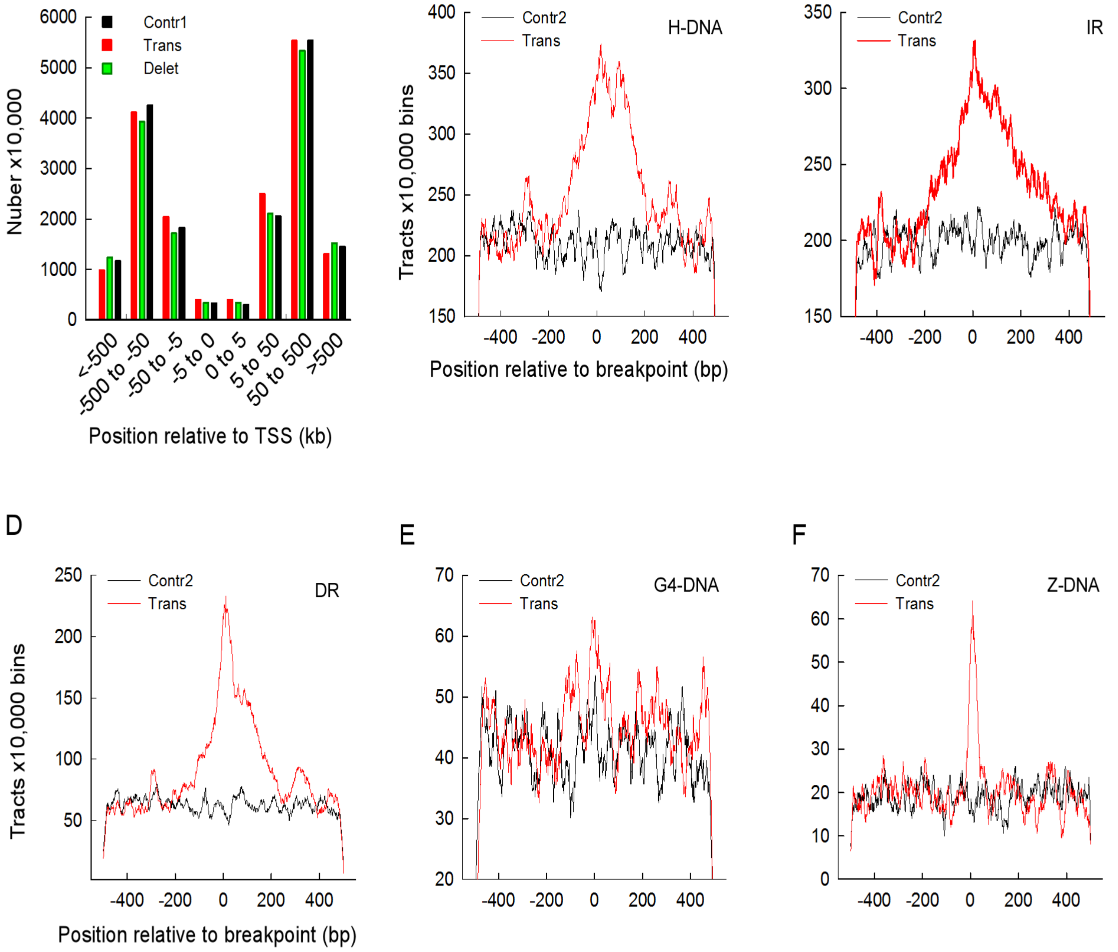
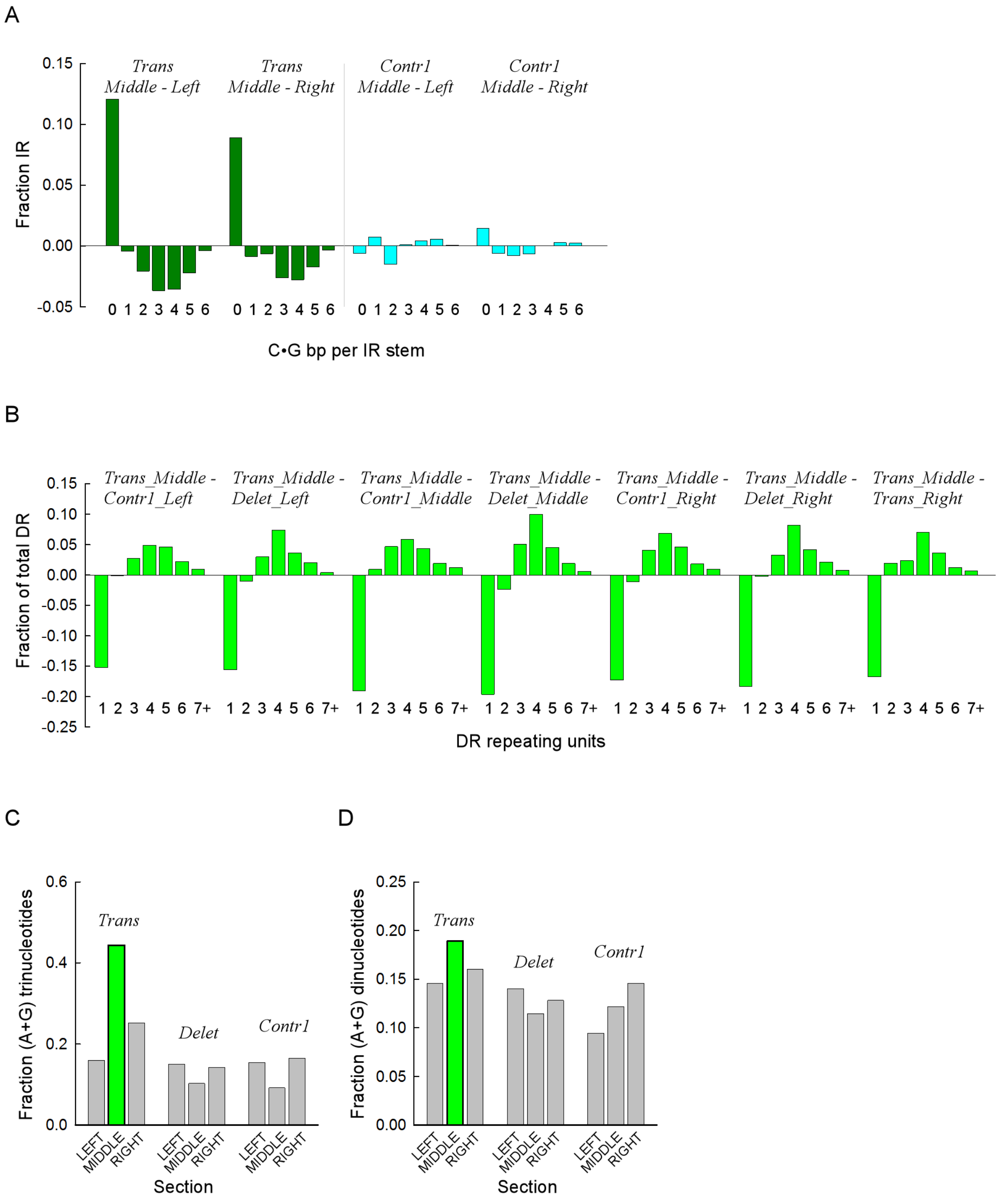
Translocation and Deletion Breakpoints in Cancer Genomes Are Associated with Potential Non-B DNA-Forming Sequences

