The NCI-60 Methylome and its Integration into CellMiner

William C. Reinhold[1*], Sudhir Varma[1,2,3], Margot Sunshine[1,2], Vinodh Rajapakse[1], Augustin Luna[1,4], Kurt W. Kohn[1], Holly Stevenson[5], Yonghong Wang[5], Holger Heyn[6], Vanesa Nogales[6], Sebastian Moran[6], David J. Goldstein[7] James H. Doroshow[1,8,] Paul S. Meltzer[5], Manel Esteller[6,9,10], and Yves Pommier[1*]

[1]Developmental Therapeutics Branch, CCR, NCI, NIH, Bethesda, MD 20892; [2]Systems Research and Applications Corp., Fairfax, VA 22033, [3]HiThru Analytics LLC, Laurel, MD 20707, [4]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, and Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA, [5]Genetics Branch, Developmental Therapeutic Program, CCR, NCI, NIH, Bethesda, MD 20892; [6]Cancer Epigenetics and Biology Program, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Catalonia, Spain; [7]Office of the Director, CCR, NCI, NIH, Bethesda, MD 20892; [8]Genetics Branch, CCR, NCI, NIH, Bethesda, MD 20892; [8]Divison of Cancer Treatment and Diagnosis, CCR, NCI, NIH, Bethesda, MD 20892; [9]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain. [10]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

*Corresponding authors: William C. Reinhold, E-mail: wcr@mail.nih.gov, Yves Pommier, E-mail pommier@nih.gov

Short title: NCI-60 methylome

Keywords: mutations, CpGs, epigenetics, transcriptional regulation, precision medicine, pharmacogenomics

There are no conflicts of interest to report.

Abstract

A unique resource for systems pharmacology and genomic studies is the NCI-60 cancer cell line panel, which provides data for the largest publicly available library of compounds with cytotoxic activity (~21,000 compounds), including 108 FDA-approved and 70 clinical trial drugs as well as genomic data, including whole-exome sequencing, gene and microRNA transcripts, DNA copy number, and protein levels. Here we provide the first readily usable genome-wide DNA methylation database for the NCI-60, including 485,577 probes from the Infinium HumanMethylation450k BeadChip array, which yielded DNA methylation signatures for 17,559 genes integrated into our open access CellMiner version 2.0 (https://discover.nci.nih.gov/cellminer). Among new insights, transcript versus DNA methylation correlations revealed the epithelial/mesenchymal gene functional category as being influenced most heavily by methylation. DNA methylation and copy number integration with transcript levels yielded an assessment of their relative influence for 15,798 genes, including tumor suppressor, mitochondrial, and mismatch repair genes. Four forms of molecular data were combined, providing rationale for microsatellite instability for 8 out of the 9 cell lines in which it occurred. Individual cell line analyses showed global methylome patterns with overall methylation levels ranging from 17 to 84%. A six-gene model including PARP1, EP300, KDM5C, SMARCB1 and UHRF1 matched this pattern. Additionally, promoter methylation of two translationally relevant genes Schlafen 11 (SLFN11) and methylguanine methyltransferase (MGMT) served as indicators of therapeutic resistance or susceptibility, respectively. Overall, our database provides a resource of pharmacological data that can reinforce known therapeutic strategies and identify novel drugs and drug targets across multiple cancer types.

Introduction

DNA methylation is an heritable epigenetic event occurring at cytosines 5’ of guanosines (CpG’s) and catalyzed by DNA methyltransferases (DNMT), which transfer a methyl group from S-adenosyl methionine to the 5 position of the

cytosine ring (1). DNA methylation is involved in multiple epigenetic processes ranging from transcriptional down-regulation, X chromosome inactivation, embryonic development, and genomic imprinting (2,3). “CpG islands” in the 5’ regulatory regions of many genes (~56%), are involved in transcription down-regulation and histone deacetylase recruitment (4,5). The pattern of methylation has both allele-specific and evolutionary conservation, as well as tissue-specific and cell-specific variations (6-8).

DNA methylation defects have been associated with multiple diseases including i) global methylation defects for diabetes, obesity, fetal alcohol syndrome and aging, ii) imprinting disorders for Angelman syndrome, Prader-Wili syndrome, Beckwith-Wiedemann syndrome and autism, iii) genetically-driven methylation defects for Fragile X syndrome, dyslexia, Rett syndrome, centromere instability, Sotos, Weaver, and Kleefstra syndomes, and iv) candidate gene methylation defects for obesity and Type 1 and 2 diabetes (9,10).

DNA methylation defects are also linked to cancers (9). Global loss of DNA methylation has been associated with genomic instability, loss of imprinting, and reactivation of transposable elements (11). Loss of genomic imprinting of IGF2 has been associated with increased risk of liver, lung, intestinal and colon cancers (3). Concurrent with global hypomethylation, specific hypermethylation of tumor suppressor genes, including CDKN2A, MLH1, VHL, and CDH1 occurs leading to their transcriptional repression (3,9).

The NCI-60 was the first cancer cell line database established and it remains the largest drug and most complete source of molecular data (12,13). In the current study we provide whole genome methylation levels from 485,577 probes across the NCI-60 cancer cell line panel. We detail probes that are associated with genes, and provide easy access to them using our CellMiner \ Cell line signature web-based application (14). This allows direct comparison and integration with other molecular and activity data, examples of which are included. We also provide a novel form of visualization of the level of influence on gene transcript levels of DNA methylation and copy number, and examples of the relevance of DNA methylation for predicting the activity of DNA-targeted agents.

Materials and Methods

Cell lines, growth, and DNA purification

We combined data for the NCI-60 cell lines from both the National Cancer Institute (NCI) and the Cancer Epigenetics and Biology Program (CEBP) (15). Both the NCI and CEPB obtained the cell lines from the Developmental Therapeutics Program (DTP) (12,16,17). For the NCI dataset, cells were grown and DNA isolated as described previously (18,19). For the CEPB dataset, cells were grown and DNA isolated as described previously (15).

