A comparison of metabolic labeling and statistical methods to study genome-wide dynamics of RNA turnover
Supplementary File 4

This file includes a comparison of decay rates estimates obtained with the 'pulseR workflow' for the nucleotide conversion protocols using a different variant caller. In the manuscript, SNP calling was done with GRAND-SLAM on all samples of each of the SLAM-, TLS-, and TUC-seq protocols (Methods).

To determine the influence of using a different variant caller on the pulseR estimates, we used BCFtools [1, 2] to find SNPs on all samples of each of the SLAM-, TLS-, and TUC-seq protocols: bcftools mpileup -Ou -d 800 -f fasta.fa -b bamlist | bcftools call -mv -Ob -o call.bcf, followed by bcftools view -m2 -M2 --types snps -e'QUAL<40' call.bcf -Ov -o snps.vcf. The resulting file 'snps.vcf' was then used in the 'pulseR workflow' by modifying the configuration file, specifying 'vcf: True', and running the main script with --trim5p 5 --trim3p 5, and all other default parameters, as described in the Methods.

We compared the estimates using correlation on a common set of genes for which estimates were available for all methods. Details are given in the manuscript (Methods). The pulseR estimates were not significantly affected, whether GRAND-SLAM or BCFtools were used to identify SNPs, when compared to the GRAND-SLAM estimates or pulse BSA, and ERCC results (Fig. 1). When looking at pulseR results only (n=12,350 instead of 11,603 genes since GRAND-SLAM results are not used), there was essentially no difference in estimates between methods (high correlation, small deviation), whether GRAND-SLAM (pulseRa) or BCFtools (pulseRb) were used to identify SNPs (Fig. 2). Although the correlation was lower, and the deviation marginally higher, when compared to pulseR BSA and ERCC results, estimates from pulseRa and pulseRb were systematically consistent and comparable.

We also included in subsequent pages mismatch rates for all first and second reads that were used by featureCounts, per sample, and all mismatches for each position in a read per sample. The first 100 positions correspond to the first read, and positions 100 to 200 to the second read. Genomic bases are on the top, and the color encodes bases in the reads. Mismatch rates were comparable between protocols, whether they were calculated with the pulseR workflow, using GRAND-SLAM or BCFtools to identify SNPs, as shown here, or with GRAND-SLAM (Supplementary Files 2 and 3). The rate increase with labeling time was clearly distinguishable, and was also consistent along the read length, with artefactual patterns of mismatches at read ends being less pronounced at later time points. In all computational workflows, mismatches at read ends were discarded (Methods).

References

[1] Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, DOI: 10.1093/bioinformatics/btp352 (2009). https://academic.oup.com/bioinformatics/article-pdf/25/16/2078/531810/btp352.pdf.

[2] Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993, DOI: 10.1093/bioinformatics/btr509 (2011). https://academic.oup.com/bioinformatics/article-pdf/27/21/2987/577342/btr509.pdf.

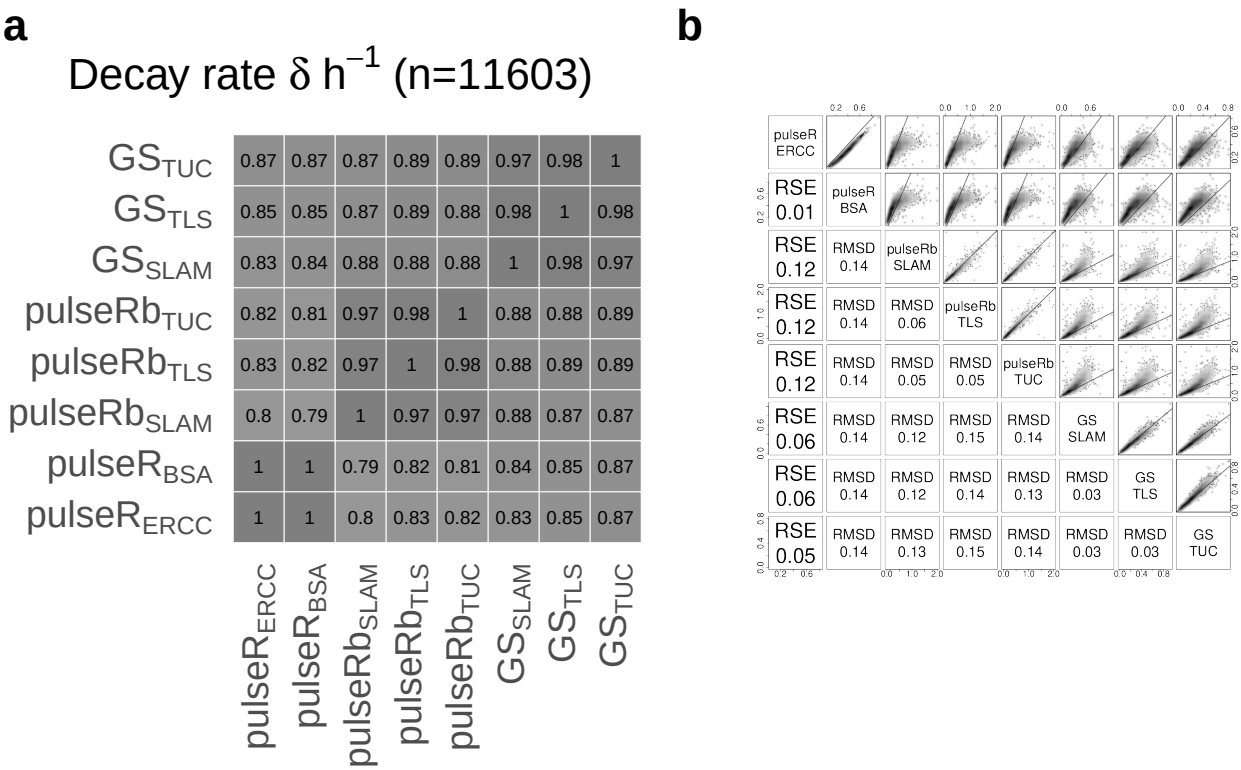


Figure 1: **a** Pearson's correlation coefficient for RNA decay rate estimates δ between any pair of methods for 11,603 common genes. **b** Scatter matrix of decay rate estimates δ between any pair of methods for 11,603 common genes with, on the lower diagonal elements, residual standard error (RSE) from a fitted regression model using the ERCC results, and root-mean-square deviation (RMSD) for all other comparisons. Estimates for GRAND-SLAM, pulseR BSA, and ERCC are those obtained as described in the mansucript. The 'pulseRb' estimates for the nucleotide conversion protocols are those obtained using BCFtools to identify SNPs. All time points (0, 1, 2, 4, and 8 h samples) were used to estimate parameters in pulseR. The maximum a posteriori estimator for δ in GRAND-SLAM was computed on the 1, 2, 4, and 8 h samples.

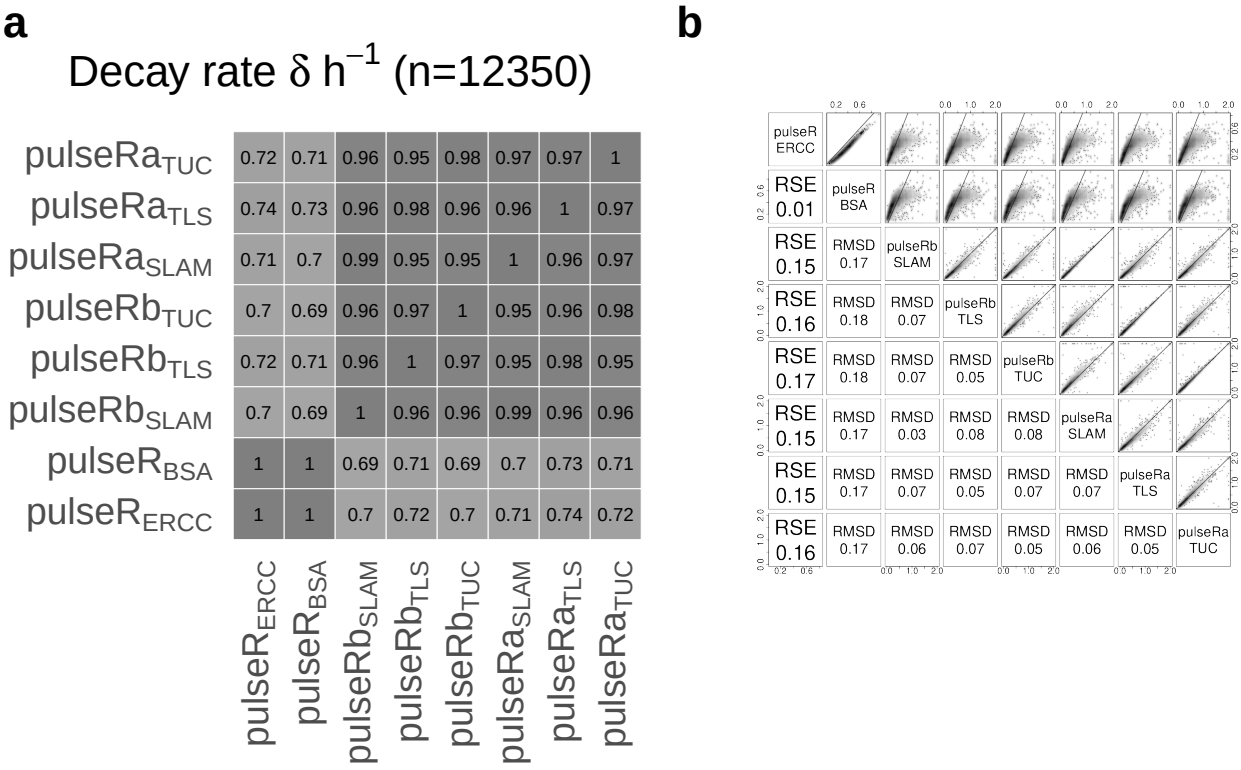
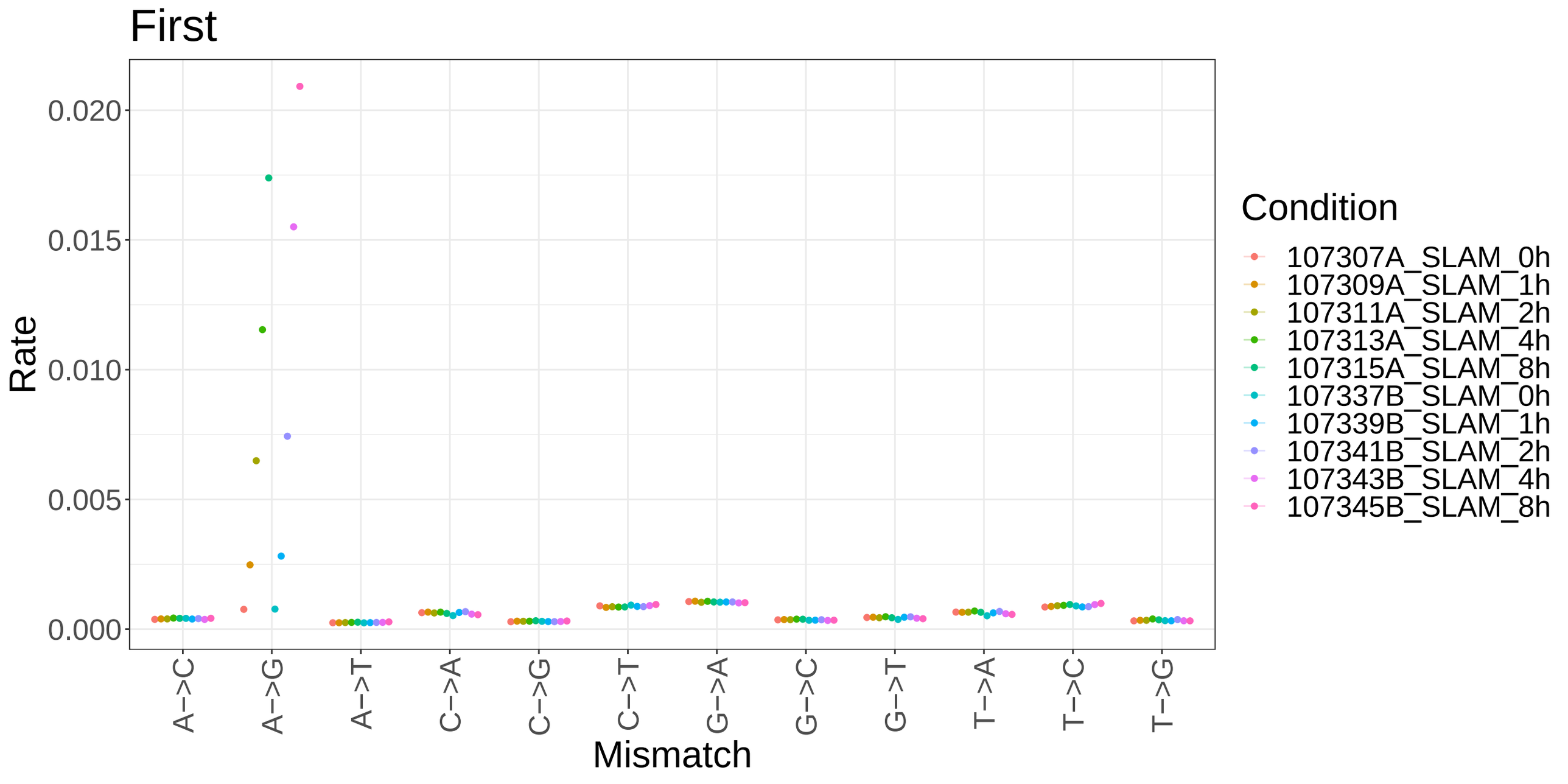
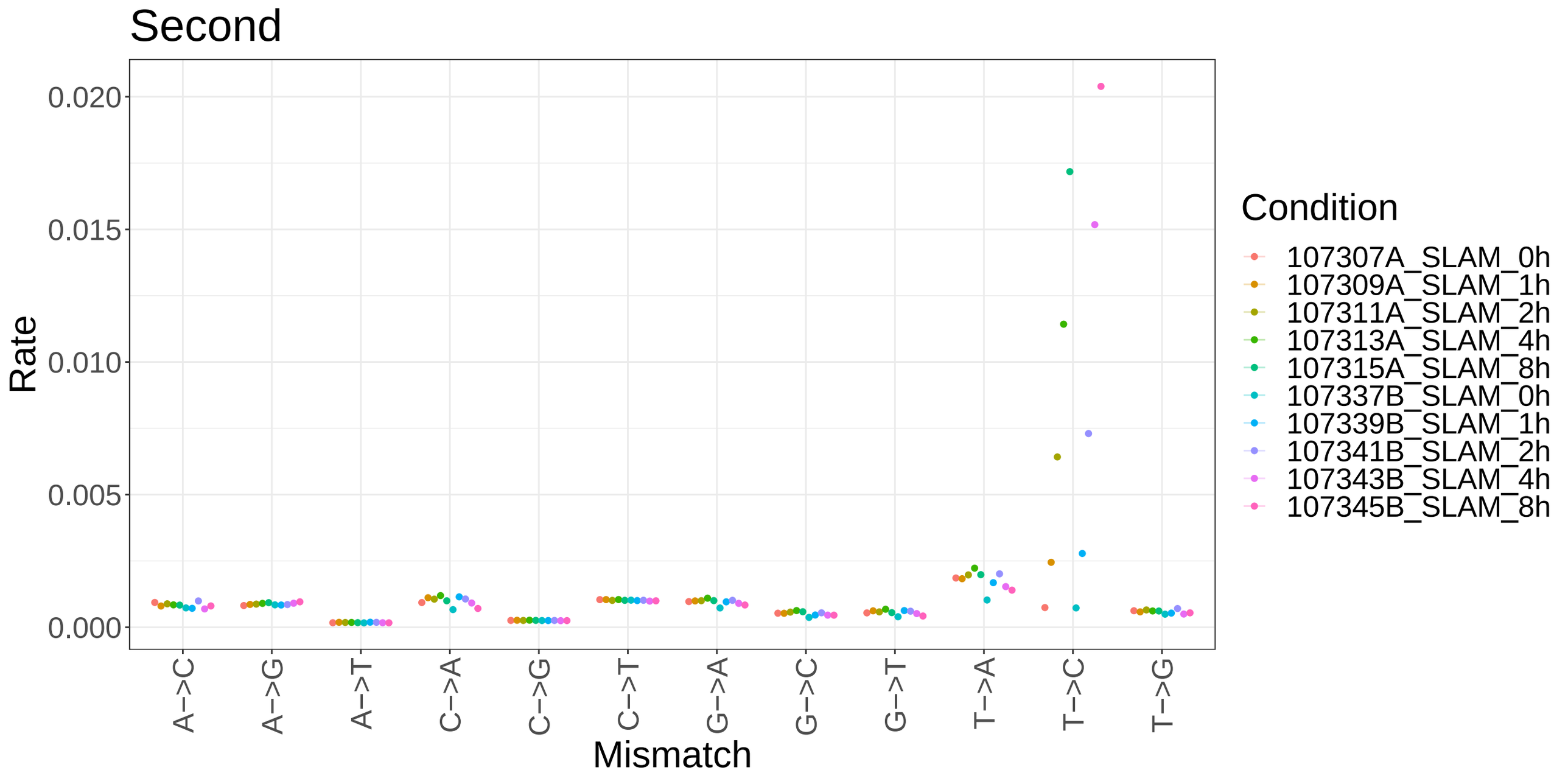
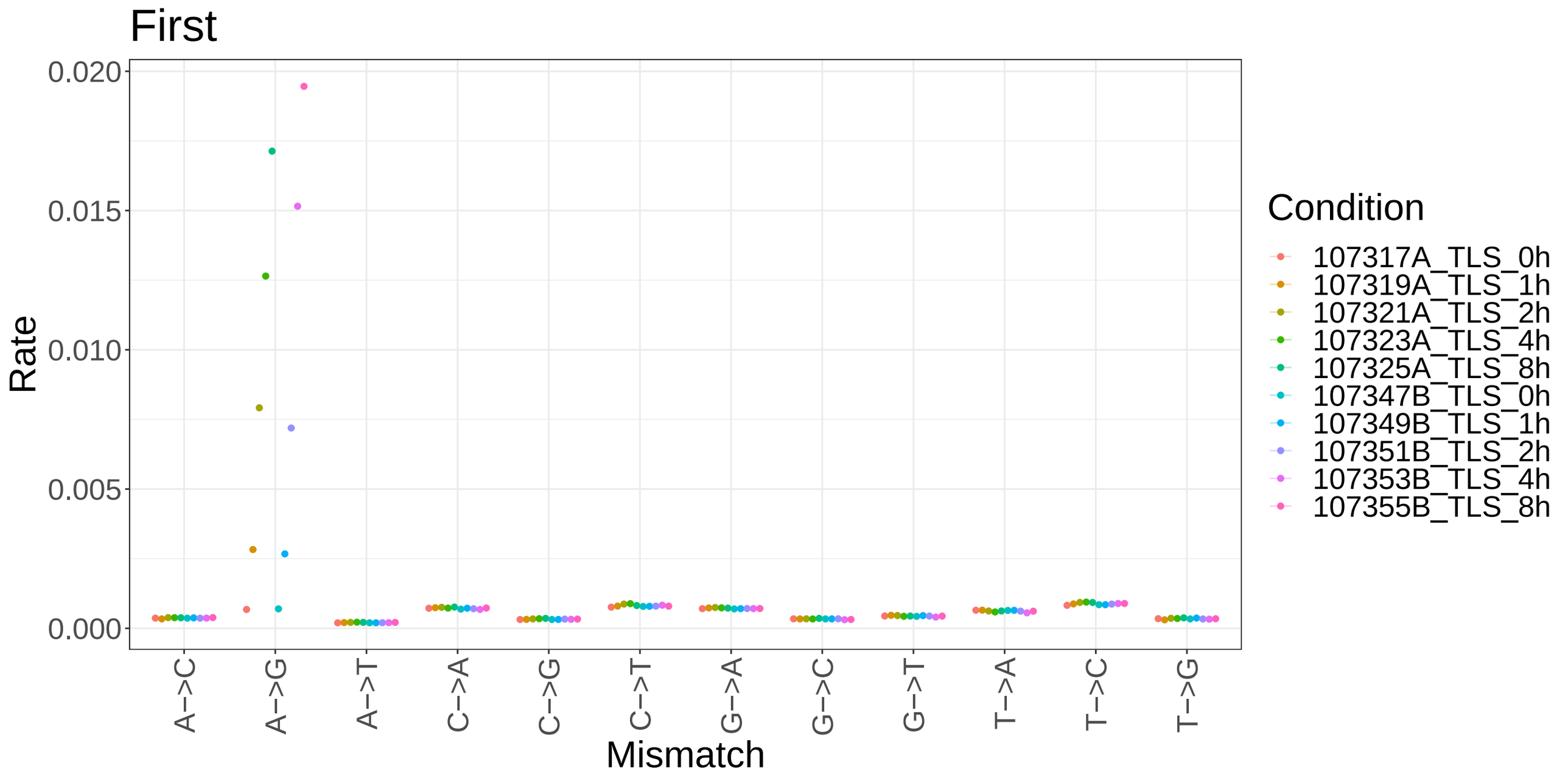
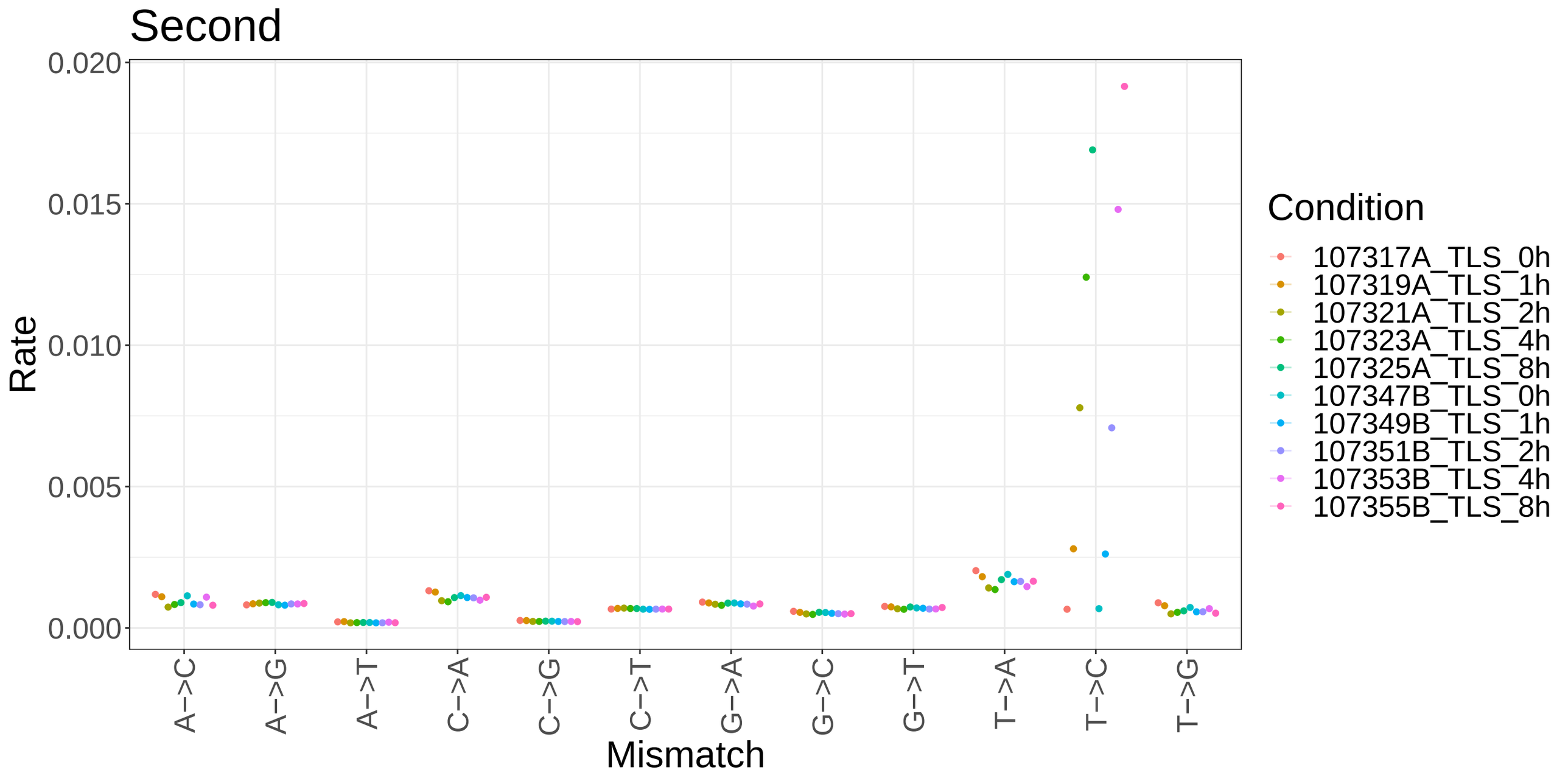


