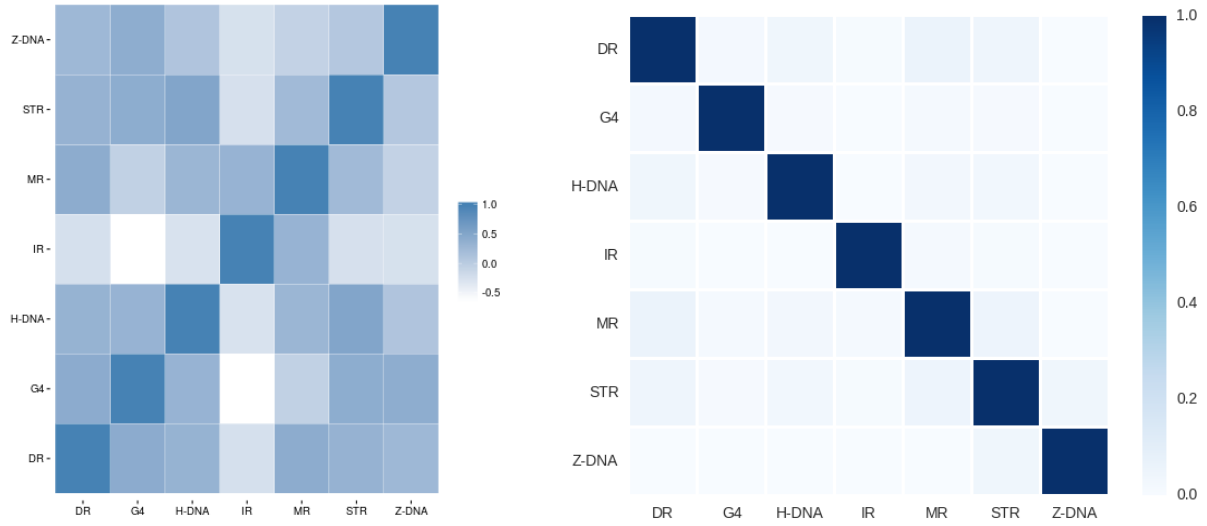
# Supplemental Figures




a. b.

**Fig. S1:** *a.* Correlation between different non-B DNA motifs in the human genome (Pearson correlation, genomic bins of 500kb). *b.* Jaccard Index heatmap reporting the amount of overlap between different non-B DNA motifs. 1 represents complete overlap, whereas 0 represents no overlap.

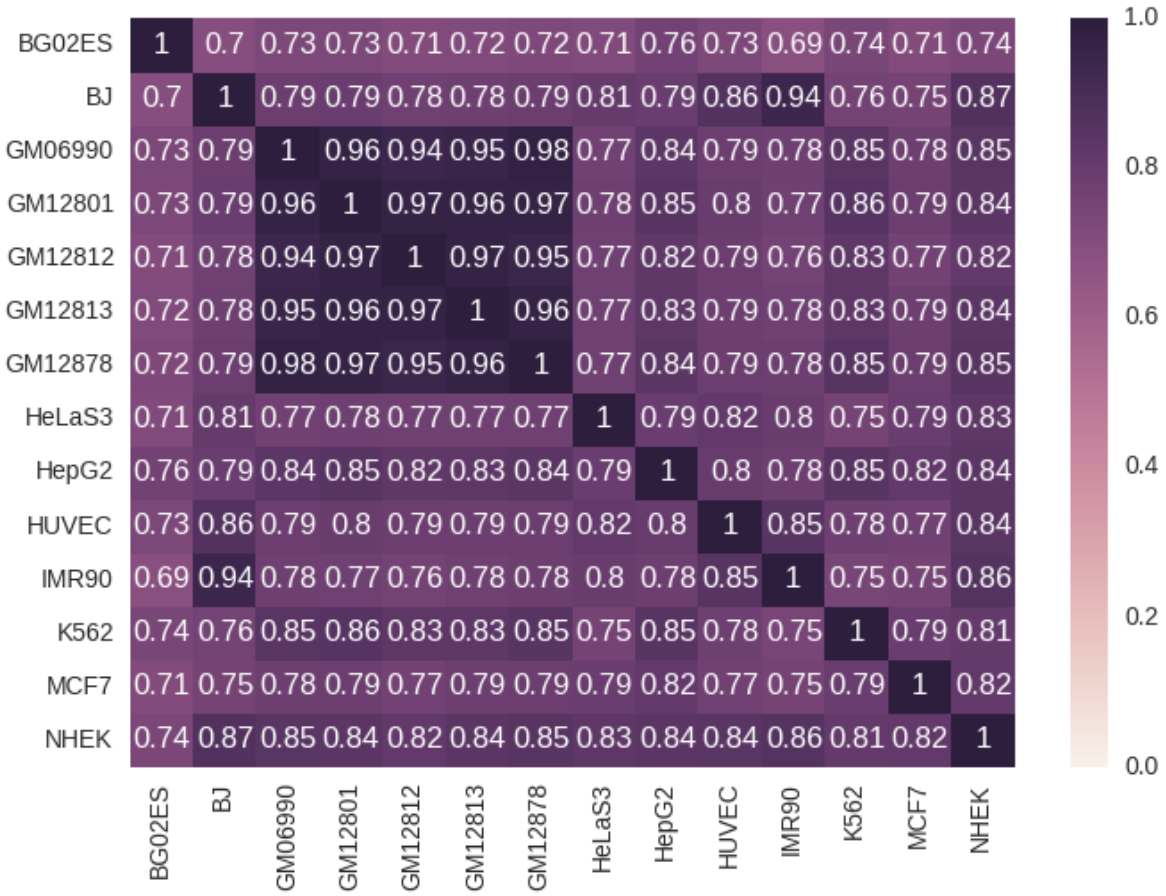



**Fig. S2:** Correlations between replication timing domains of different cell lines (Pearson, 500kb genome windows).

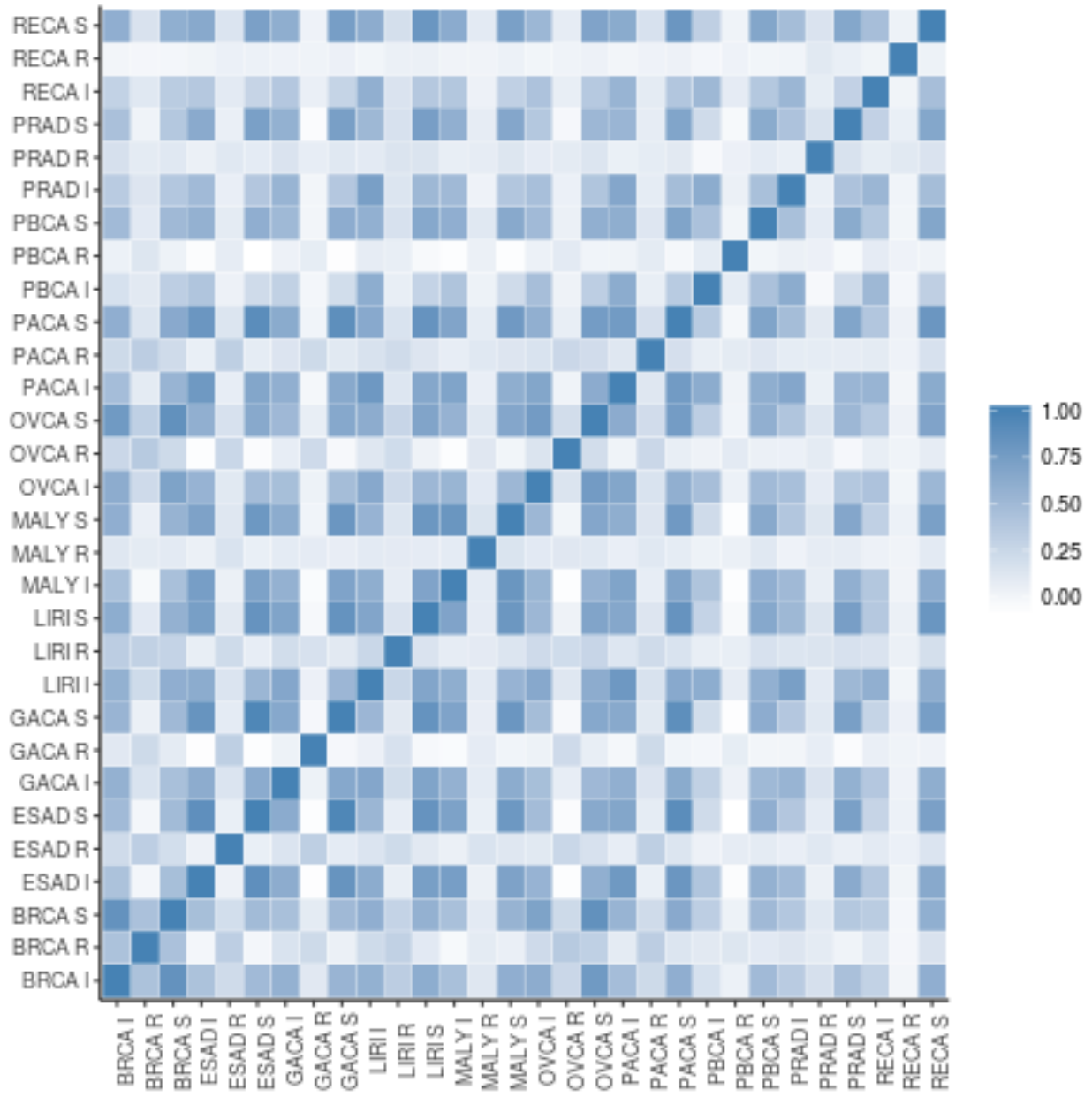
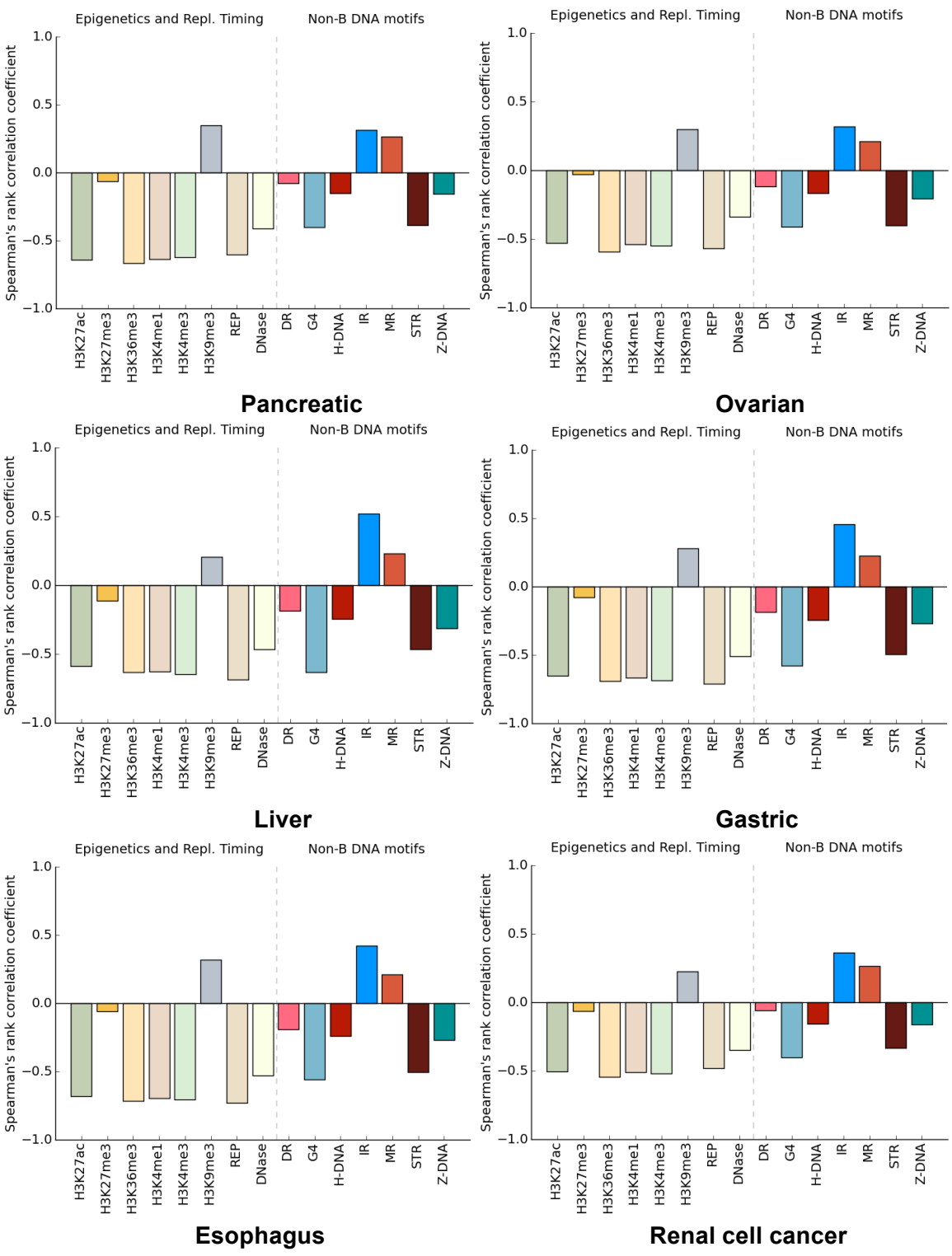
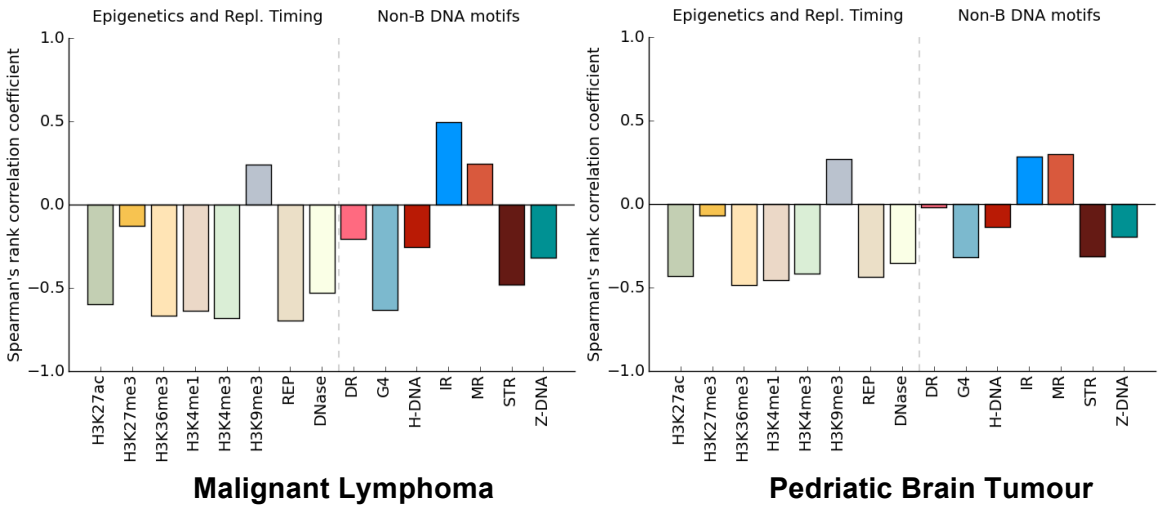


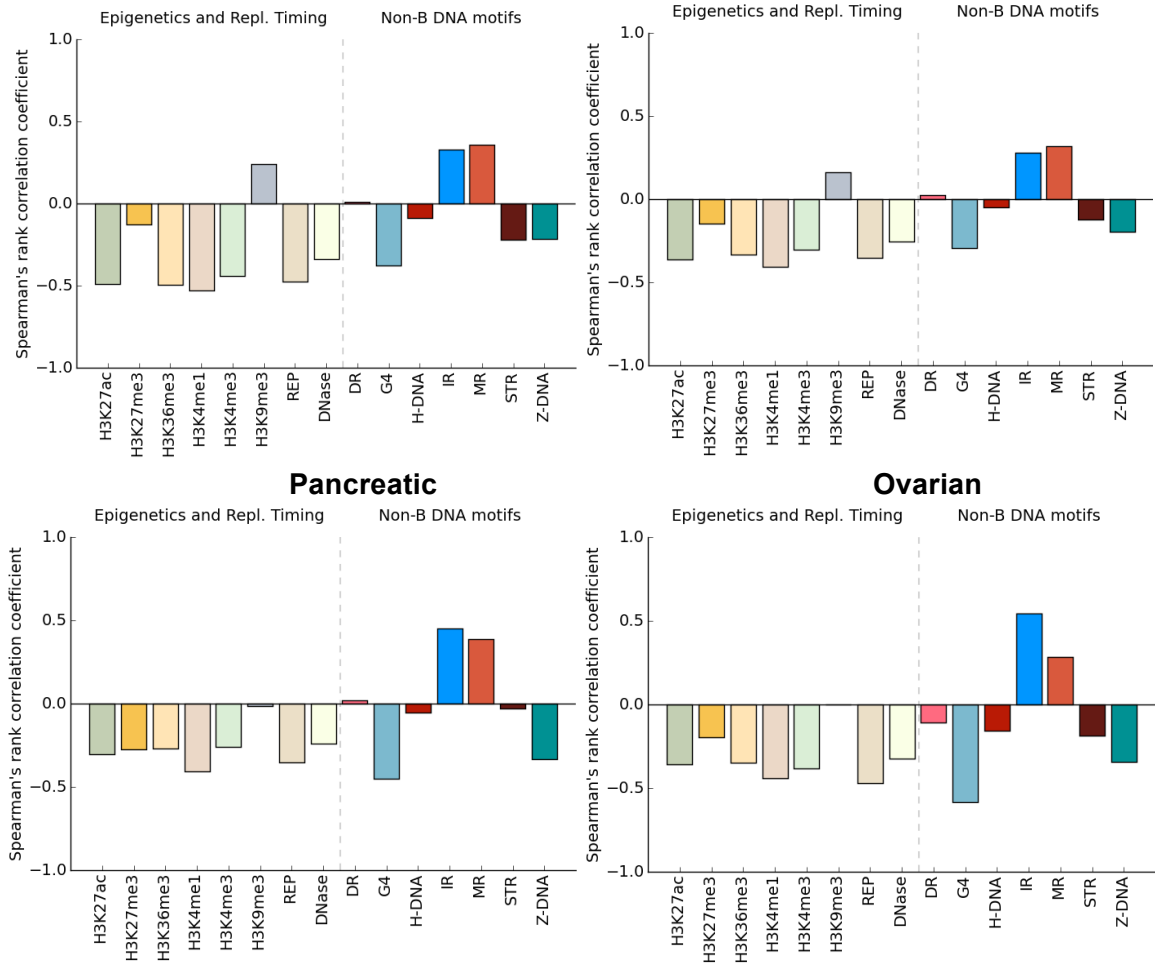

**Fig. S3:** Pearson correlation between the number of substitutions (S), indels (I) and rearrangements (R) found in non-overlapping 500 kb bins across the ten tumour types.

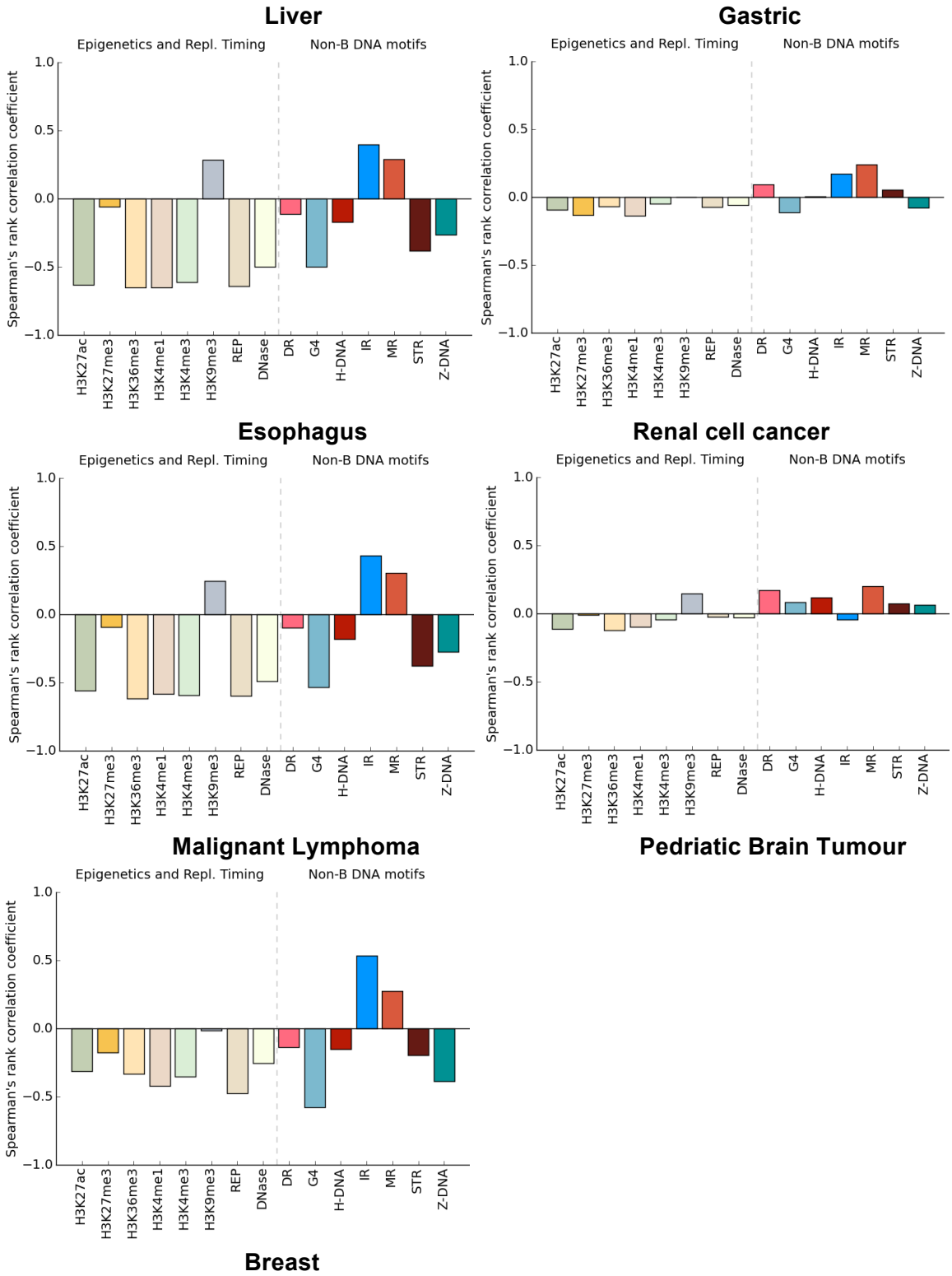
## A. SUBSTITUTIONS



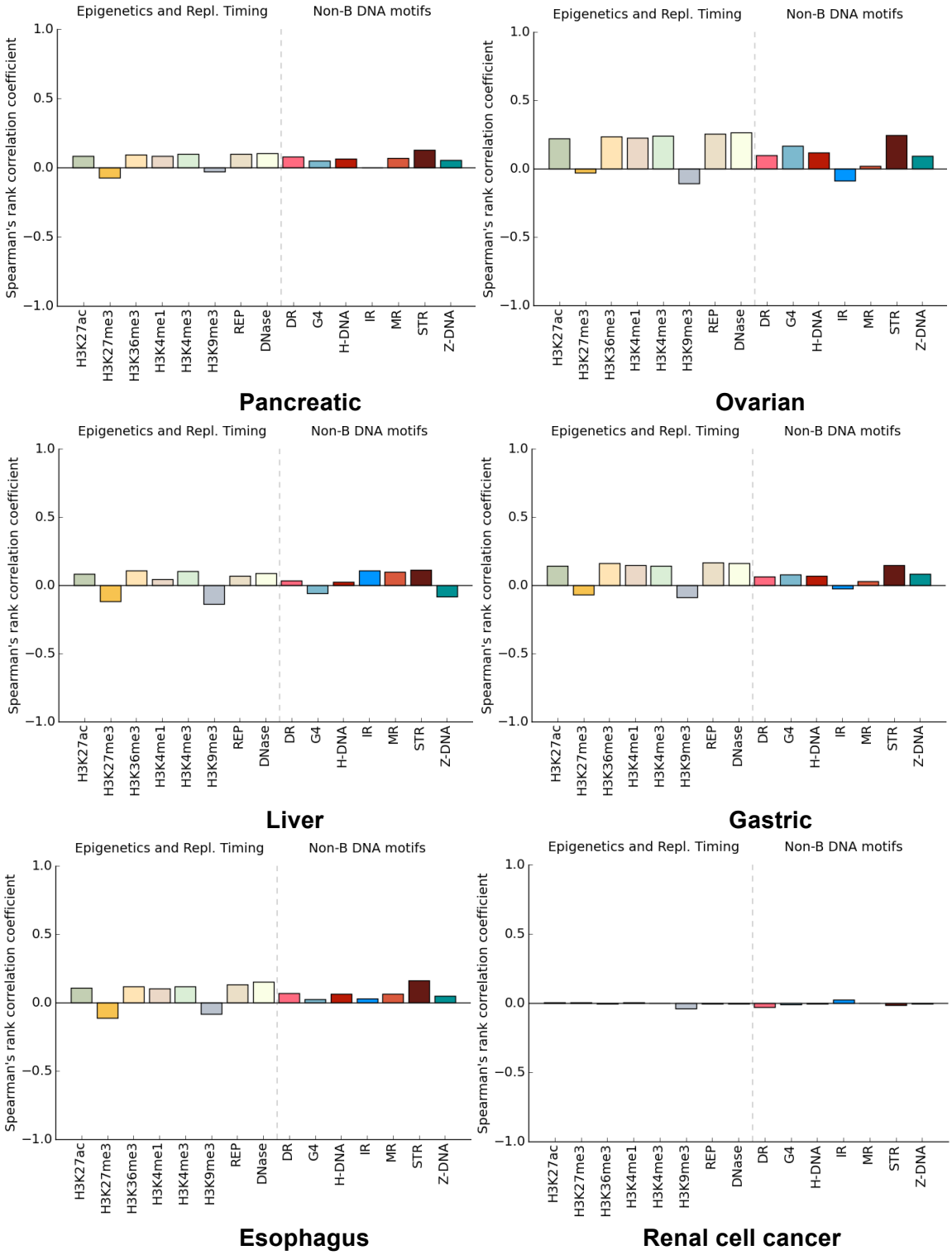

Epigenetics and Repl. Timing
Non-B DNA motifs
Spearman's rank correlation coefficient
1.0
0.5
0.0
−0.5
−1.0
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me3
REP
DNase
DR
G4
H-DNA
IR
MR
STR
Z-DNA
Malignant Lymphoma
Pedriatic Brain Tumour