Albino Bacolla, John A. Tainer, Karen M. Vasquez, and David N. Cooper

Supplementary Information



Supplementary Figure 1. Translocation sequence-dependent context does not account for the colocalization of non-B DNA-forming motifs at translocation breakpoints. (**A**) Distribution of breakpoints (number x 10,000 instances) with respect to transcription start sites (TSS) genome-wide for translocations (*red*), deletions (*green*) and Contr1 (*black*). (**B**) Number of DNA triplex-forming repeats (*H-DNA*) for 10,000 bins found near translocation (*red*), and Contr2 (*black*) breakpoints. (**C**) Same as in B, but for cruciform-forming inverted repeats (*IR*). (**D**) Same as in B, but for loop DNA-forming tandem repeats (*DR*). (**E**) Same as in B, but for quadruplex-forming repeats (*G4-DNA*). (**F**) Same as in B, but for left-handed DNA-forming repeats (*Z-DNA*). Numbers refer to the counts of bases belonging to each repeat type at every position; for H-DNA and IR, any bases separating a pair of repeats were excluded from the count.



Supplementary Figure 2. Repeat types enriched in the middle sections of translocation bins. (**A**) For each upstream (lowest genomic coordinate) IR sequence containing from 0 to 6 C|G bases, the fraction of the total number of IR found in the left, middle and right sections was computed for the translocation and Contr1 1-kb bins. The fractions obtained for the left and right sections were subtracted from those of the middle sections (*Middle – Left* and *Middle – Right*) for translocations (dark green, *Trans*) and Contr1 (cyan, *Contr*). Negative values indicate overrepresentation of IR sequences in left/right sections, whereas positive values indicate overrepresentation of IR sequences in middle sections. (**B**) For DR, the fractions of mono-, di-, tri-, tetra-, penta-, hexa-, and >hexa-nucleotides were computed separately for the translocation, deletion and Contr1 left, middle and right sections. Data for various sections were subtracted from those of the translocation middle section and plotted. Negative values indicate underrepresentation in the translocation middle section; positive values indicate overrepresentation in the translocation middle section. (**C**) and (**D**) For DR found in either the left, middle or right sections of translocation, deletion and Contr1 1-kb bins, the fraction of tri- (Panel C) and di-nucleotides (Panel D) whose strand sequence composition contained only purines (or pyrimidines, i.e. R•Y tracts) relative to all tri- (Panel C) or di-nucleotides (Panel D) in the respective section was computed and plotted. *Green bars* highlight the overrepresentation of R•Y-containing repeats in the middle section of translocations.

Supplementary Table 1.

```
A - Ranked p-values for the number of non-B DNA-forming repeats at translocation,
deletion and control breakpoint sites.

------------------------------------------------------------------------------------------
Critical value:        0.05           Number of tests:       20000
                       Corrected critical value:     0.0000025
------------------------------------------------------------------------------------------
Label                  P-values       Bonferroni     Bonferroni
                                      corrected      corrected
                                      significance   P-values
------------------------------------------------------------------------------------------

i_middle_C2_middle_T   6.93E-185      significant    1.3854E-180
i_middle_C1_middle_T   1.89E-183      significant    3.786E-179
i_left_T_middle_T      2.35E-176      significant    4.706E-172
i_middle_C2_middle_D   2.34E-154      significant    4.68E-150
i_middle_C1_middle_D   7.55E-152      significant    1.5092E-147
h_middle_C2_middle_T   2.43E-150      significant    4.856E-146
h_middle_C1_middle_T   7.05E-147      significant    1.4104E-142
h_middle_D_middle_T    9.60E-142      significant    1.9208E-137
h_left_T_middle_T      1.60E-139      significant    3.194E-135
d_middle_C1_middle_D   3.91E-120      significant    7.824E-116
h_right_T_middle_T     1.09E-116      significant    2.176E-112
q_middle_C1_middle_T   8.01E-112      significant    1.6012E-107
d_middle_C1_middle_T   3.06E-111      significant    6.126E-107
d_middle_C2_middle_T   2.09E-104      significant    4.188E-100
d_left_T_middle_T      3.87E-103      significant    7.746E-99
d_middle_C2_middle_D   1.31E-99       significant    2.626E-95
i_right_T_middle_T     2.27E-93       significant    4.532E-89
d_right_T_middle_T     2.27E-85       significant    4.54E-81
d_left_C1_left_D       2.06E-80       significant    4.116E-76
q_middle_D_middle_T    2.02E-78       significant    4.042E-74
i_right_D_middle_D     2.90E-75       significant    5.804E-71
i_left_D_middle_D      4.45E-71       significant    8.89E-67
```

