

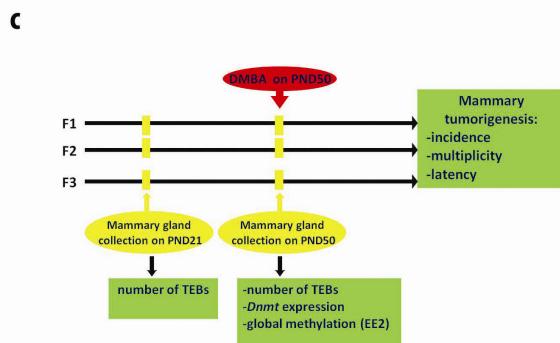
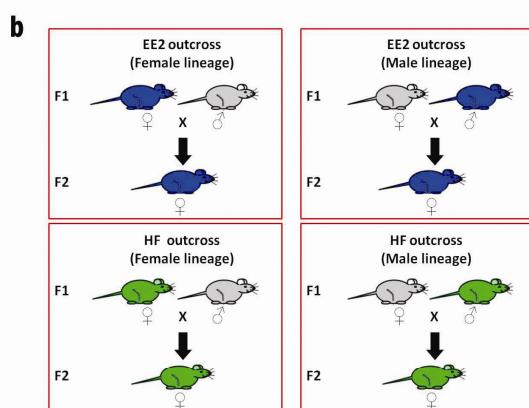
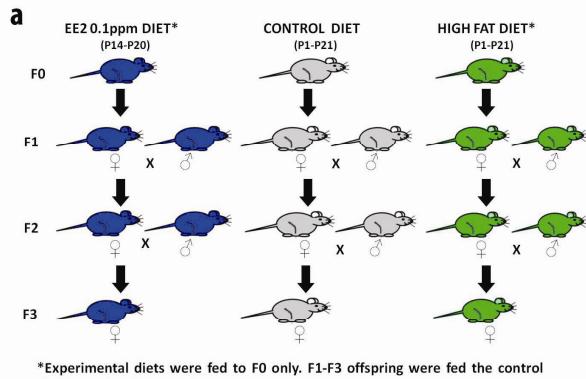
Supplementary Information

for

**High-fat or ethinyl-oestradiol intake during pregnancy increases
mammary cancer risk in several generations of offspring**

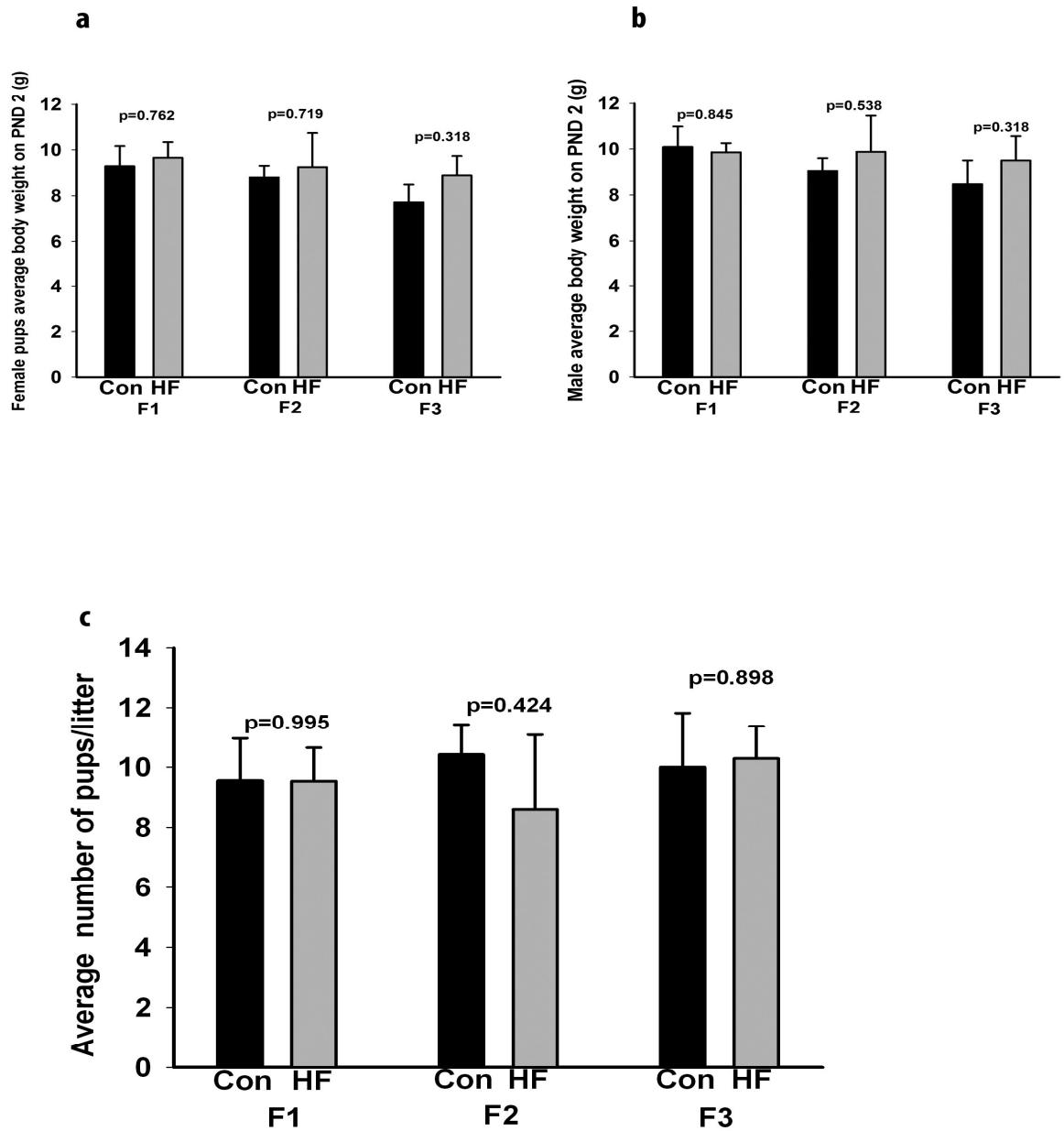
Sonia de Assis, Anni Warri, M. Idalia Cruz, Olusola Laja, Ye Tian, Bai Zhang, Yue Wang, Tim Hui-Ming Huang and Leena Hilakivi-Clarke

