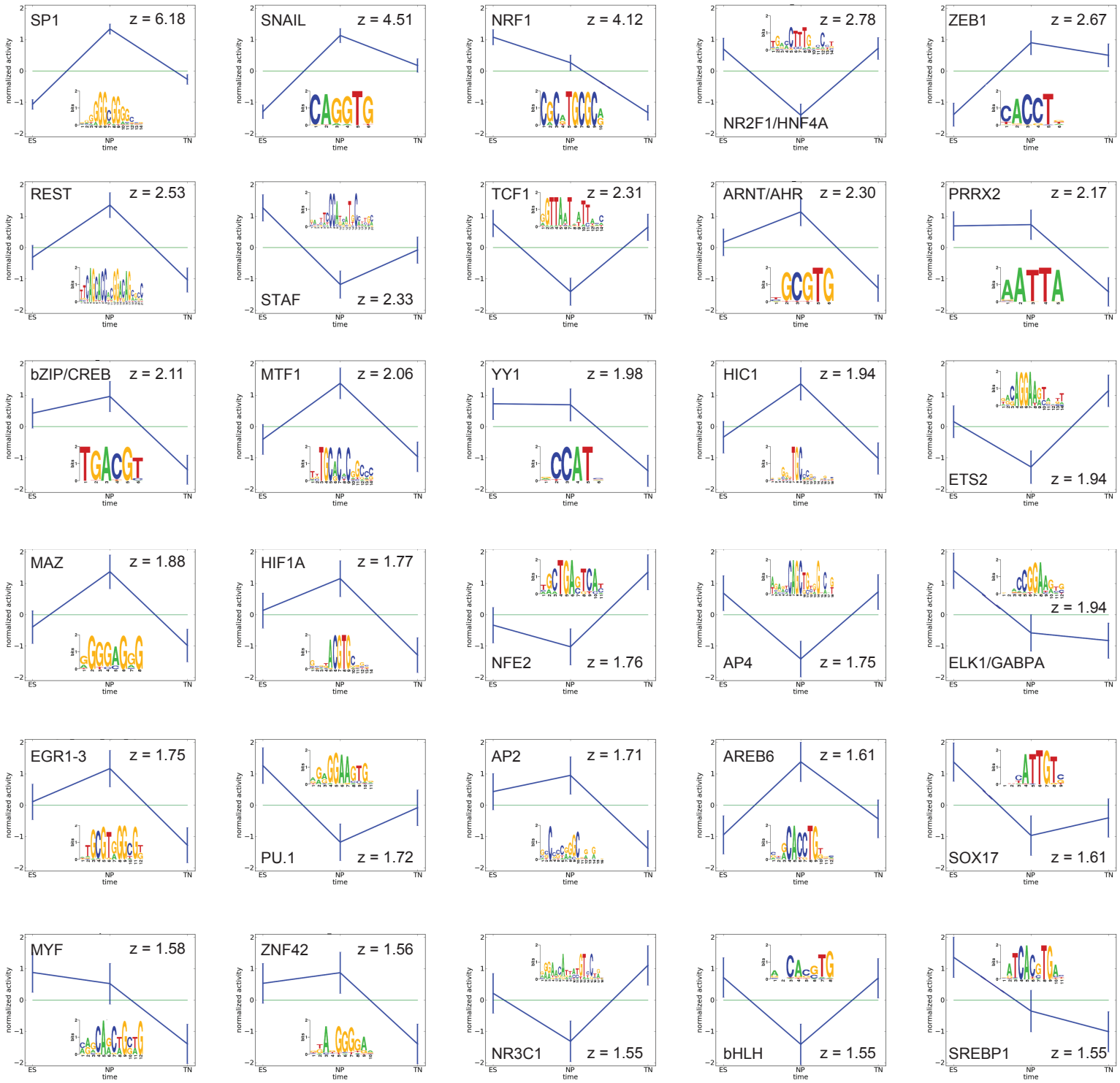
Supplementary Figure 1

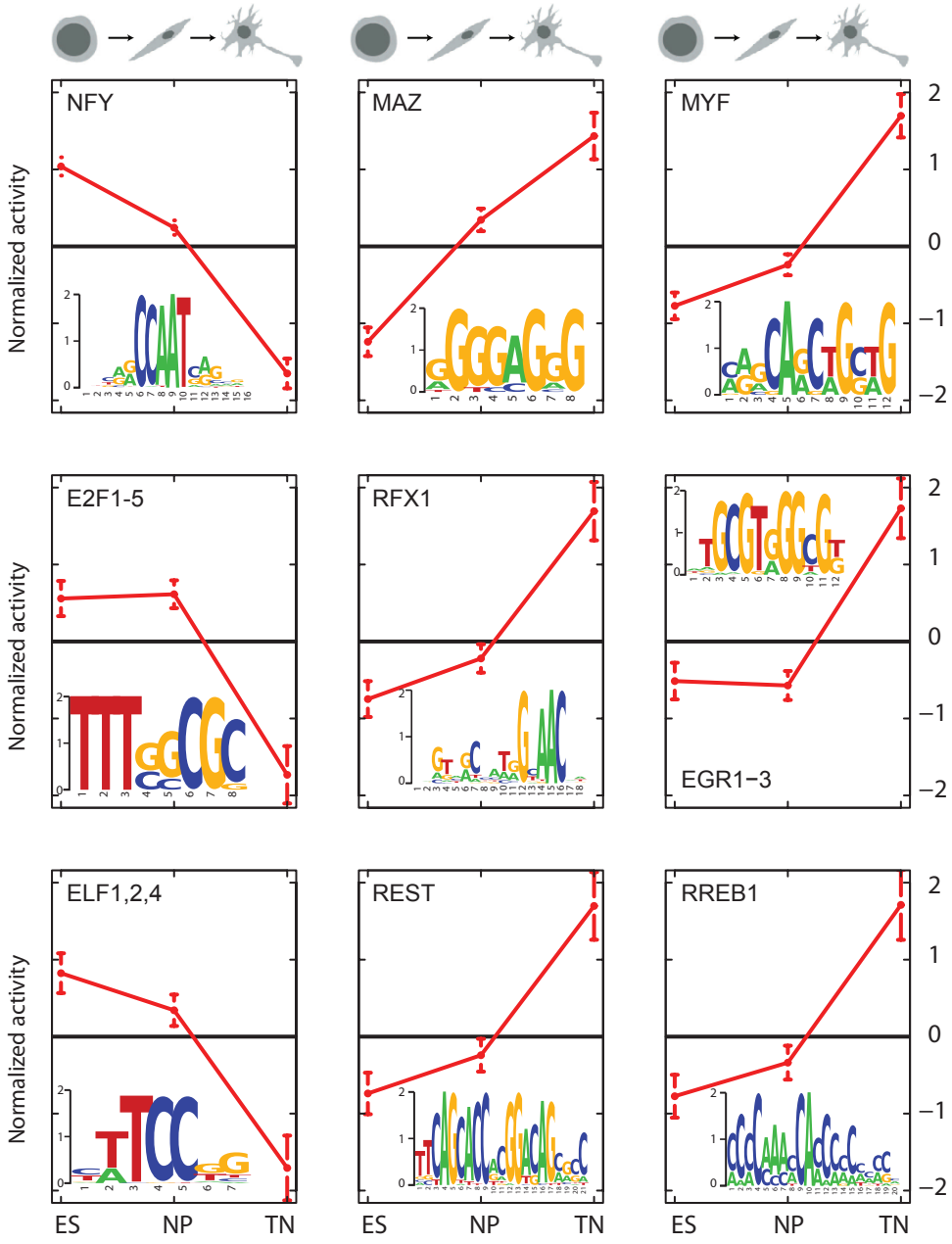
Arnold, Schöler et al.



Supplementary Figure 1: Epi-MARA predicts transcription factor activities that explain dynamics in H3K27me3 levels during neuronal differentiation: Depicted are the normalized activity profiles of the top 30 motifs (blue lines, with standard errors indicated) with their respective z-values. The three time points correspond to the embryonic stem cell (ES), neuronal progenitor (NP), and terminal neuron (TN) stage. Sequence logos of each of the motifs and the transcription factors thought to bind to them are shown as insets.

