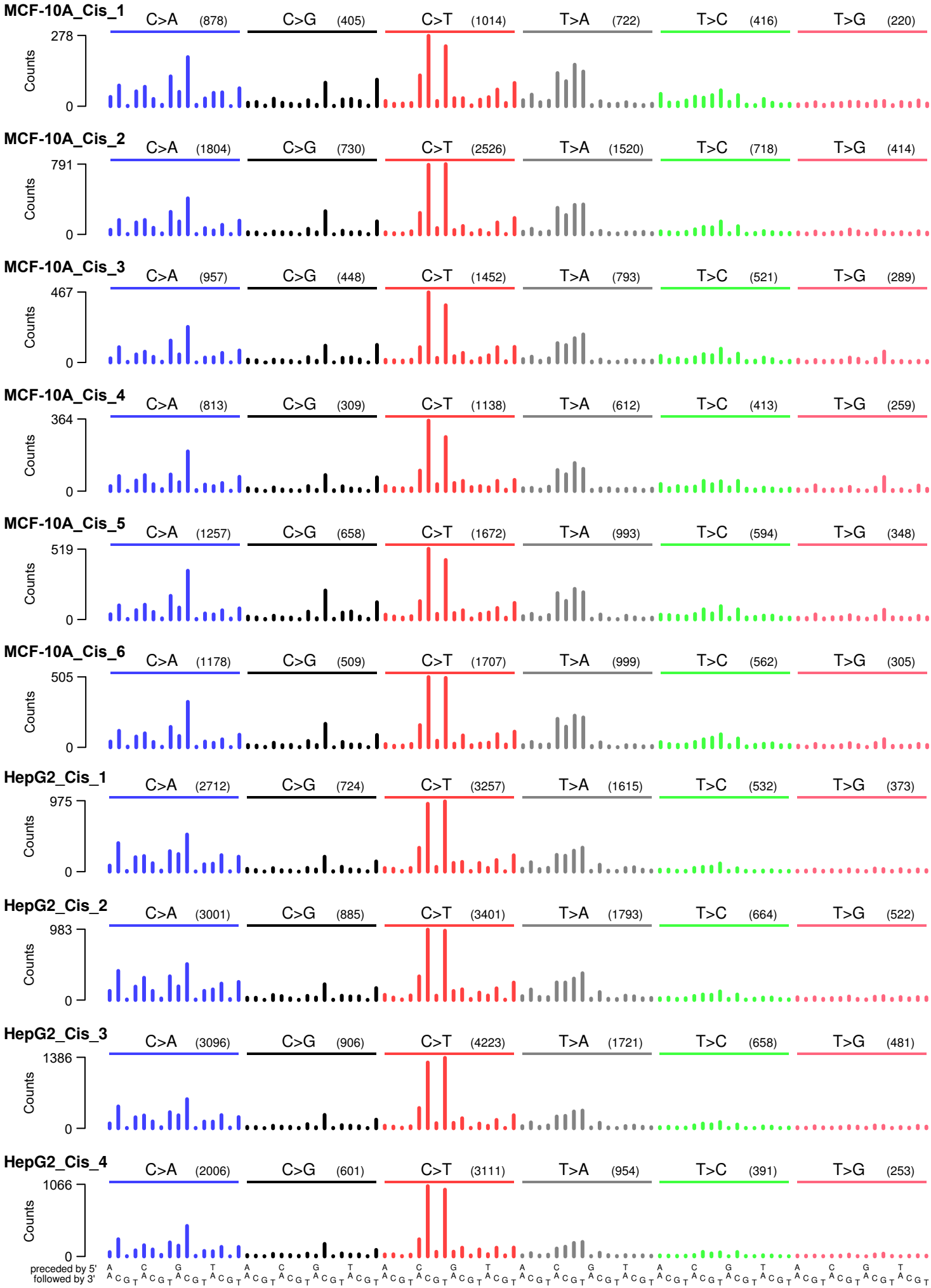
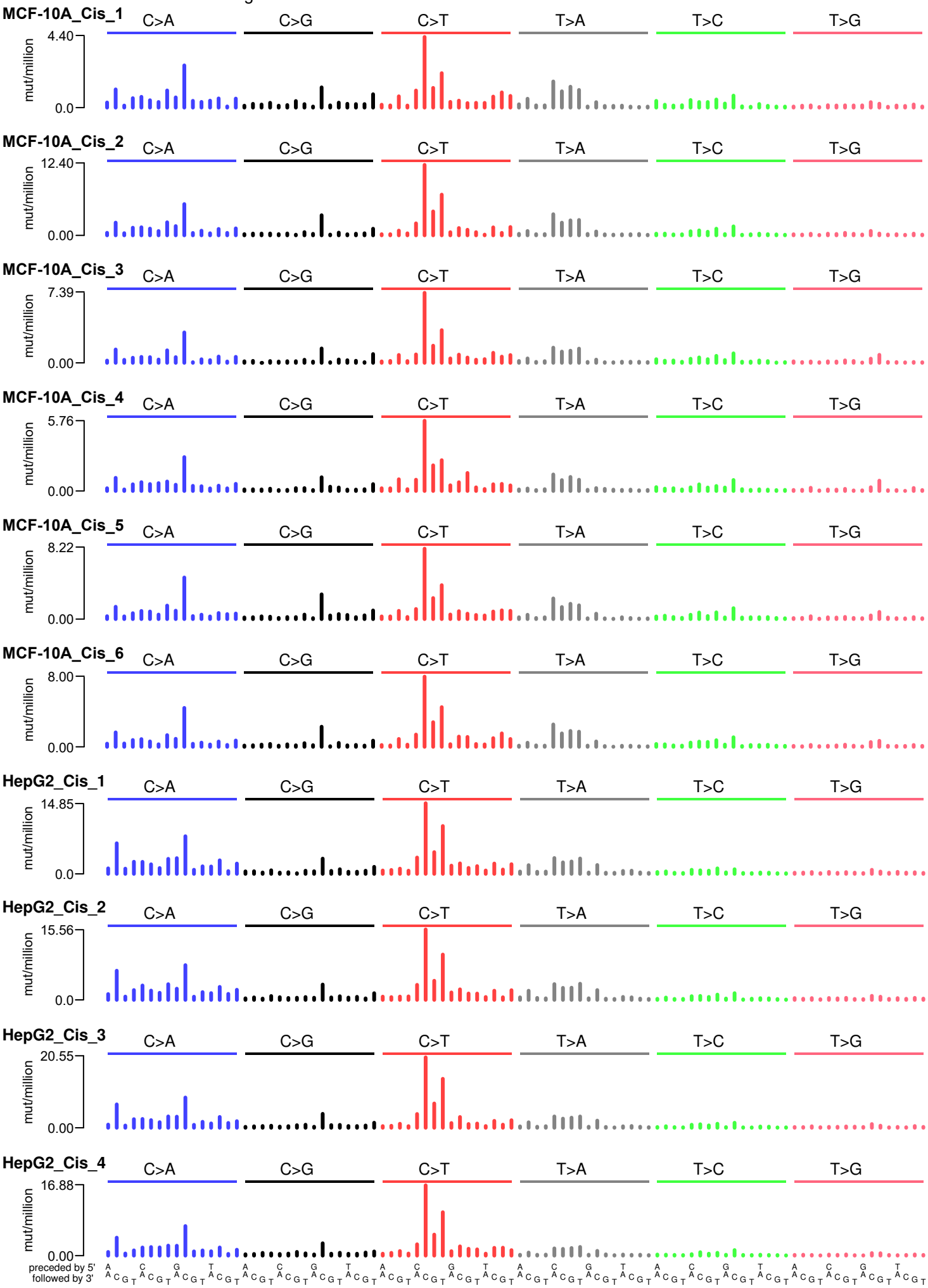
Table of Contents

Supplemental Fig. S1A: SNS mutation spectra of cisplatin treated MCF-10A and HepG2 clones

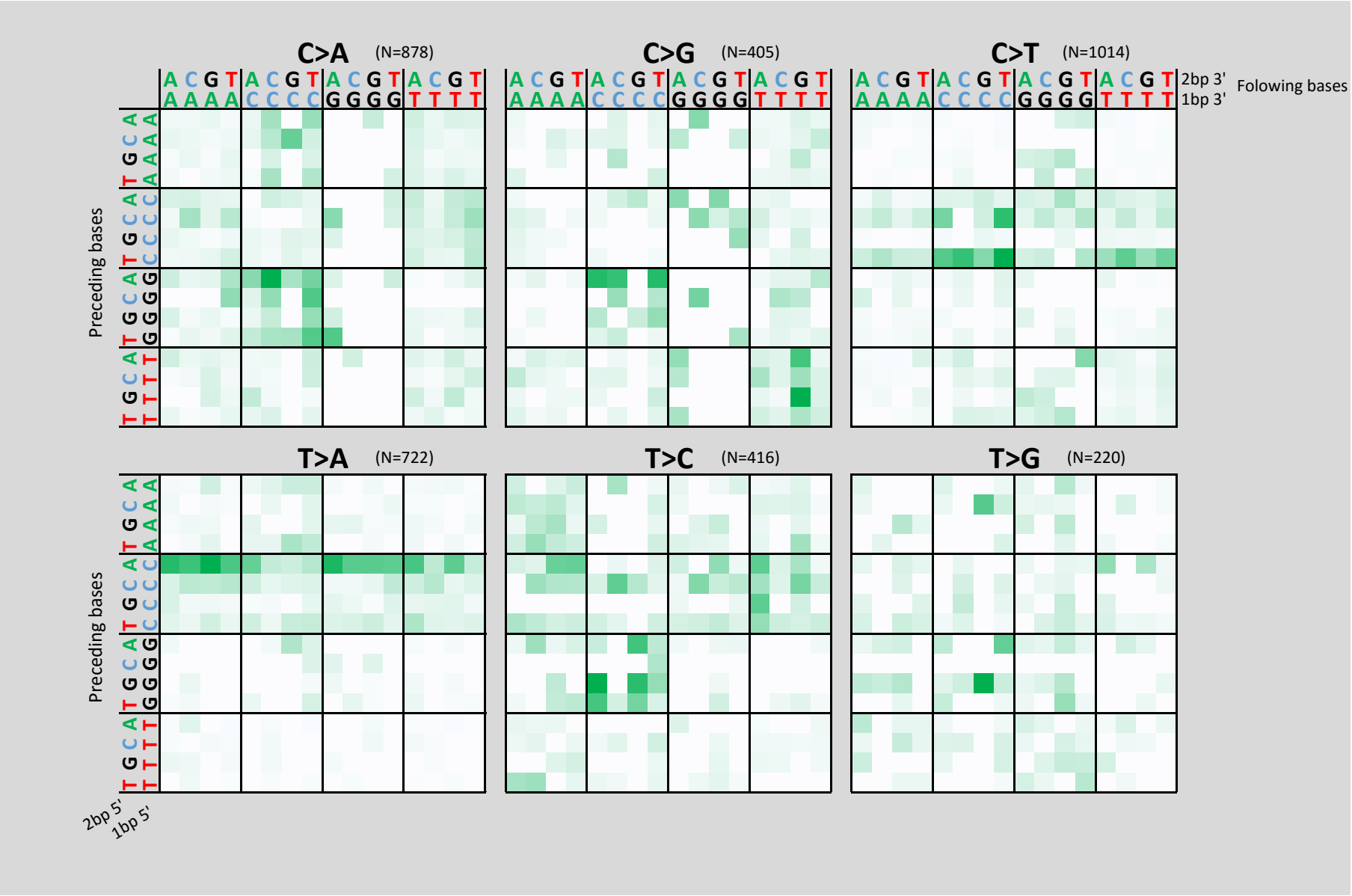


Supplemental Fig. S1B: SNS mutation spectra of cisplatin treated MCF-10A and HepG2 clones, normalized for trinucleotide occurrence in the genome.

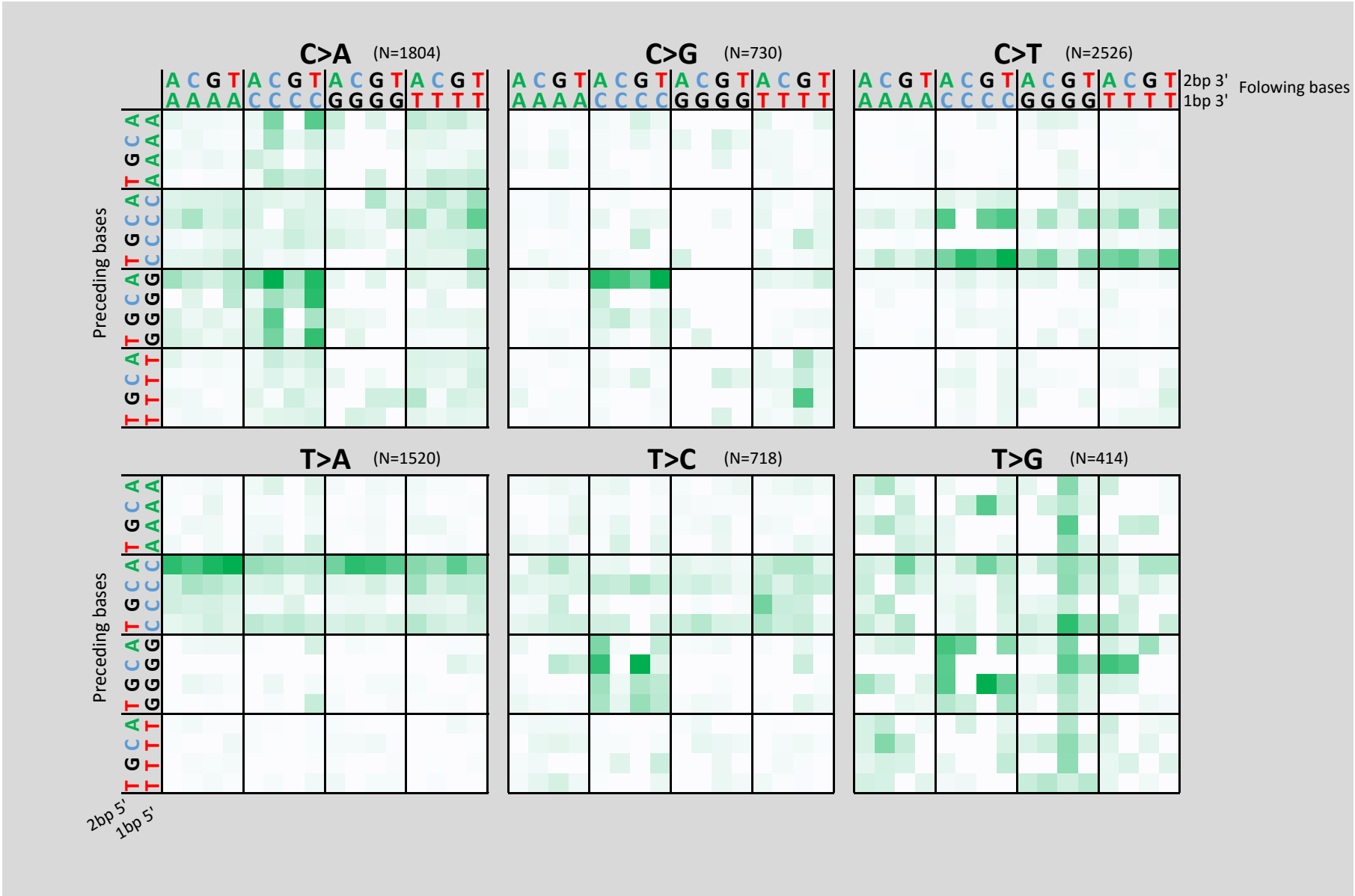


Supplemental Fig. S2: Pentanucleotide sequence contexts for all cisplatin treated MCF 10A clones (A-F) and HepG2 (G-J) clones, normalized for pentanucleotide occurrence in the genome.

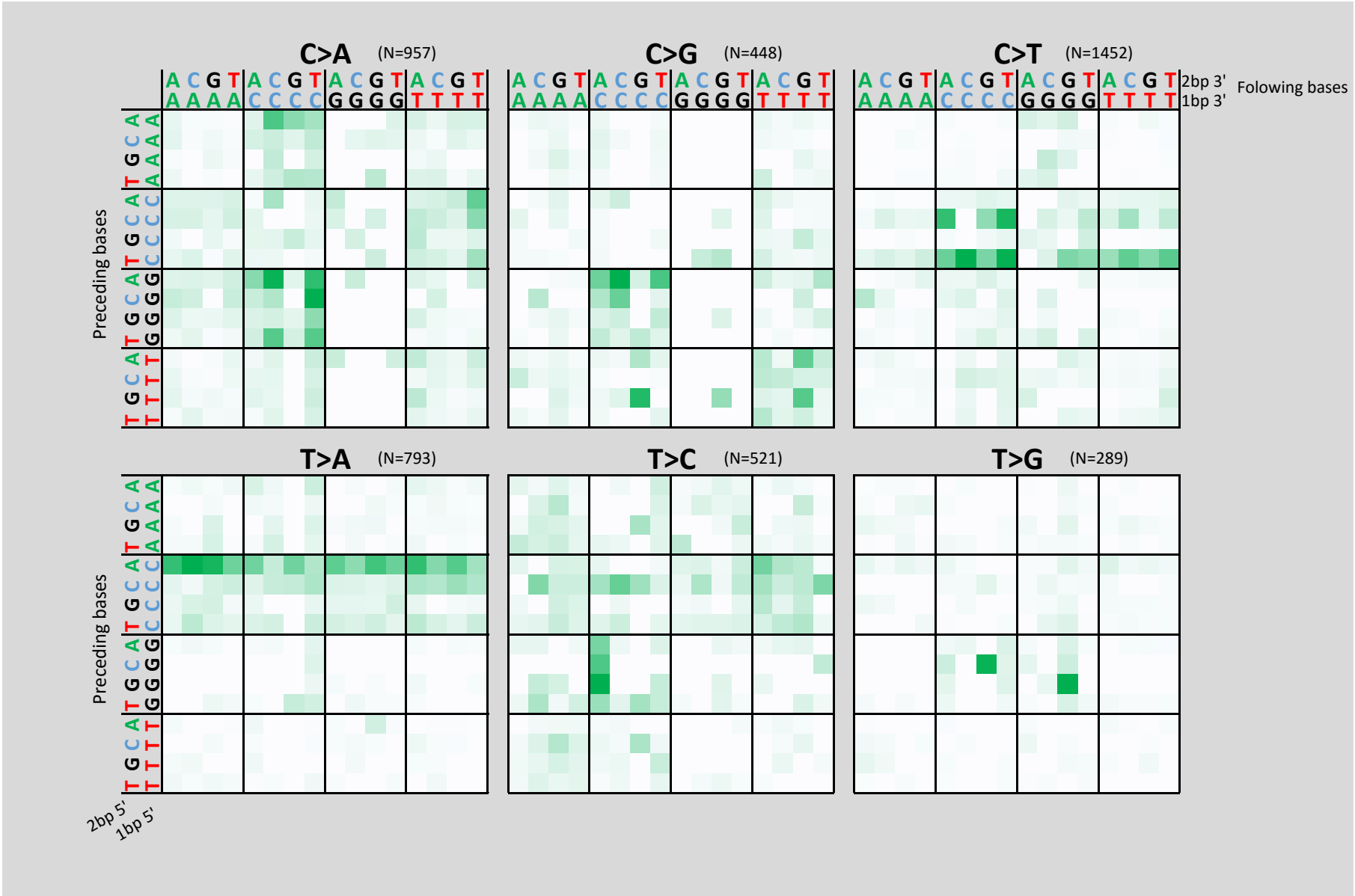
Supplemental Fig. S2A: Pentanucleotide context in MCF-10A_Cis_1



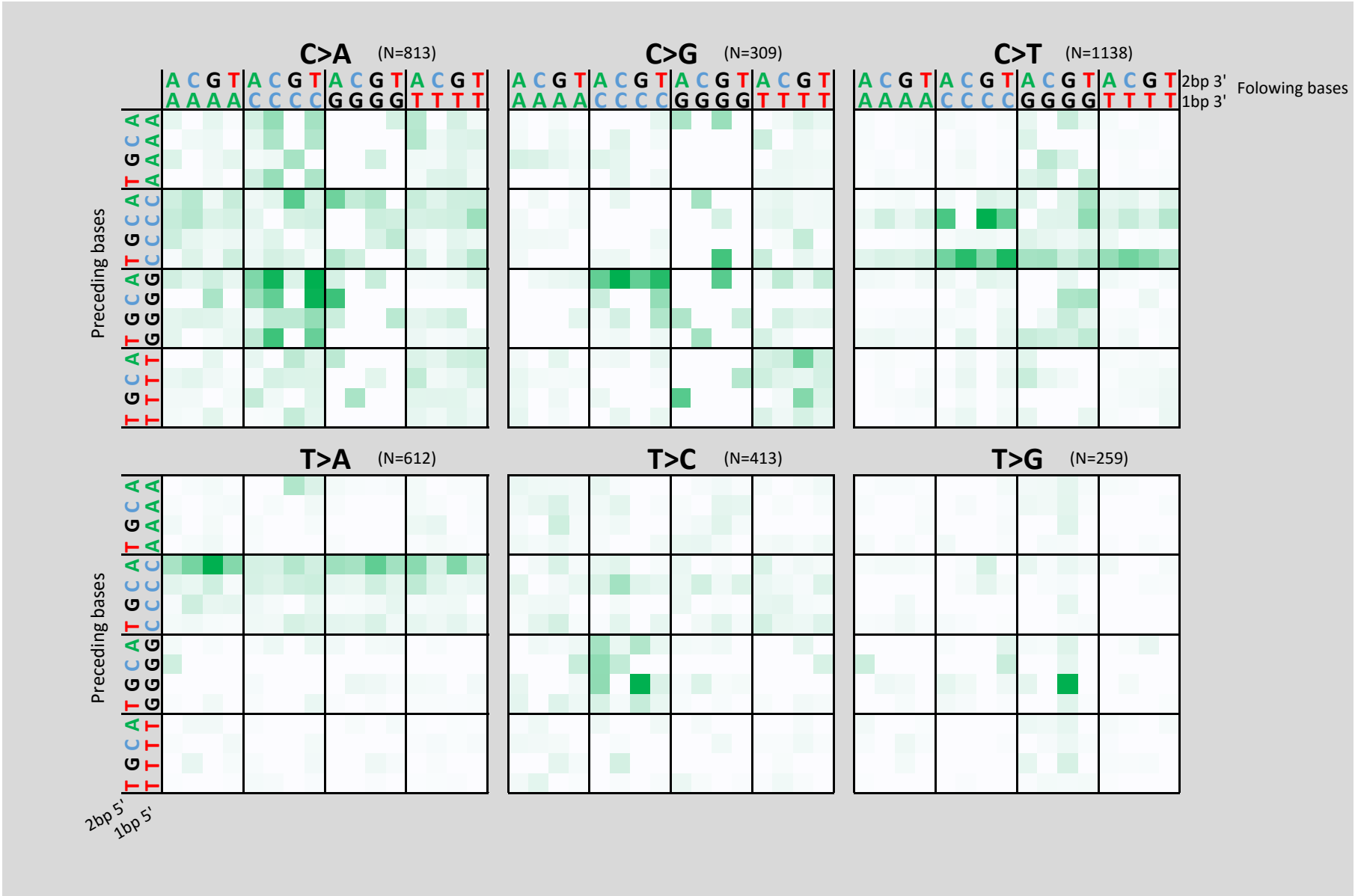
Supplemental Fig. S2B: Pentanucleotide context in MCF-10A_Cis_2



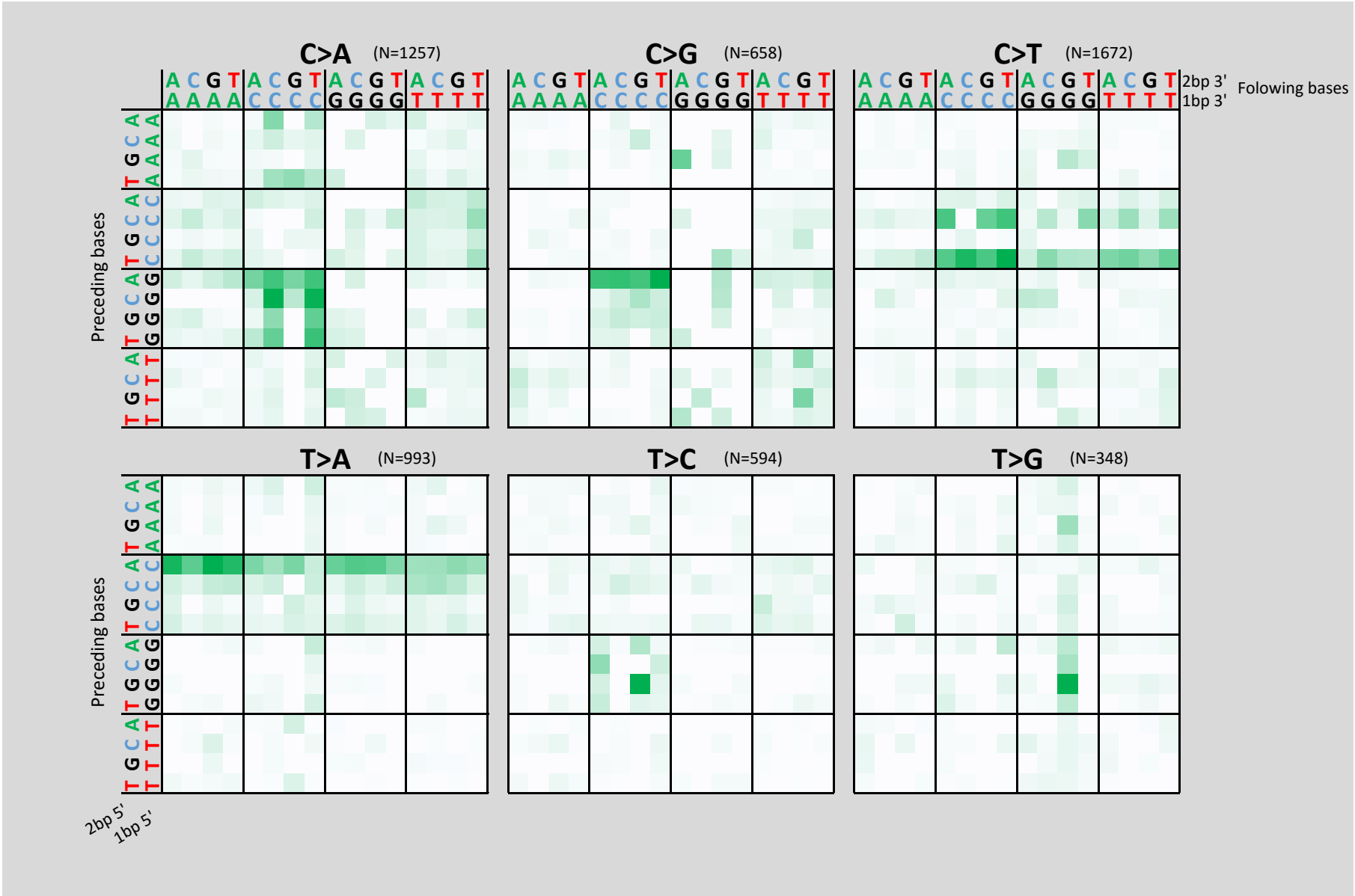
Supplemental Fig. S2C: Pentanucleotide context in MCF-10A_Cis_3



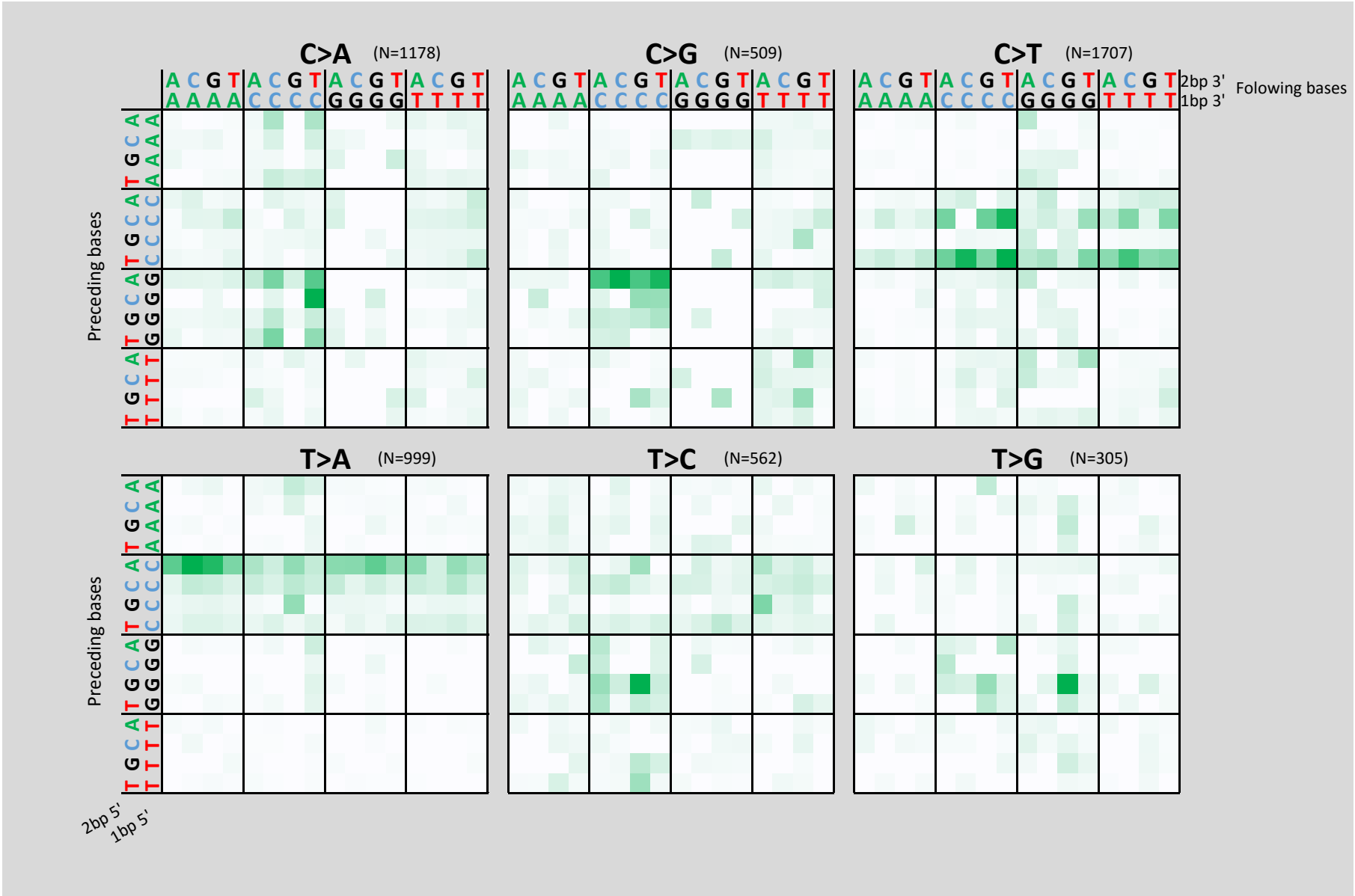
Supplemental Fig. S2D: Pentanucleotide context in MCF-10A_Cis_4



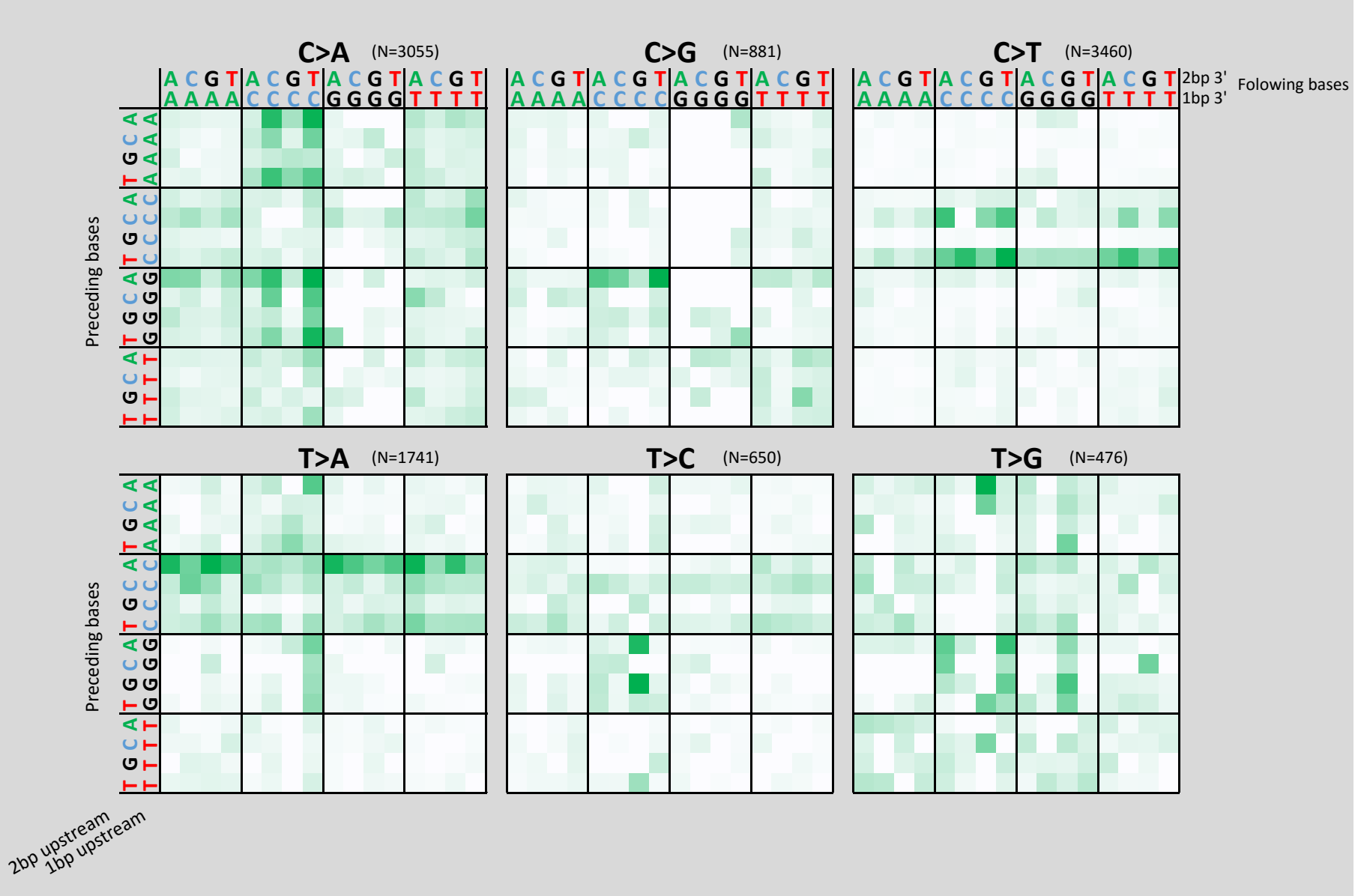
Supplemental Fig. S2E: Pentanucleotide context in MCF-10A_Cis_5



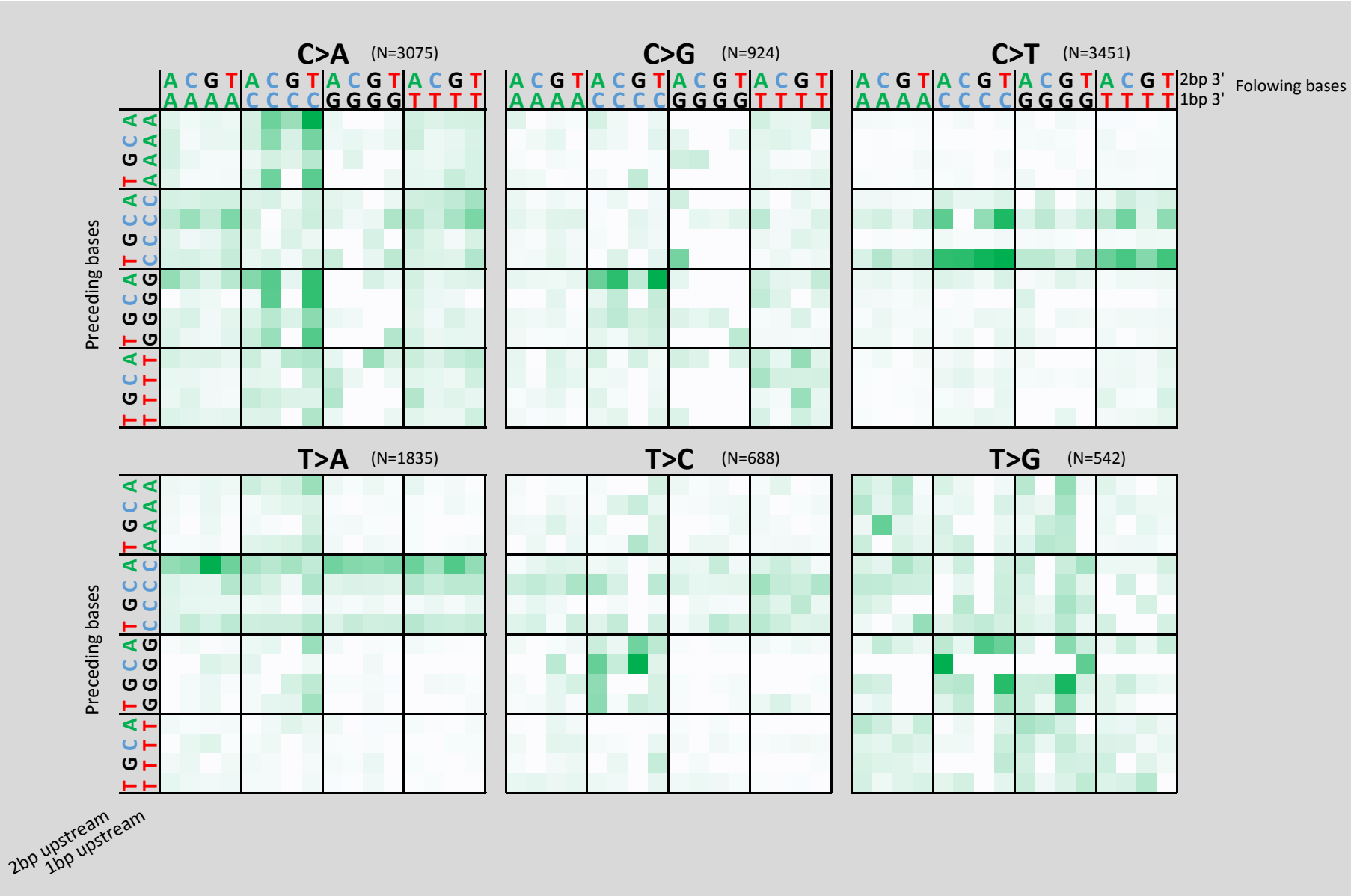
Supplemental Fig. S2F: Pentanucleotide context in MCF-10A_Cis_6



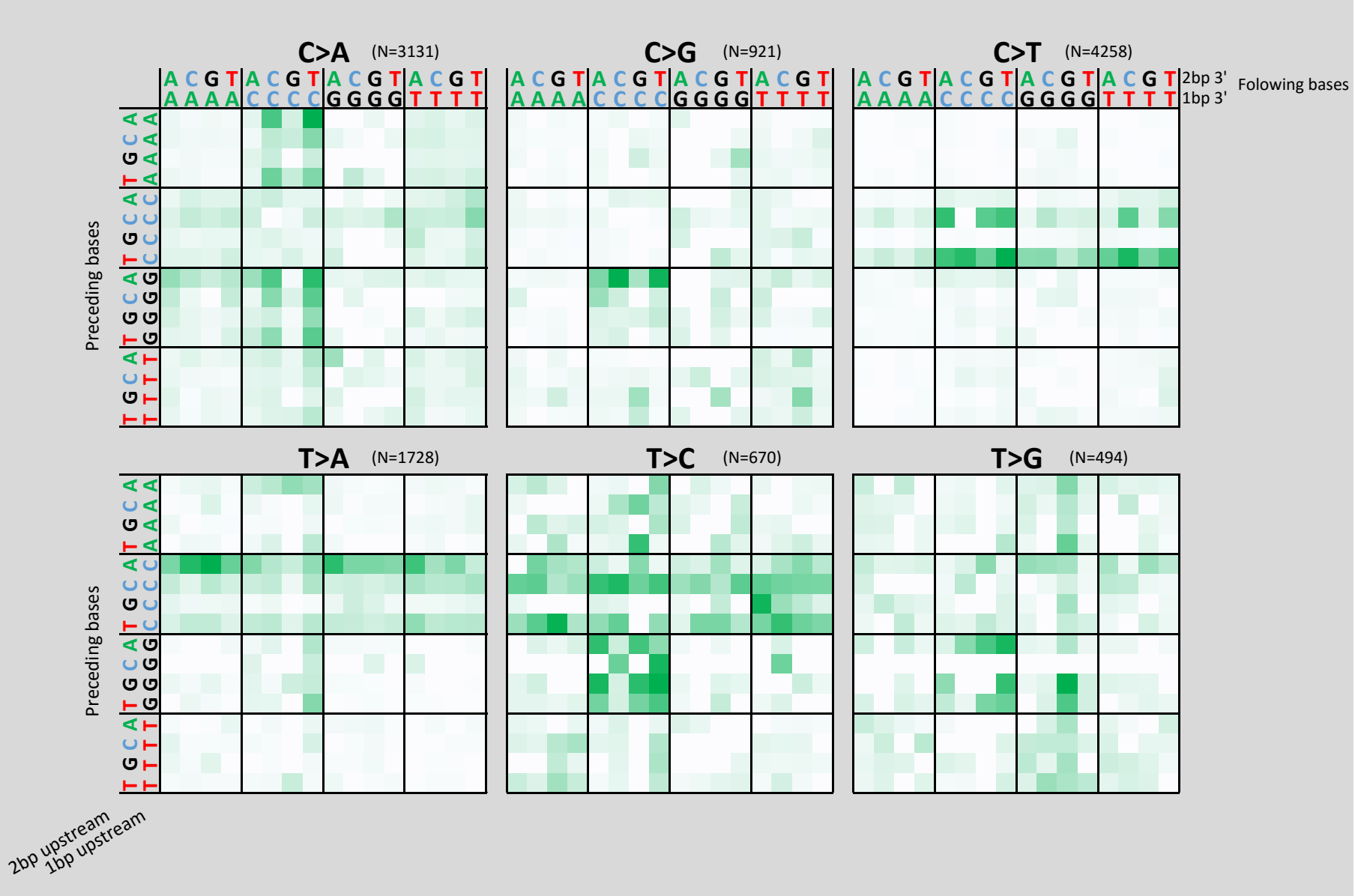
Supplemental Fig. S2G: Pentanucleotide context in HepG2_Cis_1



