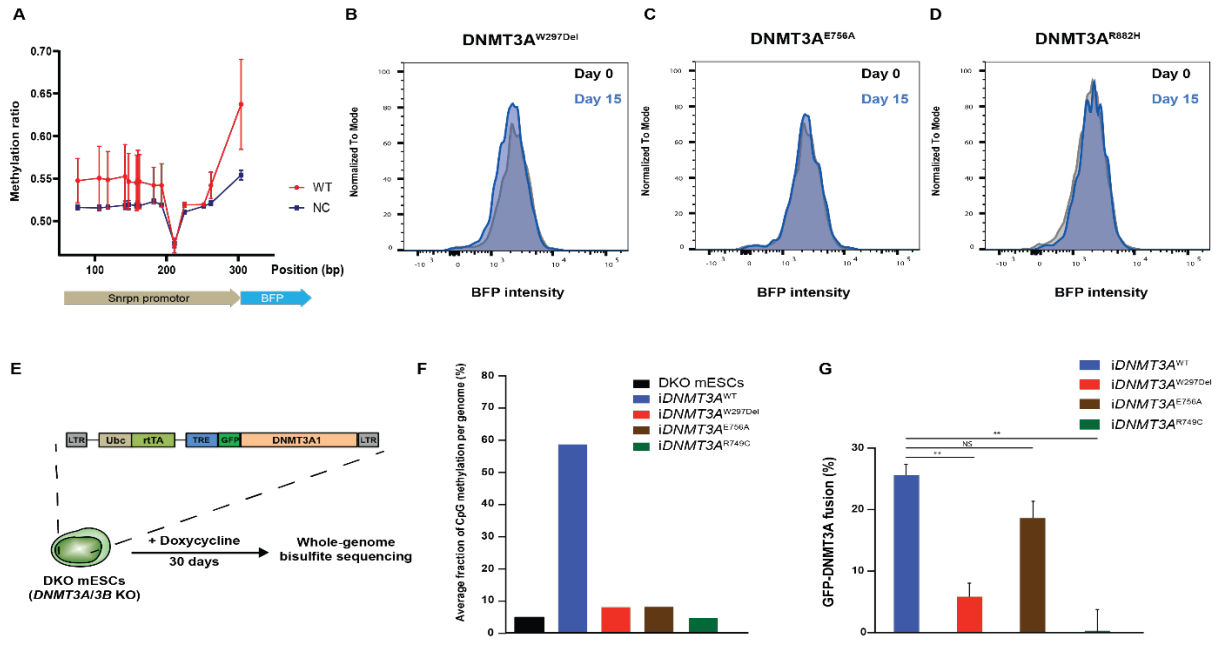
Supplementary data

Systematic profiling of *DNMT3A* variants reveals protein instability mediated by the DCAF8 E3 ubiquitin ligase adaptor

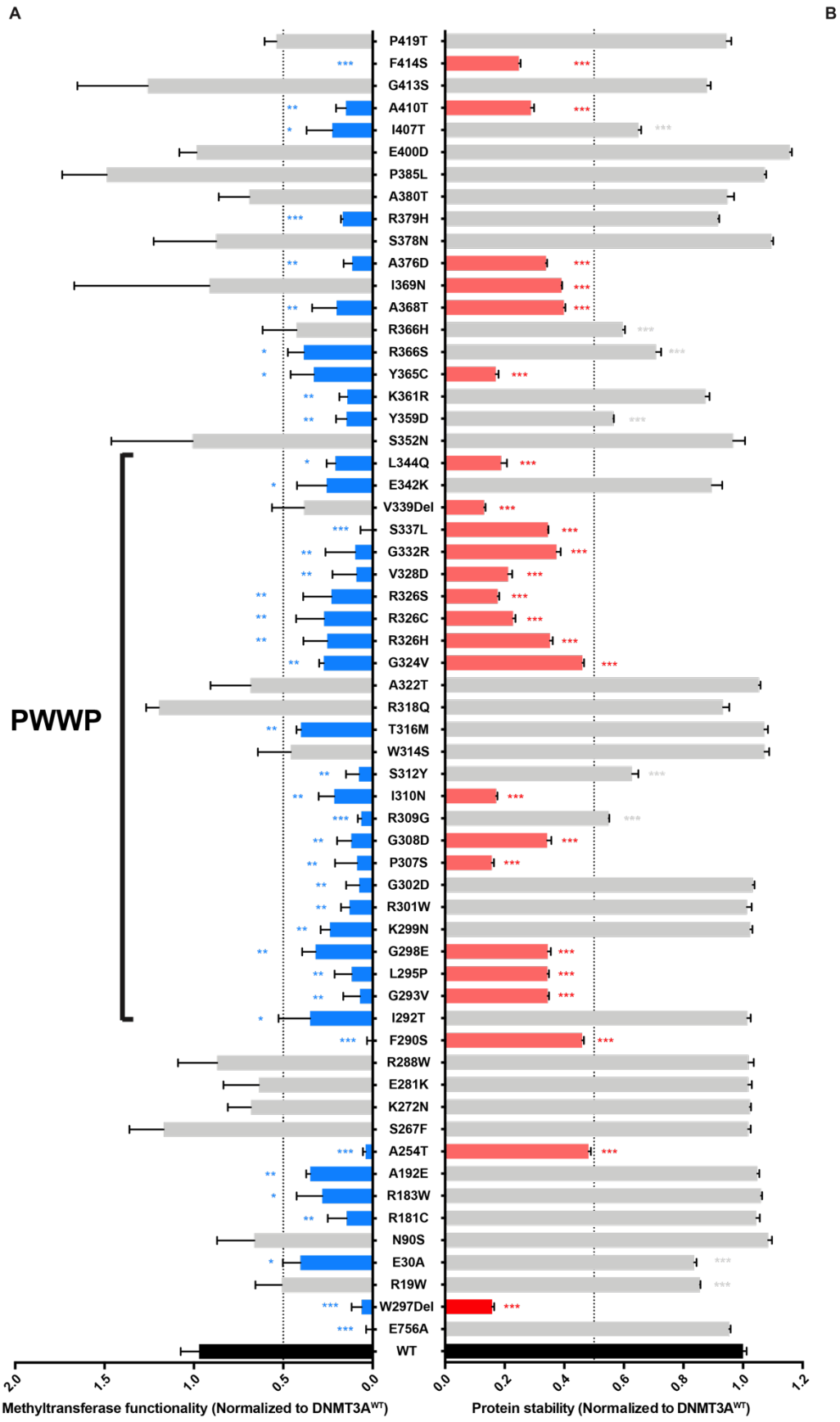
Authors and affiliations:

