Supplementary data

The effect of different binning methodology

We have used two binning methods. In both methods we started by ranking the genes according to their methylation levels and divided them into ten groups. In the first method (used in Figures 1-4 and S1-S9) the number of genes in each bin was kept constant, whereas in the second method (used in Figures S10-S12) the methylation span was constant (1.8 and 2 for C^{Me}pGs and $C^{Un\text{-}Me}$pGs respectively). The results were essentially the same with minor changes between the binning methods in the actual range in which the effect of the C^{Me}pGs and $C^{Un\text{-}Me}$pGs were seen. For example, the effect of C^{Me}pGs was seen for $C^{Un\text{-}Me}$pGs levels of 8-18 using the first binning method and 6.5-18.5 using the second method.

Supplementary figure legends

Figure S1 - ***Association between C^{Me}pG count and gene expression levels***

Shown are the cumulative distributions of the expression levels of genes in ES cells in ten bins (each containing 420 genes) of the $C^{Un\text{-}Me}$pG, each divided into two equal subsets according to the number of the C^{Me}pGs. The subsets with lower C^{Me}pGs counts are plotted with solid lines and those with higher values with dotted lines. The range of C^{Me}pGs in each bin and the P value (Mann-Whitney; one side) comparing the expression levels of the two subsets are shown for each graph.

Figure S2 - ***Association between $C^{Un\text{-}Me}$pG count and gene expression levels***

Shown are the cumulative distributions of the expression levels of genes in ES cells in ten bins (each containing 420 genes) of the $C^{Me}pG$, each divided into two equal subsets according to the number of the $C^{Un\text{-}Me}pGs$. The subsets with lower $C^{Un\text{-}Me}pGs$ counts are plotted with solid lines and those with higher values with dotted lines. The range of $C^{Me}pGs$ in each bin and the P value (Mann-Whitney; one side) comparing the expression levels of the two subsets are shown for each graph.

Figure S3 - ***Association between $C^{Me}pG$ percentage and gene expression level in ES cells***

Shown are the cumulative distributions of the expression levels of genes in ES cells in ten bins (each containing 420 genes) of the $C^{Me}pG$ percentage, each divided into two equal subsets according to the number of the CpGs. The subsets with lower CpG count are plotted with solid lines and those with higher values with dotted lines. The range of CpG percentage in each bin and the P value (Mann-Whitney; two sided) estimating the statistical significance of the difference between the two curves are shown for each graph.

Figure S4– Association between expression levels and promoter methylation in non CGI genes in primary fibroblast cells

All CpG island genes in primary fibroblasts cells were divided into 10 equally sized bins according to the number of $C^{Un\text{-}Me}$ pGs(A) and C^{Me} pGs (B)) in their promoter sequence. Cumulative distributions of the expression data are shown. The bins are numbered according to the increasing number of $C^{Un\text{-}Me}$ pGs (A) and C^{Me} pGs (B). The median

expression value is plotted as a function of the median methylation level in each bin for $C^{Un\text{-}Me}$ pGs(C) and C^{Me} pGs (D). Spearman Rho and P values are presented.

Figure S5 – Association between $C^{Un\text{-}Me}$pG count and gene expression levels in primary fibroblast cells

Shown are the cumulative distributions of the expression levels of genes in primary fibroblast cells in ten bins (each containing 480 genes) of the C^{Me}pG, each divided into two equal subsets according to the number of the $C^{Un\text{-}Me}$pGs. The subsets with lower $C^{Un\text{-}Me}$pGs counts are plotted with solid lines and those with higher values with dotted lines. The range of C^{Me}pGs in each bin and the P value (Mann-Whitney; one sided) estimating the statistical significance of the difference between the two curves are shown for each graph.

Figure S6 - ***Association between C^{Me}pG count and gene expression level in primary fibroblast cells***

Shown are the cumulative distributions of the expression levels of genes in ES cells in ten bins (each containing 480 genes) of the $C^{Un\text{-}Me}$pG, each divided into two equal subsets according to the number of the C^{Me}pGs. The subsets with lower C^{Me}pGs counts are plotted with solid lines and those with higher values with dotted lines. The range of $C^{Un\text{-}Me}$pGs in each bin and the P value (Mann-Whitney; one sided) estimating the statistical significance of the difference between the two curves are shown for each graph.

Figure S7 - ***Association between C^{Me}pG percentage and gene expression level in primary fibroblast cells***

Shown are the cumulative distributions of the expression levels of genes in primary fibroblast cells in ten bins (each containing 480 genes) of the C^{Me}pG percentage, each divided into two equal subsets according to the number of the CpGs. The subsets with lower CpG count are plotted with solid lines and those with higher values with dotted lines. The range of CpG percentage in each bin and the P value (Mann-Whitney; two sided) estimating the statistical significance of the difference between the two curves are shown for each graph.

Figure S8 – Association between expression levels and promoter methylation in non CGI genes in ES-F cells

All CpG island genes in ES-F cells were divided into 10 equally sized bins according to the number of $C^{Un\text{-}Me}$ pGs(A) and C^{Me} pGs (B)) in their promoter sequence. Cumulative distributions of the expression data are shown. The bins are numbered according to the increasing number of $C^{Un\text{-}Me}$ pGs (A) and C^{Me} pGs (B). The median expression value is plotted as a function of the median methylation level in each bin for $C^{Un\text{-}Me}$ pGs(C) and C^{Me} pGs (D). Spearman Rho and P values are presented.

Figure S9 – Association between $C^{Un\text{-}Me}$pG count and gene expression levels in ES-F cells

Shown are the cumulative distributions of the expression levels of genes in primary fibroblast cells in ten bins (each containing 460 genes) of the C^{Me}pG, each divided into two equal subsets according to the number of the $C^{Un\text{-}Me}$pGs. The subsets with lower $C^{Un\text{-}Me}$pGs counts are plotted with solid lines and those with higher values with dotted lines. The range of C^{Me}pGs in each bin and the P value (Mann-Whitney; one sided) estimating

the statistical significance of the difference between the two curves are shown for each graph.

Figure S10 - ***Association between $C^{Me}pG$ count and gene expression level in ES-F cells***

Shown are the cumulative distributions of the expression levels of genes in ES cells in ten bins (each containing 460 genes) of the $C^{Un\text{-}Me}pG$, each divided into two equal subsets according to the number of the $C^{Me}pGs$. The subsets with lower $C^{Me}pGs$ counts are plotted with solid lines and those with higher values with dotted lines. The range of $C^{Un\text{-}Me}pGs$ in each bin and the P value (Mann-Whitney; one sided) estimating the statistical significance of the difference between the two curves are shown for each graph.

Figure S11 - ***Association between $C^{Me}pG$ percentage and gene expression level in ES-F cells***

Shown are the cumulative distributions of the expression levels of genes in ES-F cells in ten bins (each containing 460 genes) of the $C^{Me}pG$ percentage, each divided into two equal subsets according to the number of the CpGs. The subsets with lower CpG count are plotted with solid lines and those with higher values with dotted lines. The range of CpG percentage in each bin and the P value (Mann-Whitney; two sided) estimating the statistical significance of the difference between the two curves are shown for each graph.

Figure S12 – Association between expression levels and promoter methylation in the Lister et al., data

Exactly the same analysis represented in Figure 1 were performed on an independent dataset of human ES (h1) cells published in Lister et al. (2009). Spearman Rho and P values were not significant in A,B and D (r=-0.5,P=0.1; r=0.6, P=0.07 and r=-0.24, P=0.5 respectively) and were highly significant (r=0.95,P= 0.00002) in C.

Figure S13 – Association between expression levels and promoter methylation status in non CGI genes in the Lister et al., data

Exactly the same analysis represented in Figure 2 were performed on an independent dataset of human ES (h1) cells published in Lister et al. (2009). Note that the results are essentially the same.

Figure S14 – Association between expression levels and promoter methylation in non CGI genes in ES cells using a different binning method

Same as figure 1c and 1d beside the binning method that used a constant methylation span (1.8 and 2 for C^{Me}pGs and $C^{Un\text{-}Me}$pGs respectively) instead of a constant number of genes. The methylation range and the number of genes in each bin are shown in figures S14 and S13 for A and B respectively.

