We will explore different standard pipelines and especially different parameters to detect CRE-gene linkages optimal for eQTL detection. However, as a preliminary study here, we focused on the CRE-gene linkages detected by a standard pipeline with default parameters from capture Hi-C data of 17 human blood cell types (11). Using these results, we tested the fraction of CREs outside of the 1 Mb threshold, referred to as distal CREs, to roughly estimate the abundance of lost eQTLs. These CRE-gene linkages we used are from capture Hi-C, which has a higher resolution than DNase Hi-C. And these CRE-gene linkages were detected by the standard pipeline, CHiCAGO, with stringent statistical cutoffs. By contrast, other capture Hi-C data of the samples in Table 1 are not appropriate for this preliminary study because they were processed by a pipeline used “in house” by the authors, and the distances between the CREs and their target gene are ambiguous. In the future, such capture Hi-C data will be processed with standard computational pipelines. At the moment, we found that in the 17 blood cell types, the fraction of distal CREs averaged over all of the expressed genes ranged from 4% to 13% with a mean of 10% (Fig. 1). The eQTLs in and near these distal CREs would be missed using the 1 Mb threshold.

Although the boundaries of a CRE are arbitrarily determined by the footprints of the restriction enzymes used, the distance between the boundaries of the CRE is still indicative of the actual CRE length. For each gene, we calculated the length of the distal CREs over the length of the total CREs. The average of these ratios is calculated over all of the genes of a sample, and this average ranges also from 4% to 13% for the 17 cell types. In addition, we divided the total length of all the identified CREs of a gene by 2 Mb. For each of the 17 blood cell types, the average percentage was consistently low, at around 2.1% (Fig. 1). This small percentage suggests that potentially many parts of the ±1Mb regions are not involved in regulating the target gene and thus many random genetic variants may be involved in eQTL detection, potentially rendering weak eQTLs difficult to be detected. We note that these numbers are crude estimations, and the proposed study will better estimate the actual improvements of using CREs for eQTL detection.