# FASTQ QC Report

| | |
|---|---|
| Report Date | 10-02-16 |
| Run ID | 160930_D00796_0121_AC9MR4ANXX |
| Project ID | EC-EL-3883 |
| Sample | Sample_YD1_R1 |
| FASTX-Toolkit Version | 0.0.13.2 |
| FastQC Version | 0.10.1 |
| Dupest Version | 0.1.0 |

This report was automatically generated by the WCMC Epigenomics Core QC pipeline and contains information for assessing the quality of FASTQ sequencing data.

The QC Pipeline executes the following analysis:

1. All FASTQ files for the sample are concatenated to a single file. For paired-end sequences, FASTQ files for each read are concatenated and processed separately, with an ”R1” or ”R2” appended to the sample name.
2. To identify genomic sequencing bias or low sequence diversity k-length oligonucleotide enrichment is calculated and plotted from the combined FASTQ file using FastQC. *Note:* FastQC only analyses the top 2% of the reads in the FASTQ file and the results are extrapolated over the remainder.
3. Duplication level is estimated from the combined FASTQ file as $(N - U)/N$ where $N$ is total reads and $U$ is the number of unique sequences.
4. Sequencing base call quality statistics are calculated from the combined FASTQ file using FASTX-Toolkit FASTQ Quality Filter.

The report contains the following figures:

1. Sequence Duplication - Estimate of duplication level as a percentage of total reads.
2. Base sequence quality - Calculated from FASTX-Toolkit FASTQ Quality Filter.

   Distribution of base quality scores (Q scores) per sequencing cycle. In a reasonably good sequencing run the majority of the signal should be above Q30. Quality scores are divided into three ranges: green indicates calls of very good quality; orange indicates calls of reasonable quality and red indicates calls of poor quality.

   Yellow boxes represent the inter-quartile range. Upper and lower whiskers represent the maximum and minimum excluding outliers. The red line represents the median quality and the blue line represents the mean quality.
3. Sequence base content - Percentage of bases represented at each position in the read; calculated from FASTX-Toolkit FASTQ Quality Filter.
4. K-mer content - calculated and plotted by FastQC. From FastQC Help:

   The k-mer analysis checks if there are short fragments of k-length nucleotides that are over represented at certain positions in the reads. In a diversified library there should not be positional bias in its appearance of k-mers. There may be biological reasons why certain k-mers are enriched or depleted overall, but these biases should affect all positions within a sequence equally. In contrast, if certain k-mers are over represented in certain positions then this could indicate issues with library preparation, quality of the input material or sequencing of the adaptors. This analysis measures the number of each 5-mer at each position in the library and then uses a binomial test to look for significant deviations from an even coverage at all positions. Any k-mer with positionally biased enrichment are reported. The top 6 most biased k-mers are additionally plotted to show their distribution. Note that because of the computational overhead associated with calculating k-mer content this analysis is performed on 2% of the reads.
5. Overrepresented sequences - Calculated and plotted by FastQC. From FastQC Help:

   A normal high-throughput library will contain a diverse set of sequences, with no individual sequence making up a tiny fraction of the whole. Finding that a single sequence is very overrepresented in the set either means that it is highly biologically significant, or indicates that the library is contaminated, or not as diverse as you expected.

   This analysis lists all of the reads which make up more than 0.1% of the total. To limit memory use only sequences which appear in the first 200,000 sequences are evaluated for their occurrences in the entire library. It is possible that a sequence which is overrepresented but doesn’t appear at the start of the file for some reason could be missed by this analysis. However, this is unlikely since library preparation and sequencing randomize the genomic elements and therefore the first 200,000 reads are sufficient to represent the diversity in the entire library.

For each overrepresented sequence the program will look for matches in a database of common contaminants and will report the best hit it finds. Hits must be at least 20bp in length and have no more than 1 mismatch. Finding a hit doesn't necessarily mean that this is the source of the contamination, but may provide clues about the true source of contamination. It's also worth pointing out that many adapter sequences are similar in sequence so a match to an adaptor sequence may not represent the true source of the adaptor.