Yung-Hsin Huang[1-3,16], Chun-Wei Chen[2-4], Venkatasubramaniam Sundaramurthy[2,3,5], Mikołaj Słabicki[6], Dapeng Hao[7,17], Caroline J. Watson[8], Ayala Tovy[2-3], Jaime M. Reyes[2,3,5], Olga Dakhova[9], Brielle R. Crovetti[2-3], Christina Galonska[10], Minjung Lee[11], Lorenzo Brunetti[2-3], Yubin Zhou[11], Katrina Tatton-Brown[12], Yun Huang[11], Xiaodong Cheng[13], Alexander Meissner[10], Peter J.M. Valk[14], Lionel Van Maldergem[15], Mathijs A. Sanders[14], Jamie R. Blundell[8], Wei Li[7,18], Benjamin L. Ebert[6], Margaret A. Goodell[1-5,*]

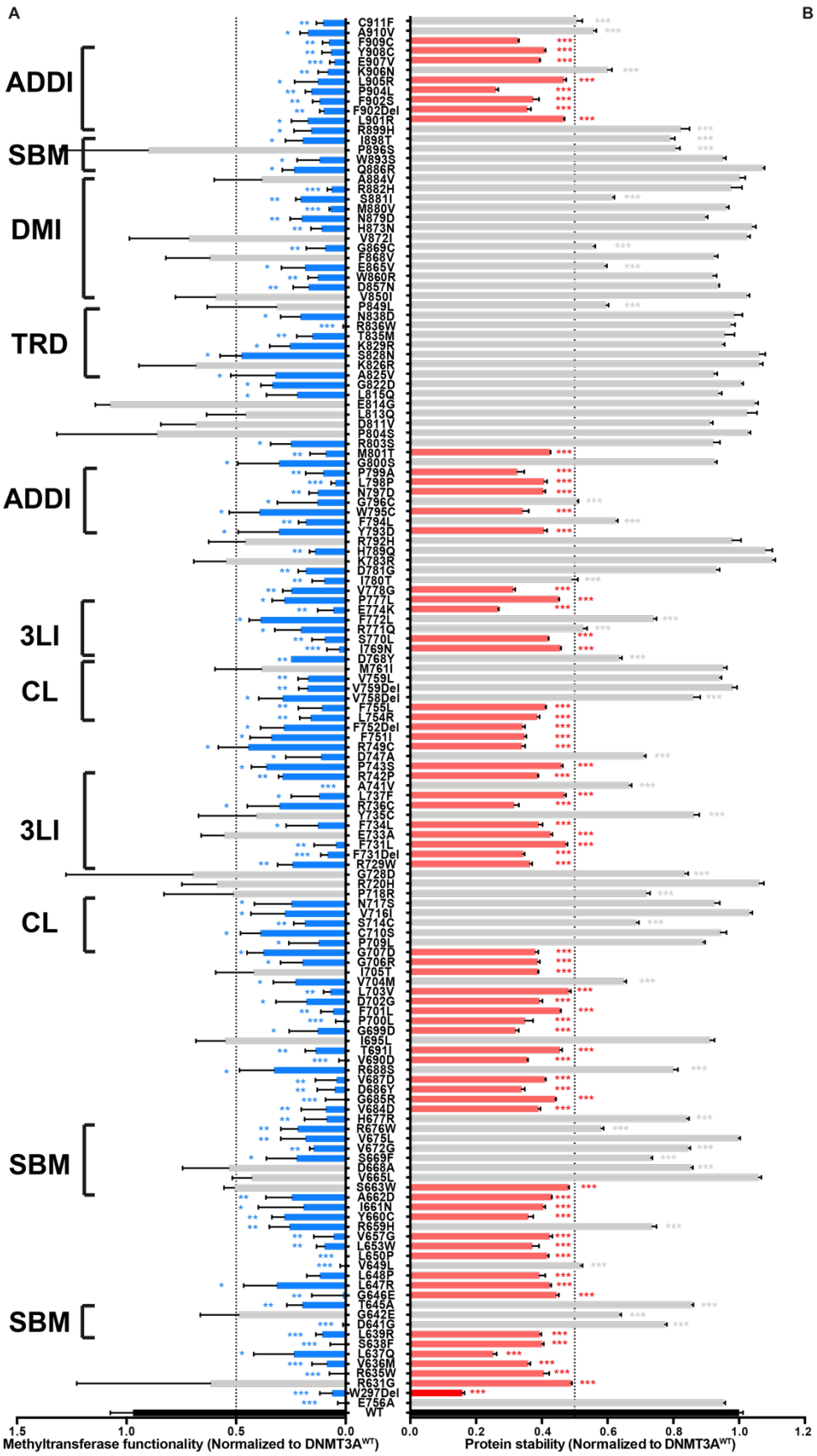


Supplemental Figure 1: Methyltransferase activity assay and *DNMT3A* mutant functional assay using whole-genome bisulfite sequencing in methylation-deficient ESCs. (A) Targeted methylation assay across *Snrpn* promotor region. DNMT3A^{WT}-transduced cells (red) demonstrated an increased methylation level compared to the vector-only negative control (blue). (B) Methyltransferase activity assay. The graph depicts blue fluorescence intensity in the DNMT3A^{W297Del}-transduced cells as measured by flow cytometry on day 0 (grey) and day 15 (blue). (C) Methyltransferase activity assay. The graph depicts blue fluorescence