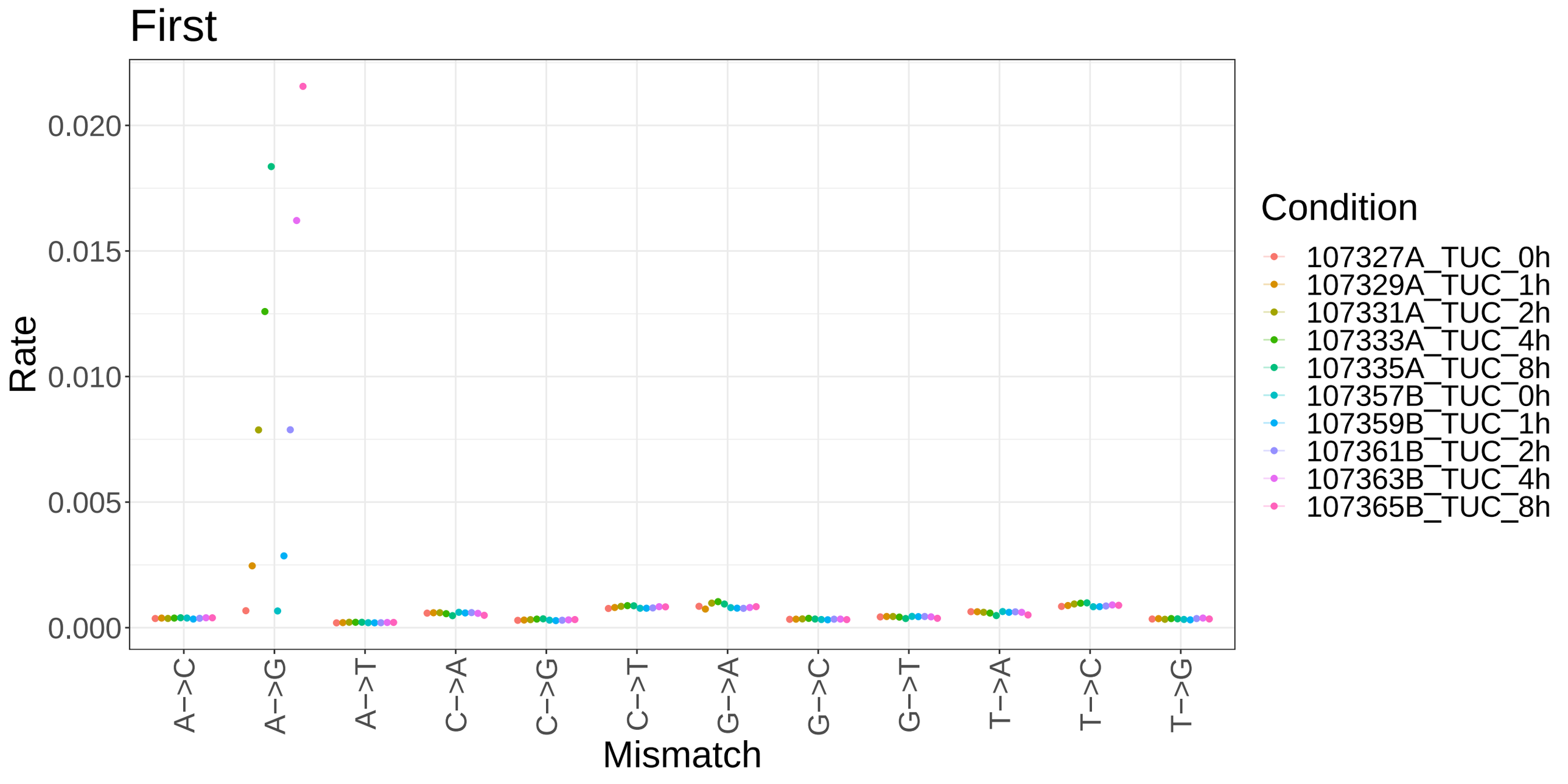
Figure 2: **a** Pearson's correlation coefficient for RNA decay rate estimates δ between any pair of methods (pulseR only) for 12,350 common genes. **b** Scatter matrix of decay rate estimates δ between any pair of methods (pulseR only) for 12,350 common genes with, on the lower diagonal elements, residual standard error (RSE) from a fitted regression model using the ERCC results, and root-mean-square deviation (RMSD) for all other comparisons. The 'pulseRa' estimates for the nucleotide conversion protocols are those obtained using GRAND-SLAM, while the 'pulseRb' are those obtained using BCFtools to identify SNPs. All time points (0, 1, 2, 4, and 8 h samples) were used to estimate parameters.

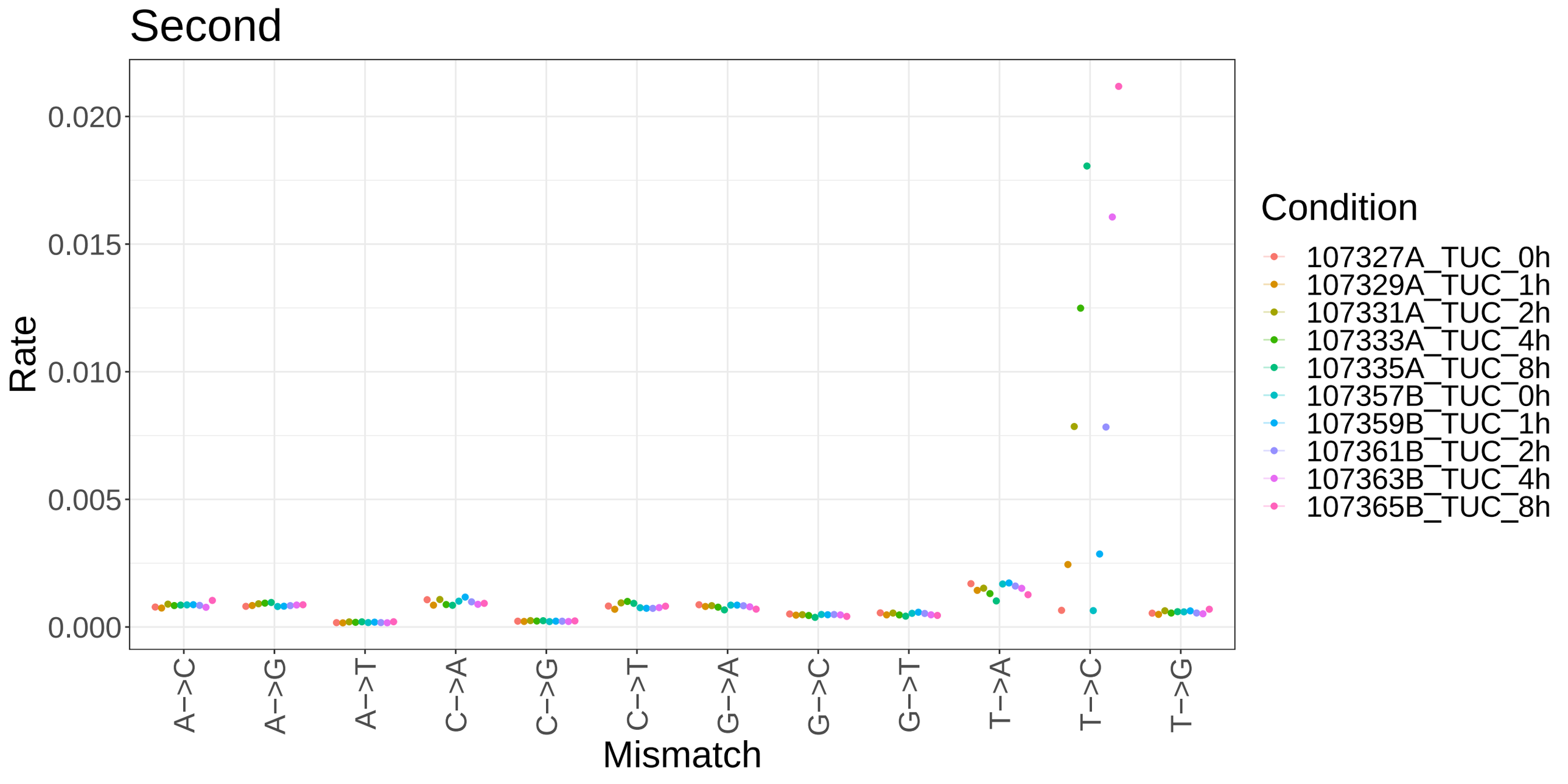

First
Rate
0.020
0.015
0.010
0.005
0.000
A−>C
A−>G
A−>T
C−>A
C−>G
C−>T
G−>A
G−>C
G−>T
T−>A
T−>C
T−>G
Mismatch
Condition
107307A_SLAM_0h
107309A_SLAM_1h
107311A_SLAM_2h
107313A_SLAM_4h
107315A_SLAM_8h
107337B_SLAM_0h
107339B_SLAM_1h
107341B_SLAM_2h
107343B_SLAM_4h
107345B_SLAM_8h

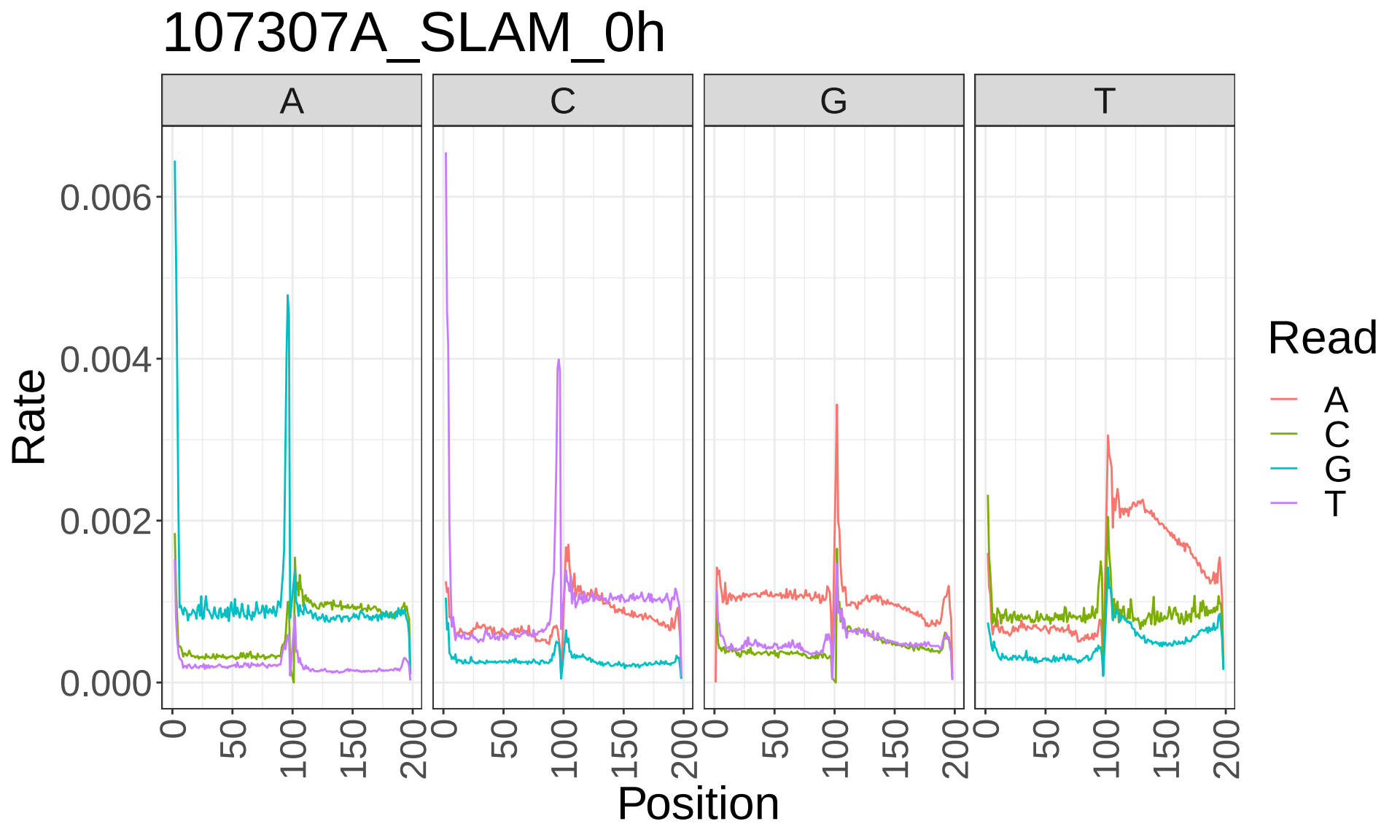

Second
Rate
0.020
0.015
0.010
0.005
0.000
A−>C
A−>G
A−>T
C−>A
C−>G
C−>T
G−>A
G−>C
G−>T
T−>A
T−>C
T−>G
Mismatch
Condition
107307A_SLAM_0h
107309A_SLAM_1h
107311A_SLAM_2h
107313A_SLAM_4h
107315A_SLAM_8h
107337B_SLAM_0h
107339B_SLAM_1h
107341B_SLAM_2h
107343B_SLAM_4h
107345B_SLAM_8h

