored protospacer sequences with no high scoring off target sites using the CHOPCHOP web tool[40] and at as 5′ most a coding exon as possible given these constraints. Protospacer sequences were input into the following oligonucleotide primer pair and used to amplify off of the pX300 plasmid (Addgene): Forward primer, AGTCAGTTAATACGACTCACTATA**G**N_{19}GTTTTAGAGCTAGAAATAGCAAG; 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 gRNAs 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 off of a similarly designed, T7 promoter driven template amplified off of the pX300 plasmid 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 comprised of 5 mM Tris-HCl, 0.1 mM EDTA, 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/ul of all three targeted sgRNAs pooled 1:1:1 and 200 ng/ul Cas9 mRNA. At E3.5, cavitated blastocyts 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 prior. To account for the ~1 day offset in developmental progression that results from uterine transfer, appropriately E6.5 stage conceptuses were isolated four days after uterine transfer and epiblast and extraembryonic ectoderm tissue were isolated as described above prior to snap freezing in minimal tissue. Each replicate consisted of 4 embryos and all experimental series include replicates generated from at least two 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 where 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 in phase.

Dual RRBS and RNA-seq profiling

Genomic DNA and mRNA purifications from low input samples were performed as described by Maculay *et al.* with modifications[41]. Briefly, the cells were mixed with 15 μl of RLT plus buffer (QIAGen) containing 1 U/μl of RNase inhibitor (SUPERase·In,

ThermoFisher Scientific), 1% β-mercaptoethanol (Sigma-Aldrich), and were then transferred to one well in a 96-well DNA LoBind plate (Eppendorf). After adding 10 μl of M-280 streptavidin bead-conjugated RT 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 (ThermoFisher Scientific). 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 RNA-seq library was generated according to the Smart-seq 2 protocol[42]. IndE×Ed RNA-seq libraries were pooled and sequenced in an Illumina Hiseq2500 sequencer.

Genomic DNA was isolated utilizing 1× Agencourt AMPure beads (Beckman Coulter) and was eluted to 15 μl of low TE buffer. RRBS library was generated as reported previously with modifications[43]. We utilized 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, DNA purification step was eliminated after each enzymatic reaction. Briefly, the genomic DNA was digested by 16 units of MspI (New England Biolabs) for 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 exo- (New England Biolabs), 0.03 mM dCTP, 0.03 mM dGTP and 0.3 mM dATP; ant the reaction was carried out at 30°C for 25 min; 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 indE×Ed adapters overnight at 16 °C by adding 2,000 units of T4 DNA ligase and 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 adapter dimers, the library pool was cleaned up using 1.8X AMPure beads and the adapter-tagged DNA fragments were eluted to 30 μl of low TE buffer. The bisulfite conversion of the adapter-tagged DNA fragments ware 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 1.3X AMPure beads and the RRBS libraries were paired-end sequenced for 2×100 cycles. Only instances where the matched pool of Epiblast and E×E from a given replicate both had over 1 million CpGs covered at 5× were included for downstream analysis.

For each sample, 10 μl of M-280 streptavidin beads (ThermoFisher Scientific) were prepared per the manufacturer's recommendations. Specifically, after washing with Solution A (0.1 N NaOH, 0.05 M NaCl) and B (0.1 M NaCl) sequentially, the beads were resuspended in 10 μl of 2× Washing and Binding buffer (10 mM Tris-HCl, 1 mM EDTA, 2 M NaCl) and then mixed with an equal volume of 2 μM of RT primer[41]. The mixture was incubated for 15 min at room temperature with gentle rotation. The bead-bound RT primer was collected in a

magnate 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 (CGI) 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 Fig. 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 TMP 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 by DSS, which use 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. An FDR cutoff 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 tumor and normal samples, Wilcoxon rank-sum test was used to identify differentially methylated CpGs, with a FDR cutoff 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 cutoff 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 sum of expression of associated isoforms. Refseq mRNA sequences were downloaded from UCSC genome browser. Expression levels were reported as Transcripts Per Million (TPM).

Pathway Enrichment

Pathway enrichment was performed by hypergeometric test using the GSEA online tool. P-value was adjusted for multiple hypothesis testing according to Benjamini and Hochberg, with 5% as a cutoff.

Connectivity analysis

We used GRAIL (Gene Relationships Among Implicated Loci, Ref 45) to test whether a query gene is functionally related a set of seed genes. GRAIL utilizes 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 who function in the same pathway tend to distribute in a coherent sub-network. In this study, we built a sub-network using E×E Hyper CGI-associated genes, which were significantly enriched in several pathways. To predict whether a query gene is functionally related to the E×E 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 function MarkDuplicates from Picard tool kit. Reads with low mapping quality (< 10) or in mitochondrial chromosome were removed. NucleoATAC was used to generate Insert density, which was normalized by the total number insertions in each sample[46].

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.
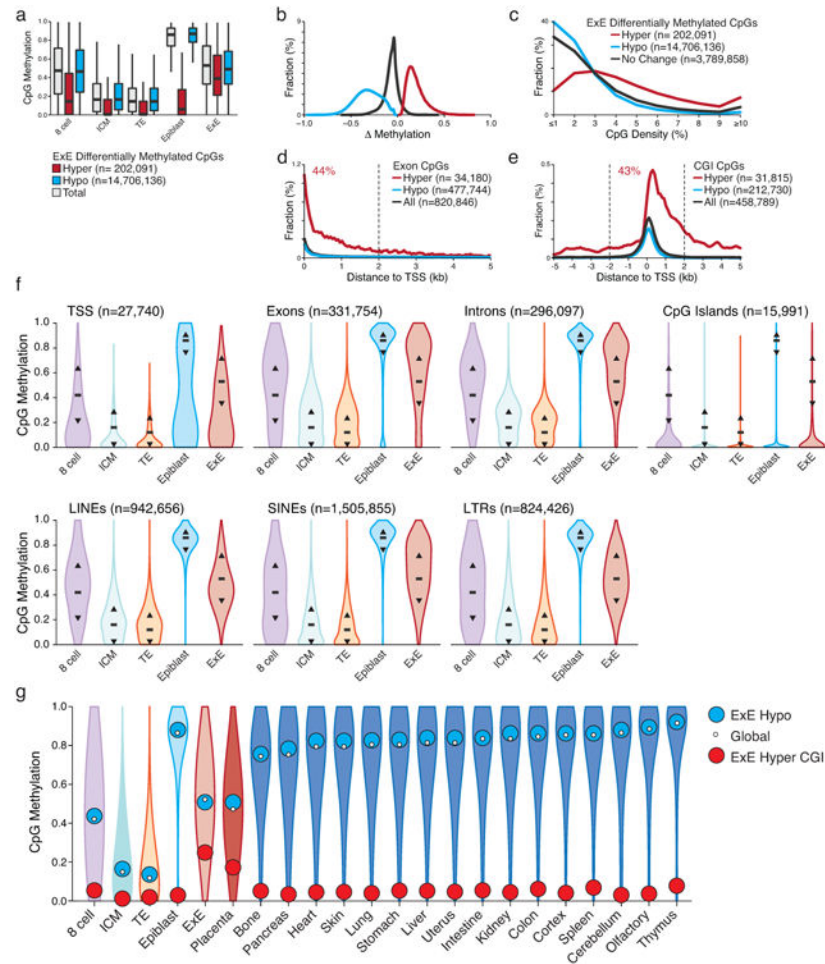
Extended Data
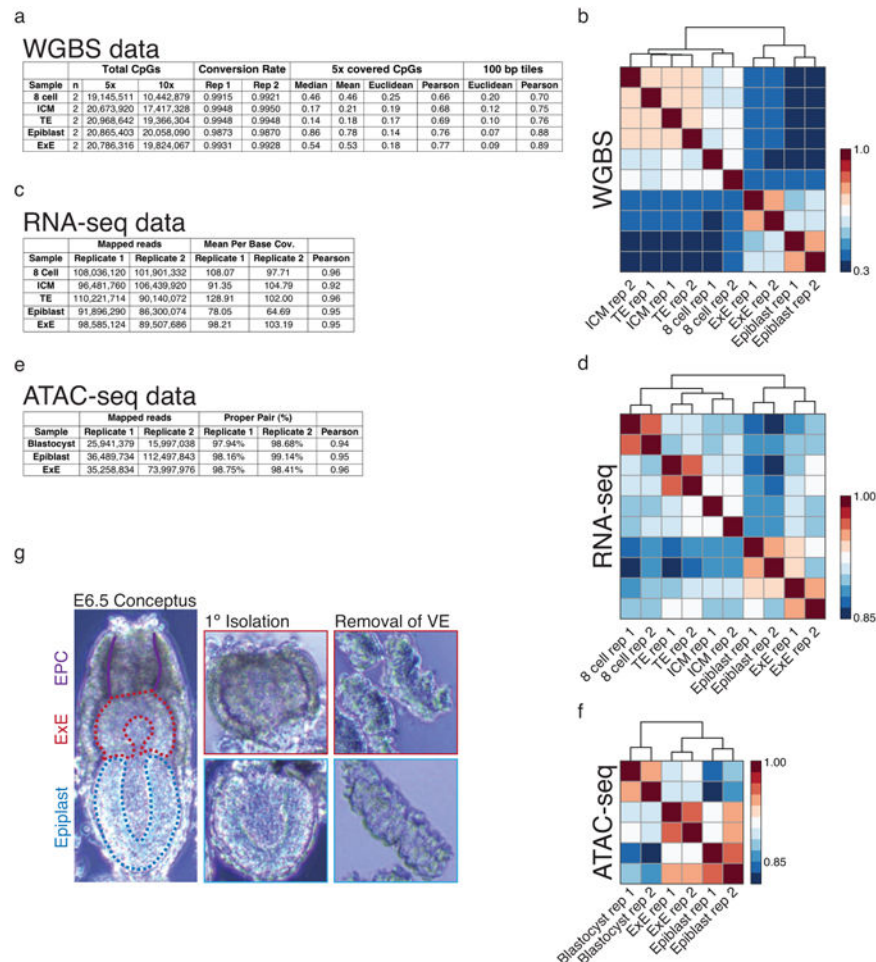
Extended Data Figure 1. Tracking divergence in DNA methylation landscapes during mouse implantation

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 post-implantation WGBS data, while tissues of the E3.5 blastocyst cluster together but not as discrete ICM and TE compartments. In general, there is minimal variation between the methylation status of the ICM and TE, with only slight global deviations around the minimal global value that is reached during this developmental period.

g. Isolation of the Epiblast and Extraembryonic Ectoderm (E×E) from the E6.5 post-implantation embryo. The conceptus is first removed from maternal decidual tissue and portioned into Epiblast and E×E fractions, taking care to remove the apical Ectoplacental

Cone (EPC). Then, outer visceral endoderm and trophoblast cells are enzymatically digested and mechanically removed using a thin glass capillary.