intensity in the DNMT3A^{E756A}-transduced cells as measured by flow cytometry on day 0 (grey) and day 15 (blue). (D) Methyltransferase activity assay. The graph depicts blue fluorescence intensity in the DNMT3A^{R882H}-transduced cells as measured by flow cytometry on day 0 (grey) and day 15 (blue). (E) Schematic of experimental procedure. *DNMT3A* variants were fused to GFP in a doxycycline-inducible lentiviral vector and expressed in DKO mESCs for 30 days prior to whole genome bisulfite sequencing (WGBS). These ES cells were knocked out for *Dnmt3a/b* and

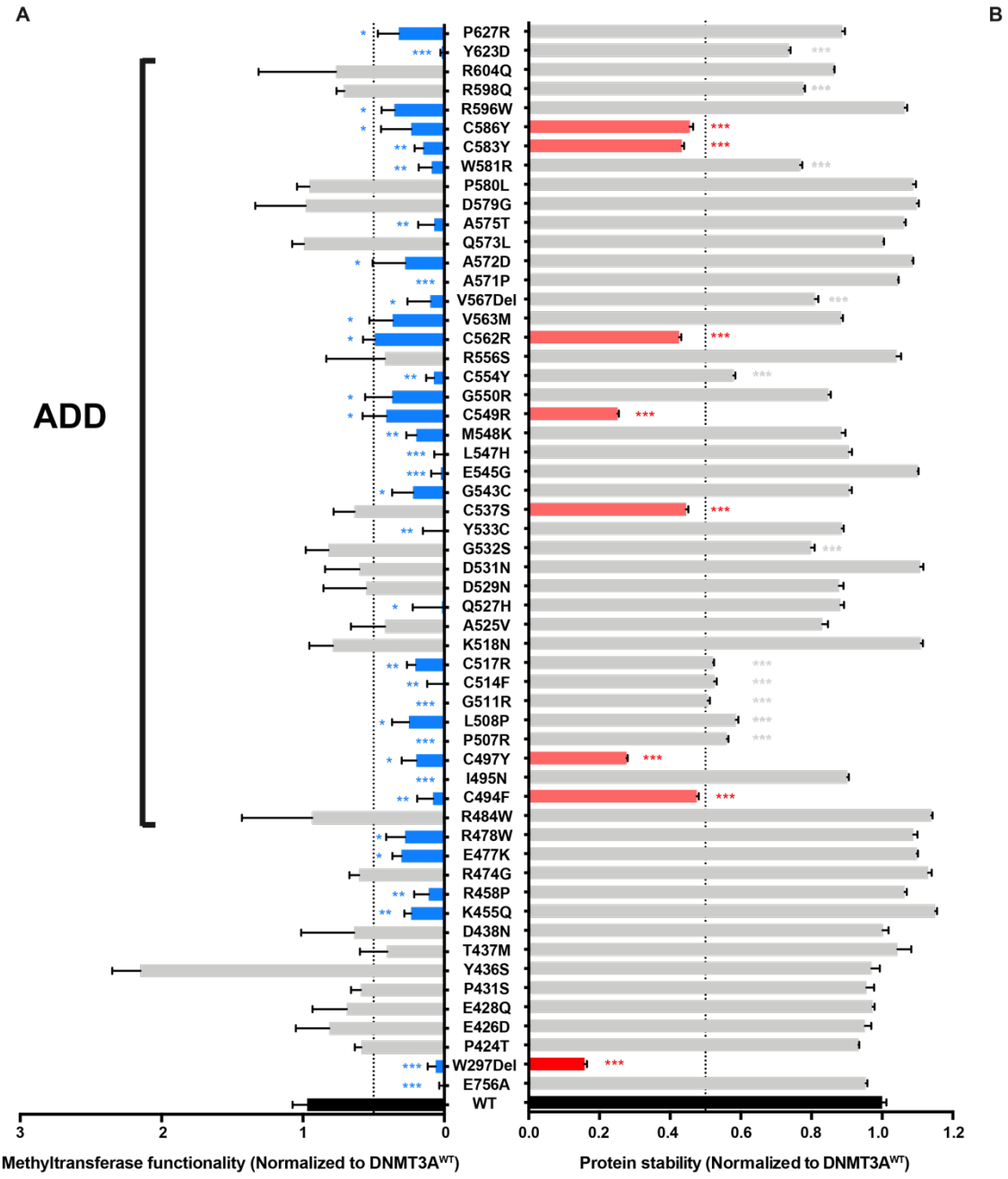
knocked down for *Dnmt1* to achieve very low background DNA methylation levels. DNMT1 expression was then restored (23) which allows DNA methylation to accumulate and stabilize when *de novo* DNA methyltransferases are re-introduced. (F) Average DNA methylation ratio across millions of CpGs measured by WGBS in DKO mESCs. (n=1) (G) The percentage of cells positive for GFP-DNMT3A fusion protein measured by flow cytometry with three replicates. All *DNMT3A* mutations shown here are compared to DNMT3A ^{WT} by unpaired t test. Data are presented as mean ± s.e.m. * P<0.05, ** P<0.01 and *** P< 0.001.


