MultiQC

Sijan_Aphanomyces_Main

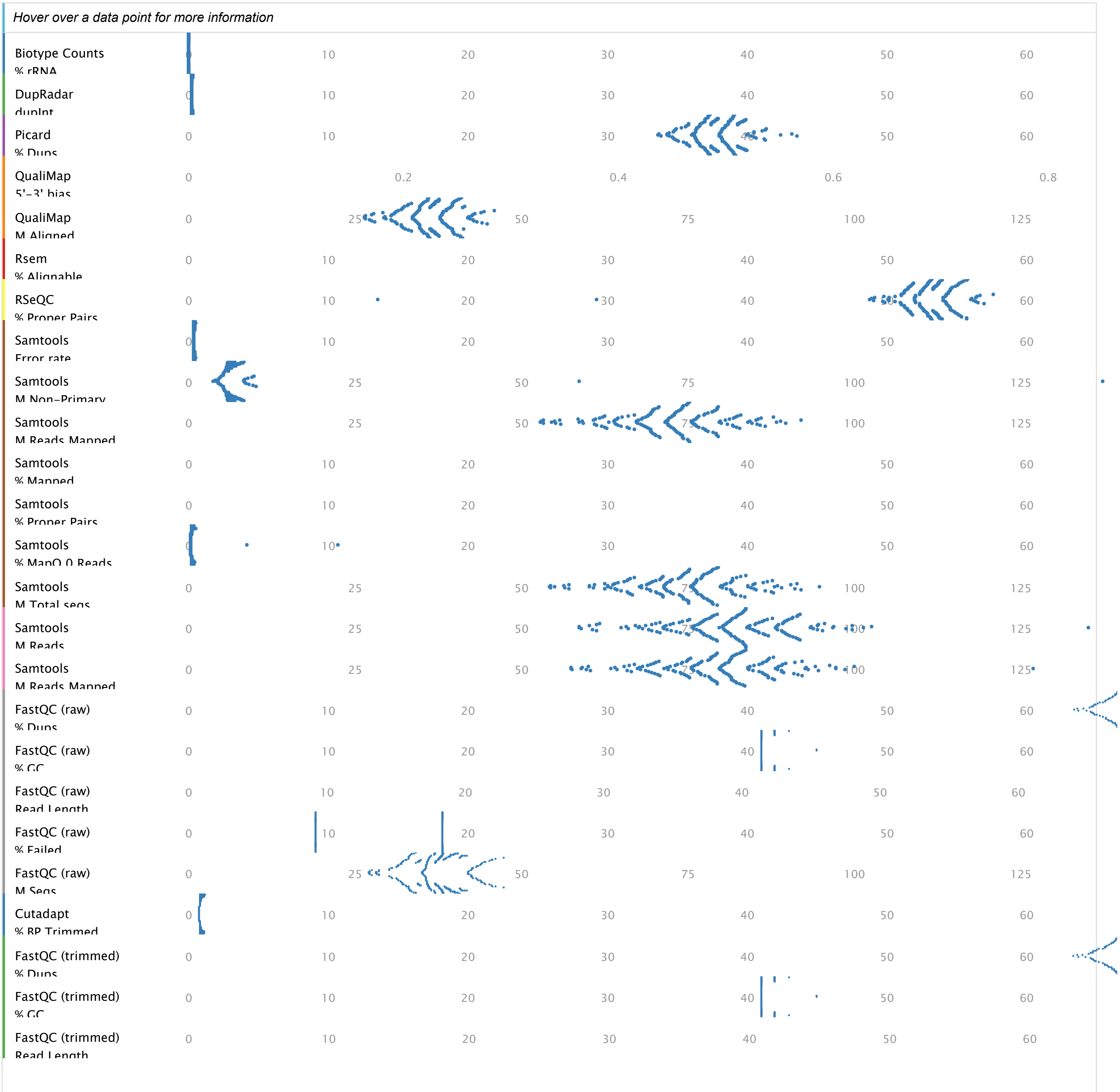
A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

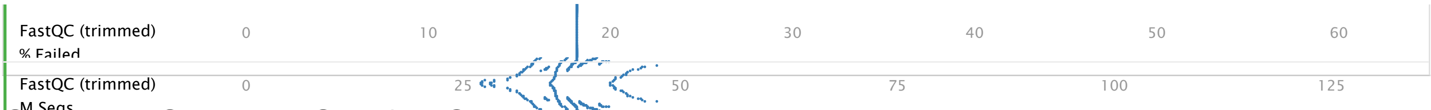
> This report has been generated by the nf-core/rnaseq analysis pipeline. For information about how to interpret these results, please see the documentation.

Report generated on 2023-01-29, 08:50 EST based on data in: `/isilon/projects/J-002368_BeeCSI/sijan_rnaseq_aphanomyces/work/1f/a1e3a9eb464adcdaa9bef6d4561b3e`

General Statistics

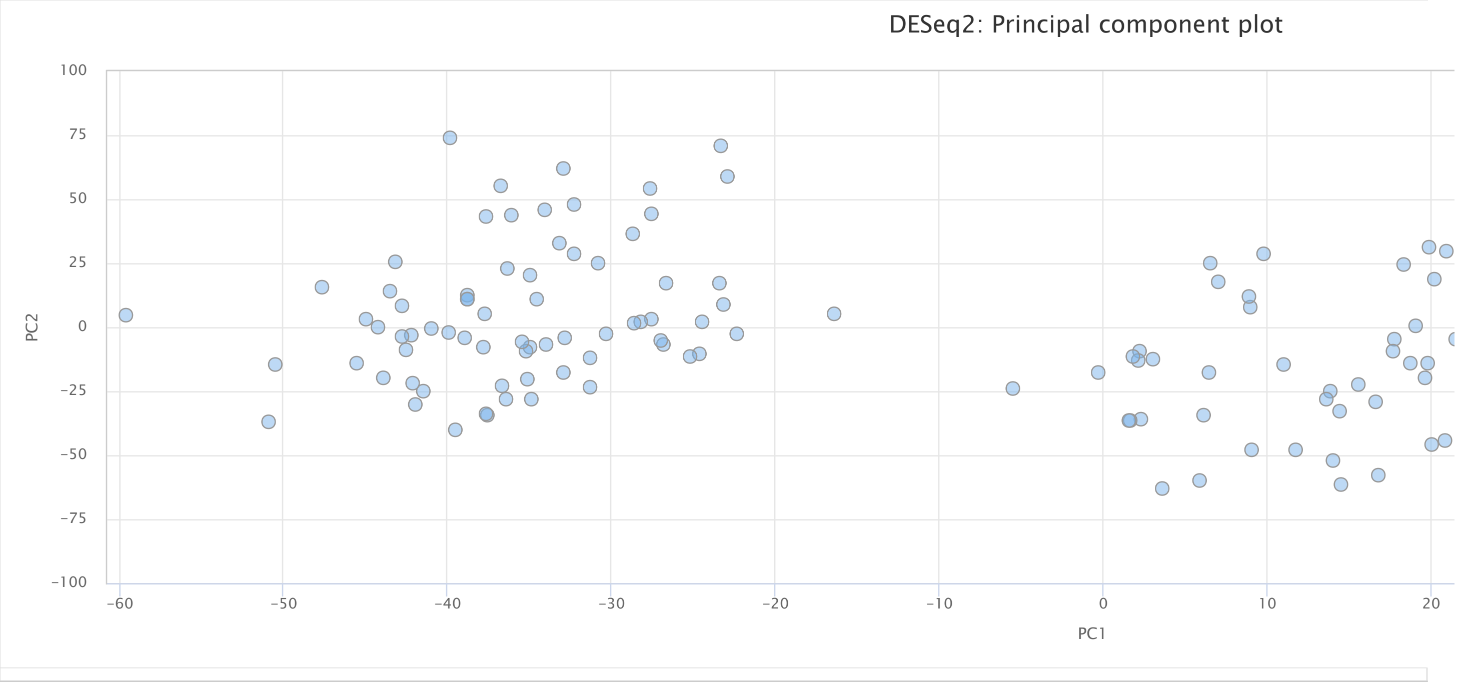
Showing 504 samples.





STAR_RSEM DESeq2 PCA plot

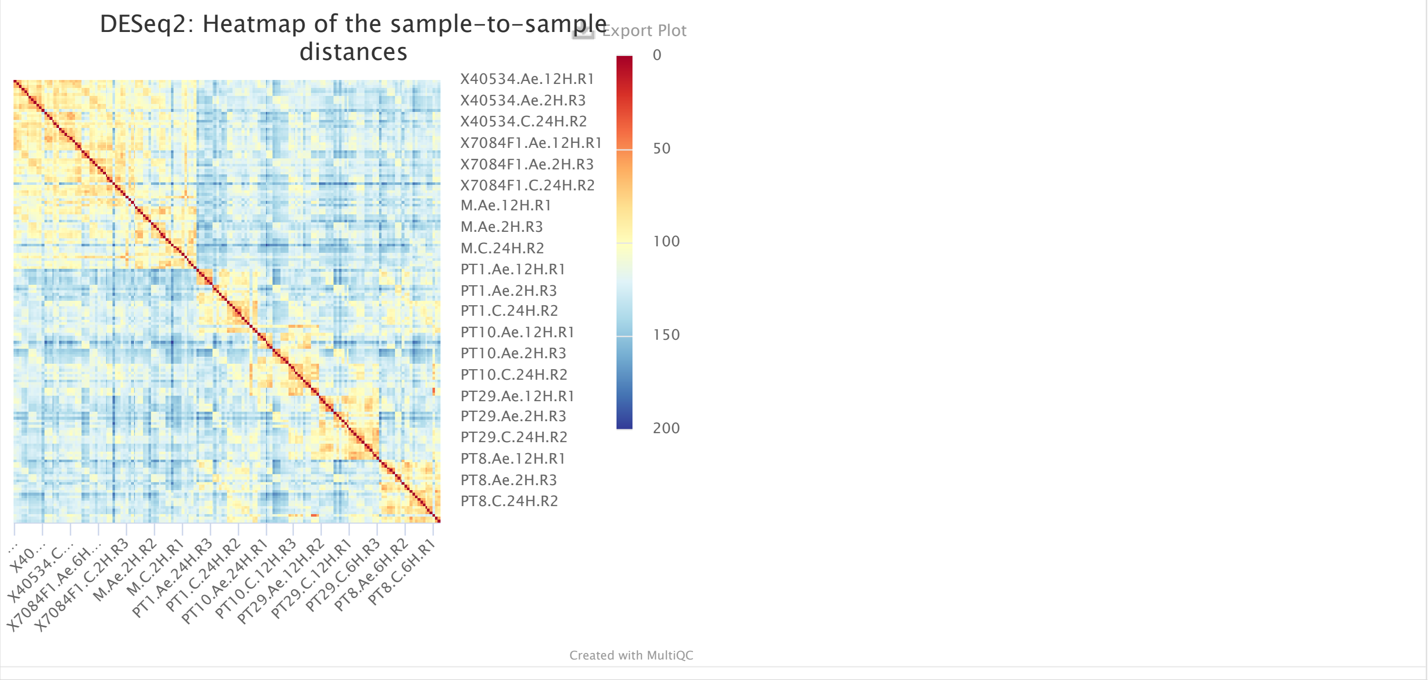
PCA plot between samples in the experiment. These values are calculated using DESeq2 in the `deseq2_qc.r` script.



STAR_RSEM DESeq2 sample similarity

is generated from clustering by Euclidean distances between DESeq2 rlog values for each sample in the `deseq2_qc.r` script.

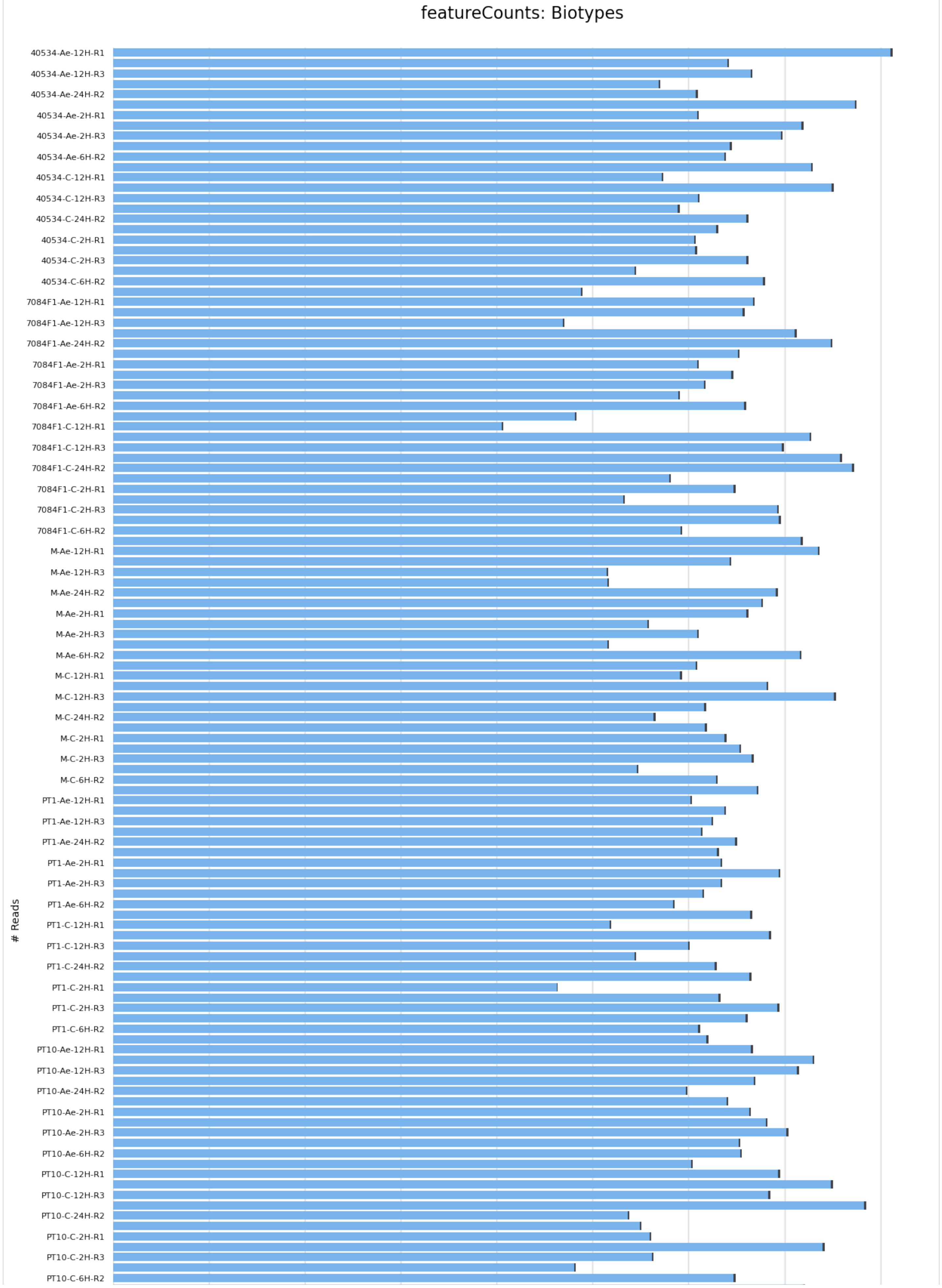
Sort by highlight **Min:** 0.0 **Max:** 196.4937

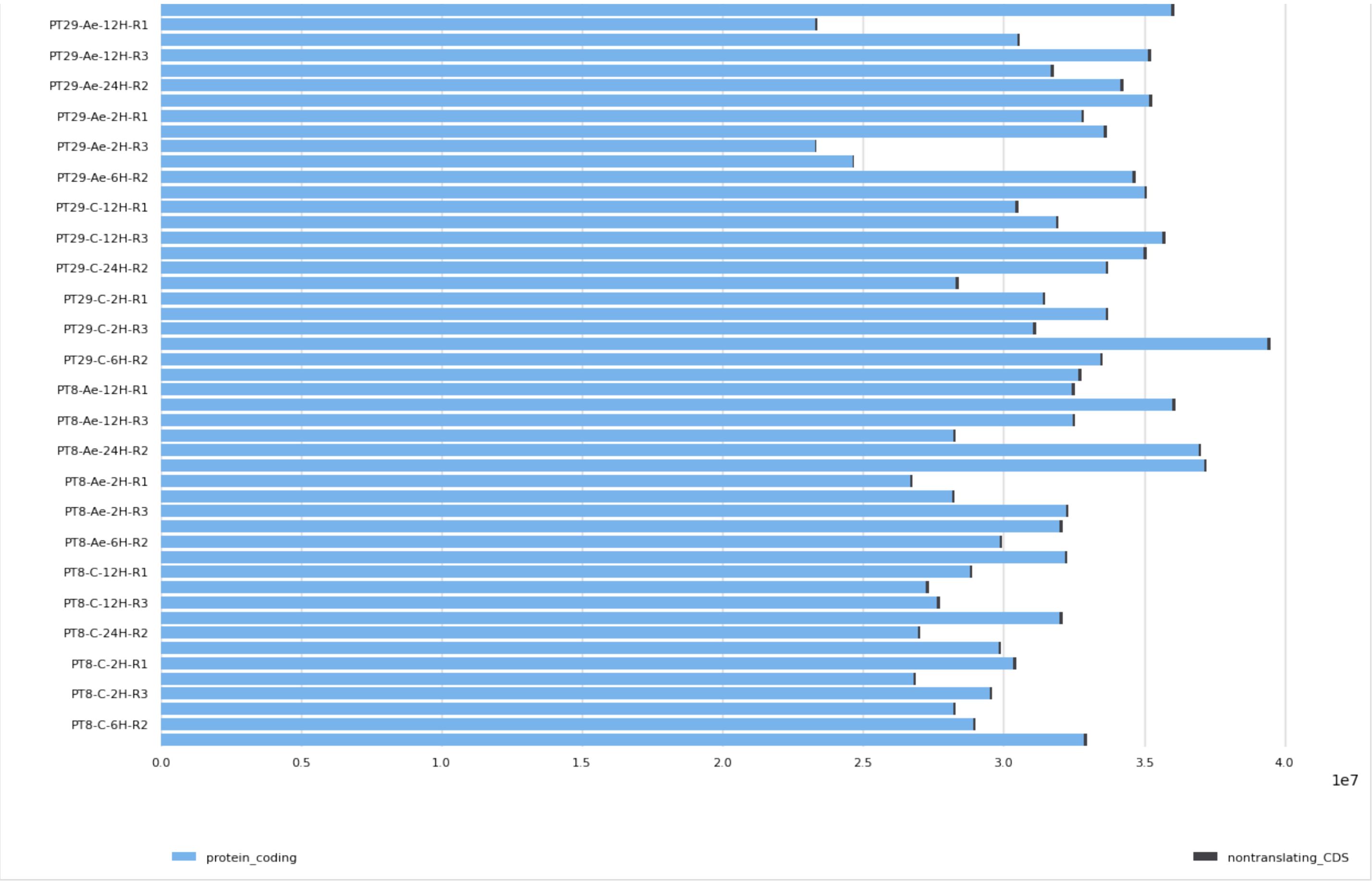


Biotype Counts

shows reads overlapping genomic features of different biotypes, counted by featureCounts.

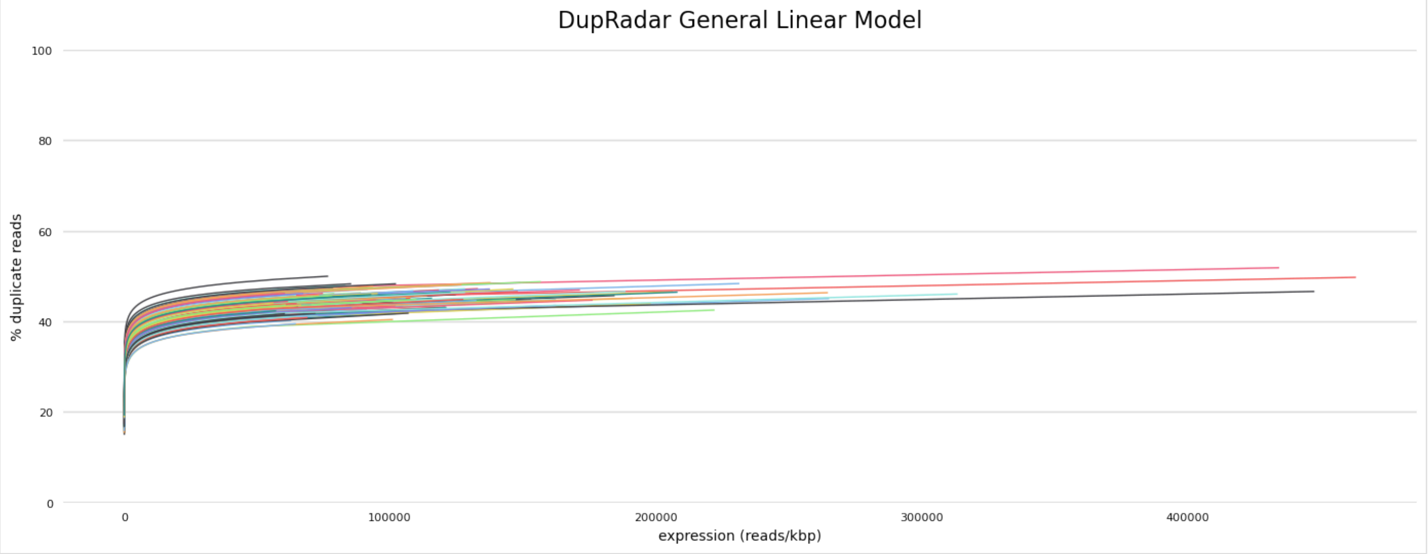
Number of Reads | Percentages


featureCounts: Biotypes
Reads
40534-Ae-12H-R1
40534-Ae-12H-R3
40534-Ae-24H-R2
40534-Ae-2H-R1
40534-Ae-2H-R3
40534-Ae-6H-R2
40534-C-12H-R1
40534-C-12H-R3
40534-C-24H-R2
40534-C-2H-R1
40534-C-2H-R3
40534-C-6H-R2
7084F1-Ae-12H-R1
7084F1-Ae-12H-R3
7084F1-Ae-24H-R2
7084F1-Ae-2H-R1
7084F1-Ae-2H-R3
7084F1-Ae-6H-R2
7084F1-C-12H-R1
7084F1-C-12H-R3
7084F1-C-24H-R2
7084F1-C-2H-R1
7084F1-C-2H-R3
7084F1-C-6H-R2
M-Ae-12H-R1
M-Ae-12H-R3
M-Ae-24H-R2
M-Ae-2H-R1
M-Ae-2H-R3
M-Ae-6H-R2
M-C-12H-R1
M-C-12H-R3
M-C-24H-R2
M-C-2H-R1
M-C-2H-R3
M-C-6H-R2
PT1-Ae-12H-R1
PT1-Ae-12H-R3
PT1-Ae-24H-R2
PT1-Ae-2H-R1
PT1-Ae-2H-R3
PT1-Ae-6H-R2
PT1-C-12H-R1
PT1-C-12H-R3
PT1-C-24H-R2
PT1-C-2H-R1
PT1-C-2H-R3
PT1-C-6H-R2
PT10-Ae-12H-R1
PT10-Ae-12H-R3
PT10-Ae-24H-R2
PT10-Ae-2H-R1
PT10-Ae-2H-R3
PT10-Ae-6H-R2
PT10-C-12H-R1
PT10-C-12H-R3
PT10-C-24H-R2
PT10-C-2H-R1
PT10-C-2H-R3
PT10-C-6H-R2



DupRadar

provides duplication rate quality control for RNA-Seq datasets. Highly expressed genes can be expected to have a lot of duplicate reads, but high numbers of duplicates at low read counts can indicate low library complexity with technical duplication. This plot shows the general linear models - a summary of the gene duplication distributions. .



