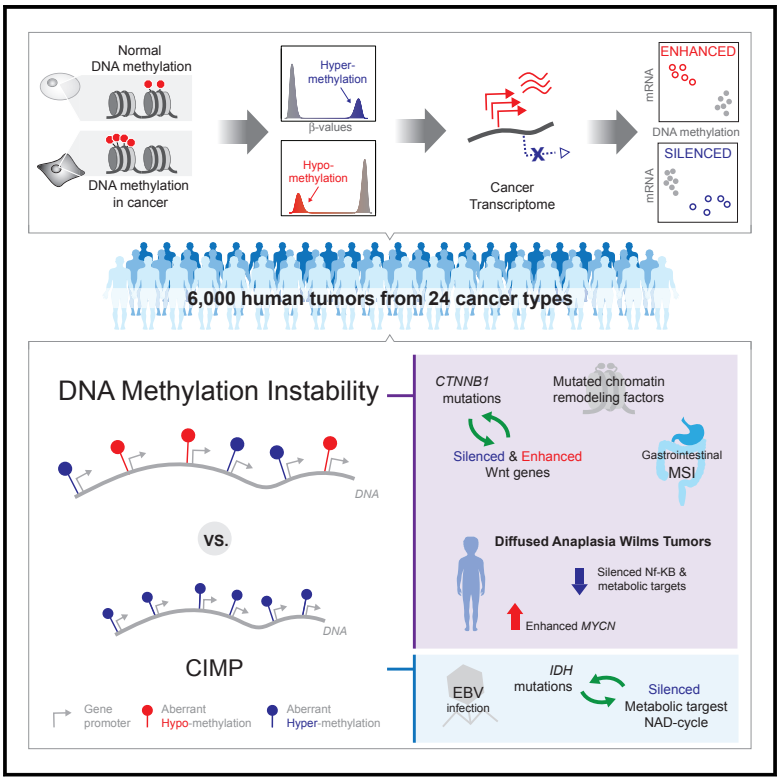
Cell Reports

Resource

Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors

Graphical Abstract



Highlights

- Mutations in WNT and chromatin modifiers are associated with DNA methylation instability
- Concurrent genetic and epigenomic changes converge on the same pathways
- Cancer-germline antigen expression correlates with response to anti-PD1
- Diffuse anaplasia Wilms tumors exhibit DNA methylation instability

Authors

Sadegh Saghafinia, Marco Mina, Nicolo Riggi, Douglas Hanahan, Giovanni Ciriello

Correspondence

giovanni.ciriello@unil.ch

In Brief

Saghafinia et al. present an algorithmic approach to identify cancer-associated DNA methylation changes affecting gene expression. By analyzing >6,000 adult and pediatric tumors, the authors identify numerous DNA methylation changes at gene promoters that coalesce into a few pathways and that were associated with patient prognosis and response to therapy.

Saghafinia et al., 2018, Cell Reports *25*, 1066–1080
October 23, 2018
https://doi.org/10.1016/j.celrep.2018.09.082

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors

Sadegh Saghafinia,[1,3,4,5] Marco Mina,[1,3,5] Nicolo Riggi,[2] Douglas Hanahan,[4] and Giovanni Ciriello[1,3,6,*]
[1]Department of Computational Biology, University of Lausanne (UNIL), Lausanne, Switzerland
[2]Department of Experimental Pathology, University of Lausanne (UNIL), Lausanne, Switzerland
[3]Swiss Institute of Bioinformatics (SIB), University of Lausanne (UNIL), Lausanne, Switzerland
[4]Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: giovanni.ciriello@unil.ch
https://doi.org/10.1016/j.celrep.2018.09.082

SUMMARY

The discovery of cancer-associated alterations has primarily focused on genetic variants. Nonetheless, altered epigenomes contribute to deregulate transcription and promote oncogenic pathways. Here, we designed an algorithmic approach (RESET) to identify aberrant DNA methylation and associated *cis*-transcriptional changes across >6,000 human tumors. Tumors exhibiting mutations of chromatin remodeling factors and Wnt signaling displayed DNA methylation instability, characterized by numerous hyper- and hypo-methylated loci. Most silenced and enhanced genes coalesced in specific pathways including apoptosis, DNA repair, and cell metabolism. Cancer-germline antigens (CG) were frequently epigenomically enhanced and their expression correlated with response to anti-PD-1, but not anti-CTLA4, in skin melanoma. Finally, we demonstrated the potential of our approach to explore DNA methylation changes in pediatric tumors, which frequently lack genetic drivers and exhibit epigenomic modifications. Our results provide a pan-cancer map of aberrant DNA methylation to inform functional and therapeutic studies.

INTRODUCTION

The cell epigenome provides a fundamental infrastructure to coordinate spatiotemporal gene expression and cell-type-specific patterns (Shen and Laird, 2013). This organization ranges from directly decorating DNA molecules and histone proteins (Jones, 2012; Zhou et al., 2011) to shaping long-range DNA contacts between regulatory elements (Bonev and Cavalli, 2016; Fatica and Bozzoni, 2014; Long et al., 2016). Acting through multiple hierarchical layers, epigenetic modifications enable and repress transcription and stably determine cell identity. Cancer cells exhibit profound modifications of this epigenetic infrastructure (Baylin and Jones, 2011; Esteller, 2008), starting from its foundation, DNA methylation (Esteller, 2007; Klutstein et al., 2016). Aberrant DNA methylation in cancer has primarily been observed in the form of global hypo-methylation within intergenic regions, especially those enriched for Alu and LINE-1 repeats (Ehrlich, 2009), punctuated by hyper-methylation of CpG dense regions, referred to as CpG islands (CGI). Unlike most of DNA hypo-methylated regions, hyper-methylated CGI are frequently located at gene promoters and have been associated with silencing of tumor suppressors (Esteller, 2007) and incomplete differentiation (Widschwendter et al., 2007), directly linking DNA methylation changes to oncogenic transformation. In turn, the development of pharmacological inhibitors of DNA methylation has provided novel therapeutic opportunities (Swisher et al., 2017).

Aberrant DNA methylation in human cancers has mostly been investigated through two main approaches: (1) analysis of global DNA methylation patterns (Noushmehr et al., 2010; Toyota et al., 1999) and differentially methylated regions (Amabile et al., 2015), and (2) identification of mRNA silencing of genes of interest by hyper-methylation (Esteller, 2007; Cancer Genome Atlas Research Network, 2012). The first approach has gained momentum with the application of high-throughput arrays to thousands of human tumors. Importantly, clustering of cancer samples based on variably DNA methylated loci led to the identification of molecularly and clinically relevant subtypes, most notably tumors characterized by frequent hyper-methylation at CGI, also known as CpG island methylator phenotype (CIMP) (Issa, 2004). Recurrently hyper-methylated loci in cancer were found enriched for polycomb targets involved in cell differentiation, thus likely to sustain a stem cell phenotype (Widschwendter et al., 2007), and associated with a mitotic signature upregulated in cancer and pre-cancerous lesions (Yang et al., 2016). The oncogenic role of specific targets within hyper-methylated regions has been highlighted by targeted profiling of known and candidate tumor suppressors. Gene silencing by promoter hyper-methylation have been found for established cancer genes, such as *VHL* (Herman et al., 1994) and *CDKN2A* (Cancer Genome Atlas Research Network, 2012) and has provided markers of therapeutic response, such as for *MGMT* (Weller et al., 2010) and *BRCA1* (Cancer Genome Atlas Research Network, 2011). More recently, unbiased approaches have been

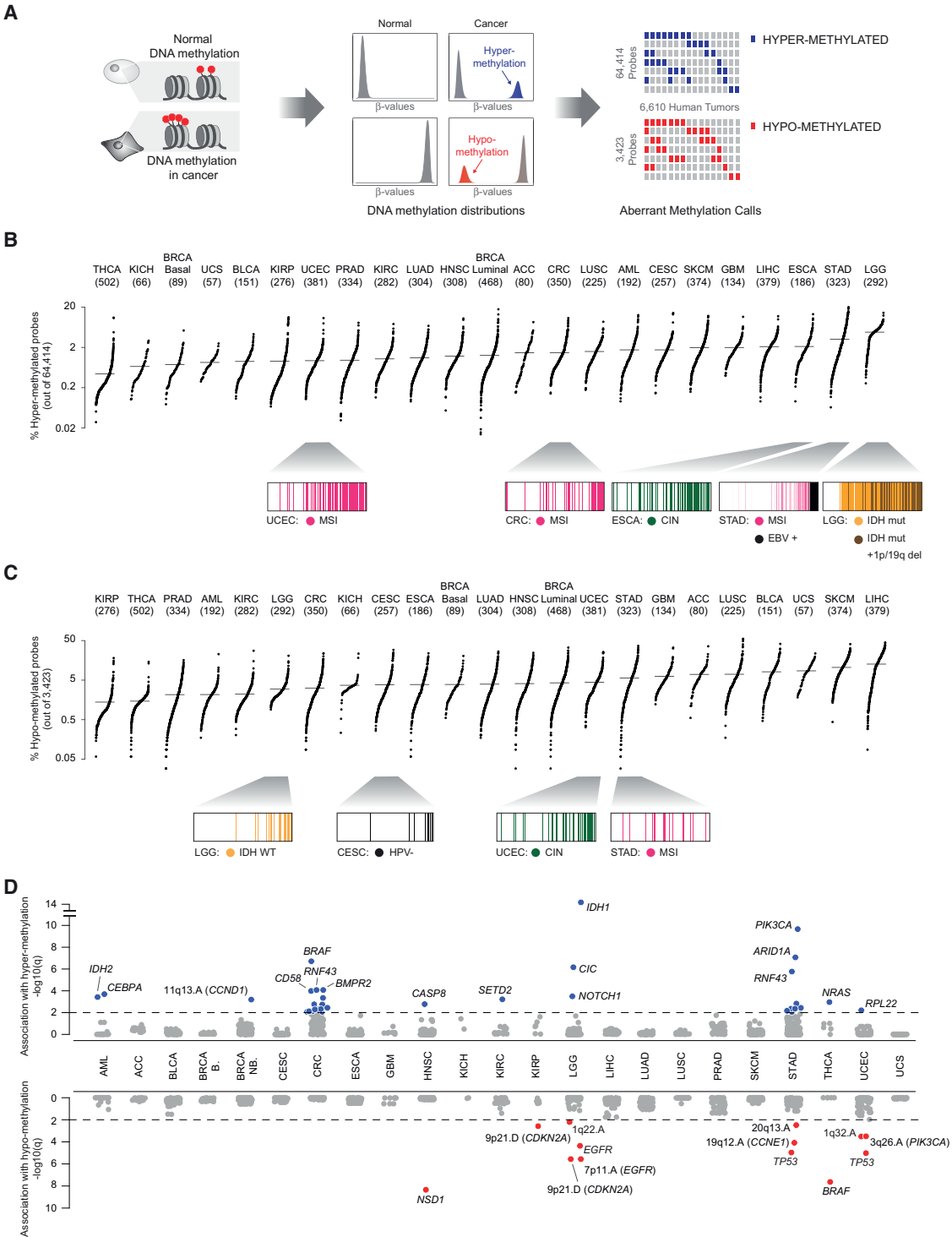


Figure 1. RESET Identifies Hyper- and Hypo-Methylation Events across Human Cancers

(A) Schematic pipeline of the first step of the RESET algorithm.

(B and C) Hyper-methylation (B) and hypo-methylation (C) event frequencies in 6,010 human tumors (black dots) across 23 cancer types. Frequencies are estimated as the percentage of probes found hyper-methylated (hypo-methylated) compared to normal tissues. Tumor types are sorted by increasing hyper-methylation (hypo-methylation) mean frequency. Molecular subtypes associated with hyper-methylation (hypo-methylation) event frequencies are shown at the bottom (p value < 0.05), samples are sorted by aberrant methylation frequency, samples with the indicated annotation are color-coded, otherwise they are left white).

(legend continued on next page)

proposed to analyze DNA methylation changes in cancer. Examples are methods integrating multiple data types to study hyper-methylation within pathway modules (Jiao et al., 2014) or develop cancer risk predictors (Teschendorff et al., 2012). Finally, explorations of silenced genes by DNA hyper-methylation have utilized statistical tests to assess gene expression differences between methylated and unmethylated tumors in an unbiased and systematic way (Hinoue et al., 2012; Sepulveda et al., 2016; Cancer Genome Atlas Network, 2015). However, these approaches often suffer from sample size biases that favor frequently altered genes (see STAR Methods), and have focused on silencing events within specific tumor types.

Overall, in contrast to systematic pan-cancer investigations of genetic alterations (Bailey et al., 2018; Beroukhim et al., 2010; Lawrence et al., 2013), analyses of aberrant DNA methylation and associated transcriptional changes have so far largely been limited to specific tumor types or genes of interest. As such, several questions remains outstanding: How does the extent of aberrantly methylated loci compare among different tumor types? Which genes and pathways are frequently aberrantly methylated? How genetic alterations and DNA methylation events associate in cancer? And what are the functional and therapeutic implications of such alterations? Here, we explore these questions by systematically and unbiasedly investigating the genetic, DNA methylation, and transcriptomic profiles of more than 6,000 human tumors (Table S1), using a newly designed algorithmic approach.

RESULTS

To systematically identify candidate functional DNA methylation changes in cancer, we (1) first, identify gene transcription start sites with aberrant methylation states in cancer samples compared to normal tissue (hyper- and hypo-methylation calling) and (2) second, we query whether aberrant methylation states associate with silenced or enhanced mRNA expression of the corresponding genes (silencing and enhancing events identification). This two-step approach models the distributions of DNA methylation at specific loci in normal and tumor samples and quantifies the strength (or effect size) of the associations between DNA methylation and mRNA expression changes (see STAR Methods). Our computational algorithm is thus a resource to detect epigenetically silenced and enhanced targets in cancer (RESET). In the following, we will separately present the results associated to each step (1 and 2) and, unless stated otherwise, the term ''epigenetic'' will be used to connotate processes associated to DNA methylation.

Hyper- and Hypo-Methylation across 24 Human Cancer Types

We applied RESET to 23 cancer cohorts (Table S1) molecularly profiled by The Cancer Genome Atlas (TCGA) for a total of 6,010 human tumors with available DNA methylomes generated by the Illumina Infinium HM450 array. For 5,633 samples, we had matched RNA-sequencing data. In addition, we separately analyzed a cohort of 600 serous ovarian tumors that were profiled by TCGA using the HM27 array, 265 of which were analyzed by RNA-sequencing. To estimate the methylation status in normal conditions, we collected DNA methylation data for 702 normal tissue samples from 14 TCGA cancer cohorts. In this study, we focused on pan-cancer DNA methylation events and, thus, we studied only probes that exhibited a consistently high or low methylation status across all the normal tissues that we analyzed (Figures S1A and S1B). Probes were further filtered to include only those proximal to gene transcription start sites (TSS probes) (Figures S1C–S1H; STAR Methods). In total, we identified 64,414 probes—mapping to 12,053 genes—that have low DNA methylation across all normal samples, and 3,423 probes—mapping to 2,006 genes—that have high DNA methylation across all normal samples (Table S2).