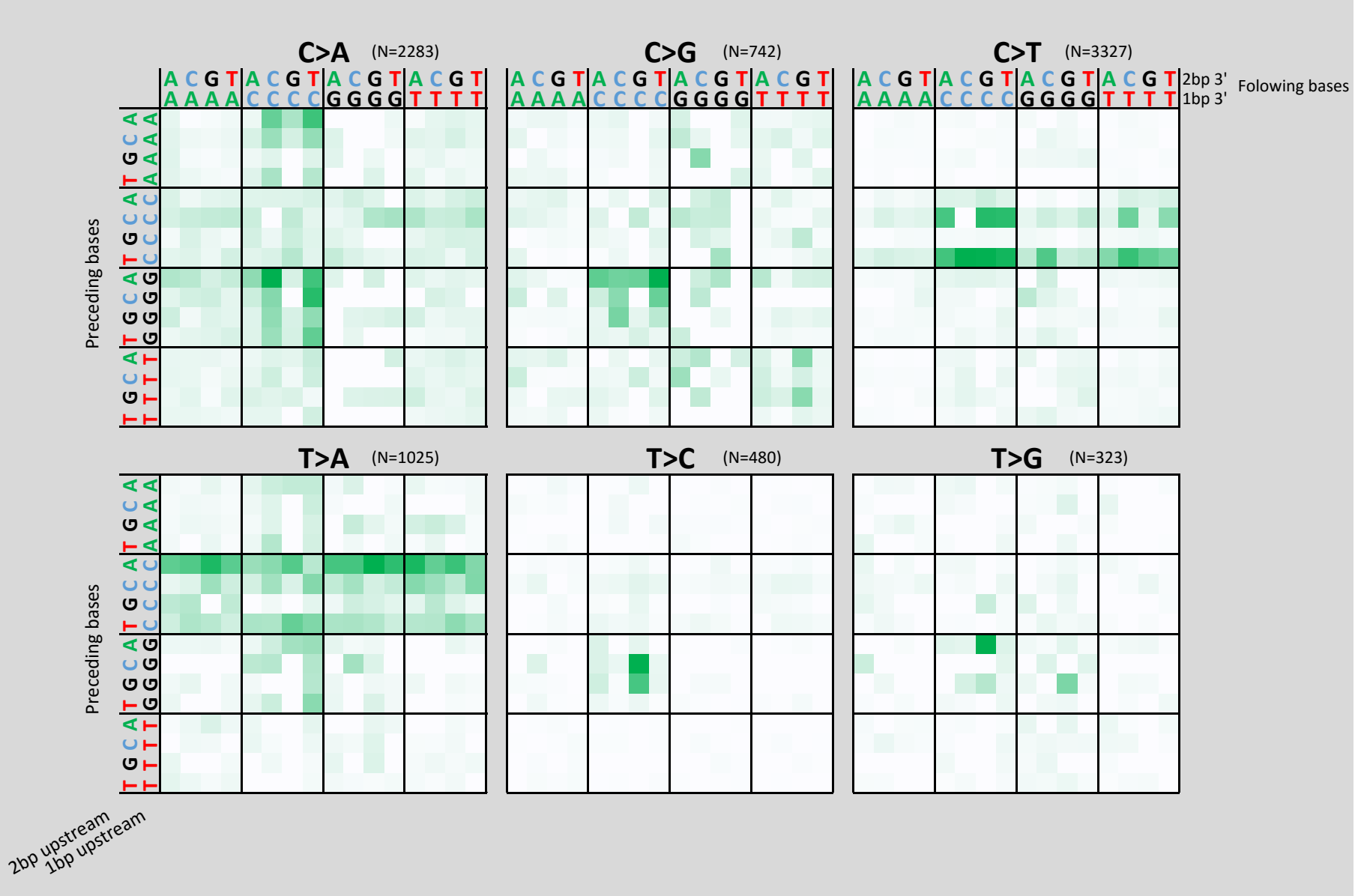
Supplemental Fig. S2H: Pentanucleotide context in HepG2_Cis_2



Supplemental Fig. S2I: Pentanucleotide context in HepG2_Cis_3

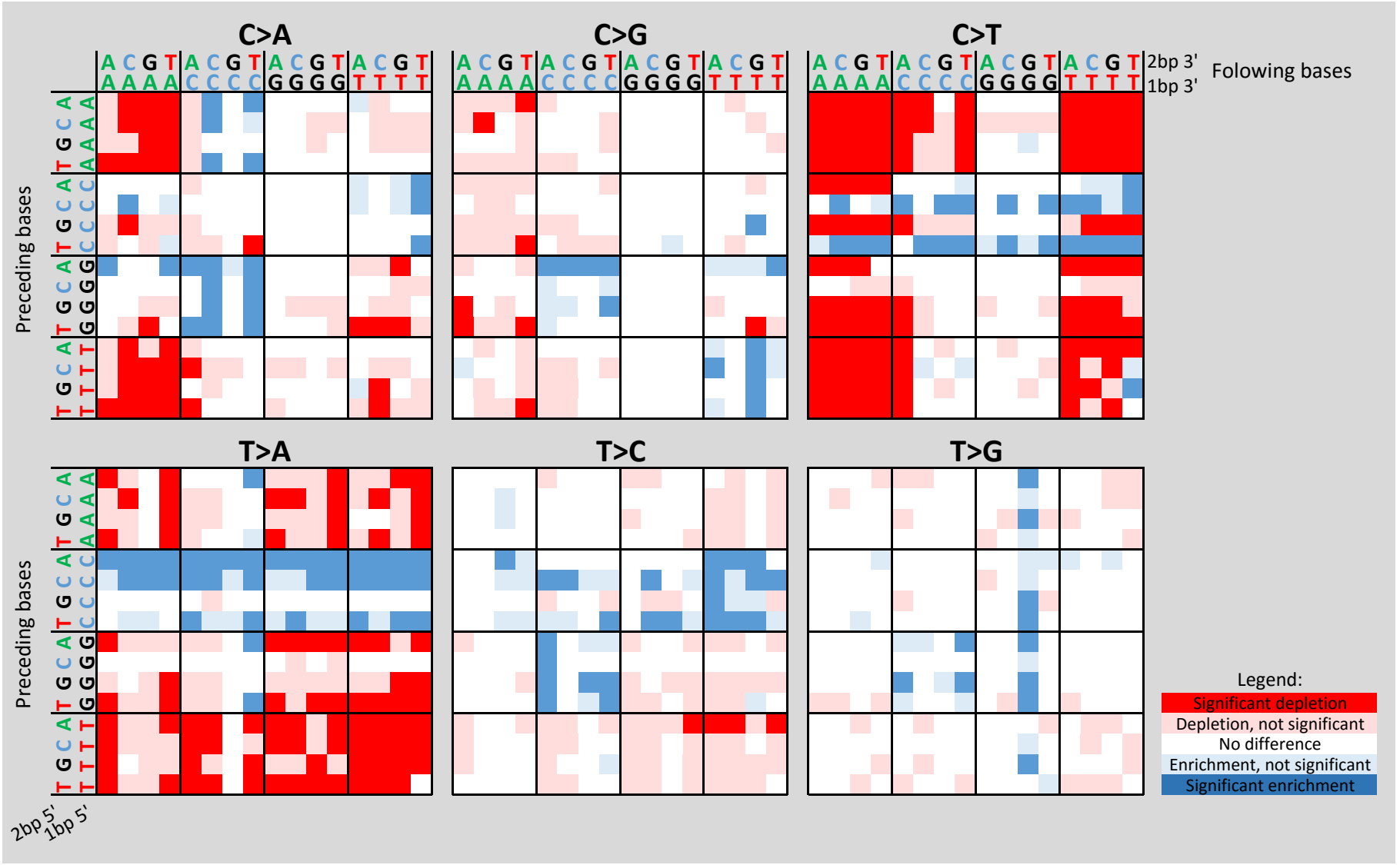


Supplemental Fig. S2J: Pentanucleotide context in HepG2_Cis_4

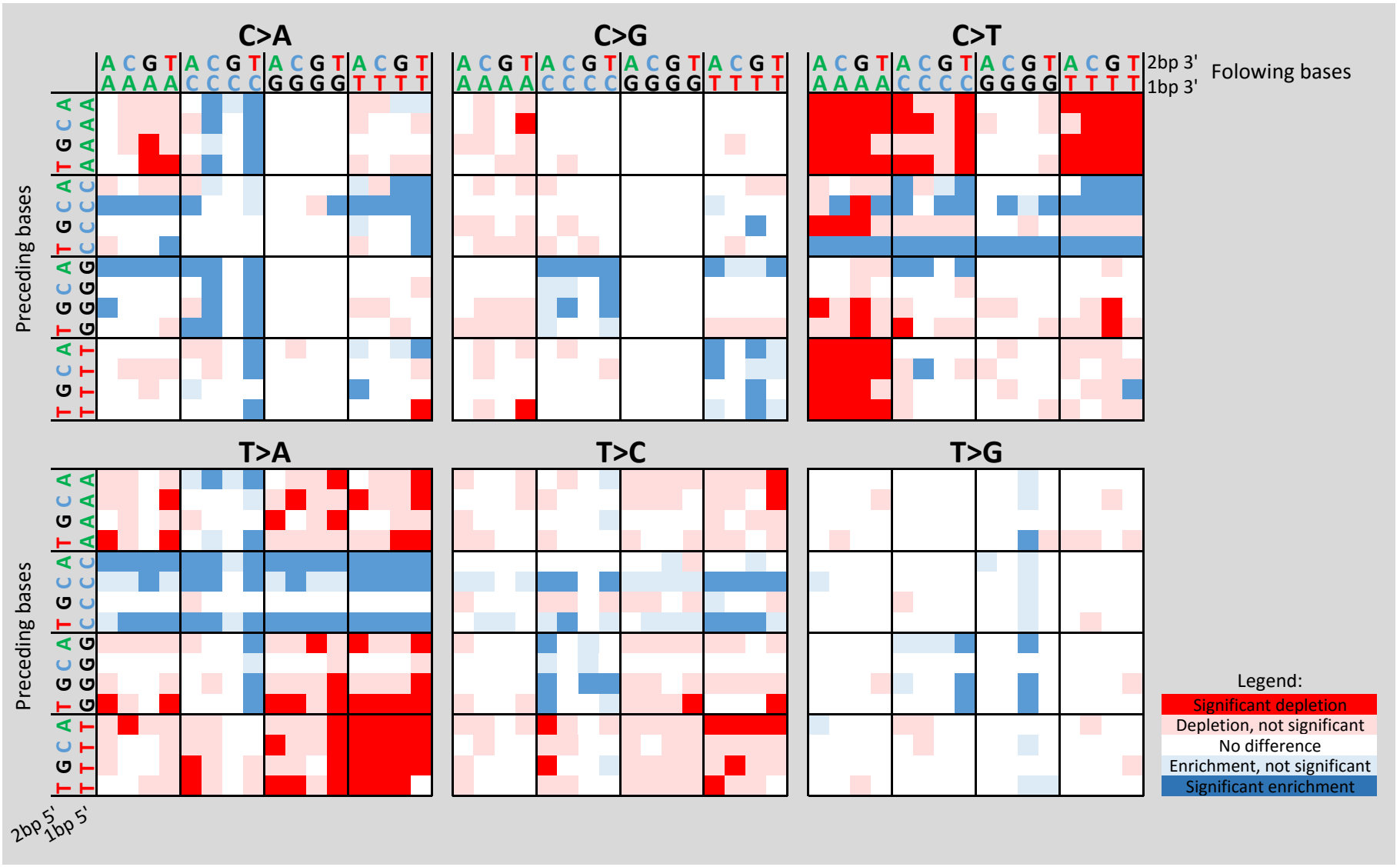


Supplemental Fig. S3: Graphical display of statistically significant enrichments and depletions of SNSs in pentanucleotide contexts in the combined MCF-10A (A) and HepG2 (B) data. Binomial tests were performed as described in Materials and Methods. Results were summarized as either enriched for mutations (blue) or depleted for mutations (red) or no difference from expected (white). Dark green and red indicate sequence contexts that were significant after Bonferroni multiple testing correction (i.e. $p < 0.05/1536$). Light blue and pink indicate sequence contexts that were not significant after correction for multiple testing (i.e. $0.05 > p > 0.05/1536$).

Supplemental Fig. S3A: Graphical display of pentanucleotides significantly enriched or depleted for SNSs in MCF-10A clones

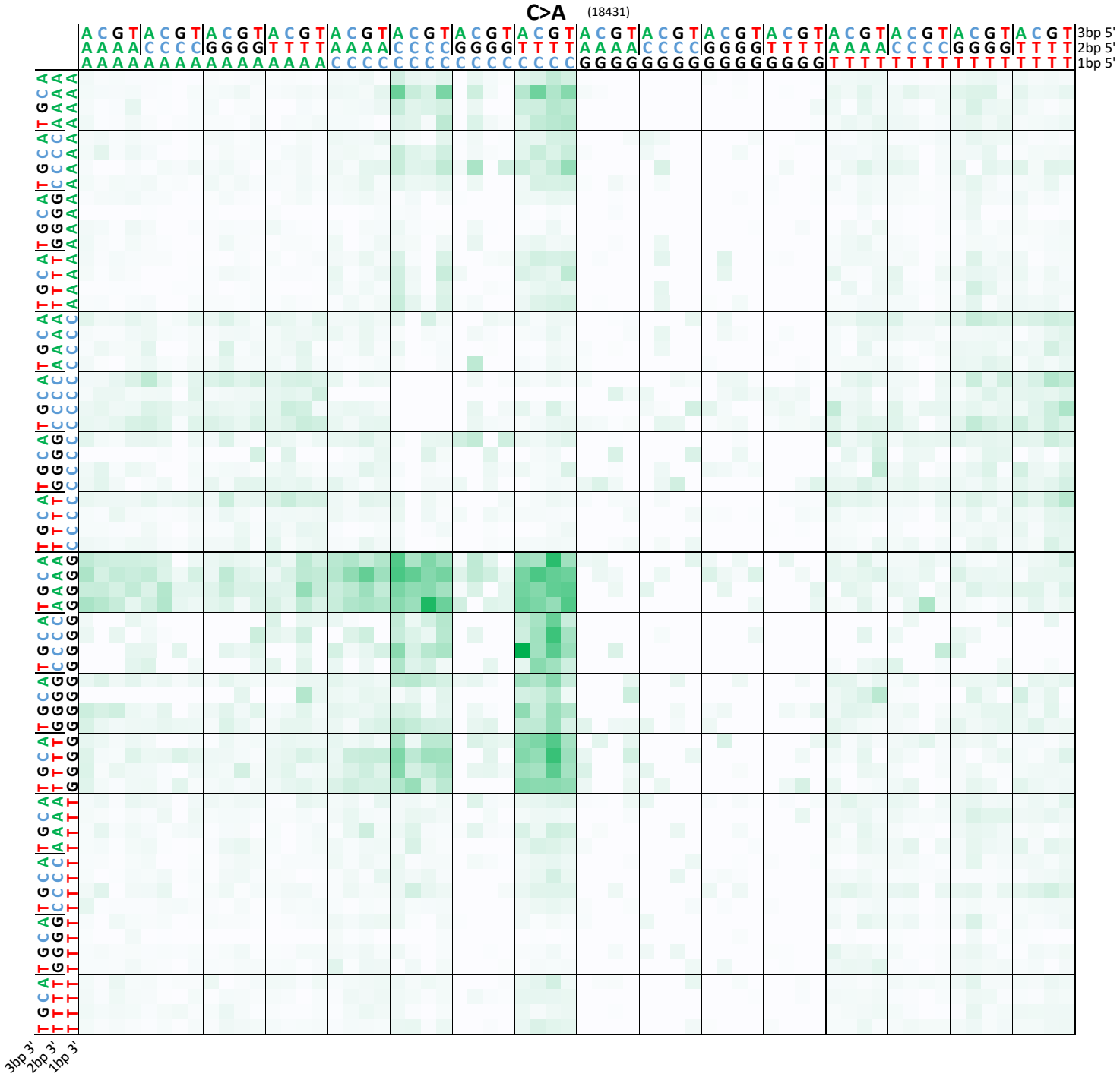


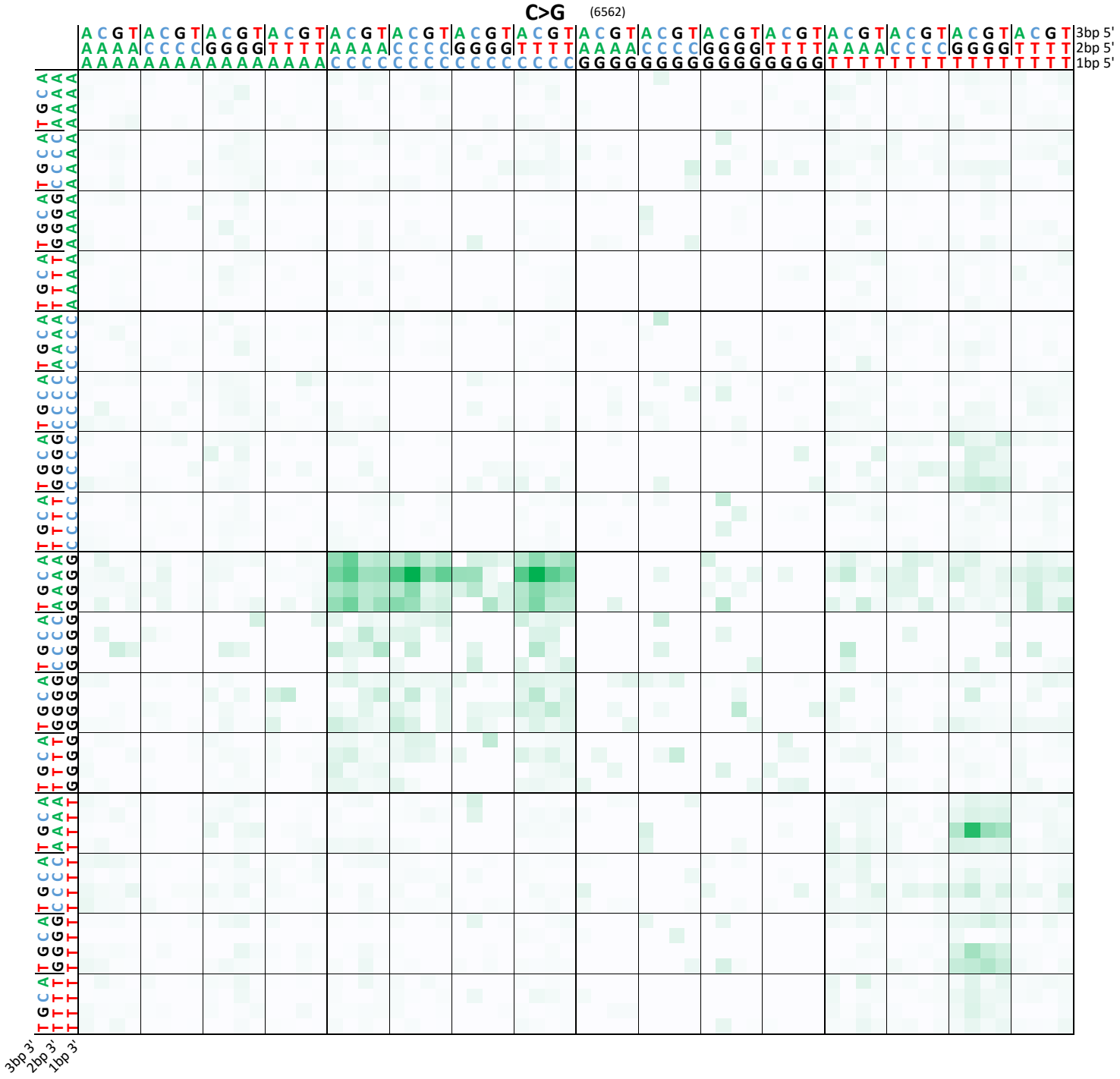
Supplemental Fig. S3B: Graphical display of pentanucleotides significantly enriched or depleted for SNSs in HepG2 clones

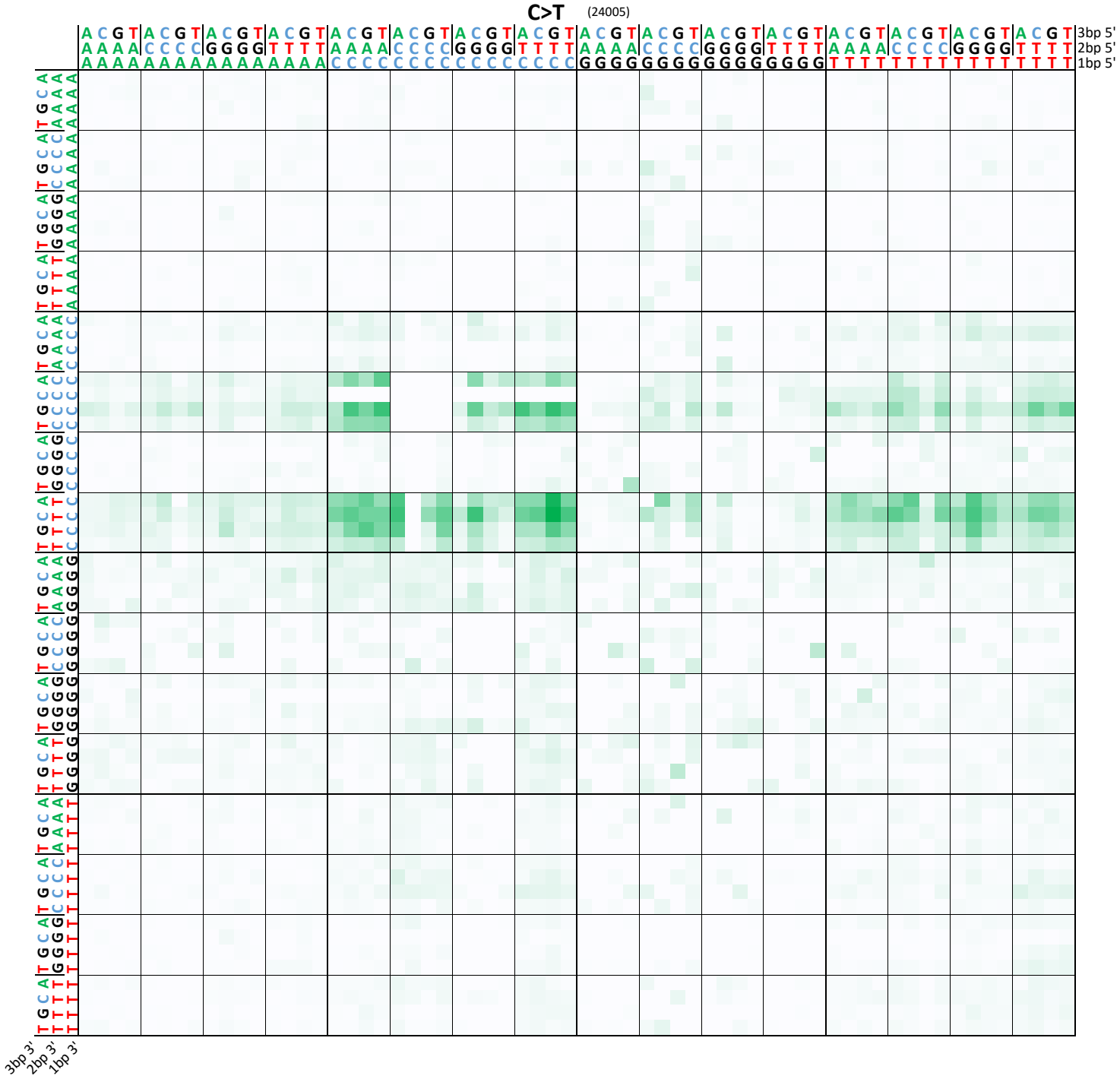


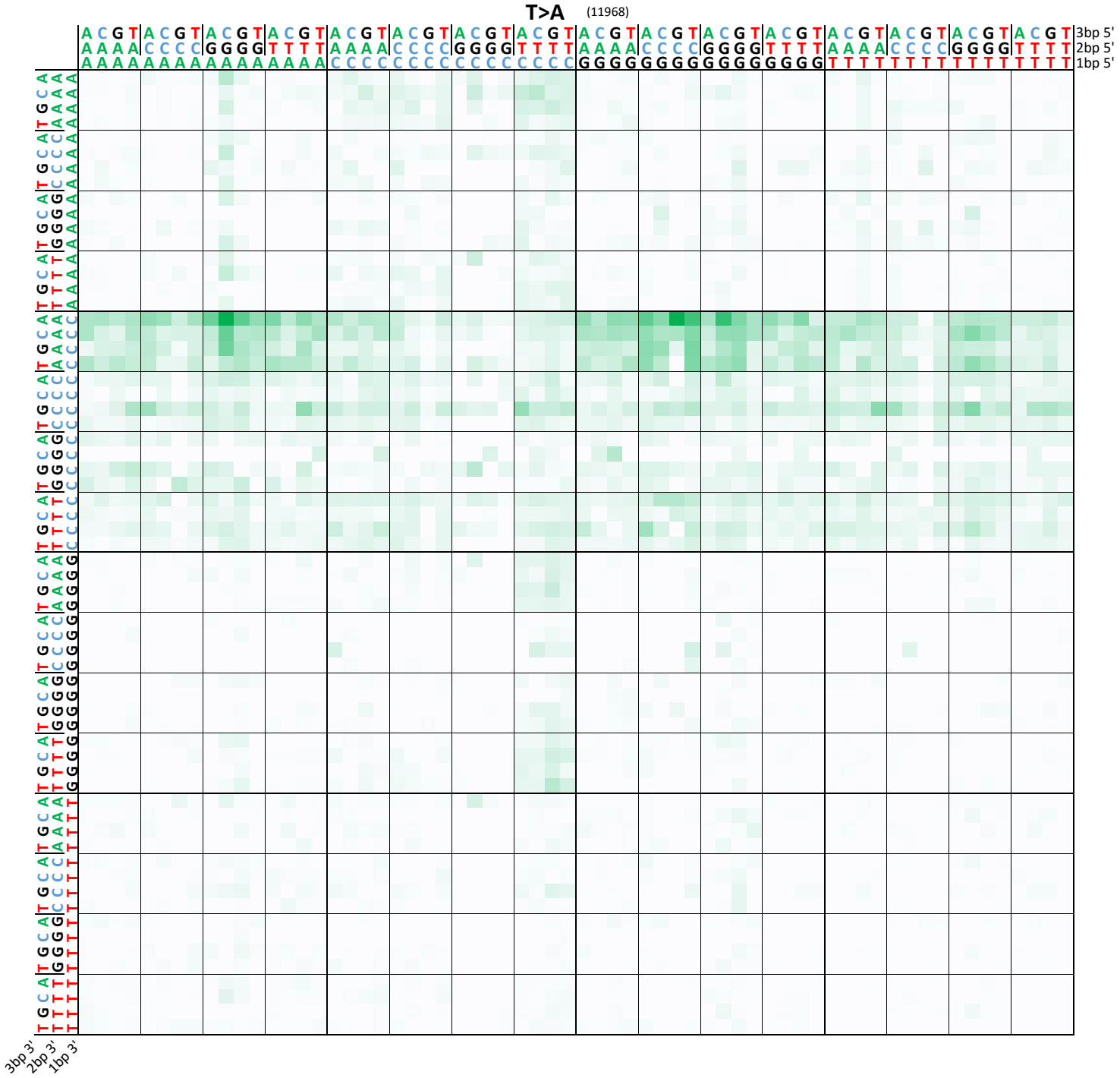
Supplemental Fig. S4: Heptanucleotide sequence context of cisplatin mutations in all HepG2 and MCF-10A clones combined (total 70,313 SNSs), normalized for heptanucleotide occurrence in the genome (A). Graphical display of statistically significant enrichments and depletions of SNSs in heptanucleotide contexts (B). Binomial tests were performed as described in Materials and Methods. Results were summarized as either enriched for mutations (blue) or depleted for mutations (red) or no difference from expected (white). Dark green and red indicate sequence contexts that were significant after Bonferroni multiple testing correction (i.e. $p < 0.05/24576$). Light blue and pink indicate sequence contexts that were not significant after correction for multiple testing (i.e. $0.05 > p > 0.05/24576$).

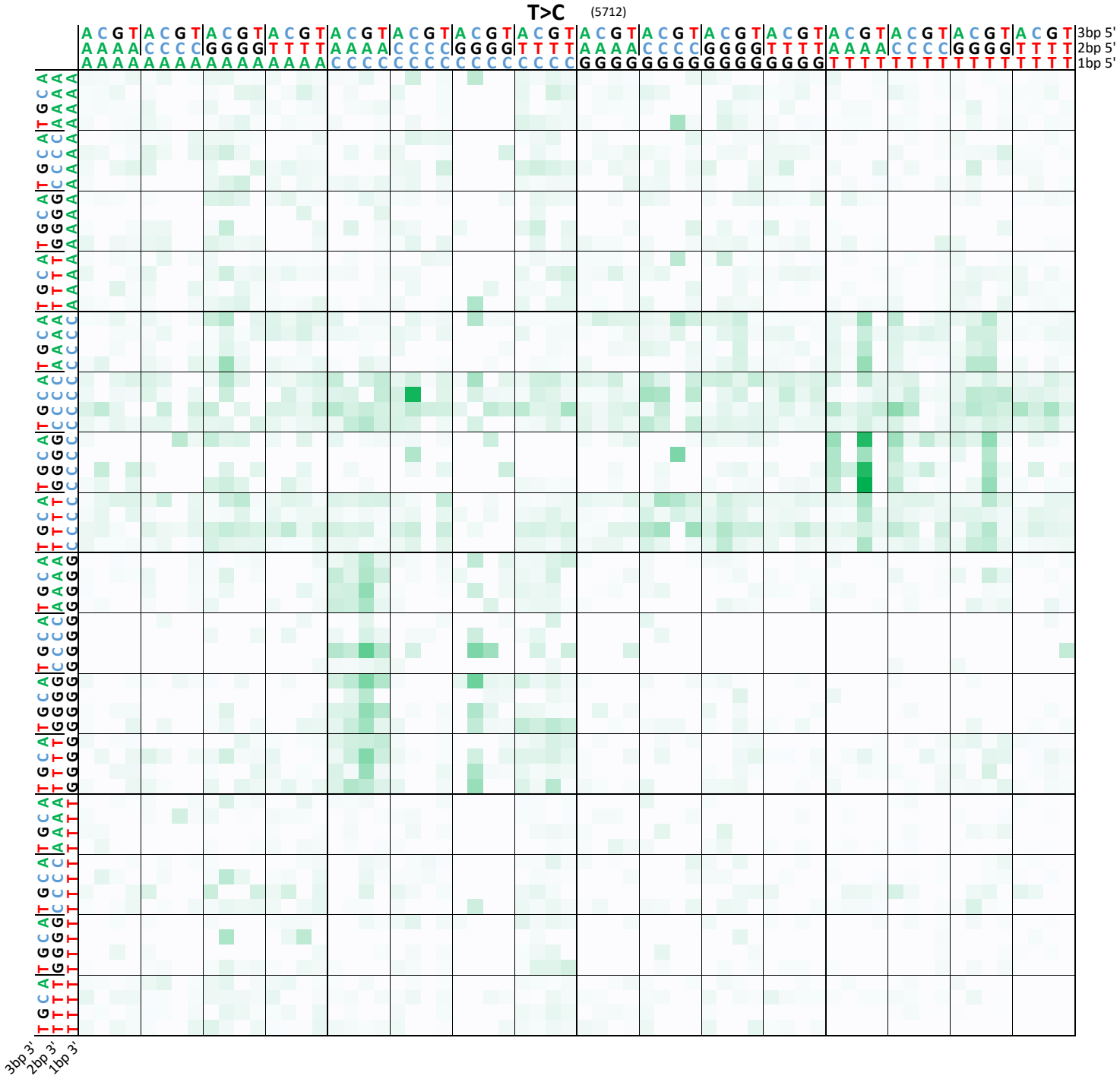
Supplemental Fig. S4A: Graphical representation of heptanucleotide context in cisplatin treated cell line clones.

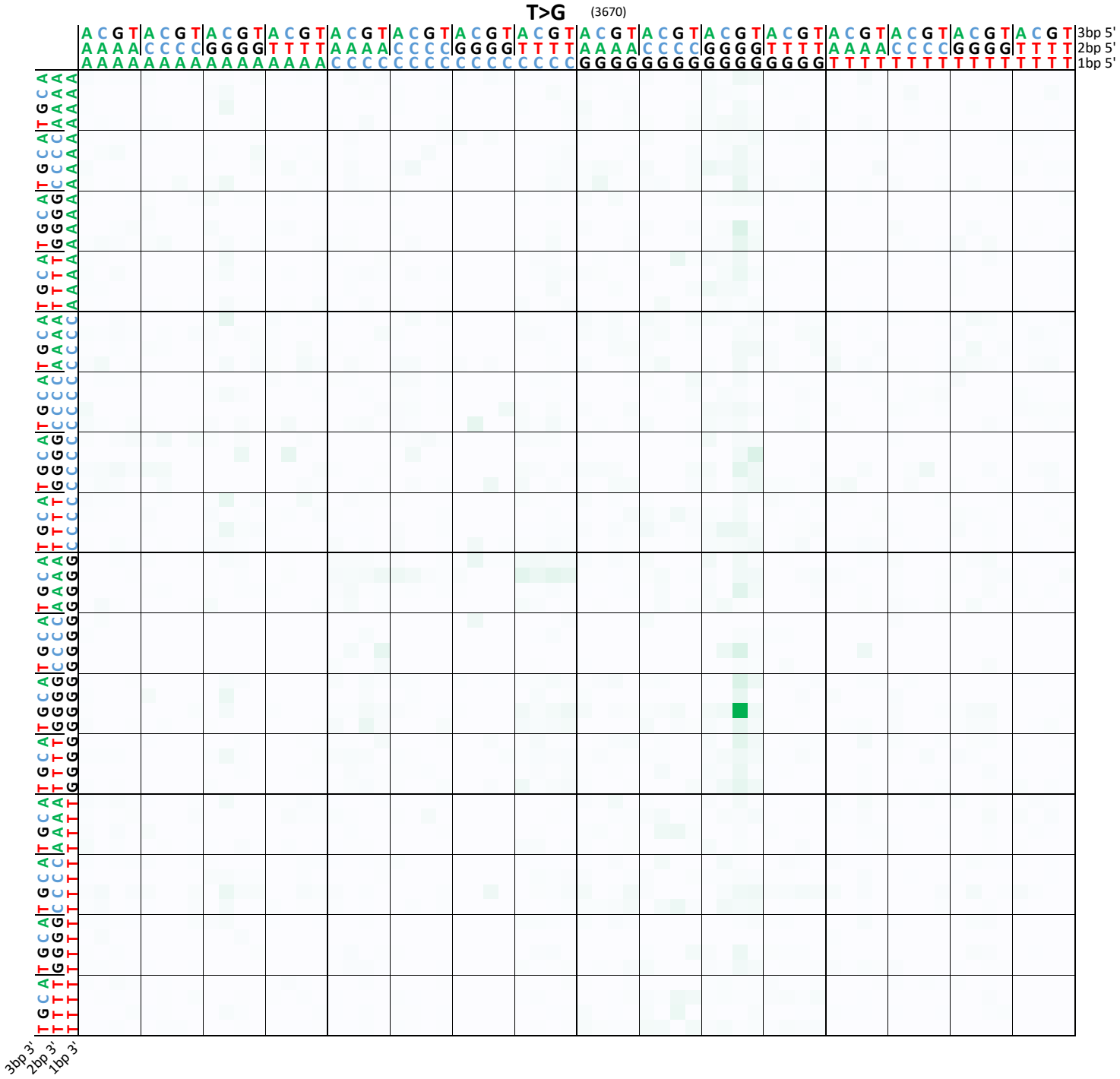



C>G
(6562)
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'

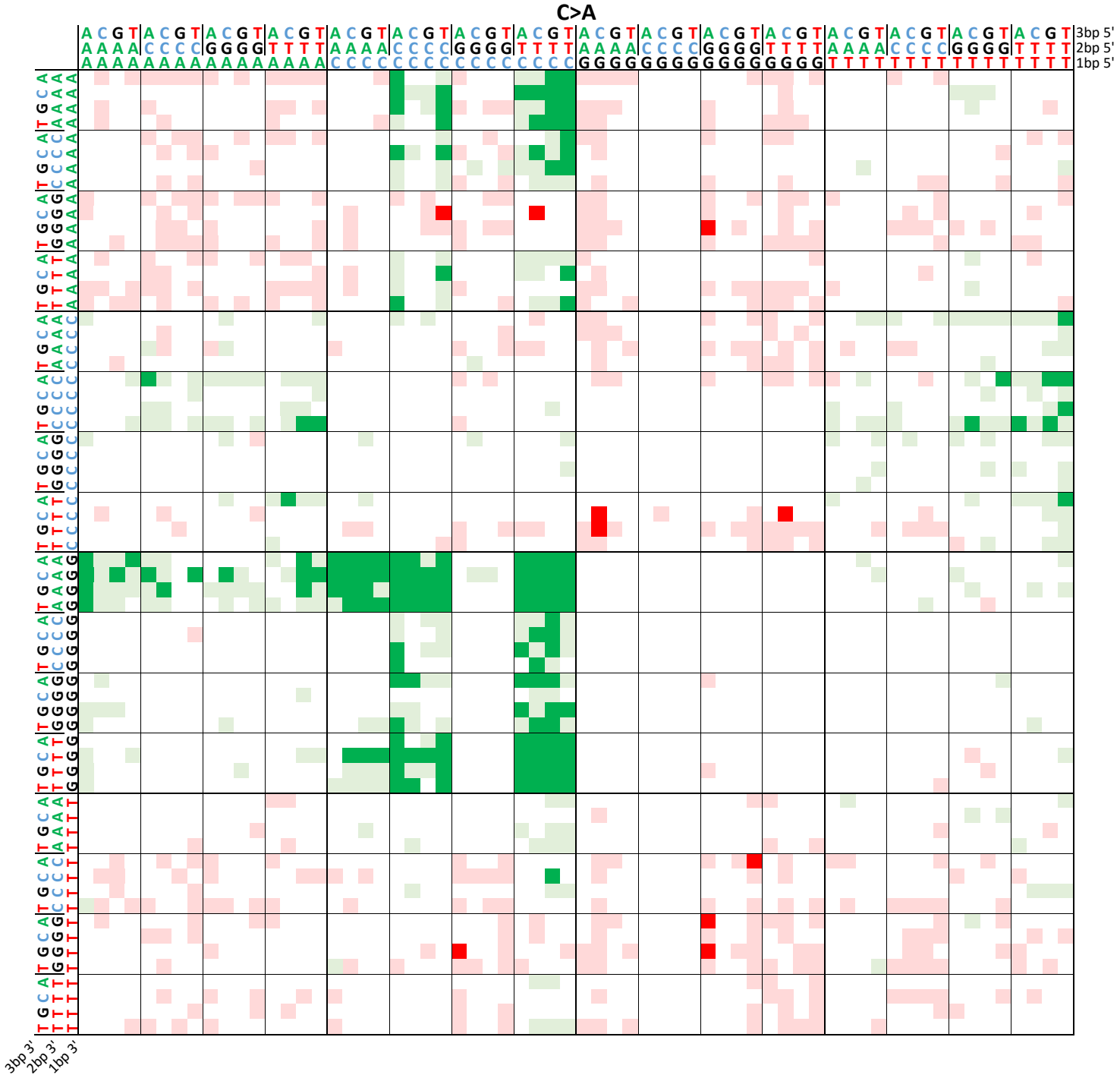

C>T
(24005)
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'

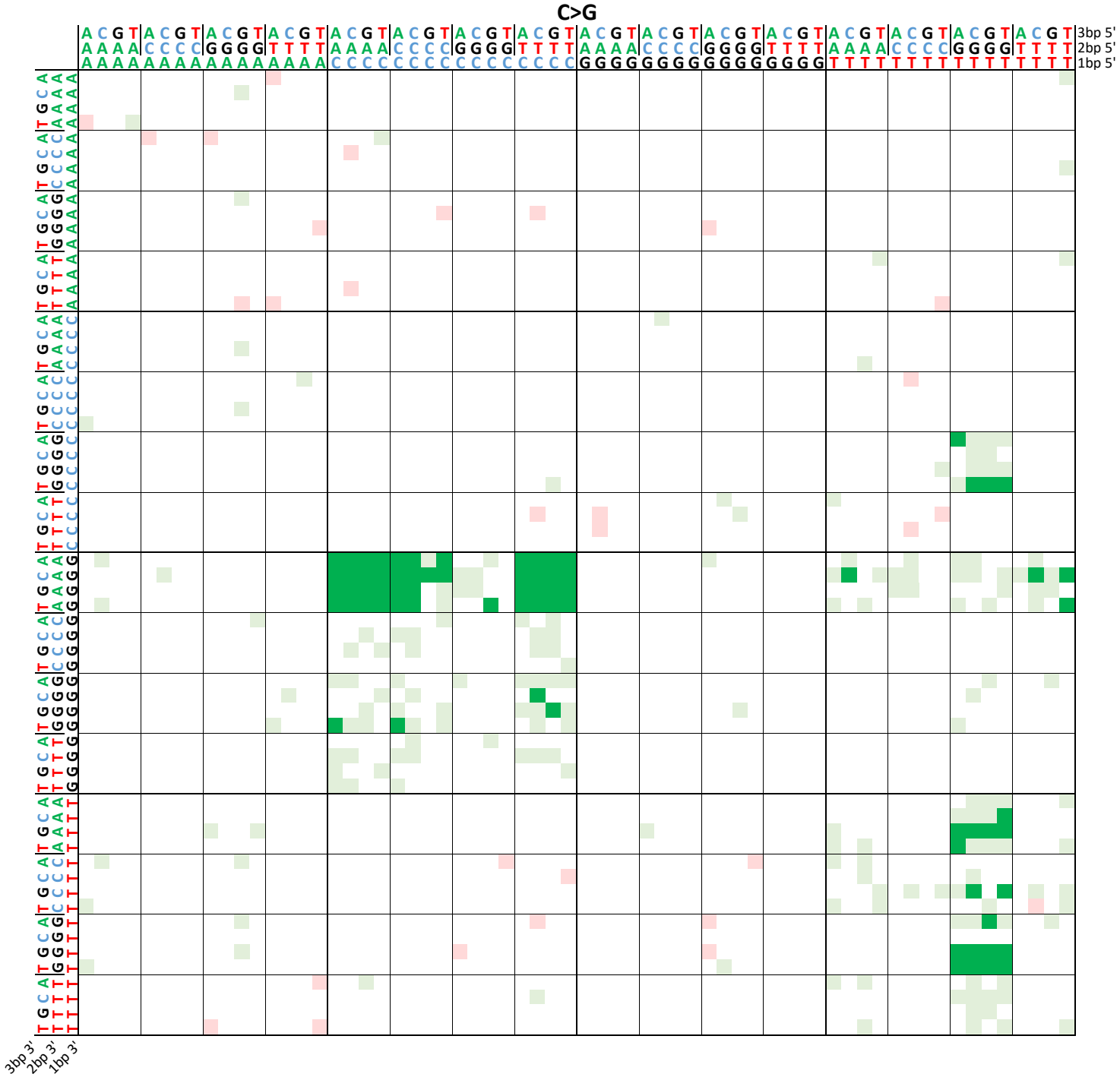

T>A
(11968)
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'

