

G Methylated

GTTA 3' bisulfite treatment->

GTTA 3'
CAAT 5'

Split top->

GTTA 3'
CAAT 5'

3' alignment->

GTTA 3'
GTTA 3'

Split bottom->

GTTA 3'
CAAT 5'

G Unmethylated

GTTA 3' bisulfite treatment->

GTTA 3'
TAAT 5'

Split top->

GTTA 3'
CAAT 5'

3' alignment->

GTTA 3'
ATT A 3'

Split bottom->

ATT A 3'
TAAT 5'

C Methylated

CTTA 3' bisulfite treatment->

CTTA 3'
GAAT 5'

Split top->

CTTA 3'
GAAT 5'

3' alignment->

CTTA 3'
CTTA 3'

Split bottom->

CTTA 3'
GAAT 5'

C Unmethylated

CTTA 3' bisulfite treatment->

TTTA 3'
GAAT 5'

Split top->

TTTA 3'
AAAT 5'

3' alignment->

TTTA 3'
CTTA 3'

Split bottom->

CTTA 3'
GAAT 5'

G>A mutation

GTTA 3'
CAAT 5'

mutation->

ATT A 3'
TAAT 5'

bisulfite treatment->

ATT A 3'
TAAT 5'

Split top->

ATT A 3'
TAAT 5'

3' alignment->

ATT A 3'
ATT A 3'

Split bottom->

ATT A 3'
TAAT 5'

G>T mutation

GTTA 3'
CAAT 5'

mutation->

TTTA 3'
AAAT 5'

bisulfite treatment->

TTTA 3'
AAAT 5'

Split top->

TTTA 3'
AAAT 5'

3' alignment->

TTTA 3'
TTTA 3'

Split bottom->

TTTA 3'
AAAT 5'

G>C mutation + Methylation

GTTA 3'

mutation->

CTTA 3'

bisulfite treatment->

CTTA 3'

CAAT 5'

GAAT 5'

GAAT 5'

Split top->

CTTA 3'
GAAT 5'

3' alignment->

CTTA 3'
CTTA 3'

Split bottom->

CTTA 3'
GAAT 5'

G>C mutation + Unmethylation

GTTA 3'
CAAT 5'

mutation->

CTTA 3'
GAAT 5'

bisulfite treatment->

TTTA 3'
GAAT 5'

Split top->

TTTA 3'
AAAT 5'

3' alignment->

TTTA 3'
CTTA 3'

Split bottom->

CTTA 3'
GAAT 5'

C>A mutation

CCTA 3'
GAAT 5'

mutation->

ATTA 3'
TAAT 5'

bisulfite treatment->

ATTA 3'
TAAT 5'

Split top->

ATTA 3'
TAAT 5'

3' alignment->

ATTA 3'
ATTA 3'

Split bottom->

ATTA 3'
TAAT 5'

C>T mutation


CTTA 3' mutation-> TTTA 3' bisulfite treatment-> TTTA 3'
 GAAAT 5' AAAT 5' AAAT 5'

Split top->

TTTA 3'

AAAT 5'

3' alignment->



TTTA 3'

TTTA 3'

Split bottom->

C>G mutation + Methylation

Split top->

TTA 3'

CAAT 5'

3' alignment->

G	T	T	A	3'
G	T	T	A	3'

Split bottom->

GTTA 3'

CAAT 5'

C>G mutation + Unmethylation

The diagram illustrates the conversion of a CpG site to a TpG site. It starts with a DNA sequence: a top strand with 'CTTA 3'' and a bottom strand with 'GAAT 5''. A red box highlights the 'C' in the top strand and the 'G' in the bottom strand, which form a CpG dinucleotide. An arrow labeled 'mutation->' points to the next state where the top strand is 'GTTA 3'' and the bottom strand is 'CAAT 5'' (the 'G' has become a 'C'). A second arrow labeled 'bisulfite treatment->' points to the final state where the top strand is 'GTTA 3'' and the bottom strand is 'TAAT 5'' (the 'C' has been converted to a 'T').