Extended Data Figure 2. Unique features of the extraembryonic methylation landscape

a. CpG methylation boxplots for all covered CpGs (gray) as well as those that are significantly hyper (red) or hypo (blue) methylated within E×E compared to Epiblast. E×E-hypomethylated CpGs largely reflect differential remethylation compared to Epiblast across the genome. Alternatively, E×E hypermethylated CpGs are mostly unmethylated in ICM and TE and remain so in Epiblast, indicating an E×E-specific mechanism. Edges refer to the 25th and 75th percentiles, and whiskers the 2.5th and 97.5th percentiles, respectively.

b. Distribution of significantly hypermethylated and hypomethylated CpGs between E×E and Epiblast (E×E hyper and E×E hypo, respectively). Hypomethylation appears to be a global feature of the E×E 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 hypomethylated within the Epiblast and subsequent embryonic and adult somatic tissues.

c. Alternate CpG density distributions for E×E hypomethylated and hypermethylated CpGs indicate differential enrichment within distinct genomic features. While E×E

hypomethylated CpGs resemble the global average, hypermethylated CpGs occur within higher CpG densities.

d. The fraction dynamically methylated CpGs that fall within annotated exons as a function of distance to their assigned transcription start site (TSS). 44% of exonal E×E hypermethylated CpGs fall within 2 kb of their associated TSS.

e. The fraction of dynamically methylated CpGs that fall within annotated CpG islands based upon their proximity to the nearest TSS. E×E hypermethylated CpGs are generally TSS proximal and skew downstream of the TSS, with 43% falling within + or – 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 25th/75th percentiles. Globally, all features exhibit the expected passage through minimal DNA methylation values within the ICM and TE of the E3.5 blastocyst prior to remethylation at implantation. Compared to its global distribution, E×E 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's refer to the number of annotated features of a given type.

g. Violin plots of 100 bp methylation data for early embryonic, placental, and adult tissues demonstrate general epigenetic retention of either the somatic Epiblast or extraembryonic E×E architecture throughout subsequent development. White dot highlights the global median, while blue and red reflect the median of E×E-hypomethylated tiles and E×E hyper CGIs, respectively. Notably, extraembryonic placenta largely preserve the hypomethylated global landscape and targeted methylation of otherwise canonically hypomethylated CGI promoters after they are established by E6.5. We show tiles and islands for E×E-specific hypomethylation and hypermethylation respectively to restrict CpGs to a notable feature where they change as a group. WGBS data of adult tissues taken from Ref 11.
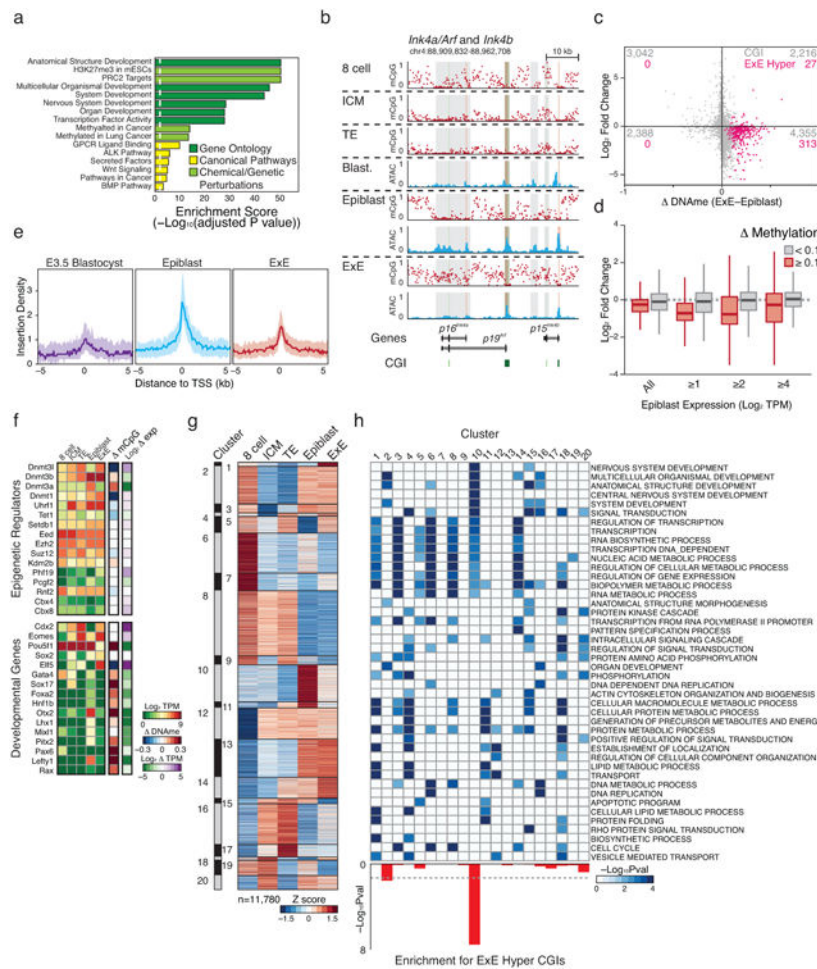
Extended Data Figure 3. Transcriptional differences between Epiblast and E×E are directed in part through DNA methylation

a. Select gene set enrichment analysis of E×E-hypermethylated transcription start sites including Gene Ontology, Canonical Pathways, and Genetic and Chemical Perturbations. E×E-hypermethylated promoters are highly enriched for transcription factors and signaling pathways involved in patterning the early embryo. Moreover, these CGI regulated 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 tumor suppressors $p16^{Ink4a}$, $p19^{Arf}$, and $p15^{Ink4b}$. While these loci are either basally or non-transcribed during early development, three regions are dynamically methylated in E×E (highlighted in gray), including a >10 kb region that encompasses the entirety of the $p16^{Ink4a}$ locus and is either wholly unmethylated in Epiblast or extensively methylated in E×E. CpG islands 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 Epiblast and E×E. While most dynamically methylated CGI-promoter containing genes have functions in later embryonic development and are not yet highly expressed, *de novo* methylation in E×E is generally associated with transcriptional repression. E×E hyper CGIs

are highlighted in pink. Promoter CGIs are assigned to the most proximal gene within a + or – 2 kb boundary.

d. Boxplots demonstrating the relationship between promoter methylation and expression in the restriction of extraembryonic and embryonic compartments. Promoters are defined as + or – 1 kb of an annotated TSS and scored as dynamically methylated in E×E if the difference with Epiblast is 0.1. Expression changes between dynamically methylated and background promoter sets are provided over increasing thresholds according to their expression in Epiblast. While many CGI promoters are not dynamically expressed in either Epiblast or E×E and are associated with genes that have 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 E×E Hyper CGI-associated TSSs in the transition from pre- to postimplantation. E×E 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 E×E, 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 regulators and master regulators over early embryonic and extraembryonic development. Most epigenetic regulators exhibit minimal expression differences between Epiblast and E×E, with the Dnmts being notable exceptions. Key isoforms of *Dnmt3a* and *Dnmt3b* are upregulated in Epiblast in conjunction with global remethylation, while the suppression of *Dnmt3a* in E×E 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 E×E, with reciprocal methylation of the *Dnmt3l* promoter in Epiblast. The histone 3 lysine 36 demethylase *Kdm2b* displays differential expression of catalytically active and inactive isoforms within Epiblast and E×E, respectively, with isoform switching seemingly imposed by *de novo* methylation around the somatically utilized CGI promoter. The E×E is characterized by persistent expression of the master regulators *Cdx2*, *Eomes*, and *Elf5* (Refs 47-50), while the still pluripotent epiblast remains *Pou5f1* (*Oct4*) positive. Many additional regulators of subsequent developmental stages are basally expressed within the Epiblast and their promoters *de novo* methylated in E×E. The difference in promoter methylation refers to the annotated TSS that exhibits the greatest absolute difference between E×E and Epiblast. TPM: Transcripts Per Million. Additional high-resolution genome-browser tracks are displayed for select transcriptional and epigenetic regulators in Extended Data Figure 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 E×E. Heatmap intensity reflects the row-normalized Z score.