Figure S15 – Association between $C^{Un\text{-}Me}$pG count and gene expression levels in ES cells using a different binning method

Same as figure S2 beside the binning method that used a constant methylation span (1.8 C^{Me}pGs) instead of a constant number of genes. In each graph we show the number of the genes in the bin (N) the C^{Me}pGs range and the Mann-Whitney P value.

Figure S16 - ***Association between $C^{Me}pG$ count and gene expression level in ES cells using a different binning method***

Same as figure S1 beside the binning method that used a constant methylation span (2 $C^{Un\text{-}Me}pGs$) instead of a constant number of genes. In each graph we show the number of the genes in the bin (N) the $C^{Un\text{-}Me}pGs$ range and the Mann-Whitney P value.

Figure S17 - ***Association between $C^{Un\text{-}Me}pG$ count and gene expression levels in the WT data of Blackledge et al., 2010***

Shown are the cumulative distributions of the expression levels of genes in ES cells(average of the expression data taken from GSM530254 and GSM530255) in ten bins (each containing 710 genes) of the $C^{Me}pG$, each divided into two equal subsets according to the number of the $C^{Un\text{-}Me}pGs$. The subsets with lower $C^{Un\text{-}Me}pGs$ counts are plotted with solid lines and those with higher values with dotted lines. The range of $C^{Me}pGs$ in each bin and the P value (Mann-Whitney; one side) comparing the expression levels of the two subsets are shown for each graph.

Figure S18 - ***Association between $C^{Un\text{-}Me}pG$ count and gene expression levels in the KDM2a KD data of Blackledge et al., 2010***

Shown are the cumulative distributions of the expression levels of genes in ES cells(average of the expression data taken from GSM530256 and GSM530257) in ten bins (each containing 710 genes) of the $C^{Me}pG$, each divided into two equal subsets according to the number of the $C^{Un\text{-}Me}pGs$. The subsets with lower $C^{Un\text{-}Me}pGs$ counts are plotted with solid lines and those with higher values with dotted lines. The range of

C^{Me}pGs in each bin and the P value (Mann-Whitney; one side) comparing the expression levels of the two subsets are shown for each graph.

Figure S19 - ***Interplay between C^{Me}pG and $C^{Un\text{-}Me}$pG levels***

Same as figure 3, beside using 20 bins instead of 10.

Figure S1

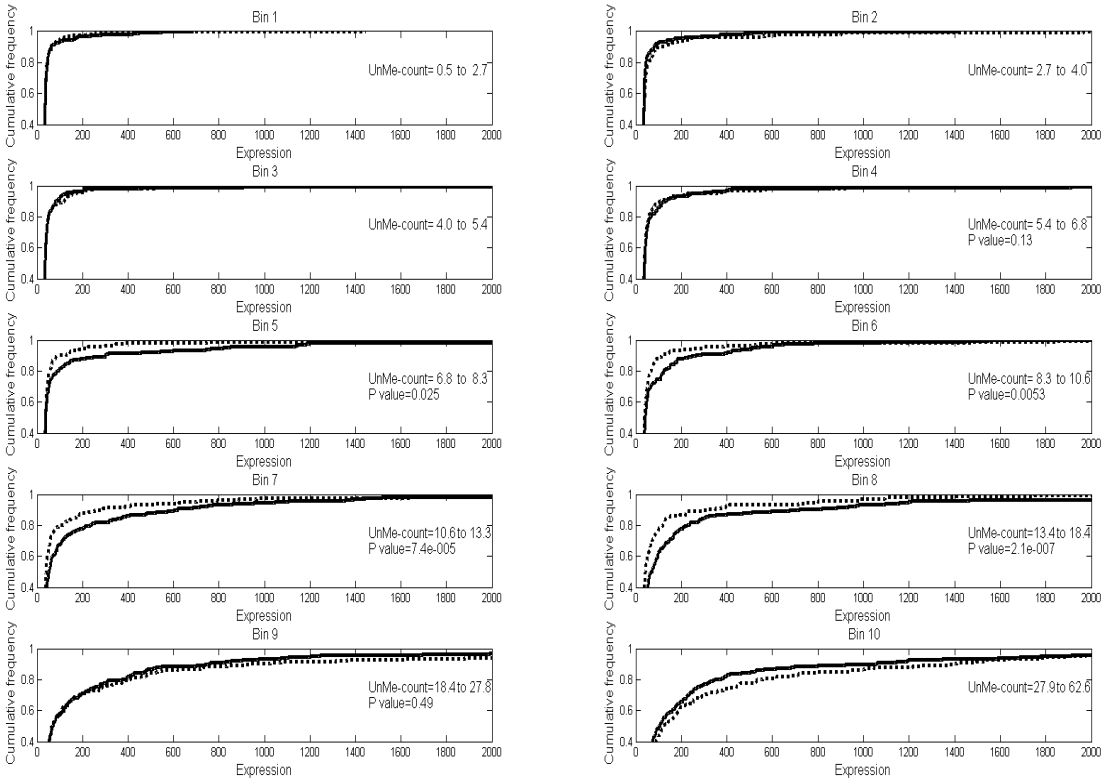


Figure S2

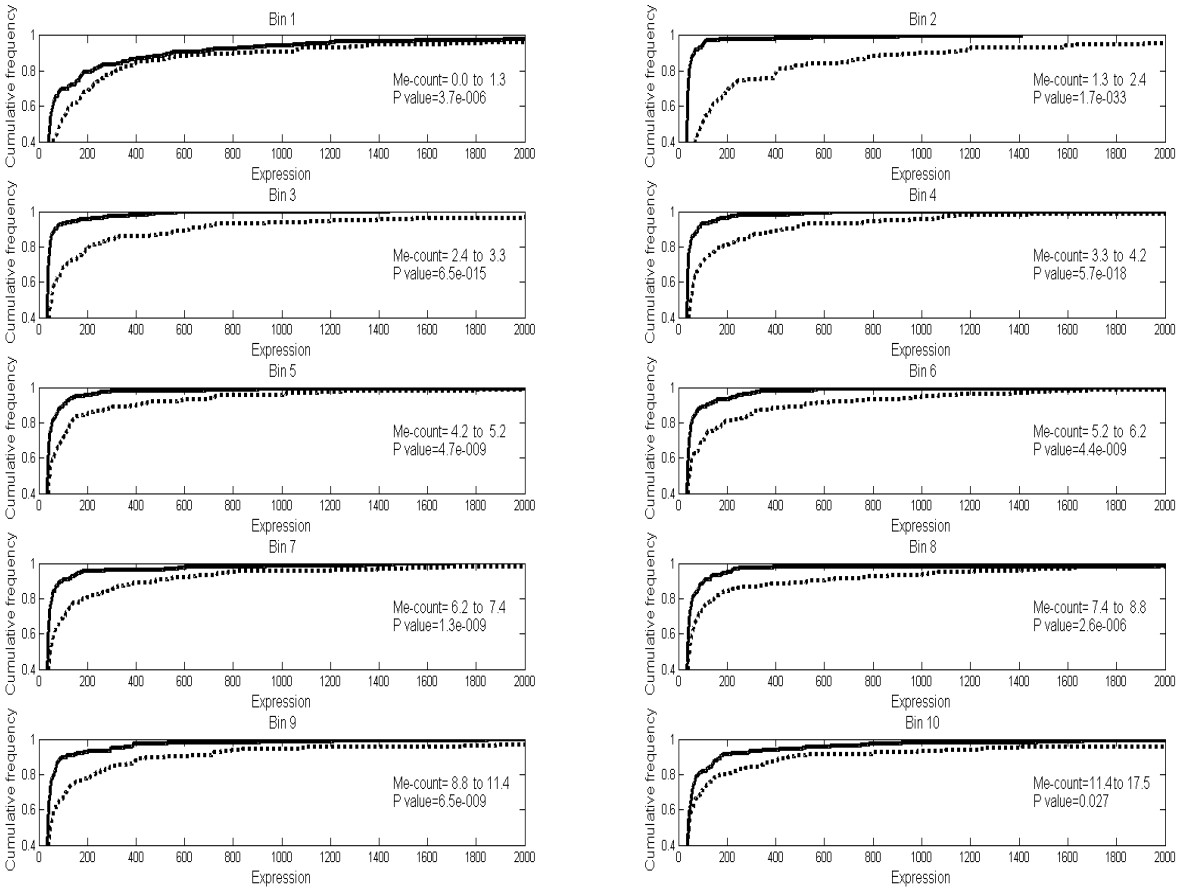

Bin 1
Me-count= 0.0 to 1.3
P value=3.7e-006
Bin 2
Me-count= 1.3 to 2.4
P value=1.7e-033
Bin 3
Me-count= 2.4 to 3.3
P value=6.5e-015
Bin 4
Me-count= 3.3 to 4.2
P value=5.7e-018
Bin 5
Me-count= 4.2 to 5.2
P value=4.7e-009
Bin 6
Me-count= 5.2 to 6.2
P value=4.4e-009
Bin 7
Me-count= 6.2 to 7.4
P value=1.3e-009
Bin 8
Me-count= 7.4 to 8.8
P value=2.6e-006
Bin 9
Me-count= 8.8 to 11.4
P value=6.5e-009
Bin 10
Me-count=11.4to 17.5
P value=0.027
Cumulative frequency
Expression