DNA methylation, comparison of CEBP and DTB datasets, and probe beta values

NCI bisulfite conversion and DNA sample handling was done as described previously (20). The Infinium HumanMethylation450k BeadChip kit (Illumina, Inc, San Diego, CA, 92122, catalog #WG-314-1001) was used with standard protocol for both the CEBP and NCI studies (20,21). CEPB bisulfite conversion and DNA sample handling was done as described (15).

NCI samples used p-values of detection as filters for each probe. Probes with values of $p>0.01$ were treated as missing. For dataset comparison, we did average linkage hierarchical clustering of all cell line probe intensity profiles using 1- Pearson correlation distance. This and all subsequent statistical analysis was done in the R statistical environment (22). We used the probe-wise average of the cell line replicates from the two datasets for all subsequent analysis.

Probe beta values by cell line are calculated as:
(Intensity of the methylated probe) / [(Intensity of the methylated probe) + (Intensity of the unmethylated probe)]

Selection of probes associated to genes for comparison to gene expression

Probes were assessed for location with respect to genes and proximal CpG islands defined by Illumina. Probes were designated as category-1 or 2 with category-1 considered to be most informative. For subsequent comparisons, category-1 probes were used if available, and if not then category-2. In either case, gene methylation values resulted from averages of those probes. For genes with multiple transcriptional start sites, the transcript with the most negative correlation to the methylation probes was identified and used.

DNA methylation versus transcript expression, and gene group definition

For comparisons between methylation levels and transcript expression z scores, we used the methylation and transcript "Cell line signatures" (16). Expression versus methylation correlations, heat-map and histogram were generated using The R Project for Statistical Computing (22). For the transcript versus methylation analyses by gene category, genes were divided based on correlation value of $r \leq -0.5$. Enrichment of Gene Ontology Consortium classifications were accessed using GoMiner with a false discovery rate cut-off off < 0.05 and a minimum ten identified genes per category (23,24).

Gene DNA copy number determinations

DNA copy number patterns were determined as described previously (25). Their gene "Cell line signatures" can be accessed at CellMiner by gene using "CellMiner \ NCI-60 Analysis Tools \ Cell line signature" (16).

Linear regressions for testing the predictive power of DNA copy number and methylation on transcript expression

For each of the 15,798 genes with all three forms of data available (transcript, methylation and copy number levels) a linear regression model was fit, with both copy number and methylation as independent variables, and transcript expression as the dependent variables. The model provided coefficients for the copy number and methylation that gave the lowest squared error between fitted

values and true expression. We separated individual contributions of these two factors for gene expression prediction using the method of relative importance (26), using the *lmg* method (27) from the R package *relaimpo* to compute individual R^2 values. Total (or combined) R^2 is the summation of these two. Square roots of the R^2 values were multiplied by the sign of the coefficients of the factors in the combined model to get the value of R.

Mutational status of genes

Genetic variants were assessed using the "CellMiner\NCI-60 Analysis Tools \ Cell line signature \ Genetic variant summation" tool (12,13,16). The form of the data used for SMARCB1 in the linear regression analysis is the "Amino acid changing". The form of the data used for MLH1, MSH2, and MSH6 is the "Protein function affecting".

Global methylation and PARP1 and EP300 protein expression levels

The box and whisker plots and linear regression analysis were done using The R Project for Statistical Computing (22). For the linear regression analysis, the form of KDM2B, CREBBP, and SMARCB1 "Genetic variant summation" data used is the "Amino acid changing". Protein expression for PARP1 and EP300 were accessed using "CellMiner \ NCI-60 Analysis Tools \ Cell line signature \ Protein mean values" (14,16).

Pattern comparisons and drug activity determinations

Correlations between median methylation values and cytogenetic measurements of instability were determined using "Pattern comparison", accessed at "CellMiner \ NCI-60 Analysis Tools \ Pattern comparison", within "Miscellaneous phenotypic parameters". Drug "Cell line signatures" can be accessed at CellMiner by National Service Center (NSC) number using "CellMiner \ NCI-60 Analysis Tools \ Cell line signature \ Drug activity z scores" (14,28).

Results

Concordance and merging of two datasets for establishing the NCI-60 methylome database

A comparison of the two NCI-60 whole genome DNA methylation datasets independently prepared at both the NCI and the Cancer Epigenetics and Biology Program (CEBP, Barcelona, Spain) was done (15). Cross correlations of probes-specific methylation levels for matched cell lines in the two datasets yielded very high correlations for all cell lines, with a range between 0.919 and 0.995, and an average of 0.978. Three representative cell lines probe set comparisons are shown in Figure 1A. Also, clustering analyses of all cell lines across the two datasets (NCI vs. CEBP) showed that in all cases, the nearest neighbor cell line was the duplicate cell line from the independent dataset (Figure 1B). Only two of the cell lines ME:MDA-N and RE:CAKI-1 did not have duplicates in the CEBP analysis. Together, these analyses demonstrated the consistency of the NCI-60 methylome data. Because of their high concordance, the NCI and CEBP NCI-60 whole genome DNA methylation datasets were merged by taking the matched cell line average methylation. This form of the data was used for the remaining analyses in this manuscript, as well as for uploading the data to CellMiner.

Identification of methylation probes for individual genes

In the NCI dataset analysis, independent from the CEBP team (15), methylation probes that were gene-specific were selected. To this end, the 485,577 methylation probes on the Infinium HumanMethylation450k BeadChip array were assessed based on location with respect to CpG islands and genes (Supplemental Figure 1A, B, and C) and their relationship with gene expression (Supplemental Figure 1D). This resulted in a 2-tier probe categorization (Supplemental Figure 1C) yielding DNA methylation signatures for 17,552 genes. The complete gene-probe information is included in Supplemental Table 1. For the purpose of comparing gene transcript levels with methylation, category-1

probes were the most informative. Category-2 probes were found to be less informative (less well correlated), but still relevant for some genes. As a result, the NCI-60 methylation data for 17,552 genes may be directly retrieved from the CellMiner website (updated to version 2.0) in several forms as described previously (14,16,28). A link providing online instructions for generating DNA methylation Cell line signatures is available (29). The graphical output “Cell line signature’s” for 3 examples, VHL, SLFN11, and IRF6, are presented in Figure 2.