d_right_D_middle_D	1.77E-70	significant	3.534E-66
q_right_C1_right_T	1.46E-68	significant	2.926E-64
d_middle_D_middle_T	3.32E-62	significant	6.63E-58
q_right_D_right_T	5.00E-54	significant	1.0006E-49
d_right_C1_right_T	2.77E-49	significant	5.544E-45
q_left_C1_left_T	1.66E-46	significant	3.328E-42
q_middle_C2_middle_T	1.97E-41	significant	3.938E-37
d_left_D_middle_D	1.17E-40	significant	2.34E-36
d_right_C2_right_T	2.13E-39	significant	4.254E-35
d_right_C1_right_D	6.04E-35	significant	1.2076E-30
h_left_D_middle_D	1.92E-32	significant	3.83E-28
z_middle_C2_middle_D	2.03E-32	significant	4.062E-28
i_left_C2_left_D	3.13E-32	significant	6.254E-28
q_left_T_middle_T	1.57E-31	significant	3.144E-27
q_middle_C1_middle_D	4.40E-31	significant	8.802E-27
h_right_D_right_T	6.63E-31	significant	1.3266E-26
i_right_C2_right_T	1.67E-30	significant	3.344E-26
d_left_D_left_T	2.57E-29	significant	5.146E-25
d_left_C2_left_D	2.83E-29	significant	5.656E-25
z_left_C1_left_T	6.02E-29	significant	1.2042E-24
i_right_C2_right_D	1.48E-28	significant	2.96E-24
z_right_C2_middle_C2	3.23E-28	significant	6.458E-24
h_right_C1_right_T	1.92E-27	significant	3.842E-23
h_left_D_left_T	6.76E-27	significant	1.351E-22
q_left_C2_left_D	5.38E-26	significant	1.076E-21
h_left_C1_left_T	3.14E-25	significant	6.288E-21
q_middle_C2_middle_D	1.94E-24	significant	3.888E-20
q_right_D_middle_D	1.47E-23	significant	2.944E-19
z_left_C1_middle_C	1.68E-23	significant	3.364E-19
i_left_C1_left_D	4.15E-23	significant	8.304E-19
h_right_D_middle_D	5.24E-23	significant	1.0478E-18
i_middle_D_middle_T	5.39E-23	significant	1.078E-18
h_left_C2_left_D	3.22E-22	significant	6.438E-18
d_left_T_right_T	4.08E-22	significant	8.154E-18
d_right_C2_right_D	5.13E-22	significant	1.0266E-17
z_left_C1_left_D	1.19E-20	significant	2.378E-16
h_right_C2_right_T	1.48E-19	significant	2.964E-15
i_left_D_left_T	4.65E-19	significant	9.306E-15
i_left_T_right_T	5.34E-19	significant	1.067E-14
q_left_D_left_T	6.25E-18	significant	1.2504E-13
z_middle_C2_middle_T	7.85E-18	significant	1.5704E-13
q_right_C2_right_T	4.83E-17	significant	9.658E-13
d_left_C2_right_C2	1.17E-16	significant	2.342E-12
z_left_D_middle_D	3.12E-16	significant	6.242E-12
q_right_C2_right_D	1.57E-15	significant	3.144E-11
z_right_T_middle_T	1.80E-15	significant	3.606E-11
d_left_D_right_D	3.83E-15	significant	7.666E-11
h_middle_C1_middle_D	4.88E-15	significant	9.754E-11
z_right_C2_right_T	5.86E-15	significant	1.1724E-10
q_left_C1_left_D	1.91E-14	significant	3.814E-10
h_left_C1_middle_C	2.60E-14	significant	5.198E-10
d_left_C1_left_T	1.64E-13	significant	3.282E-09
i_right_C1_right_T	1.27E-12	significant	2.53E-08
d_right_D_right_T	2.76E-12	significant	5.52E-08
z_left_C1_right_C	9.93E-12	significant	1.985E-07
q_left_D_middle_D	1.08E-11	significant	2.158E-07
z_left_C2_right_C2	4.51E-11	significant	9.028E-07
q_right_T_middle_T	7.14E-11	significant	1.4282E-06
z_left_T_right_T	1.54E-10	significant	0.000003086
z_right_D_right_T	2.06E-10	significant	0.000004128
z_middle_C1_middle_T	5.54E-10	significant	0.000011086
i_right_C1_right_D	1.00E-09	significant	0.00002
i_right_C2_middle_C2	3.82E-09	significant	0.00007648
h_middle_C2_middle_D	6.82E-09	significant	0.0001364
i_left_C2_middle_C2	7.30E-09	significant	0.00014604
