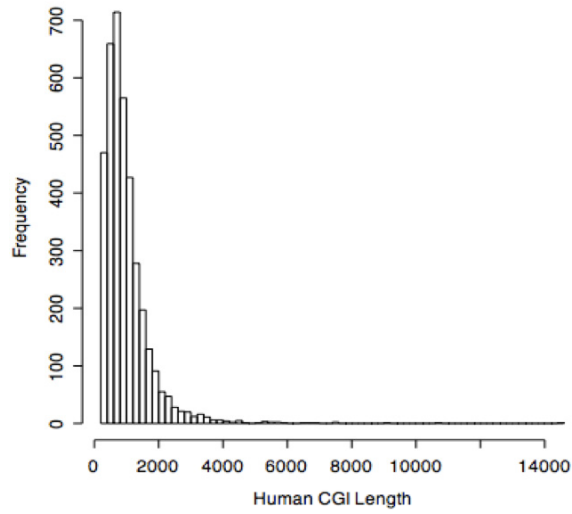


A. Human promoter-CGIs



B. Mouse promoter-CGIs

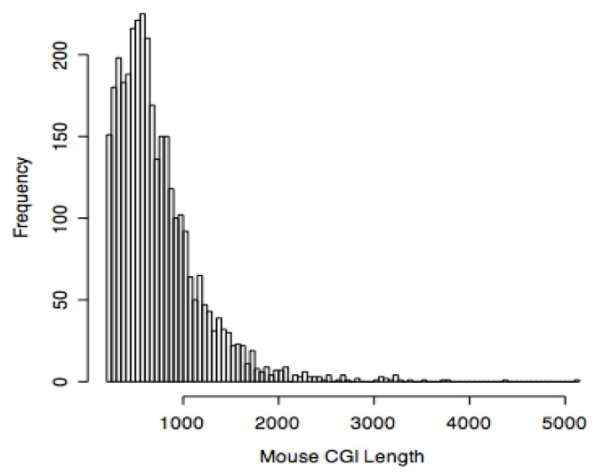


FIGURE S1.—Distribution of promoter-associated CGI lengths. Distribution of lengths of promoter-associated CpG islands (CGIs). Lengths distributions of human and mouse promoter-CGIs are shown in Figure S1. Mouse CGIs are shorter than human CGIs in general. In both species, the length distributions exhibit long tails. In other words, there are very long CGIs. The median CGI lengths are 921 and 800 nucleotides for human and mouse, respectively.

