

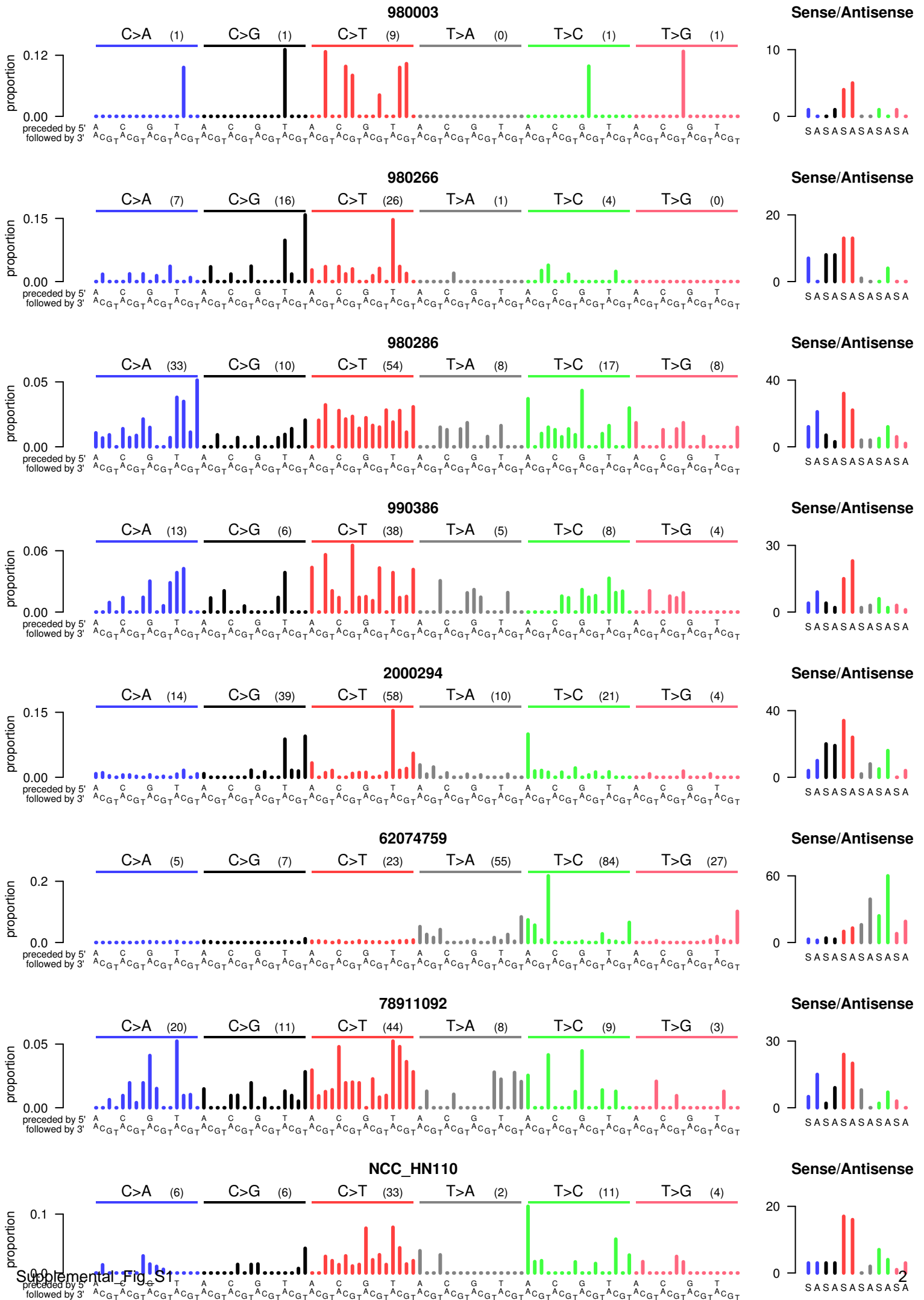
# Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types

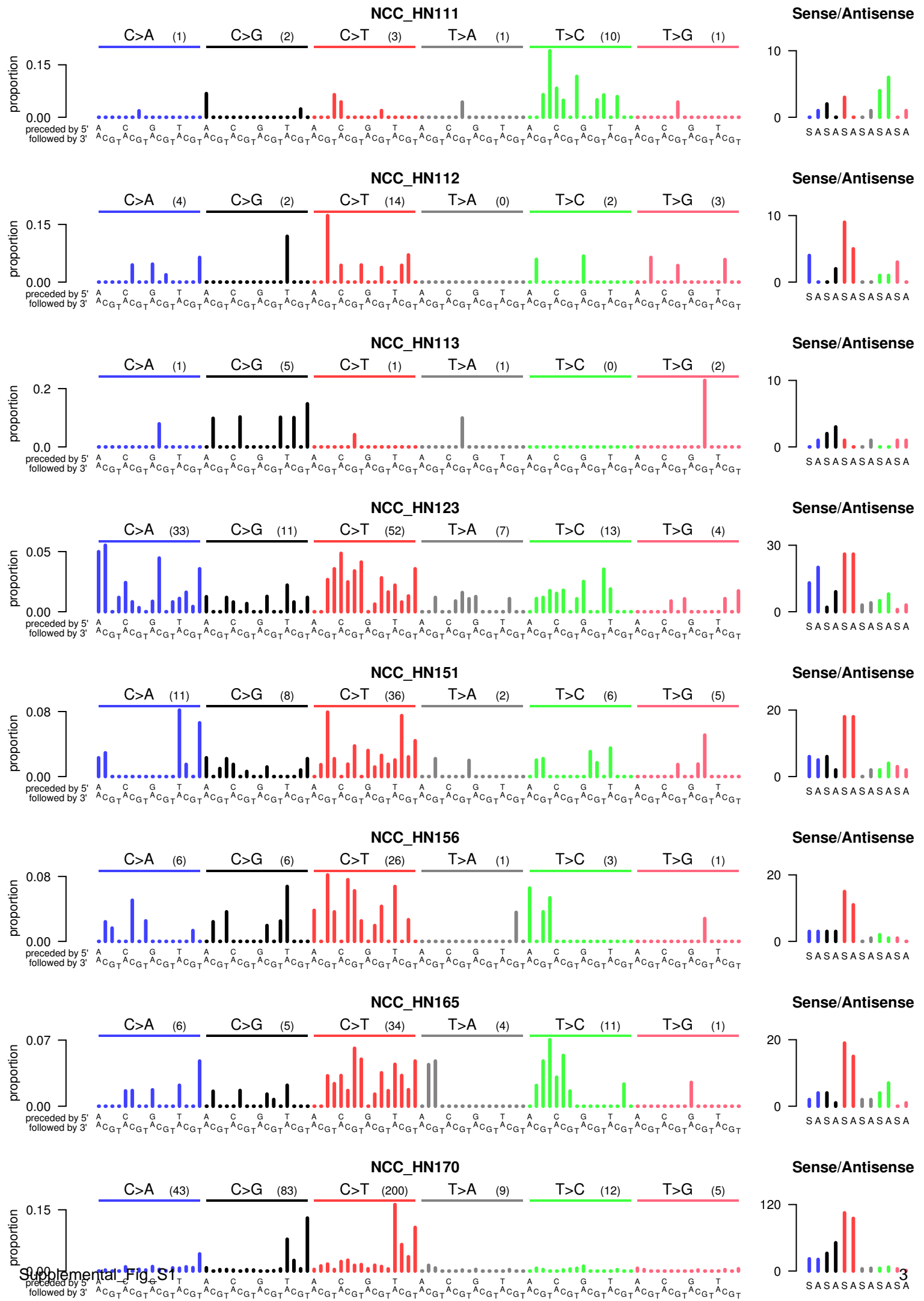
Boot et al., Genome Research, 2020

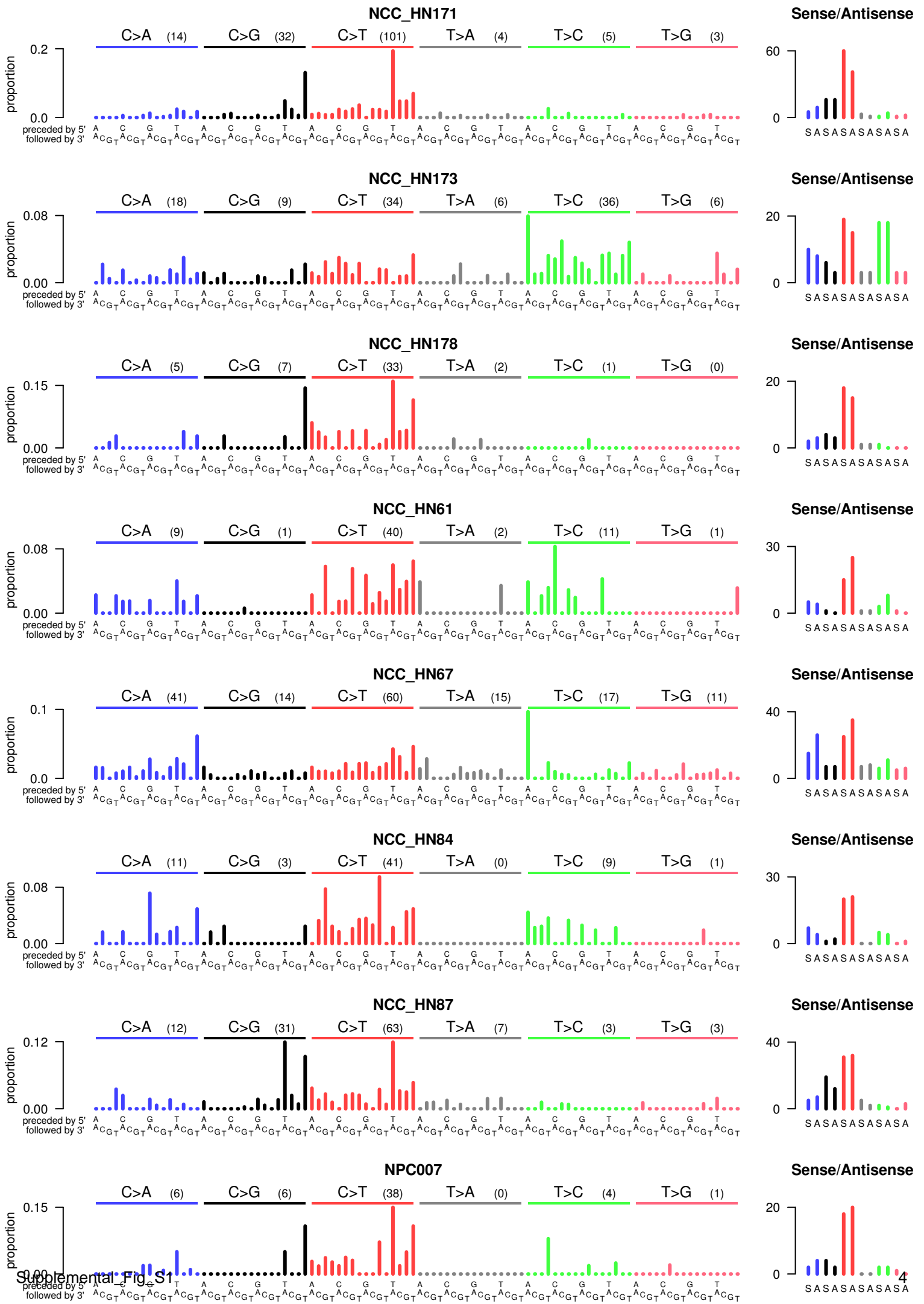
## Supplementary Figures

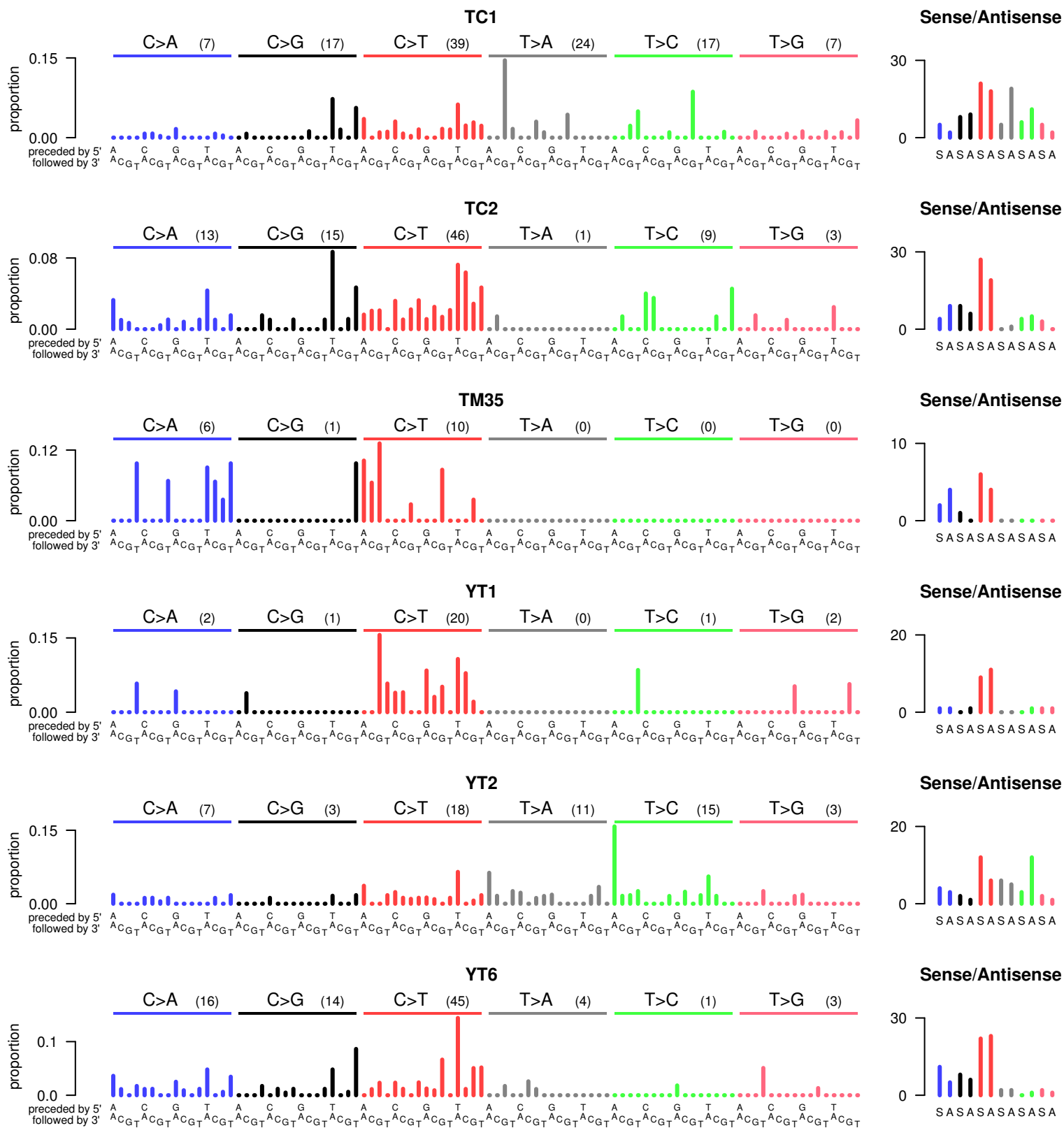
	Page
Supplemental_Fig_S1: SNS mutational spectra of the exome sequencing data from 30 OSCCs analyzed in this study.	2
Supplemental_Fig_S2: Mutational signature analysis of 30 OSCCs.	6
Supplemental_Fig_S3: T-SNE clustering of singaporean OSCCs and TCGA HNSCCs.	7
Supplemental_Fig_S4: SNS sequence context of all (top), all T>N (bottom left) and all C>N (bottom right) mutations in sample 62074759.	8
Supplemental_Fig_S5: Variant allele frequency analysis of mutations in sample 62074759 suggestive of ongoing mutagenesis.	9
Supplemental_Fig_S6: Sequence context of deletions of single thymines in thymine-repeats in sample 62074759.	10
Supplemental_Fig_S7: Transcriptional strand bias of deletions of a single thymine in thymine-repeats ranging from 1 to 8	11
Supplemental_Fig_S8: Dinucleotide substitution spectrum of sample 62074759.	12
Supplemental_Fig_S9: Mutational spectra of tumours from publicly available sequencing data showing likely exposure to SBS_A <sup>n</sup> T identified by examining the sequence context of thymine mutations.	13
Supplemental_Fig_S10: Indel spectra for SBS_AnT positive PCAWG tumours.	20
Supplemental_Fig_S11: Sequence context of deletions of thymines by repeat length for SBS_AnT positive PCAWG tumours	21
Supplemental_Fig_S12: SNS mutational spectra of duocarmycin SA treated HepG2 clones.	27
Supplemental_Fig_S13: Indel mutational spectra of duocarmycin SA treated HepG2 clones.	28
Supplemental_Fig_S14: Double base substitution spectrum of Duocarmycin SA treated HepG2 clones.	29
Supplemental_Fig_S15: Characterization of the TC1 mutational spectrum.	30
Supplemental_Fig_S16: Identifying samples whose mutational spectrum resembles that of TC1.	31
Supplemental_Fig_S17: Alignment of non-human reads from WGS data of patients 62074759 and TC1 to bacterial genomes	38

Supplemental\_Fig\_S1: SNS mutational spectra of the exome sequencing data from 30 OSCCs analyzed in this study.

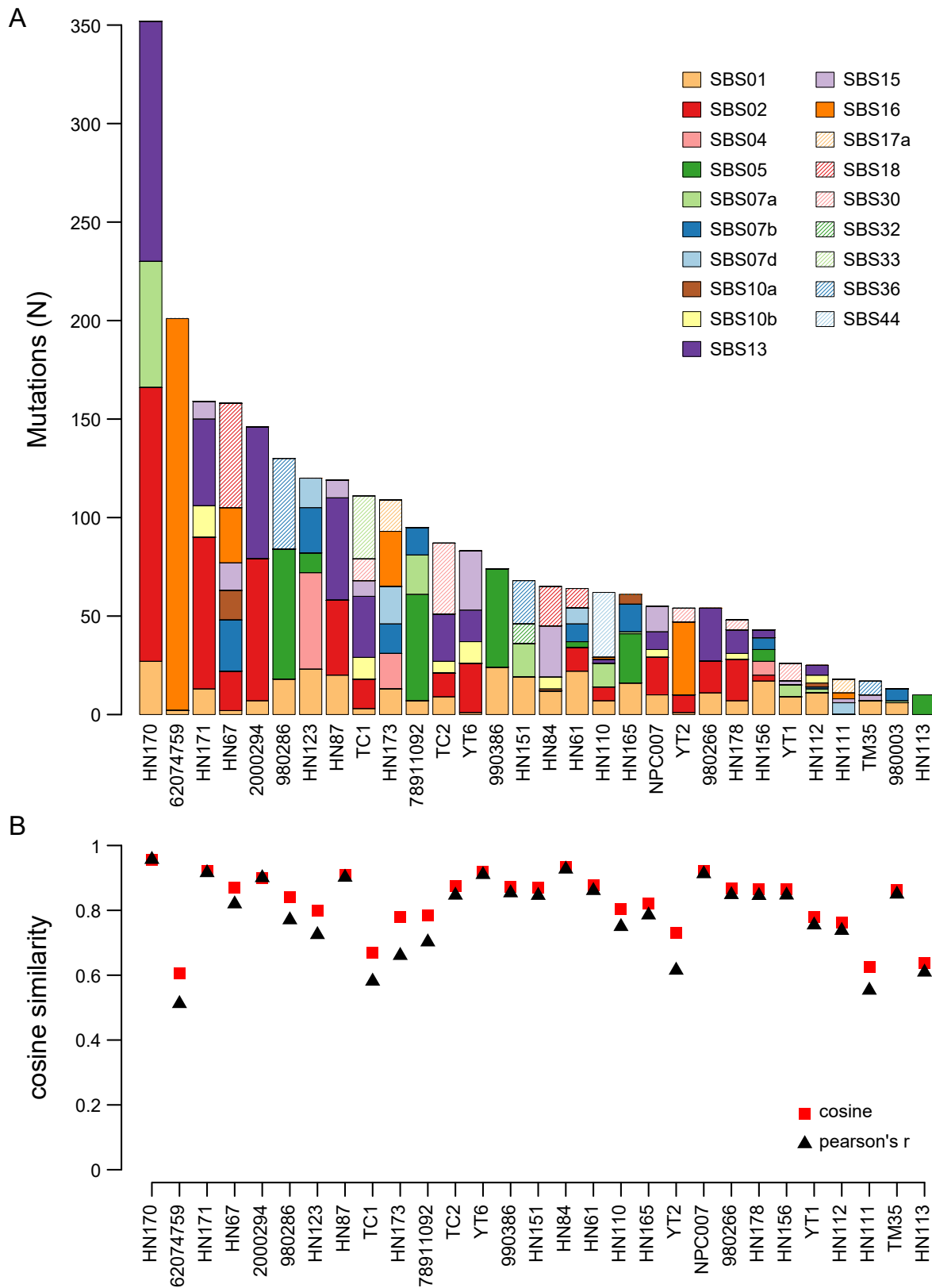




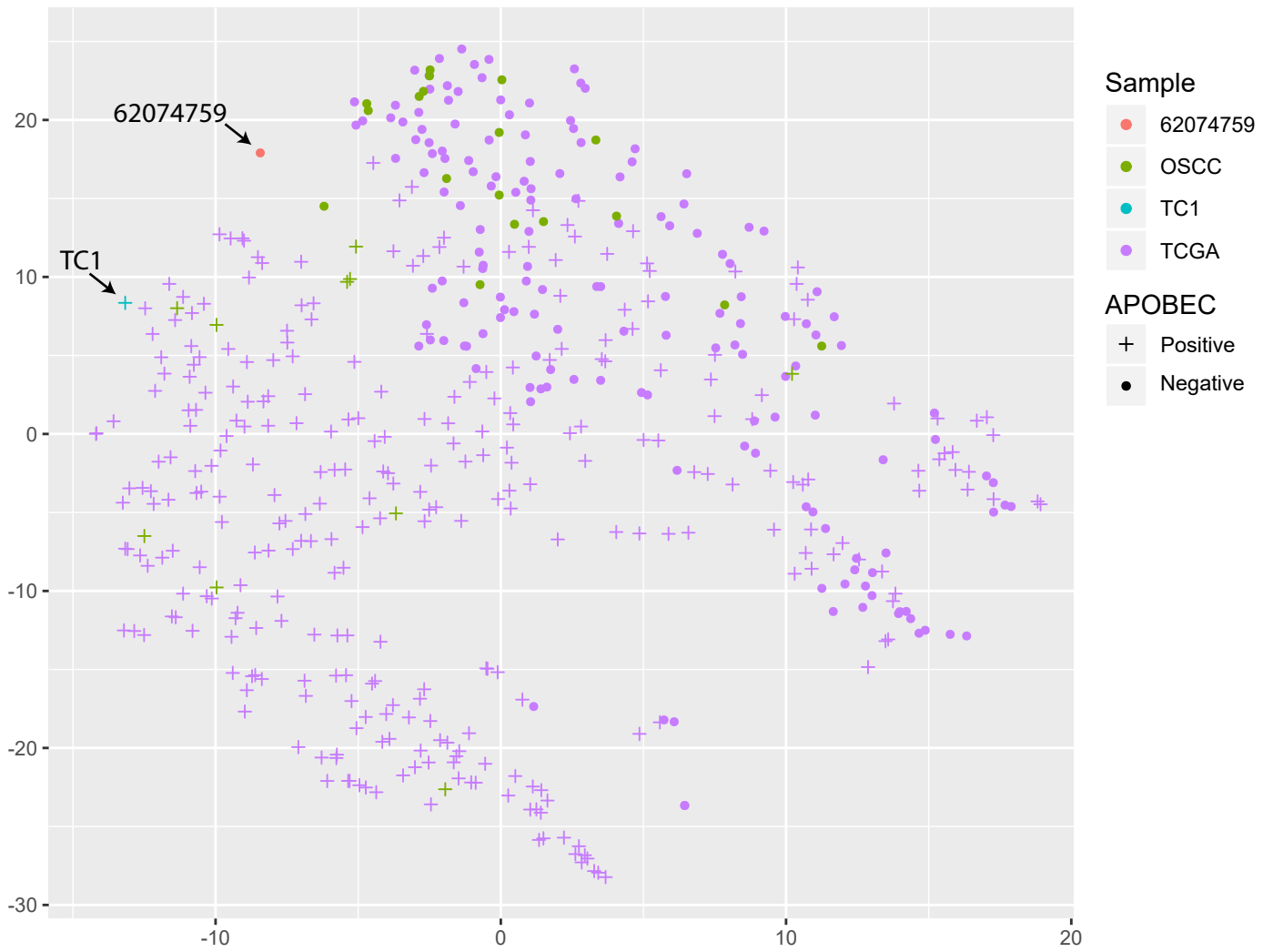




Supplemental\_Fig\_S2: **Mutational signature analysis of 30 OSCCs.** **A:** Reconstruction of the patient mutation spectra using the mutational signatures established by the PCAWG consortium 1. **B:** Evaluation of reconstruction by cosine similarity and Pearson's  $r$ .

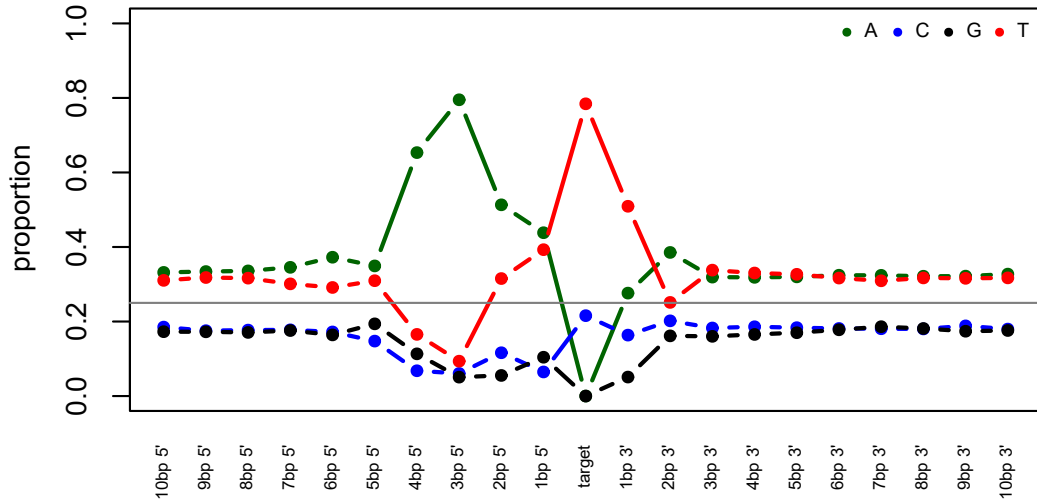


Supplemental\_Fig\_S3: T-SNE clustering of singaporean OSCCs and TCGA HNSCCs.

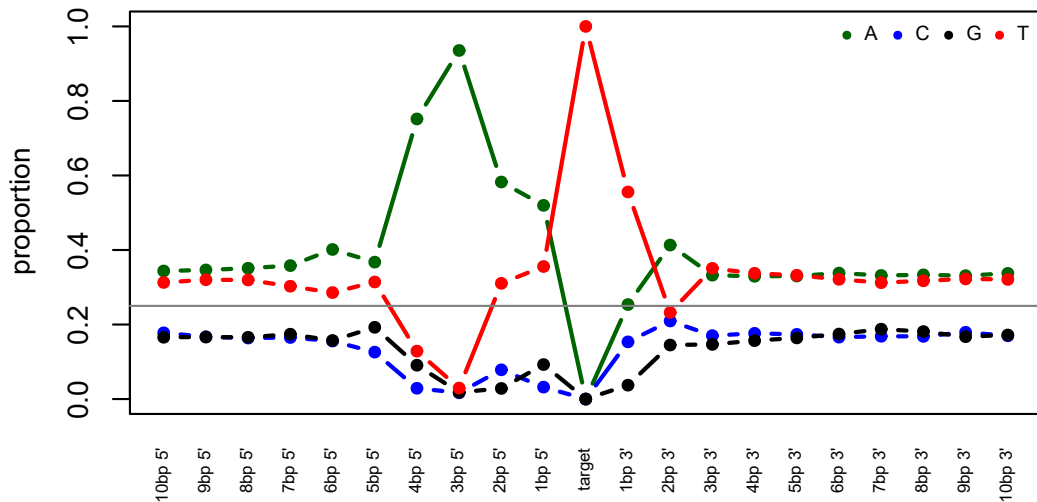


Supplemental\_Fig\_S4: SNS sequence context of all (top), all T>N (bottom left) and all C>N (bottom right) mutations in sample 62074759.