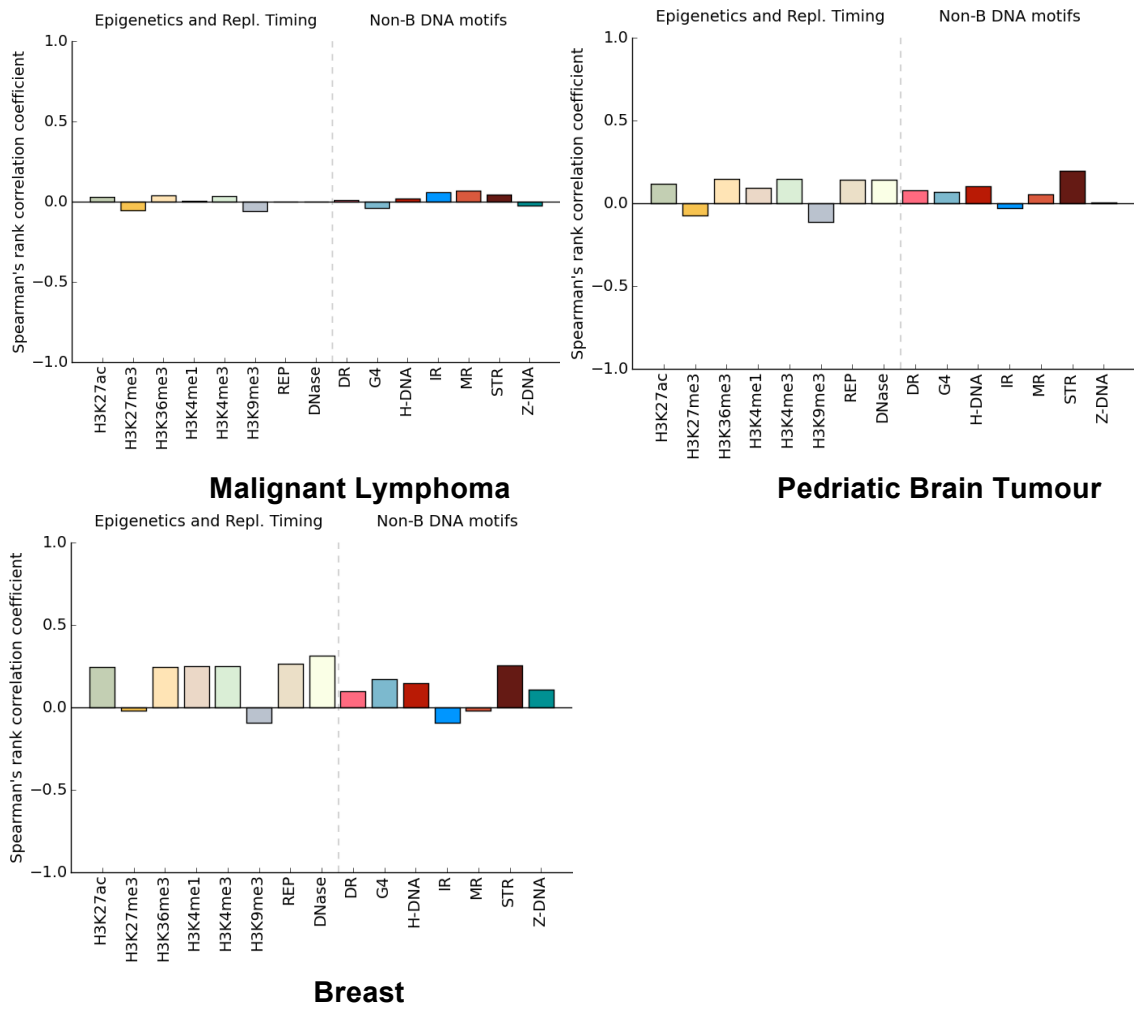
## B. INDELS


Epigenetics and Repl. Timing
Non-B DNA motifs
Spearman's rank correlation coefficient
1.0
0.5
0.0
−0.5
−1.0
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me3
REP
DNase
DR
G4
H-DNA
IR
MR
STR
Z-DNA
Pancreatic
Ovarian


Liver
Epigenetics and Repl. Timing
Non-B DNA motifs
Spearman's rank correlation coefficient
1.0
0.5
0.0
−0.5
−1.0
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me3
H3K9me3
REP
DNase
DR
G4
H-DNA
IR
MR
STR
Z-DNA
Gastric
Esophagus
Renal cell cancer
Malignant Lymphoma
Pedriatic Brain Tumour
Breast

## C. REARRANGEMENTS






c.

**Fig. S4: Spearman's rank correlation coefficient between non-B motifs and epigenetic markers/replication timing for:** *a.* substitutions, *b.* indels, *c.* rearrangements across multiple tumour types. *For prostate cancer, epigenetic marks for a cell of origin were not available and was excluded.

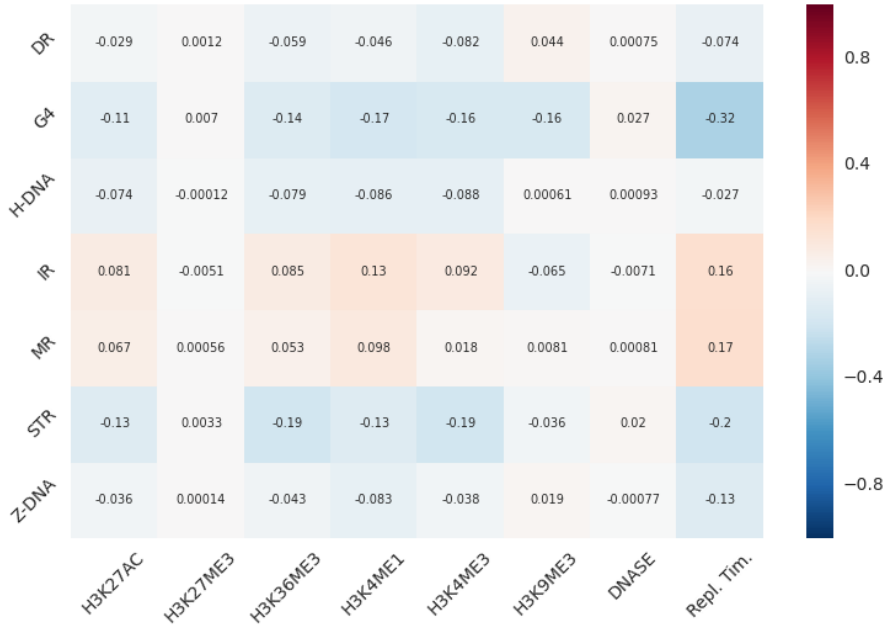



**Fig. S5: Results following partial correlation analyses.** Remaining correlations (Pearson - partial) for each non-B motif (y-axis) when controlling for epigenetic features and replication timing (x-axis).

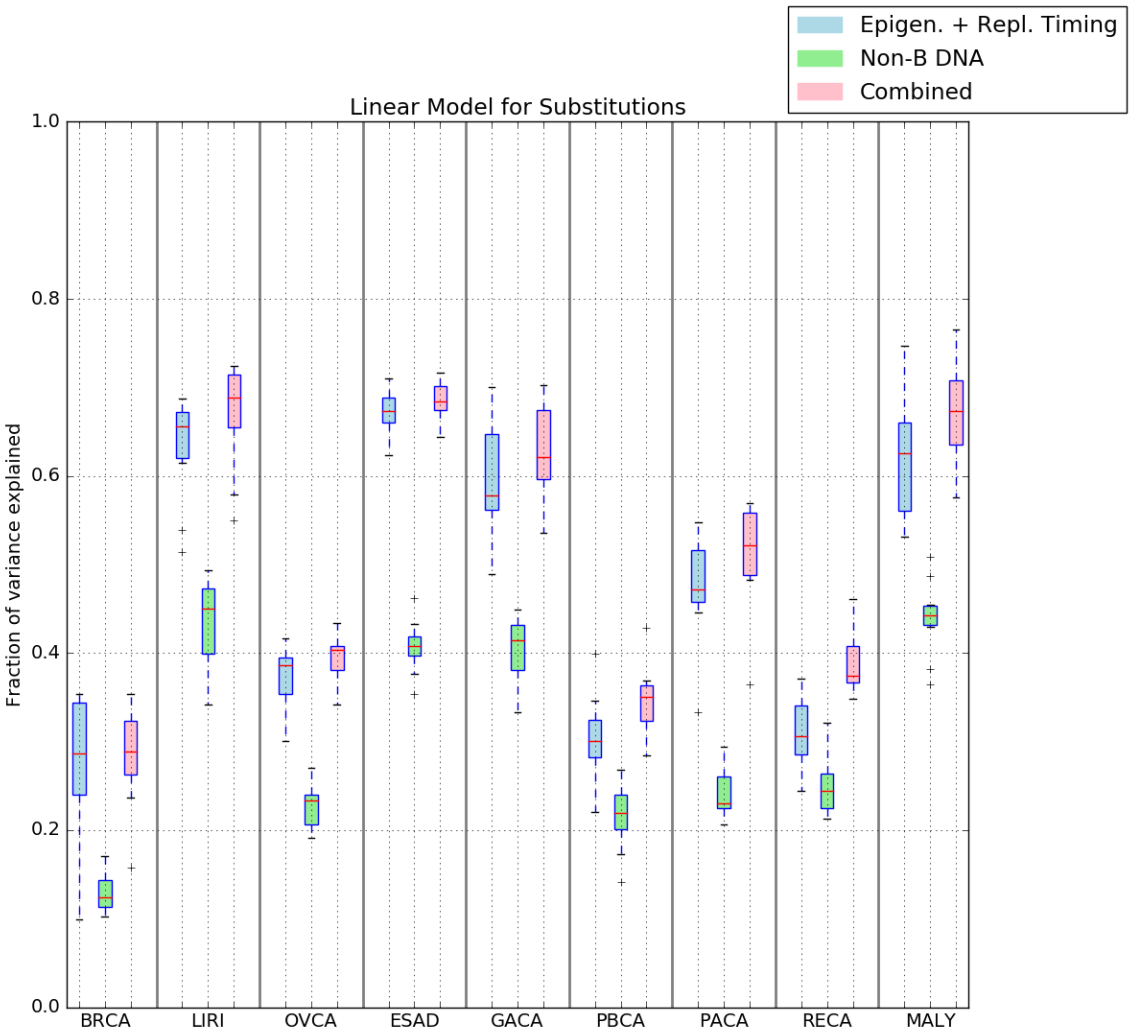
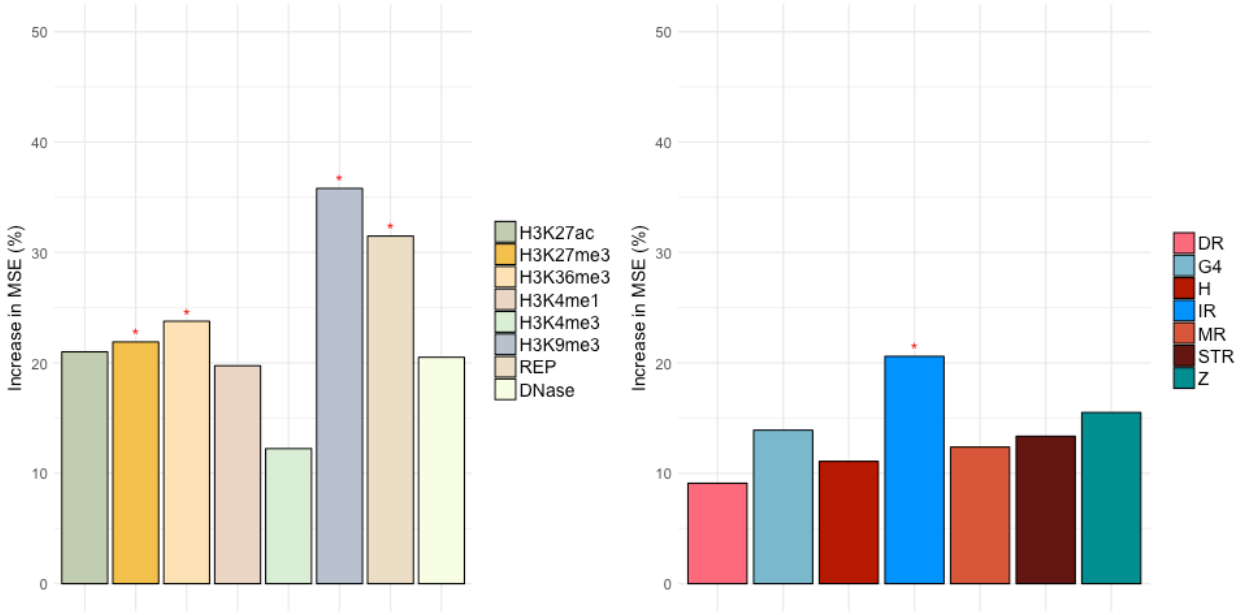
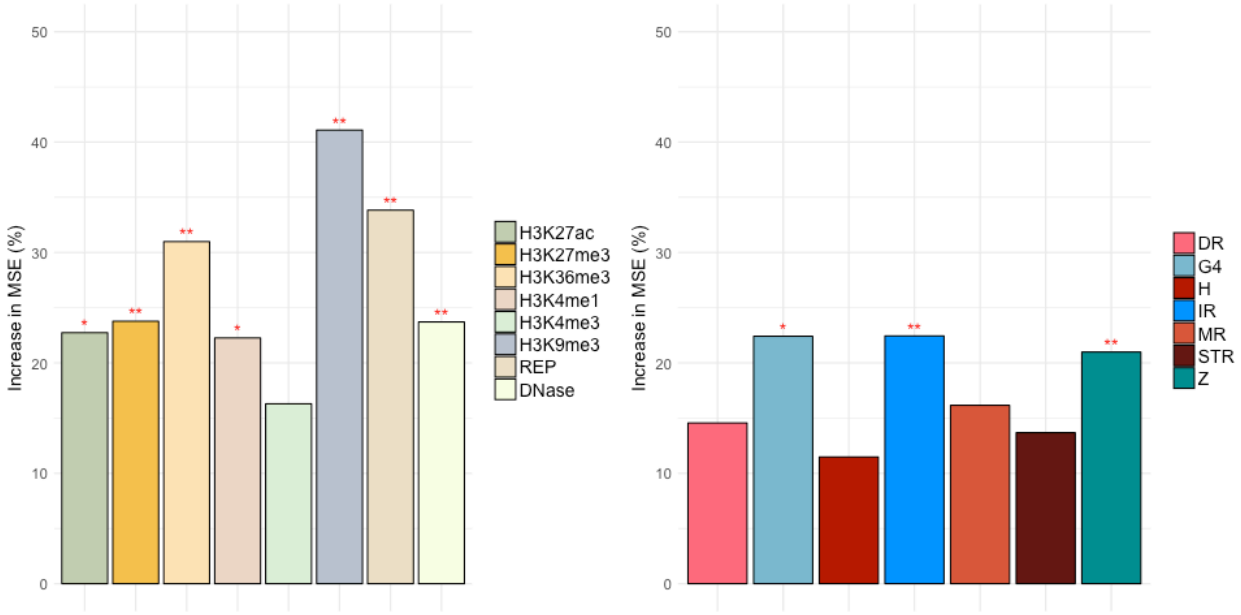



**Fig. S6: Linear model predicts mutability of base substitutions.** Results depicted across nine tumour types using the following factors: non-B DNA motifs (green), epigenetics/replication timing (blue), or a combination of both (pink).