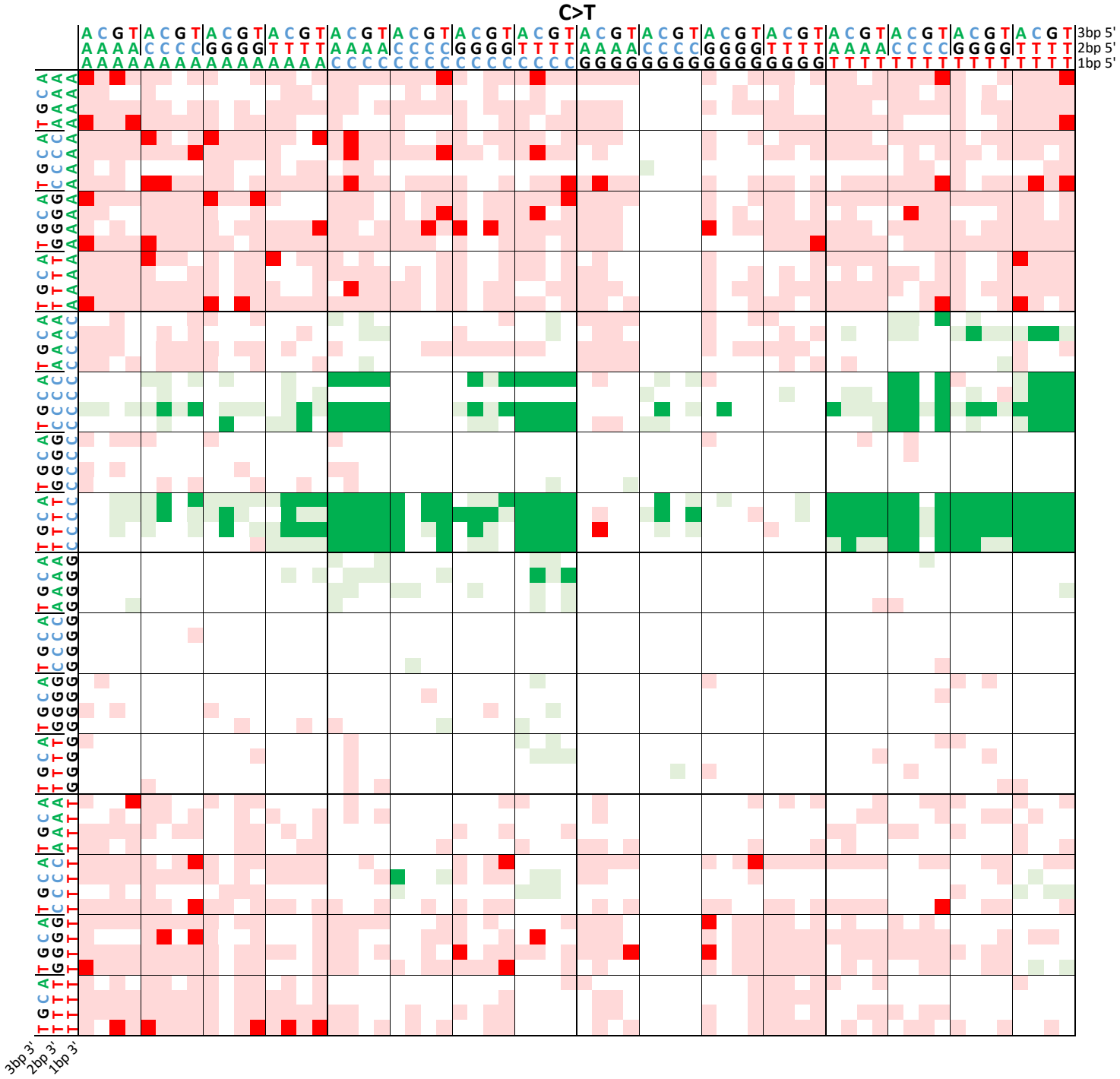

T>C
(5712)
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'

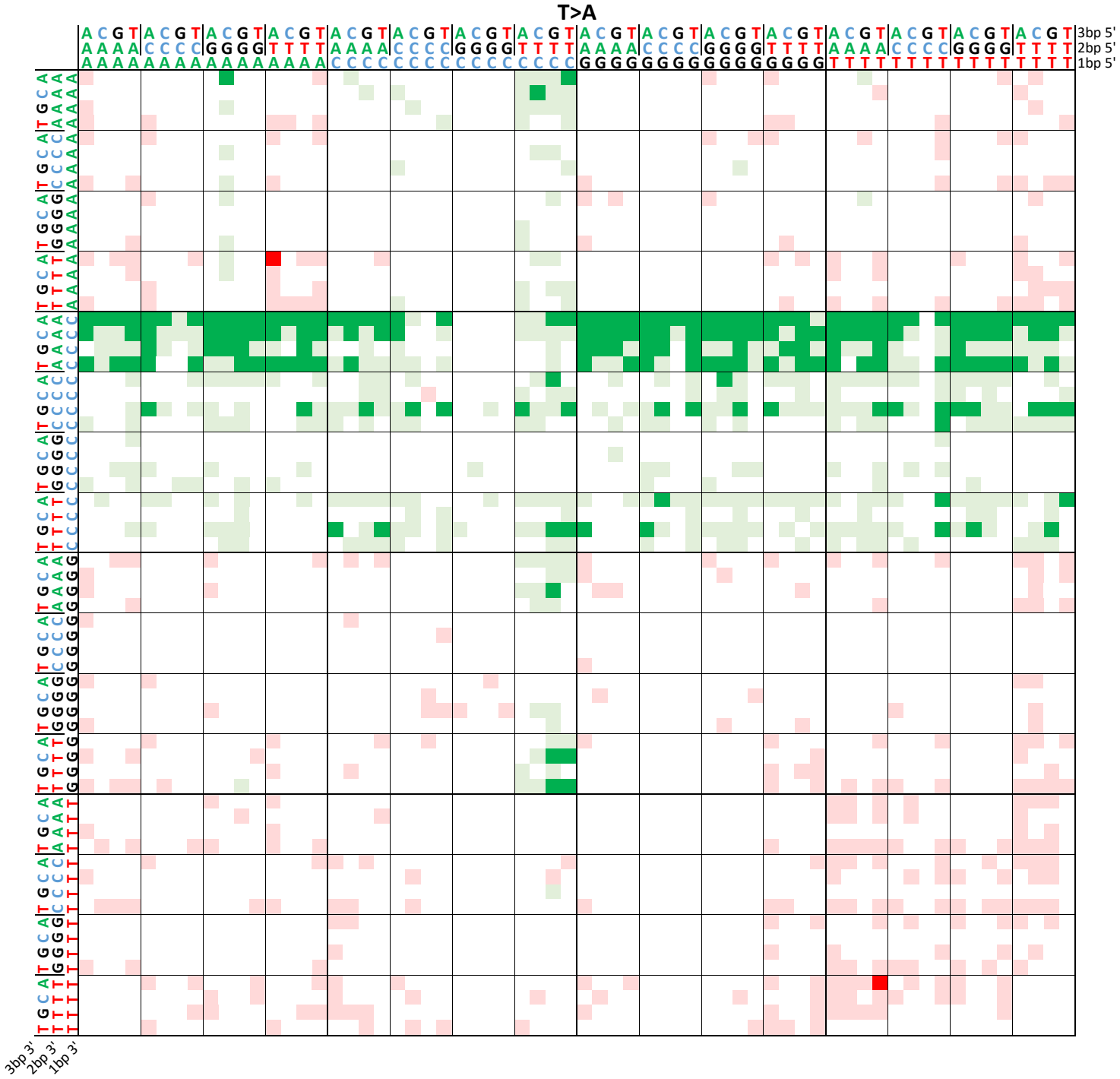

T>G
(3670)
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'

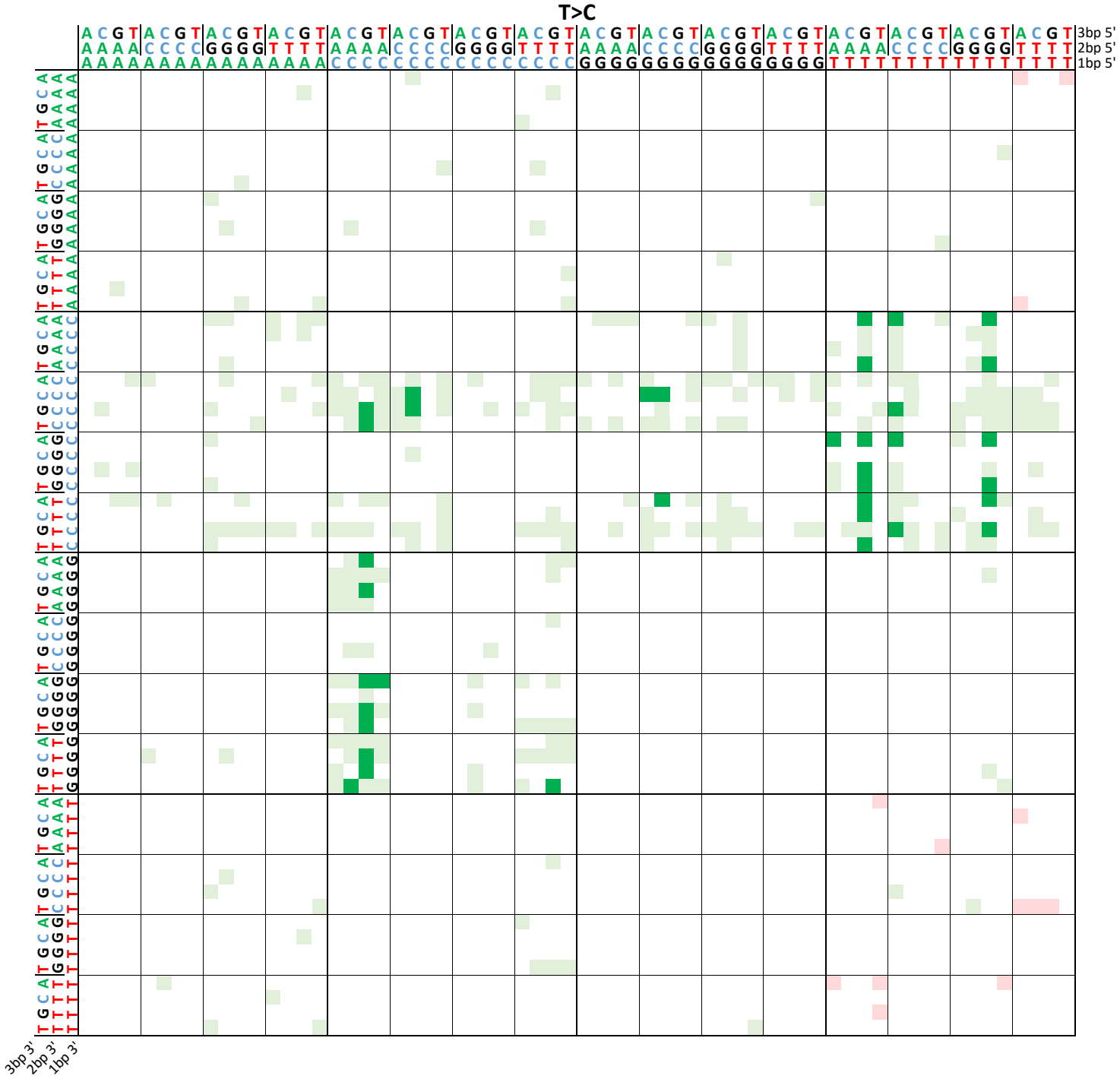
Supplemental Fig. S4B: Graphical display of heptanucleotides significantly enriched or depleted for SBSs in cisplatin treated cell line clones

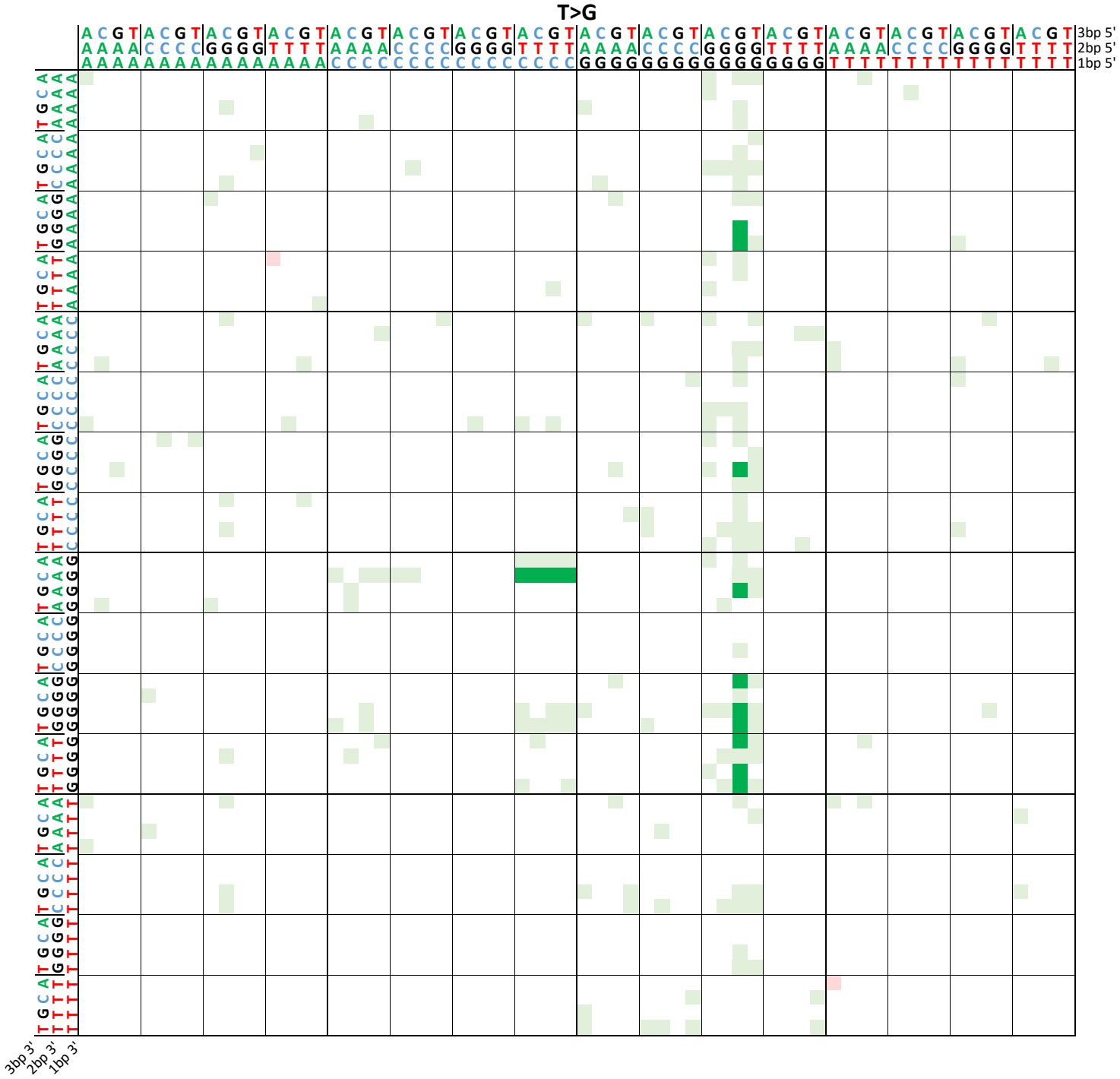



C>G
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'


C>T
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'

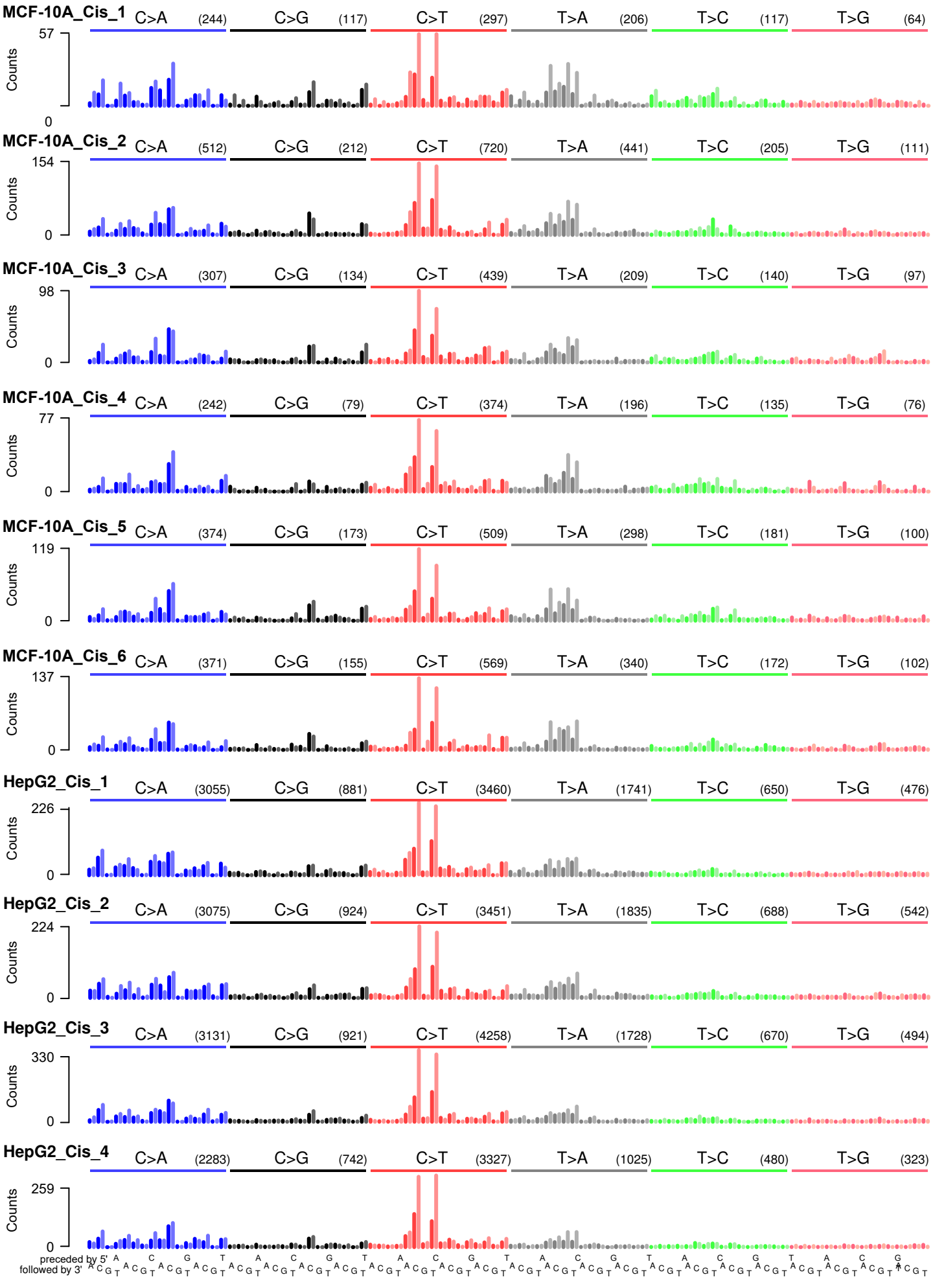

T>A
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'


T>C
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'

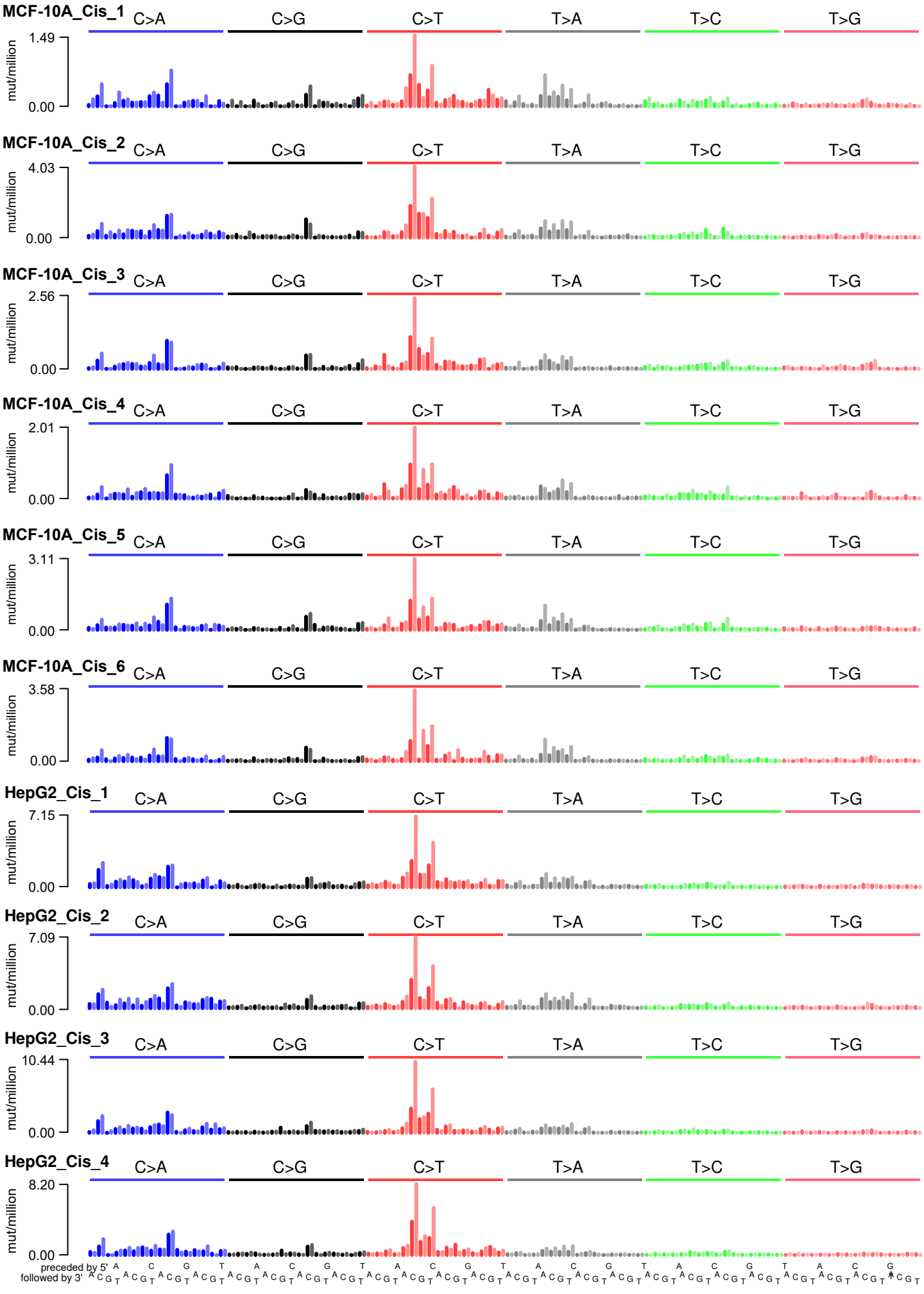

T>G
3bp 5'
2bp 5'
1bp 5'
3bp 3'
2bp 3'
1bp 3'

Supplemental Fig. S5: Transcription strand bias for each of the 96-channels of the mutation spectrum for each of the cisplatin treated MCF-10A and HepG2 clones, both displayed as raw counts (A) and mutations per million trinucleotides (B). Dark = antisense (transcribed) strand, light = sense (untranscribed) strand.

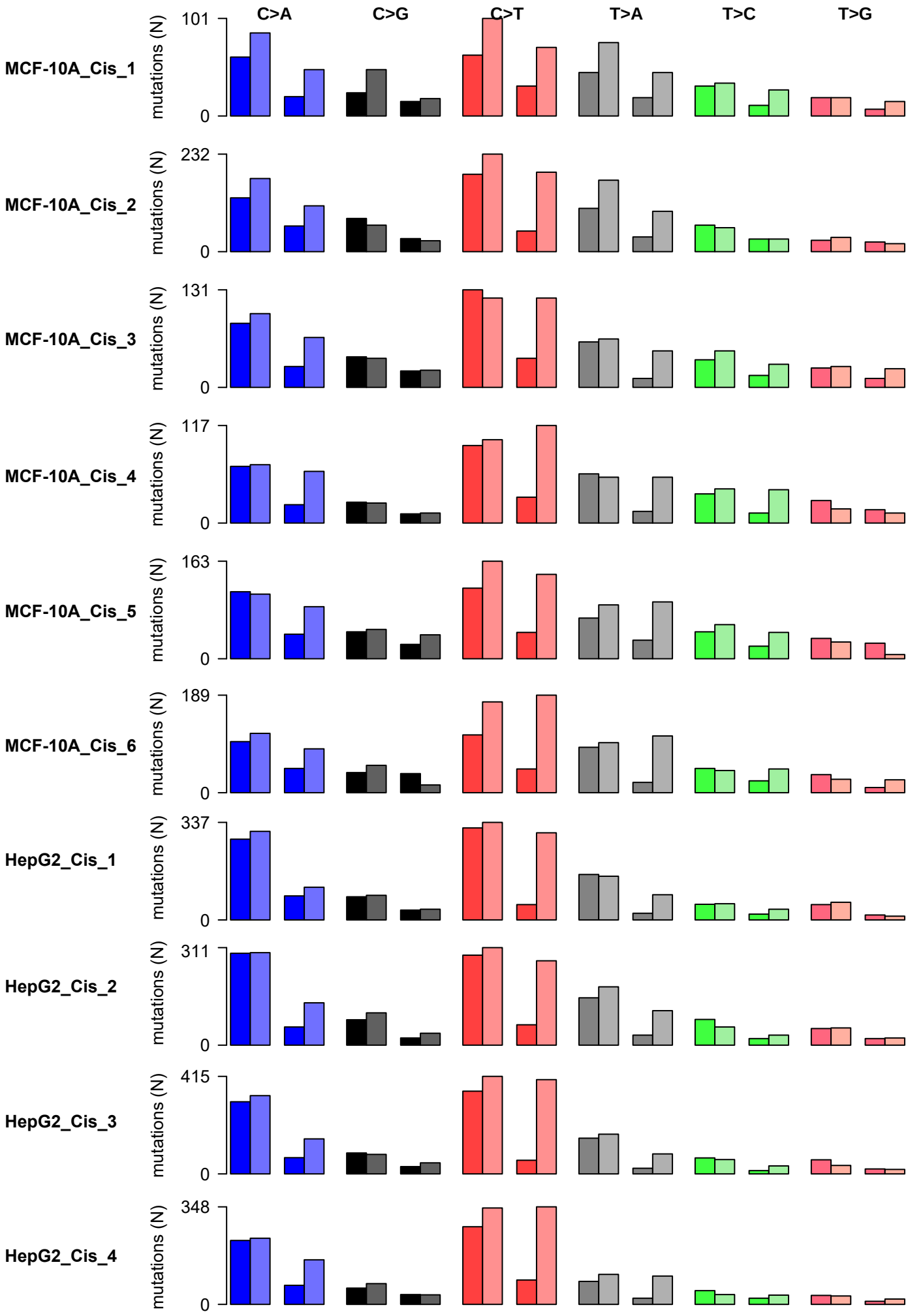
Supplemental Fig. S5A: Transcription strand bias in cisplatin treated cell line clones



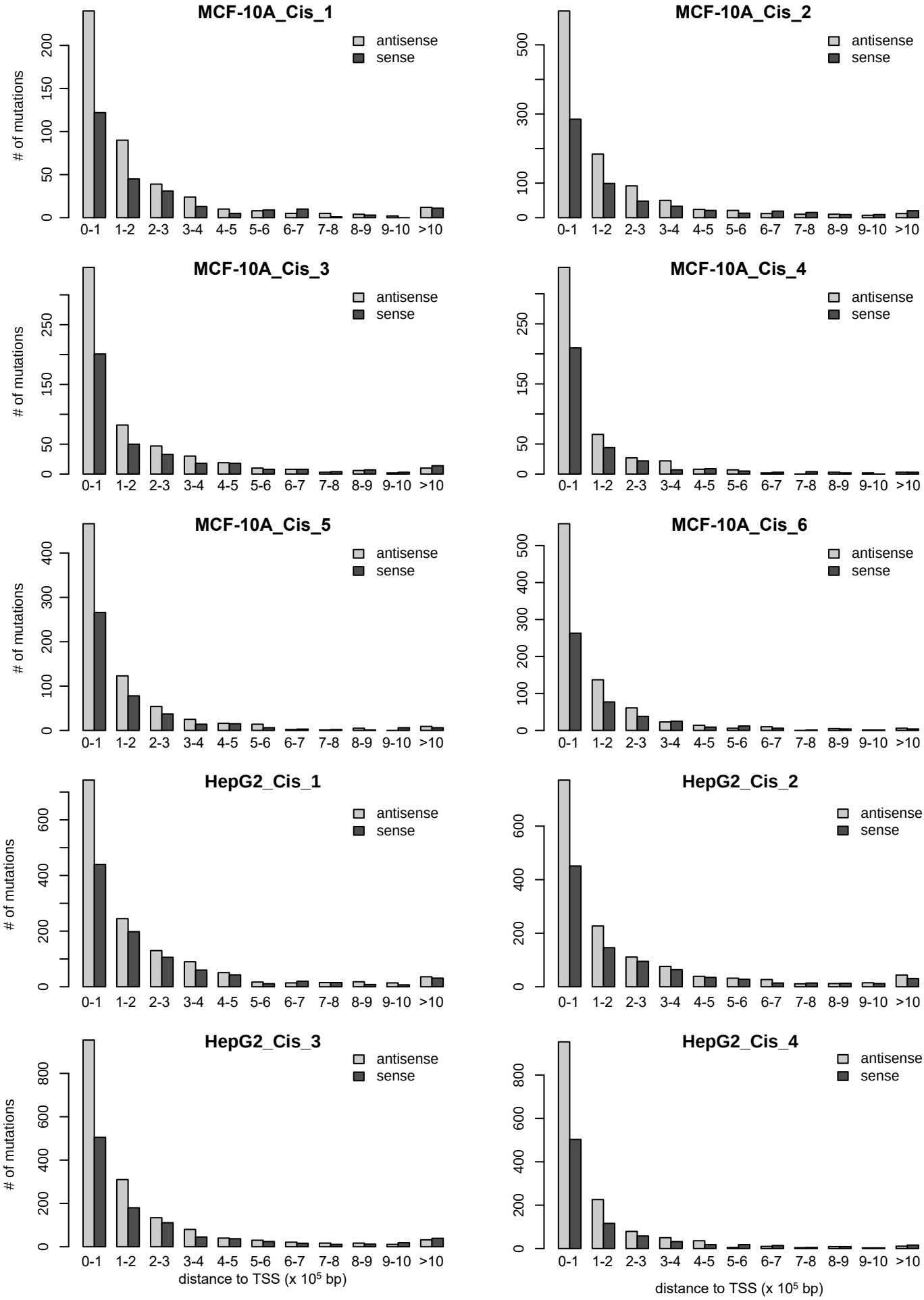
Supplemental Fig. S5B: Transcription strand bias in cisplatin treated cell line clones, mutations per million trinucleotides



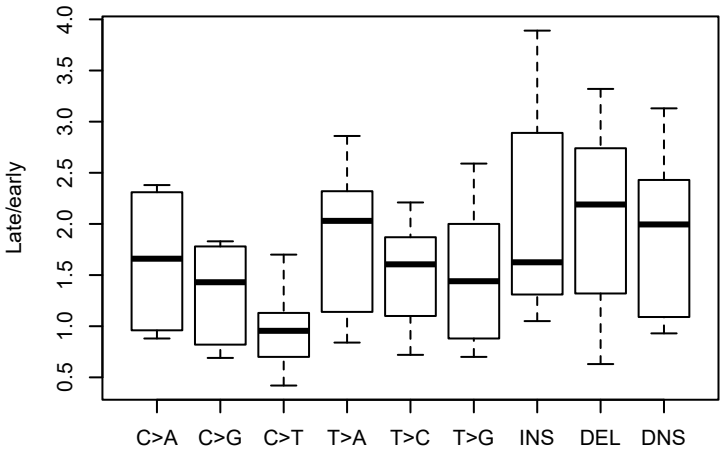
Supplemental Fig. S6: Transcription strand bias as a function of gene expression for each of the cisplatin treated MCF-10A and HepG2 clones. Genes were divided in either low or highly expressed (using the median). Transcription strand bias was plotted for each of the mutation classes. Dark = sense (untranscribed) strand, light = antisense (transcribed) strand.



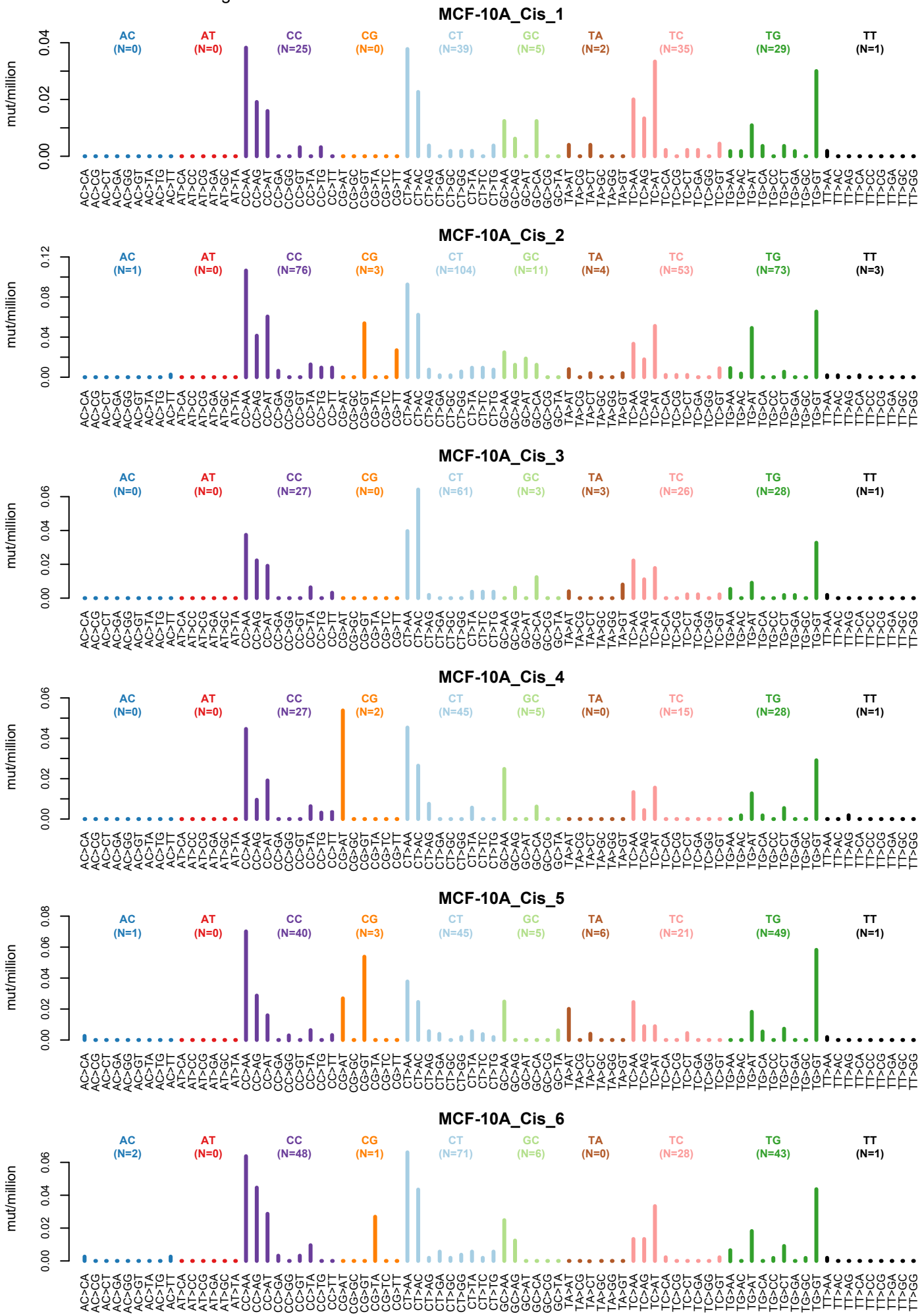
Supplemental Fig. S7: Transcription strand bias as a function of distance to the transcription start site in cisplatin treated MCF-10A and HepG2 clones. Mutations were binned according to their distance to the respective TSS of their respective genes. Bin 1 is 0 to 100,000 bp away from the TSS, bin 2 is 100,001 to 200,000 bp away from the TSS and so forth. Dark = sense (untranscribed) strand, light = antisense (transcribed) strand.

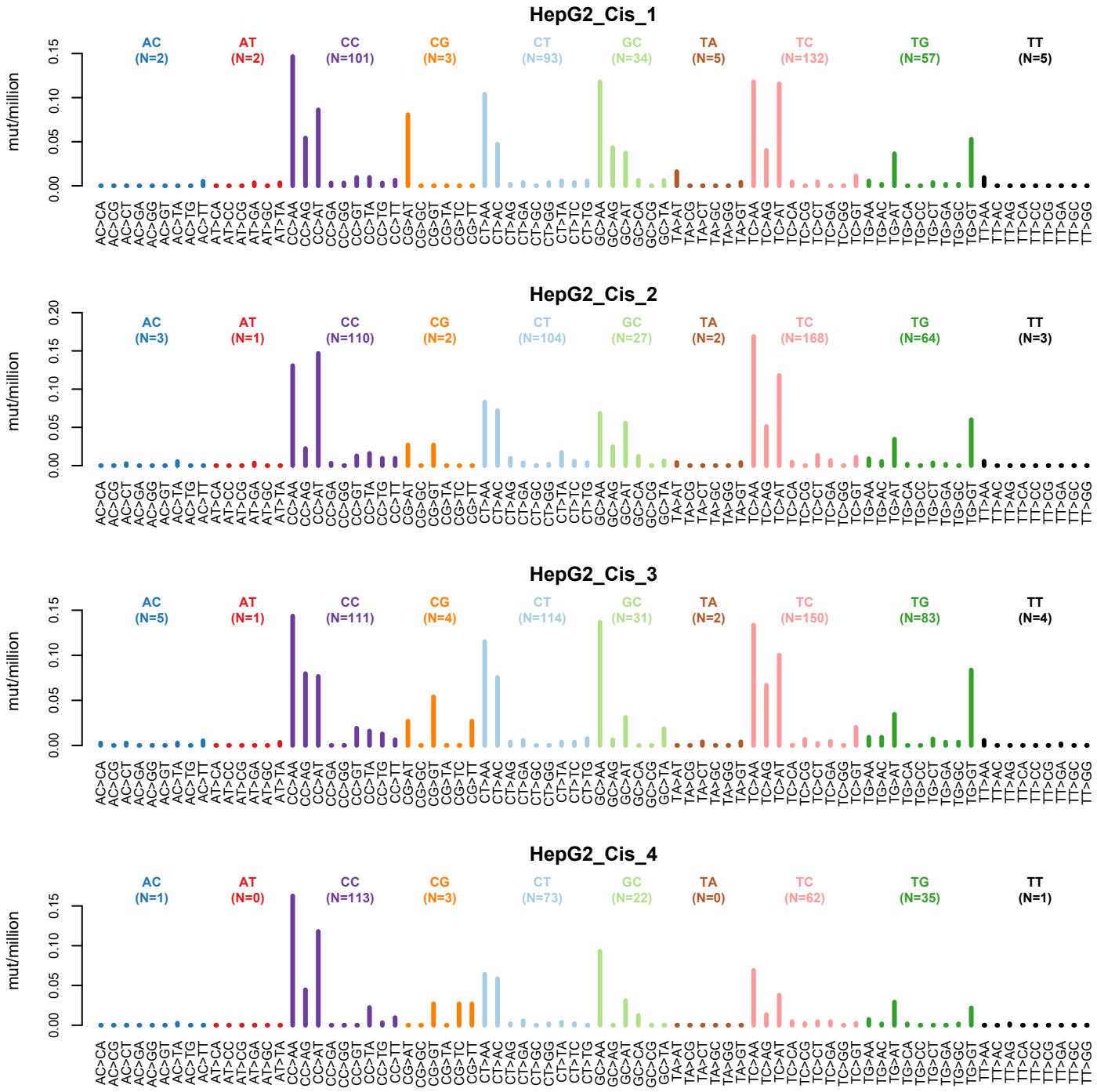


Supplemental Fig. S8: Replication timing bias of cisplatin mutations in all 10 cisplatin treated cell line clones



Supplemental Fig. S9: DNS mutational spectra of cisplatin treated MCF-10A and HepG2 clones per mutations per million dinucleotides in the genome.