Figure S3

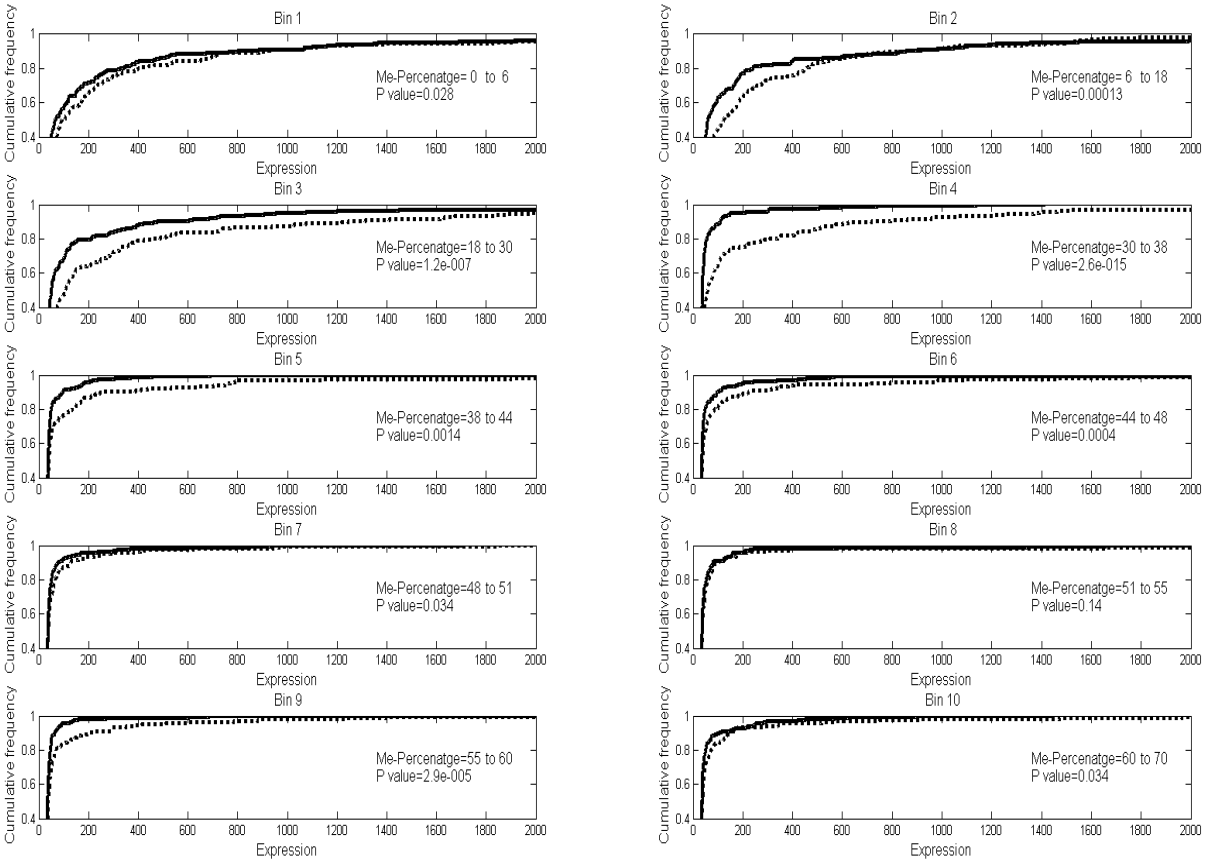

Bin 1
Me-Percenatge= 0 to 6
P value=0.028
Bin 2
Me-Percenatge= 6 to 18
P value=0.00013
Bin 3
Me-Percenatge=18 to 30
P value=1.2e-007
Bin 4
Me-Percenatge=30 to 38
P value=2.6e-015
Bin 5
Me-Percenatge=38 to 44
P value=0.0014
Bin 6
Me-Percenatge=44 to 48
P value=0.0004
Bin 7
Me-Percenatge=48 to 51
P value=0.034
Bin 8
Me-Percenatge=51 to 55
P value=0.14
Bin 9
Me-Percenatge=55 to 60
P value=2.9e-005
Bin 10
Me-Percenatge=60 to 70
P value=0.034
Cumulative frequency
Expression