h. Significant Gene Ontology enrichment for the 20 gene expression dynamics characterized in **e**, including for those regulated by E×E-methylated CGI promoters, as calculated using the binomial test. Cluster 10 is enriched for both developmental functions and E×E promoter methylation.
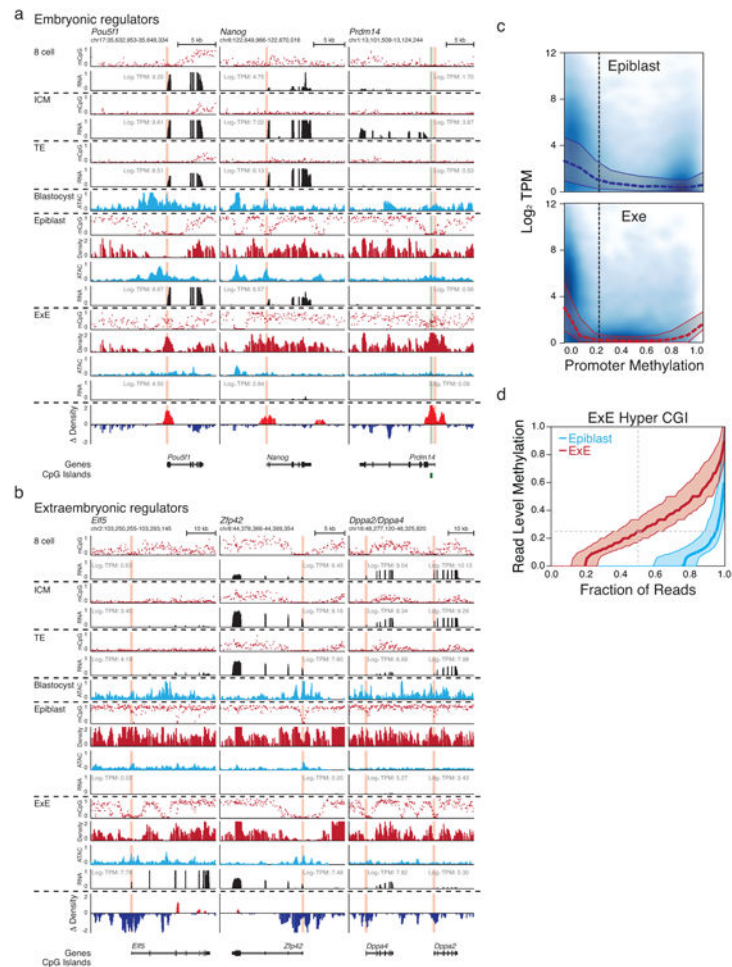
Extended Data Figure 4. Unique bifurcation and epigenetic reinforcement of transcriptional regulators during postimplantation development
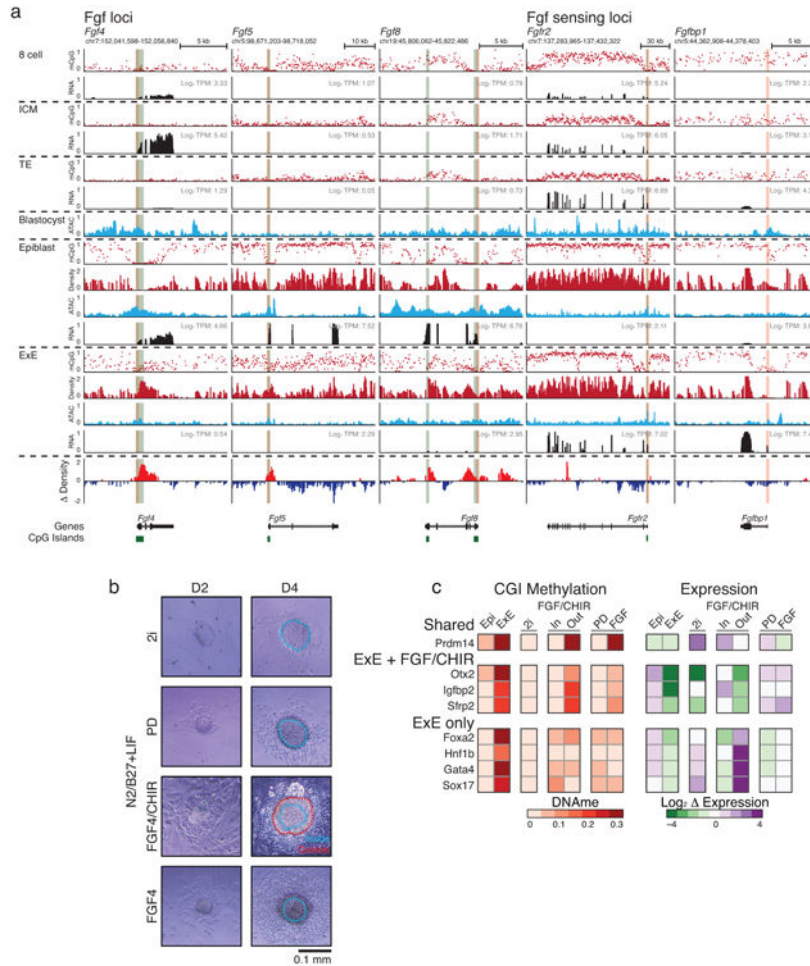
a. Genome browser tracks for WGBS, ATAC-seq and RNA-seq data for transcriptional regulators associated with embryonic or extraembryonic development. CGIs are highlighted in green, and the positions of included TSSs are highlighted in red. Embryonic regulators include Pou5f1, Nanog, and Pdrm14, which are progressively expressed over preimplantation and for which Pou5f1 and Nanog remain expressed in the Epiblast. For these genes, repression in ExE is accompanied by differential methylation of their TSSs, which is strikingly apparent as a local hypermethylation "peak" at the *Pou5f1* locus within an ~5kb region that is otherwise hypomethylated in Epiblast. At the *Nanog* locus, an upstream region remains hypomethylated in both tissues. Finally, *de novo* methylation of the *Prdm14* promoter is representative of the unique ExE-specific target of CGIs promoters 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 ExE specifically (Density refers to the difference compared to 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 Epiblast. Intriguingly, many transcriptional regulators associated with pluripotency and germline development

persist within the E×E, including Zfp42 and the paralogs Dppa2 and Dppa4. As with *Elf5*, the promoters for these genes are differentially methylated in Epiblast and frequently characterized by broad kilobase-scale hypomethylation surrounding their TSS in E×E.

c. Scatterplots for $\log_2$ Transcripts Per Million (TPM) as a function of promoter methylation reveals a higher sensitivity to low methylation levels in E×E in comparison to epiblast. Median and 25th, 75th percentiles for expression is overlayed over bins of 0.1. The fraction of unmethylated promoters is very similar between each tissue and exhibit comparable expression values. Promoters are calculated as + or – 1 kb of an annotated TSS.

d. Read level methylation of E×E Hyper CGIs in E×E and Epiblast. The methylation status for every sequencing read within a given CGI was ranked and binned into percentiles. Plotted are the median and 25th/75th percentile for these ranks across E×E Hyper CGIs for both E×E and Epiblast. In general, ~80% of reads falling within these regions are methylated in E×E, with a median methylation value of 0.25, very close to the average, unphased measurement for the CGI entirely, indicating that *de novo* methylation occurs within a high fraction of cells within the E×E and to a similar extent.

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 Fgf growth factors, receptors, and potentiators that are dynamically regulated during early post-implantation development. Fgf's such as the ICM-specific Fgf4 and Epiblast-specific Fgf5 and Fgf8 are all regulated by CGI containing promoters that are *de novo* methylated in E×E. Alternatively, Fgf sensing genes such as Fgfr2 the potentiating protein Fgfbp1 become specific to the E×E and are 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 prior to the epigenetic restriction that is imposed by DNA methylation during implantation[51,52]. 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 E×E specifically (Density refers to the difference compared to epiblast).

b. Bright field images of ICM outgrowths after 2 or 4 days under disparate growth factor or small molecule conditions. All ICM were cultured on irradiated feeders in a basal N2/B27 media supplemented with Leukemia Inhibitory Factor (LIF). 2i refers to the canonical FGF-inhibited, WNT-active condition supplemented with 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 signaling in the absence of an additional WNT input[53]. 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, ICM were cultured in FGF4 alone. Outlines highlight the specific components of each outgrowth that were subsequently purified for analysis by dual RRBS and RNA-seq profiling (see Methods). Scale bar shown on the bottom right.

c. Differential methylation of CGIs *in vitro* differs from E×E according to developmental trajectory. Shown are specific TSS-associated CGIs that are either methylated in E×E and both conditions, E×E and FGF/CHIR, or E×E-only and the corresponding mean adjusted Log$_2$ fold change in gene expression, respectively. 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[54]. FGF differs from the E×E and FGF/CHIR conditions in the methylation of CGI's associated with either the epiblast or 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, E×E 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 Transcripts Per Million, 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 E×E Hyper CGI methylation is not observed with PD 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.