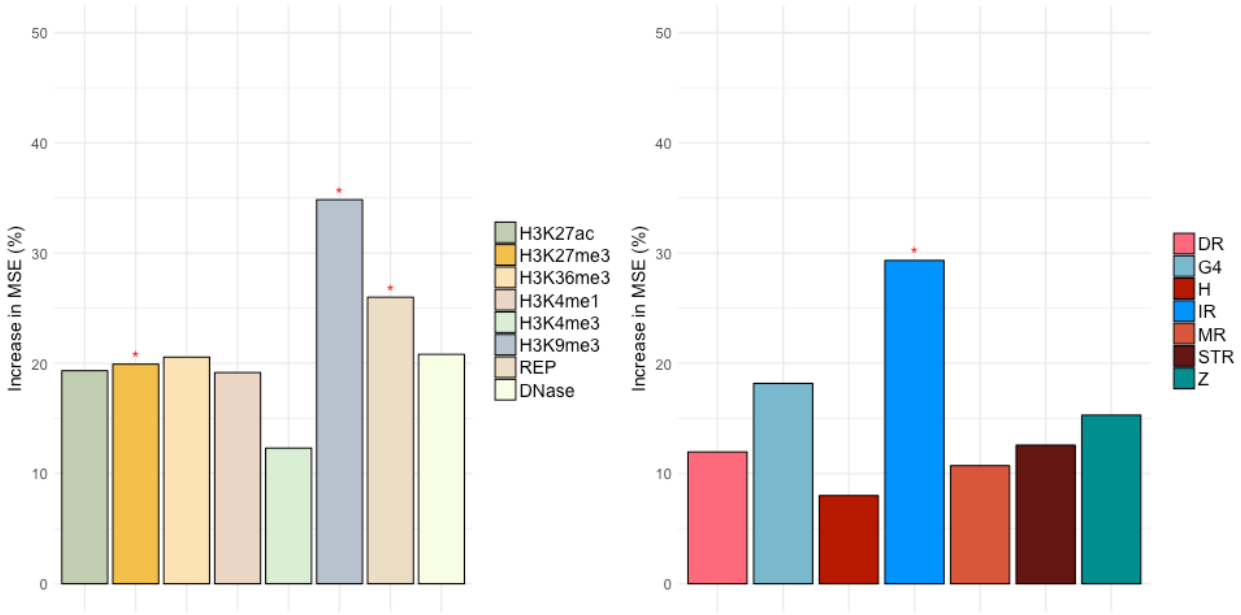
## A. SUBSTITUTIONS



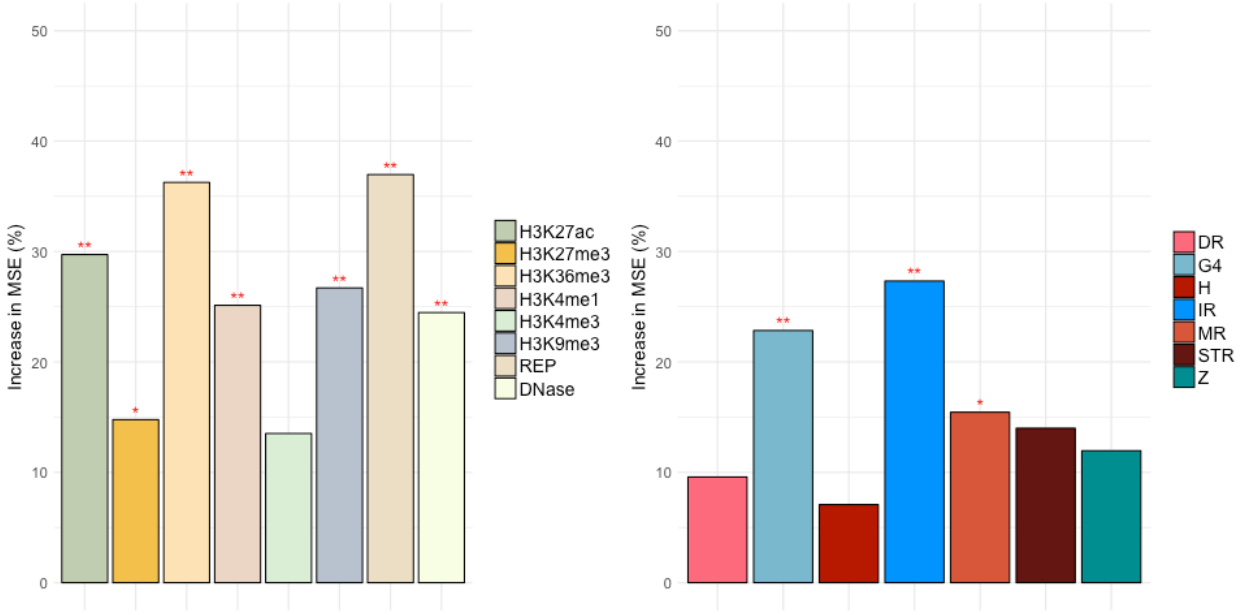

**Ovarian**



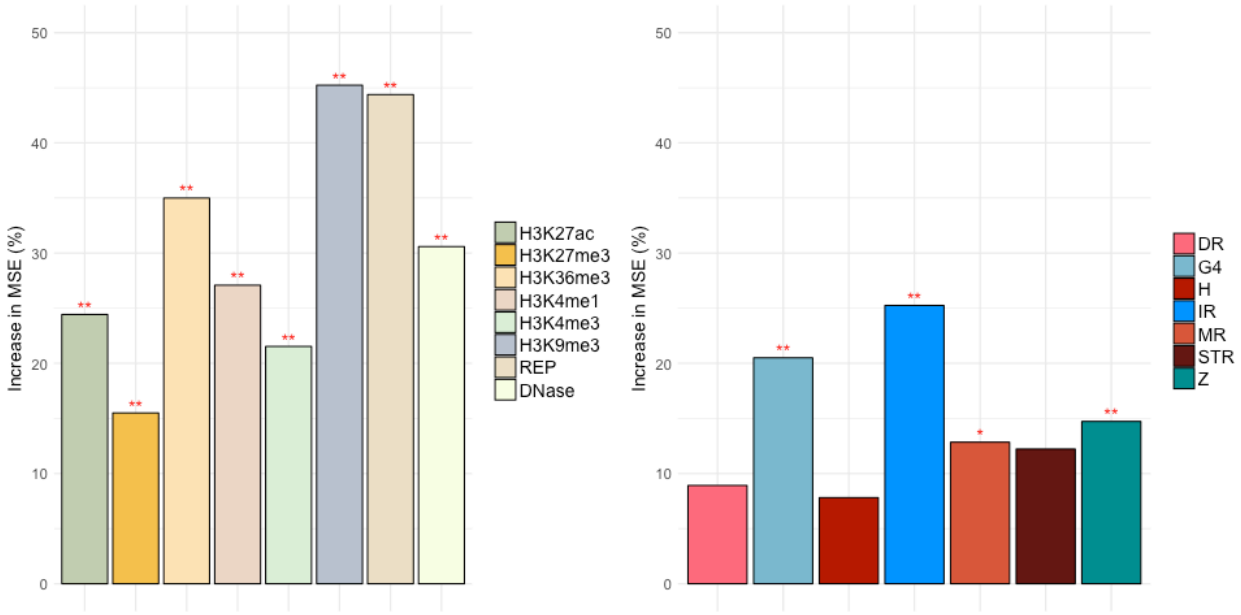

**Pancreatic**



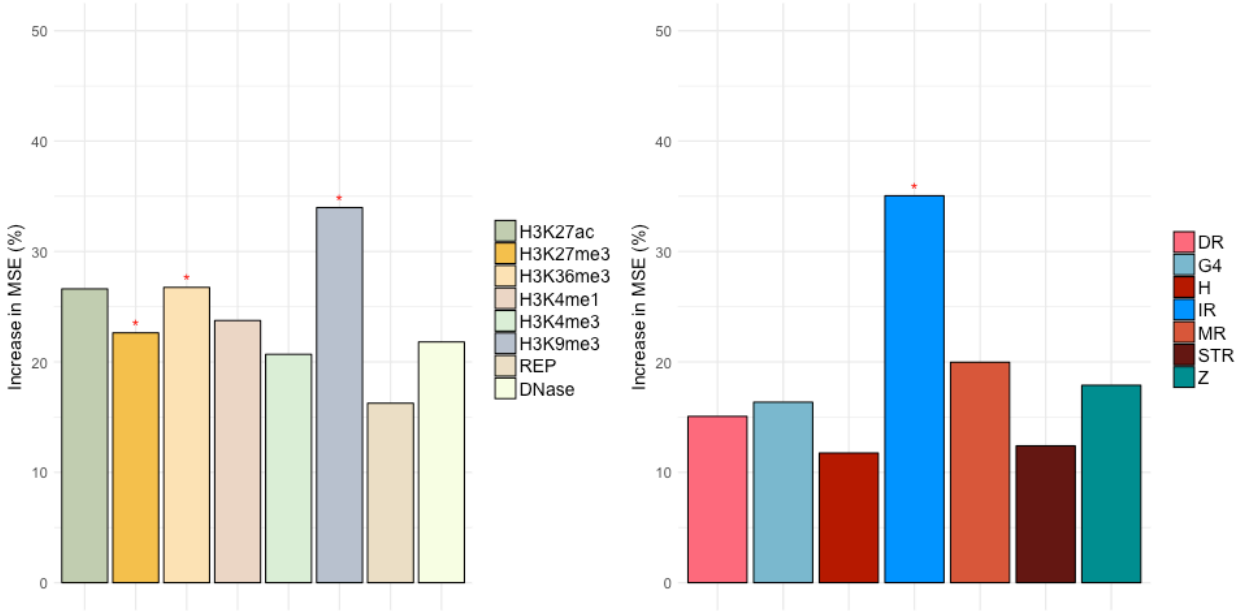

**Liver**



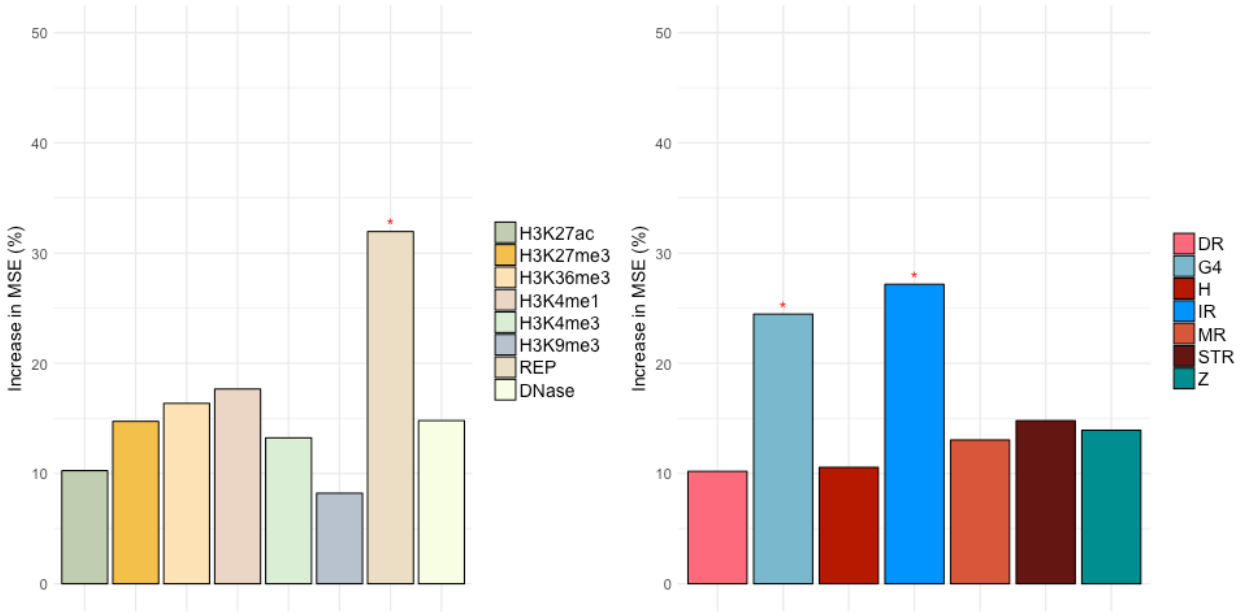

**Gastric**



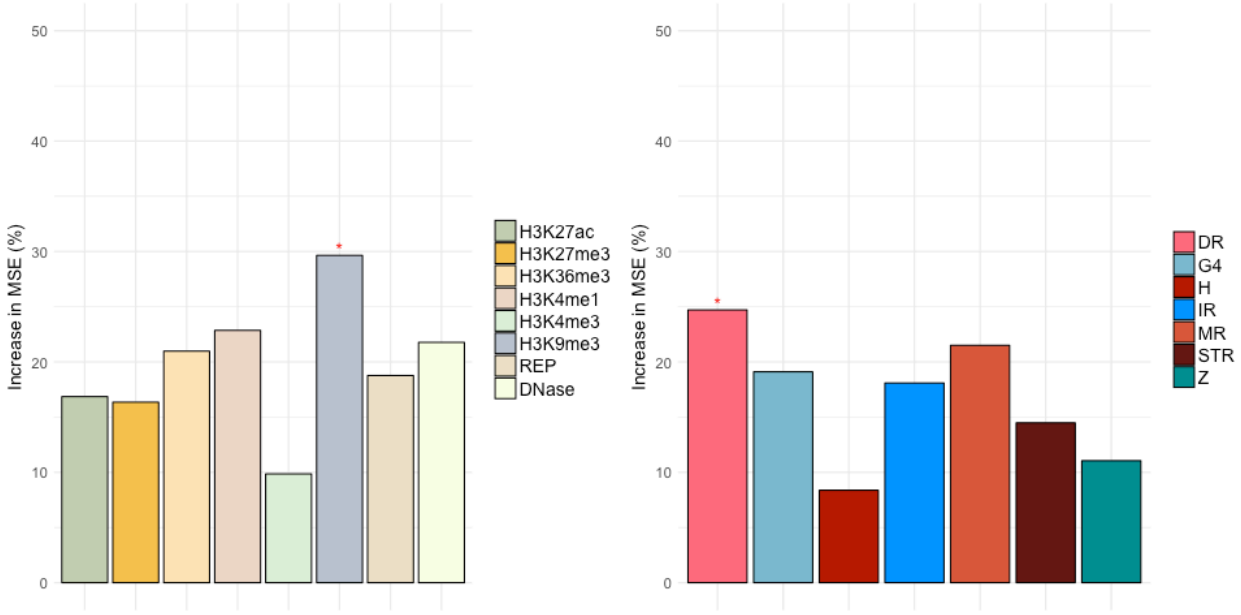

**Esophagus**




**Renal cell cancer**

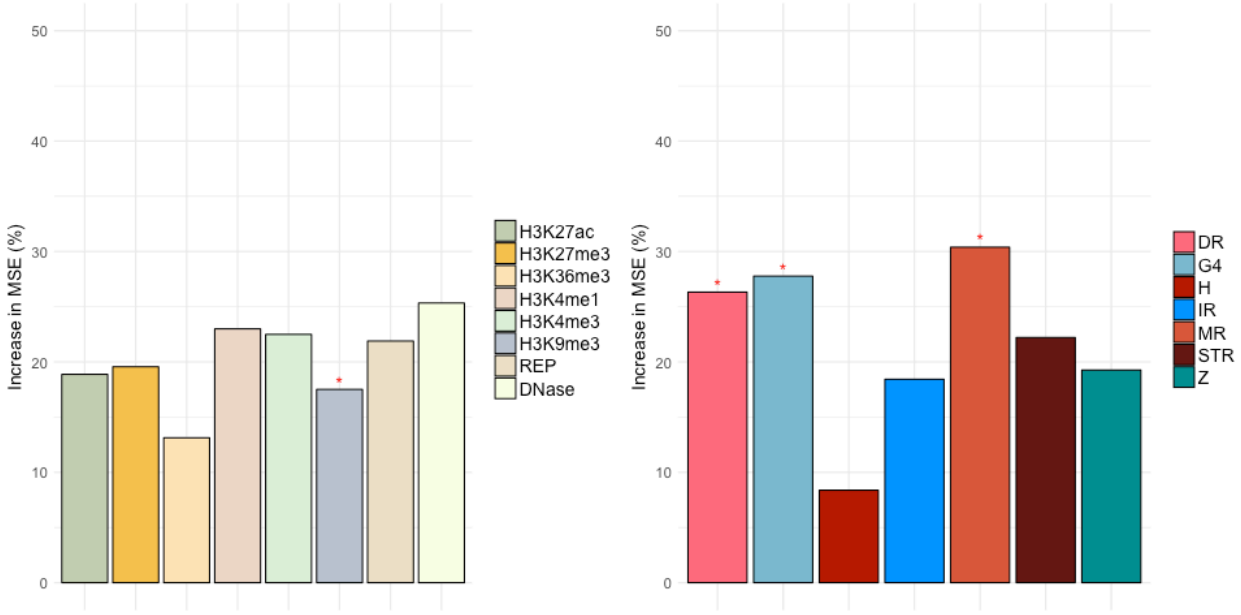



**Malignant Lymphoma**

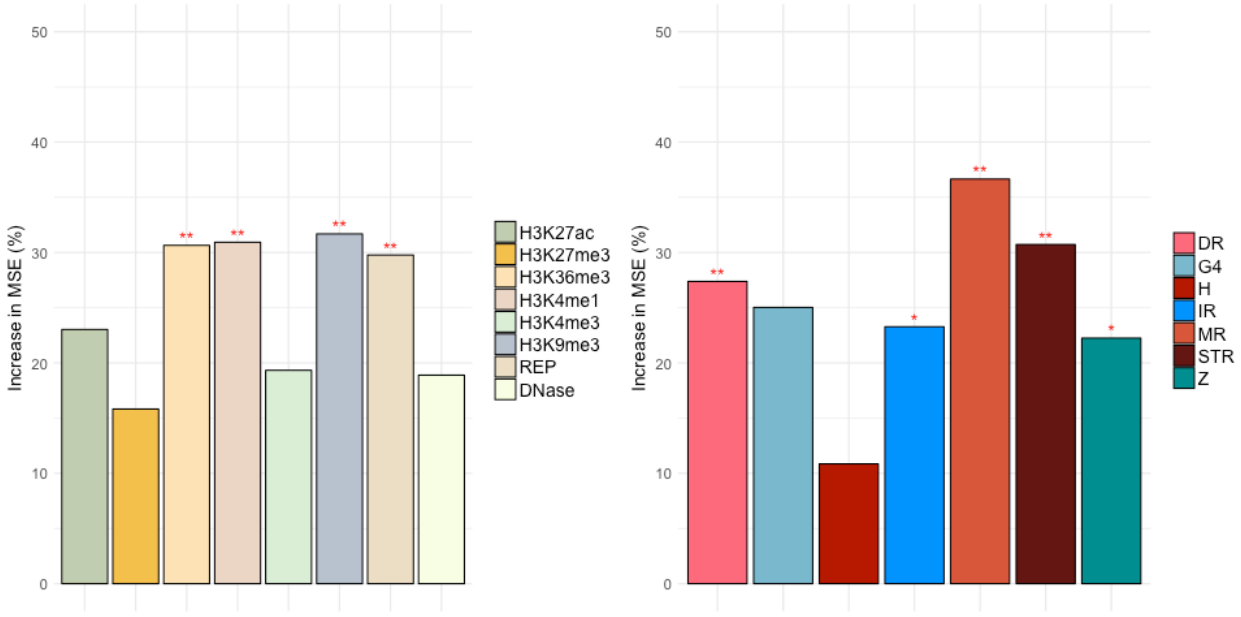



**Pedriatic Brain Tumour**

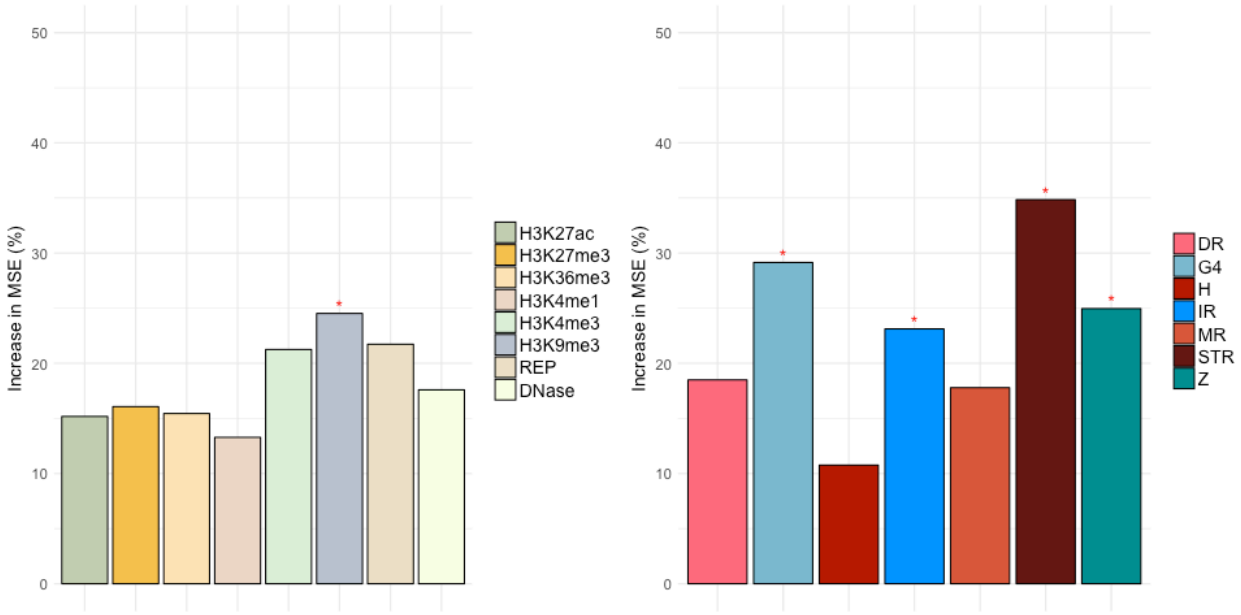
**B. INDELS**



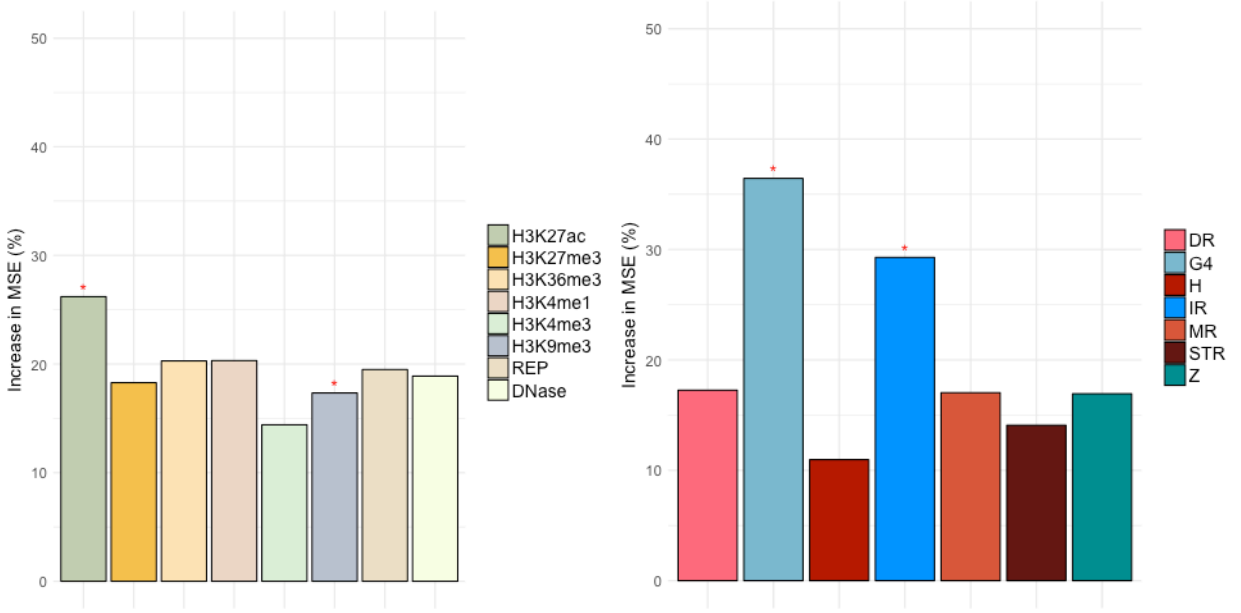

**Ovarian**



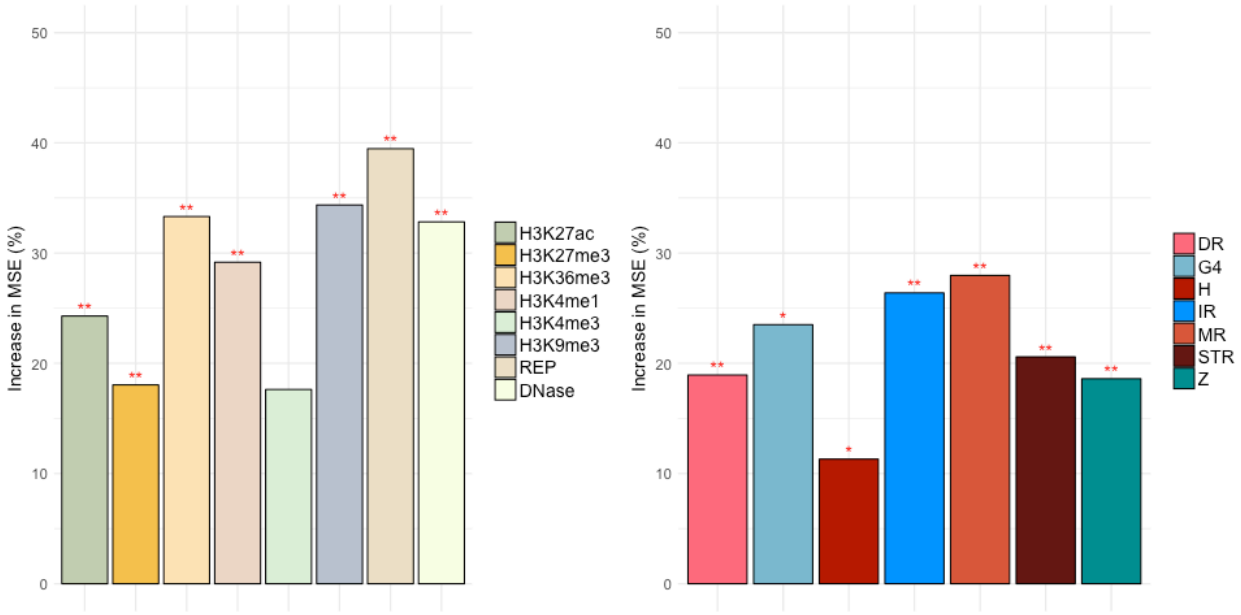

**Pancreatic**



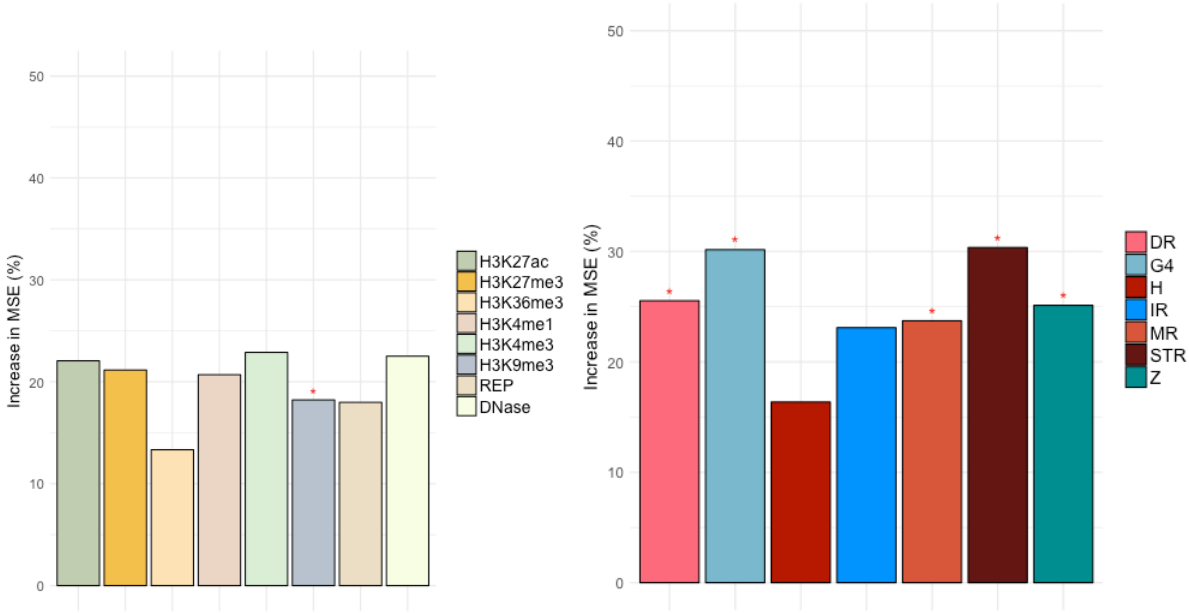

