Supplementary Table 5. Proportions of unique start sites and genome coverage (by any unique tag) for nucleotide-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. 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)	**30(1)**	**35(1)**	**50(2)**	**60(3)**	**75(4)**	**90(5)**
Unique start sites	17.7%	24.4%	28.5%	36.2%	38.9%	41.5%	43.0%
Maximum possible start sites	20.0%	27.5%	32.0%	38.8%	41.0%	43.1%	44.4%
Coverage by unique tags	86.6%	87.5%	88.1%	95.0%	95.7%	96.5%	96.9%