A
B
PWWP
P419T
F414S
G413S
A410T
I407T
E400D
P385L
A380T
R379H
S378N
A376D
I369N
A368T
R366H
R366S
Y365C
K361R
Y359D
S352N
L344Q
E342K
V339Del
S337L
G332R
V328D
R326S
R326C
R326H
G324V
A322T
R318Q
T316M
W314S
S312Y
I310N
R309G
G308D
P307S
G302D
R301W
K299N
G298E
L295P
G293V
I292T
F290S
R288W
E281K
K272N
S267F
A254T
A192E
R183W
R181C
N90S
E30A
R19W
W297Del
E756A
WT
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Methyltransferase functionality (Normalized to DNMT3A^WT)
Protein stability (Normalized to DNMT3A^WT)

Supplemental Figure 2: Methyltransferase activity and protein stability assay around the PWWP domain. (A) The graph depicts the methyltransferase activity of DNMT3A missense mutations around the PWWP domain as measured by flow cytometry and as normalized with BFP-negative % of DNMT3A^{WT} infected cells at post-infection day 15. (B) The graph depicts the protein stability represented by the ratio of normalized DNMT3A-GFP and DsRed expression in the bicistronic construct. All *DNMT3A* mutations shown here were chosen from the TBRS, CH and COSMIC databases and are compared to DNMT3A ^{WT} by unpaired t test. Data are presented as mean ± s.e.m. * P<0.05, ** P<0.01 and *** P< 0.001.