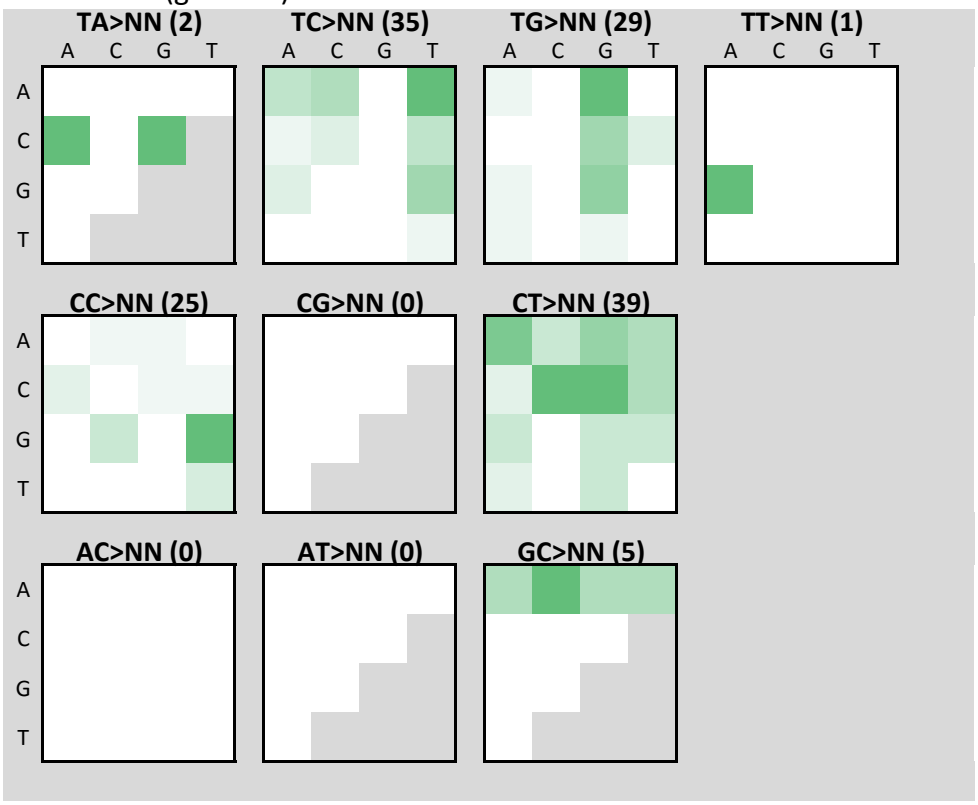



HepG2_Cis_1
mut/million
0.00
0.05
0.10
0.15
AC (N=2)
AT (N=2)
CC (N=101)
CG (N=3)
CT (N=93)
GC (N=34)
TA (N=5)
TC (N=132)
TG (N=57)
TT (N=5)
AC>CA AC>CG AC>CT AC>GA AC>GG AC>GT AC>TA AC>TG AC>TT AT>CA AT>CC AT>CG AT>GA AT>GC AT>TA CC>AA CC>AG CC>AT CC>GA CC>GG CC>GT CC>TA CC>TG CC>TT CG>AT CG>GC CG>GT CG>TA CG>TC CG>TT CT>AA CT>AC CT>AG CT>GA CT>GC CT>GG CT>TA CT>TC CT>TG GC>AA GC>AG GC>AT GC>CA GC>CG GC>TA TA>AT TA>CG TA>CT TA>GC TA>GG TA>GT TC>AA TC>AG TC>AT TC>CA TC>CG TC>CT TC>GA TC>GG TC>GT TG>AA TG>AC TG>AT TG>CA TG>CC TG>CT TG>GA TG>GC TG>GT TT>AA TT>AC TT>AG TT>CA TT>CC TT>CG TT>GA TT>GC TT>GG
HepG2_Cis_2
mut/million
0.00
0.05
0.10
0.15
0.20
AC (N=3)
AT (N=1)
CC (N=110)
CG (N=2)
CT (N=104)
GC (N=27)
TA (N=2)
TC (N=168)
TG (N=64)
TT (N=3)
AC>CA AC>CG AC>CT AC>GA AC>GG AC>GT AC>TA AC>TG AC>TT AT>CA AT>CC AT>CG AT>GA AT>GC AT>TA CC>AA CC>AG CC>AT CC>GA CC>GG CC>GT CC>TA CC>TG CC>TT CG>AT CG>GC CG>GT CG>TA CG>TC CG>TT CT>AA CT>AC CT>AG CT>GA CT>GC CT>GG CT>TA CT>TC CT>TG GC>AA GC>AG GC>AT GC>CA GC>CG GC>TA TA>AT TA>CG TA>CT TA>GC TA>GG TA>GT TC>AA TC>AG TC>AT TC>CA TC>CG TC>CT TC>GA TC>GG TC>GT TG>AA TG>AC TG>AT TG>CA TG>CC TG>CT TG>GA TG>GC TG>GT TT>AA TT>AC TT>AG TT>CA TT>CC TT>CG TT>GA TT>GC TT>GG
HepG2_Cis_3
mut/million
0.00
0.05
0.10
0.15
AC (N=5)
AT (N=1)
CC (N=111)
CG (N=4)
CT (N=114)
GC (N=31)
TA (N=2)
TC (N=150)
TG (N=83)
TT (N=4)
AC>CA AC>CG AC>CT AC>GA AC>GG AC>GT AC>TA AC>TG AC>TT AT>CA AT>CC AT>CG AT>GA AT>GC AT>TA CC>AA CC>AG CC>AT CC>GA CC>GG CC>GT CC>TA CC>TG CC>TT CG>AT CG>GC CG>GT CG>TA CG>TC CG>TT CT>AA CT>AC CT>AG CT>GA CT>GC CT>GG CT>TA CT>TC CT>TG GC>AA GC>AG GC>AT GC>CA GC>CG GC>TA TA>AT TA>CG TA>CT TA>GC TA>GG TA>GT TC>AA TC>AG TC>AT TC>CA TC>CG TC>CT TC>GA TC>GG TC>GT TG>AA TG>AC TG>AT TG>CA TG>CC TG>CT TG>GA TG>GC TG>GT TT>AA TT>AC TT>AG TT>CA TT>CC TT>CG TT>GA TT>GC TT>GG
HepG2_Cis_4
mut/million
0.00
0.05
0.10
0.15
AC (N=1)
AT (N=0)
CC (N=113)
CG (N=3)
CT (N=73)
GC (N=22)
TA (N=0)
TC (N=62)
TG (N=35)
TT (N=1)
AC>CA AC>CG AC>CT AC>GA AC>GG AC>GT AC>TA AC>TG AC>TT AT>CA AT>CC AT>CG AT>GA AT>GC AT>TA CC>AA CC>AG CC>AT CC>GA CC>GG CC>GT CC>TA CC>TG CC>TT CG>AT CG>GC CG>GT CG>TA CG>TC CG>TT CT>AA CT>AC CT>AG CT>GA CT>GC CT>GG CT>TA CT>TC CT>TG GC>AA GC>AG GC>AT GC>CA GC>CG GC>TA TA>AT TA>CG TA>CT TA>GC TA>GG TA>GT TC>AA TC>AG TC>AT TC>CA TC>CG TC>CT TC>GA TC>GG TC>GT TG>AA TG>AC TG>AT TG>CA TG>CC TG>CT TG>GA TG>GC TG>GT TT>AA TT>AC TT>AG TT>CA TT>CC TT>CG TT>GA TT>GC TT>GG

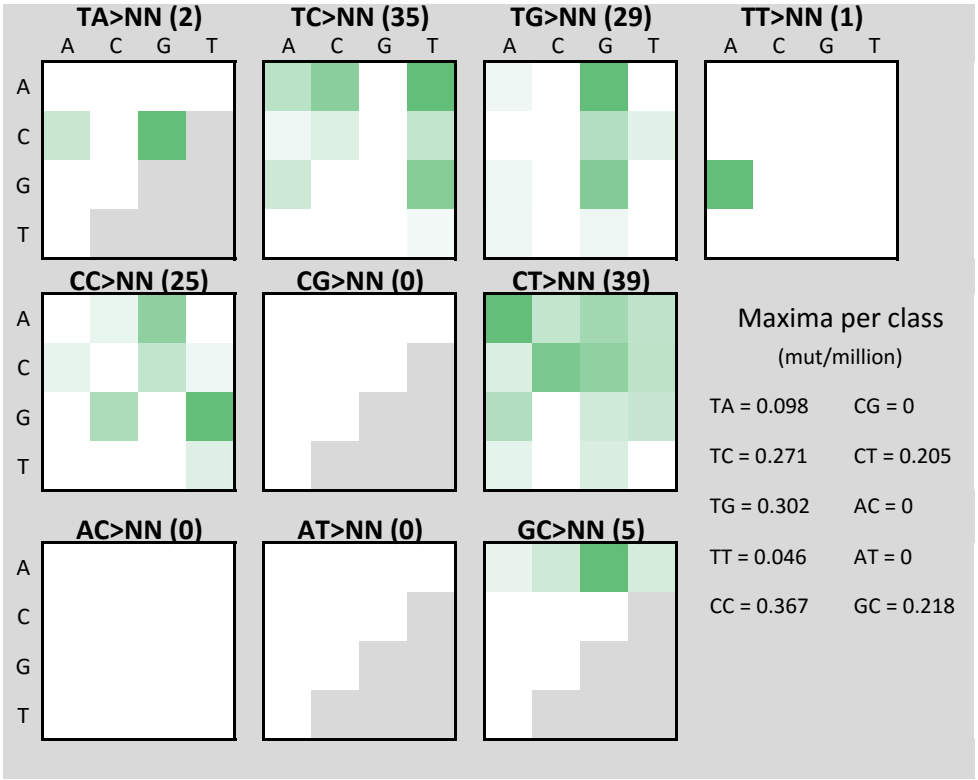
Supplemental Fig. S10: ±1bp sequence context preference of DNSs in the cisplatin treated MCF-10A and HepG2 clones. Sequence context preference is displayed both as raw counts, and normalized for tetranucleotide abundance in the genome. The total number of DNSs per mutation class is shown between brackets above each plot. Color intensity is relative to the number of mutations with that sequence context, normalized for tetranucleotide abundance in the genome. The vertical axis is the preceding base; the horizontal axis is the following base.

MCF-10A_Cis_1

Raw counts (genome)

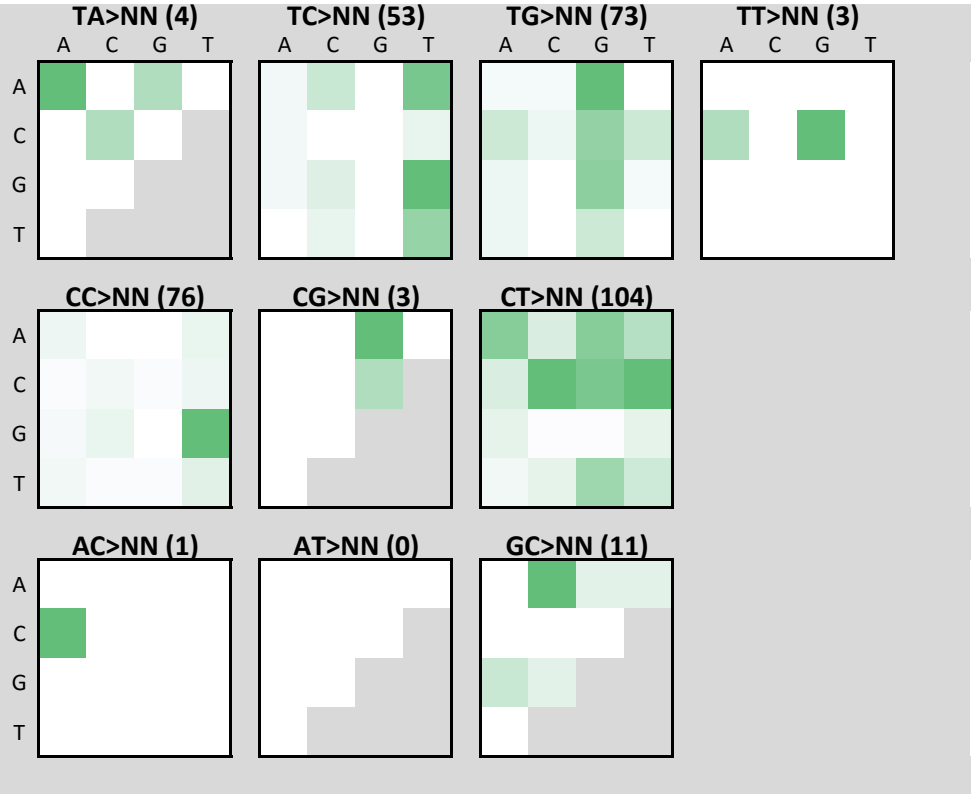


Flat (as if genome opportunity was equal for all tetranucleotides)

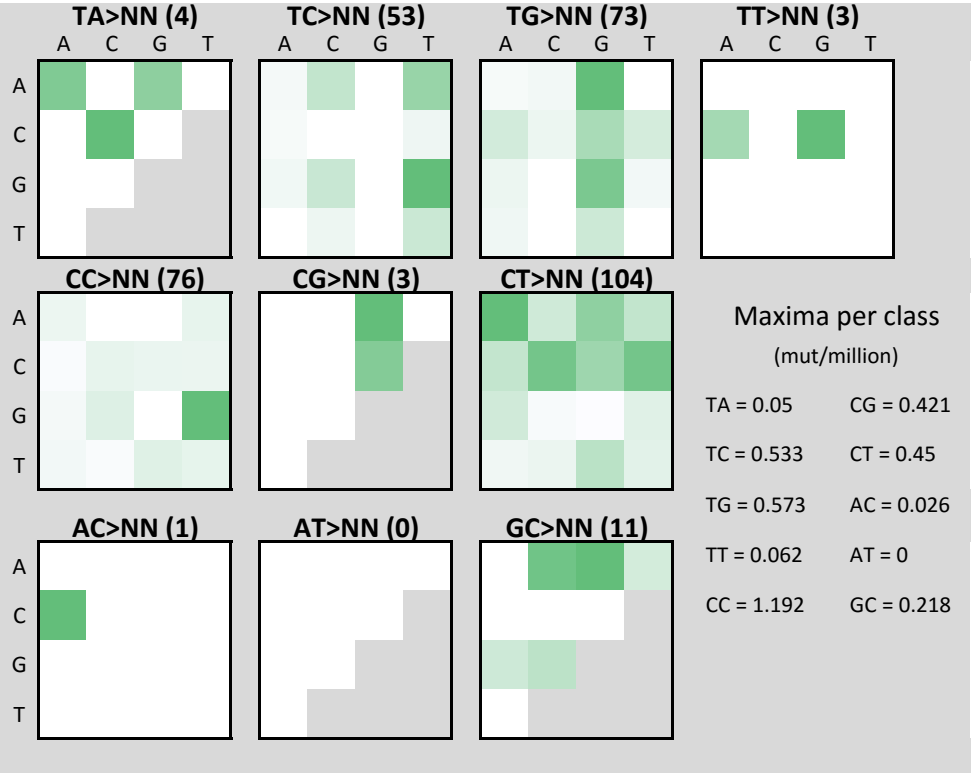


MCF-10A_Cis_2

Raw counts (genome)

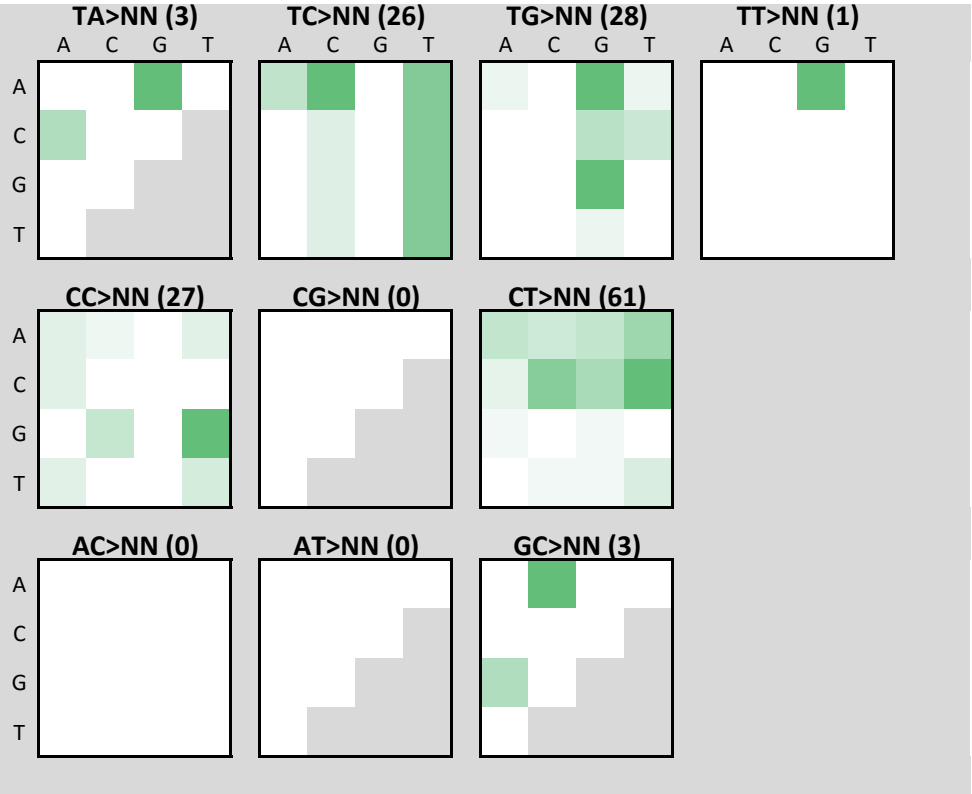


Flat (as if genome opportunity was equal for all tetranucleotides)

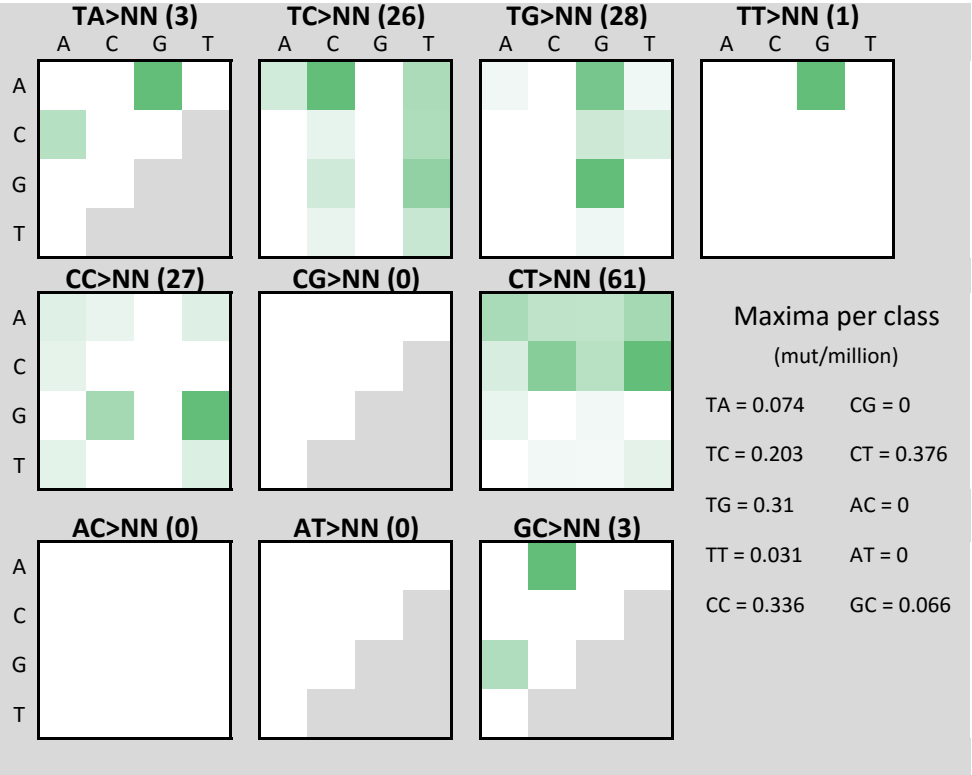


MCF-10A_Cis_3

Raw counts (genome)

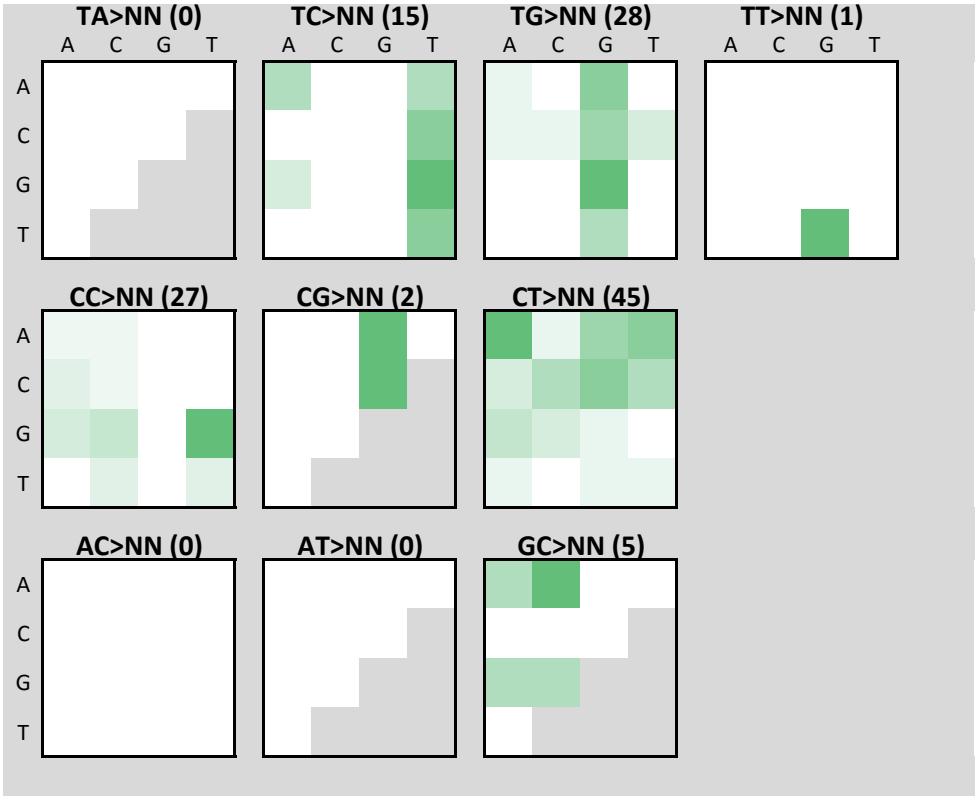


Flat (as if genome opportunity was equal for all tetranucleotides)

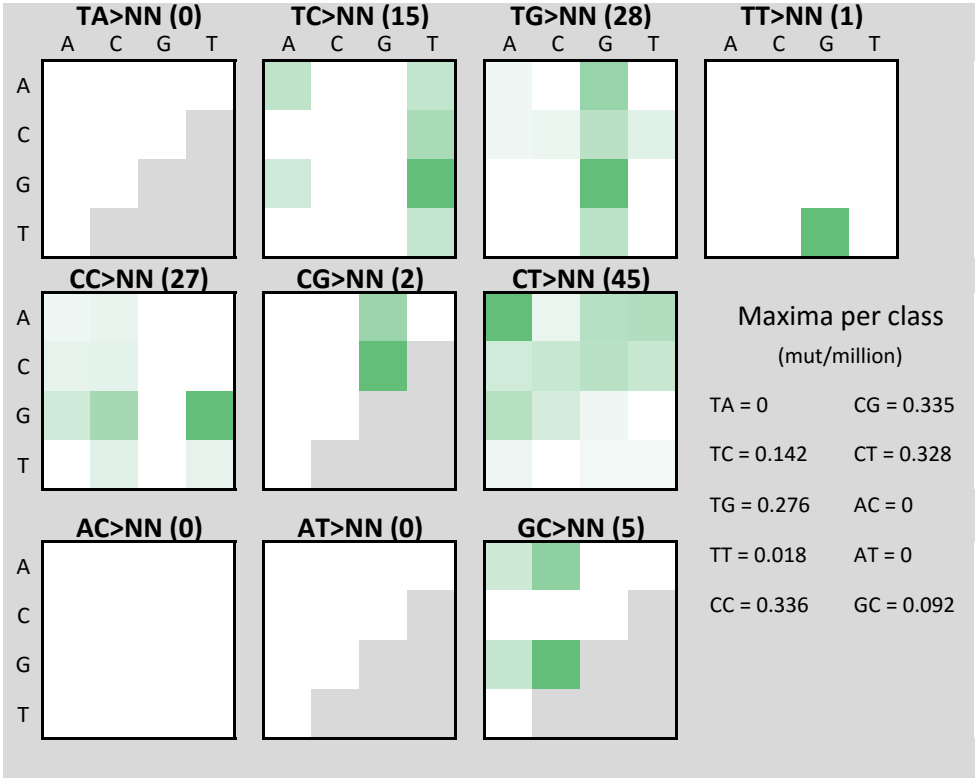


MCF-10A_Cis_4

Raw counts (genome)

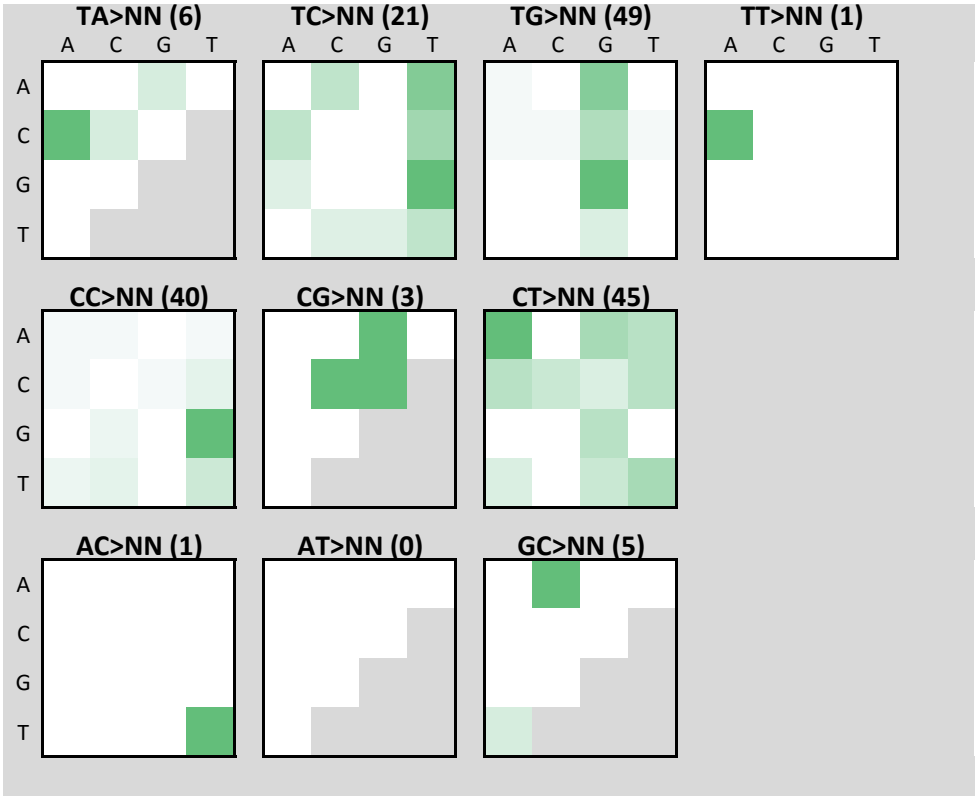


Flat (as if genome opportunity was equal for all tetranucleotides)

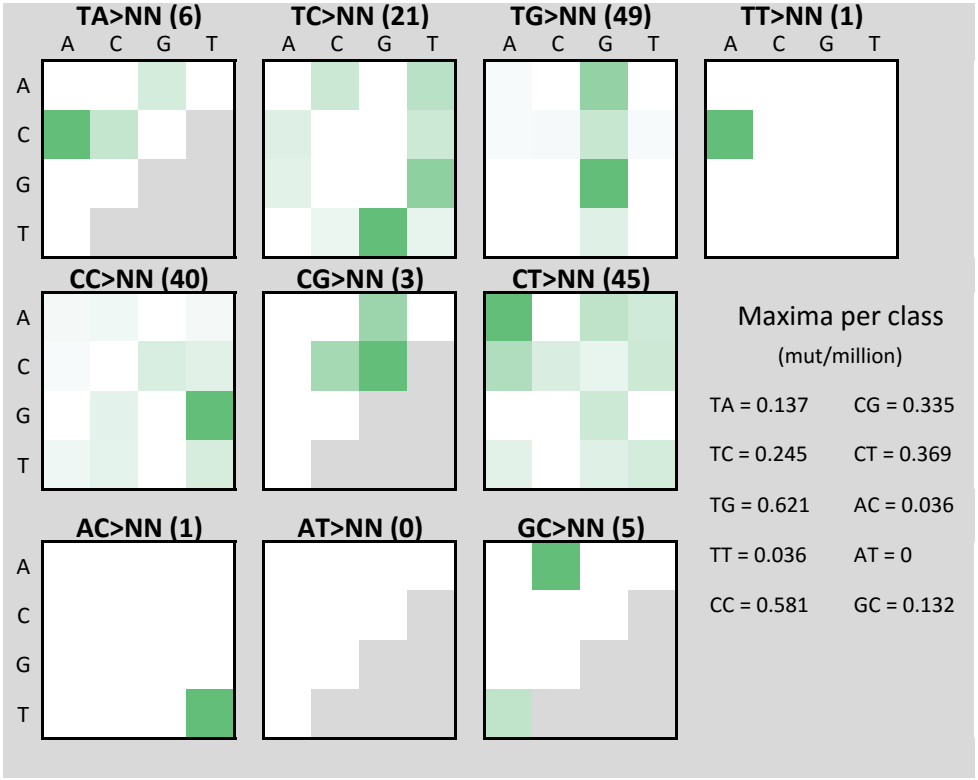


MCF-10A_Cis_5

Raw counts (genome)

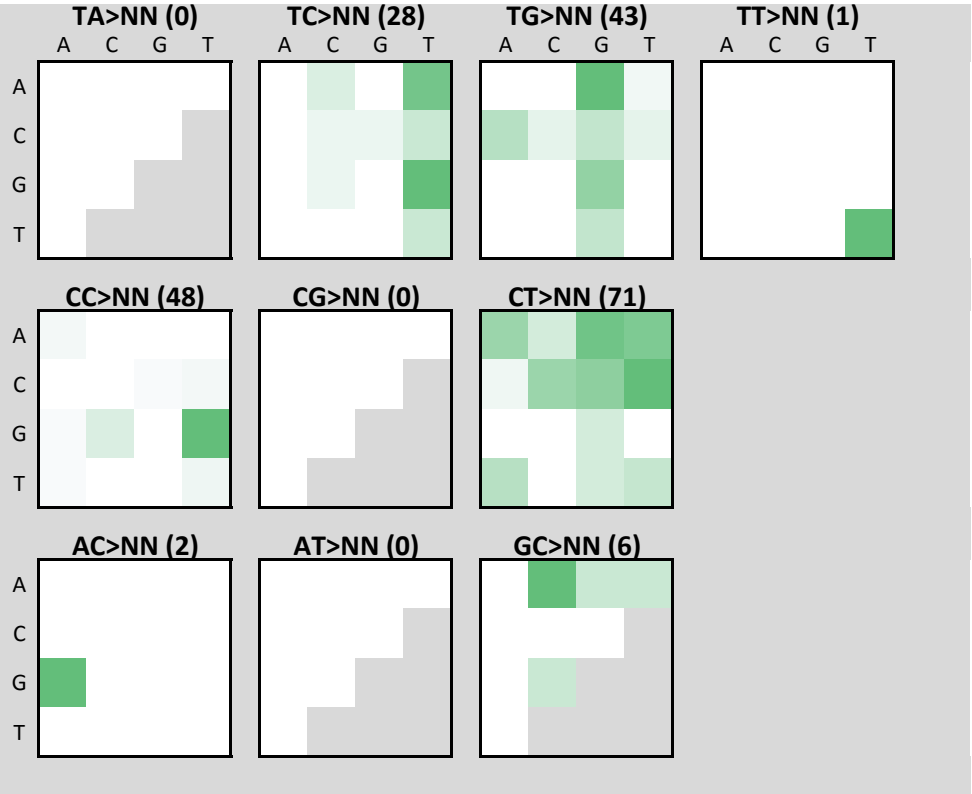


Flat (as if genome opportunity was equal for all tetranucleotides)

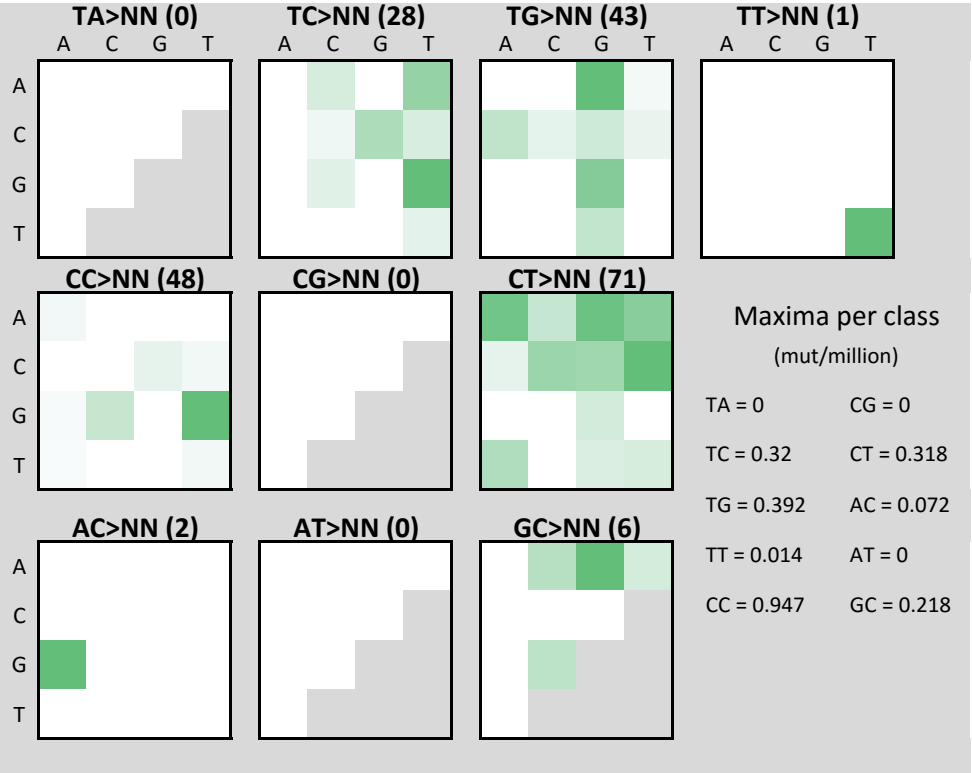


MCF-10A_Cis_6

Raw counts (genome)

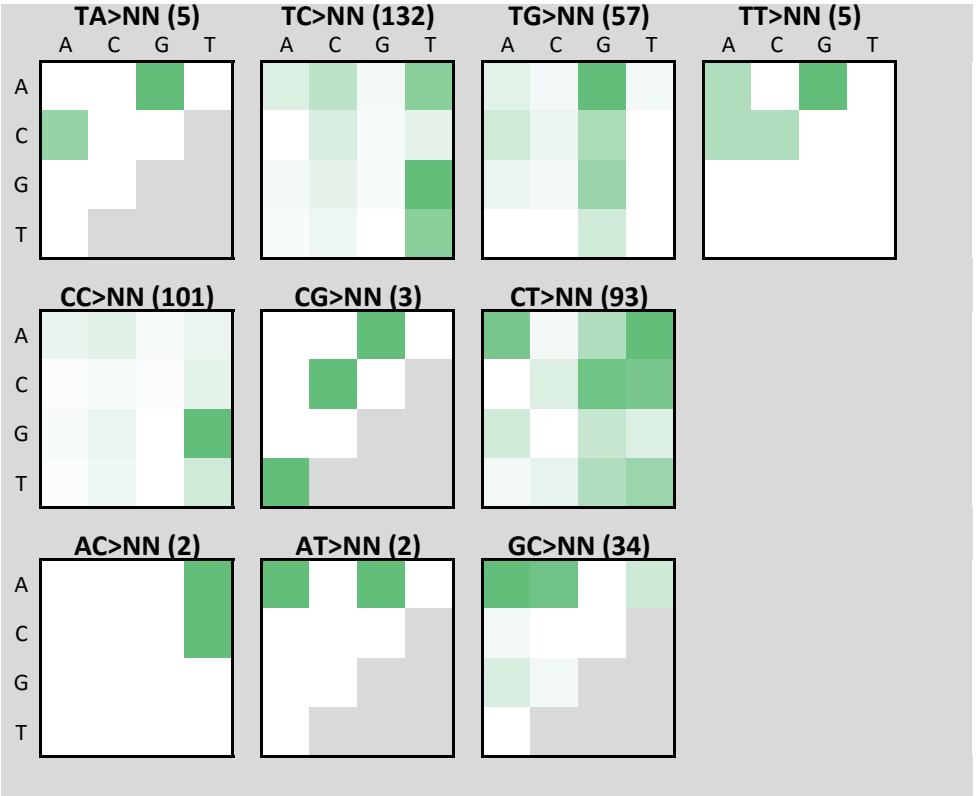


Flat (as if genome opportunity was equal for all tetranucleotides)

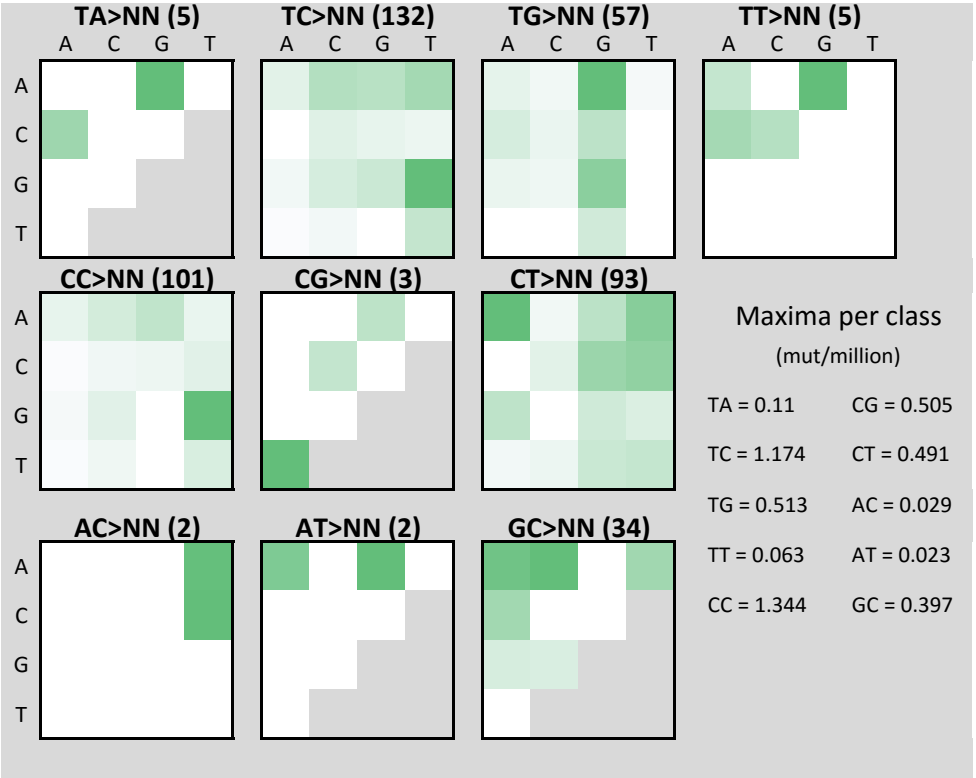


HepG2_Cis_1

Raw counts (genome)

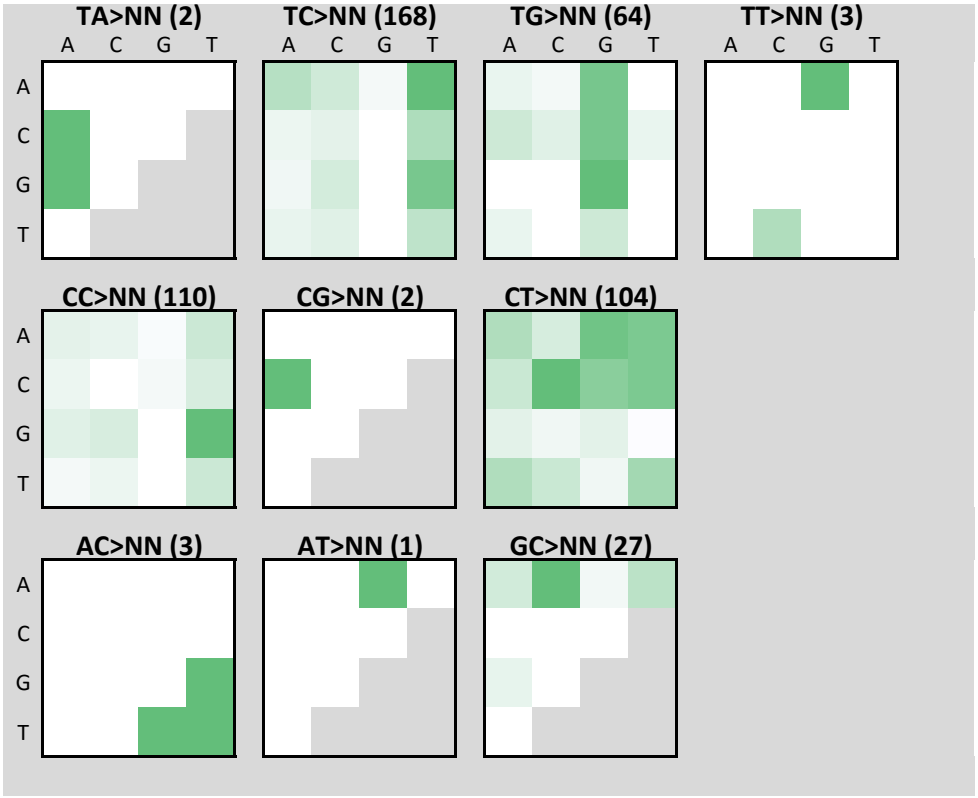


Flat (as if genome opportunity was equal for all tetranucleotides)

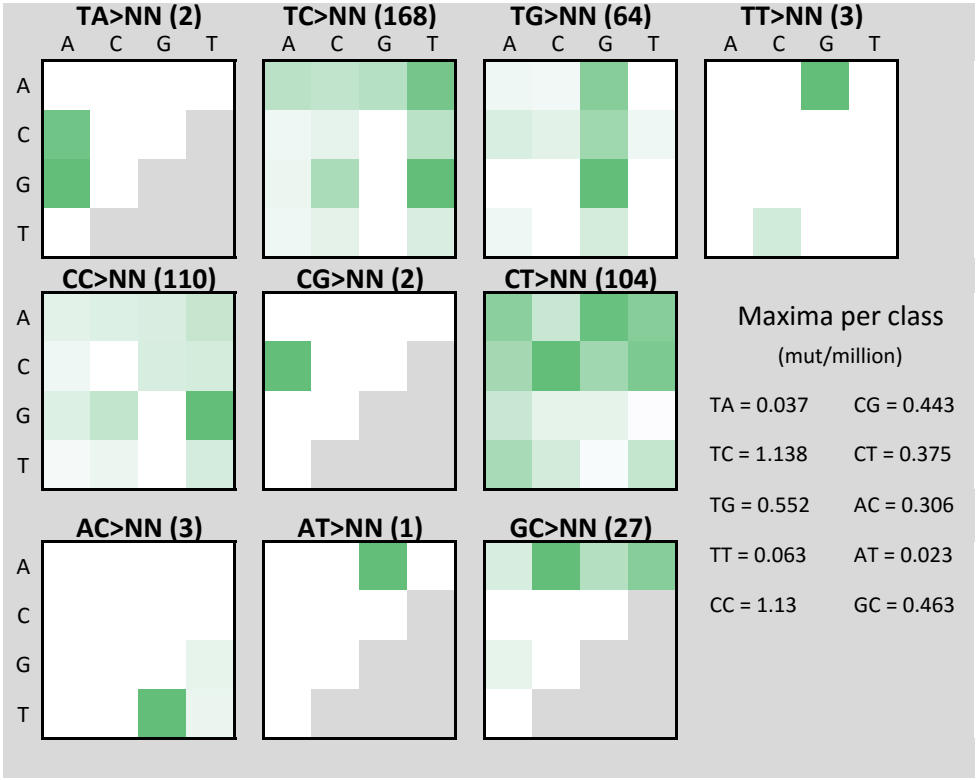


HepG2_Cis_2

Raw counts (genome)

