

# RNA-seq Metrics

## Read Count Metrics

The following summary statistics are calculated by counting the number of reads that have the given characteristics.

### Total Reads

Sample	Note	Total Purity Filtered Reads Sequenced	Alternative Alignments	Failed Vendor QC Check	Read Length	Estimated Library Size
XXXX	XXXX	90,442,064	17,590,890	NA	75	36,238,570

**Total Purity Filtered Reads Sequenced** are filtered for vendor fail flags and exclude alternative alignment reads. **Alternative Alignments** are duplicate read entries providing alternative coordinates. **Failed Vendor QC Check** are reads which have been designated as failed by the sequencer. **Read Length** is the maximum length found for all reads. **Estimated Library Size** is the number of expected fragments based upon the total number of reads and duplication rate assuming a Poisson distribution.

### Mapped Reads

Sample	Note	Mapped	Mapping Rate	Mapped Unique	Mapped Unique Rate of Total	Unique Rate of Mapped	Duplication Rate of Mapped	Base Mismatch Rate	rRNA	rRNA Rate
XXXX	XXXX	85,795,762	0.949	50,290,104	0.556	0.586	0.414	0.003	1,392,126	0.015

**Mapped** reads are those that were aligned. **Mapping Rate** is per total reads. **Mapped Unique** are both aligned as well as non-duplicate reads. **Mapped Unique Rate of Total** is per total reads. **Unique Rate of Mapped** are unique reads divided by all mapped reads. **Duplication Rate of Mapped** is the duplicate read divided by total mapped reads. **Base Mismatch Rate** is the number of bases not matching the reference divided by the total number of aligned bases. **rRNA** reads are non-duplicate and duplicate reads aligning to rRNA regions as defined in the transcript model definition. **rRNA Rate** is per total reads.

### Mate Pairs

Sample	Note	Mapped Pairs	Unpaired Reads	End 1 Mapping Rate	End 2 Mapping Rate	End 1 Mismatch Rate	End 2 Mismatch Rate	Fragment Length Mean	Fragment Length StdDev	Chimeric Pairs
XXXX	XXXX	42,871,992	NA	0.949	0.949	0.002	0.003	163	441	NA

**Mapped Pairs** is the total number of pairs for which both ends map. **Unpaired Reads** are the number of reads that are lacking a mate. **End 1/2 Mapping Rate** is the number of mapped divided by the total number of End1/End2 reads. **End 1/2 Mismatch Rate** is the number of End 1 and 2 bases not matching the reference divided by the total number of mapped End 1 and 2 bases. **Fragment Length Mean/StdDev** is the mean distance, standard deviation between the start of an upstream read and the end of the downstream one. Only fragments contained within single exons are used. **Chimeric Pairs** are pairs whose mates map to different genes.

### Transcript-associated Reads

Sample	Note	Intragenic Rate	Exonic Rate	Intronic Rate	Intergenic Rate	Split Reads	Expression Profiling Efficiency	Transcripts Detected	Genes Detected
XXXX	XXXX	0.706	0.087	0.619	0.293	701,003	0.083	50,287	14,038

All of the above rates are per mapped read. **Intragenic Rate** refers to the fraction of reads that map within genes (within introns or exons). **Exonic Rate** is the fraction mapping within exons. **Intronic Rate** is the fraction mapping within introns. **Intergenic Rate** is the fraction mapping in the genomic space between genes. **Split Reads** is the number of reads spanning an exon exon junction. **Expression Profile Efficiency** is the ratio of exon reads to total reads. **Transcripts/Genes Detected** is the number of transcripts/Genes with at least 5 reads.

### Strand Specificity

Sample	Note	End 1 Sense	End 1 Antisense	End 2 Sense	End 2 Antisense	End 1 % Sense	End 2 % Sense
XXXX	XXXX	10,646,982	16,941,235	16,965,719	10,658,302	38.592	61.417

**End 1/2 Sense** are the number of End 1 or 2 reads that were sequenced in the sense direction. Similarly, **End 1/2 Antisense** are the number of End 1 or 2 reads that were sequenced in the antisense direction. **End 1/2 Sense %** are percentages of intragenic End 1/2 reads that were sequenced in the sense direction.

## Coverage Metrics for Bottom 1000 Expressed

## Transcripts

The metrics in this table are calculated across the transcripts that were determined to have the highest expression levels.

Sample	Note	Mean Per Base Cov.	Mean CV	No. Covered 5'	5'200 Base Norm	No. Covered 3'	3' 200Base Norm	Num. Gaps	Cumul. Gap Length	Gap %
<a href="#">XXXX</a>	XXXX	2.51	3.22	281	1.25	295	1.623	5398	1838384	75.9

It is important to note that these values are restricted to the bottom 1000 expressed transcripts. 5' and 3' values are per-base coverage averaged across all top transcripts. 5' and 3' ends are 200 base pairs. Gap % is the total cumulative gap length divided by the total cumulative transcript lengths.

## Coverage Metrics for Middle 1000 Expressed Transcripts

The metrics in this table are calculated across the transcripts that were determined to have the highest expression levels.

Sample	Note	Mean Per Base Cov.	Mean CV	No. Covered 5'	5'200 Base Norm	No. Covered 3'	3' 200Base Norm	Num. Gaps	Cumul. Gap Length	Gap %
<a href="#">XXXX</a>	XXXX	8.55	2.95	322	0.72	375	1.257	5888	1556762	64.9

It is important to note that these values are restricted to the middle 1000 expressed transcripts. 5' and 3' values are per-base coverage averaged across all top transcripts. 5' and 3' ends are 200 base pairs. Gap % is the total cumulative gap length divided by the total cumulative transcript lengths.

## Coverage Metrics for Top 1000 Expressed Transcripts

The metrics in this table are calculated across the transcripts that were determined to have the highest expression levels.

Sample	Note	Mean Per Base Cov.	Mean CV	No. Covered 5'	5'200 Base Norm	No. Covered 3'	3' 200Base Norm	Num. Gaps	Cumul. Gap Length	Gap %
--------	------	--------------------	---------	----------------	-----------------	----------------	-----------------	-----------	-------------------	-------

<a href="#">XXXX</a>	XXXX	241.68	1.4 0	670	0.42	721	0.681	200 3	318996	24 .7
----------------------	------	--------	----------	-----	------	-----	-------	----------	--------	----------

It is important to note that these values are restricted to the top 1000 expressed transcripts. **5'** and **3'** values are per-base coverage averaged across all top transcripts. 5' and 3' ends are 200 base pairs. Gap % is the total cumulative gap length divided by the total cumulative transcript lengths.

## Summary of Runtime Parameters

Option	Description	Value
Samples	Samples/Sample File used	XXXX.star.dd.bam XXXX
Transcript Model	GTF formatted file containing the transcript definitions	gencode.v35lift37.transcript.gtf
Reference Genome	The genome version to which the BAM is aligned	hg19.fa
Downsampling	For Coverage Metrics, the number of reads is randomly reduced to the given level	none
Detailed Report	The optional detailed report contains coverage metrics for every transcript	no details
rRNA Intervals	Genomic coordinates of rRNA loci	ribosomal_interval.list

Generated by [RNA-SeQC](#) v1.1.8.1 07/11/14

Run on Sat Dec 05 19:55:03 EST 2020