Picard

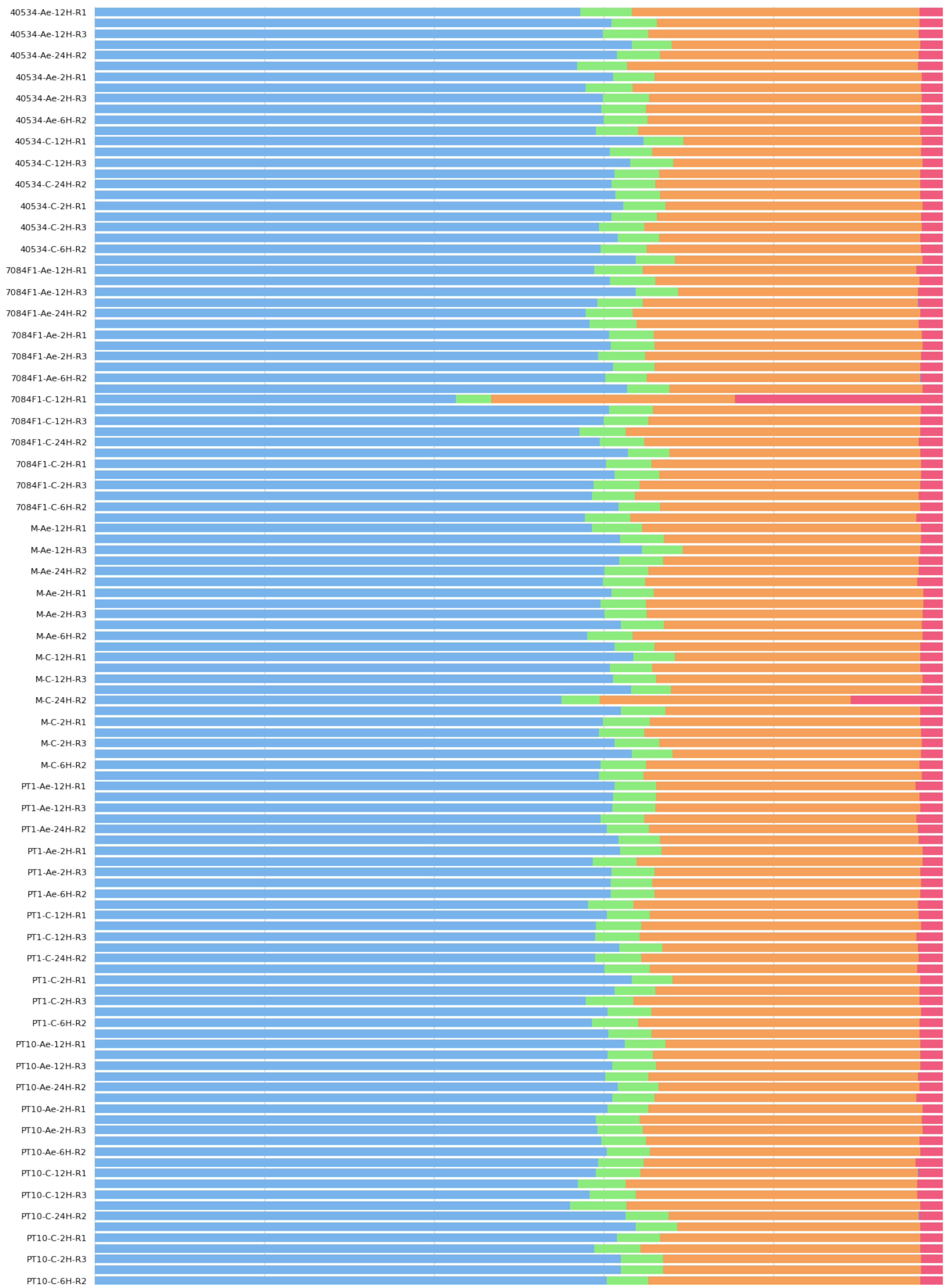
Picard is a set of Java command line tools for manipulating high-throughput sequencing data.

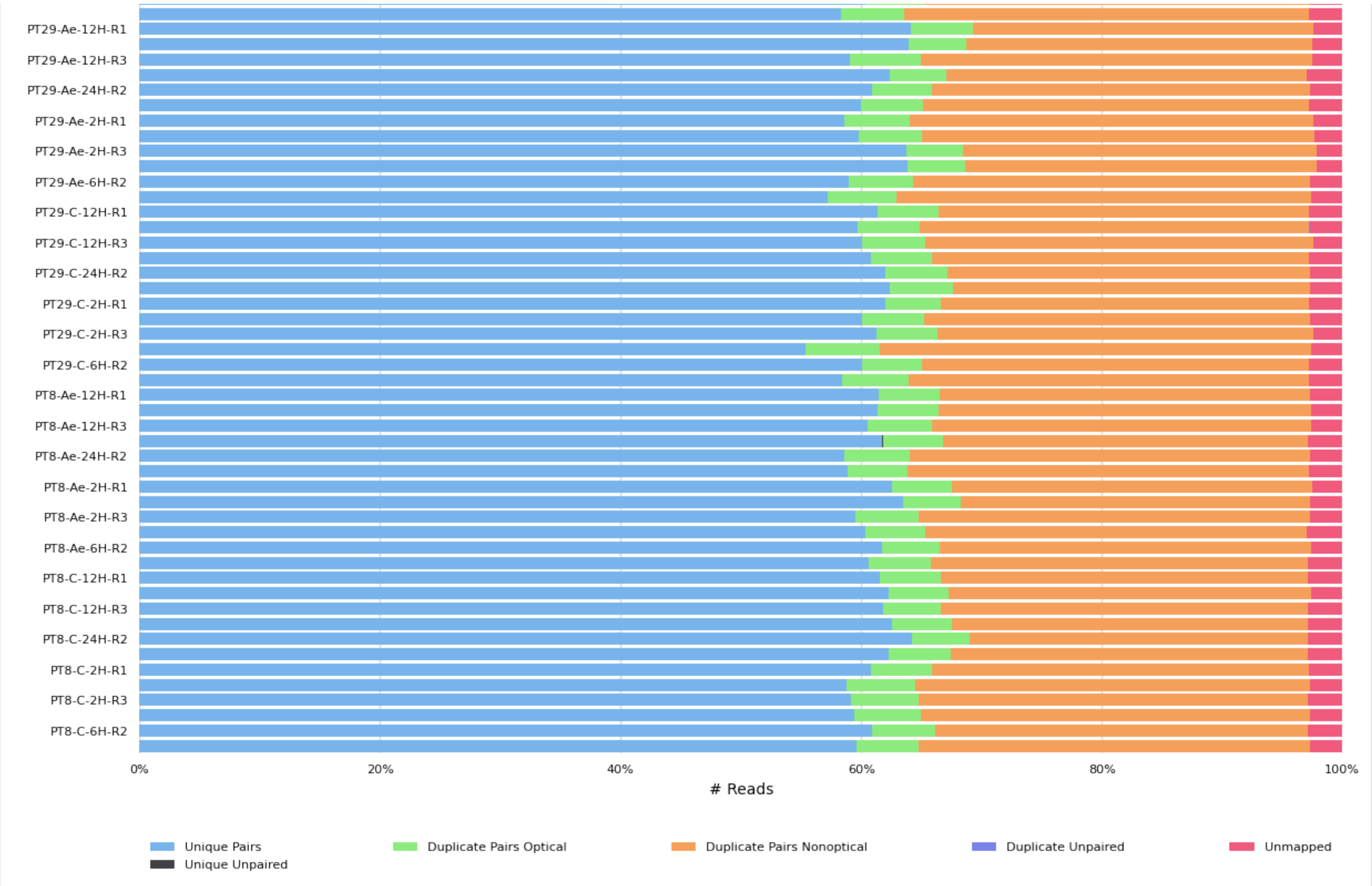
Mark Duplicates

Number of reads, categorised by duplication state. **Pair counts are doubled** - see help text for details.

Number of Reads | Percentages

Picard: Deduplication Stats




PT29-Ae-12H-R1
PT29-Ae-12H-R3
PT29-Ae-24H-R2
PT29-Ae-2H-R1
PT29-Ae-2H-R3
PT29-Ae-6H-R2
PT29-C-12H-R1
PT29-C-12H-R3
PT29-C-24H-R2
PT29-C-2H-R1
PT29-C-2H-R3
PT29-C-6H-R2
PT8-Ae-12H-R1
PT8-Ae-12H-R3
PT8-Ae-24H-R2
PT8-Ae-2H-R1
PT8-Ae-2H-R3
PT8-Ae-6H-R2
PT8-C-12H-R1
PT8-C-12H-R3
PT8-C-24H-R2
PT8-C-2H-R1
PT8-C-2H-R3
PT8-C-6H-R2
0%
20%
40%
60%
80%
100%
Reads
Unique Pairs
Duplicate Pairs Optical
Duplicate Pairs Nonoptical
Duplicate Unpaired
Unmapped
Unique Unpaired

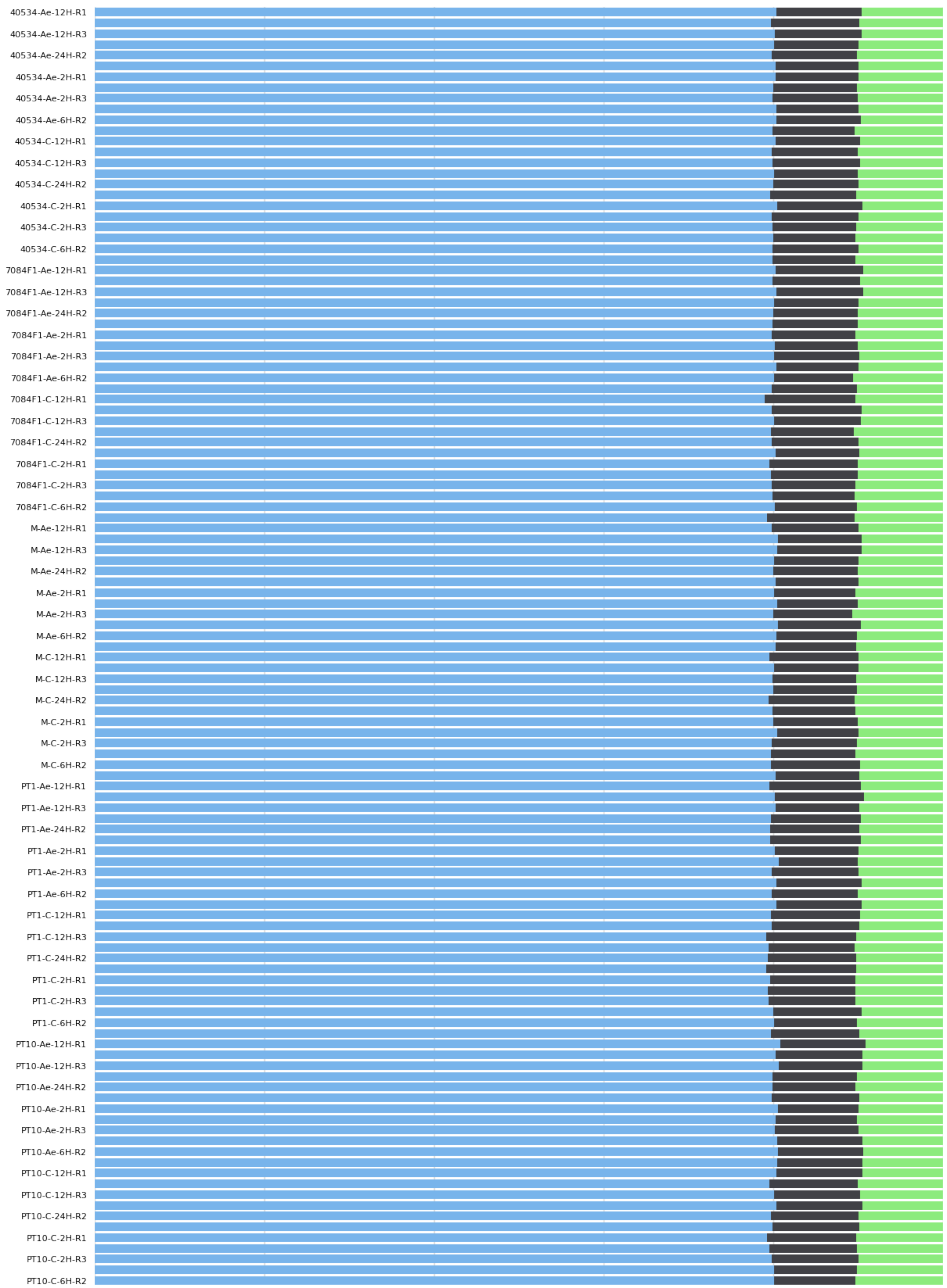
QualiMap

QualiMap is a platform-independent application to facilitate the quality control of alignment sequencing data and its derivatives like feature counts. *DOI: 10.1093/bioinformatics/btv566; 10.1093/bioinformatics/bts503.*

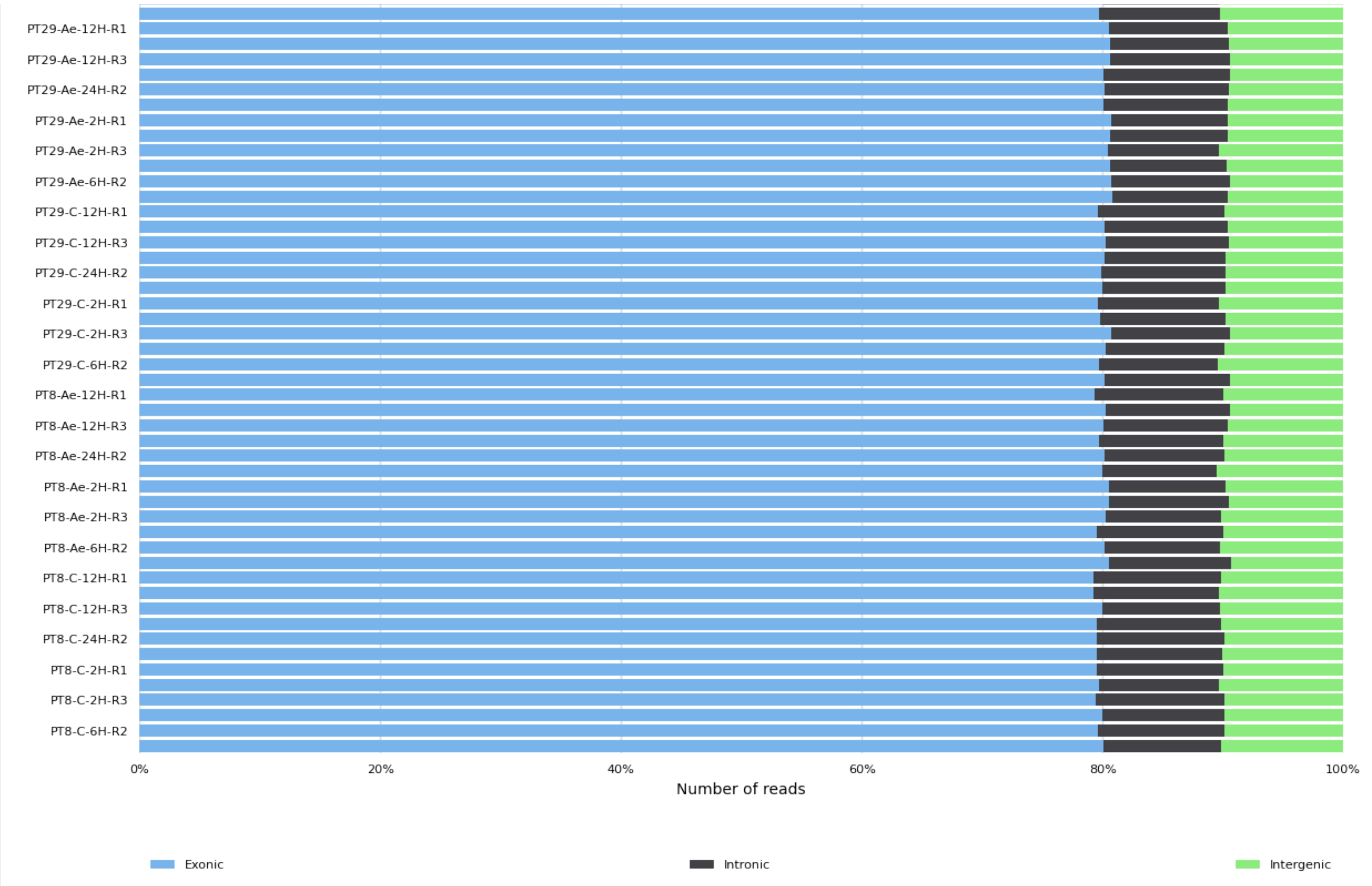
Genomic origin of reads

Classification of mapped reads as originating in exonic, intronic or intergenic regions. These can be displayed as either the number or percentage of mapped reads.

Counts | Percentages



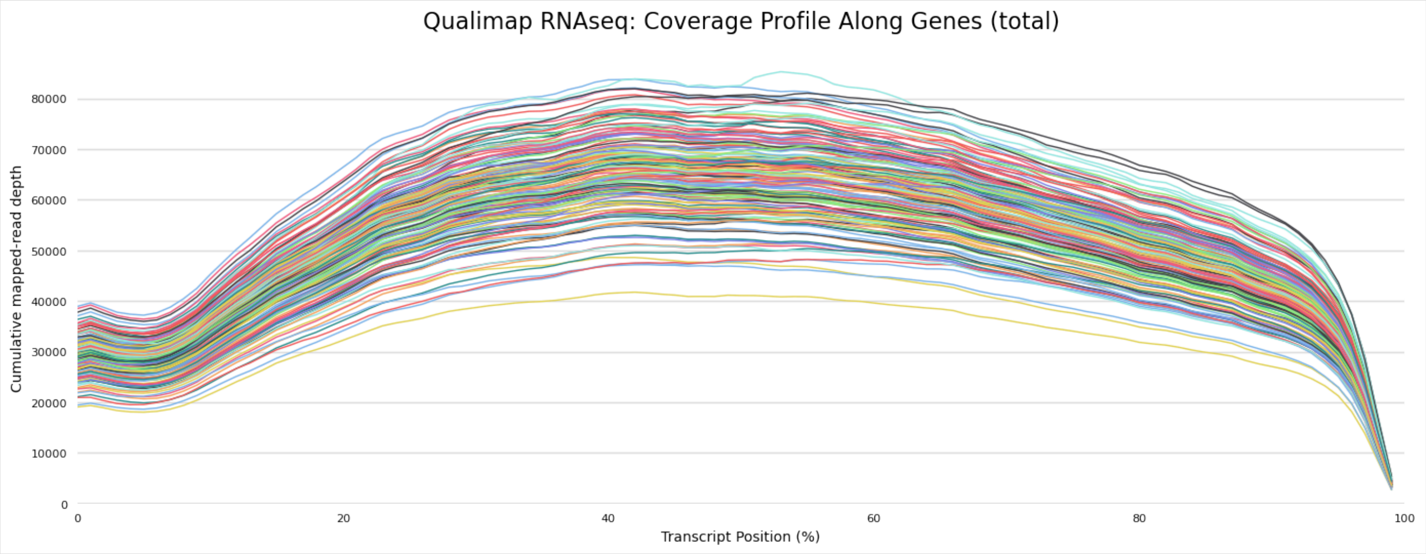
Qualimap RNAseq: Genomic Origin



Gene Coverage Profile

Mean distribution of coverage depth across the length of all mapped transcripts.