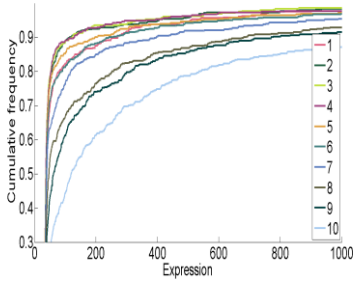
Figure S4

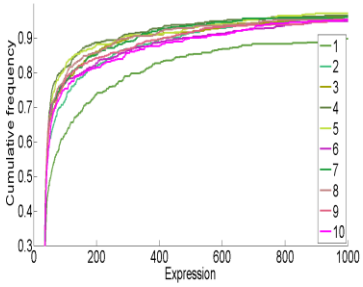
Fibroblasts

$C^{Un\text{-}Me}pG$ count

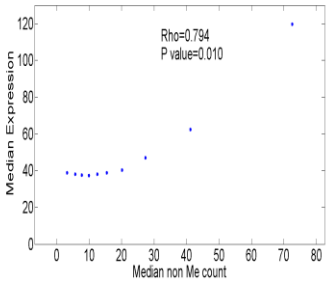
$C^{Me}pG$ count

A





B



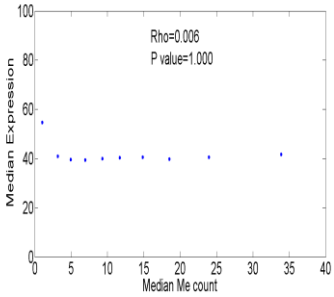


Figure S5

Fibroblasts

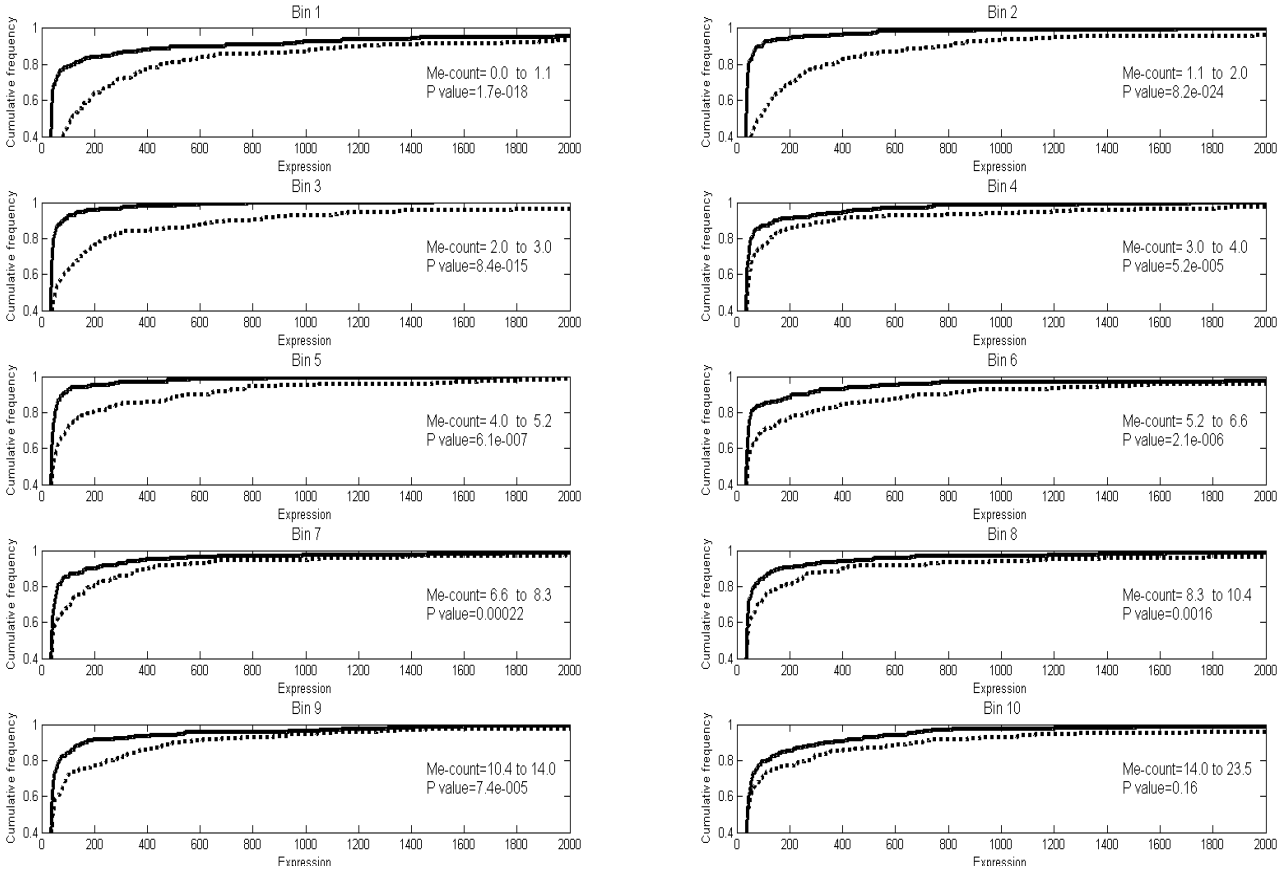


Figure S6