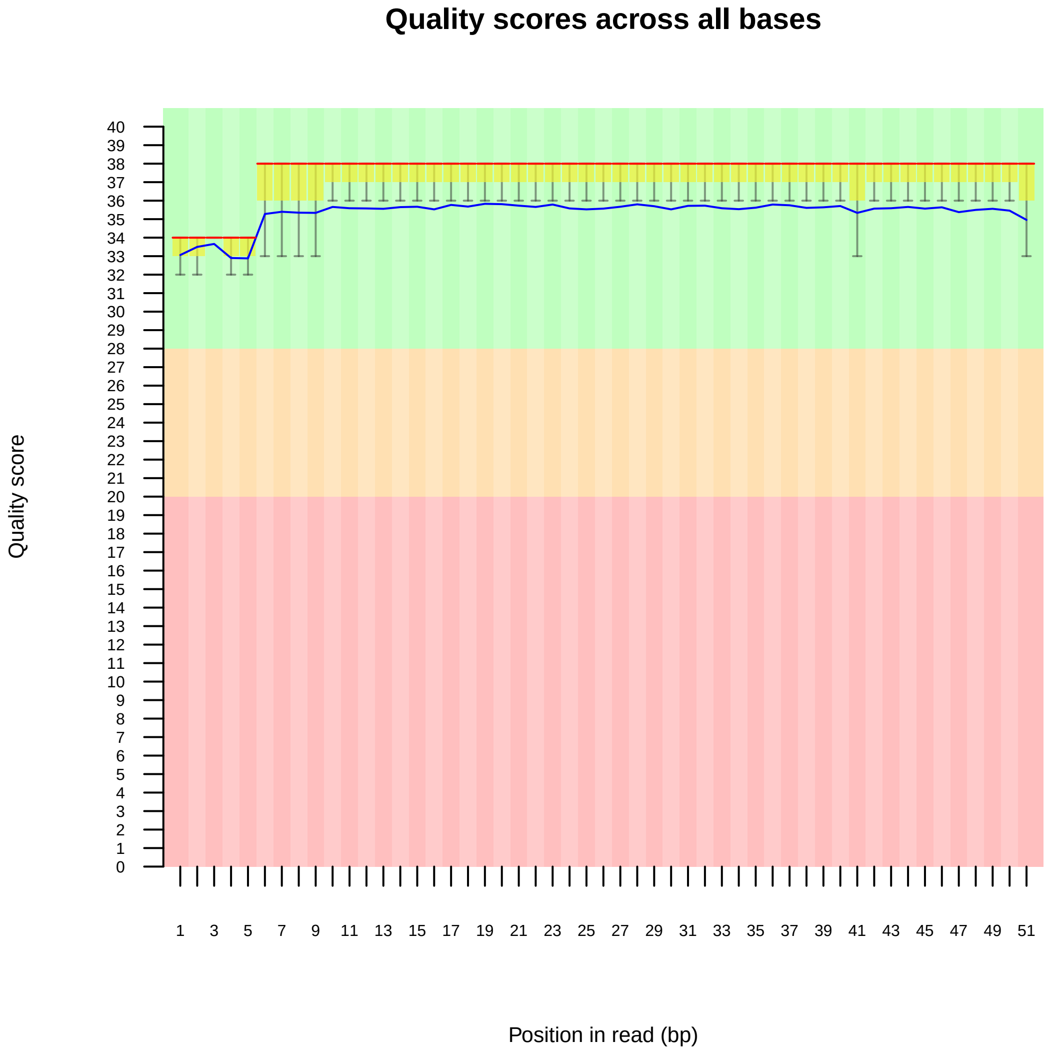
Because the duplication detection requires an exact sequence match over the whole length of the sequence. Reads over 75bp in length are truncated to 50bp for the purposes of this analysis.

```
FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
FASTX-Toolkit:  http://hannonlab.cshl.edu/fastx_toolkit
```