Global methylation levels for the NCI-60

The distribution of global methylation levels for 485,577 probes showed marked differences among individual cell lines (Figure 3A). Median values ranged from 17% for melanoma MALME-3M to 84% for colon carcinoma HCT-116. To determine genes that are major contributors in such global differences, the global methylation pattern was compared to epigenetic, chromatin, and histone functional group’s (defined in Supplemental Table 2) for: i) amino acid changing genetic variants, ii) protein function affecting genetic variants, iii) gene transcript levels, and iv) protein levels, using the “Pattern comparison” web application (14,28).

Significant correlations ($p<0.01$) were found for 24 genes, including five with literature connection to DNA methylation: UHRF1, MTA1, HIST1H1A, PARP1, and EP300. However, none of the 24 significant correlations found could individually predict more than 21% of the median methylation pattern, implying multivariate causation. Using linear regression, it was found that all of these except HIST1H1A made significant contribution to modeling a relationship to the median methylation pattern. After removing HIST1H1A, four other genes (KDM5C, KDM2B, CREBBP, and SMARCB1) were added back, one at a time, based on their appearing in both the epigenetic and chromatin functional categories (Supplemental Table 2). Of these, KDM5C and SMARCB1 were found to contribute to the model significantly ($p=0.00046$) by comparison to the four-gene model. The resulting scatter plot of true versus fitted values in Figure 3B showed an r^2 of 0.62. The six-gene model for prediction of global methylation

status included PARP1 (p value= 3.7 x 10^{-6}, r= -0.062), MTA1 (p value= 5.9 x 10^{-6}, r= 0.066), EP300 (p value= 1.1 x 10^{-3}, r= -0.030), KDM5C (p value= 1.2 x 10^{-3}, r= -0.051), SMARCB1 (p value= 0.017, r= -0.0049), and UHRF1 (p value= 0.041, r= 0.021).

Pattern comparison using the median probe methylation values (Figure 3A) identified significant negative correlation to six parameters of genomic instability from cytogenetics (30). These were modal chromosomal number, numerical complexity, structural heterogeneity, fraction of abnormal chromosomes that experience numerical heterogeneity, fraction of normal chromosomes that experience numerical heterogeneity, and numerical heterogeneity, with correlations of -0.341, -0.362, -0.362, -0.377, -0.389, and -0.425 respectively. These correlations indicate that as the levels of DNA methylation decrease, the levels of genomic instability increase.

Comparison of methylation levels to transcript expression for functional gene groups

Comparison of gene methylation to transcript expression levels across the NCI-60 identified 44 GO categories enriched for genes with significant methylation versus expression correlations (see rows in Figure 4A) (23,24). As these 44 categories had overlapping genes, they were organized into the seven groups shown on the right of Figure 4A, including: metabolic processes, blood coagulation, cell migration and mobility, cell adhesion and assembly, white blood cell proliferation, activation of immune response, cell death and signaling.

Curated gene lists generated from literature for different pathways related to cancer (including epithelial mesenchymal transition, tumor suppressors, oncogenes, apoptosis, DNA repair chromatin and mitochondria; see Supplemental Table 2) were also tested. For all the genes in each gene category, the correlation between transcript and methylation levels was computed. The data presented in Figure 4B show the median Pierson correlation for each curated gene category (vertical lines), as well as for the GO categories shown in panel A (red fonts). The histogram of correlation distribution for all the 17,144

transcripts with expression and DNA methylation data is shown as the shaded area (see Supplemental Table 2 for individual values).

The epithelial and mesenchymal gene categories, consisting of 25 and 27 genes respectively, were assembled previously (31). These categories showed the most significant correlation between DNA methylation and transcript levels, with medians of -0.639 and -0.525 for the epithelial and mesenchymal genes, respectively. A wide gap was found prior to the next groups of genes with significant correlation. Those included the genes found by GO analysis (see Figure 4A), and additional categories including tumor suppressor genes, which were tightly grouped as a cluster of 23 functional groups, within a range of -0.259 to -0.148. By contrast, the microRNAs median value showed a lack of significant correlation to expression (r=0.010). Together, these analyses demonstrate that DNA methylation drives gene expression for selected pathways, such as the epithelial mesenchymal transition (EMT), and to a lesser extend for tumor suppressors in the NCI-60.

Integration of the NCI-60 methylome with transcriptome profiles

Supplemental Table 2 contains the Pearson's correlations between DNA methylation and transcript expression for 17,144 transcripts, including 16,155 genes, 494 open reading frames, 167 loci, 256 microRNAs, 67 long intergenic non-protein coding RNA, and five long non-coding RNA's that contain microRNAs in their introns. The ability to retrieve the methylation and transcript expression patterns for individual genes (through CellMiner) enables their comparison. Figure 5 shows representative transcript versus methylation scatter plots constructed from the CellMiner data, including their correlations, for genes from the functional categories listed in Figure 4B.