Fibroblasts

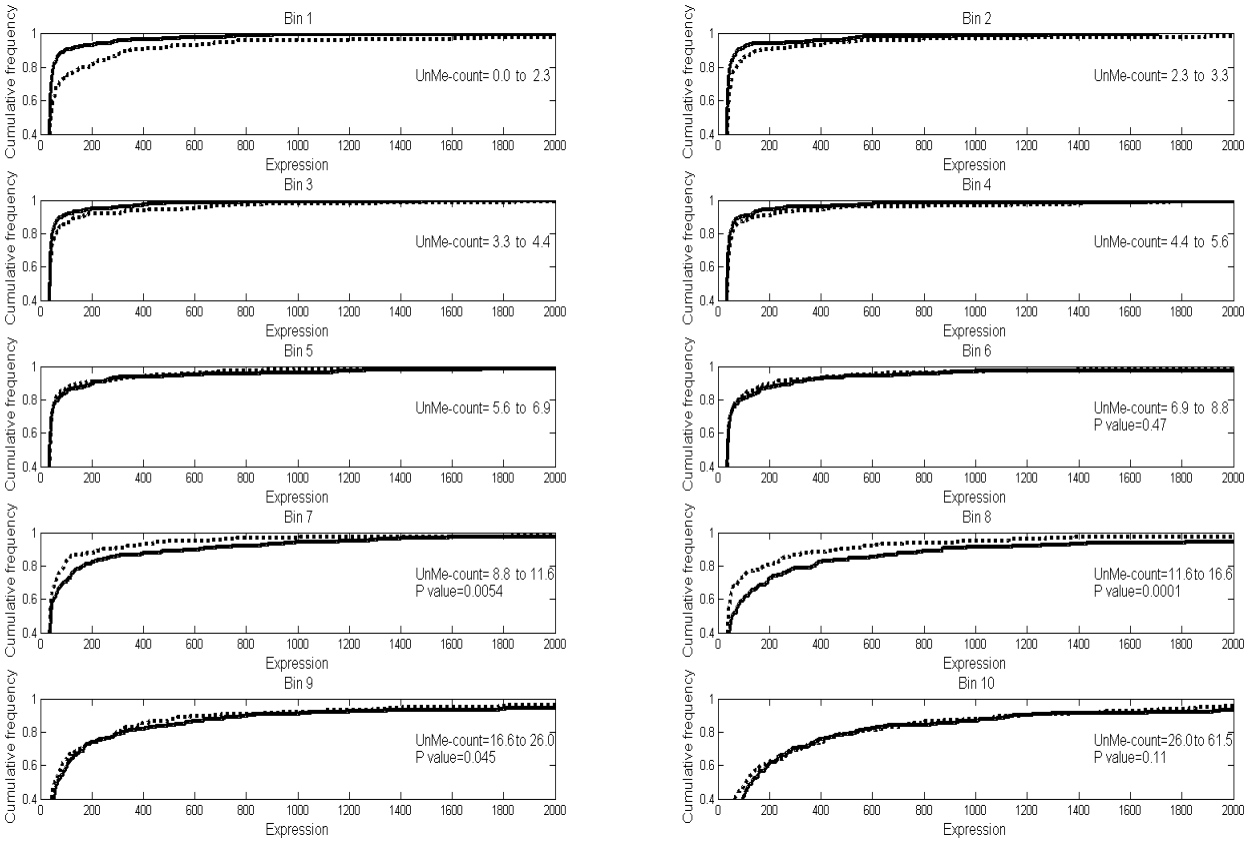


Figure S7

Fibroblasts

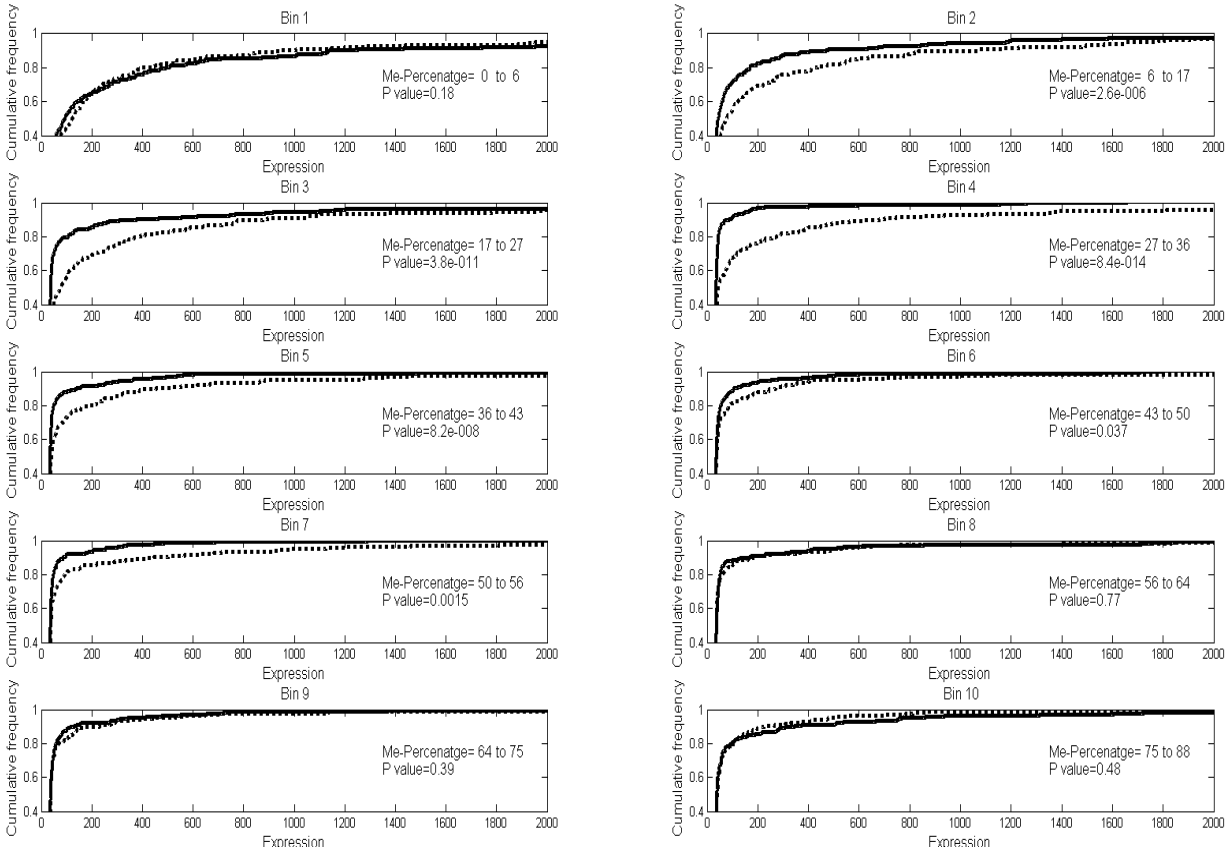


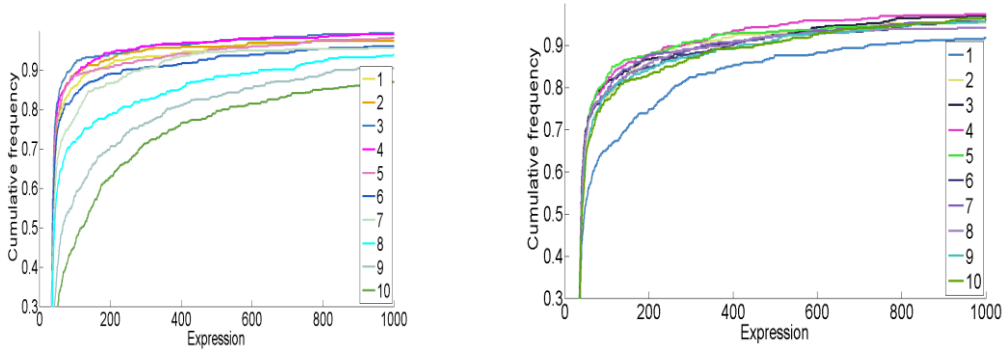
Figure S8

ES_F

$C^{Un\text{-}Me}pG$ count

$C^{Me}pG$ count

A



B

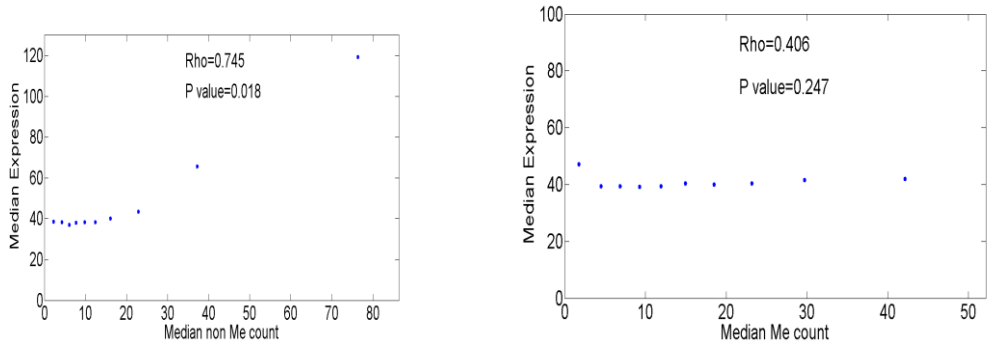


Figure S9

ES_F

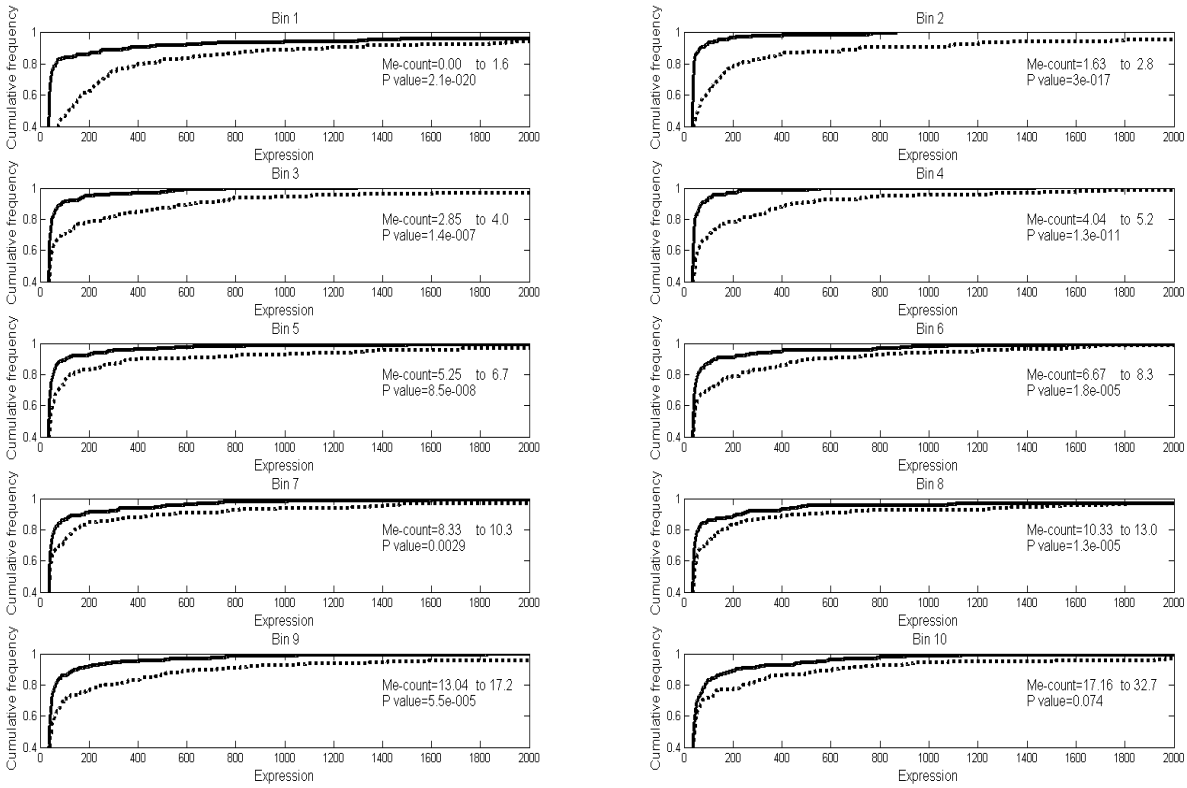