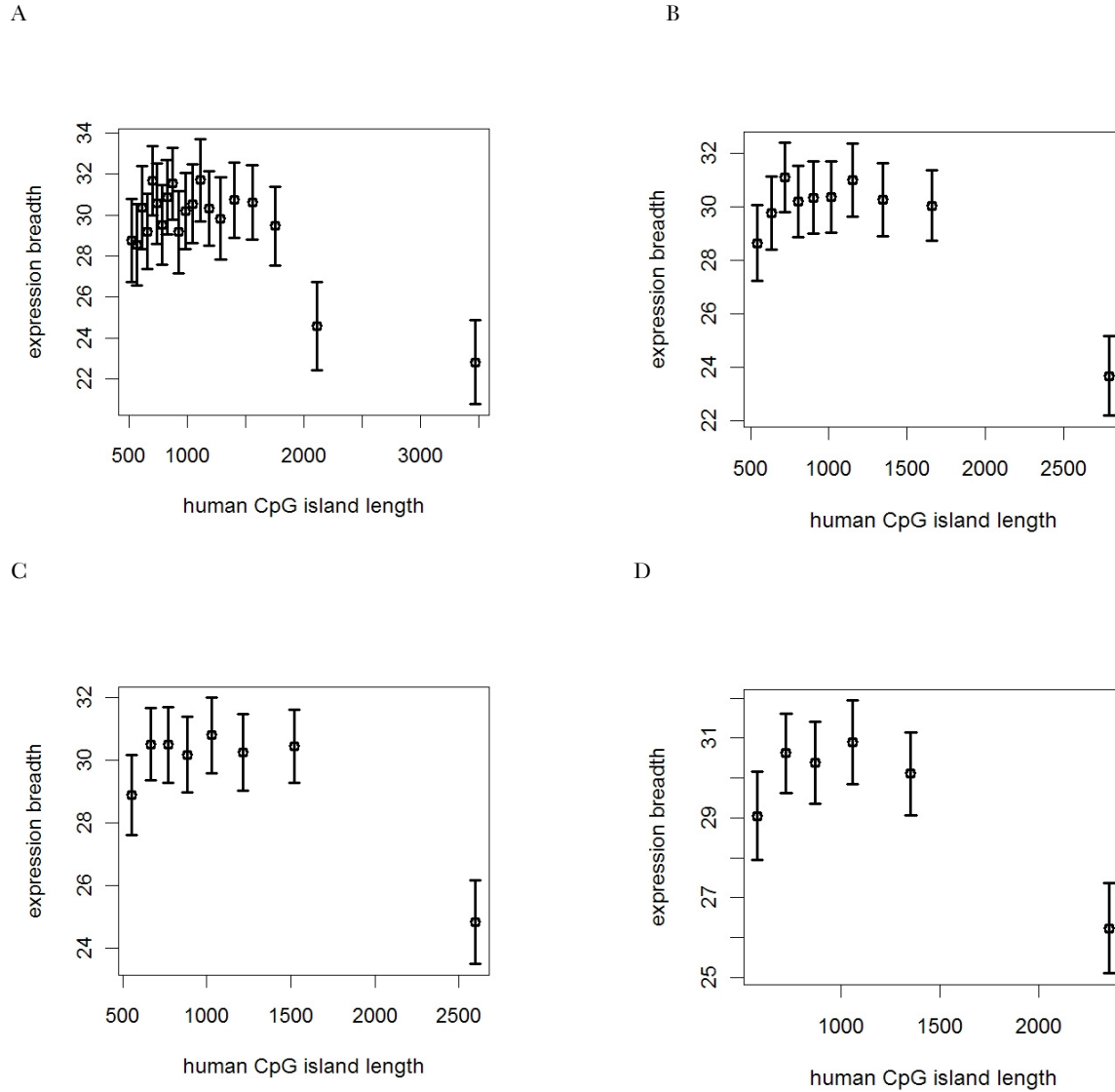


FIGURE S2. — Procedures Differentiating LCGI- and SCGI- promoters. We found that among genes harboring CpG islands in their promoters (CGI-promoters), a subset of genes with unusually long CpG islands in their promoters exhibit distinctively narrower gene expression breadths. This pattern was obvious when we used several different cutoff values to define ‘long-CpG island’ promoters (see below).

In Figure S2, we show results from human promoter analyses, using several cutoff values (5%, 10%, 12.5%, 15% in each bin). From the analyses of 20 bins (5% of the promoters in each bin; Figure S2A), we found that bins with mean $> \sim 2\text{Kb}$ showed reduced expression breadth compared to other bins. This value of $\sim 2\text{Kb}$ did not change when different bin sizes were used (10 bins, 8 bins, 6 bins, representing 10%, 12.5%, 17% cutoff: Figures S2B-D). We chose to use the cutoff value of 2kb, and defined CGI promoters with longer than 2kb CGI as ‘long-CGI promoters’. Promoters with CGI lengths shorter than 2kb are defined as ‘short-CGI promoters’.

The difference in expression breadths between SCGI and LCGI are highly significant ($P < 10^{-14}$, t -test). Using slightly different cutoff values near 2kb also provides highly significant results (for example, cutoff = 1600bps; $P < 10^{-10}$, 1800bps; $P < 10^{-12}$, 2200bps; $P < 10^{-10}$). However, the P -value was the smallest when 2kb was used, supporting that 2kb represents the value closest to the actual cutoff between LCGI and SCGI.