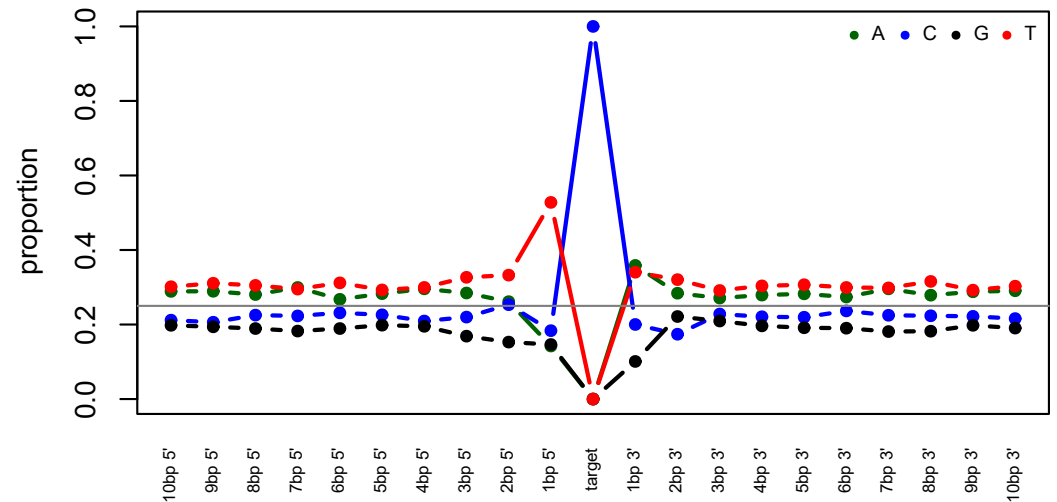
Sequence context of all mutations in 62074759



Sequence context of T>N mutations in 62074759



Sequence context of C>N mutations in 62074759



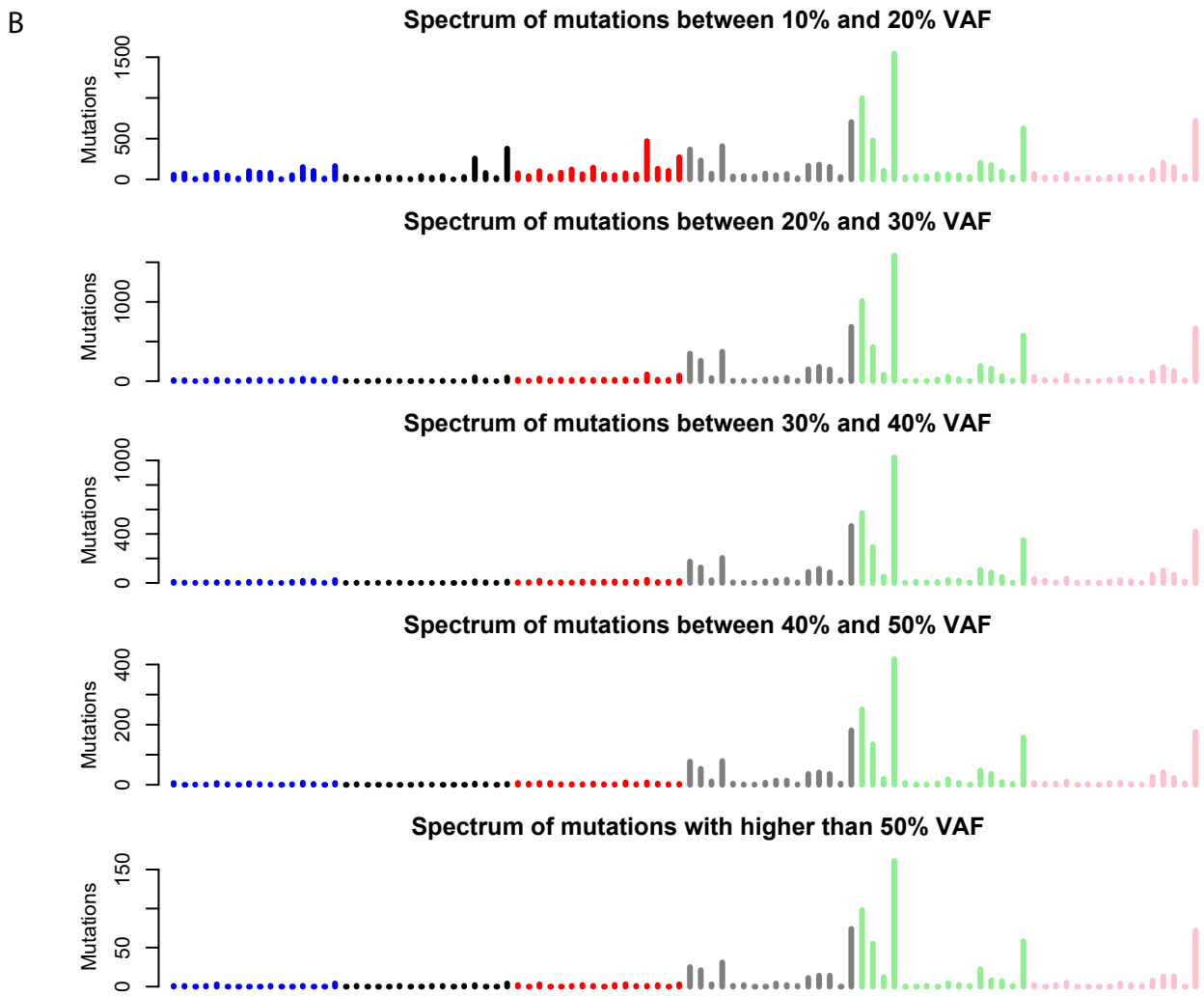
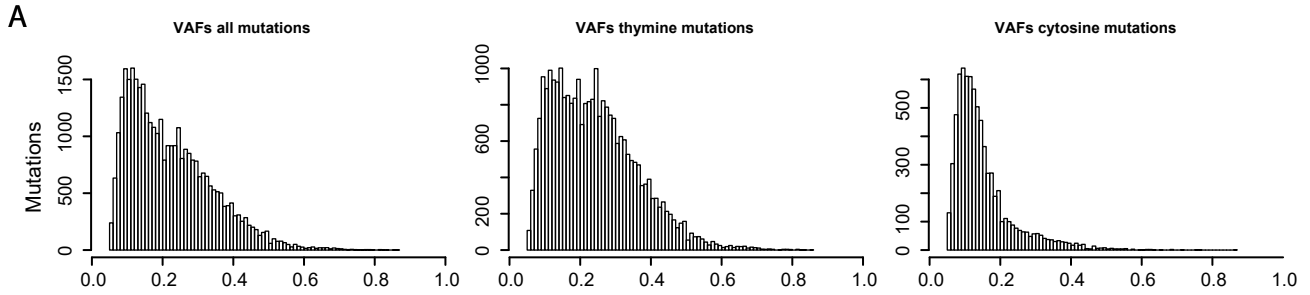


Supplemental\_Fig\_S5: Variant allele frequency analysis of mutations in sample 62074759 suggestive of ongoing mutagenesis.

(A) Variant allele frequency of mutations in sample 62074759.

(B) The mutational spectrum of sample 62074759 is similar across different variant allele frequencies.

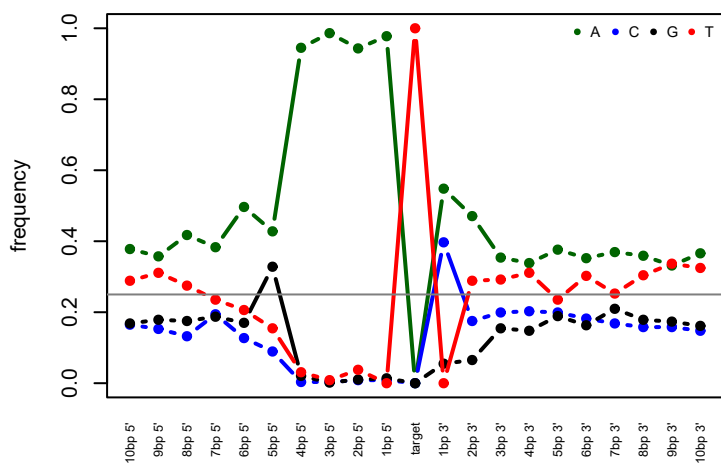
(C) Table of cosine similarities of the different spectra shown in B



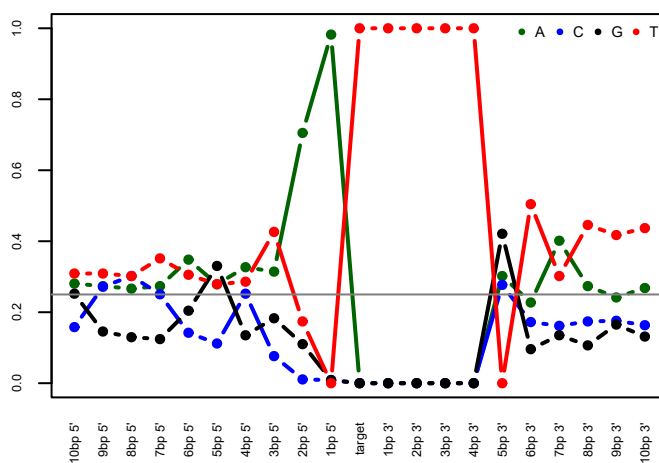
C Similarity of spectra of mutations with different VAFs (cosine similarity)

all	10-20%	20-30%	30-40%	40-50%	>50%	
	0.988	0.990	0.985	0.983	0.981	all
		0.958	0.948	0.945	0.943	10-20%
			0.997	0.996	0.994	20-30%
				0.998	0.997	30-40%
					0.997	40-50%
						>50%

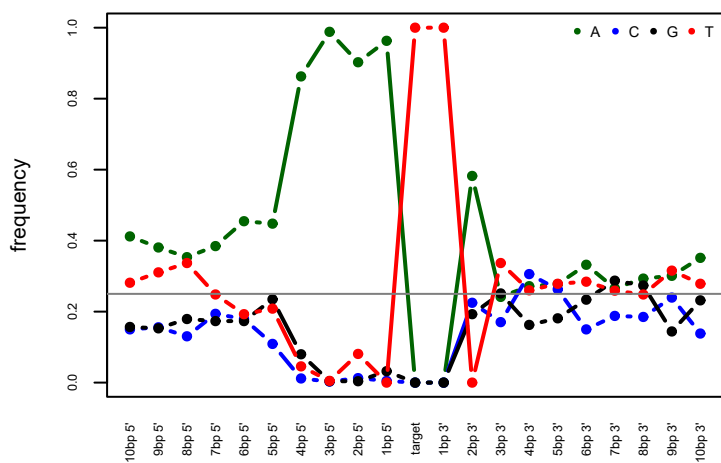
Sequence context of delT in single T (N=582)



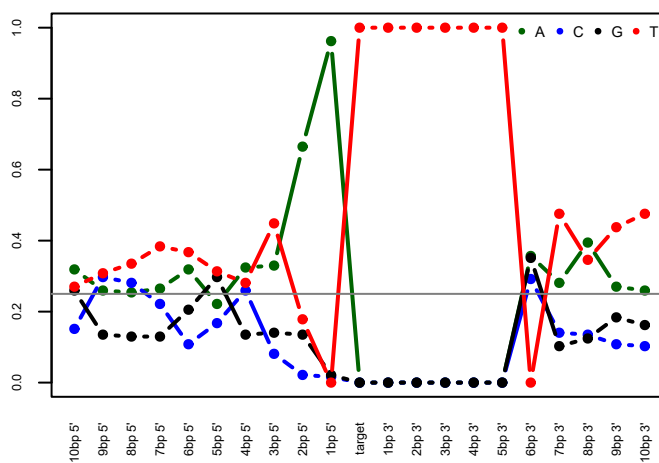
Sequence context of delT in TTTTT (N=563)



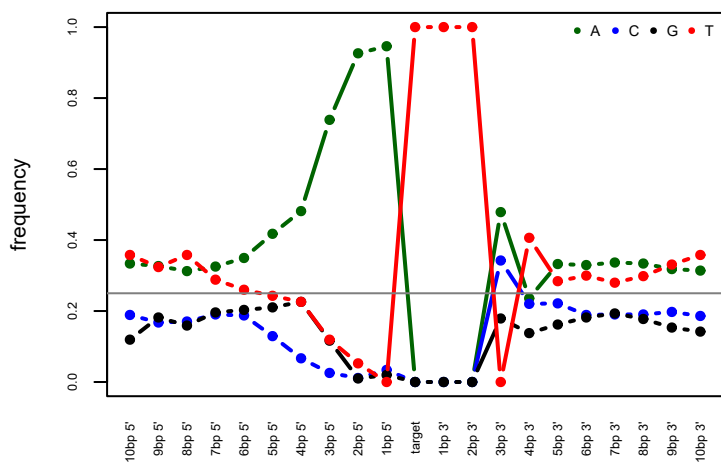
Sequence context of delT in TT (N=1027)



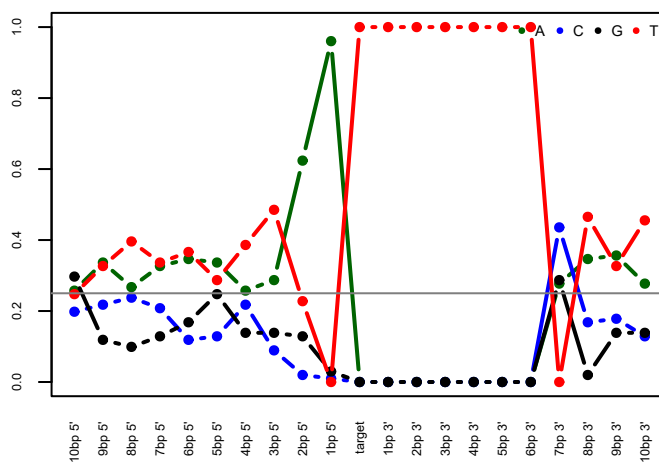
Sequence context of delT in TTTTTT (N=185)



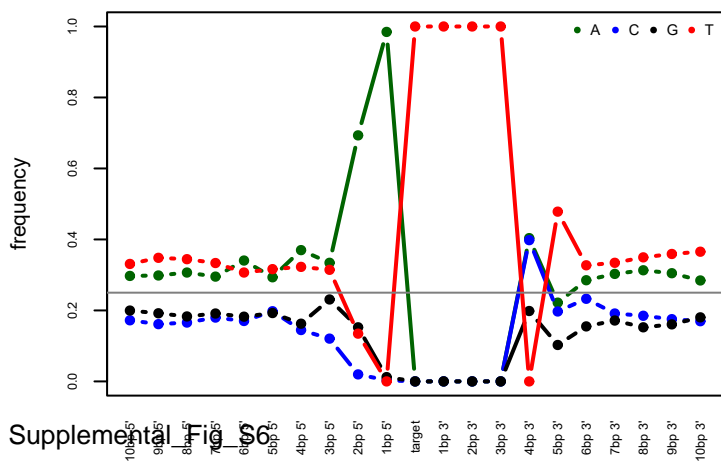
Sequence context of delT in TTT (N=704)



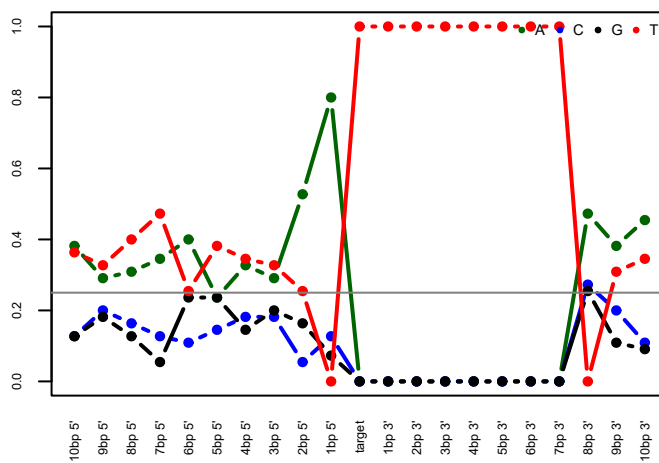
Sequence context of delT in TTTTTTT (N=101)



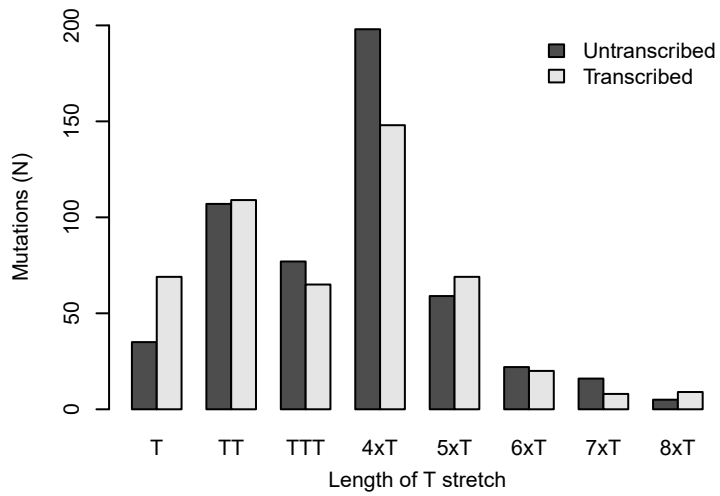
Sequence context of delT in TTTT (N=1568)



Sequence context of delT in TTTTTTTT (N=55)



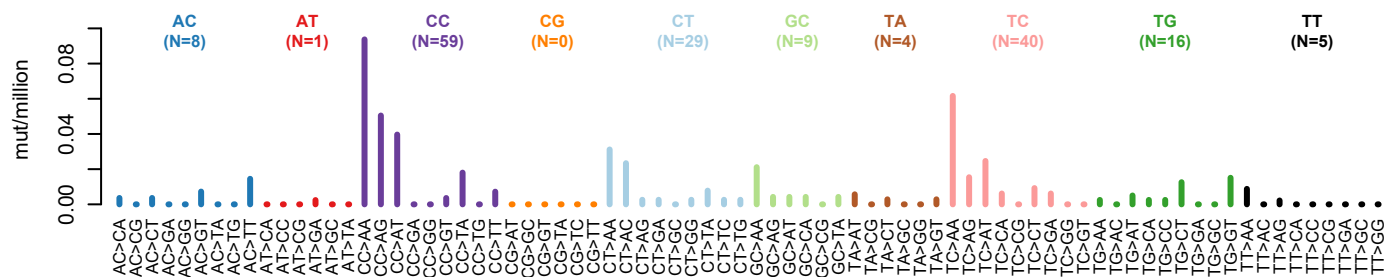
Supplemental\_Fig\_S7: **Transcriptional strand bias of deletions of a single thymine in thymine-repeats ranging from 1 to 8 thymines.**



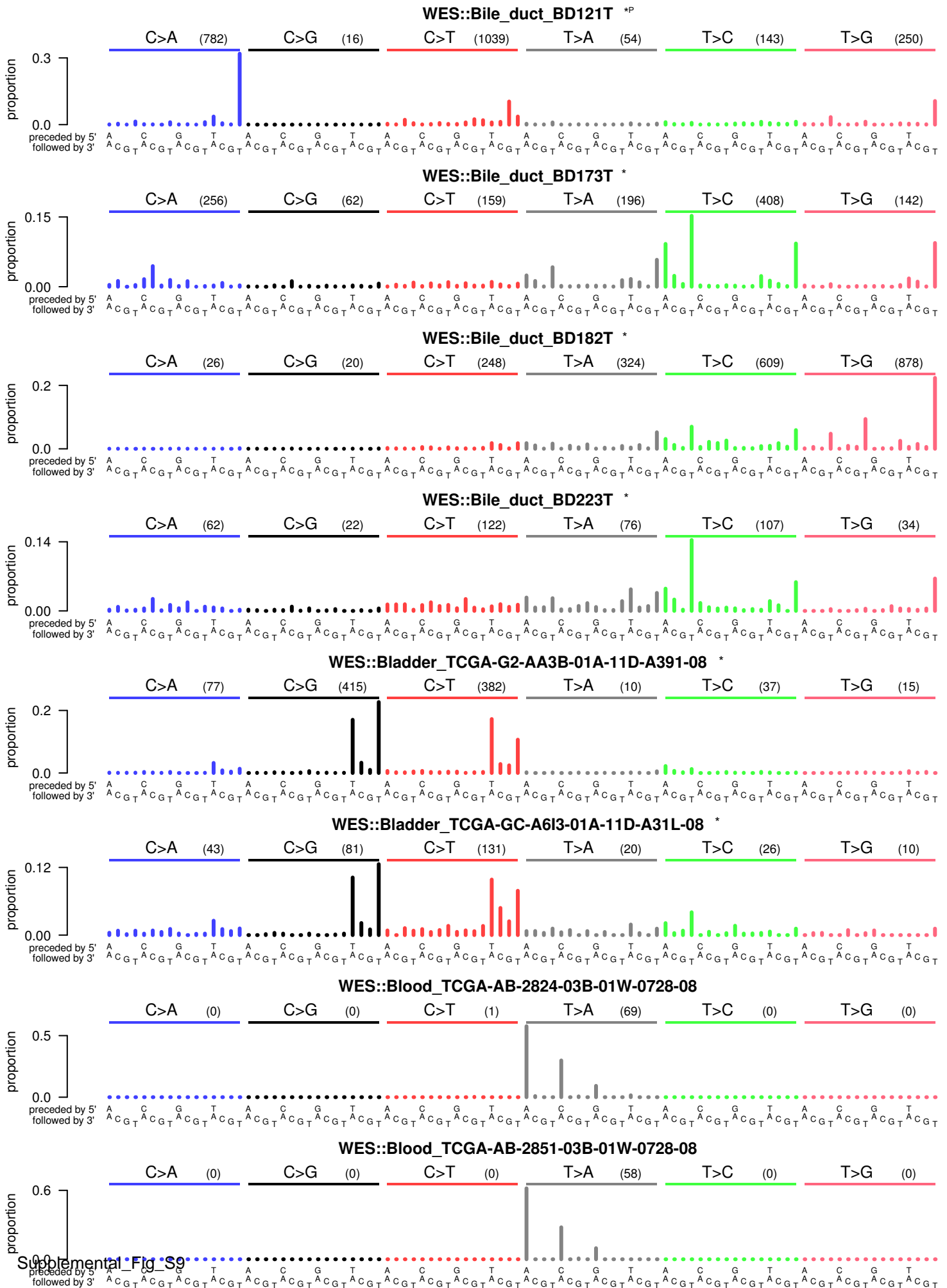
Binomial test results for transcriptional strand bias of deletions of thymines in sample 62074759

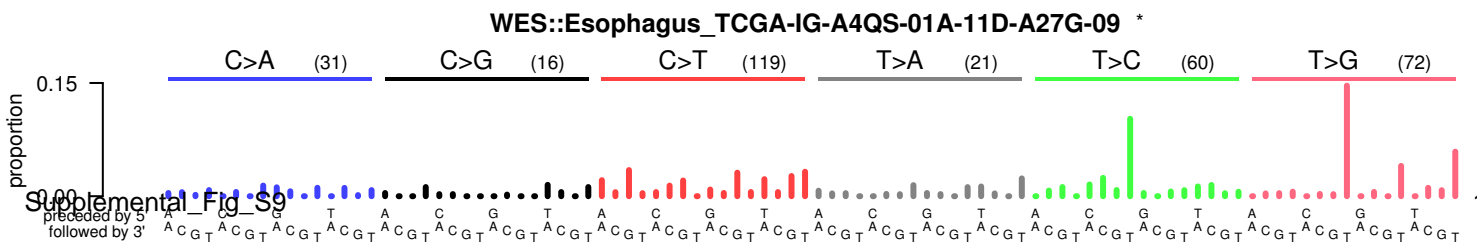
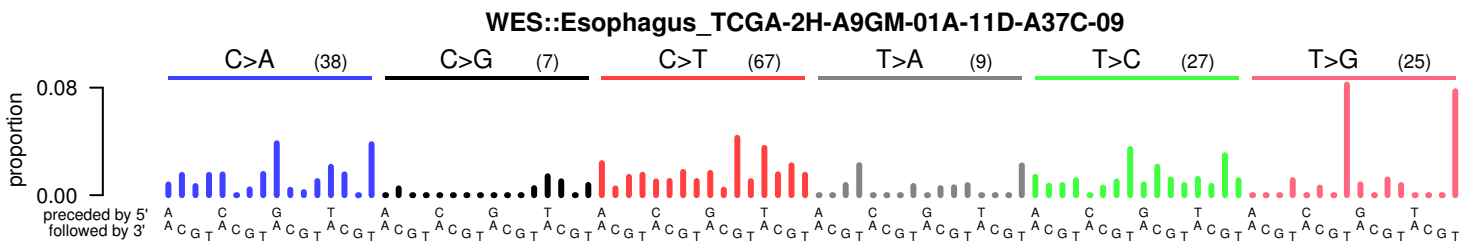
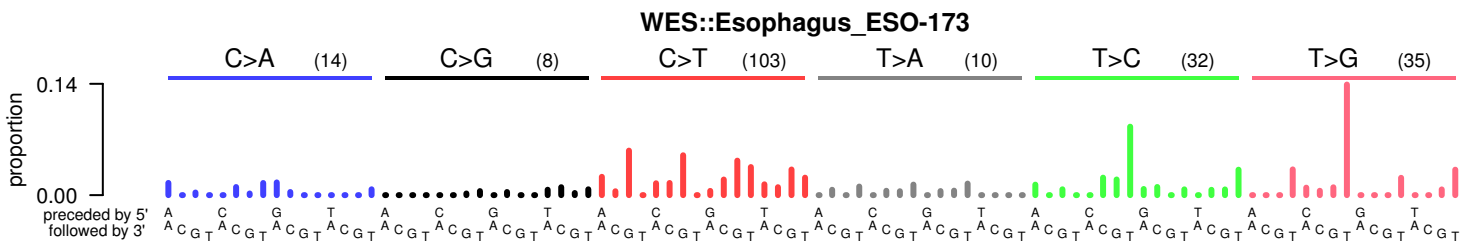
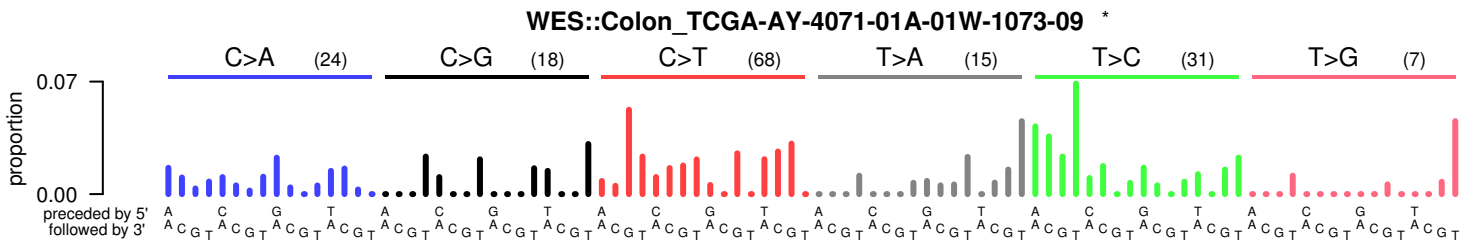
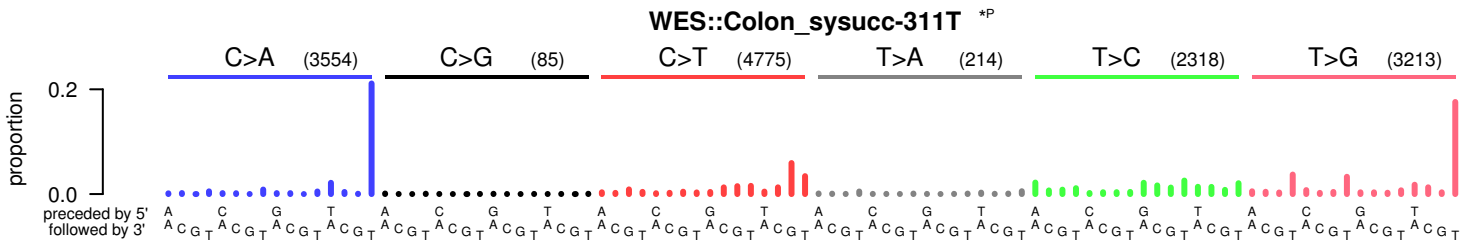
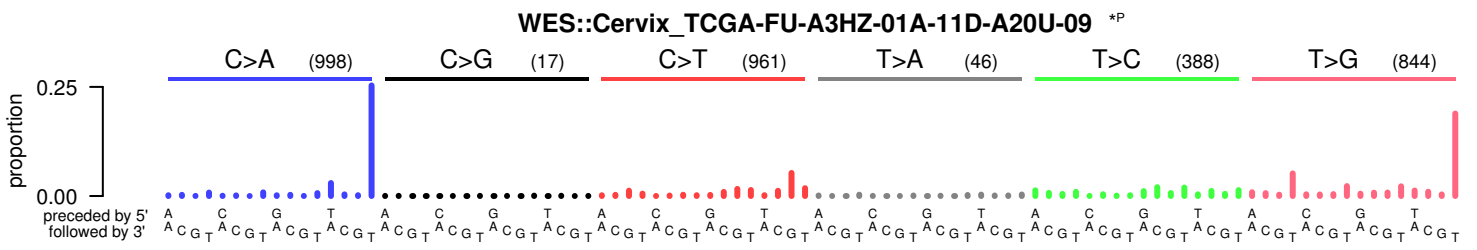
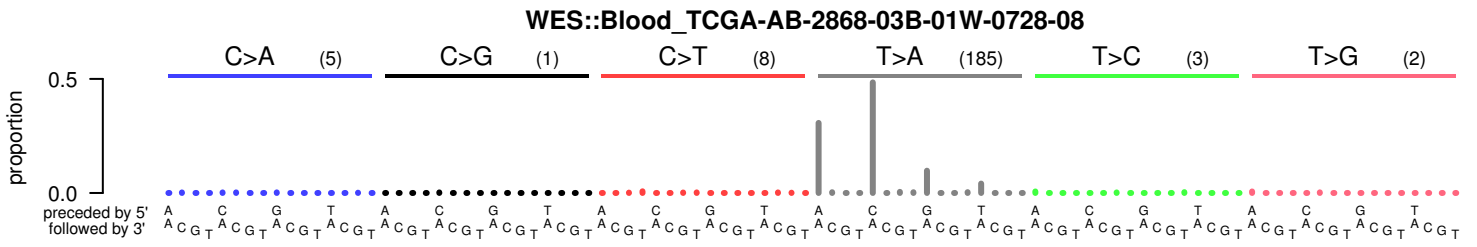
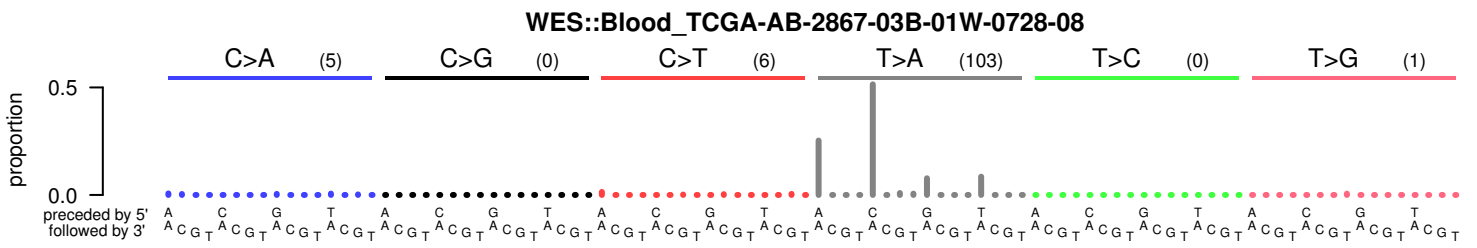
	Transcribed	Untranscribed	pval
T	68	35	0.001
TT	109	106	0.892
TTT	65	77	0.356
TTTT	148	198	0.008
TTTTT	69	59	0.426
TTTTTT	20	22	0.878
TTTTTTT	8	15	0.210
TTTTTTTT	9	5	0.424