**Liver**



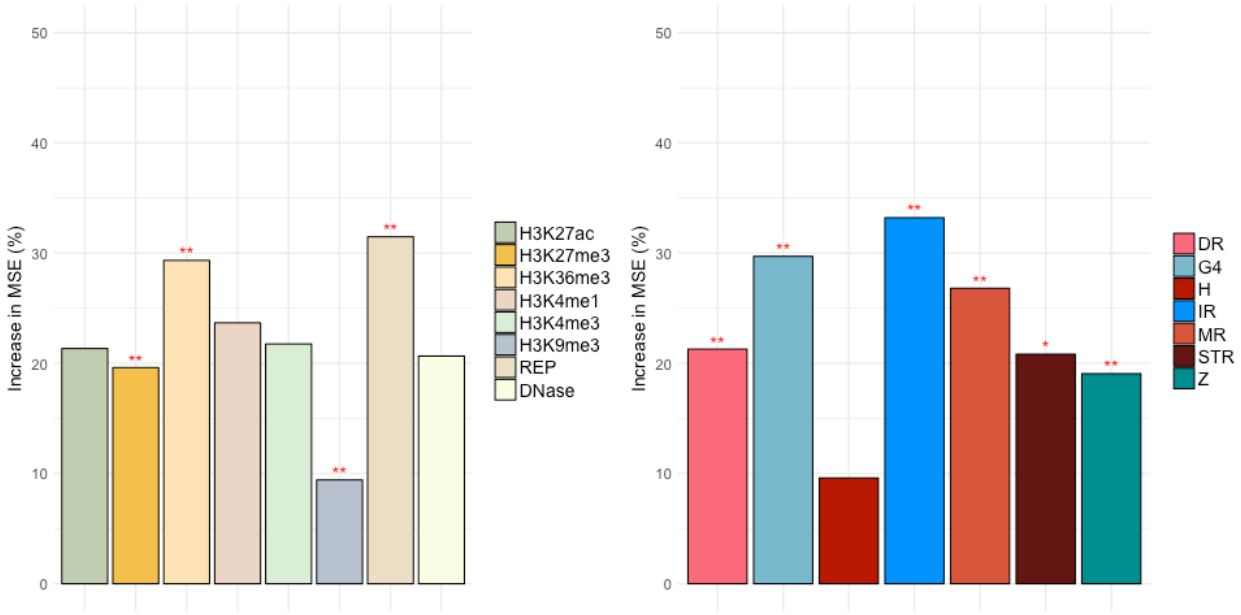

**Gastric**



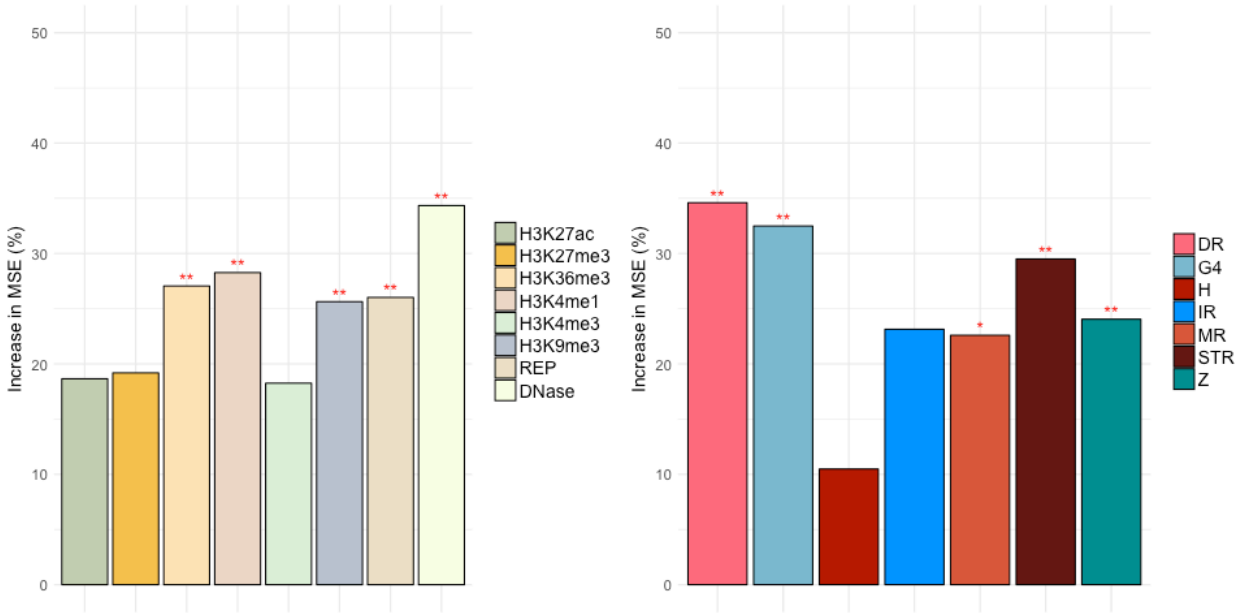

**Esophagus**




**Renal cell cancer**

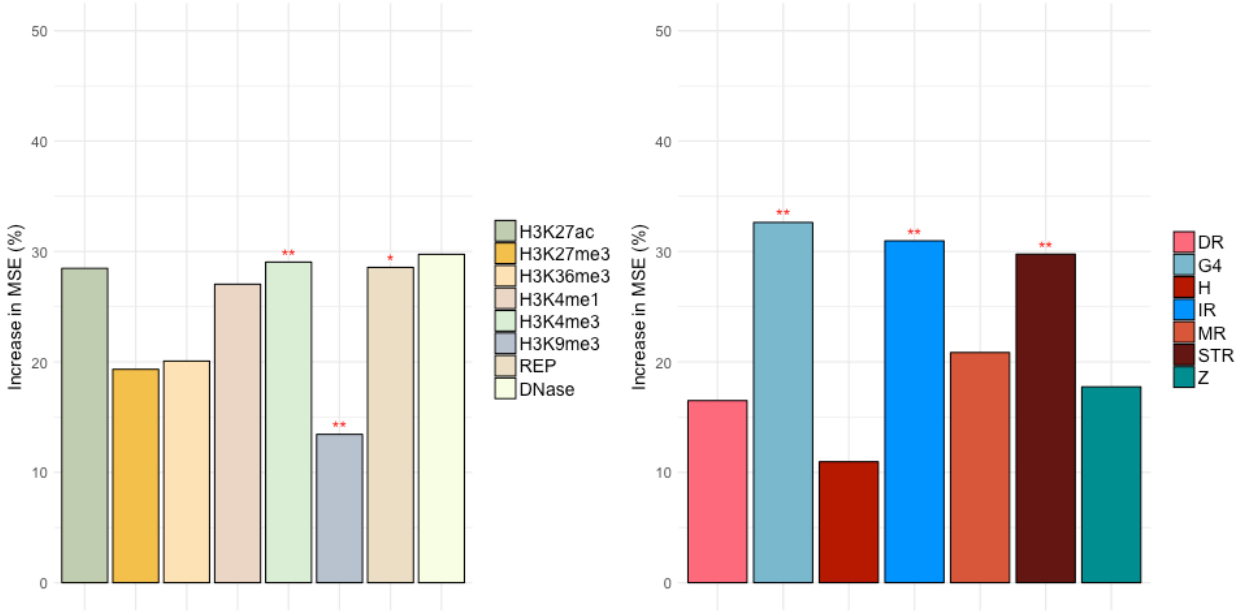



**Malignant Lymphoma**

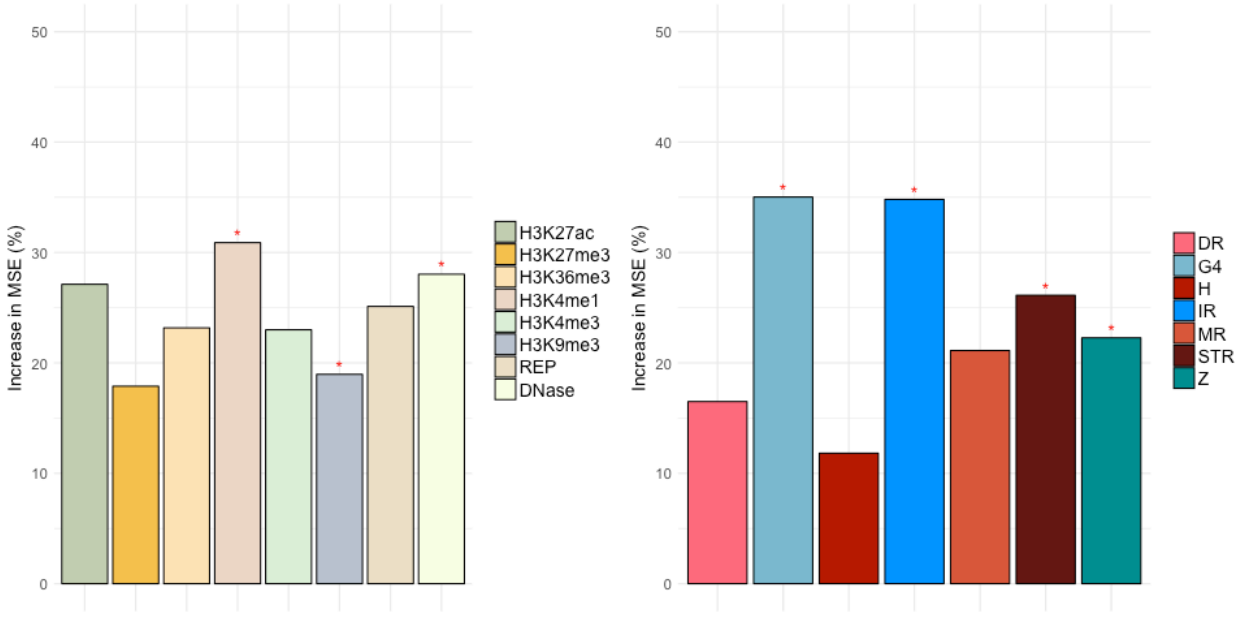



**Pedriatic Brain Tumour**

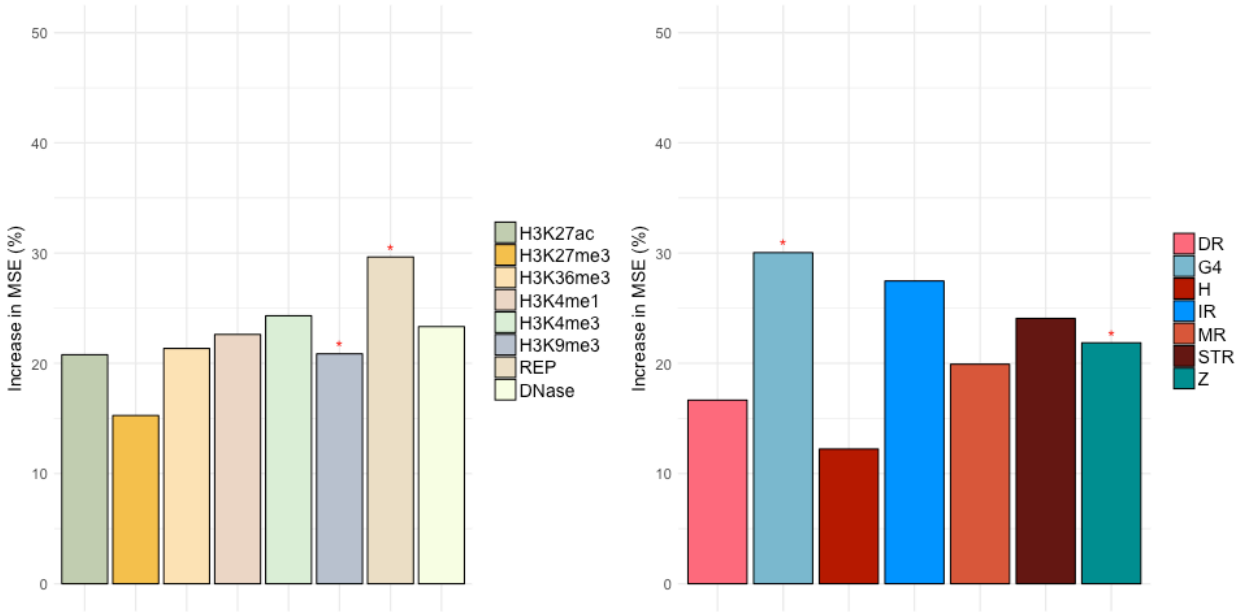
## C. REARRANGEMENTS



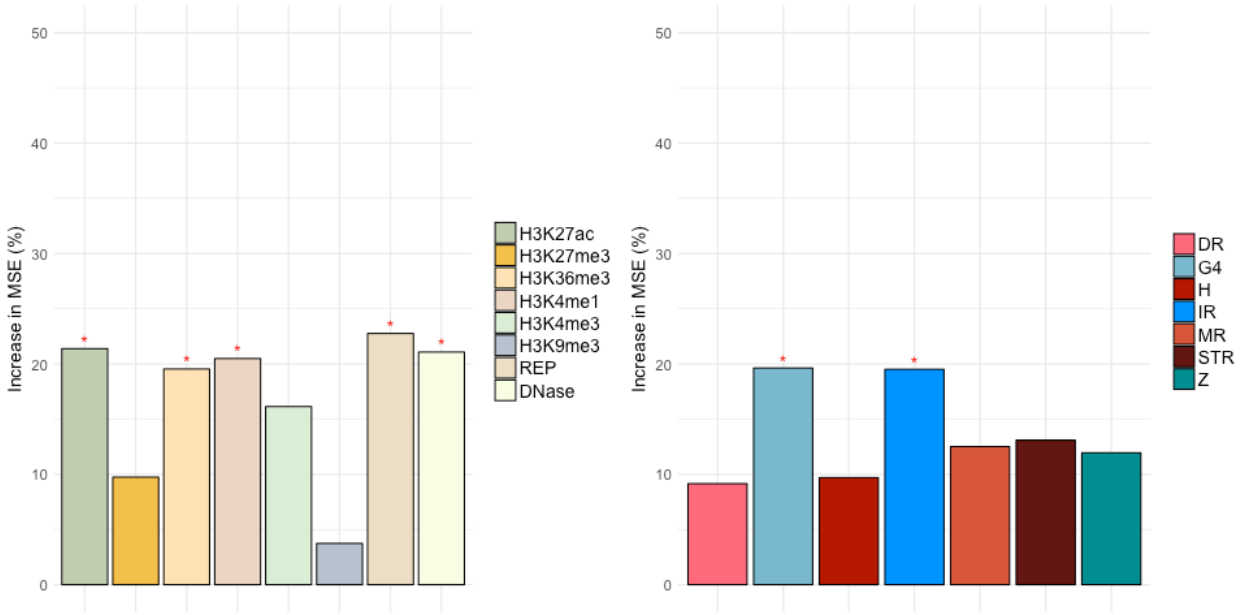

**Ovarian**



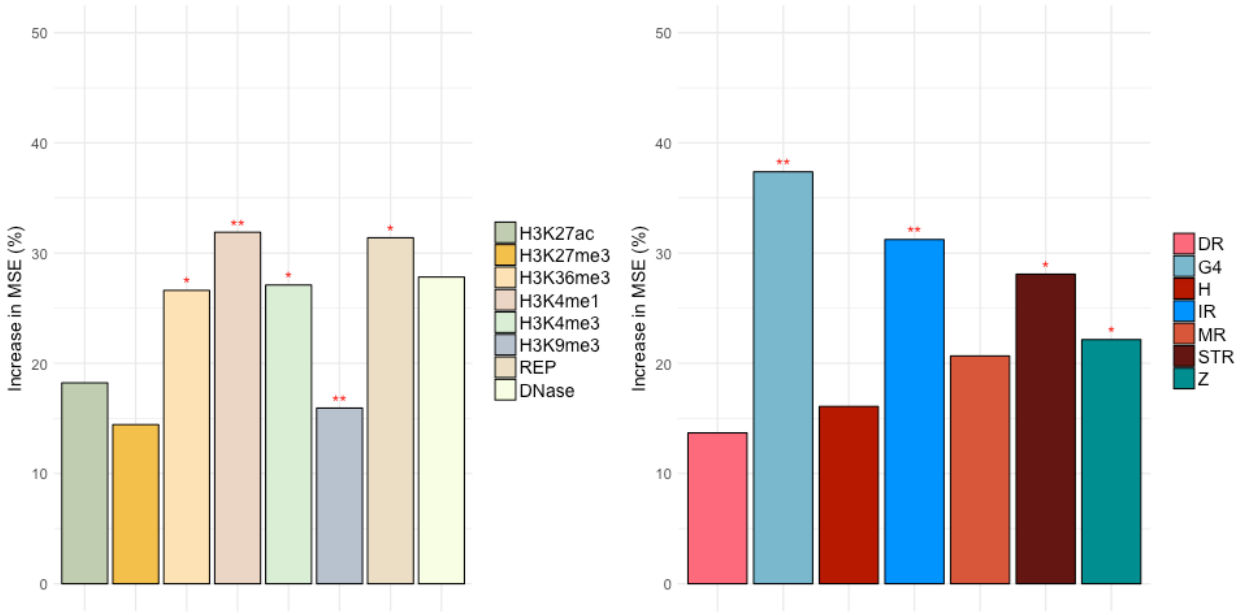

**Pancreatic**



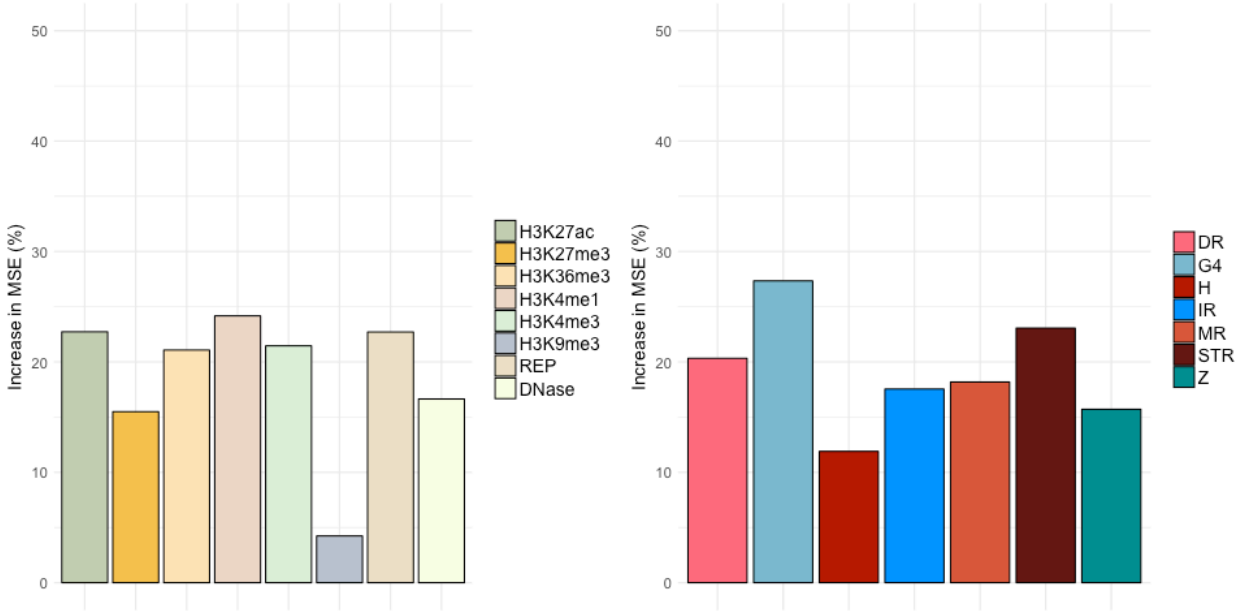

**Liver**




**Gastric**




**Esophagus**




**Renal cell cancer**

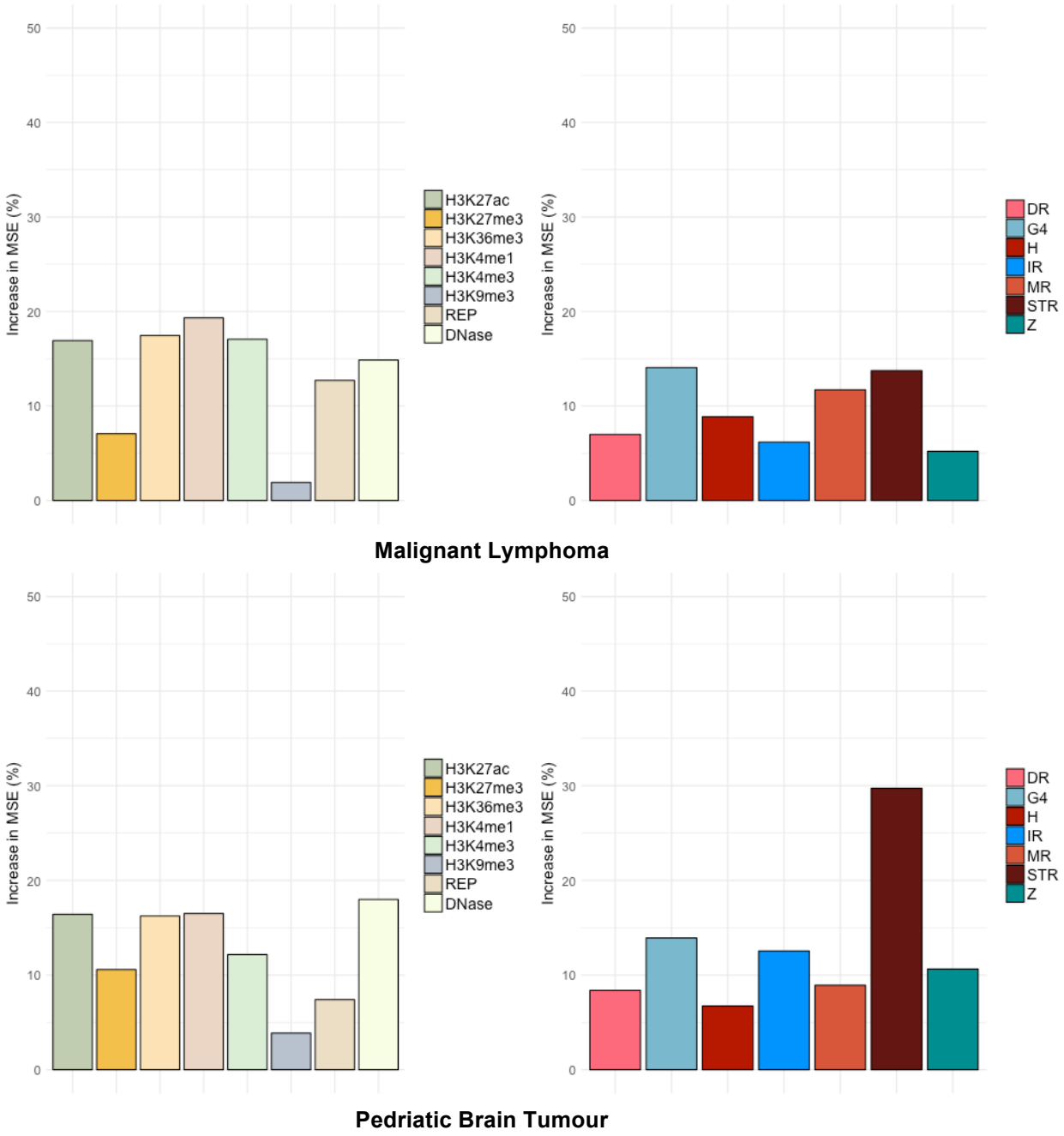
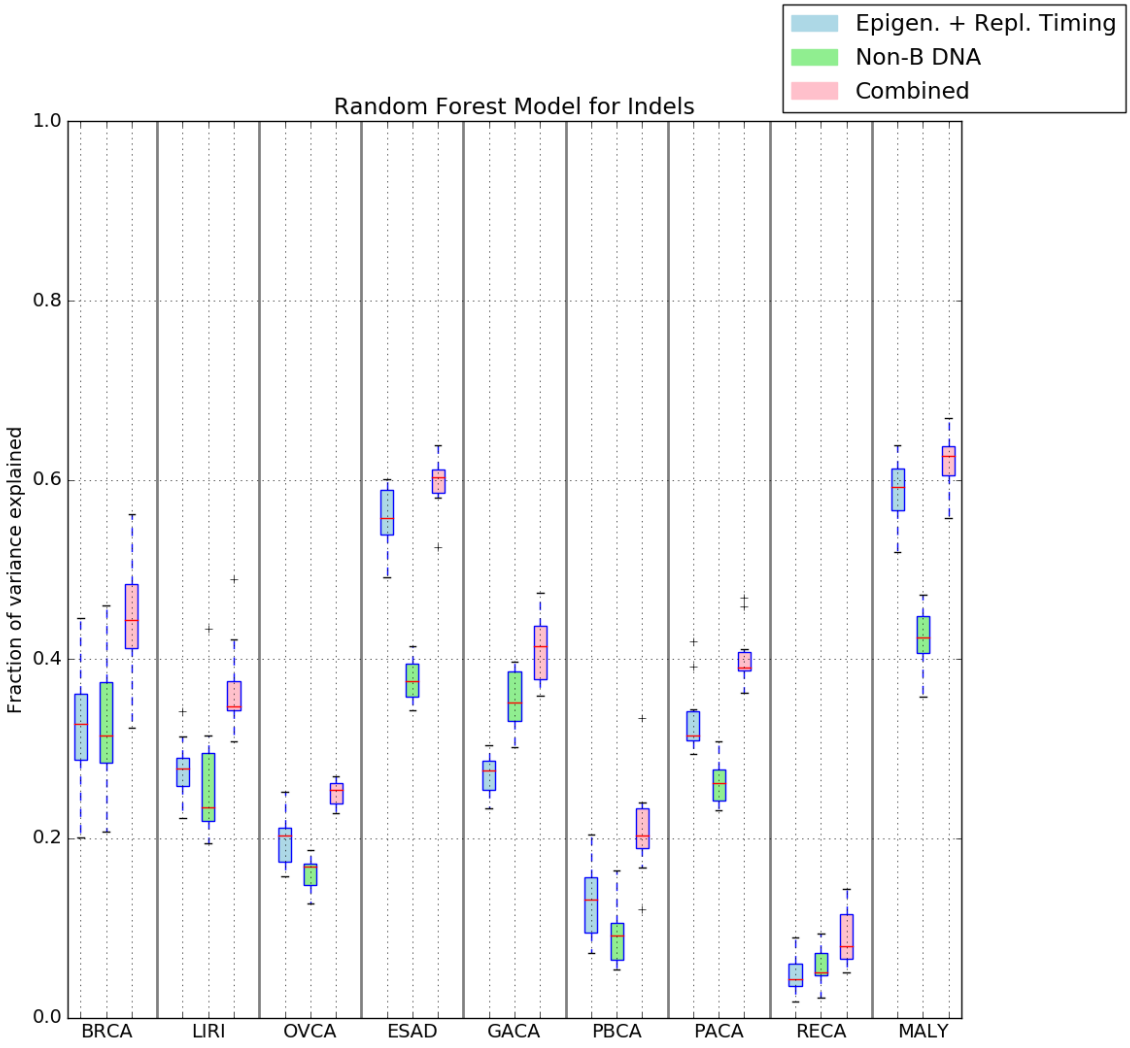
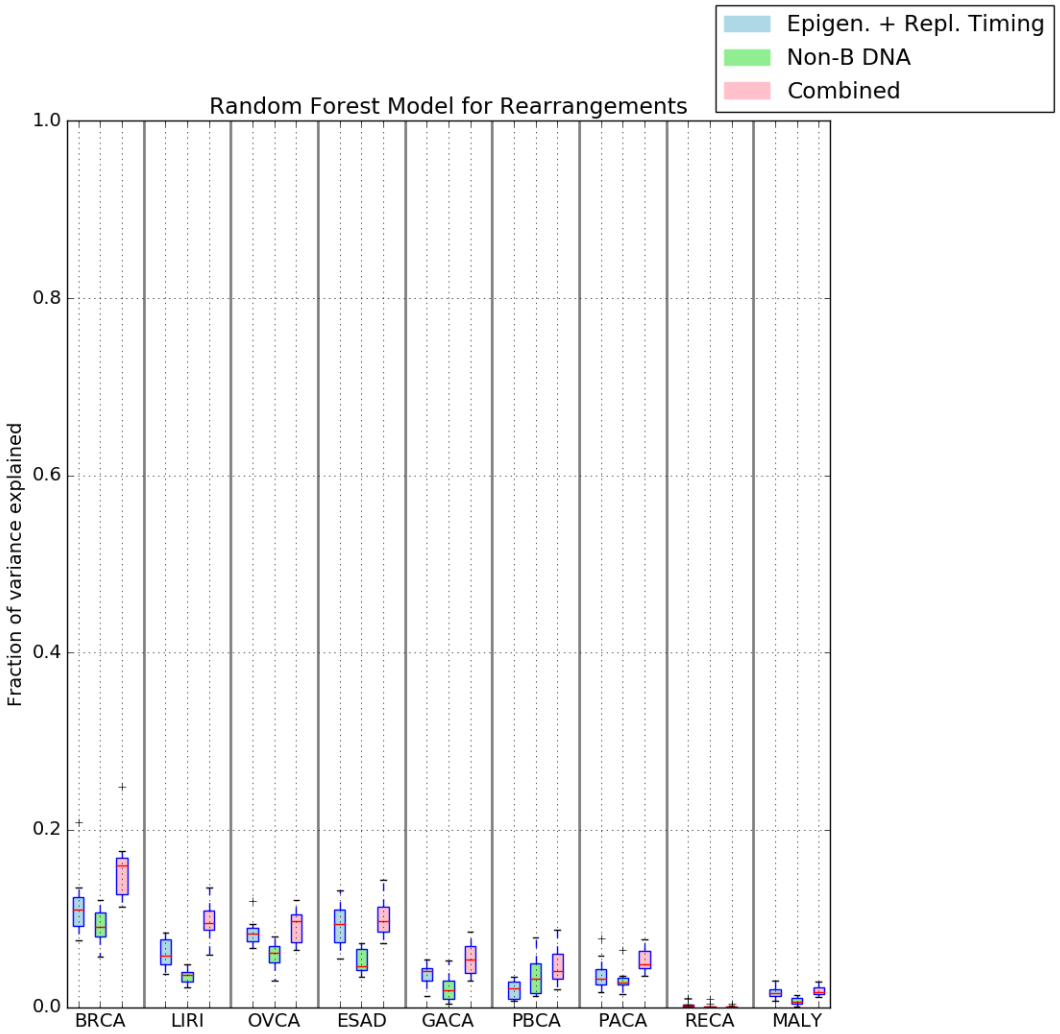