d_left_C2_middle_C2	1.30E-08	significant	0.0002594
z_left_T_middle_T	6.06E-08	significant	0.0012126
q_left_C1_middle_C	7.50E-08	significant	0.0015006
q_left_T_right_T	1.21E-07	significant	0.002418
z_left_D_right_D	1.11E-06	not significant	0.02228
z_left_C2_left_T	1.57E-06	not significant	0.03144
z_middle_D_middle_T	1.58E-06	not significant	0.03164
q_right_C2_middle_C2	1.69E-06	not significant	0.03388

```
z_left_D_left_T       3.36E-06       not significant       0.06718
z_middle_C1_middle_D  3.67E-06       not significant       0.07346
z_right_D_middle_D    3.82E-06       not significant       0.07632
z_left_C2_middle_C2   5.27E-06       not significant       0.10538
z_right_C1_right_D    5.68E-06       not significant       0.11366
h_right_C1_middle_C   1.43E-05       not significant       0.2852
h_left_C2_right_C2    1.83E-05       not significant       0.3652
q_left_C2_right_C2    2.81E-05       not significant       0.5612
q_right_C1_right_D    4.29E-05       not significant       0.8586
h_left_T_right_T      1.11E-04       not significant       1
i_left_C1_middle_C    5.54E-04       not significant       1
q_right_C1_middle_C   7.88E-04       not significant       1
z_right_C1_middle_C   3.50E-03       not significant       1
i_left_C1_right_C     5.44E-03       not significant       1
z_right_C1_right_T    1.32E-02       not significant       1
q_left_D_right_D      3.63E-02       not significant       1
q_left_C1_right_C     6.12E-02       not significant       1
h_left_C1_right_C     6.40E-02       not significant       1
h_left_D_right_D      7.45E-02       not significant       1
i_right_D_right_T     9.26E-02       not significant       1
h_left_C1_left_D      3.13E-01       not significant       1
h_right_C1_right_D    3.70E-01       not significant       1
i_left_C1_left_T      3.97E-01       not significant       1
i_left_D_right_D      4.41E-01       not significant       1
d_left_C1_middle_C    5.49E-01       not significant       1
i_right_C1_middle_C   6.59E-01       not significant       1
d_left_C1_right_C     7.34E-01       not significant       1
d_right_C1_middle_C   8.06E-01       not significant       1
h_right_C2_right_D    1.75E-04       not significant       1
d_right_C2_middle_C2  7.31E-04       not significant       1
h_left_C2_middle_C2   9.12E-04       not significant       1
i_left_C2_left_T      1.05E-03       not significant       1
z_right_C2_right_D    1.34E-03       not significant       1
d_left_C2_left_T      3.16E-02       not significant       1
h_left_C2_left_T      3.24E-02       not significant       1
h_right_C2_middle_C2  7.60E-02       not significant       1
z_left_C2_left_D      3.32E-01       not significant       1
q_left_C2_left_T      4.09E-01       not significant       1
i_left_C2_right_C2    7.26E-01       not significant       1
q_left_C2_middle_C2   9.80E-01       not significant       1

-----------------------------------------------------------------------------------------
d, DR; h, H-DNA; i, IR; q, G4-DNA; z, Z-DNA; C1, Contr1; C2, Contr2; D, deletions; T,
Translocations
-----------------------------------------------------------------------------------------
```

