Figure S1

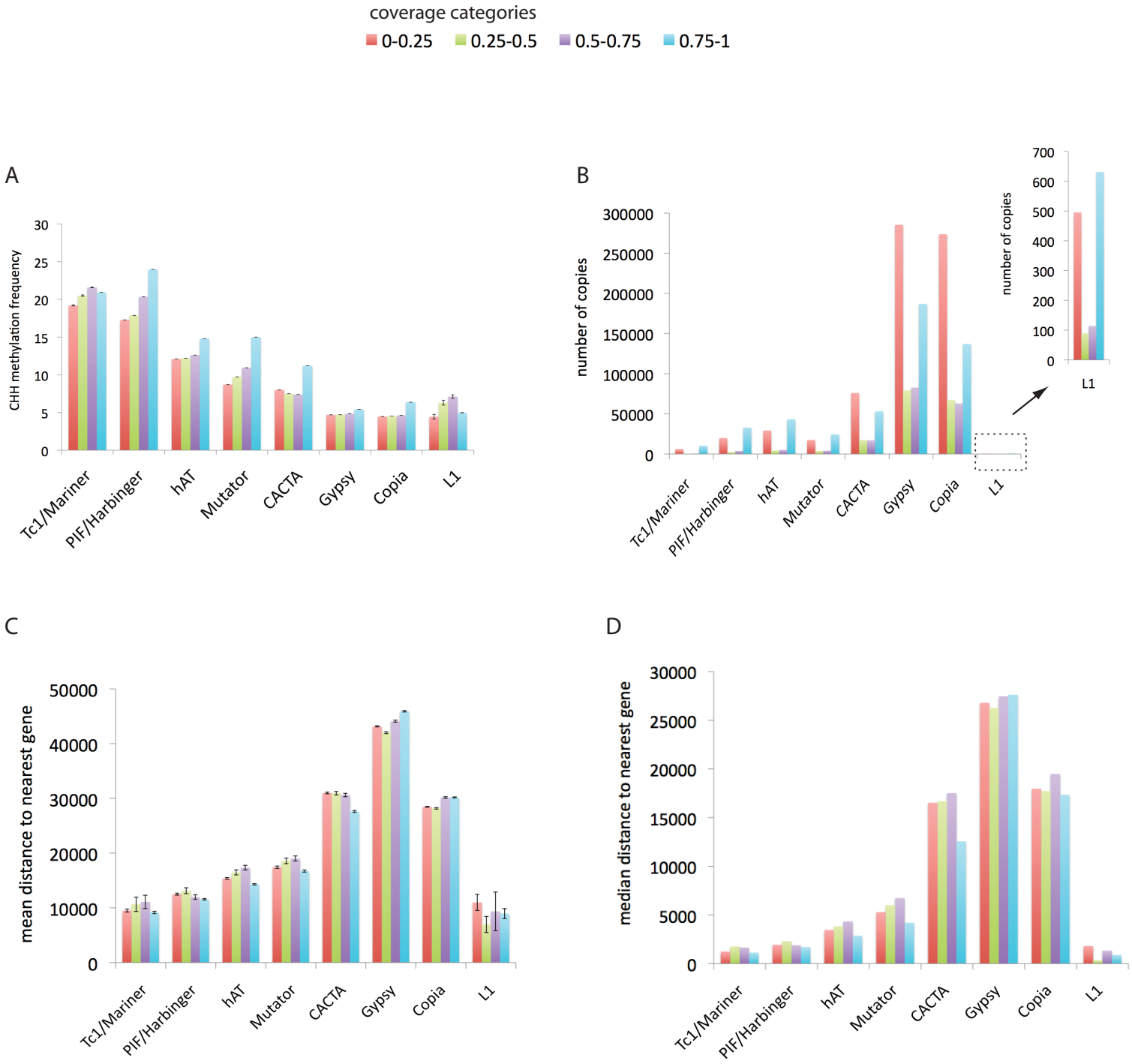
coverage categories
0-0.25
0.25-0.5
0.5-0.75
0.75-1
A
CHH methylation frequency
30
25
20
15
10
5
0
Tc1/Mariner
PIF/Harbinger
hAT
Mutator
CACTA
Gypsy
Copia
L1
B
number of copies
300000
250000
200000
150000
100000
50000
0
number of copies
700
600
500
400
300
200
100
0
L1
Tc1/Mariner
PIF/Harbinger
hAT
Mutator
CACTA
Gypsy
Copia
L1
C
mean distance to nearest gene
50000
40000
30000
20000
10000
0
Tc1/Mariner
PIF/Harbinger
hAT
Mutator
CACTA
Gypsy
Copia
L1
D
median distance to nearest gene
30000
25000
20000
15000
10000
5000
0
Tc1/Mariner
PIF/Harbinger
hAT
Mutator
CACTA
Gypsy
Copia
L1