The expression of the mesenchymal gene Vimentin *(VIM*) was significantly driven by promoter methylation, as were 4 epithelial genes (Claudins 7 and 4, the Epithelial Splicing Regulator Protein 2 gene *ESRP2*, and *RAB25*, a member of the RAS oncogene family). Among tumor suppressors, examples of highly significant correlations include *APC* (the Adenomatous Polyposis Coli gene),

VHL (the Von Hippel-Lindau gene, inactivated in RE:RXF-393), *BRCA1* (inactivated in 3 of the ovarian cell lines: OVCAR-4, OVCAR-8 and NCI/ADR-RES), *IRF6* (the Interferon Regulatory Factor 6) and *CDKN2A* (Cyclin-Dependent Kinase Inhibitor 2A). Both *IRF6* and *CDKN2A* are inactivated by promoter methylation in multiple NCI-60 cell lines. For the apoptotic pathway, promoter methylation was found significant for some key genes such as BIM (*BCL2L11*), *BID*, the cell surface death receptor (*FAS*), Caspase 8 and Heat Shock 70kDa Protein 1B (*HSPA1B*).

Notably, for the genes encoding epigenetic factors, only the expression of TET1, the gene encoding Ten-Eleven Translocation Methylcytosine Dioxygenase 1, a key demethylating enzyme, was found driven by promoter methylation (Figure 5, 4th row, on left). Examples of DNA damage response genes driven by promoter methylation include *MLH3* (MutL Homolog 3), *HLTF* (Helicase-Like Transcription Factor/SMARCA3), *TDP1* (Tyrosyl-DNA phosphodiesterase 1), *MGMT* (O-6-Methylguanine-DNA Methyltransferase) and *SLFN11* (Schlafen 11). *MGMT* and *SLFN11* (Figure 5, 5th row, on right) will be discussed below in the context of their pharmacological relevance (32-34).

Integration of DNA methylation and gene expression with gene copy number and mutations across the NCI-60

The majority of the genes (15,798 genes) can be queried for concurrent analyses of DNA methylation, transcript expression, and DNA copy number in the NCI-60 (14) (Supplemental Table 3). In Figure 6A, “Cell line signatures” obtained directly from CellMiner for DNA methylation, copy number (14,25), and transcript expression (28), are presented for three exemplary genes for which transcript levels are driven by DNA methylation, copy number of both.

Expression of RAB25, a member of the RAS family involved in membrane trafficking and cell polarity and epithelial phenotype, showed high correlation with methylation (as do other EMT genes see Figure 4B and Figure 5, top row), but not DNA copy number (r = -0.978 and 0.069, respectively). By contrast, expression of *POLG*, a housekeeping gene encoding the mitochondrial

replicative DNA polymerase, showed no correlation with methylation but instead a high correlation with copy number (r = 0.042 and 0.758, respectively). The high impact of copy is probably related to the frequent gain or loss of the locus 15q25/26, which also encodes important repair genes including *FANCI* and *BLM* as well as *IDH2* and *CHD2* (readily detected by the "Pattern comparison" tool of CellMiner). Expression of the tumor suppressor gene *CDKN2A* encoding $p16^{INK4A}$ and $p14^{ARF}$ showed high correlation to both DNA copy number and methylation (r = 0.622 and -0.558, respectively). Moreover, the methylation and copy number profiles across the NCI-60 showed remarkable mirror images (Figure 6A, compare 2nd and 3rd bar graphs from the right), with the same cell lines lacking a gene copy also showing hypermethylation on the other allele and r=0.849.

Linear regression assessment of influence of both DNA methylation and copy number on transcript expression at the whole genome scale is shown in Figure 6B. In the scatter plots of R (the coefficients of determination times the sign of the coefficient) for both DNA copy number (the x-axis), and methylation (the y-axis), values of 1 or -1 indicate perfect prediction in the positive or negative sense, respectively, and 0 no predictive value. The "All genes" plot (Figure 6B, left) illustrates the cumulative importance of both DNA methylation and copy number, as indicated by the presence of 59% of the points in the bottom right quadrant. These points are both negative for DNA methylation (indicating negative predictive power for transcript expression), and positive for DNA copy number (indicating positive predictive power for transcript expression).

Figure 6B (2nd panel from left) shows that restricting the plot to the epithelial genes (including *RAB25, CLDN4, ESRP2, CLDN7*; see Figures 4B and 5) demonstrates that methylation is their predominant regulator. The mitochondrial and tumor suppressors plots indicate more balanced influences from both DNA methylation and copy number for these categories. The R values are pre-calculated for any gene of interest (with data) in Supplemental Table 3. These analyses demonstrate which genes transcript levels in what cell lines are driven by copy number and/or methylation levels in particular cell lines.

Figure 6C extends the integrative approach combining DNA copy number, methylation, transcript expression, and mutational data obtained from whole sequencing of the NCI-60 (12,14) for MLH1, MSH2, and MSH6, to provide the genomic underpinnings for the reported presence of microsatellite instability in 8/9 cell lines (12,35). For MLH1, there is DNA copy number loss in SKOV3, DNA methylation and reduced transcript expression in KM12, reduced transcript expression in IGROV1 and SKOV3, and predicted function affecting amino acid changes in HCT116, HCT15, CCRF-CEM, IGROV1, and DU145. For MSH2, there are predicted function affecting amino acid changes in SK-MEL2, NCI-H522, and DU145. For MSH6, there is reduced transcript expression in IGROV1, and predicted function affecting amino acid changes in IGROV1 and DU145. Only the instability of LE:MOLT-4 remains unexplained. The example of the mismatch repair pathway demonstrates the converging contribution of the four genomic parameters now available in the NCI-60 (methylation, copy number, expression and deleterious mutations), and the importance of data integration.