Figure S10

ES-F

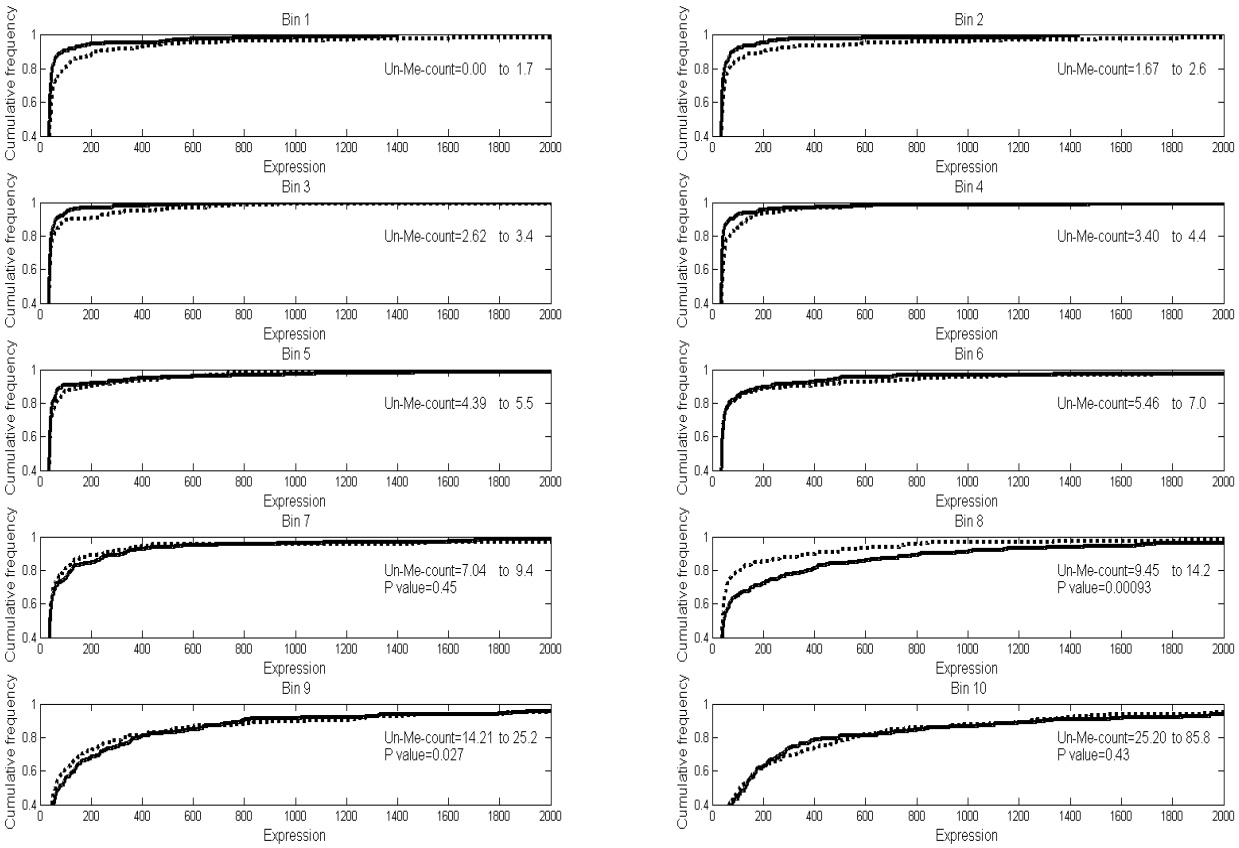


Figure S11

ES-F

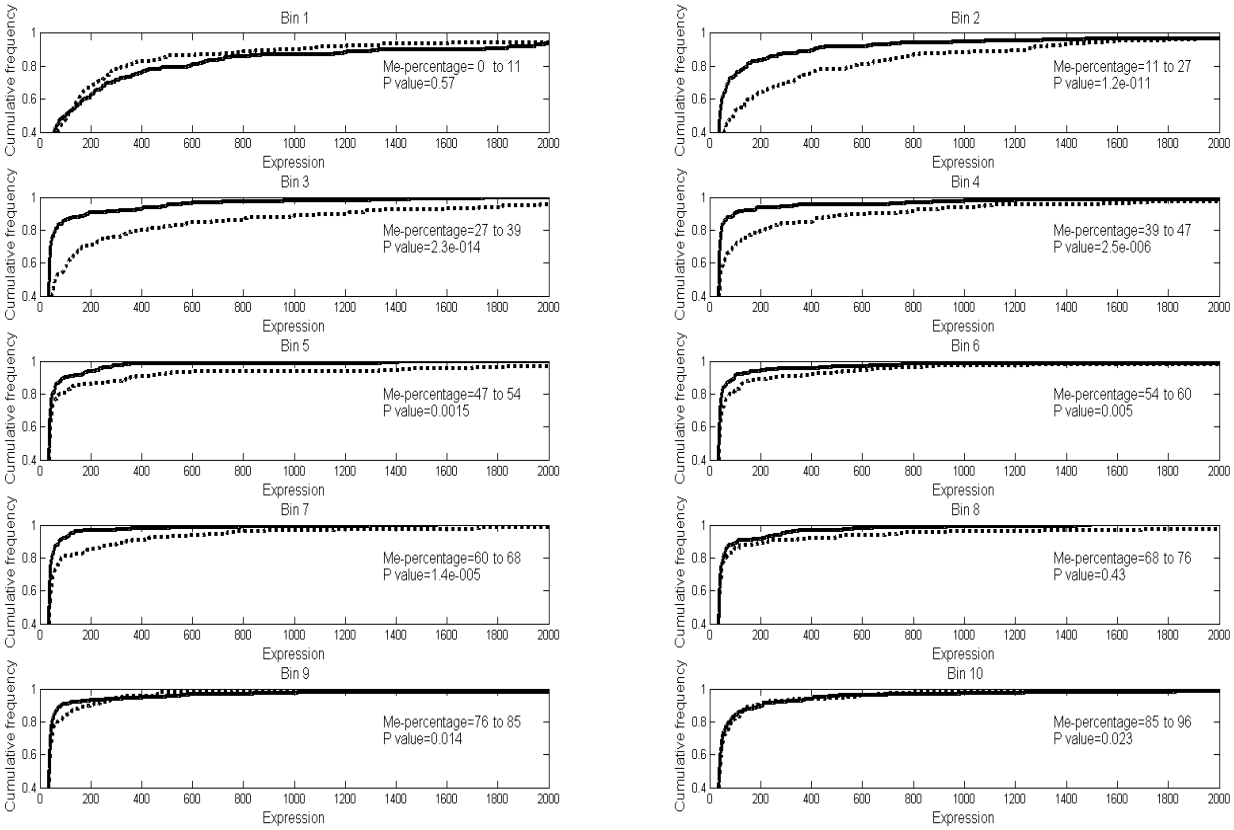


Figure S12

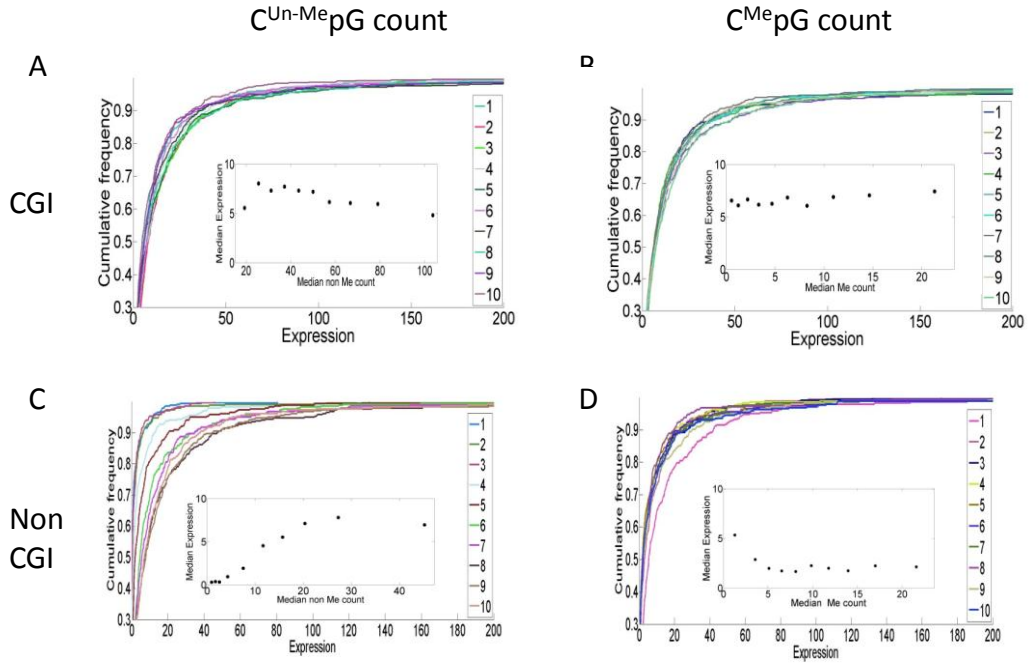

C^Un-Me pG count
C^Me pG count
A
B
CGI
C
D
Non CGI
Cumulative frequency
Expression
Median Expression
Median non Me count
Median Me count