Figure S1. Comparisons of transposon read coverage with CHH methylation and with distance to genes

(A) Read coverage vs CHH methylation. Transposons copies were categorized into quartiles based on how well each one was covered by bisulfite reads, from no coverage (a coverage value of zero) to full length coverage (reads spanning the entire copy, a coverage value of one). Copy number and sequence identity are both reflected in the number of reads that can be uniquely aligned to a particular transposon. The average CHH methylation (5-methylcytosine per total cytosines in the CHH context) was calculated for each transposon superfamily in each quartile. Error bars depict standard errors of the means.

(B) Read coverage vs transposon copy number. The number of transposon copies in each coverage quartile is shown. Since the number of L1 LINE copies was too small to be visible at the same scale as the others, it is blown up as an inset.

(C) Read coverage vs mean distance to the nearest gene. The distance to the nearest gene was obtained for each transposon copy, then averaged for each coverage quartile and superfamily. Distances are measured in bp. Error bars depict standard errors of the means.

(D) Read coverage vs median distance to the nearest gene. Since mean distances can be inflated by large intergenic areas, median distances between transposon copies and nearest genes were also obtained and displayed as in (C).

Figure S2

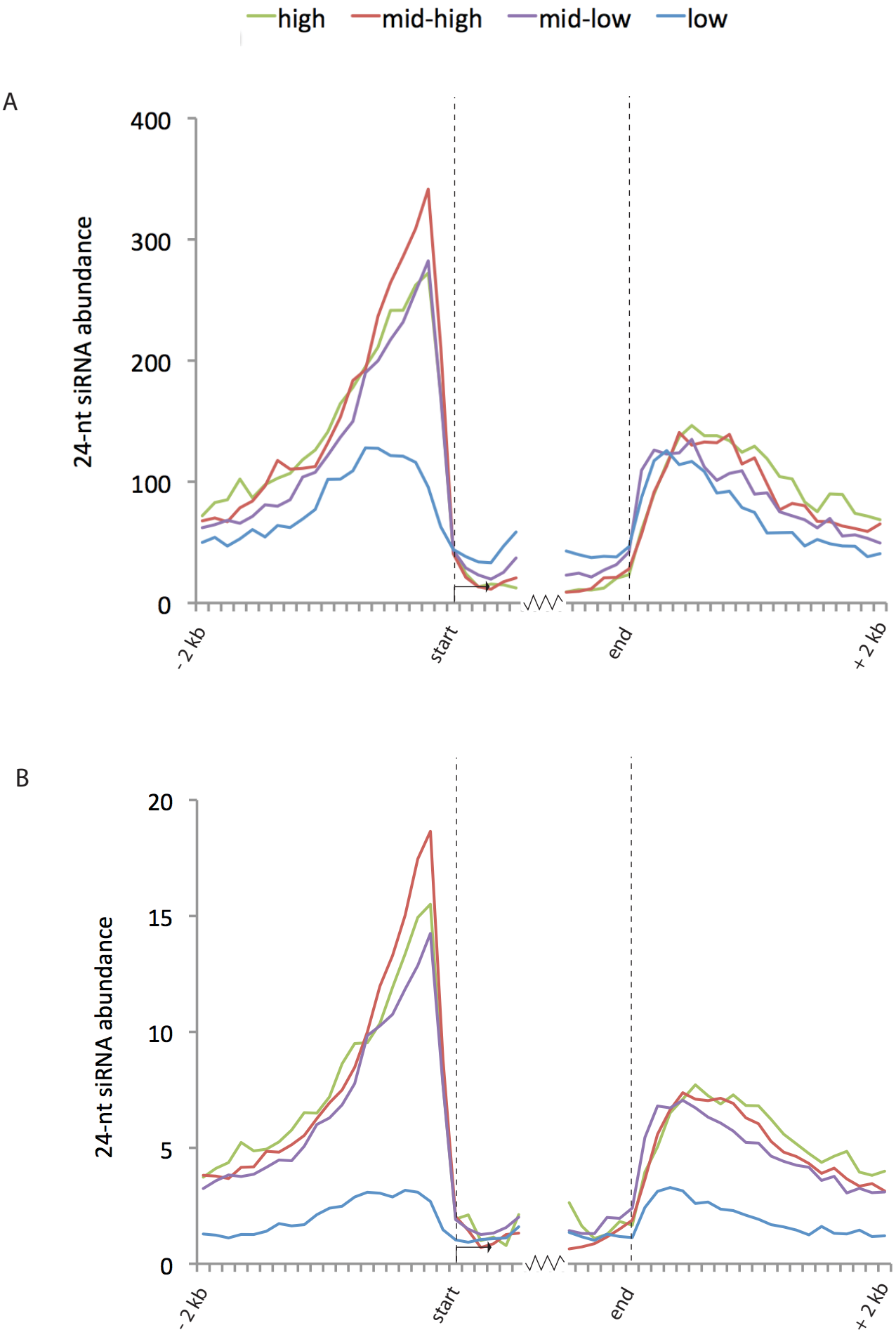
high
mid-high
mid-low
low
A
24-nt siRNA abundance
400
300
200
100
0
- 2 kb
start
end
+ 2 kb
B
24-nt siRNA abundance
20
15
10
5
0
- 2 kb
start
end
+ 2 kb