Split top->

TTA 3'

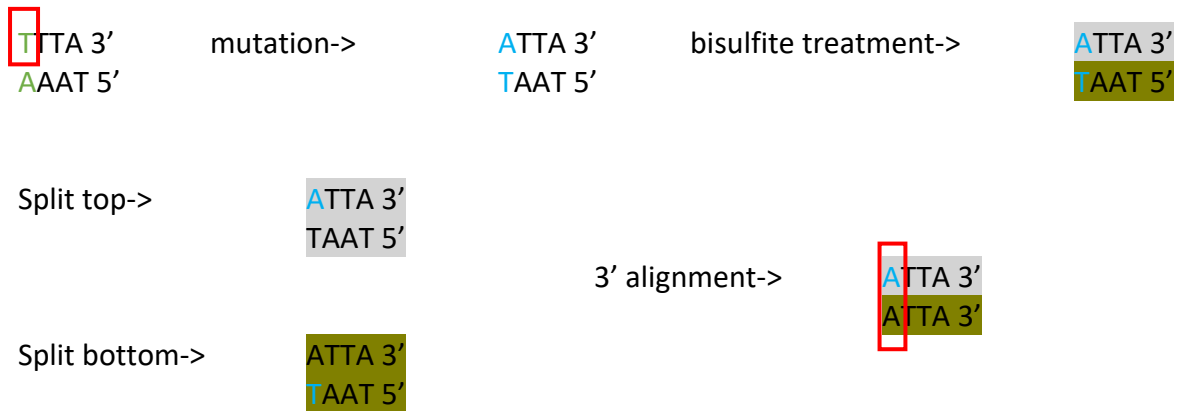
CAAT 5'

3' alignment->

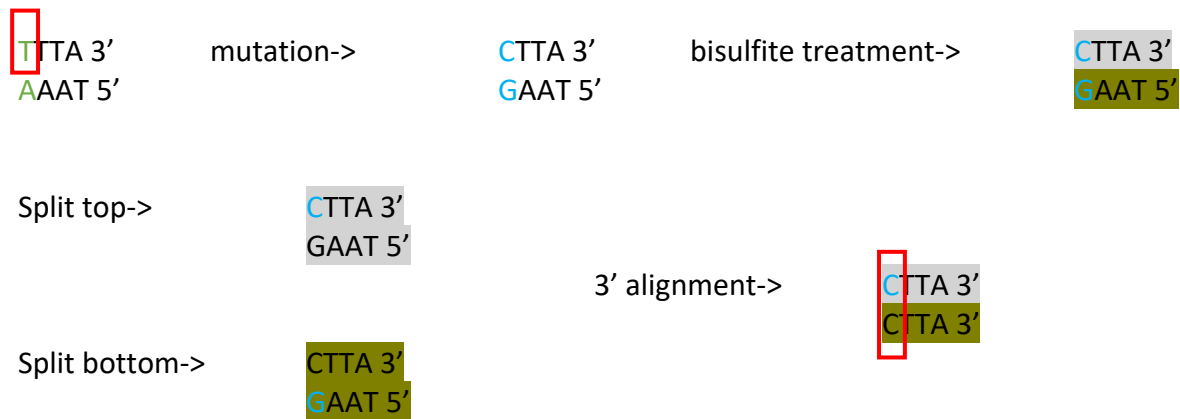
G	T	T	A	3'
A	T	T	A	3'

Split bottom->

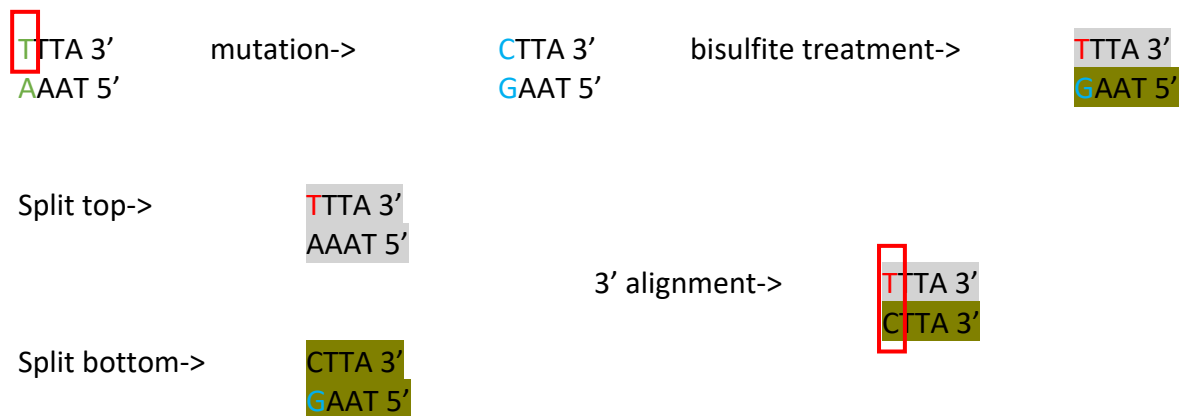
T>A mutation



T>C mutation + Methylation



T>C mutation + Unmethylation



T>G mutation + Methylation

TTTA 3'
AAAT 5'

mutation->

GTTA 3'
CAAT 5'

bisulfite treatment->

GTTA 3'
CAAT 5'

