**Supplementary Notes.**

**Supplementary Note 1: Heuristics for the discovery of TREs that change between species.**

We defined two types of changes in TRE activities between species. First, we used deSeq2 to identify species-specific changes in the abundance of Pol II at TREs that were active in all species. Second, we developed heuristic tests based on dREG scores to identify complete lineage-specific gains or losses of TREs. The rationale for, and validation of, this second set of tests is described in detail in this Supplementary Note.

A significant weakness of relying on deSeq2 alone is that this approach is overly conservative for identifying bona fide differences between species in TREs, particularly when levels of transcription are low (as with many eRNAs). Moreover, previous reports [(Villar et al. 2015; Vierstra et al. 2014)](https://paperpile.com/c/HXfSJ9/Rd8f+518w) suggest that evolutionary changes in TREs are common in mammals. Thus, filtering on false discovery rates from deSeq2 would result in very large numbers of false negatives. We therefore developed alternative criteria based on the dREG scores themselves, which are more sensitive to the transcriptional signatures of TREs than the raw read counts considered by deSeq2. This strategy can be considered analogous to the peak-calling strategy adopted in several previous studies [(Villar et al. 2015; Schmidt et al. 2012; Schmidt et al. 2010)](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) i[n](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) [t](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)h[a](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)ti[t](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) [d](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)e[p](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)e[n](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)d[s](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) [o](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)nac[u](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)s[t](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)o[m](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)i[z](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)e[d](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW),s[p](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)e[c](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)i[e](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)s[-](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)s[p](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)e[c](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)i[f](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)i[c](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) [p](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)r[o](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)c[e](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)s[s](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)i[n](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)go[f](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) [t](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)h[e](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) [d](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)a[t](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)a[,](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) [r](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)a[t](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)h[e](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)rt[h](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)a[n](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) [o](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)nt[h](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)e [raw](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) r[e](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)a[d](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW) [c](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)o[u](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)n[t](https://paperpile.com/c/HXfSJ9/Rd8f+tQdY+1lRW)s.

Our strategy selects putative differences between species using two thresholds in order to minimize errors where both species are near a single selected threshold. In particular, we select sites that both (1) exceed a stringent dREG score threshold (>0.3) in one or more of the species, indicating high-confidence presence of a TRE, and (2) fall below a permissive dREG threshold in at least one of the other species (<0.05), indicating high-confidence absence of a TRE. The stringent threshold (0.3) was selected because at this cutoff more than 93% of dREG sites identified in human CD4+ T-cells overlap another mark of active promoters and enhancers, including DNase-I, H3K27ac, H3K9ac, H4K14ac, H3K4me3, H3K4me1, or promoters and enhancers annotated using chromHMM [(Ernst and Kellis 2010)](https://paperpile.com/c/HXfSJ9/UIEk). The permissive threshold (0.05) was selected because at this threshold less than 6% of DNase-I hypersensitive sites that are also marked by H3K27ac remain to be identified, suggesting that relatively few true TREs exist below this threshold. By applying both of these thresholds in combination, we obtain a relatively stringent test for TREs that are present in at least one species and absent in at least one species, which require a gain or a loss event to explain.

Several lines of evidence support the idea that differences between species discovered using these two thresholds are highly enriched for bona fide evolutionary changes. First, differences between species on the basis of these criteria were highly enriched for statistically significant p-values estimated by deSeq2 based on the abundance of Pol II loading at these sites (**Supplementary Fig. 5**). Second, we used two independent statistical strategies [(Mosig et al. 2001; Nettleton et al. 2006; Phipson 2013)](https://paperpile.com/c/HXfSJ9/Lboi+IlUI+Mw8D) to estimate that, in not more than 10-15% of these cases, the null hypothesis of no evolutionary change is true, suggesting that 85-90% of these differences reflect bona fide gains and losses. Third, based on these criteria, gains and losses of TREs between treated and untreated samples were rare (though many TREs changed the abundance of transcription at TREs) (**Fig. 2a**) demonstrating that our criteria are robust when applied within a species. Fourth, the predicted gains and losses exhibit correlated changes in active histone modifications measured using orthogonal forms of genomic data in human CD4+ T-cells (**Fig. 2b**). Fifth, several differences between species at the *SGPP2* locus were validated experimentally using reporter gene assays (**Fig. 3; Supplementary Fig. 10**). These results suggest that our heuristic criteria are largely comprised of bona-fide differences between species.

Lastly, we also tested the sensitivity of our major results to the specific values of these thresholds. All of the major covariates that correlate with rates of change in enhancer activity between species were robust to reasonable changes in the threshold, or even to whether or not we filter on the false discovery rate in Pol II loading based on p-values estimated by deSeq2 (see **Supplementary Table 2**).