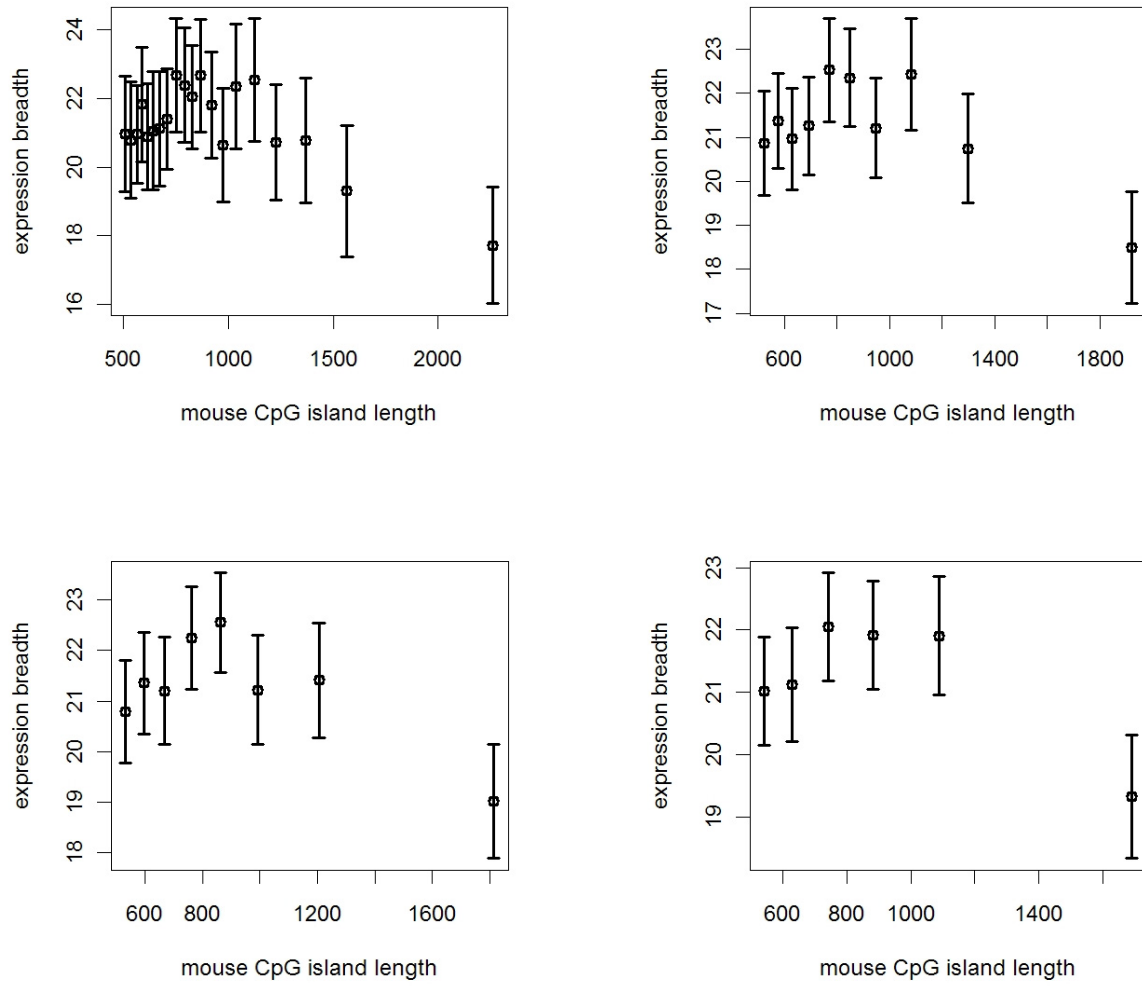


FIGURE S3.—20, 10, 8 and 6 bin analyses of mouse promoters. We performed a similar analysis using mouse promoters. Namely, we investigated distributions of gene expression breadths using different bin sizes (5%, 10%, 12.5%, 15%). We again observed that a subset of promoters with particularly long CpG islands, roughly longer than 1400bps, tend to exhibit narrower patterns of gene expression compared to promoters with short CGIs (Figure S3).