B - Ranked p-values for the number for GC contents at translocation, deletion, and control breakpoint sites.

```
-----------------------------------------------------------------------------------------
Critical value:       0.05           Number of tests:      20000
                      Corrected critical value:     0.0000025
-----------------------------------------------------------------------------------------
Label                 P-values       Bonferroni     Bonferroni
                                     corrected      corrected
                                     significance   P-values
-----------------------------------------------------------------------------------------
gc_T_right_t_  c      5.1743E-143    significant    1.0349E-138
gc_I_middle_t_c       4.0299E-137    significant    8.0597E-133
gc_T_left_t_c         1.3076E-134    significant    2.6152E-130
gc_D_middle_t_c       2.131E-103     significant    4.2621E-99
gc_T_middle_t_c       1.01675E-93    significant    2.03351E-89
gc_H_middle_t_d       4.98331E-82    significant    9.96662E-78
gc_I_middle_d_c       1.33301E-79    significant    2.66603E-75
gc_I_middle_t_d       4.0778E-57     significant    8.15561E-53
gc_D_left_d_c         2.6158E-47     significant    5.2316E-43
gc_D_middle_t_d       3.52376E-42    significant    7.04752E-38
gc_D_ left_t_d        2.33256E-39    significant    4.66512E-35
gc_H_middle_d_c       1.7776E-32     significant    3.5552E-28
gc_I_ left_t_d        3.85601E-28    significant    7.71202E-24
gc_I_right_d_c        2.63271E-20    significant    5.26542E-16
gc_H_middle_t_c       4.09102E-19    significant    8.18205E-15
gc_I_left_d_c         3.81487E-16    significant    7.62974E-12
```

```
gc_D_right_t_d         1.54356E-08    significant    0.000308712
gc_H_left_d_c          5.35458E-08    significant    0.001070916
gc_I_right_t_c         1.40841E-07    significant    0.002816819
gc_D_right_d_c         4.18205E-07    significant    0.008364098
gc_H_left_t_c          2.00076E-06    significant    0.040015296
gc_H_right_d_c         0.000245314    not significant       1
gc_H_right_t_d         0.000282709    not significant       1
gc_I_left_t_c          0.000486602    not significant       1
gc_I_right_t_d         0.006738116    not significant       1
gc_D_middle_d_c        0.174775767    not significant       1
gc_H_right_t_c         0.649415628    not significant       1
gc_D_left_t_c          0.672781808    not significant       1
gc_D_right_t_c         0.758824601    not significant       1
gc_H_ left_t_d         0.851981696    not significant       1
---------------------------------------------------------------------------------------------
T, whole COSMIC dataset; I, IR; D, DR; H, H-DNA; t, translocations; d, deletions; c,
controls;
left, left section; middle, middle section; right, right section
```

