Supplementary Table 6. Proportions of unique start sites and genome coverage (by any unique tag) for color-space short tag alignments in hg19 RefSeq exon-exon junctions. Columns shown are length of tag matched; numbers in parentheses represent the number of mismatches allowed. Valid-adjacent errors are counted as 2 mismatches for all tag lengths except 25, where valid-adjacent errors are counted as a single mismatch. RefSeq libraries were constructed as described in Cloonan *et al*., 2009, where each side of the junction is the tag length -10nt (ie. for the 50nt tag lengths, 40nt either side of the exon-exon junction was concatenated to force a minimum of 10nt overlap of tags). The subsequent proportions of unique starts sites are lower than for genome based matching, due to the constraint on fixed length tags matching within a short window. The maximum possible proportion of start sites for each length is detailed below.

	25(1)	**35(3)**	**50(5)**	**60(6)**
Unique start sites	17.7%	28.4%	35.7%	38.6%
Maximum possible start sites	20.0%	32.0%	38.8%	41.0%
Coverage by unique tags	86.6%	87.9%	94.4%	95.6%