Counts Normalised



Rsem

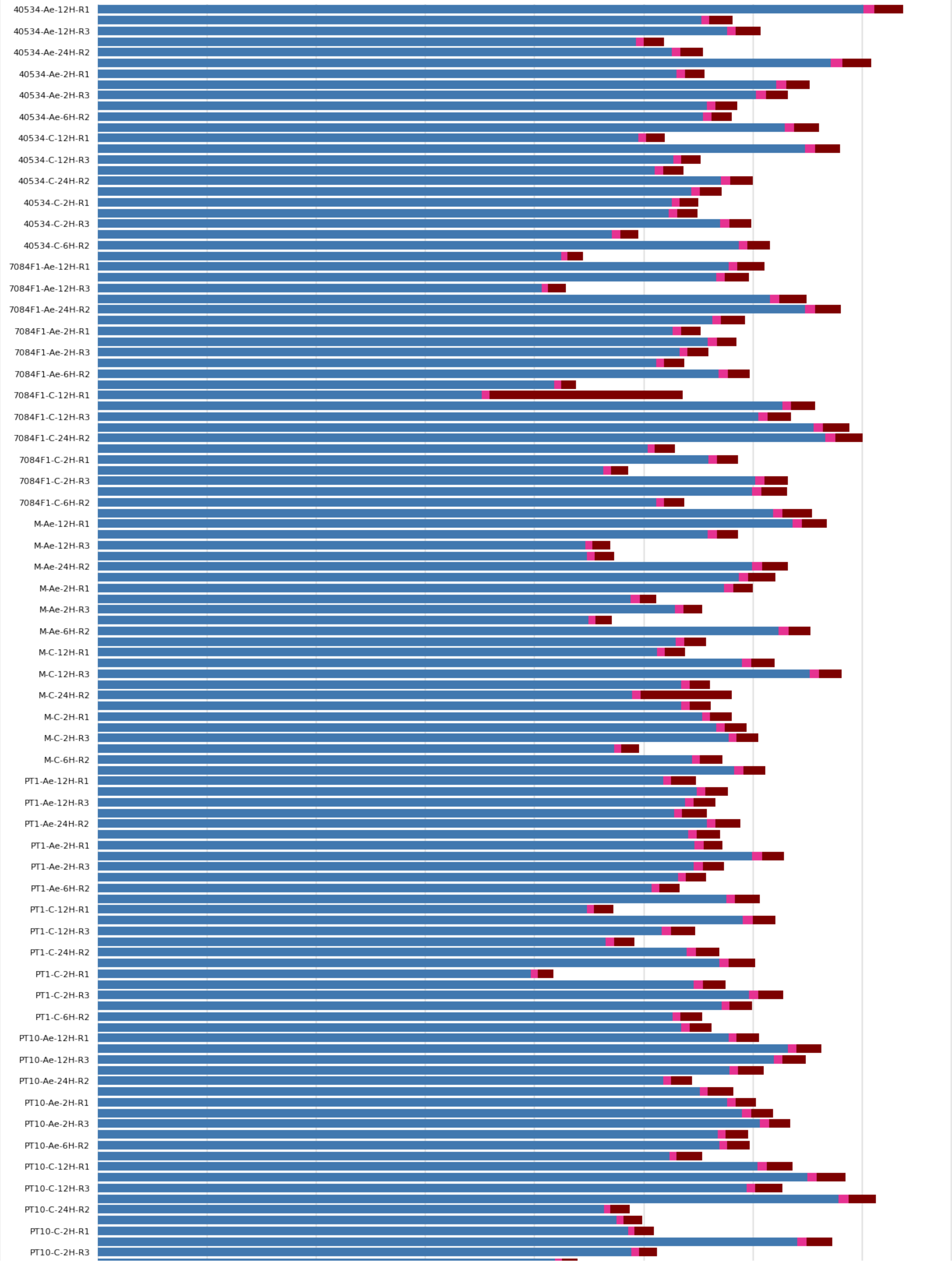
Rsem RSEM (RNA-Seq by Expectation-Maximization) is a software package forestimating gene and isoform expression levels from RNA-Seq data. *DOI: 10.1186/1471-2105-12-323.*

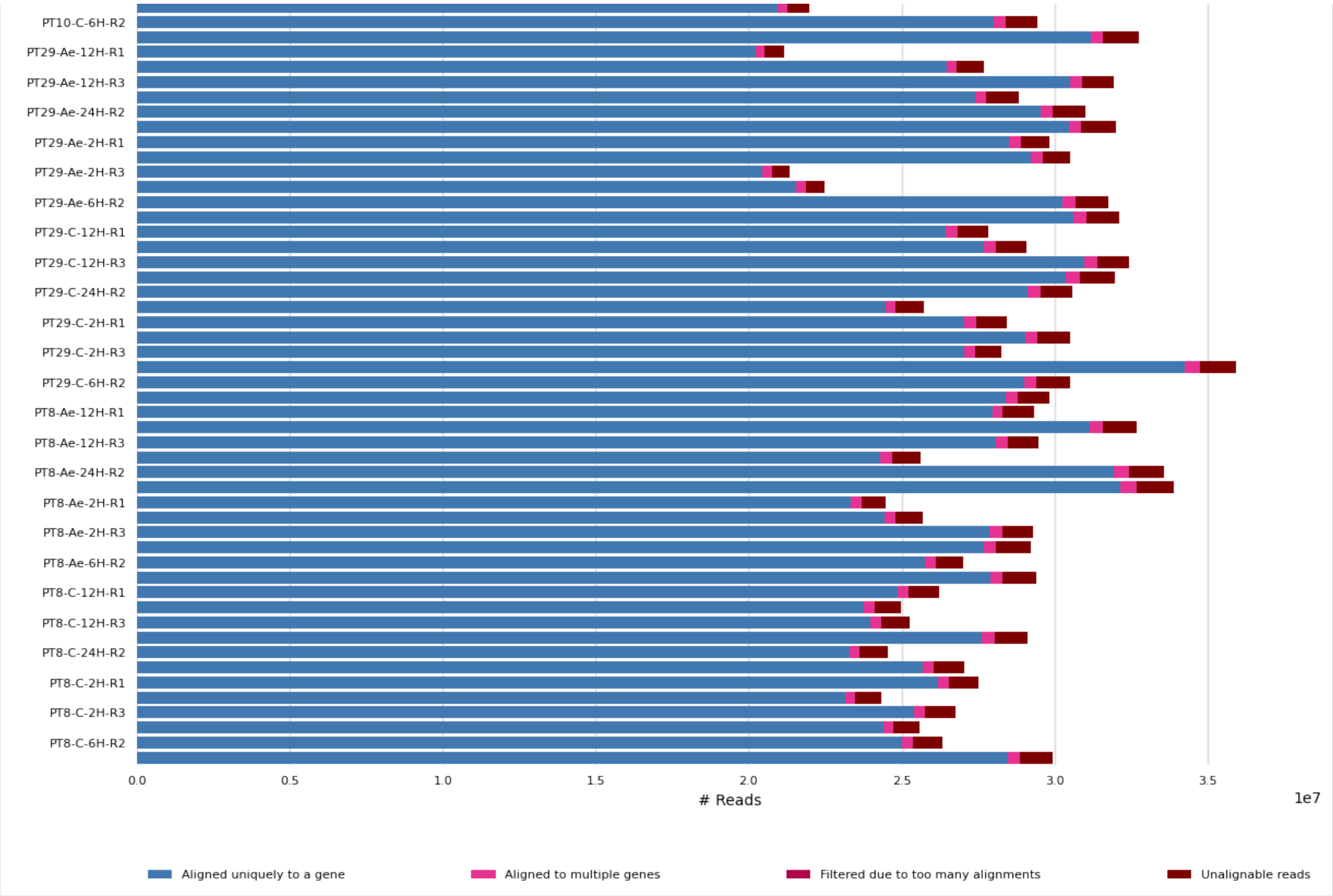
Mapped Reads

A breakdown of how all reads were aligned for each sample.

Number of Reads | Percentages

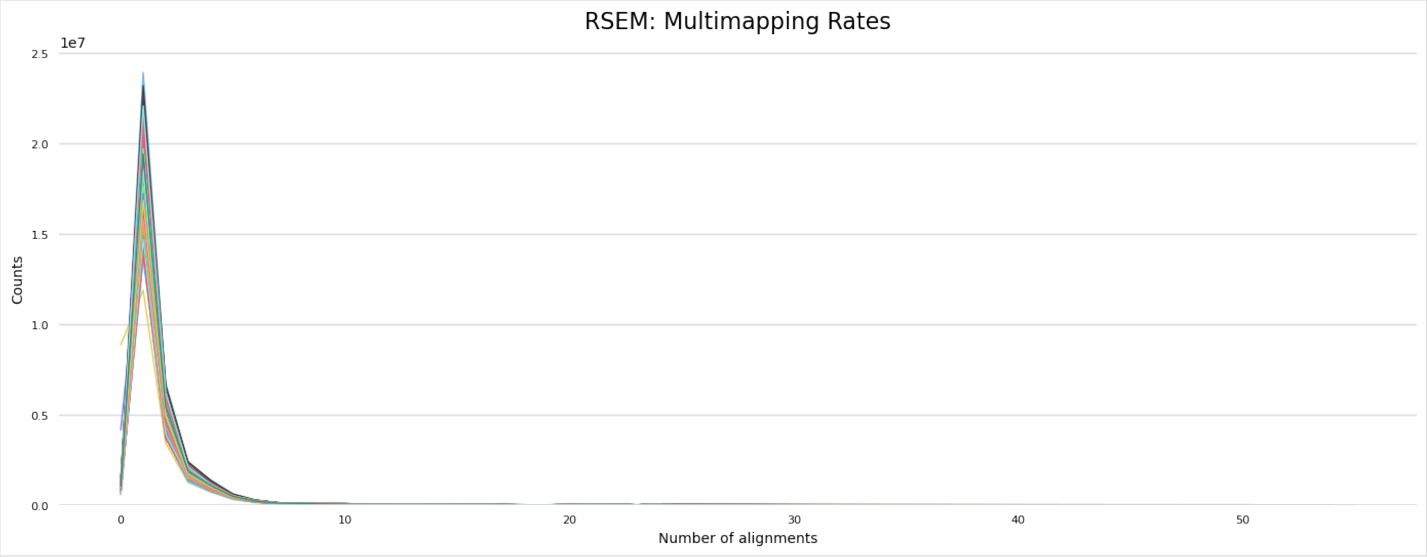
RSEM: Mapped reads





Multimapping rates

A frequency histogram showing how many reads were aligned to `n` reference regions.

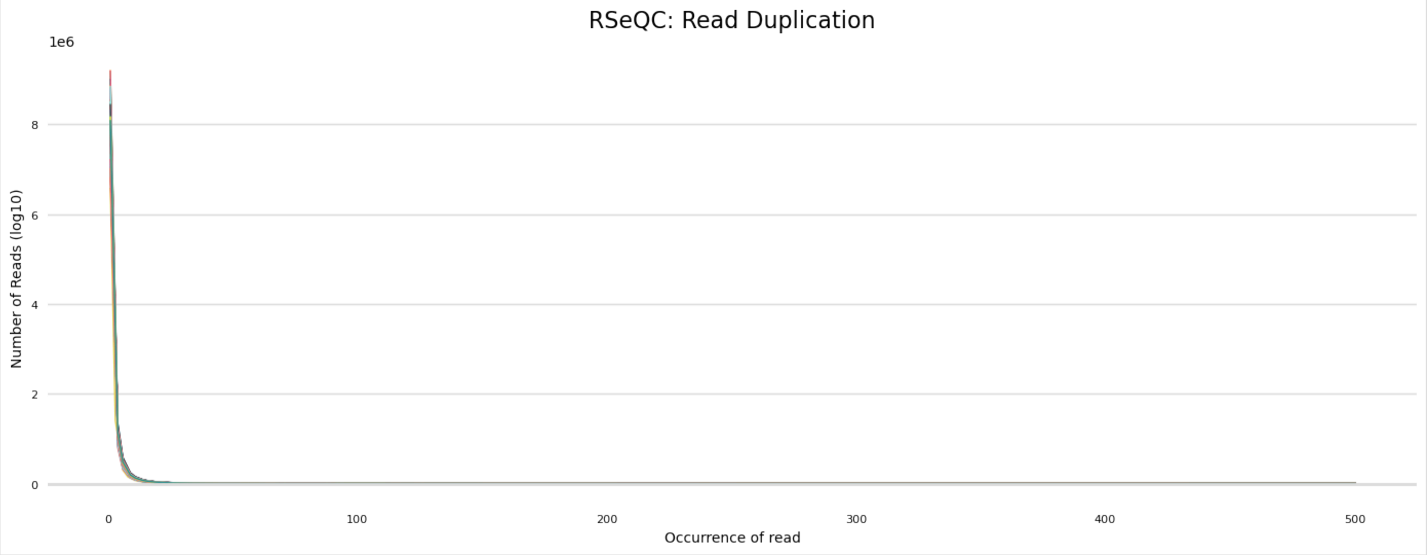


RSeQC

RSeQC package provides a number of useful modules that can comprehensively evaluate high throughput RNA-seq data. *DOI: 10.1093/bioinformatics/bts356.*

Read Duplication

read_duplication.py calculates how many alignment positions have a certain number of exact duplicates. Note - plot truncated at 500 occurrences and binned.

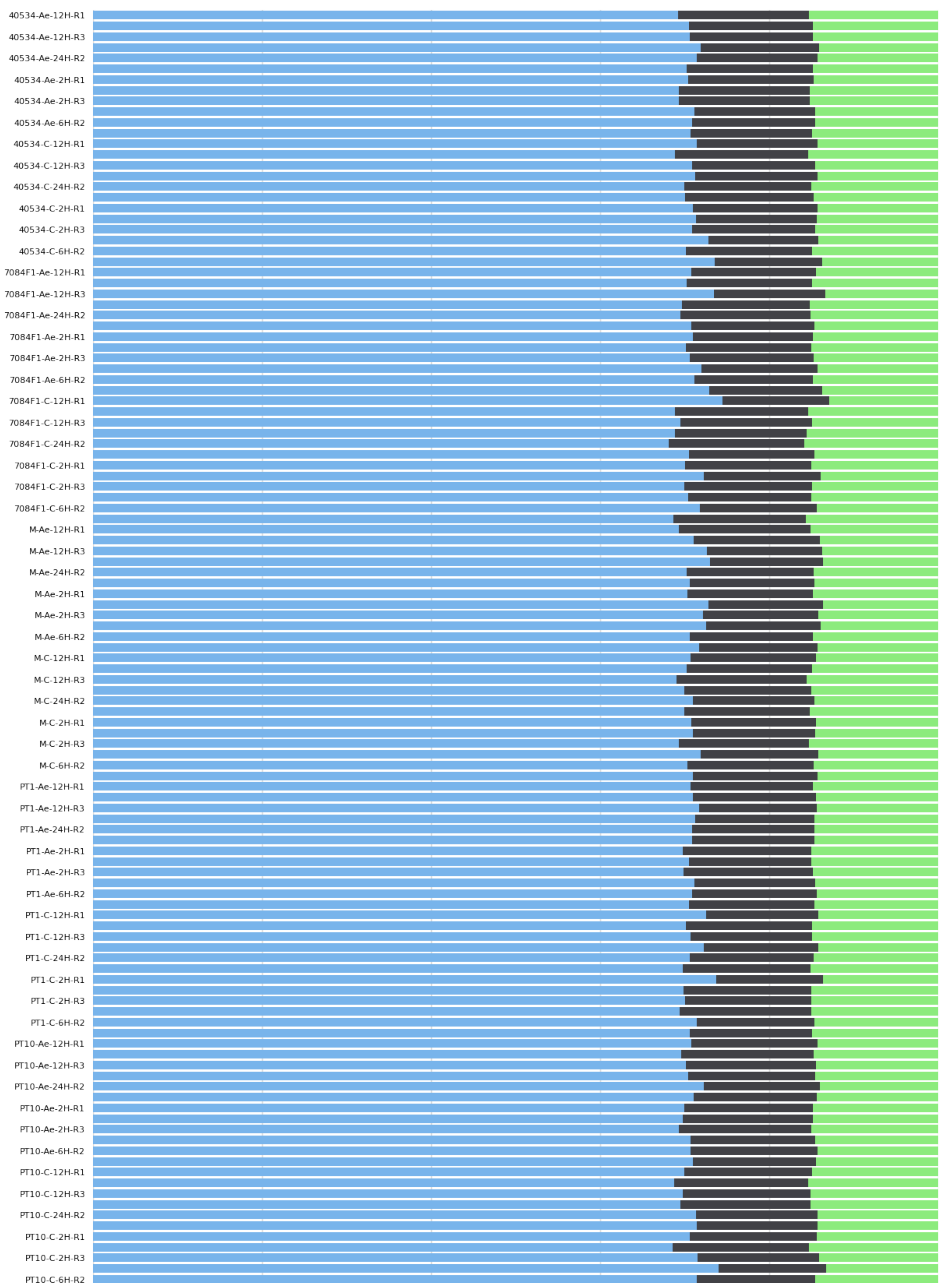


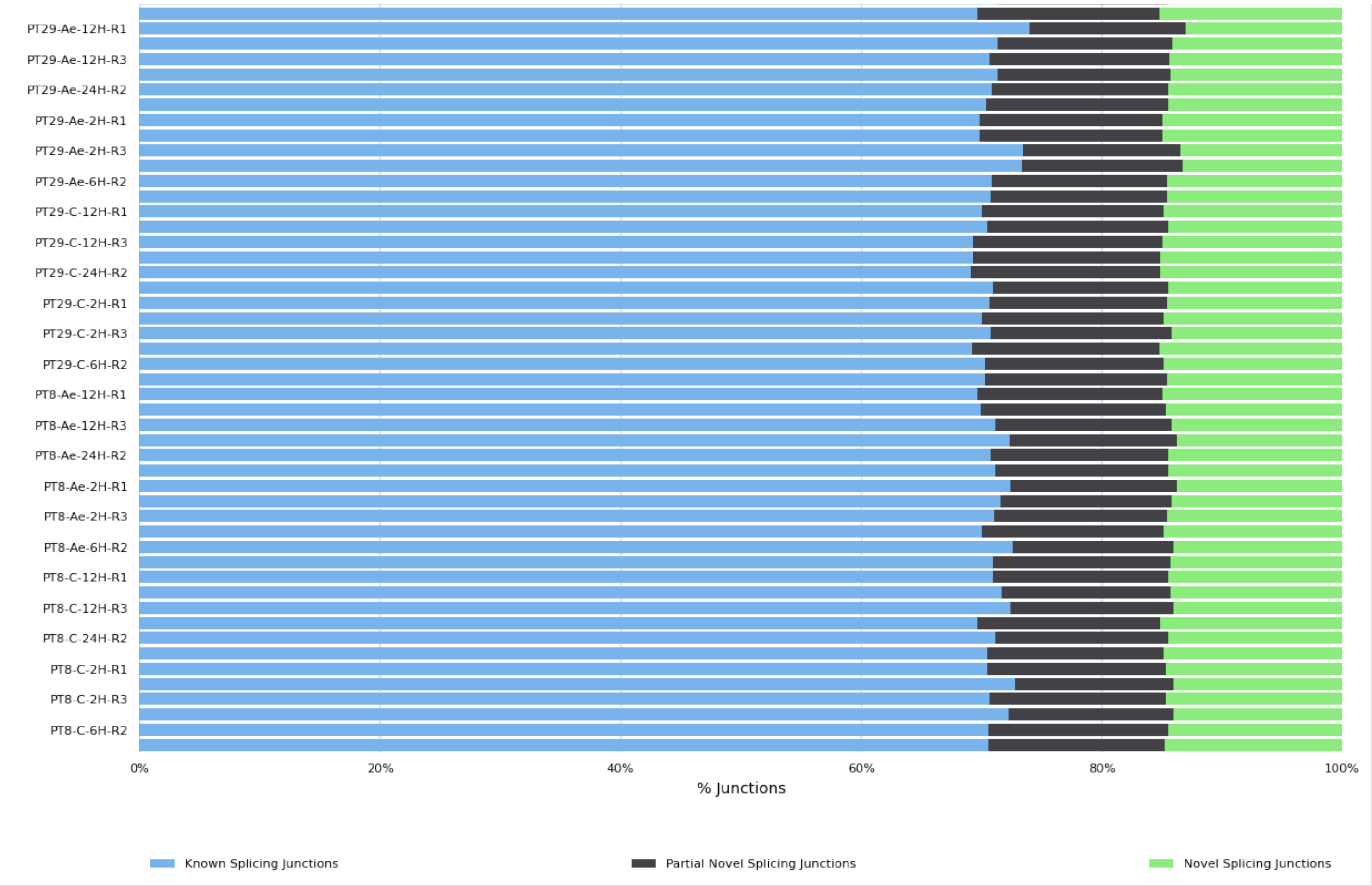
Junction Annotation

Junction annotation compares detected splice junctions to a reference gene model. An RNA read can be spliced 2 or more times, each time is called a splicing event.

Counts | Percentages   Junctions | Events

RSeQC: Splicing Junctions



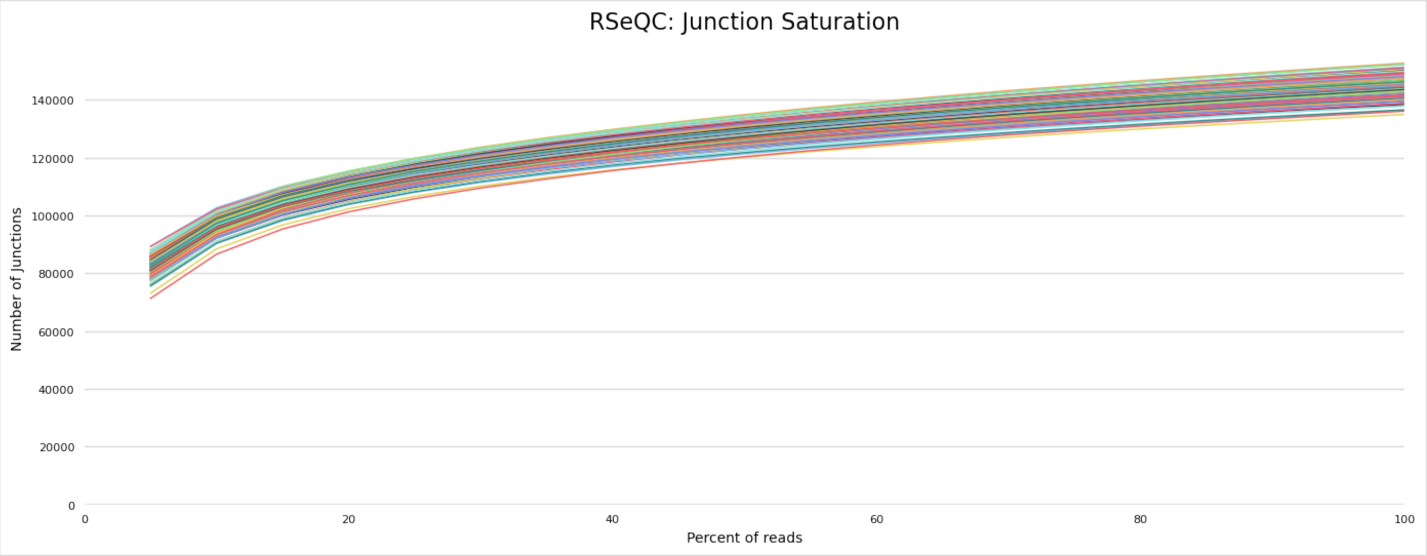


Junction Saturation

Junction Saturation counts the number of known splicing junctions that are observed in each dataset. If sequencing depth is sufficient, all (annotated) splice junctions should be rediscovered, resulting in a curve that reaches a plateau. Missing low abundance splice junctions can affect downstream analysis.

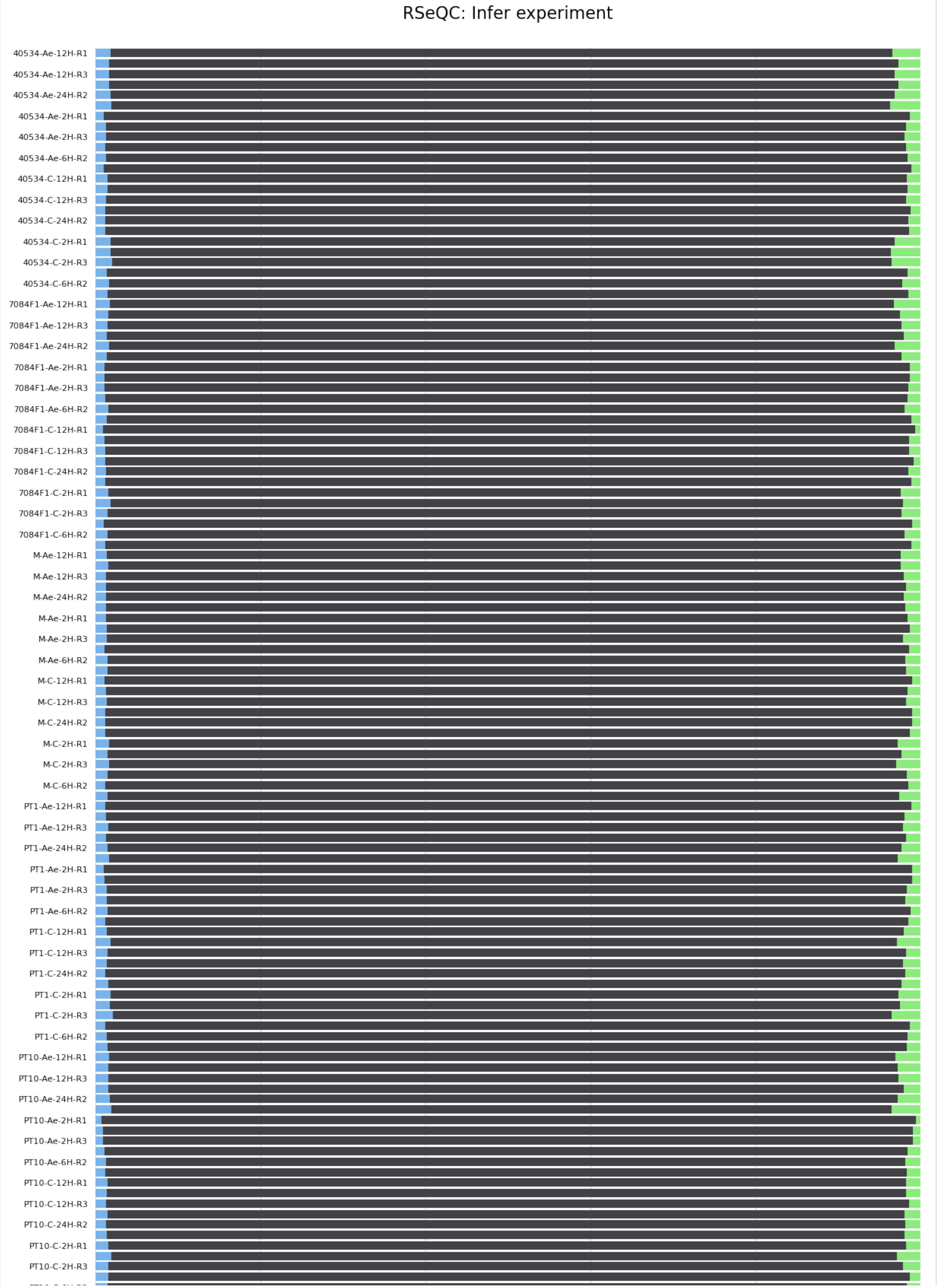
Click a line to see the data side by side (as in the original RSeQC plot).

