Supplementary Information:

Supplementary Methods

Preparation of samples and genome-scale libraries

Reduced representation and alignment was performed essentially as described [3]. Gamete and zygotic data sets were taken from a parallel study where they contribute to a complete pre-implantation timeline [2] but in this study were used exclusively as an *in vivo* counterpart to examine the global epigenetic consequences of NT on somatic cells.

Our comparative oocyte-based reprogramming data set consisted of donor fibroblasts and those same donor fibroblasts submitted to nuclear transfer. Donor fibroblasts were isolated of tail tips from two BDF x CAST/EiJ or two 129X1/SvJ donor males. BDF1 oocytes for NT were obtained 13 h post hCG injection and the spindle was removed under 5 μg/ml cytochalasin B. Donor fibroblasts were injected into enucleated oocytes and the meiosis I polar body was eliminated as a 1x genomic contaminant via manual XY Clone Infra-red laser assisted biopsy (Hamilton Thorne), as were meiosis I/II polar bodies in fertilized zygotes. Activation was done 1-3 hours post injection by 1h incubation in 1 mM $SrCl_2$ in Ca-free MZCB supplemented with cytochalasin B, followed by 4h incubation in cytochalasin B only. Embryos were then cultured in KSOM-AA (Chemicon) and collected 12-14 hours after the onset of $SrCl_2$ (hours post activation, hpa) in accordance with later zygotic stages, and before completion of the first interphase. Zygotes used for comparison were staged within pronuclear stages 2 and 4, after the reported completion of active DNA demethylation by immunohistochemical detection of methylated cytosine[4].

Analysis

For 100 bp tiles, reads for the CpGs that were covered more than 5x were pooled and used to estimate the methylation level by taking the number of reads reporting a C, divided by the total number of reads reporting a C or T. Outliers were removed using Dixon's Q-test with a confidence level of 90% for tiles that had more than 3 replicate values before calculating mean methylation values for a sample. The CpG density for a 100 bp tile is the average number of CpGs within 50 bp of contributing CpGs.

High, intermediate, and low CpG density promoter (HCP, ICP, LCP), and ICR annotations were taken from [1]. LINE, LTR, and SINE annotations were downloaded from the UCSC browser (mm9) RepeatMasker tracks. The methylation level of an individual feature is estimated by pooling read counts for CpGs within the feature that are covered more than 5x, and levels are reported if a feature contains ≥5 CpGs with such coverage.

Calculation of paternal methylation levels in zygote

Zygote methylation is the mean methylation of the maternal and paternal genomes:

Zygote=0.5(Oocyte)+0.5(Sperm)

Assuming methylation on the oocyte genome does not change in the zygote:

Zy(Sp)=(Zygote-0.5(Oocyte))/0.5= 2*Zygote – Oocyte

The variance for Zy(Sp) is then:

Var(Zy(Sp)) = 4*Var(Zygote) + Var(Oocyte)

Methylation values for Zy(Sp) are estimated using the mean methylation value of the zygote and oocyte samples. Predicted values <0 and >1 are set to their respective boundaries if that boundary is within one standard deviation for that locus.

We are able to track 565 methylation values of individual CpGs using SNPs at 10x paternal coverage in BDF1xX1 zygotes. The Pearson correlation coefficient between the inferred sperm in zygote value and the SNP tracked paternal methylation value is 0.8.

The inferred paternal methylation values in 2-cell embryos are calculated in the same manner by swapping 2-cell methylation values for zygote methylation values.

Estimation of residual host oocyte DNA in NT embryos

We used the proportion of SNPs that showed any recipient oocyte allele as an estimate for the proportion of loci affected by contamination in 129X1 NT embryos. SNPs affected by contamination are represented by the left peak of the bimodal log odds distribution and contain 40% and 32% of the SNPs for each replicate respectively (**Supplementary Fig. 1c**). To estimate the magnitude of contamination, we binned all SNPs that showed any contamination by their coverage into 5x intervals. As coverage increases, we should achieve a better estimate of the true contamination level and indeed, we see contamination leveling out around 16% and 14% for each replicate, respectively. This works out to 0.4*0.16=0.064 total contamination. We believe the true level of contamination is lower as this estimate is confounded by amplification noise. Estimating contamination in BDF1/Cast NT embryos is more difficult as the recipient BDF1 oocyte cannot be readily distinguished from the donor genome. To estimate the proportion of loci affected by contamination, we assume that the distribution for an ideal, uncontaminated sample follows the same shape but is shifted such that its peak is over a log odds ratio of zero. Then, the estimate of loci affected is the density of SNPs that remain with a log odds ratio less than zero after the density of the ideal distribution has been subtracted. This is ~13% for both BDF1/Cast NT embryo replicates. To estimate the average magnitude of contamination for these samples, we take the weighted mean level of contamination for the affected loci, which is ~33% for both replicates. This results in an overall contamination of 0.13*0.33=0.043.