Supplementary Figure 2

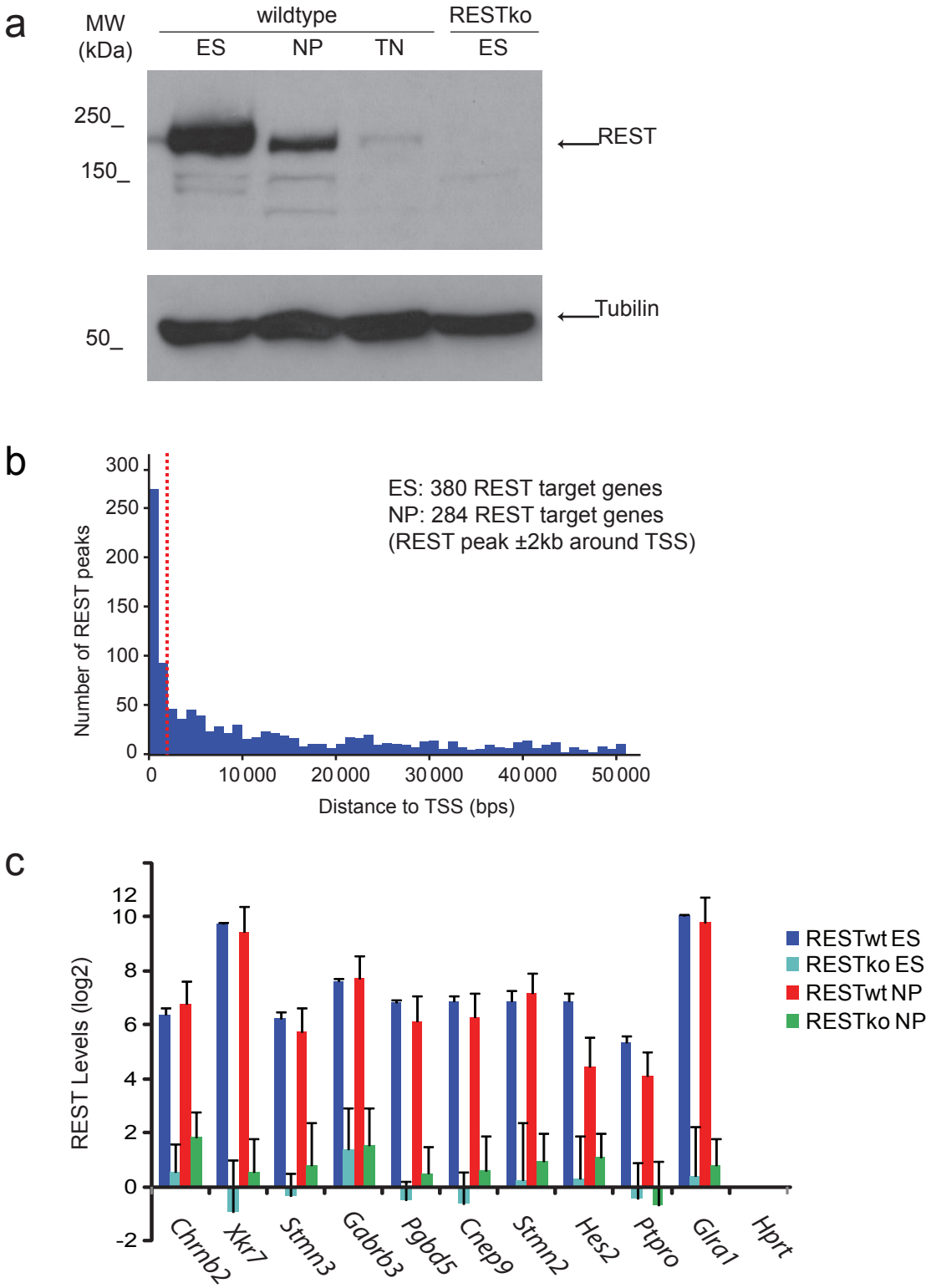
Arnold, Schöler et al.



Supplementary Figure 2: MARA expression analysis predicts the activities of transcription factors driving gene expression dynamics during neuronal differentiation: Normalized activity profiles of the nine most significant motifs that explain changes in gene expression during the differentiation process (red lines, with standard errors indicated). The three time points correspond to the ES, NP and TN stage. Sequence logos of each of the motifs and the transcription factors thought to bind to them are shown as insets.

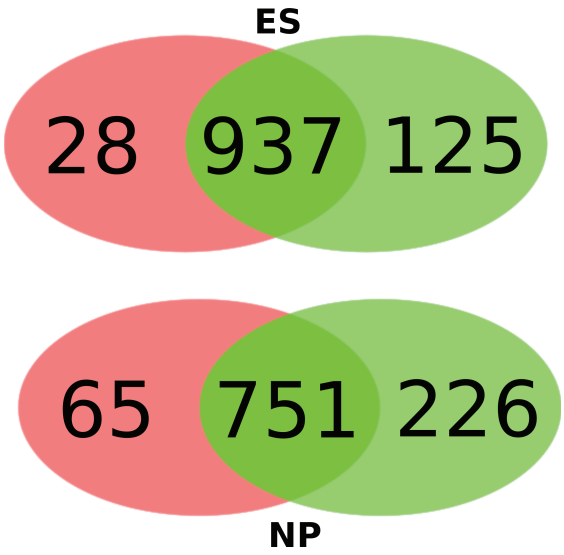
Supplementary Figure 3

Arnold, Schöler et al.



Supplementary Figure 3: REST expression and binding during differentiation: a) Protein levels of REST as detected by Western Blot in extracts from the ES, NP and TN stages in wildtype as well as RESTko background (upper panel). Tubulin serves as a loading control (lower panel). **b)** Distribution of the distance between REST binding peaks and the nearest transcription start site (TSS). Genes with REST binding within +/- 2000 bp of the TSS were classified as REST targets (cut-off indicated by dashed red line). This resulted in a total of 380 target genes in ES cells and 284 in progenitors, with a 96% overlap. **c)** Quantitative PCR of REST ChIPs at the ES and NP stage confirms all 10 tested sites of REST binding as identified by ChIP-seq. Enrichments are normalized to a negative control (*Hprt*). RESTko cells show no signal confirming the specificity of the antibody. Shown are mean enrichments. Error bars show the standard deviation of three biological replicates.

Supplementary Figure 4

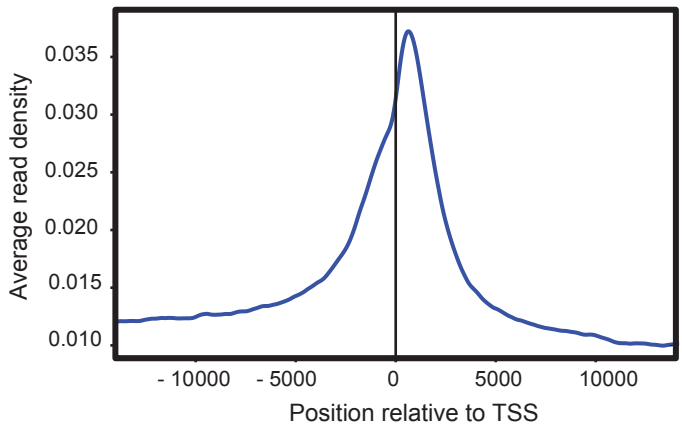


Supplementary Figure 4: Comparison of the REST binding sites identified in our study with those identified by Johnson et al. We first gathered all regions that were identified by Johnson et al. as REST binding regions at either the ES or NP stage. For each of these regions, and separately for both the ES and NP stages, we then calculated a REST enrichment z-value according to both the data of Johnson et al. and our data. The Venn diagrams show the number of these regions that pass a z-value cut-off of 2 for both the ES (upper panel) and NP (lower panel) stages. On the left (red) are regions that only pass the cut-off according to the data of of Johnson et al., on the right (green) regions that only pass the cut-off according to our ChIP-seq data and in the middle the regions that pass the cut-off in both data-sets. Note that while both studies use ES cells as starting point distinct differentiation protocols are used leading to different neuronal populations, which explains larger variation in the differentiated state.