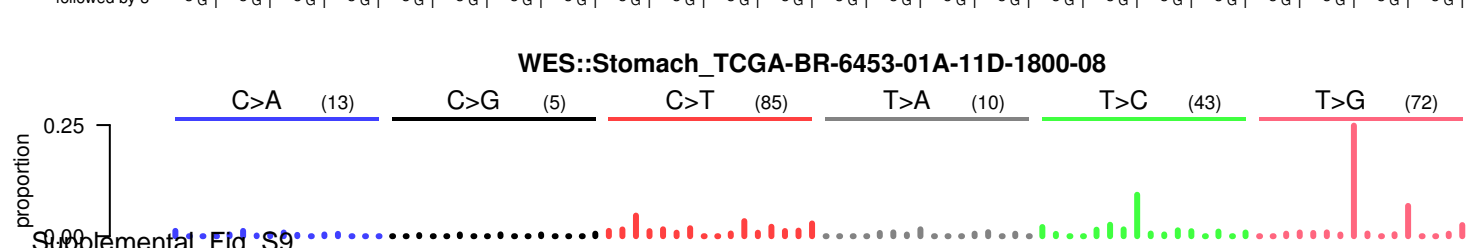
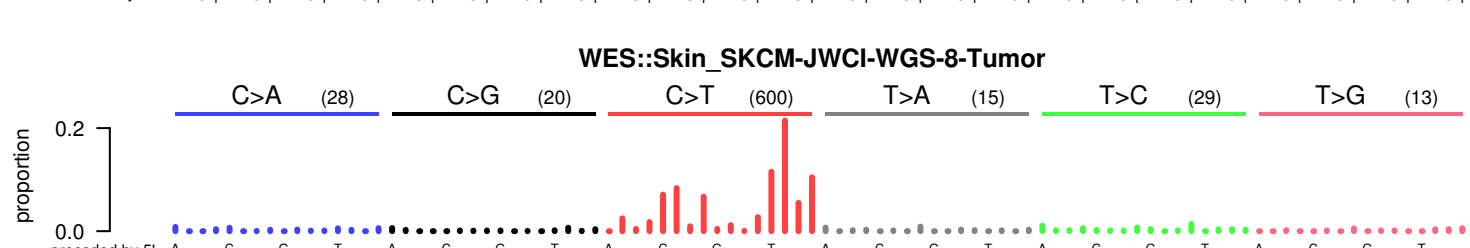
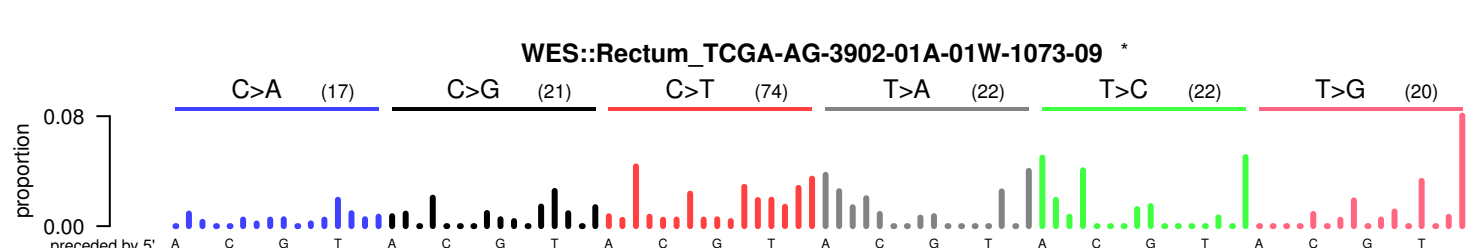
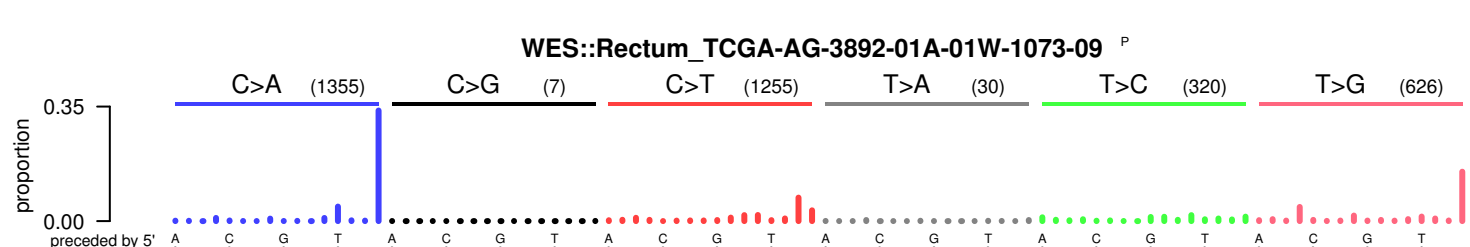
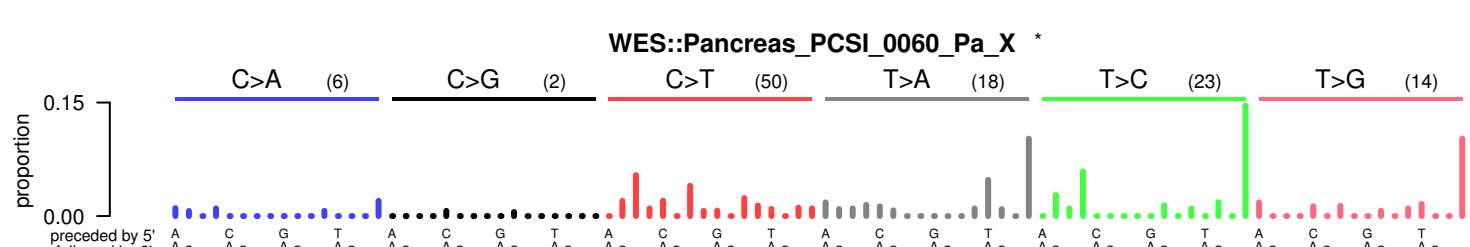
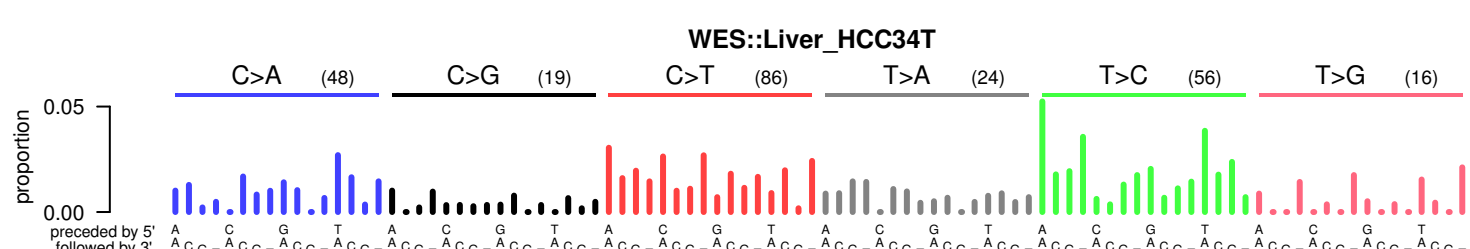
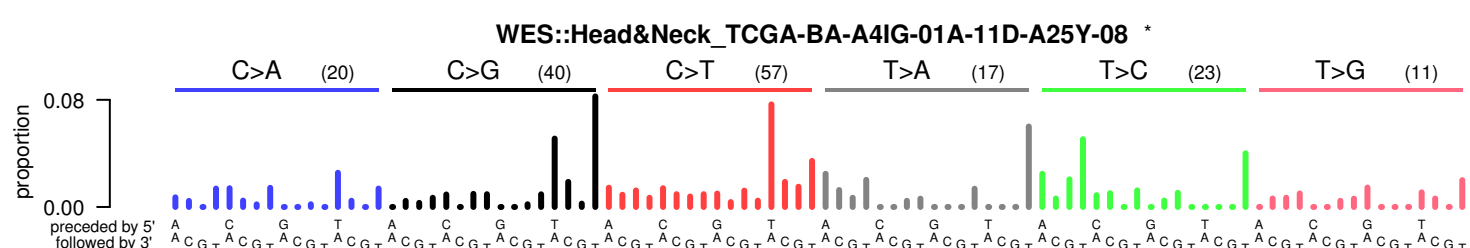
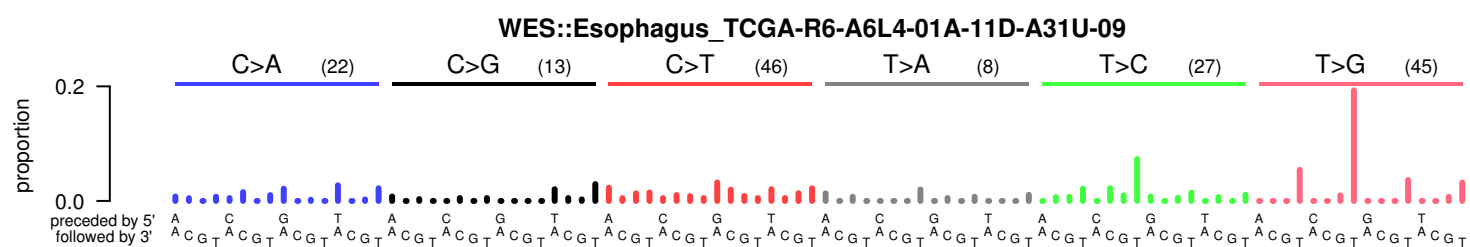
Supplemental\_Fig\_S8: Dinucleotide substitution spectrum of sample 62074759.



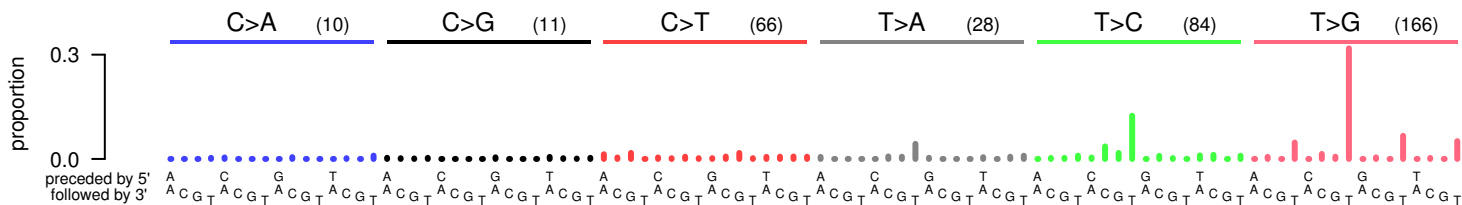
Supplemental\_Fig\_S9: Mutational spectra of tumours from publicly available sequencing data showing likely exposure to SBS\_AnT identified by examining the sequence context of thymine mutations in these samples. Samples in which SBS\_AnT was detected using mSigAct are marked with a \*, samples with SBS10a presence are marked with a <sup>P</sup>



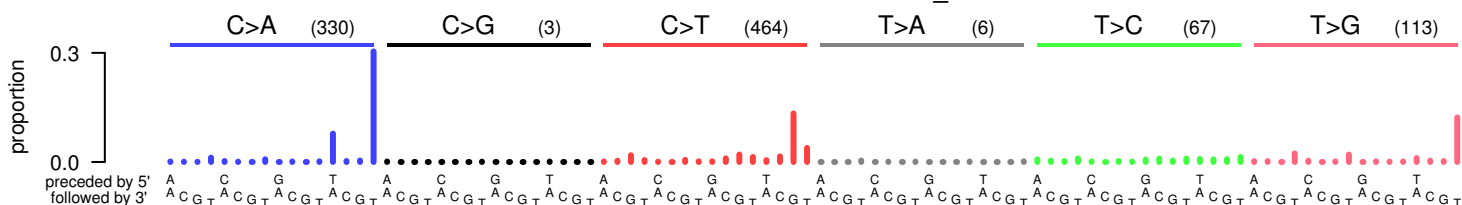




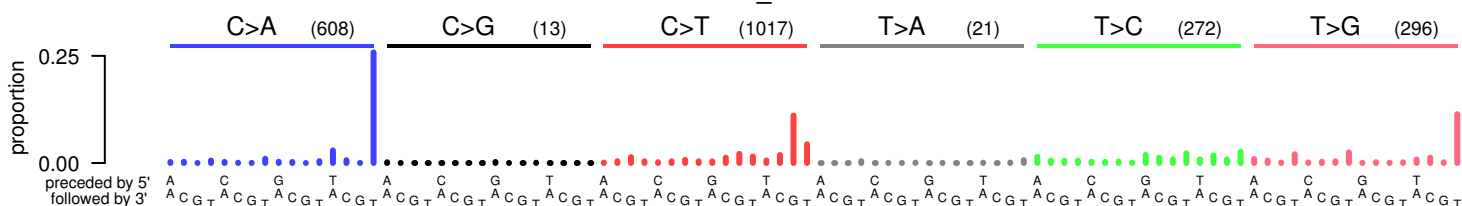
**WES::Stomach\_TCGA-D7-A4Z0-01A-22D-A25D-08**



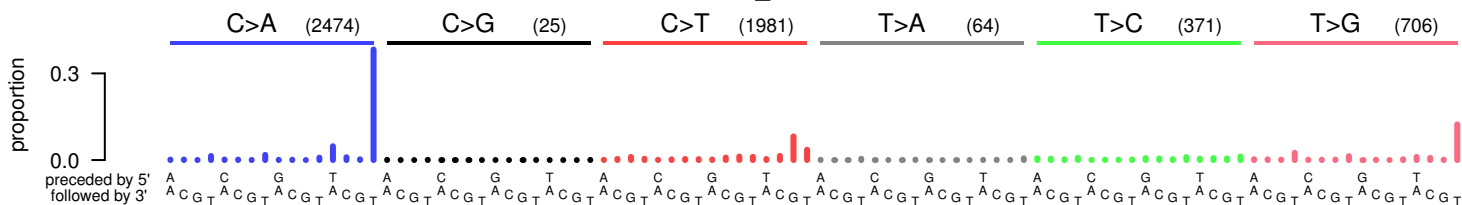
**WES::Uterus/Endometrium\_T155<sup>P</sup>**



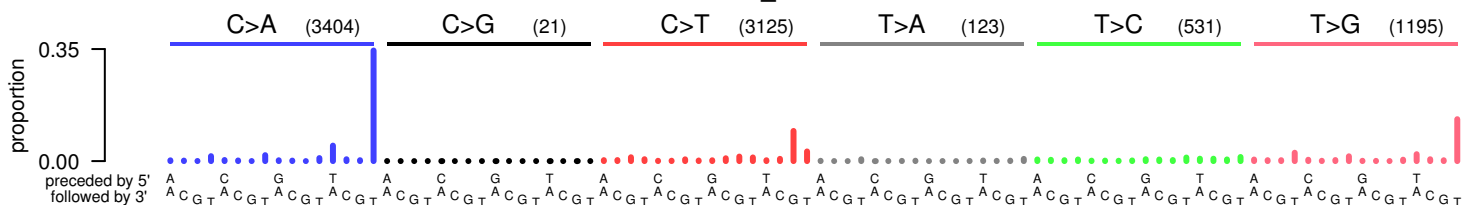
**WES::Uterus/Endometrium\_TCGA-A5-A0GP-01A-11W-A062-09<sup>P</sup>**



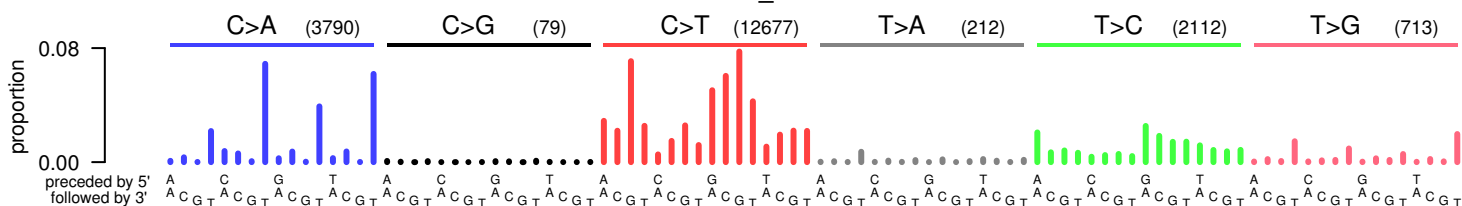
**WES::Uterus/Endometrium\_TCGA-AJ-A5DW-01A-11D-A27P-09<sup>P</sup>**



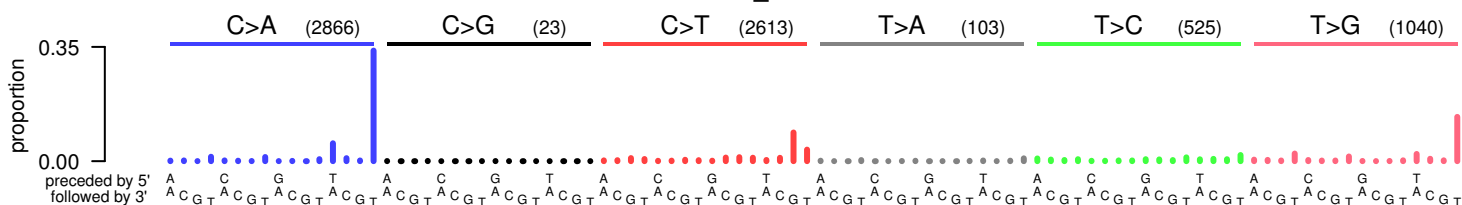
**WES::Uterus/Endometrium\_TCGA-AP-A1E0-01A-11D-A135-09<sup>P</sup>**



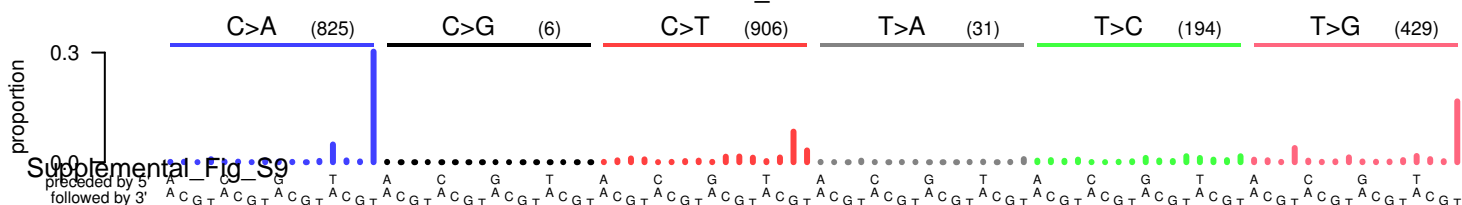
**WES::Uterus/Endometrium\_TCGA-AX-A1CE-01A-11D-A135-09**



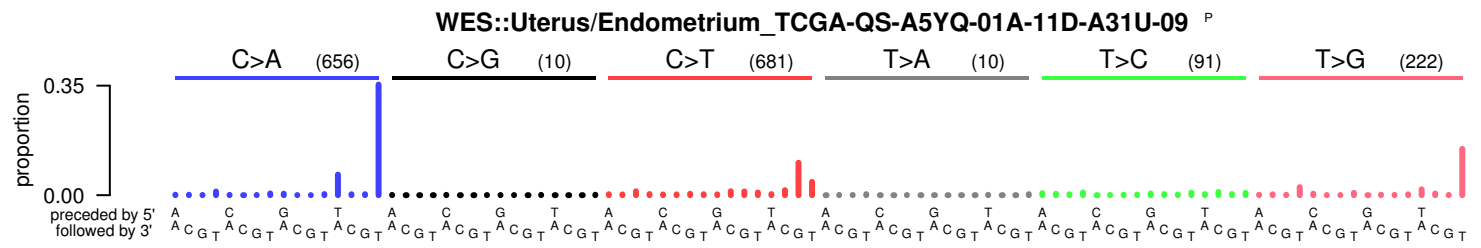
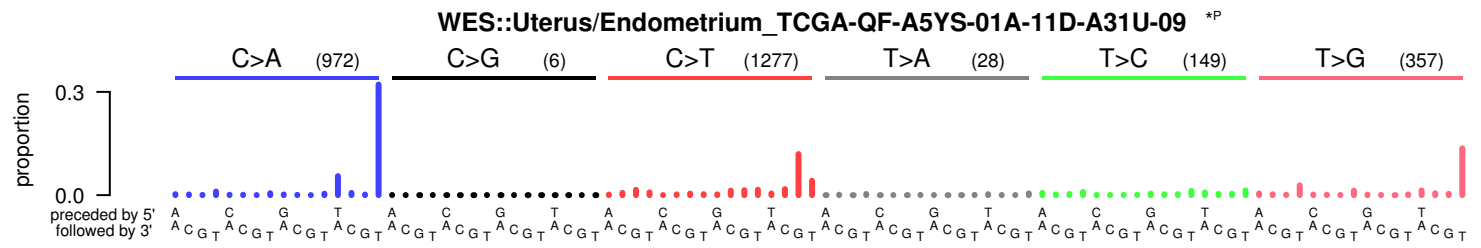
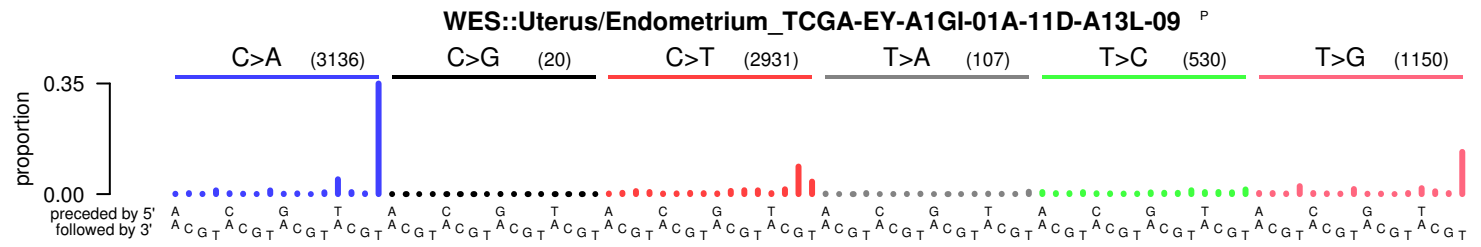
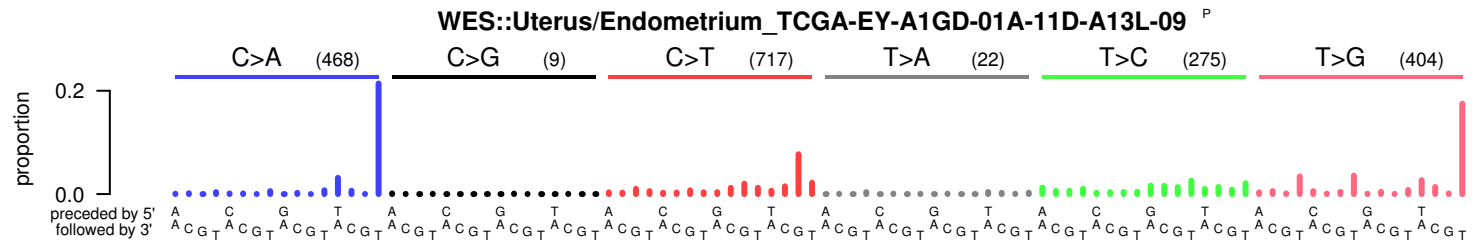
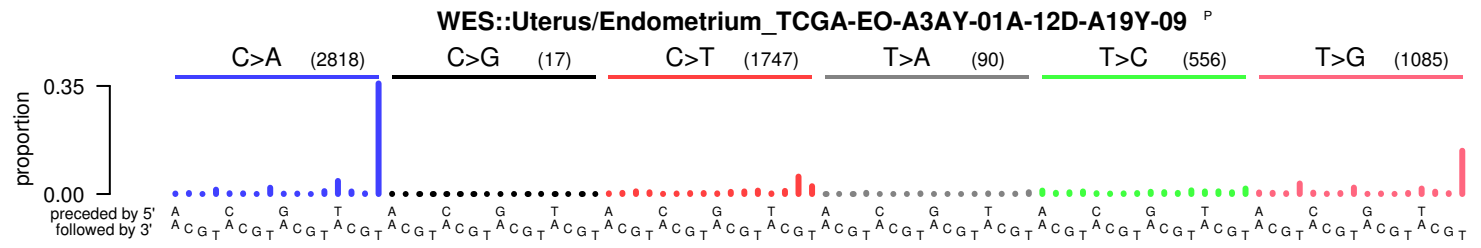
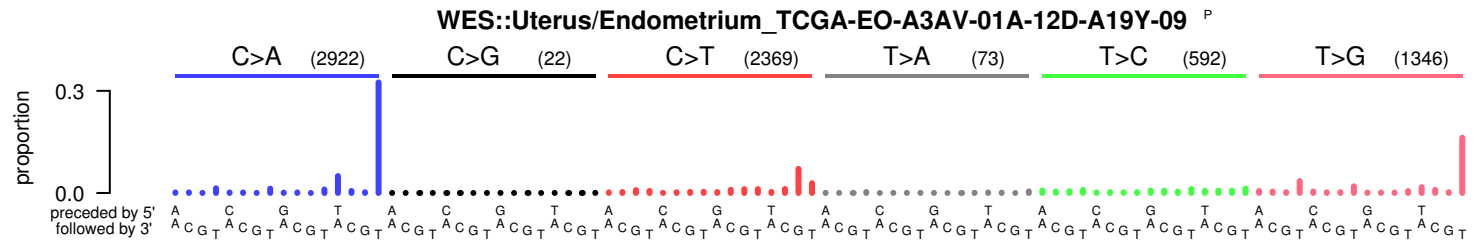
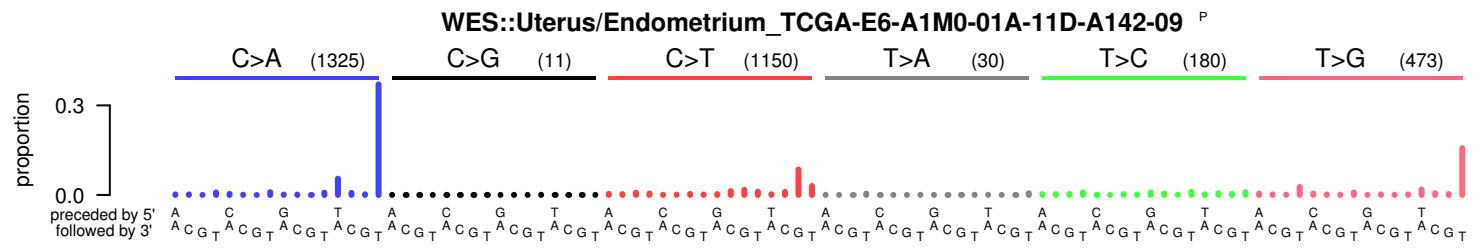
**WES::Uterus/Endometrium\_TCGA-BK-A6W3-01A-12D-A34Q-09<sup>\*P</sup>**