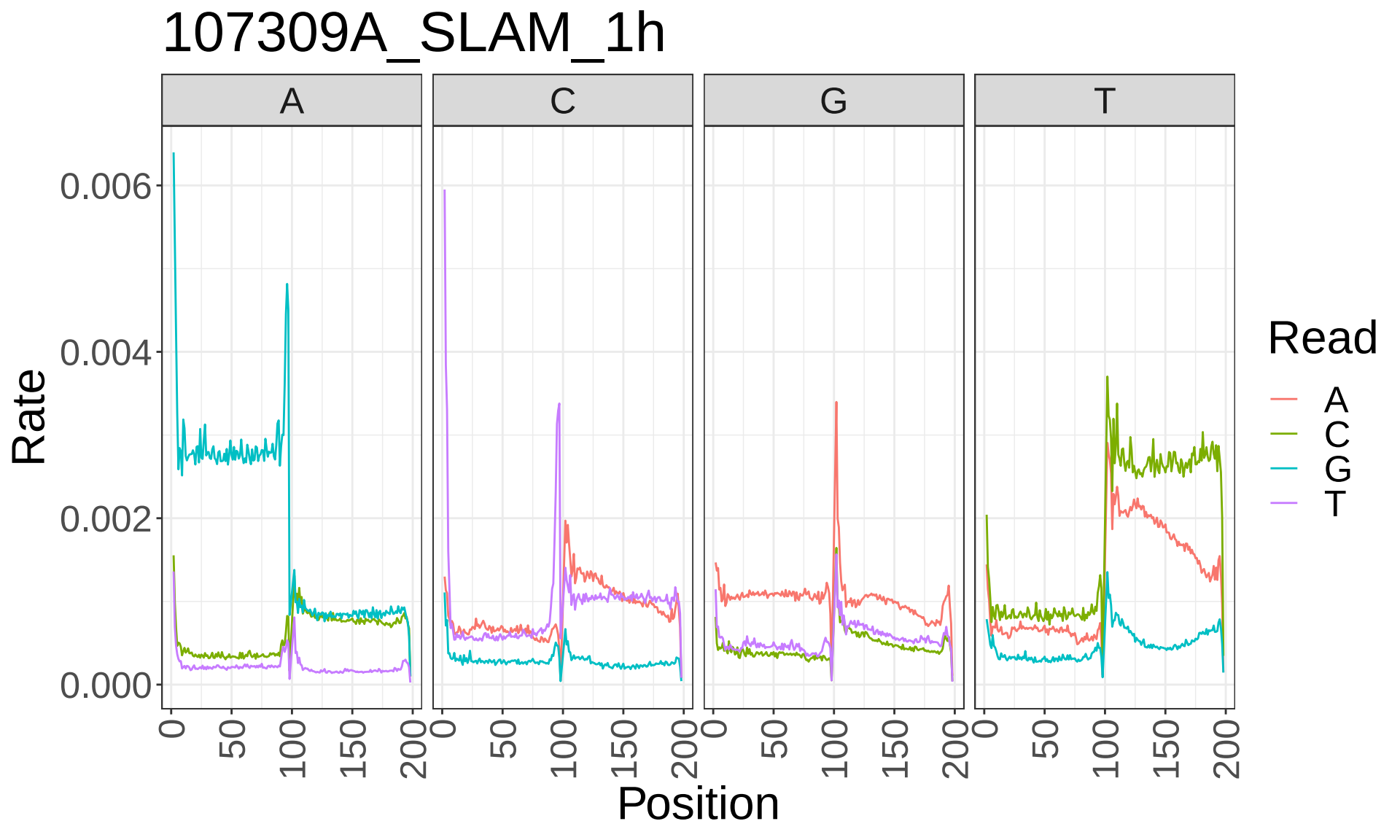

First
0.020
0.015
0.010
0.005
0.000
Rate
A−>C
A−>G
A−>T
C−>A
C−>G
C−>T
G−>A
G−>C
G−>T
T−>A
T−>C
T−>G
Mismatch
Condition
107317A_TLS_0h
107319A_TLS_1h
107321A_TLS_2h
107323A_TLS_4h
107325A_TLS_8h
107347B_TLS_0h
107349B_TLS_1h
107351B_TLS_2h
107353B_TLS_4h
107355B_TLS_8h

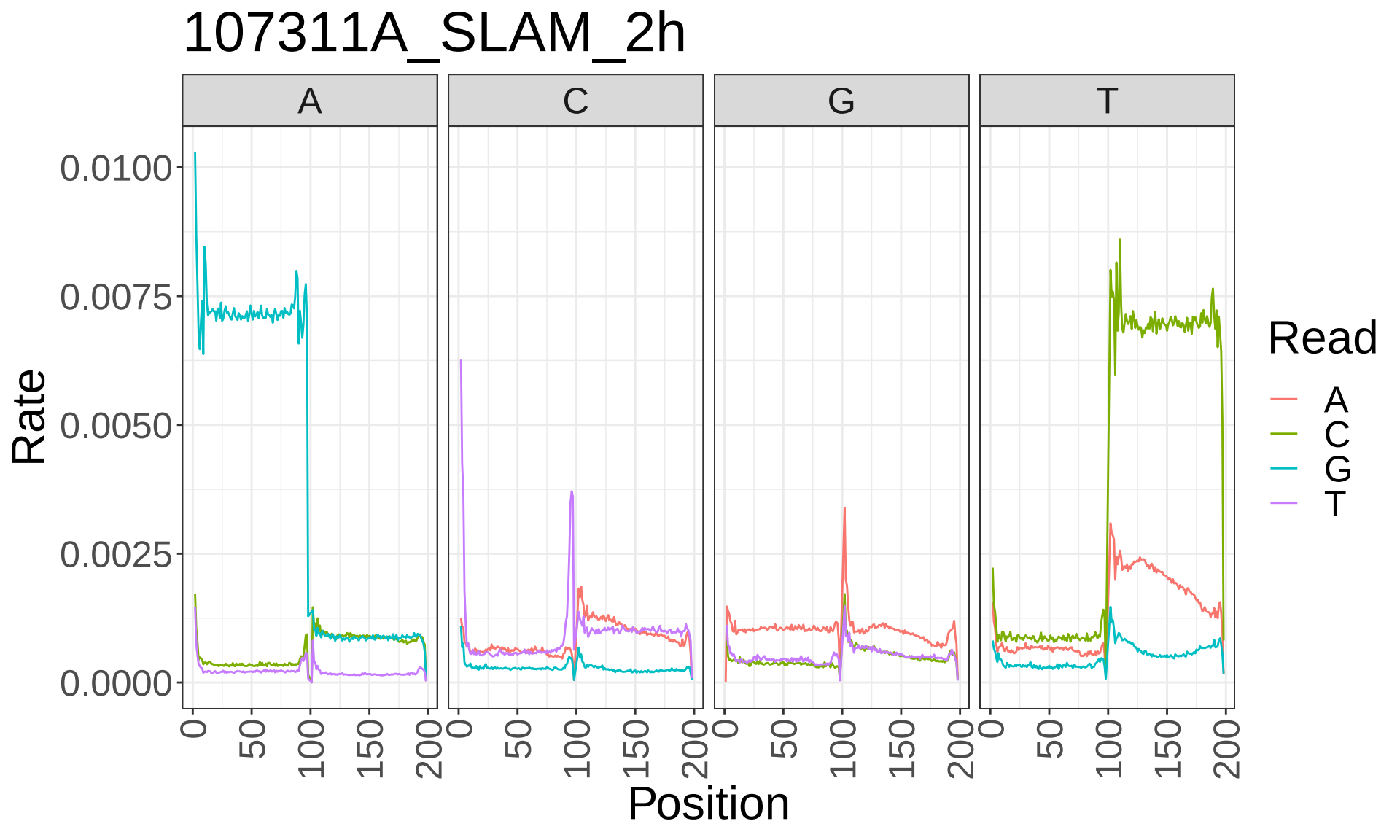

Second
0.020
0.015
0.010
0.005
0.000
Rate
A−>C
A−>G
A−>T
C−>A
C−>G
C−>T
G−>A
G−>C
G−>T
T−>A
T−>C
T−>G
Mismatch
Condition
107317A_TLS_0h
107319A_TLS_1h
107321A_TLS_2h
107323A_TLS_4h
107325A_TLS_8h
107347B_TLS_0h
107349B_TLS_1h
107351B_TLS_2h
107353B_TLS_4h
107355B_TLS_8h

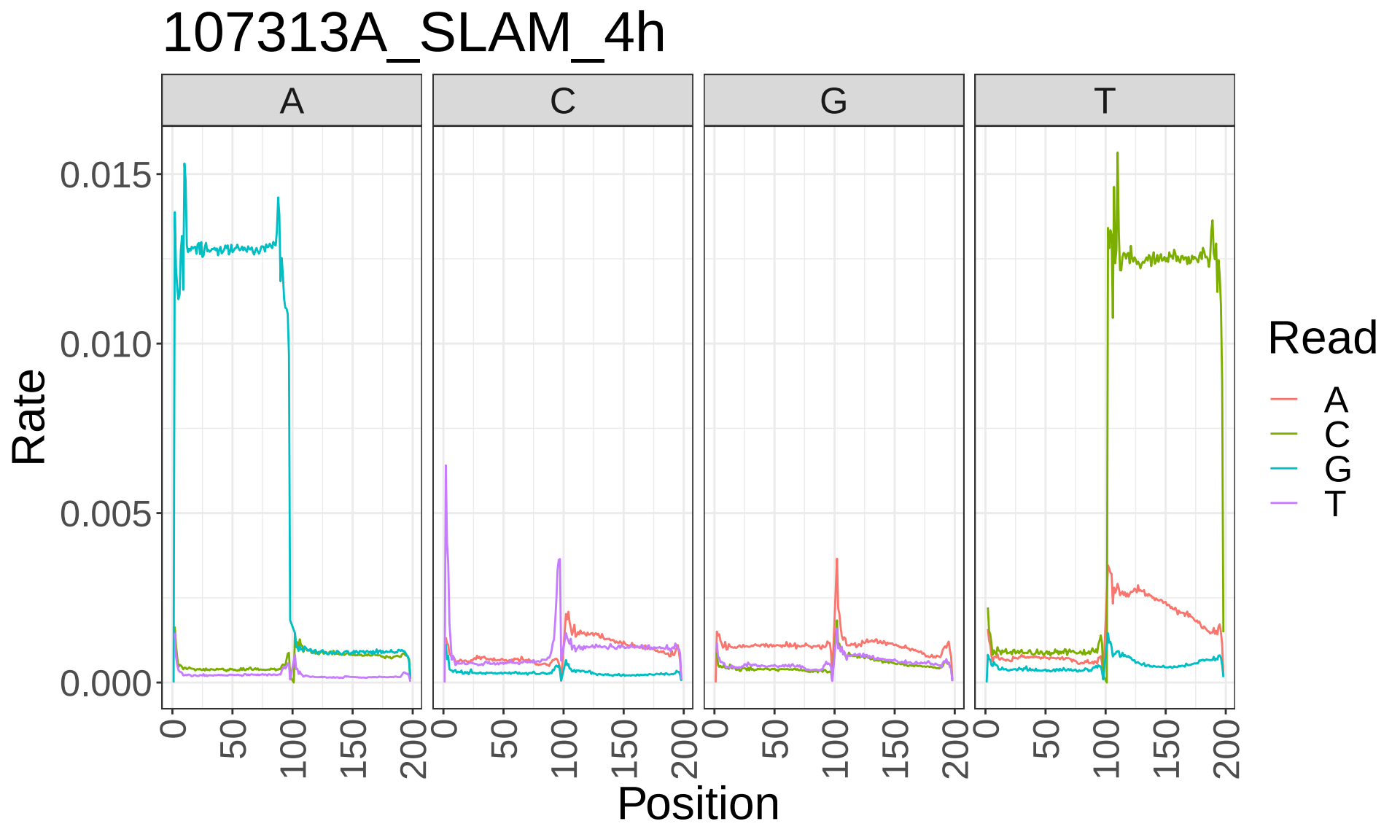

First
Rate
0.020
0.015
0.010
0.005
0.000
A−>C
A−>G
A−>T
C−>A
C−>G
C−>T
G−>A
G−>C
G−>T
T−>A
T−>C
T−>G
Mismatch
Condition
107327A_TUC_0h
107329A_TUC_1h
107331A_TUC_2h
107333A_TUC_4h
107335A_TUC_8h
107357B_TUC_0h
107359B_TUC_1h
107361B_TUC_2h
107363B_TUC_4h
107365B_TUC_8h

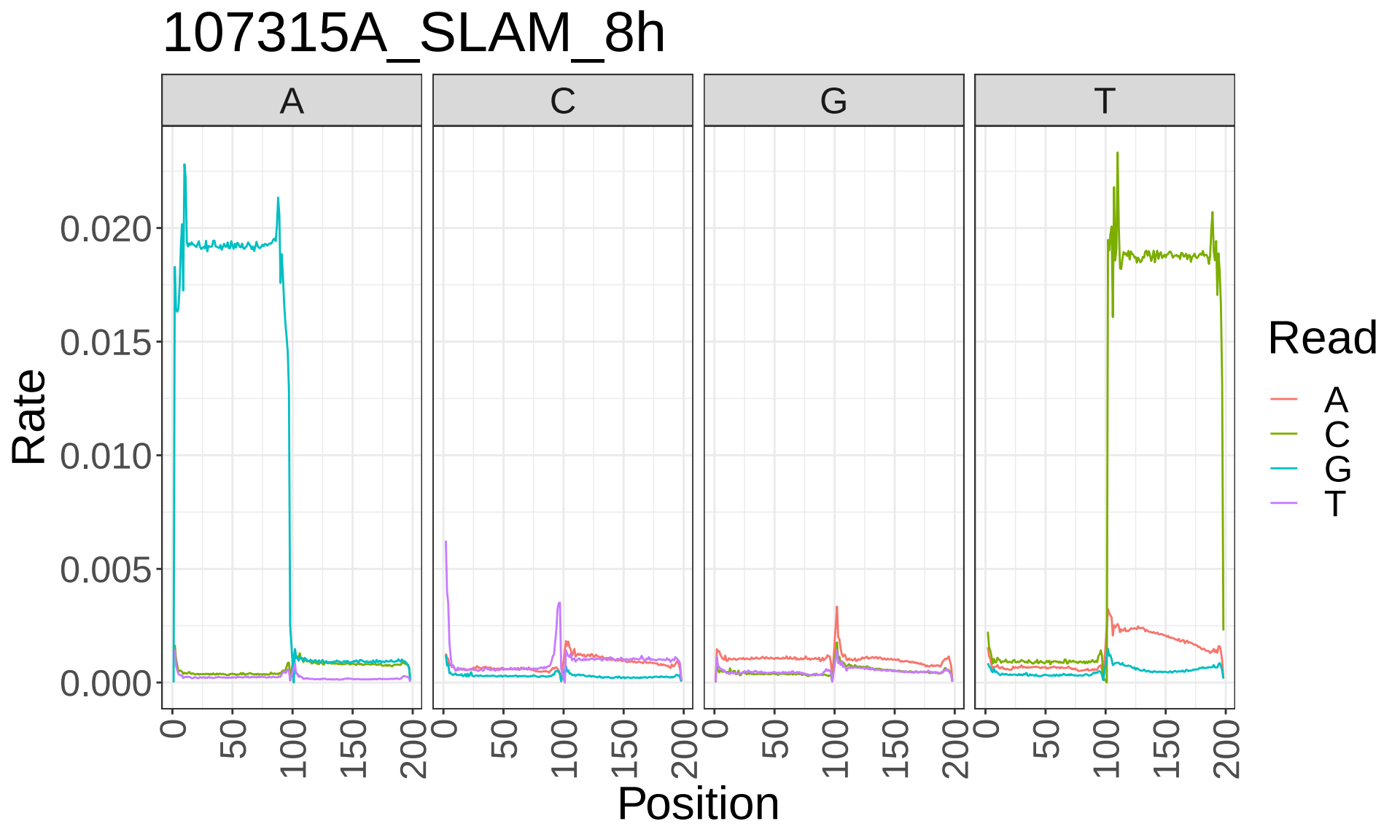

Second
Rate
0.020
0.015
0.010
0.005
0.000
A−>C
A−>G
A−>T
C−>A
C−>G
C−>T
G−>A
G−>C
G−>T
T−>A
T−>C
T−>G
Mismatch
Condition
107327A_TUC_0h
107329A_TUC_1h
107331A_TUC_2h
107333A_TUC_4h
107335A_TUC_8h
107357B_TUC_0h
107359B_TUC_1h
107361B_TUC_2h
107363B_TUC_4h
107365B_TUC_8h