Analysis of dynamic genomic feature sets

In order to equalize the statistical power across feature sets that had different numbers of elements, we subsampled the feature set with replacement 1000 times and took the average p-value from the Wilcoxon signed rank test on the mean value across replicates using 150 elements, which is the approximate coverage of SINE elements. We used the Bonferroni correction with FDR=0.05 to identify significantly changing features in each comparison. We considered accounting for potential oocyte DNA contamination by modifying the fibroblast methylation levels to include the appropriate proportion of oocyte methylation (thereby creating the appropriate background to measure NT embryo dynamics against) but decided that both overestimation or underestimation of contamination levels would lead to biased results.

Analysis of dynamic promoters

Promoters are considered dynamic if they show a change >0.2 between the means of the replicates in fertilization (Sp – Zy(Sp)) or NT (Fib– NT(Fib)). A promoter is robustly changing during fertilization if the p-value from a one-tail t-test comparing the Zy(Sp) to the sperm values is significant after multiple hypothesis test correction using the Benjamini-Hochberg method. The samples in NT are paired such that each donor has a matching NT embryo except in the case of the two 129X1 replicates where the fibroblasts from the same animal were used for both NT embryos (X1 tail tip replicate 1). Nonetheless, we assume that the two replicates are independent because the samples were cultured independently and subjected to separate rounds of nuclear transfer. Promoters must satisfy three criteria in order to be considered robustly changing in NT. First, promoters must change >0.2 in all replicates measured and must be present in at least three replicates when the change is calculated on raw values. For a given replicate, a change >0.2 corresponds to an empirical p-value of 0.0372 - 0.0479 when the null

distribution is determined by permuting the relationship between promoters in each donor-NT embryo pair. The null distribution for each pair was generated 1000 times and the average proportion of promoters >0.2 was taken to determine the empirical p-value. The requirement of change in three replicates corresponds to a p-value of $9.7*10^{-5}$. Unfortunately, this would not be significant after multiple hypothesis correction but we did not want to limit our results due to missing values. This results in 49 promoters. Secondly, we were concerned that oocyte contamination would bias results towards demethylation so we require promoters to change >0.2 in contamination-normalized 129X1 replicates. They must change in all replicates measured and must be present in one of the replicates. To normalize the data for comparison, we modified all fibroblast promoter methylation levels to include an appropriate amount of oocyte contamination (16% and 14% in each replicate respectively). For example, Fibroblast(normalized)= (1-0.16)*Fibroblast + 0.16*Oocyte.

An empirical p-value for change >0.2 in contamination-normalized comparisons is 0.0344-0.0412. We now have 31 promoters. Finally, if Cast allelo-typed CpG methylation data at 5x exists for the promoter, then it must also show the same dynamic. This supported two promoters and eliminated two other promoters. Promoters were then assigned to three categories (NT specific, fertilization specific, or shared) depending on their dynamics. In order to be shared, a promoter must be robustly changing in both NT and fertilization. DAVID was used to ascribe functional enrichment [5].

Identification of SNPs

Known SNPs between 129X1 and BDF1 (C57BL/6N × DBA/2) and Cast and BDF1 were taken from mouse genome resources[6,7]. The sets were filtered such that SNPs that fell into the following categories were removed: (1) SNPs that had inconsistent entries for the same position; (2) SNPs not trackable by RRBS (C/T or A/G); (3) SNPs between

C57BL/6N and DBA/2; and (4) SNPs that were not covered by X1 and BDF1, or Cast and BDF1 in an *in silico* digest. The log odds ratio [$\log_2$(X1 count +0.01/C57 count+0.01)] was calculated for each SNP that was covered in the data set.

Parent-of-origin methylation tracking

Reads were segregated into either Cast or BDF1 according to SNP type, and CpG methylation levels were called in the same manner described above. SNP normalized methylation values (**Supplementary Fig. 1d**) are the average of the methylation values derived from each haplotype.

Data availability

RRBS data is deposited at the Gene Expression Omnibus under accession number GSE38711.