The fractions of hyper- and hypo-methylated TSS probes determined by RESET (step 1, Figure 1A) varied across tumor types and were associated with molecular subtypes (Figures 1B, 1C, and S1I for ovarian cancer). Our results recapitulated known associations between recurrent hyper-methylation and tumor subtypes characterized by *IDH1* (R132) and *IDH2* (R140 and R172) mutations in glioma (Noushmehr et al., 2010), micro-satellite instability (MSI) in uterine, colorectal, and stomach cancers (Cancer Genome Atlas Network, 2012; Kandoth et al., 2013; Cancer Genome Atlas Research Network, 2014), and viral infection (EBV) in stomach cancer (Figure 1B). Conversely, IDH-wild-type glioma and HPV-negative cervical tumors exhibited high percentages of hypo-methylated loci. Interestingly, MSI stomach tumors also exhibited a significantly higher percentage of hypo-methylated loci than micro-satellite stable tumors (Figure 1C). Overall, the fraction of aberrantly methylated probes did not correlate with tumor purity, ploidy, or infiltration of specific immune cell populations (Figure S2A; Table S2). Instead, the fraction of hypo-methylated probes correlated with a recently proposed stemness signature based on hypo-methylation of specific loci (Malta et al., 2018) and with the extent of chromosomal alterations in lung, prostate, stomach, and uterine carcinoma, supporting a link between DNA hypo-methylation and chromosomal instability (Gaudet et al., 2003).

(D) We tested 505 genetic alterations for their association with increased hyper- (top) and hypo-methylation (bottom) within each tumor type. Wilcoxon FDR q values are reported as –log10(q) and significant associations ($q < 0.01$) are color-coded (blue/red) and labeled by gene symbol for somatic mutations or cytoband for amplifications (A) and deletions (D). Candidate targets of focal amplifications and deletions are reported in brackets.
ACC, adrenocortical carcinoma; AML, acute myeloid leukemia; BLCA, bladder carcinoma; BRCA basal/B. basal, basal breast invasive carcinoma; BRCA luminal/B. luminal, luminal breast invasive carcinoma; CESC, cervix squamous cell carcinoma; CRC, colon and rectum carcinoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe carcinoma; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LGG, low grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian carcinoma; PRAD, prostate adenocarcinoma; STAD, stomach adenocarcinoma; SKCM, skin cutaneous melanoma; THCA, thyroid papillary carcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma. The number of samples for each tumor type are available in Table S1.
See also Figure S1 and Table S2.

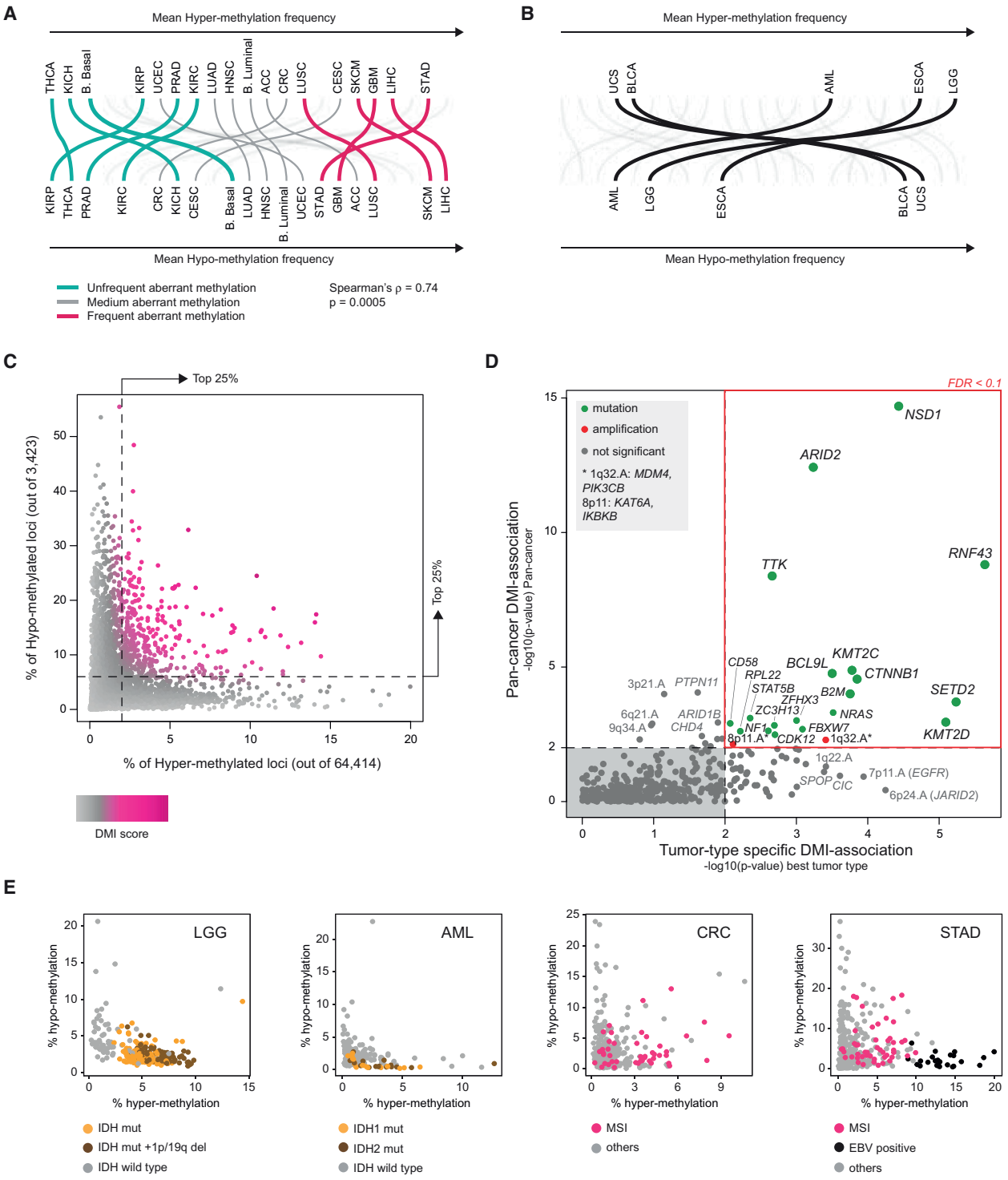


Figure 2. DNA Methylation Instability

(A and B) Comparison of rank positions obtained by each tumor type with respect to hyper-methylation (top) or hypo-methylation (bottom) event frequencies. (A) For 19 tumor types the rank positions based on hyper- and hypo-methylation shift by at most 8 positions (corresponding to one-third of the total) showing a high rank correlation (Spearman's correlation = 0.74, p = 0.0005). These 19 cancer types were grouped into cancers with low, medium, and high aberrant methylation frequency, as indicated by the color key.

(B) Only 5 tumor types scored at the top of one rank and at the bottom of the other.

(legend continued on next page)

Next, we tested a set of 505 recurrent genetic alterations in cancer (Mina et al., 2017) for association with aberrant DNA methylation (Table S2). Many of the alterations that scored as significant ($q < 0.01$) were reported enriched in the molecular subtypes highlighted above. These include *IDH1*, *CIC*, and *NOTCH1* mutations in glioma, which characterize IDH mutant tumors with 1p/19q co-deletion (Brat et al., 2015), *BRAF* mutations in MSI colorectal cancer (Cancer Genome Atlas Network, 2012), and *PIK3CA* and *ARID1A* mutations in EBV^+ stomach tumors (Cancer Genome Atlas Research Network, 2014) (Figure 1D). Significant associations with hypo-methylation events were found for *NSD1* mutations in head and neck cancer, *BRAF* mutations in thyroid carcinoma, and alterations associated with chromosomal instability, such as *CDKN2A* deletions in glioma and papillary kidney tumors, and *TP53* mutations and copy number amplifications in stomach and uterine cancer (Figure 1D). The latter were consistent with the observed correlation between the fraction of hypo-methylated probes and chromosomal instability in these tumor types (Figure S2A).

The majority of tumor types (19 out of 24) with low (high) mean hyper-methylation frequency also exhibited low (high) mean hypo-methylation frequency (Figure 2A, Spearman rank correlation = 0.74, p = 0.0005). Notable exceptions were low grade glioma (LGG) and acute myeloid leukemia (AML), which were both enriched for *IDH1/2* mutations and hyper-methylated loci (Figure 2B). A considerable number of samples exhibited high numbers of both hyper- and hypo-methylated TSS probes (Figure 2C). We refer to these samples as characterized by DNA methylation instability (DMI), to distinguish them from cases where only hyper-methylation or hypo-methylation is frequent. We scored the extent of DMI in each sample by combining the percentages of hyper- and hypo-methylated loci using the F_2-measure (color-coded in Figure 2C). DMI scores did not correlate with tumor purity, ploidy, or immune infiltration (Figure S2B).

Using these scores, we assessed the association between recurrent genetic alterations and DMI, both within each tumor type and at a pan-cancer level (Table S3). The two analyses concordantly identified a set of recurrent mutations as enriched in samples with high DMI scores (Figure 2D). DMI-associated mutated genes were enriched for chromatin remodeling factors (p = 4.2E−8, false discovery rate [FDR] <2E−4, Table S3), such as H3K36 methyltransferases *NSD1* and *SETD2*, SWI/SNF components *ARID2* and *ARID1B*, and lysine methyl-transferases *KMT2C* and *KMT2D*. Additionally, alterations of Wnt signaling were significantly associated with DMI (p = 3.3E−4, FDR <0.05), especially mutations of β-catenin (*CTNNB1*) and the Wnt-inhibitor *RNF43*. These mutations were previously found associated with micro-satellite instability (MSI) in gastric tumors (Cancer Genome Atlas Network, 2012; Cancer Genome Atlas Research Network, 2014), in apparent contradiction with reported associations between MSI and CIMP (Hinoue et al., 2009). This observation prompted us to re-evaluate tumors subtypes in the TCGA cohort that were previously classified CIMP. Tumors with *IDH1/2* mutations (IDH mutant) and EBV-positive stomach adenocarcinoma were characterized by high fractions of hyper-methylated and low fractions of hypo-methylated TSS probes (Figure 2E). Conversely, MSI tumors and FH mutant renal papillary carcinoma exhibited high fractions of both hyper- and hypo-methylated TSS probes (Figures 2E and S2C). These results indicate that, by accounting for both aberrant hyper- and hypo-methylation, we can distinguish between "pure" CIMP (i.e., tumors exhibiting highly recurrent hyper-methylation of gene TSS but not hypo-methylation), and tumors characterized by DMI (i.e., frequent hyper and hypo-methylation of gene TSS).

Epigenetically Silenced and Enhanced Targets across 24 Human Cancers

RESET was run independently in each of the 24 cancer cohorts (Figure 3A), and genes scoring as significant in at least one tumor type were retained and ranked based on the sum of their significant scores (Figures 3B and 3C; Table S4). In total, we identified 581 epigenetically silenced and 85 epigenetically enhanced genes (Figures 3B and 3C). Silenced targets included previously reported targets of recurrent hyper-methylation such as *MLH1*, *BRCA1*, *FANCF*, and *CHFR* (Lahtz and Pfeifer, 2011; Cancer Genome Atlas Research Network, 2011; Cancer Genome Atlas Network, 2012; Toyota et al., 2003) and other cancer-associated genes including the cell-cycle-regulator *CDKN1C*, the transforming growth factor β (TGF-β) signaling genes *TGIF1* and *ACVR1C*, and the pro-apoptotic genes *FAS*, *BIRC3*, *TNFRSF1A*, and *TNFRSF10A*. Similarly, epigenetically enhanced targets included positive controls such as *MAGEC2* (Van Tongelen et al., 2017) and *SYCP2* (Degli Esposti et al., 2017), and known oncogenes, such as *MYCN*, *BCL2L10*, *CTNNB1*, *IRS2*, and *IGF2*. Only 3 out of 581 silenced and 3 out 85 enhanced target genes showed significant associations between their alteration patterns and the estimated fraction of infiltrated non-tumor cells (q value <0.1, fold-change >1.5, Table S4) and, for each gene, this association was found only within one tumor type.

The roster of silenced and enhanced targets identified by RESET included several genes whose roles in cancer are still largely unexplored, such as *PXMP4* and *PNLDC1*, that scored as significant across multiple tumor types (Figures 3D and 3E). *PXMP4* is a peroxisome component and was previously shown to become silenced in an androgen-insensitive prostate cancer subline (AI-LnCAP), whereas its transient re-activation was able to impair cell proliferation (Wu and Ho, 2004). Here, *PXMP4* was found to be hyper-methylated and silenced in 8 tumor types,

(C) Percentage of probes in each tumor sample that are hyper- (x axis) and hypo-methylated (y axis). Samples are color-coded by their DNA methylation instability (DMI) score (see STAR Methods).

(D) We tested 505 genetic alterations for their association with DMI scores both within each single tumor type (top scoring association for each event on the x axis, Wilcoxon p value) and at a pan-cancer level (y axis, ANOVA p value). Alterations significant in both tests (FDR <0.1 for both analyses) are color-coded (mutations in green, amplifications in red) and labeled as in Figure 1D.

(E) Normalized fraction of hyper- (H) and hypo-methylated (h) probes in low grade glioma (LGG), acute myeloid leukemia (AML), colorectal cancer (CRC), and stomach adenocarcinoma (STAD). The following tumor subtypes are highlighted: LGG IDH mutant (yellow) and LGG IDH-mut with 1p/19q co-deletion (brown), AML samples with *IDH1* (yellow) and *IDH2* (brown) mutations, EBV-positive STAD (black), and STAD and CRC tumors with micro-satellite instability (MSI, pink). See also Figure S2 and Table S3.