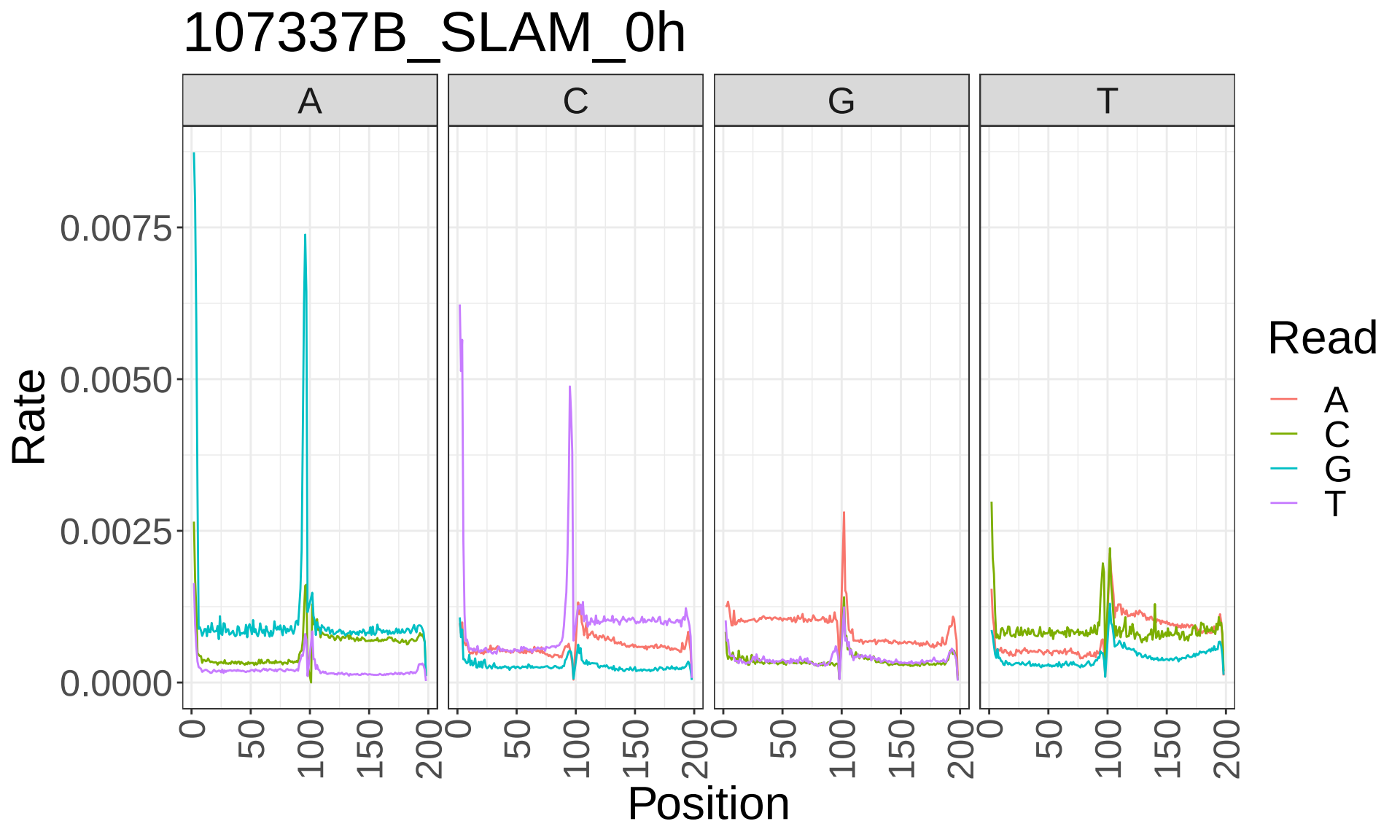

107307A_SLAM_0h
A
C
G
T
Rate
0.006
0.004
0.002
0.000
0
50
100
150
200
Position
Read
A
C
G
T

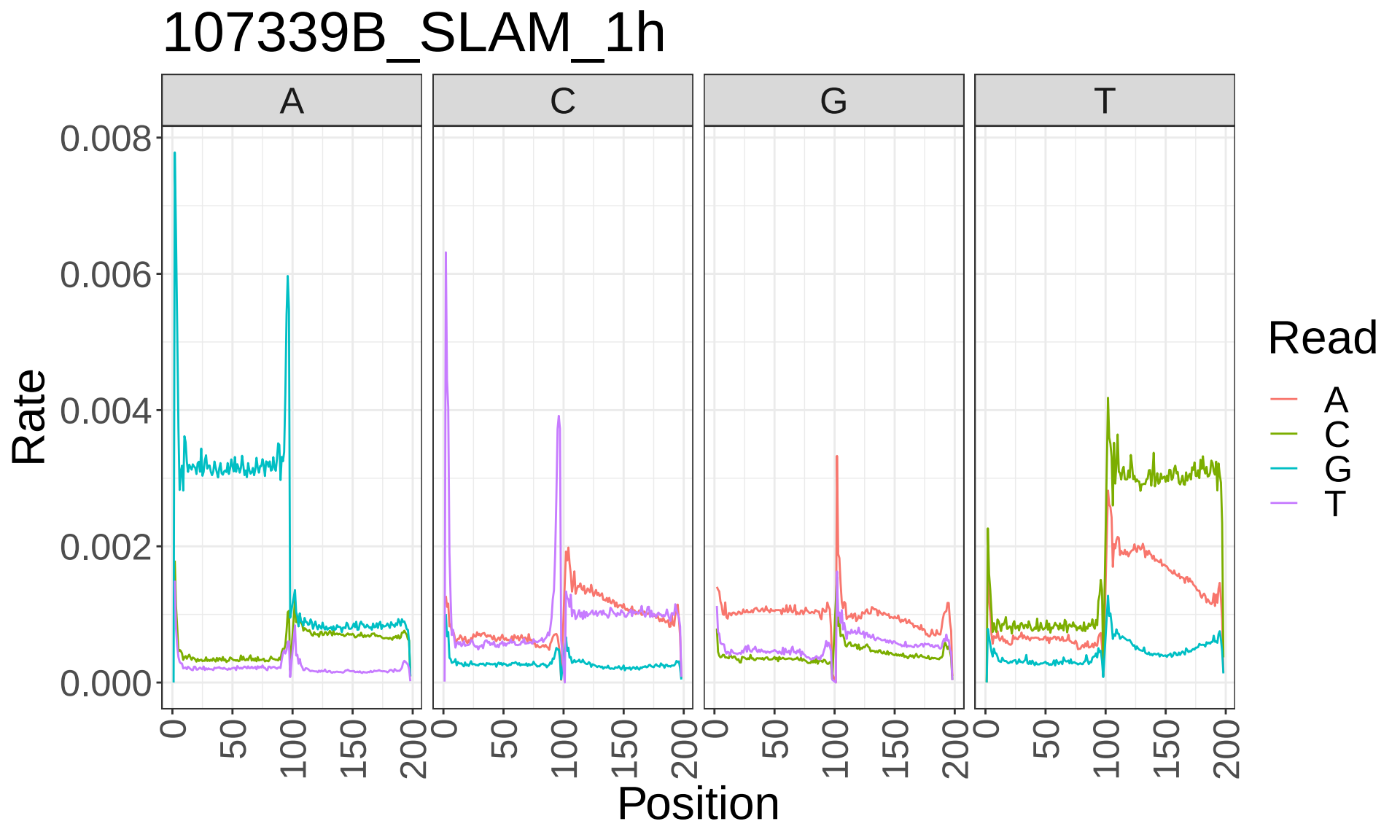

107309A_SLAM_1h
A
C
G
T
Rate
0.006
0.004
0.002
0.000
0
50
100
150
200
Position
Read
A
C
G
T

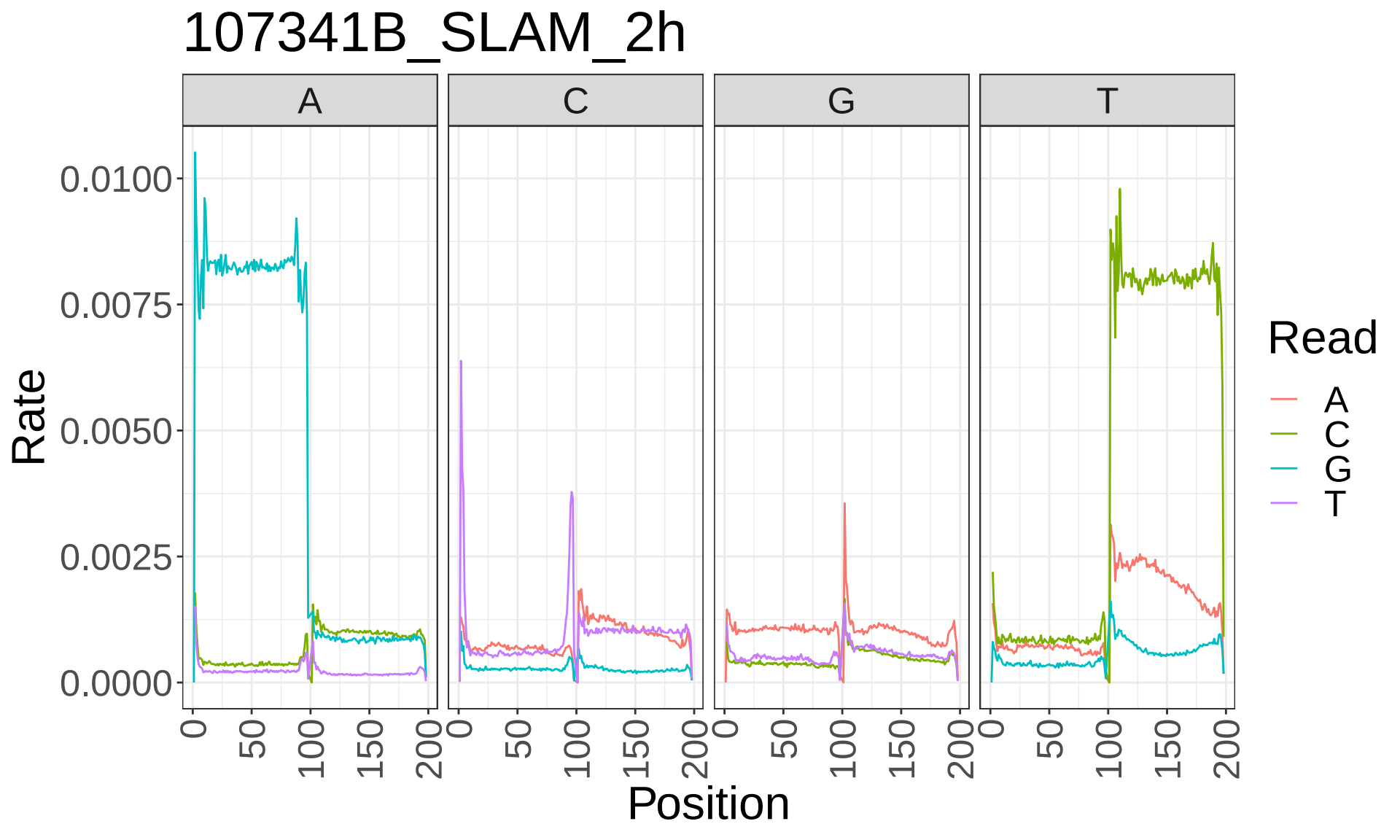

107311A_SLAM_2h
A
C
G
T
Rate
0.0100
0.0075
0.0050
0.0025
0.0000
0
50
100
150
200
Position
Read
A
C
G
T

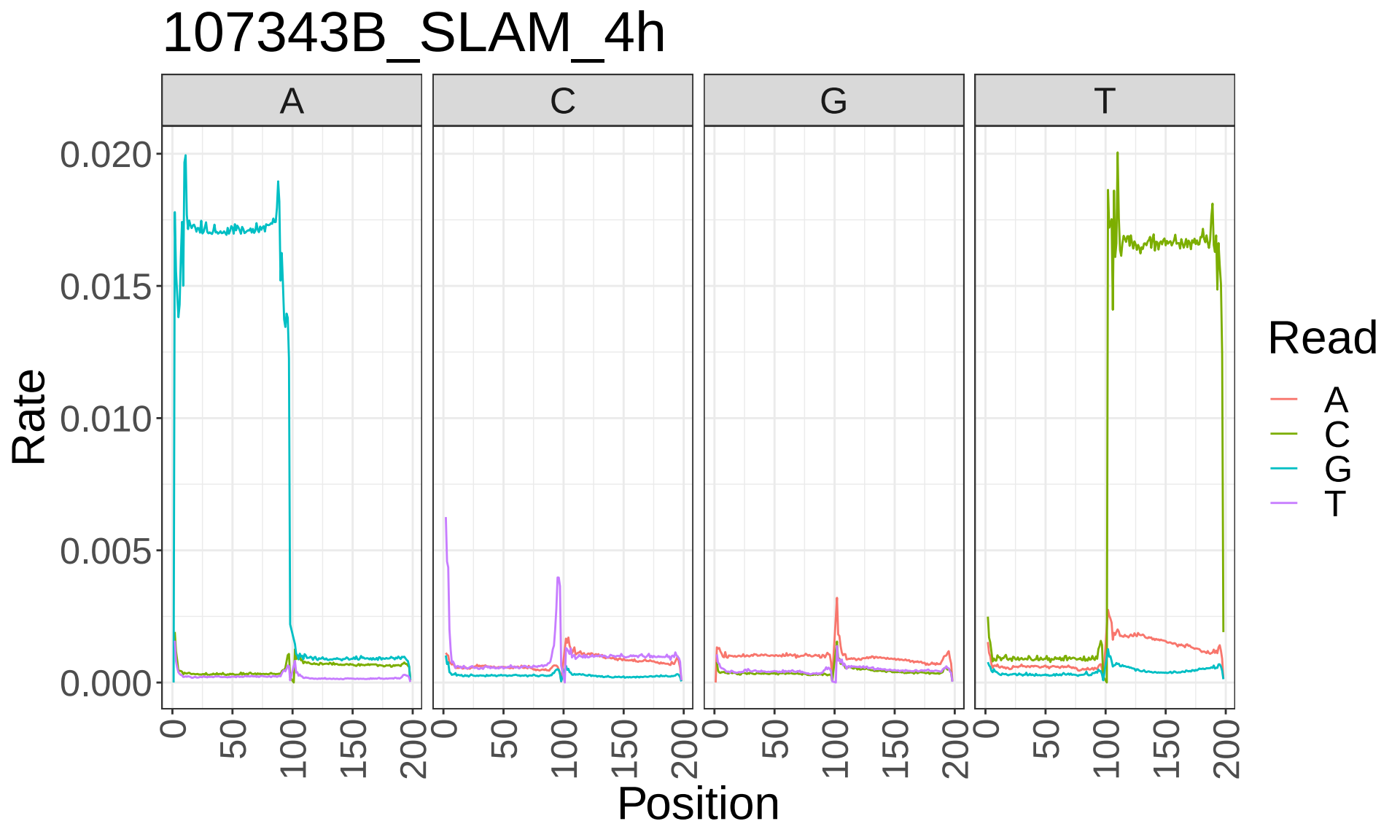

107313A_SLAM_4h
A
C
G
T
Rate
0.015
0.010
0.005
0.000
0
50
100
150
200
Position
Read
A
C
G
T

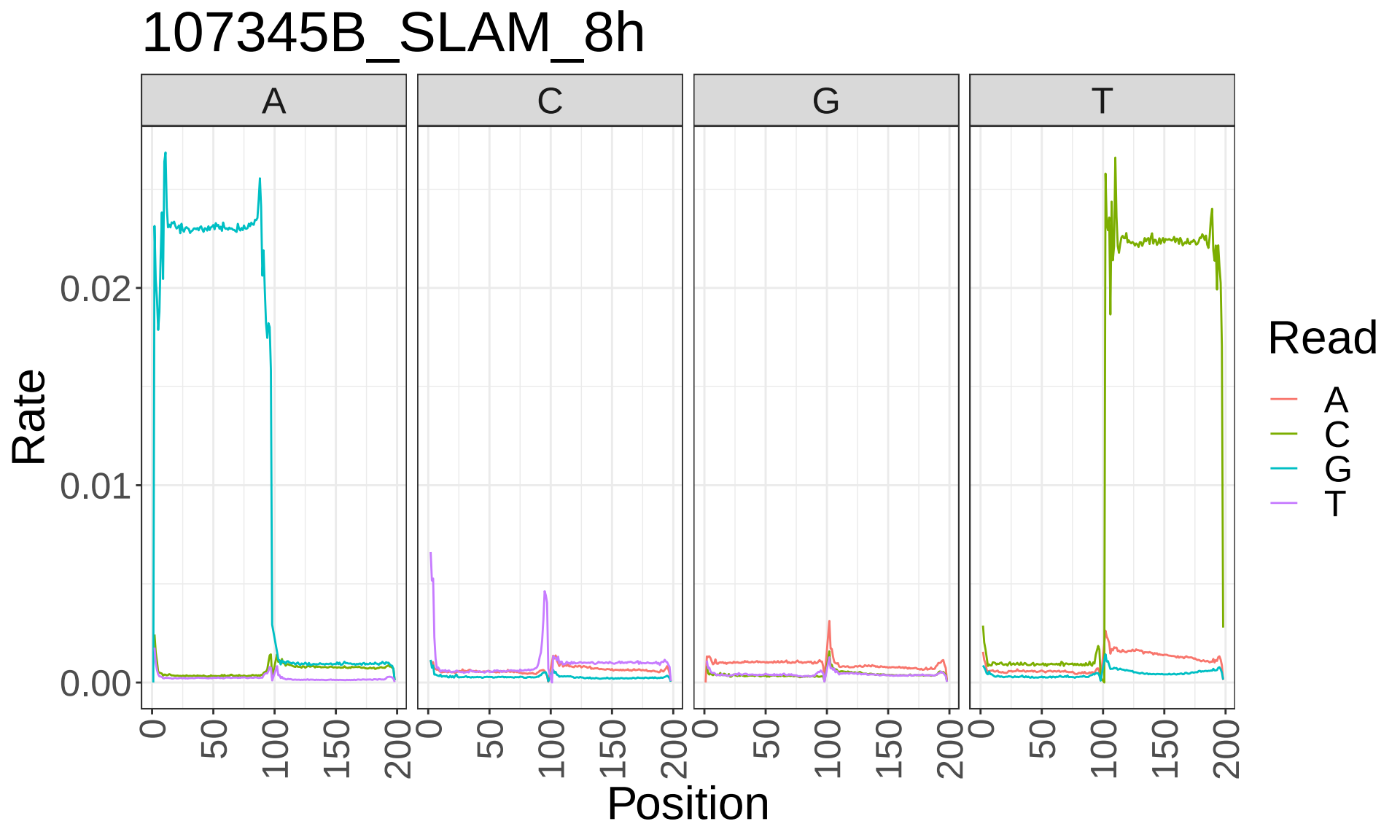

107315A_SLAM_8h
A
C
G
T
Rate
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
Position
Read
A
C
G
T