References:

1 Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770 (2008).

2 Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, doi:10.1038/nature10960 (2012).

3 Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature protocols* **6**, 468-481, doi:nprot.2010.190 [pii] 10.1038/nprot.2010.190 (2011).

4 Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* **241**, 172-182. (2002).

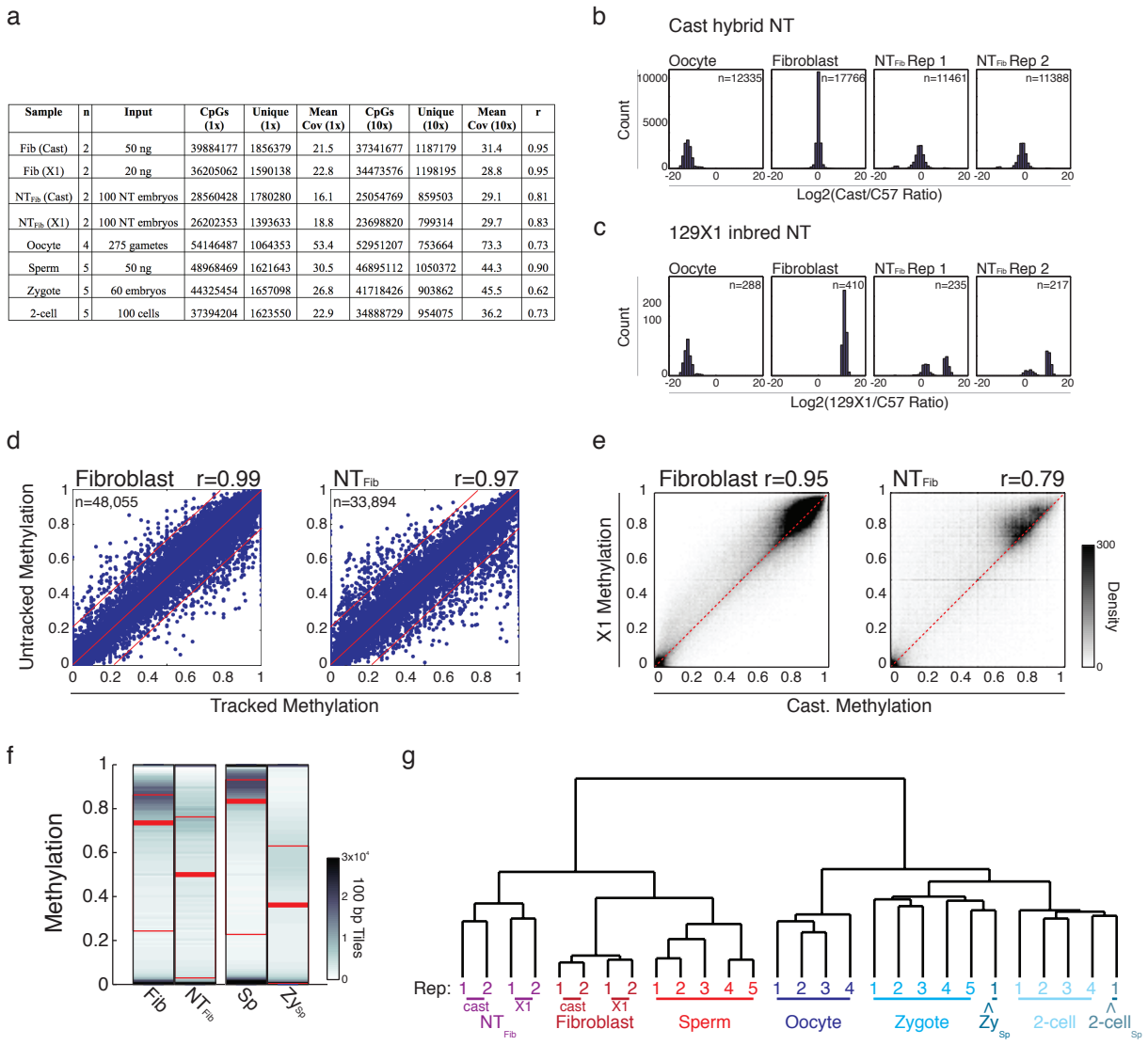
5 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:nprot.2008.211 [pii] 10.1038/nprot.2008.211 (2009).

6 Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E. & Eppig, J. T. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res* **39**, D842-848, doi:gkq1008 [pii] 10.1093/nar/gkq1008 (2011).