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
PD	6	2,708,588	2,448,519	0.50	0.43	0.88	0.18
FGF	5	2,666,145	2,416,074	0.90	0.72	0.97	0.09
FGF/CHIR inside	7	2,668,463	2,424,931	0.44	0.39	0.88	0.19
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
Dnmt1	5	2,585,520	2,341,277	0.43	0.38	0.81	0.17
Dnmt3a	5	2,774,163	2,542,382	0.77	0.60	0.97	0.11
Dnmt3b	6	2,765,347	2,539,808	0.61	0.51	0.94	0.15
Dnmt3l	5	2,490,838	2,242,846	0.73	0.58	0.86	0.22
Eed	4	2,506,687	2,246,193	0.80	0.62	0.92	0.16
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
Dnmt1	5	2,520,255	2,274,229	0.16	0.19	0.62	0.16
Dnmt3a	4	2,602,481	2,356,566	0.43	0.42	0.85	0.17
Dnmt3b	6	2,543,780	2,302,780	0.20	0.25	0.82	0.14
Dnmt3l	4	2,745,180	2,479,070	0.33	0.36	0.84	0.17
Eed	4	2,460,919	2,178,707	0.50	0.45	0.88	0.16
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
PD	6	17,195,224	0.91	11.38	0.96
FGF	5	19,735,502	0.92	13.40	0.98
FGF/CHIR inside	3	25,130,944	0.91	17.51	0.99
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
Dnmt1	5	16,323,776	0.90	10.93	0.94
Dnmt3a	5	16,092,316	0.91	10.64	0.97
Dnmt3b	6	17,284,693	0.91	11.62	0.96
Dnmt3l	5	19,409,738	0.91	13.25	0.97
Eed	4	28,232,447	0.90	18.23	0.96
Kdm2b	3	26,003,962	0.93	18.14	0.98
ExE Wild Type	4	11,537,966	0.93	8.25	0.98
Dnmt1	5	16,171,034	0.93	11.42	0.95
Dnmt3a	4	18,319,262	0.93	13.10	0.96
Dnmt3b	6	20,682,531	0.92	15.17	0.96
Dnmt3l	4	20,708,730	0.92	14.81	0.95
Eed	4	25,794,839	0.92	17.24	0.97
Kdm2b	3	26,295,000	0.93	18.60	0.91

c

gRNA sequences

Target	gRNA	Coordinates (mm10)	Sequence
Dnmt1	1	chr9:20936581-20936600:−	GAAACTGGAAGAGGTAACAG
Dnmt1	2	chr9:20929059-20929078:−	GACTCCCGAGGACAGAGACG
Dnmt1	3	chr9:20926483-20926502:+	AGTATCAAACCACGTCGAGG
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 and **b.** Sequencing metrics and coverage information for dual RRBS and RNA-seq 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 (gRNA) sequences specific to early exons that are shared across different isoforms. The genomic coordinates and protospacer sequences are provided (see Methods).
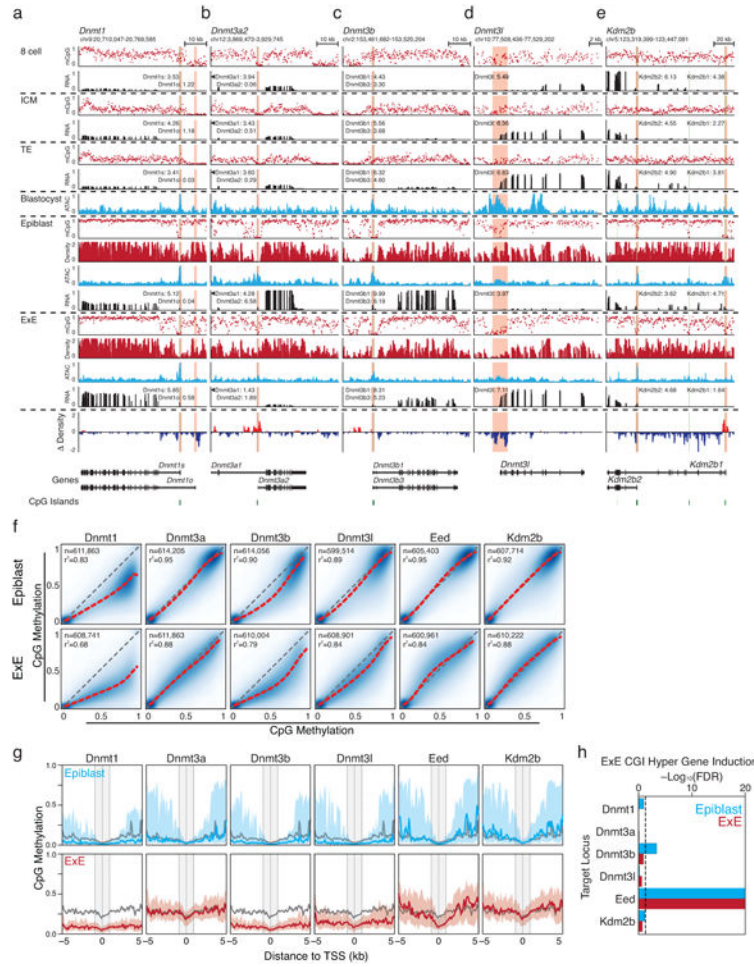
Extended Data Figure 7. Dynamic expression and epigenetic regulation of key epigenetic regulators during early implantation

Genome browser tracks for WGBS, ATAC-seq and RNA-seq data for the *Dnmt1*, *Dnmt3a*, and *Dnmt3b* loci, as well as corresponding RNA-seq data (in Log$_2$ Transcripts Per Million, TPM, to highlight the expression of 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 due to a transient maternal imprint over the somatically-utilized TSS (Dnmt1s)[33,55], but shows moderate induction within the ICM. Then, at implantation, it is induced within both the Epiblast and E×E. Dnmt1 is expressed at higher levels within the E×E 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 during implantation and is also expressed within embryonic stem cells (ESCs). Alternatively, the CGI-containing promoter of Dnmt3a2 is methylated in E×E and its transcription is suppressed.

c. Like Dnmt1, the Dnmt3b promoter contains a CGI that is maternally imprinted during preimplantation[33,55]. 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[56]. During implantation, Dnmt3l is initially expressed in both ICM and TE, but it remains expressed in E×E and is silenced by *de novo* methylation in the Epiblast.

e. The Histone-3 lysine 36 (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[57-60]. A catalytically-inactive isoform, *Kdm2b2*, initiates from an alternate TSS downstream of exons encoding the demethylating Jumonji (JMJ) domain of the catalytically active Kdm2b1[17]. Kdm2b2 is the most prevalent isoform during preimplantation development and remains expressed in the E×E. Alternatively, *Kdm2b1* is only induced during implantation within the Epiblast, while its CGI-containing promoter gains methylation in E×E. Like *Dnmt1s* and *Dnmt3b*, the CGI promoter of *Kdm2b1* is a maternally-methylated imprint that resolves to hypomethylation during implantation[33,55].

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 E×E (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 E×E, presumably because of compensation from Dnmt3a. Intriguingly, the decrease in global methylation levels when Dnmt1 is deleted is greater for E×E than epiblast, indicating a higher dependence on maintenance and less efficient *de novo* methyltransferase activity in this tissue. The identity line is included in gray and the best fit by LOESS regression in red. The number of 100 bp tiles used in each comparison and the r^2 are included in the upper left of each plot.

g. Composite plots of E×E Hyper CGI-containing promoters in CRISPR/Cas9 targeted Epiblast and E×E respectively. In general, only limited effects are observed in 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, 3b, and 3l null samples. Specificity for diminished methylation at the TSS is observed in *Eed*-null E×E, particularly downstream within the first kilobase. In both Epiblast and E×E, the wild type median is included in gray for comparison. Line represents the median and shaded area the 25th/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 E×E Hyper CGI associated genes demonstrates a comparable requirement for Eed in both epiblast and E×E. Gene expression of KO samples were compared to matched WT samples using DESeq2 with raw counts as input. Enrichment for E×E 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.
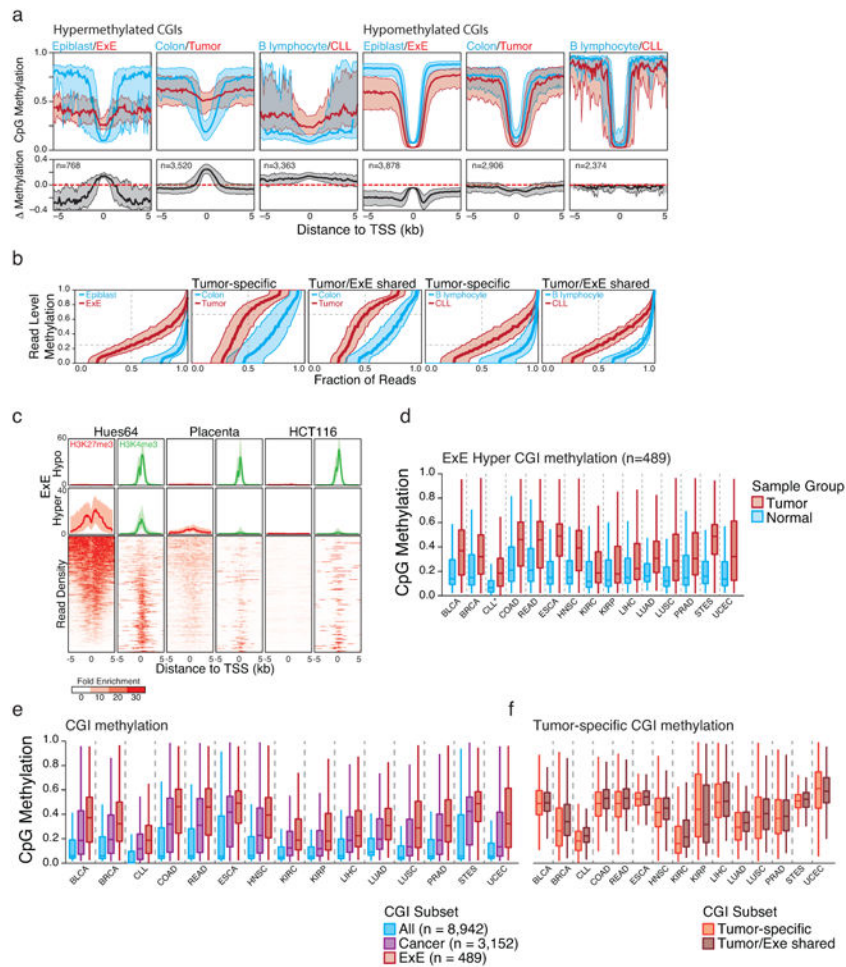
Extended Data Figure 8. General features of the cancer methylome and of CGI DMRs