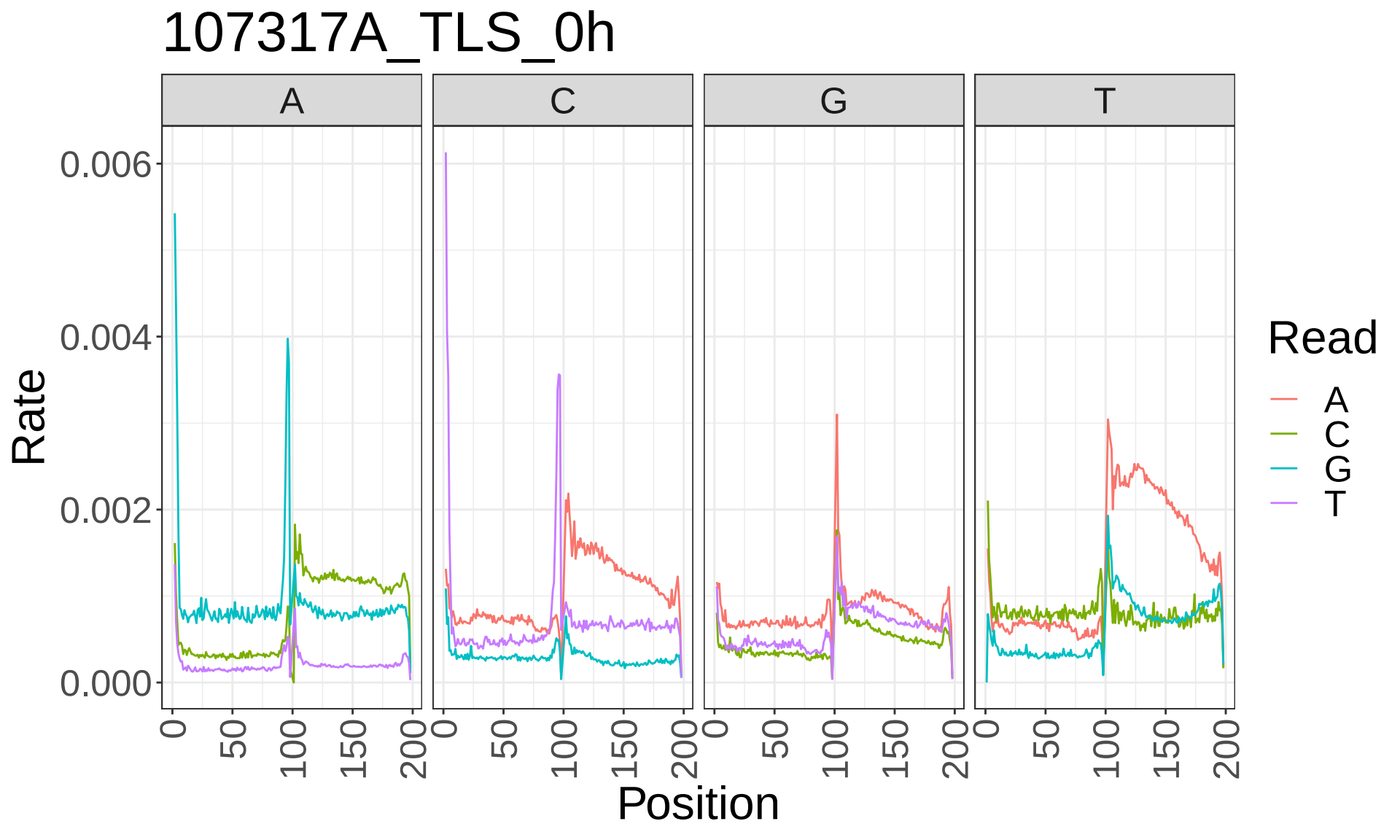

107337B_SLAM_0h
A
C
G
T
Rate
0.0075
0.0050
0.0025
0.0000
0
50
100
150
200
Position
Read
A
C
G
T

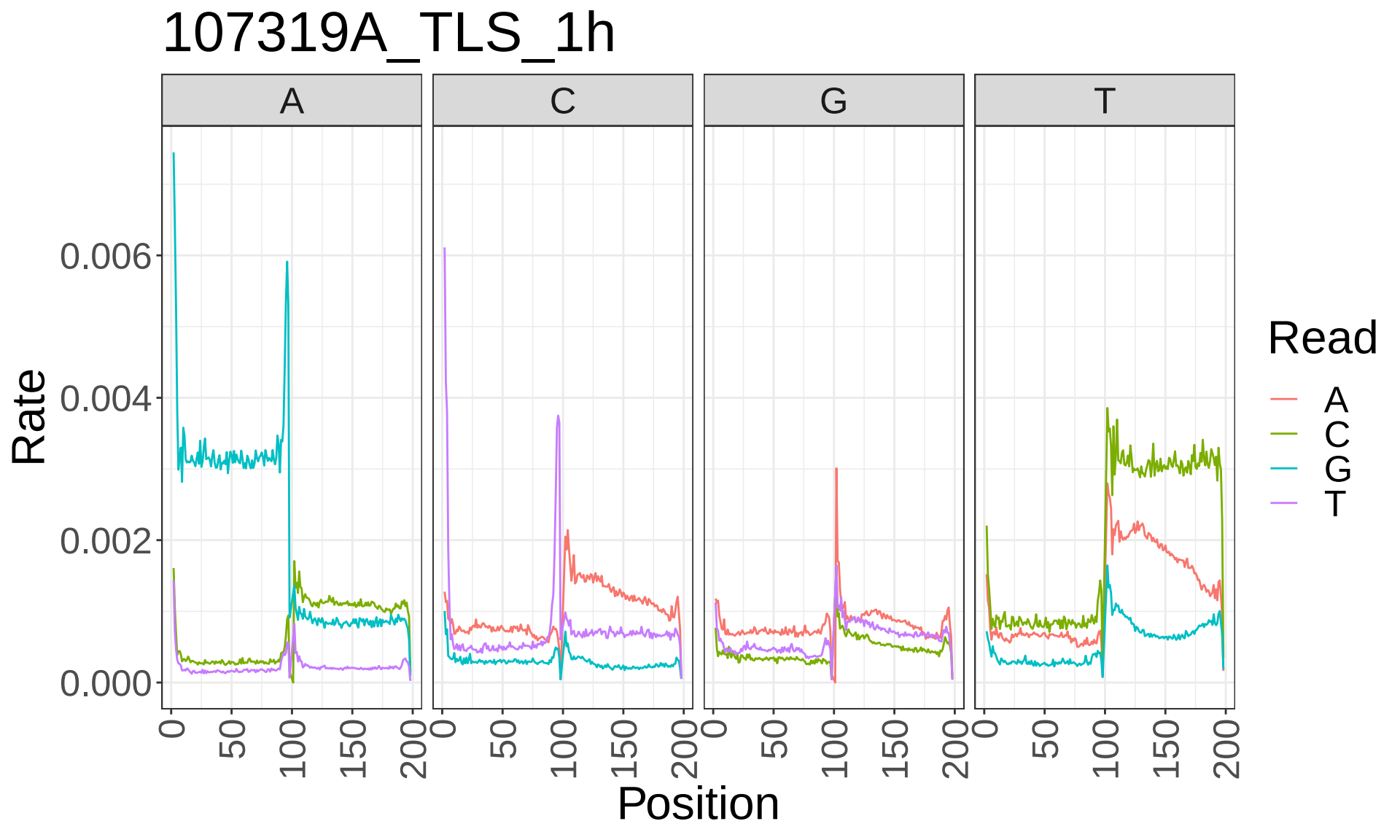

107339B_SLAM_1h
A
C
G
T
Rate
0.008
0.006
0.004
0.002
0.000
0
50
100
150
200
Position
Read
A
C
G
T

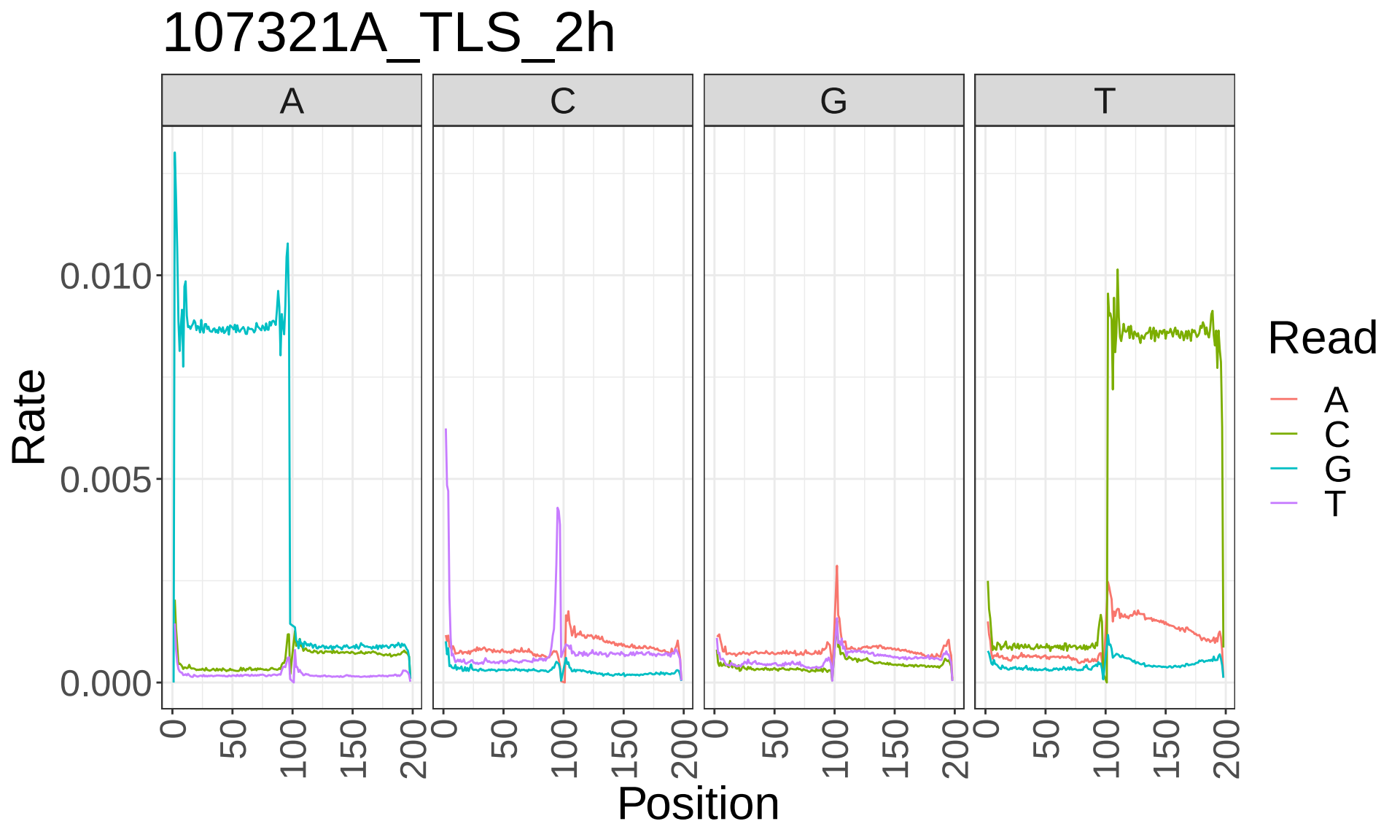

107341B_SLAM_2h
A
C
G
T
Rate
0.0100
0.0075
0.0050
0.0025
0.0000
0
50
100
150
200
Position
Read
A
C
G
T

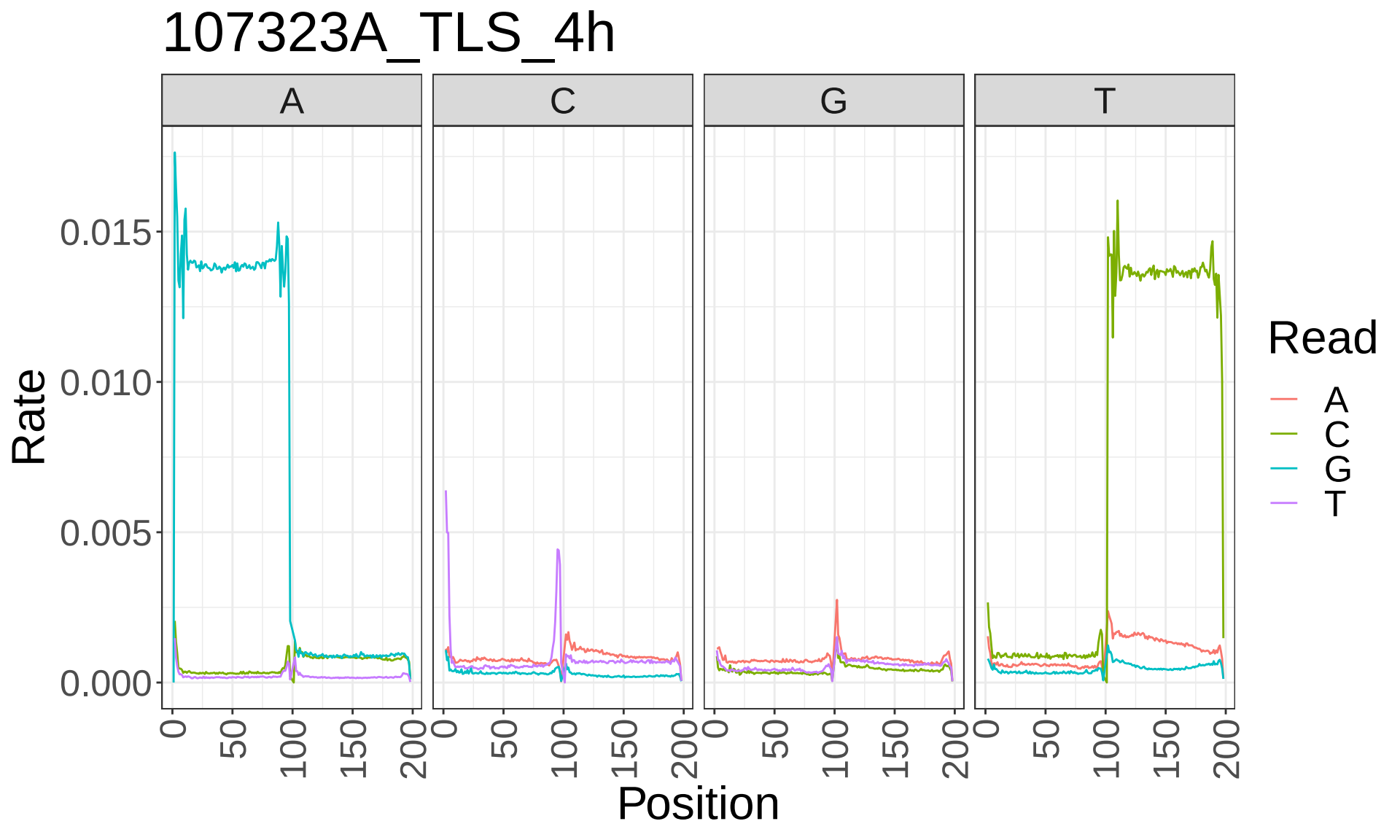

107343B_SLAM_4h
A
C
G
T
Rate
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
Position
Read
A
C
G
T

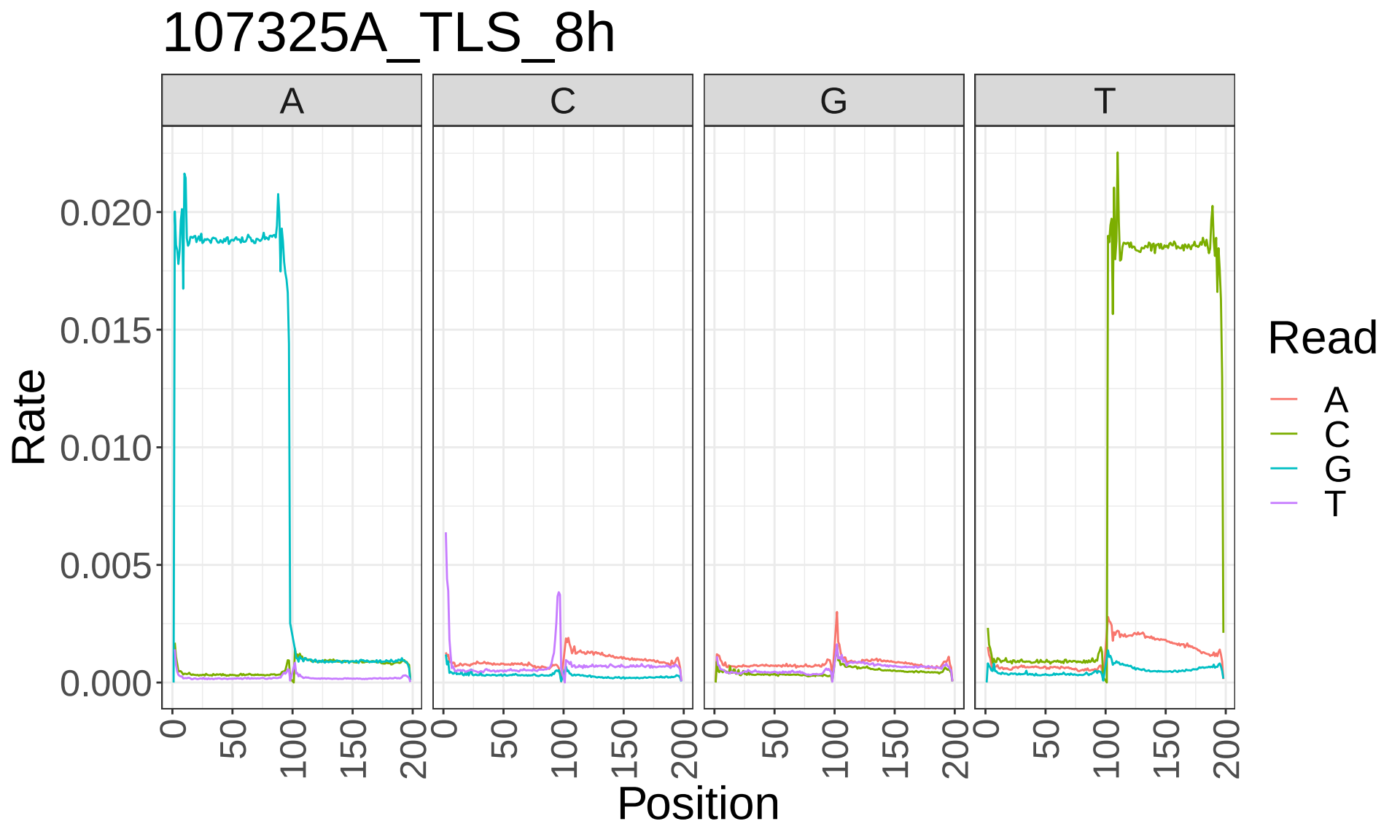

107345B_SLAM_8h
A
C
G
T
Rate
0.02
0.01
0.00
0
50
100
150
200
Position
Read
A
C
G
T