# 1 Sequence Duplication

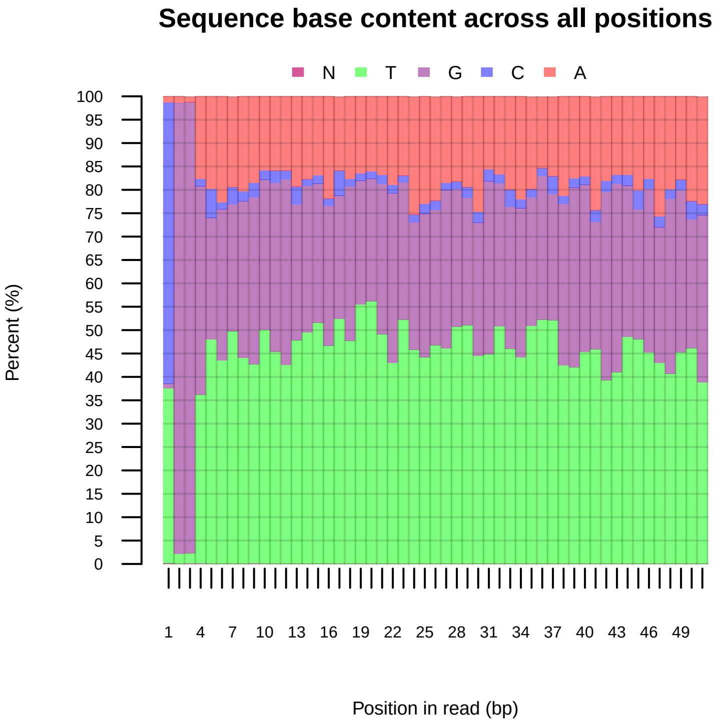
- Estimated Duplication rate 73.7312%

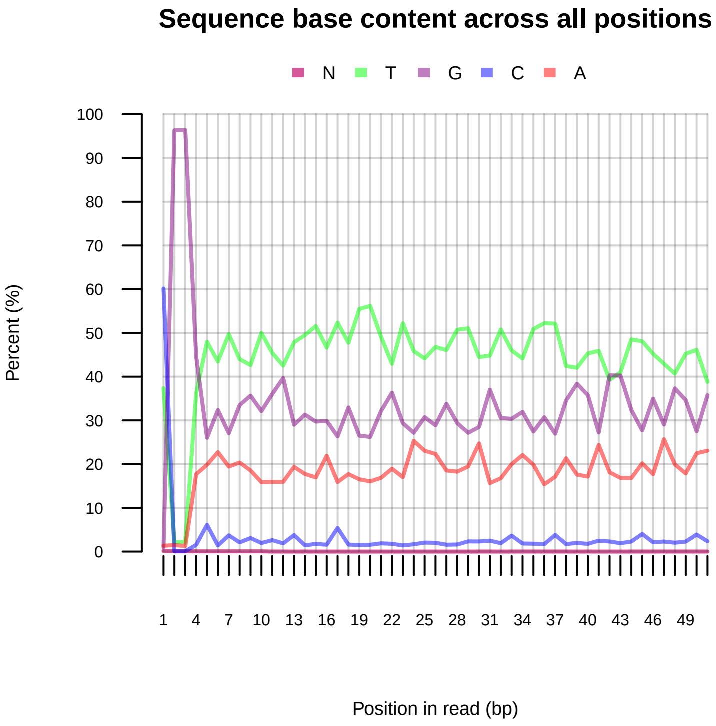
# 2 Per base sequence quality



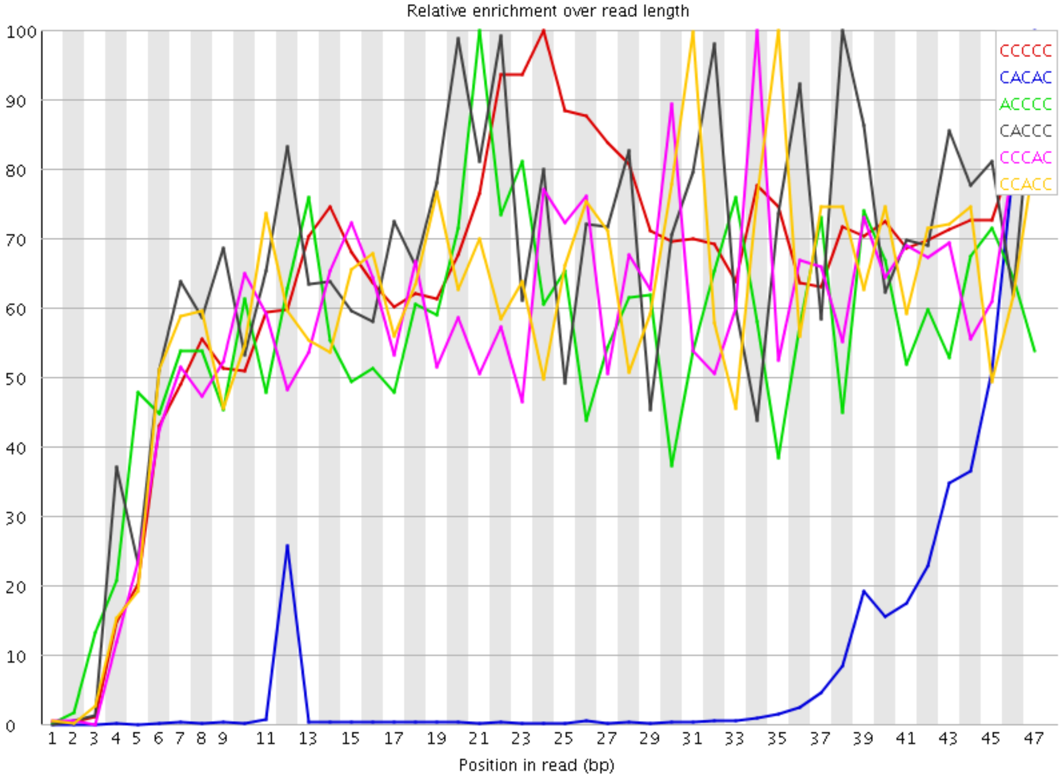
| | |
|---|---|
| Background colors | Green - calls of very good quality<br>Orange - calls of reasonable quality<br>Red - calls of poor quality |
| Yellow boxes | Inter-quartile range |
| Upper and lower whiskers | Maximum and minimum quality excluding outliers |
| Red line | Median quality |
| Blue line | Mean quality |

# 3 Sequence base content





# 4 Sequence K-mer content



Note: FastQC analyses 2% of the sequence data and results are extrapolated to the rest of the sequence.