NCI-60 methylome and anticancer pharmacology

MGMT encodes methylguanine methyltransferase, an enzyme that removes O6-methylguanine, the most cytotoxic DNA methylation adduct produced by temozolomide, a commonly used oral drug in glioblastomas (36). Cancer cells deficient for MGMT are exquisitely sensitive to temozolomide (34) and *MGMT* promoter methylation is a positive prognostic indicator for temozolomide treatment in glioblastoma (36). The scatter plot of *MGMT* methylation versus expression levels (see Figure 5 bottom right) showed significant correlation (r=-0.48). DNA promoter methylation levels above 40% was found associated with *MGMT* expression levels at background levels (<-0.6) for 81% of those cell lines (Figure 5 and Supplemental Figure 2B). DNA promoter methylation levels less than 40% are associated with (76%) of expressed cell lines (>-0.6). Promoter hypermethylation leading to transcriptional inactivation extended beyond the 6 glioma (CNS) cell lines, with colon (KM12) and the leukemia (SR) both having high methylation and background expression. However, only 5 of the 21 NCI-60

cell lines with low *MGMT* expression (<-.06) have significant methylation (above 50%). Thus, the incomplete association between *MGMT* expression and methylation parameters imply the use of epigenetic silencing by DNA methylation level as a prognostic indicator for temozolomide treatment is useful, but incomplete.

A second gene, which has recently been causally linked with response to a broad spectrum of DNA damaging drugs (including topoisomerase inhibitors, PARP inhibitors as single agentes or in combination with temozolomide, cisplatin, alkylating agents and DNA synthesis inhibitors) is *SLFN11* (Schlafen 11) (15,32,33,37,38). SLFN11 encodes a nuclear protein with putative helicase activity that blocks cell cycle progression and dampens DNA repair (32,39). *SLFN11* was among the genes with the highest correlation (at the top 94th percentile) between methylation and expression (see Figure 5 bottom right). Notably, approximately 38% of the NCI-60 cell lines do not express SLFN11 above background level (32). Of these, lack of SLFN11 expression is linked to methylation in approximately half of the cell lines (Figure 5). Neither *SLFN11* or *MGMT* expression had significant association to DNA copy number ($p<0.01$, Supplemental Table 3).

To test whether *SLFN11* promoter methylation was linked with resistance to DNA damaging agents, the whole NCI-60 drug database (≈ 21,000 compounds including 108 FDA-approved and 70 clinical trial drugs) was tested. Drug activity correlations revealed that *SLFN11* methylation was significantly correlated with resistance to multiple clinically relevant drugs that cause DNA damage, including alkylating agents (cisplatin, carboplatin, melphalan), topoisomerase I (topotecan, LMP400) and II (etoposide) inhibitors, DNA synthesis inhibitors (gemcitabine, fludarabine, cytarabine, hydroxyurea), PARP inhibitors (talazoparib, olaparib) and bleomycin (Table 1). As expected (32,37), no correlation was observed for tubulin inhibitors (paclitaxel, docetaxel) and protein kinase inhibitors (erlotinib, crizotinib, vemurafenib), consistent with the selective implication of SLFN11 for cytotoxic response to DNA damaging drugs.

Discussion

Here we provide readily usable and accessible genome-wide data for the NCI-60 cancer cell line methylome based on two independent determinations (Figure 1). Assignment of salient CpG sites for each gene (Supplemental Table 1) enables the extraction of those data using our CellMiner tools (14,29). We provide examples of how the DNA methylation data may be integrated with the extensive genomic and pharmacological databases for the NCI-60 (Figures 4, 5 and 6). We anticipate that these data and tools will enable exploration of i) the relevance of DNA methylation as a key regulator of gene expression, ii) its relationships with other forms of molecular data, and iii) genomic biomarkers for precision therapeutics. Determination of methylation status is a preferred approach for clinical samples because it provides robust information with which to evaluate cancer genomes (as it uses DNA rather than RNA, which tends to be unstable).

The examples we provide demonstrate that genome-wide access to DNA methylation gives new insights in tumor biology and regulatory mechanisms for gene expression. Figures 4, 5 and 6 demonstrate that promoter methylation is a prominent regulator for EMT (epithelial mesenchymal transition), a major determinant for tumor invasion and resistance to therapy, which is being targeted by DNA methyl transferase inhibitors (40). On the other hand, for tumor suppressors, both methylation and gene deletion drive gene expression. A salient example is the cyclin-dependent kinase inhibitor 2A gene *CDKN2A*, which encodes the two major tumor suppressors, $p16^{INK4A}$ and $p14^{ARF}$. As shown in Figures 5 and 6, approximately 40% of the NCI-60 fail to express *CDKN2A*. That almost half of cancer cells suppress *CDKN2A* expression is consistent with the larger MIT-Broad-CCLE dataset where ~40% of the 1000 cell lines only express background (no) *CDKN2A* (33). In addition, examination of *CDKN2A* gene copy number shows that 20 of the NCI-60 cell lines have 9p21 deletions (Figure 6A, right). Furthermore, integrating the NCI-60 data for *CDKN2A* methylation and gene copy number shows a high correlation between the two genomic parameters ($r=-0.85$; $p=9.7 \times 10^{-18}$), This demonstrates that, in the NCI-60,

cancer cells commonly inactivate *CDKN2A* by biallelic loss of *CDKN2A* (one allele by promoter methylation and the other by 9p21 chromosome deletion).

DNA methylation is also relevant to precision therapeutics, and, in this study, examples of two genes are presented: *SLFN11* and *MGMT*. Both genes encode key factors that determine response to widely used DNA damaging agents, which remain a major component of the cancer armamentarium but lag behind protein kinase inhibitors in terms of predictive biomarkers. A relationship between *SLFN11* transcript levels and pharmacological response has been demonstrated in both the NCI-60 and CCLE cancer cell lines (15,32,33,41,42). This relationship is both causal and broad in scope, affecting topoisomerase I inhibitors, topoisomerase II inhibitors, alkylating agents and DNA synthesis inhibitors (32,37). *SLFN11* expression has recently been shown to be driven by ETS transcription factors (43), which explain high SLFN11 expression in Ewing's sarcoma (33,43). Epigenetic inactivation of *SLFN11*, which accounts at least in part for frequent lack of expression of *SLFN11* in many cancer cell lines including commonly used ones such as HeLa, U2OS and HCT116 (38), and has recently been shown to have a robust and causal influence on resistance to platinum-derived drugs such as cisplatin and carboplatin (15). The present study expands this finding to eleven clinically relevant drugs (Table 1). Correlations between the SLFN11 methylation levels and DNA damaging drug activities demonstrate the potential for using DNA methylation of SLFN11 in addition or in place of RNA- and immunofluorescence-based assays for measuring SLFN11 expression, and testing the usefulness of SLFN11 as a novel predictive biomarker for drug activity in the clinical setting.