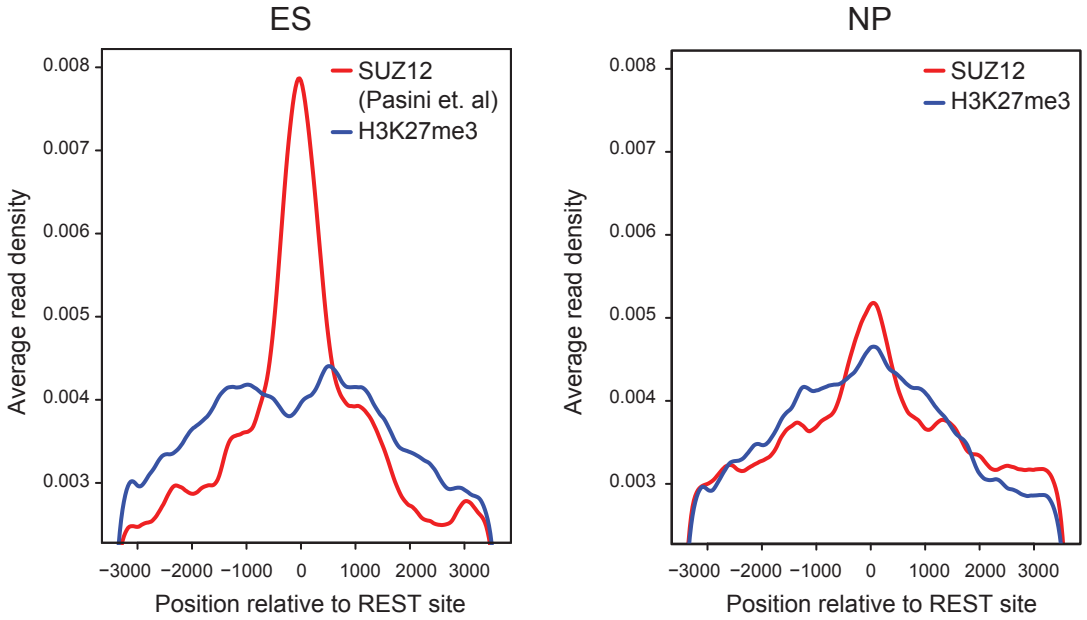
Supplementary Figure 5

Arnold, Schöler et al.

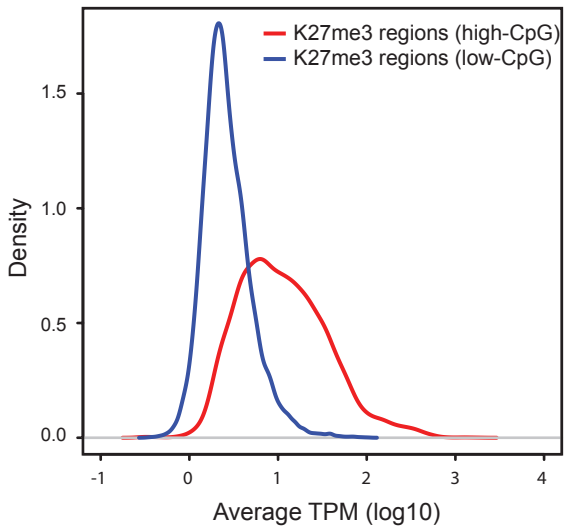
a



b



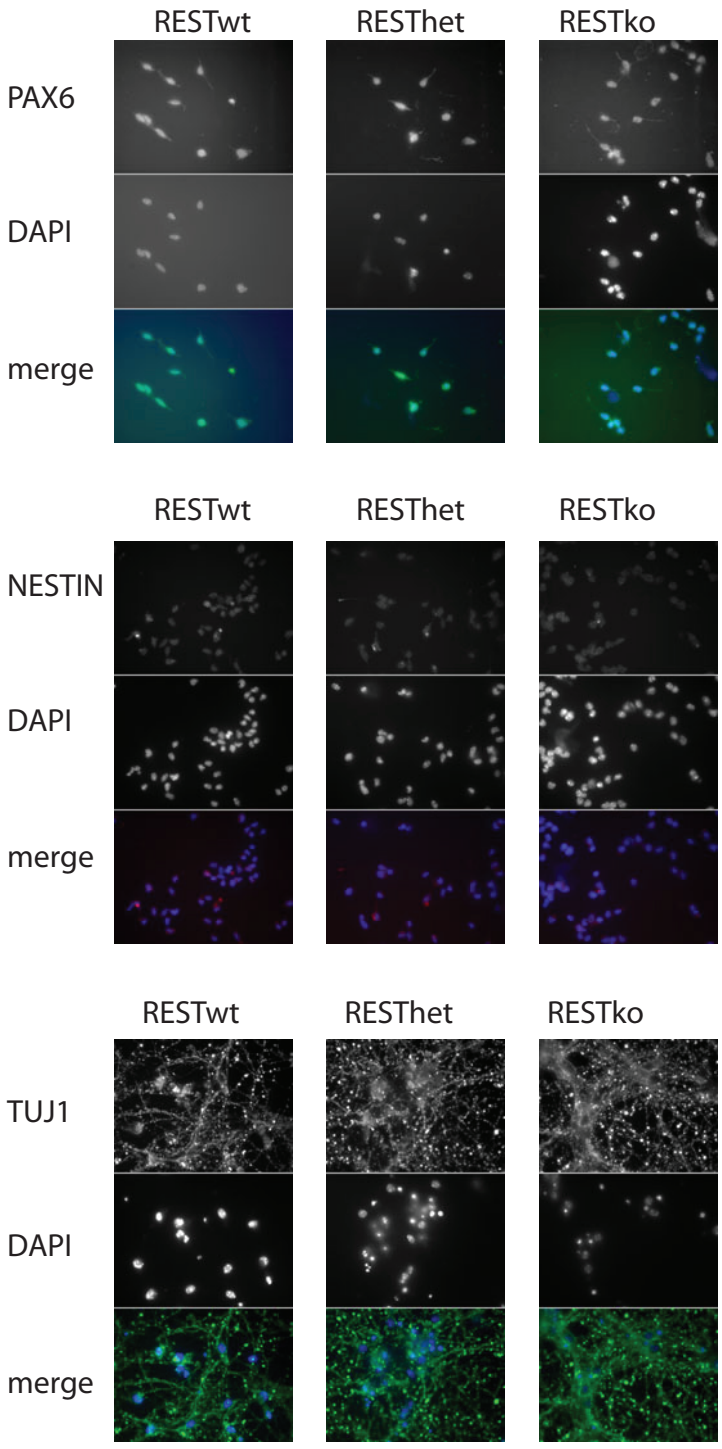
c



Supplementary Figure 5: a) H3K27me3 signal peaks downstream of transcription start sites: Spatial distribution of H3K27me3 signal relative to TSS. All H3K27me3 ChIP-seq samples were pooled and the total read density was plotted relative to position around TSS (±14000bp). This reveals a H3K27me3 peak in the first 1000bps downstream of TSS. **b)** H3K27me3 and SUZ12 levels peak around REST binding sites. Shown is the normalized average read density of SUZ12 (red) and H3K27me3 (blue) in ES cells (Pasini *et. al.*) and neuronal progenitors. **c)** Distributions of the absolute H3K27me3 levels (log ChIP-seq reads per million averaged across the three stages) at all high-CpG (red) and low-CpG regions (blue) that are significantly enriched for H3K27me3.