| Sequence | Count | Obs/Exp Overall | Obs/Exp Max | Max Obs/Exp Position |
|---|---|---|---|---|
| CCCCC | 117105 | 745.79395 | 1162.6891 | 24 |
| CACAC | 429625 | 95.411575 | 1029.5717 | 47 |
| ACCCC | 41715 | 49.610054 | 89.69723 | 21 |
| CACCC | 40955 | 48.706215 | 74.33079 | 38 |
| CCCAC | 39970 | 47.534794 | 83.27361 | 34 |
| CCACC | 39640 | 47.142338 | 79.64081 | 35 |
| CCCCA | 38665 | 45.982807 | 69.29922 | 24 |
| CGGGC | 4330065 | 27.410915 | 975.74524 | 1 |
| CCCGC | 24605 | 15.638508 | 32.25727 | 26 |
| GCCCC | 24285 | 15.435122 | 30.914362 | 47 |
| CGCCC | 23670 | 15.044238 | 25.238813 | 32 |
| CCGCC | 23525 | 14.952079 | 28.225252 | 27 |
| CCCCG | 23410 | 14.878988 | 30.912964 | 25 |
| CGCGC | 231830 | 14.7051735 | 60.83841 | 13 |
| AGCAC | 618040 | 13.697984 | 106.35402 | 45 |
| CGCGG | 2043440 | 12.935732 | 288.67917 | 5 |
| GCGCG | 1963470 | 12.429492 | 285.97922 | 4 |
| GCACA | 524300 | 11.620368 | 106.505646 | 46 |
| CGGCG | 1679615 | 10.632585 | 284.0674 | 1 |
| CGGAA | 4453825 | 9.851488 | 207.75189 | 1 |
| GGCGC | 1407840 | 8.912148 | 286.51587 | 3 |
| CGGGA | 6366400 | 7.5258765 | 224.30069 | 1 |
| ACACG | 325645 | 7.2174616 | 83.51301 | 47 |
| TCGCG | 1478785 | 7.101227 | 24.347147 | 30 |
| TCCCC | 13705 | 6.607698 | 15.188635 | 5 |
| CGGGT | 12501355 | 5.991203 | 230.86655 | 1 |
| AGACG | 2638945 | 5.8371253 | 63.666187 | 27 |
| CTCCC | 12060 | 5.8145814 | 10.422622 | 47 |
| CCTCC | 12015 | 5.792885 | 11.781795 | 40 |
| ACGCG | 481050 | 5.6980314 | 21.776299 | 6 |
| CGCGT | 1175310 | 5.643918 | 22.292543 | 31 |
| CGTCG | 1175210 | 5.6434383 | 25.275799 | 41 |
| CCCCT | 11590 | 5.5879774 | 9.062976 | 38 |
| CCCTC | 11580 | 5.583156 | 9.742635 | 39 |
| CGGAG | 4632135 | 5.4757595 | 160.73503 | 1 |
| CGCGA | 426125 | 5.047445 | 18.67137 | 5 |
| AGATC | 2995555 | 5.0262623 | 25.113857 | 43 |
| CGGAC | 418020 | 4.951442 | 152.06451 | 1 |
| CGGTC | 1021390 | 4.904784 | 171.37505 | 1 |
| CGGTT | 13466895 | 4.8957987 | 168.12839 | 1 |
| CGGGG | 7560375 | 4.776404 | 115.447655 | 1 |
| CGACG | 387210 | 4.5864973 | 32.14517 | 24 |
| GGGCG | 6961865 | 4.398285 | 102.263916 | 2 |

