



Comments and definitions:

- 1) A mapping with 2 or more introns is oriented as follows. If the mapping has no reverse canonical junctions and at least twice as many forward as reverse junctions, it is assigned to the plus strand. The analogous criterion is used to assign a mapping to the minus strand. If a mapping fits neither of these criteria, it is considered to have inconsistent splice signals and is not assigned an orientation.
- 2) A poly-A-tail in a transcript sequence is defined as a contiguous initial or terminal stretch of at least 6 Ts or As, respectively, not mapped to the genome, but within 1 bp from the start of the genomic mapping. This poly-A tail criterion was found to give the correct orientation in 99.4% of cases when tested on mappings oriented by their splice signals (GT-AG, GC-AG, AT-AC).
- 3) A poly-A-signal is defined as AATAAA or ATAAAA within the interval [-31, -10] relative to a 3' end. The directional annotation+poly-A signal criterion was found to give the correct orientation in 99.7% of cases when tested on mappings oriented by their splice signals.
- 4) Libraries with reliable annotation were identified by comparing splice signals to directional annotation for all spliced ESTs. A library is considered to have reliable directional annotation if the fraction of correctly oriented ESTs in the library was estimated to be >99% at the 99% confidence level.
- 5) The additional checks for signs of reversal are:
 - a) Check for reverse, possibly internal, poly-A-tail in transcript sequence (a tail can be internal in a transcript sequence if bordered by retained vector sequence). Defined as a stretch of at least 10 Ts, not mapped to the genome, but within 1 bp from the start of the genomic mapping.
 - b) Check for reverse canonical splices undetected due to low mapping quality. Such a splice is defined as a gap without recognized junctions that approximately (junctions shifted +1 bp) colocalizes with a canonical splice evidenced by a different mapping. Applied only to mappings with no previously detected introns and no poly-A-tail support.
 - c) Check for evidence of reversal combined with internal priming, defined as at least 10 Ts within a 14-base window in the region region [-11, +14] relative to the 5'-end of a mapping. Applied to unspliced EST-only mappings lacking poly-A-tail and -signal support.

Figure S1 Flowchart describing the procedure used to assign a mapping to a genomic strand. Accuracy rates for orientation of unspliced mappings were estimated as described in Methods. Ninety-five percent confidence intervals, calculated using the binomial distribution, for the estimates were (99.80%, 99.85%) for human cDNA mappings, (99.75%, 99.77%) for human EST mappings, (99.73%, 99.79%) for mouse cDNA mappings and (99.81%, 99.83%) for mouse EST mappings.