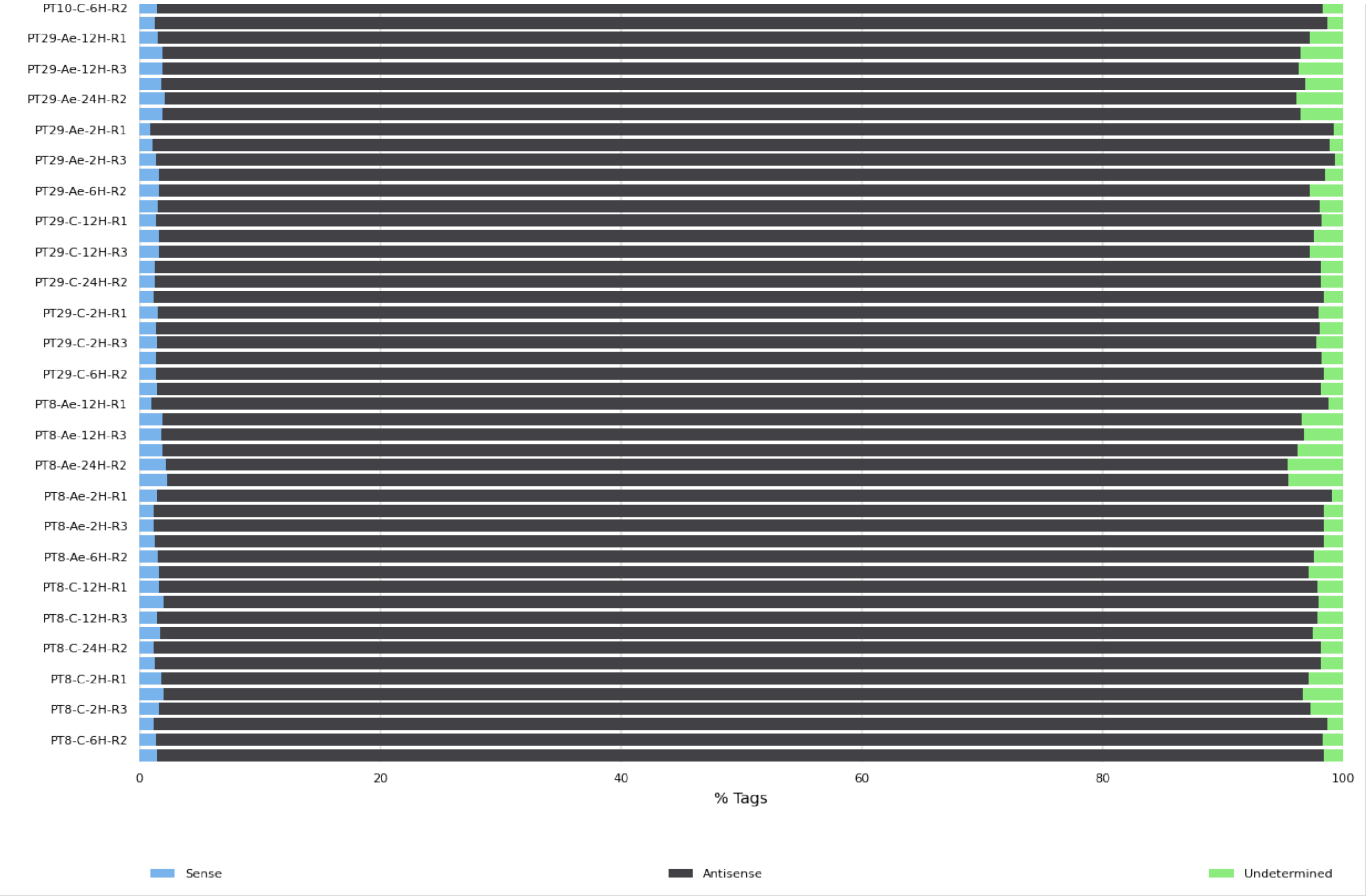
All Junctions | Known Junctions | Novel Junctions



Infer experiment

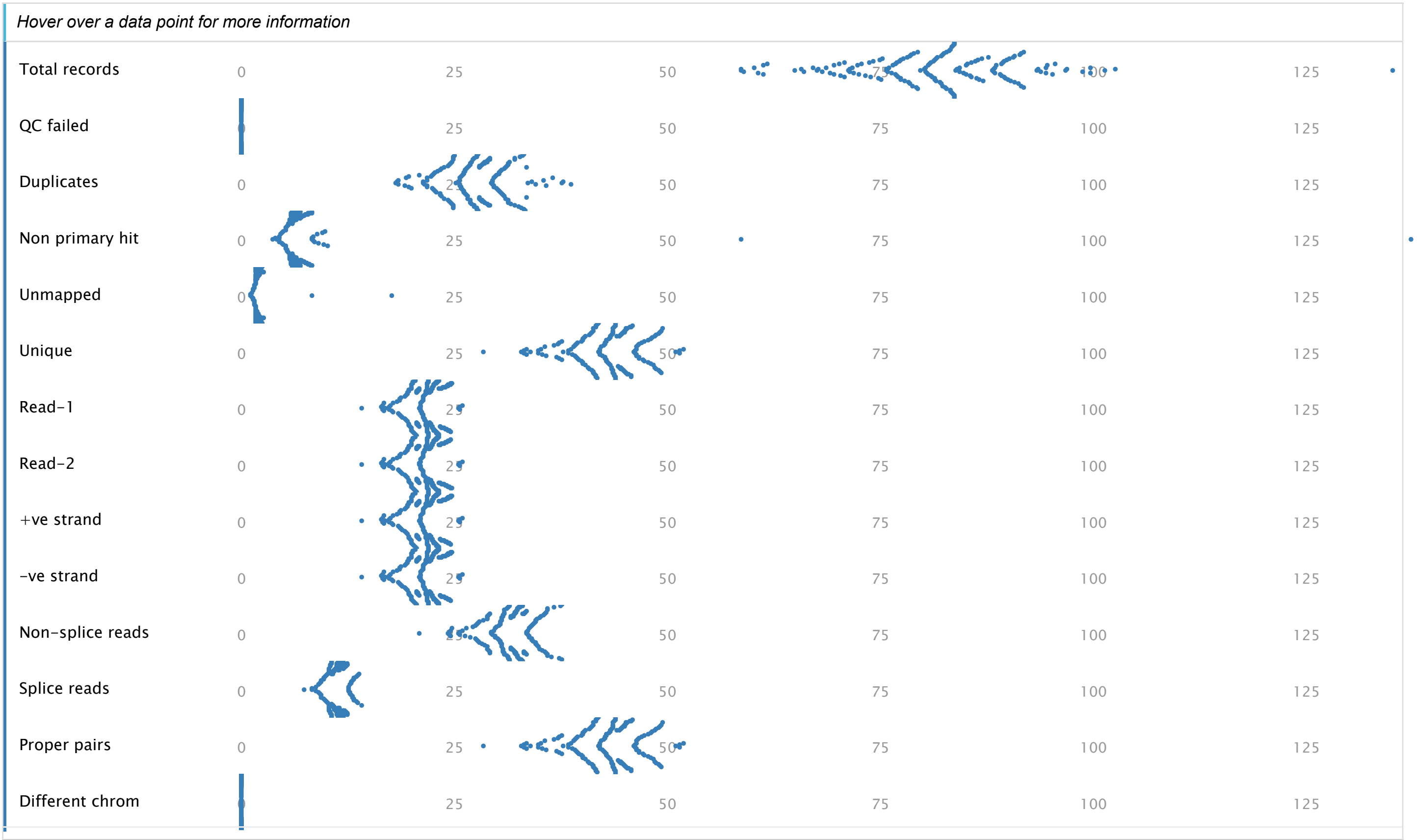
Infer experiment counts the percentage of reads and read pairs that match the strandedness of overlapping transcripts. It can be used to infer whether RNA-seq library preps are stranded (sense or antisense).


RSeQC: Infer experiment
40534-Ae-12H-R1
40534-Ae-12H-R3
40534-Ae-24H-R2
40534-Ae-2H-R1
40534-Ae-2H-R3
40534-Ae-6H-R2
40534-C-12H-R1
40534-C-12H-R3
40534-C-24H-R2
40534-C-2H-R1
40534-C-2H-R3
40534-C-6H-R2
7084F1-Ae-12H-R1
7084F1-Ae-12H-R3
7084F1-Ae-24H-R2
7084F1-Ae-2H-R1
7084F1-Ae-2H-R3
7084F1-Ae-6H-R2
7084F1-C-12H-R1
7084F1-C-12H-R3
7084F1-C-24H-R2
7084F1-C-2H-R1
7084F1-C-2H-R3
7084F1-C-6H-R2
M-Ae-12H-R1
M-Ae-12H-R3
M-Ae-24H-R2
M-Ae-2H-R1
M-Ae-2H-R3
M-Ae-6H-R2
M-C-12H-R1
M-C-12H-R3
M-C-24H-R2
M-C-2H-R1
M-C-2H-R3
M-C-6H-R2
PT1-Ae-12H-R1
PT1-Ae-12H-R3
PT1-Ae-24H-R2
PT1-Ae-2H-R1
PT1-Ae-2H-R3
PT1-Ae-6H-R2
PT1-C-12H-R1
PT1-C-12H-R3
PT1-C-24H-R2
PT1-C-2H-R1
PT1-C-2H-R3
PT1-C-6H-R2
PT10-Ae-12H-R1
PT10-Ae-12H-R3
PT10-Ae-24H-R2
PT10-Ae-2H-R1
PT10-Ae-2H-R3
PT10-Ae-6H-R2
PT10-C-12H-R1
PT10-C-12H-R3
PT10-C-24H-R2
PT10-C-2H-R1
PT10-C-2H-R3



Bam Stat

All numbers reported in millions.


Hover over a data point for more information

Samtools

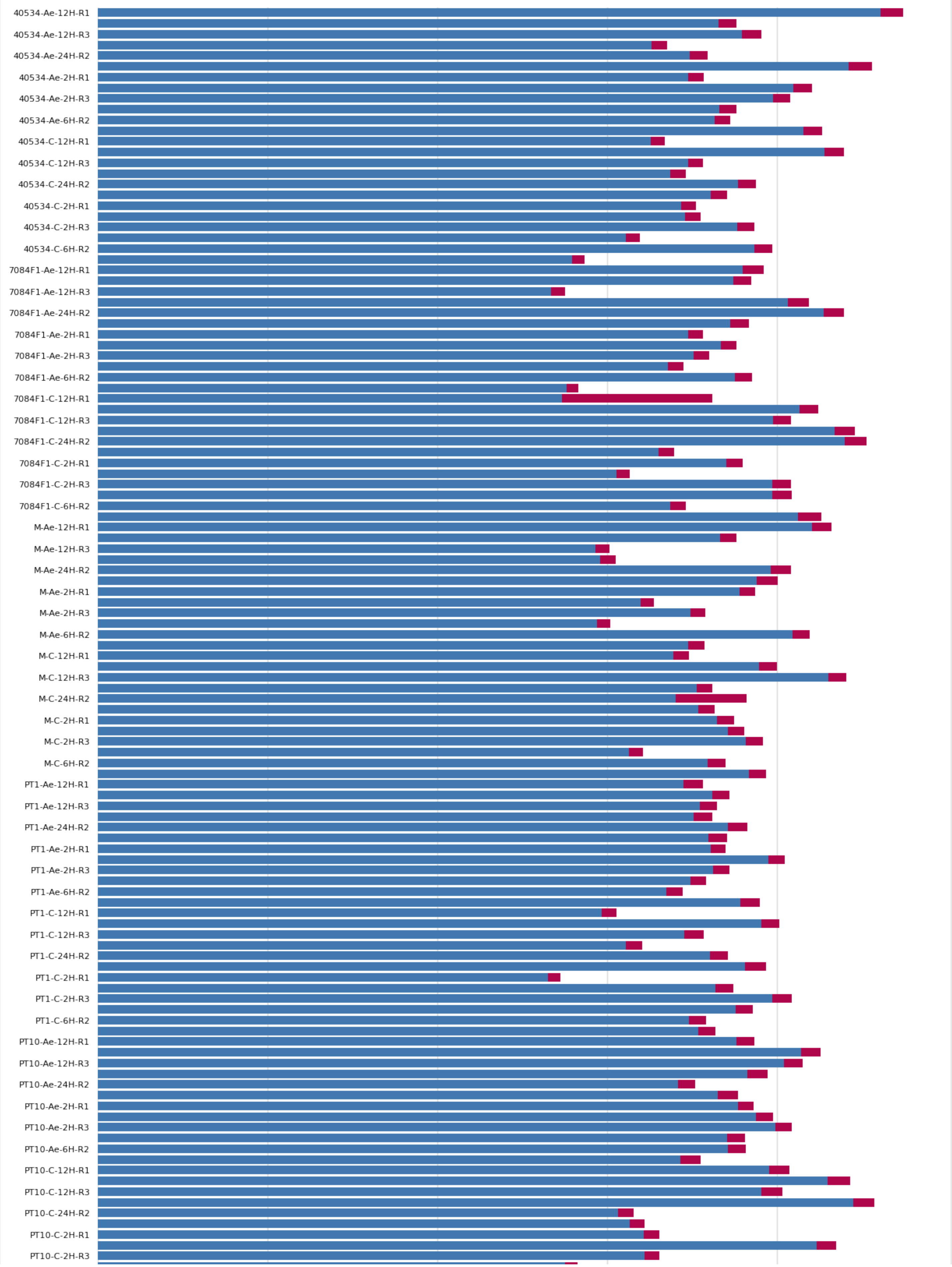
Samtools is a suite of programs for interacting with high-throughput sequencing data. *DOI: 10.1093/bioinformatics/btp352.*

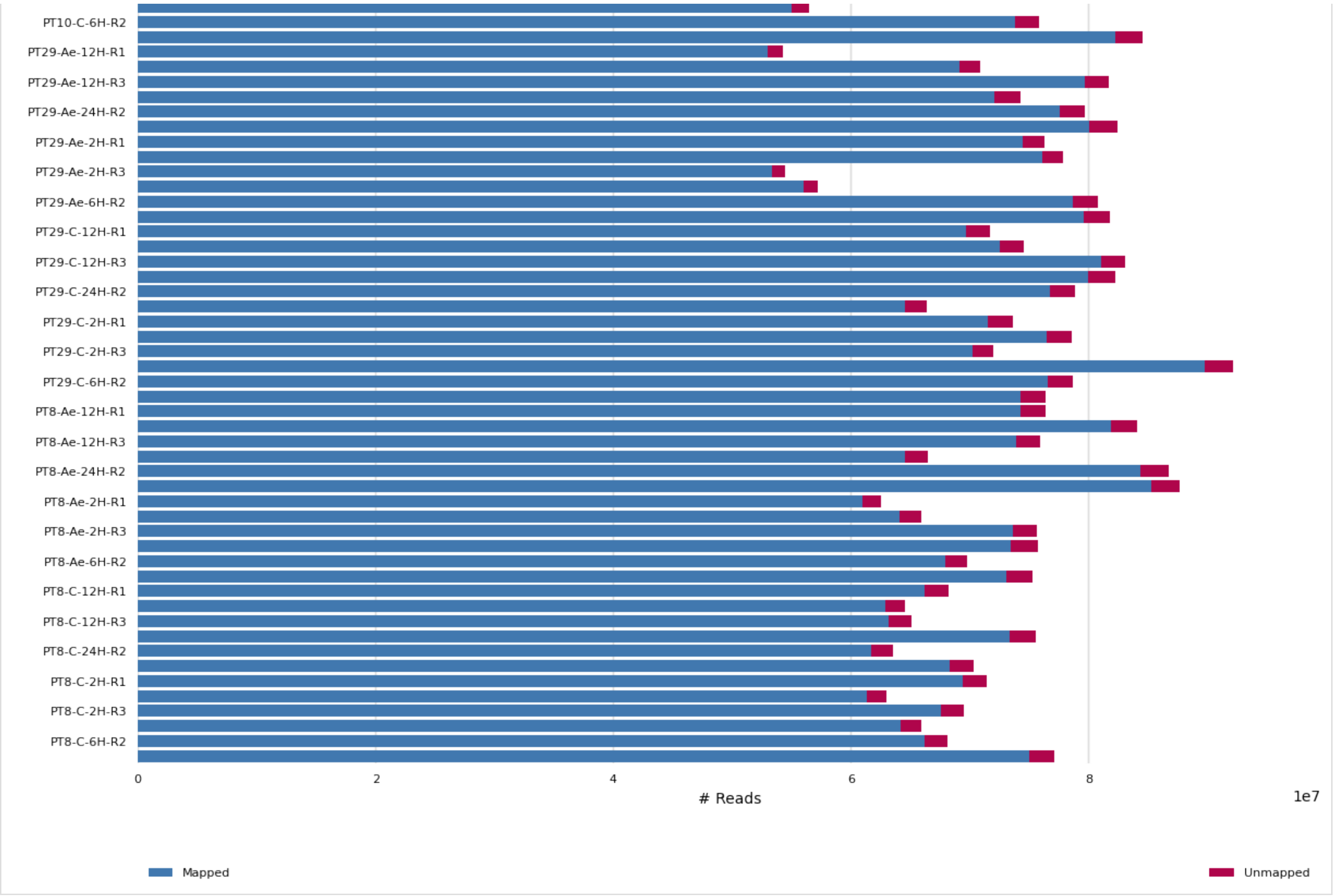
Percent Mapped

Alignment metrics from `samtools stats`; mapped vs. unmapped reads.

Number of Reads | Percentages

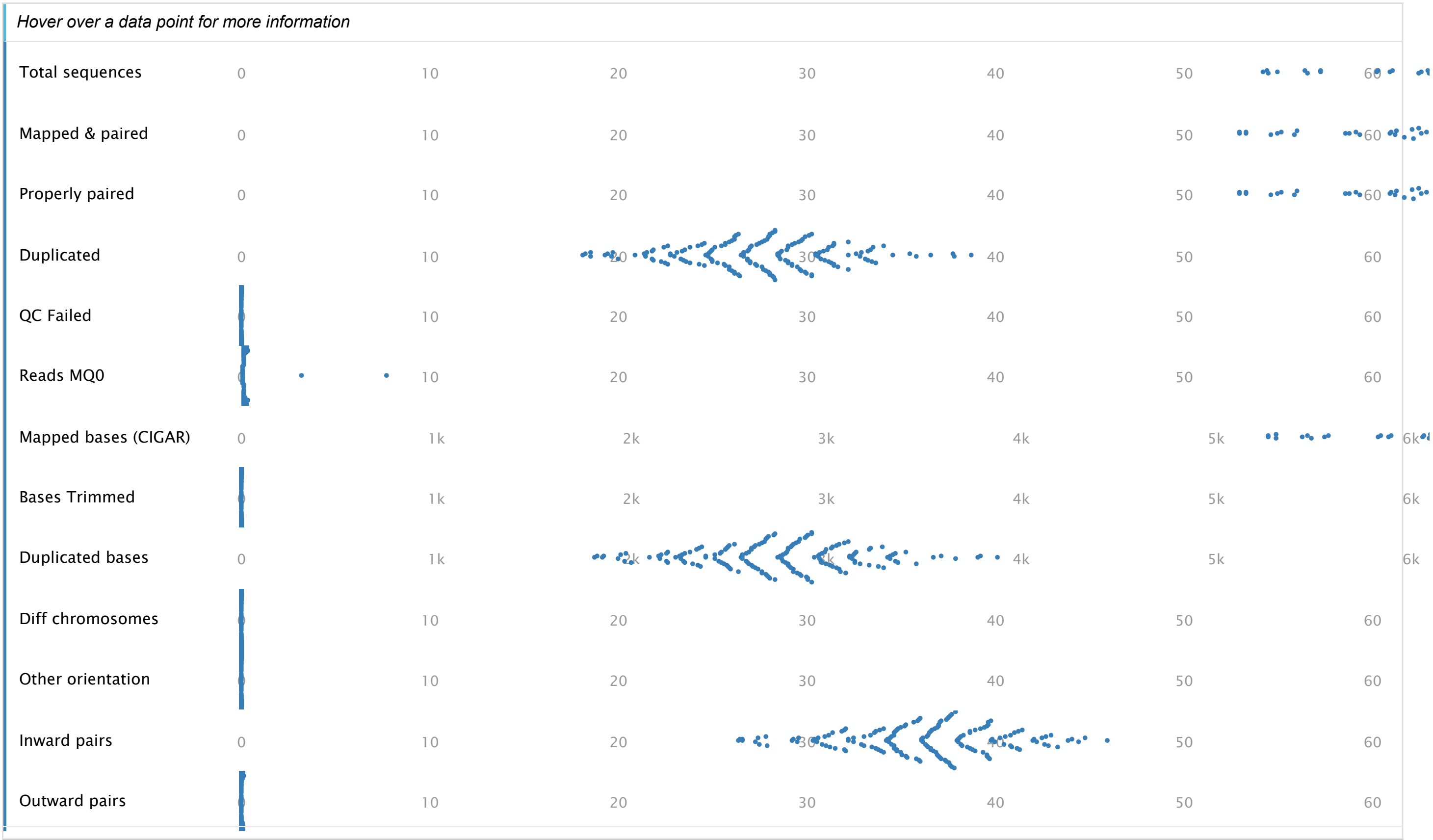
Samtools stats: Alignment Scores




PT10-C-6H-R2
PT29-Ae-12H-R1
PT29-Ae-12H-R3
PT29-Ae-24H-R2
PT29-Ae-2H-R1
PT29-Ae-2H-R3
PT29-Ae-6H-R2
PT29-C-12H-R1
PT29-C-12H-R3
PT29-C-24H-R2
PT29-C-2H-R1
PT29-C-2H-R3
PT29-C-6H-R2
PT8-Ae-12H-R1
PT8-Ae-12H-R3
PT8-Ae-24H-R2
PT8-Ae-2H-R1
PT8-Ae-2H-R3
PT8-Ae-6H-R2
PT8-C-12H-R1
PT8-C-12H-R3
PT8-C-24H-R2
PT8-C-2H-R1
PT8-C-2H-R3
PT8-C-6H-R2
0
2
4
6
8
Reads
1e7
Mapped
Unmapped