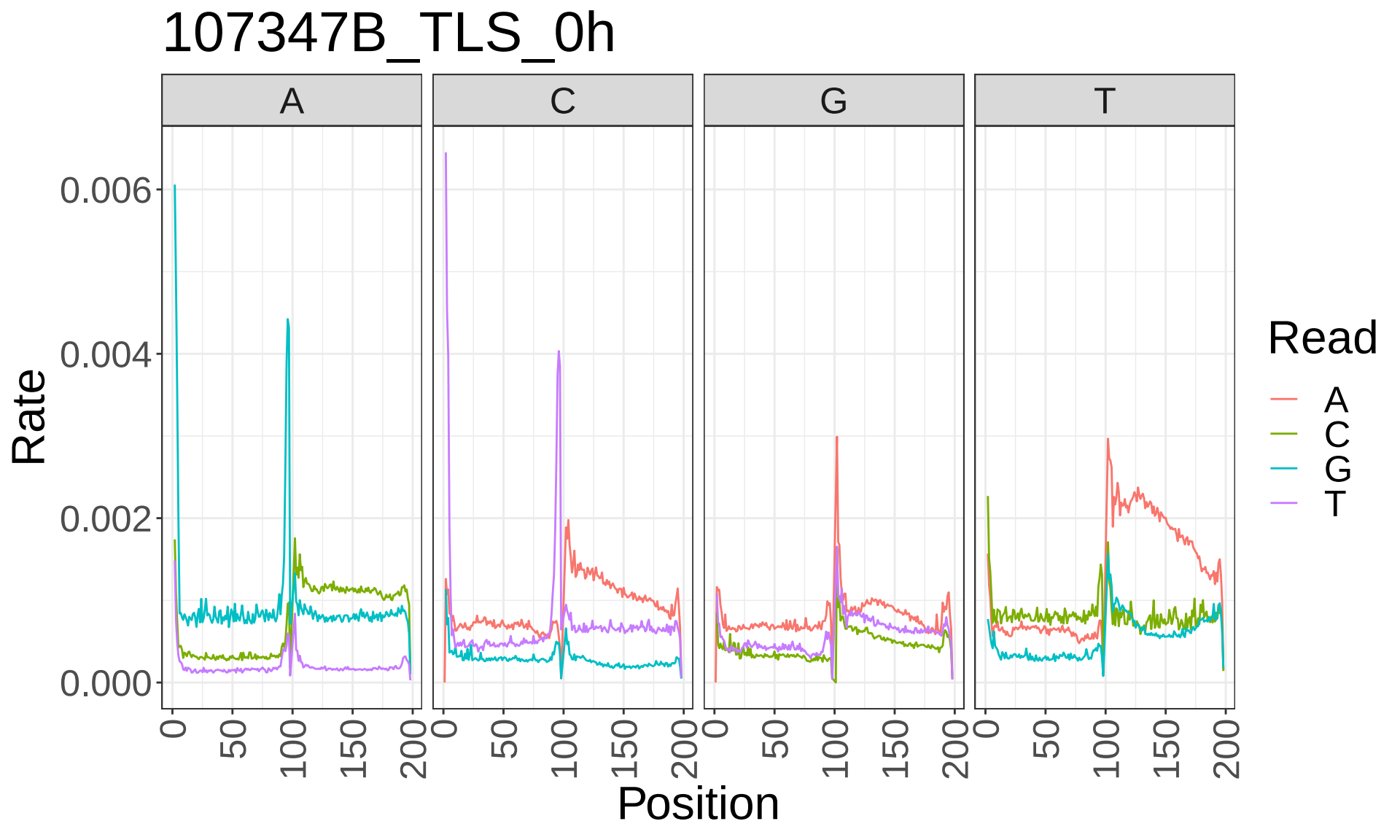

107317A_TLS_0h
A
C
G
T
0.006
0.004
0.002
0.000
Rate
0
50
100
150
200
Position
Read
A
C
G
T

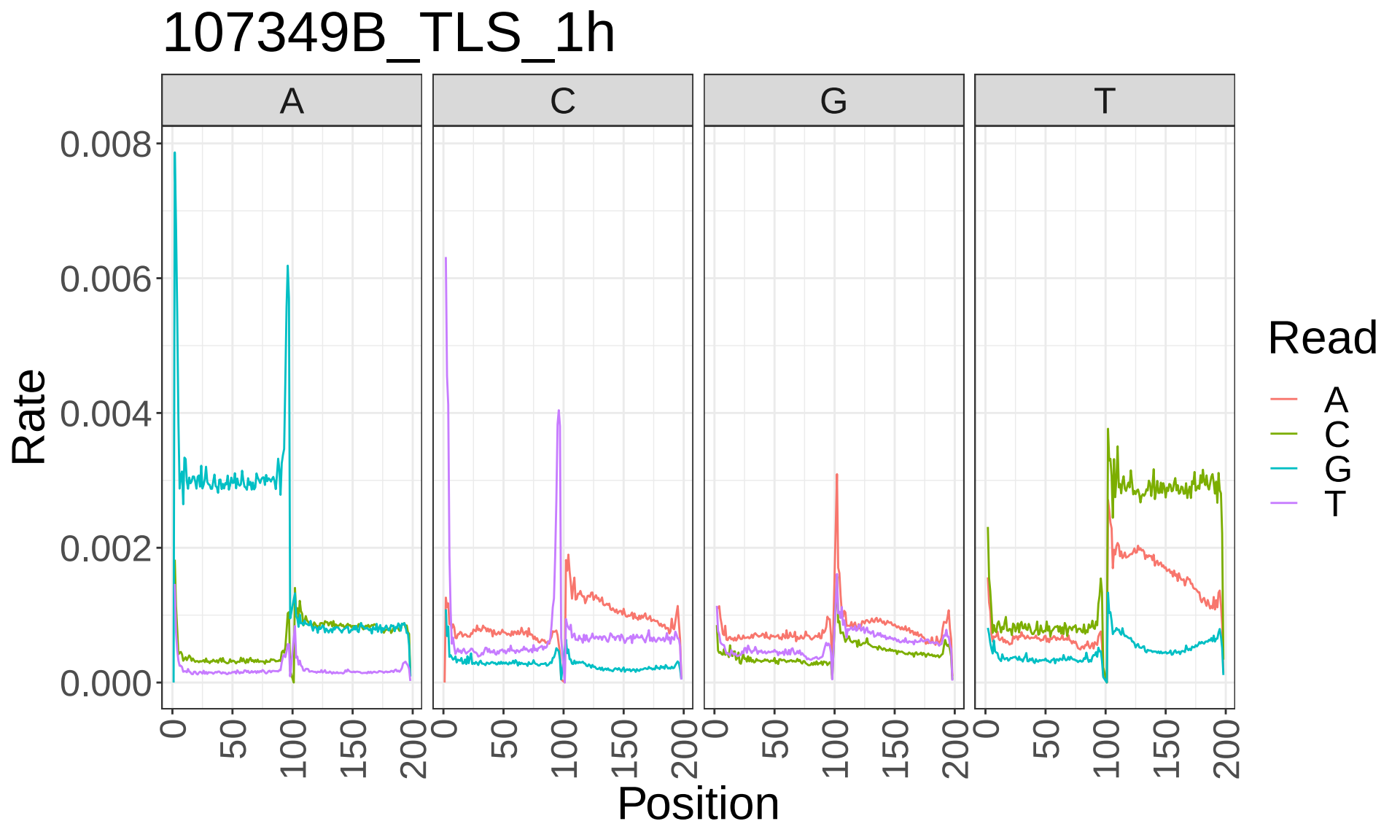

107319A_TLS_1h
A
C
G
T
Rate
0.006
0.004
0.002
0.000
0
50
100
150
200
Position
Read
A
C
G
T

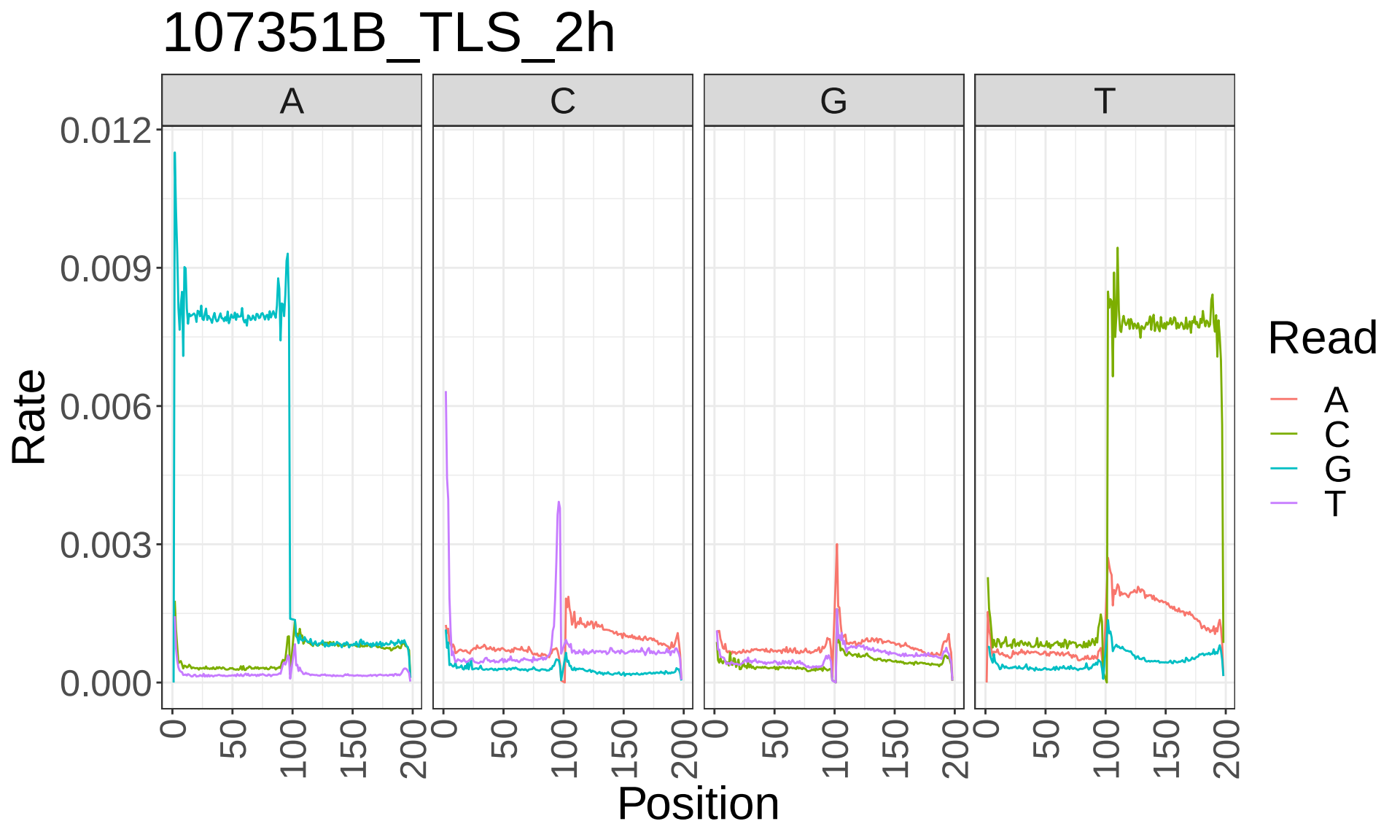

107321A_TLS_2h
A
C
G
T
Rate
0.010
0.005
0.000
0
50
100
150
200
Position
Read
A
C
G
T

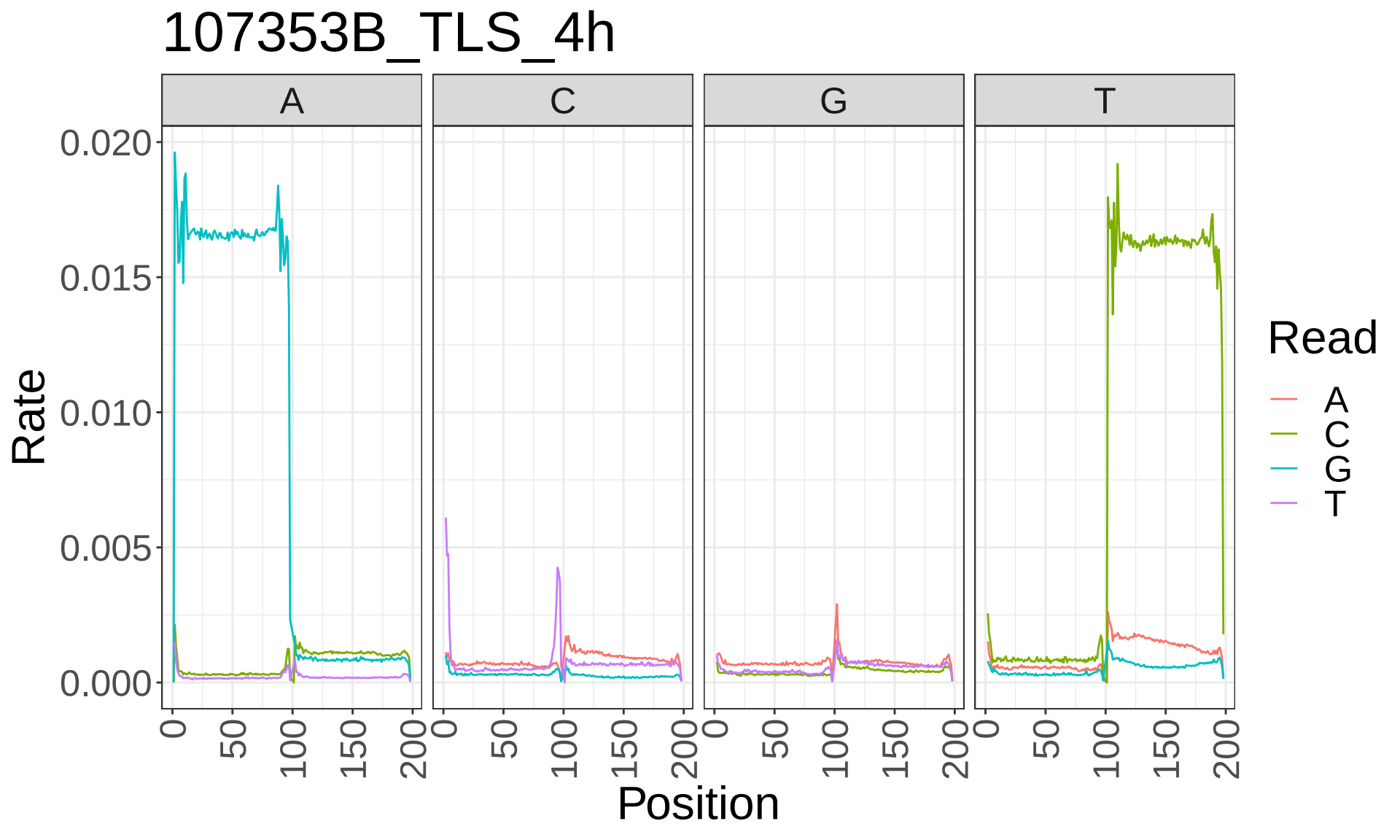

107323A_TLS_4h
A
C
G
T
Rate
0.015
0.010
0.005
0.000
0
50
100
150
200
Position
Read
A
C
G
T

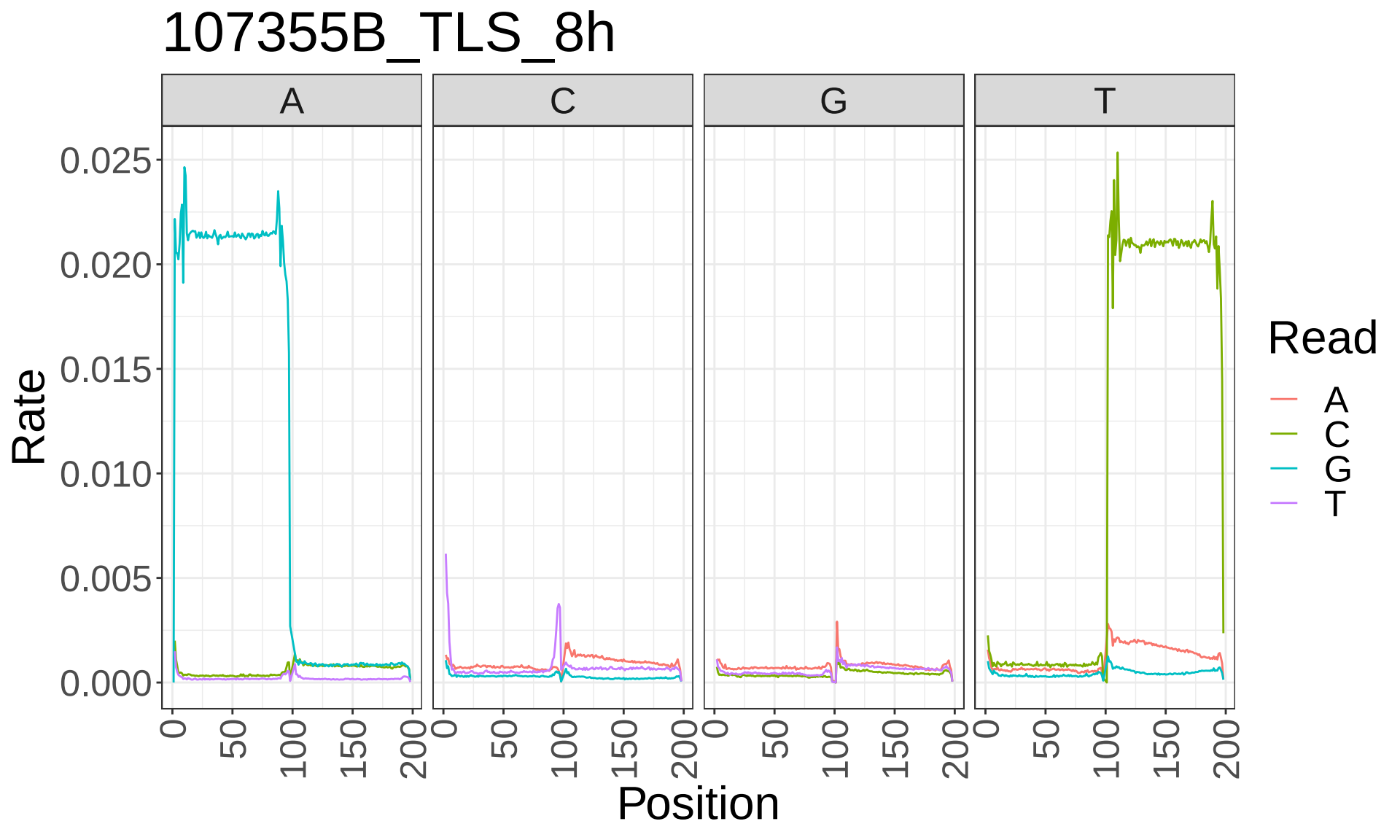

107325A_TLS_8h
A
C
G
T
Rate
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
Position
Read
A
C
G
T

