

## SUPPLEMENTARY FIGURE LEGENDS

**Supplementary Figure 1.** Tagging effectiveness and effects of minor allele frequency (MAF) in low LD regions of MHC region using three different tagging methods at same level of tagging efficiency. Tagging effectiveness was defined as the percentage of hidden SNPs which had haplotype correlations with tagging SNPs over a threshold (haplotype  $r^2 = 0.80$ ). The results indicate average efficiency and effectiveness obtained across all low LD regions in the 1 SNP/5 kb marker set of the entire 4.46 Mb human MHC region. The comparison was carried out with tagging efficiency at 1.2 (haplotype diversity = 97%, haplotype  $r^2 = 0.92$  and pair-wise  $r^2 = 0.50$ . White bars denote haplotype diversity based tagging method, grey bars denote minimum haplotype  $r^2$  based method, and dark bars denote pair-wise  $r^2$  based method.

**Supplementary Table 1** Tagging space of different tagging methods on hidden SNPs in low LD regions of ENCODE regions.

Tag set selected by	Tag set selected by	Tagging efficiency of 1 <sup>st</sup> sets from both methods	% tags shared <sup>a</sup>	% hidden SNPs detected <sup>b</sup>	% detected SNPs shared <sup>c</sup>	Tagging efficiency of “max shared” sets <sup>d</sup>	% tags shared <sup>ad</sup>	% hidden SNPs detected <sup>bd</sup>	% detected SNPs shared <sup>cd</sup>	% hidden SNPs detected by (1 <sup>st</sup> + ran) <sup>e</sup>	% detected SNPs shared <sup>ce</sup>
Hap $r^2$ 0.75	HapDiv 0.92	1.28 vs. 1.28	93.3	59.3 vs. 61.6	99.0	1.27 vs. 1.28	96.3	60.1 vs. 61.6	99.3	59.7 vs. 61.6	99.0
Hap $r^2$ 0.75	HapDiv 1.00	1.28 vs. 1.18	98.5	59.3 vs. 62.5	100	1.27 vs. 1.18	100	60.0 vs. 62.5	100	59.7 vs. 62.5	100
Hap $r^2$ 1.00	HapDiv 1.00	1.18 vs. 1.18	100	62.5 vs. 62.5	100	1.18 vs. 1.18	98.1	62.5 vs. 62.5	100	62.5 vs. 62.5	100
Hap $r^2$ 0.75	$r^2$ 0.50	1.28 vs. 1.28	91.1	59.3 vs. 58.3	97.4	1.27 vs. 1.28	95.6	59.7 vs. 58.9	98.6	59.7 vs. 58.3	98.1
Hap $r^2$ 0.75	$r^2$ 0.70	1.28 vs. 1.18	98.5	59.3 vs. 62.5	100	1.26 vs. 1.18	100	60.0 vs. 62.4	100	59.7 vs. 62.5	100
Hap $r^2$ 1.00	$r^2$ 0.70	1.18 vs. 1.18	95.9	62.5 vs. 62.5	100	1.18 vs. 1.18	98.1	62.5 vs. 62.5	100	62.5 vs. 62.5	100

<sup>a</sup> The percentage of tags in the smaller set shared by the larger set.<sup>b</sup> “Detected” means haplotype  $r^2 \geq 0.80$  between a hidden SNP and a set of tagging SNPs.<sup>c</sup> Each tagging SNP set produced a set of “detected” hidden SNPs. The value here indicated the percentage of SNPs in the smaller “detected” set shared by the larger “detected” set.<sup>d</sup> By comparing each set of tagging SNPs produced by two different tagging methods, the pair of tagging SNP sets sharing maximum proportion of tags was obtained.<sup>e</sup> By selecting random markers from non-tagging SNPs and adding them to the first tagging SNP set (column 3), equal values of tagging efficiency could be reached to those of the column 7. Values shown were the averages of five assessments (standard deviations found to be zero or close to zero).

**Supplementary Table 2** Tagging space of different tagging methods on hidden SNPs in high LD regions of MHC region

Tag set selected by	Tag set selected by	Tagging efficiency of 1 <sup>st</sup> sets from both methods	% tags shared <sup>a</sup>	% hidden SNPs detected <sup>b</sup>	% detected SNPs shared <sup>c</sup>	Tagging efficiency of “max shared” sets <sup>d</sup>	% tags shared <sup>ad</sup>	% hidden SNPs detected <sup>bd</sup>	% detected SNPs shared <sup>cd</sup>	% hidden SNPs detected by (1 <sup>st</sup> + ran) <sup>e</sup>	% detected SNPs shared <sup>ee</sup>
Hap $r^2 0.75$	HapDiv 0.91	3.0 vs. 3.0	73.6	72.8 vs. 76.1	93.9	2.95 vs. 2.94	91.5	75.3 vs. 74.9	96.4	(74.1±0.7) vs. (76.0±0.2)	93.9±0.1
Hap $r^2 0.75$	HapDiv 0.96	3.0 vs. 2.5	83.7	72.8 vs. 81.6	99.4	2.95 vs. 2.47	97.6	75.0 vs. 81.2	99.4	(74.1±0.7) vs. 81.6	99.4
Hap $r^2 0.89$	HapDiv 0.96	2.5 vs. 2.5	89.4	82.3 vs. 81.6	99.9	2.46 vs. 2.48	96.7	82.2 vs. 81.6	99.8	82.3 vs. 81.6	99.9
Hap $r^2 0.89$	HapDiv 1.00	2.5 vs. 2.0	92.3	82.3 vs. 83.1	100	2.43 vs. 2.0	98.8	82.2 vs. 83.1	100	82.3 vs. 83.1	100
Hap $r^2 0.75$	$r^2 0.52$	3.0 vs. 3.0	67.8	72.8 vs. 73.2	91.1	2.97 vs. 3.0	88.3	74.1 vs. 70.2	95.6	(74.1±0.7) vs. 73.2	91.5±0.5
Hap $r^2 0.75$	$r^2 0.68$	3.0 vs. 2.5	79.9	72.8 vs. 78.5	96.0	2.97 vs. 2.5	95.6	75.6 vs. 75.9	96.0	(74.1±0.7) vs. 78.5	95.0±0.1
Hap $r^2 0.89$	$r^2 0.68$	2.5 vs. 2.5	72.6	82.3 vs. 78.5	100	2.46 vs. 2.5	89.6	82.3 vs. 77.9	100	82.3 vs. 78.5	100
Hap $r^2 0.89$	$r^2 0.89$	2.5 vs. 2.0	92.8	82.3 vs. 83.0	100	2.45 vs. 2.0	100	82.3 vs. 82.9	100	82.3 vs. 83.0	100

<sup>a</sup> The percentage of tags in the smaller set shared by the larger set.

<sup>b</sup> “Detected” means haplotype  $r^2 \geq 0.80$  between a hidden SNP and a set of tagging SNPs.

<sup>c</sup> Each tagging SNP set produced a set of “detected” hidden SNPs. The value here indicated the percentage of SNPs in the smaller “detected” set shared by the larger “detected” set.

<sup>d</sup> By comparing each set of tagging SNPs produced by two different tagging methods, the pair of tagging SNP sets sharing maximum proportion of tags was obtained.

<sup>e</sup> By selecting random markers from non-tagging SNPs and adding them to the first tagging SNP set (column 3), equal values of tagging efficiency could be reached to those of the column 7. Values shown were the averages of five assessments. Standard deviations were only shown where results were different among the five assessments.

Supplementary Figure 1