Temozolomide is approved for the treatment of glioblastomas because of the frequent inactivation of *MGMT* by promoter methylation in those tumors (34,36,44). Methylome and gene expression analyses of the NCI-60 reveals that *MGMT* expression is frequently suppressed beyond CNS cancer cell lines (2 out of the 6 breast cancer cell lines, 2 out of the 7 colon, 2 or the 6 leukemia, 4 of the 10 melanomas, three of the 9 lung, and one of the 8 kidney cancer cell lines) (Figure 5 and Supplemental Figure 2B). Yet, promoter methylation explained only

5 of these 20 cell lines with low or no *MGMT* expression. These observations suggest that MGMT promoter methylation (36,44) is actually insufficient to predict temozolomide activity and that other assays (transcripts or protein measurements) should be used to monitor patient candidates for temozolomide not only for glioblastomas but also outside of brain tumors.

In summary, the NCI-60 methylome adds to the preexisting molecular and pharmacological databases, which are publicly available and usable by non-informaticists at the Cellminer website (14). They provide an additional translationally relevant piece in the molecular puzzle of understanding and predicting transcriptional regulation, and for the broader interplay among cancer-associated molecular and pharmacological parameters.

References

1. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 1999;99(3):247-57.
2. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nature methods 2016.
3. Hirst M, Marra MA. Epigenetics and human disease. The international journal of biochemistry & cell biology 2009;41(1):136-46.
4. Antequera F, Bird A. Number of CpG islands and genes in human and mouse. Proceedings of the National Academy of Sciences of the United States of America 1993;90(24):11995-9.
5. Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature genetics 1998;19(2):187-91.
6. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, et al. Amount and distribution of 5-methylcytosine in human DNA from

different types of tissues of cells. Nucleic acids research 1982;10(8):2709-21.

7. Tang A, Huang Y, Li Z, Wan S, Mou L, Yin G, et al. Analysis of a four generation family reveals the widespread sequence-dependent maintenance of allelic DNA methylation in somatic and germ cells. Scientific reports 2016;6:19260.
8. Zhang M, Wang CC, Yang C, Meng H, Agbagwa IO, Wang LX, et al. Epigenetic Pattern on the Human Y Chromosome Is Evolutionarily Conserved. PloS one 2016;11(1):e0146402.
9. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. Nature reviews Genetics 2012;13(10):679-92.
10. Schenkel LC, Rodenhiser DI, Ainsworth PJ, Pare G, Sadikovic B. DNA methylation analysis in constitutional disorders: Clinical implications of the epigenome. Critical reviews in clinical laboratory sciences 2016:1-19.
11. Esteller M. Epigenetics in cancer. The New England journal of medicine 2008;358(11):1148-59.
12. Abaan OD, Polley EC, Davis SR, Zhu YJ, Bilke S, Walker RL, et al. The Exomes of the NCI-60 Panel: A Genomic Resource for Cancer Biology and Systems Pharmacology. Cancer Res 2013;73(14):4372-82.
13. Reinhold WC, Varma S, Sousa F, Sunshine M, Abaan OD, Davis SR, et al. NCI-60 whole exome sequencing and pharmacological CellMiner analyses. PloS one 2014;9(7):e101670.
14. CellMiner. https://discover.nci.nih.gov/cellminer.
15. Nogales V, Reinhold WC, Varma S, Anna Martinez-Cardus A, Moutinho C, Moran S, et al. Epigenetic Inactivation of the Putative DNA/RNA Helicase SLFN11 in Human Cancer Confers Resistance to Platinum Drugs Oncotarget 2015;Accepted Online.
16. Reinhold WC, Sunshine M, Varma S, Doroshow JH, Pommier Y. Using CellMiner 1.6 for Systems Pharmacology and Genomic Analysis of the NCI-60. Clinical cancer research : an official journal of the American Association for Cancer Research 2015;21(17):3841-52.

17. Developmental Therapeutics Program (DTP). http://dtp.nci.nih.gov/.
18. Reinhold W, Reimers M, Maunakea A, Kim S, Lababidi S, Scherf U, et al. Detailed DNA methylation profiles of the E-cadherin promoter in the NCI-60 cancer cells. Mol Cancer Ther 2007;6:391-403.
19. Liu H, Petula D'Andrade, Stephanie Fulmer-Smentek, Philip Lorenzi, Kurt W. Kohn, John N. Weinstein, et al. mRNA and microRNA expression profiles of the NCI-60 integrated with drug activities. MCT 2010;9(5):1080-1091.
20. Killian JK, Kim SY, Miettinen M, Smith C, Merino M, Tsokos M, et al. Succinate dehydrogenase mutation underlies global epigenomic divergence in gastrointestinal stromal tumor. Cancer discovery 2013;3(6):648-57.
21. Illumina HumanMethylation450K Documentation and Literature. http://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit/documentation.html.
22. The R Project for Statistical Computing. http://www.r-project.org/.
23. Gene Ontology Consortium (GO). http://geneontology.org/.
24. GoMiner. http://discover.nci.nih.gov/gominer/index.jsp.
25. Varma S, Pommier Y, Sunshine M, Weinstein JN, Reinhold WC. High resolution copy number variation data in the NCI-60 cancer cell lines from whole genome microarrays accessible through CellMiner. PloS one 2014;9(3):e92047.
26. Gromping U. Relative Importance for Linear Regression in R: The Package relaimpo. Journal of Statistical Software 2006;17.
27. Lindeman RH, Merenda PF, Gold RZ. Introduction to Bivariate and Multivariate Analysis 1980.
28. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, et al. CellMiner: A Web-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set. Cancer Res 2012(72):13.