Supplementary Figure 6

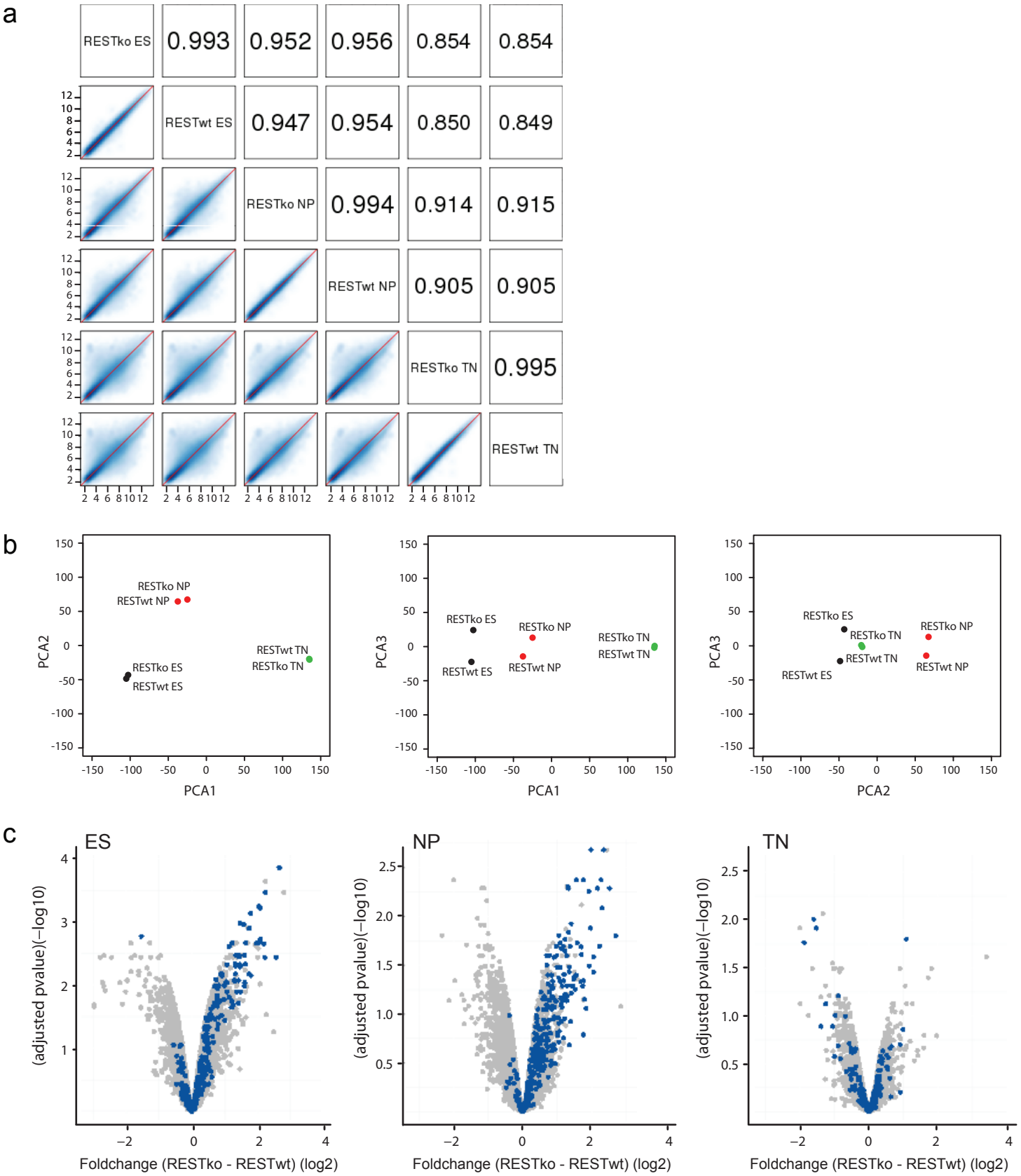
Arnold, Schöler et al.



Supplementary Figure 6: REST knockout ES cells form neurons similar to wildtype ES cells Part I: Marker proteins show similar staining patterns in immuno-cytochemistry in REST knockout and wildtype (WT) cells: REST wildtype (RESTwt), heterozygous (RESThet) and homozygous knockout (RESTko) neuronal progenitors and terminal neurons were fixed and stained for several marker proteins specific for the neuronal progenitor stage (PAX6 (top panel) and NESTIN (middle panel)) and terminal neuron stage (TUJ1 (bottom panel)), respectively. The cells shown are representative for the population.

Supplementary Figure 7

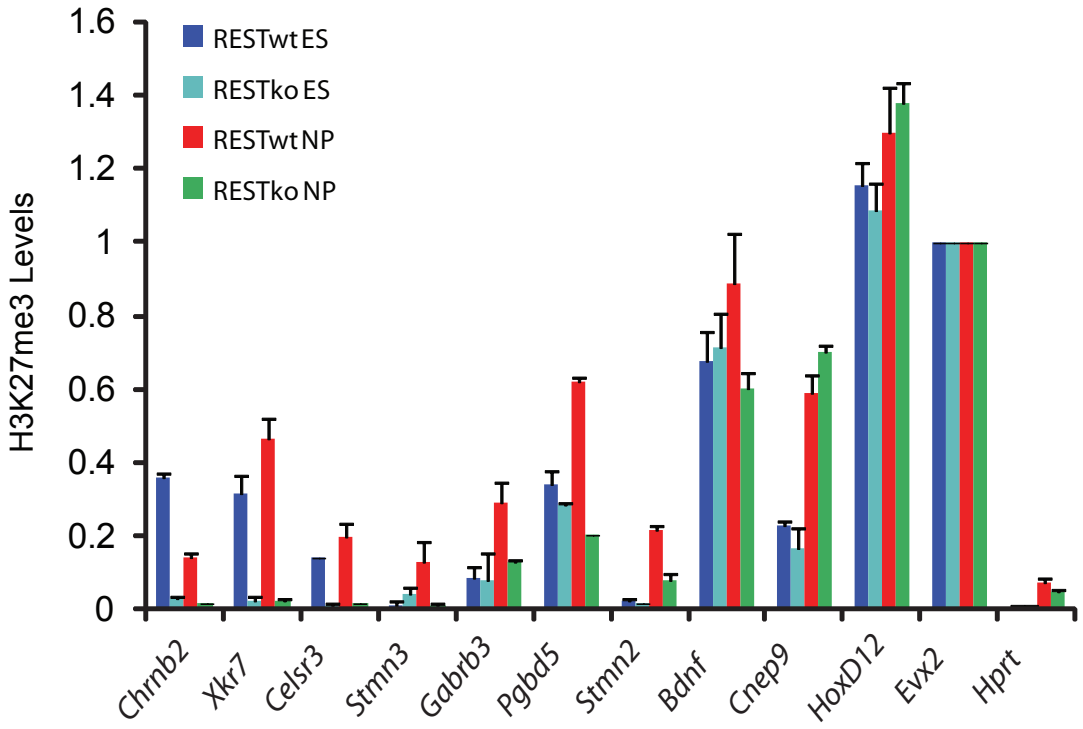
Arnold, Schöler et al.



Supplementary Figure 7: REST knockout ES cells form neurons similar to wildtype ES cells Part II: a) Pairwise correlations of all gene expression microarrays. Shown are the normalized mean expression levels of three biological replicates each. RESTko cells show the highest correlation to each corresponding REST wildtype stage illustrating that the effect of REST knockout on gene expression is small relative to the changes across the differentiation. **b)** Principal component analysis of the gene expression profiles shows that RESTko cells cluster with the corresponding wildtype stage. **c)** Volcano plots depict, for each gene, the fold-change in gene expression in RESTko vs RESTwt cells and the corresponding adjusted p-value for all three stages of differentiation. REST target genes are colored in blue. At the TN stage only very few genes significantly change expression. At the ES and NP stages the number of significantly affected genes is also relatively small and is dominated by direct REST target genes that are upregulated in the RESTko cells.