Figure S2. Distributions of 24-nt siRNAs near genes

(A) Distributions of unique 24-nt siRNAs near genes (from the same tissue type as the methylation and mRNA analyses, outer developing ear). Unique-aligning siRNAs were counted in each 100-bp interval in a 2-kb region upstream and downstream of gene ends for all annotated genes in the filtered gene set (version 5b). siRNAs were also counted in the first 600-bp inside genes on each end. Genes were divided into four sets based on expression level. The average number of 24-nt siRNAs per gene that aligned uniquely within each 100-bp interval is displayed for each set of genes. Unique-aligning siRNAs are those with a single, high-confidence alignment position, here defined as a MAPQ value greater than or equal to 30.

(B) Distributions of total 24-nt siRNAs near genes (from root tips). siRNA reads derived from seedling root tips were obtained from a previous study (Gent et al. 2012). The average number of 24-nt siRNAs per gene that aligned within each 100-bp interval is displayed for each set of genes. Both uniquely aligning and repetitive siRNAs were included.

Figure S3

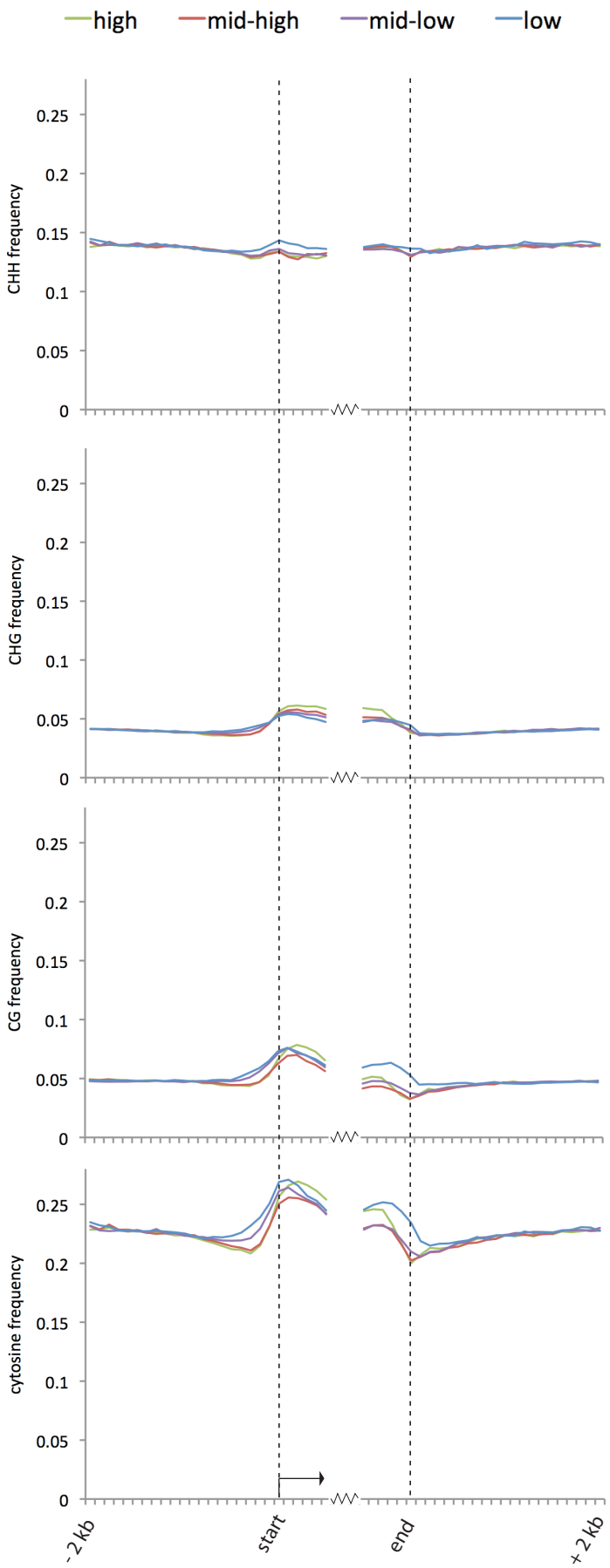
high
mid-high
mid-low
low
CHH frequency
CHG frequency
CG frequency
cytosine frequency
0.25
0.2
0.15
0.1
0.05
0
- 2 kb
start
end
+ 2 kb

Figure S3. Genomic CG, CHG, and CHH distributions near genes

Cytosine frequency was measured for each 100-bp interval in a 2-kb region upstream and downstream of gene ends for all annotated genes in the filtered gene set (version 5b). Cytosine frequency was also measured for the first 600-bp inside genes on each end. Genes were divided into four sets based on expression level. To measure the frequency of cytosine in each sequence context, the absolute methylation values in Figure 2B, D, and F (5-methylcytosines per total nucleotides) were divided by the relative methylation values in Figure 2A, C, and E, respectively (5-methylcytosines per cytosines in the specific sequence context), to yield cytosines per nucleotide. Also shown is the total cytosine frequency, the cumulative value of all three sequence contexts.