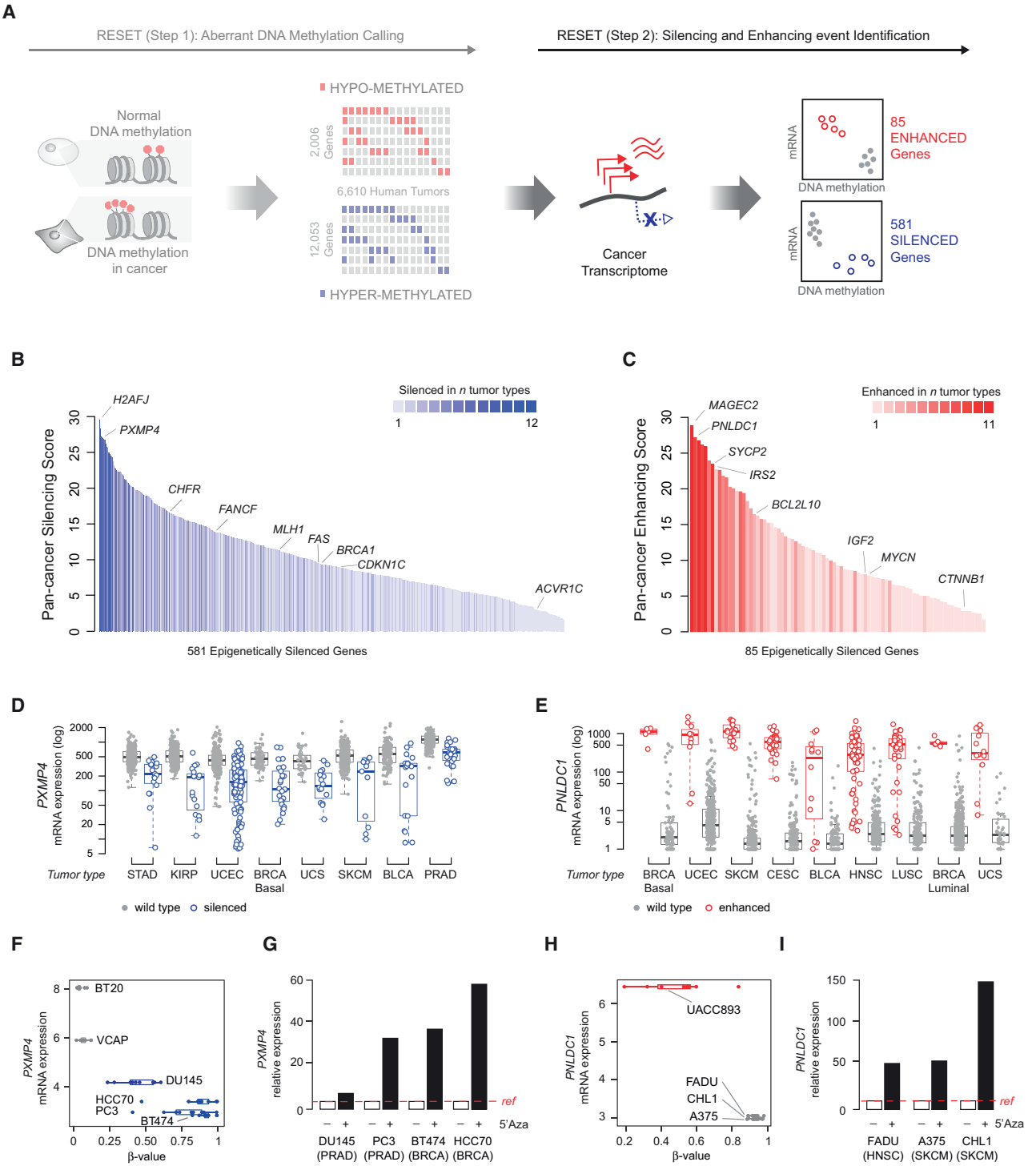


Figure 3. Epigenetic Silencing and Enhancing Events in Cancer

(A) Schematic pipeline of the complete RESET algorithm. The first step (shaded panel on the left) performs aberrant DNA methylation calling, the second step (highlighted panel on the right) identifies epigenetically silenced and enhanced genes.

(B and C) RESET identifies 581 epigenetically silenced genes (B) and 85 epigenetically enhanced genes (C) across the pan-cancer dataset. Genes are ranked based on the sum of scores obtained in each tumor type and color-coded based on the number of cancer types where they scored as significant (shades of blue for silenced, shades of red for enhanced). Representative silenced and enhanced targets are labeled by their official gene symbol.

(legend continued on next page)

with incidences varying between 3% (skin melanoma) to 31% (endometrial cancer) (Figure 3D). *PNLDC1* encodes for a deadenylase enzyme expressed in mouse embryonic stem cells and suppressed by DNA methylation during cell differentiation (Anastasakis et al., 2016). *PNLDC1* was altered in 9 tumor types, with incidences varying from 1% (luminal breast cancer) to 22% (head and neck carcinoma) (Figure 3E). Both genes are thus altered across multiple tumor types, albeit only epigenetically, and are associated with potential oncogenic pathways such as cell proliferation and differentiation.

To support the association between aberrant DNA methylation and mRNA expression for our set of epigenetically altered genes, we analyzed colorectal and lung cancer cell lines that were profiled by RNA-sequencing before and after treatment with the DNA-demethylating agent 5′azacitidine (5′AZAdC). In all cases, the vast majority of RESET-targets (n = 581) that were methylated in the analyzed cell lines increased expression upon treatment with 5′AZAdC (Figure S3D). Next, we selected 4 top scoring genes (*PXMP4* and *H2AFJ* among silenced targets, *PNLDC1* and *SYCP2* among enhanced targets) and verified their methylation and expression status in a broader panel of cancer cell lines (Iorio et al., 2016). Reported mRNA expression of these genes was negatively correlated with promoter DNA methylation (Figures 3F, 3H, and S3F). We validated mRNA levels of the 4 genes by quantitative PCR (qPCR) (Figure S3G) and re-assessed them after 7 days of treatment with 5′AZAdC. The treatment restored expression of all tested genes in the cell lines where they were silenced, and this effect was more pronounced in cell lines where the targets exhibited the most extreme DNA methylation and transcriptional downregulation (Figures 3G, 3I, and S3H). These results indicate that RESET can identify bona fide epigenetically regulated genes in cancer.

Interdependencies between Genetic and Epigenetic Events

Out of the 581 silenced and 85 enhanced targets identified by RESET, only a handful had been previously found as recurrently mutated, deleted, or amplified in cancer (Mina et al., 2017) (Figures 4A and 4B). A few targets in our datasets exhibited patterns of co-occurrence between promoter hyper-methylation and loss of heterozygosity, among which genes in chromosomes 1p and 19q, which are frequently deleted in IDH mutant LGG. These included the candidate tumor suppressors *CD58* (Challa-Malladi et al., 2011) and *EMP3* (Alaminos et al., 2005) that exhibited significantly lower expression when copy number losses were accompanied by promoter hyper-methylation (Figures S4A and S4B). Similarly, we found synergistic *CTNNB1* upregulation in cervical carcinomas with concurrent promoter hypo-methylation and gain-of-function mutations (Figure S4C). Within each tumor type, genetic alterations associated with specific molecular subtypes (e.g., MSI, CIN, EBV$^+$, and IDH mutant gliomas) were associated with an increased number of epigenetic silencing and/or enhancing events (Figure 4C; Table S5). To correct for tumor types and subtypes, we assessed co-occurrence and mutual exclusivity between genetic and epigenetic alterations within the pan-cancer cohort using the SELECT algorithm (Mina et al., 2017) (Table S5).

BRAF mutations were found frequently mutually exclusive with multiple silencing events (Figure 4D), especially occurring in skin melanoma, whereas, *IDH1* mutations co-occurred with the highest number of epigenetic silencing events (Figure 4E). Silenced genes co-occurrent with *IDH1* mutations included several metabolic regulators involved in oxidation-reduction processes (p value = 1.8E−5) and nicotinamide adenine dinucleotide (NAD) cycle (Chiarugi et al., 2012), such as *NMRAL1*, *ACADS*, *CYB5R1*, and *NMNAT3*. The latter two were also concurrent with *IDH2* mutations. In addition, silencing of the candidate tumor suppressors *HTATIP2* (Dong et al., 2015) and *SH2D4A* (Ploeger et al., 2016) (Figure 4F) was co-occurring with *IDH1* mutations in both glioma and melanoma. Interestingly, mutations of *CTNNB1* and *NSD1* were significantly associated with DMI (Figure 2D) and, consistently, co-occurrent with both silencing and enhancing events (Figure 4E). *CTNNB1* was significantly co-occurrent with 5 enhancing and 10 silencing events (Figure 4G), most of them prevalent in liver hepatocellular carcinoma and targeting Wnt-related genes. These included enhanced expression of the glutamine synthetase *GLUL*, a direct Wnt and Hippo target (Cox et al., 2016; Lachenmayer et al., 2012), and of the synovial sarcoma associated gene *SSX1*, recently implicated in Wnt-target activation (Cironi et al., 2016). Similarly, the Wnt repressor *NPHP4* (Borgal et al., 2012) was found frequently silenced in *CTNNB1* mutant tumors. In addition, *CTNNB1* mutations significantly co-occurred with silencing of the heat-shock factor *DNAJA4* in skin melanoma, liver, adrenocortical, and stomach carcinoma. *DNAJA4* has been shown to promote ApoE expression, which in turn suppresses invasion and metastasis in melanoma (Pencheva et al., 2012). These results suggest DMI is associated with both the emergence and selection of epigenetic alterations promoting oncogenic processes.

(D) *PXMP4* mRNA expression (y axis) is compared in epigenetically silenced (blue-contoured dots) and wild-type (gray dots) cases in 8 tumor types where epigenetic silencing of PXMP4 was significant.
(E) PNLDC1 mRNA expression (y axis) is compared in epigenetically enhanced (red-contoured dots) and wild-type (gray dots) cases in 9 tumor types where epigenetic enhancement of *PNLDC1* was significant.
(F) DNA methylation (x axis) and mRNA expression (y axis) of *PXMP4* in 6 cancer cell lines. DNA methylation values are reported for each probe that maps to the gene and summarized by boxplots.
(G) Relative *PXMP4* expression measured by qPCR in cell lines with high methylation of *PXMP4* is strongly enhanced upon treatment with 5′AZAdC for 7 days (black filled bars) compared to untreated cells (black contoured bars).
(H) DNA methylation (x axis) and mRNA expression (y axis) of *PNLDC1* in 3 cancer cell lines. DNA methylation values are reported for each probe that maps to the gene and summarized by boxplots.
(I) Relative *PNLDC1* expression measured by qPCR in cell lines with high methylation of *PNLDC1* is strongly enhanced upon treatment with 5′AZAdC for 7 days (black filled bars) compared to untreated cells (black contoured bars).
In all the analyses, qPCR experiments were repeated three times and averaged.
See also Figure S3 and Table S4.

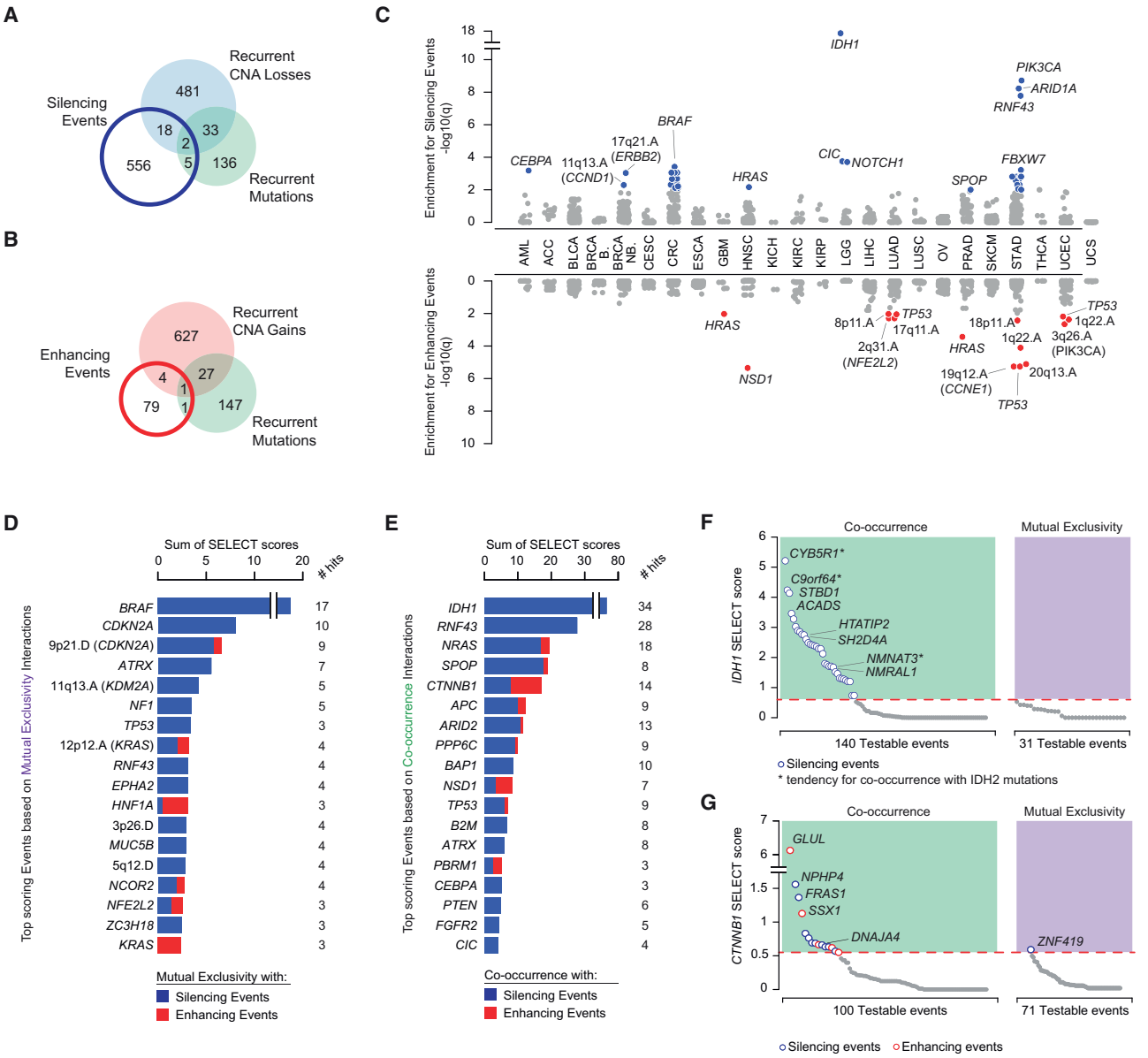


Figure 4. Interdependencies between Genetic and Epigenetic Alterations

(A and B) Gene targets of epigenetic silencing (blue contoured circle, A) and enhancing (red contoured circle, B) events are intersected with gene targets of recurrent somatic mutations (green circles), copy number deletions (blue circle, A) and copy number amplifications (red circle, B).