**Fig. S7: Importance of the different predictors for the random forest regression for different tumor types.** The y-axis shows the increase in mean square error (MSE) when the variable is excluded. Bars with * have an FDR<.05 and ** have FDR<.01 as determined by a

permutation test. Results depicted are for different mutation classes: *a)* substitutions, *b)* indels and *c)* rearrangements.

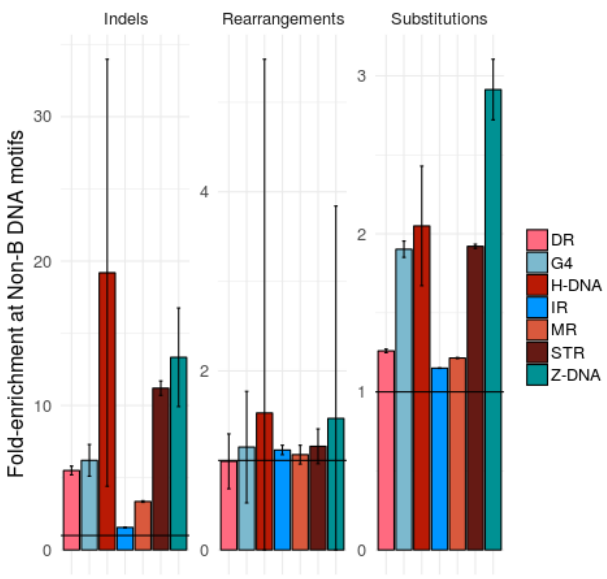



a.

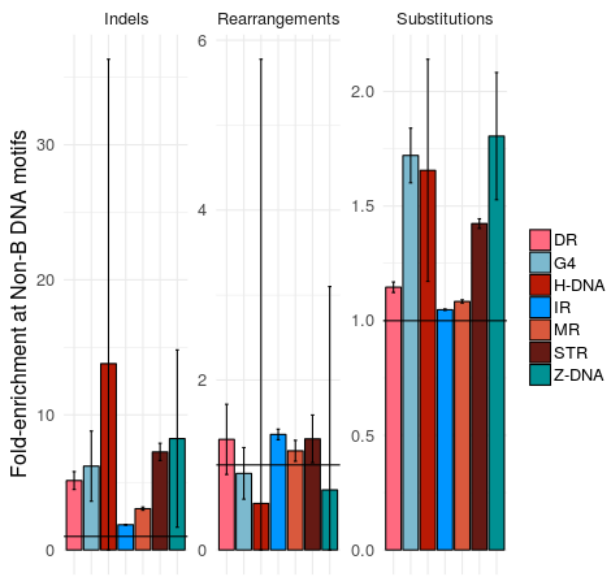



b.

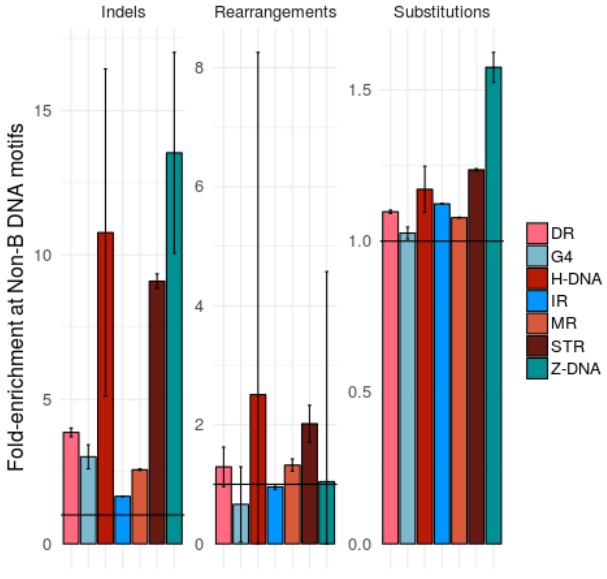
**Fig. S8: Random forest model predicts mutability.** *(a)* Indels and *(b)* rearrangements. Results depicted across nine tumour types using the following factors: non-B DNA motifs (green), epigenetics/replication timing (blue), or a combination of both (pink).



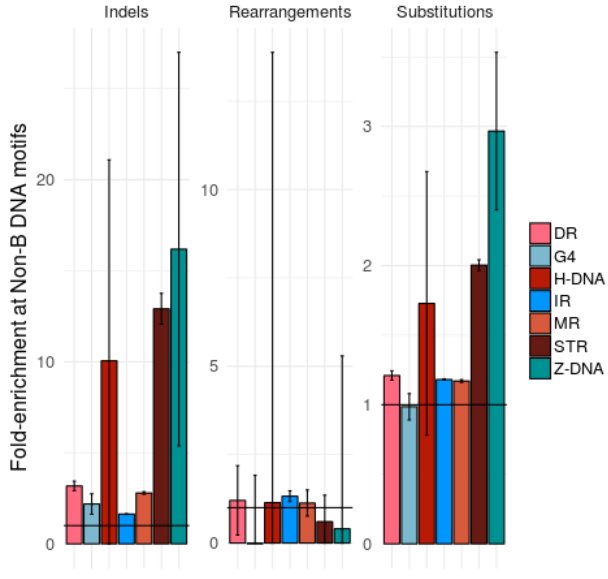

**Pancreatic**



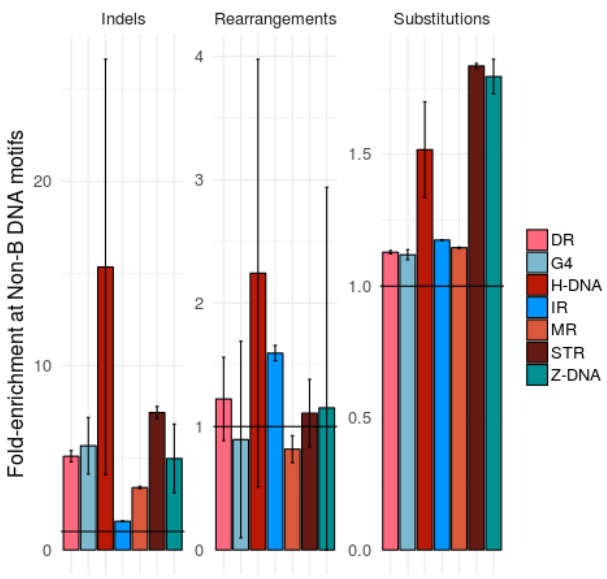

**Ovarian**



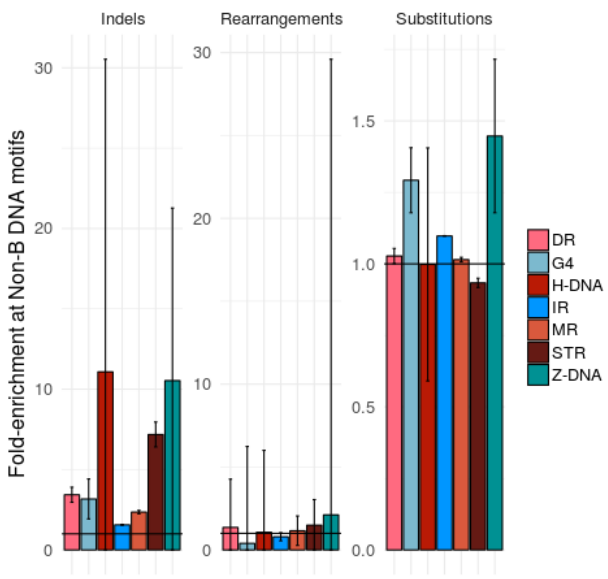

**Liver**



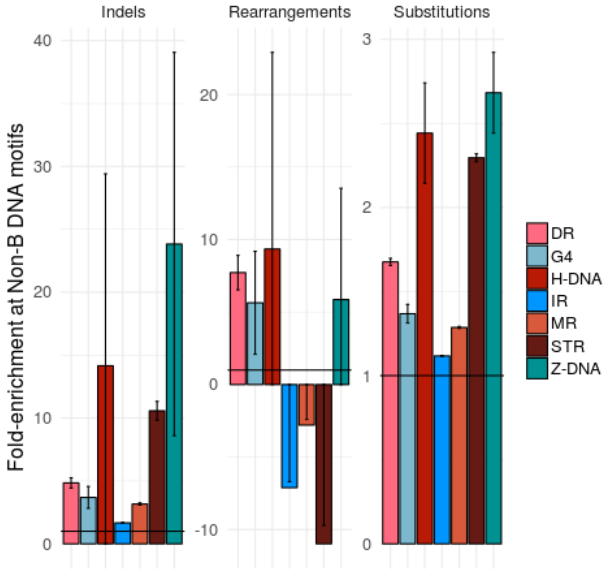

**Gastric**



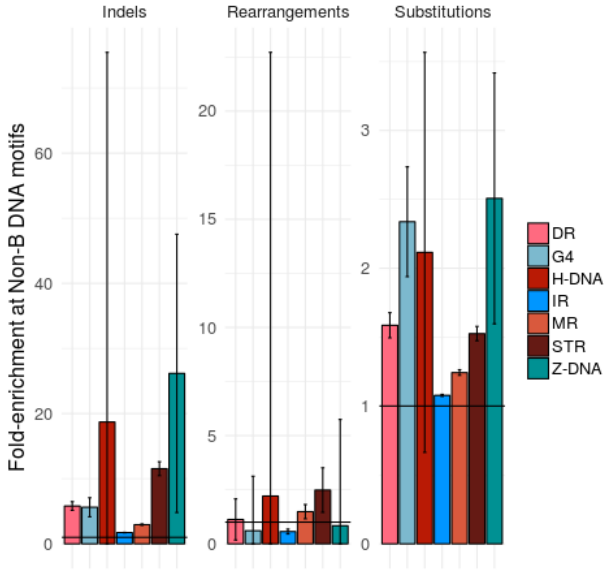

**Esophagus**



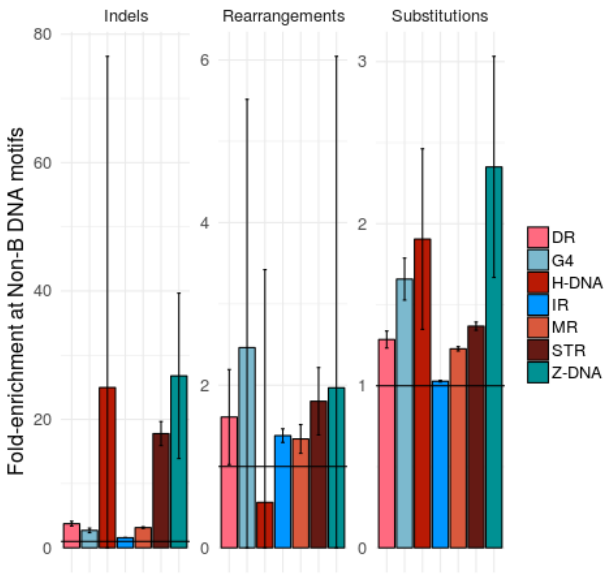

**Renal cell cancer**




**Malignant Lymphoma**

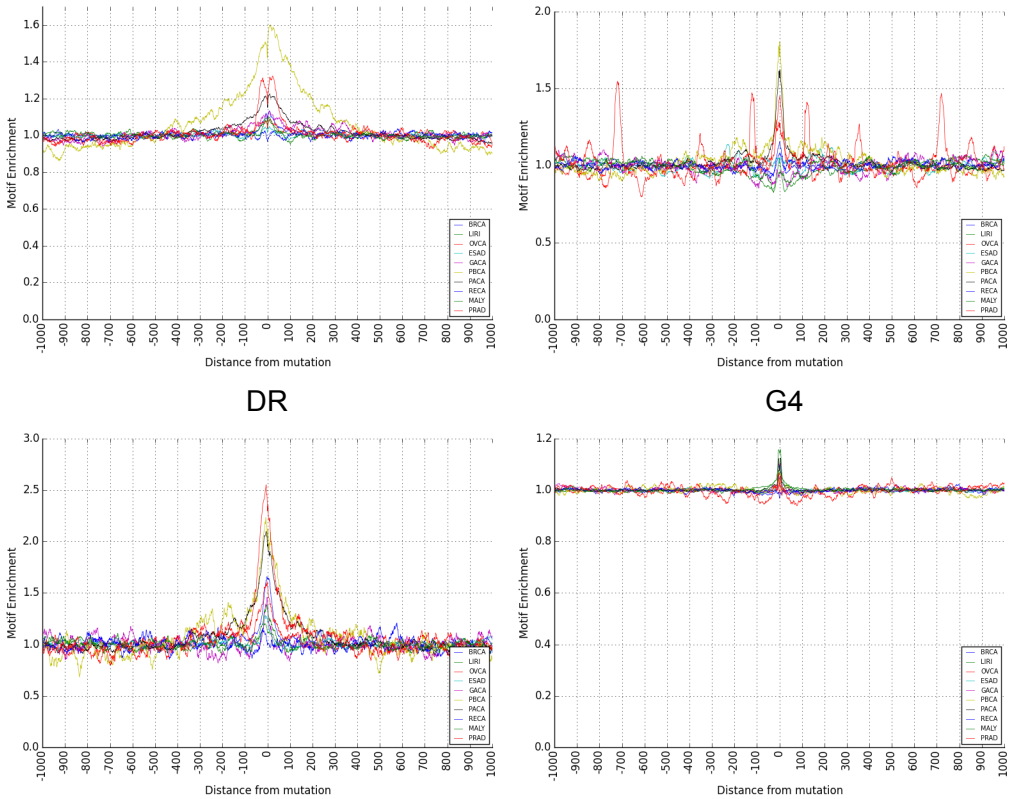


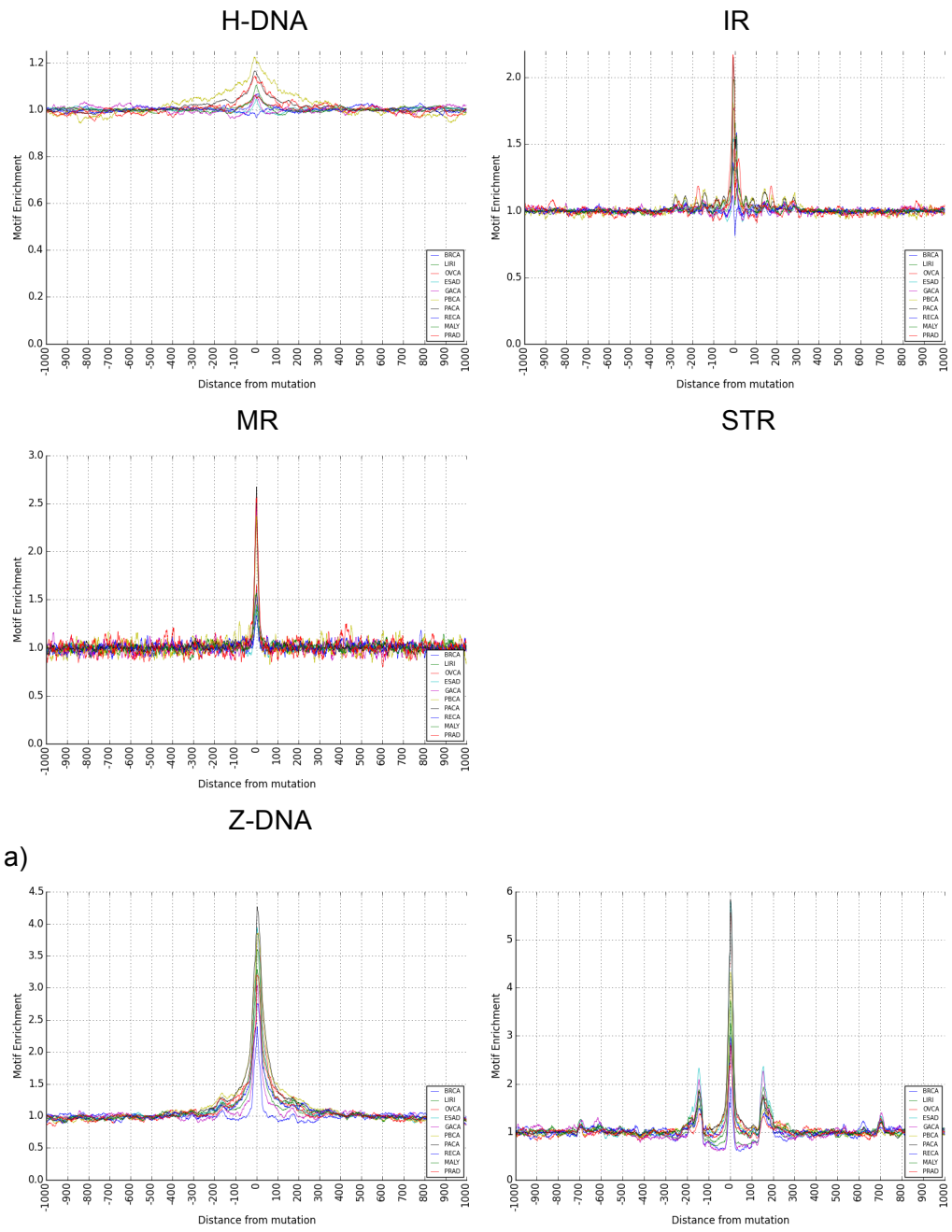

**Pedriatic Brain Tumour**




**Prostate**

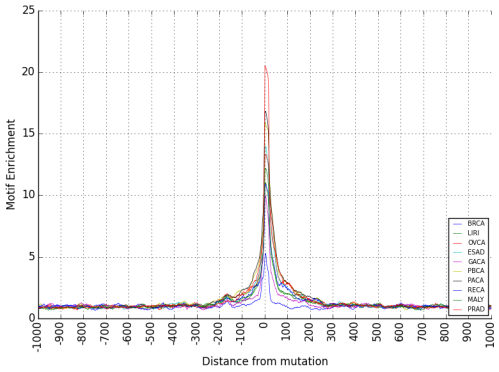
**Fig. S9: Enrichment of mutations overlapping non B-DNA motifs across nine tumour types.** *Negative enrichment in malignant lymphoma IR, MR, STR for rearrangements is due to correction (see Methods).