Supplementary Figures

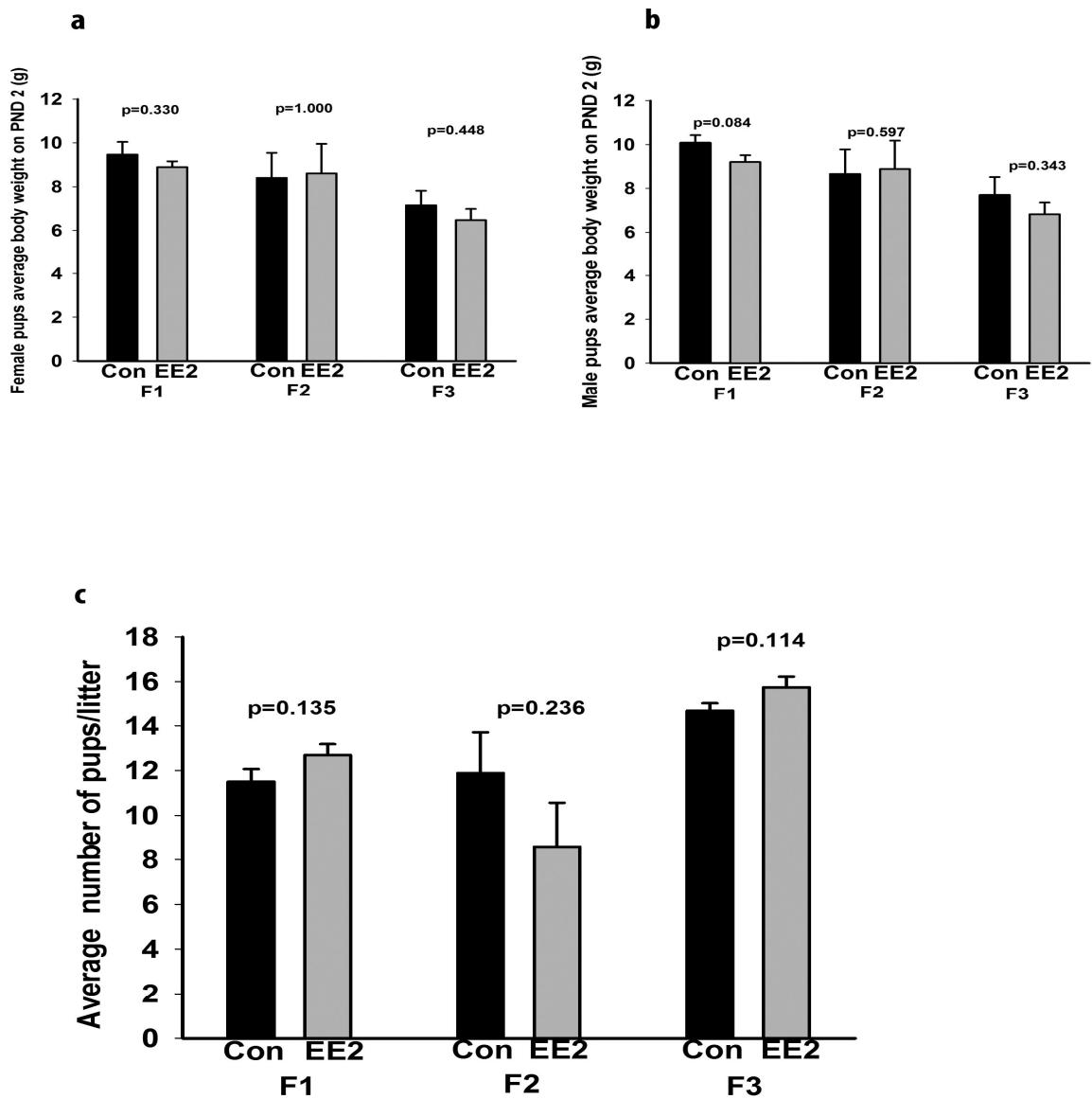


Supplementary Figure S1. Study design. **a**, Pregnant Sprague-Dawley rat dams (F0, n=10) were fed one of the following experimental diets: High Fat (HF), EE2-supplemented (EE2) or control. The HF group was fed this experimental diet throughout pregnancy while the EE2 group

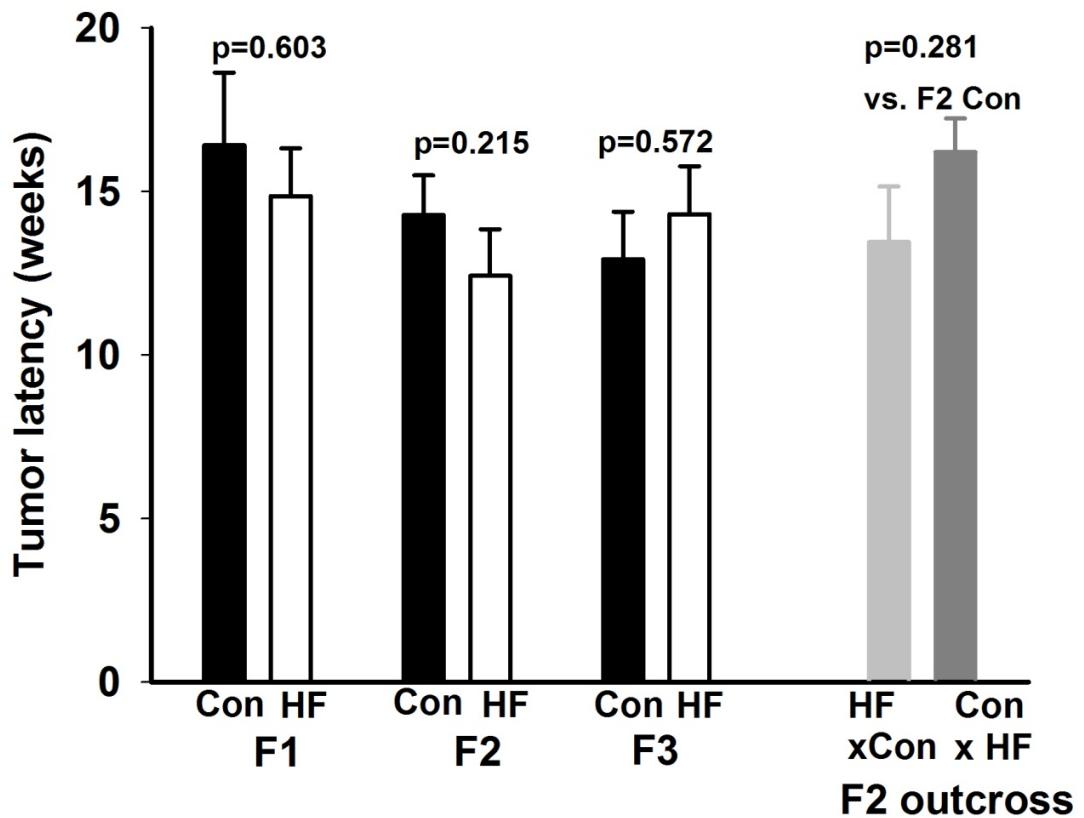
was fed this experimental diet from pregnancy day 14 to 20. F1 and F2 HF or EE2 exposed females were mated on PND 60 with males from the same group to produce the F2 and F3 generations. All F1 and F2 pregnant dams were fed control diet for the extent of pregnancy. **b**, Outcross experiments were performed by mating F1 females or males exposed to EE2 or HF to control males or females. All F1 pregnant dams were fed control diet for the extent of pregnancy. **c**, Mammary gland tissues were collected on PND21 and PND50 for morphological analysis, DNA and mRNA extraction. Mammary tumors were induced on PND50 in F1, F2 and F3 generation females by administration by oral gavage of 10 mg of 9,12-dimethylbenz[a]anthracene (DMBA). Rats were examined for mammary tumors by palpation once per week, starting on 3 weeks post-DMBA and continued for 20 weeks post-DMBA.