7 Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-294, doi:nature10413 [pii] 10.1038/nature10413 (2011).

a

Sample	n	Input	CpGs (1x)	Unique (1x)	Mean Cov (1x)	CpGs (10x)	Unique (10x)	Mean Cov (10x)	r
Fib (Cast)	2	50 ng	39884177	1856379	21.5	37341677	1187179	31.4	0.95
Fib (X1)	2	20 ng	36205062	1590138	22.8	34473576	1198195	28.8	0.95
NT_{Fib} (Cast)	2	100 NT embryos	28560428	1780280	16.1	25054769	859503	29.1	0.81
NT_{Fib} (X1)	2	100 NT embryos	26202353	1393633	18.8	23698820	799314	29.7	0.83
Oocyte	4	275 gametes	54146487	1064353	53.4	52951207	753664	73.3	0.73
Sperm	5	50 ng	48968469	1621643	30.5	46895112	1050372	44.3	0.90
Zygote	5	60 embryos	44325454	1657098	26.8	41718426	903862	45.5	0.62
2-cell	5	100 cells	37394204	1623550	22.9	34888729	954075	36.2	0.73



Supplementary Figure 1: Genome-scale methylation profiling of nuclear transfer embryos

a. CpG coverage captured by reduced representation bisulfite sequencing (RRBS)[1]. Shown are the number of replicates (n) and mean input for each sample, the numbers of total and unique CpGs, and their mean coverage (Cov) at 1X or 10X. Mean replicate Pearson correlation coefficient (r) is included for each sample. Oocyte, sperm, zygote, and 2-cell data is from Ref 2.

b. Log odds ratio histograms for allelic frequencies for C57/B6 and CAST/EiJ SNPs for experimental round 1 with BDF1xCast hybrid tail tip fibroblasts. The shift of the log odds ratio away from 0 suggests possible recipient oocyte contamination (~13%) in the NT_{Fib} sample. n is the number of SNPs captured for each sample.

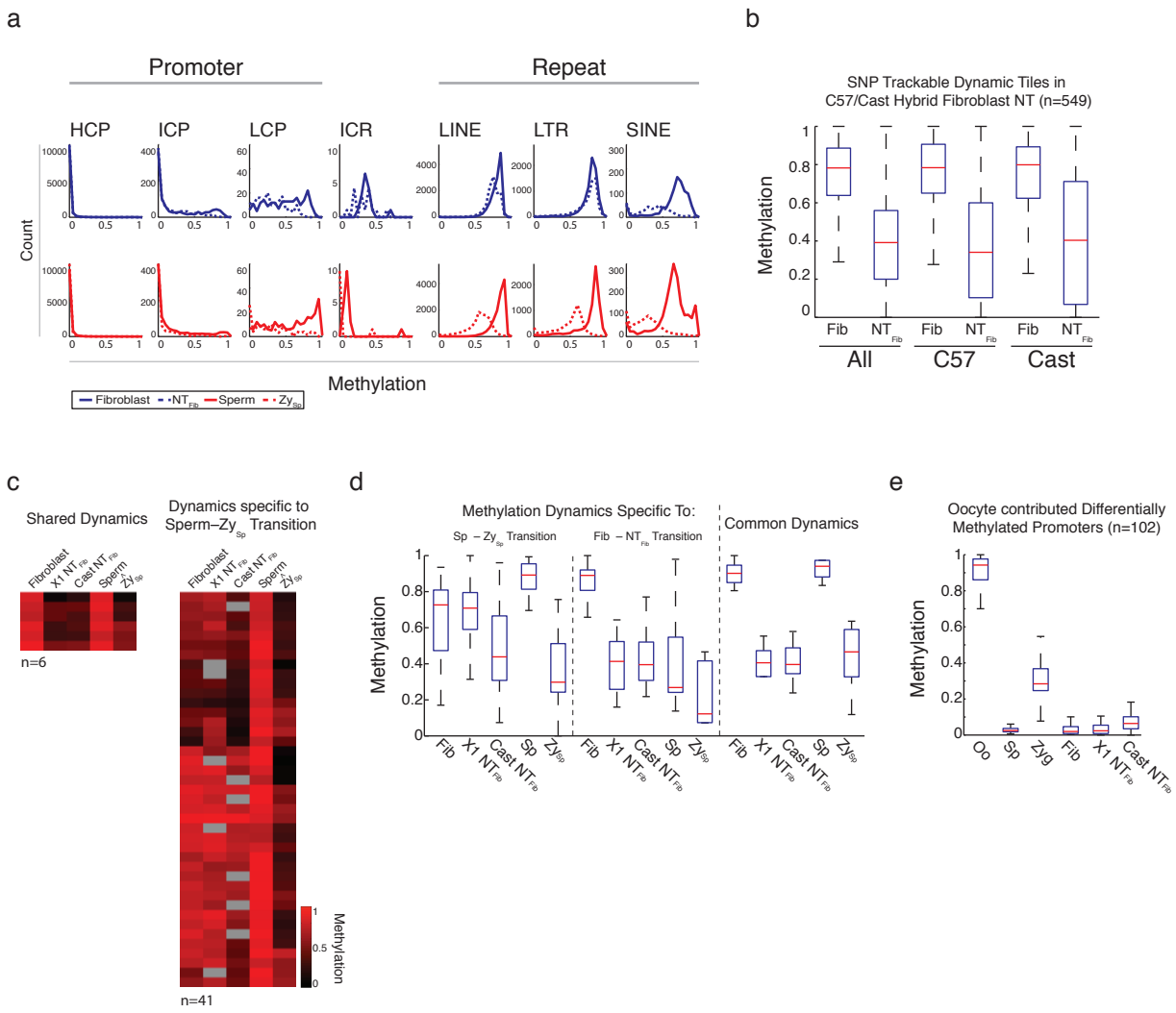
c. Log odds ratio histograms for allelic frequencies for C57/B6 and 129X1 SNPs for experimental round 2 using 129X1 inbred tail tip fibroblasts. The bimodal distribution shows recipient oocyte contamination affected 40% and 32% of loci in each NT_{Fib} replicate, respectively. The level of contamination is low (~15%, see Supplementary Methods). n is the number of SNPs captured for each sample.