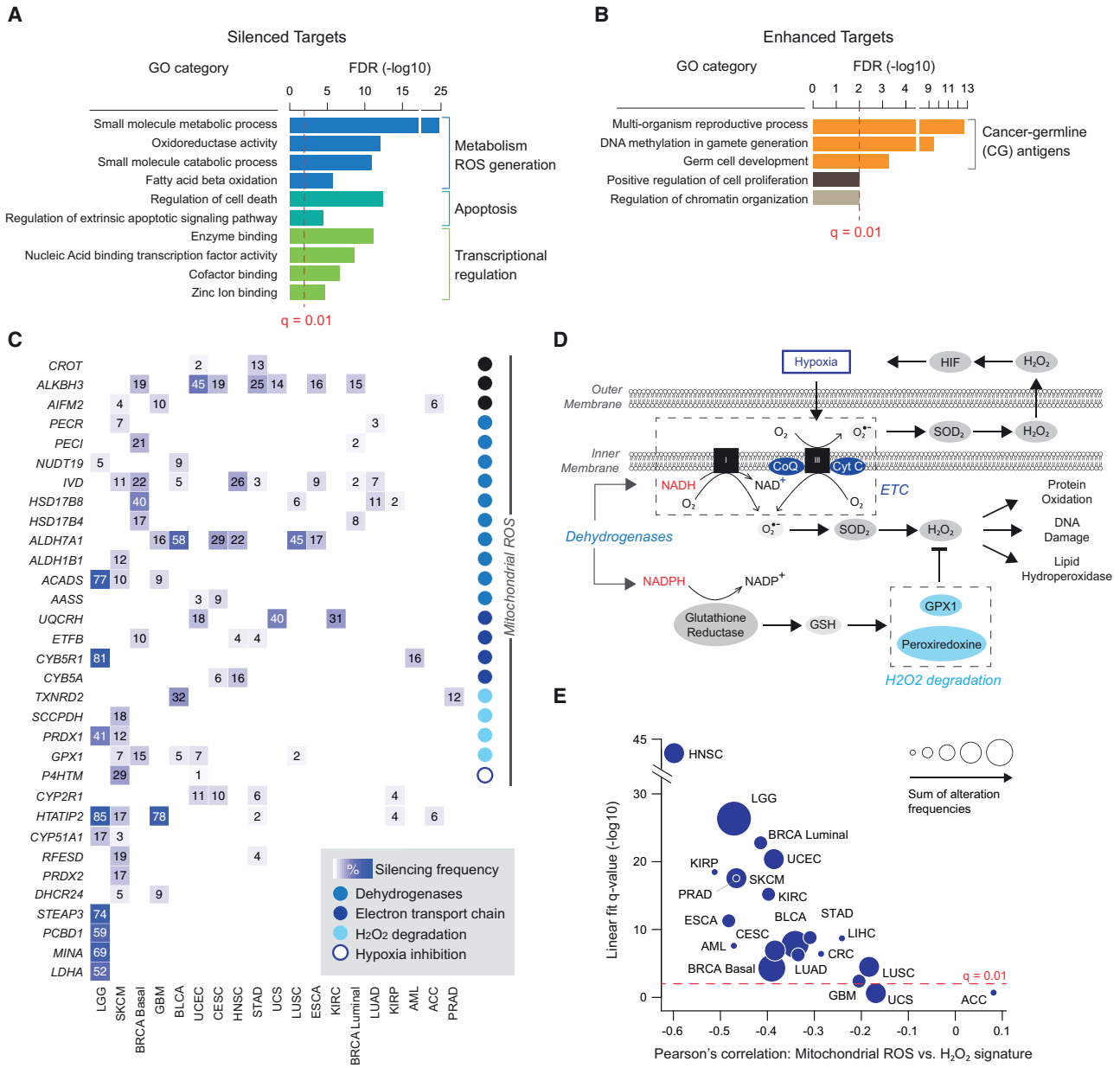
(C) We tested 505 genetic alterations for their association with the incidence of epigenetic silencing (top) and enhancing (bottom) events within each tumor type. Wilcoxon FDR q values are reported as –log10(q) and significant associations ($q < 0.01$) are color-coded (blue/red) and labeled as in Figure 1D.

(D and E) Using SELECT, we tested 505 recurrent genetic alterations in cancer for mutual exclusivity and co-occurrence with our set of 581 epigenetic silencing and 85 enhancing events in cancer. Genetic alterations that are mutually exclusive (D) or co-occurrent (E) with multiple epigenetic silencing (blue) or enhancing (red) events. Genetic alterations are ranked by the sum of SELECT scores and the actual number of significant associations found by SELECT are reported (# hits). Genetic alterations are labeled as in Figure 1D.

(F and G) Co-occurrence (left) and mutual exclusivity (right) interactions between *IDH1* (F) or *CTNNB1* (G) and silencing and enhancing events. Significant events (above the red dotted line) are color-coded (silencing events, blue contoured dots; enhancing events, red contoured dots) and representative targets are labeled. See also Figure S4 and Table S5.

Cancer Pathways Affected by Epigenetically Silencing and Enhancing Events

Functional enrichment analysis performed on epigenetically silenced genes found transcriptional regulation, apoptosis, and cell metabolism as the most enriched categories (Figure 5A; Table S4). In addition, frequently methylated polycomb targets (Widschwendter et al., 2007) were also moderately enriched within our set of silenced genes (p value = 0.025). In contrast,



(legend on next page)

epigenetically enhanced genes were highly enriched for germ cell development, corresponding to a high presence of cancer-germline (CG) antigen genes (CG genes), as well as cell proliferation and chromatin organization, both associated with epigenetic enhancing of oncogenes and transcriptional regulators such as *MYCN*, *CTNNB1*, and *IGF2* (Figure 5B).

Cell metabolism is characteristically altered in most cancers (Cairns et al., 2011), and mutations of the isocitrate dehydrogenases *IDH1* and *IDH2* establish a critical link between altered metabolic processes and increased DNA methylation (Lu and Thompson, 2012). IDH-mutations in our dataset characterized 77% of LGG, and LGG tumors exhibited the highest incidence of silenced metabolic genes. However, most of these silencing events were rarely found in other tumor types (Figure 5C). The majority of metabolic genes that were silenced across multiple tumor types were associated with mitochondrial reactive oxygen species (ROS) generation including a broad class of dehydrogenases, inhibitors of H_2O_2, and components of the electron transport chain (ETC) (Figures 5C and 5D). Because the predicted downstream effect of silencing these targets is an increased generation of ROS, we scored each tumor using (1) an mRNA expression signature derived from silenced mitochondrial ROS genes identified by RESET, and (2) an experimentally derived signature composed of genes overexpressed upon H_2O_2 induction (Parikh et al., 2010). Scores from these signatures were significantly anti-correlated in almost all tumor types, supporting the association between silencing of ROS inhibitors and generation of mitochondrial ROS (Figure 5E).

CG antigens, or CG-genes, are predominantly expressed in germ cells and trophoblasts, but upregulated in tumors by epigenetic mechanisms (Simpson et al., 2005) (Figure S5A). Not all tumor types presented an equal extent of upregulated CG-genes, which were either enhanced across multiple cancers or specifically in a single tumor type (Figure S5B). While their oncogenic role remains unclear (Simpson et al., 2005), CG antigens have been recognized as highly immunogenic and have been identified as targets of antitumor T cell immune response in patients receiving tumor-infiltrating adoptive T cell therapy (Stevanović et al., 2017). To assess the prognostic value of epigenetically enhanced CG-genes, we collected mRNA expression data for a cohort of skin melanoma patients that were treated with an anti-PD-1 therapeutic antibody (Hugo et al., 2016). The majority of CG-genes that RESET found epigenetically enhanced in melanoma were more highly expressed in patients that either completely (CR) or partially (PR) responded to therapy than in patients with a progressive disease (PD) (Figure 5F). Moreover, an mRNA expression signature comprising an extended set of CG-genes (Almeida et al., 2009) (CG-signature) was associated with beneficial responses to anti-PD-1. Indeed, complete responders exhibited significantly higher values of the CG-signature than non-responding patients ($p = 0.017$, Figure 5G), and patients stratified based on this signature had significantly different overall survival (Figure 5H). Vice versa, in skin melanoma patients treated with another immune-checkpoint inhibitor, anti-CTLA4 (Van Allen et al., 2015), high CG-scores were associated with worse prognosis, even though not reaching statistical significance (Figure S5C). Recently, expression of the CG-antigen MAGE-A was found predictive of resistance to anti-CTLA4, but not to anti-PD1 (Shukla et al., 2018). Together, these results warrant further investigations on the prognostic value of CG-genes expression for patients treated with immune checkpoint inhibitors.

Aberrant DNA Methylation in Pediatric Tumors

Pediatric tumors typically display a low mutation burden (Lawrence et al., 2013), and their pathogenesis have been frequently associated with specific alterations in epigenetic regulatory mechanisms (McKenna and Roberts, 2009). Therefore, we explored the potential of using RESET to investigate aberrant DNA methylation in pediatric cancer. Recently, the NCI's Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative has profiled a large cohort of Wilms tumors, the most common form of kidney cancer in children (Gadd et al., 2017). This cohort included 84 favorable histology cases that relapsed

Figure 5. Cancer Pathways Affected by Epigenetic Silencing and Enhancing Events

(A) Gene Ontology (GO) categories significantly enriched (FDR q-value on the y axis) for epigenetically silenced genes. GO categories are color-coded based on 3 main groups: metabolism and ROS generation, apoptosis, and transcriptional regulation.

(B) GO categories significantly enriched (FDR q-value on the y axis) for epigenetically enhanced genes. GO categories are color-coded based on 3 main groups: cancer-germline (CG) antigens, cell proliferation, and chromatin organization.

(C) Frequency of silencing genes involved in the ROS generation pathway segregated by cancer type. Silenced targets include several genes with mitochondrial activity (annotated on the right) that can be grouped into dehydrogenases (blue), promoters of H_2O_2 degradations (light blue), components of the electron transport chain (ETC, dark blue), and hypoxia inhibitors (blue contoured).

(D) Pathway schematic of mitochondrial ROS generation highlighting the functional categories that group most of the epigenetically silenced genes shown in (C).

(E) Correlation between a gene expression signature composed of genes overexpressed in the presence of H_2O_2 and a gene expression signature composed of silenced targets involved in mitochondrial ROS generation (x axis). The two signatures are significantly anti-correlated (linear fit q-value on the y axis) in almost all tumor types (here represented by blue dots with size proportional to the incidence of epigenetic silencing of mitochondrial ROS inhibitors).

(F) Normalized mean expression differences (y axis) of CG- and CG-like genes (x axis) in melanoma patients responding versus non-responding to anti-PD-1 (CR, complete responders; PR, partial responders; PD, progressive disease). CG- and CG-like genes were tested if they were found epigenetically enhanced by RESET in the TCGA melanoma cohort. mRNA expression differences come from an independent skin melanoma patient cohort that underwent treatment with anti-PD-1 therapeutic antibody. Expression differences are normalized as the sum of expression values in responders (CR and PR) minus the sum of expression values in PD cases divided by the total sum of expression values.

(G) Comparison of CG signature scores between melanoma patients with a progressive disease (PD, gray dots) or exhibiting a partial response (PR, red contoured dots) or a complete response (CR, red-filled dots) to anti-PD-1.

(H) Overall survival of melanoma patients treated with anti-PD-1. Patients were partitioned in two groups based on their CG signature scores with patients having high scores (score >0.5, red line) showing significant better survival than those with low scores (score <0.5, gray line).

See also Figure S5 and Table S4.

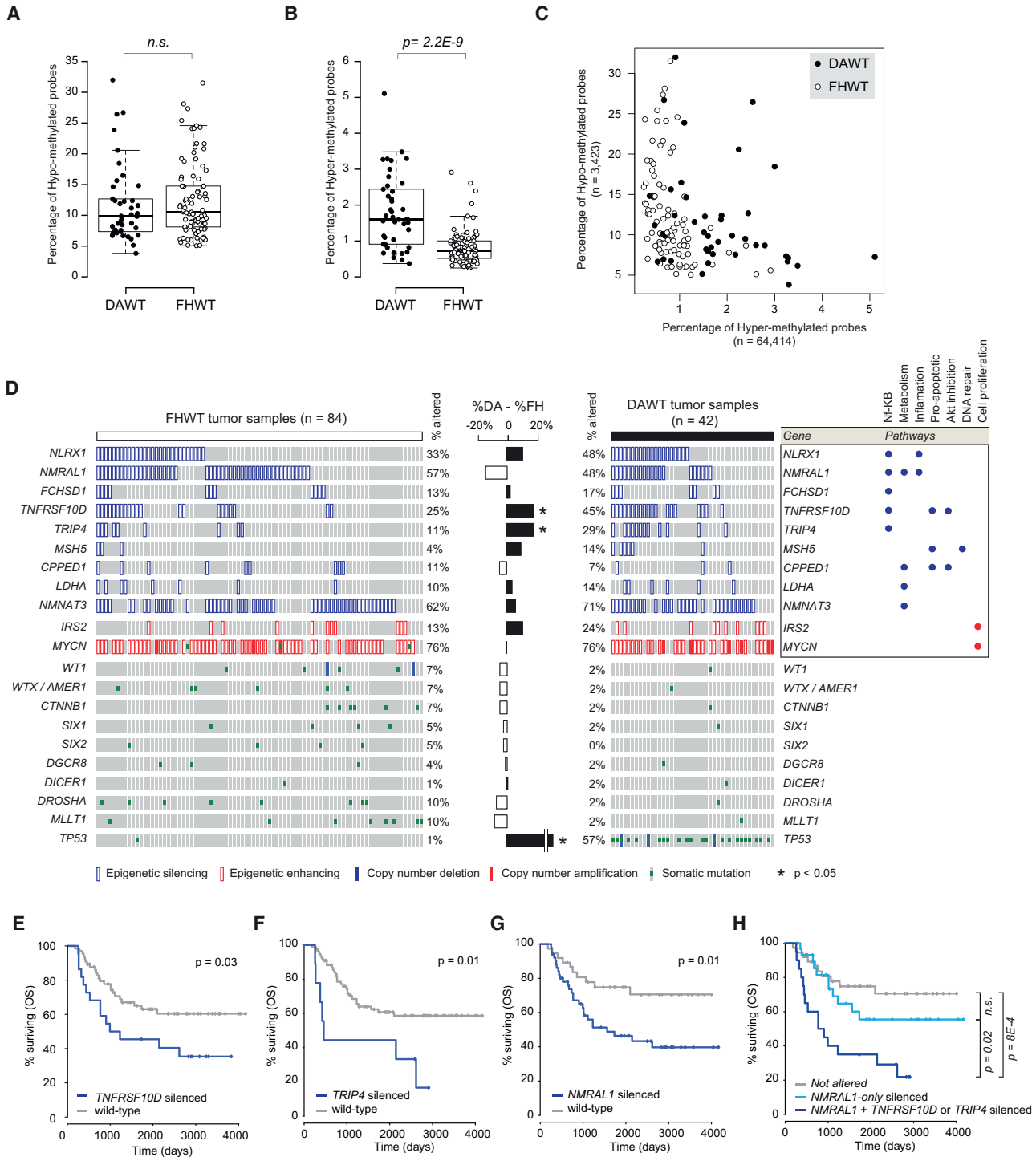


Figure 6. Aberrant DNA Methylation in Pediatric Wilms Tumors

(A) Percentage of probes for each tumor sample that are hyper- (x axis) and hypo-methylated (y axis). Tumor samples are color-coded based on their histological subtype: favorable histology (FHWT, black contoured dots) and diffuse anaplasia (DAWT, black filled dots).