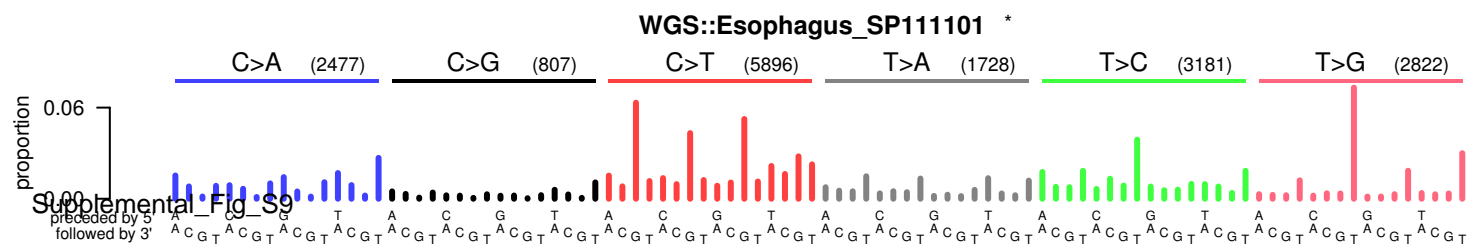
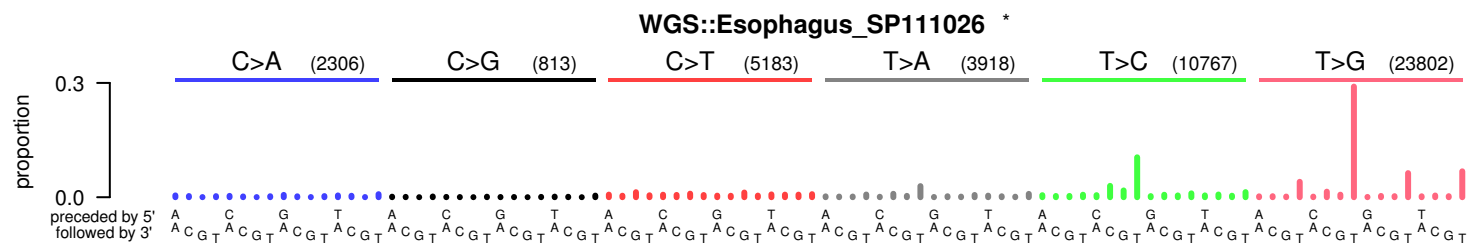
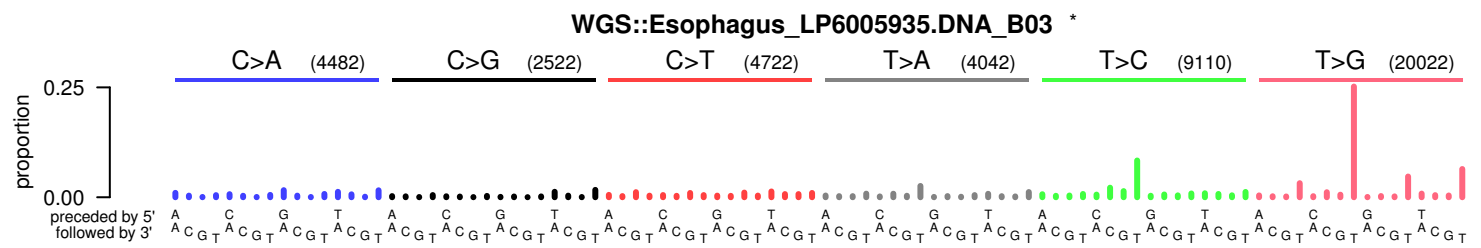
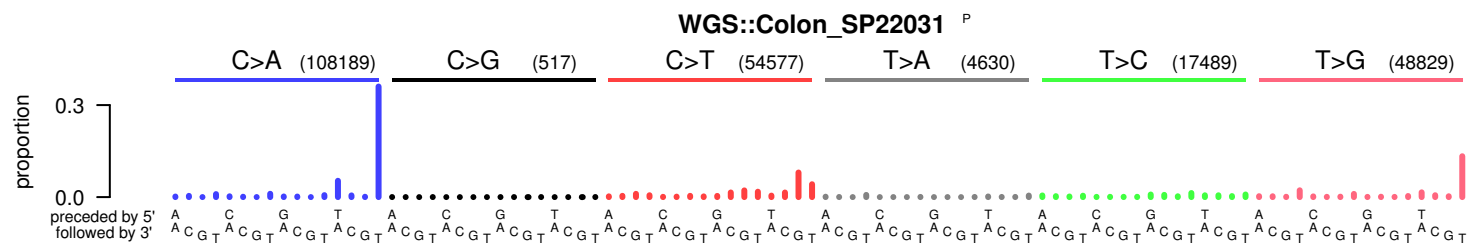
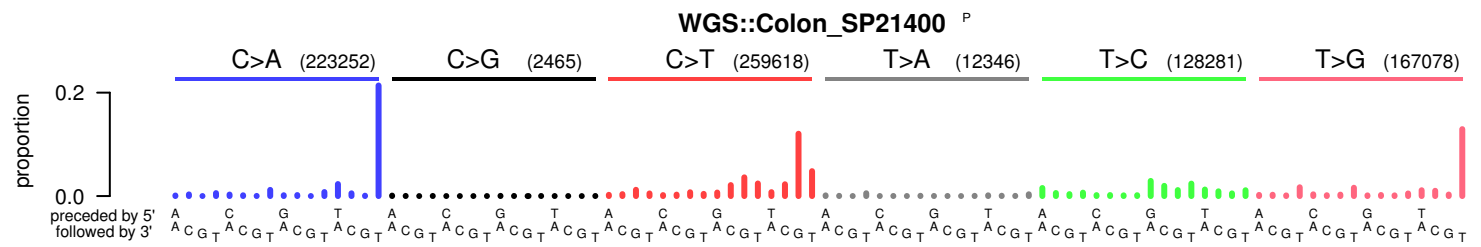
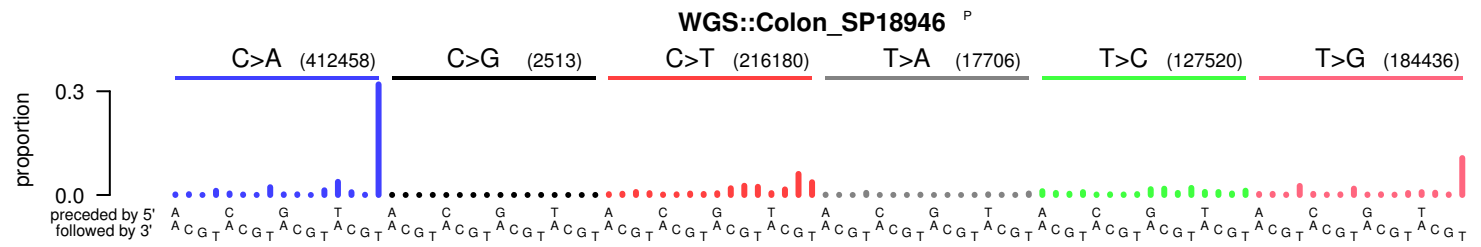


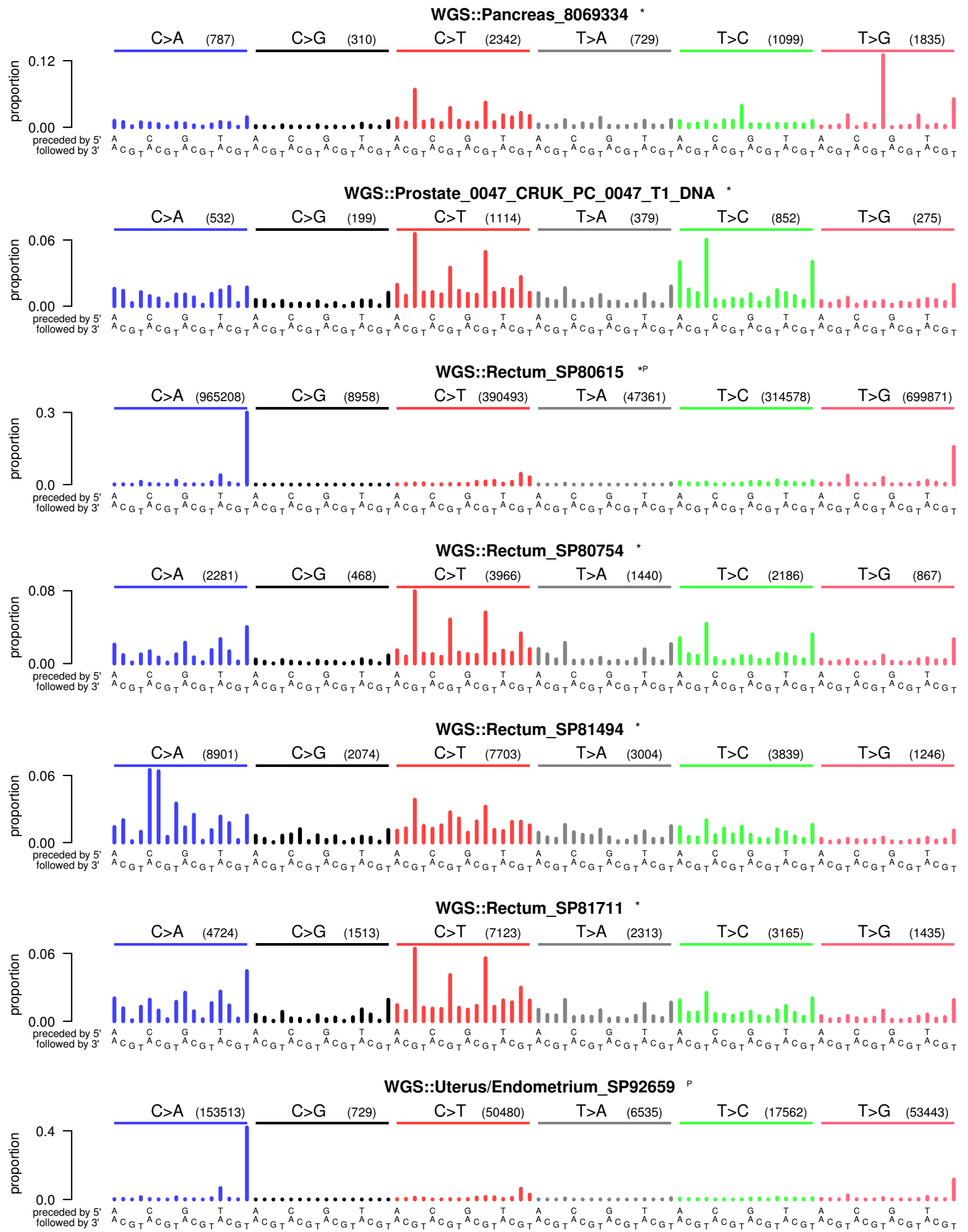
**WES::Uterus/Endometrium\_TCGA-DF-A2KV-01A-11D-A17W-09<sup>\*P</sup>**



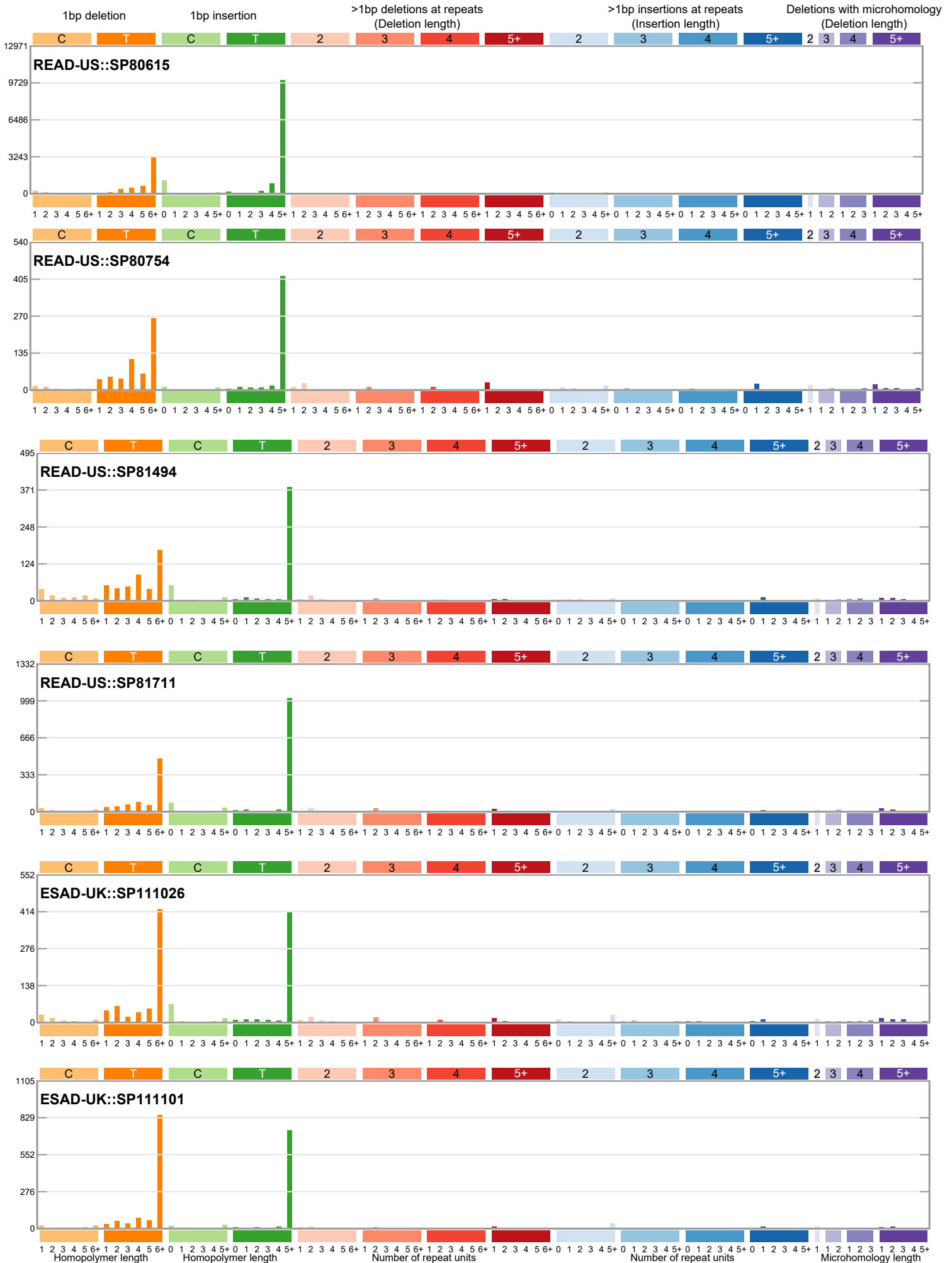




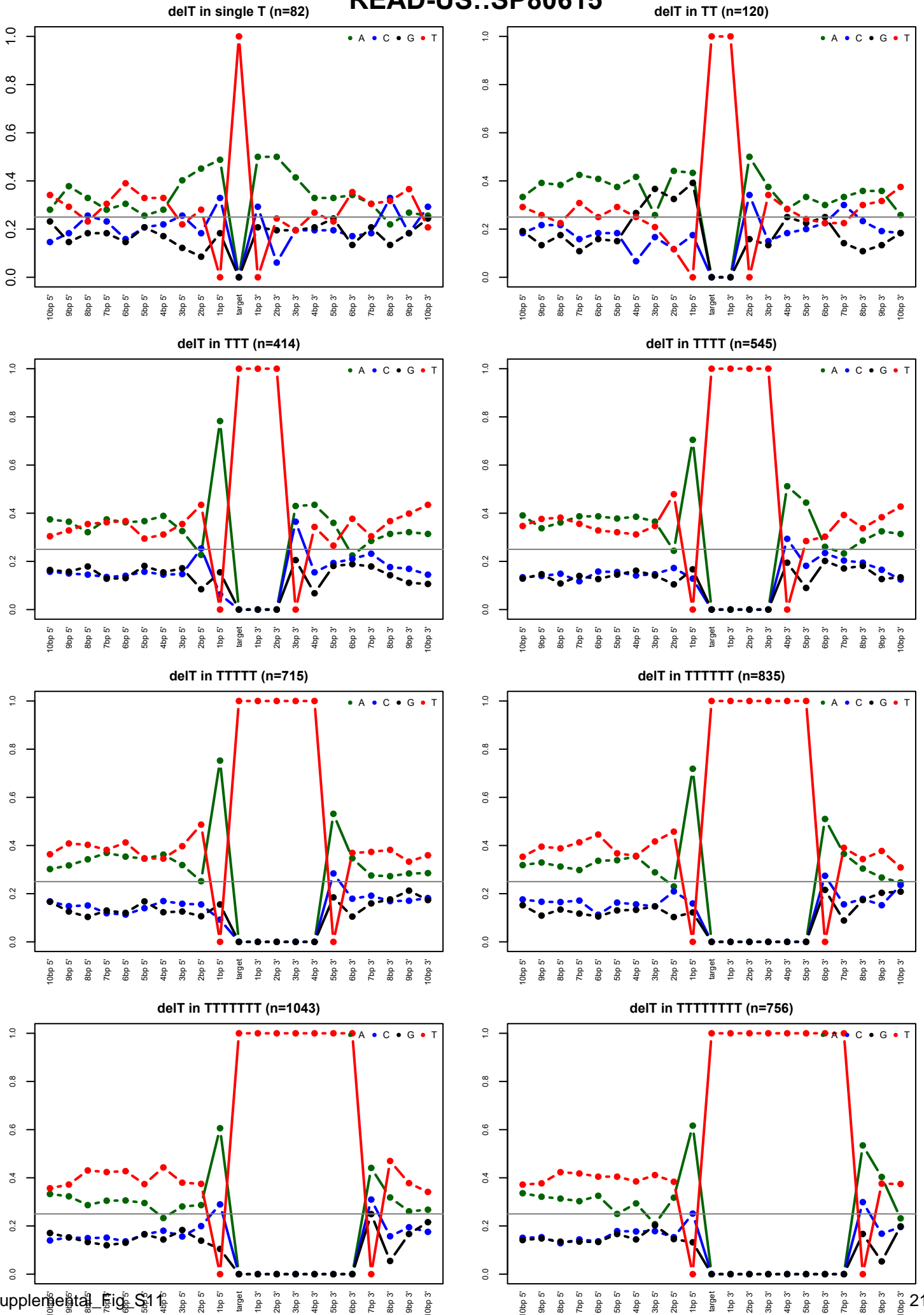




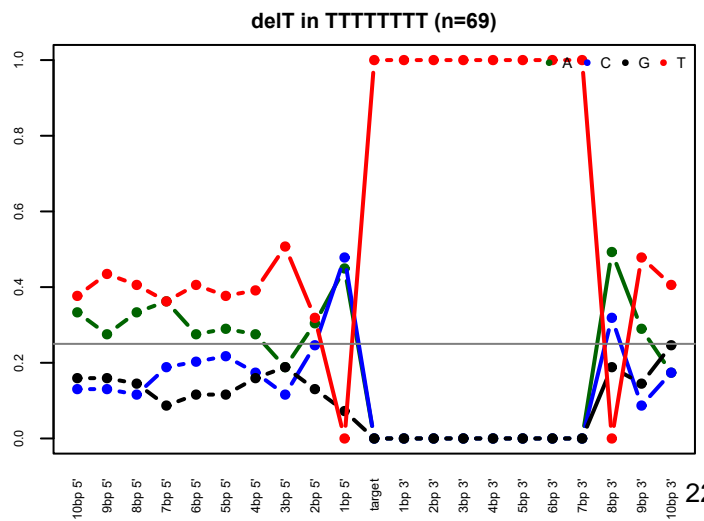
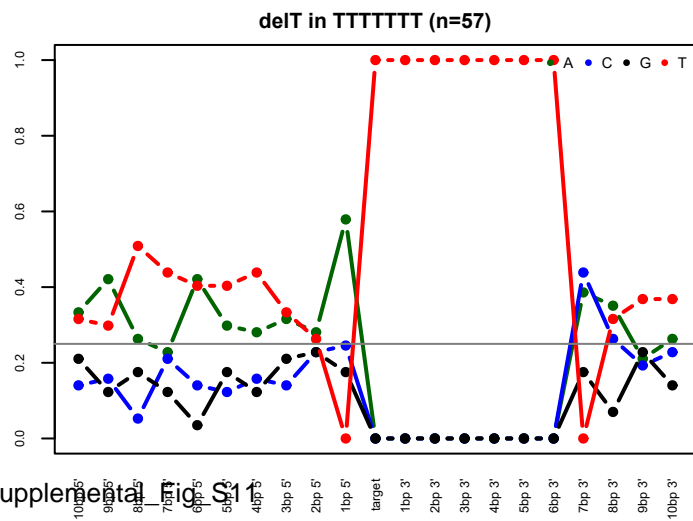
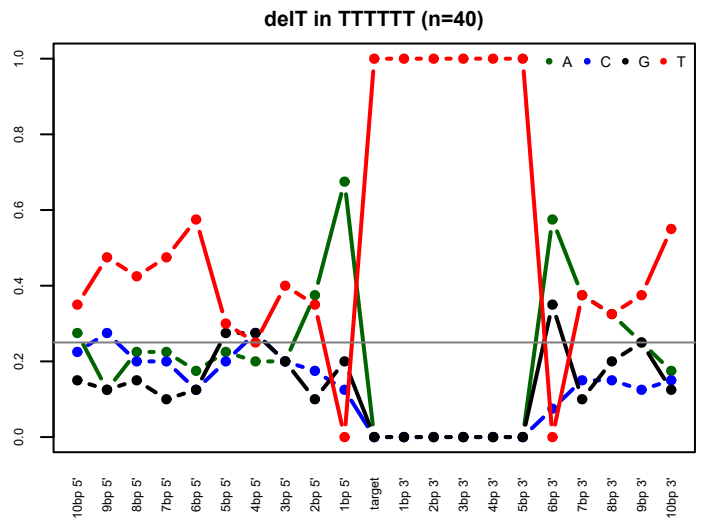
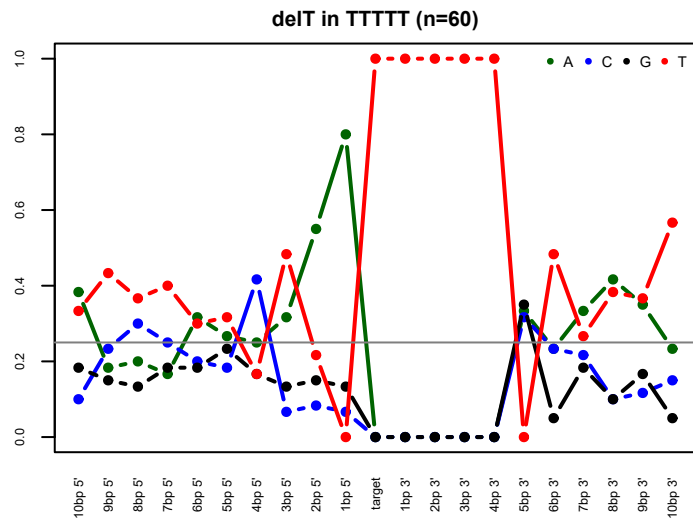
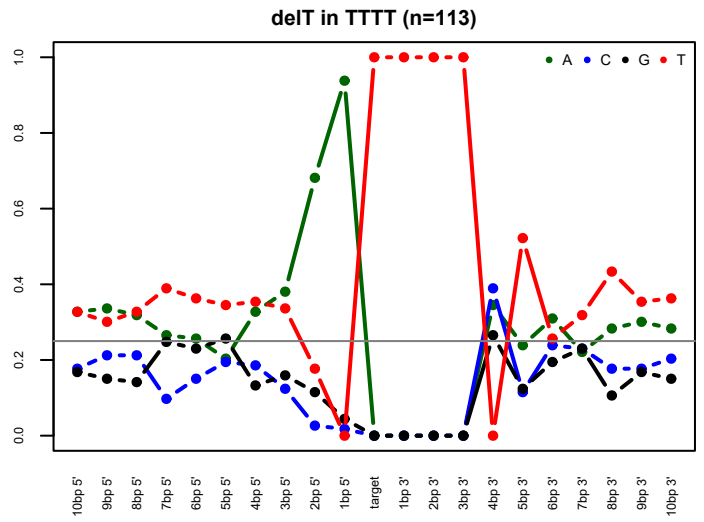
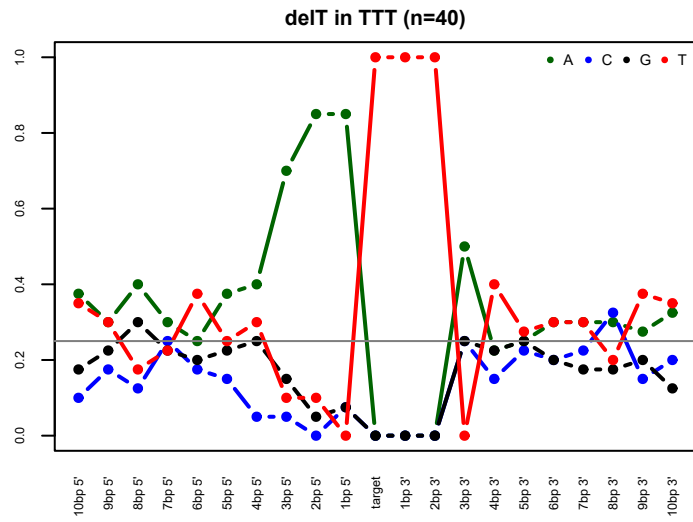
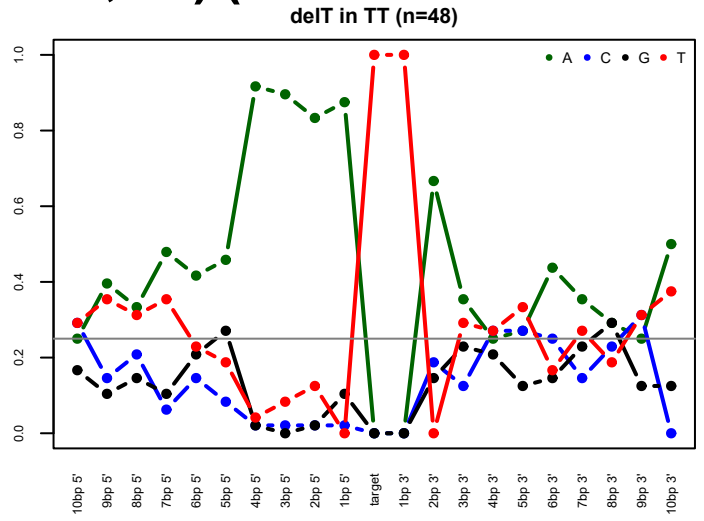
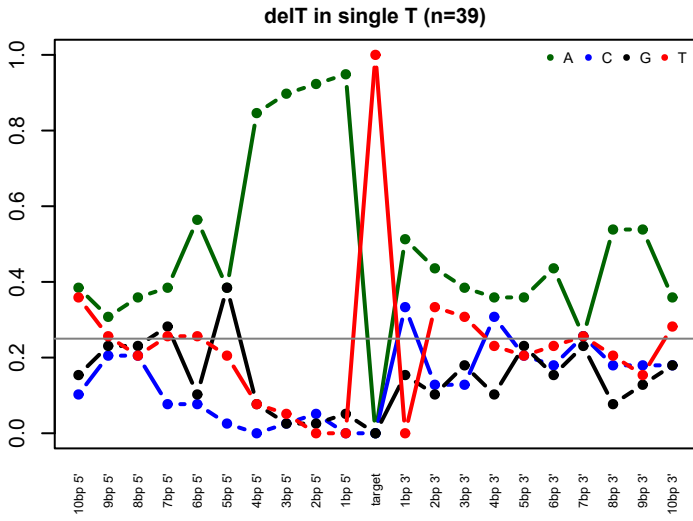
Supplemental\_Fig\_S10: Indel spectra for SBS\_A<sup>n</sup>T positive PCAWG tumours. In addition to the A<sup>n</sup>T indel signature, these tumours also show clear evidence of signatures ID1 and ID2 (Alexandrov et al. 2019).