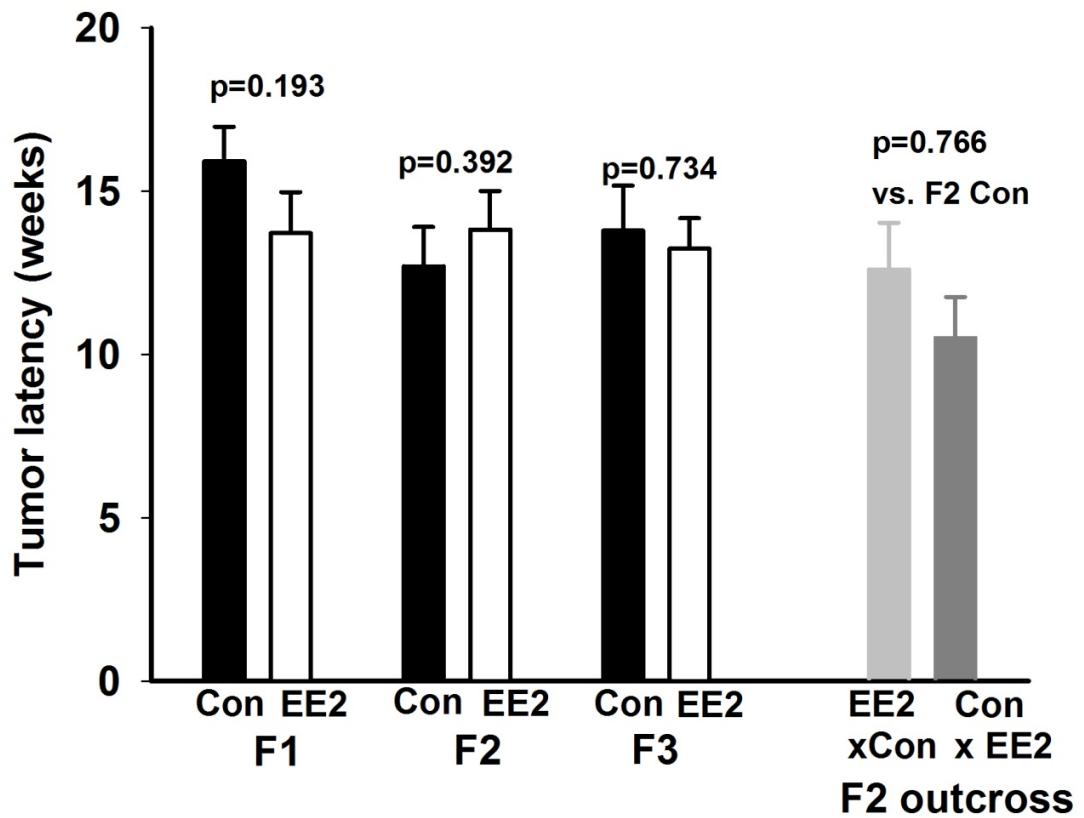
Supplementary Figure S2. HF offspring body weight by gender and litter size in F1-F3 generations. **a-b,** Body weight (g) by gender of pups on PND2 offspring **c,** Number of pups per litter in **F1**(Control: n=11, HF: n=15), **F2**(Control: n=12, HF: n=5) and **F3**(Control: n=8, EE2: n=7). All data are mean \pm s.e.m. Significant differences versus the control group by t-test. $p<0.05$ is considered significant; exact p values are shown in each plot



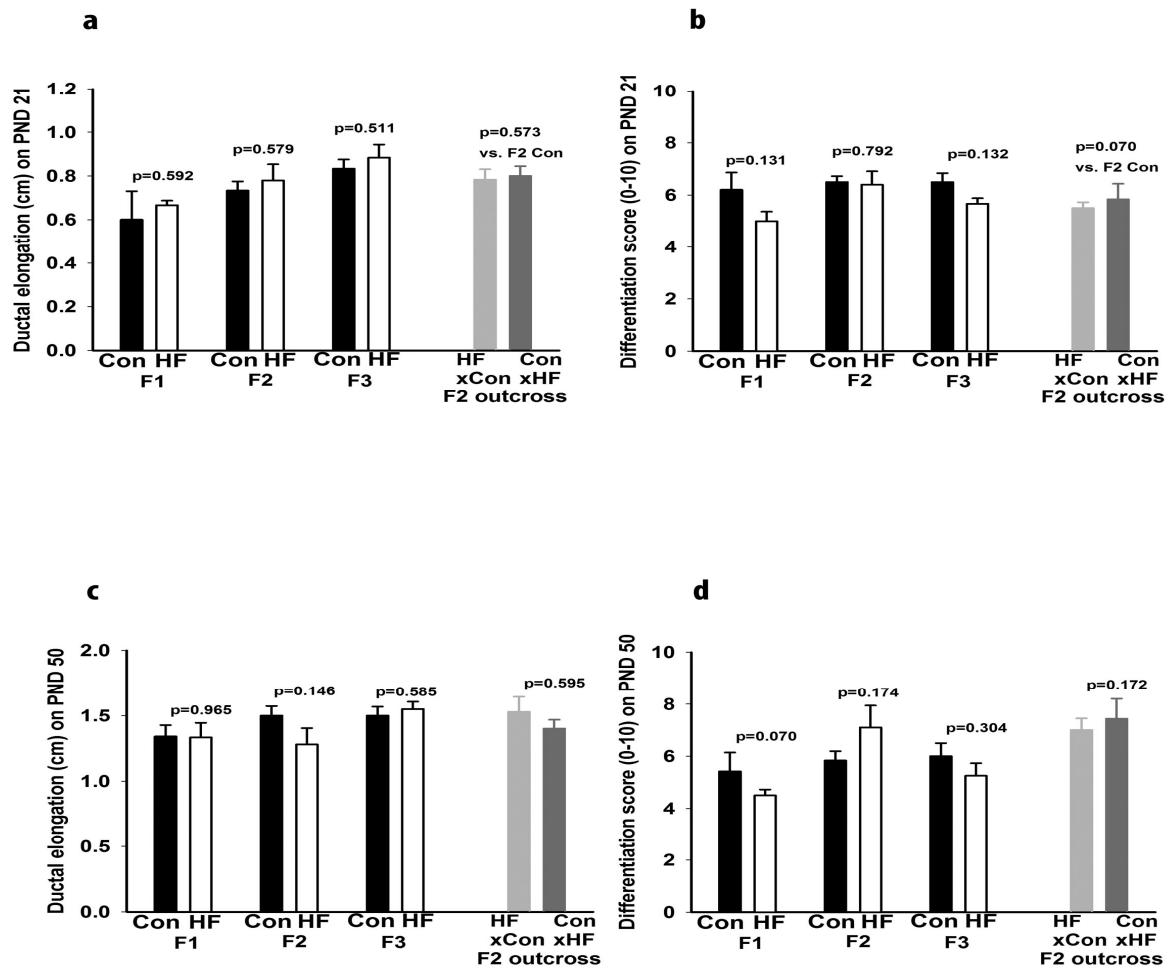
Supplementary Figure S3. EE2 offspring body weight by gender and litter size in F1-F3 generations. **a-b**, Body weight (g) by gender of pups on PND2 offspring **c**, Number of pups per litter in **F1**(Control: n=8, EE2: n=13), **F2**(Control: n=10, EE2: n=11) and **F3**(Control: n=6, EE2: n=6). All data are mean \pm s.e.m. Significant differences versus the control group by t-test. $p<0.05$ is considered significant; exact p values are shown in each plot.