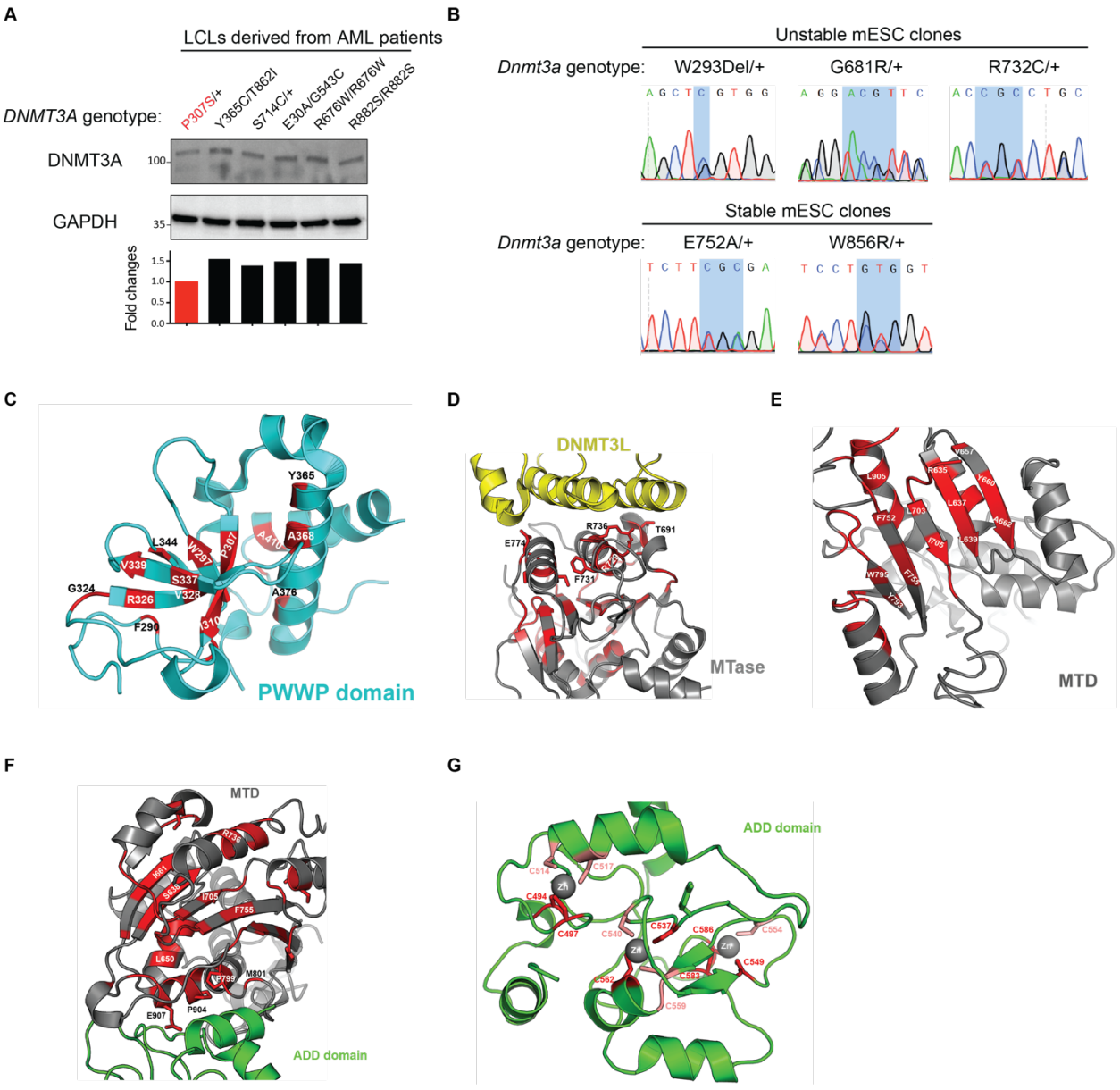

A
B
ADDI
SBM
DMI
TRD
ADDI
3LI
CL
3LI
CL
SBM
SBM
C911F
A910V
F909C
Y908C
E907V
K906N
L905R
P904L
F902S
F902Del
L901R
R899H
I898T
P896S
W893S
Q886R
A884V
R882H
S881I
M880V
N879D
H873N
V872I
G869C
F868V
E865V
W860R
D857N
V850I
P849L
N838D
R836W
T835M
K829R
S828N
K826R
A825V
G822D
L815Q
E814G
L813Q
D811V
P804S
R803S
M801T
G800S
P799A
L798P
N797D
G796C
W795C
F794L
Y793D
R792H
H789Q
K783R
D781G
I780T
V778G
P777L
E774K
F772L
R771Q
S770L
I769N
D768Y
M761I
V759L
V759Del
V758Del
F755L
L754R
F752Del
F751I
R749C
D747A
P743S
R742P
A741V
L737F
R736C
Q735C
F734L
E733A
F731L
F731Del
R729W
G728D
R720H
P718R
N717S
V716I
S714C
C710S
P709L
G707D
G706R
I705T
V704M
L703V
D702G
F701L
P700L
G699D
I695L
T691I
V690D
R688S
V687D
D686Y
G685R
V684D
H677R
R676W
V675L
V672G
S669F
D668A
V665L
S663W
A662D
I661N
V660C
R659H
V657G
L653W
L650P
V649L
L648P
L647R
G646E
T645A
G642E
D641G
L639R
S638F
L637Q
V636M
R635W
R631G
W297Del
E756A
WT
1.5
1.0
0.5
0.0
Methyltransferase functionality (Normalized to DNMT3A^WT)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Protein stability (Normalized to DNMT3A^WT)

Supplemental Figure 3: Methyltransferase activity and protein stability assay at the MTase domain. (A) The graph depicts the methyltransferase activity of *DNMT3A* missense mutations around the MTase domain as measured by flow cytometry and as normalized with BFP-negative % of $DNMT3A^{WT}$ infected cells at post-infection day 15. (B) The graph depicts the protein stability represented by the ratio of normalized DNMT3A-GFP and DsRed expression in the bicistronic construct. SBM represents the substrate (S-adenosyl-L-methionine) binding motif; CL represents the catalytic loop; 3LI represents the interface of DNMT3A-DNMT3L; ADDI represents the interface between the ADD domain and the catalytic domain; TRD represents the target recognition domain (TRD) and DMI represents the DNMT3A homodimer interfaces. All *DNMT3A* mutations shown here were chosen from the TBRS, CH and COSMIC databases and are compared to $DNMT3A^{WT}$ by unpaired t test. Data are presented as mean ± s.e.m. * $P<0.05$, ** $P<0.01$ and *** $P< 0.001$.