Alignment metrics

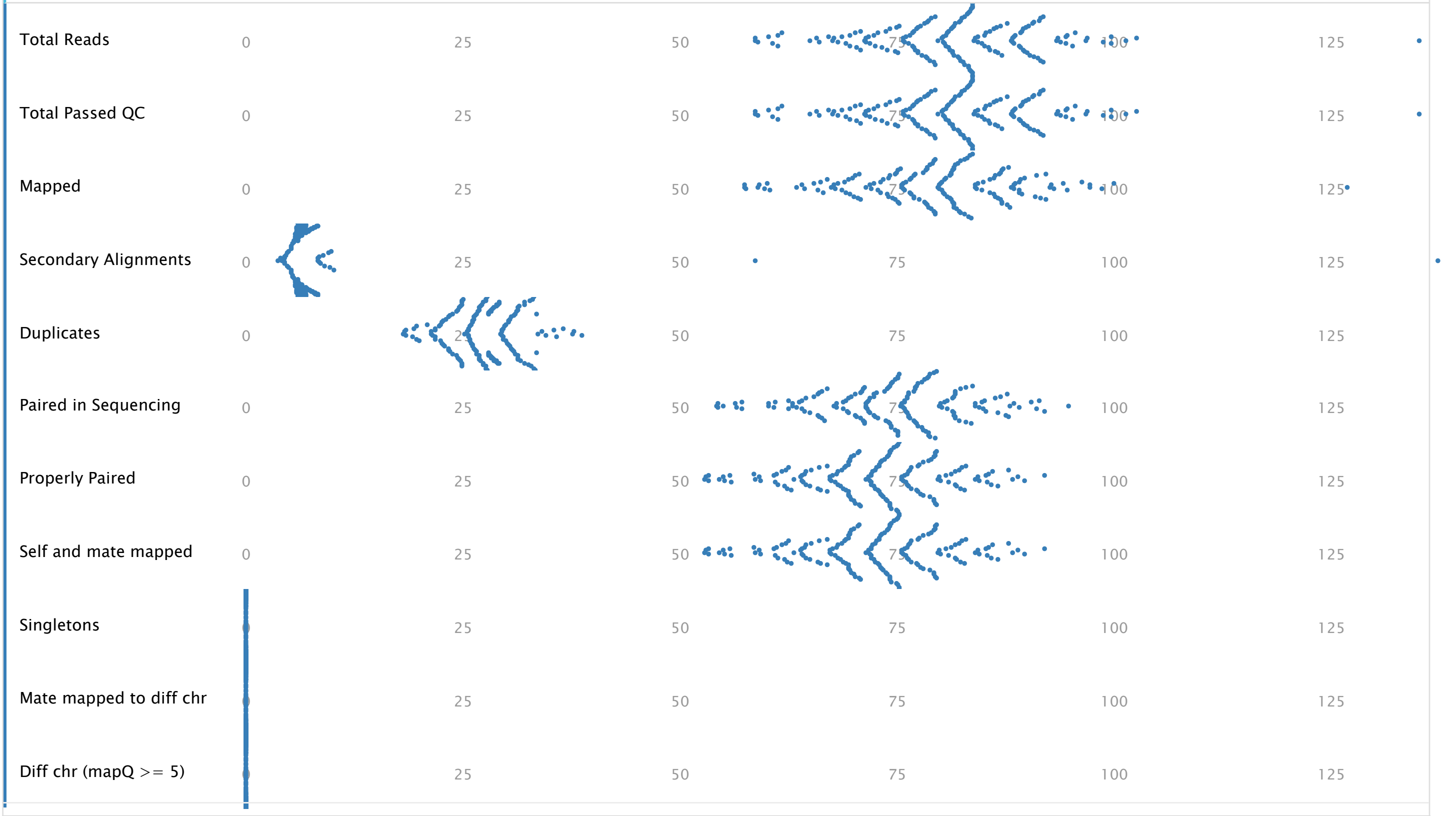
This module parses the output from `samtools stats`. All numbers in millions.



Samtools Flagstat

This module parses the output from `samtools flagstat`. All numbers in millions.

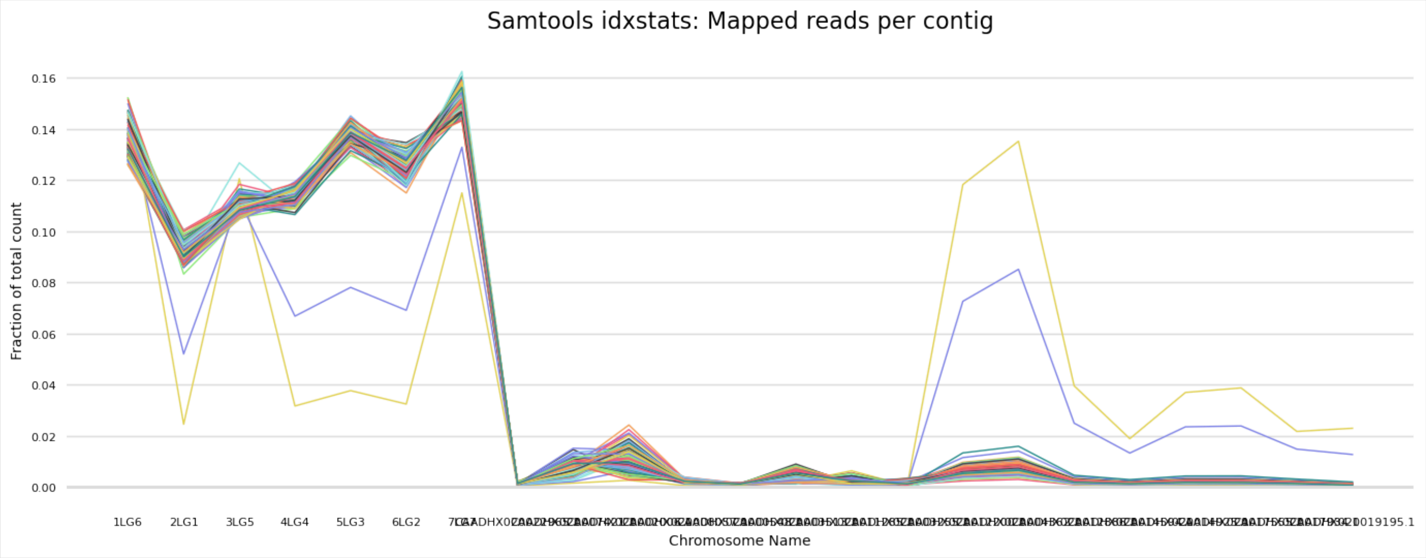
Hover over a data point for more information



Mapped reads per contig

The `samtools idxstats` tool counts the number of mapped reads per chromosome / contig. Chromosomes with < 0.1% of the total aligned reads are omitted from this plot.

Normalised Counts | Observed over Expected Counts | Raw Counts



FastQC (raw)

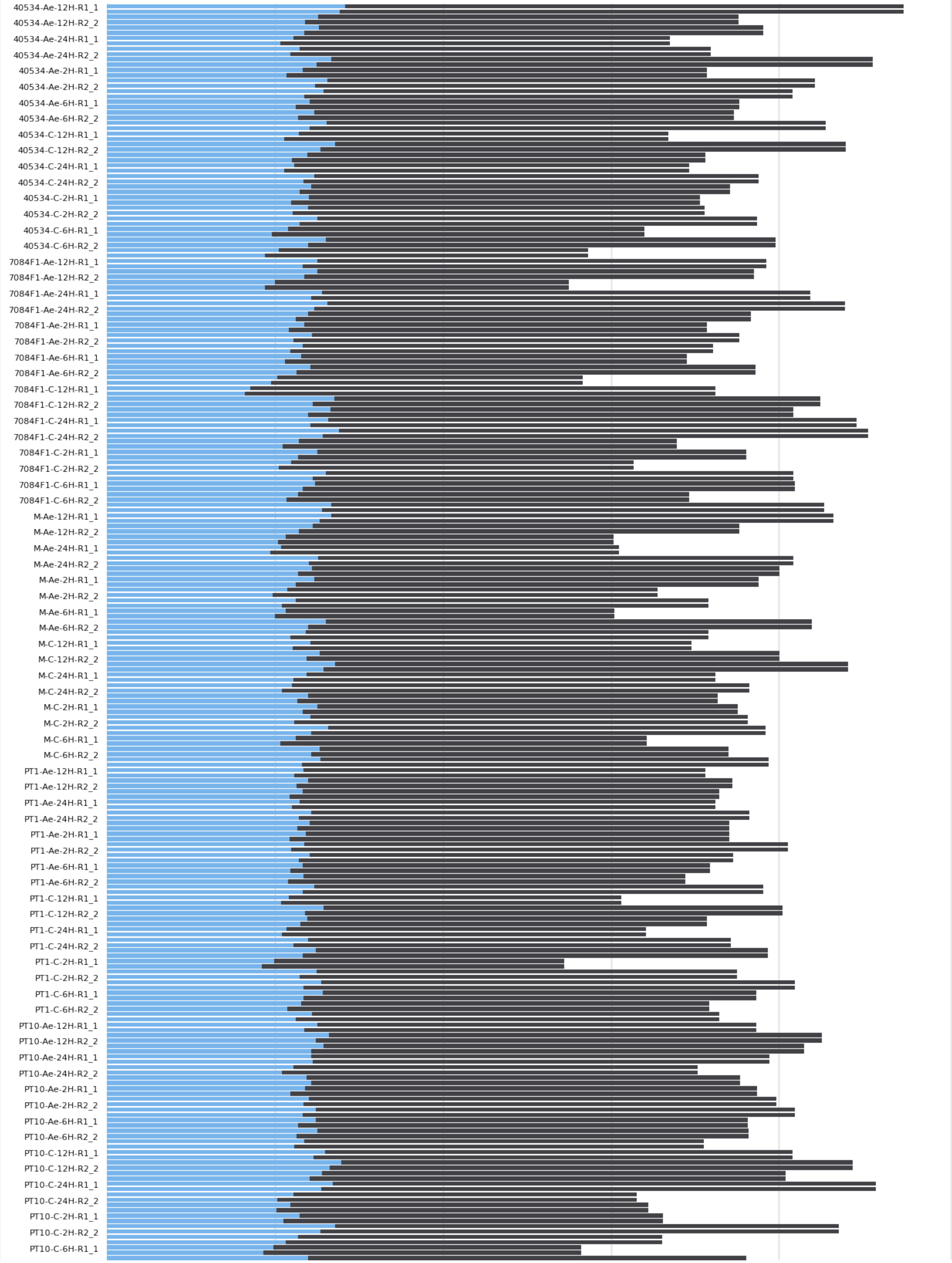
FastQC (raw) This section of the report shows FastQC results before adapter trimming.

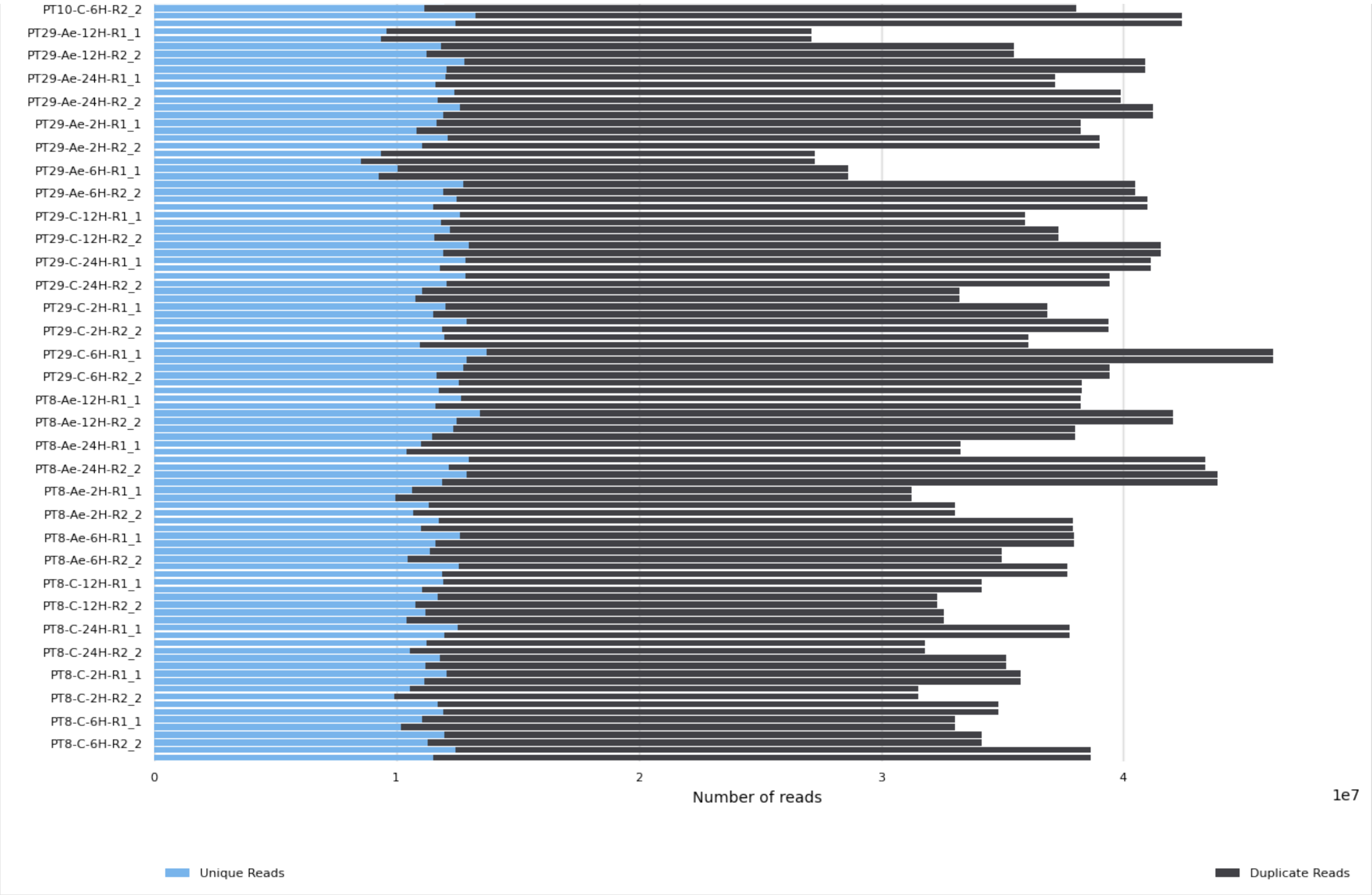
Sequence Counts

Sequence counts for each sample. Duplicate read counts are an estimate only.

Number of reads | Percentages

FastQC: Sequence Counts





Sequence Quality Histograms

336

The mean quality value across each base position in the read.

Per Sequence Quality Scores 336

The number of reads with average quality scores. Shows if a subset of reads has poor quality.

Per Base Sequence Content 83

The proportion of each base position for which each of the four normal DNA bases has been called.

Click a sample row to see a line plot for that dataset.

Rollover for sample name

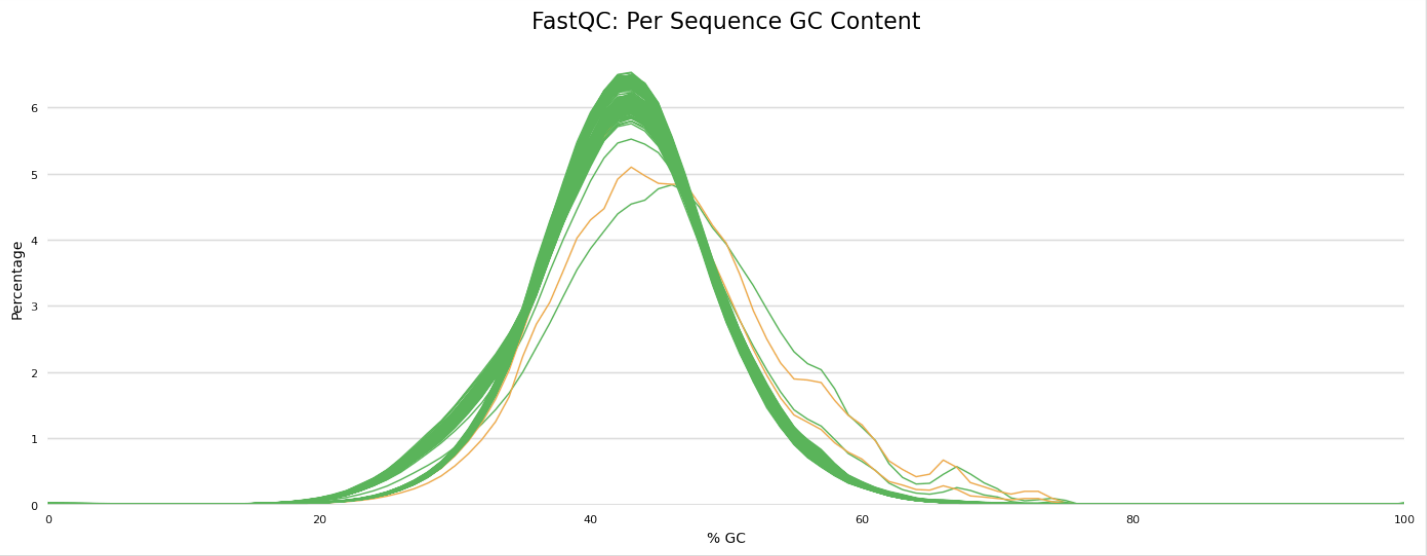
Position: - %T: - %C: - %A: - %G: -

Export Plot

Per Sequence GC Content 334

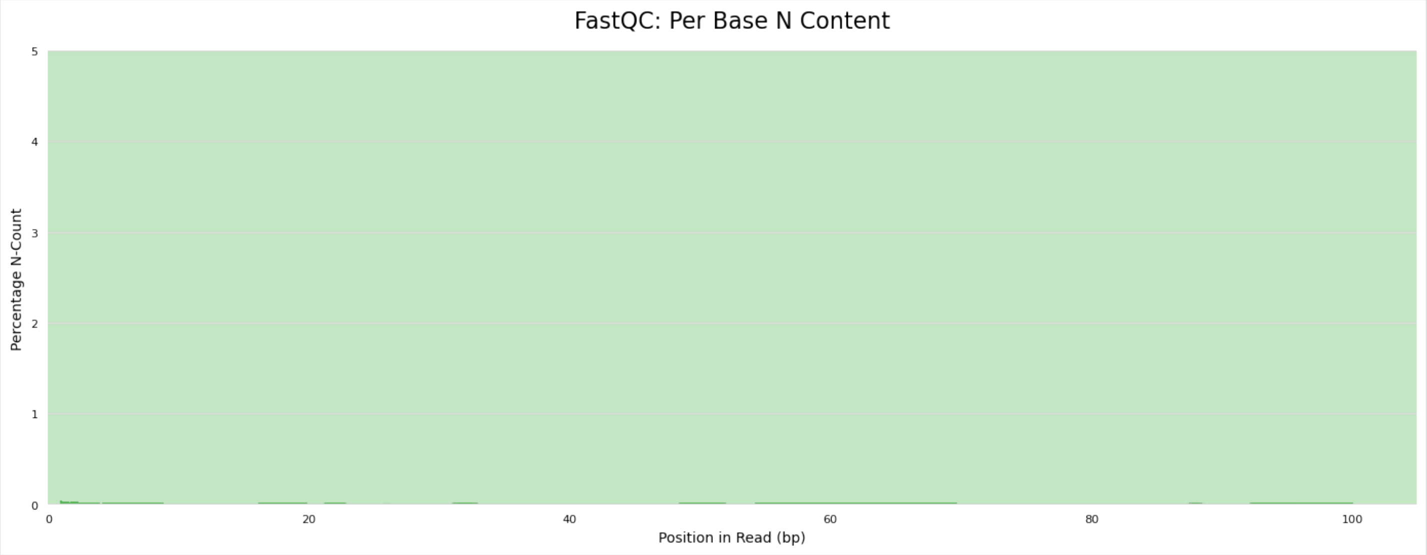
The average GC content of reads. Normal random library typically have a roughly normal distribution of GC content.

Percentages Counts



Per Base N Content 336

The percentage of base calls at each position for which an `N` was called.

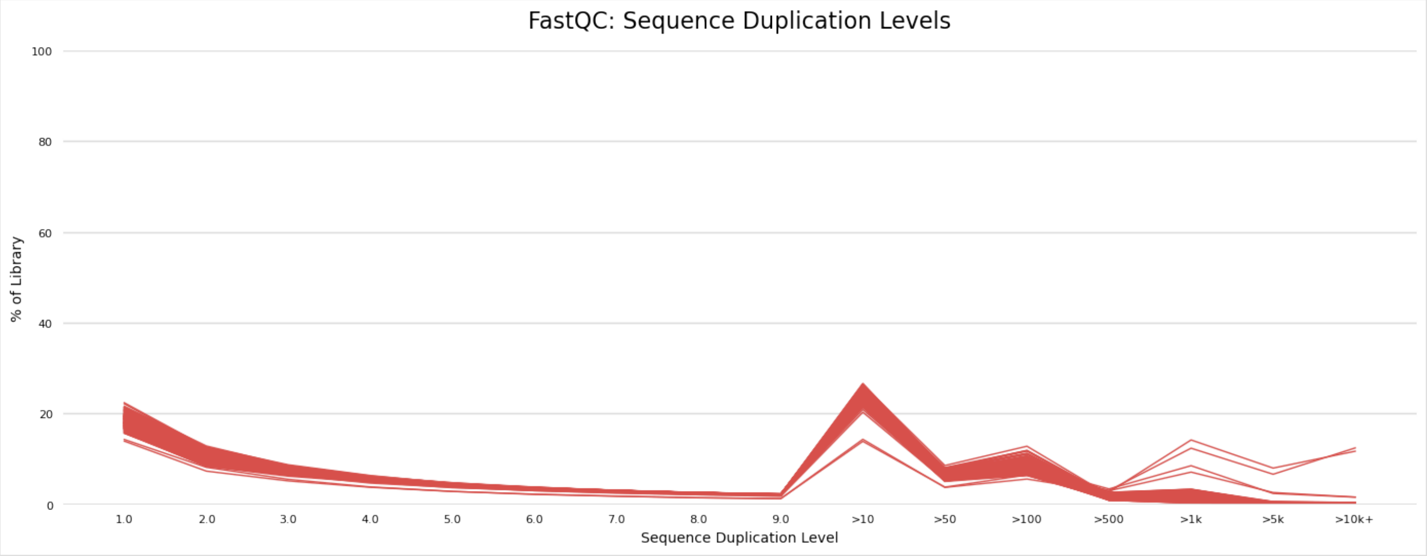


Sequence Length Distribution 336

All samples have sequences of a single length (101bp).