a. Median methylation of differentially regulated CGI-containing promoters in a primary colon tumor isolate and Chronic Lymphocytic Leukemia (CLL) compared to colon and B lymphocytes, respectively. E×E Hyper CGIs as identified in this study and shown in Figure 1 are included for reference. The median methylation difference between extraembryonic or cancerous tissue compared to Epiblast or normal tissue is also included. The general features of both cancer methylomes are similar to that of the E×E, with a maximal increase in DNA methylation centered 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[61]. Shaded area represents the 25th and 75th percentiles per 100 bp bin.
b. Read level methylation of hypermethylated CGIs in E×E vs Epiblast, Colon Tumor vs Colon, and CLL vs B lymphocyte, with those islands 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/75th percentile for these ranks across CGIs called as hypermethylated in each pairwise comparison. The E×E-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 can only be obtained by dispersive *de novo* methylation across the majority of alleles within the population. Alternatively, Colon Tumor exhibits substantially higher read level methylation, with a median per read methylation 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 Figure 9. Moreover, the extent to which E×E Hyper CGIs are methylated within each tumor reflects the read-level methylation distribution for the tumor. As such, the targeting to E×E 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 embryonic stem cell line HUES64, a proxy for the pluripotent epiblast, displays notable enrichment for both repressive, PRC2 deposited H3K27me3 and activating H3K4me3 modifications at orthologous E×E Hyper CGIs. Alternatively, human placenta exhibits diminished enrichment for both modifications at these regions, as does HCT116. Both systems display substantial methylation over these islands as presented in Figure 4, Extended Data Figure 9, and Supplementary Table 7. As a control, "E×E hypo" demonstrates uniformly high H3K4me3 levels across all three tissues. Enrichment density heatmaps are provided for the full E×E-hyper set and are ranked across plots according to their enrichment for H3K27me3 in HUES64. Normalized enrichment represents the fold ChIP-enrichment against sample matched Whole Cell Extract (WCE).

d. Boxplots of mean methylation for 489 E×E-methylated, orthologous CGIs (E×E Hyper CGIs) across the 14 tissue-matched TCGA tumor types that display disregulated DNA methylation landscapes and for CLL. Note: 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 data sets and CLL for the absolute methylation values of All orthologously mapped CGIs, those methylated across **Cancer**, and those that are specifically methylated in mouse E×E. In all 15 cancer types that exhibit general global hypomethylation and CGI methylation as part of their departure from somatic cells, E×E Hyper CGIs are specifically enriched, more so than for CGIs that are observed as hypermethylated in any tumor.

f. Boxplots for the same TCGA data for tumor-specific CGIs and those that are also methylated in mouse E×E. Notably, the extent to which mouse E×E Hyper CGIs are methylated reflect the tumor, with some cancer types exhibiting higher absolute methylation values than others. However, in 14 out of 15 cases, the absolute methylation status of tumor-specific CGI DMRs and those that are also methylated in E×E are nearly identical, and often slightly greater. Absolute methylation values therefore appear to be determined by the cancer type, while targeting of extraembryonically methylated CGIs is a general feature.
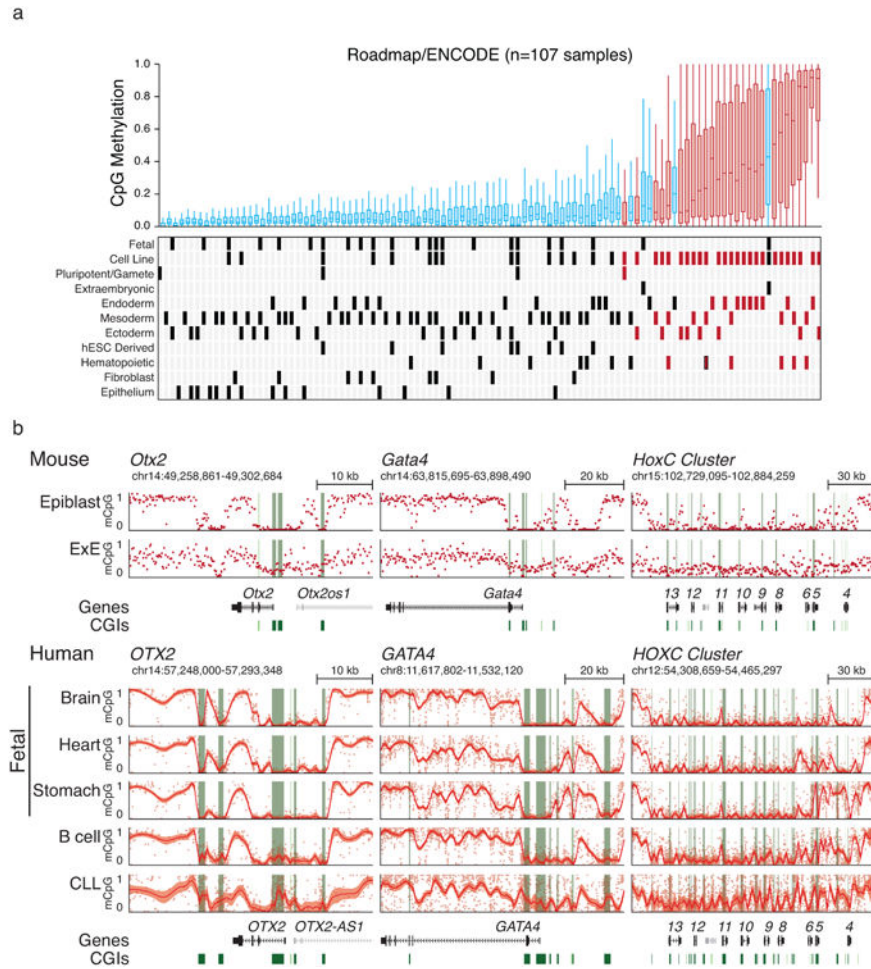
Extended Data 9. Broad conservation of extraembryonic methylation patterns across cancer types and cell lines

a. Boxplots of orthologous E×E Hyper CGIs across 107 ENCODE/Roadmap samples as presented in Figure 4, with notable additional features of each sample highlighted below. Notably, 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 E×E Hyper CGIs include numerous endodermal lineages, such as colonic mucosa, stomach and liver (mean methylation of 0.275, 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 of 0.199, 0.173 and 0.173, respectively). In contrast, ectodermal and epithelial cells are comparatively less methylated than other somatic tissues, although cancer cell lines and primary tumors derived from these tissues remain sensitive to hypermethylation.

b. Genome browser tracks for orthologous loci as originally presented for mouse development in Figure 1 display high similar transitions during transformation. Loci include *OTX2*, *GATA4*, and the entire *HOXC* cluster in three human fetal tissues that represent each germ layer (Brain, Ectoderm; Heart, Mesoderm; Stomache, Endoderm), primary human B lymphocytes, and a Chronic Lymphocytic Leukemia (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 islands while methylation values drop in the surrounding areas. Red line and shaded area reflect the local mean and standardized deviation as calculated by local regression (LOESS) to compensate for the greater number of CpGs within the human orthologs versus mouse, which can complicate visual estimates of local methylation at these scales. CGIs are highlighted in green, and the positions of included TSSs are highlighted in red.
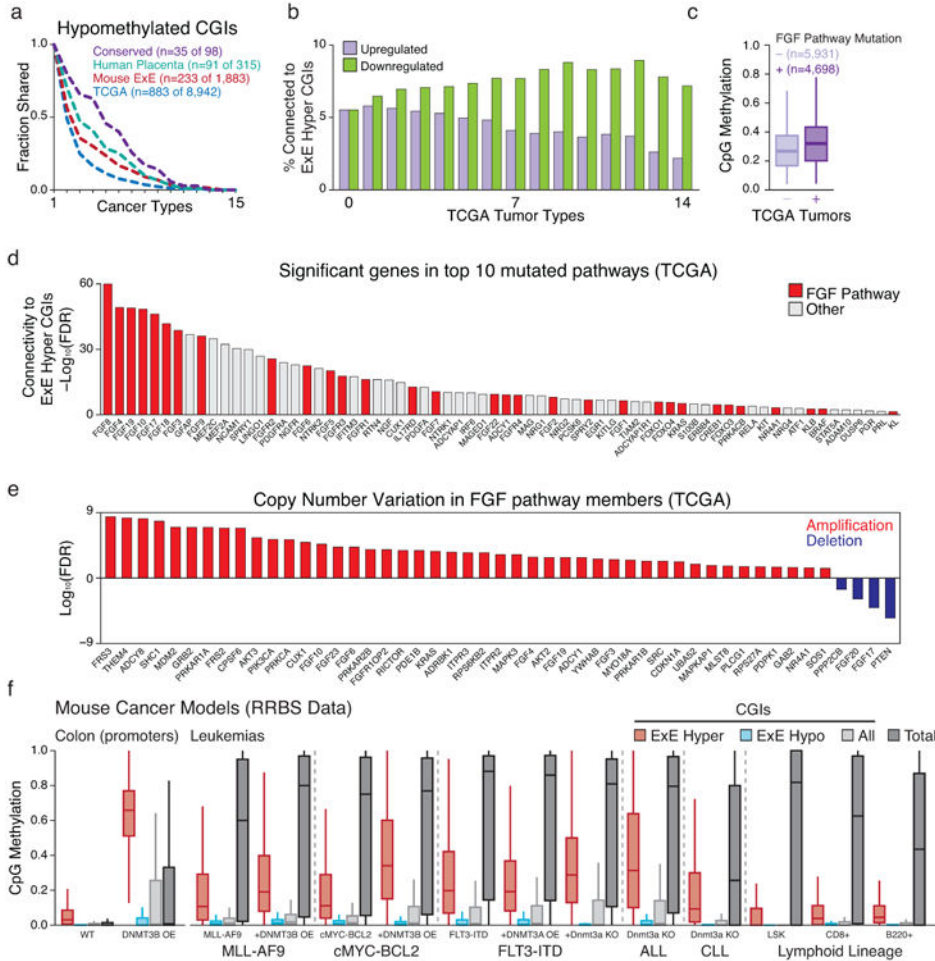
Extended Data Figure 10. Genetic features of E×E CGI methylation in cancers
a. Intersection analysis as presented in Figure 4 for cancer-hypomethylated CGIs across the 14 TCGA tumor types and CLL that exhibit global loss of methylation in tandem with CGI hypermethylation. Generally, CGI hypomethylation is more specific, such that the intersection across cancers decays exponentially. Notably, even for hypomethylated CGI, the intersection across cancer types remains higher for those that are also hypomethylated in E×E, human placenta, or both (conserved).
b. Intersection analysis for cancer-dysregulated genes across TCGA tumor types. Of genes significantly dysregulated in at least n (0 – 14) TCGA cancer types, the fraction of genes