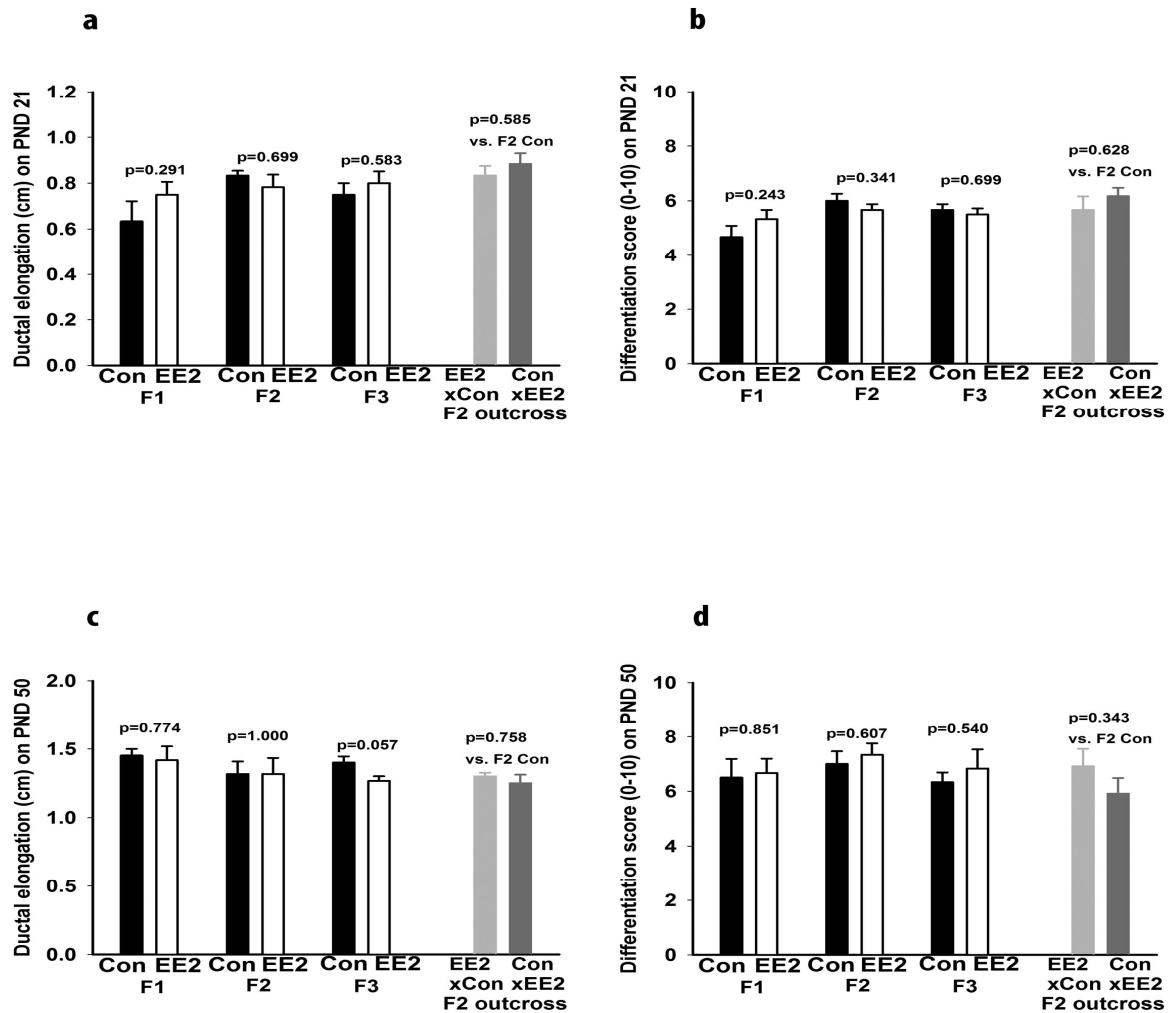
Supplementary Figure S4. Tumor latency (weeks) in HF offspring in F1-F3 generations.
Mammary tumor latency (mean \pm s.e.m.) in **F1**(Control: n=7, HF: n=18), **F2**(Control: n=11, HF: n=12) , **F3**(Control: n=12, HF: n=10) and F2 outcross groups(Control: n=11, HFxControl: n=9, ControlxHF: n=15. All data are mean \pm s.e.m. Significant differences versus the control group were determined as follows: t-test and one-way ANOVA followed by Dunn's *post-hoc* test (**outcross groups**). $p<0.05$ is considered significant; exact p values are shown in each plot.