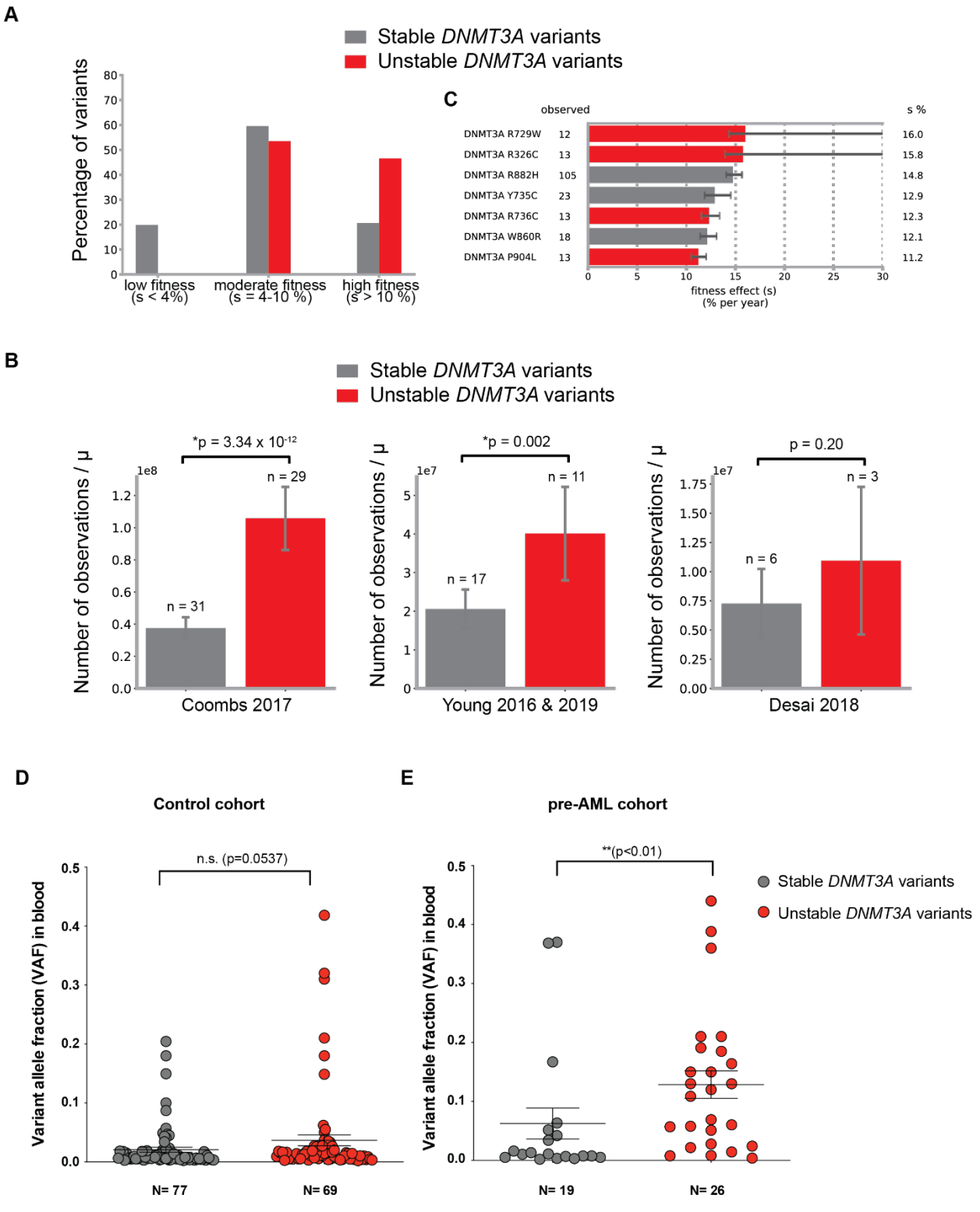
Supplemental Figure 4: Methyltransferase activity and protein stability assay around the ADD domain. (A) The graph depicts the methyltransferase activity of *DNMT3A* missense mutations around the ADD domain as measured by flow cytometry and as normalized with BFP-negative % of DNMT3A^{WT} infected cells at post-infectoin day 15. (B) The graph depicts the protein stability represented by the ratio of normalized DNMT3A-GFP and DsRed expression in

the bicistronic construct. All *DNMT3A* mutations shown here were chosen from the TBRS, CH and COSMIC databases and are compared to DNMT3A ^{WT} by unpaired t test. Data are presented as mean ± s.e.m. * P<0.05, ** P<0.01 and *** P< 0.001.