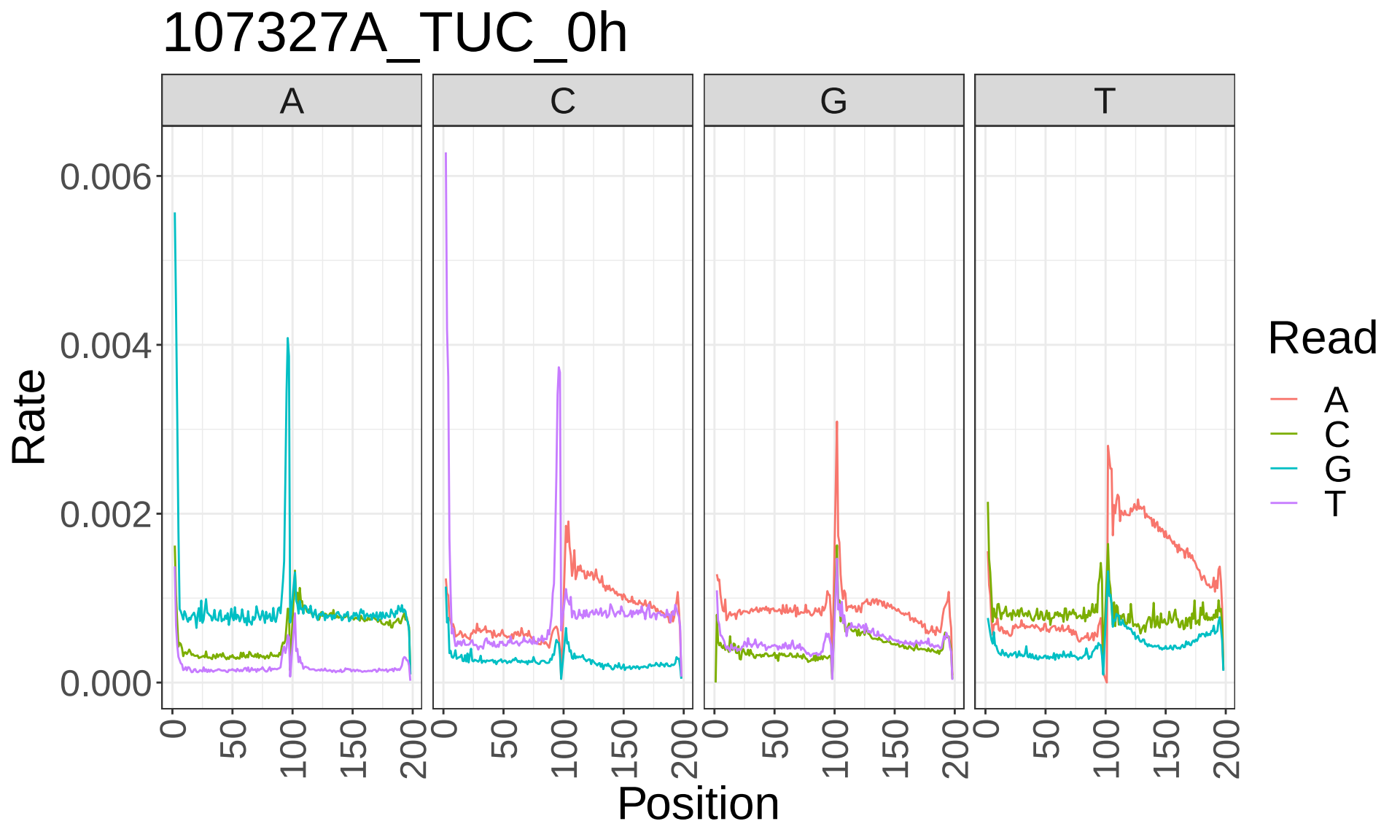

107347B_TLS_0h
A
C
G
T
Rate
0.006
0.004
0.002
0.000
0
50
100
150
200
Position
Read
A
C
G
T

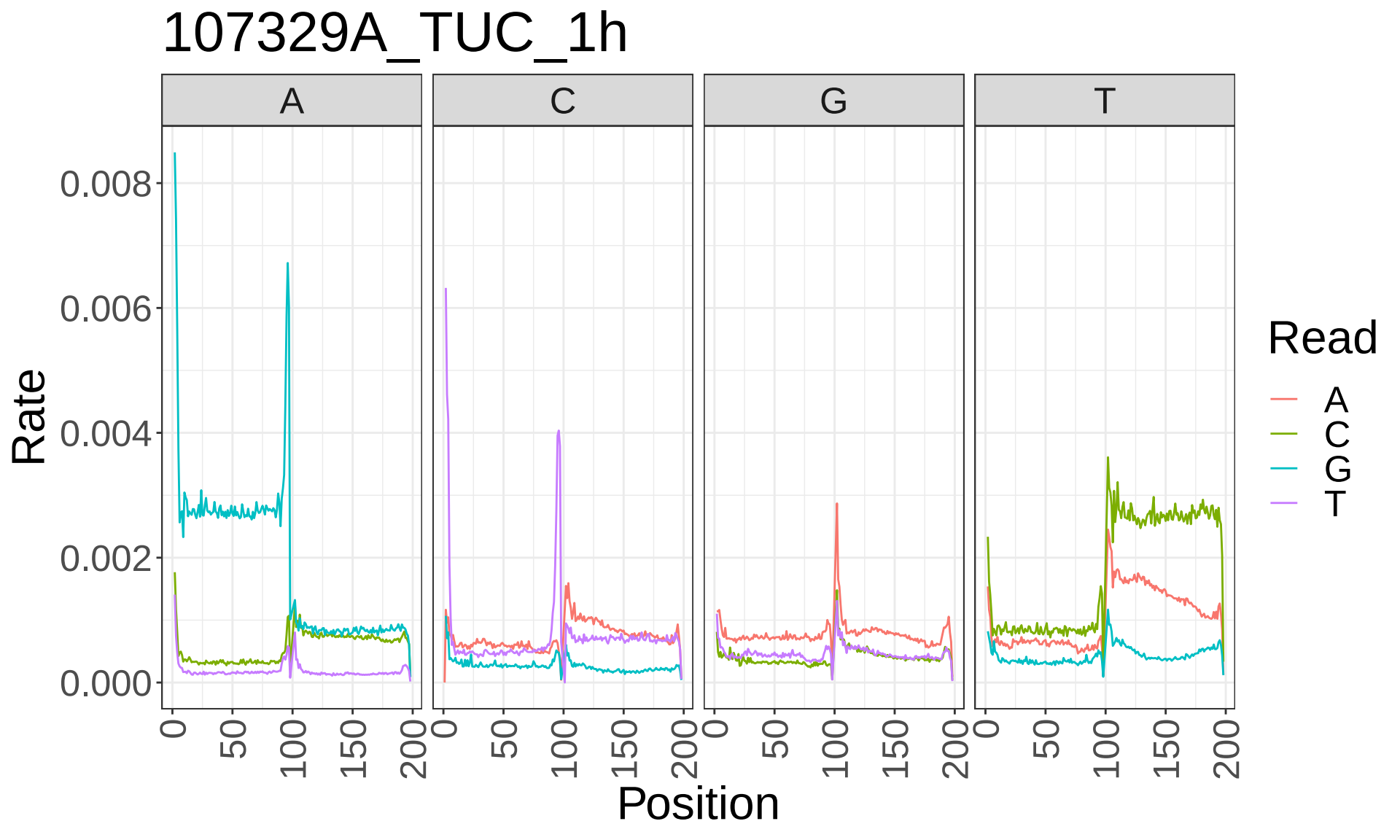

107349B_TLS_1h
A
C
G
T
Rate
0.008
0.006
0.004
0.002
0.000
0
50
100
150
200
Position
Read
A
C
G
T

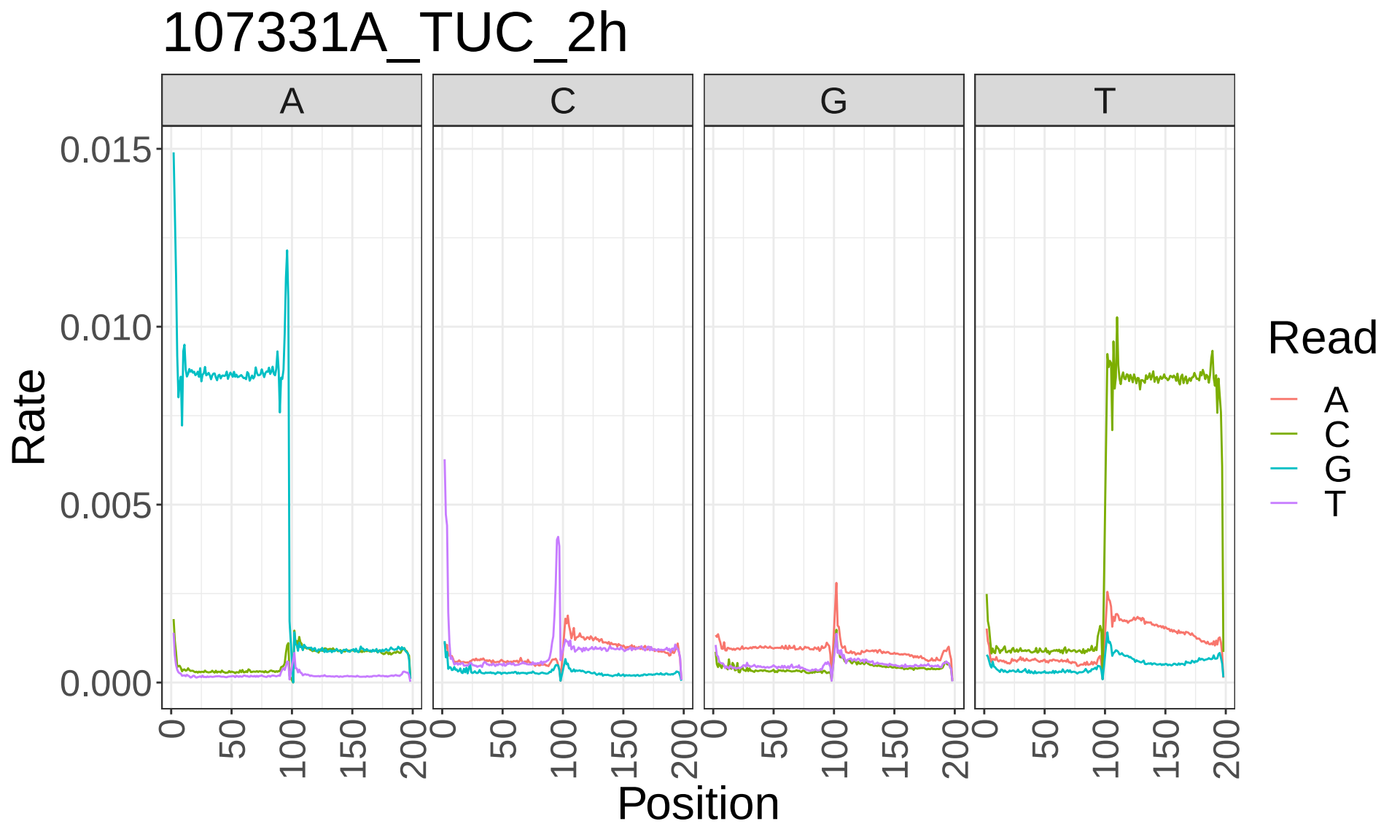

107351B_TLS_2h
A
C
G
T
0.012
0.009
0.006
0.003
0.000
Rate
0
50
100
150
200
Position
Read
A
C
G
T

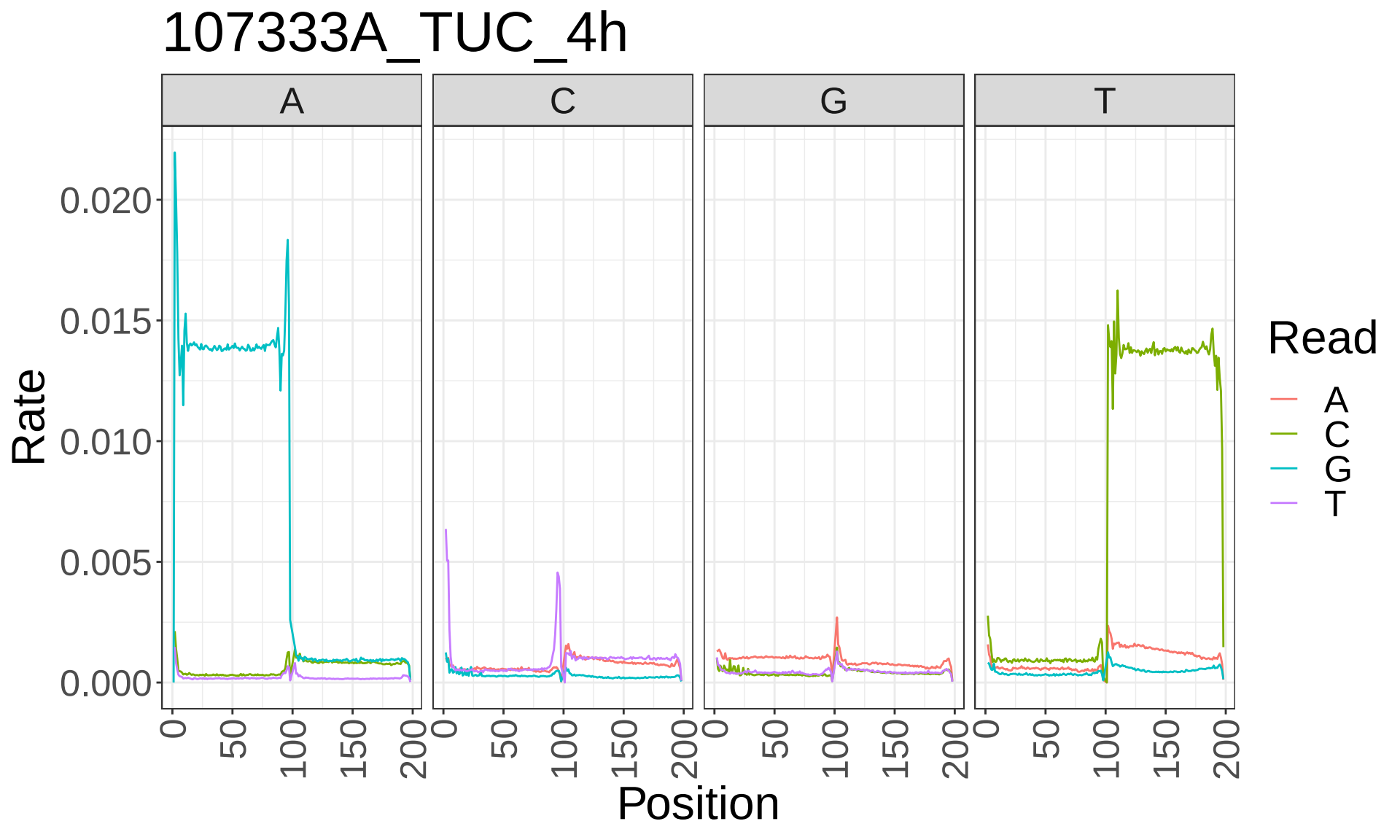

107353B_TLS_4h
A
C
G
T
Rate
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
Position
Read
A
C
G
T

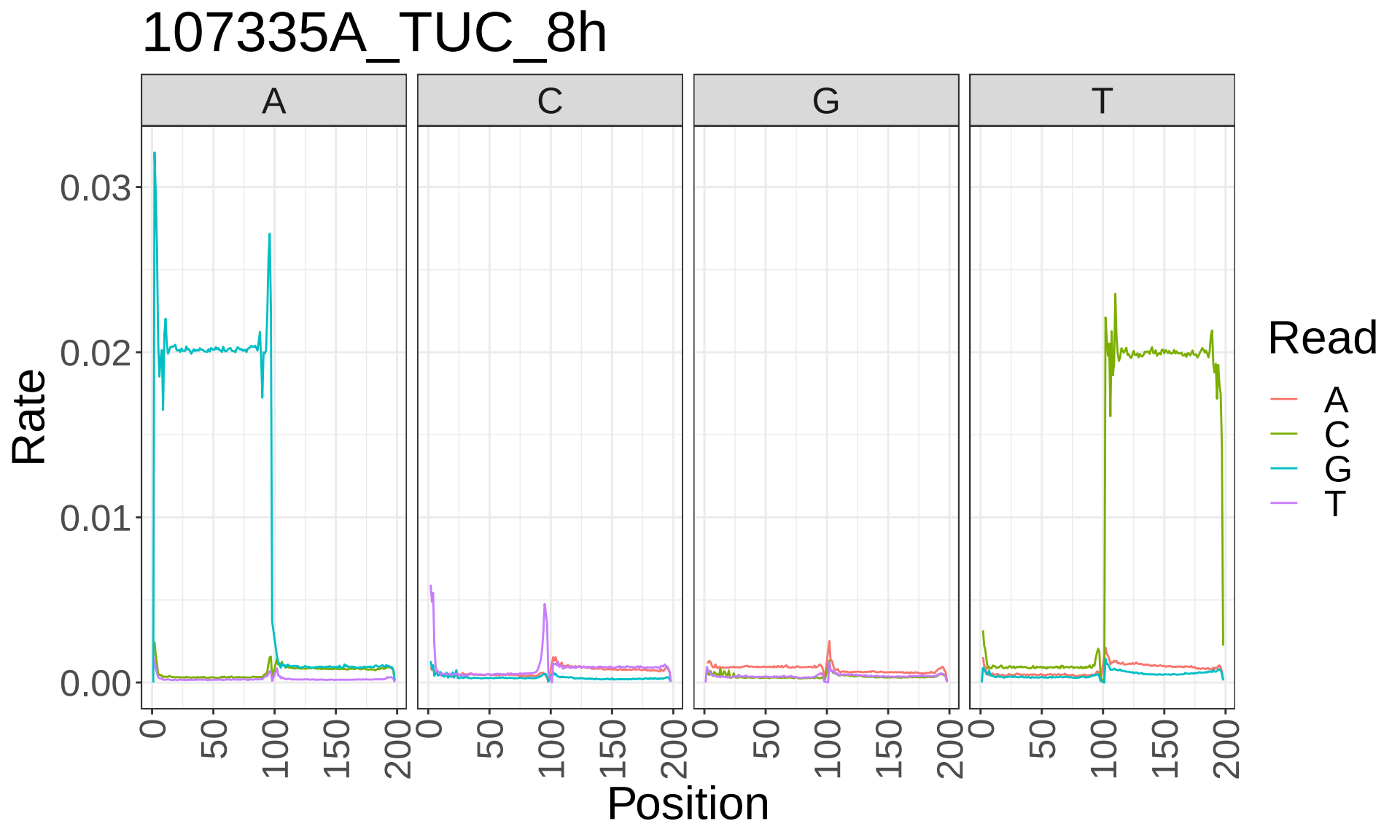

107355B_TLS_8h
A
C
G
T
Rate
0.025
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
Position
Read
A
C
G
T