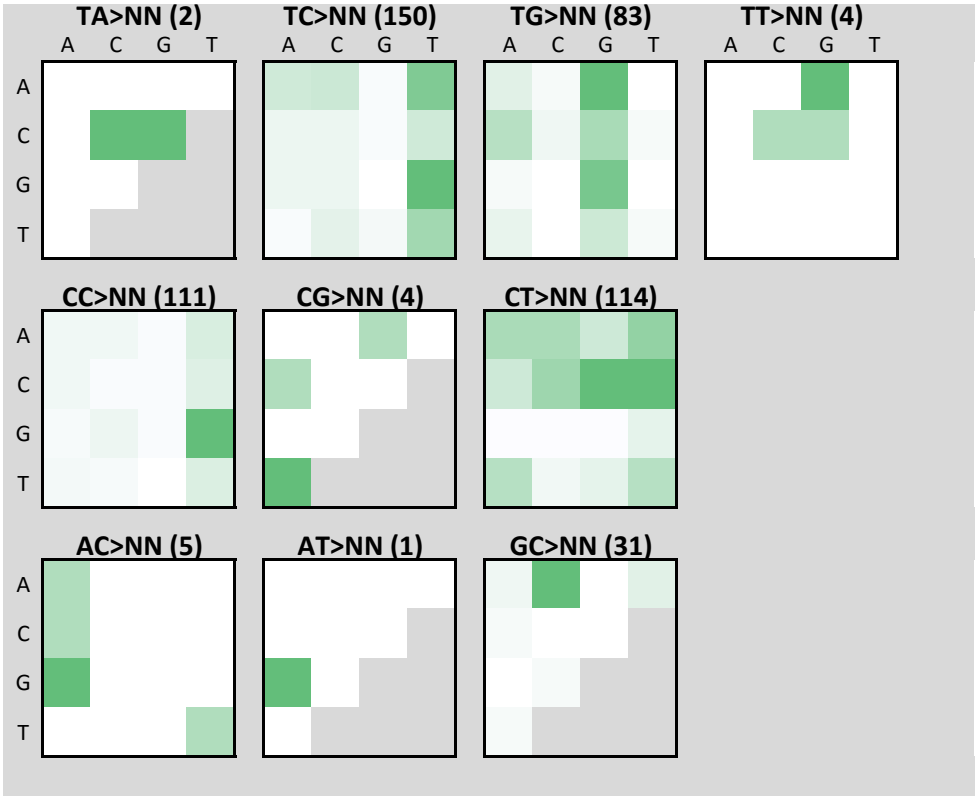


Flat (as if genome opportunity was equal for all tetranucleotides)

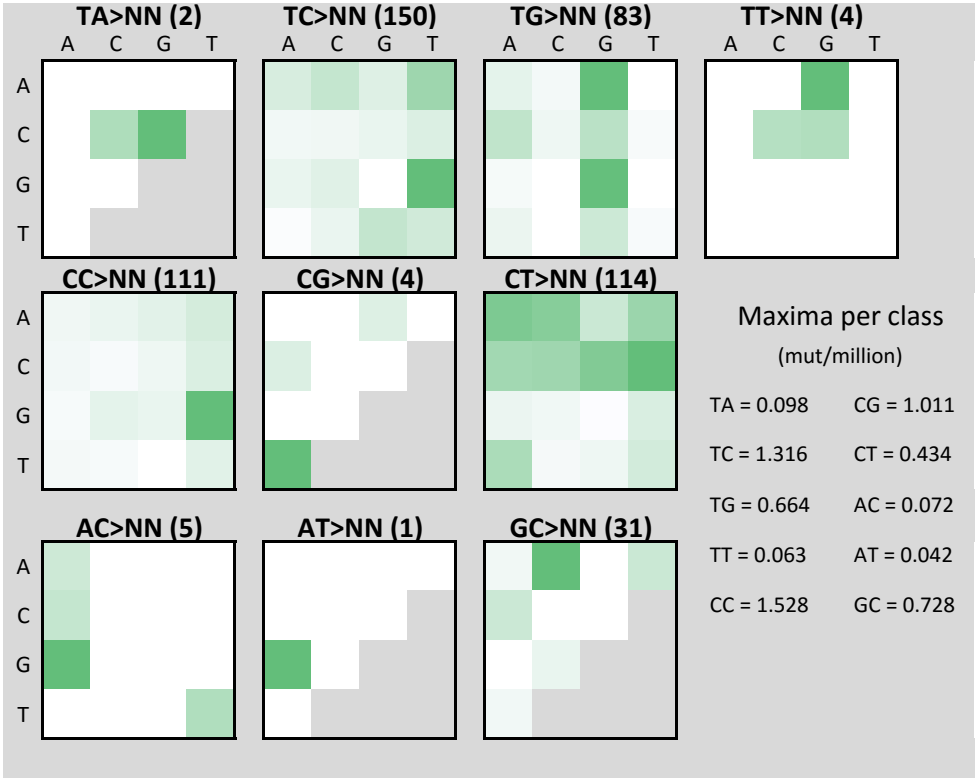


HepG2_Cis_3

Raw counts (genome)

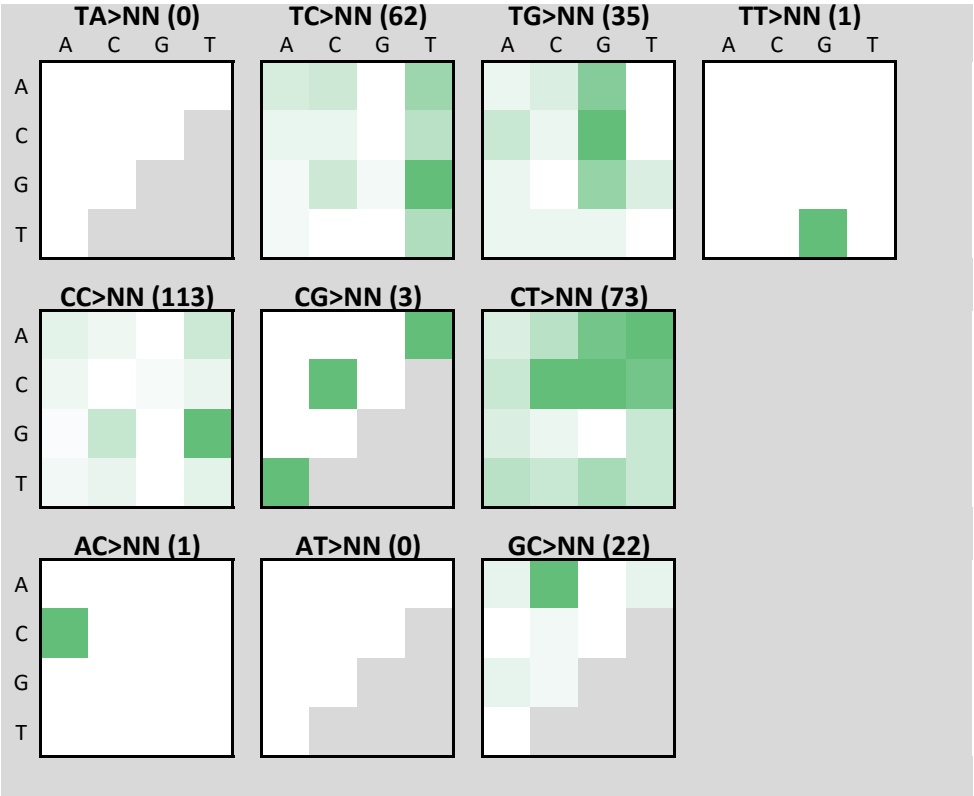


Flat (as if genome opportunity was equal for all tetranucleotides)

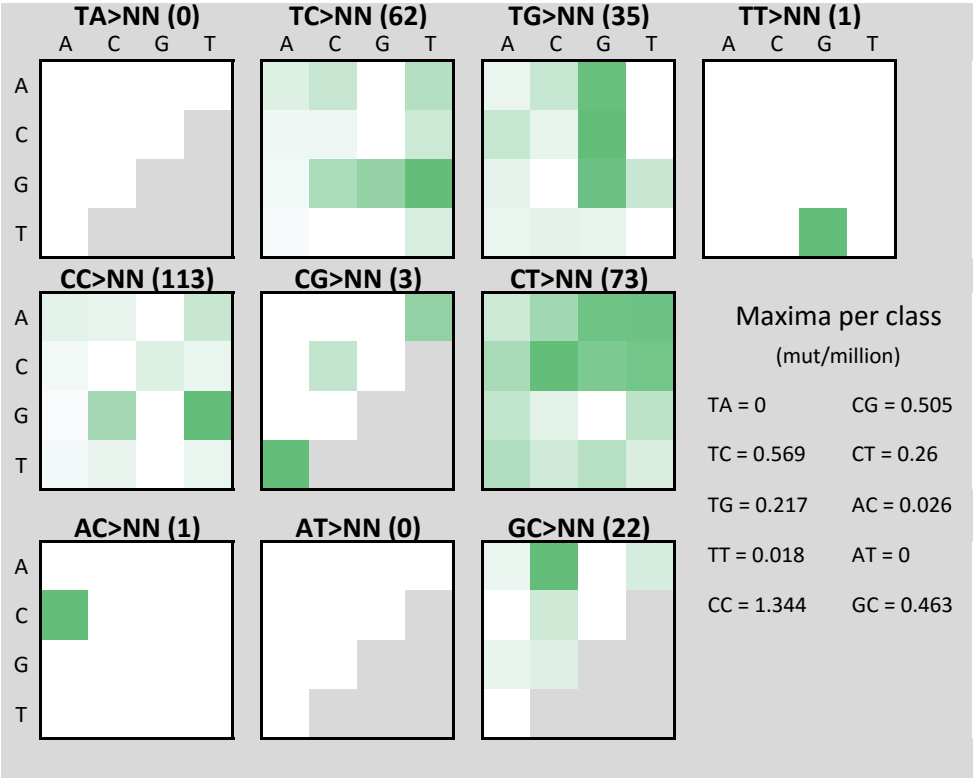


HepG2_Cis_4

Raw counts (genome)

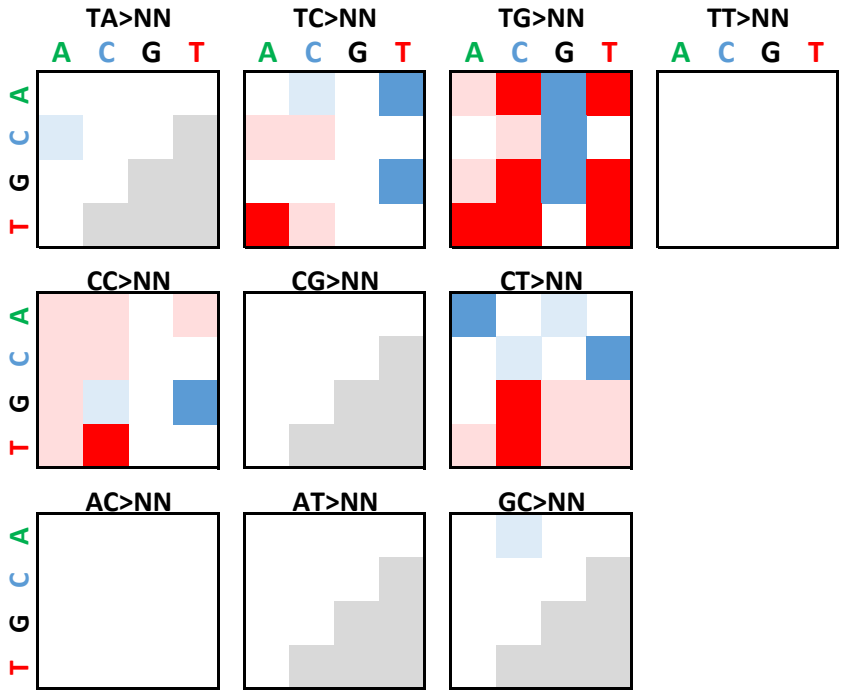


Flat (as if genome opportunity was equal for all tetranucleotides)

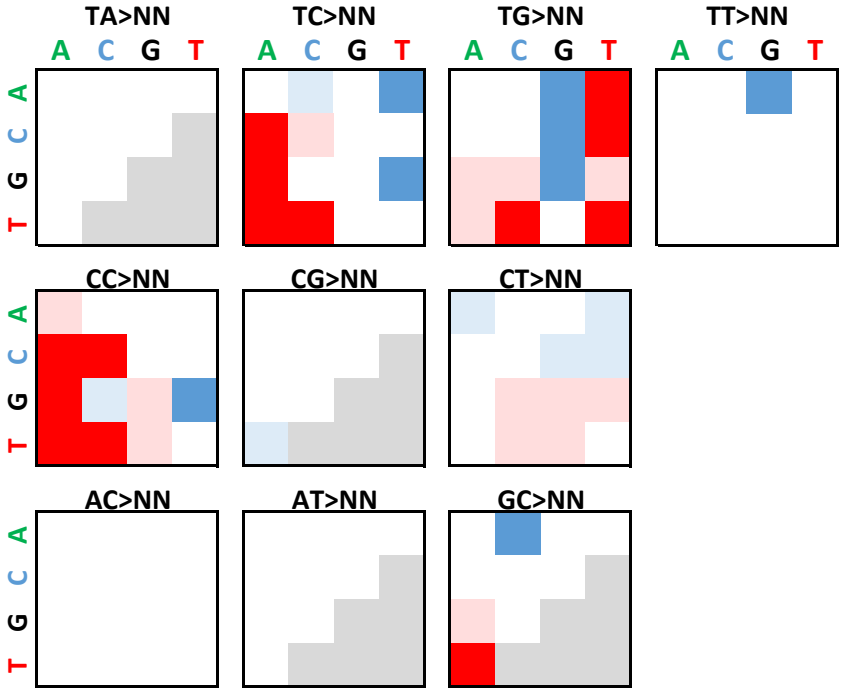


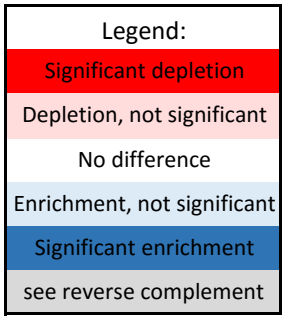
Supplemental Fig. S11: Graphical display of statistically significant enrichments and depletions of DNSs in tetranucleotide contexts for all MCF-10A (A) and HepG2 (B) clones combined. Binomial tests were performed as described in Materials and Methods. Results were summarized as either enriched for mutations (blue) or depleted for mutations (red) or no difference from expected (white). Dark green and red indicate sequence contexts that were significant after Bonferroni multiple testing correction (i.e. $p < 0.05/136$). Light blue and pink indicate sequence contexts that were not significant after correction for multiple testing (i.e. $0.05 > p > 0.05/136$).

A: MCF-10A



B: HepG2

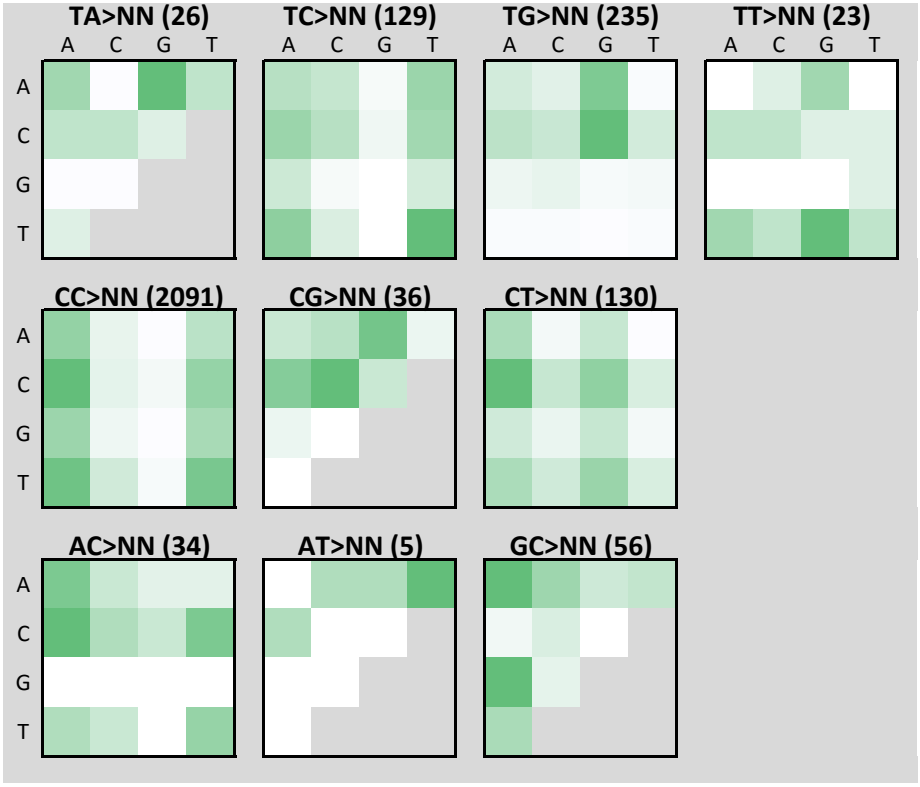




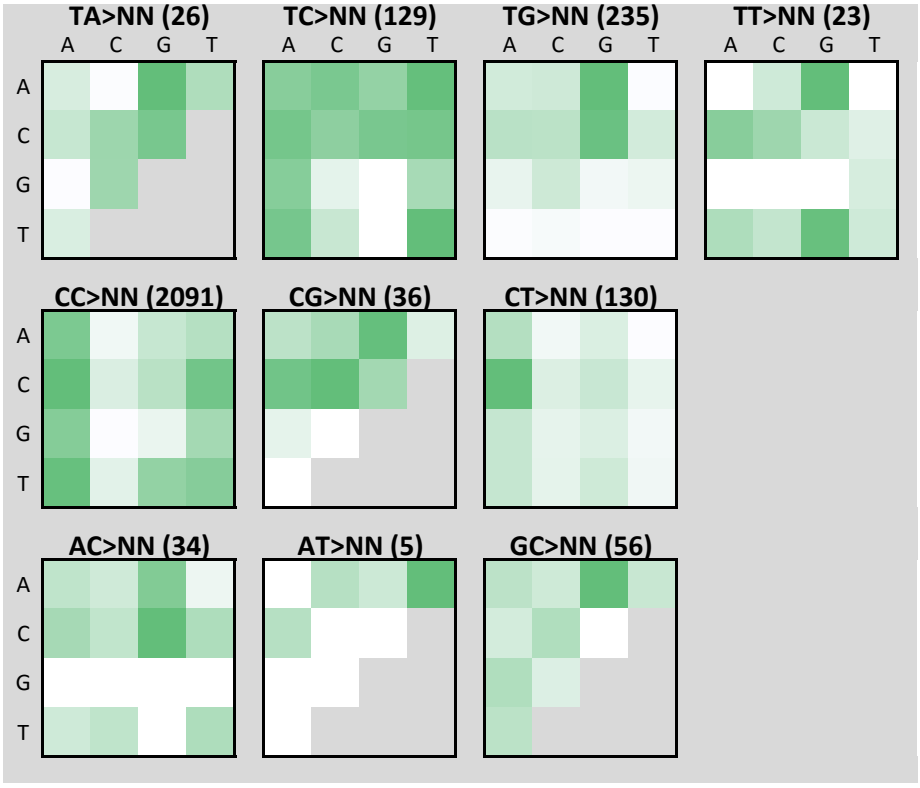
Supplemental Fig. S12: Example plots of DNS sequence context preference in 3 lung adenocarcinomas and 3 melanomas. The lung adenocarcinomas did not display very strong preference sequence context of DNSs. Conversely, in the melanomas most dinucleotides showed sequence context preference. For example, 52.8% of all CC mutations occurred TCCN context. Similarly, TT mutations preferentially occur in NTTA context (52.1%), and CT mutations showed extremely strong preference for a T either preceding or following the mutated dinucleotide (79.5%). The strongest tetranucleotide context preference was observed for TA mutations, which prefer ATAA context (46.9%), and CG mainly occur in TCGA context (50.4%).
Figure 10: Dinucleotide substitution sequence context for lung adenocarcinomas and melanomas

LUAD-E00934

Raw counts (genome)

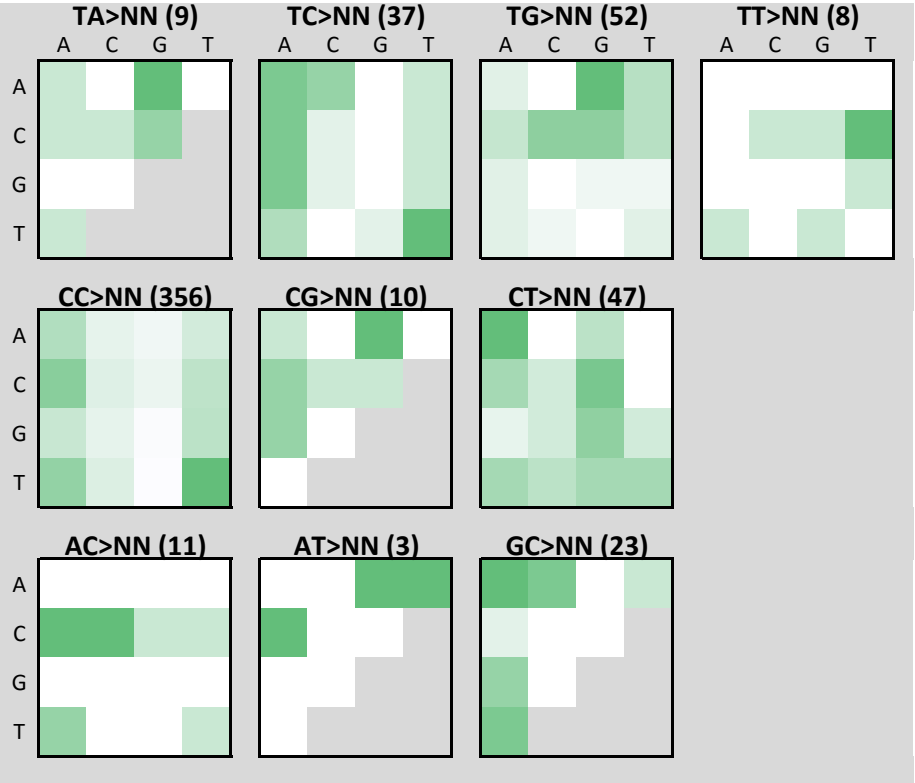


Flat (per million tetranucleotides)

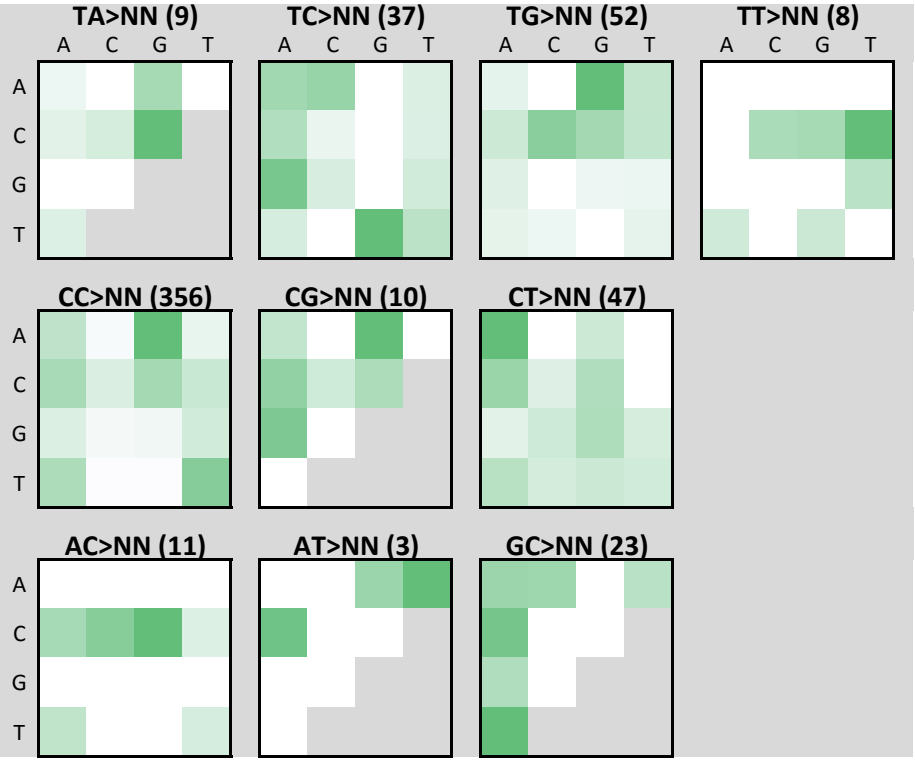


LUAD-E01317

Raw counts (genome)

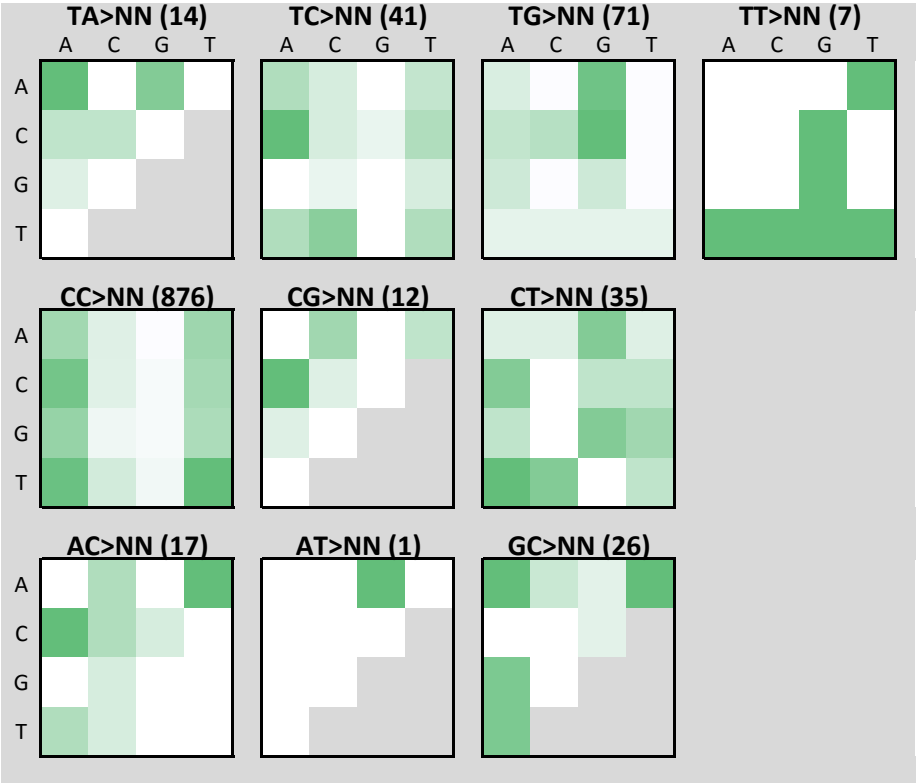


Flat (per million tetranucleotides)

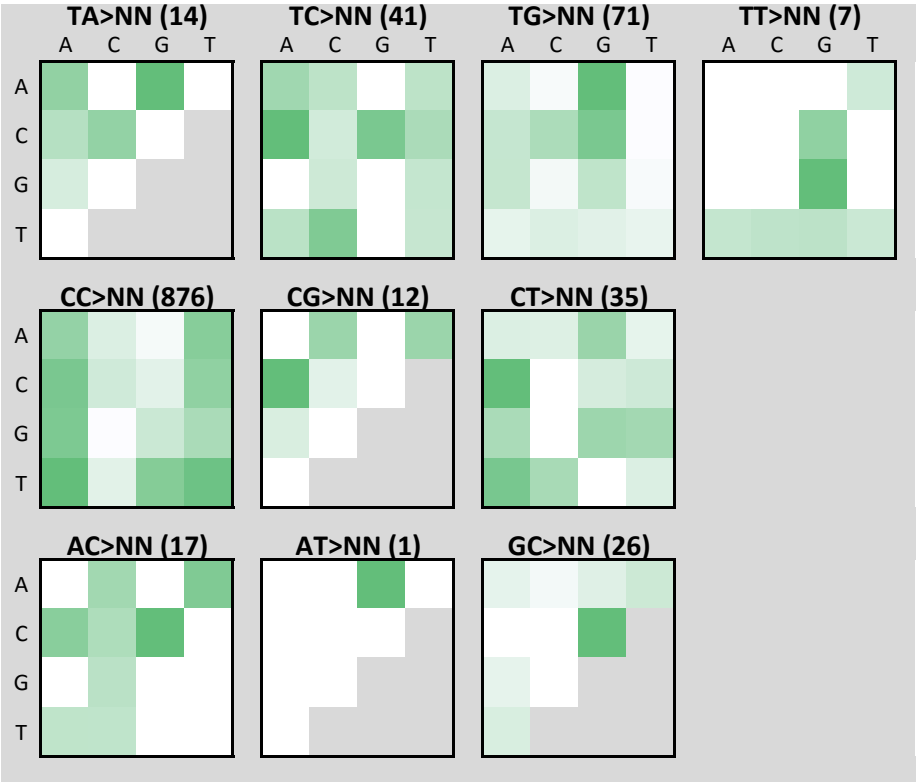


LUAD-S01405

Raw counts (genome)

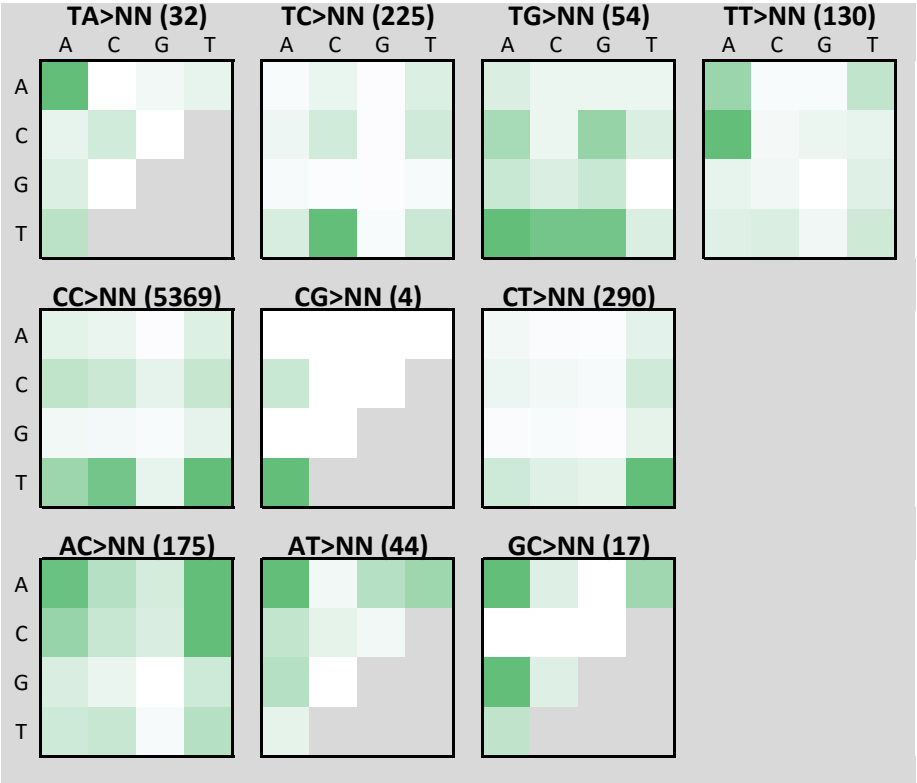


Flat (per million tetranucleotides)

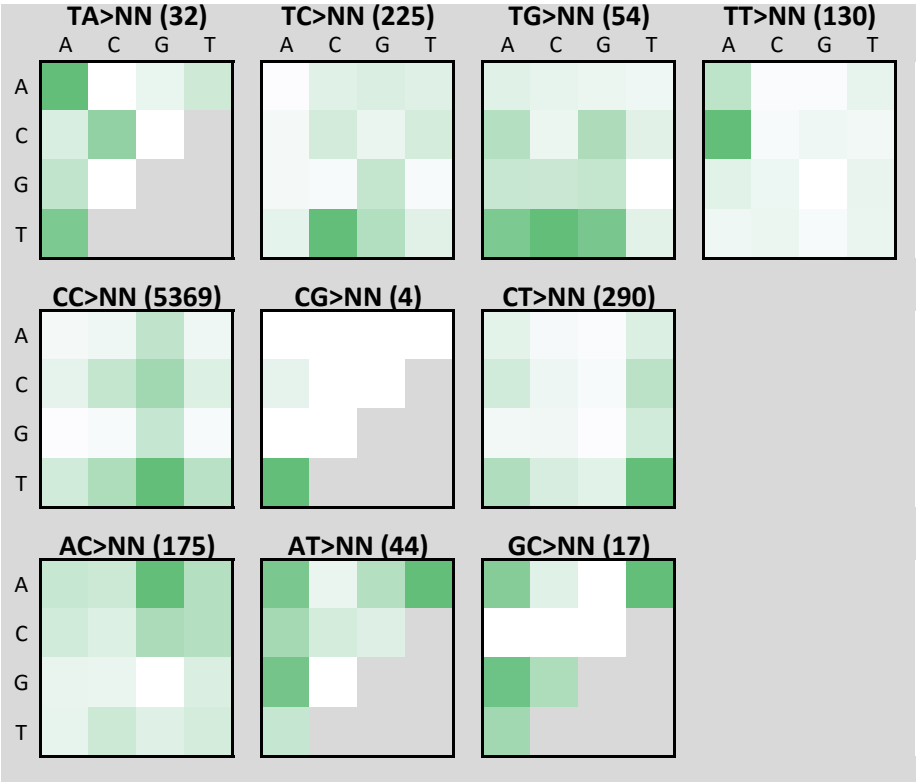


MELA_0165

Raw counts (genome)

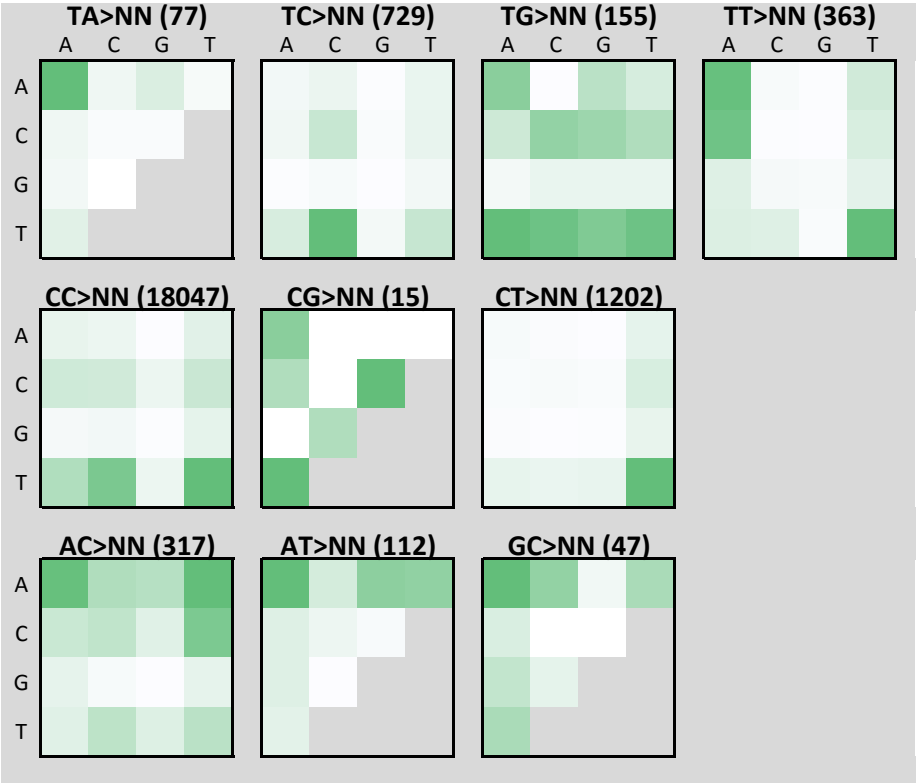


Flat (per million tetranucleotides)



MELA_0247

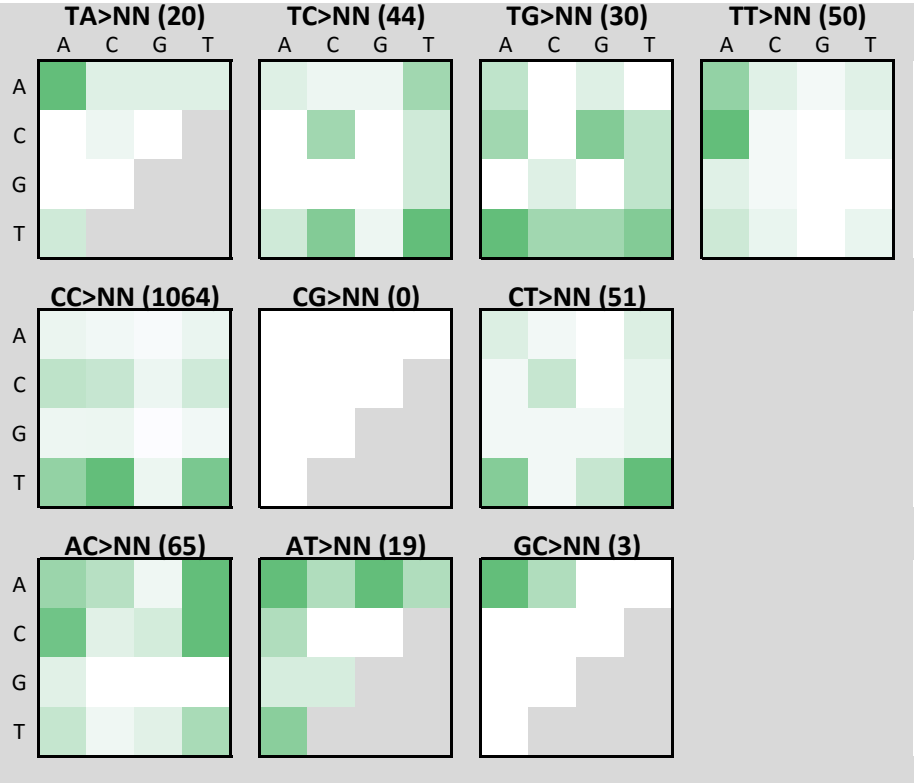
Raw counts (genome)



Flat (per million tetranucleotides)

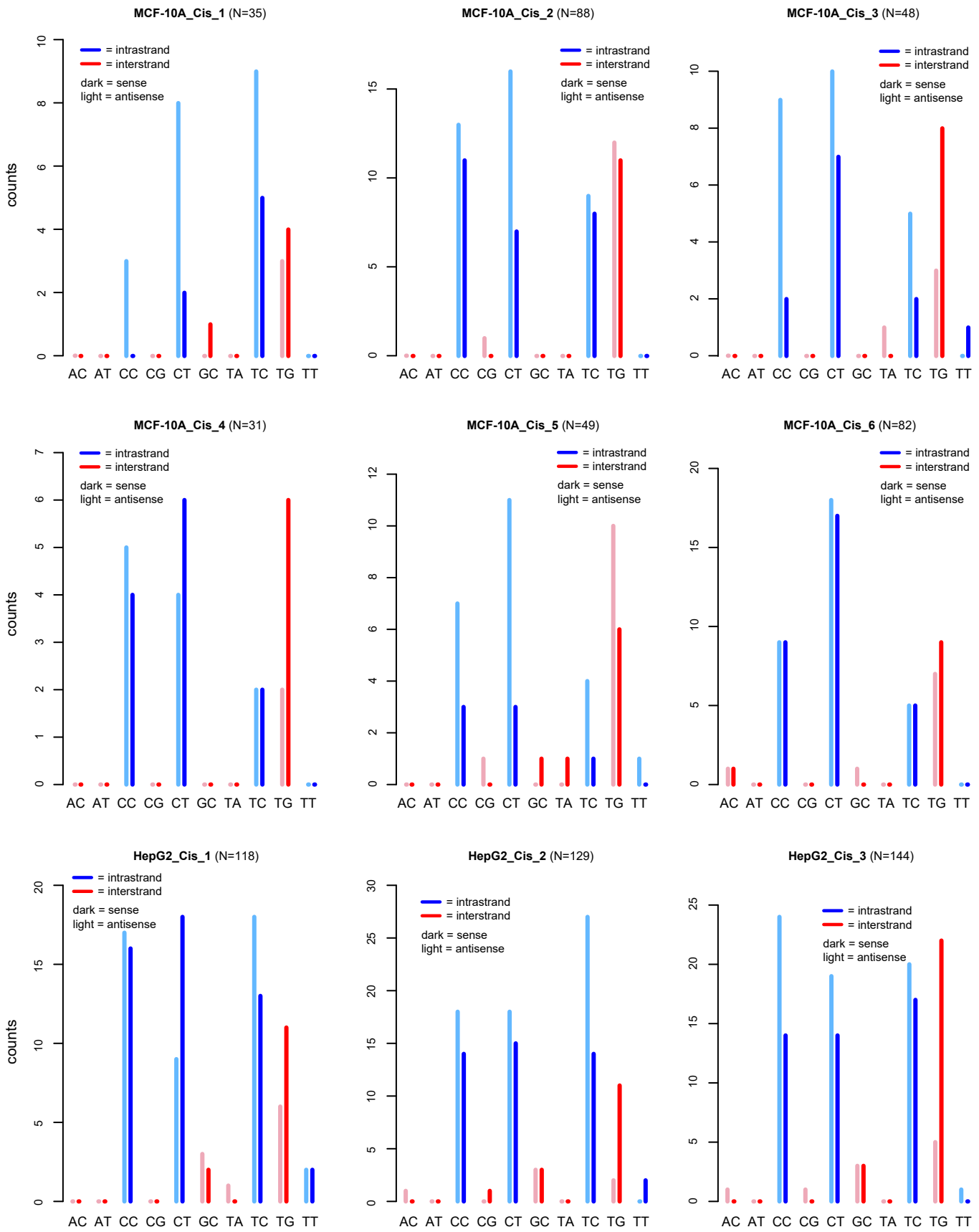
MELA_0255

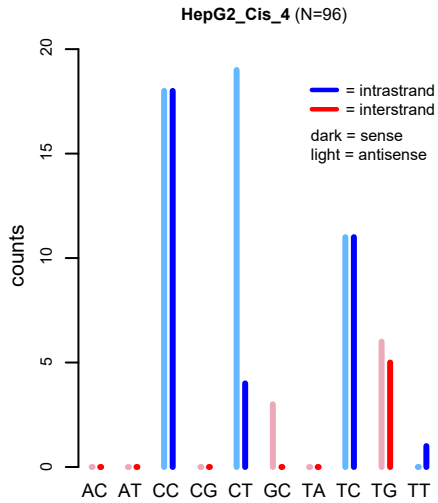
Raw counts (genome)



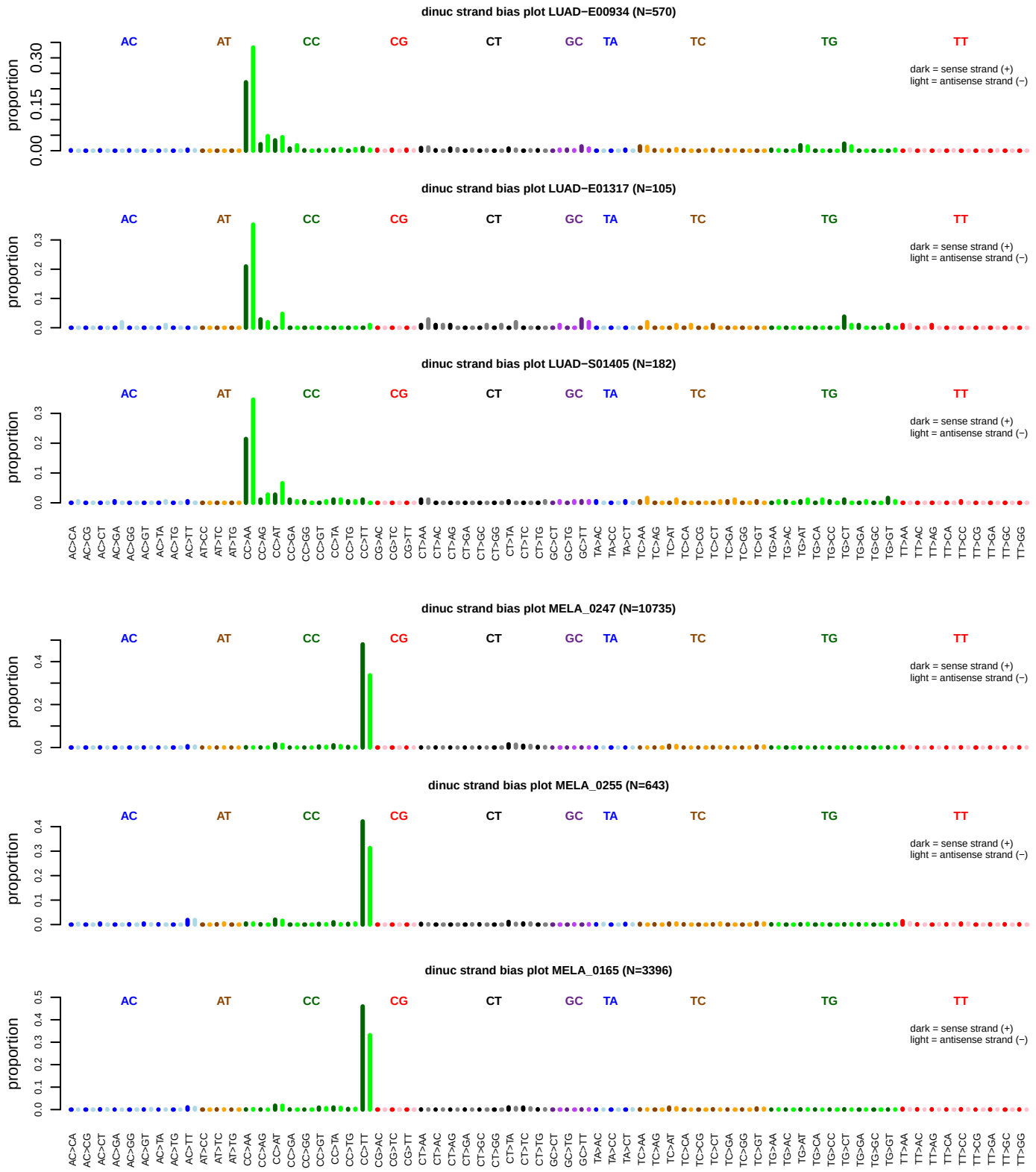
Flat (per million tetranucleotides)

Supplemental Fig. S13: Transcription strand bias of DNSs in cisplatin treated MCF-10A and HepG2 clones. The total number of dinucleotides eligible for transcription strand bias analysis is displayed in parentheses.