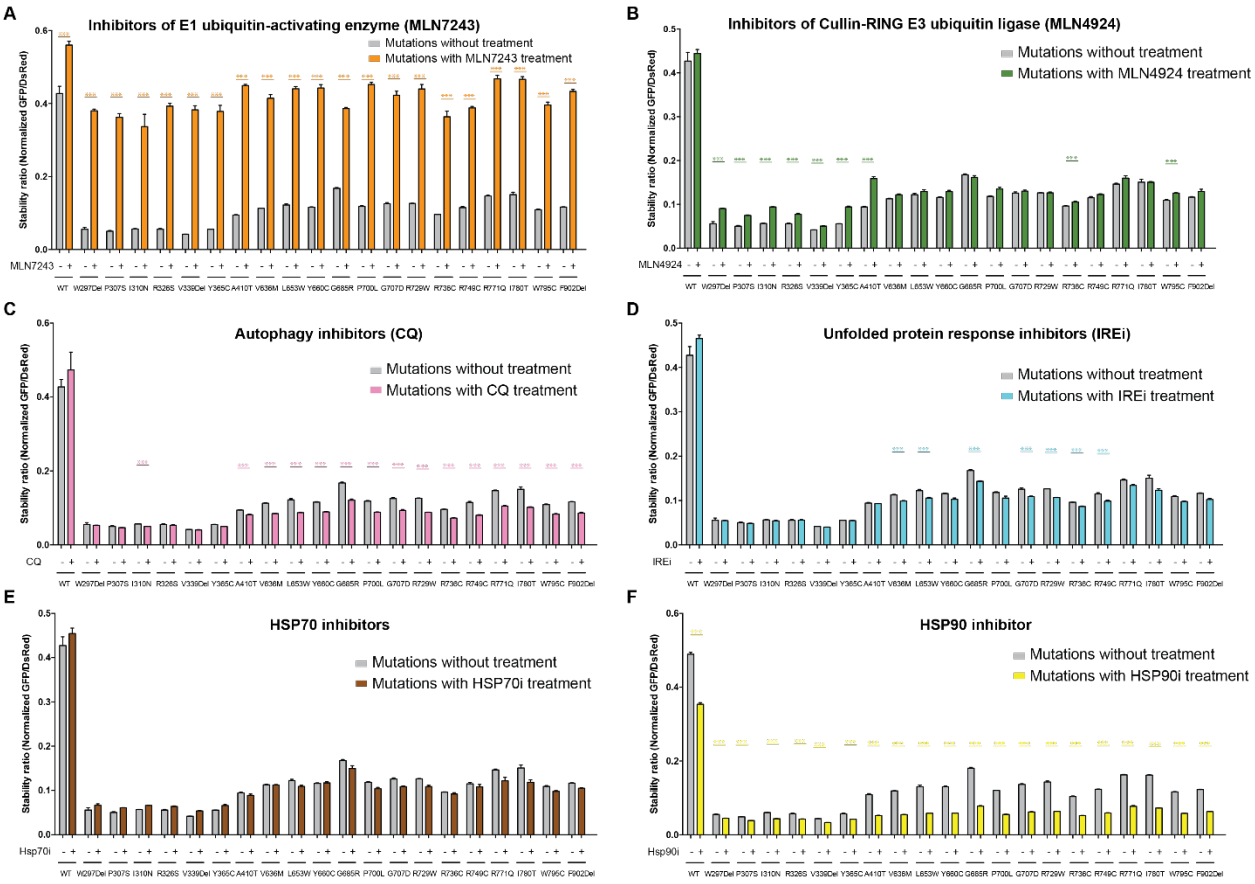
Supplemental Figure 5: Structural insights into *DNMT3A* missense mutations causing protein instability. (A) DNMT3A protein expression (by western blot) in lymphoblastoid cell lines (LCLs) derived from acute myeloid leukemia (AML) patients. (B) Sanger sequencing tracks showing the heterozygous knock-ins and the successful generation of unstable mutants (*Dnmt3a*$^{W293Del/+}$, *Dnmt3a*$^{G681R/+}$, and *Dnmt3a*$^{R732C/+}$) and stable mutants (*Dnmt3a*$^{E752A/+}$ and *Dnmt3a*$^{W856R/+}$). (C) The structural modeling depicts the unstable mutations (red) located in the

PWWP domain (cyan). The structure of the PWWP domain was from PDB 3llr. (D) The structural modeling depicts the unstable mutations (red) located in the methyltransferase domain (MTD) (grey) interacting withDNMT3L (yellow). (E, F) The structural modeling depicts the unstable mutations (red) located in the methyltransferase domain (MTD). The methyltransferase domain is labeled in grey and the ADD domain is labeled in green. (G) The structural modeling depicts the unstable mutations located in the ADD domain labeling as red. The zinc atoms are labeled in grey and the ADD domain is labeled in green. The cysteine residues labeled in salmon are essential for zinc finger formation. The structure of the MTase and DNMT3L was from PDB 4U7T. See also supplemental Figures 1-3.