Sequence Duplication Levels

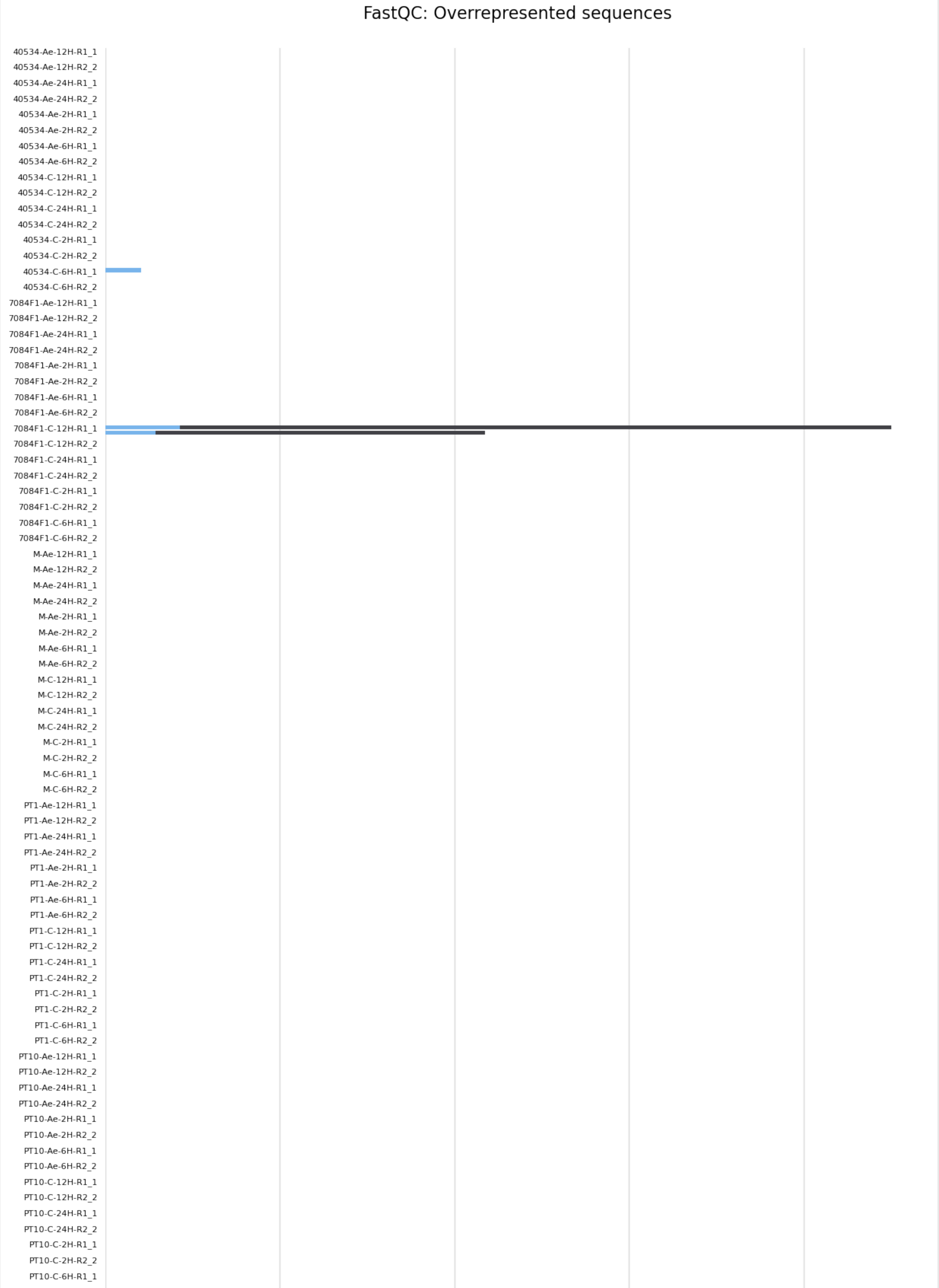
The relative level of duplication found for every sequence.

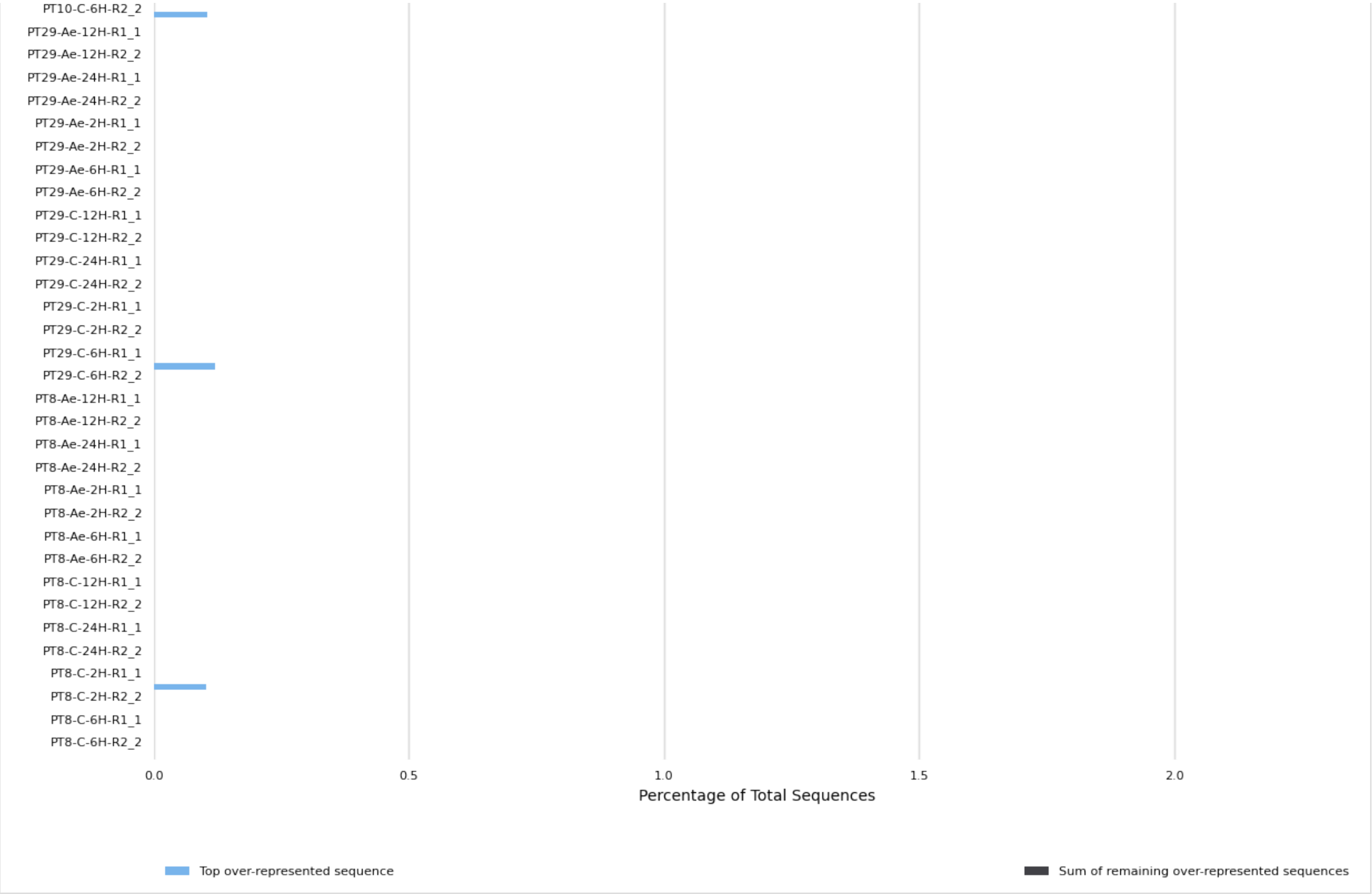


Overrepresented sequences

330

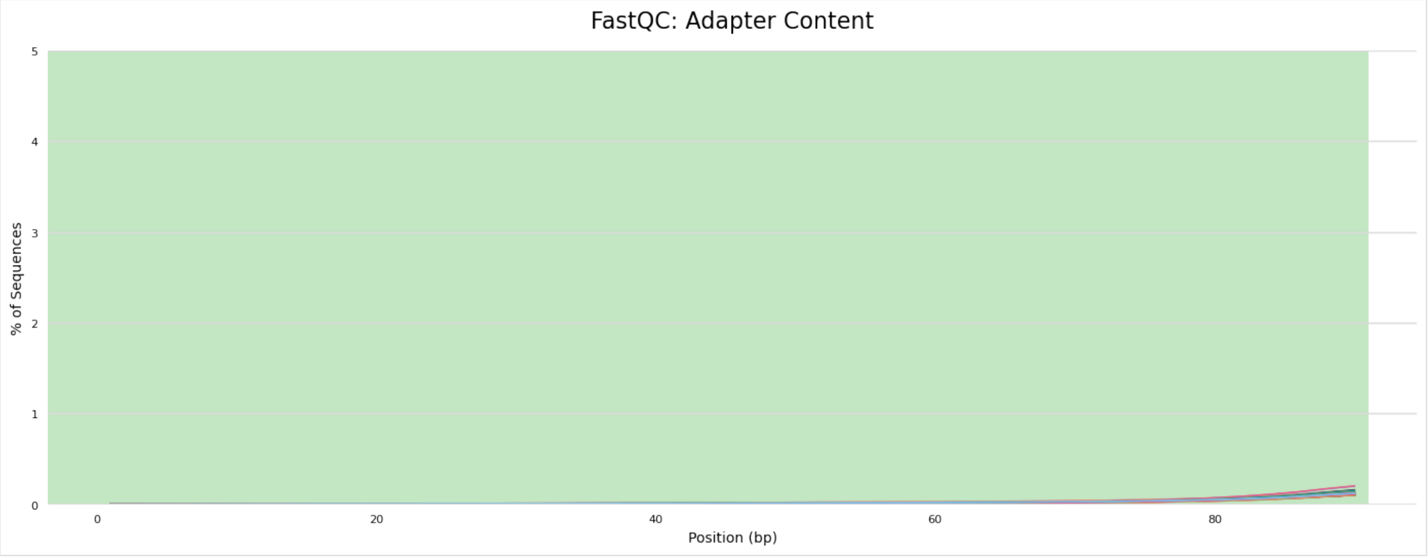
The total amount of overrepresented sequences found in each library.


FastQC: Overrepresented sequences
40534-Ae-12H-R1_1
40534-Ae-12H-R2_2
40534-Ae-24H-R1_1
40534-Ae-24H-R2_2
40534-Ae-2H-R1_1
40534-Ae-2H-R2_2
40534-Ae-6H-R1_1
40534-Ae-6H-R2_2
40534-C-12H-R1_1
40534-C-12H-R2_2
40534-C-24H-R1_1
40534-C-24H-R2_2
40534-C-2H-R1_1
40534-C-2H-R2_2
40534-C-6H-R1_1
40534-C-6H-R2_2
7084F1-Ae-12H-R1_1
7084F1-Ae-12H-R2_2
7084F1-Ae-24H-R1_1
7084F1-Ae-24H-R2_2
7084F1-Ae-2H-R1_1
7084F1-Ae-2H-R2_2
7084F1-Ae-6H-R1_1
7084F1-Ae-6H-R2_2
7084F1-C-12H-R1_1
7084F1-C-12H-R2_2
7084F1-C-24H-R1_1
7084F1-C-24H-R2_2
7084F1-C-2H-R1_1
7084F1-C-2H-R2_2
7084F1-C-6H-R1_1
7084F1-C-6H-R2_2
M-Ae-12H-R1_1
M-Ae-12H-R2_2
M-Ae-24H-R1_1
M-Ae-24H-R2_2
M-Ae-2H-R1_1
M-Ae-2H-R2_2
M-Ae-6H-R1_1
M-Ae-6H-R2_2
M-C-12H-R1_1
M-C-12H-R2_2
M-C-24H-R1_1
M-C-24H-R2_2
M-C-2H-R1_1
M-C-2H-R2_2
M-C-6H-R1_1
M-C-6H-R2_2
PT1-Ae-12H-R1_1
PT1-Ae-12H-R2_2
PT1-Ae-24H-R1_1
PT1-Ae-24H-R2_2
PT1-Ae-2H-R1_1
PT1-Ae-2H-R2_2
PT1-Ae-6H-R1_1
PT1-Ae-6H-R2_2
PT1-C-12H-R1_1
PT1-C-12H-R2_2
PT1-C-24H-R1_1
PT1-C-24H-R2_2
PT1-C-2H-R1_1
PT1-C-2H-R2_2
PT1-C-6H-R1_1
PT1-C-6H-R2_2
PT10-Ae-12H-R1_1
PT10-Ae-12H-R2_2
PT10-Ae-24H-R1_1
PT10-Ae-24H-R2_2
PT10-Ae-2H-R1_1
PT10-Ae-2H-R2_2
PT10-Ae-6H-R1_1
PT10-Ae-6H-R2_2
PT10-C-12H-R1_1
PT10-C-12H-R2_2
PT10-C-24H-R1_1
PT10-C-24H-R2_2
PT10-C-2H-R1_1
PT10-C-2H-R2_2
PT10-C-6H-R1_1



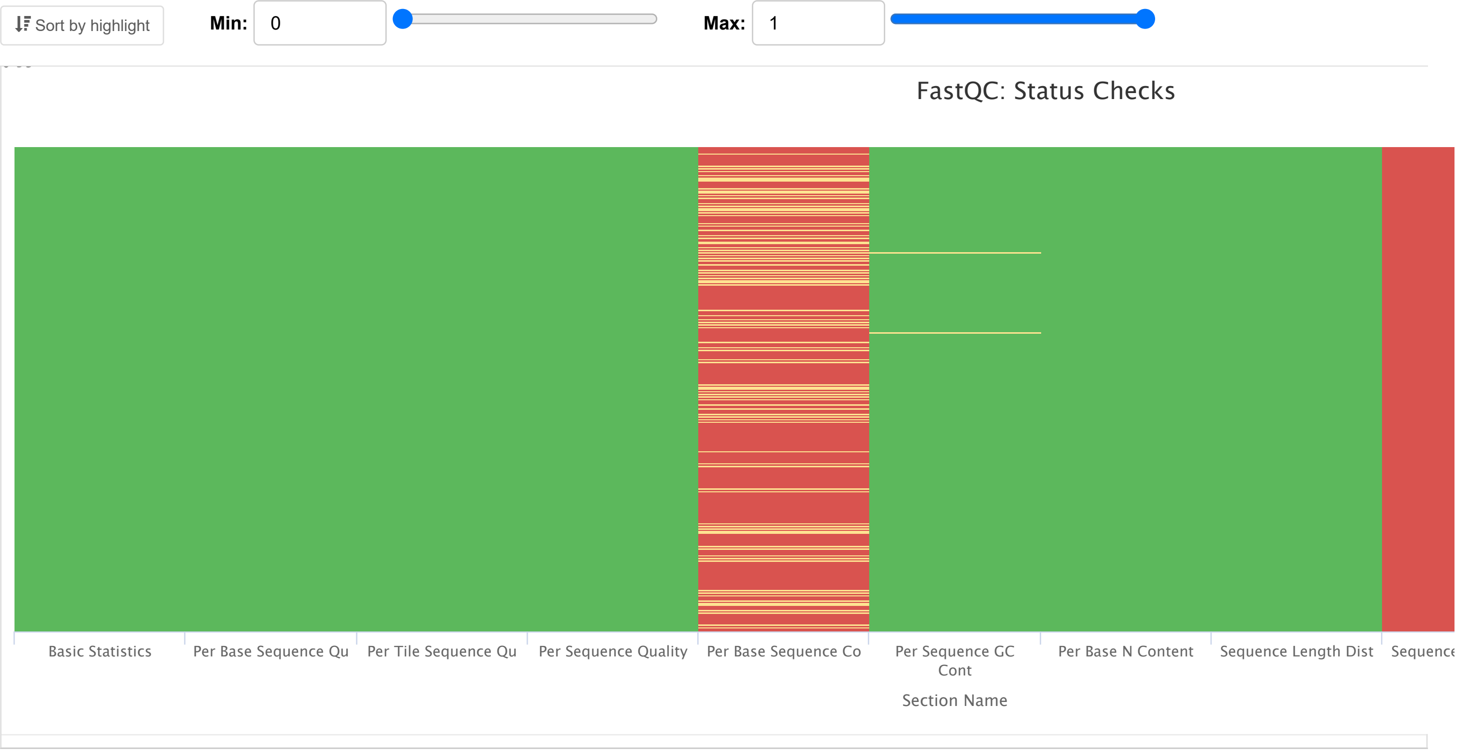
Adapter Content 336

The cumulative percentage count of the proportion of your library which has seen each of the adapter sequences at each position.



Status Checks

Status for each FastQC section showing whether results seem entirely normal (green), slightly abnormal (orange) or very unusual (red).



Cutadapt

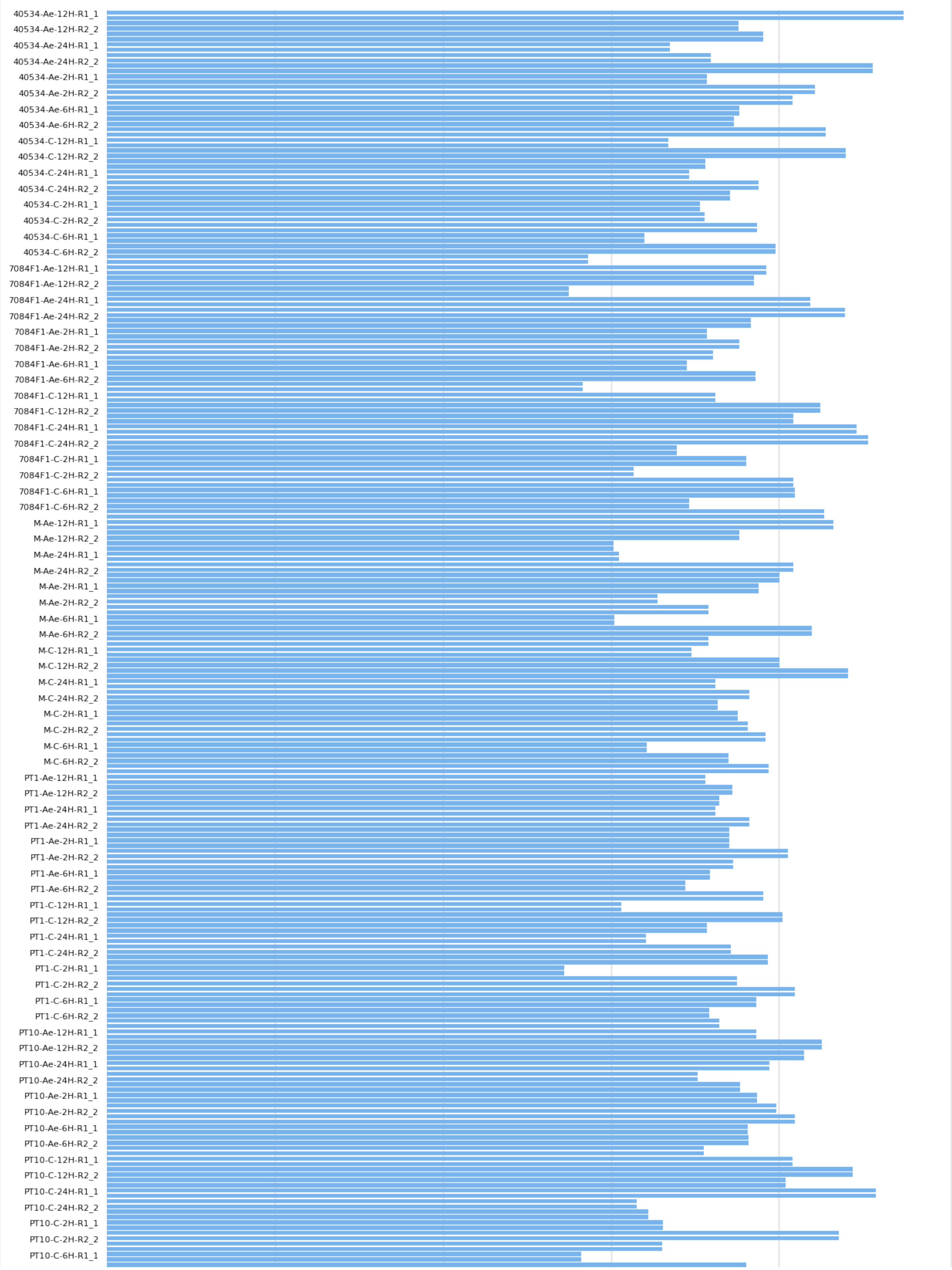
Cutadapt is a tool to find and remove adapter sequences, primers, poly-A tails and other types of unwanted sequence from your high-throughput sequencing reads. *DOI: 10.14806/ej.17.1.200.*

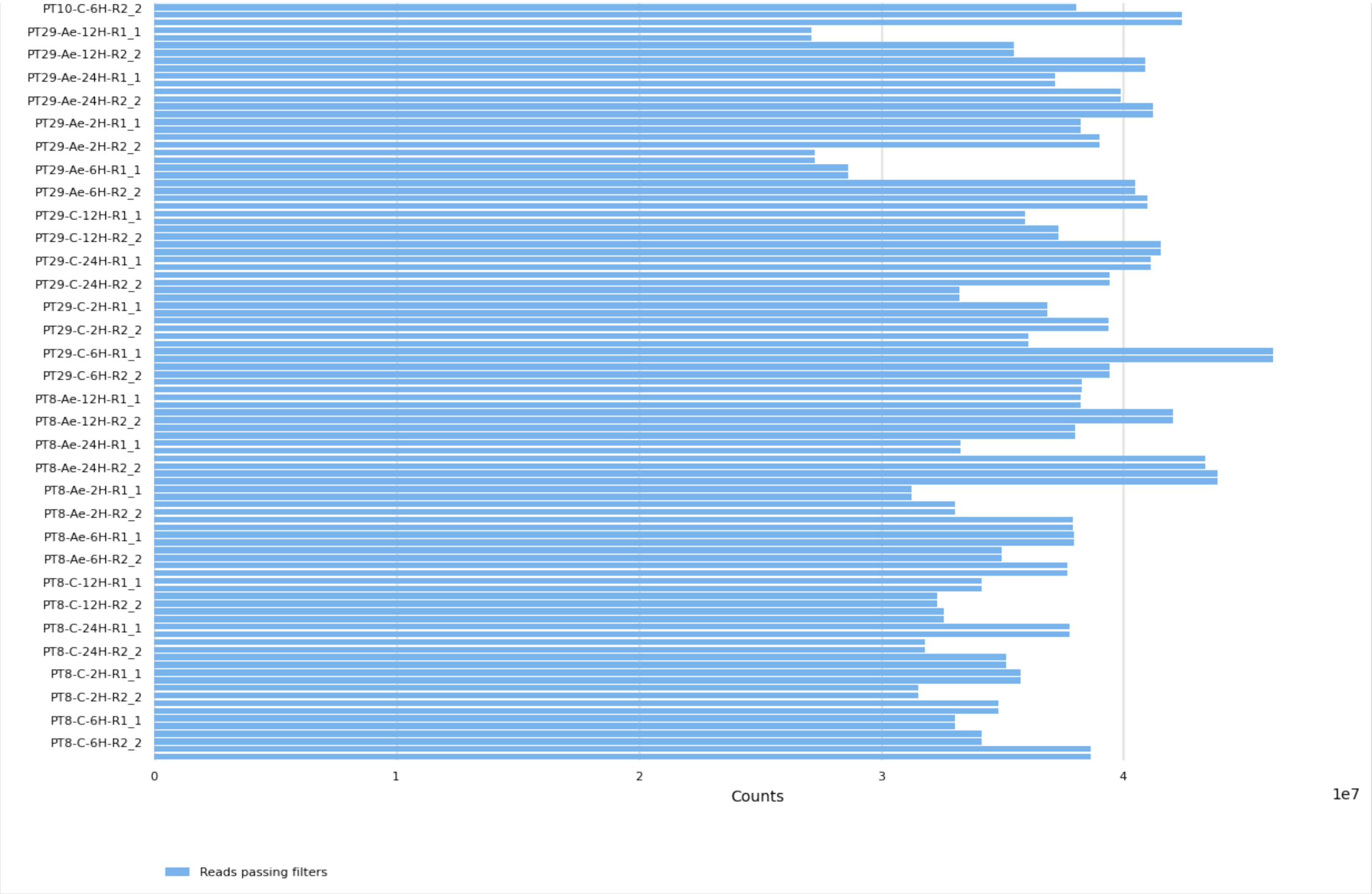
Filtered Reads

This plot shows the number of reads (SE) / pairs (PE) removed by Cutadapt.