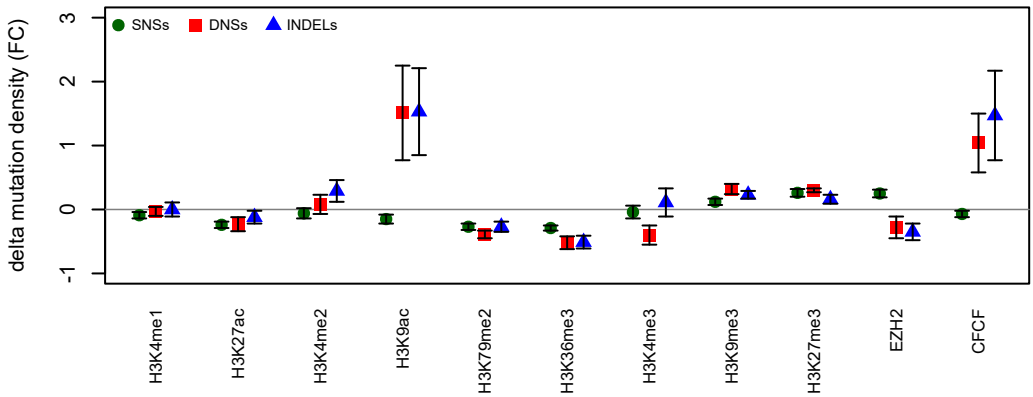



HepG2_Cis_4 (N=96)
counts
0
5
10
15
20
= intrastrand
= interstrand
dark = sense
light = antisense
AC
AT
CC
CG
CT
GC
TA
TC
TG
TT

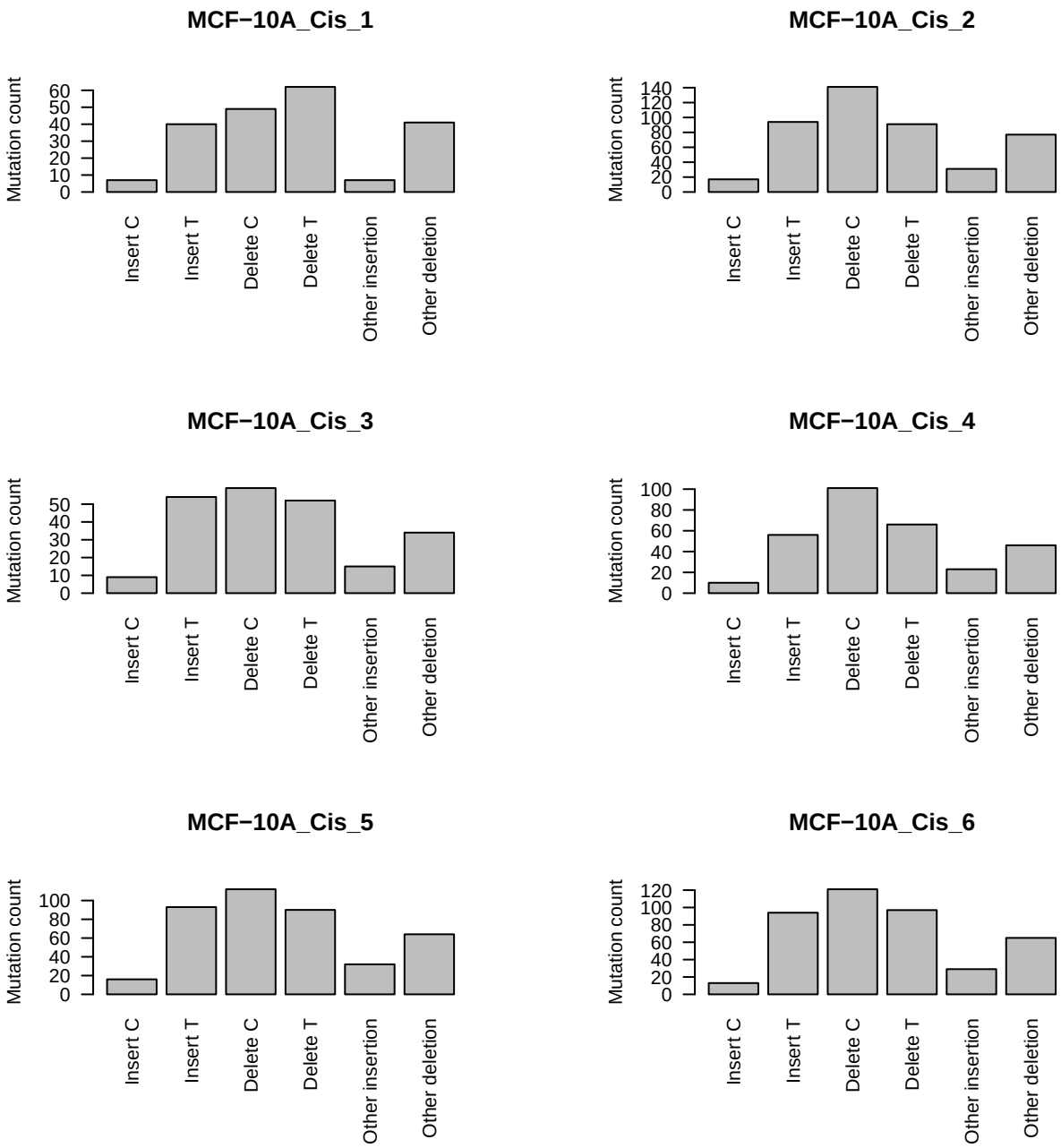
Supplemental Fig. S14: Transcription strand bias of DNSs in 3 lung adenocarcinomas and 3 melanomas. Contrary to the cisplatin DNSs, both the smoking and UV associated DNSs displayed transcription strand bias. The UV associated dinucleotides showed a decrease of CC>TT mutations on the transcribed strand, as CC crosslinks induced by UV are repaired by TC-NER. Similarly, smoking associated dinucleotides showed transcription strand bias with a decrease of CC>AA mutations on the untranscribed strand. This fits the prior knowledge that smoking causes GG intrastrand crosslinks, which are repaired by TC-NER if they are located on the transcribed strand. As we display the DNSs as CC>AA, the strand bias is reversed.



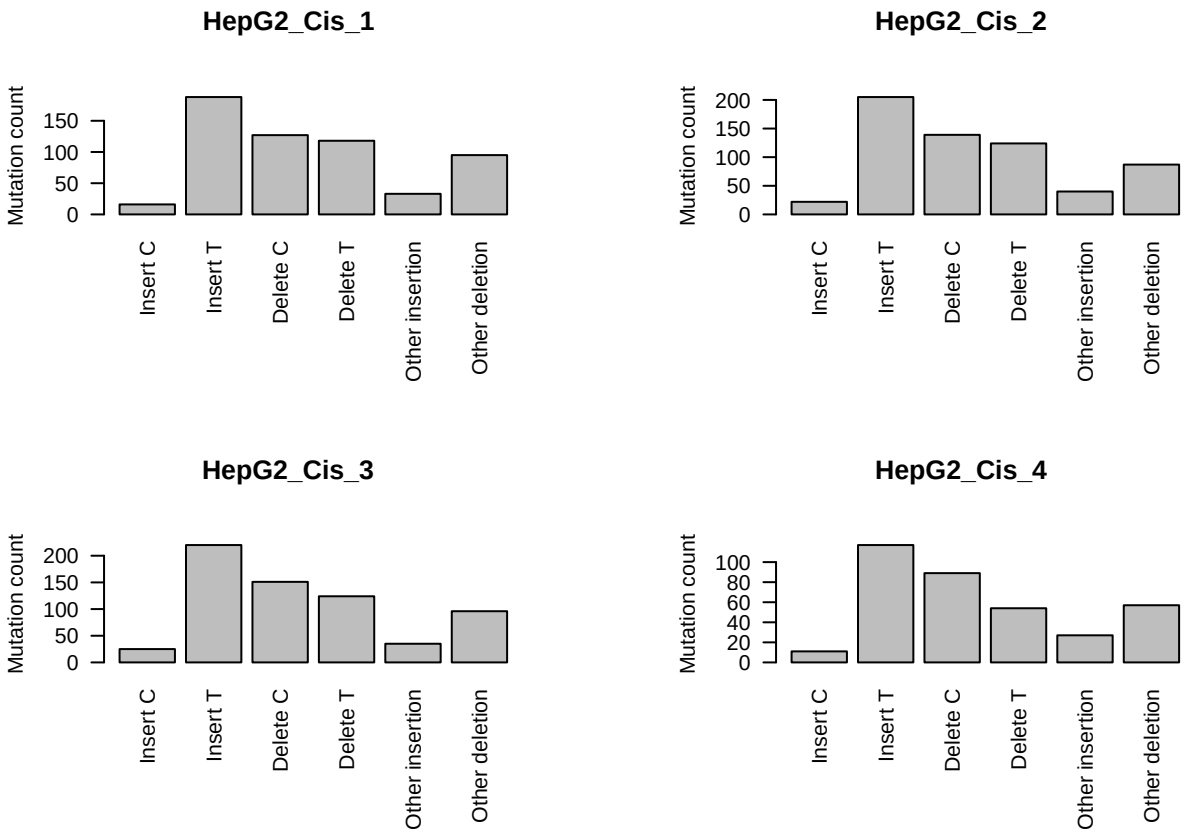
Supplemental Fig. S15: Mutation density in regions with histone modifications, relative to the mutation density of each respective sample (Supplementary Table 1). Plotted is the average for all 10 clones ± SEM.



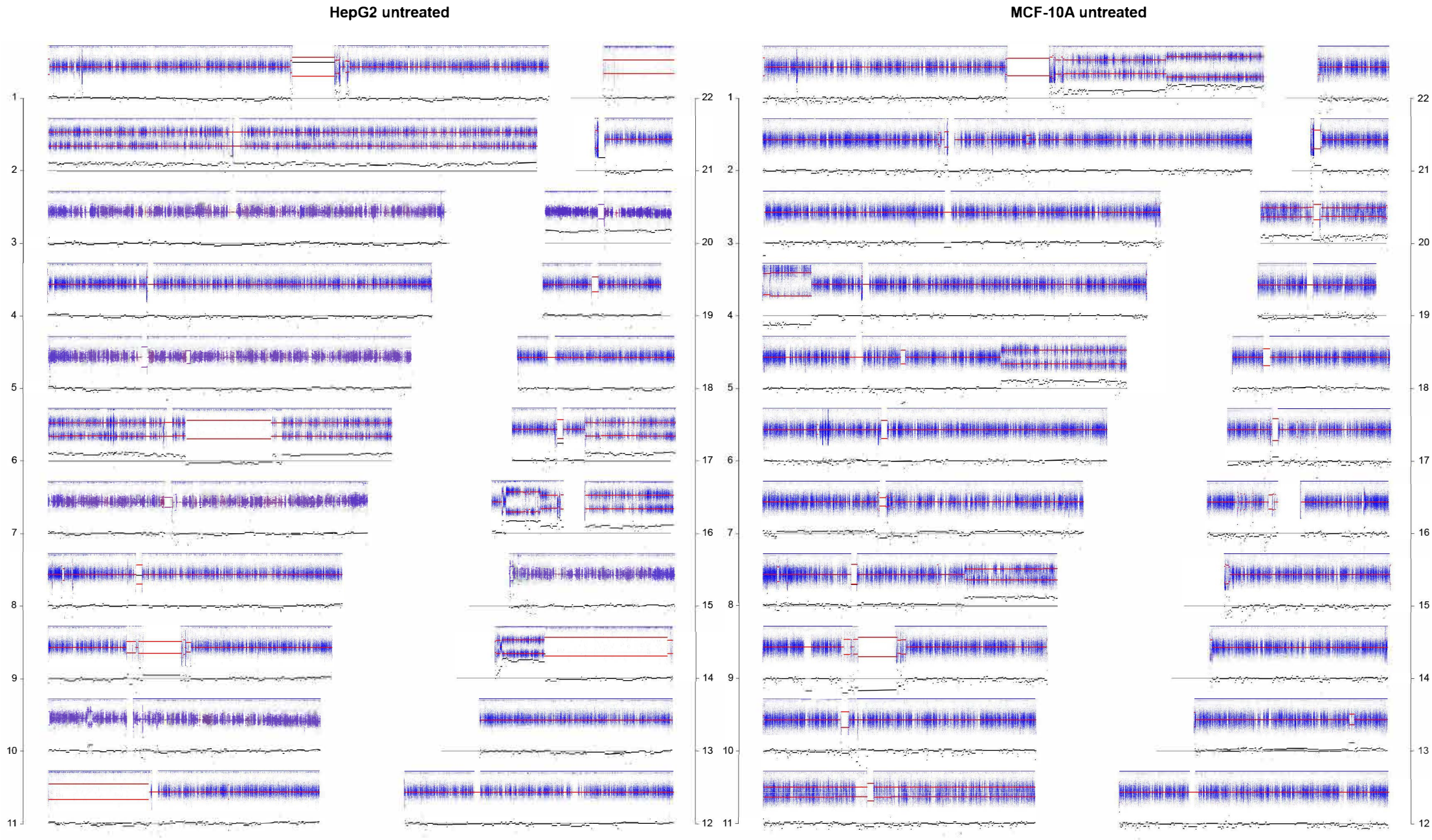
Supplemental Fig. S16: Indel mutation patterns.

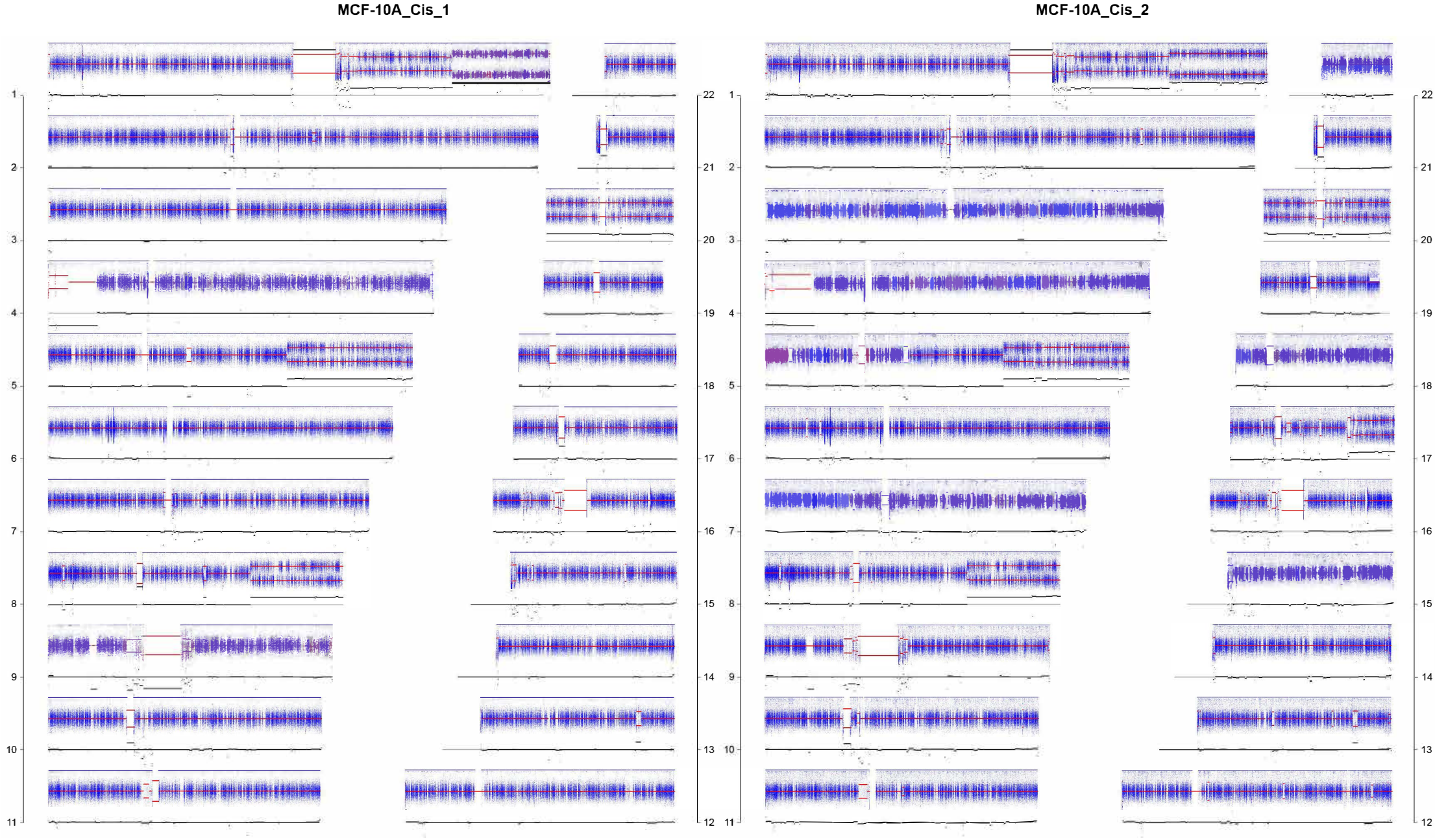


Supplemental Fig. S16: Indel mutation patterns, continued.

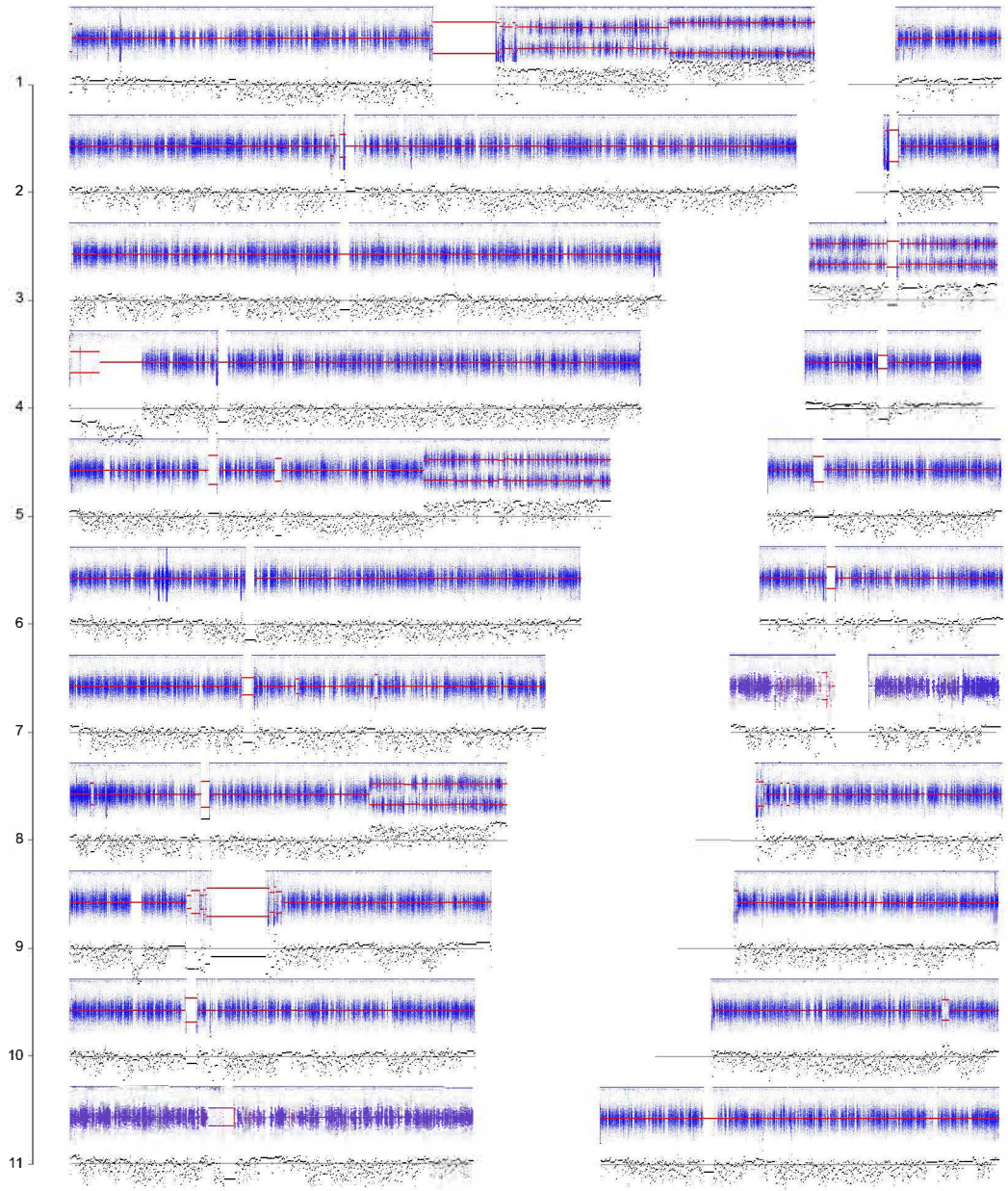


Supplemental Fig. S17: Copy number analysis of cisplatin treated cell line clones. For each cisplatin treated cell line clone and the untreated cell lines, we plotted the copy number (black) and beta-allele frequency (individual datapoints in blue and segmented in red).