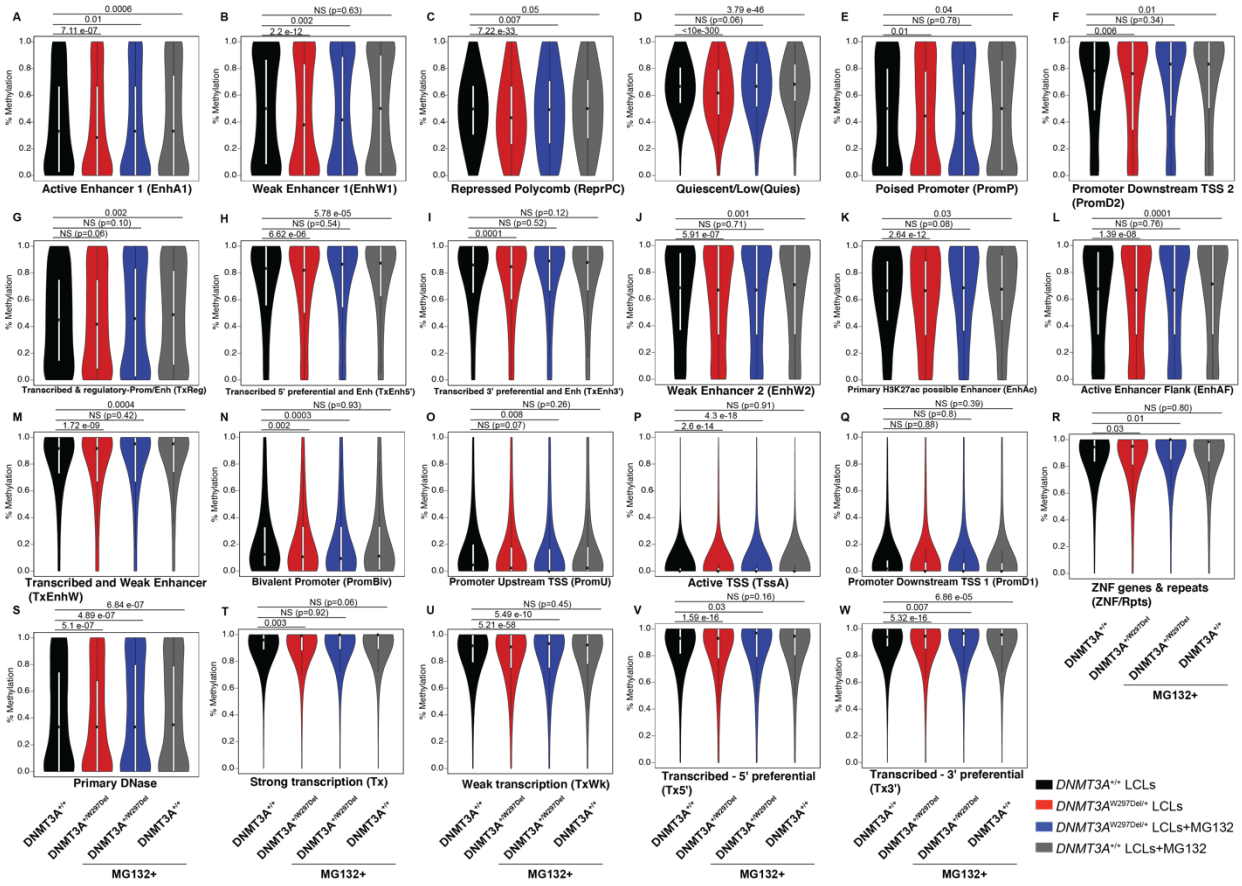


Supplemental Figure 6: *DNMT3A* variants conferring protein instability are associated with increased clonal expansion in patients. (A) Inferred distribution of fitness effects for

stable and unstable variants from the maximum likelihood fits after dividing fitness up into three fitness classes (low <4% per year, moderate 4%-10% per year and high >10% per year). A higher proportion of unstable variants confer moderate-to-high fitness (red) relative to stable variants (grey) (B) Number of observed stable and unstable DNMT3A variants after controlling for mutation rate shows that unstable variants are significantly enriched relative to stable variants ($p = 2\times10^{-13}$, Poisson exact test followed by Fisher's method, with 0.05 significance point at p=0.0019 for 3 combined p-values). Error bars are one standard deviation assuming Poisson sampling noise. (C) Fitness effects of 7 variants inferred from VAF-density histograms of individual variants (29). Error bars represent 95% confidence intervals for the inferred fitness effect (s). (D) The graph depicts the variant allele fractions (VAFs) in the blood of the cohorts of the control group in two categories. *DNMT3A* mutations not affecting protein stability are labeled in gray; *DNMT3A* mutations causing protein instability are labeled in red. (E) The graph depicts a similar categorization of the variant allele fractions (VAFs) in the blood of the cohorts of the pre-AML group (33).



Supplemental Figure 7: Inhibitor of E1 ubiquitin-activating enzymes can rescue DNMT3A mutant protein expression while HSP90i further destabilizes DNMT3A mutant protein expression. All graphs depict the stability ratio of mean fluorescence intensity (MFI) of DNMT3A-GFP versus MFI of DsRed with the indicated inhibitor treatments. Treated with in (A) inhibitors of E1 ubiquitin-activating enzymes, in (B) with inhibitors of Cullin-RING E3 ubiquitin ligases, in (C) with inhibitors of autophagy (CQ), in (D) with inhibitors of the unfolded protein response, in (E) with a HSP70i inhibitor, in (F) with an HSP90i inhibitor. In all cases, flow cytometry was performed 48 hours after transfection.



Supplemental Figure 8: Active/weak enhancers, heterochromatin, poised promotor and regions with repressed polycomb complexes were sensitive while regions closed to TSS and transcribed regions were inert to proteasome inhibitors administration. All graphs depict the methylation ratio of the indicated genomic regions: (A) 8869 active enhancer regions as defined in the epigenomic roadmap project, (B) 15029 weak enhancer regions, (C) 13701 regions with repressed polycomb complexes, (D) 101293 quiescent regions, (E) 6312 poised promotor regions, (F) 6085 regions downstream of promotors, (G) 7870 transcribed and regulatory promotor/enhancer regions, (H) 8904 5' preferentially transcribed enhancers, (I) 7231 3' preferentially transcribed enhancers, (J) 35253 weak enhancers, (K) 7540 putative H3K27ac enhancers, (L) 13756 active enhancer flanking regions, (M) 19613 transcribed and weak enhancers, (N) 7053 bivalent promotors, (O) 15379 promotors upstream of the TSS, (P) 6305

active TSS regions, (Q) 10492 promotors downstream of the TSS, (R) 2308 zinc finger genes and repeats, (S) 17820 primary DNase sensitive regions, (T) 12003 regions with strong transcription, (U) 52563 regions with weak transcription, (V) 22059 5' preferential transcribed regions, (W) 39659 3' preferential transcribed regions. All regions were as defined in the epigenome roadmap project.