| | | | | |
|---|---|---|---|---|
| TCGAG | 4897360 | 4.391617 | 54.415215 | 44 |
| AGGCG | 3698655 | 4.372271 | 62.62182 | 47 |
| AAAAA | 2993330 | 4.328808 | 10.733278 | 31 |
| AACCC | 19270 | 4.279502 | 7.2531023 | 20 |
| GAGAC | 1914295 | 4.23426 | 61.119755 | 26 |
| ACACC | 18465 | 4.1007266 | 6.3660183 | 21 |
| ACCAC | 18235 | 4.0496483 | 6.157601 | 46 |
| CCACA | 16955 | 3.7653844 | 6.1053524 | 35 |
| CAACC | 16900 | 3.7531698 | 6.7836776 | 31 |
| ACCCA | 16660 | 3.6998703 | 6.6271496 | 33 |
| CACCA | 16530 | 3.6709998 | 5.8966517 | 45 |
| TTACG | 5294155 | 3.6012914 | 45.50449 | 14 |
| CCCAA | 16100 | 3.575505 | 6.731326 | 15 |
| CCAAC | 16075 | 3.569953 | 6.0009212 | 30 |
| CGAGG | 2955480 | 3.4937449 | 68.2526 | 45 |
| CGGTA | 3894840 | 3.4926252 | 119.910835 | 1 |
| GACGG | 2884115 | 3.4093833 | 34.144657 | 28 |
| CGTTT | 12337185 | 3.4022903 | 37.057297 | 17 |
| ACGTC | 372205 | 3.3443832 | 15.7567005 | 47 |
| GGCGG | 5266250 | 3.3270488 | 35.28336 | 11 |
| TACGT | 4818405 | 3.2776678 | 47.17076 | 15 |
| ACGTT | 4810745 | 3.2724571 | 48.710873 | 16 |
| ACTCC | 35960 | 3.237613 | 103.02234 | 23 |
| ACGGG | 2717540 | 3.2124708 | 34.1087 | 29 |
| GATCG | 3513570 | 3.1507285 | 14.582109 | 44 |
| GGCGT | 6508210 | 3.1190224 | 47.570316 | 3 |
| CGGAT | 3447235 | 3.0912437 | 104.69618 | 1 |
| CTCCA | 33505 | 3.0165799 | 102.684326 | 24 |
| AGAGA | 7271610 | 3.0035393 | 22.296162 | 25 |
| GCGGC | 473640 | 2.9983165 | 9.698414 | 9 |
| CGTTC | 815695 | 2.97136 | 30.565315 | 33 |
| TTTCG | 10486165 | 2.891825 | 14.7146015 | 30 |
| GTCGA | 3159135 | 2.8328955 | 53.921783 | 43 |
| AAGCG | 1260635 | 2.788419 | 51.360832 | 8 |
| CGAGA | 1253665 | 2.7730017 | 32.007812 | 25 |
| TTCGA | 4009730 | 2.7275753 | 32.597553 | 31 |
| AGCGA | 1232720 | 2.7266734 | 51.904068 | 9 |
| ATCGC | 301910 | 2.7127597 | 35.75201 | 29 |
| TTCGC | 733165 | 2.6707256 | 7.7958245 | 33 |
| AGAGC | 1174475 | 2.5978403 | 15.381723 | 47 |
| TTTTT | 161718920 | 2.5611978 | 5.493463 | 16 |
| GAGGC | 2137520 | 2.5268147 | 51.50613 | 46 |
| GCGGG | 3984565 | 2.517321 | 36.10282 | 12 |
| TCGTT | 9030555 | 2.4904037 | 6.0224857 | 36 |
| GGAGG | 21041230 | 2.4823496 | 28.47198 | 39 |
| CGTTA | 3601690 | 2.45001 | 28.532904 | 9 |
| GTCGC | 493865 | 2.3715734 | 11.411354 | 3 |
| CGGTG | 4946290 | 2.370481 | 44.57898 | 1 |
| TCGGA | 2626320 | 2.3551037 | 13.23687 | 46 |
| TTCGT | 8485865 | 2.3401916 | 5.893085 | 35 |
| TCGTC | 637185 | 2.3210957 | 10.211157 | 40 |
| ATCGG | 2578175 | 2.3119307 | 12.888706 | 45 |
| ATTCG | 3376910 | 2.2971063 | 36.9754 | 34 |
| TTTTA | 58240300 | 2.2751613 | 11.968877 | 26 |
| GGGAG | 19271630 | 2.27358 | 25.2672 | 38 |
| GCGGA | 1913075 | 2.2614925 | 21.101871 | 7 |
| ACGGA | 1021250 | 2.2589195 | 11.937143 | 30 |
| GACGC | 188140 | 2.228516 | 19.548552 | 5 |
| CACGT | 246800 | 2.2175782 | 21.491013 | 47 |
| GGTCG | 4625500 | 2.2167444 | 30.767872 | 42 |
| TTTAG | 42963810 | 2.2125452 | 15.0582695 | 27 |
| AGTAG | 13081310 | 2.1905224 | 20.057009 | 35 |
| CGTAG | 2428230 | 2.17747 | 23.029205 | 5 |
| GGAAG | 9851410 | 2.1746817 | 11.459644 | 2 |
| CGAGT | 2420395 | 2.170444 | 39.61435 | 33 |
| TACGC | 239970 | 2.1562088 | 9.90384 | 13 |
| ACGGC | 180270 | 2.135296 | 14.962502 | 32 |
| CGAGC | 179910 | 2.1310315 | 8.149218 | 32 |
| GAGGT | 23344550 | 2.089183 | 21.814428 | 40 |
| GCGTT | 5611430 | 2.0399976 | 23.10239 | 16 |
| ATTTT | 51750150 | 2.0216231 | 8.002555 | 25 |
| AGGAG | 9152015 | 2.0202916 | 8.651783 | 38 |
| GCGGT | 4214200 | 2.0196311 | 28.328953 | 6 |
| GAAGA | 4839630 | 1.9990097 | 6.1074204 | 46 |
| TTTAC | 3858930 | 1.9912587 | 33.575264 | 13 |
| AAACG | 473835 | 1.9611071 | 8.896878 | 7 |
| AGCGC | 164080 | 1.9435258 | 13.044966 | 35 |
| ATCGT | 2795270 | 1.9014521 | 14.693673 | 39 |
| TAGAG | 11345915 | 1.899923 | 9.977442 | 24 |
| AATTT | 19618000 | 1.8903822 | 17.447817 | 24 |
| GAGAT | 11255010 | 1.8847005 | 8.804213 | 26 |
| GAAAA | 2390455 | 1.8475186 | 5.3486495 | 3 |
| GCGAC | 155305 | 1.8395858 | 17.920506 | 23 |
| GCGTC | 382790 | 1.8381836 | 11.27538 | 40 |
| AGGTC | 2044340 | 1.833224 | 51.158577 | 41 |
| TAGTT | 35384765 | 1.8222404 | 8.471415 | 29 |
| TACGG | 2011790 | 1.8040354 | 14.072233 | 5 |
| AAGAG | 4355535 | 1.7990543 | 6.106158 | 47 |
| GGACG | 1518185 | 1.7946838 | 15.897143 | 2 |
| GAGCG | 1506005 | 1.7802855 | 9.820176 | 28 |
| GCGTA | 1983225 | 1.7784201 | 22.643696 | 4 |
| TTAGT | 34452090 | 1.7742096 | 14.457446 | 28 |
| GGAAA | 4293585 | 1.7734658 | 12.053032 | 2 |
| TAGTA | 13939075 | 1.7706374 | 15.723802 | 29 |
| TATCG | 2580860 | 1.755602 | 15.066344 | 38 |
| CACGC | 14585 | 1.7310613 | 6.302785 | 47 |
| AGGTA | 10249210 | 1.716275 | 25.29407 | 47 |
| GGAGA | 7760470 | 1.7131103 | 10.818168 | 2 |
| CGTAC | 190510 | 1.7117941 | 8.337296 | 13 |
| AGCGG | 1440540 | 1.7028978 | 5.132185 | 6 |
| GAGCA | 760965 | 1.6831908 | 12.167128 | 47 |
| TCGAC | 185790 | 1.6693836 | 8.484946 | 23 |
| GTAGA | 9958745 | 1.6676353 | 9.663302 | 23 |
| CGATT | 2445285 | 1.6633782 | 18.705809 | 11 |
| AACGG | 746665 | 1.6515604 | 10.443926 | 29 |
| GTCGT | 4529950 | 1.6468327 | 10.3899145 | 3 |
| CGCAC | 13860 | 1.6450127 | 5.940235 | 47 |