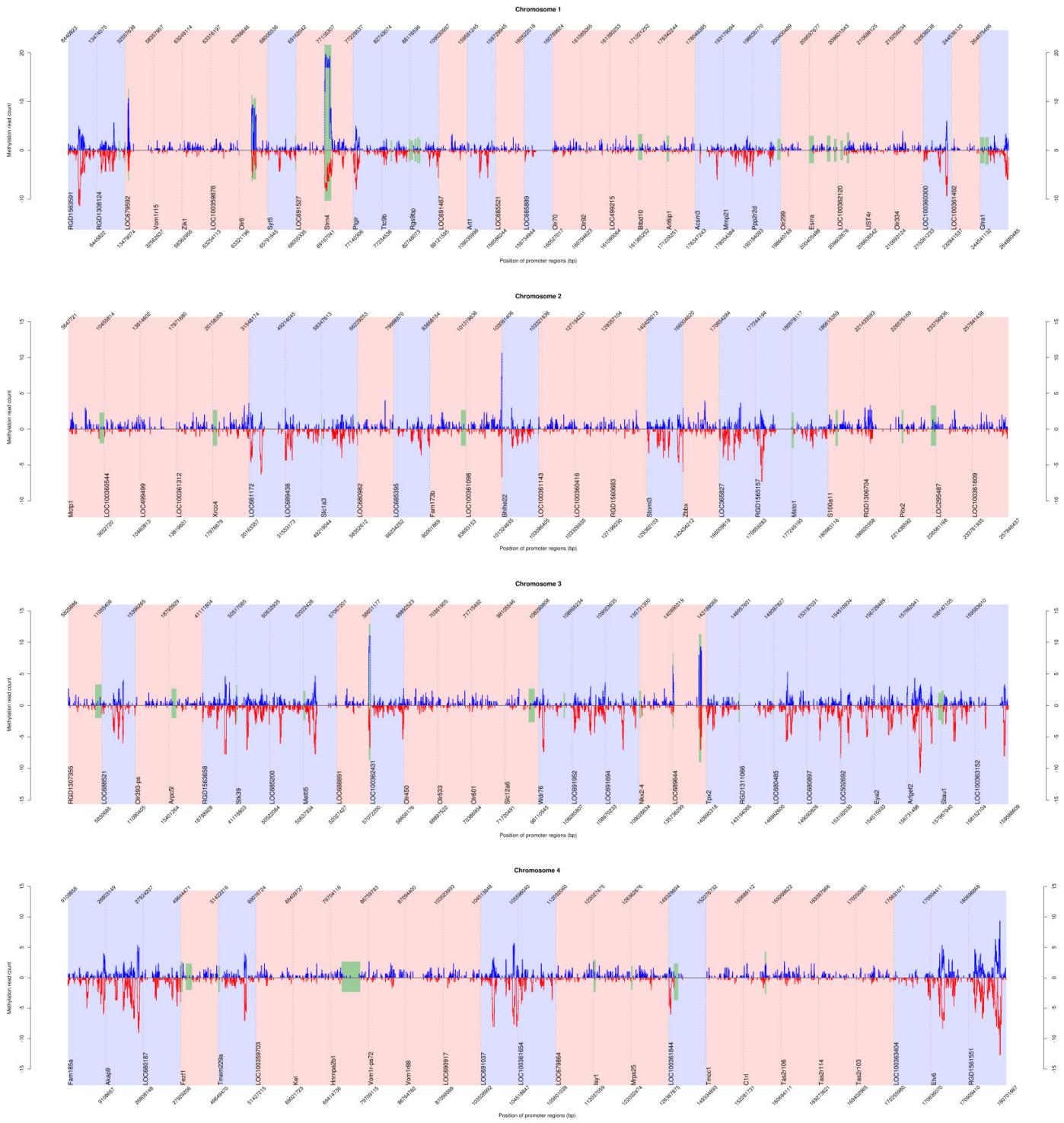
Supplementary Figure S5. Tumor latency (weeks) in EE2 offspring in F1-F3 generations. Mammary tumor latency (mean \pm s.e.m.) in **F1**(Control: n=15, EE2: n=14), **F2**(Control: n=13, EE2: n=11), **F3**(Control: n=14, EE2: n=21) and **F2** outcross groups (Control: n=24, EE2xControl: n=16, ControlxEE2: n=12) and tumor multiplicity (Control: n=13, HFxControl: n=10, ControlxHF: n=4). All data are mean \pm s.e.m. Significant differences versus the control group were determined as follows: t-test and one-way ANOVA followed by Dunn's *post-hoc* test (**outcross groups**). $p < 0.05$ is considered significant; exact p values are shown in each plot.