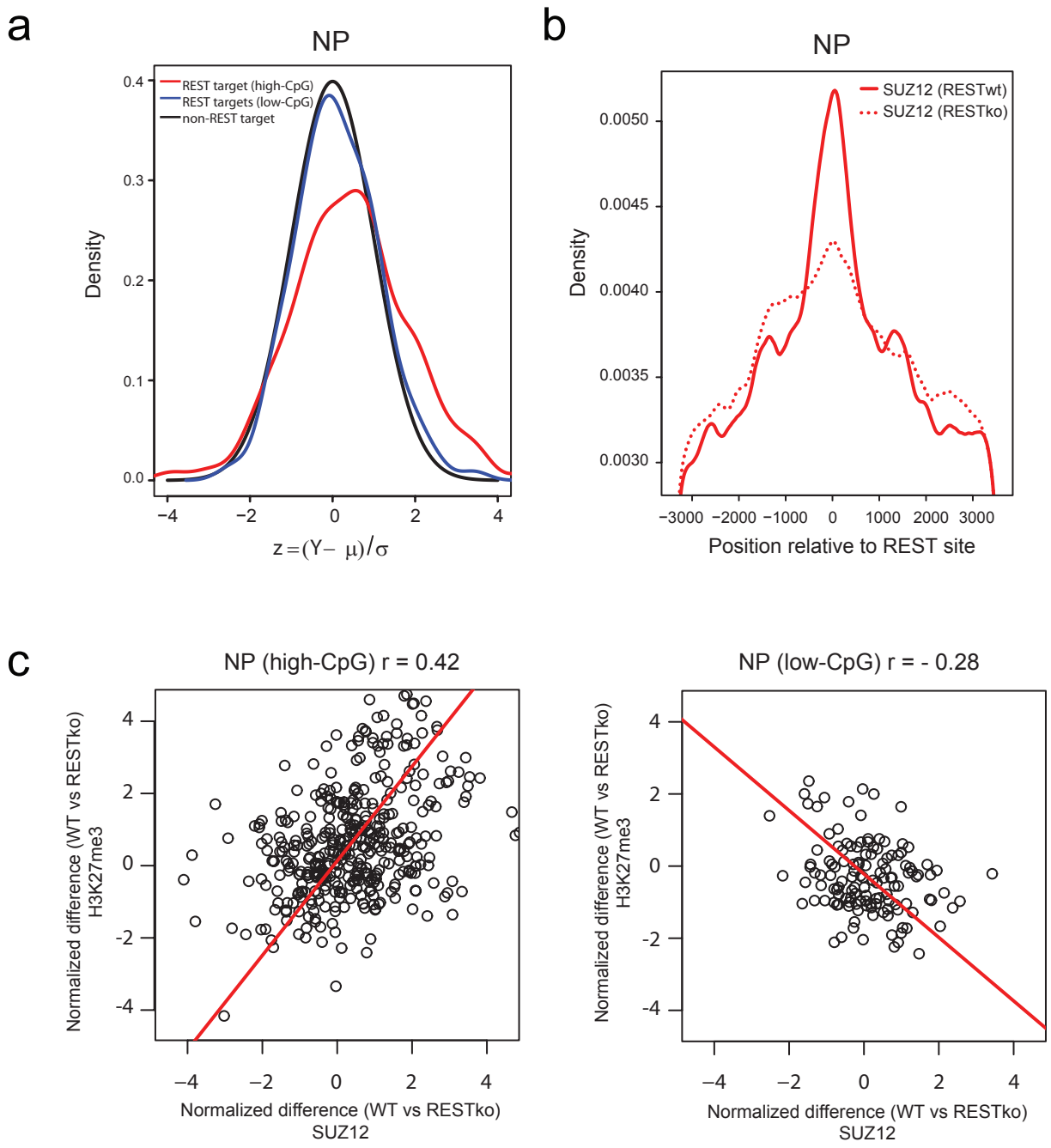
Supplementary Figure 8

Arnold, Schöler et al.

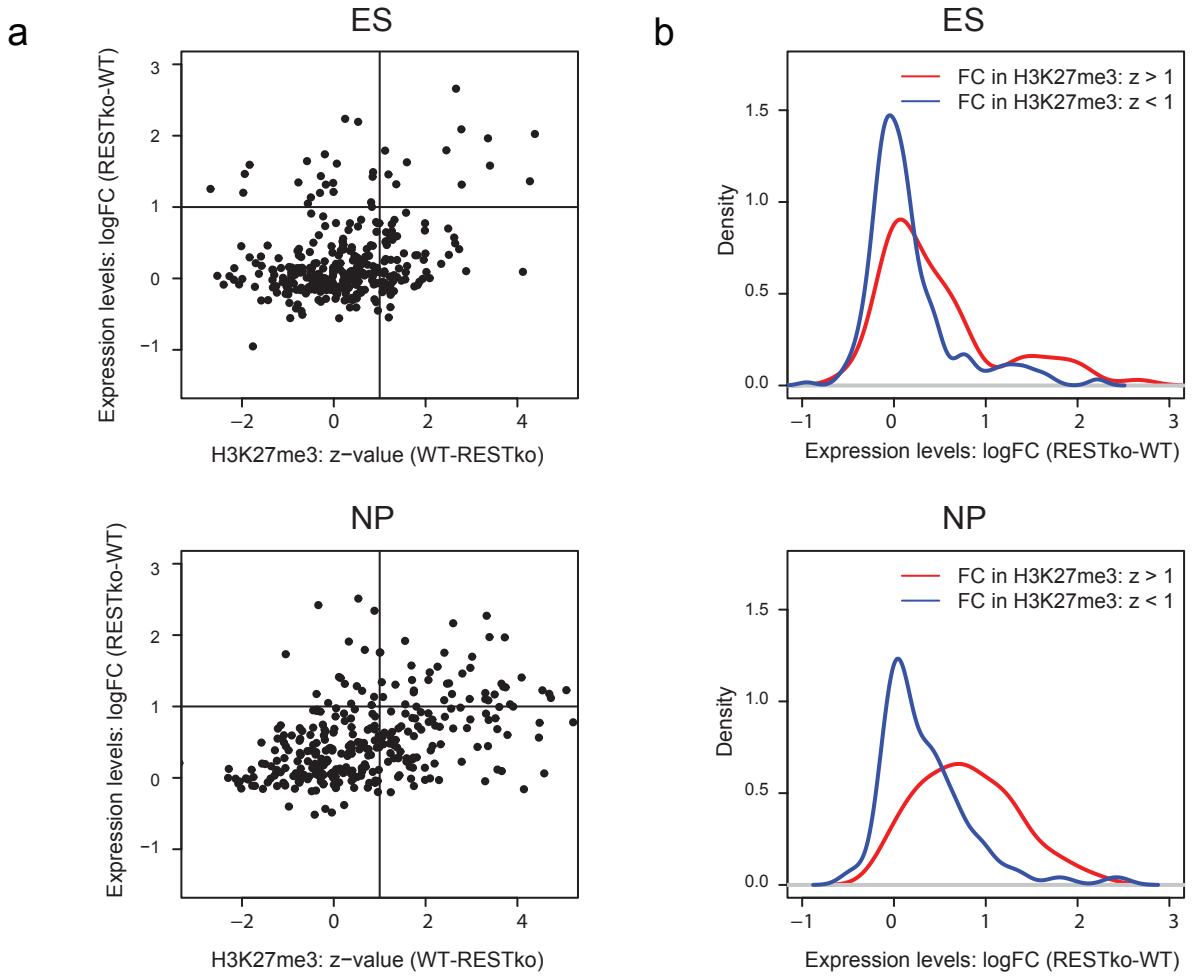


Supplementary Figure 8: Single gene validation of REST targets that lose H3K27 methylation in RESTko at the NP stage: Quantitative PCR of H3K27me3 ChIPs at the ES and NP stage confirms the loss of H3K27me3 in RESTko cells. Enrichments are normalized to a positive control (*Evx2*). All three genes (*Chrnb2*, *Xkr7* and *Celsr3*) already loosing H3K27me3 in RESTko ES ChIP-seq data could be validated as well as all genes loosing H3K27me3 in RESTko NPs ChIP-seq data (*Chrnb2*, *Xkr7*, *Stmn3*, *Gabrb3*, *Pgbd5*, *Celsr3*, *Stmn2*, *Bdnf*) compared to control regions (*Cpne9* and *HoxD12*), which did not change. *Hprt* serves as a negative control. Enrichments show the mean of three biological replicates. Error bars indicate standard deviation.