| | | | | |
|---|---|---|---|---|
| TCGAA | 980300 | 1.6448523 | 5.299588 | 32 |
| AGTTT | 31786925 | 1.6369593 | 7.96578 | 26 |
| AGTTA | 12868785 | 1.6346818 | 17.969532 | 30 |
| AGTCG | 1814850 | 1.627433 | 13.148586 | 22 |
| CGAAA | 392405 | 1.6240847 | 6.633343 | 32 |
| TATTT | 41244745 | 1.6112287 | 5.956659 | 32 |
| CGTGG | 3329835 | 1.5958043 | 31.093039 | 5 |
| AACGC | 71875 | 1.5930079 | 5.3221636 | 23 |
| GCGTG | 3300700 | 1.5818416 | 31.26083 | 4 |
| TGGGA | 17636485 | 1.5783488 | 14.398709 | 37 |
| GTACG | 1743895 | 1.5638053 | 13.780388 | 4 |
| TGGCG | 3258130 | 1.5614401 | 30.487566 | 10 |
| CGAAC | 69795 | 1.5469077 | 6.983545 | 29 |
| AGCGT | 1720915 | 1.5431987 | 7.771758 | 29 |
| ACGAG | 695710 | 1.5388522 | 5.198354 | 32 |
| GAACG | 695065 | 1.5374256 | 10.264055 | 28 |
| GGGAA | 6896040 | 1.5222888 | 13.339953 | 2 |
| TAGGA | 9015090 | 1.5096161 | 6.7754054 | 37 |
| TAATT | 15600425 | 1.5032504 | 17.09925 | 23 |
| ACGGT | 1674675 | 1.5017338 | 13.36789 | 6 |
| TTCGG | 4124240 | 1.4993398 | 20.296938 | 35 |
| GGAAT | 8915000 | 1.4928557 | 10.46273 | 2 |
| GGTTT | 54031235 | 1.487061 | 9.094247 | 2 |
| AGGTT | 21422060 | 1.4542915 | 13.288554 | 41 |
| TTATT | 36514170 | 1.4264286 | 6.4196515 | 32 |
| AAGTA | 4547485 | 1.4248633 | 10.492579 | 34 |
| TTTAA | 14770880 | 1.4233156 | 7.9397097 | 5 |
| GCGAT | 1570610 | 1.4084152 | 22.23787 | 10 |
| GTAGT | 20619250 | 1.3997905 | 8.757728 | 36 |
| GGCGA | 1178430 | 1.3930508 | 8.716326 | 2 |
| TGGAA | 8302795 | 1.3903393 | 8.776133 | 1 |
| TTAAG | 10940310 | 1.3897135 | 9.542501 | 6 |
| TTGAG | 20392045 | 1.3843663 | 12.2611065 | 44 |
| TATAG | 10871930 | 1.3810275 | 15.536783 | 47 |
| AGATA | 4399265 | 1.3784215 | 5.459907 | 26 |
| ACGAC | 61855 | 1.3709288 | 6.0199814 | 23 |
| TTATA | 14173755 | 1.365777 | 12.052828 | 46 |
| GTTTA | 26434405 | 1.3613158 | 7.6605024 | 4 |
| GACGT | 1509240 | 1.3533831 | 6.1541524 | 3 |
| TCGGG | 2737045 | 1.3117131 | 25.528843 | 36 |
| GGTTA | 19249115 | 1.3067753 | 16.94351 | 2 |
| GGTAG | 14469610 | 1.2949345 | 7.266741 | 2 |
| TCGTG | 3556965 | 1.2931107 | 7.2033567 | 40 |
| GGAGT | 14367260 | 1.285775 | 10.052305 | 2 |
| GTTAA | 10095345 | 1.2823803 | 20.916756 | 3 |
| GGGTT | 35334115 | 1.2819732 | 13.805705 | 2 |
| TTGTA | 24568995 | 1.2652513 | 13.792803 | 20 |
| TAAGC | 752065 | 1.2618952 | 37.171265 | 7 |
| GAGTA | 7489825 | 1.2542039 | 14.365621 | 34 |
| AAAAC | 161720 | 1.2524009 | 14.124469 | 6 |
| ATTAT | 12941425 | 1.2470301 | 11.918967 | 45 |
| GGGGA | 10345520 | 1.2205178 | 10.250622 | 2 |
| TCGAT | 1781470 | 1.2118256 | 7.3154535 | 11 |