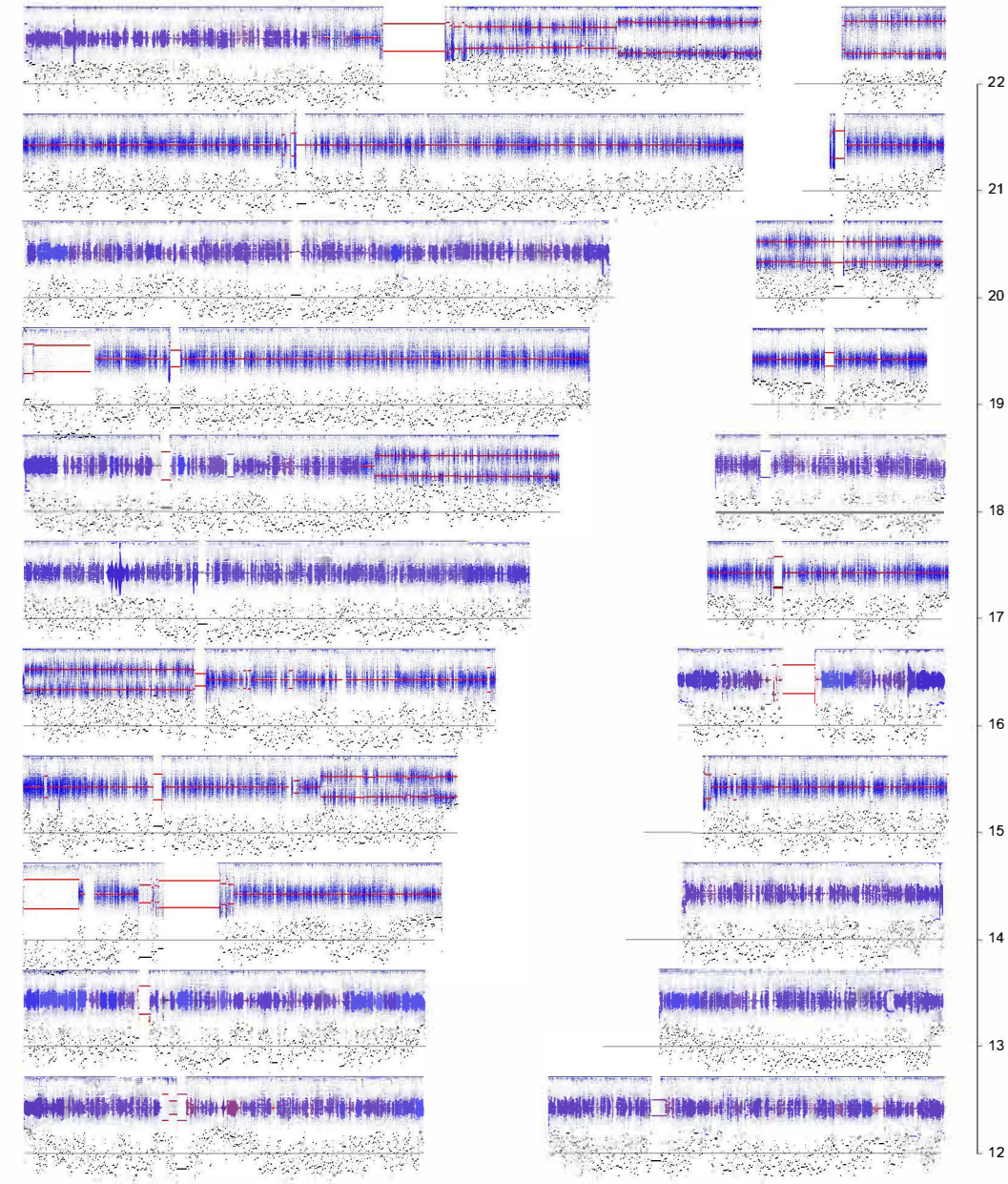



MCF-10A_Cis_1
MCF-10A_Cis_2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

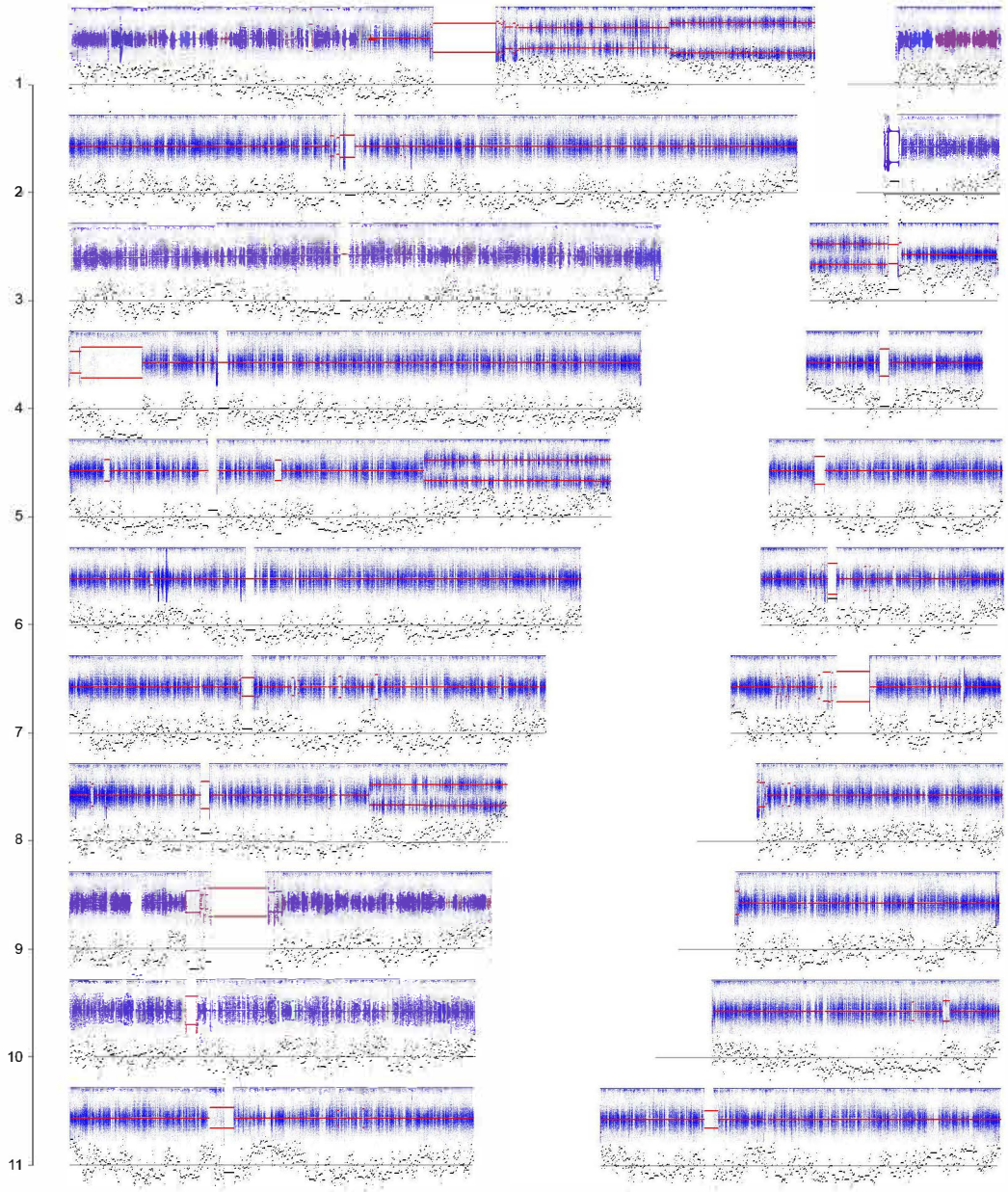
MCF-10A_Cis_3



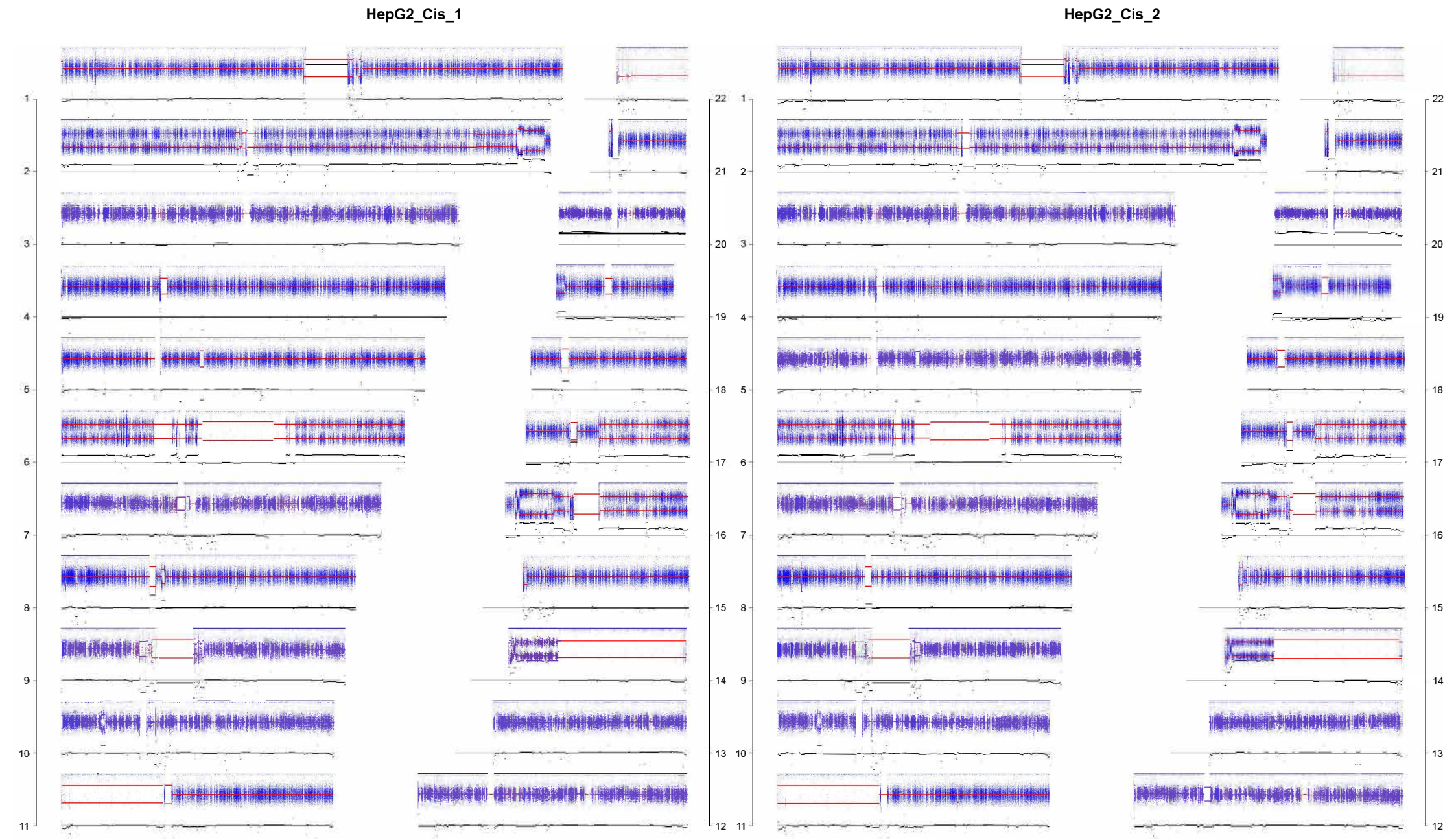
MCF-10A_Cis_4



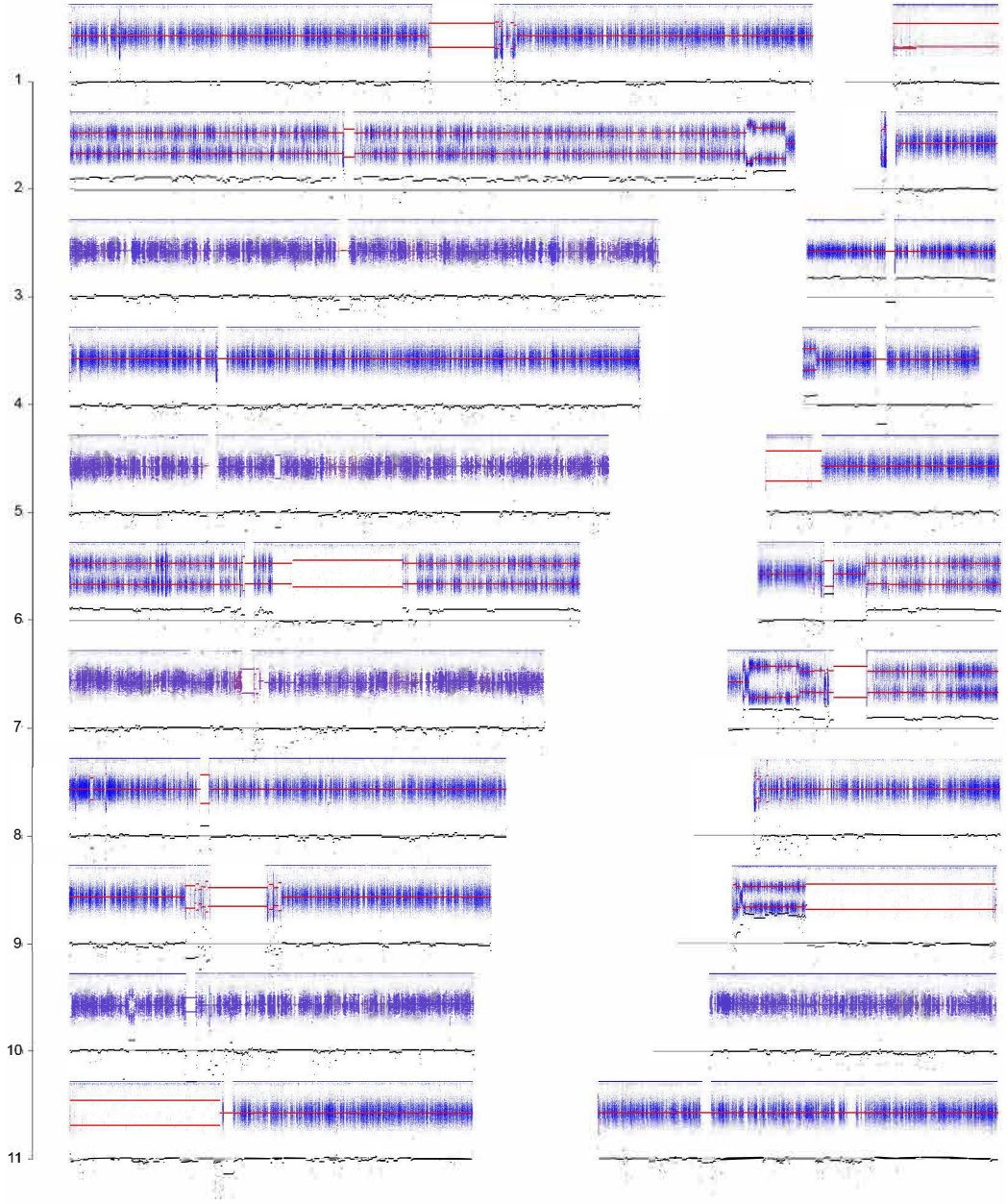
MCF-10A_Cis_5



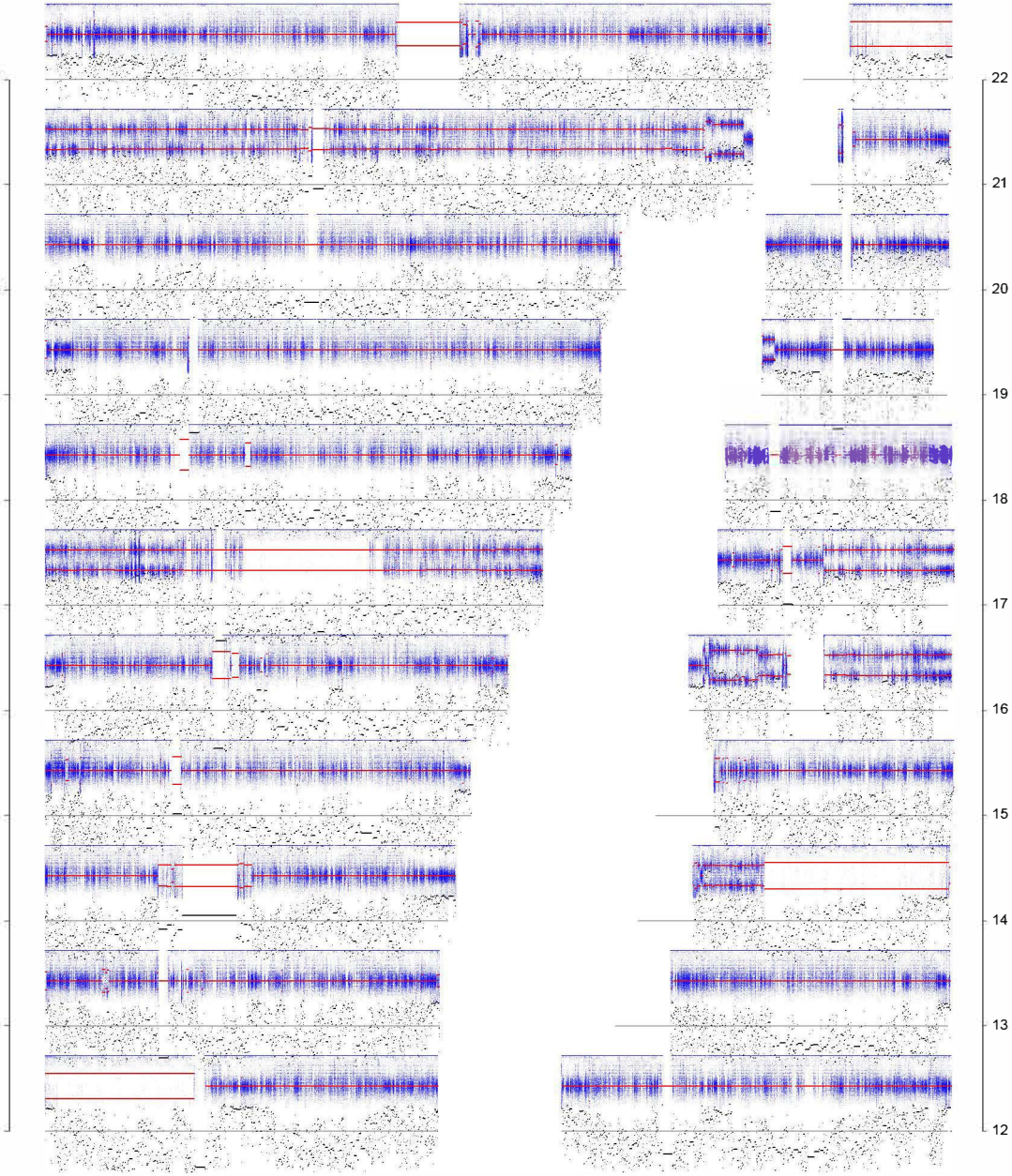
MCF-10A_Cis_6


HepG2_Cis_1
HepG2_Cis_2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

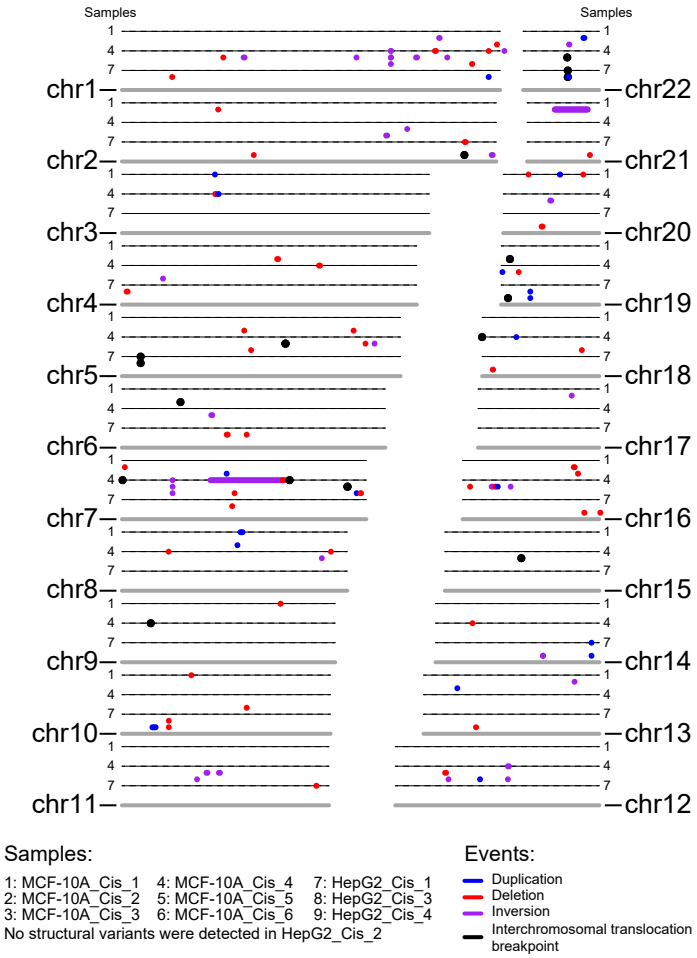
HepG2_Cis_3



HepG2_Cis_4



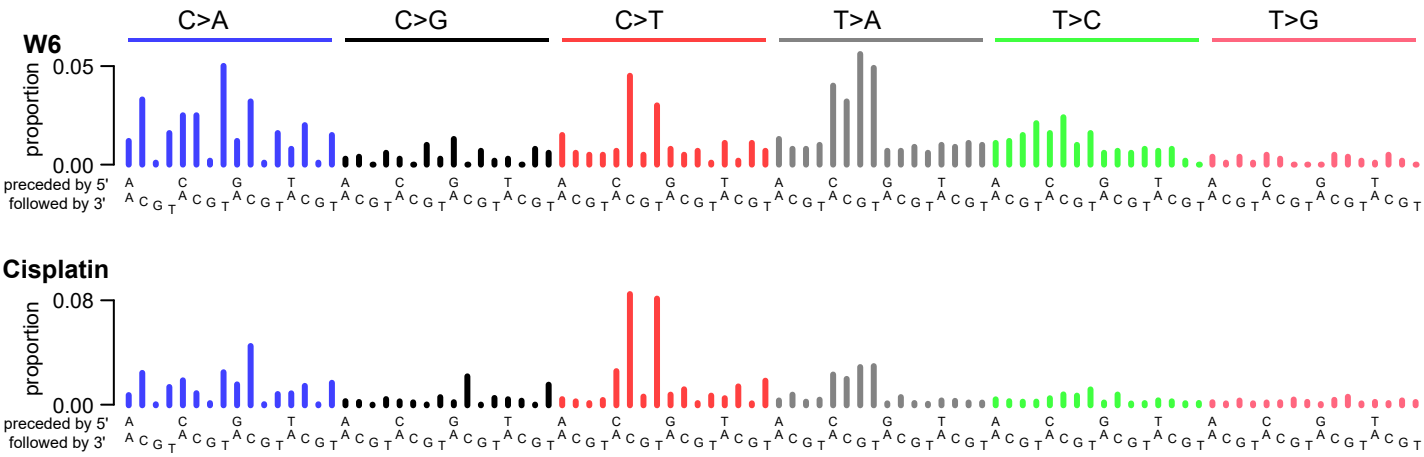
Supplemental Fig. S18: Structural variants detected in cisplatin treated cell line clones. Cell line clones are plotted in separate levels above the chromosomes (in grey). Guide-lines are shown indicating samples 1, 4 and 7 (thin black lines). All interchromosomal translocations were balanced translocations, the details of which are listed in the table below.



Balanced translocations:

Sample_ID	Chrom_start	POS_Start	Chrom_End	POS_End
MCF10A_Cis_3	6	38009460	19	5796141
MCF10A_Cis_3	19	5796141	6	38009460
MCF10A_Cis_4	7	430688	9	18747165
MCF10A_Cis_4	7	108813393	18	11150
MCF10A_Cis_4	9	18747165	7	430688
MCF10A_Cis_4	18	11150	7	108813393
MCF10A_Cis_5	5	106162330	7	146375568
MCF10A_Cis_5	5	106162582	7	146375620
MCF10A_Cis_5	7	146375570	5	106162328
MCF10A_Cis_5	7	146375622	5	106162580
MCF10A_Cis_5	15	49695419	22	28720903
MCF10A_Cis_5	22	28720908	15	49695414
HepG2_Cis_1	5	12159881	22	29065383
HepG2_Cis_1	22	29065382	5	12159882
HepG2_Cis_3	5	12159881	22	29065383
HepG2_Cis_3	22	29065382	5	12159882
HepG2_Cis_4	2	222260650	19	4551288
HepG2_Cis_4	19	4551288	2	222260650

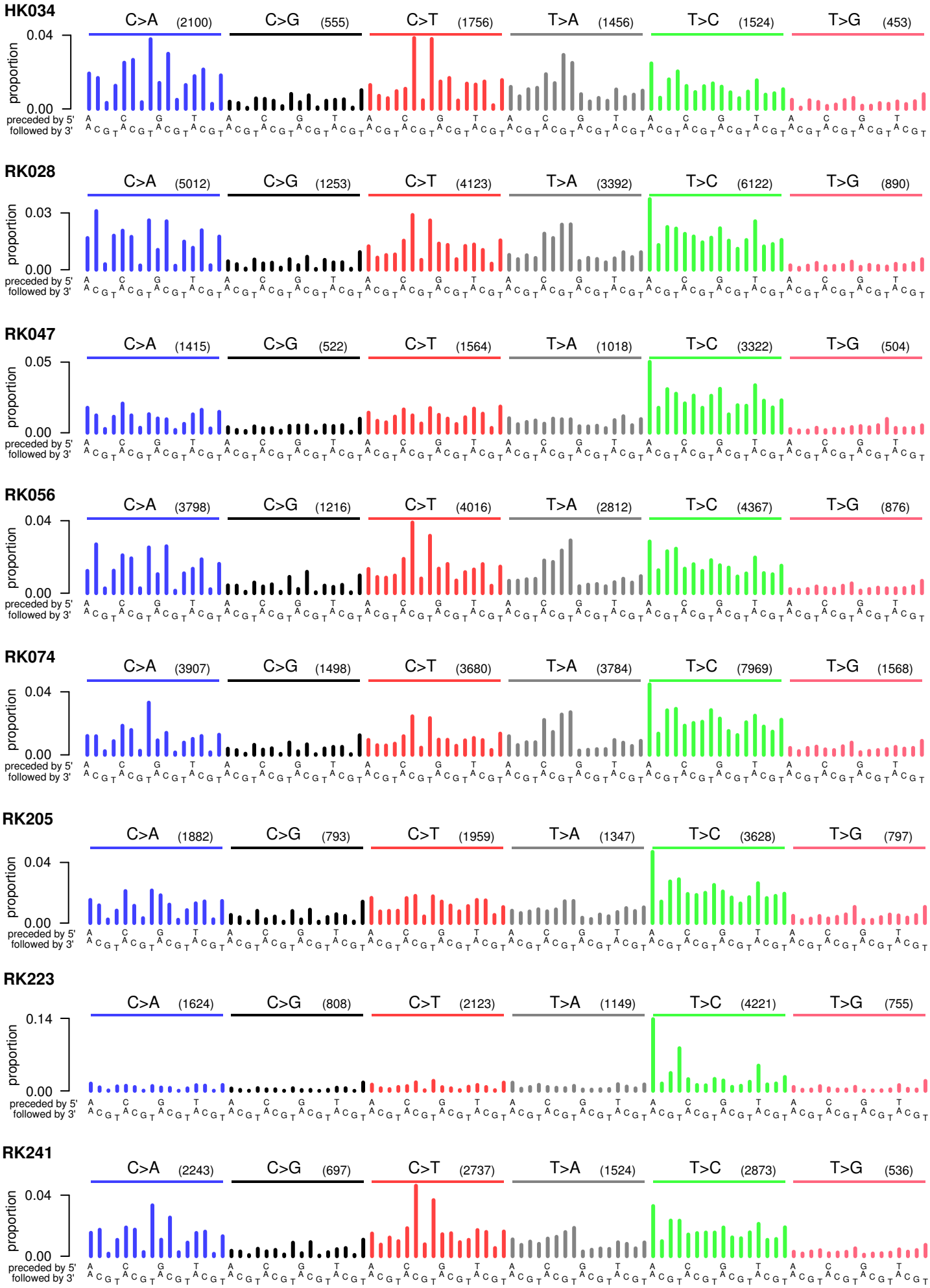
Supplemental Fig. S19: Comparison of signature W6 with experimental cisplatin signature



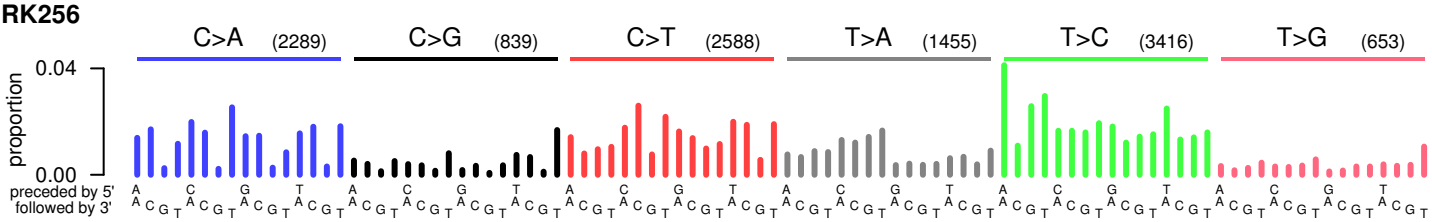
Cosine similarities between W6 and cisplatin

Mutation class	Cosine similarity
Overall	0.781
C>A	0.915
C>G	0.357
C>T	0.917
T>A	0.981
T>C	0.850
T>G	0.767

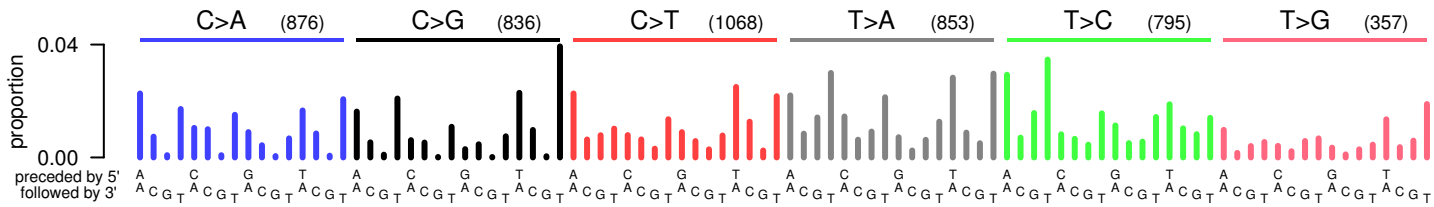
Supplemental Fig. S20: SNS spectrum of HCCs positive for cisplatin in the SNS analysis



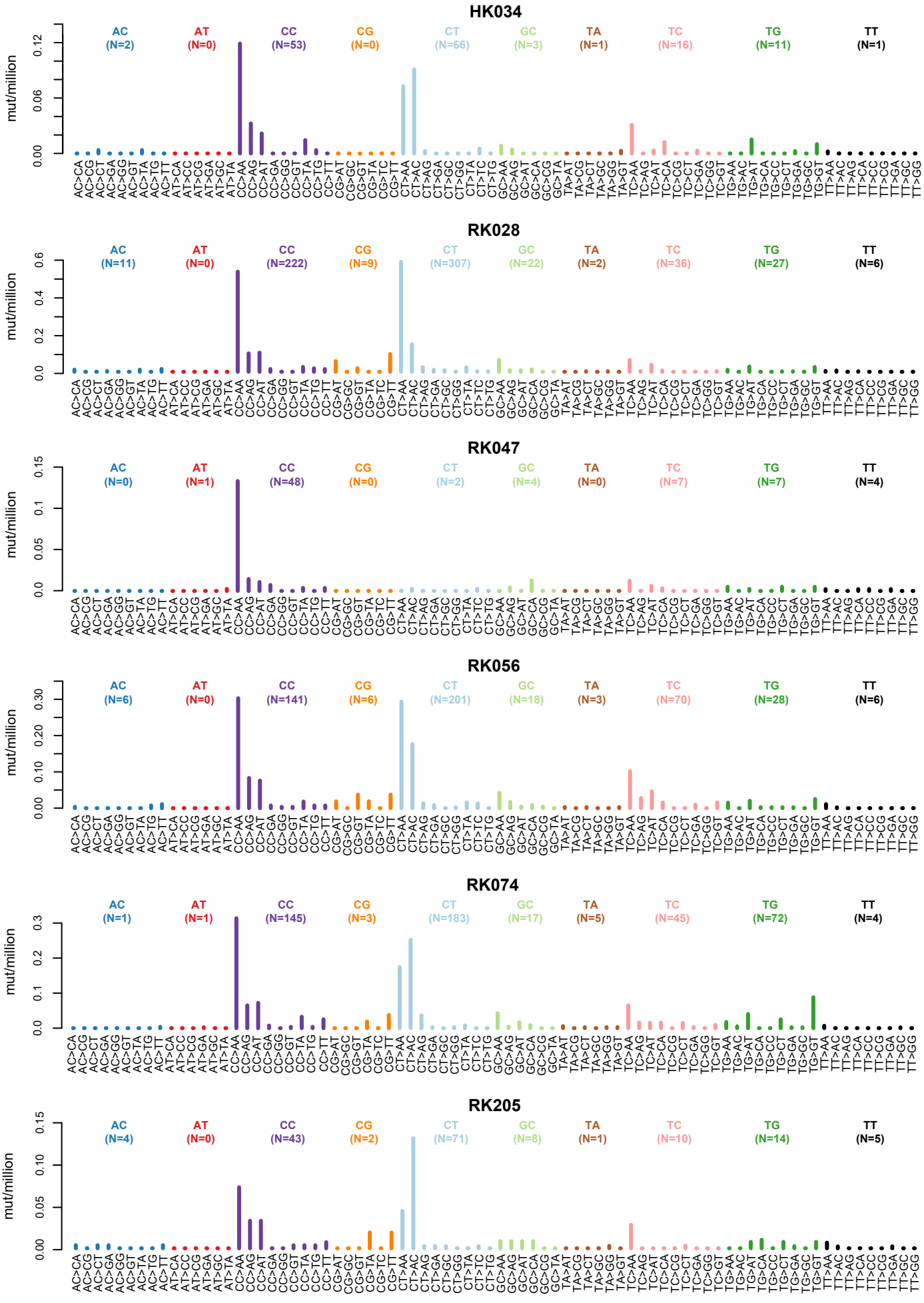
RK256

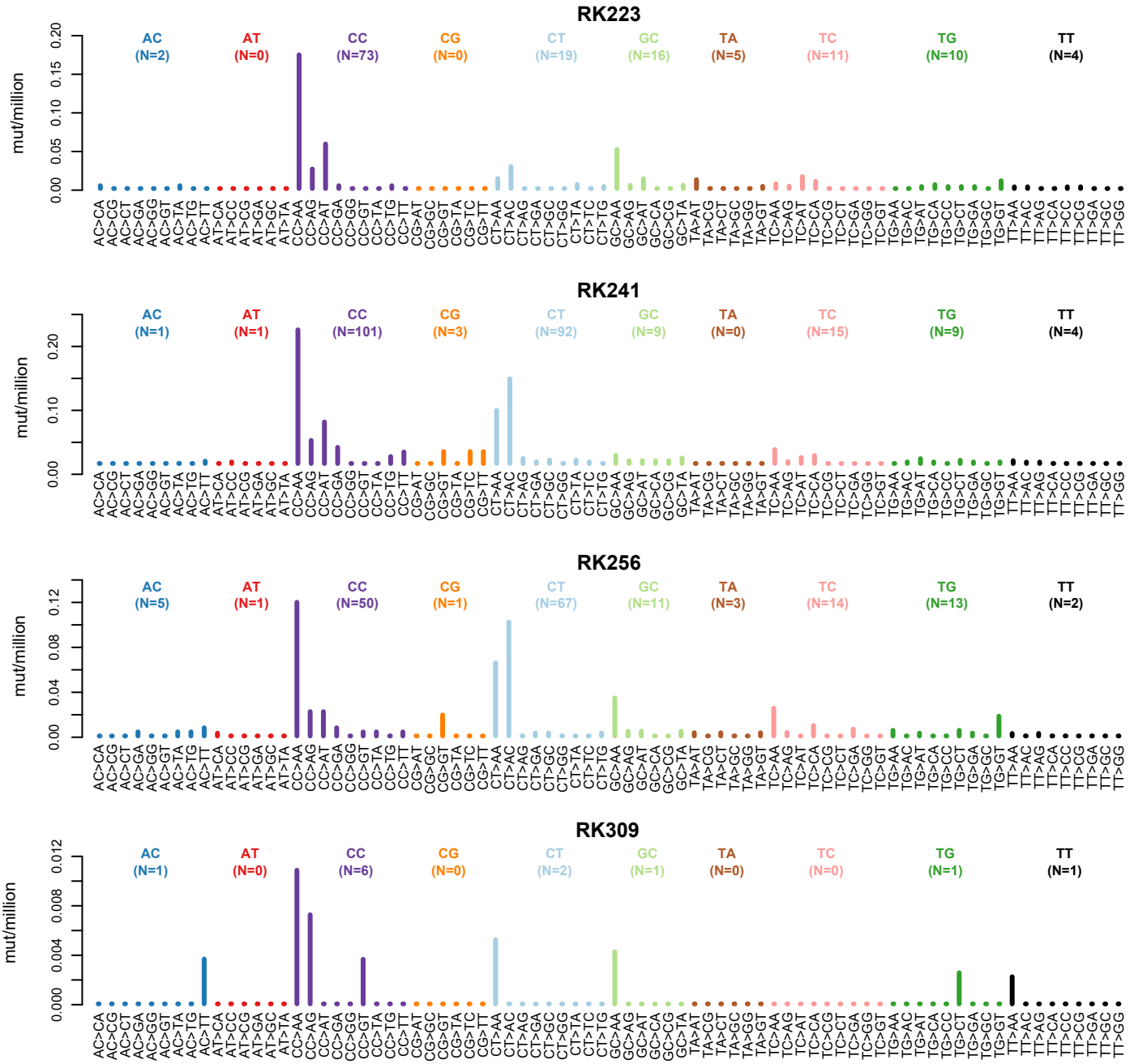

C>A (2289)
C>G (839)
C>T (2588)
T>A (1455)
T>C (3416)
T>G (653)
proportion
0.04
0.00
preceded by 5'
followed by 3'

RK309

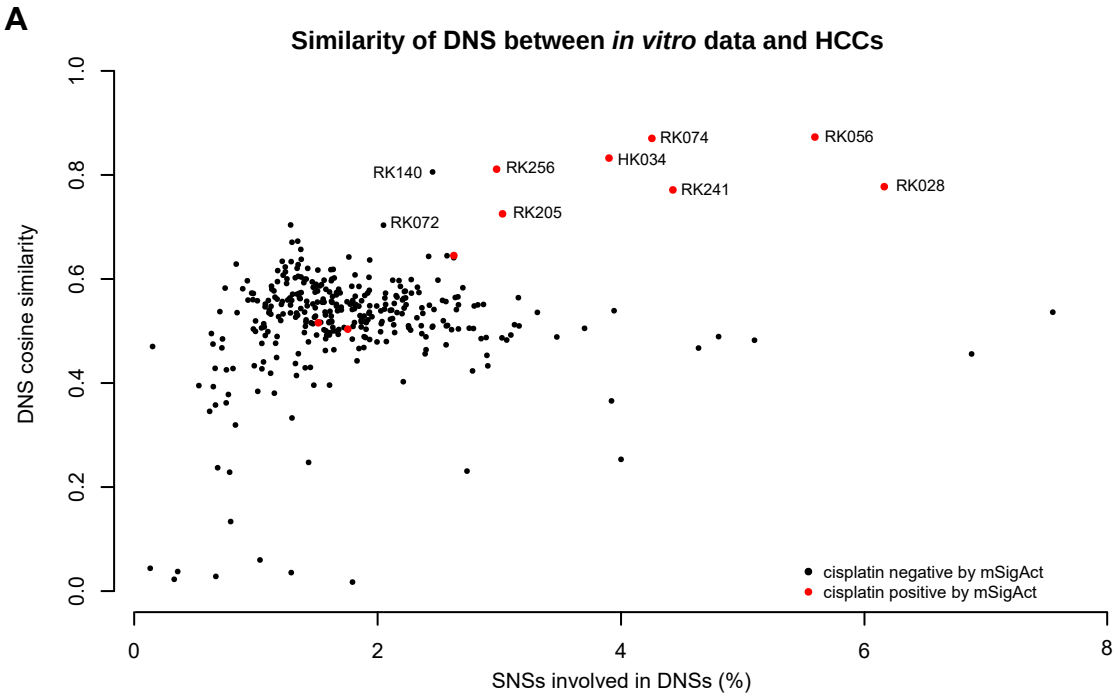

C>A (876)
C>G (836)
C>T (1068)
T>A (853)
T>C (795)
T>G (357)
proportion
0.04
0.00
preceded by 5'
followed by 3'

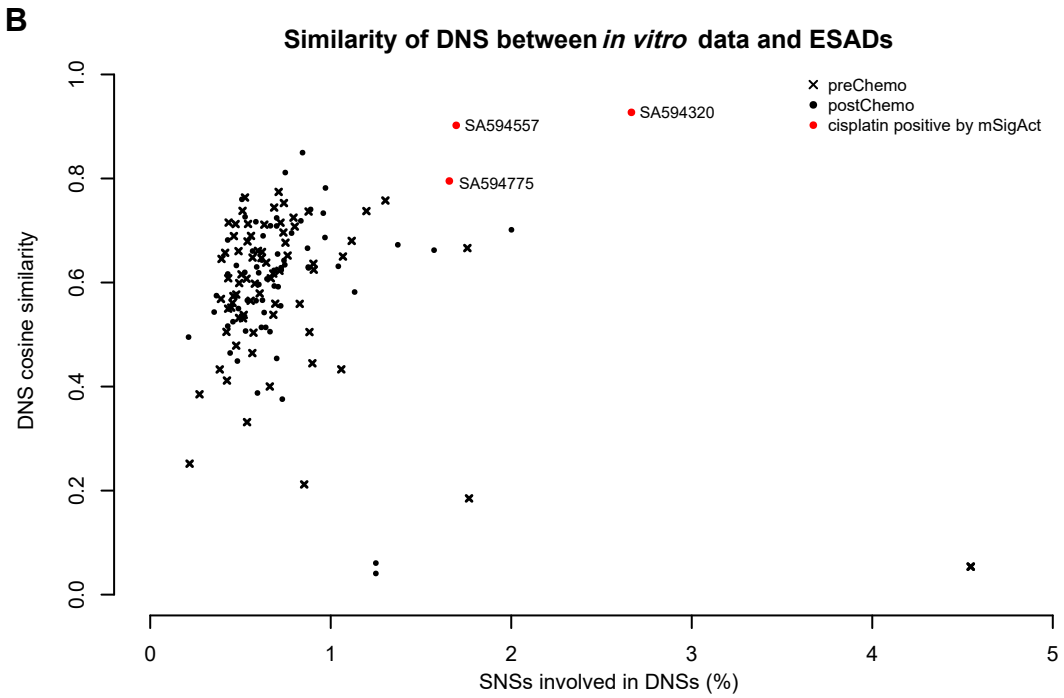
Supplemental Fig. S21: Dinucleotide substitution spectra of HCCs positive for cisplatin in the SNS analysis



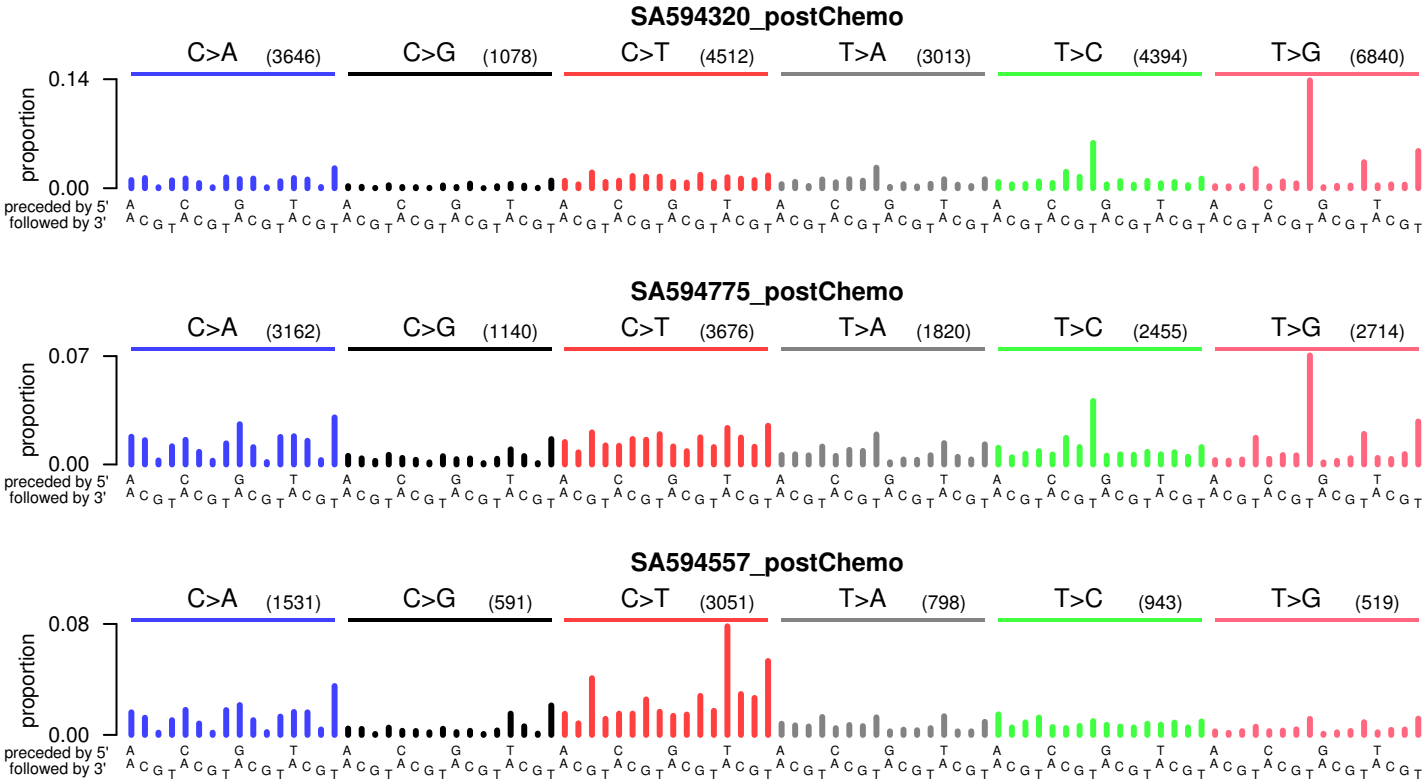

RK223
AC (N=2)
AT (N=0)
CC (N=73)
CG (N=0)
CT (N=19)
GC (N=16)
TA (N=5)
TC (N=11)
TG (N=10)
TT (N=4)
mut/million
0.00
0.05
0.10
0.15
0.20
RK241
AC (N=1)
AT (N=1)
CC (N=101)
CG (N=3)
CT (N=92)
GC (N=9)
TA (N=0)
TC (N=15)
TG (N=9)
TT (N=4)
mut/million
0.00
0.10
0.20
RK256
AC (N=5)
AT (N=1)
CC (N=50)
CG (N=1)
CT (N=67)
GC (N=11)
TA (N=3)
TC (N=14)
TG (N=13)
TT (N=2)
mut/million
0.00
0.04
0.08
0.12
RK309
AC (N=1)
AT (N=0)
CC (N=6)
CG (N=0)
CT (N=2)
GC (N=1)
TA (N=0)
TC (N=0)
TG (N=1)
TT (N=1)
mut/million
0.000
0.004
0.008
0.012
AC>CA AC>CG AC>CT AC>GA AC>GG AC>GT AC>TA AC>TG AC>TT AT>CA AT>CC AT>CG AT>GA AT>GC AT>TA CC>AA CC>AG CC>AT CC>GA CC>GG CC>GT CC>TA CC>TG CC>TT CG>AT CG>GC CG>GT CG>TA CG>TC CG>TT CT>AA CT>AC CT>AG CT>GA CT>GC CT>GG CT>TA CT>TC CT>TG GC>AA GC>AG GC>AT GC>CA GC>CG GC>TA TA>AT TA>CG TA>CT TA>GC TA>GG TA>GT TC>AA TC>AG TC>AT TC>CA TC>CG TC>CT TC>GA TC>GG TC>GT TG>AA TG>AC TG>AT TG>CA TG>CC TG>CT TG>GA TG>GC TG>GT TT>AA TT>AC TT>AG TT>CA TT>CC TT>CG TT>GA TT>GC TT>GG

Supplemental Fig. S22: Clustering of patients according to the percentage of SNSs involved in DNSs and the cosine similarity of their DNS spectrum with that of the experimental cisplatin DNS signature. Samples that were identified to be positive for the cisplatin mutational signature in the mSigAct SNS analysis are displayed in red. A: HCCs, B: ESADs. For the ESADs, samples with known prior exposure to platinum-based chemotherapeutics are shown as circles; samples without prior platinum-based treatment are shown as crosses.

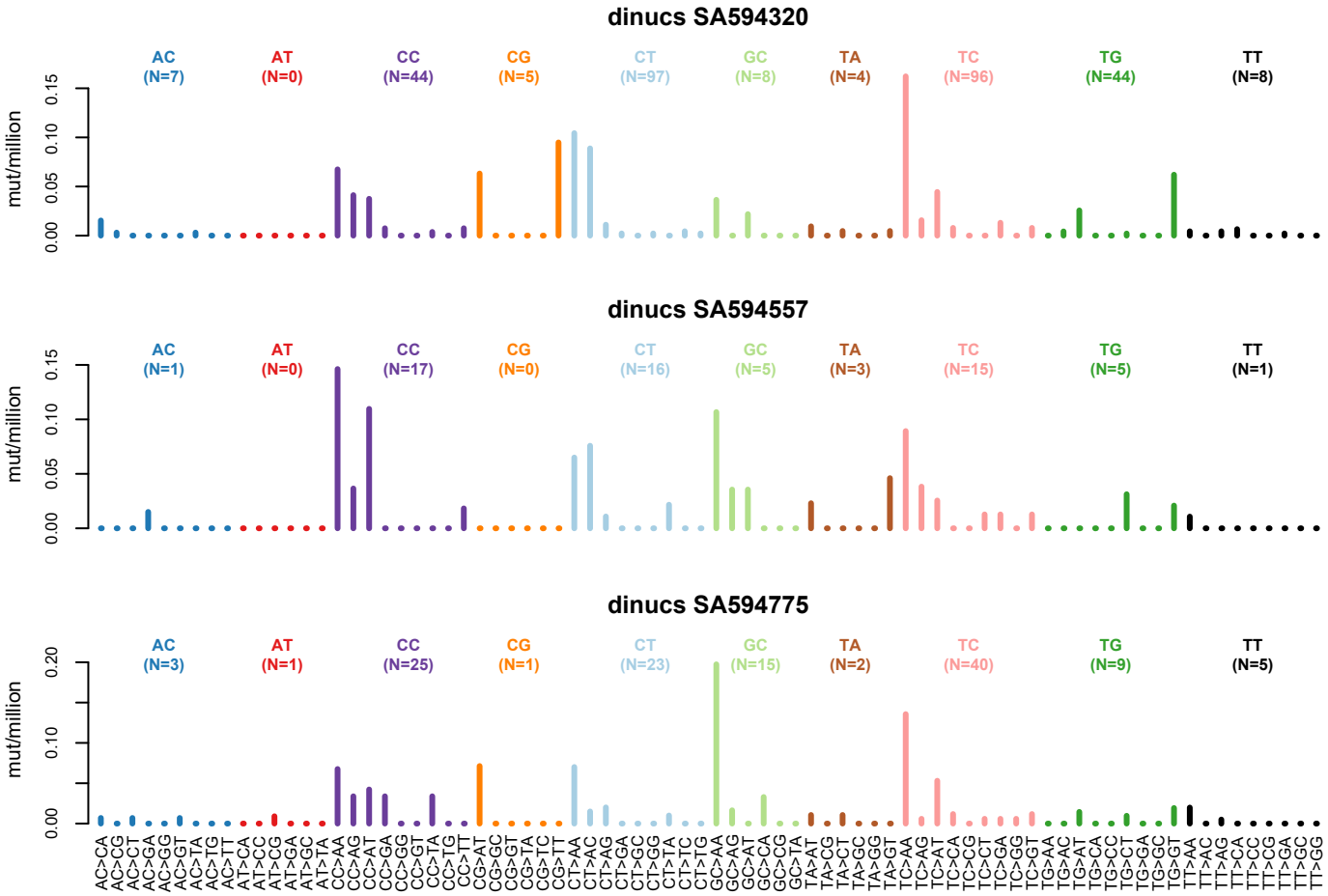




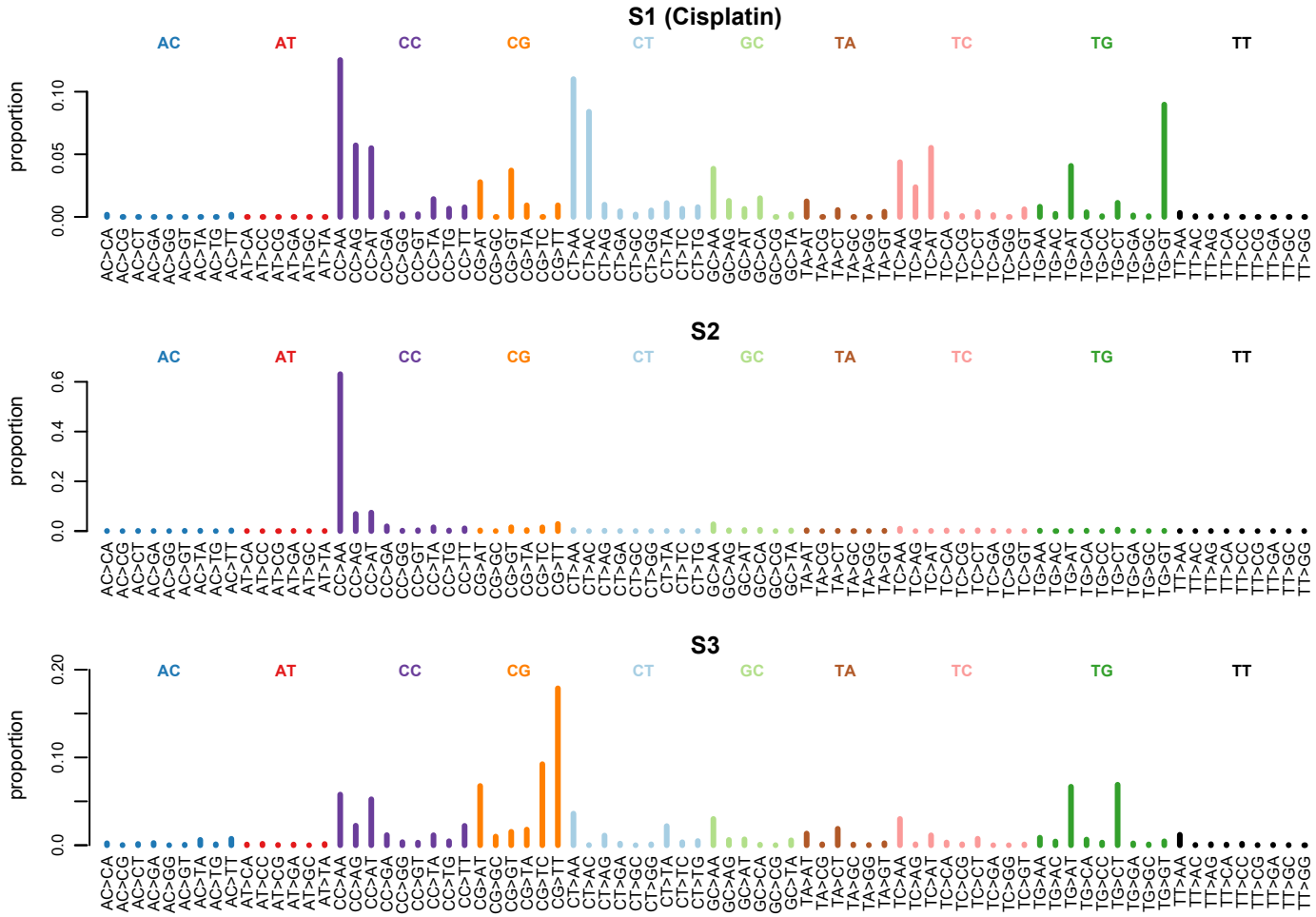
Supplemental Fig. S23: SNS spectrum of esophageal adenocarcinomas positive for cisplatin in the SNS analysis.