that are functionally related to E×E 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 cutoff. As the number of TCGA tumor types increases, the fraction of E×E Hyper CGI-associated genes within the downregulated set generally increases, while those that are upregulated decreases substantially.

c. Boxplots of the average methylation for the 489 orthologous E×E Hyper CGI feature set for 10,629 tumors available in TCGA with matched mutational and methylation data, segregated by mutational status of genes that function as part of the FGF signaling pathway. In aggregate, tumors with FGF pathway mutations have a median average E×E 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 E×E Hyper CGI-associated genes (FDR < 5%), as predicted by GRAIL using text-mining database. Genes in FGF-signaling pathway are highlighted in red. In general, FGF signaling pathway genes have high connectivity scores to E×E 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 E×E Hyper CGIs across colonic and hematopoietic mouse cancer models where *de novo* methyltransferase activity has been perturbed. All samples are measured by RRBS. Data sets include: primary colon tissue in which Dnmt3b has been overexpressed (promoter methylation status reported, Ref 62); genetic models of acute myeloid leukemia (AML) including those transformed by the MLL-AF9 fusion (Ref 63), cMyc and Bcl2 overexpression (Ref 63), and FLT3 internal tandem duplication (FLT3-IDT, Ref 64); and Acute and Chronic Lymphoblastic Leukemia models driven by Dnmt3a knock out alone (Refs 65 and 66). Methylation of E×E Hyper CGIs is observed in both colonic Dnmt3b overexpression and hematopoietic Dnmt3a knockout, though additional oncogenic drivers appear sufficient to induce *de novo* methylation of these regions in the presence or absence of Dnmt3 expression, indicating the potential of numerous drivers to activate this pathway. Wild type hematopoietic tissues are included for reference and taken from Ref 65 and 66. Edges refer to the 25th and 75th percentiles, whiskers the 2.5th and 97.5th percentiles, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013; 14:204–220. DOI: 10.1038/nrg3354 [PubMed: 23400093]

2. Ohm JE, et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet. 2007; 39:237–242. DOI: 10.1038/ng1972 [PubMed: 17211412]

3. Schlesinger Y, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet. 2007; 39:232–236. DOI: 10.1038/ng1950 [PubMed: 17200670]

4. Widschwendter M, et al. Epigenetic stem cell signature in cancer. Nat Genet. 2007; 39:157–158. DOI: 10.1038/ng1941 [PubMed: 17200673]

5. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. Nat Rev Genet. 2006; 7:21–33. DOI: 10.1038/nrg1748 [PubMed: 16369569]

6. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. Science. 2017; 357

7. Schroeder DI, et al. The human placenta methylome. Proc Natl Acad Sci U S A. 2013; 110:6037–6042. DOI: 10.1073/pnas.1215145110 [PubMed: 23530188]

8. Branco MR, et al. Maternal DNA Methylation Regulates Early Trophoblast Development. Dev Cell. 2016; 36:152–163. DOI: 10.1016/j.devcel.2015.12.027 [PubMed: 26812015]

9. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011; 25:1010–1022. DOI: 10.1101/gad.2037511 [PubMed: 21576262]

10. Arnold SJ, Robertson EJ. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Biol. 2009; 10:91–103. DOI: 10.1038/nrm2618 [PubMed: 19129791]

11. Hon GC, et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat Genet. 2013; 45:1198–1206. DOI: 10.1038/ng.2746 [PubMed: 23995138]

12. Ziller MJ, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013; 500:477–481. DOI: 10.1038/nature12433 [PubMed: 23925113]

13. Landan G, et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat Genet. 2012; 44:1207–1214. DOI: 10.1038/ng.2442 [PubMed: 23064413]

14. Landau DA, et al. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. Cancer Cell. 2014; 26:813–825. DOI: 10.1016/j.ccell.2014.10.012 [PubMed: 25490447]

15. Arman E, Haffner-Krausz R, Chen Y, Heath JK, 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 U S A. 1998; 95:5082–5087. [PubMed: 9560232]

16. Leitch HG, et al. Naive pluripotency is associated with global DNA hypomethylation. Nat Struct Mol Biol. 2013; 20:311–316. DOI: 10.1038/nsmb.2510 [PubMed: 23416945]

17. Boulard M, Edwards JR, Bestor TH. Abnormal X chromosome inactivation and sex-specific gene dysregulation after ablation of FBXL10. Epigenetics Chromatin. 2016; 9:22. [PubMed: 27252784]

18. Meissner A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008; 454:766–770. DOI: 10.1038/nature07107 [PubMed: 18600261]

19. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. DOI: 10.1038/nature11247 [PubMed: 22955616]

20. Hoadley KA, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 2014; 158:929–944. DOI: 10.1016/j.cell.2014.06.049 [PubMed: 25109877]

21. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–330. DOI: 10.1038/nature14248 [PubMed: 25693563]

22. MacLeod AR, Rouleau J, Szyf M. Regulation of DNA methylation by the Ras signaling pathway. J Biol Chem. 1995; 270:11327–11337. [PubMed: 7744770]

23. Lu CW, et al. Ras-MAPK signaling promotes trophectoderm formation from embryonic stem cells and mouse embryos. Nat Genet. 2008; 40:921–926. DOI: 10.1038/ng.173 [PubMed: 18536715]

24. Serra RW, Fang M, Park SM, Hutchinson L, Green MRA. KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. Elife. 2014; 3:e02313. [PubMed: 24623306]

25. Rhee I, et al. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature. 2002; 416:552–556. DOI: 10.1038/416552a [PubMed: 11932749]

26. Lin H, et al. Suppression of intestinal neoplasia by deletion of Dnmt3b. Mol Cell Biol. 2006; 26:2976–2983. DOI: 10.1128/MCB.26.8.2976-2983.2006 [PubMed: 16581773]

27. Ley TJ, et al. DNMT3A mutations in acute myeloid leukemia. N Engl J Med. 2010; 363:2424–2433. DOI: 10.1056/NEJMoa1005143 [PubMed: 21067377]

28. Walter MJ, et al. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. Leukemia. 2011; 25:1153–1158. DOI: 10.1038/leu.2011.44 [PubMed: 21415852]

29. Novakovic B, Saffery R. Placental pseudo-malignancy from a DNA methylation perspective: unanswered questions and future directions. Front Genet. 2013; 4:285. [PubMed: 24368911]

30. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–674. DOI: 10.1016/j.cell.2011.02.013 [PubMed: 21376230]

31. Smith ZD, et al. DNA methylation dynamics of the human preimplantation embryo. Nature. 2014; 511:611–615. DOI: 10.1038/nature13581 [PubMed: 25079558]

32. Chenoweth JG, Tesar PJ. Isolation and maintenance of mouse epiblast stem cells. Methods Mol Biol. 2010; 636:25–44. DOI: 10.1007/978-1-60761-691-7_2 [PubMed: 20336514]

33. Smith ZD, et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature. 2012; 484:339–344. DOI: 10.1038/nature10960 [PubMed: 22456710]

34. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013; 10:1213–1218. DOI: 10.1038/nmeth.2688 [PubMed: 24097267]

35. Lara-Astiaso D, et al. Immunogenetics. Chromatin state dynamics during blood formation. Science. 2014; 345:943–949. DOI: 10.1126/science.1256271 [PubMed: 25103404]

36. Yoshida N, Perry AC. Piezo-actuated mouse intracytoplasmic sperm injection (ICSI). Nat Protoc. 2007; 2:296–304. DOI: 10.1038/nprot.2007.7 [PubMed: 17406589]

37. Ying QL, et al. The ground state of embryonic stem cell self-renewal. Nature. 2008; 453:519–523. DOI: 10.1038/nature06968 [PubMed: 18497825]

38. Ying QL, 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. 2003; 115:281–292. [PubMed: 14636556]

39. Wang H, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013; 153:910–918. DOI: 10.1016/j.cell.2013.04.025 [PubMed: 23643243]

40. Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res. 2016; 44:W272–276. DOI: 10.1093/nar/gkw398 [PubMed: 27185894]

41. Macaulay IC, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015; 12:519–522. DOI: 10.1038/nmeth.3370 [PubMed: 25915121]

42. Picelli S, et al. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014; 9:171–181. DOI: 10.1038/nprot.2014.006 [PubMed: 24385147]