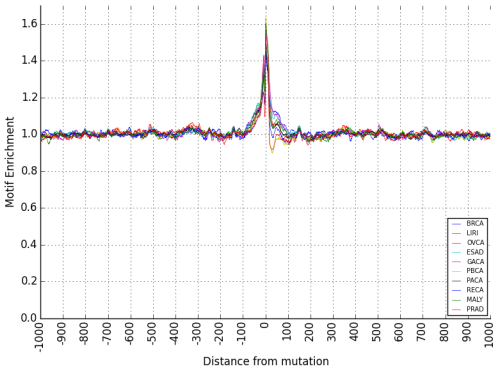

H-DNA
IR
MR
STR
Z-DNA
Motif Enrichment
Distance from mutation
BRCA
LIRI
OVCA
ESAD
GACA
PBCA
PACA
RECA
MALY
PRAD

a)

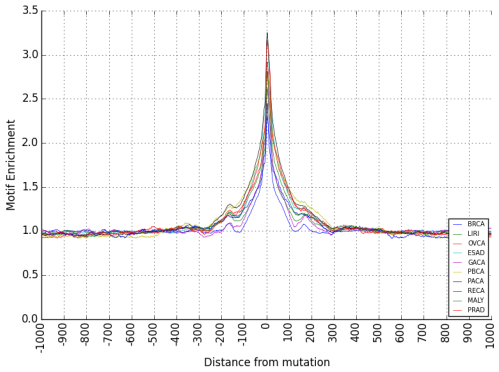

DR
G4
Motif Enrichment
Distance from mutation
BRCA
LIRI
OVCA
ESAD
GACA
PBCA
PACA
RECA
MALY
PRAD



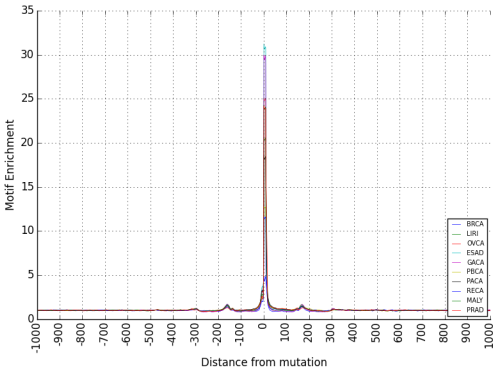

H-DNA



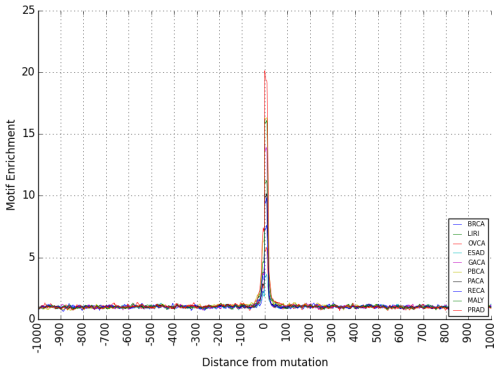

IR




MR

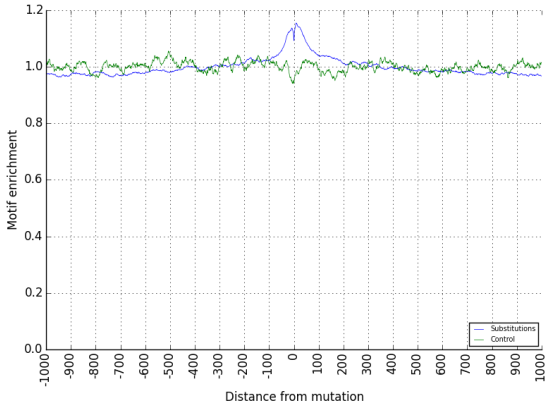



STR

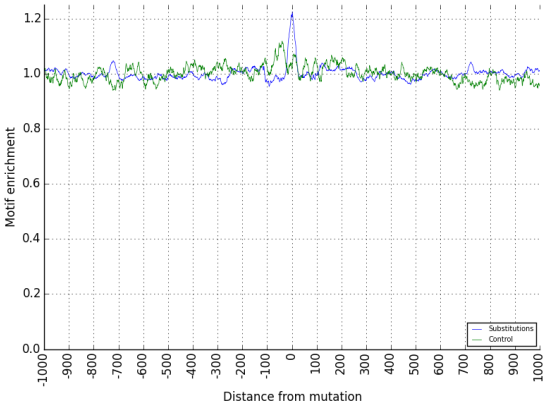



Z-DNA

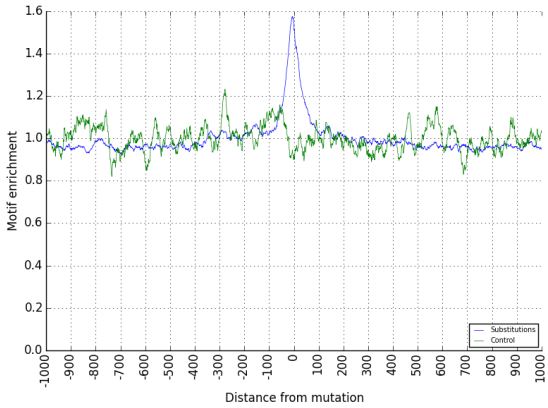
b)



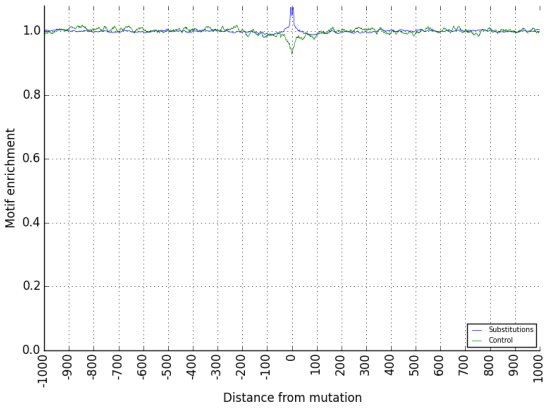

DR



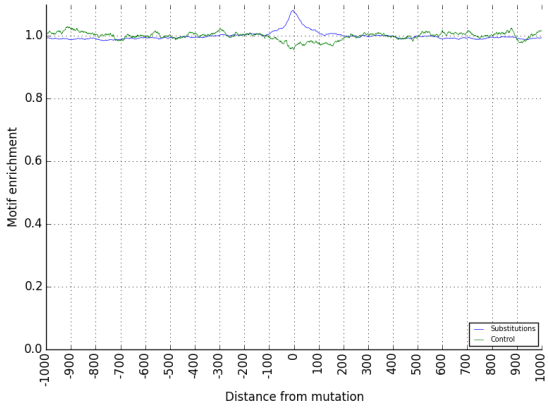

G4



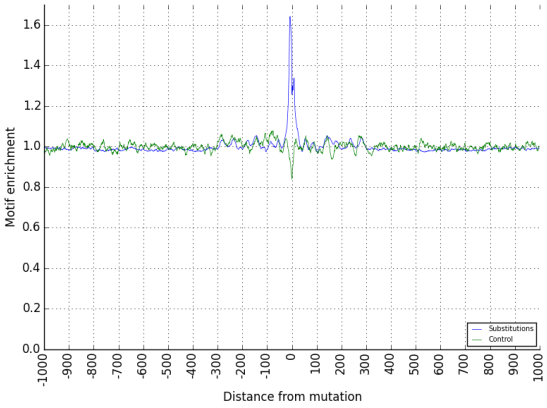

H-DNA



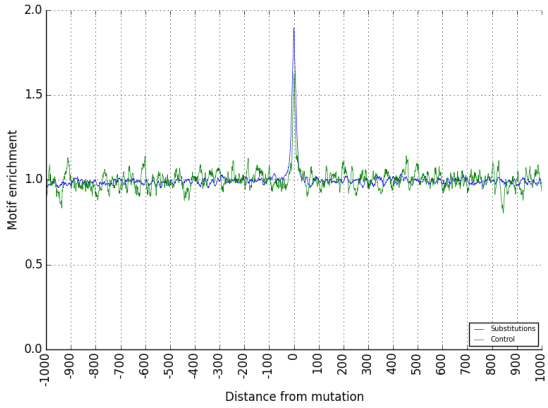

IR




MR

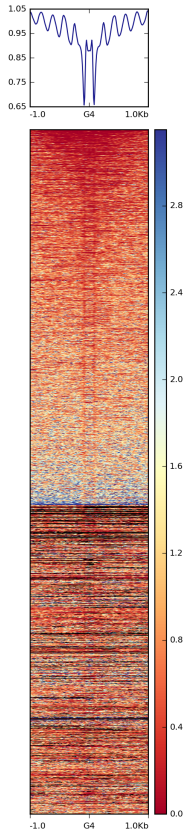



STR



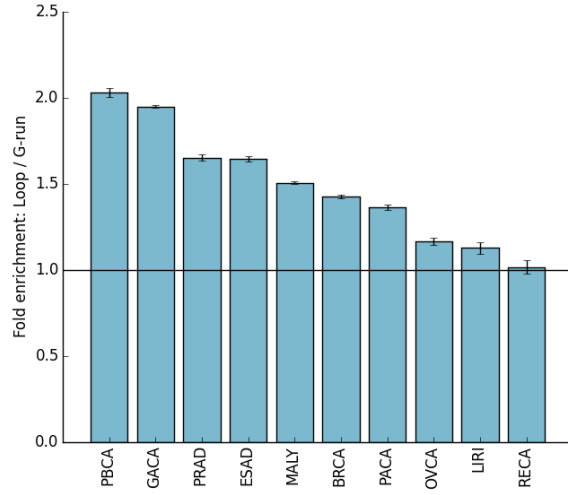

Z-DNA

c)

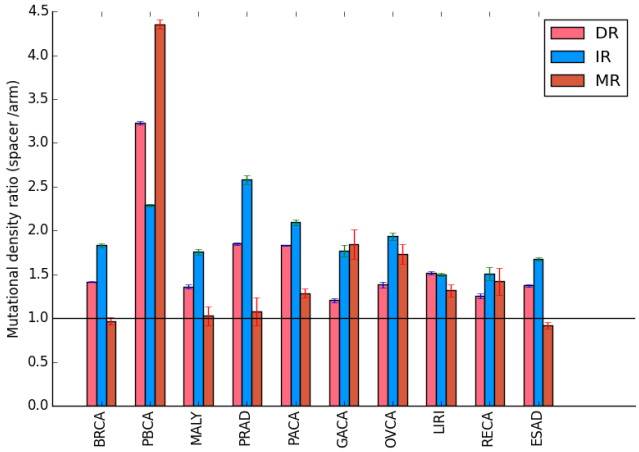



d)

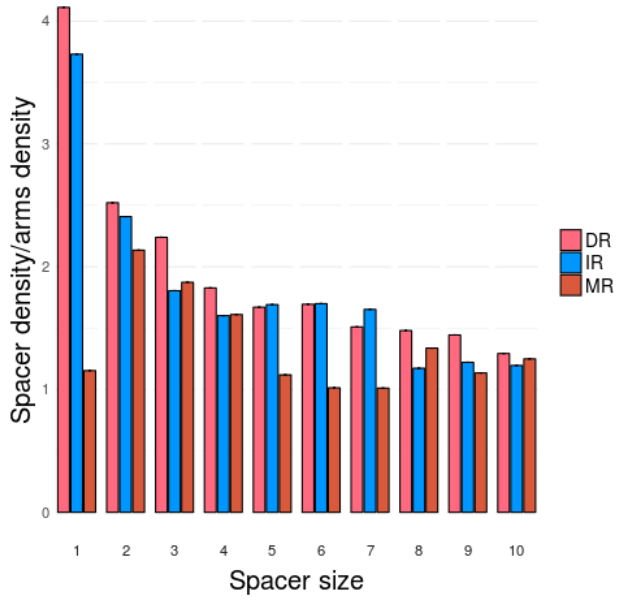
**Fig. S10: Relationship between non-B DNA motifs and mutability.** Mutational enrichment at 2kb window (+/- 1kb) showing the distribution of non-B DNA motifs for: a) substitutions, centered at position 0 on x-axis, b) indels, centered at position 0 on x-axis, c) substitutions and controls generated with same trinucleotide content, both centered at position 0 on x-axis. d) Nucleosome density at 2kb window centered at G4s (position 0, x-axis), indicating the relationship between G4s and nucleosome occupancy (MNase signal).



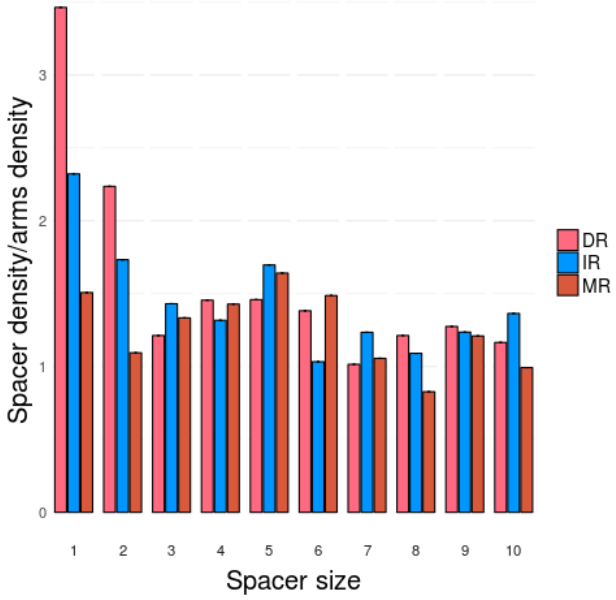

a)



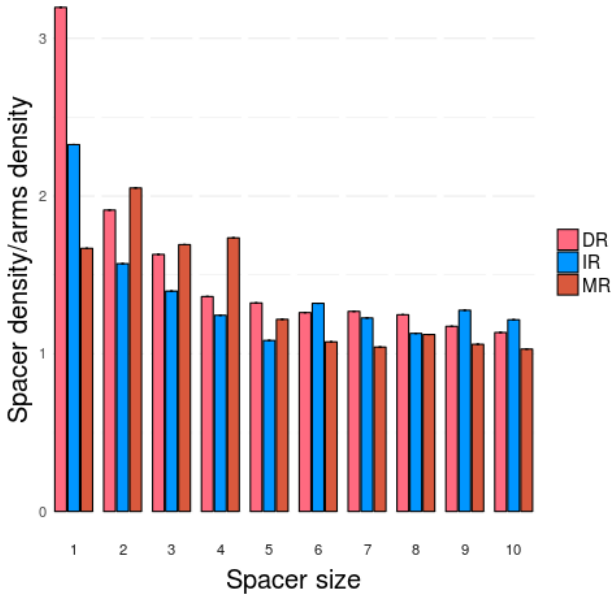

b)



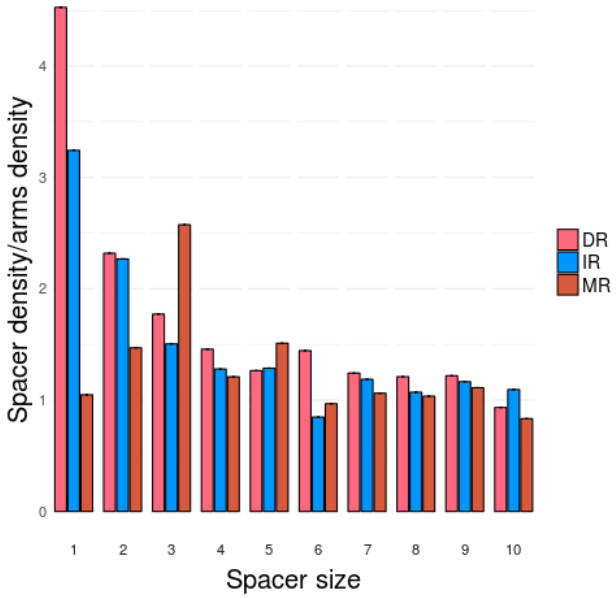

**Pancreatic**



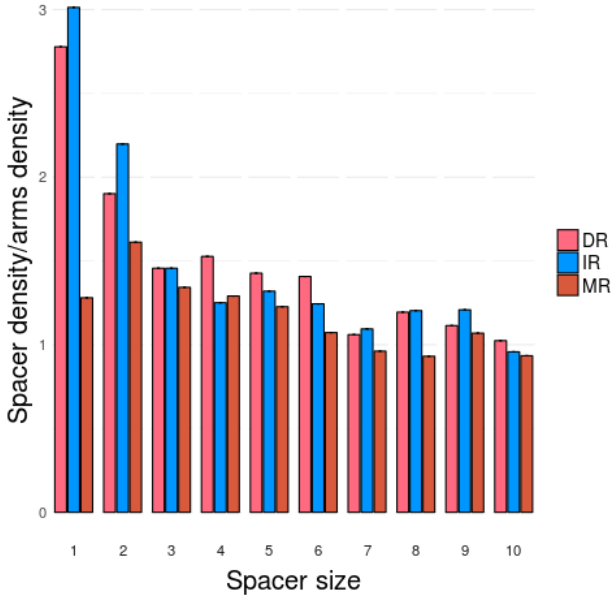

**Ovarian**



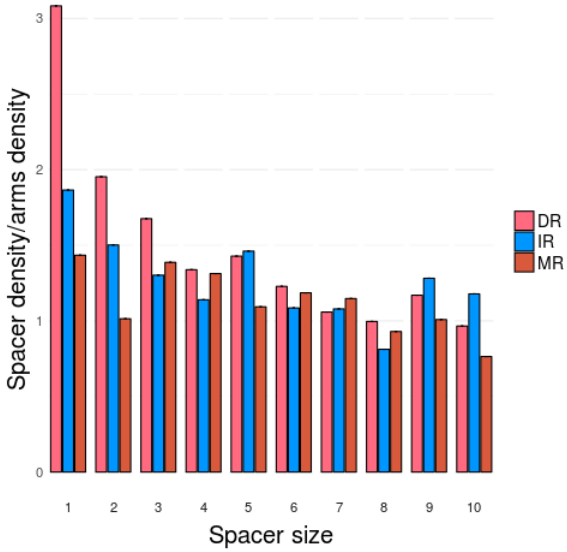

**Liver**



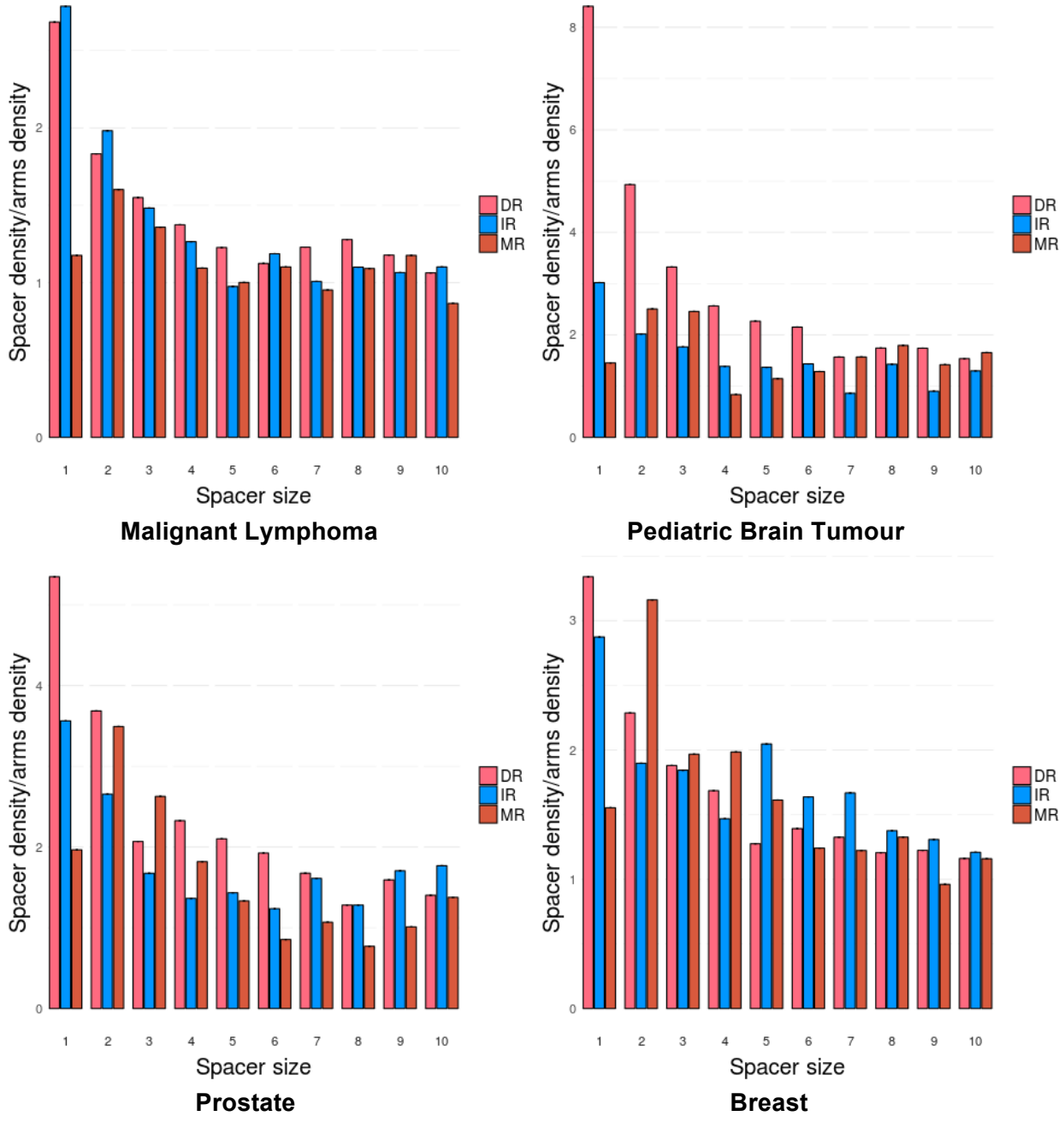

**Gastric**




**Esophagus**



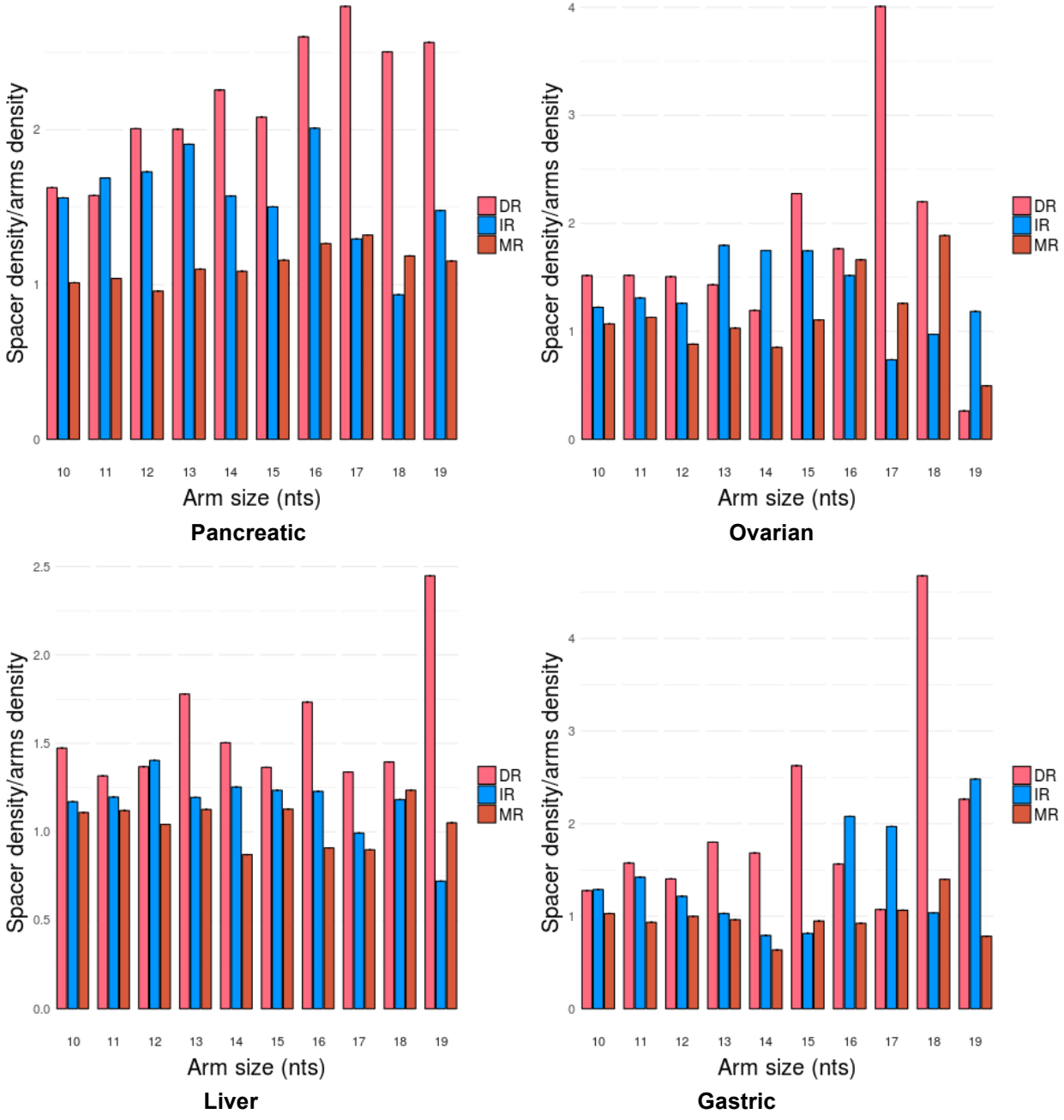

**Renal cell cancer**

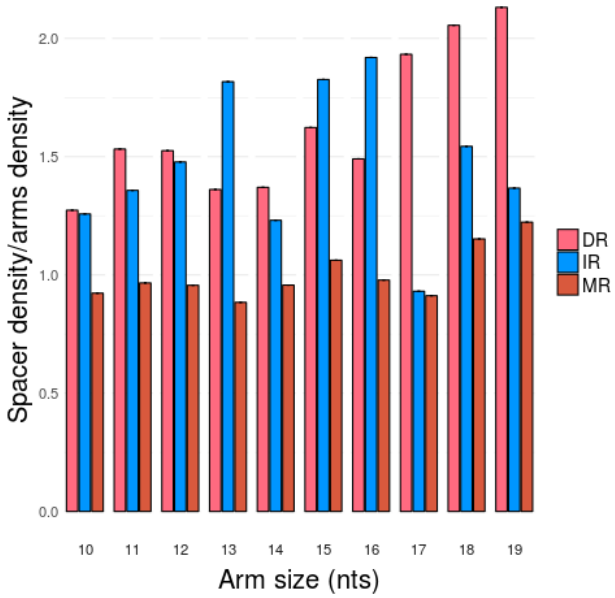



c)

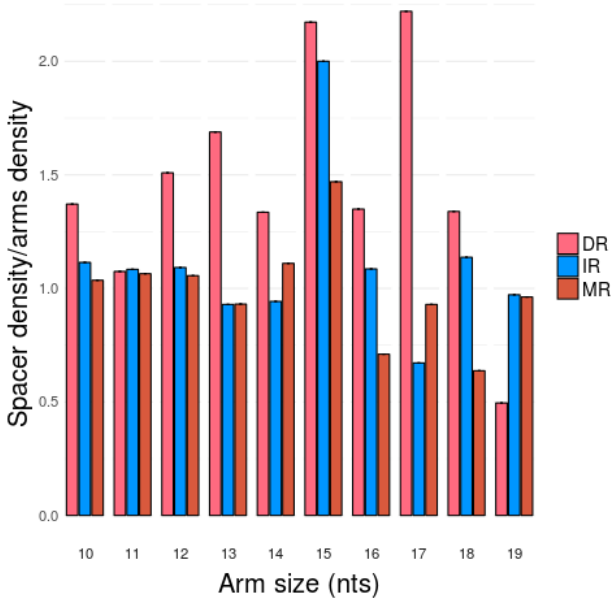
**Fig. S11: Increased mutability is domain-specific for particular non-B DNA motifs.** a) Mutational density in loops compared to G-runs for G-quadruplexes, corrected for trinucleotide content. b) Mutational density in spacers compared to arms for direct repeats, inverted repeats and mirror repeats, corrected for trinucleotide content. Error bars represent standard error from bootstrapping in a-b. c) Mutational density in spacers compared to arms for direct repeats,

inverted repeats and mirror repeats as a function of the length of the spacer. Error bars represent the standard error. Results shown for ten tumour types.