(B and C) Distribution comparison of the percentage of hypo- (B) and hyper-methylated (C) probes between the two histological subtypes (Wilcoxon test).

(D) Map overview of selected genetic and epigenetic alterations in Wilms tumors. Columns are tumor samples and each row is one selected alteration including epigenetic silencing (blue contoured bars) and enhancing (red contoured bars) events, copy number amplifications (red bars), and somatic mutations (green squares). Wild-type cases are in gray. Two separate maps are generated for FHWT and DAWT cases. The alteration frequencies of each event in the two subtypes are compared by taking the difference between the percentage of altered DAWT cases (%DA) and the percentage of altered FHWT cases (%FH) (bar plot, black

(legend continued on next page)

(FHWT) (Figure S6A) and 42 cases characterized by diffuse anaplasia (DAWT) and associated with unfavorable prognosis and *TP53* mutations (Figure S6B).

Using RESET, we first estimated the extent of hyper- and hypo-methylation in these tumors (RESET-step 1). DAWT and FHWT had comparable extent of hypo-methylation (Figure 6A), but DAWT exhibited significantly higher hyper-methylation (Figure 6B) and combined percentages of hyper and hypo-methylated TSS probes indicated that these tumors exhibited DMI (Figure 6C). Concordantly, epigenetic silencing and enhancing events (RESET-step 2, Table S6) were in general more frequent in DAWT cases that in FHWT (Figure 6D) and affected genes involved in nuclear factor κB (NF-κB) and AKT-signaling as well as apoptosis, DNA repair, cell metabolism, and cell proliferation (Figure 6D). In particular, we found that virtually all cases exhibited hypo-methylation of the *MYCN* oncogene, sometimes concomitant with copy number amplifications or somatic mutations. *MYCN* has been reported as frequently upregulated in Wilms tumors and altered through different mechanisms (Williams et al., 2015). Our results indicate that hypo-methylation of this locus is a hallmark of this tumor type.

Given that the unfavorable prognosis subtype, DAWT, had an overall higher incidence of epigenetic events than tumors with favorable histology (FHWT), we asked whether the occurrence of epigenetic alterations was associated with worse outcome in FHWT. Epigenetic silencing of *TNFRSF10D*, *TRIP4*, and *NMRAL1* (Figure S6C) was associated with significantly worse survival (Figures 6E–6G). Because *TNFRSF10D* and *TRIP4* almost always co-occurred with silencing of *NMRAL1*, we tested whether they had independent prognostic value. Interestingly, we found that FHWT tumors exhibiting silencing of *NMRAL1* and of either *TNFRSF10D* or *TRIP4* had significantly worse outcome than *NMRAL1*-only altered samples (Figure 6H). These targets have all been implicated in NF-κB signaling (Degli-Esposti et al., 1997; Jung et al., 2002; Lian and Zheng, 2009), making this pathway a candidate driver of Wilms tumor development and progression.

DISCUSSION

Over the last decade, several large-scale cancer genomics studies have explored and characterized the landscape of genetic modifications across thousands of human tumors. The resulting catalog of pan-cancer associated variants misses an important second dimension represented by cancer-selected epigenetic alterations. Here, we explored DNA methylation changes occurring across 24 human cancers. Our results stem from an approach, RESET, that couples the concept of aberrant methylation calling with an unbiased and systematic identification of silencing and enhancing events. To draw an analogy with computational methods to analyze DNA sequencing, the first step of our approach resembles somatic variant calling, whereas the second step is analogous to the search for candidate functional mutations.

By focusing of both hyper- and hypo-methylation at gene TSS, we could identify a class of tumors exhibiting both numerous hyper-methylated TSS and numerous hypo-methylated TSS. We refer to this phenotype as DNA methylation instability or DMI. In particular, we showed that tumor subtypes previously reported exhibiting a CpG island methylator phenotype (CIMP) are actually characterized by DMI. DMI was significantly associated with mutations of chromatin remodelers, strengthening a link between histone and DNA methylation and suggesting that both need to be considered to understand the impact of recurrent mutations of chromatin remodeling complexes in cancer (Lu and Allis, 2017; Suvà et al., 2013).

By exploring the interplay between genetic and epigenetic alterations, we found examples of functional convergence. Indeed, silencing and enhancing events co-occurrent with *IDH1* and *CTNNB1* mutations frequently affected genes involved in their respective pathways: cell metabolism and Wnt-signaling. Silenced and enhanced events were further enriched for genes involved in apoptosis, transcriptional regulation, especially Zinc-finger proteins, cell metabolism, and for CG antigens. Inhibitors of mitochondrial ROS generation were frequently silenced across multiple tumor types, potentially favoring survival signals in stress conditions (López-Otín et al., 2013). Interestingly, epigenetically enhanced expression of CG-antigens was significantly associated with sensitivity to anti-PD-1, but not anti-CTLA4, in skin melanoma. In light of recent studies on the use of demethylating agents in the clinic, these results lead to the suggestive hypothesis that therapeutic reactivation of CG-antigens could potentiate the effectiveness of the anti-PD1 checkpoint inhibitor (Siebenkäs et al., 2017), whereas it could be deleterious in combination with anti-CTLA4 (Shukla et al., 2018), and offer cancer-specific targets to therapeutic approaches based on T cells engineering.

The pilot study that we conducted on Wilms tumors demonstrates the potential for using our approach to explore clinically relevant epigenetic modifications in pediatric tumors. Pediatric tumors typically harbor few genetic alterations and are mostly associated with epigenetic abnormalities affecting their differentiation potential. Accordingly, they may represent a particularly relevant setting for systematic investigation of aberrant DNA methylation patterns.

Finally, here, we explored modification of DNA methylation with respect to normal tissues and associated with *in cis* changes of gene expression. However, adjacent normal tissues not necessarily represent a tumor's cell of origin, and expression changes might not be direct consequence of DNA methylation

contoured bars are for negative values, i.e., %FH > %DA, black filled bars are for positive values, i.e., %DA > %FH). Significant differences (Fisher's exact test p value < 0.05) are indicated by a star. Pathway annotations for silencing and enhancing events are shown on the right (blue dots for annotations of silenced genes and red dots for annotations of enhanced genes).

(E–G) Kaplan-Meier survival analysis comparing FHWT cases with epigenetic silencing of *TNFRSF10D* (E), *TRIP4* (F), or *NMRAL1* (G) (blue curves) versus wild-type cases (gray curves). Altered and wild-type groups are tested for significant different survival by log-rank test (p values are reported).

(H) Kaplan-Meier survival analysis comparing FHWT cases with epigenetic silencing of *NMRAL1* and at least one of *TNFRSF10D* and *TRIP4* (blue curve), FHWT cases with epigenetic silencing of *NMRAL1* only (light blue curve), and wild-type cases (gray curve). Each group pair is tested for significant different survival by log-rank test (p values are reported).

See also Figure S6 and Table S6.

changes and/or not be a necessary condition for DNA methylation changes to be functional (Widschwendter et al., 2007). While functional investigations of specific targets will be required to validate the cancer specificity and causal relationships of epigenetic and transcriptional changes, DNA methylation signatures could be designed to infer the cell of origin of different tumor subtypes enabling a more robust distinction between cancer-driven and cancer-predisposing events (Capper et al., 2018).

Overall, DNA methylation constitutes a simple mechanism to control gene function in both normal and malignant conditions. Systematic investigations of how cancer cells exploit this mechanism to deregulate specific targets and processes can fill a major gap in our understanding of disease manifestation, by capturing and functionally implicating cancer-associated methylation events and exploiting the therapeutic opportunities they offer.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines used in this study
  - Cell culture and experimental conditions
  - 5-AZAdC treatment
  - RNA isolation and RT-PCR
- METHODS DETAILS
  - Data collection
  - The RESET algorithm
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Association between genomic alterations and percentage of hyper/hypo methylated probes
  - DNA Methylation Inversion (DMI) score calculation and genomic alteration enrichment in DMI samples
  - Association between EPIC cell type composition and epigenetic silencing and enhancing events
  - Gene Ontology (GO) Terms enrichment analysis
  - Gene expression signature enrichment analysis
  - Correlation Analysis between gene expression signatures
  - Survival analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and six tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.09.082.

ACKNOWLEDGMENTS

We wish to thank Elisa Oricchio and Ping-Chin Ho for insightful feedback on both the content and form of this manuscript. We also thank Elizabeth J. Perlman for her insight on the analysis of the Wilms tumor cohort and Eliezer M. Van Allen for providing the data on anti-CTLA4 treated skin melanomas. This work was supported by the Swiss League Against Cancer Foundation (KFS-3983-08-2016). M.M. is supported by the Swiss National Science Foundation (SNSF) (310030_169519) and G.C. is also supported by the Gabriella Giorgi-Cavaglieri Foundation.

AUTHOR CONTRIBUTIONS

Conceptualization, S.S., M.M., and G.C.; Methodology, S.S., M.M., and G.C.; Software, S.S. and M.M.; Formal Analysis, S.S. and M.M.; Data Curation, S.S., M.M., N.R., and G.C.; Writing, S.S., M.M., N.R., D.H., and G.C.; Supervision, D.H. and G.C.; Project Administration, G.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 24, 2018
Revised: August 29, 2018
Accepted: September 25, 2018
Published: October 23, 2018

REFERENCES

Alaminos, M., Dávalos, V., Ropero, S., Setién, F., Paz, M.F., Herranz, M., Fraga, M.F., Mora, J., Cheung, N.-K.V., Gerald, W.L., and Esteller, M. (2005). EMP3, a myelin-related gene located in the critical 19q13.3 region, is epigenetically silenced and exhibits features of a candidate tumor suppressor in glioma and neuroblastoma. Cancer Res. *65*, 2565–2571.

Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C.C., Mundstein, A.S., Cohen, T., Chen, Y.-T., Chua, R., Gurung, S., Gnjatic, S., et al. (2009). CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res. *37*, D816–D819.

Amabile, G., Di Ruscio, A., Müller, F., Welner, R.S., Yang, H., Ebralidze, A.K., Zhang, H., Levantini, E., Qi, L., Martinelli, G., et al. (2015). Dissecting the role of aberrant DNA methylation in human leukaemia. Nat. Commun. *6*, 7091.

Anastasakis, D., Skeparnias, I., Shaukat, A.-N., Grafanaki, K., Kanellou, A., Taraviras, S., Papachristou, D.J., Papakyriakou, A., and Stathopoulos, C. (2016). Mammalian PNLDC1 is a novel poly(A) specific exonuclease with discrete expression during early development. Nucleic Acids Res. *44*, 8908–8920.

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al.; MC3 Working Group; Cancer Genome Atlas Research Network (2018). Comprehensive characterization of cancer driver genes and mutations. Cell *173*, 371–385.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. Nat. Rev. Cancer *11*, 726–734.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678.

Borgal, L., Habbig, S., Hatzold, J., Liebau, M.C., Dafinger, C., Sacarea, I., Hammerschmidt, M., Benzing, T., and Schermer, B. (2012). The ciliary protein nephrocystin-4 translocates the canonical Wnt regulator Jade-1 to the nucleus to negatively regulate β-catenin signaling. J. Biol. Chem. *287*, 25370–25380.

Brat, D.J., Verhaak, R.G., Aldape, K.D., Yung, W.K., Salama, S.R., Cooper, L.A., Rheinbay, E., Miller, C.R., Vitucci, M., Morozova, O., et al.; Cancer

Genome Atlas Research Network (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N. Engl. J. Med. *372*, 2481–2498.

Cairns, R.A., Harris, I.S., and Mak, T.W. (2011). Regulation of cancer cell metabolism. Nat. Rev. Cancer *11*, 85–95.

Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

Cancer Genome Atlas Network (2015). Genomic classification of cutaneous melanoma. Cell *161*, 1681–1696.

Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.

Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature *489*, 519–525.

Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of gastric adenocarcinoma. Nature *513*, 202–209.

Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., et al. (2018). DNA methylation-based classification of central nervous system tumours. Nature *555*, 469–474.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.

Challa-Malladi, M., Lieu, Y.K., Califano, O., Holmes, A.B., Bhagat, G., Murty, V.V., Dominguez-Sola, D., Pasqualucci, L., and Dalla-Favera, R. (2011). Combined genetic inactivation of β2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. Cancer Cell *20*, 728–740.

Chiarugi, A., Dölle, C., Felici, R., and Ziegler, M. (2012). The NAD metabolome–a key determinant of cancer cell biology. Nat. Rev. Cancer *12*, 741–752.

Cironi, L., Petricevic, T., Fernandes Vieira, V., Provero, P., Fusco, C., Cornaz, S., Fregni, G., Letovanec, I., Aguet, M., and Stamenkovic, I. (2016). The fusion protein SS18-SSX1 employs core Wnt pathway transcription factors to induce a partial Wnt signature in synovial sarcoma. Sci. Rep. *6*, 22113.

Cox, A.G., Hwang, K.L., Brown, K.K., Evason, K., Beltz, S., Tsomides, A., O'Connor, K., Galli, G.G., Yimlamai, D., Chhangawala, S., et al. (2016). Yap reprograms glutamine metabolism to increase nucleotide biosynthesis and enable liver growth. Nat. Cell Biol. *18*, 886–896.

Degli-Esposti, M.A., Dougall, W.C., Smolak, P.J., Waugh, J.Y., Smith, C.A., and Goodwin, R.G. (1997). The novel receptor TRAIL-R4 induces NF-kappaB and protects against TRAIL-mediated apoptosis, yet retains an incomplete death domain. Immunity *7*, 813–820.

Degli Esposti, D., Sklias, A., Lima, S.C., Beghelli-de la Forest Divonne, S., Cahais, V., Fernandez-Jimenez, N., Cros, M.-P., Ecsedi, S., Cuenin, C., Bouaoun, L., et al. (2017). Unique DNA methylation signature in HPV-positive head and neck squamous cell carcinomas. Genome Med. *9*, 33.