Supplementary Table 2. Fractions of H-DNA repeats containing various numbers of C•G base pairs in the stem

Section	Number of C•G base-pairs							Total
	0	1	2	3	4	5	6	
Left_T	0.3462	0.2055	0.1455	0.1153	0.0891	0.0605	0.0258	0.9878
Middle_T	0.2910	0.1908	0.1503	0.1311	0.1010	0.0624	0.0292	0.9558
Right_T	0.3385	0.1926	0.1385	0.1267	0.0927	0.0682	0.0269	0.9840
Left_D	0.3467	0.2140	0.1375	0.1129	0.0935	0.0591	0.0220	0.9858
Middle_D	0.3717	0.2148	0.1321	0.1085	0.0855	0.0527	0.0197	0.9850
Right_D	0.3397	0.2144	0.1419	0.1206	0.0877	0.0592	0.0234	0.9870
Left_C1	0.3228	0.2191	0.1551	0.1156	0.0918	0.0600	0.0223	0.9867
Middle_C1	0.3301	0.2162	0.1423	0.1188	0.0886	0.0651	0.0275	0.9886
Right_C1	0.3298	0.2151	0.1448	0.1142	0.0924	0.0668	0.0226	0.9857

Target	P-values (1-sample Student's t-test: Target vs. other 8 sections)						
Middle_T	3.3E-05	2.4E-04	1.2E-02	1.4E-04	1.3E-05	6.3E-01	6.9E-04
Middle_D	3.5E-04	1.5E-01	5.8E-04	2.0E-03	3.0E-03	1.0E-04	9.7E-04

T, translocations; *D*, deletions; *C1*, Contr1