Supplementary Figure 9



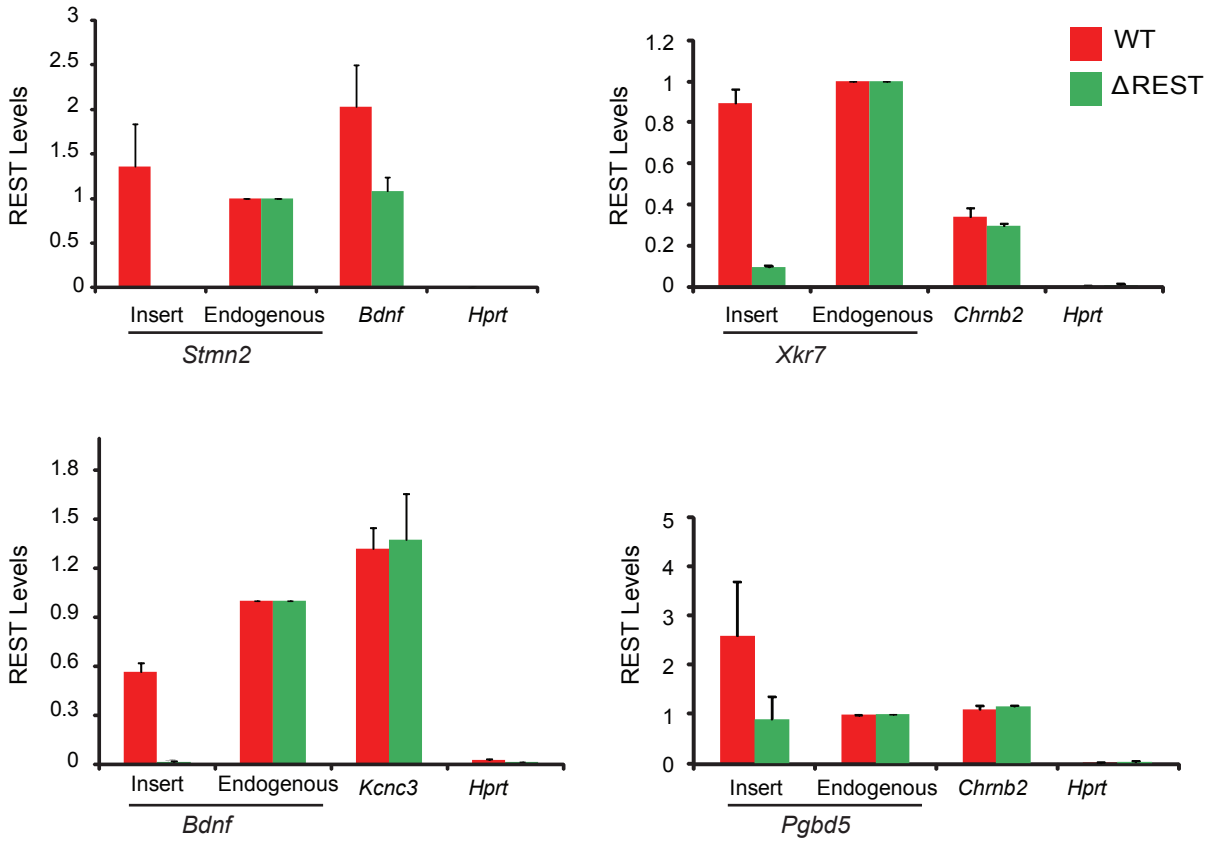
Supplementary Figure 9: a) Comparison of SUZ12 levels between WT and RESTko cells in H3K27me3-enriched regions at the NP stage. Shown are the distributions of the normalized difference in SUZ12 levels (represented as a z-statistic, see Methods) in WT versus RESTko for non-target regions (black line) and for REST targets in either low-CpG (blue line) or high-CpG (red line) regions. Few low-CpG REST targets significantly change, whereas a considerable fraction of high-CpG REST targets show evidence of losing SUZ12 in the RESTko cells.
b) Absence of REST reduces the localization of SUZ12 at REST peaks. Shown is the normalized average read-density of SUZ12 in WT and RESTko neuronal progenitors. **c)** Comparison of the effects on SUZ12 and H3K27me3 levels of REST knockout at REST target regions. For each REST target the normalized difference (z-statistic) between WT and RESTko levels at the NP stage for SUZ12 (x-axis) and H3K27me3 (y-axis) are shown. High-CpG regions are shown in the left panel and low-CpG regions in the right panel. Significant correlation is observed between loss of H3K27me3 and loss of SUZ12 cells for high-CpG REST targets in RESTko cells (r=0.42, p-value < 2.2e-16). A very weak but still significant anti-correlation (i.e. gain of H3K27me3 and loss SUZ12 in the RESTko cells) is observed for low-CpG REST targets (r=-0.28, p-value < 0.001).



Supplementary Figure 10: Independent changes in H3K27me3 levels and gene expression levels at many REST targets. a) Pairwise comparison of changes in the significance of H3K27me3 levels (z-value wildtype minus RESTko, horizontal axis) and changes in transcription (log fold-change RESTko minus wildtype, vertical axis) at both the ES (top panel) and NP stages (bottom panel). The horizontal and vertical lines correspond to a z-value of 1 and a log fold-change of 1 (i.e. two-fold upregulation). In general there is only a weak correlation between the amount of H3K27me3 loss and transcriptional up-regulation. At the ES stage many REST targets are transcriptionally up-regulated without showing a loss of H3K27me3. At the NP stage a significant fraction (33%) of REST targets that significantly lose H3K27me3 (z>1) are not significantly up-regulated. **b)** Distribution of expression log fold-changes under RESTko for REST targets that significantly lose H3K27me3 (z>1, red lines) and REST targets that do not significantly lose H3K27me3 (z<1, blue lines), both at the ES (top panel) and NP (bottom panel) stages. As expected there is an overall association between loss of H3K27me3 and transcriptional up-regulation, especially at the NP stage.

Supplementary Figure 11

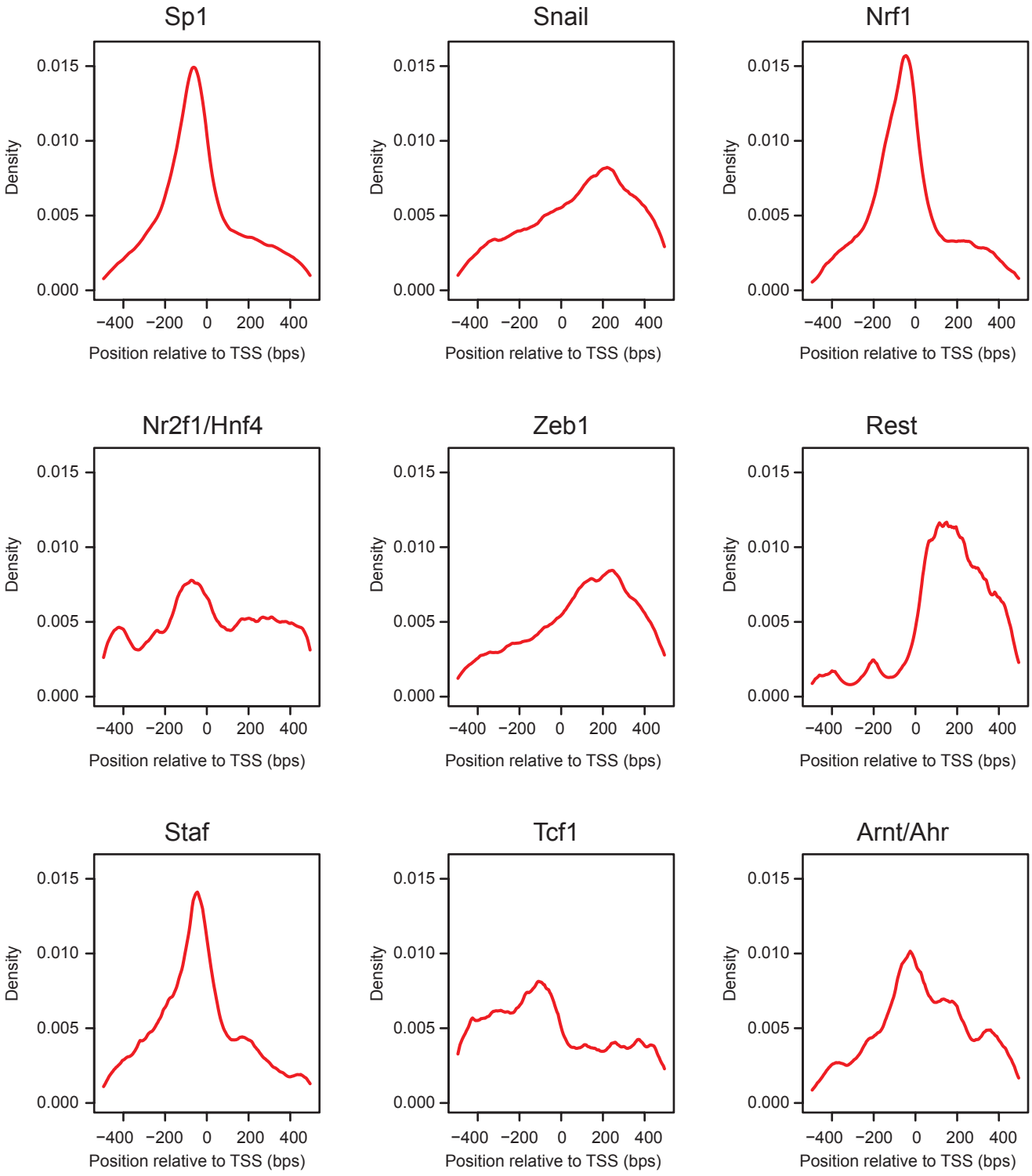
Arnold, Schöler et al.