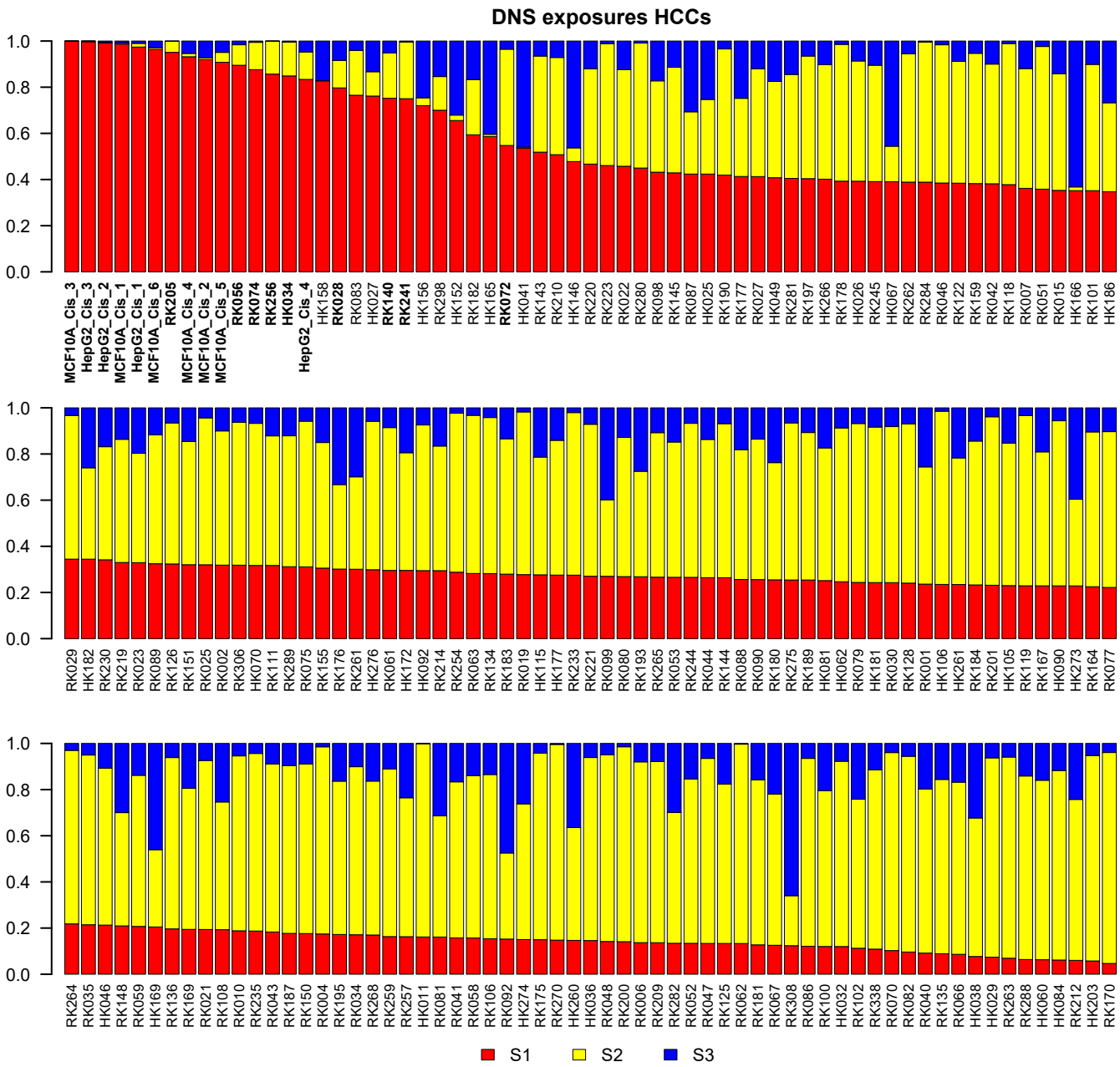
Supplemental Fig. S24: Dinucleotide substitution spectra of esophageal adenocarcinomas positive for cisplatin in the SNS analysis.



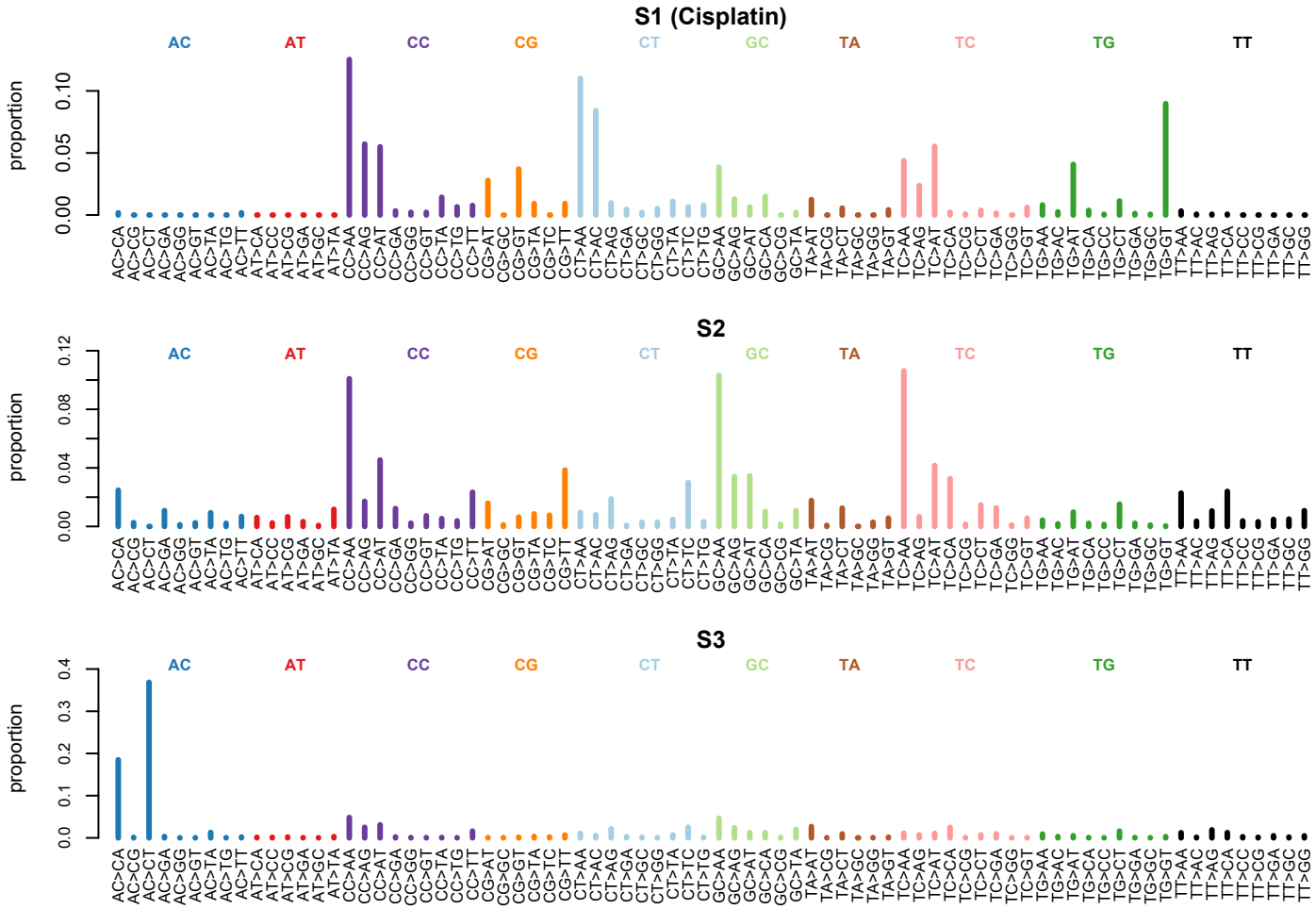
Supplemental Fig. S25: ssNMF result on DNS spectra of all HCCs with at least 25 DNSs. A: DNS signatures extracted by NMF



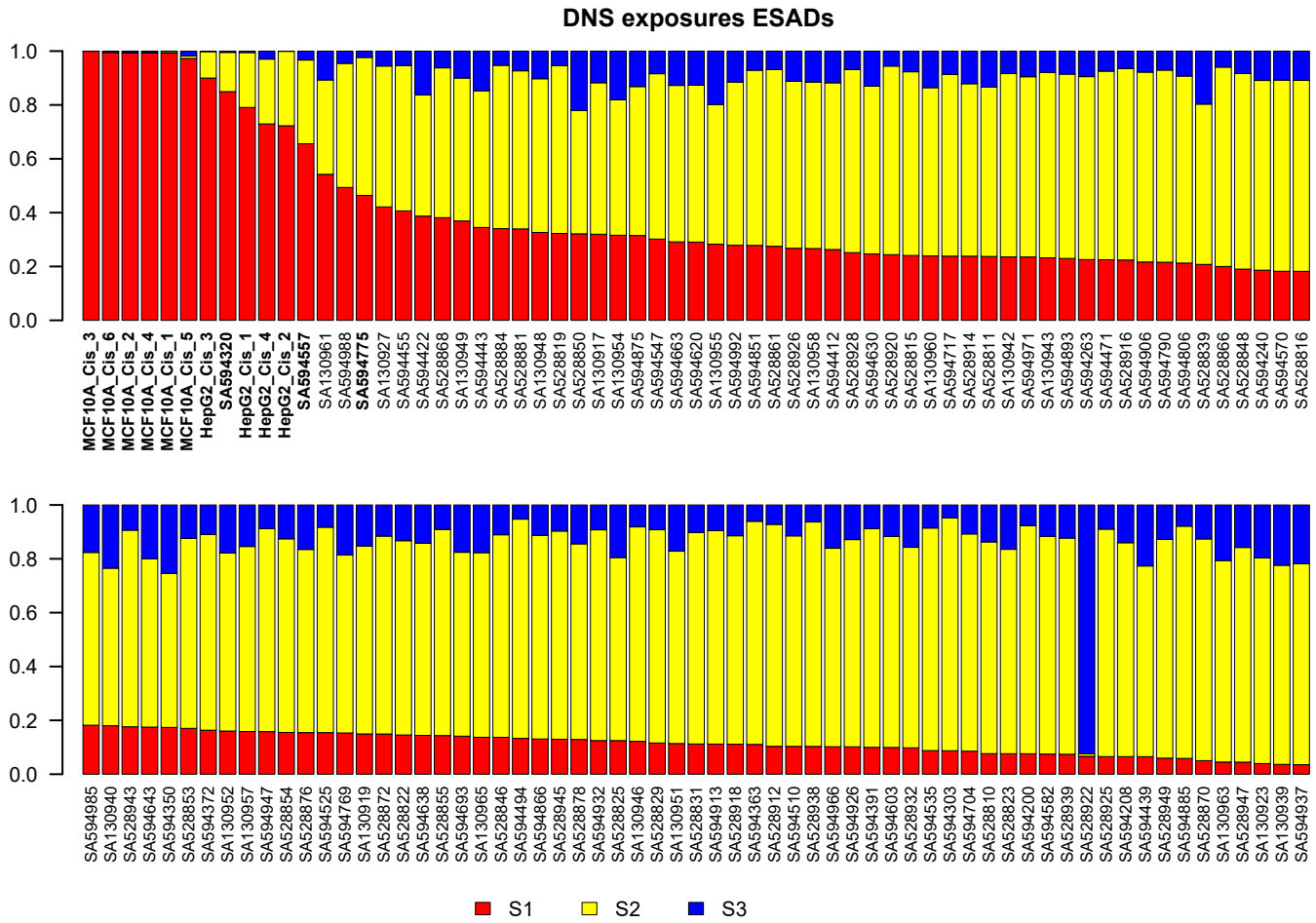
B: DNS signature exposures in HCCs



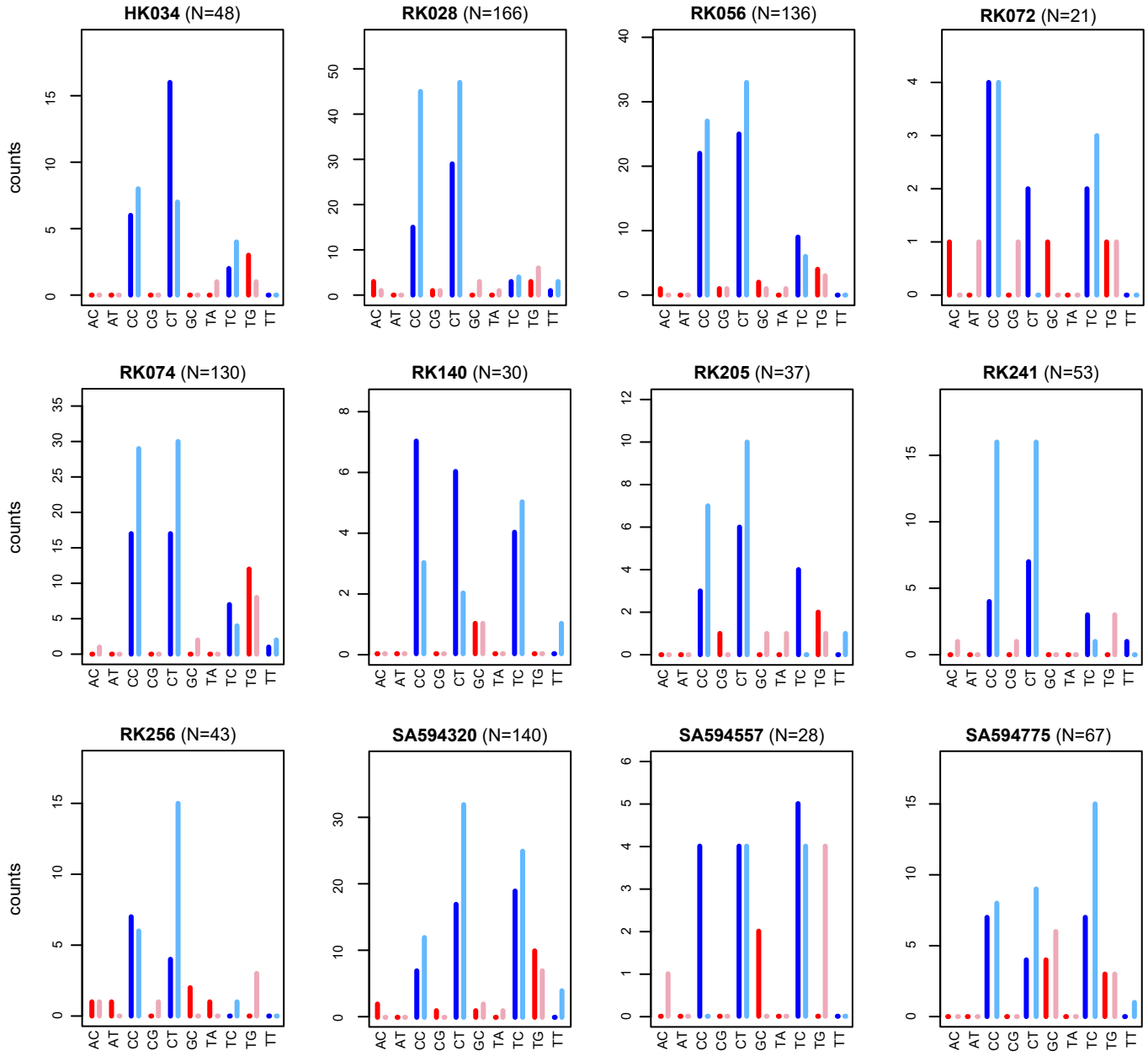
Supplemental Fig. S26: ssNMF result on DNS spectra of all ESADs with at least 25 DNSs. A: DNS signatures extracted by NMF



B: DNS signature exposures in ESADs



Supplemental Fig. S27: DNS transcription strand bias for tumors positive for cisplatin mutagenesis. The total number of DNSs eligible for transcription strand bias analysis is displayed in parentheses.

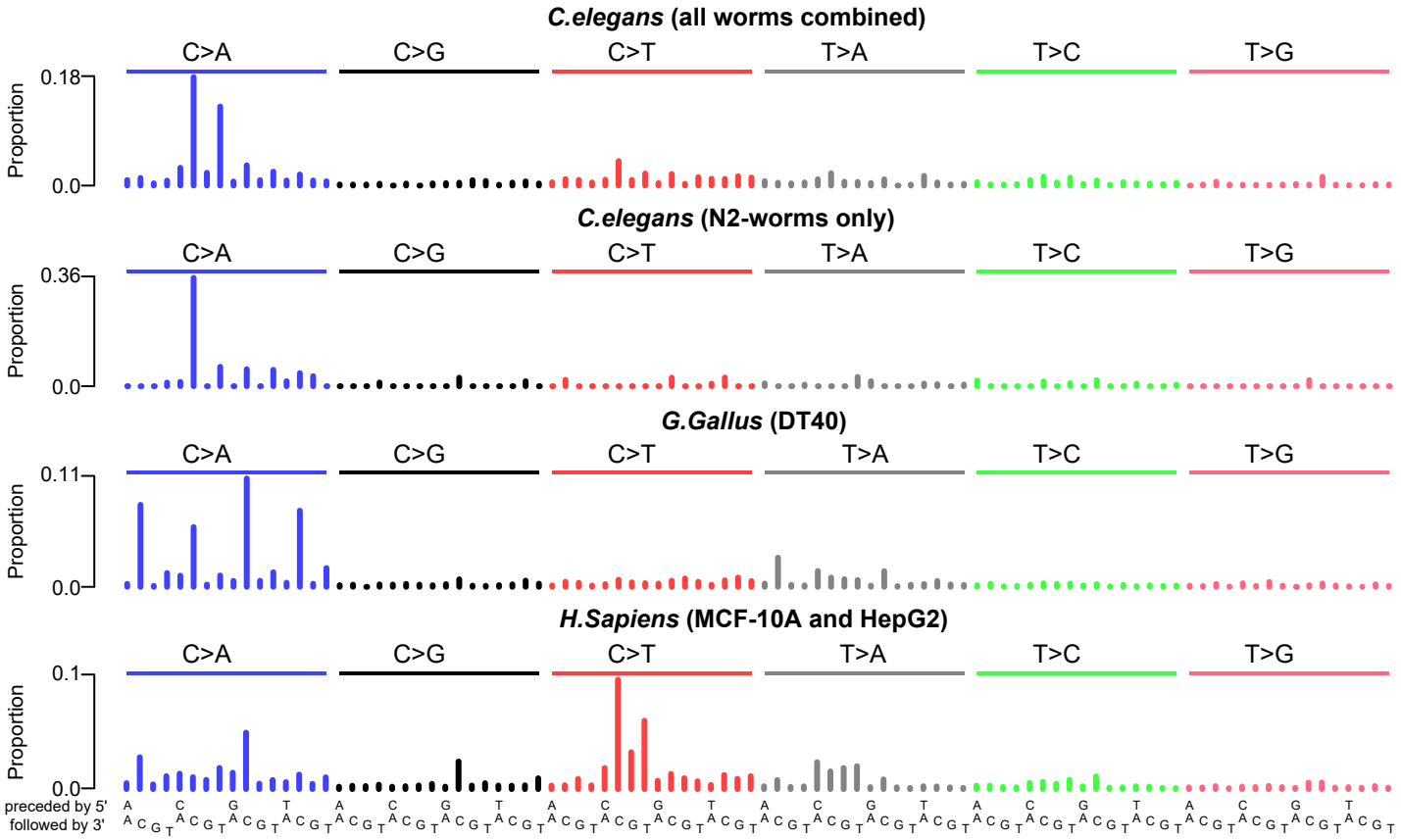


Transcriptional strand bias of dinucletotides in HCCs and ESADs*

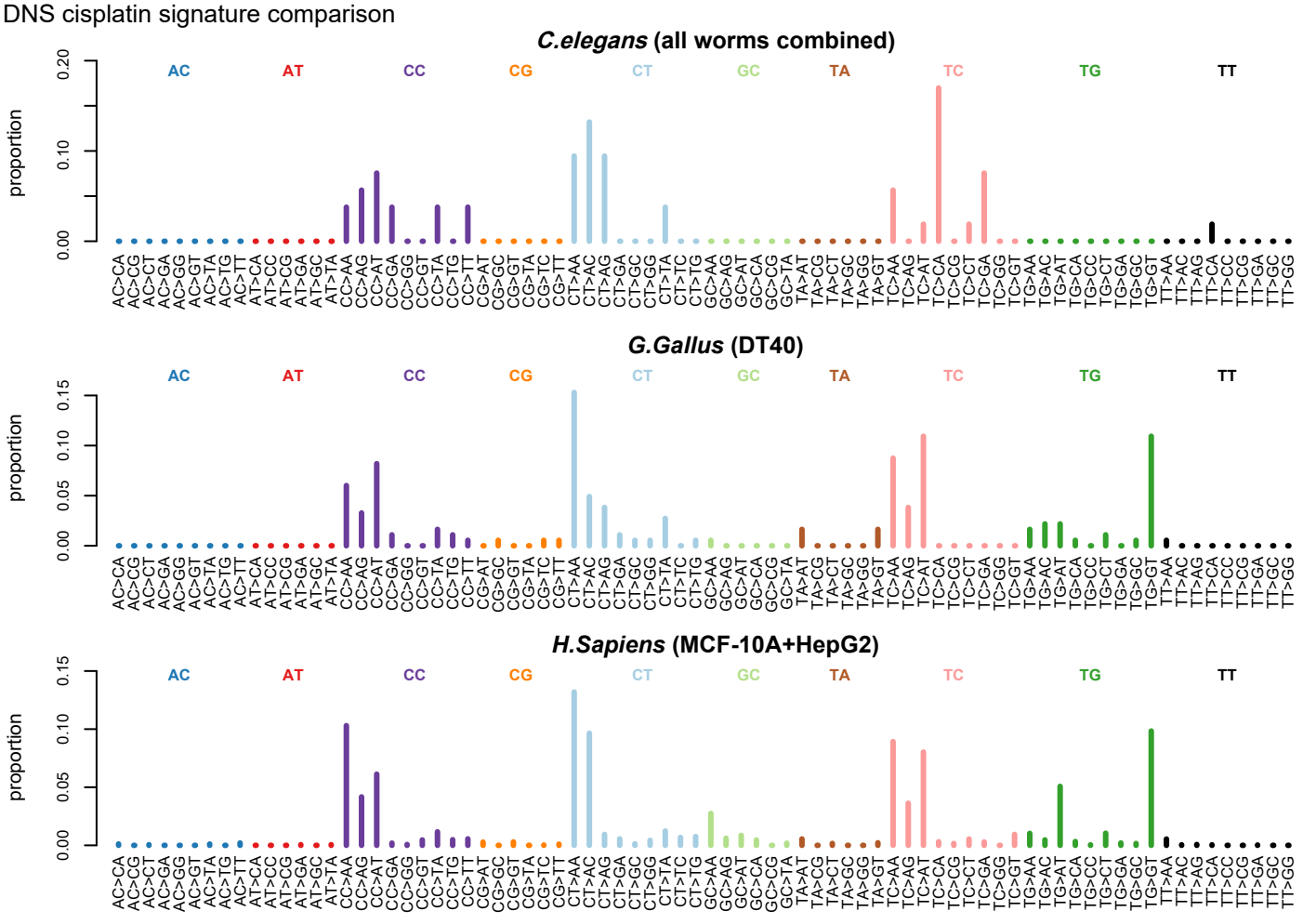
Sample	Intrastrand	Interstrand
HK034	NS	NS
RK056	2.1E-01	NS
RK072	NS	NS
RK074	1.6E-02	NS
RK140	NS	NS
RK205	NS	NS
RK241	6.6E-03	NS
RK256	4.0E-02	NS
RK028	1.6E-05	NS
SA594320	3.4E-03	NS
SA594557	NS	NS
SA594775	2.4E-02	NS

*: NS = not significant

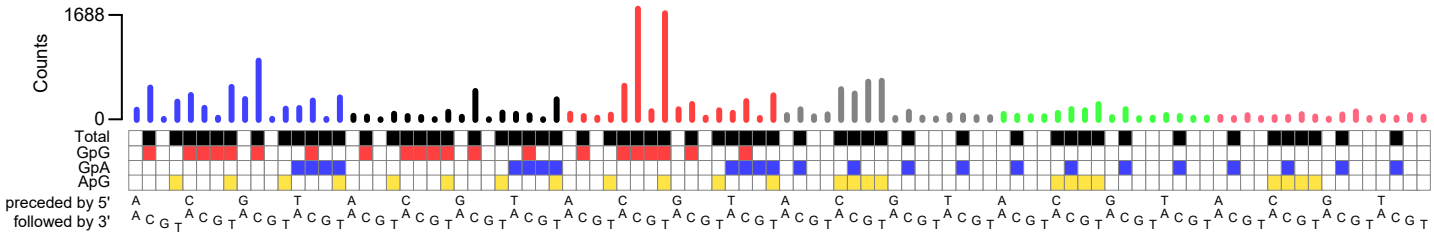
Supplemental Fig. S28: Comparison of cisplatin SNS (A) and DNS (B) mutational signature with previously published cisplatin signatures. *C. elegans* (all worms combined) is the combined mutation spectrum of all mutations of worms treated with cisplatin, regardless of DNA repair deficiency (total 681 SNSs and 53 DNSs). *C. elegans* (N2-worms only) is the mutation spectrum of the mutations of N2 (DNA repair proficient) worms treated with cisplatin (total 51 SNSs, there is no DNS plot, as there was only a single TC>CA DNS). *G. gallus* (DT40) is the mutation spectrum of the 3 DT40 clones sequenced (total 2,436 SNSs and 183 DNSs). *H. sapiens* (MCF-10A and HepG2) is the consensus mutation spectrum of all MCF-10A and HepG2 clones combined (70,313 SNSs and 2,839 DNSs).



B: DNS cisplatin signature comparison

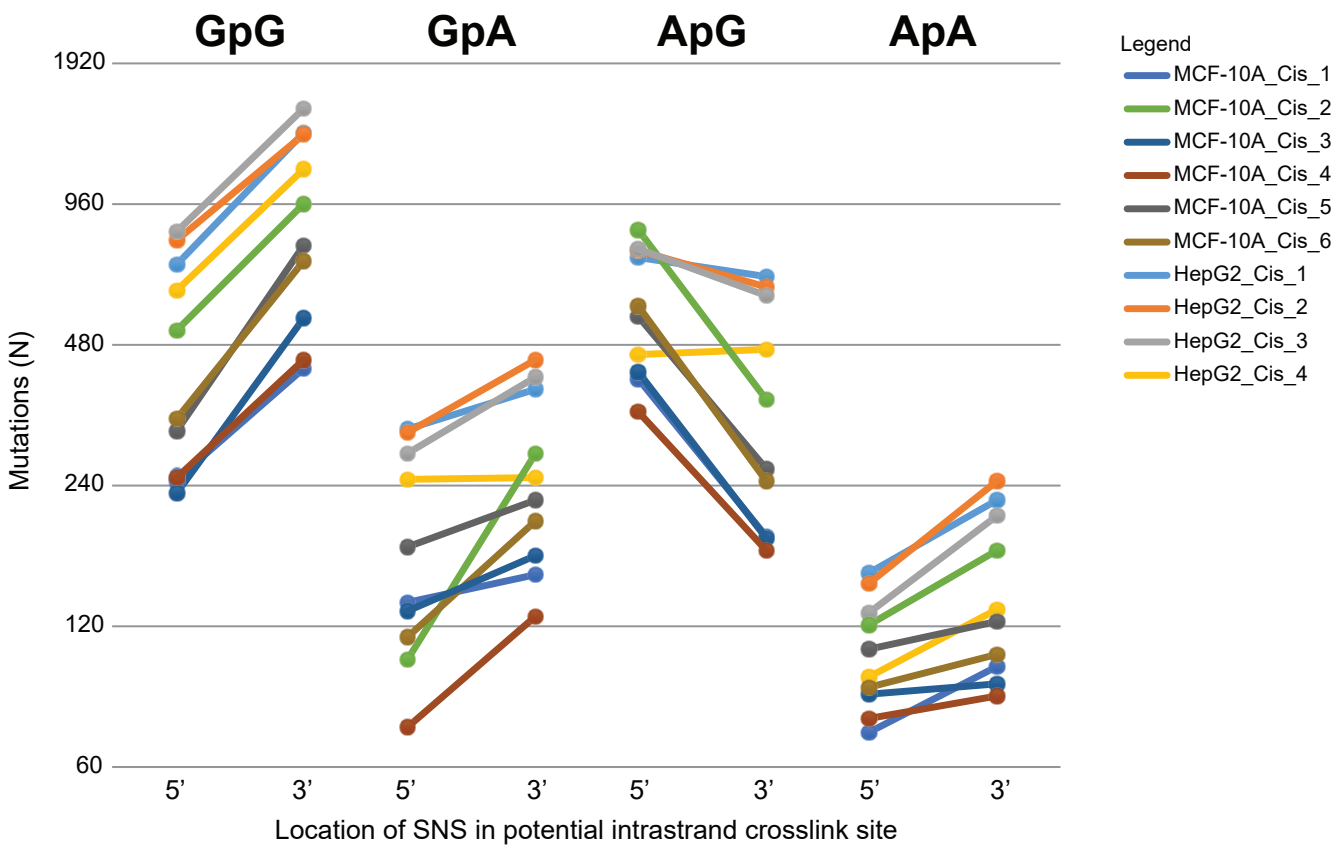


Supplemental Fig. S29: Schematic display of which peaks from the SNS mutation spectra are inside GpG, GpA or ApG dinucleotides.

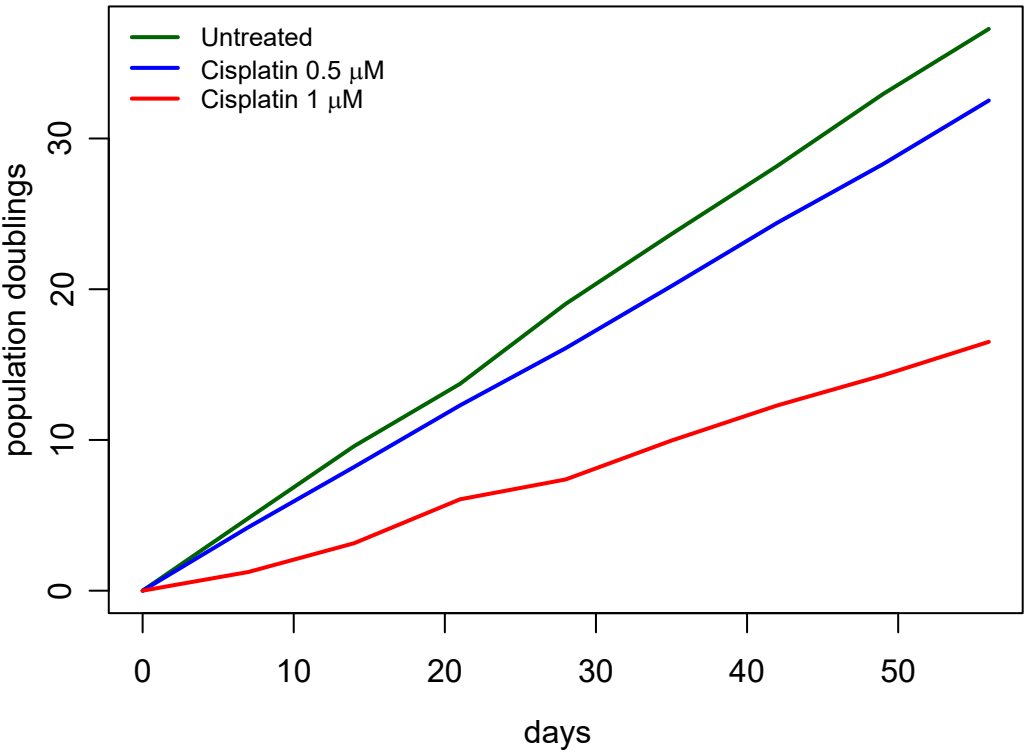


*: Total indicates trinucleotides containing any potential cisplatin intrastrand crosslink sites (GpG, GpA or ApG)

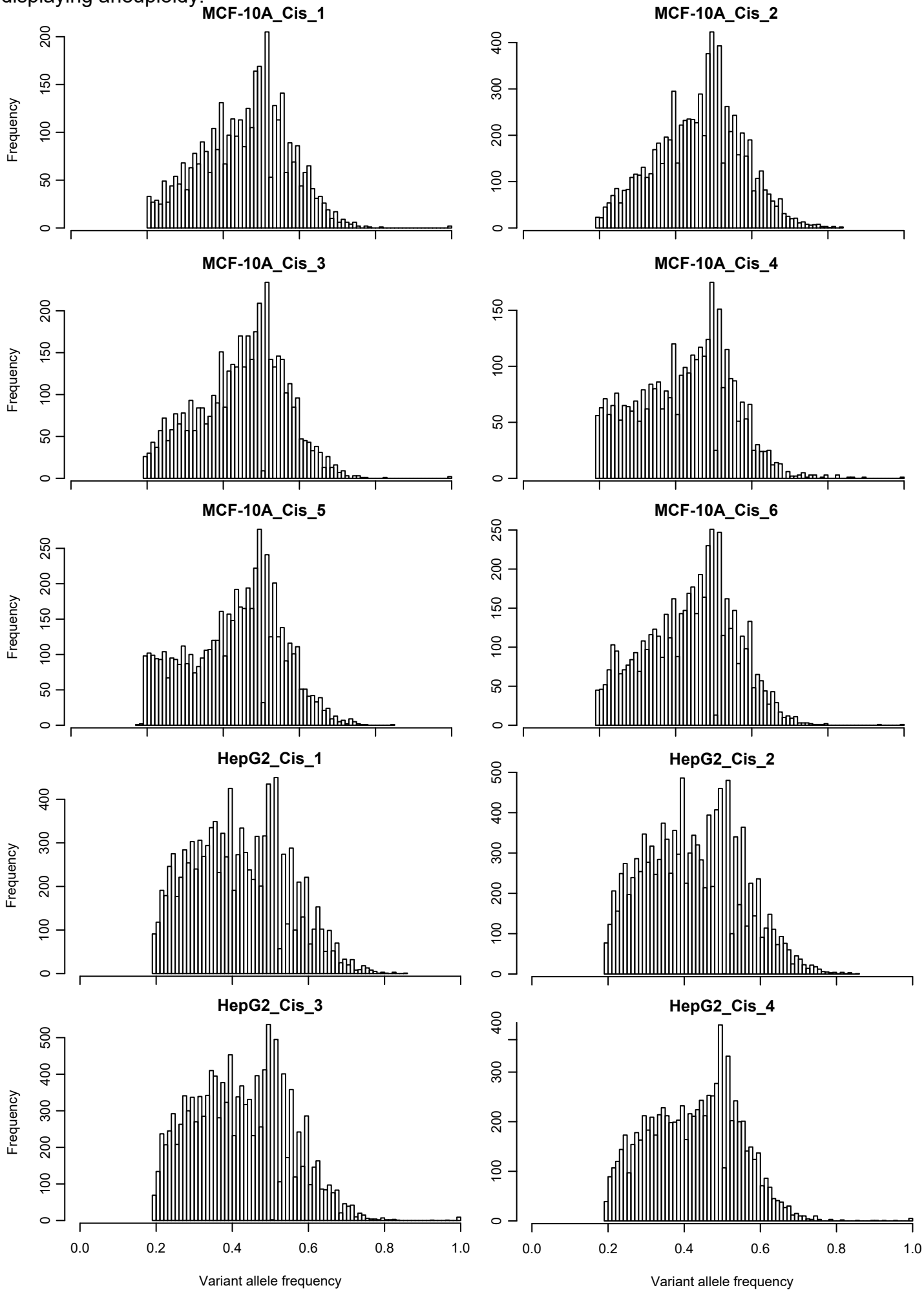
Supplemental Fig. S30: Bias of translesion synthesis towards misincorporation across the 5’ or 3’ adducted-base in the 4 types of intrastrand adducts induced by cisplatin.
(note the log scale of the Y-axis)



Supplemental Fig. S31: MCF-10A proliferation rate during cisplatin exposure



Supplemental Fig. S32: Variant allele frequency distribution of the cisplatin mutations. The tail towards the lower end of the histogram represents variants inside regions of the MCF-10A genome displaying aneuploidy.



Supplemental Table S1: Sequencing statistics and substitutions in cisplatin treated cell line clones

	MCF10A							HepG2				
Sample_ID	**Parental**	**1**	**2**	**3**	**4**	**5**	**6**	**Parental**	**1**	**2**	**3**	**4**
Cisplatin concentration	--	0.5µM	1µM	0.5µM	0.5µM	1µM	1µM	--	0.75µM	0.75µM	0.75µM	0.75µM
Exposure time (weeks)	--	4	4	8	8	8	8	--	8	8	8	8
Sequencing statistics												
Reads mapped (%)	99.65	99.67	99.67	99.84	99.61	99.63	99.74	99.47	99.89	99.88	99.89	99.92
Mean mapping quality	50.73	50.73	50.81	50.69	49.19	49.50	50.10	50.86	50.03	49.86	50.51	50.65
mean coverage	35.06	34.16	35.69	36.31	41.3	40.70	33.13	40.98	35.96	37.62	34.55	34.60
Substitutions												
Single nucleotide substitutions	--	3655	7712	4460	3544	5522	5260	--	10263	10515	11202	8180
Dinucleotide substitutions	--	136	328	149	123	171	199	--	434	484	505	310
Trinucleotide substitutions	--	2	10	2	2	6	6	--	8	10	6	1
Deletions	--	152	309	145	213	266	283	--	340	350	371	200
Insertions	--	54	142	78	89	141	136	--	237	267	280	155
Substitutions												
Mutation density (mut/mb)	--	1.26	2.66	1.54	1.22	1.91	1.82	--	3.54	3.63	3.87	2.82
SNSs in DNSs	--	7.4%	8.5%	6.7%	6.9%	6.2%	7.6%	--	8.5%	9.2%	9.0%	7.6%
ENA accession	ERS1836384	ERS1836385	ERS1836386	ERS1928084	ERS1928085	ERS1928086	ERS1928087	ERS2004856	ERS2004855	ERS2004854	ERS2004853	ERS2004852

#Supplemental Table S2: Consensus cisplatin SNS signature; weighted average of the 6 MCF 10A clones and the 4 HepG2 clones.

ACA>AAA	0.00936678
ACC>AAC	0.02965292
ACG>AAG	0.000941593
ACT>AAT	0.016762243
CCA>CAA	0.022843424
CCC>CAC	0.011100618
CCG>CAG	0.001999424
CCT>CAT	0.030228879
GCA>GAA	0.019135663
GCC>GAC	0.054234893
GCG>GAG	0.00102869
GCT>GAT	0.010337297
TCA>TAA	0.010960297
TCC>TAC	0.017780008
TCG>TAG	0.000941647
TCT>TAT	0.020611633
ACA>AGA	0.003811876
ACC>AGC	0.003077642
ACG>AGG	0.000613865
ACT>AGT	0.005699248
CCA>CGA	0.003608814
CCC>CGC	0.002712672
CCG>CGG	0.000792245
CCT>CGT	0.007482123
GCA>GGA	0.002866315
GCC>GGC	0.026493711
GCG>GGG	0.000667852
GCT>GGT	0.006739952
TCA>TGA	0.005566101
TCC>TGC	0.004222515
TCG>TGG	0.000676115
TCT>TGT	0.018870884
ACA>ATA	0.005645534
ACC>ATC	0.003625505
ACG>ATG	0.002040361
ACT>ATT	0.004857312
CCA>CTA	0.031259149
CCC>CTC	0.101702502
CCG>CTG	0.007967736
CCT>CTT	0.09757019
GCA>GTA	0.009773178
GCC>GTC	0.014642504
GCG>GTG	0.002052975
GCT>GTT	0.008869208
TCA>TTA	0.006516012

TCC>TTC	0.017279439
TCG>TTG	0.001824739
TCT>TTT	0.022530916
ATA>AAA	0.004450587
ATC>AAC	0.009735718
ATG>AAG	0.003330778
ATT>AAT	0.005155571
CTA>CAA	0.028151266
CTC>CAC	0.024159493
CTG>CAG	0.034982991
CTT>CAT	0.035847387
GTA>GAA	0.001804298
GTC>GAC	0.007525187
GTG>GAG	0.002111257
GTT>GAT	0.001389473
TTA>TAA	0.004471737
TTC>TAC	0.004086108
TTG>TAG	0.002584201
TTT>TAT	0.00251771
ATA>ACA	0.005473727
ATC>ACC	0.003774145
ATG>ACG	0.003317564
ATT>ACT	0.003408848
CTA>CCA	0.006473984
CTC>CCC	0.009982651
CTG>CCG	0.008871106
CTT>CCT	0.01454855
GTA>GCA	0.002571098
GTC>GCC	0.009870109
GTG>GCG	0.001700346
GTT>GCT	0.002077657
TTA>TCA	0.00456193
TTC>TCC	0.003657906
TTG>TCG	0.001961637
TTT>TCT	0.002185522
ATA>AGA	0.002837065
ATC>AGC	0.001783529
ATG>AGG	0.004379581
ATT>AGT	0.001849858
CTA>CGA	0.00262128
CTC>CGC	0.002812001
CTG>CGG	0.005300296
CTT>CGT	0.003160225
GTA>GGA	0.001186192
GTC>GGC	0.005022697
GTG>GGG	0.007821102

GTT>GGT	0.001775102
TTA>TGA	0.00313576
TTC>TGC	0.002073962
TTG>TGG	0.004483479
TTT>TGT	0.003032136

#Supplemental Table S3: Consensus cisplatin DNS signature; weighted average of the 6 MCF 10A clones and the 4 HepG2 clones.

AC>CA	0.001192677
AC>CG	0
AC>CT	0.000577034
AC>GA	0
AC>GG	0
AC>GT	0
AC>TA	0.001154068
AC>TG	0
AC>TT	0.002058228
AT>CA	0
AT>CC	0
AT>CG	0
AT>GA	0.000577034
AT>GC	0
AT>TA	0.000577034
CC>AA	0.102979447
CC>AG	0.041684709
CC>AT	0.06126526
CC>GA	0.001933273
CC>GG	0.000740597
CC>GT	0.004654881
CC>TA	0.011647375
CC>TG	0.004857051
CC>TT	0.005597647
CG>AT	0.002798824
CG>GC	0
CG>GT	0.002962387
CG>TA	0
CG>TC	0.000288517
CG>TT	0.001029114
CT>AA	0.131921669
CT>AC	0.096527465
CT>AG	0.009377847
CT>GA	0.005597648
CT>GC	0.001356239
CT>GG	0.004318625
CT>TA	0.012301625
CT>TC	0.006828931
CT>TG	0.007569529
GC>AA	0.027468072
GC>AG	0.006174682
GC>AT	0.008569164
GC>CA	0.004607142
GC>CG	0
GC>TA	0.001894665

TA>AT 0.005511301
TA>CG 0
TA>CT 0.001644756
TA>GC 0
TA>GG 0
TA>GT 0.00222179
TC>AA 0.08924228
TC>AG 0.036230277
TC>AT 0.080279992
TC>CA 0.003087341
TC>CG 0.001606148
TC>CT 0.005434085
TC>GA 0.002923779
TC>GG 0
TC>GT 0.009386977
TG>AA 0.010329744
TG>AC 0.004693489
TG>AT 0.050735432
TG>CA 0.003289512
TG>CC 0.00045208
TG>CT 0.010445568
TG>GA 0.002058227
TG>GC 0.001606148
TG>GT 0.098365261
TT>AA 0.005434085
TT>AC 0.00045208
TT>AG 0.000740597
TT>CA 0.00045208
TT>CC 0
TT>CG 0
TT>GA 0.000288517
TT>GC 0
TT>GG 0

Supplemental Table S4: Variant allele frequencies of SNSs and DNSs in cisplatin positive HCCs.

	SNSs		DNSs		CC>AA DNSs	
	median	sd	median	sd	median	sd
RK028	37.8%	11.5%	36.7%	11.2%	38.3%	11.3%
RK056	36.4%	12.9%	35.7%	12.6%	35.8%	10.7%
RK072	37.5%	17.4%	32.7%	17.5%	39.1%	18.3%
RK074	45.5%	17.1%	41.0%	16.7%	42.5%	17.7%
RK140	22.7%	8.2%	22.1%	9.1%	20.0%	6.6%
RK205	33.3%	16.7%	28.6%	13.3%	27.4%	21.2%
RK241	36.8%	12.0%	31.2%	10.3%	30.6%	11.9%
RK256	44.1%	12.8%	44.6%	12.4%	40.9%	10.5%