Dong, X., Deng, Q., Nie, X., Zhang, M., Jia, W., Chen, C., Xu, C., and Xu, R. (2015). Downregulation of HTATIP2 expression is associated with promoter methylation and poor prognosis in glioma. Exp. Mol. Pathol. *98*, 192–199.

Ehrlich, M. (2009). DNA hypomethylation in cancer cells. Epigenomics *1*, 239–259.

Esteller, M. (2007). Epigenetic gene silencing in cancer: the DNA hypermethylome. Hum. Mol. Genet. *16*, R50–R59.

Esteller, M. (2008). Epigenetics in cancer. N. Engl. J. Med. *358*, 1148–1159.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. *15*, 7–21.

Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al.; FANTOM Consortium and the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

Gadd, S., Huff, V., Walz, A.L., Ooms, A.H.A.G., Armstrong, A.E., Gerhard, D.S., Smith, M.A., Auvil, J.M.G., Meerzaman, D., Chen, Q.-R., et al. (2017). A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. Nat. Genet. *49*, 1487–1494.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. Science *300*, 489–492.

Herman, J.G., Latif, F., Weng, Y., Lerman, M.I., Zbar, B., Liu, S., Samid, D., Duan, D.S., Gnarra, J.R., Linehan, W.M., et al. (1994). Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc. Natl. Acad. Sci. USA *91*, 9700–9704.

Hinoue, T., Weisenberger, D.J., Pan, F., Campan, M., Kim, M., Young, J., Whitehall, V.L., Leggett, B.A., and Laird, P.W. (2009). Analysis of the association between CIMP and BRAF in colorectal cancer by DNA methylation profiling. PLoS ONE *4*, e8357.

Hinoue, T., Weisenberger, D.J., Lange, C.P.E., Shen, H., Byun, H.-M., Van Den Berg, D., Malik, S., Pan, F., Noushmehr, H., van Dijk, C.M., et al. (2012). Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res. *22*, 271–282.

Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2016). Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell *165*, 35–44.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A landscape of pharmacogenomic interactions in cancer. Cell *166*, 740–754.

Issa, J.-P. (2004). CpG island methylator phenotype in cancer. Nat. Rev. Cancer *4*, 988–993.

Jiao, Y., Widschwendter, M., and Teschendorff, A.E. (2014). A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Bioinformatics *30*, 2360–2366.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jung, D.-J., Sung, H.-S., Goo, Y.-W., Lee, H.M., Park, O.K., Jung, S.-Y., Lim, J., Kim, H.-J., Lee, S.-K., Kim, T.S., et al. (2002). Novel transcription coactivator complex containing activating signal cointegrator 1. Mol. Cell. Biol. *22*, 5203–5211.

Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., et al.; Cancer Genome Atlas Research Network (2013). Integrated genomic characterization of endometrial carcinoma. Nature *497*, 67–73.

Klutstein, M., Nejman, D., Greenfield, R., and Cedar, H. (2016). DNA methylation in cancer and aging. Cancer Res. *76*, 3446–3450.

Lachenmayer, A., Alsinet, C., Savic, R., Cabellos, L., Toffanin, S., Hoshida, Y., Villanueva, A., Minguez, B., Newell, P., Tsai, H.-W., et al. (2012). Wnt-pathway activation in two molecular classes of hepatocellular carcinoma and experimental modulation by sorafenib. Clin. Cancer Res. *18*, 4997–5007.

Lahtz, C., and Pfeifer, G.P. (2011). Epigenetic changes of DNA repair genes in cancer. J. Mol. Cell Biol. *3*, 51–58.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Lian, M., and Zheng, X. (2009). HSCARG regulates NF-kappaB activation by promoting the ubiquitination of RelA or COMMD1. J. Biol. Chem. *284*, 17998–18006.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell *167*, 1170–1187.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217.

Lu, C., and Allis, C.D. (2017). SWI/SNF complex in cancer. Nat. Genet. *49*, 178–179.

Lu, C., and Thompson, C.B. (2012). Metabolic regulation of epigenetics. Cell Metab. *16*, 9–17.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al.; Cancer Genome Atlas Research Network (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell *173*, 338–354.

McKenna, E.S., and Roberts, C.W.M. (2009). Epigenetics and cancer without genomic instability. Cell Cycle *8*, 23–26.

Mina, M., Raynaud, F., Tavernari, D., Battistello, E., Sungalee, S., Saghafinia, S., Laessle, T., Sanchez-Vega, F., Schultz, N., Oricchio, E., and Ciriello, G. (2017). Conditional selection of genomic alterations dictates cancer evolution and oncogenic dependencies. Cancer Cell *32*, 155–168.

Noushmehr, H., Weisenberger, D.J., Diefes, K., Phillips, H.S., Pujara, K., Berman, B.P., Pan, F., Pelloski, C.E., Sulman, E.P., Bhat, K.P., et al.; Cancer Genome Atlas Research Network (2010). Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell *17*, 510–522.

Parikh, J.R., Klinger, B., Xia, Y., Marto, J.A., and Blüthgen, N. (2010). Discovering causal signaling pathways through gene-expression patterns. Nucleic Acids Res. *38*, W109-17.

Pencheva, N., Tran, H., Buss, C., Huh, D., Drobnjak, M., Busam, K., and Tavazoie, S.F. (2012). Convergent multi-miRNA targeting of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis. Cell *151*, 1068–1082.

Ploeger, C., Waldburger, N., Fraas, A., Goeppert, B., Pusch, S., Breuhahn, K., Wang, X.W., Schirmacher, P., and Roessler, S. (2016). Chromosome 8p tumor suppressor genes SH2D4A and SORBS3 cooperate to inhibit interleukin-6 signaling in hepatocellular carcinoma. Hepatology *64*, 828–842.

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D.E., and Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. eLife *6*, e26476.

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafinia, S., et al.; Cancer Genome Atlas Research Network (2018). Oncogenic signaling pathways in The Cancer Genome Atlas. Cell *173*, 321–337.

Sepulveda, J.L., Gutierrez-Pajares, J.L., Luna, A., Yao, Y., Tobias, J.W., Thomas, S., Woo, Y., Giorgi, F., Komissarova, E.V., Califano, A., et al. (2016). High-definition CpG methylation of novel genes in gastric carcinogenesis identified by next-generation sequencing. Mod. Pathol. *29*, 182–193.

Shen, H., and Laird, P.W. (2013). Interplay between the cancer genome and epigenome. Cell *153*, 38–55.

Shukla, S.A., Bachireddy, P., Schilling, B., Galonska, C., Zhan, Q., Bango, C., Langer, R., Lee, P.C., Gusenleitner, D., Keskin, D.B., et al. (2018). Cancer-germline antigen expression discriminates clinical outcome to CTLA-4 blockade. Cell *173*, 624–633.

Siebenkäs, C., Chiappinelli, K.B., Guzzetta, A.A., Sharma, A., Jeschke, J., Vatapalli, R., Baylin, S.B., and Ahuja, N. (2017). Inhibiting DNA methylation activates cancer testis antigens and expression of the antigen processing and presentation machinery in colon and ovarian cancer cells. PLoS ONE *12*, e0179501.

Simpson, A.J.G., Caballero, O.L., Jungbluth, A., Chen, Y.-T., and Old, L.J. (2005). Cancer/testis antigens, gametogenesis and cancer. Nat. Rev. Cancer *5*, 615–625.

Stevanović, S., Pasetto, A., Helman, S.R., Gartner, J.J., Prickett, T.D., Howie, B., Robins, H.S., Robbins, P.F., Klebanoff, C.A., Rosenberg, S.A., and Hinrichs, C.S. (2017). Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. Science *356*, 200–205.

Suvà, M.L., Riggi, N., and Bernstein, B.E. (2013). Epigenetic reprogramming in cancer. Science *339*, 1567–1570.

Swisher, E.M., Harrell, M.I., Lin, K., Coleman, R.L., Konecny, G.E., Tinker, A.V., O'Malley, D.M., McNeish, I., and Kaufmann, S.H. (2017). BRCA1 and RAD51C promoter hypermethylation confer sensitivity to the PARP inhibitor rucaparib in patients with relapsed, platinum-sensitive ovarian carcinoma in ARIEL2 Part 1. Gynecol. Oncol. *145*, 5.

Teschendorff, A.E., Jones, A., Fiegl, H., Sargent, A., Zhuang, J.J., Kitchener, H.C., and Widschwendter, M. (2012). Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. Genome Med. *4*, 24.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.-P.J. (1999). CpG island methylator phenotype in colorectal cancer. Proc. Natl. Acad. Sci. USA *96*, 8681–8686.

Toyota, M., Sasaki, Y., Satoh, A., Ogi, K., Kikuchi, T., Suzuki, H., Mita, H., Tanaka, N., Itoh, F., Issa, J.-P.J., et al. (2003). Epigenetic inactivation of CHFR in human tumors. Proc. Natl. Acad. Sci. USA *100*, 7818–7823.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA *98*, 5116–5121.

Van Allen, E.M., Miao, D., Schilling, B., Shukla, S.A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M.H.G., Goldinger, S.M., et al. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science *350*, 207–211.

Van Tongelen, A., Loriot, A., and De Smet, C. (2017). Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes. Cancer Lett. *396*, 130–137.

Weller, M., Stupp, R., Reifenberger, G., Brandes, A.A., van den Bent, M.J., Wick, W., and Hegi, M.E. (2010). MGMT promoter methylation in malignant gliomas: ready for personalized medicine? Nat. Rev. Neurol. *6*, 39–51.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., and Laird, P.W. (2007). Epigenetic stem cell signature in cancer. Nat. Genet. *39*, 157–158.

Williams, R.D., Chagtai, T., Alcaide-German, M., Apps, J., Wegert, J., Popov, S., Vujanic, G., van Tinteren, H., van den Heuvel-Eibrink, M.M., Kool, M., et al. (2015). Multiple mechanisms of MYCN dysregulation in Wilms tumour. Oncotarget *6*, 7232–7243.

Wu, M., and Ho, S.-M. (2004). PMP24, a gene identified by MSRF, undergoes DNA hypermethylation-associated gene silencing during cancer progression in an LNCaP model. Oncogene *23*, 250–259.

Yang, Z., Wong, A., Kuh, D., Paul, D.S., Rakyan, V.K., Leslie, R.D., Zheng, S.C., Widschwendter, M., Beck, S., and Teschendorff, A.E. (2016). Correlation of an epigenetic mitotic clock with cancer risk. Genome Biol. *17*, 205.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. *12*, 7–18.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
5-Aza-2′-deoxycytidine	Sigma	A3656
Critical Commercial Assays		
RNeasy kit	QIAGEN	74106
PrimeScript RT Master Mix	Takara Bio Europe SAS	RR036A
Rotor-Gene SYBR Green PCR Kit	QIAGEN	204076
Deposited Data		
TCGA cohort - PanCanAtlas	TCGA PanCanAtlas data; Sanchez-Vega et al., 2018	https://portal.gdc.cancer.gov/
TCGA cohort - FireHose	TCGA Consortium	https://gdac.broadinstitute.org/
TCGA cohort - cBioPortal	TCGA Consortium	http://www.cbioportal.org/
FANTOM5 TSS data	Forrest et al., 2014	http://fantom.gsc.riken.jp/5/
TARGET cohort	TARGET initiative website	https://ocg.cancer.gov/programs/target/data-matrix
Gene Ontology and GO Annotation	Gene Ontology Consortium	http://www.geneontology.org/
GDSC dataset (GSE32323, GSE29060, GSE5816)	Genomics of Drug Sensitivity in Cancer	https://www.cancerrxgene.org/
GTEx database	Gene Expression Database	http://gtexportal.org/home/datasets
Genome Annotation (hg19)	UCSC Genome Browser	http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/
RNA-seq cell line datasets	Gene Expression Omnibus	GEO: GSE32323, GSE29060, and GSE5816
Experimental Models: Cell Lines		
MDA.MB.453	ATCC	HTB-131; RRID:CVCL_0418
HCC70	ATCC	CRL-2315; RRID:CVCL_1270
UACC.893	ATCC	CRL-1902; RRID:CVCL_1782
BT.20	ATCC	HTB-19; RRID:CVCL_0178
BT.474	ATCC	HTB-20; RRID:CVCL_0179
NCI.H522	ATCC	CRL-5810; RRID:CVCL_1567
SW1573	ATCC	CRL-2170; RRID:CVCL_1720
PC.3	ATCC	CRL-1435; RRID:CVCL_0035
DU.145	ATCC	HTB-81; RRID:CVCL_0035
VCAP	ATCC	CRL-2876; RRID:CVCL_2235
FADU	ATCC	HTB-43; RRID:CVCL_1218
CHL.1	ATCC	CRL-1935; RRID:CVCL_4585
A375	ATCC	CRL-1619; RRID:CVCL_0132
Oligonucleotides		
PXMP4 Frw Primer: ACCTGGCACGGTTTGTGTT	This paper	N/A
PXMP4 Rev Primer: CCACCTGGGTTCAGGGATG	This paper	N/A
SYCP2 Frw Primer: TTCTGCTGGTCATACATGATGTC	This paper	N/A
SYCP2 Rev Primer: GAACAAATGCGAGGTACGAAAC	This paper	N/A
PNLDC1 Frw Primer: GGCAGGTCTGGACATAGAGTT	This paper	N/A
PNLDC1 Rev Primer: CGGGTCTTTAGATACCACTCCG	This paper	N/A
H2AFJ Frw Primer: ACGAGGAGTTAAACAAGCTGC	This paper	N/A
H2AFJ Rev Primer: TCATTTGCTCTTCGTCTTCTGAC	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GAPDH Frw Primer: GGAGCGAGATCCCTCCAAAAT	This paper	N/A
GAPDH Rev Primer: GGCTGTTGTCATACTTCTCATGG	This paper	N/A
Software and Algorithms		
RESET	This paper	http://ciriellolab.org/reset/reset.html Mendeley Database: https://doi.org/10.17632/xgfm8y6mmf.1.
SELECT	Mina et al., 2017	http://ciriellolab.org/select/select.html http://ciriellolab.org/
EPIC	Racle et al., 2017	N/A
xCELL	Aran et al., 2017	N/A
ABSOLUTE	Carter et al., 2012	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and software should be directed to and will be fulfilled by the Lead Contact, Giovanni Ciriello (giovanni.ciriello@unil.ch).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines used in this study

All the human cancer cell lines used in this study (Breast Cancer cell lines: MDA.MB.453, HCC70, UACC.893,BT.20, BT.474; Lung adenocarcinoma cell lines: NCI.H522, SW1573; Prostate adenocarcinoma cell lines: PC.3, DU.145, VCAP; Head and Neck squamous carcinoma cell lines: FADU; Melanoma cell lines: CHL.1, A375) were purchased from ATCC (supplied by LGC Standards GmbH, Wesel, Germany).

