QC Alerts

The following table contains information explaining the "Alert Text" warnings that may be included in Illumina Data files

QC alert #	Alert text	Metric	Explanation of alert text
QC_1	**Portion of read pairs having SMART linker sequence is higher than expected.**	>50%	SMART linker is used in TranscritomeLite protocol to amplify mRNA out of small amount of total RNA to make a RNA-seq library. The portion of read pairs having SMART linker sequence at the beginning of either reads is reported, allowing up to two mismatches. Such reads may not align. Extremely high level of such reads may affect sequencing quality as well.
QC_2	**Portion of read pairs having Whole Genome Amplification primer sequence is higher than expected.**	>50%	The portion of read pairs having the GSC version of Whole Genome Amplification (WGA) primer sequence at the beginning of either reads is reported, allowing up to one mismatch. Such reads may not align. Extremely high level of such reads may affect sequencing quality as well.
QC_3	**Portion of read pairs partially mapped to known reagent sequences is higher than expected.**	>10%	Reagents are detected in reads through a BLAT search against a list of known reagents used at the GSC, excluding SMART and Nextera mate-pair junction. If a reagent appears within the 50bp of the start of the read, the read is classified as containing reagent. All reads containing reagents as a portion of chastity-passed reads are reported as reagent leftover. Such reads may not align.
QC_4	**Portion of read pairs with short or empty inserts is higher than expected.**	>10%	In a single-end mode non-miRNA library, it is the portion of reads containing 3' adapter sequence (i.e. AGATCGGA). In paired-end mode, it is the portion of read pairs having AGATCGGA in both forward reads and reverse reads, where AGATCGGA is the A tail added to the insert plus the beginning of 3' adapter. Some aligners may not be able to align such reads. The current test threshold is applicable to ~200bp insert fragments sequenced up to 100bp.
QC_5	**Portion of reads containing adapter sequence is higher than expected.**	>50%	Reads starting with ATCTCG, which is a sequence of A tail added to the insert plus the beginning of 3' miRNA adapter, is reported as adapter dimer. Such reads do not align.
QC_6	**Portion of read pairs with rare artifact is higher than expected.**	>10%	After filtering reads with polymers, SMART linker, WGA primer, mate-pair junction and other reagent sequences, the portion of read pairs is reported as shadow reads if the forward read share the same first 10bp with the paired reverse read.

QC_9	**Portion of 21bp tags mapped to human rRNA is higher than expected.**	>10% in human RNA-seq and RiboZero or >50% in human Ribominus	21 bp segments taken from the sequence data are compared to a collection of 21 base sequences from human ribosomal RNA. Excessive amount of rRNA decreases the yield of mRNA sequences, and it also reduces the power to detect inter-species contamination if rRNA is not filtered out.
QC_10	**Portion of 27bp tags mapped to mitochondrial DNA is higher than expected.**	Different depending on protocol and species	The portion of reads with their first 27bp matching expected mitochondrial genome is reported as mitochondrial content. Excessive amounts of mitochondrial DNA may decrease the yield of non-mitochrondial content.
QC_12	**Portion of 21bp tags mapped to an unexpected species is higher than expected.**	>4% in a non-xenograft	To detect contamination from other species, 21 bp segments taken from the sequence data are compared against a database of 21 base tags from multiple species. Should a high portion (>4% in non-xenograft) map to an unexpected species, possible contamination is indicated. For xenograft libraries, the host species reads are not considered in this category (see alert 18).
QC_13	**Portion of 18bp tags mapped to miRNAs is lower than expected.**	<30% in human	The portion of reads whose first 18bp match the first 18bp in an annotated miRNA of the expected species is reported.
QC_14	**Portion of reads not allocated to any indices is higher than expected.**	>20%	The portion of index reads matching none of the expected indices in a pool is reported, allowing one base mismatch between reads and an expected index.
QC_15	**Portion of reads allocated to multiple indices is higher than expected.**	>1% of the expected share of each library in a pool	The portion of index reads matching more than one expected indices in a pool is reported, allowing one base mismatch between reads and an expected index.
QC_17	**Raw read yield is lower than 10% of expected.**	< 10% of its expected share in a pool	A library's read yield is too low when it is less than 1/10 of its expected share. This applies to both uniform and intentionally unevenly pooled libraries. Libraries are assumed to be evenly pooled unless specified during submission.
QC_18	**Portion of 21bp tags mapped to host species is higher than expected**	>15% host-species reads in a xenograft library	To detect over-representation of the host species in a xenograft library, 21bp segments taken from the sequence data are compared against a database of 21 base tags from the host species. Should a high portion (>15%) map to the host species, possible contamination is indicated.
QC_19	**Bisulfite conversion rate of spiked-in lambda genome is lower than expected**	<97%	To measure bisulfite conversion rate, unmethylated lambda genomic DNA is spiked into sample DNA before bisulfite treatment. The portion of converted Cs in lambda reads is then reported as bisulfite conversion rate, only if

			>1000 lambda read pairs are detected.
QC_20	**Portion of reads containing adapter sequence is higher than expected.**	>50%	Reads containing 23bp of adapter sequence in a miRNA run are treated as adapter dimers. Such reads do not align.
QC_21	**Portion of reads mapping to target species ribosomal RNA is higher than expected.**	>10% in RNA-seq	Reads are compared against a set of known ribosomal sequences (rRNA) for the target species. Excessive amounts of rRNA decreases the yield of mRNA sequences, and it also reduces the power to detect inter-species contamination if rRNA is not filtered out.
QC_22	**Portion of reads mapping to mitochondrial DNA is higher than expected.**	Different depending on protocol and species	Reads are compared against a set of known mitochrondial sequences (mtDNA) for the target species. Excessive amounts of mtDNA may decrease the yield of non-mitochrondrial content.
QC_23	**Portion of reads classified as an unexpected species is higher than expected.**	>4% in a non-xenograft	To detect contamination from other species, reads that are not classified as target species, known reagents or known spike-in/vectors are compared against the genomes of multiple species. Should a high portion (>4% in a non-xenograft library) map to one or more unexpected species, possible contamination is indicated. For xenograft libraries, the host species reads are not considered in this category (see alert 25).
QC_24	**Portion of reads mapped to known miRNAs is lower than expected.**	<30% in human	The portion of reads whose first 18bp match the first 18bp in an annotated miRNA of the expected species is reported.
QC_25	**Portion of reads classified as host species is higher than expected.**	>15% host-species reads in a xenograft library	To detect over-representation of the host species in a xenograft library, reads are classified as either host or xenograft species. Should a high portion (>15%) map to the host species, possible contamination is indicated.
QC_26	**Too few lambda reads detected**	< 1000 lambda reads detected by Novoalign or < 10 lambda reads detected by BBT	Lambda is usually spiked into bisulfite libraries to measure the bisulfite conversion rate. If too few lambda reads are detected, it is difficult to accurately assess the success of the bisulfite conversion reaction.
QC_27	**Percent of reads matching target genome is below threshold**	Low Matched target genome	The number of chastity-passed, reagent-removed reads classified by BBT to match the target species is below the threshold value. The exact value of the threshold will depend on the species and protocol