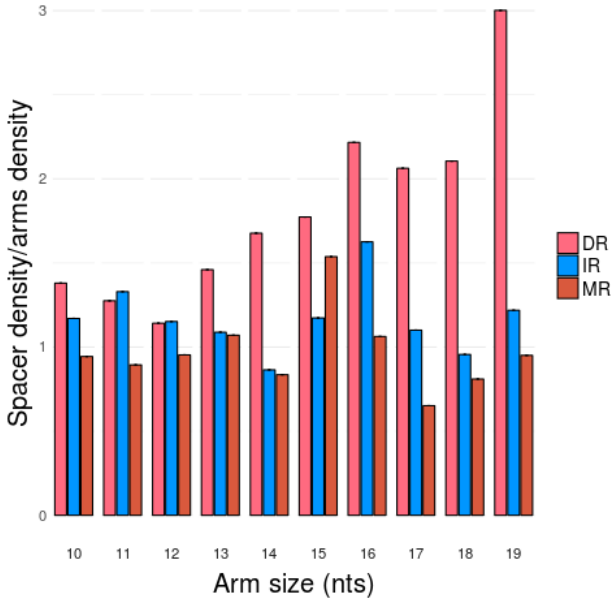




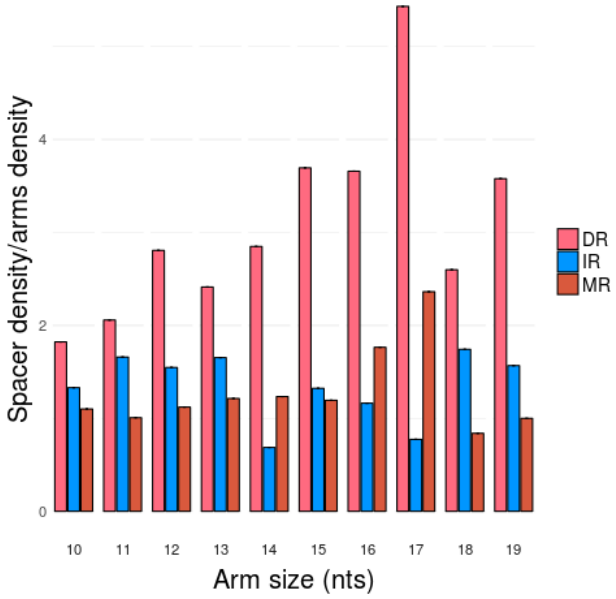

**Esophagus**




**Renal cell cancer**




**Malignant Lymphoma**




**Pediatric Brain Tumour**

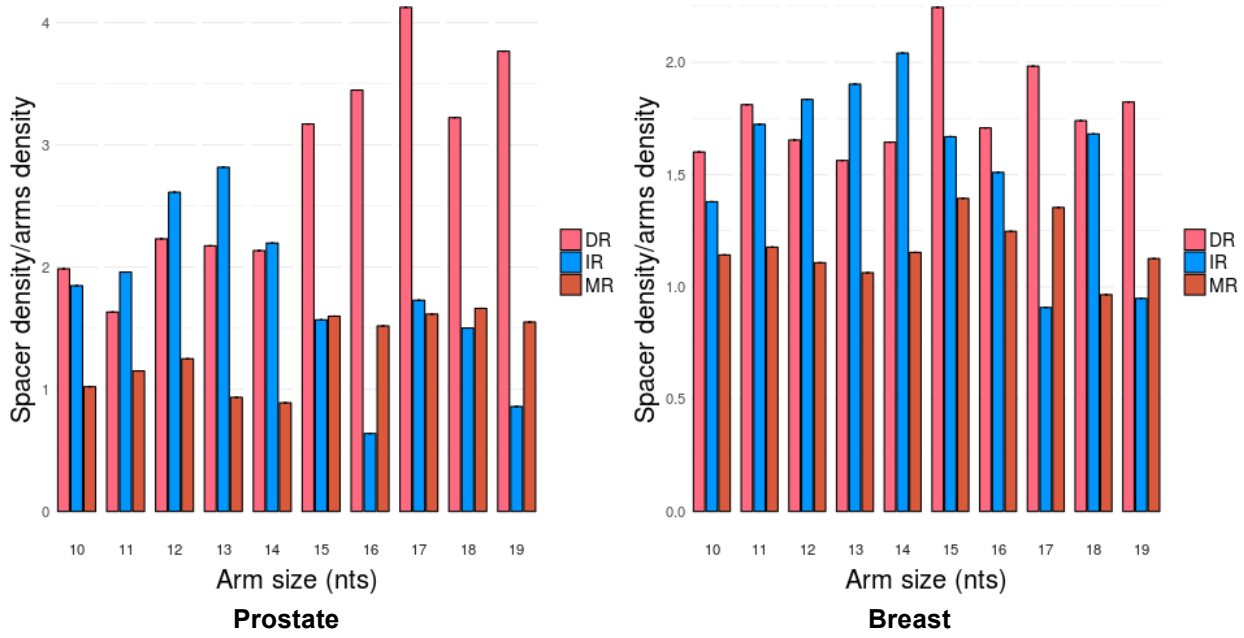



**Fig. S12: Increased mutability is domain-specific for particular non-B motifs.** Mutation density in spacers compared to arms for direct repeats, inverted repeats and mirror repeats as a function of the length of the arm. Error bars represent the standard error. Results shown for ten tumour types.

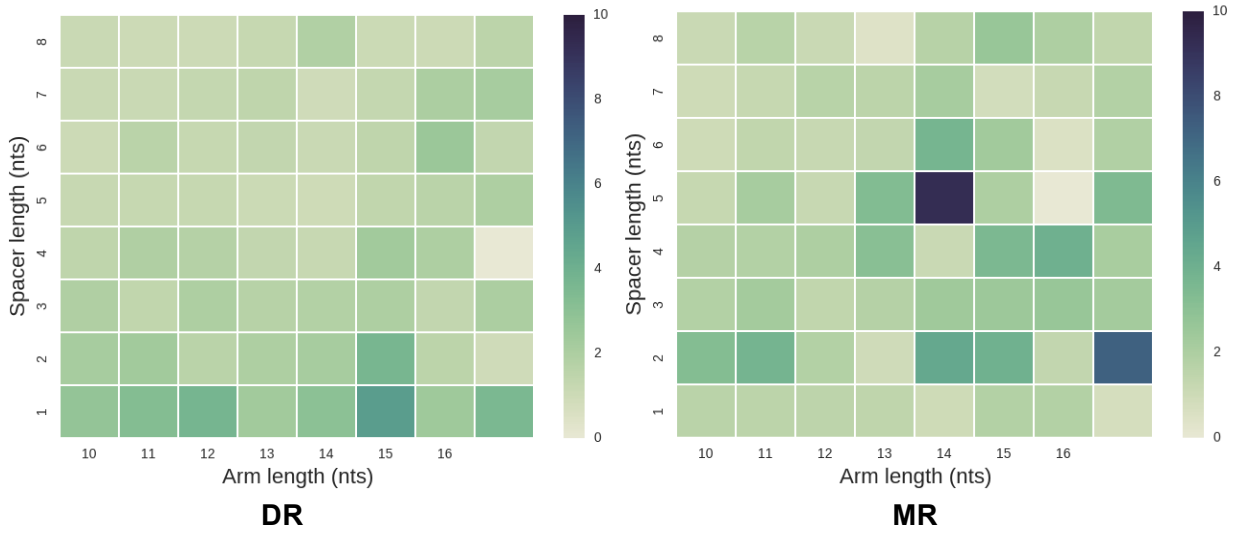
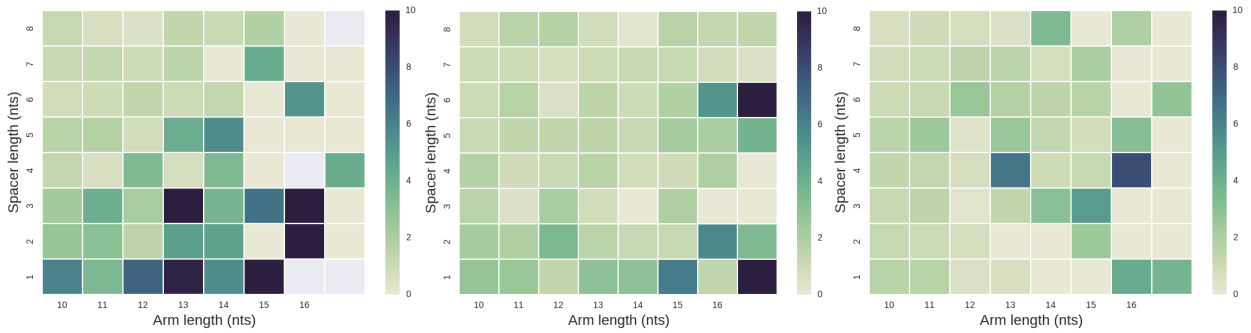
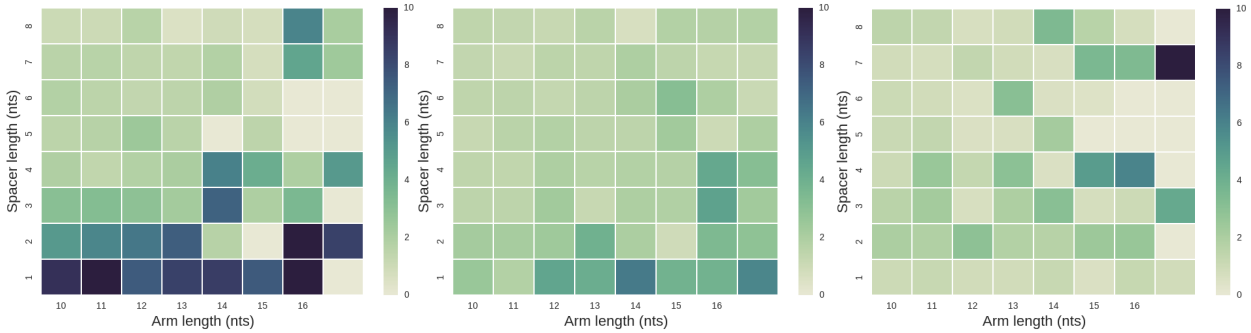



**Fig. S13: Differences in spacer to arm mutability in non-B motifs for different spacer and arm lengths variable from one non-B motif to another in breast cancer.** Heatmap of

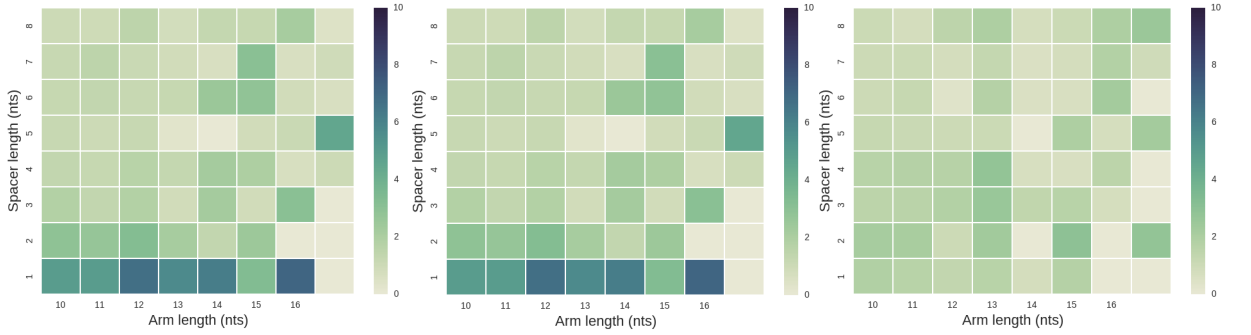
mutational density at spacer versus arms for different spacer and arm lengths for direct repeats (DR) and mirror repeats (MR) for breast cancer showing differences to inverted repeats (IR, as seen in Fig. 3d).