Figure S4

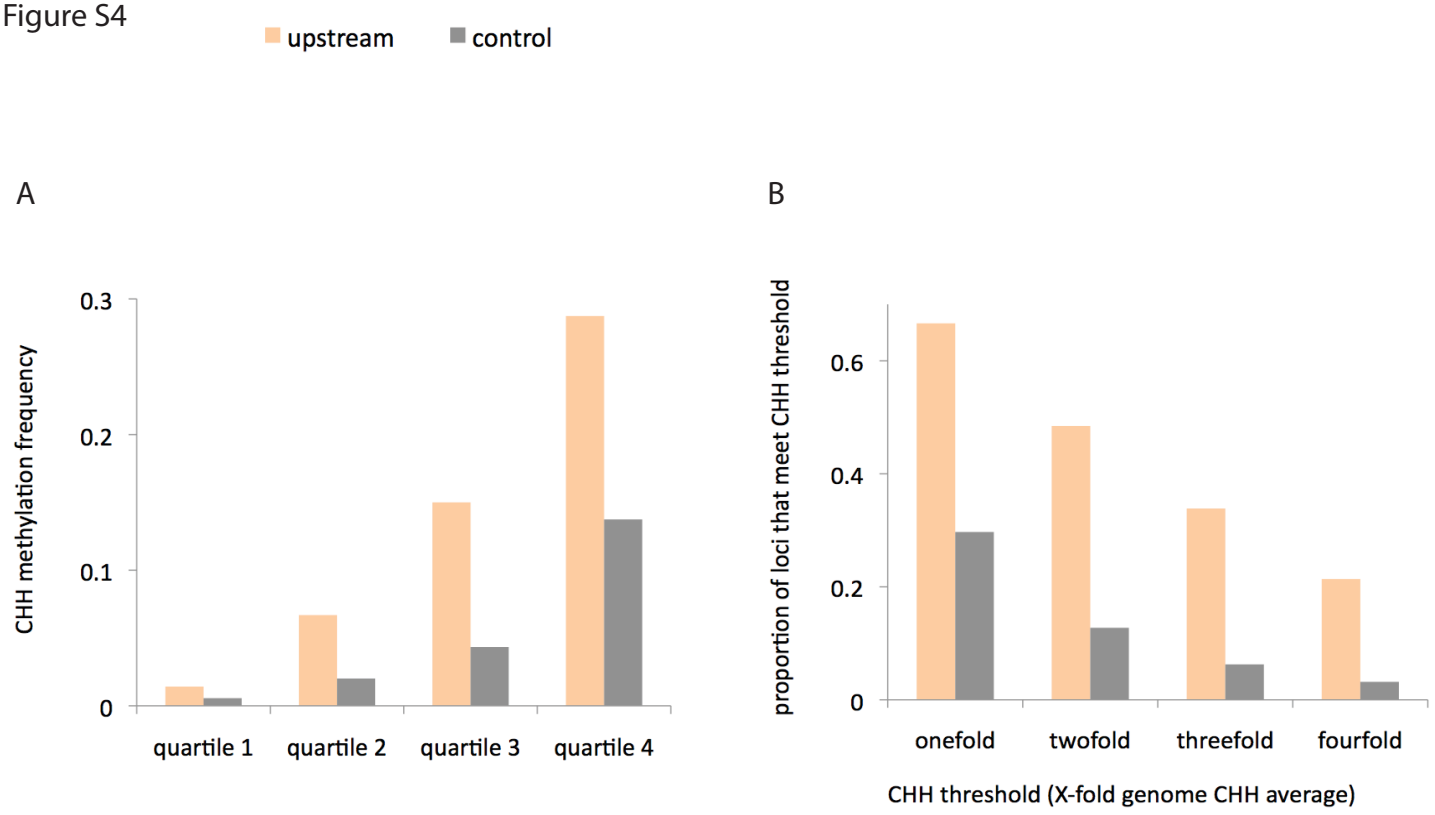
upstream
control
A
CHH methylation frequency
0.3
0.2
0.1
0
quartile 1
quartile 2
quartile 3
quartile 4
B
proportion of loci that meet CHH threshold
0.6
0.4
0.2
0
onefold
twofold
threefold
fourfold
CHH threshold (X-fold genome CHH average)

Figure S4. Comparison between CHH methylation levels upstream of genes and from a random sampling of the genome

(A) Average CHH methylation per quartile. Control 1-kb loci were sampled from across the genome. For both the upstream loci and controls, only those with greater than zero coverage of bisulfite reads were included in the analysis (a total of 38,212 upstream loci and 37,258 control loci). Each set of loci was split into four quartiles based on CHH methylation, with quartile 1 having the lowest and quartile 4 the highest. The height of each bar indicates the average CHH methylation in 5-methylcytosines per total cytosines in the CHH context. The upstream values were significantly larger than the controls for all four quartiles (p-value << 0.005)

(B) Proportion of loci whose CHH methylation exceeded the genomic average. The height of each bar depicts the proportion of loci whose average CHH methylation was at least as high as the indicated threshold value on the X axis (multiples of the genomic average, 0.0544 5-methylcytosines per total cytosines in the CHH context).