d. Methylation values of single CpGs that can be allelotyped according to either a C57 or Cast background show a high concordance between the untracked calculation and the allele-normalized methylation value.

e. Scatterplot of methylation values for 100 bp tiles between donor fibroblasts and NT embryos of each experimental round and strain confirm the reproducibility of both RRBS and the NT assay even at minimal (~100 cell) inputs. The methylation levels in X1 NT_{Fib} are slightly higher than in BDF1/Cast NT_{Fib} which is evident by the cloud above the red line which represents equal methylation.

f. Boxplot overlaid on a density heatmap showing the distribution of methylation values for 100 bp tiles for donor fibroblast, SCNT reconstructed embryos, sperm and the paternal value in zygote (Zy_{sp}). There is retention of more methylation after nuclear transfer than during fertilization. Bold red bar indicates the median, red line the 25th/75th percentile and outer border the 2.5th/97.5th percentile.

g. Sample clustering using Euclidean distance on methylation values for 100 bp tiles shows that NT embryos most closely resemble, but are distinct from, donor fibroblasts. This trend holds regardless of the experimental round. Raw zygote values are used in clustering in addition to inferred Zy_{sp} values used in analysis. 2-cell and inferred paternal values for 2-cell embryos (2-cell_{sp}) cluster with their respective biological samples.



Supplementary Figure 2: Genomic and promoter dynamics in nuclear transfer

a. Histogram of methylation values for genomic features. Features show very similar global dynamics during fertilization and NT except for LINE and LTR repetitive elements.

b. Boxplots for SNP trackable CpGs (n=549) falling into dynamic 100 bp tiles (>0.2 change) after NT in BDF1xCast hybrid fibroblast experiments. "All" refers to the SNP normalized methylation value before segregation into either C57 or Cast allelotypes.

c. Heatmap of promoter dynamics that are shared in fertilization and NT and those that are fertilization specific. Rounds of NT using different background strains are shown separately to highlight the fidelity of these trends. Grey represents missing values.

d. Boxplot of promoter methylation for the dynamic sets. Promoters specific to the sperm-zygote transition are mostly hypermethylated in fibroblasts and retained higher methylation levels after NT. Similarly, promoters specific to NT are predominantly hypomethylated in sperm.

e. The early embryo receives methylation information from the oocyte genome in the form of 102 hypermethylated promoter DMRs. These methylated regions are pre-implantation specific and erased upon embryonic specification. They remain hypomethylated in both fibroblasts and in NT_{Fib} experiments.

Supplementary Table 1. Promoter methylation values for SCNT and fertilization data sets