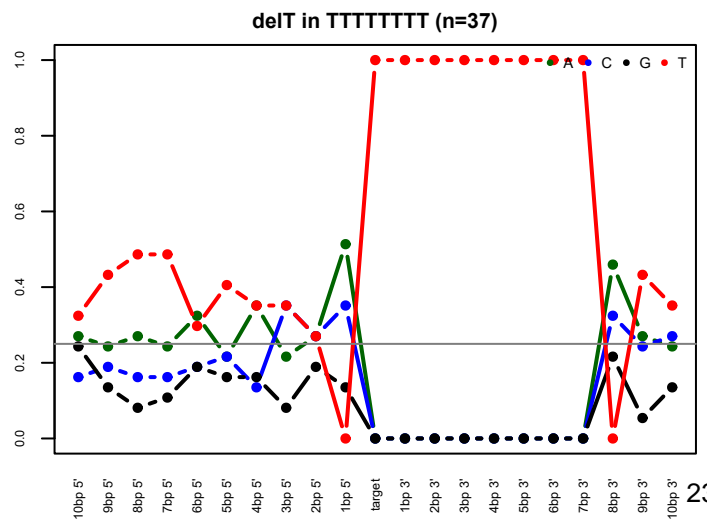
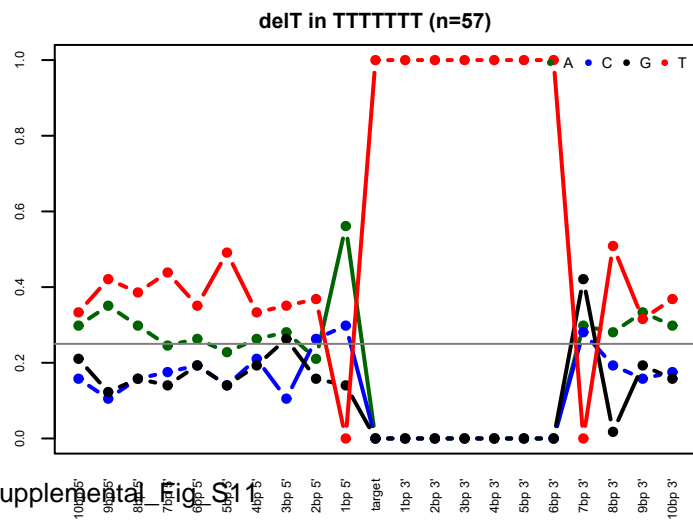
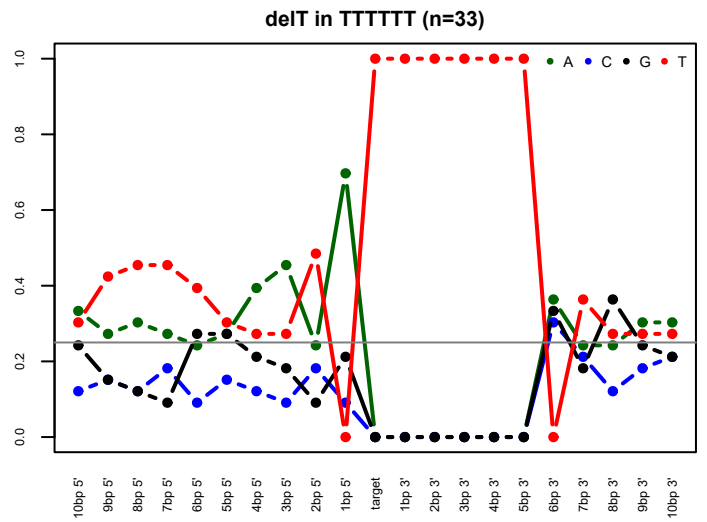
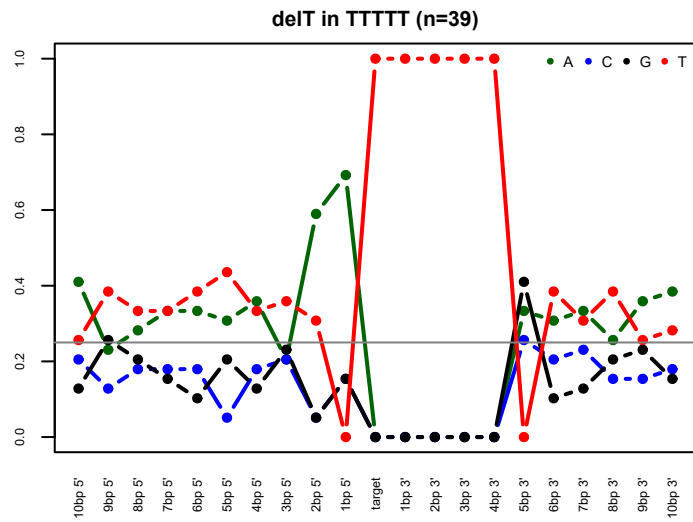
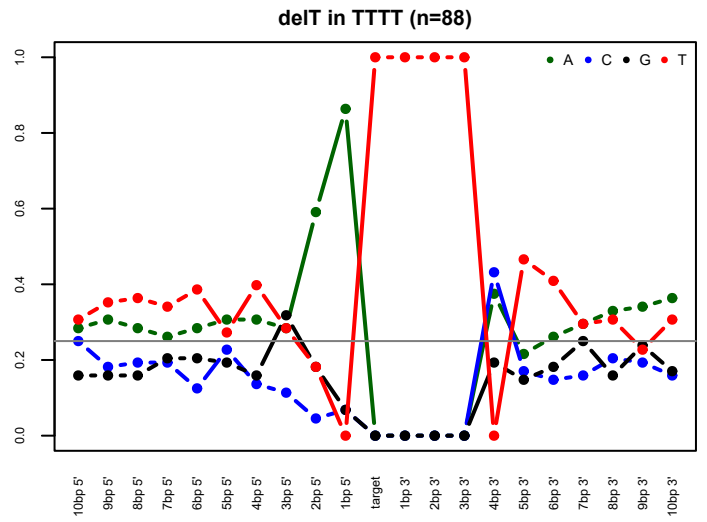
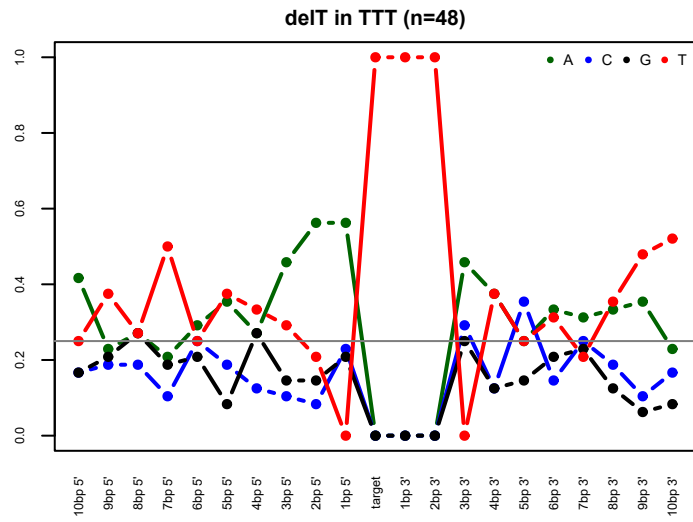
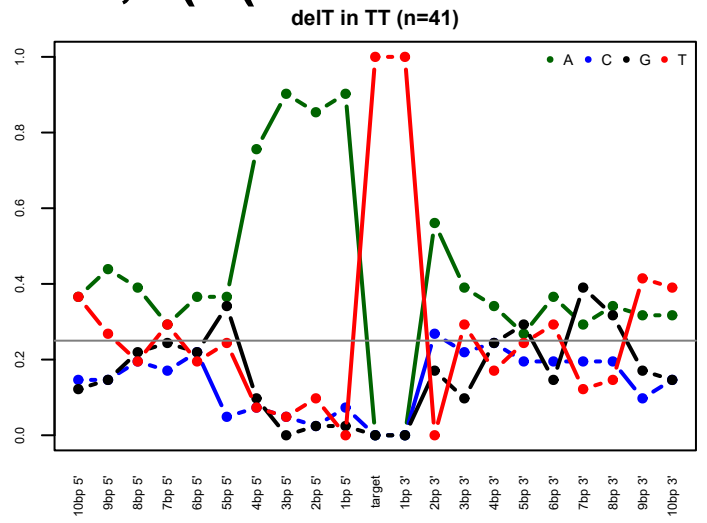
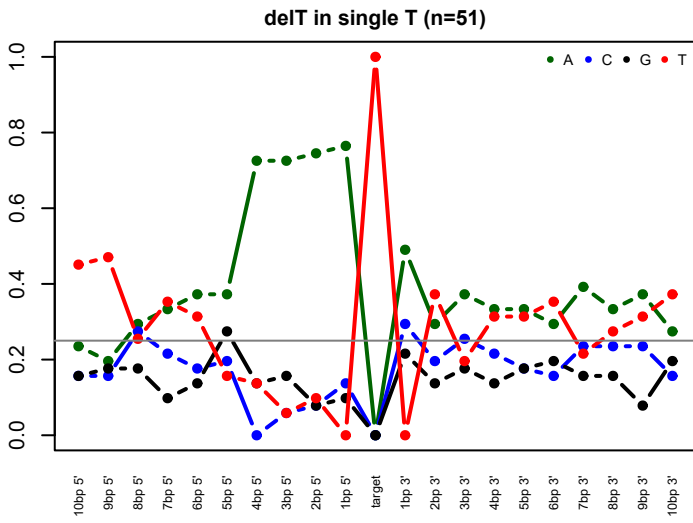
## READ-US::SP80615



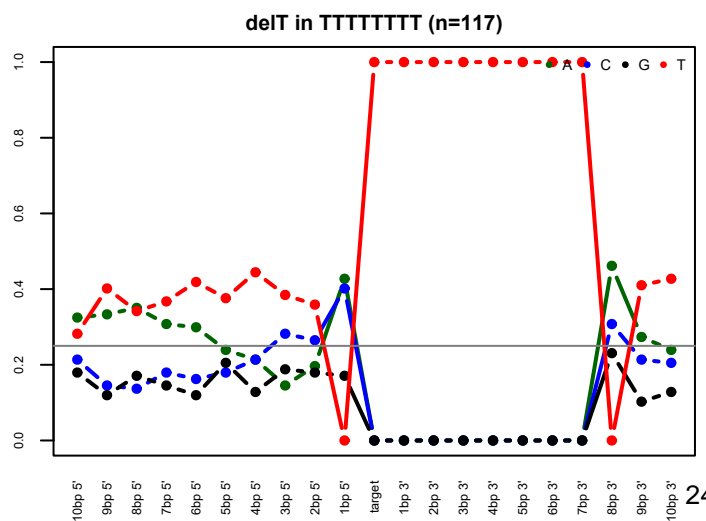
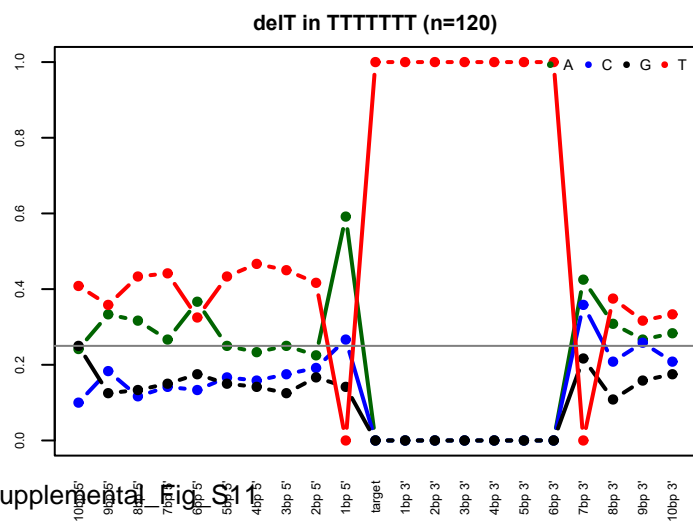
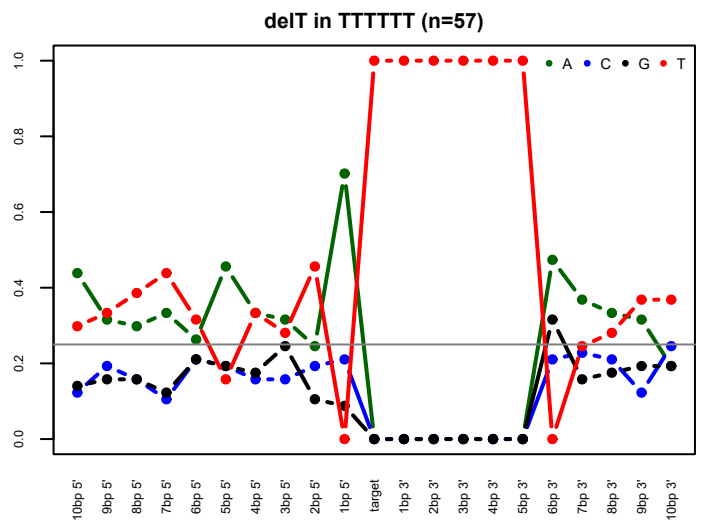
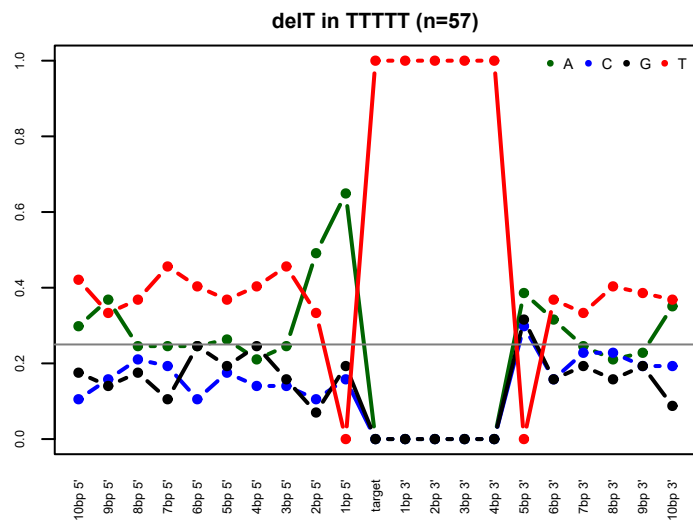
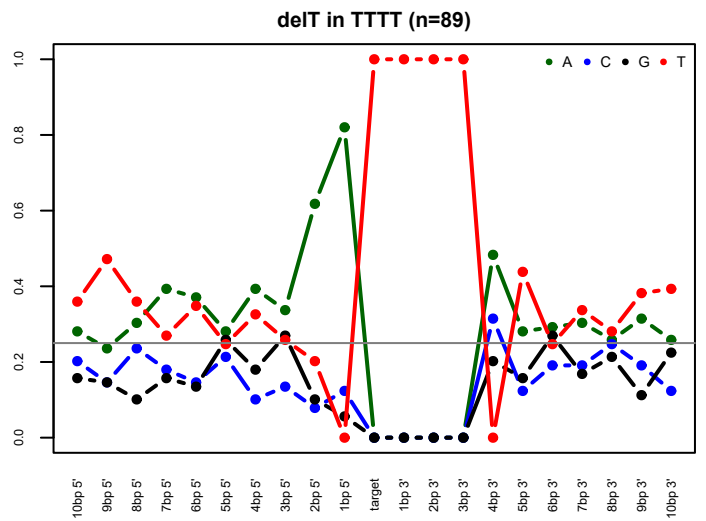
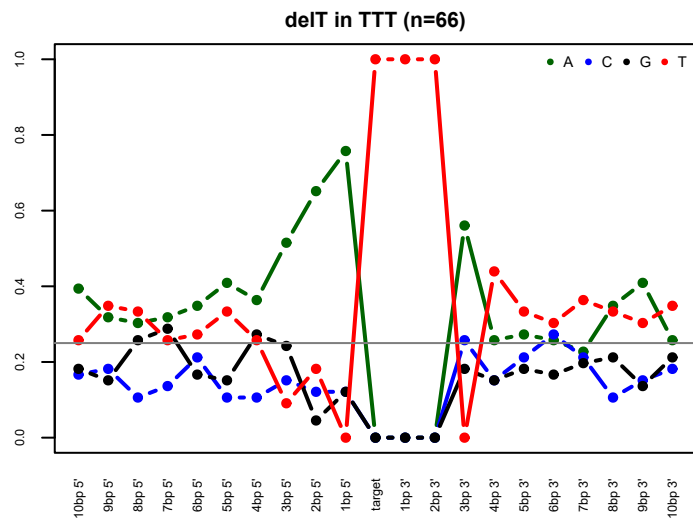
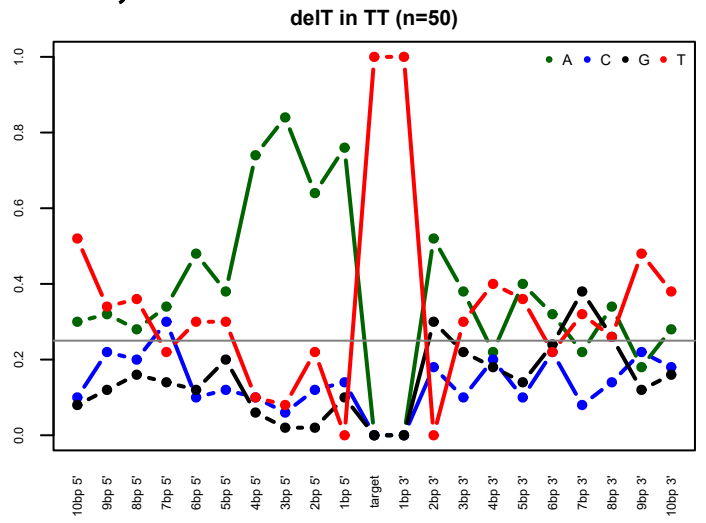
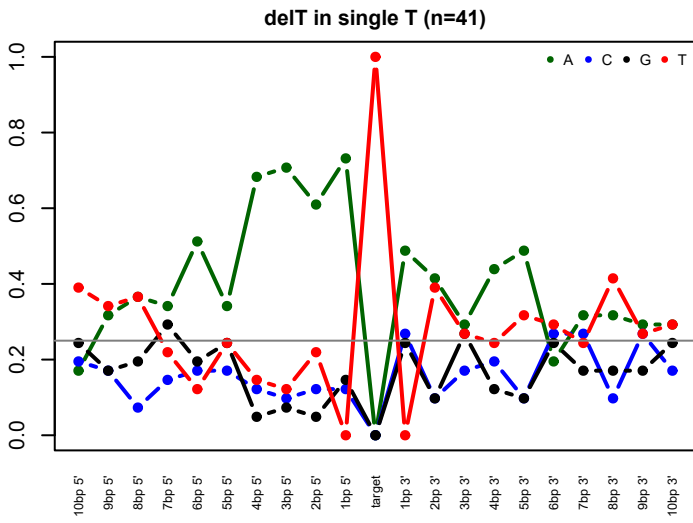
# READ-US::GD, \$(+) (



# READ-US::GD, % - (

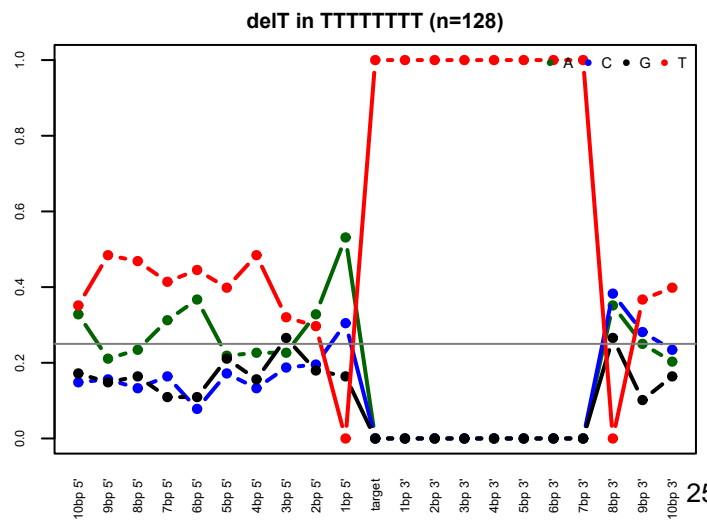
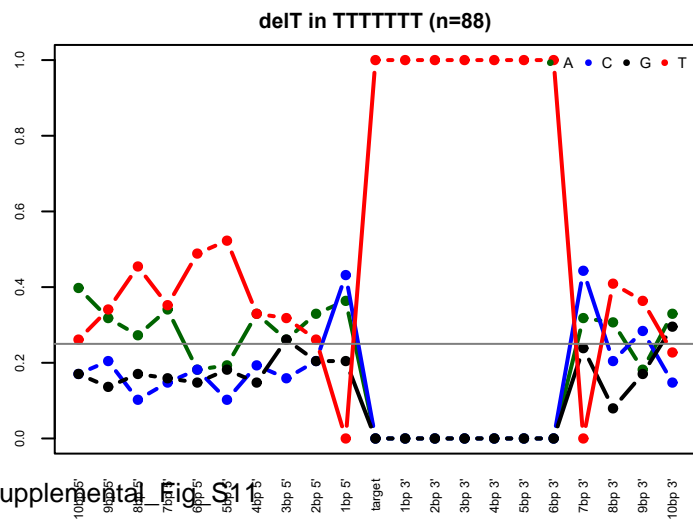
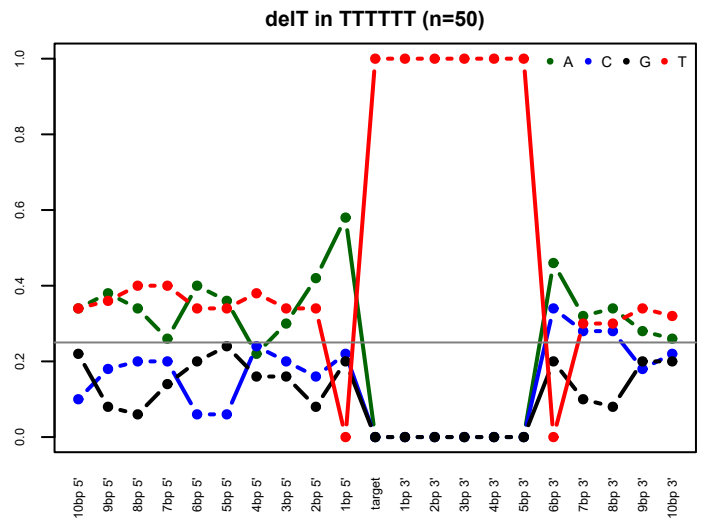
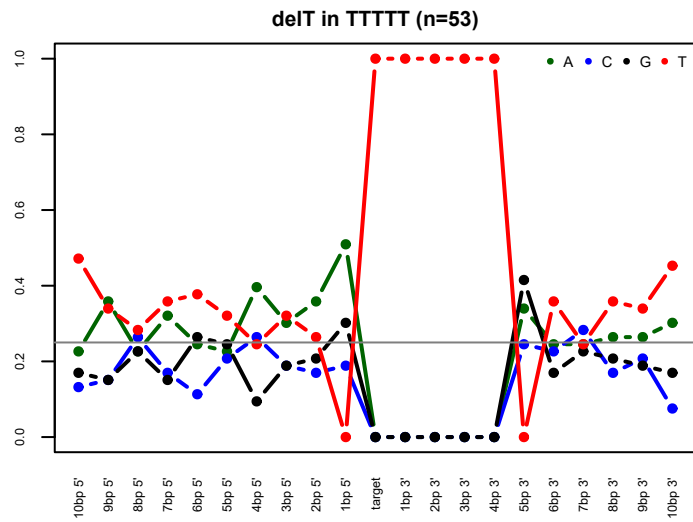
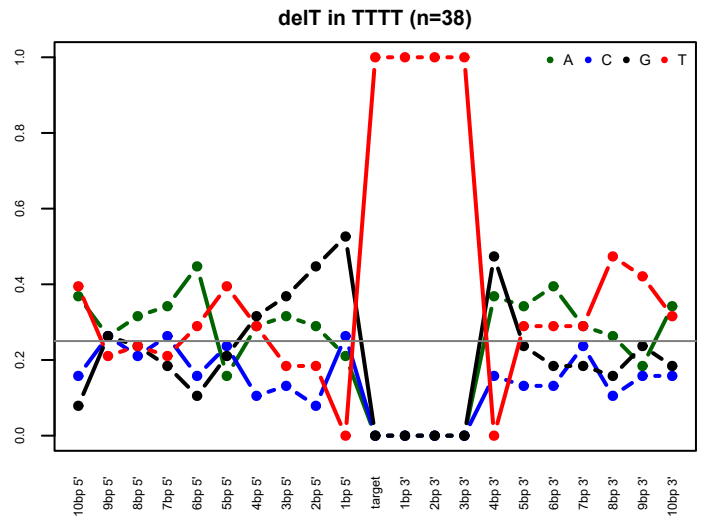
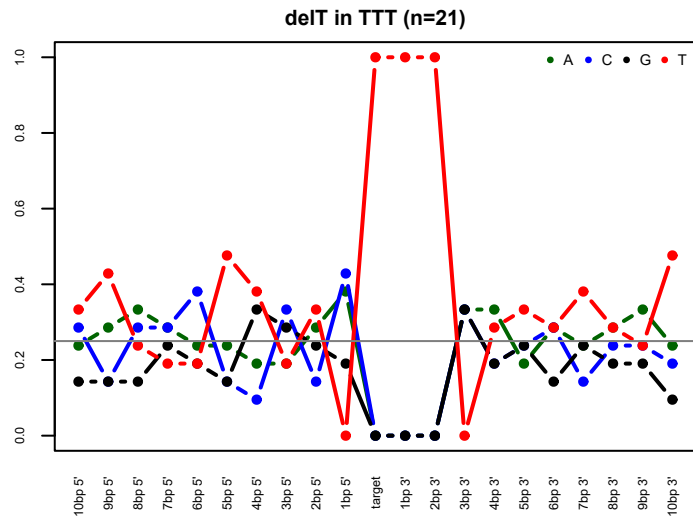
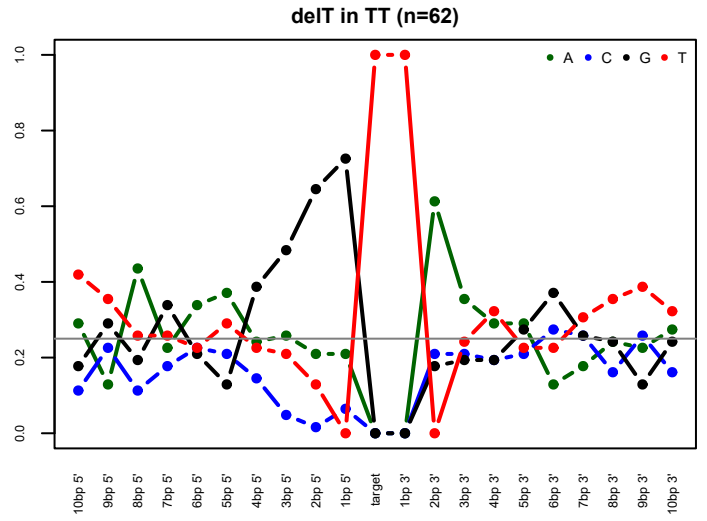
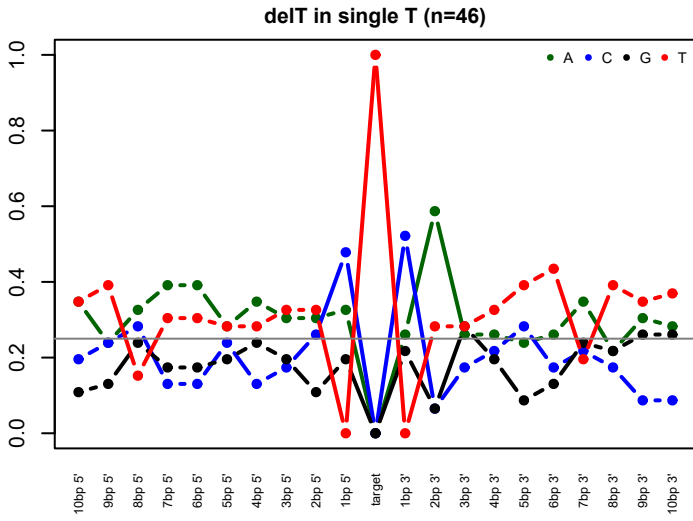