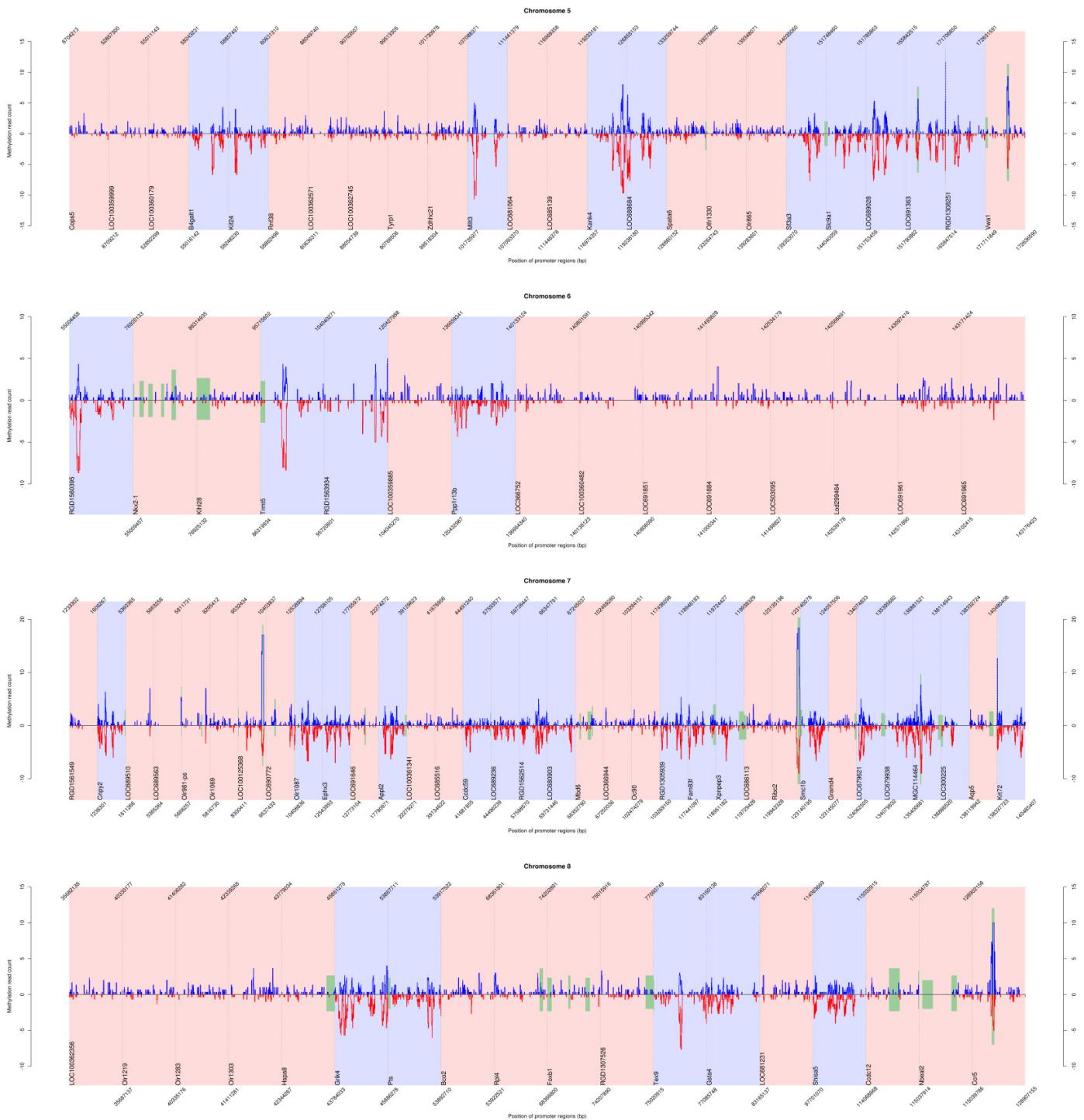


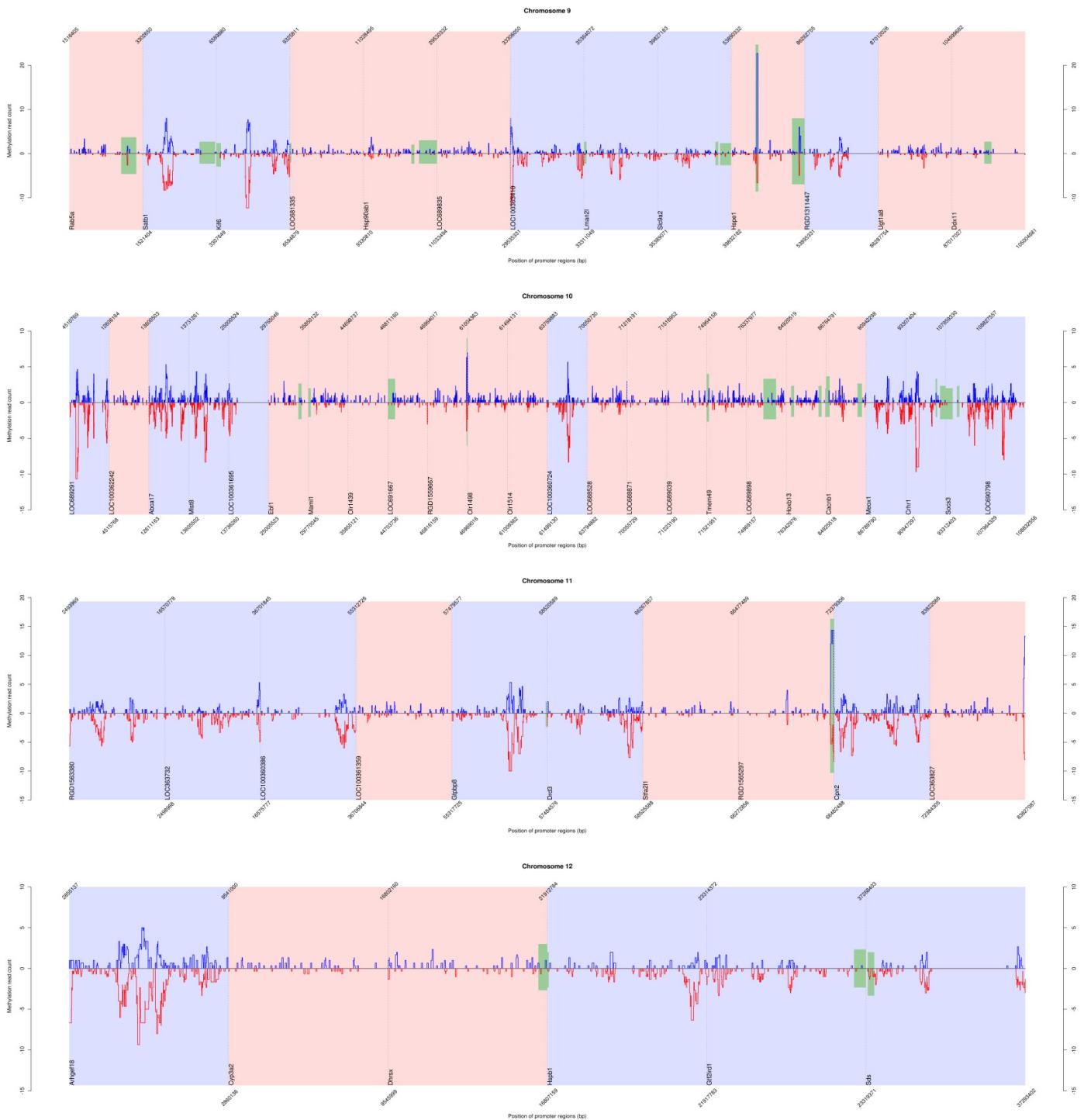
Supplementary Figure S6. Mammary gland ductal elongation (cm) and differentiation (score 0-10) in F1-F3 generation HF female offspring. **a, c**, Ductal elongation **b,d** and Differentiation score on PND 21 (**F1**,Control, HF: n=5), (**F2**,Control: n=6, HF: n=5; outcrosses: n=6) and (**F3**, Control, HF: n=5) and PND 50 (**F1**, Control, HF: n=5), (**F2**, Control: n=6, HF: n=5, outcrosses: n=5) and (**F3**, Control: n=6, HF: n=5) . All data are mean \pm s.e.m. Significant differences versus the control group were determined as follows: t-test and one-way ANOVA (**outcross groups**) followed by Holm-Sidak *post-hoc* test. *p*<0.05 is considered significant; exact *p* values are shown in each plot.