Counts | Percentages

Cutadapt: Filtered Reads

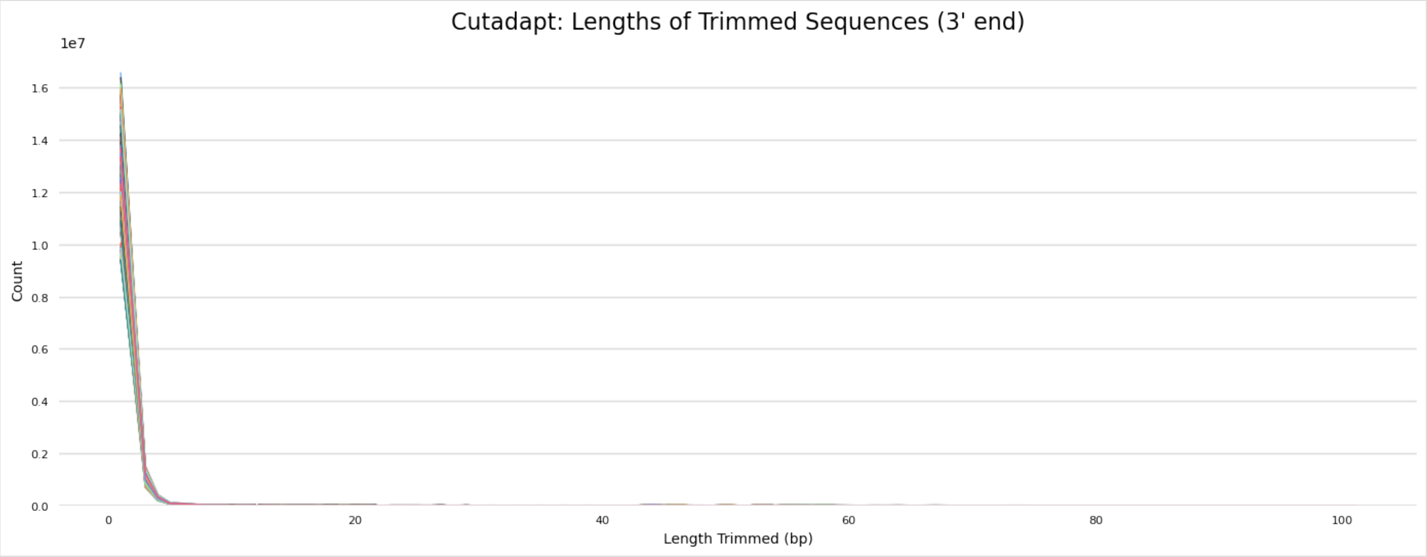




Trimmed Sequence Lengths (3')

This plot shows the number of reads with certain lengths of adapter trimmed for the 3' end.

Counts Obs/Exp



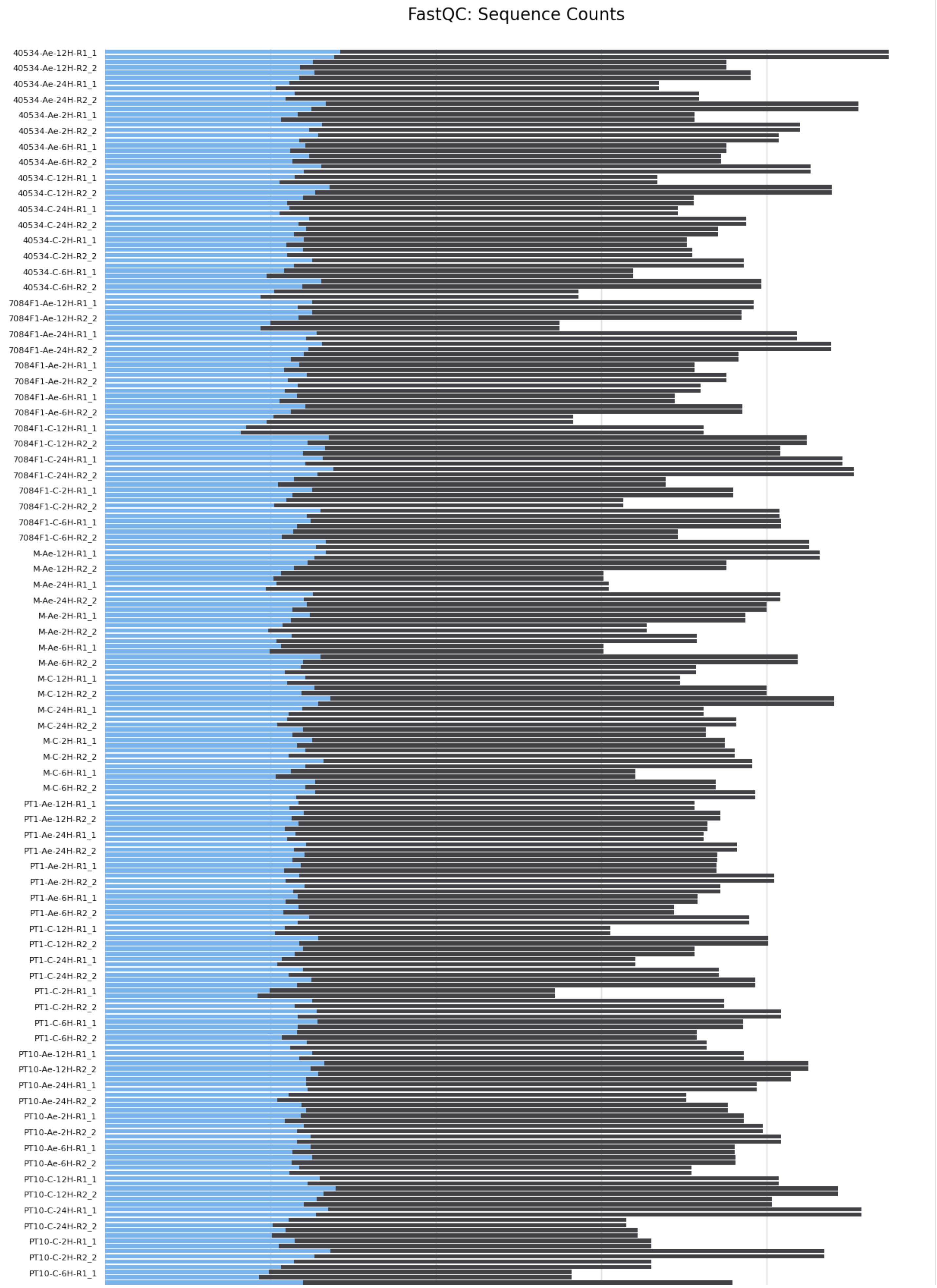
FastQC (trimmed)

FastQC (trimmed) This section of the report shows FastQC results after adapter trimming.

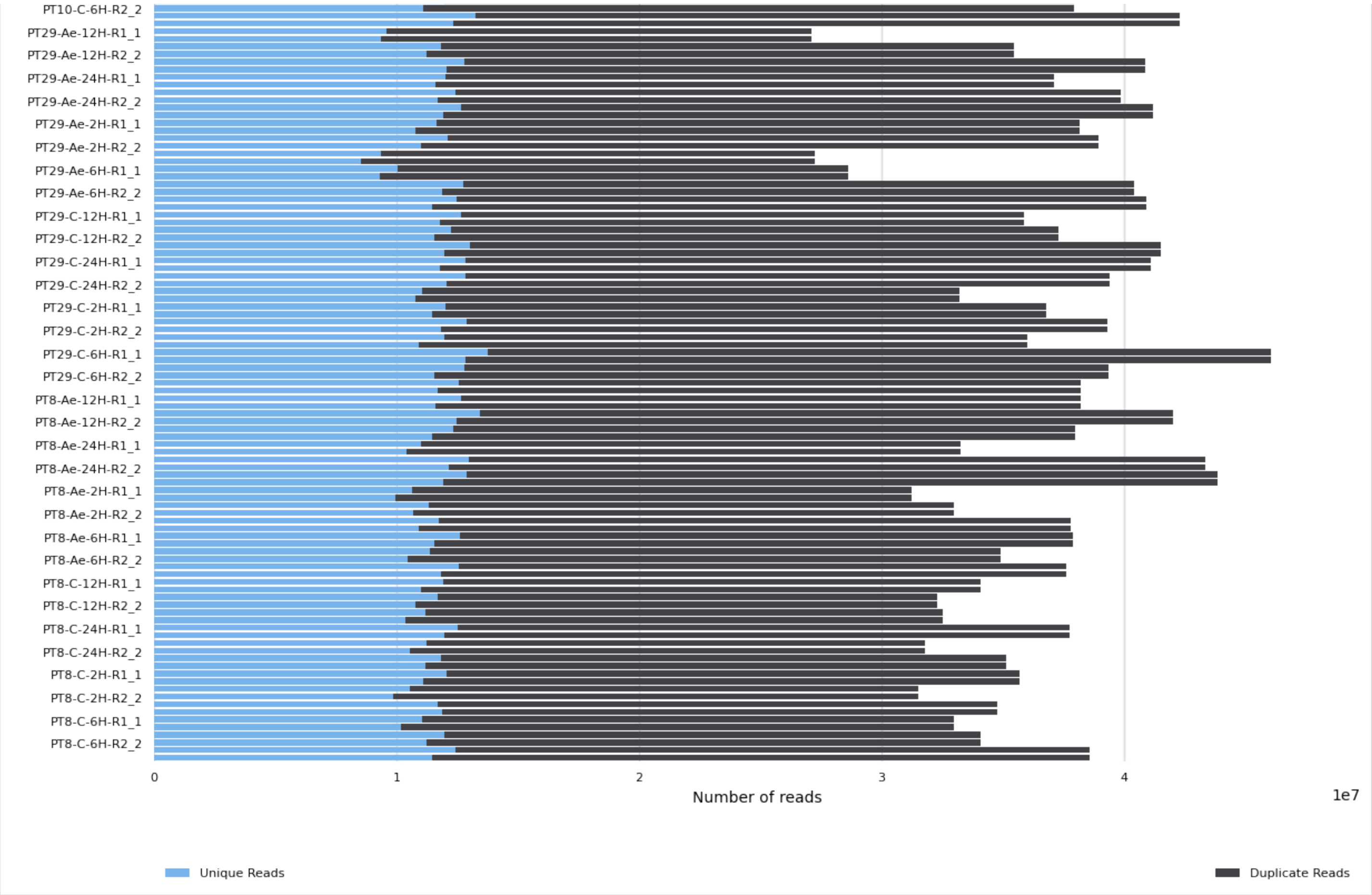
Sequence Counts

Sequence counts for each sample. Duplicate read counts are an estimate only.

Number of reads | Percentages



FastQC: Sequence Counts



Sequence Quality Histograms

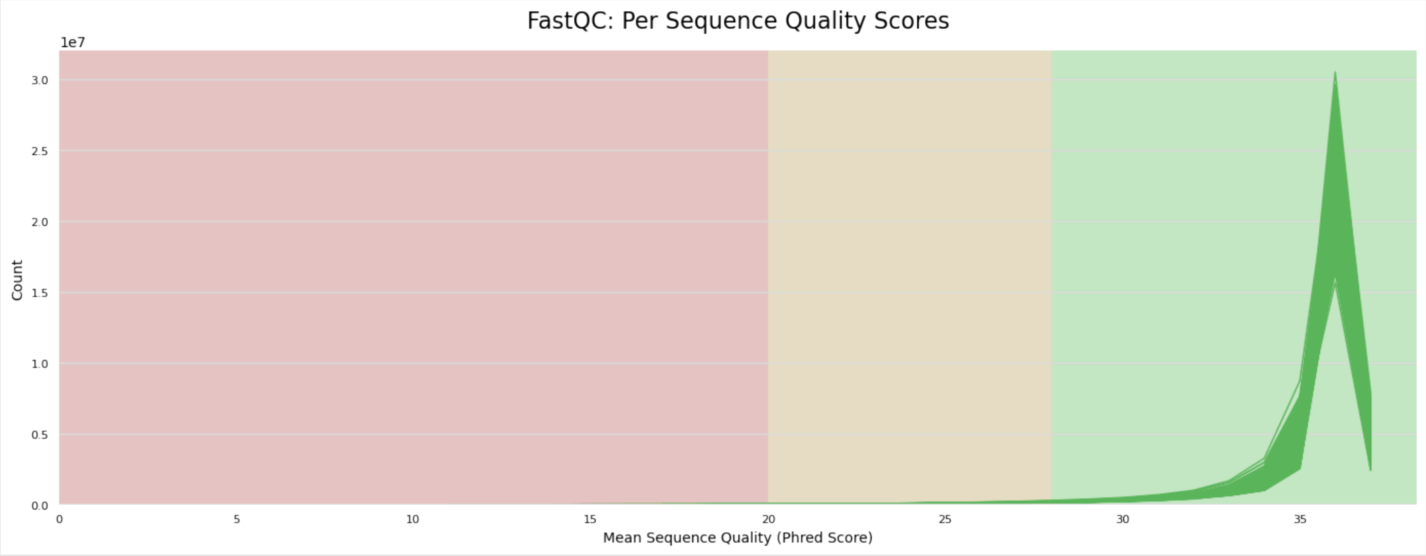
336

The mean quality value across each base position in the read.

Per Sequence Quality Scores

336

The number of reads with average quality scores. Shows if a subset of reads has poor quality.

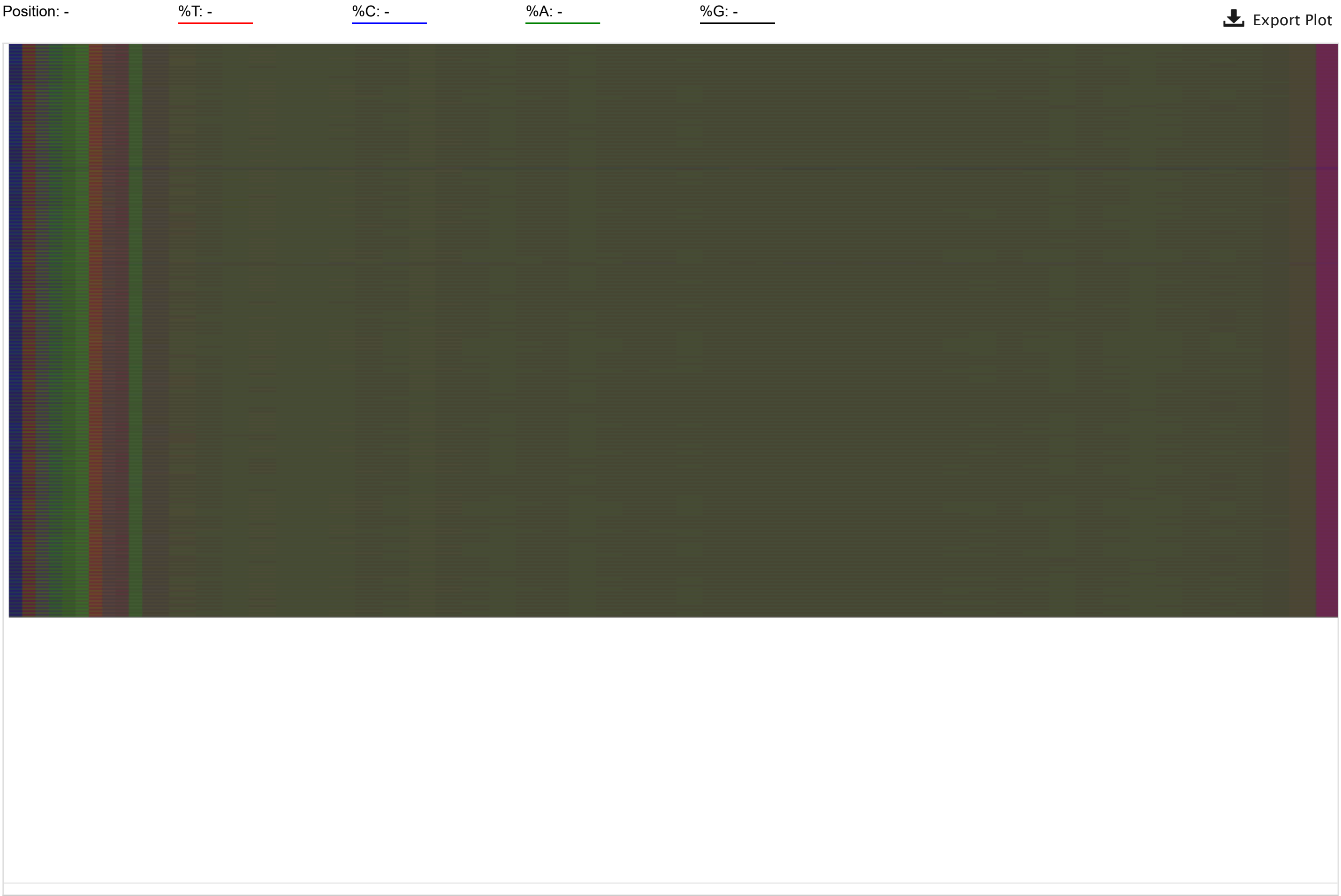


Per Base Sequence Content

The proportion of each base position for which each of the four normal DNA bases has been called.

Click a sample row to see a line plot for that dataset.

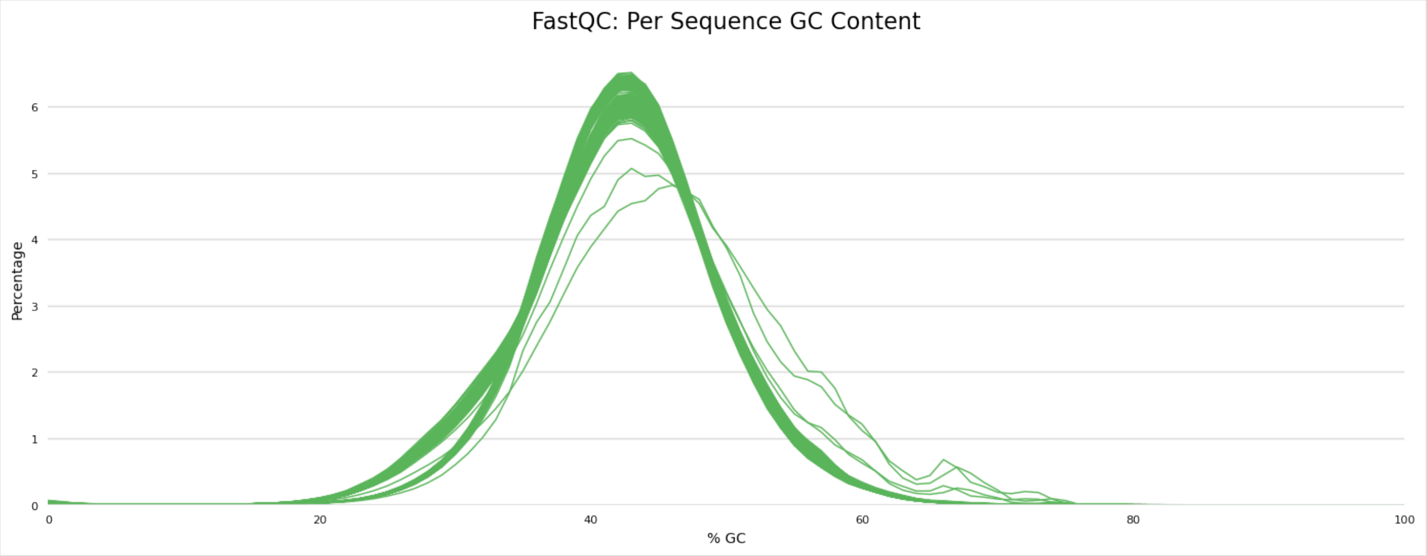
Rollover for sample name



Per Sequence GC Content 336

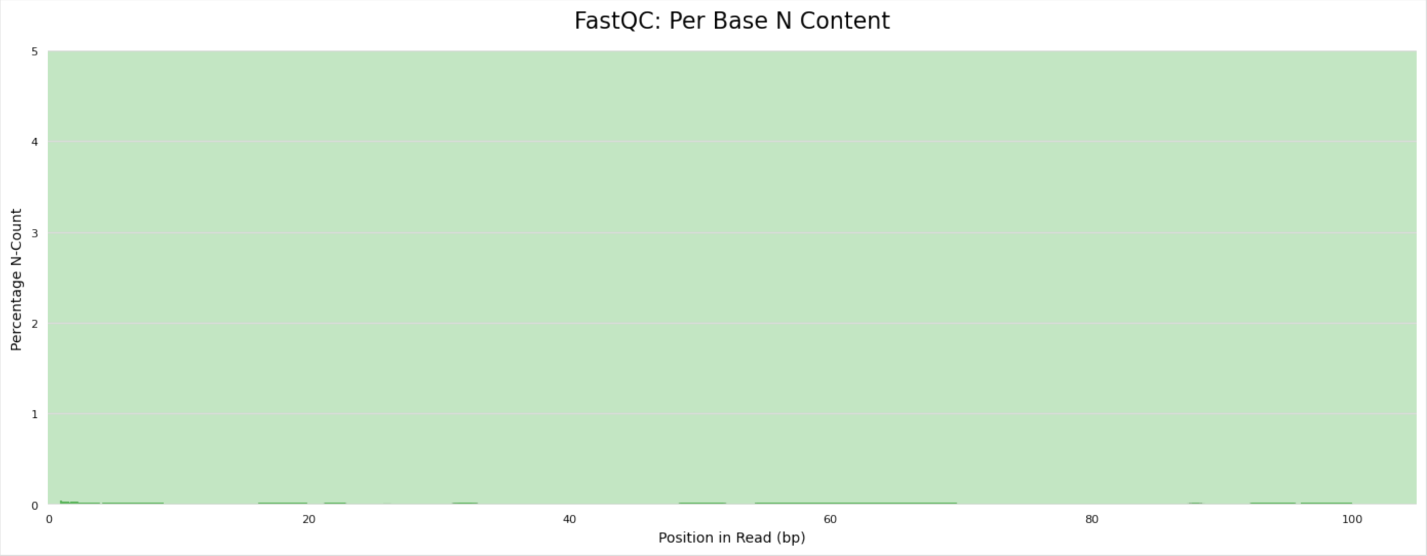
The average GC content of reads. Normal random library typically have a roughly normal distribution of GC content.

