

BED - Insert File

The insert file provides information on the fragments (so-called "inserts") described by the mappings of paired-end read data. Inserts are reported after annotation mapping, with their insert sizes according to the processed (spliced) transcript coordinate distances in the reference annotation; however, as read-pair mappings before deconvolution are not always unique (neither they necessarily are after deconvolution), one read pair mapping can be represented by multiple lines in the insert file, one for each different insert size that can be hypothesized from the mapping of the read pair. On the other hand, mapping pairs that exceed the annotated transcript boundaries at one or both ends are ignored.

The insert file format is an adapted (truncated) BED format, sub-divided into the following the tab-separated columns:

Col. Nr	Tag	Description
1	chr	chromosome where all the following things are happening
2	start	left-most position of fragment on the reference genome
3	end	right edge of fragment marked by the read-pair
4	ID	read pair identifier
5	score	the <i>length</i> of the insert, on the processed underlying transcript structure
6	strand	transcription directionality of the annotated RNA from which the insert size has been derived, either '+' or '-'
7	(thick Start)	always 0 (not used)
8	(thick End)	always 0 (not used)
9	itemRGB	<p>3 comma-separated values: (1) number of current insert size estimation from a given read pair, (2) total number of different insert sizes estimated from that read pair, (3) number of different transcript annotations that are compatible with the read-pair mapping. Naturally, (1) (2) (3) holds true. A more unique transcriptome mapping describes (close to) one insert size: (1)~ (2)~ (3)~1.</p> <p><i>Remark:</i> when loaded in the UCSC genome browser, the 3 comma-separated values are interpreted as an RGB (red, green, blue) component vector of the color used for rendering the item; IMHO on "normal" transcript annotations/mappings you should always get values close to black (or darkdarkdarkdarkdarkgrey), however, I wouldn't take responsibility for any freaky color effects that can be produced by that.</p>