29. DNA methylation "Cell line signitures" instructions. https://discover.nci.nih.gov/cellminer/methylation.html.
30. Roschke AV, Tonon G, Gehlhaus KS, McTyre N, Bussey KJ, Lababidi S, et al. Karyotypic complexity of the NCI-60 drug-screening panel. Cancer Res 2003;63(24):8634-47.
31. Kohn KW, Zeeberg BM, Reinhold WC, Pommier Y. Gene expression correlations in human cancer cell lines define molecular interaction networks for epithelial phenotype. PloS one 2014;9(6):e99269.
32. Zoppoli G, Regairaz M, Leo E, Reinhold WC, Varma S, Ballestrero A, et al. Putative DNA/RNA helicase Schlafen-11 (SLFN11) sensitizes cancer cells to DNA-damaging agents. Proceedings of the National Academy of Sciences of the United States of America 2012;109(37):15030-5.
33. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483(7391):603-307.
34. Zhang J, Stevens MF, Bradshaw TD. Temozolomide: mechanisms of action, repair and resistance. Curr Mol Pharmacol 2012;5(1):102-14.
35. Catalogue of somatic mutations in cancer (COSMIC). http://www.sanger.ac.uk/genetics/CGP/MSI/table1.shtml: Accessed 2012, March 26.
36. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. The New England journal of medicine 2005;352(10):997-1003.
37. Sousa FG, Matuo R, Tang SW, Rajapakse VN, Luna A, Sander C, et al. Alterations of DNA repair genes in the NCI-60 cell lines and their predictive value for anticancer drug activity. DNA repair 2015;28:107-15.
38. Murai J, Feng Y, Yu GK, Ru Y, Tang SW, Shen Y, et al. Resistance to PARP inhibitors by SLFN11 inactivation can be overcome by ATR inhibition. Oncotarget 2016.

39. Mu Y, Lou J, Srivastava M, Zhao B, Feng XH, Liu T, et al. SLFN11 inhibits checkpoint maintenance and homologous recombination repair. EMBO Rep 2016;17(1):94-109.
40. Zahnow CA, Topper M, Stone M, Murray-Stewart T, Li H, Baylin SB, et al. Inhibitors of DNA Methylation, Histone Deacetylation, and Histone Demethylation: A Perfect Combination for Cancer Therapy. Adv Cancer Res 2016;130:55-111.
41. Gmeiner WH, Reinhold WC, Pommier Y. Genome-wide mRNA and microRNA profiling of the NCI 60 cell-line screen and comparison of FdUMP[10] with fluorouracil, floxuridine, and topoisomerase 1 poisons. Mol Cancer Ther 2010;9(12):3105-14.
42. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat Chem Biol 2016;12(2):109-16.
43. Tang SW, Bilke S, Cao L, Murai J, Sousa FG, Yamade M, et al. SLFN11 Is a Transcriptional Target of EWS-FLI1 and a Determinant of Drug Response in Ewing Sarcoma. Clinical cancer research : an official journal of the American Association for Cancer Research 2015;21(18):4184-93.
44. Wick W, Weller M, van den Bent M, Sanson M, Weiler M, von Deimling A, et al. MGMT testing--the challenges for biomarker-based glioma treatment. Nature reviews Neurology 2014;10(7):372-85.

Figure Legends

Figure 1. Comparisons of the NCI and CEPB DNA methylation datasets for the NCI-60. **A**, Scatter plots of the NCI versus CEPB methylation data for three representative cell lines. Methylation levels for 30,000 randomly selected probes were used in each plot. The x and y-axes are the CEPB and NCI methylation levels, respectively. The values of 0 to 1 for both axes correspond to 0 to 100% methylation. “r” is the Pearson’s correlation between datasets. **B**, Hierarchical

clustering of the NCI and CEPB data for all the NCI-60 cell lines using 1-Pearson correlation distance, and average linkage. A randomly selected set of 10,000 probes was used. The x-axis is the 1- Pearson correlation distance. The y-axis contains the cell lines with colors corresponding to tissue of origin (as in the CellMiner tools) (14).

Figure 2. Methylation data for three representative genes across the NCI-60. The graphical output for the "Cell line signature" tool, for the average gene methylation values for each cell line from the CellMiner website (http://discovery.nci.gov). The x-axis is the average gene methylation level (percentage / 100). The y-axis lists the cell lines, color-coded by tissue of origin. Note the high DNA methylation of VHL in one and only one of the 60 cell lines.

Figure 3. Global methylation levels across the NCI-60. **A**, Box and whisker plots of average methylation for the total set of 485,577 probes. For each cell line, the data is broken into four quartiles of equal number. The first quartile is distributed from the top of the dotted line to the top of the colored bar, the second from the top of the colored bar to the black median line, the third from the median line to the bottom of the colored bar, and the forth from the bottom of the colored bar to the bottom of the dotted line. The x-axis lists the cell lines, color-coded by tissue of origin, and the y-axis is the methylation level (percentage / 100). **B**, Linear regression analysis using six genes to predict the median methylation pattern in Figure 6A. R is the correlation coefficient.

Figure 4. Functional categories with significant correlation between gene transcript expression and DNA methylation. **A**, Heat-map for GO categories, including gene overlap between categories enriched for genes with high correlations between expression and methylation. These fall into the seven color-coded partially overlapping functional groupings identified on the right side of the figure. The x-axis has the same GO categories identified for the y-axis (on the left of the figure), going from left to right instead of top to bottom. The fraction of

genes in common between different categories is indicated by the red-yellow-white color code, with the red diagonal line of boxes indicating each category versus itself. **B**, Histogram of the distribution of correlations of 17,144 transcript expression and DNA methylation data. Median values are shown for the transcript expression versus DNA methylation level correlations of the seven significant GO functional groupings from panel A, each of which is the composite of its component GO categories and is preceded by a red “GO”. In addition, the median values are included for 17 additional functional groups (defined in Supplemental Table 2), an “All genes” category of the 16,888 transcripts excluding the 256 microRNAs. The x-axis is the correlations of the transcript expression versus the DNA methylation values, and the y-axis is the frequency.