Figure S13

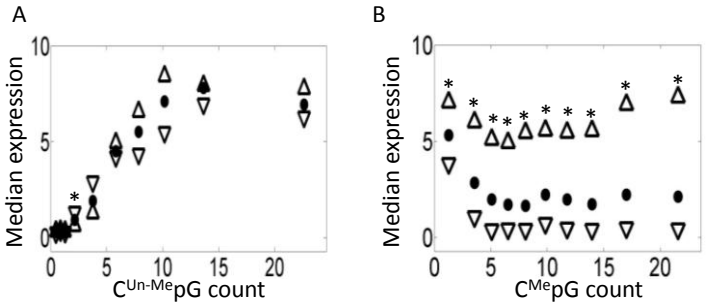


Figure S14

ES-different binning method

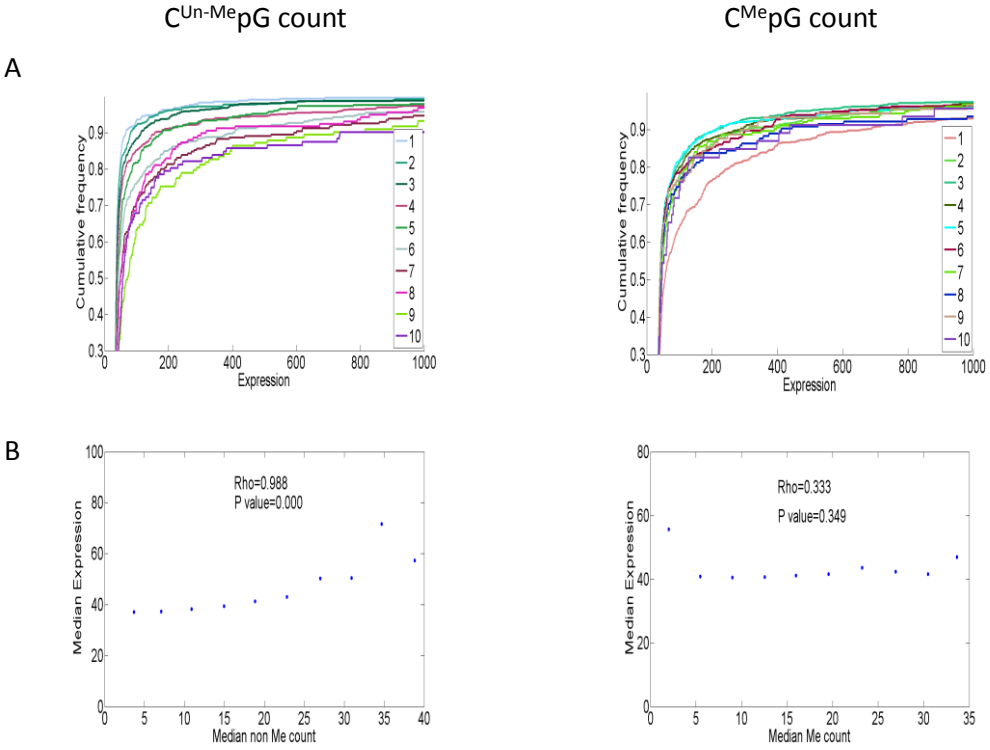


Figure S15

ES-different binning method

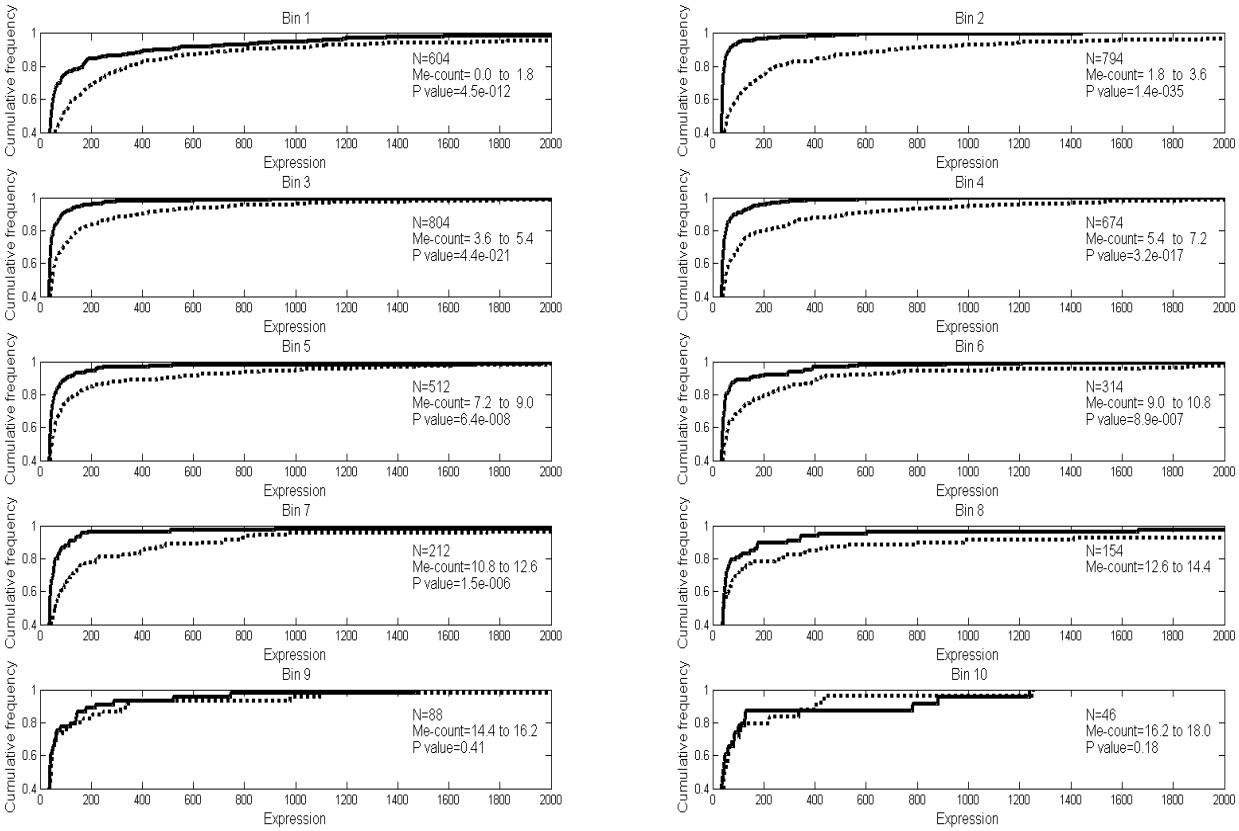


Figure S16

ES-different binning method