Cell culture and experimental conditions

The cell lines were treated according to their recommended medium (http://www.atcc.org/products?geo_country=us) and all the culture media contained 10% FBS. For 5-Aza-2′-deoxycytidine (5-AZAdC, A3656 Sigma) treatment, cells were plated at 10^5 cells/ml in 100mm culture plate and allowed to attach and grow for 24 hours before the treatment.

5-AZAdC treatment

Cells were subjected to a 7-day treatment during which the medium was changed every day, and 5-AZAdC was added in aqueous solution at a final concentration of 1mM.

RNA isolation and RT-PCR

mRNA extractions from cultured cell lines were performed with the QIAGEN RNeasy kit (74106). Reverse transcription into cDNA was performed with PrimeScript RT Master Mix (Takara Bio Europe SAS). RT-PCR was carried out with probes designed according to PrimerBank database (https://pga.mgh.harvard.edu/primerbank/index.html) and synthesized by Microsynth AG, Balgach, Switzerland.

Primers

PXMP4 Frw Primer: ACCTGGCACGGTTTGTGTT
PXMP4 Rev Primer: CCACCTGGGTTCAGGGATG
SYCP2 Frw Primer: TTCTGCTGGTCATACATGATGTC
SYCP2 Rev Primer: GAACAAATGCGAGGTACGAAAC
PNLDC1 Frw Primer: GGCAGGTCTGGACATAGAGTT
PNLDC1 Rev Primer: CGGGTCTTTAGATACCACTCCG
H2AFJ Frw Primer: ACGAGGAGTTAAACAAGCTGC
H2AFJ Rev Primer: TCATTTGCTCTTCGTCTTCTGAC
GAPDH Frw Primer: GGAGCGAGATCCCTCCAAAAT
GAPDH Rev Primer: GGCTGTTGTCATACTTCTCATGG

All RT-PCR reactions were performed in triplicates using the 7900HT Fast RT-QPCT System (Applied Biosystems™), and the results were normalized to the housekeeping gene *GAPDH*.

METHODS DETAILS

Data collection

TCGA cohort data collection

Molecular and clinical data for 24 tumor types profiled by The Cancer Genome Atlas (TCGA) consortium were collected in July 2017 from FireHose (https://gdac.broadinstitute.org/) and cBioPortal (http://www.cbioportal.org/) (Cerami et al., 2012) data repositories. Only the data publicly available at that time were used in our study (Table S1). The dataset includes somatic point mutation (whole exome sequencing), copy number changes (Illumina SNP6 array), gene expression profiling (Illumina HiSeq, RSEM normalized counts), and DNA methylation data (Illumina Infinium array). DNA methylation data were generated by Infinium HM450 array (485,577 CpG site targeting probes), except for the ovarian cancer cohort that was analyzed with the Infinium HM27 array (27,578 CpG site targeting probes). All data was generated and processed by The Cancer Genome Atlas research network as described in the corresponding manuscripts (see Table S1). The latest update regarding tumor types and subtype annotations were taken from (Sanchez-Vega et al., 2018).

TARGET cohort data collection

Molecular and clinical data for the cohort of Wilms tumors were collected from TARGET initiative website (https://ocg.cancer.gov/programs/target/data-matrix). It includes somatic mutation (whole exome sequencing and whole genome sequencing), copy number changes (Affymetrx 6.0 SNP arrays), gene expression profiling (Illumina HiSeq, RPKM normalized counts), and DNA methylation data (Illumina Infinium HM450 array). Methods, including data generation and processing, and validate somatic mutations used in Figure 6 were reported in Gadd et al. (2017).

Other datasets

DNA methylation and mRNA expression data for the cancer cell lines used in this manuscript (Figures 3F–3I and S3D–S3H) were obtained from the GDSC database (https://www.cancerrxgene.org) and Gene Expression Omnibus (GEO): GSE32323, GSE29060, and GSE5816. mRNA expression values for normal tissue samples (Figure S5A) was obtained from the GTEx database (http://gtexportal.org/home/datasets). Genetic alterations (somatic point mutations, copy number amplifications and deletions) and the list of 505 recurrent events used in the enrichment had been previosly derived from the same TCGA datasets used in this manuscript (Mina et al., 2017). Genome annotations used for the methylation probes characterization (Figures S1C–S1E) was obtained from UCSC Genome Browser (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/).

DNA methylation probe selection and analysis

DNA methylation probe selection

Probes were further filtered to include only those mapping to TSSs including both canonical TSS within 5′UTRs and unconventional exonic TSS, as defined by cap analysis of gene expression (CAGE) by the FANTOM5 consortium (Forrest et al., 2014) (Table S2 and Figure S1A). The genomic intervals flanking the TSSs (300 nt upstream to 300 nt downstream) were considered as gene promoter regions (Figure S1A). Only probes mapping to a gene promoter region were considered in the downstream analysis.

Analysis of DNA methylation probe distribution

Probed loci were mapped to the canonical isoform of the corresponding gene according to FANTOM annotations or genomic proximity to generate unique location distributions. Probes genomic locations were characterized by defining 7 classes: 1) upstream: up to 1kb upstream of the TSS, 2) 5′UTR: interval between the TSS and coding start site, 3) exonic: inside the exons, 4) intronic: inside the introns, 5) 3′UTR: interval between coding end site and transcription end site (TES), 6) downstream: up to 1kb downstream of TES, 7) intergenic: more than 1kb upstream or more than 1kb downstream of the TSS or TES, respectively. We manually verified that all probes corresponding to silenced and enhanced genes identified by RESET and classified as intergenic were actually mapping to non-canonical isoforms, hence the ‘‘intergenic’’ category has been renamed to ‘‘non-canonical’’ for these distributions. The set of selected TSS probes included the majority of probes less than 1KB upstream or at 5′UTR of the canonical isoform of the corresponding gene, whereas only a minor fraction was downstream (< 1KB) or within the 3′UTR of the canonical isoform (Figure S1C). Selected probes that fell within the gene body were preferentially exonic, rather than intronic, despite the latter type was overall more frequent (Figure S1C). Probes associated with silenced and enhanced targets followed the same distribution (Figures S1D and S1E). Overall, less than 1% of the selected probes mapped to Alu and LINE-1 repeats.

The RESET algorithm

RESET analyzes large-scale DNA methylation sample cohorts to (i) test for the presence of aberrant DNA methylation at gene TSS and (ii) assess whether aberrant methylation states lead to epigenetic silencing or enhancing of gene expression. RESET is based on four steps: 1) probe selection and modeling, 2) aberrant methylation states analysis, 3) epigenetic silencing and enhancing (ESE) score evaluation, and 4) statistical significance analysis of the ESE scores. The output consists of (i) sample-level methylation states for each probe mapping to a gene promoter, i.e., normal, hyper-methylated, or hypo-methylated, and (ii) the collection of silencing and enhancing events with a statistically significant effect on gene transcription. A schematic representation of the pipeline is shown in Figures 1A and 3A.

Input data

RESET requires methylation data of both normal and tumor samples (unmatched), and transcriptomic profiles of the tumor samples. Additionally, a map between DNA methylation array probes and gene TSS must be provided.

Step 1: Probe selection and modeling of DNA methylation in normal samples

In bisulfite-genomic DNA sequencing analysis, the extent of DNA methylation in a CpG locus is reported as the ratio of the number of methylated probed molecules over the total number of probed molecules. These ratios assume values between 0 and 1, can be characterized by a β-distribution and are thus referred to as β-values. Actual β-values for all probes in our dataset followed a bimodal distribution with modes in the low (< 0.1) and the high (> 0.8) extremes of the possible range. To ensure changes of DNA methylation in cancer samples are not due to the expected variability of DNA methylation measured at a specific locus, RESET discards probes with high variability in normal samples (Figure S2B), and defines two probe sets:

1. Low DNA Methylation probes: probes with mean β-values lower than 0.1 and standard deviation lower than 0.005 in normal samples.
2. High DNA Methylation probes: probes with mean β -values higher than 0.8 and standard deviation lower than 0.005 in normal samples.

RESET then uses these sets of DNA methylation probes with consistent status across normal samples to infer the expected distribution of values at each locus (probe specific distribution) and at all loci with the same status, i.e., either low or high DNA methylation (probe set distribution). The density function of a beta distribution is described by the two positive parameters, α and β:

$$f(x;\alpha,\beta)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},$$

where $B(a,b)=\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ and $\Gamma(z)$ is the gamma distribution. To determine both probe specific and probe set distributions, RESET estimates the α and β parameters for each probe x from the observed mean and variance of beta values observed in normal samples, according to the following formulas:

$$\text{Mean}_x=\frac{\alpha}{\alpha+\beta}$$

$$Var_x=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

The R library MASS to fit the global beta distribution models to the methylation beta values. The parameter space of the estimated β-distributions in cancer samples for our set of selected probes is consistent with peaked distributions around 0 for probes with low methylation in normal samples and around 1 for probes with high methylation in normal samples (Figures S1F–S1H).

Step 2. Methylation status analysis

To evaluate the DNA methylation status of a probe in a tumor sample, RESET compares its β-value to the probe specific and probe set distributions estimated in step 1. A probe x with methylation value y in a specific sample is considered hyper-methylated if the following conditions are met:

$$x \in \text{Low DNA methylation probe set}$$

$$\text{CDF}_{x-\text{probe specific } \beta \text{ distribution}}(y)\geq 0.995$$

$$\text{CDF}_{\text{Low}-\text{probe set } \beta \text{ distribution}}(y)\geq 0.995$$

where CDF is the cumulative density function of the distribution fitted in step 1.

Similarly, a probe x with methylation value y is considered hypo-methylated if

$$x \in \text{High DNA methylation probe set}$$

$$\text{CDF}_{x-\text{probe specific } \beta \text{ distribution}}(y)\leq 0.005$$

$$\text{CDF}_{\text{High}-\text{probe set } \beta \text{ distribution}}(y)\leq 0.005$$

The output of this step is a discrete description of the DNA methylation state of each probe in each tumor sample, where a probe can be normal-methylated, hyper-methylated, or hypo-methylated, with the latter two also defined as aberrant DNA methylation states.

Step 3. Epigenetic silencing and enhancing score

After determining the methylation status of each probe in each tumor sample, RESET evaluates the effect of aberrant DNA methylation states on gene transcription. Specifically, RESET quantifies the associations between hyper- (hypo-) methylation of a probe *p* and a significant decrease (increase) of mRNA expression of the gene *g* whose TSS maps to *p*. In all the analyses that follow, RNA-seq RSEM normalized counts have been rank transformed using a qq-normalization (R function qqnorm). To evaluate the association between DNA methylation and gene transcription, we defined a score as the product of two distances: the centroid distance and the rank distance.

The Centroid Distance (CD) is meant to capture the global difference in gene expression between tumor samples where a probe is hyper-/hypo-methylated (altered samples) and tumor samples where the probe is normal-methylated (wild-type or WT samples). CD is defined as the geometric mean of the centroids of altered and wild-type tumor samples in the 2D space defined by the β-values (x) of *p* and mRNA expression values (y) of *g* across all the samples:

$$CD_p = \sqrt{|\text{centroid}(x_{WT}) - \text{centroid}(x_{altered})| \times |\text{centroid}(y_{WT}) - \text{centroid}(y_{altered})|}$$

RNA-seq qq-transformed values are here rescaled to the [0,1] interval to be within the same range of values as beta values.

The Rank Distance (RD) is designed to evaluate the inverse correlation between DNA methylation and mRNA expression. Here, samples are first binned into four classes ($B_{1\text{-}4}$) according to their beta values falling within the intervals: [0,0.25], (0.25,0.5], (0.5,0.75], and (0.75,1]. In an ideal situation, where gene expression is perfectly anti-correlated with DNA methylation values, all samples in the bin *i* have lower mRNA expression values than samples in bin *i-1* and higher mRNA expression values than samples in bin *i+1*. Formally, if N_k, $k \in \{1,2,3,4\}$, is the number of samples in bin k, r_s the ranking of sample *s* according to its expression level (from higher to lower) and $b_s^*$the ideal bin [1-4] to which *s* is assigned based on its DNA methylation value, then:

$$r_s \in [1,\ N_1]\ \forall\ s \,|\, b_s^* = 1$$

$$r_s \in [N_1 + 1,\ N_1 + N_2]\ \forall\ s \,|\, b_s^* = 2$$

$$r_s \in [N_1 + N_2 + 1,\ N_1 + N_2 + N_3]\ \forall\ s \,|\, b_s^* = 3$$

$$r_s \in [N_1 + N_2 + N_3 + 1,\ N_1 + N_2 + N_3 + N_4]\ \forall\ s \,|\, b_s^* = 4$$

Note that the rank of samples within a bin is irrelevant. The Rank Distance quantifies the distance of the observed associations between r_s and b_s for all samples *s* from the ideal scenario just described and it is defined as follow:

$$RD_p = \sum_{k=1}^{4}\left(1 + \frac{|I_k|}{N_k + \alpha} - \frac{\sum_{i=1}^{|O_k|}|b_i^* - b_i|}{N_k + \alpha}\right),$$

where I_k is the set of samples in B_k when B_k corresponds to their ideal bin $(b_i^* = b_i)$, whereas O_k is the set of samples in B_k when B_k is not their ideal bin $(b_i^* \neq b_i)$. Finally, a positive constant is added to the denominator of each component, to limit the variance of RD when N_k is small. In this study we set $\alpha = 4$.

Step 4. Statistical significance analysis

To determine the significance of the observed scores, we adopted the strategy previously proposed in the Significance Analysis of Microarrays (SAM) (Tusher et al., 2001). This technique is based on a permutation of the observed data to estimate the false discovery rate (FDR) associated to a significance threshold of the measured effect size. Precisely, for a threshold t, the number of False Positive (FP_t) out of the observed number of measurements exceeding t (set of positive solution or P_t) is estimated as the average number of measurements exceeding t after multiple random permutations of the input values, and the FDR of t is defined as:

$$FDR_t = \frac{FP_t}{P_t}.$$

In our analysis, P_t is determined by counting the number of probes that obtain a score higher than a given threshold t. To permute the expression data, we pooled the normalized expression values of all genes in samples belonging to the same DNA methylation bin B_k, with k = 1,2,3,4. The distributions of normalized mRNA expression values in each bin confirmed an overall inverse correlation between DNA methylation and mRNA expression (data not shown). For a given probe p and sample s, the corresponding expression value is then sampled from the pool corresponding to the bin associated to s (b_s). Once expression values are sampled for all samples, the score of p is re-computed. We repeat this process 100 times to generate 100 random distributions of scores and from these distributions we determine FP_t as described above. The smallest t that guarantees an FDR < 0.1 in all our analyses is chosen as the significance cut-off threshold t*, and genes matched to probes with a score greater than t* are considered as epigenetically silenced

(if p is a hyper-methylated probe) or epigenetically enhanced (if p is a hypo-methylated probe). In this work, we additionally required the score to be always greater than 1.5, to guarantee a minimum effect size.