Split top->

GTTA 3'
CAAT 5'

3' alignment->

GTTA 3'

Split bottom->

GTTA 3'
CAAT 5'

T>G mutation + Unmethylation

TTTA 3'
AAAT 5'

mutation->

GTTA 3'
CAAT 5'

bisulfite treatment->

GTAA 3'
TAAT 5'

Split top->

GTTA 3'
CAAT 5'

3' alignment->

GTTA 3'
ATTA 3'

Split bottom->

ATTA 3'
TAAT 5'

A>T mutation

ATTA 3'
 TAAAT 5'

mutation->

TTTA 3'
AAAT 5'

bisulfite treatment->

TTTA 3'
AAAT 5'

Split top->

TTTA 3'
AAAT 5'

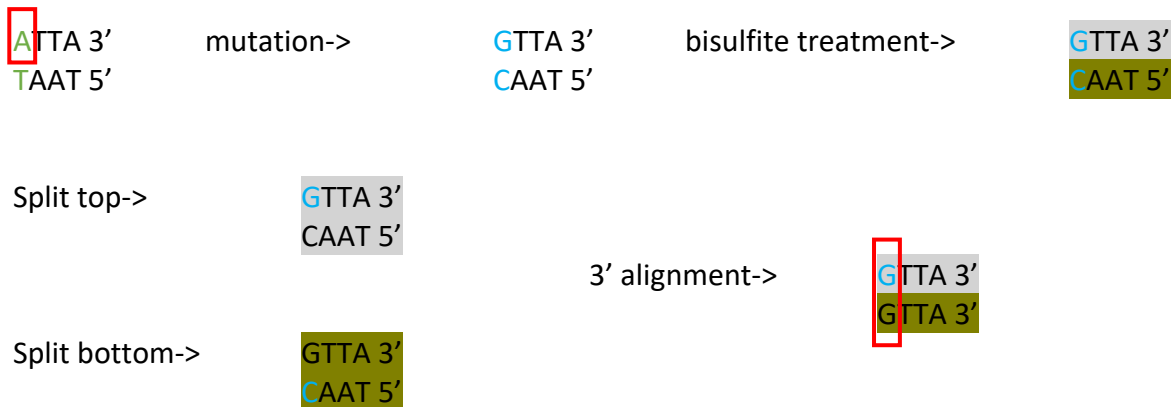
3' alignment->

TTTA 3'

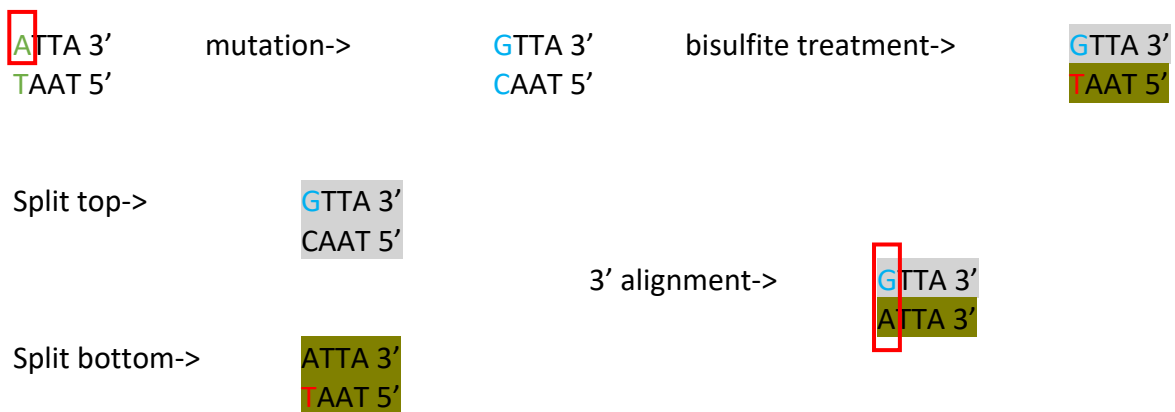
Split bottom->

TTTA 3'
AAAT 5'