| CGTAA | 714080 | 1.1981597 | 7.3699846 | 21 |
| TTTGT | 57173090 | 1.1936439 | 6.678857 | 19 |
| GTAAT | 9391625 | 1.1929888 | 21.243431 | 22 |
| GGGAT | 13225750 | 1.1836174 | 10.633817 | 42 |
| GGTGG | 24221750 | 1.1584866 | 10.759412 | 8 |
| TATTC | 2239090 | 1.1553999 | 26.332382 | 33 |
| GATTA | 9078880 | 1.1532619 | 15.09729 | 44 |
| TTTTC | 5381940 | 1.1258833 | 10.259242 | 29 |
| AGTAT | 8794440 | 1.1171302 | 14.8747 | 30 |
| TGAGG | 12470360 | 1.1160148 | 14.679997 | 45 |
| GGGGT | 23317390 | 1.1152326 | 8.075184 | 2 |
| CGTGA | 1231800 | 1.1045938 | 8.490784 | 26 |
| GGATT | 16076770 | 1.0914128 | 8.320286 | 43 |
| GTATT | 21180460 | 1.0907488 | 6.5250435 | 31 |
| TAGGC | 1216240 | 1.0906405 | 7.8441896 | 13 |
| CGATC | 120765 | 1.0851127 | 8.483165 | 44 |
| CGTGT | 2982840 | 1.0843914 | 6.8762293 | 41 |
| AGTAA | 3413875 | 1.0696694 | 6.695455 | 9 |
| TGTAA | 8393565 | 1.0662085 | 20.703653 | 21 |
| GGGTA | 11860285 | 1.0614172 | 14.418836 | 2 |
| TTAAT | 10994935 | 1.0594672 | 14.275443 | 4 |
| AAGGC | 474810 | 1.0502399 | 16.376509 | 46 |
| TGGAG | 11655005 | 1.043046 | 9.343786 | 1 |
| GTTAT | 19690195 | 1.0140033 | 7.5767136 | 31 |
| ATTTC | 1946110 | 1.0042185 | 5.891459 | 22 |
| GTGGC | 2086910 | 1.0001395 | 28.684414 | 9 |
| TGTAG | 14575290 | 0.9894808 | 7.1868305 | 21 |
| TGCGG | 2034805 | 0.97516865 | 6.533266 | 5 |
| TAAGT | 7641245 | 0.9706436 | 6.0562787 | 7 |
| AGTTG | 14273405 | 0.9689866 | 8.73196 | 38 |
| GGTTG | 26500970 | 0.9614938 | 6.4538026 | 42 |
| TTATC | 1863085 | 0.9613764 | 11.048589 | 37 |
| GTTGA | 14024960 | 0.95212024 | 11.665486 | 43 |
| TGGGG | 19870025 | 0.9503507 | 8.222273 | 1 |
| AAGAC | 229320 | 0.94910896 | 6.9659343 | 32 |
| GGAGC | 797935 | 0.94325846 | 8.851328 | 27 |
| TTGGG | 25988500 | 0.9429007 | 6.507926 | 36 |
| GGGGG | 14883910 | 0.93843323 | 5.84855 | 2 |
| TAAGG | 5549820 | 0.92934155 | 5.1988454 | 45 |
| ATTAC | 729755 | 0.9288469 | 5.268732 | 29 |
| GGATA | 5492525 | 0.9197473 | 7.339028 | 2 |
| TTTGG | 33066480 | 0.91006386 | 5.2723346 | 35 |
| GTTTG | 33010200 | 0.90851486 | 6.6433578 | 18 |
| GGAAC | 408330 | 0.90319175 | 9.221982 | 27 |
| TAGAC | 538010 | 0.9027307 | 9.761506 | 25 |
| TGGTT | 32663145 | 0.89896315 | 7.0166154 | 1 |
| GTGGT | 24549610 | 0.8906955 | 7.772997 | 9 |
| AGTGA | 5230765 | 0.87591445 | 5.245468 | 18 |
| GGGTG | 18258780 | 0.8732876 | 8.215834 | 2 |
| GGTAC | 952970 | 0.8545581 | 13.815373 | 3 |
| GTGCG | 1726650 | 0.8274871 | 6.114503 | 4 |
| GAAGC | 368010 | 0.8140073 | 7.6075315 | 4 |
| GGTAT | 11953930 | 0.8115231 | 5.8684883 | 2 |
| GGTAA | 4618395 | 0.7733704 | 6.104903 | 2 |