# READ-US::GD, %&#x2013;%&#x2013;

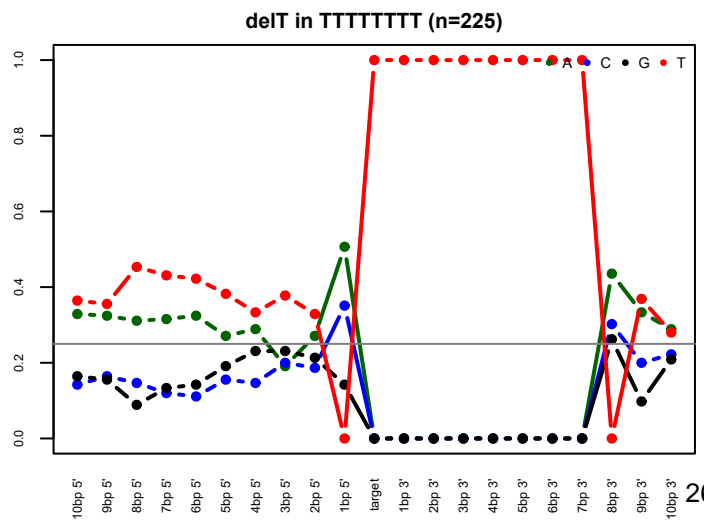
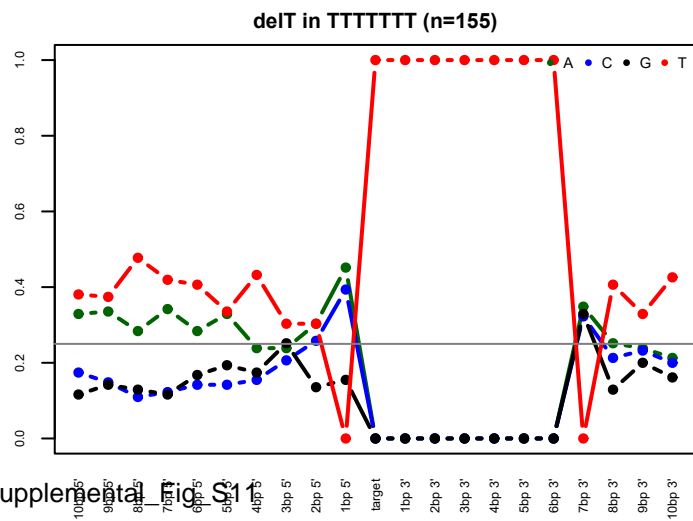
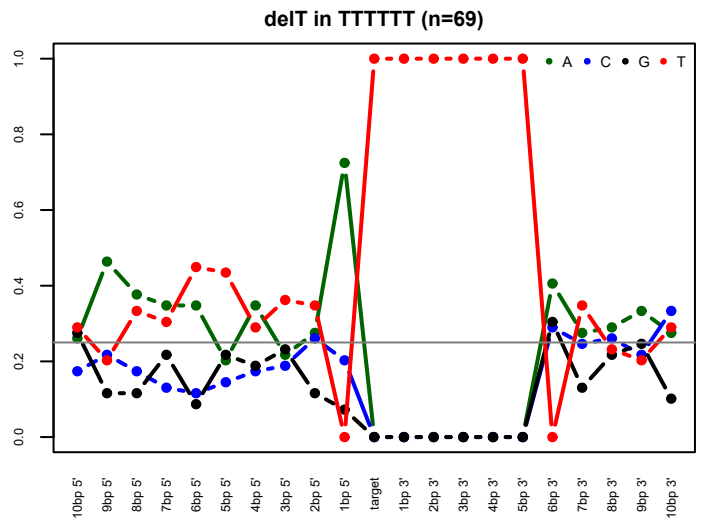
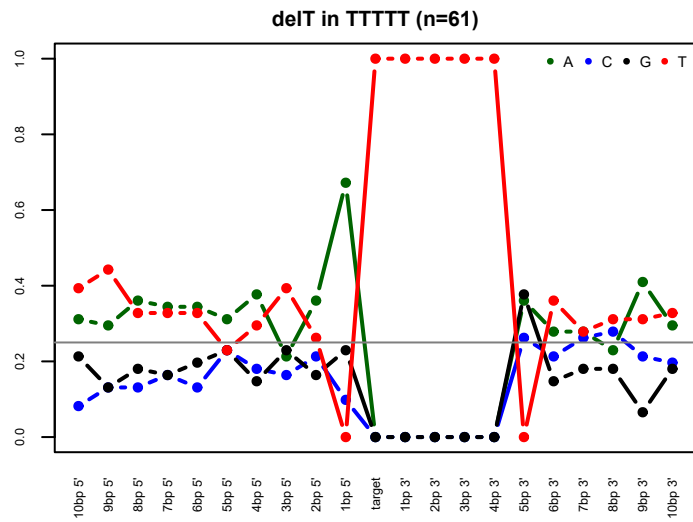
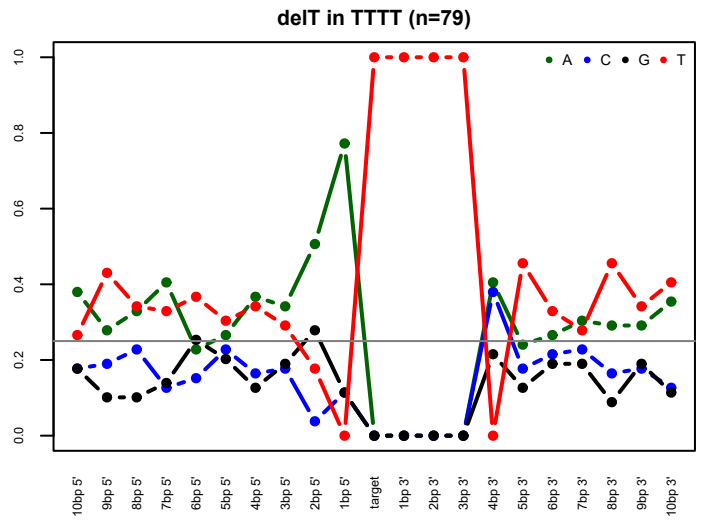
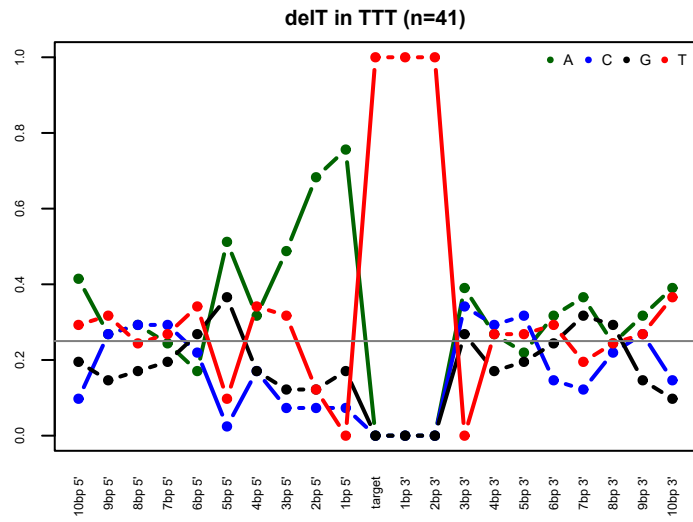
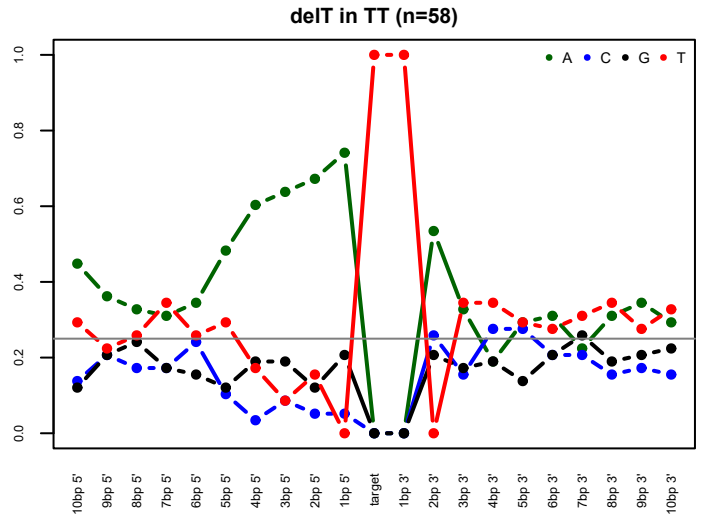
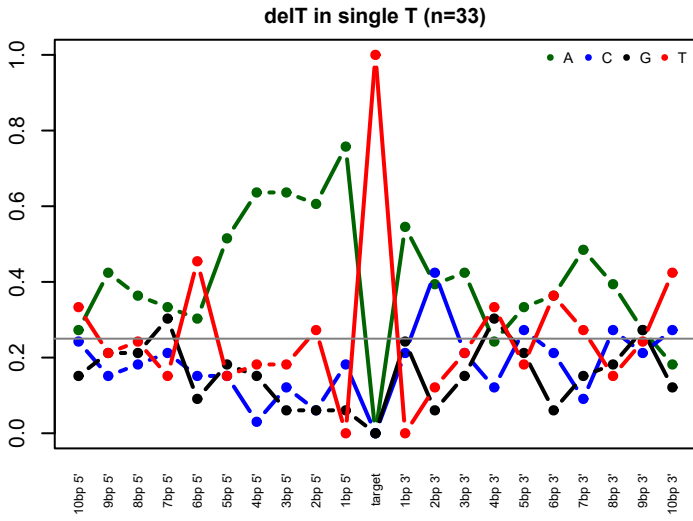




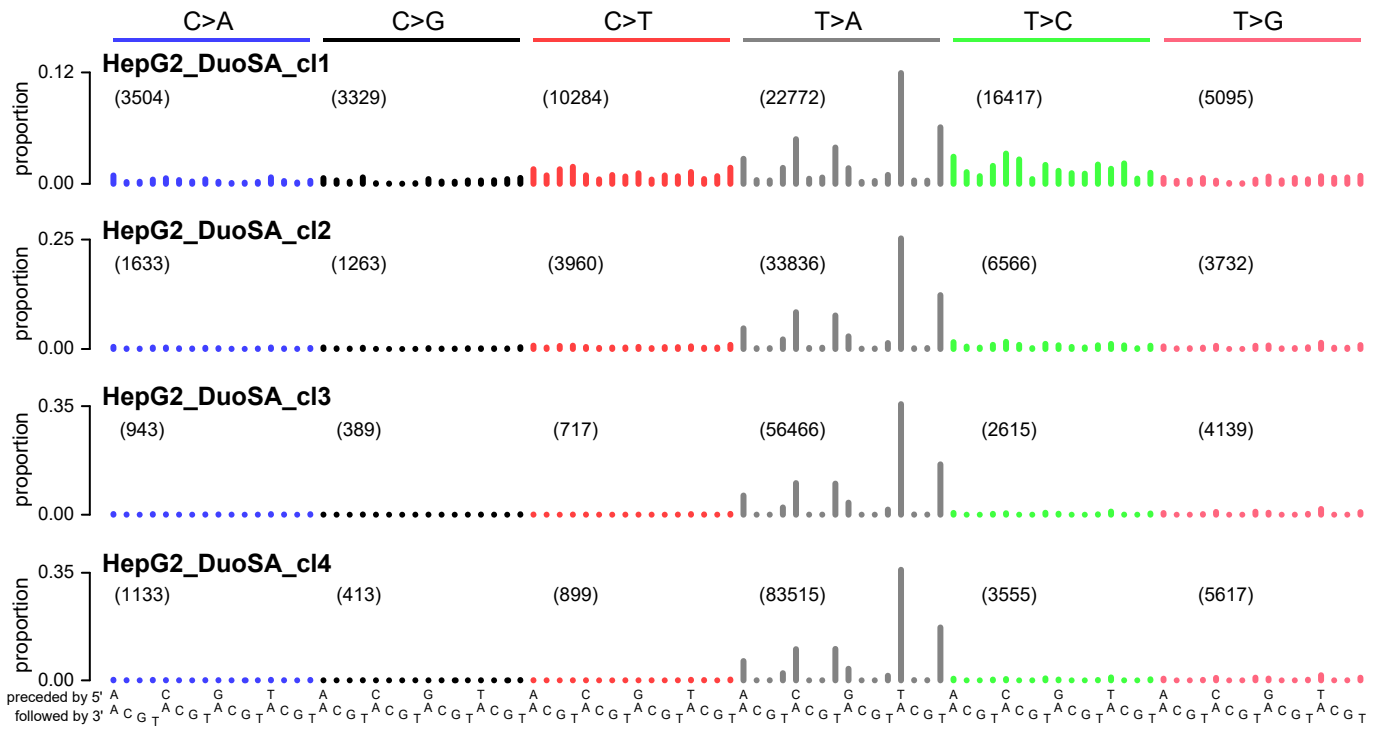
# ESAD-UK::GD%/%/\$&\*



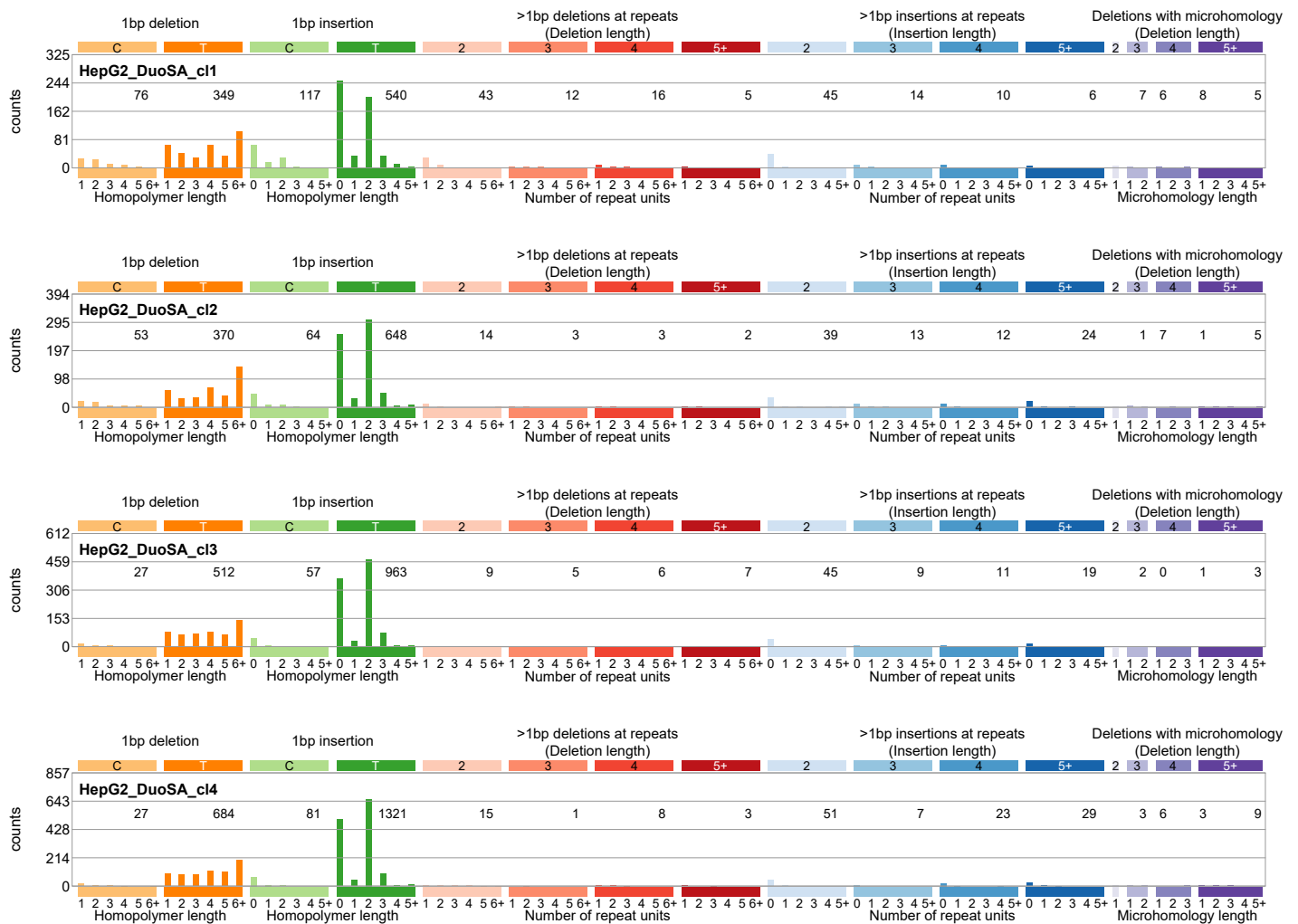
# ESAD-UK::GD%/%/%\$%



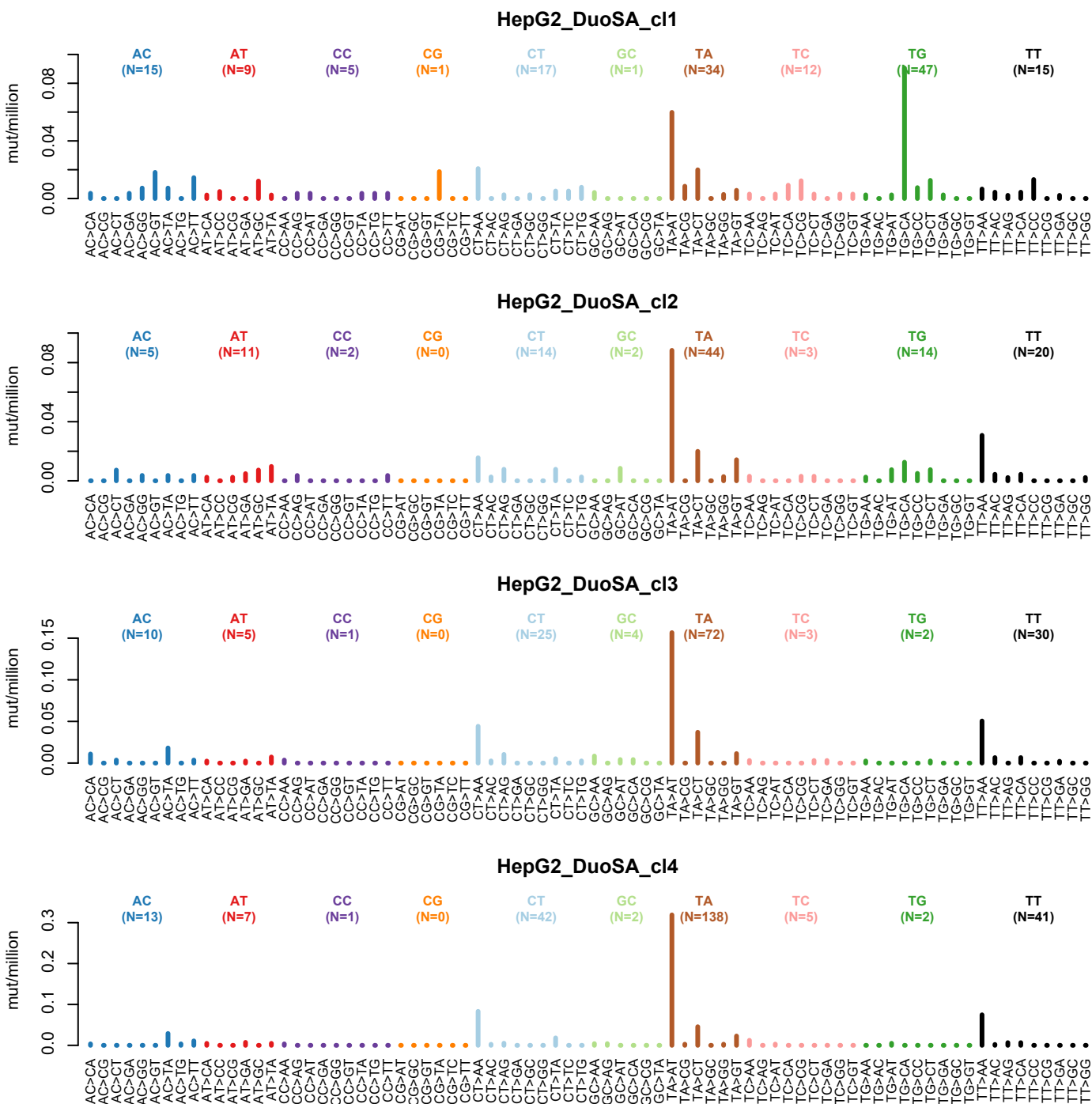
Supplemental\_Fig\_S12: **SNS mutational spectra of duocarmycin SA treated HepG2 clones.** Clones 1 and 2 were exposed to 100pM duocarmycin SA, clones 3 and 4 to 250pM.



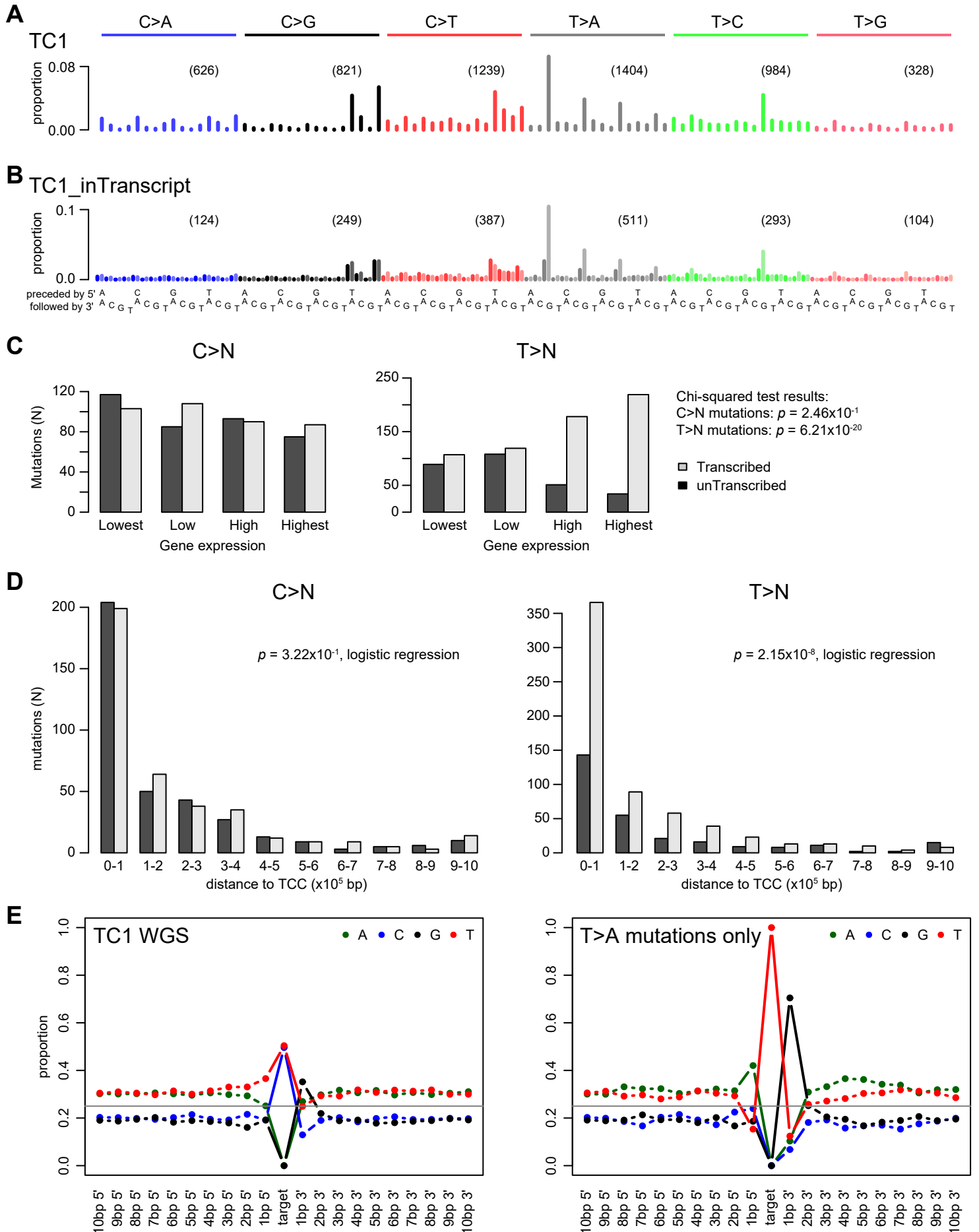
Supplemental\_Fig\_S13: Indel mutational spectra of duocarmycin SA treated HepG2 clones. Clones 1 and 2 were exposed to 100pM duocarmycin SA, clones 3 and 4 to 250pM.



Supplemental\_Fig\_S14: **Double base substitution spectrum of Duocarmycin SA treated HepG2 clones.** Clones 1 and 2 were treated with 100pM duocarmycin SA, clones 3 and 4 were treated with 250pM duocarmycin SA.



Supplemental\_Fig\_S15: **Characterization of the TC1 mutational spectrum.** (A) Whole-genome sequencing mutation spectrum of TC1. (B) Visualization of transcriptional strand bias in TC1 by plotting mutations on the transcribed (light) and untranscribed strand (dark) side by side. (C) Transcriptional strand bias as a function of gene expression for C>N (left) and T>N mutations (right). (D) Transcriptional strand bias as a function of distance to the transcription start site for C>N (left) and T>N mutations (right). (E) Extended sequence context specificity of TC1 mutations.



Supplemental\_Fig\_S16: **Identifying samples whose mutational spectrum resembles that of TC1.**

Background: Sample TC1 is not very highly mutated, and as a consequence we cannot extract a pure 'TC1 mutational signature' from the whole-genome sequencing data which could be used for mSigAct or sigProfiler analysis for presence of this signature in the publicly available sequencing data. Therefore we screened the publicly available sequencing data for tumors with presence of the same mutational processes by looking for tumors in which either the thymine mutations, or the T>A mutations specifically, bore high resemblance to that of TC1.

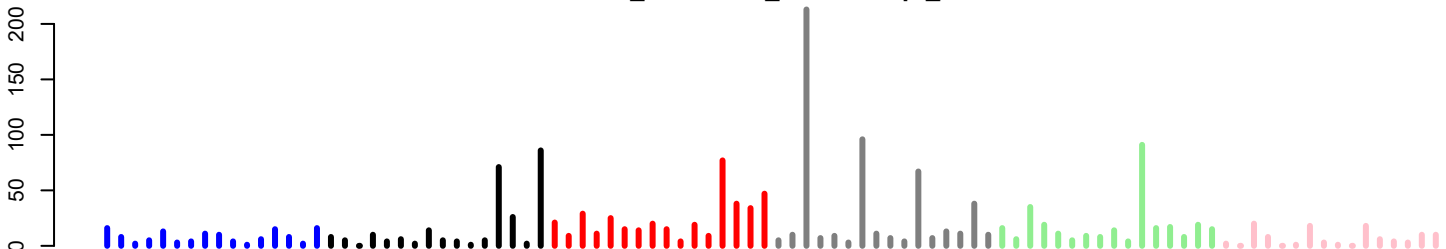
Below tables list the top 20 samples with the highest cosine similarity of T>N (left) or T>A (right) mutations compared to the TC1 spectrum from the 23,829 tumors investigated.

The mutation spectra of these tumors are found attached below.

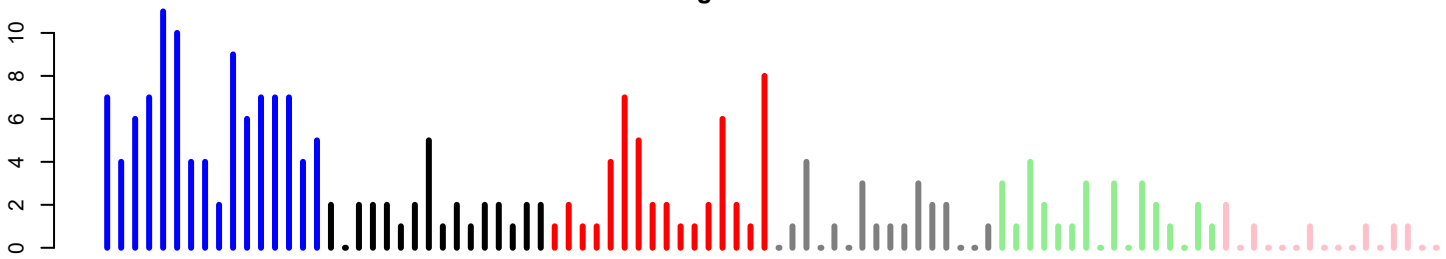
Although we observe some similarity to the TC1 mutation spectrum in the samples below, the mutation counts are always very low. Therefore we cannot make solid conclusions on the presence of the same mutational process.