43. Gu H, et al. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc. 2011; 6:468–481. DOI: 10.1038/nprot.2010.190 [PubMed: 21412275]

44. Wu H, et al. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. Nucleic Acids Res. 2015; 43:e141. [PubMed: 26184873]

45. Raychaudhuri S, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 2009; 5:e1000534. [PubMed: 19557189]

46. Schep AN, et al. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. 2015; 25:1757–1770. DOI: 10.1101/gr. 192294.115 [PubMed: 26314830]

47. Ciruna BG, Rossant J. Expression of the T-box gene Eomesodermin during early mouse development. Mech Dev. 1999; 81:199–203. [PubMed: 10330500]

48. Ralston A, Rossant J. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. Dev Biol. 2008; 313:614–629. DOI: 10.1016/j.ydbio.2007.10.054 [PubMed: 18067887]

49. Savory JG, et al. Cdx2 regulation of posterior development through non-Hox targets. Development. 2009; 136:4099–4110. DOI: 10.1242/dev.041582 [PubMed: 19906845]

50. Donnison M, et al. Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. Development. 2005; 132:2299–2308. DOI: 10.1242/dev.01819 [PubMed: 15829518]

51. Goldin SN, Papaioannou VE. Paracrine action of FGF4 during periimplantation development maintains trophectoderm and primitive endoderm. Genesis. 2003; 36:40–47. DOI: 10.1002/gene. 10192 [PubMed: 12748966]

52. Kang M, Piliszek A, Artus J, Hadjantonakis AK. FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. Development. 2013; 140:267–279. DOI: 10.1242/dev.084996 [PubMed: 23193166]

53. Nichols J, Silva J, Roode M, Smith A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development. 2009; 136:3215–3222. DOI: 10.1242/dev. 038893 [PubMed: 19710168]

54. 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. 2014; 15:545. [PubMed: 25476147]

55. Smallwood SA, et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet. 2011; 43:811–814. DOI: 10.1038/ng.864 [PubMed: 21706000]

56. Ooi SK, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007; 448:714–717. DOI: 10.1038/nature05987 [PubMed: 17687327]

57. He J, et al. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat Cell Biol. 2013; 15:373–384. DOI: 10.1038/ncb2702 [PubMed: 23502314]

58. Wu X, Johansen JV, Helin K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. Mol Cell. 2013; 49:1134–1146. DOI: 10.1016/j.molcel. 2013.01.016 [PubMed: 23395003]

59. Blackledge NP, et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell. 2014; 157:1445–1459. DOI: 10.1016/j.cell. 2014.05.004 [PubMed: 24856970]

60. Boulard M, Edwards JR, Bestor TH. FBXL10 protects Polycomb-bound genes from hypermethylation. Nat Genet. 2015; 47:479–485. DOI: 10.1038/ng.3272 [PubMed: 25848754]

61. Irizarry RA, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009; 41:178–186. DOI: 10.1038/ng. 298 [PubMed: 19151715]

62. Steine EJ, et al. Genes methylated by DNA methyltransferase 3b are similar in mouse intestine and human colon cancer. J Clin Invest. 2011; 121:1748–1752. DOI: 10.1172/JCI43169 [PubMed: 21490393]

63. Schulze I, et al. Increased DNA methylation of Dnmt3b targets impairs leukemogenesis. Blood. 2016; 127:1575–1586. DOI: 10.1182/blood-2015-07-655928 [PubMed: 26729896]

64. Yang L, et al. DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated Leukemias. Cancer Cell. 2016; 29:922–934. DOI: 10.1016/j.ccell.2016.05.003 [PubMed: 27300438]

65. Mayle A, et al. Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. Blood. 2015; 125:629–638. DOI: 10.1182/blood-2014-08-594648 [PubMed: 25416277]

66. Haney SL, et al. Promoter Hypomethylation and Expression Is Conserved in Mouse Chronic Lymphocytic Leukemia Induced by Decreased or Inactivated Dnmt3a. Cell Rep. 2016; 15:1190–1201. DOI: 10.1016/j.celrep.2016.04.004 [PubMed: 27134162]

67. Ben-Porath I, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet. 2008; 40:499–507. DOI: 10.1038/ng.127 [PubMed: 18443585]
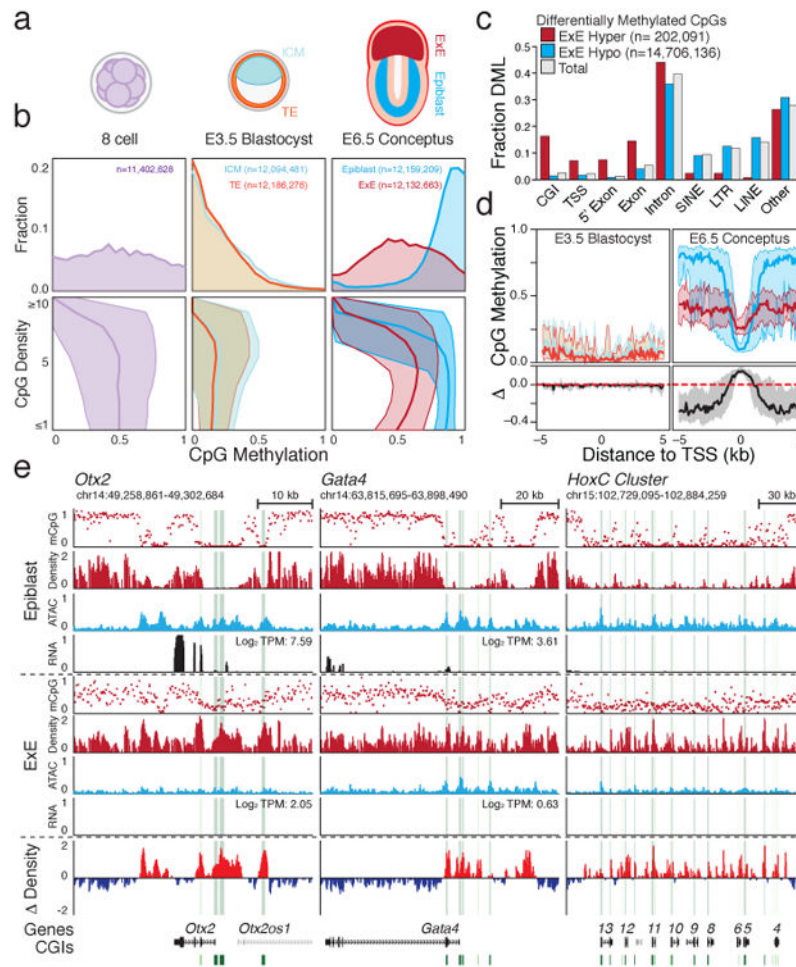
**Figure 1. Divergent postimplantation DNA methylation landscapes**

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 as well as E6.5 Extraembryonic Ectoderm (**ExE**) and **Epiblast** (see Methods).

b. Top: CpG methylation distribution for 100 bp tiles. **Bottom:** Median 100 bp tile methylation as a function of local CpG density. Shaded area represents the 25th and 75th percentiles.

c. Feature level enrichment for differentially methylated CpGs compared to genomic background. ExE-hypomethylated CpGs are prevailing found in non-genic sequences, while ExE-hypermethylated CpGs localize to CpG islands (CGI), Transcription Start Sites (TSS) and 5′ Exons. Here, TSS refers to the 1 kb upstream of an annotated TSS only, while 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 difference (methylation), which diverges considerably upon implantation. Shaded area represents the 25th and 75th percentiles per 100 bp bin.

e. Genome browser tracks for WGBS, 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 E×E specifically (Density refers to the difference compared to epiblast). For *Otx2* and *Gata4*, E×E-specific methylation and repression are concurrent, while the *HoxC* cluster is expressed later in embryonic development. CGIs and annotated TSSs are highlighted in green and red, respectively.
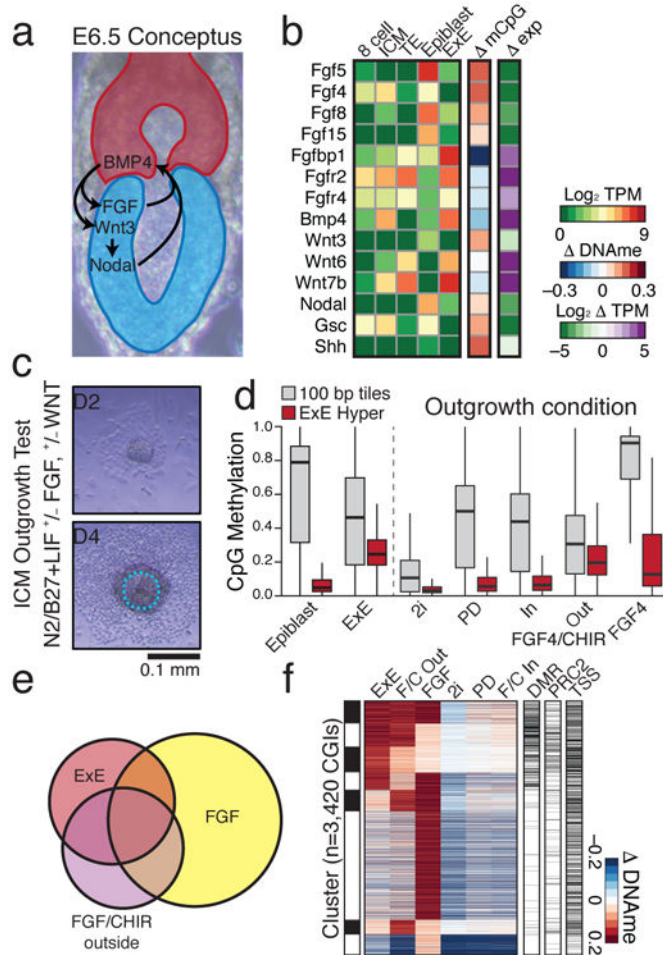
Figure 2. *De novo* methylation of early developmental gene promoters can be modulated by external conditions

a. Schematic of signaling pathway interactions between Epiblast (blue) and E×E (red). Epiblast-produced Fibroblast Growth Factors (FGFs) promote E×E development, which expresses Bone Morphogenic Protein 4 (BMP4) to induce Wingless (WNT) proteins in Epiblast. Epiblast secreted pro-Nodal is processed by the E×E to establish a proximal-distal gradient and the primitive streak[10].