Supplementary Figure 11: A REST binding site is essential for REST binding. Transgenic wildtype promoters show strong REST binding, but no or weak binding at the four REST mutant sequences. Levels were measured at, from left to right in each panel, the inserted region, the corresponding endogenous locus, a positive control, and a negative control region. All REST levels are scaled to that of the endogenous region and error-bars show the standard error of three biological replicates. Note that, of all mutant promoter fragments, the *Pgbd5* promoter shows the clearest evidence of residual REST binding.

Supplementary Figure 12

Arnold, Schöler et al.



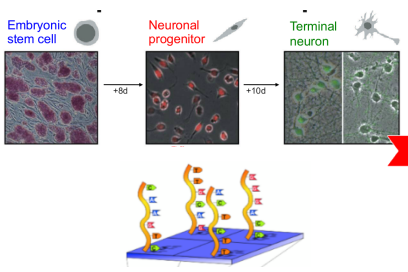
Supplementary Figure 12 Frequencies of predicted binding sites around transcription start sites for Epi-MARA's top 9 predicted motifs. Sp1, Nrf1, Staf and Arnt/Ahr show a strong binding preference around 50 bps upstream of TSS, while Snail, Zeb1, and Rest motifs are mostly found downstream of TSS. Nr2f1/Hnf4 and Tcf1 show much less pronounced positional preferences.

Supplementary Figure 13

Arnold, Schöler et al.

Promoter Analysis

Experimental Setup



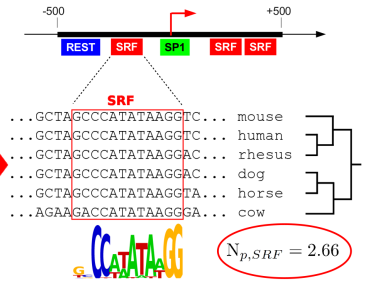
H3K27me3 measured by ChIP-chip

Microarray normalization

BIOCONDUCTOR

"Quantifying H3K27me3 levels"
(Suppl. methods p.11)

TFBS predictions



"TFBS predictions for promoters"
(Suppl. methods p.10)

Epi-MARA

$$\tilde{M}_{pt} = \text{noise} + \sum_m \tilde{N}_{pm} \cdot \tilde{A}_{mt}$$

$$z_m = \sqrt{\frac{1}{T}\sum_t \left(\frac{\bar{A}^{mt}}{\delta A^{mt}}\right)^2}$$

"Fitting the linear model"
(Suppl. methods p.11)

Promoters: Top motifs

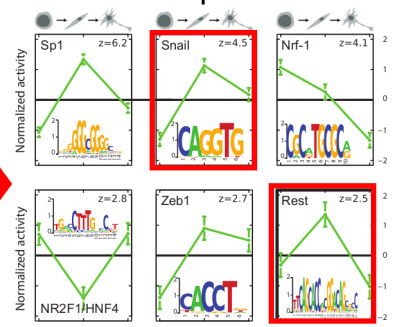
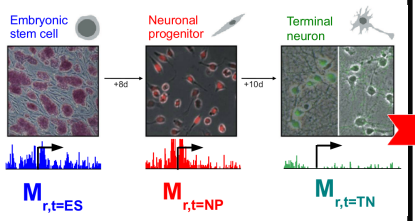


Figure 1

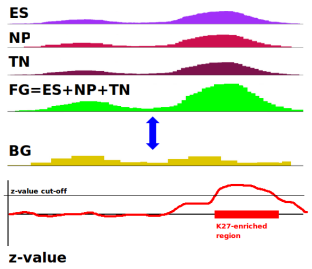
Genome-wide Analysis

Experimental Setup



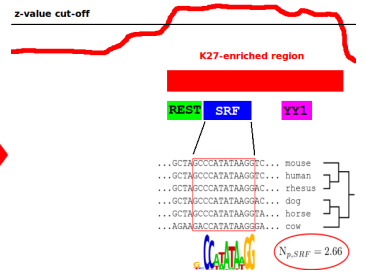
H3K27me3 measured by ChIP-seq

K27 analysis



" Determining K27 regions genome-wide"
(Suppl. methods p.15)

TFBS predictions



"Epi-MARA on K27-enriched regions"
(Suppl. methods p.14)

Epi-MARA

$$\tilde{M}_{pt} = \text{noise} + \sum_m \tilde{N}_{pm} \cdot \tilde{A}_{mt}$$

$$N_{rm} = \begin{pmatrix} N_{rm}^{\text{high-CpG}} & 0 \\ 0 & N_{rm}^{\text{low-CpG}} \end{pmatrix}$$

"Epi-MARA on K27-enriched regions"
(Suppl. methods p.14)

K27me3 regions: Top motifs

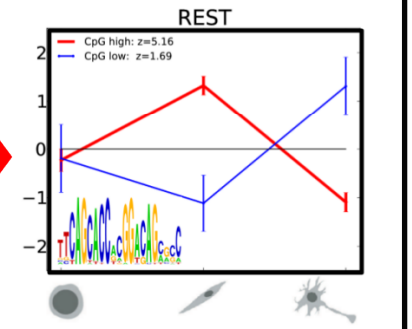
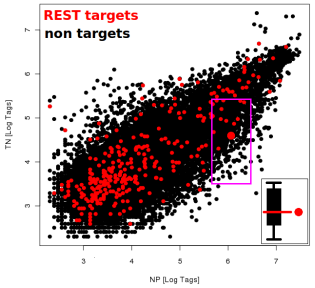


Figure 3

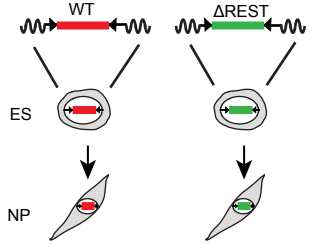
Validation

Comparison of WT and KO



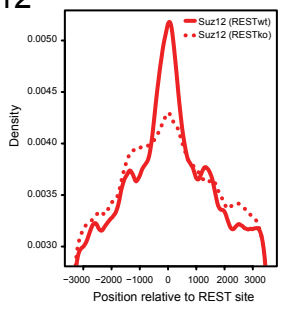
"Analysis of sequencing data"
(Main methods p.22)

Insertion experiments



"Recombinase mediated cassette exchange"
(Main methods p.26)