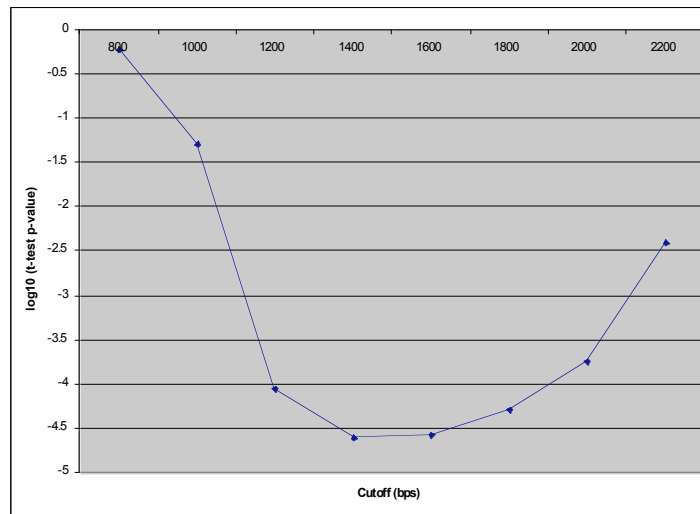


FIGURE S4.—Mouse promoters with CpG islands longer than 1400bps exhibit highly significantly different expression breadths compared to other CGI promoters. The difference between LCGI and other CGI promoters was not as conspicuous as in the case of human. Thus, to pinpoint the cutoff values between LCGI and SCGI promoters, we divided mouse CGI promoters to two groups using a specific cutoff value. Then we performed a *t*-test to determine whether the expression breadths of the two groups differ significantly. We performed this experiment using cutoff values ranging from 800bps to 2000bps, with the interval size 200bps. Figure S4 depicts the results of this analysis. At 1400bps, the two groups are highly significantly different ($P < 2 \times 10^{-5}$). Cutoff values near 1400bps provided similarly strong results (cutoff = 1200bps: $P < 10^{-4}$, 1600bps: $P = 2 \times 10^{-5}$). The *P*-value was the lowest when cutoff value of 1400bps was used. From these experiments, we chose 1400bps as the cutoff between LCGI and SCGI promoters.