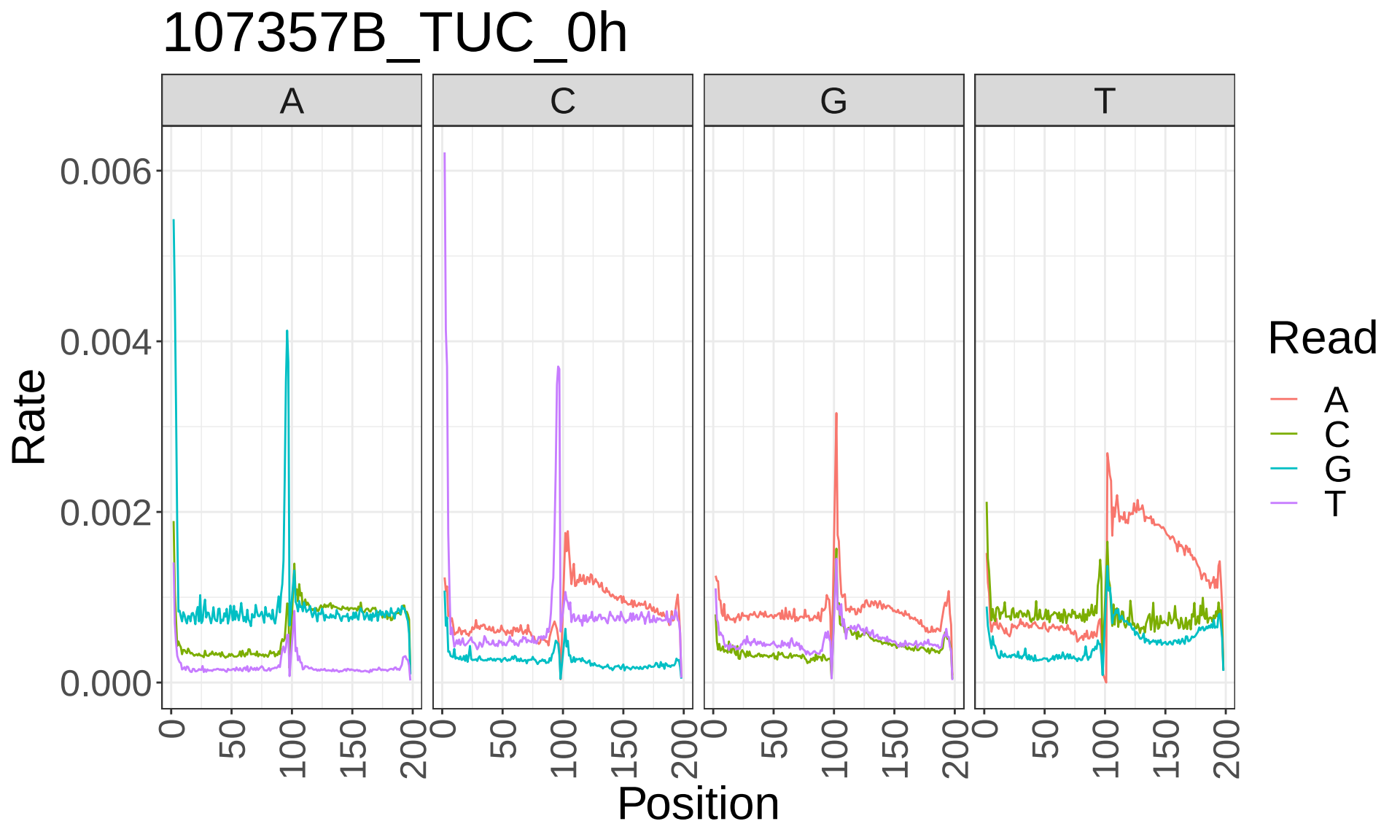

107327A_TUC_0h
A
C
G
T
0.006
0.004
0.002
0.000
Rate
0
50
100
150
200
Position
Read
A
C
G
T

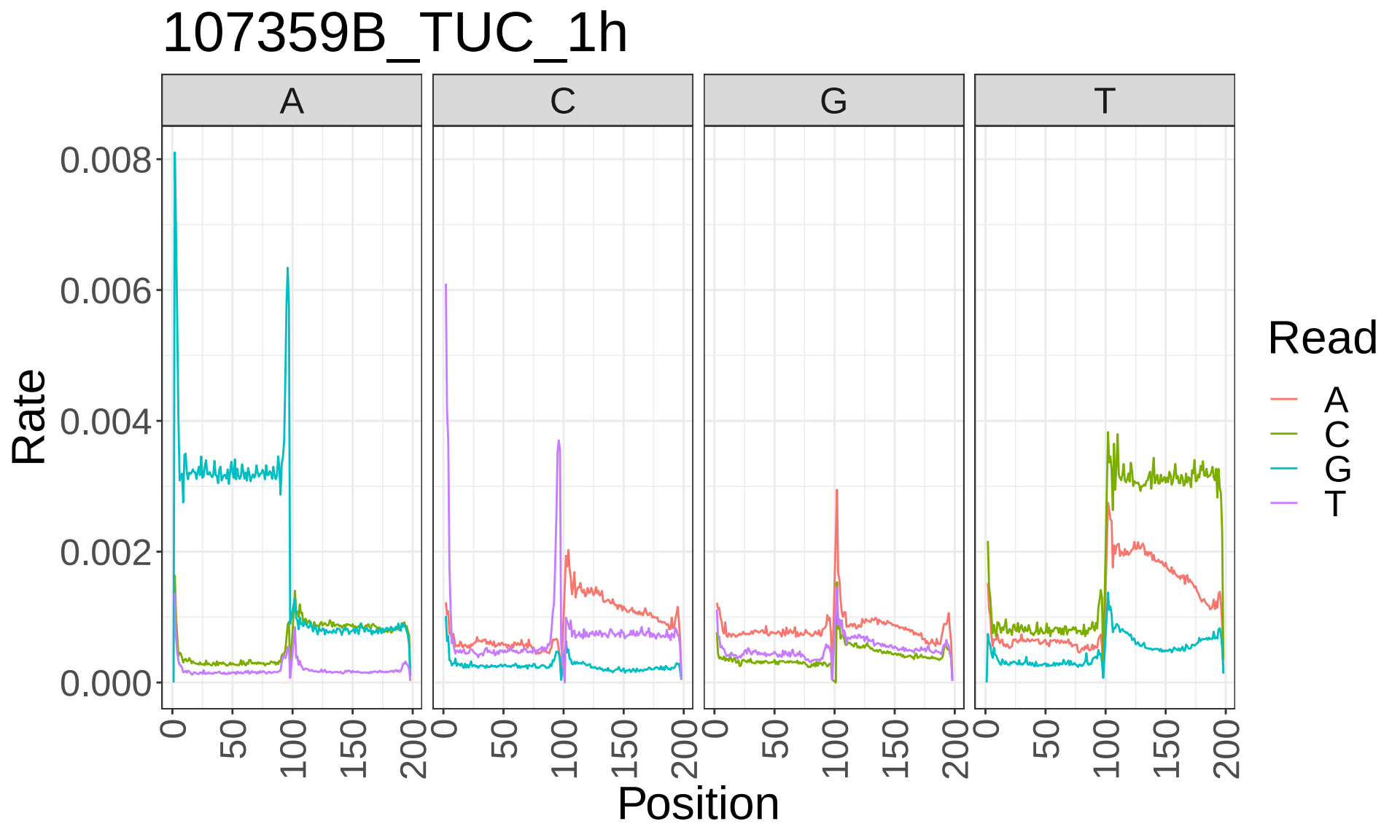

107329A_TUC_1h
A
C
G
T
Rate
0.008
0.006
0.004
0.002
0.000
0
50
100
150
200
Position
Read
A
C
G
T

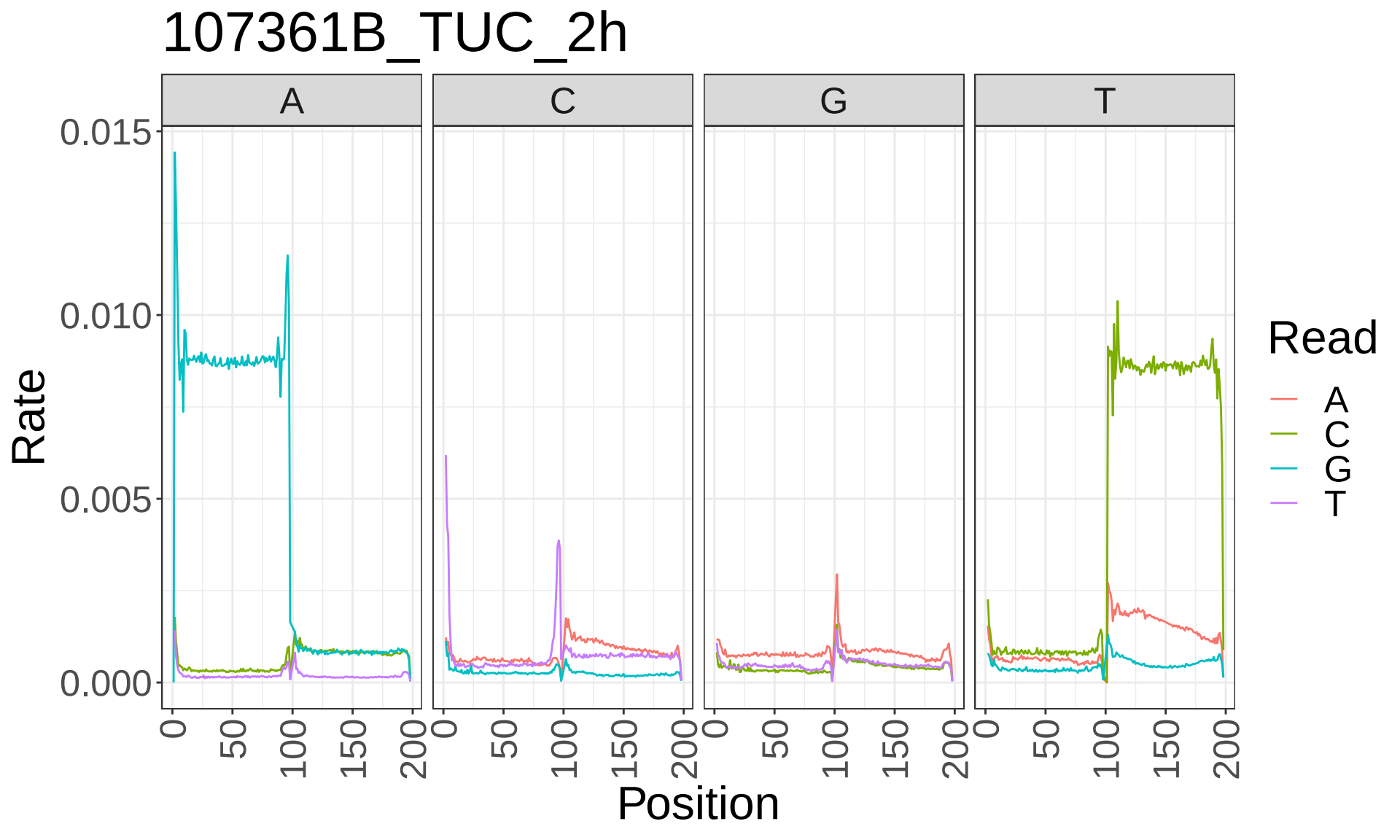

107331A_TUC_2h
A
C
G
T
0.015
0.010
0.005
0.000
Rate
0
50
100
150
200
Position
Read
A
C
G
T

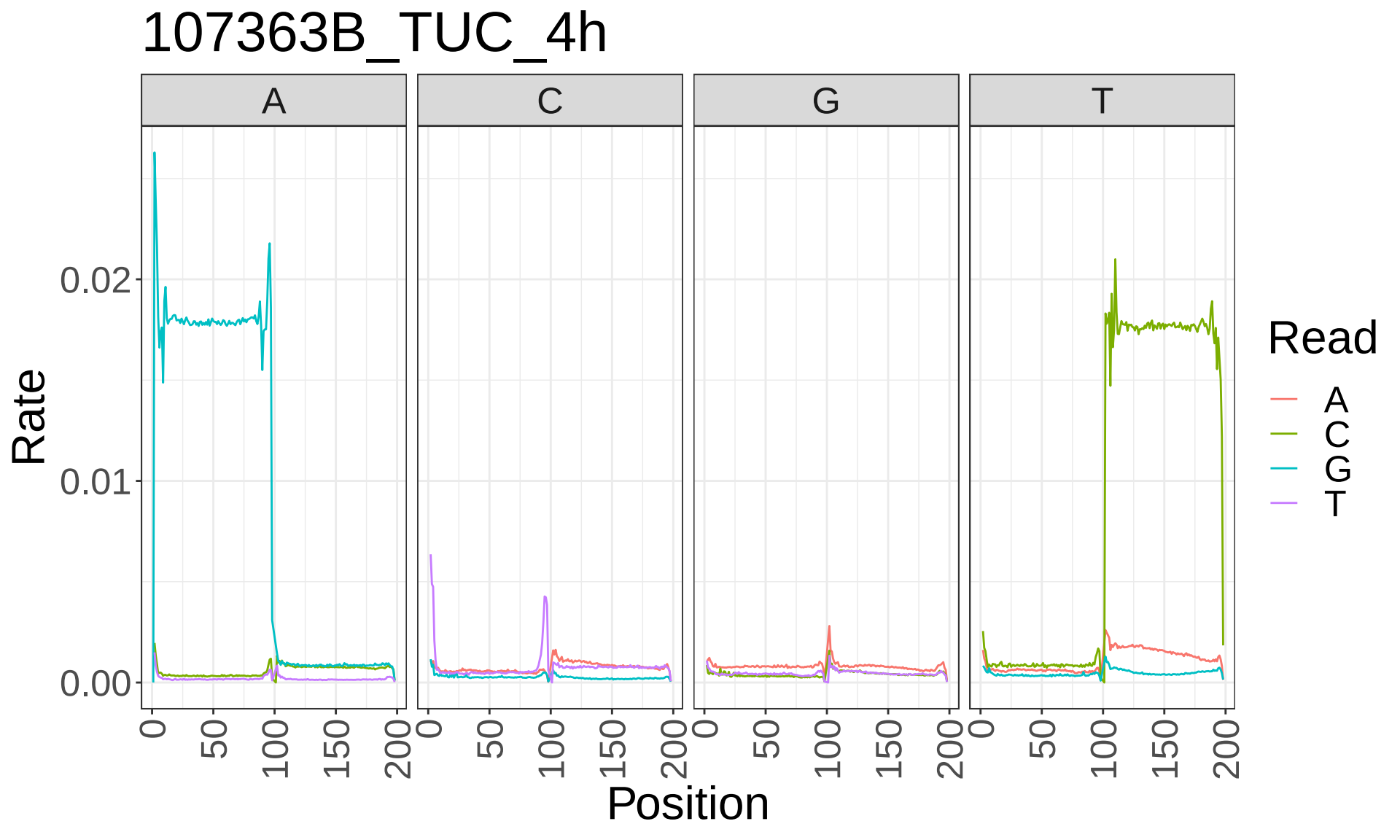

107333A_TUC_4h
A
C
G
T
Rate
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
Position
Read
A
C
G
T

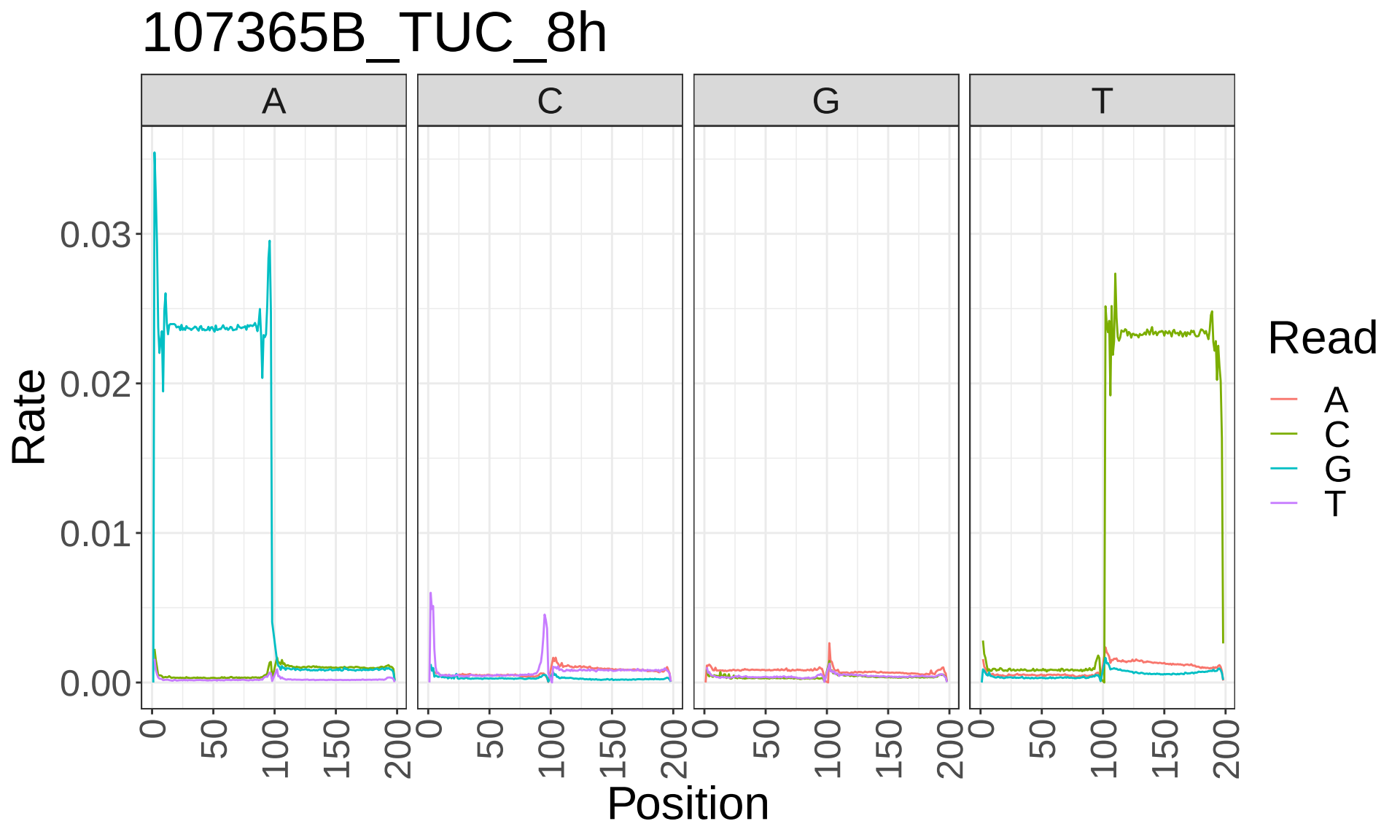

107335A_TUC_8h
A
C
G
T
Rate
0.03
0.02
0.01
0.00
0
50
100
150
200
Position
Read
A
C
G
T


107357B_TUC_0h
A
C
G
T
0.006
0.004
0.002
0.000
Rate
0
50
100
150
200
Position
Read
A
C
G
T


107359B_TUC_1h
A
C
G
T
0.008
0.006
0.004
0.002
0.000
Rate
0
50
100
150
200
Position
Read
A
C
G
T


107361B_TUC_2h
A
C
G
T
0.015
0.010
0.005
0.000
Rate
0
50
100
150
200
Position
Read
A
C
G
T


107363B_TUC_4h
A
C
G
T
Rate
0.02
0.01
0.00
0
50
100
150
200
Position
Read
A
C
G
T


107365B_TUC_8h
A
C
G
T
Rate
0.03
0.02
0.01
0.00
0
50
100
150
200
Position
Read
A
C
G
T