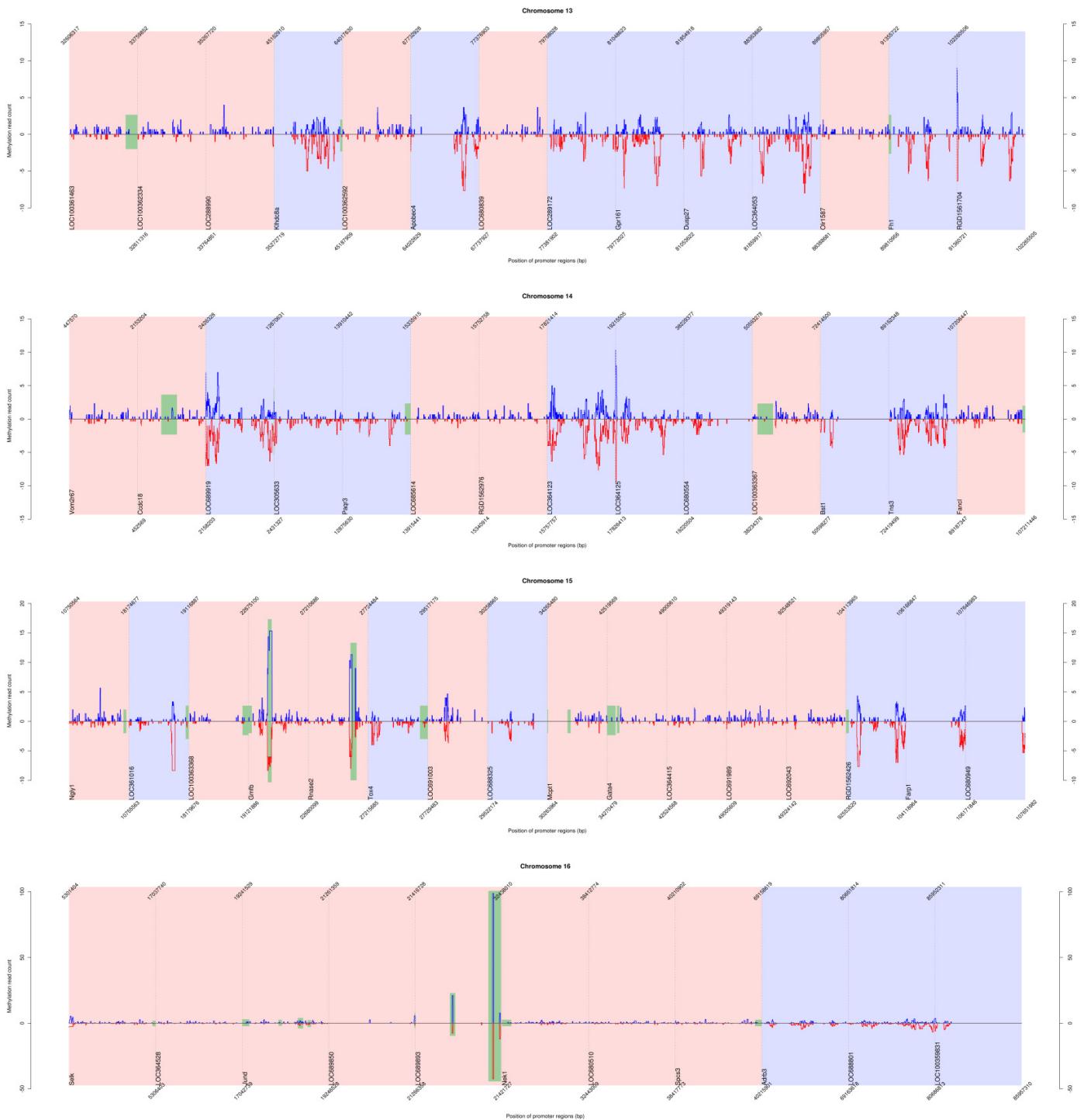


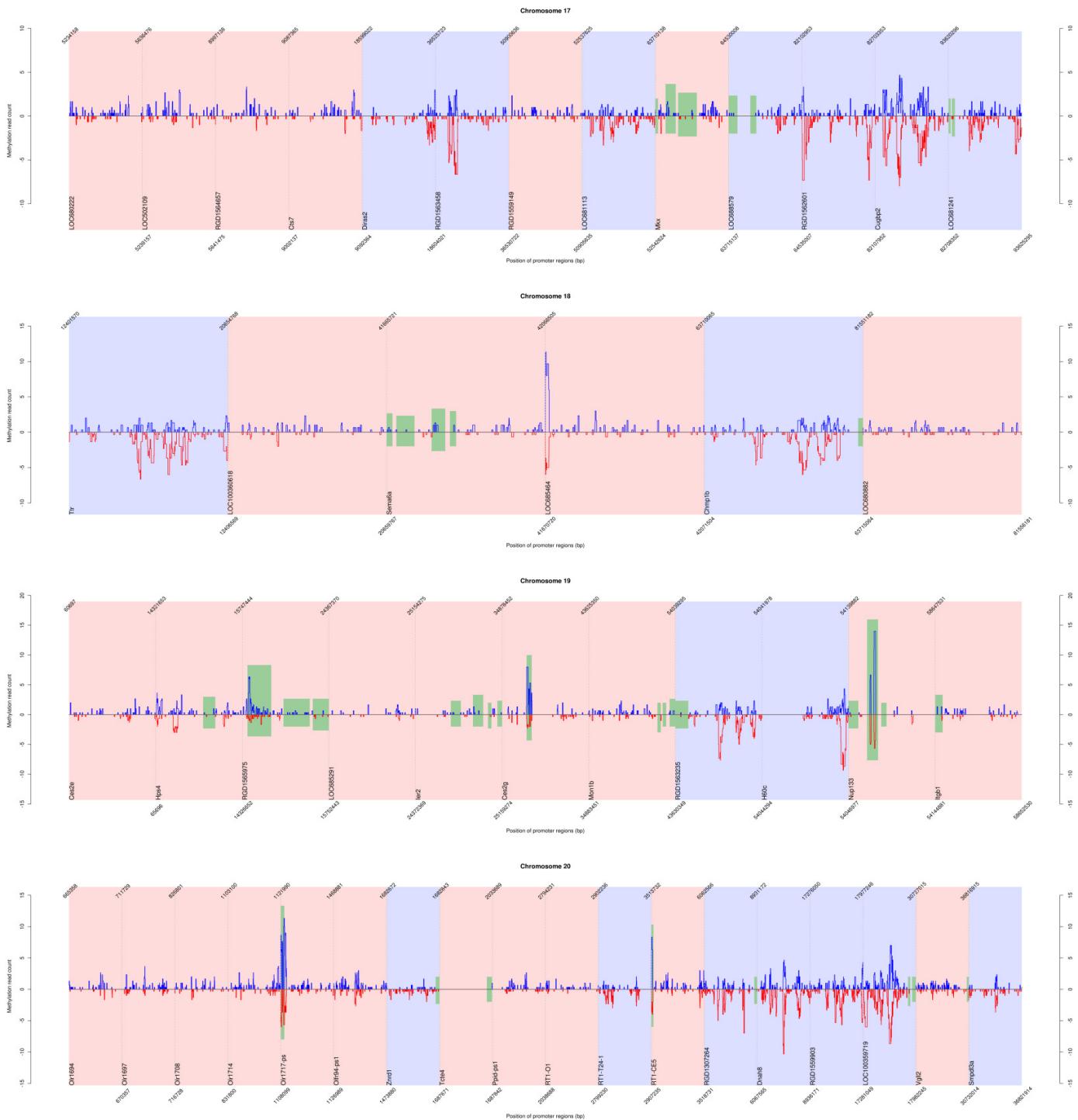
Supplementary Figure S7. Mammary gland ductal elongation (cm) and differentiation (score 0-10) in F1-F3 generation EE2 offspring . a, c, Ductal elongation b,d and Differentiation score on PND 21 (F1, Control: n=5, EE2: n=6), (F2, Control: n=6, EE2: n=5; outcrosses: n=5) and (F3, Control, EE2: n=6) and PND 50 (F1, Control, EE2: n=6), (F2, Control, EE2: n=5; outcrosses: n=6) and (F3, Control: n=5, EE2: n=6). All data are mean \pm s.e.m. Significant differences versus the control group were determined as follows: t-test and one-way ANOVA (outcross groups**) followed by Holm-Sidak *post-hoc* test. p<0.05 is considered significant; exact p values are shown in each plot.**