| | | | | |
|---|---|---|---|---|
| TGGGT | 21103580 | 0.7656685 | 8.629322 | 1 |
| TGGTG | 20151685 | 0.73113245 | 5.543472 | 7 |
| GAGTC | 777865 | 0.69753605 | 12.000475 | 21 |
| AACTC | 39530 | 0.66460884 | 19.743896 | 22 |
| TGGTA | 9659960 | 0.6557911 | 5.0151124 | 1 |
| CGTCT | 164195 | 0.5981188 | 5.8315487 | 47 |
| GATTC | 819340 | 0.55734706 | 6.2282443 | 29 |
| ACATC | 30145 | 0.506821 | 19.566685 | 38 |
| CTACA | 28805 | 0.4842918 | 19.385117 | 36 |
| CACGG | 36690 | 0.4345926 | 13.7044935 | 31 |
| GGTGC | 900915 | 0.43175828 | 6.1671543 | 3 |
| TCACG | 35325 | 0.31740662 | 10.283962 | 30 |
| TCCAG | 32215 | 0.28946224 | 10.623625 | 25 |
| ATGCC | 31185 | 0.28020737 | 11.183603 | 47 |
| CAGTC | 30880 | 0.27746686 | 10.45909 | 27 |
| CCAGT | 30010 | 0.2696496 | 10.336585 | 26 |
| GTCAC | 29905 | 0.2687061 | 10.480273 | 29 |
| GCTAC | 29325 | 0.26349464 | 10.338968 | 35 |
| CGGCT | 48495 | 0.23287627 | 5.6112323 | 33 |
| ATCTC | 33480 | 0.2282015 | 8.428948 | 40 |
| CATCT | 28990 | 0.19759741 | 7.9468784 | 39 |

# 5 Overrepresented sequences

Note: FastQC tracks sequences that appear in the first 200,000 reads to the end of the file.

| Sequence | Count | % | Possible Source |
|---|---|---|---|
| CGGGTTTACGTTATTTTTTTTGTTTTAGTTTTTCGAGTAGTTGGGATTATAG | 204085 | 0.2695049814744295 | No Hit |
| CGGGCGCGGTGGTTTACGTTTGTAATTTTAGTATTTTGGGAGGTCGAGGCG | 192369 | 0.254033386977262 | No Hit |
| CGGGTTTACGTTATTTTTTTTGTTTTAGTTTTTTAAGTAGTTGGGATTATAG | 127928 | 0.16893565558498083 | No Hit |
| CGGTTAATTTTTTGTATTTTTTAGTAGAGACGGGGTTTTATCGTGTTAGTTA | 83705 | 0.1105368570660123 | No Hit |
| CGGGATGGTTTCGATTTTTTGATTTCGTGATTCGTTCGTTTCGGTTTTTTA | 82952 | 0.10954248094307212 | No Hit |
| CGGGCGTAGTGGCGGGCGTTTGTAGTTTTAGTTATTTGGGAGGTTGAGGTA | 76218 | 0.10064987959927513 | No Hit |