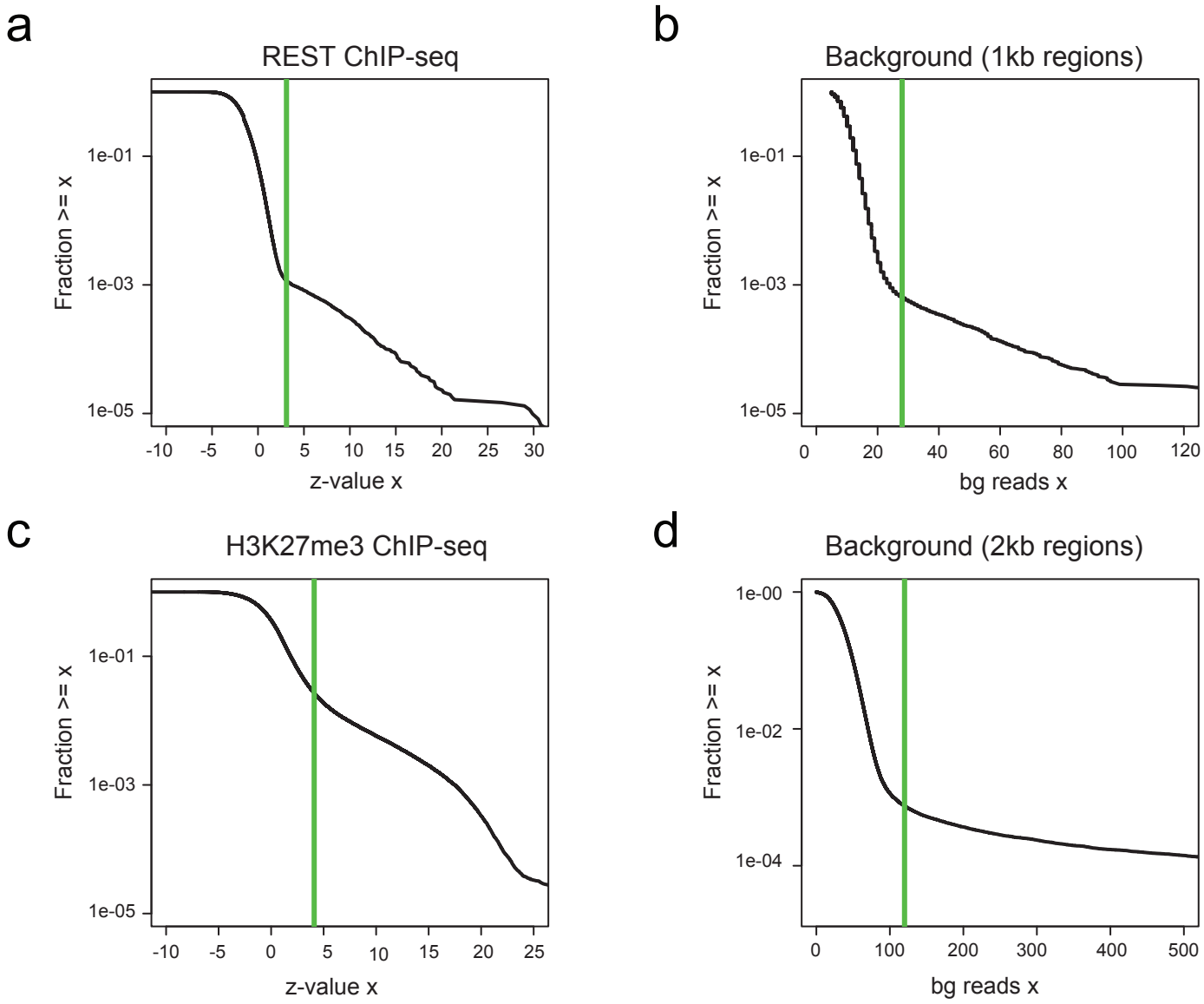
Suz12



"Analysis of sequencing data"
(Main methods p.22)

Supplementary Figure 13: Overview of the methods used in this study. Shown are the steps in the prediction of TFs that regulate H3K27me3 at promoters or genome-wide regions as well as key experiments for subsequent validation. Each panel corresponds to a subsection of the methods (corresponding subsections and page numbers are indicated).

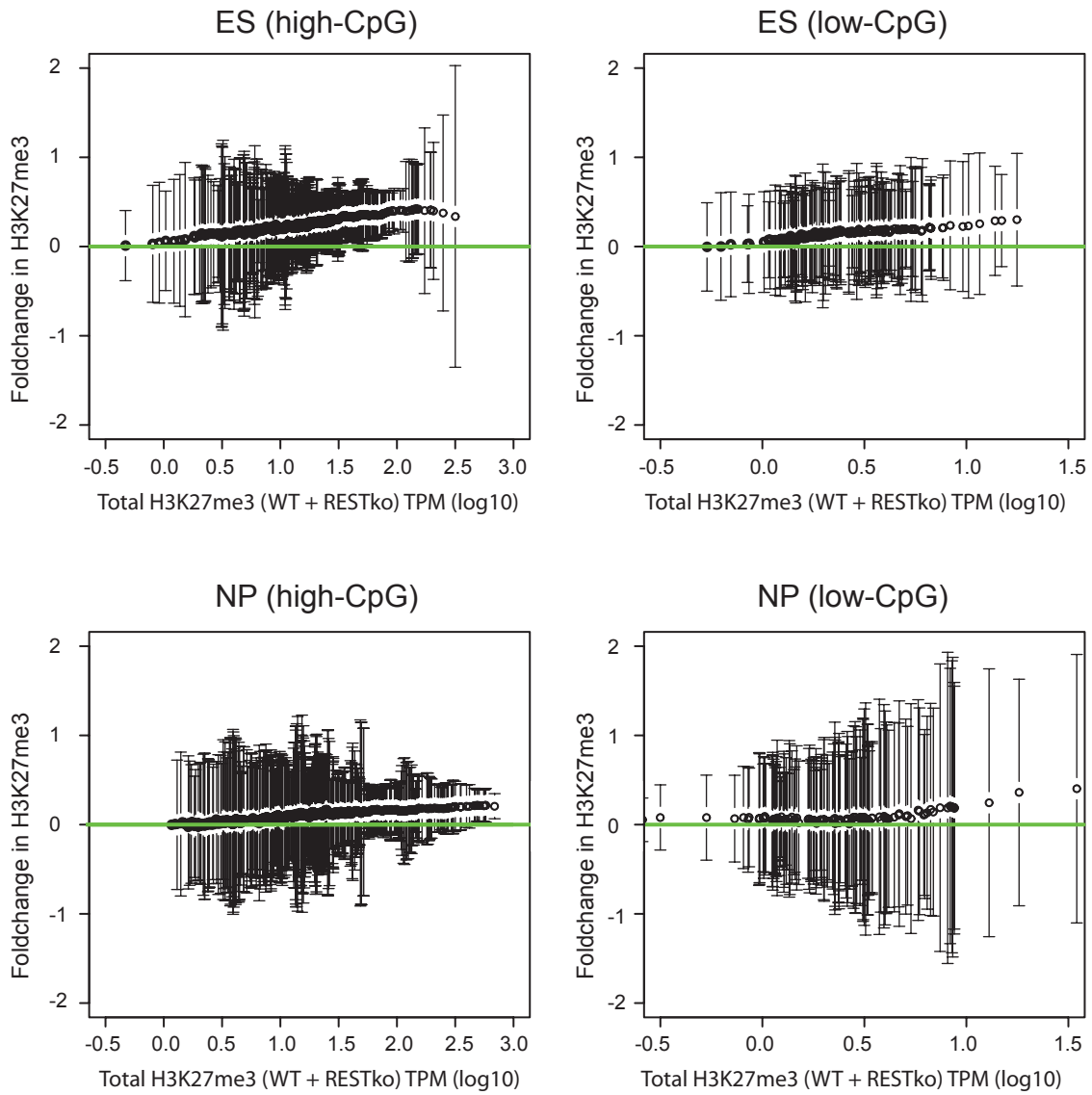
Supplementary Figure 14



Supplementary Figure 14: a) Reverse-cumulative distribution of the z-statistic for enrichment of ChIP-seq reads from the REST IPs relative to background for 1 kb windows genome-wide. The distribution clearly shows two regimes with a second tail at z-statistics larger than approximately 3. The vertical line (z=3.1) shows the cut-off that we chose for considering a window significantly enriched for REST binding. The cut-off was chosen to ensure good sensitivity in the identification of REST binding regions. **b)** Reverse-cumulative distribution of the number of background reads per 1 kb window genome-wide. We observe that the distribution drops steeply up to approximately 20 reads per window, after which it shows a long tail with some windows showing over 100 reads. We remove these genomic regions with aberrantly high background counts (vertical line). **c)** Reverse-cumulative distribution of the z-statistic for enrichment of ChIP-seq reads from the H3K27me3 IPs relative to background for 2 kb windows genome-wide. Again two (and maybe even three) regimes in the distribution are clearly evident and we chose a cut-off of z=4.0 (vertical line) to identify windows significantly enriched for H3K27me3. **d)** Reverse-cumulative distribution of the number of background reads per 2 kb window genome-wide. We observe that the distribution drops steeply up to approximately 100 reads per window, after which the distribution shows a long tail with some windows showing over 500 reads. We remove these genomic regions with aberrantly high background counts (vertical line).

Supplementary Figure 15

Arnold, Schöler et al.



Supplementary Figure 15 Shown are on the x-axis the total H3K27me3 levels (sum of wildtype and RESTko signal) and on the y-axis the average H3K27me3 fold-changes between wildtype and RESTko with standard errors (black dots with error-bars) for all non REST target regions separated into high-CpG (left) and low-CpG (right) regions and for both the ES (top) and NP (bottom) stages.