Comaprison of RESET to other methods

Previous approaches to identify epigenetically silenced and enhanced genes had been based either on defining altered and not-altered samples based on an arbitrary β-value threshold (hyper-methylated if β-value > 0.3, hypo-methylated if β-value < 0.1) and then testing the two groups for differential expression (Hinoue et al., 2012; Cancer Genome Atlas Network, 2015), or on identifying significant anti-correlation between β-values and mRNA expression (Sepulveda et al., 2016). Independently of the adopted test, common pitfalls of all these methods are that their results are invariably dependent on the alteration frequency, and, for large cohorts, highly significant p values can be obtained even for small effect sizes, leading to large numbers of significant hits even after correction for multiple testing (Tusher et al., 2001). As an example, we implemented and applied these methodologies to the luminal breast cancer cohort and test all genes for epigenetic silencing. Independently of the adopted test, they all show a string dependency with alteration frequency and return more than 1,000 significant hits with an FDR < 0.01 (Figure S3A). Overall, these methods rank candidate silenced and enhanced genes based on the obtained p value, even though altered and non-altered sample sets have variable sizes that depend on the gene being tested. As a consequence, genes altered in ~50% of the cases will be more likely to be top scoring because their sample size guarantees the maximum statistical power (Figure S3A). This is exemplified by computing the best possible p value of a Wilcoxon test as a function of the size of the altered sample set (Figure S3B). In contrast, RESET put an emphasis on the strength (or effect size) of the inverse association between DNA methylation and mRNA expression, here defined by the RESET score, and it directly estimates the false discovery rate associated with specific scores to determine a significance threshold. Importantly, this score is independent of the alteration frequency and sample size (Figure S3C). For completeness, t test and Wilcoxon test based analyses were repeated using the methylation calls generated by the step 1 of RESET, yielding similar results. The methodologies were run on both the FPKM/RSEM normalized count data and the qq-normalized gene expression. We should note that the use of t test with RSEM values or log-transformed RSEM-values should be avoided given the non-normal distribution of these data.

To evaluate the performance of RESET and other conventional methods using experimental models, we exploited publicly available RNA sequencing datasets regarding cell lines treated with 5-AZAdC (Figure S3E). Methylome data for non-treated cell lines were also downloaded from GDSC database. First, for each cell line, we obtained a list of reactivated genes upon 5-AZAdC treatment. For gene x to be considered as reactivated, it needed to be hypermethylated in non-treated condition (average beta-value of all probes mapping to gene x TSS be larger than 0.7), and have expression fold-change bigger than 1.5 (5-AZAdC treated versus non-treated). Second we applied all the methods to the cell lines' tissue-of-origins in TCGA cohort (CRC and LUAD), and derived the silenced genes proposed by each method. Percentage of reactivated genes for each cell line/method was obtained by intersecting the list of reactivated gene in each cell line and the silenced genes proposed by each method (Figure S3E).

Application of RESET to the pan-cancer TCGA dataset

In this study, we ran RESET on the pan-cancer TCGA dataset considering the 170,953 probes that map to FANTOM5 TSSs. Given that only 14 tumor types have DNA methylation data available for normal tissue samples (BLCA, BRCA-Basal, BRCA-Non Basal, CRC, ESCA, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, THCA, UCEC), we defined a set of pan-cancer probes including only the probes consistently classified as Low (or High) DNA methylation probes across all normal tissues available. The results presented in the manuscript were obtained by running RESET separately on each tumor type using the pan-cancer probe set. For tumor types without methylation data for normal samples, the pool of normal samples from the all tissue types was used to build the reference methylation models. The aggregated set of pan-cancer epigenetic silencing and enhancing events was taken as the union of the results from each tumor type. Aberrant DNA methylation at genes scoring as significant in at least one tumor type was re-assessed in all cohorts, and each gene was considered as significant in all cohorts where the score was greater than 1.5. Finally, RESET was also run on each individual tumor type using all the probes (not only the pan-cancer probe set) selected at Step 1 based on the corresponding normal tissue samples (when available). These results thus include tissue-specific methylation events, i.e., silencing and enhancing of genes whose methylation status is tissue-dependent. While the discussion of these results is beyond the scope of this manuscript, all results are made available at http://ciriellolab.org/reset/reset.html.

Application of RESET to the TARGET dataset

The Wilms tumor cohort did not include methylation data for normal samples, hence we used for this analysis the previously-defined pan-cancer Low (or High) DNA methylation probe sets. Moreover, the pool of normal samples used in TCGA dataset for the tumors without normal methylation data was used to build the reference methylation models. In this analysis, enhancing events were selected with a score associated to FDR < 0.1.

QUANTIFICATION AND STATISTICAL ANALYSIS

All the statistical analyses described in this section were performed with R.

Association between genomic alterations and percentage of hyper/hypo methylated probes

We tested whether recurrent genomic alterations (Mina et al., 2017) are enriched on samples with a higher percentage of hyper/hypo methylated probes, separately for each individual tumor type. The Wilcoxon's rank-sum test (one sided) was used in the individual tumor type analysis, contrasting wild-type and altered samples. The same procedure was applied to test the enrichment of tumor

subtypes among samples with high fraction of hyper-/hypo- methylated probes (Figures 1B–1D). P values were corrected for multiple hypothesis testing with the Benjamini-Hochberg method.

DNA Methylation Inversion (DMI) score calculation and genomic alteration enrichment in DMI samples

The DMI score, designed with the aim of capturing the concomitant increase in both hyper- and hypo-methylation events, is defined as the F_β-measure) of hyper- (H) and hypo- (h) methylation frequencies:

$$DMI = \left(1+\beta^2\right)\frac{Hh}{\beta^2 H + h}.$$

We used $\beta = 2$ in our analyses, to compensate for the asymmetric distribution of hyper- and hypo-methylation frequencies. The enrichment of each genomic alteration on samples with high DMI score was studied both within each individual tumor type and at pan-cancer level. The Wilcoxon's rank-sum test (one sided) was used for the individual tumor type analyses, contrasting wild-type and altered samples. For the pan-cancer analysis we perfomed a type II ANOVA analysis separately for each genomic alteration, testing only alterations with at least 4 occurrences. The alteration occurrence in each sample was used as binary independent variable, whereas the DMI score was used as dependent variable. For the pan-cancer analysis we considered tumor type as co-factor, and for each particular alteration *x*, the tumor types with no occurrence of the alteration *x* were excluded from the ANOVA analysis. P values were corrected for multiple hypothesis testing with the Benjamini-Hochberg method.

Association between genomic alterations and epigenetic silencing and enhancing events

We performed enrichment analysis of genomic alterations at two different levels: global and event-specific. In the global analysis, we sought to understand whether specific genomic alterations were enriched in samples with high epigenetic silencing or enhancing events. In this case we repeated the enrichment analysis performed for the hyper/hypo methylated probes, described above. The analysis was performed separately on each single tumor type, using the same procedure, detailed above, for testing the enrichment of genomic alterations in DMI samples.

For the event-specific analysis, instead, we identified the patterns of co-occurrence and mutual-exclusivity between individual epigenetic silencing and enhancing events identified by RESET and the set of 505 recurrent cancer-associated genetic alterations, using the SELECT algorithm (Mina et al., 2017). The analysis was performed in a pancan fashion, considering all the tumor samples together and using tumor type and subtype as covariates of the SELECT analysis to remove tumor type effects.

Estimation of samples' purity, cell type composition, chromosomal instability and stemness score: the following datasets/pipelines were used to evaluate different qualities of tumor samples:

- ABSOLUTE (Carter et al., 2012): estimation of samples purity and ploidy.
- EPIC (Racle et al., 2017): estimation of immune cells infiltration
- xCell (Aran et al., 2017): estimation of immune, microenvironment, and stroma scores
- Chromosomal instability was measure with (i) the number of amplified and deleted segments in each sample (normalized copy number value > 0.3, segment count), and (ii) the percentage of genes amplified or deleted (fraction AMP/DEL genes). The segment file from the TCGA Pan-Cancer Atlas cohort and the discrete gene-level copy number changes derived by GISTIC were used as input data.
- Stemness signatures were derived using a dataset provided in Malta et al. (2018).

Association between EPIC cell type composition and epigenetic silencing and enhancing events

Associations between epigenetic silencing and enhancing events and immune cell type composition estimated by EPIC were tested on each individual tumor type. For each epigenetic event, EPIC scores in altered samples were compared to score in non-altered samples by Wilcoxon one-tail test. P values were corrected for false discovery rate using the Benjamini-Hochberg procedure. Associations were considered significant if the obtain a q value < 0.1 and altered samples had score at least 1.5 fold higher than non-altered samples.

Gene Ontology (GO) Terms enrichment analysis

GO Terms enrichment analysis was performed using the online webservice based on the Molecular Signatures Database (MSigDB) (http://software.broadinstitute.org/gsea/msigdb/index.jsp). Enriched GO terms were defined as GO biological process (BP) and molecular function (MF) terms obtaining a FDR-adjusted P value < 0.01, retreiving a maximum of 100 terms.

Gene expression signature enrichment analysis

Single sample Gene Set Enrichment analysis (Barbie et al., 2009) (ssGSEA), implemented in the R package GSVA, was used to calculate an expression score for each gene expression signature and each sample. The default parameters from the GSVA package were used. In this study, we used the following gene signatures:

Mitochondrial-ROS: *GPX1, IVD, HSD17B4, ACADS, DECR1, CPT1A, CROT, PECI, NUDT19, ETFB, ALDH7A1, AASS, CYB5A, PECR, HSD17B8, NNT, ALDH1B1, CYB5R1, UQCRH, PRDX1, AIFM2, ALKBH3, P4HTM, CYB561, SCCPDH, TXNRD2.*

H_2O_2-signature: *PCF11, TMEFF1, NRIP1, SERPINB9, EFNB2, OXTR, ATP6V0E1, ATRX, LUC7L3, SUV420H1, PTPRO, SCAMP1, SP100, SOCS5, THAP10, YTHDC2, FLRT3, RAB2A, C1orf103, HOMER1, MYBL1, BAALC*.
CG-signature: *ACRBP, ACTL8, ADAM2, ADAM29, AKAP3, AKAP4, ANKRD45, ARMC3, ARX, ATAD2, BAGE, BAGE2, BAGE4, BAGE5, BRDT, CABYR, CAGE1, CALR3, CASC5, CCDC110, CCDC33, CCDC36, CCDC62, CCDC83, CEP290, CEP55, COX6B2, CPXCR1, CRISP2, CSAG1, CT45A1, CT45A3, CT45A4, CT45A5, CT45A6, CT47A1, CT47A10, CT47A11, CT47A2, CT47A3, CT47A4, CT47A6, CT47A7, CT47A8, CT47B1, CTAG1A, CTAG1B, CTAG2, CTAGE1, CTAGE5, CTCFL, CTNNA2, DCAF12, DDX43, DDX53, DKKL1, DMRT1, DNAJB8, DPPA2, DSCR8, ELOVL4, FAM133A, FAM46D, FATE1, FBXO39, FMR1NB, FTHL17, GAGE1, GAGE12B, GAGE12C, GAGE12D, GAGE12E, GAGE12F, GAGE12H, GAGE12I, GAGE12J, GAGE13, GAGE2A, GAGE4, GAGE5, GAGE6, GAGE7, GAGE8, GPAT2, GPATCH2, HORMAD1, HORMAD2, HSPB9, IGSF11, IL13RA2, KIAA0100, LDHC, LEMD1, LIPI, LUZP4, LY6K, MAEL, MAGEA1, MAGEA10, MAGEA11, MAGEA12, MAGEA2, MAGEA2B, MAGEA3, MAGEA4, MAGEA5, MAGEA6, MAGEA8, MAGEA9, MAGEB1, MAGEB2, MAGEB3, MAGEB4, MAGEB5, MAGEB6, MAGEC1, MAGEC2, MAGEC3, MORC1, NLRP4, NOL4, NR6A1, NXF2, NXF2B, ODF1, ODF2, ODF3, ODF4, OIP5, OTOA, PAGE1, PAGE2, PAGE2B, PAGE3, PAGE4, PAGE5, PASD1, PBK, PIWIL2, PLAC1, POTEA, POTEB, POTEC, POTED, POTEE, POTEG, POTEH, PRAME, PRM1, PRM2, PRSS54, PRSS55, RBM46, RGS22, ROPN1, RQCD1, SAGE1, SEMG1, SLCO6A1, SPA17, SPACA3, SPAG1, SPAG17, SPAG4, SPAG6, SPAG8, SPAG9, SPANXA1, SPANXA2, SPANXB1, SPANXC, SPANXD, SPANXN1, SPANXN2, SPANXN3, SPANXN4, SPANXN5, SPATA19, SPEF2, SPO11, SSX1, SSX2, SSX2b, SSX3, SSX4, SSX4B, SSX5, SSX6, SSX7, SSX9, SYCE1, SYCP1, TAF7L, TDRD1, TDRD6, TEKT5, TEX101, TEX14, TEX15, TFDP3, THEG, TMEFF1, TMEFF2, TMEM108, TMPRSS12, TPPP2, TPTE, TSGA10, TSPY2, TSPY3, TSSK6, TTK, TULP2, VENTXP1, XAGE1B, XAGE1E, XAGE2, XAGE3, XAGE5, ZNF165, ZNF645*.

Correlation Analysis between gene expression signatures

To evaluate the degree of correlation between two signature scores in the tumor samples, we used the R "cor" function to retrieve the Pearson correlation coefficient. To estimate the significance of the correlation value, we used R "lm" function to fit a linear model and retrieve the corresponding FDR-adjusted P values.

Survival analysis

Kaplan–Meier survival analysis was used to assess the relationship of the signature scores with overall survival. We applied the Cox proportional hazard model for multivariable analysis, to determine the associations between predictor variables and to obtain adjusted hazard-ratios. These analyses were performed with the R package "survival."

DATA AND SOFTWARE AVAILABILITY

The R implementation of the RESET algorithm, the processed TCGA data and the results of the RESET analysis are publicly available either at http://ciriellolab.org/reset or at the Mendeley Database: https://doi.org/10.17632/xgfm8y6mmf.1.