Percentages Counts



Per Base N Content 336

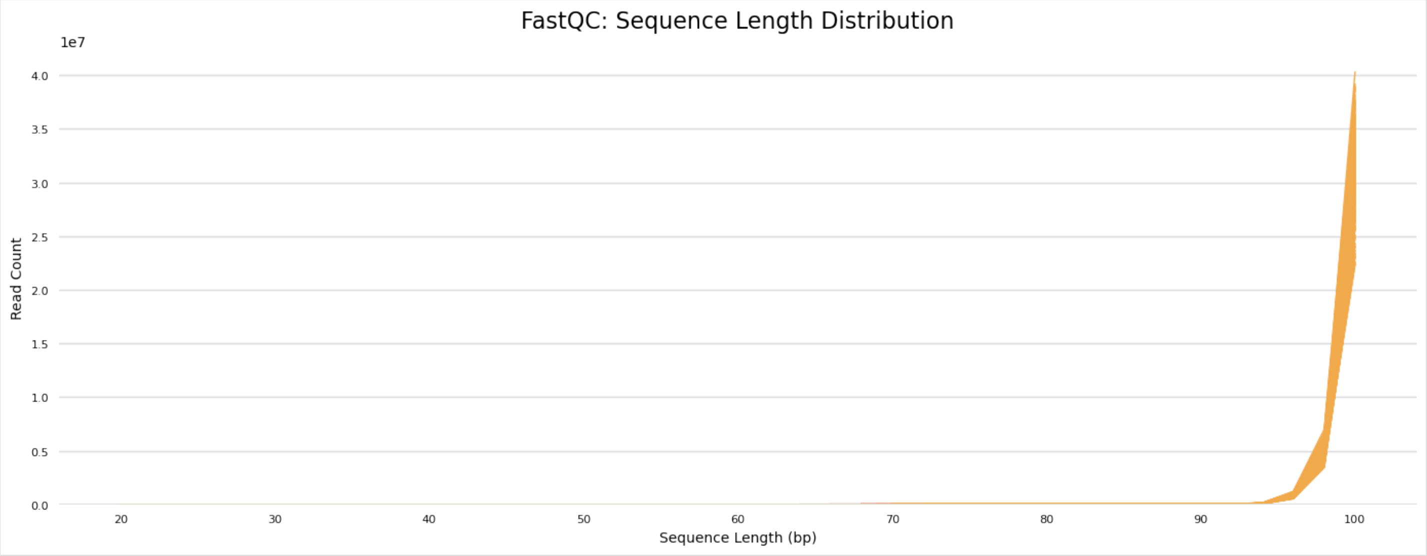
The percentage of base calls at each position for which an `N` was called.



Sequence Length Distribution

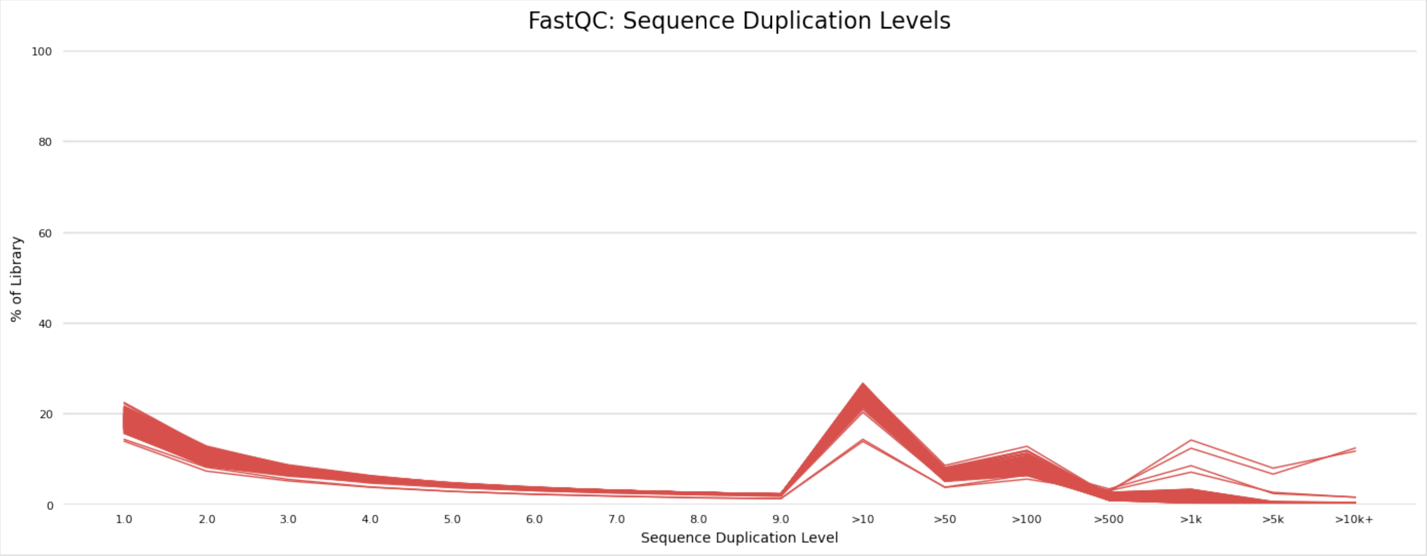
336

The distribution of fragment sizes (read lengths) found. See the FastQC help



Sequence Duplication Levels

The relative level of duplication found for every sequence.

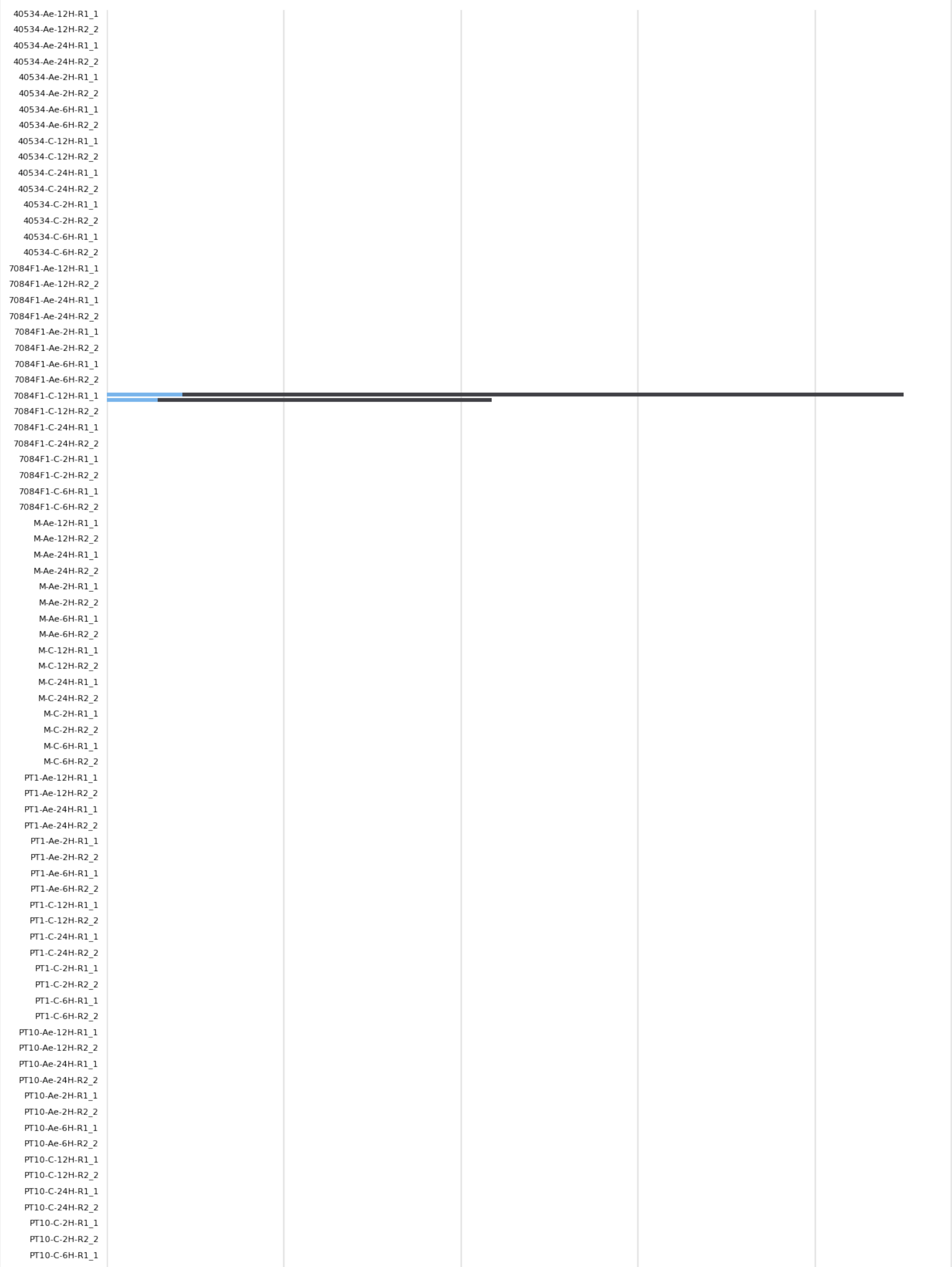


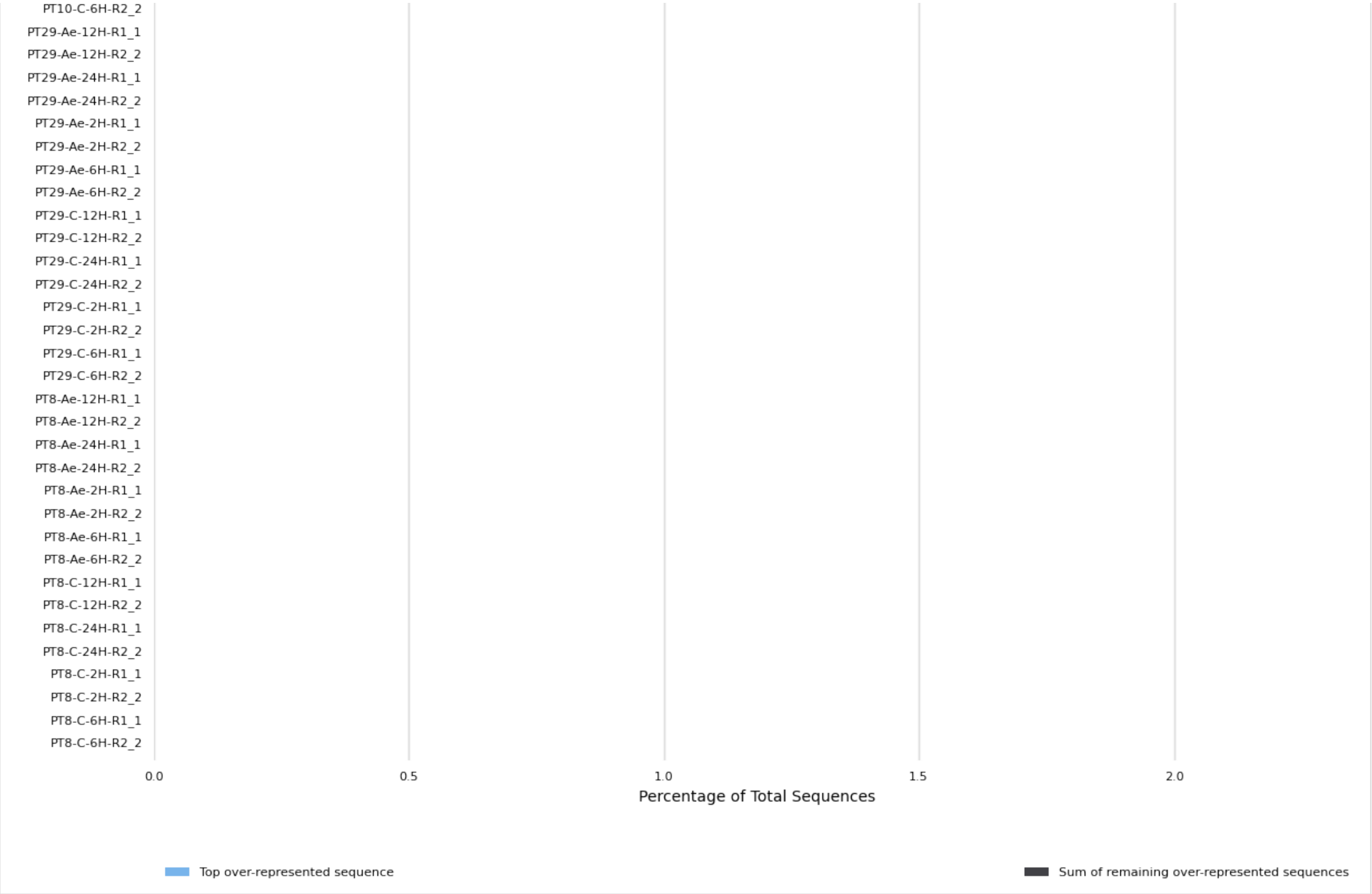
Overrepresented sequences

334

The total amount of overrepresented sequences found in each library.

FastQC: Overrepresented sequences





Adapter Content

336

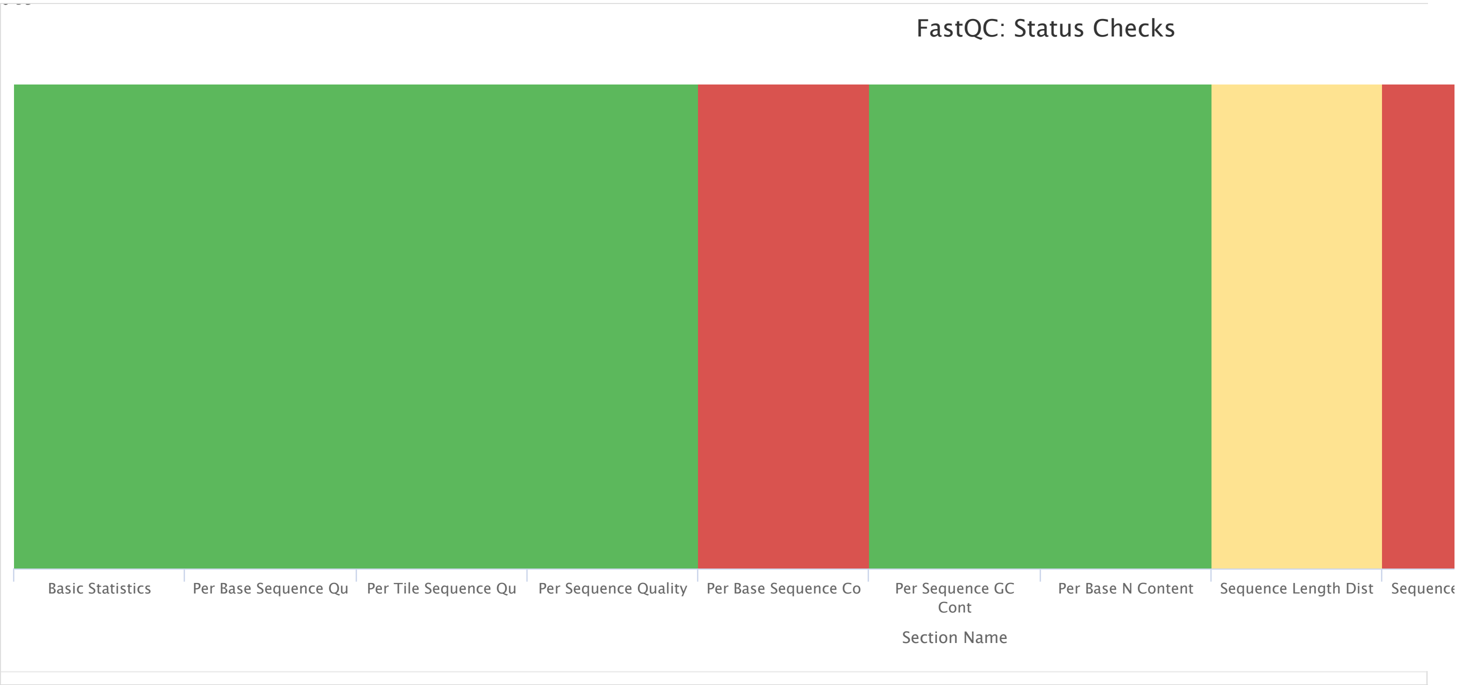
The cumulative percentage count of the proportion of your library which has seen each of the adapter sequences at each position.

No samples found with any adapter contamination > 0.1%

Status Checks

Status for each FastQC section showing whether results seem entirely normal (green), slightly abnormal (orange) or very unusual (red).

Sort by highlight **Min:** 0 **Max:** 1



nf-core/rnaseq Methods Description

Suggested text and references to use when describing pipeline usage within the methods section of a publication.

Methods

Data was processed using nf-core/rnaseq v3.10.1 (doi: https://doi.org/10.5281/zenodo.1400710) of the nf-core collection of workflows (Ewels *et al.*, 2020).

The pipeline was executed with Nextflow v22.10.4 (Di Tommaso *et al.*, 2017) with the following command:

```
nextflow run nf-core/rnaseq -r 3.10.1 -profile singularity -c /home/AAFC-AAC/limna/biocluster.conf -w /isilon/projects/J-002368_BeeCSI/sijan_rnase
q_aphanomyces/work -resume -params-file /isilon/projects/J-002368_BeeCSI/sijan_rnaseq_aphanomyces/nfcore_main/nf-params.json
```

References

- Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. Nature Biotechnology, 35(4), 316-319. https://doi.org/10.1038/nbt.3820
- Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology, 38(3), 276-278. https://doi.org/10.1038/s41587-020-0439-x

Notes:

- The command above does not include parameters contained in any configs or profiles that may have been used. Ensure the config file is also uploaded with your publication!
- You should also cite all software used within this run. Check the "Software Versions" of this report to get version information.

nf-core/rnaseq Software Versions

are collected at run time from the software output.

Process Name	Software	Version
BEDTOOLS_GENOMECOV	bedtools	2.30.0
CUSTOM_DUMPSOFTWAREVERSIONS	python	3.10.6
	yaml	6.0
CUSTOM_GETCHROMSIZES	getchromsizes	1.16.1
DESEQ2_QC_RSEM	bioconductor-deseq2	1.28.0
	r-base	4.0.3
DUPRADAR	bioconductor-dupradar	1.28.0
	r-base	4.2.1
FASTQC	fastqc	0.11.9
GTF2BED	perl	5.26.2
GTF_GENE_FILTER	python	3.9.5
GUNZIP_FASTA	gunzip	1.10
GUNZIP_GTF	gunzip	1.10
MAKE_TRANSCRIPTS_FASTA	rsem	1.3.1
	star	2.7.10a
MULTIQC_CUSTOM_BIOTYPE	python	3.9.5
PICARD_MARKDUPLICATES	picard	2.27.4-SNAPSHOT
QUALIMAP_RNASEQ	qualimap	2.2.2-dev
RSEM_CALCULATEEXPRESSION	rsem	1.3.1
	star	2.7.10a
RSEM_MERGE_COUNTS	sed	4.7
RSEM_PREPAREREFERENCE_GENOME	rsem	1.3.1
	star	2.7.10a
RSEQC_BAMSTAT	rseqc	3.0.1
RSEQC_INFEREXPERIMENT	rseqc	3.0.1
RSEQC_JUNCTIONANNOTATION	rseqc	3.0.1
RSEQC_JUNCTIONSATURATION	rseqc	3.0.1
RSEQC_READDUPLICATION	rseqc	3.0.1
SALMON_INDEX	salmon	1.9.0
SAMPLESHEET_CHECK	python	3.9.5
SAMTOOLS_FLAGSTAT	samtools	1.16.1
SAMTOOLS_IDXSTATS	samtools	1.16.1
SAMTOOLS_INDEX	samtools	1.16.1
SAMTOOLS_SORT	samtools	1.16.1
SAMTOOLS_STATS	samtools	1.16.1
STRINGTIE_STRINGTIE	stringtie	2.2.1
SUBREAD_FEATURECOUNTS	subread	2.0.1
TRIMGALORE	cutadapt	3.4

Process Name	Software	Version
	`trimgalore`	`0.6.7`
`UCSC_BEDCLIP`	`ucsc`	`377`
`UCSC_BEDGRAPHTOBIGWIG`	`ucsc`	`377`
`Workflow`	`Nextflow`	`22.10.4`
	`nf-core/rnaseq`	`3.10.1`

nf-core/rnaseq Workflow Summary

- this information is collected when the pipeline is started.

Core Nextflow options

revision	`3.10.1`
runName	`nostalgic_carlsson`
containerEngine	`singularity`
launchDir	`/isilon/projects/J-002368_BeeCSI/sijan_rnaseq_aphanomyces`
workDir	`/isilon/projects/J-002368_BeeCSI/sijan_rnaseq_aphanomyces/work`
projectDir	`/home/AAFC-AAC/limna/.nextflow/assets/nf-core/rnaseq`
userName	`limna`
profile	`singularity`
configFiles	`/home/AAFC-AAC/limna/.nextflow/assets/nf-core/rnaseq/nextflow.config, /home/AAFC-AAC/limna/biocluster.conf`

Input/output options

input	`/isilon/projects/J-002368_BeeCSI/sijan_rnaseq_aphanomyces/nfcore_metadata/sample_fastq_mapping.csv`
outdir	`/isilon/projects/J-002368_BeeCSI/sijan_rnaseq_aphanomyces/output`
email	`nathanielzhin-loong.lim@agr.gc.ca`
multiqc_title	`Sijan_Aphanomyces_Main`

Reference genome options

fasta
`/isilon/projects/J-002368_BeeCSI/sijan_rnaseq_aphanomyces/nfcore_genome/Pisum_sativum.Pisum_sativum_v1a.dna.toplevel.fa.gz`
gtf
`/isilon/projects/J-002368_BeeCSI/sijan_rnaseq_aphanomyces/nfcore_genome/Pisum_sativum.Pisum_sativum_v1a.55.gtf.gz`

Read trimming options

trim_nextseq	`20`

Alignment options

aligner	`star_rsem`
bam_csi_index	`true`
seq_center	`AAFC-AAC`

Max job request options

max_memory	`192.GB`

MultiQC v1.13 - Written by Phil Ewels, available on GitHub.
This report uses HighCharts, jQuery, jQuery UI, Bootstrap, FileSaver.js and clipboard.js.