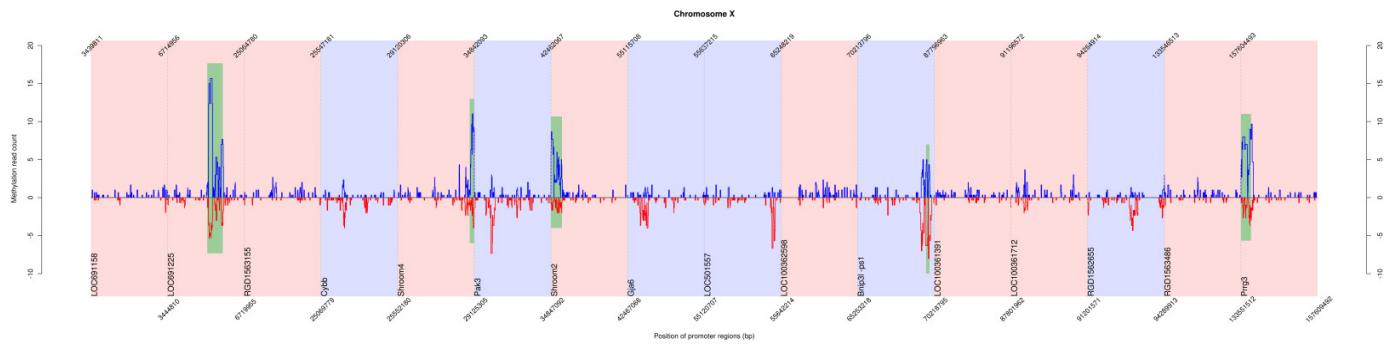












Supplementary Figure S8. Chromosomal distribution of transgenerational change in DNA methylation in promoter regions identified by MBDCap-seq analysis.

To visually compare the inherited methylation changes of EE2 (n=3) and control (n=3), we plotted the methylation profiles of the 375 significant promoter regions sorted by 21 chromosomes. The chromosomes are represented by a compact view by concatenating the significant promoters while leaving remaining genomic regions out to obtain a distinguishable resolution. The horizontal axis index is base pair, and the vertical axis index is methylation count. Promoters are separated by dashed vertical lines. The start position and end position of the promoter are marked at the top-left and bottom-right of the promoter region respectively. The gene name is marked at the bottom-left of the promoter region. To facilitate the comparison, the control group is plotted under 0 by taking the opposite value. The solid blue line is the averaged methylation profile of EE2, and the solid red line is the averaged methylation profile of control. Regions in pink background indicate hypermethylation and regions in light purple background indicate hypomethylation. The CpG islands within the promoter regions are identified using the method in Takai D and Jones PA, 2002 (ref.55) and highlighted in the plot using dark green rectangles. We can see many methylation events are overlapped with CpG islands

Supplementary Table

Supplementary Table S1. Rat primer sequences used in q-PCR experiments.

<i>Dnmt1</i>	Forward	5'-GCTAAGGACGATGATGAGACG-3'
	Reverse	5'-CTTTTGGGTGACGGCAACTC-3'
<i>Dnmt3a</i>	Forward	5'-ACGCCAAAGAACGTGTCTGCT-3'
	Reverse	5'-CTTGCCTGCTTATGGAG-3'
<i>Dnmt3b</i>	Forward	5'-GAATTGAGCAGCCCAGGTTG-3'
<i>Dnmt3b</i>	Reverse	5'-AAGAAGAGCCTCCTGTGCC-3'
<i>18S rRNA</i>	Forward	5'-TCGGAACTGAGGCCATGATT-3'
	Reverse	5'-CCTCCGACTTCGTTCTGATT-3'

Supplementary Methods

Identification of Significant Inheritable Methylation Alterations (ISIMA)

We developed a statistical approach, namely Identification of Significant Inheritable Methylation Alterations (ISIMA), to identify the consistently inherited differential methylation patterns of genomic regions. In this study, we applied this method to identify the gene promoter regions that consistently showed significant methylation changes in all three generations.

Methylation intensity fold change is used to represent the differential methylation status. The total number of bases of short reads in a promoter is used to indicate the methylation intensity of that region. Based on this measure, we calculated the fold change of methylation intensity of the j th promoter in the i th generation by

$$f_{i,j} = \frac{N_{EE2_{i,j}} + \beta L}{N_{control_{i,j}} + \beta L}, i = 1, 2, 3, \quad (\text{S1})$$

where $N_{EE2_{i,j}}$ and $N_{control_{i,j}}$ are the methylation intensities of EE2 group and control group, respectively, L is the short-read length which equals 36 in MDBCap-seq data, and β is an offset parameter that suppresses the effects of weak methylation signals in the denominator. Here we set $\beta = 10$.

Further, we assessed the significance of the inherited differential methylation based on the methylation fold changes of all generation using ISIMA, to select promoters that have consistently large fold changes in all generations than would be expected by chance.

ISIMA consists of 5 steps.

1. Calculate the transformed fold change of all promoters by taking logarithm of the result of (S1).

$$F_{i,j} = \log_2\left(\frac{N_{EE2_{i,j}} + \beta L}{N_{control_{i,j}} + \beta L}\right), i = 1, 2, 3. \quad (\text{S2})$$

After the transformation, the fold change value become symmetric: a positive value indicates hypermethylation and a negative value indicates hypomethylation.

2. Calculate a consistent methylation score (M -score) for promoter j by the sum of fold changes $F_{i,j}$,

$$M_j = \sum_{i=1}^3 F_{i,j}. \quad (\text{S3})$$

This definition encourages large fold changes in each and all generations and consistent changes. Contradictory fold changes will cancel each other to produce a smaller M -score.

3. Permute the promoter labels within each generation and recalculate the M -scores for every promoter. Suppose B permutations are used, then the B permuted M -scores of the j th promoter simulate the null distribution of M_j .

4. Due to the potential unbalanced occurrence between hypermethylation and hypomethylation, it is probable that the null distribution of M_j is asymmetric in some chromosomes. Therefore, the permutation p values of hypermethylation and hypomethylation are evaluated separately, *i.e.*, one-sided test is used in each case.

5. Find all promoters with $p < 0.002$ to form significant promoter set P . We use p value cutoff of 0.002 to select reasonable amount of promoters for visualization.

The promoters in set P represent the inherited hypermethylation and hypomethylation induced by EE2 across three generations.