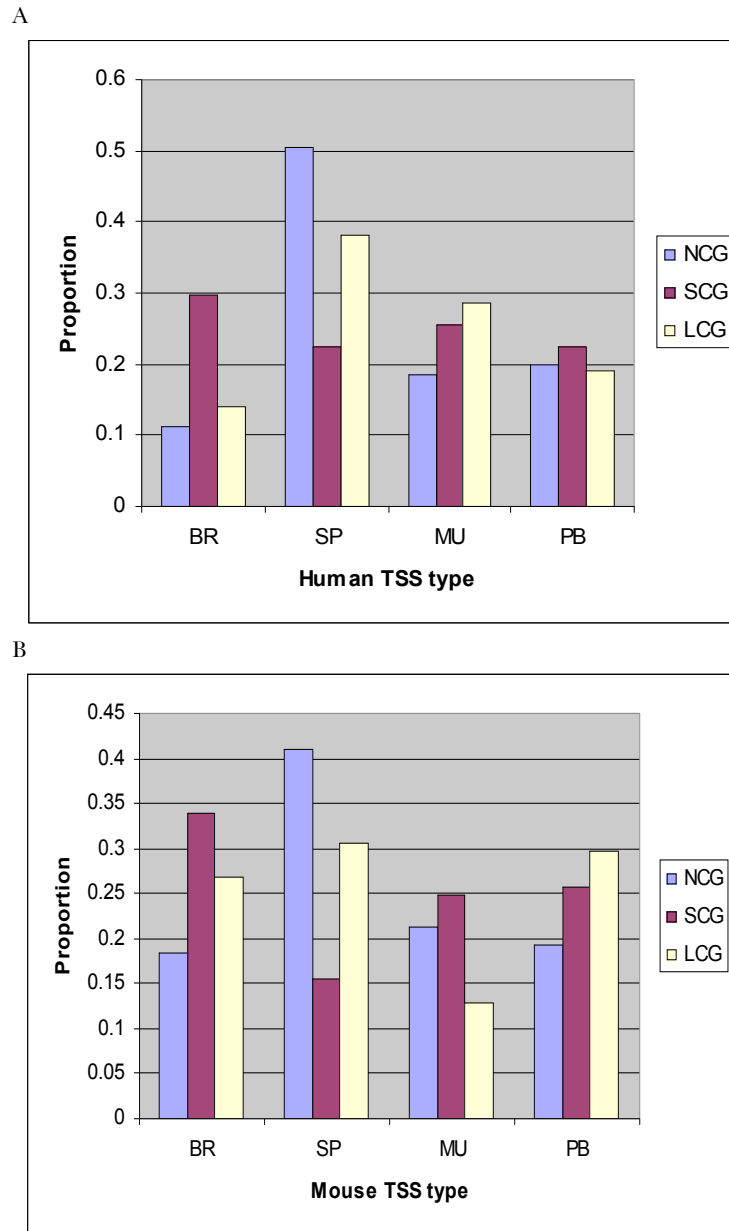


FIGURE S5.—Transcription Start Site Analyses using data from CAGE experiments. For tag clusters with number of tags > 100, Carninci et al. (2006) found that the distribution of the number of tags in a cluster falls into 4 broad categories (Broad [BR], Single dominant peak [SP], Bi- or Multi-modal [MU], and Broad Dominant Peak [PB]). We overlapped these different classes of TSSs on to the three different classes of promoters (LSCGI, NCGI, and SCGI). Within each promoter class (LCGI, SCGI, NCGI), we found the proportion of TSSs that belonged to each TSS class (BR, SP, MU, PB) in the human (Figure S5A) and the mouse (Figure S5B) genomes, respectively.

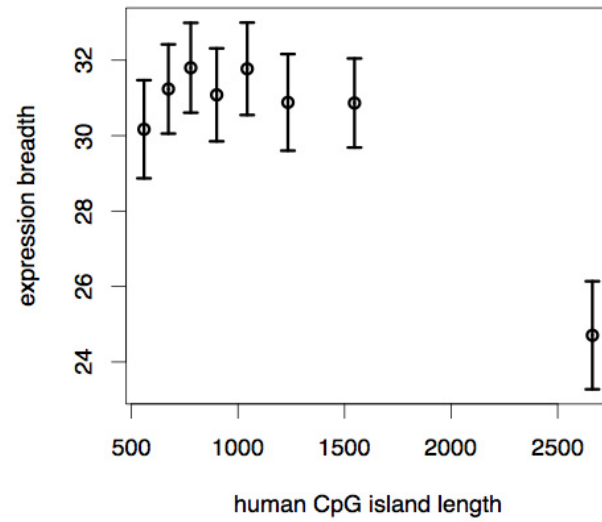


FIGURE S6.—Analyses of CGIs overlapping with TSS. To investigate whether CGIs that overlap with TSS exhibit different patterns than the one we report in the manuscript, we restricted our data set to those smacking on top of TSS. The following graph displays distribution of expression breadths re-calculated from this restricted data set. We observe that promoters with long CGIs (LCGI promoters) exhibit intermediate tissue-specificity.