Sample	cosine T>N	cosine T>A	Sample	cosine T>N	cosine T>A
Liver-HCC::TCGA-RC-A6M6-01A-11D-A32G-10	0.8133	0.8973	Lung-SCC::TCGA-33-4547-01A-01D-1267-08	0.7455	0.9605
Lung-AdenoCa::TCGA-44-8120-01A-11D-2238-08	0.8127	0.7769	Adrenal-neoplasm::TCGA-OR-A5J6-01A-31D-A29I-10	0.5917	0.9546
CNS-GBM::SK00102_P	0.7937	0.9068	Uterus-AdenoCa::TCGA-A5-A2K2-01A-11D-A18P-09	0.5914	0.9290
Uterus-AdenoCa::TCGA-AX-A3G4-01A-11D-A20S-09	0.7932	0.9136	Panc-AdenoCa::8014777	0.7065	0.9137
Lung-SCC::TCGA-68-A59J-01A-21D-A26M-08	0.7882	0.7959	Uterus-AdenoCa::TCGA-AX-A3G4-01A-11D-A20S-09	0.7932	0.9136
Lung-AdenoCa::LUAD-NYU1051S	0.7881	0.8168	Panc-AdenoCa::TCGA-IB-A5SS-01A-11D-A32N-08	0.6687	0.9108
Lung-AdenoCa::TCGA-55-7570-01A-11D-2036-08	0.7817	0.8526	Prost-AdenoCa::TCGA-CH-5765-01A-11D-1576-08	0.5500	0.9108
Lung-AdenoCa::TCGA-49-4506-01A-01D-1265-08	0.7798	0.8500	Liver-HCC::TCGA-DD-AADD-01A-11D-A40R-10	0.6629	0.9092
Lung-SCC::TCGA-98-A53H-01A-12D-A25L-08	0.7752	0.8631	CNS-GBM::SK00102_P	0.7937	0.9068
Lung-AdenoCa::TCGA-73-4658-01A-01D-1753-08	0.7723	0.8579	Transitional-cell-carcinoma::TCGA-DK-AA77-01A-11D-A	0.7243	0.9042
Lung-SCC::TCGA-43-8118-01A-11D-2395-08	0.7700	0.7591	Sarcoma::TCGA-3B-A9HO-01A-11D-A387-09	0.5967	0.9004
Lung-SCC::TCGA-56-8307-01A-11D-2293-08	0.7699	0.7707	Kidney-Papillary::TCGA-G7-6796-01A-11D-1961-08	0.6828	0.8973
Lung-AdenoCa::TCGA-49-AARQ-01A-11D-A410-08	0.7690	0.7660	Liver-HCC::TCGA-RC-A6M6-01A-11D-A32G-10	0.8133	0.8973
Lung-SCC::TCGA-39-5040-01A-21D-2122-08	0.7689	0.8090	Lymph-BNHL::TCGA-RQ-A68N-01A-11D-A31X-10	0.5947	0.8962
Lung-Small::S00050	0.7640	0.8592	Bladder-TCC::B106-Tumor	0.4696	0.8946
Lung-Small::585270	0.7630	0.8379	Head-SCC::HN_00378	0.5959	0.8916
Lung-AdenoCa::TCGA-MP-A4TK-01A-11D-A24P-08	0.7614	0.7600	Sarcoma::TCGA-DX-A8BZ-01A-11D-A37C-09	0.6456	0.8910
Lung-AdenoCa::TCGA-05-4249-01A-01D-1105-08	0.7604	0.8219	Kidney-RCC::TCGA-B0-5100-01A-01D-1421-08	0.6424	0.8889
Lung-AdenoCa::TCGA-NJ-A4YI-01A-11D-A25L-08	0.7594	0.8000	AML::CN-AML-CR-42-Dx	0.5972	0.8865
Lung-SCC::TCGA-77-8144-01A-11D-2244-08	0.7579	0.8624	Transitional-cell-carcinoma::TCGA-E5-A4U1-01A-11D-A	0.6997	0.8862

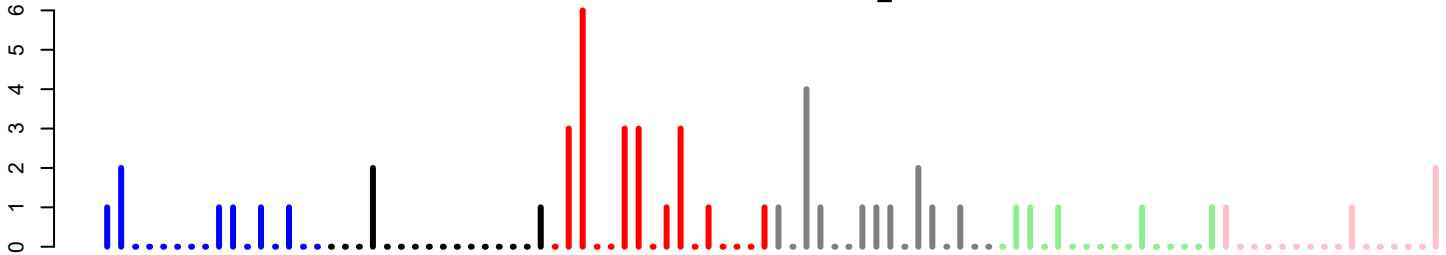
TC1.Tumor\_SNVresult\_inTranscript\_intersect



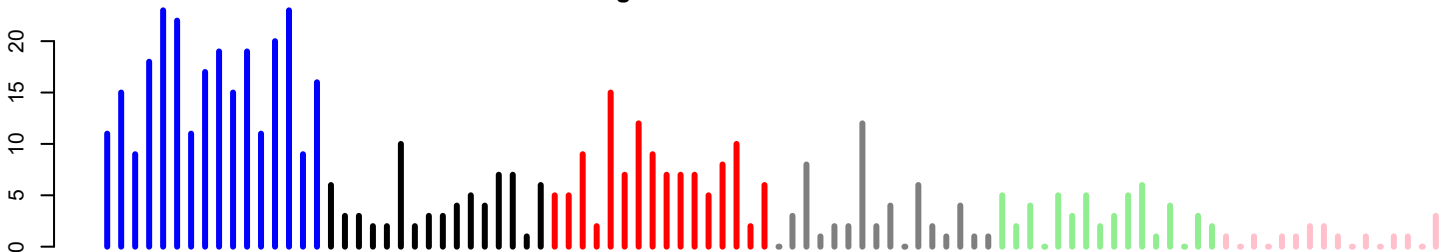
Lung-Small::S00050



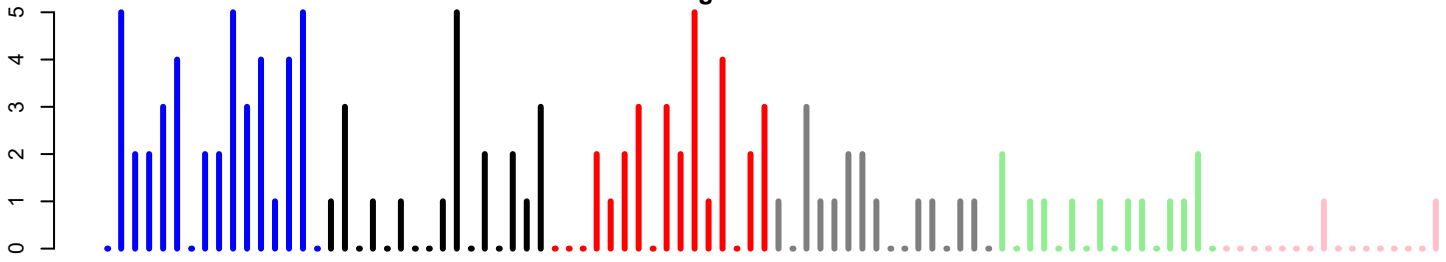
CNS-GBM::SK00102\_P



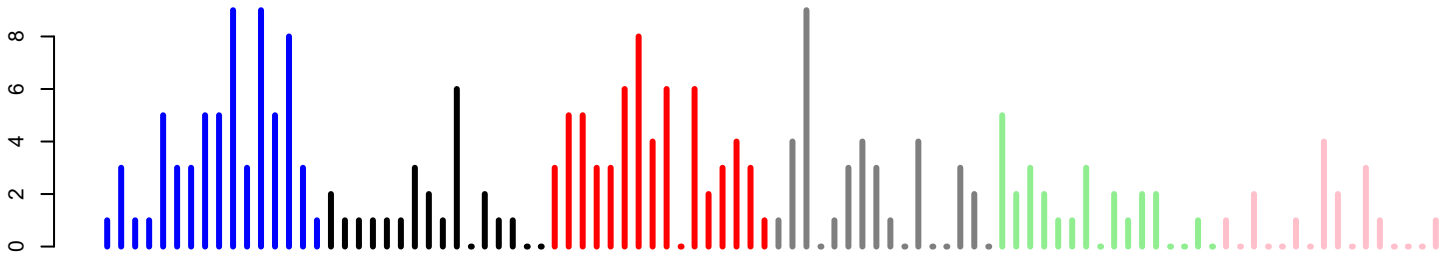
Lung-AdenoCa::LUAD-NYU1051S



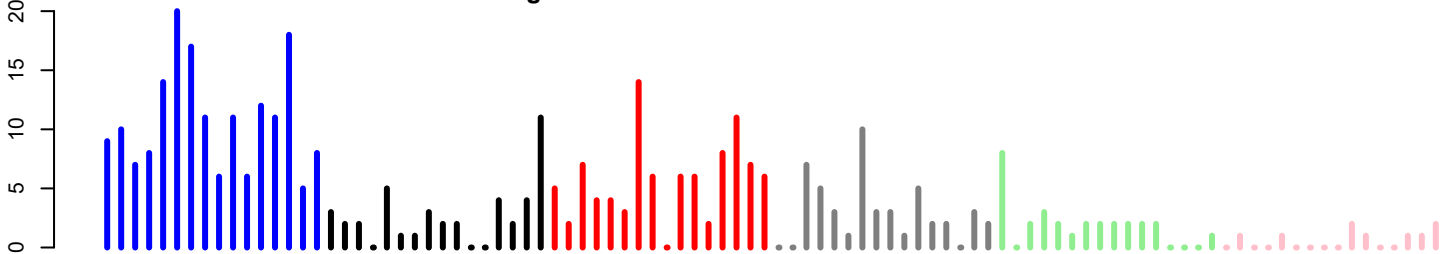
Lung-Small::585270



Liver-HCC::TCGA-RC-A6M6-01A-11D-A32G-10

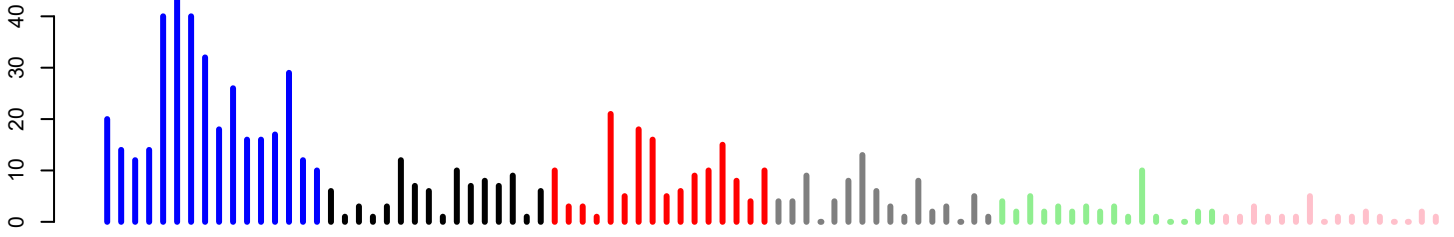


Lung-AdenoCa::TCGA-05-4249-01A-01D-1105-08

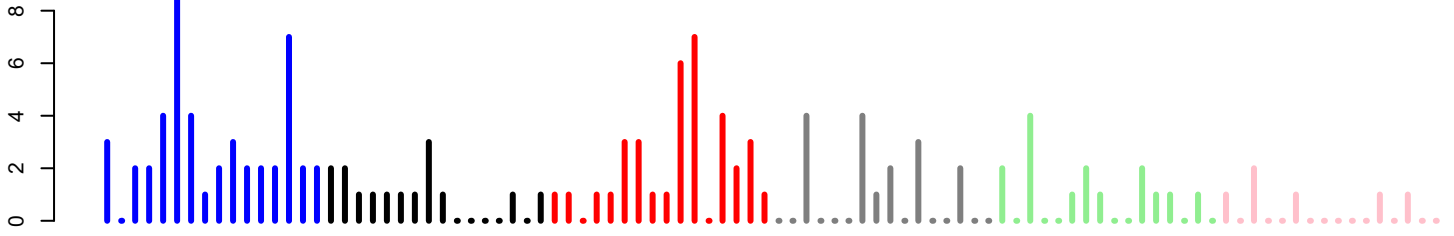




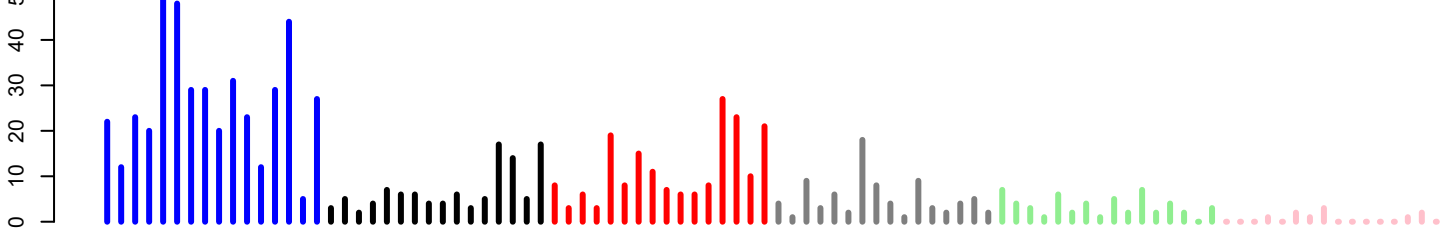
Lung-AdenoCa::TCGA-44-8120-01A-11D-2238-08



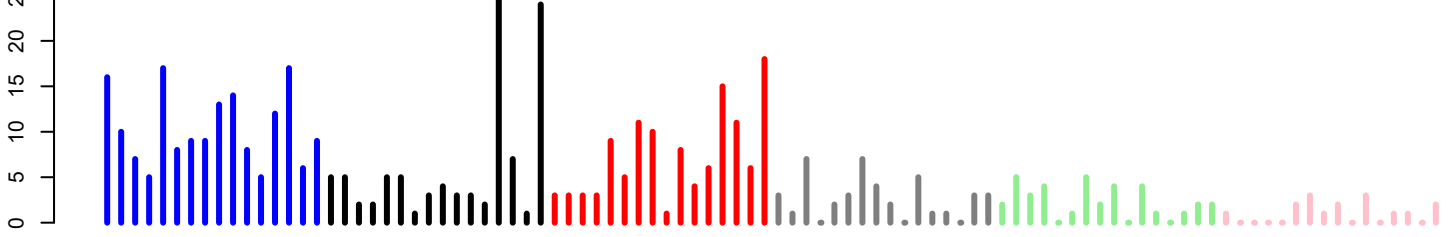
Lung-AdenoCa::TCGA-49-4506-01A-01D-1265-08



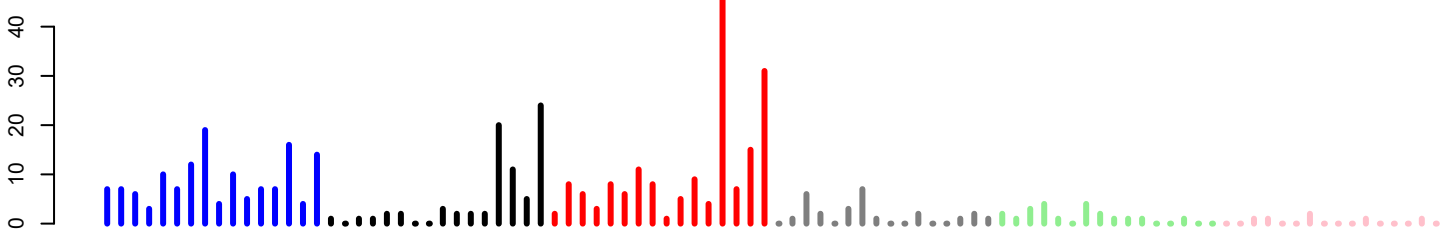
Lung-AdenoCa::TCGA-49-AARQ-01A-11D-A410-08



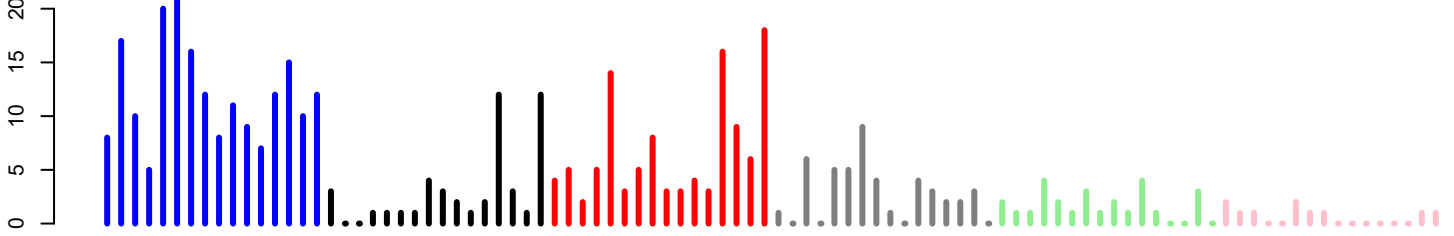
Lung-AdenoCa::TCGA-55-7570-01A-11D-2036-08



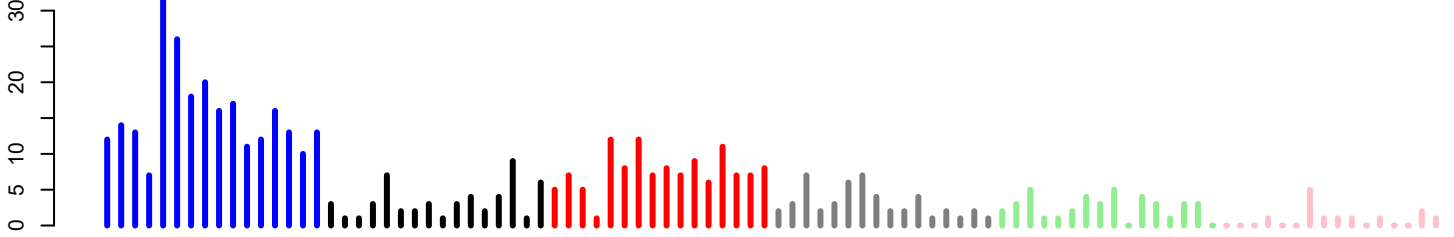
Lung-AdenoCa::TCGA-73-4658-01A-01D-1753-08



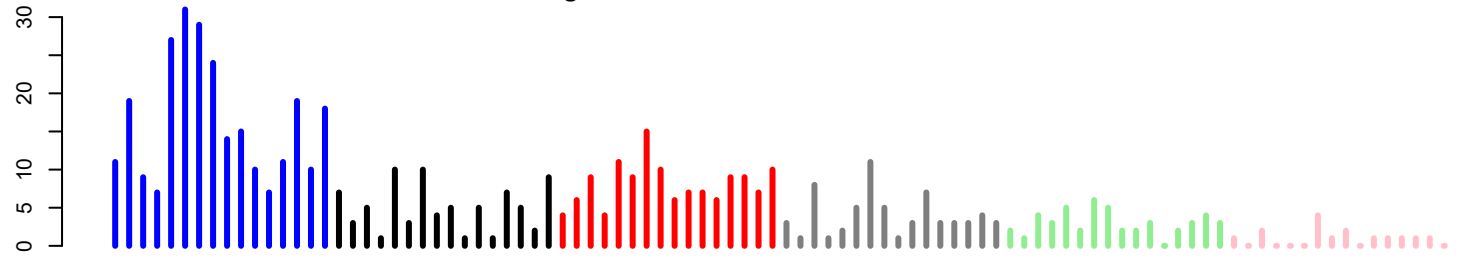
Lung-AdenoCa::TCGA-MP-A4TK-01A-11D-A24P-08



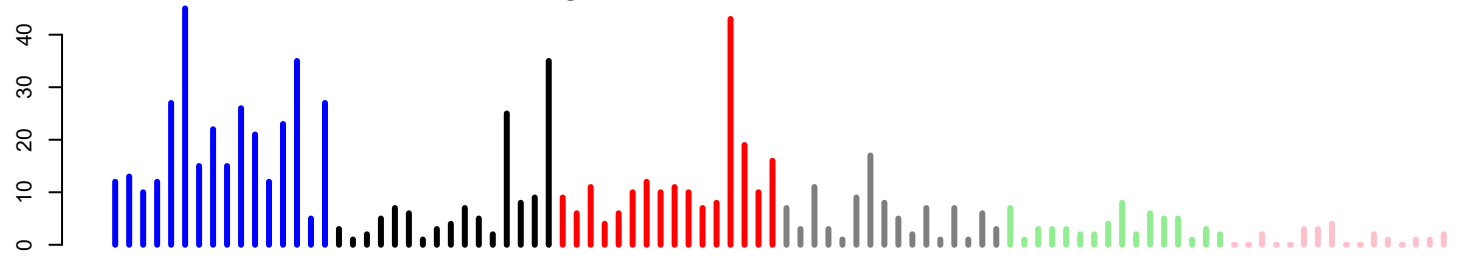
Lung-AdenoCa::TCGA-NJ-A4YI-01A-11D-A25L-08



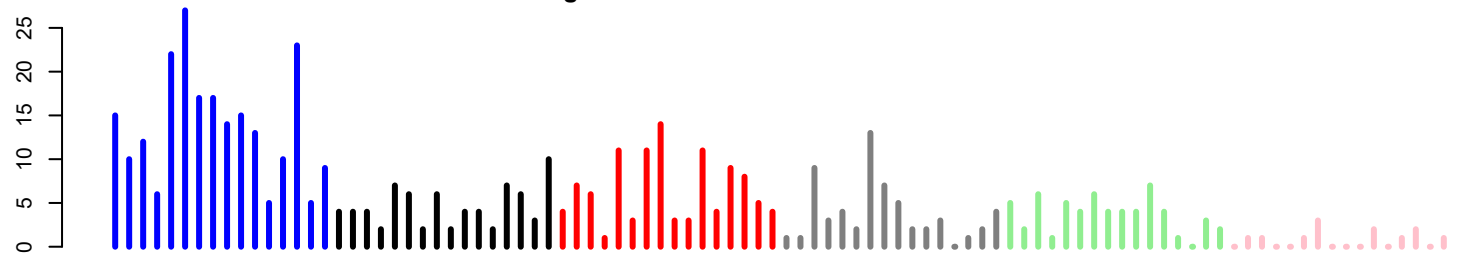
Lung-SCC::TCGA-39-5040-01A-21D-2122-08



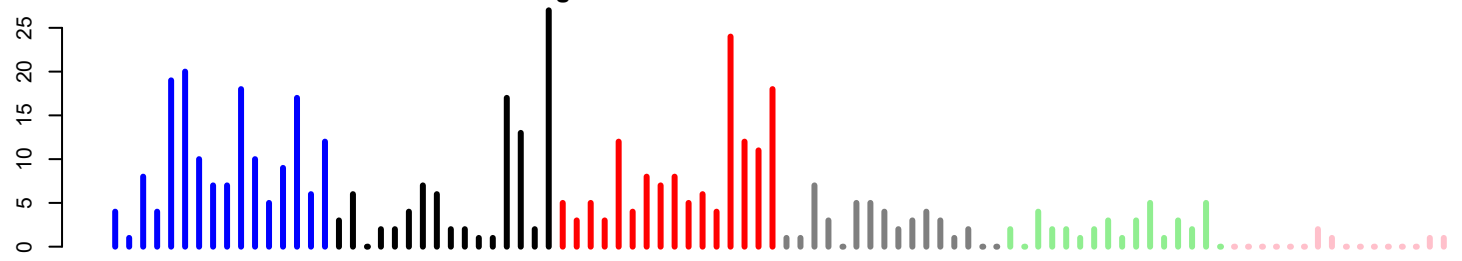
Lung-SCC::TCGA-43-8118-01A-11D-2395-08



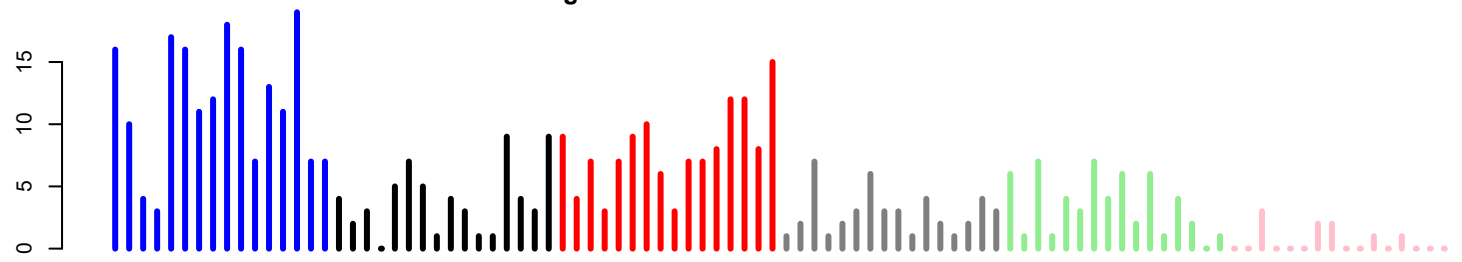
Lung-SCC::TCGA-56-8307-01A-11D-2293-08



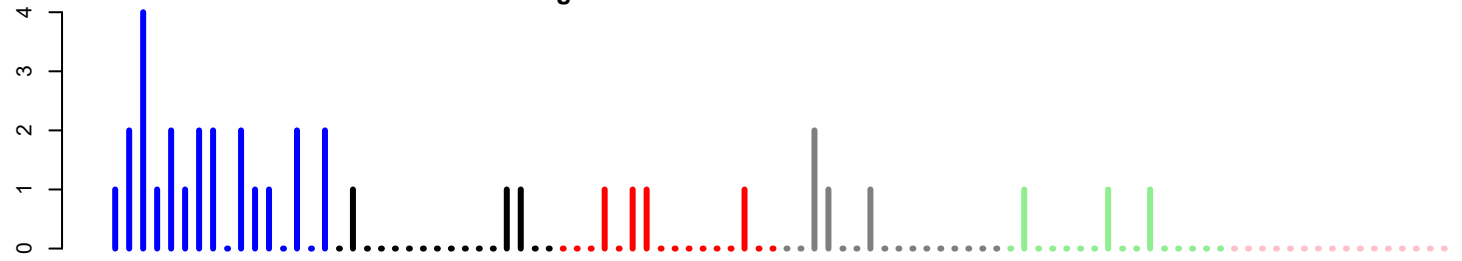
Lung-SCC::TCGA-68-A59J-01A-21D-A26M-08



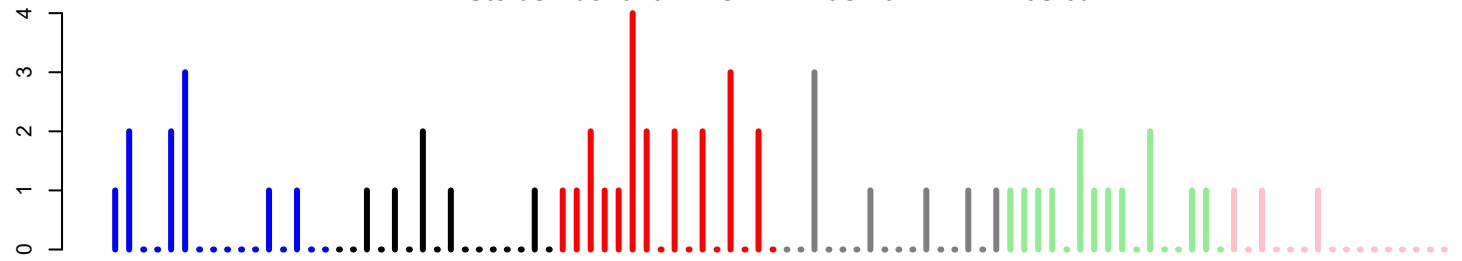
Lung-SCC::TCGA-77-8144-01A-11D-2244-08



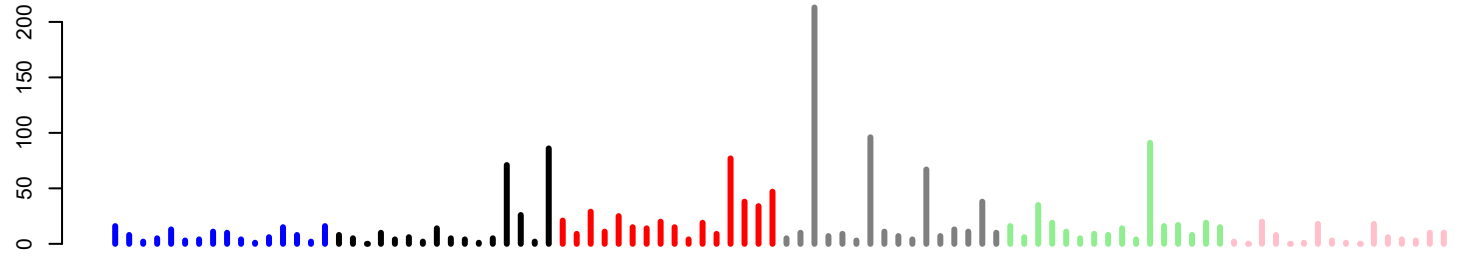
Lung-SCC::TCGA-98-A53H-01A-12D-A25L-08



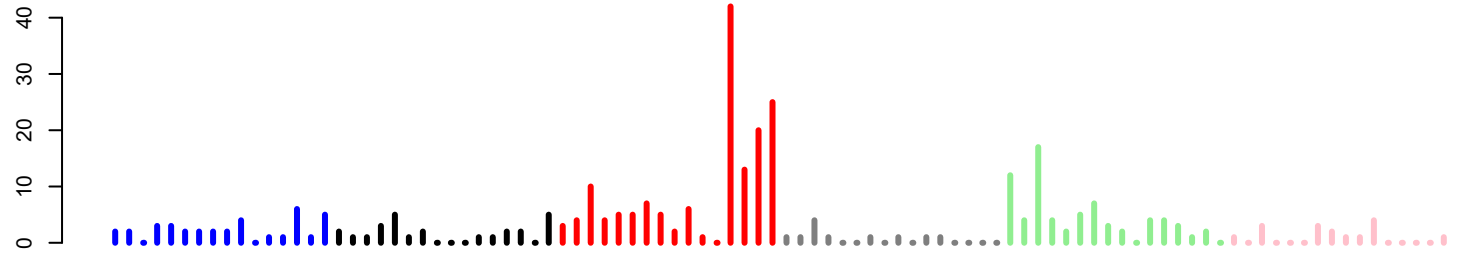
Uterus-AdenoCa::TCGA-AX-A3G4-01A-11D-A20S-09