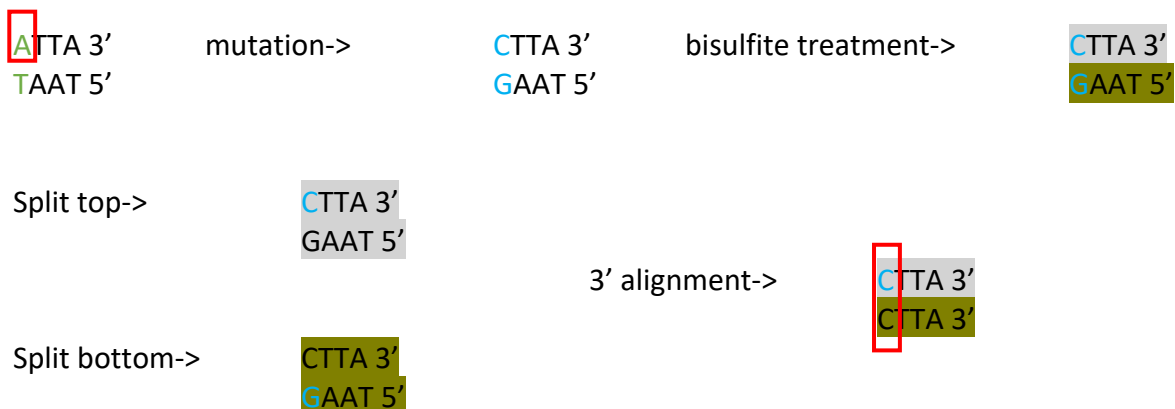
A>G mutation + Methylation



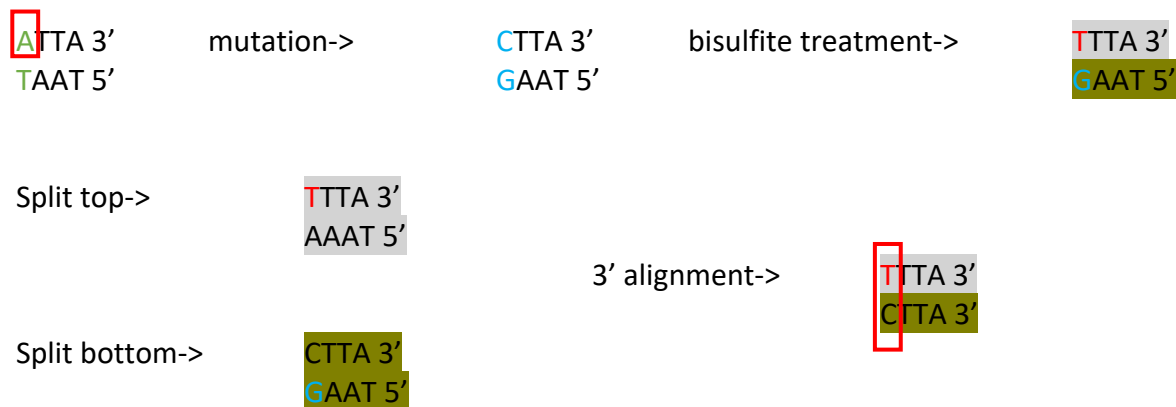
A>G mutation + Unmethylation



A>C mutation + Methylation



A>C mutation + Unmethylation



Supplementary Figure 1: Permutations of bases with possible mutations and/or methylation status. Each scenario is split into “top” and “bottom” to refer to the top strand and bottom strand which are separated and amplified during the PCR steps. “3’ alignment” represents the method in which the bases at each position are compared to the reference genome. Bases in green font refer to the initial pair being analysed. Bases in blue font refer to the pair after mutation has happened. Bases in red font refer to the change happening due to bisulfite treatment. The red box highlights the comparison being made when computing variant calling due to 3’ alignment to the reference genome. The grey shading indicates the DNA strand pair stemming off the initial forward strand, and the green shading indicates the same for the initial reverse strand. Only cytosines can be methylated, however guanine bases are complementary of cytosines and hence should be considered both in the methylated and unmethylated scenarios.