b. Expression and differential promoter methylation of key signaling components. Many FGFs and associated receptors exhibit reciprocal expression and promoter methylation between Epiblast and E×E. Wnt3 induction is apparent in epiblast, while Wnt6 and 7b are highly expressed in both TE and E×E. Differential promoter methylation refers to the annotated TSS (+ or – 1 kb) with the greatest absolute difference (Supplementary Table 2).

c. Schematic for the ICM outgrowth test. ICM outgrowths are cultured for 4 days under disparate growth factor or small molecule conditions intended to either stimulate or repress FGF and WNT activity. Outline highlights the purified component (Methods).

d. Methylation boxplots for the conditions described in c, including all RRBS-captured 100 bp tiles and E×E-targeted CGIs (E×E Hyper CGIs). Edges refer to the 25th and 75th percentiles, whiskers the 2.5th and 97.5th percentiles, respectively.

e. The E×E, 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 Epiblast (n=3,420). The FGF4 condition has the highest number of methylated CGIs, but fewer intersect with E×E than when CHIR is also present.

f. Clustering of differentially methylated CGIs from e, with methylation status in E×E, embryonic regulation by PRC2, and TSS proximity ($^+/_-$ 2 kb) included. 25% of E×E methylated CGIs overlap with both conditions, while 51% overlap with the FGF/CHIR exterior outgrowth.
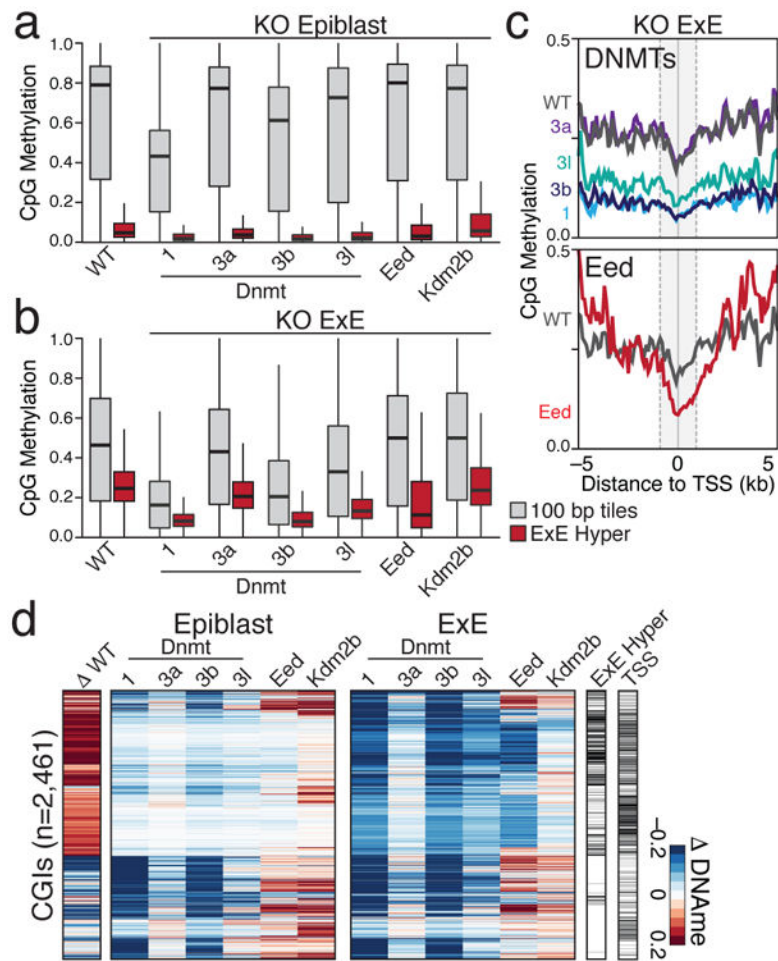
Figure 3. A novel configuration of epigenetic regulators contributes to the extraembryonic methylation landscape

a. Boxplots for E6.5 epiblast tissue for wild type and CRISPR/Cas9 disrupted samples, including for 100 bp tiles and E×E 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 E×E. In comparison to the Dnmt3a and 3b-positive epiblast, *Dnmt1* or *Dnmt3b* disruption have far greater effects on global methylation levels and result in a highly depleted genome.

c. Composite plots of E×E Hyper CGIs by knockout status. CGI methylation is disrupted in *Eed-null* E×E, particularly within +1 kb of the TSS, without affecting global levels. Gray line represents the wild type median. Composite plots map the median of 200 bp windows over 50 bp intervals from RRBS data.

d. Heatmap of the differential CGI methylation (0.1) between CRISPR/Cas9-targeted Epiblast or E×E compared to their wild type counterparts. Differential E×E methylation status and TSS proximity (+/− 2 kb) are included for reference.
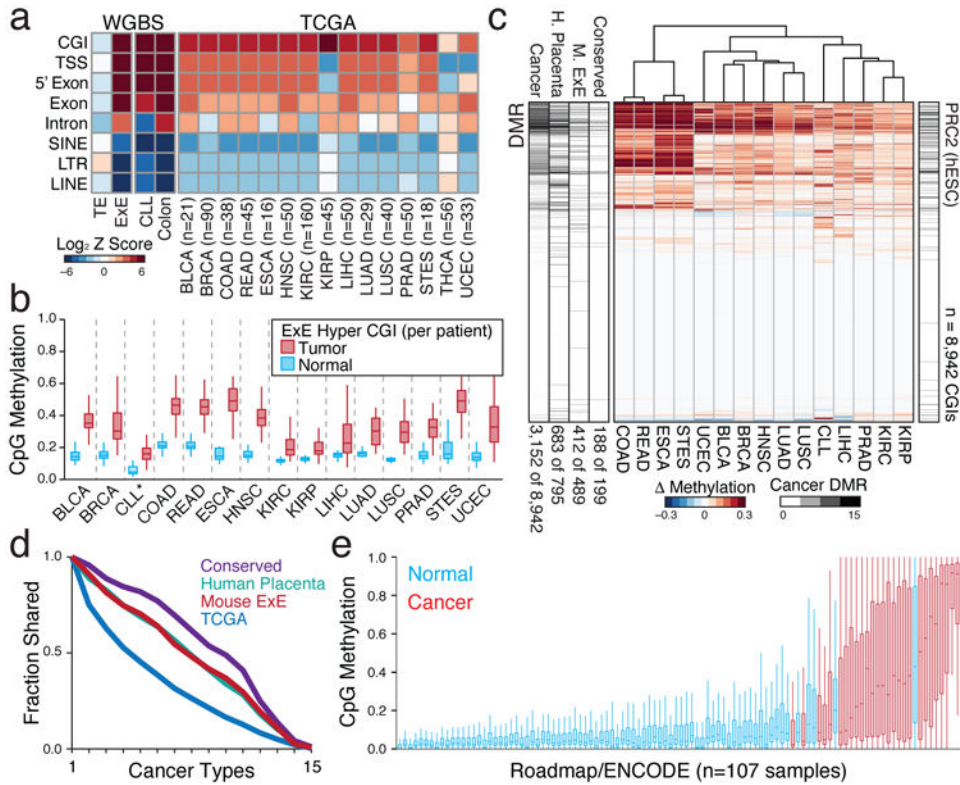
Figure 4. Extraembryonically-targeted CpG islands are pervasively methylated across human cancer types

a. Disruption of global methylation creates similar biases for CGIs and promoters between TE/ICM and E×E/Epiblast or patient- or age-matched normal/tumor tissue comparisons. Heatmap shows the log Z score enrichment for features by the binomial test. Of these 16 cancer types, only THCA does not display a notably dysregulated methylome. N's refer to the number of matched tumor/normal tissue isolates for each type. TCGA samples include Bladder Urothelial Carcinoma (BLCA), Breast Invasive Carcinoma (BRCA), Colon Adenocarcinoma (COAD), Colorectal Adenocarcinoma (READ), Esophageal 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 Esophageal Carcinoma (STES), Thyroid Carcinoma (THCA), and Uterine Corpus Endometrial Carcinoma (UCEC). Here, Chronic Lymphocytic Leukemia (CLL) to B lymphocyte comparison is between age-matched samples measured by WGBS.

b. Feature level boxplots of 489 E×E Hyper CGIs that preserve their status in human, calculated as a feature per tumor or normal tissue for the 15 cancer types where CGI methylation is generally apparent. Note: 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 heatmap for 8,942 orthologous CGIs measured in TCGA or by RRBS and clustered by Euclidean Distance. DMR bar includes the cumulative number of cancers a given island is called as differentially methylated, as well as the DMR status in

either human placenta compared to human embryonic stem cells (hESCs), mouse E×E compared to Epiblast, or shared between both comparisons (Conserved). PRC2 (hESC) denotes regulation by polycomb in hESCs. 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 E×E 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. Inter-tumor 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 E×E Hyper CGIs across 107 ENCODE/Roadmap 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. 9 or Supplementary Table 7 includes additional sample characteristics.