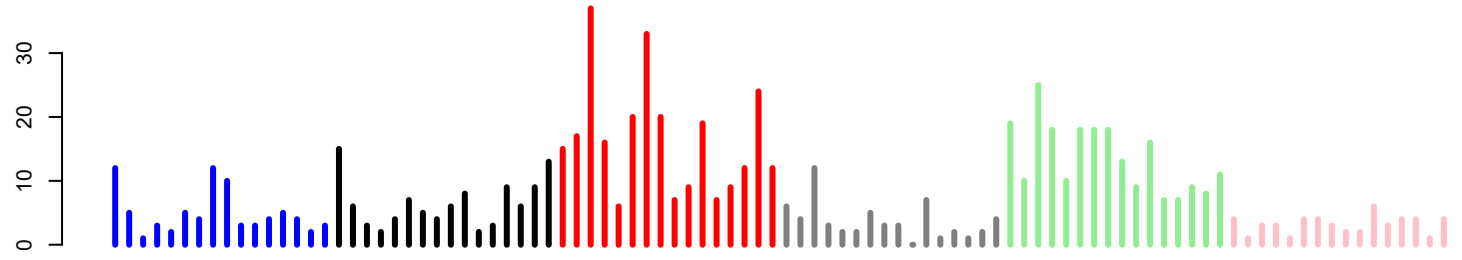
TC1.Tumor\_SNVresult\_inTranscript\_intersect



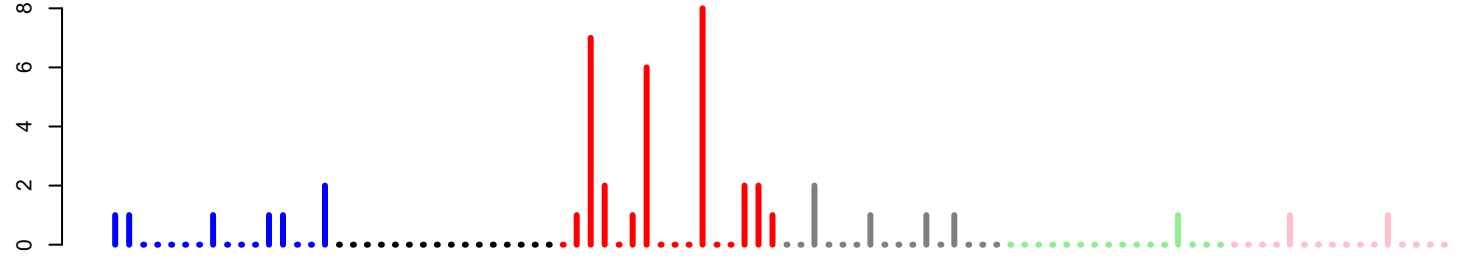
Bladder-TCC::B106-Tumor



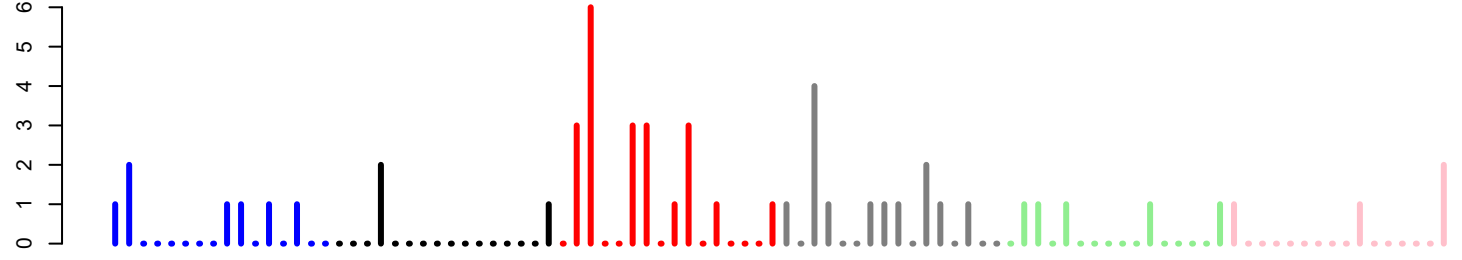
AML::CN-AML-CR-42-Dx



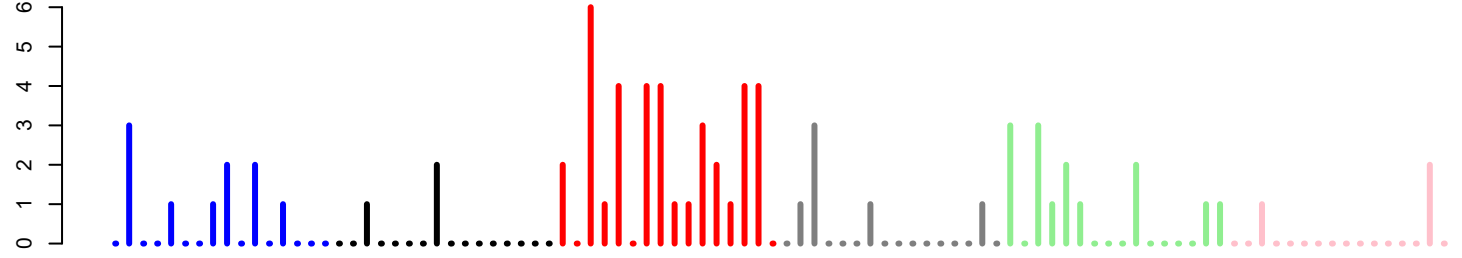
Panc-AdenoCa::8014777



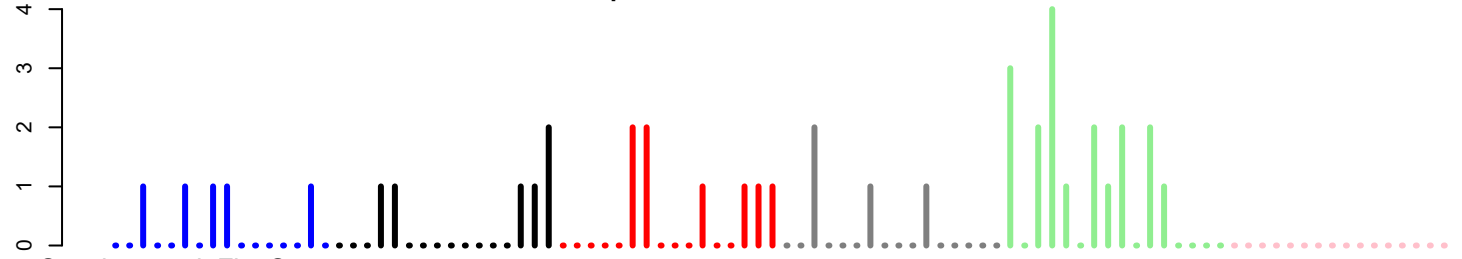
CNS-GBM::SK00102\_P



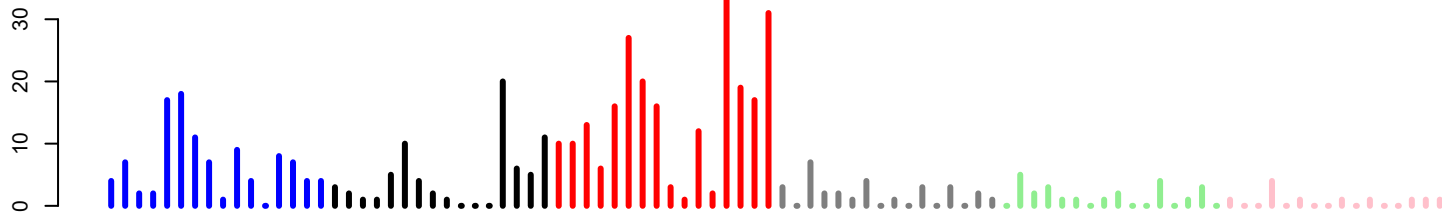
Head-SCC::HN\_00378



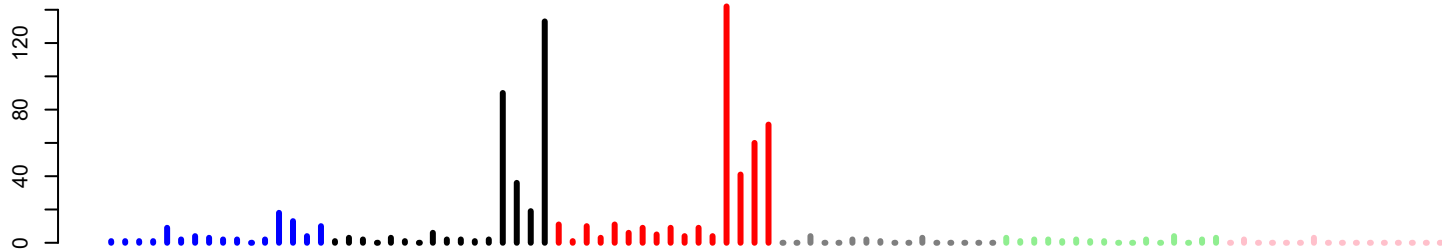
Adrenal-neoplasm::TCGA-OR-A5J6-01A-31D-A29I-10



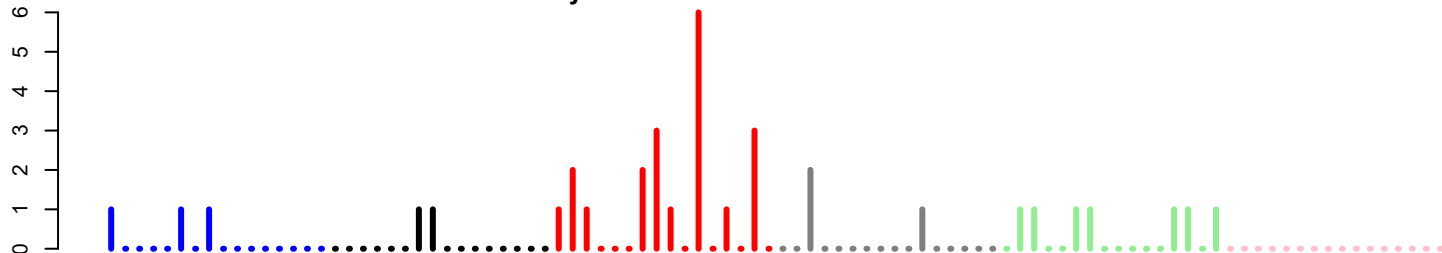
Transitional-cell-carcinoma::TCGA-DK-AA77-01A-11D-A391-08



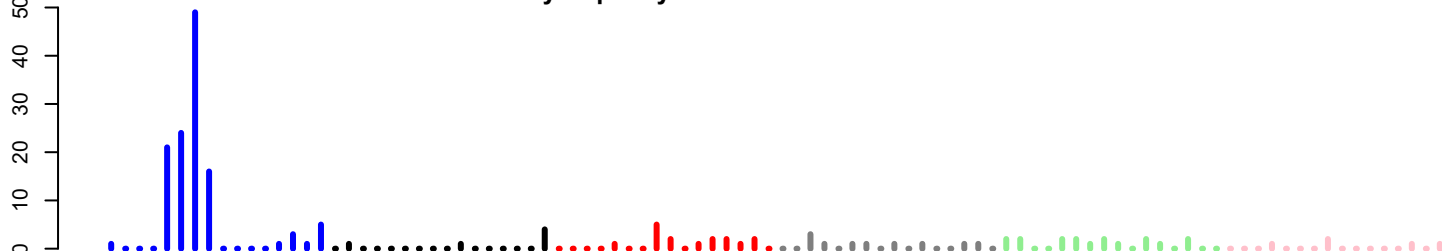
Transitional-cell-carcinoma::TCGA-E5-A4U1-01A-11D-A31L-08



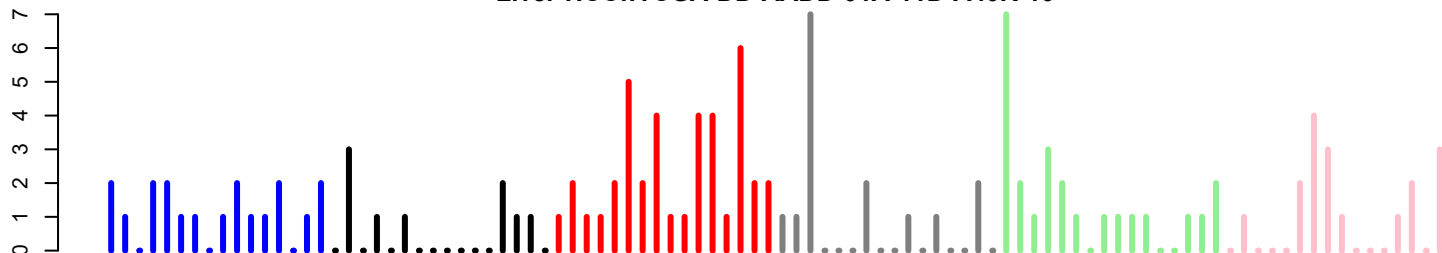
Kidney-RCC::TCGA-B0-5100-01A-01D-1421-08



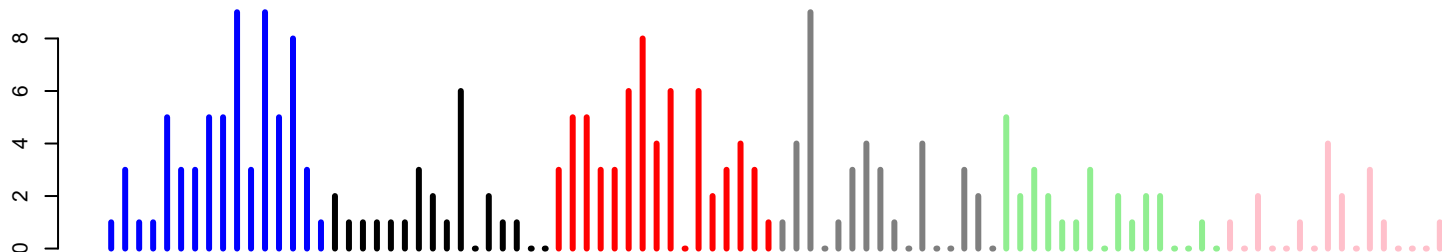
Kidney-Papillary::TCGA-G7-6796-01A-11D-1961-08



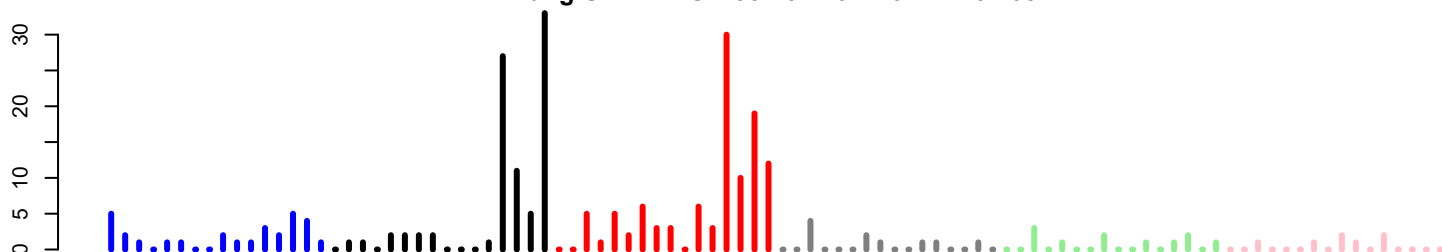
Liver-HCC::TCGA-DD-AADD-01A-11D-A40R-10

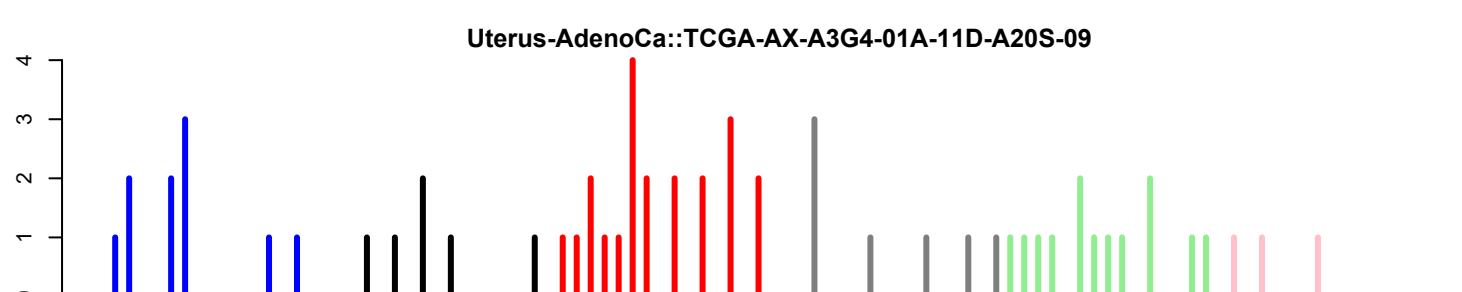
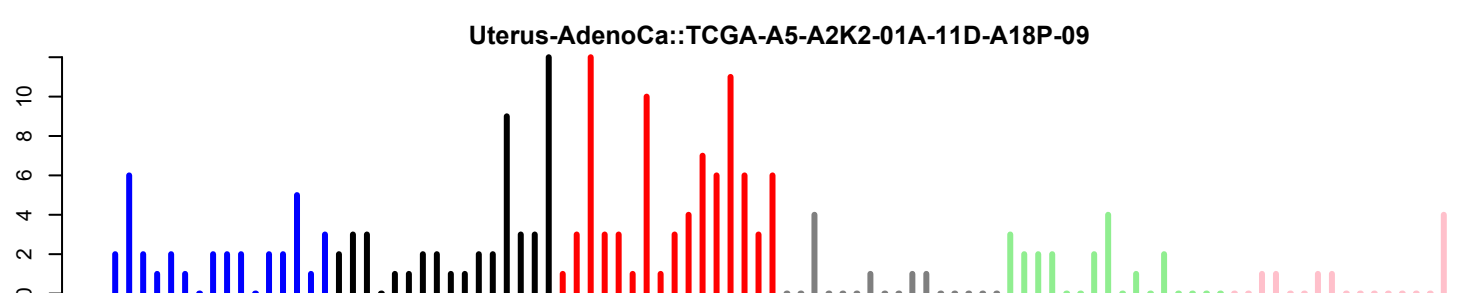
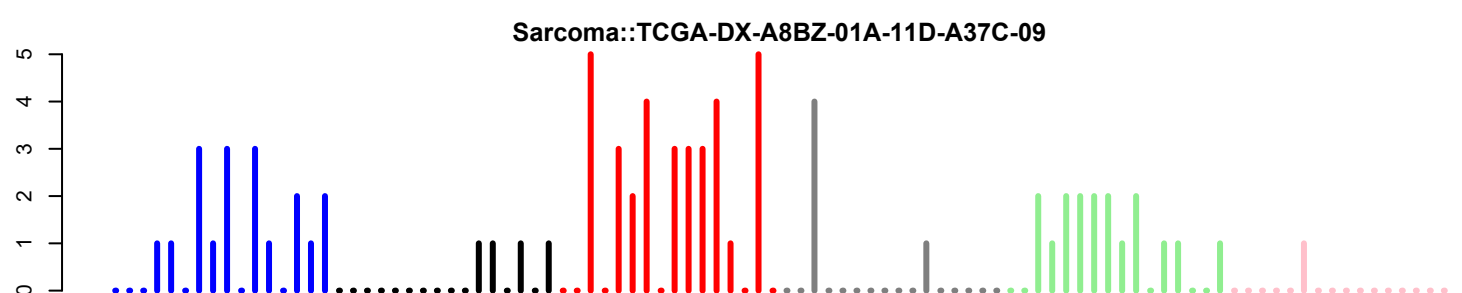
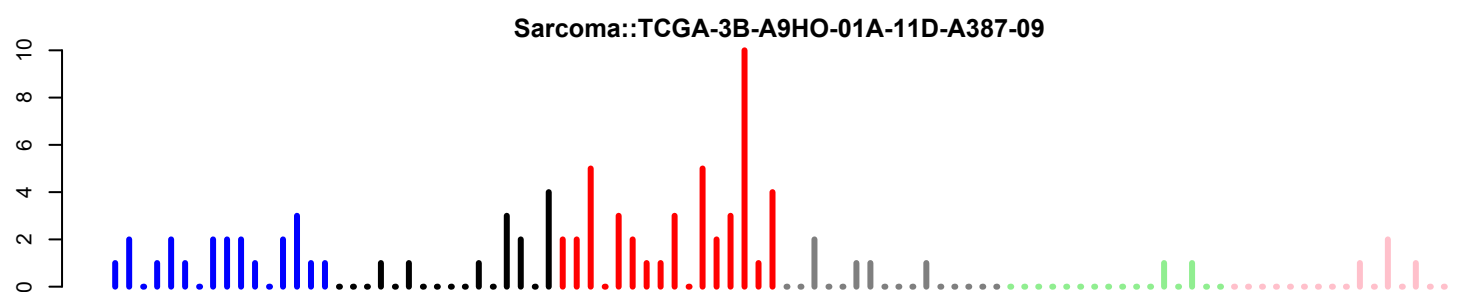
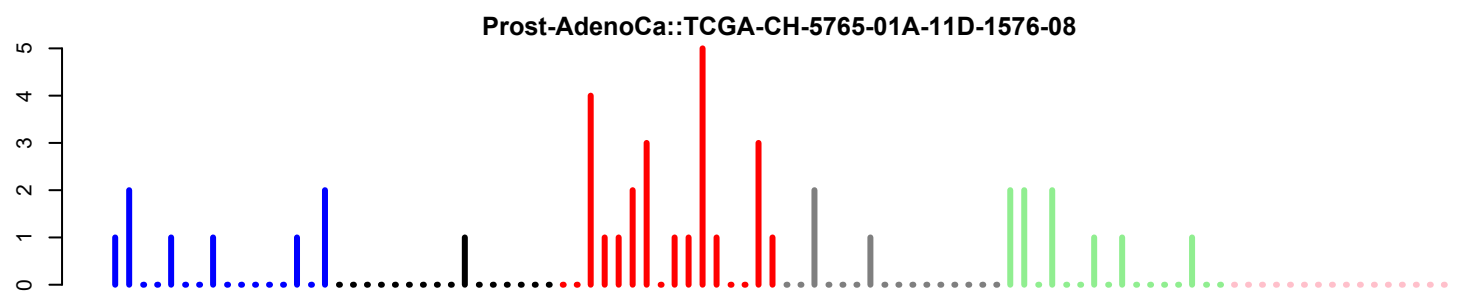
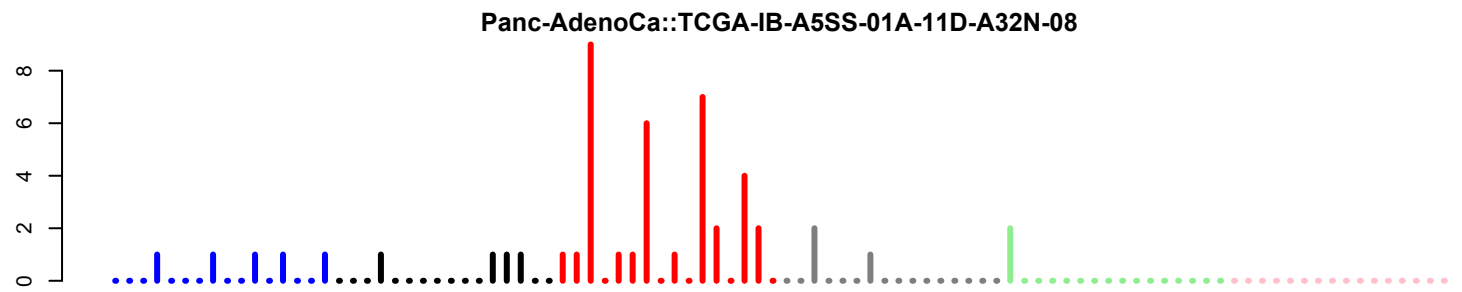
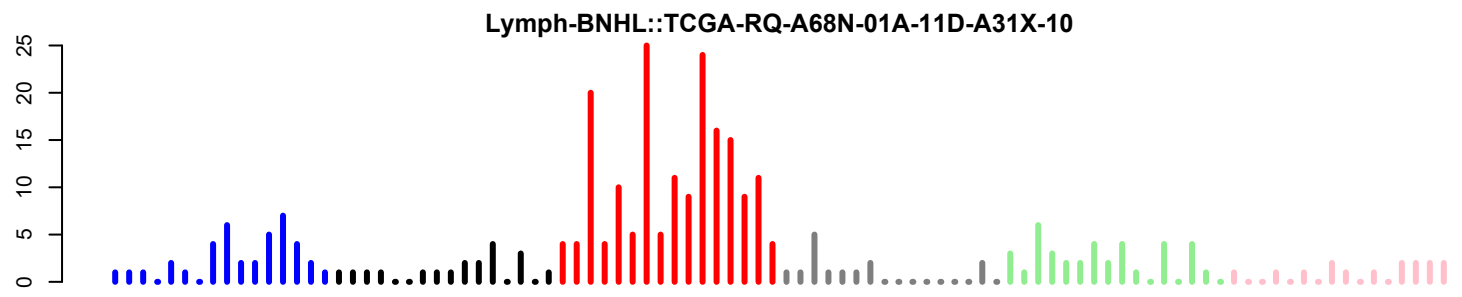


Liver-HCC::TCGA-RC-A6M6-01A-11D-A32G-10



Lung-SCC::TCGA-33-4547-01A-01D-1267-08





Supplemental\_Fig\_S17: Alignment of non-human reads from WGS data of patients 62074759 and TC1 to bacterial genomes