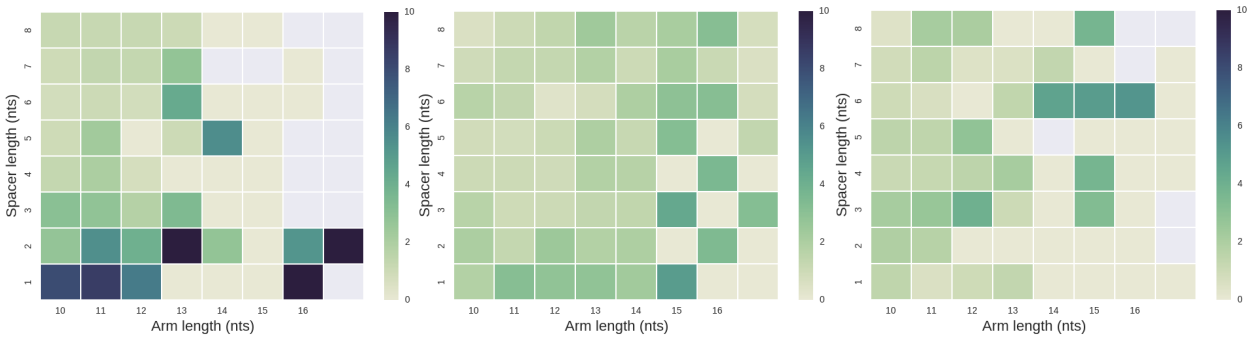

**Ovarian**




**Pancreatic**

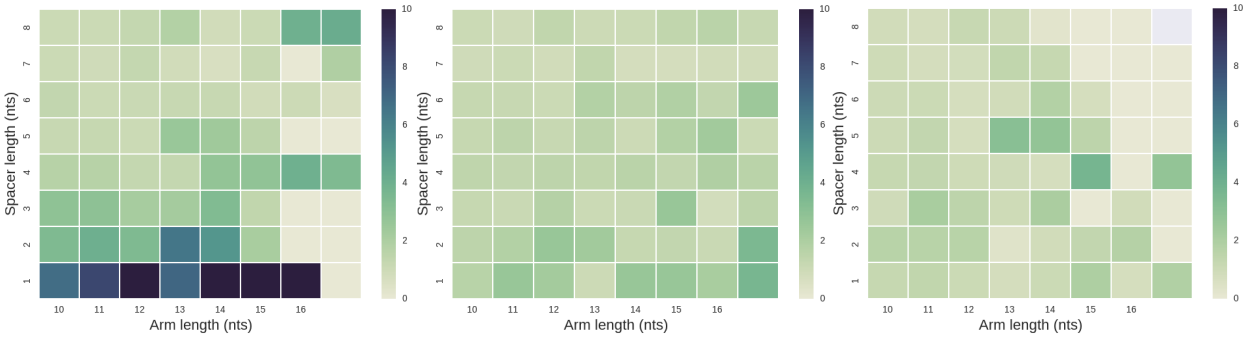



**Liver**



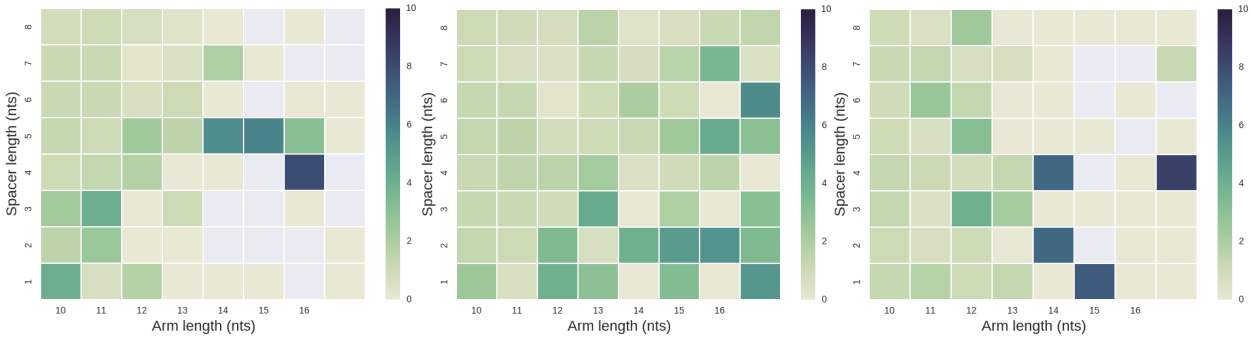

IR DR MR

**Gastric**



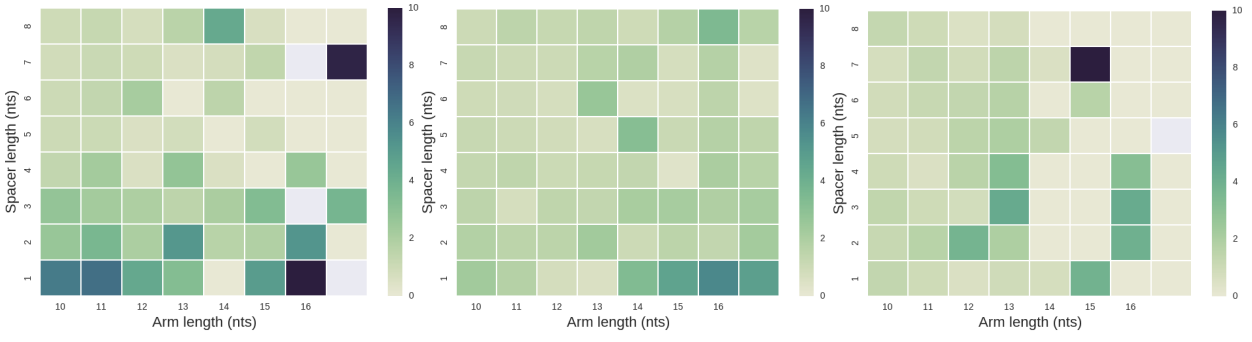

IR DR MR

**Esophagus**




IR DR MR

**Renal cell cancer**




IR DR MR

**Malignant Lymphoma**

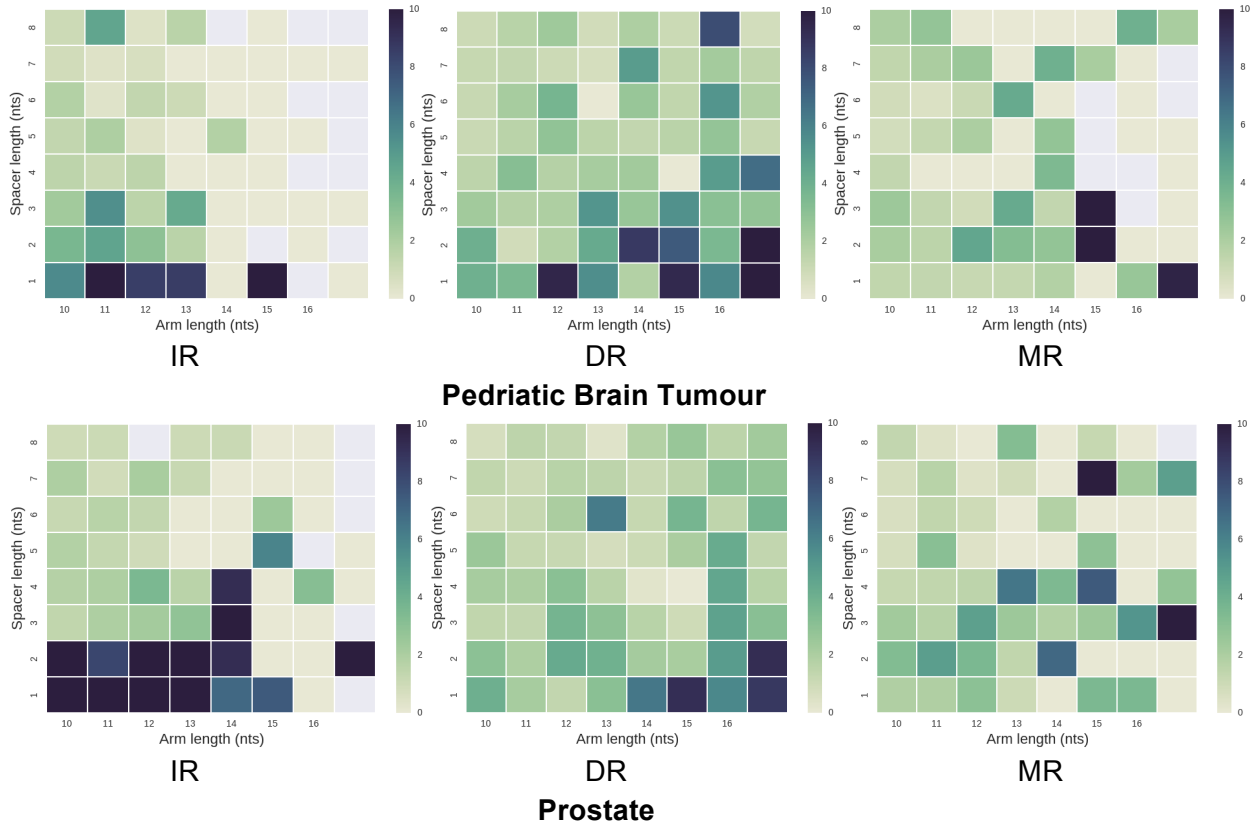
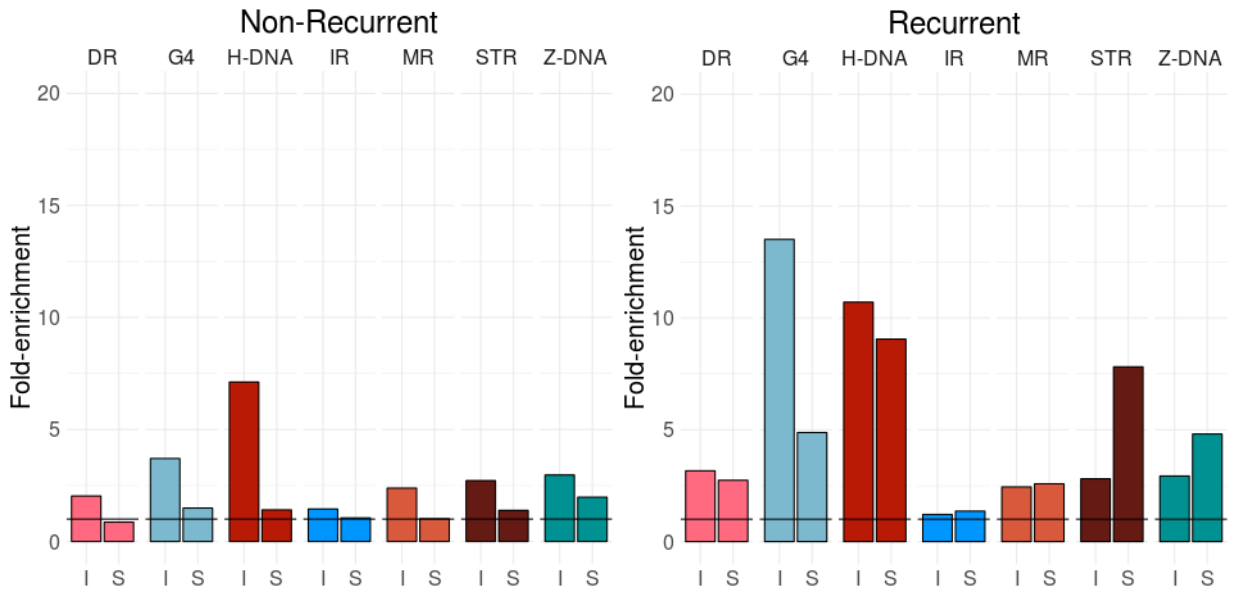
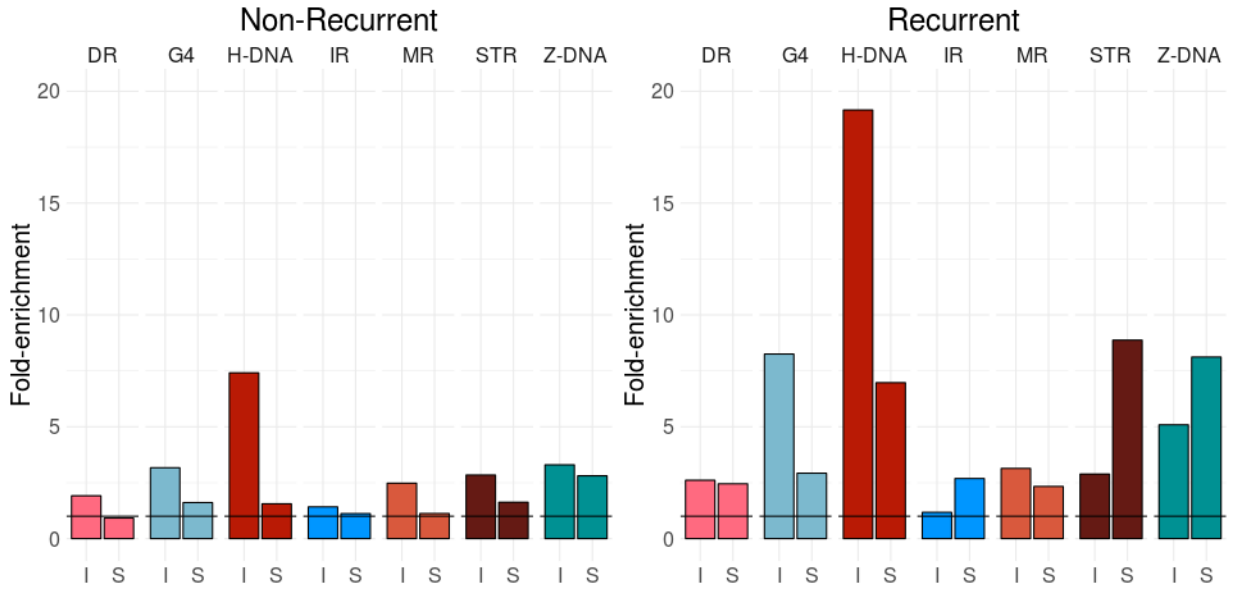


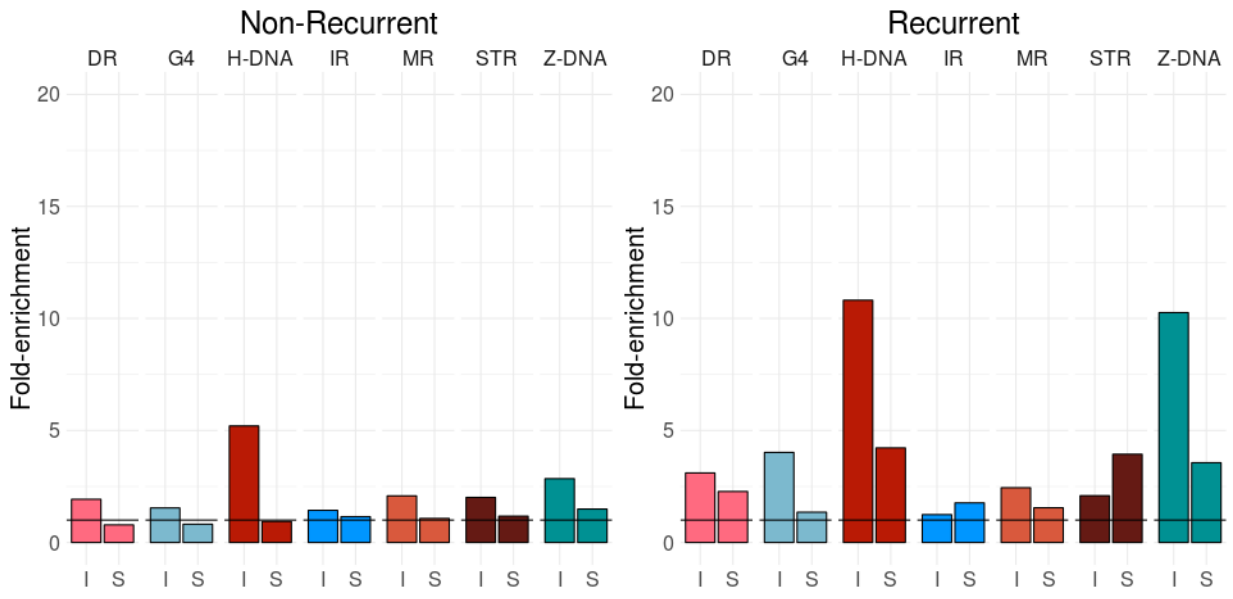

**Fig. S14: Differences in spacer to arm mutability in non-B motifs is particularly elevated for spacer and arm lengths that are associated with increased thermodynamic stability for secondary structure formation in different tumor types.** Heatmap of mutational density at spacer versus arms for different spacer and arm lengths for inverted repeats (IR, left), direct repeats (DR, middle), mirror repeats (MR right) across nine tumour types.



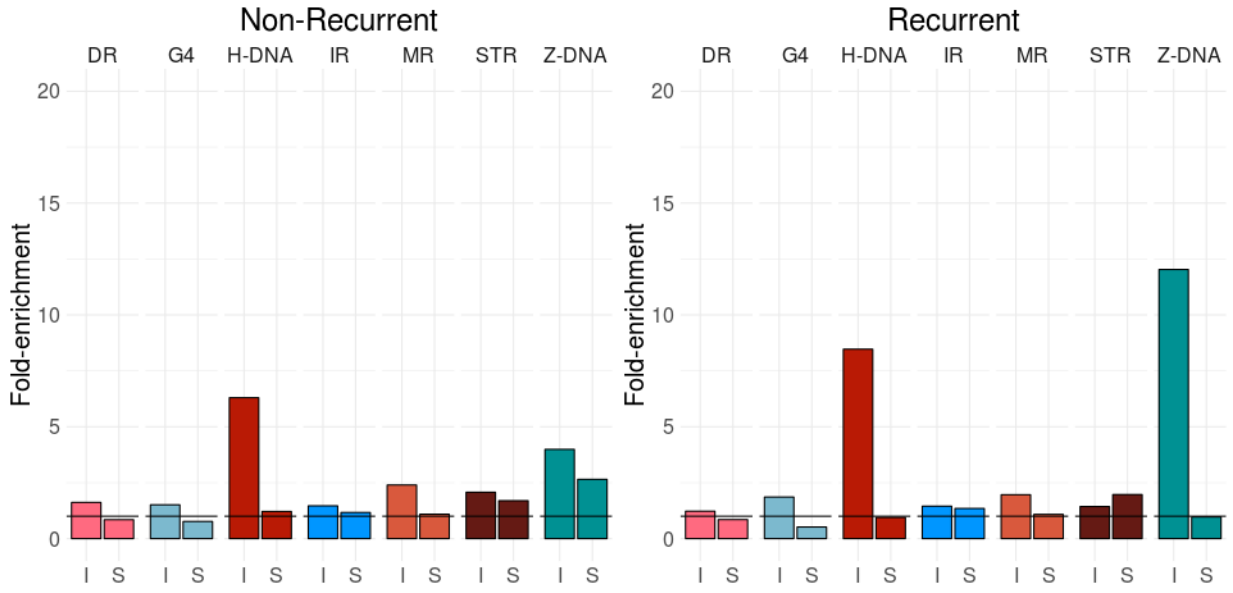

**Ovarian**



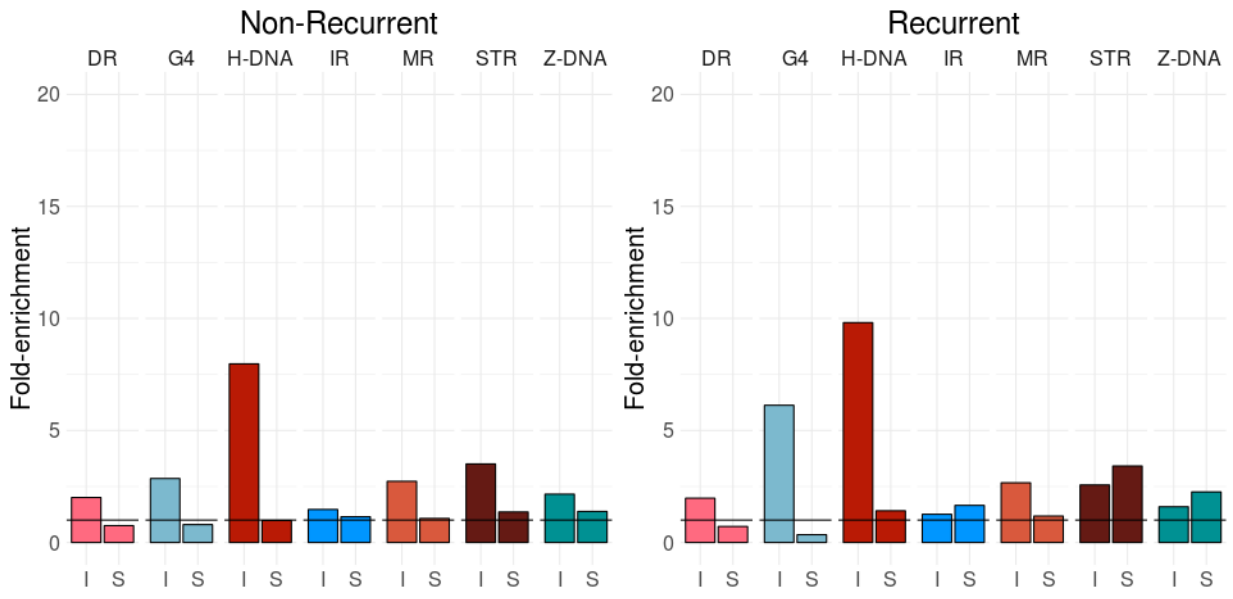

**Pancreatic**



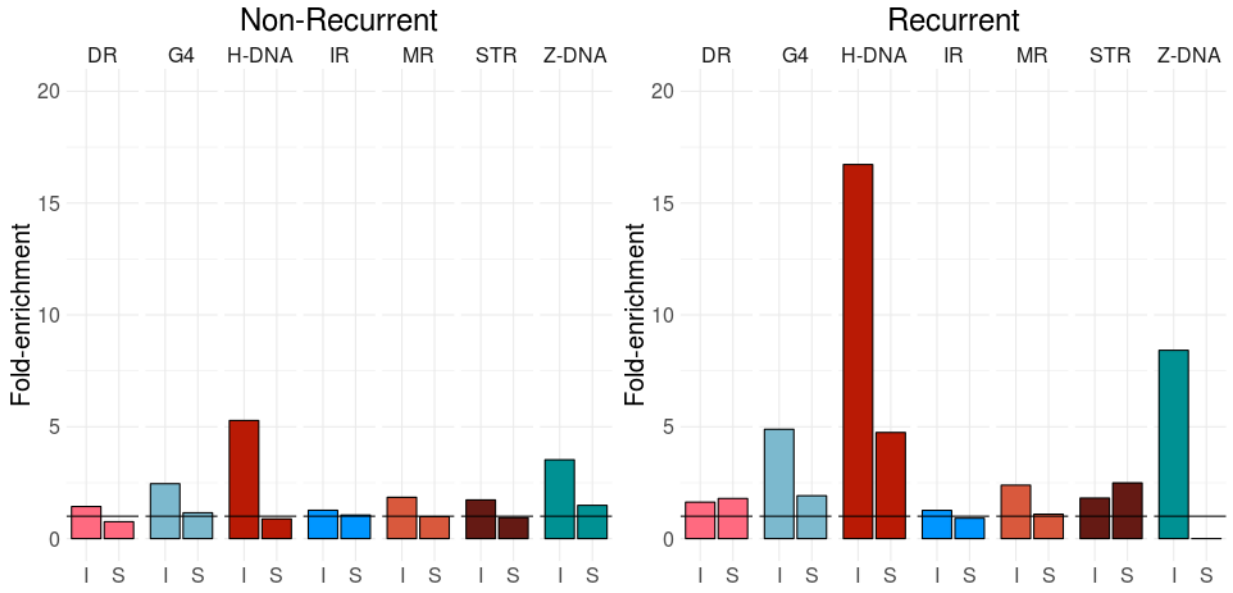

**Liver**



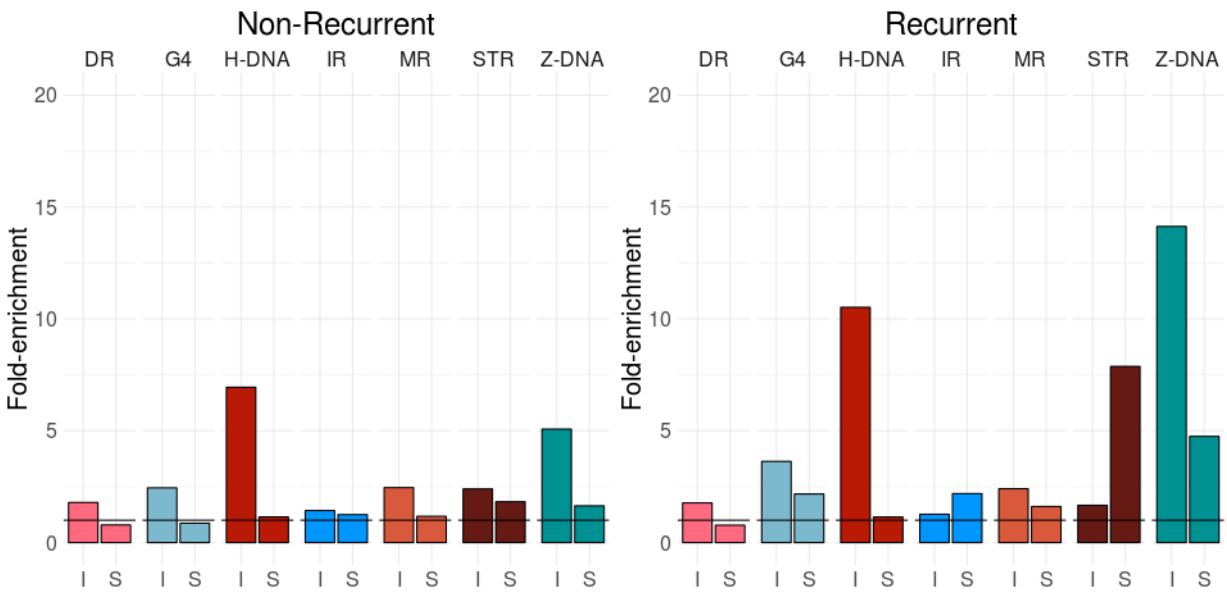

**Gastric**



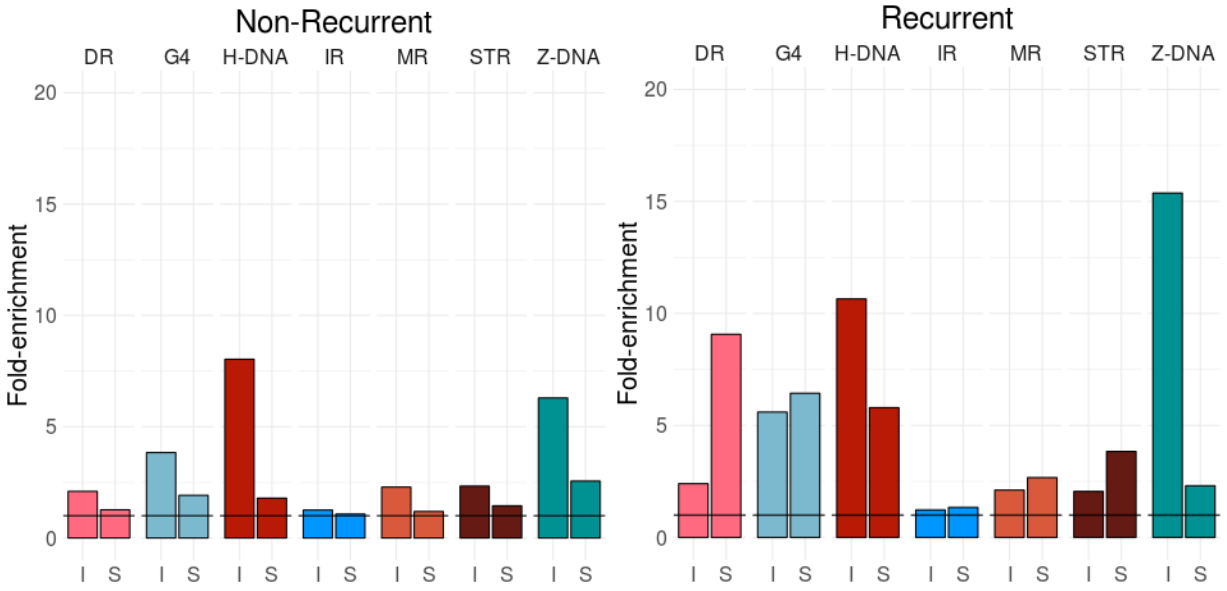

**Esophagus**




**Renal cell cancer**




**Malignant Lymphoma**




**Pedriatic Brain Tumour**

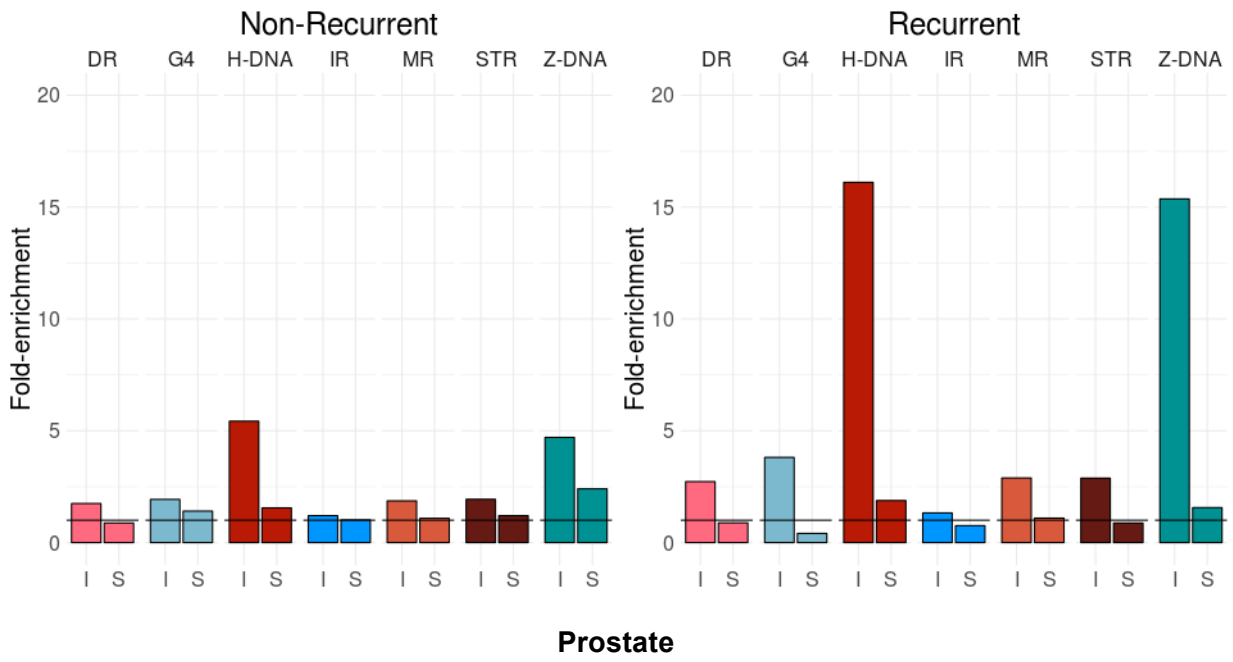



**Fig. S15: Recurrent mutations are more likely to overlap non-B DNA motifs.** Comparison of enrichment of non-recurrent mutations (left-hand panel) versus recurrent mutations (right hand panel) that are overlapping non-B-DNA motifs for indels (I) and substitutions (S) in human cancer. Mann-Whitney *U* test for: substitutions in ovarian cancer (p-value <0.001 for DR, STR, H-DNA and G4 and p-value <0.05 for IR and Z-DNA), indels in ovarian cancer (p-value < 0.001 for DR, MR, G4 and p-value <0.05 for H-DNA, IR), substitutions in pancreatic cancer (p-value <0.001 across all non-B DNA motifs), indels in pancreatic cancer (p-value <0.001 for DR, IR, H-DNA, Z-DNA, MR and G4), substitutions in liver cancer (p-value <0.001 for all non-B DNA motifs), indels in liver cancer (p-value <0.001 for DR, IR, H-DNA, Z-DNA, G4, MR), substitutions for gastric cancer (p-value <0.001 for IR, Z-DNA, MR, p-value <0.05 for G4, STR), indels for gastric cancer (p-value <0.001 for Z-DNA, p-value<0.05 for STR, DR), substitutions in esophageal cancer (p-value <0.001 for STR, IR, Z-DNA, G4 and p-value < 0.05 for H-DNA), in esophageal cancer for indels (p-value <0.001 for STR, IR, G4, MR and p-value <0.05 for H-DNA, Z-DNA), substitutions in renal cell cancer (p-value <0.001 for MR, STR, p-value < 0.05 in DR, H-DNA), indels in renal cell cancer (p-value <0.001 for H-DNA, Z-DNA, G4), substitutions in

malignant lymphoma (p-value <0.001 for STR, IR, Z-DNA, G4) and indels in malignant lymphoma (p-value <0.001 for STR, IR, H-DNA, Z-DNA, G4, MR), substitutions in pedriatic brain tumour (p-value <0.001 for DR, STR, H-DNA, G4, MR and p-value <0.05 for IRs), indels in pedriatic brain tumour (p-value<0.001 for H-DNA, Z-DNA, G4, MR and p-value <0.05 for DR, STR), substitutions in prostate (p-value <0.001 for STR, IR, G4, MR and p-value <0.05 for Z-DNA), indels in prostate (p-value <0.001 for DR, STR, IR, H-DNA, Z-DNA, G4).