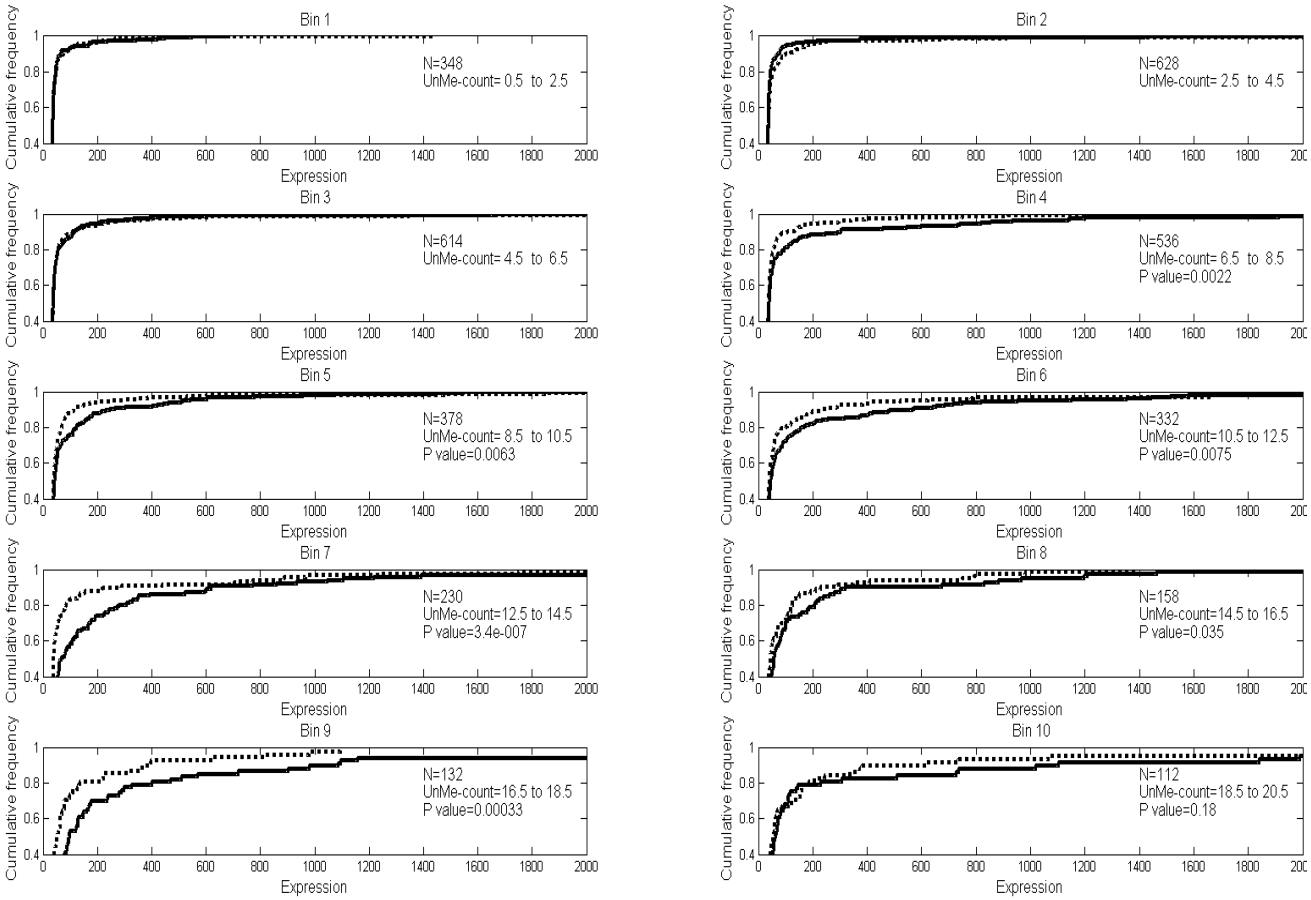


Figure S17

WT data from Blackledge et al., 2010

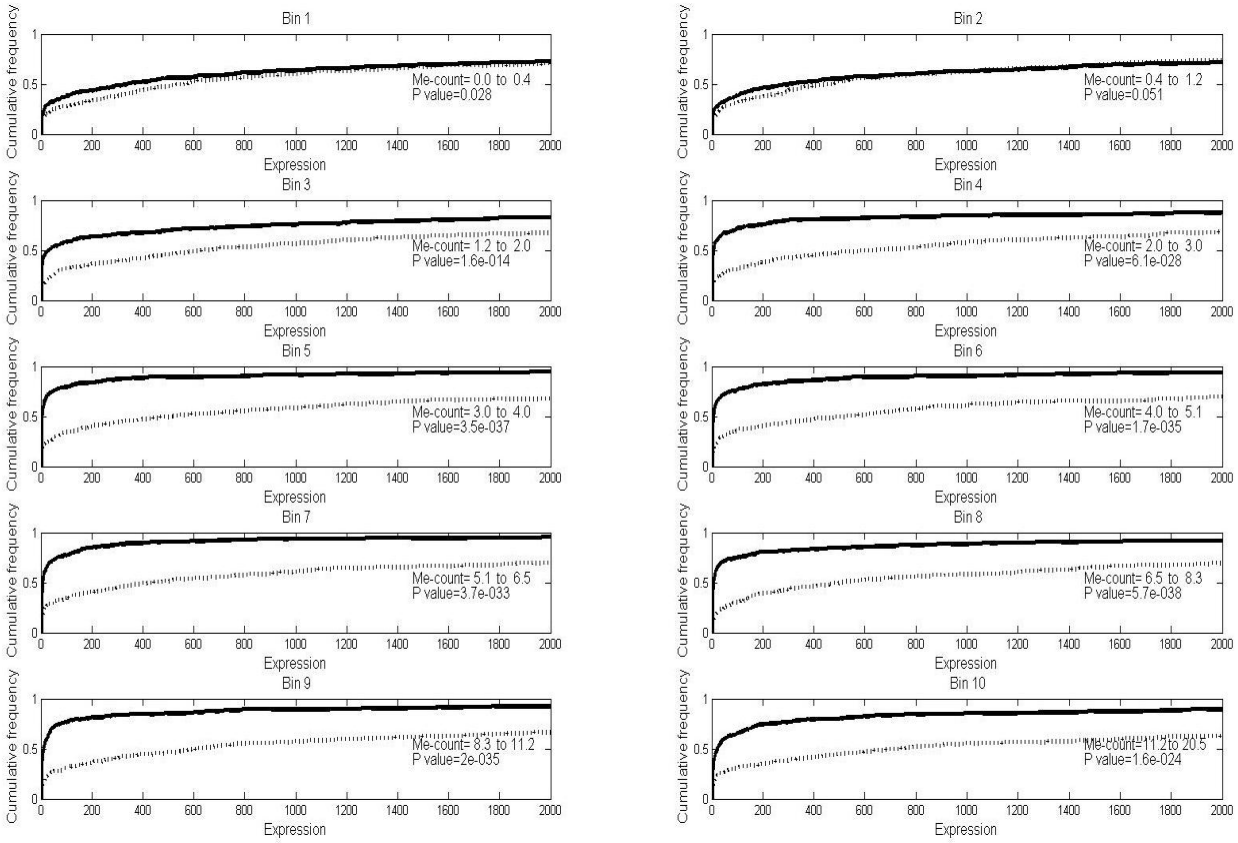


Figure S18

KD2Ma KD data from Blackledge et al., 2010

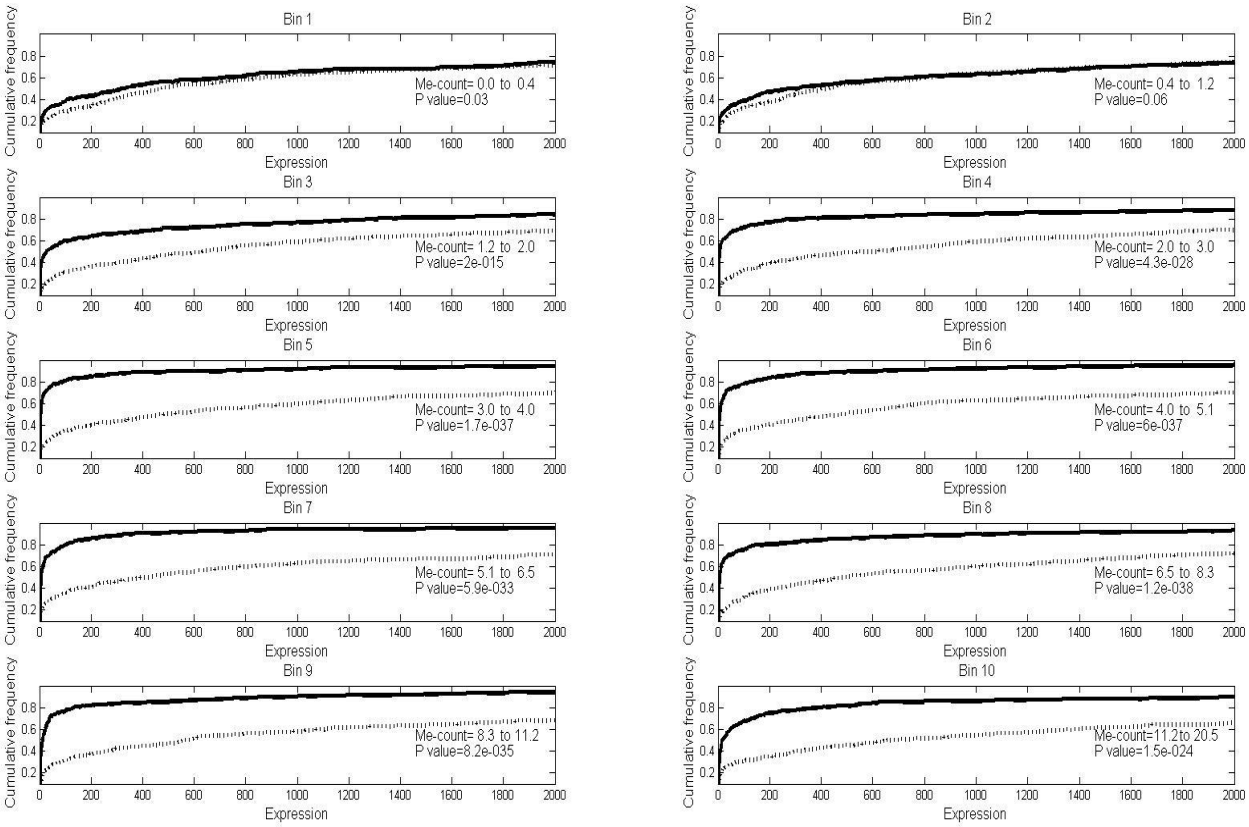


Figure S19

20 bins