Figure 5. Representative scatter plots of DNA methylation versus transcript expression levels for genes from different functional categories (see Figure 4 and Supplemental Table 2). Transcript expression z scores values were obtained from CellMiner \ NCI-60 Analysis Tools \ Cell line signature \ Gene transcript z scores and are plotted on the x-axis. The transcript values are standard deviations from the mean expression with the 0 values demarked by dashed vertical lines for each gene. Methylation values were obtained from CellMiner \ NCI-60 Analysis Tools \ Cell line signature \ Gene methylation values (see Figure 4) and are plotted on the y-axis. For the methylation data, values of 0 correspond to 0% methylation, and 1 corresponds to 100%. The Pearson’s correlation between the two is “r”. Gene categories are as described in Supplemental Table 2. EMT is the abbreviation for epithelial-mesenchymal transition. Note the cluster of cell lines at the upper left of the 4 diagrams for epithelial genes (4 diagrams on the right in the upper row); these clusters correspond to cell lines having high DNA methylation and low expression, presumably representing mesenchymal cell lines.

Figure 6. Integration of NCI-60 genomic data. **A**, Cell line signatures for three types of genomic data for *RAB25*, *POLG*, and *CDKN2A*. Genomic signatures

were generated for DNA copy number, DNA methylation, and transcript expression using “CellMiner \ NCI-60 Analysis Tools \ Cell line signature”. The x-axis bars correspond to DNA copy number, gene methylation level (percentage/100), and transcript z-score, respectively. The y-axis is the NCI-60 cell lines in all cases, color-coded by tissue of origin. **B**, **P**lots of the linear regression coefficients of determination times the sign of the coefficients (R) of both DNA copy number versus expression and methylation versus expression. In all four plots, for both axes, R ranges from -1 to 1. An R of 0 indicates no predictive power. An R of 1 or -1 indicates perfect predictive power in the positive or negative sense, respectively. “All genes” is a smoothed scatter plot for the 15,798 genes with all three forms of data (Supplemental Table 3). “Epithelial”, “DNA damage response”, and “Tumor suppressors” are scatter plots of these gene groups from Figure 3B, defined in Supplemental Table 2. Note that the epithelial genes show deviations towards low methylation and little or no deviations with respect to gene copy number. **C**, Integration of four different forms of genomic data accounting for the microsatellite instability (MSI) phenotype in the NCI-60. Cell line signatures were generated for DNA copy number, DNA methylation, transcript expression (z scores), and genetic variant summation using “CellMiner \ NCI-60 Analysis Tools \ Cell line signature”. The x-axis is DNA copy number, gene methylation level (percentage/100), transcript z-score, and summation of variants (using the “Protein function affecting” output), respectively. The y-axis is the NCI-60 cell lines in all cases. The dotted lines are a visual aid to ease alignment of data by cell line. The red stars indicate the cell lines proposed to be affected functionally by the indicated molecular parameter. Multiple red stars in the “Mutation” bar graph for a single cell line indicate deleterious variants occurring in more than one of the genes. Microsatellite instability for the cell lines is as described previously (12,35). Note the strong positive relationship between mutation of the mismatch repair genes (*MLH1*, *MSH2*, and *MSH6*) and microsatellite instability, and that this occurs in colon cell lines, as expected, and also in other cell line types.

Table 1. Correlations between SLFN11 methylation and drug activities [a]

Corr. [b]	p-value [b]	NSC [c]	Name	Mechanism [d]	FDA Status [e]
-0.592	0.000001	119875	Cisplatin	A7|AlkAg	FDA approved
-0.547	0.000006	241240	Carboplatin	A7|AlkAg	FDA approved
-0.455	0.000479	757098	Melphalan	A7|AlkAg	FDA approved
-0.452	0.000285	609699	Topotecan	T1	FDA approved
-0.438	0.000529	759878	Irinotecan	T1	FDA approved
-0.336	0.009383	724998	LMP-400	T1	Clinical trial
-0.356	0.005304	279836	Mitoxantrone	T2	FDA approved
-0.338	0.009564	246131	Valrubicin	T2	FDA approved
-0.407	0.001238	613327	Gemcitabine	Ds	FDA approved
-0.416	0.000944	312887	Fludarabine	Ds|AM	FDA approved
-0.312	0.015334	63878	Cytarabine	Ds	FDA approved
-0.318	0.013378	32065	Hydroxyurea	AM|Dr	FDA approved
-0.427	0.000659	125066	Bleomycin	Db	FDA approved
-0.23	0.091124	747856	Olaparib	PARP	Clinical trial
0.144	0.286869	758645	Paclitaxel	Tu	FDA approved
0.159	0.274977	628503	Docetaxel	Tu	FDA approved
-0.085	0.522831	718781	Erlotinib	YK|PK:EGFR	FDA approved
0.147	0.267993	756645	Crizotinib	YK|PK:MET	Clinical trial
0.066	0.619252	761431	Vemurafenib	YK|PK:BRAF	FDA approved

[a] SLFN11 DNA methylation, and drug activity data is as obtained from CellMiner\NCI-60Analysis Tools\ Cell line signature", using either the, "Gene methylation values" or "Drug activity z scores" selections at https://discover.nci.nih.gov/cellminer/. The drug activity is from the Developmental Therapeutics Program, http://dtp.nci.nih.gov/.

[b] Corr. is Correlation (Pearson's coefficients). P-values were calculated within the "CellMiner\NCI-60 Analysis Tools\Pattern comparison" tool.

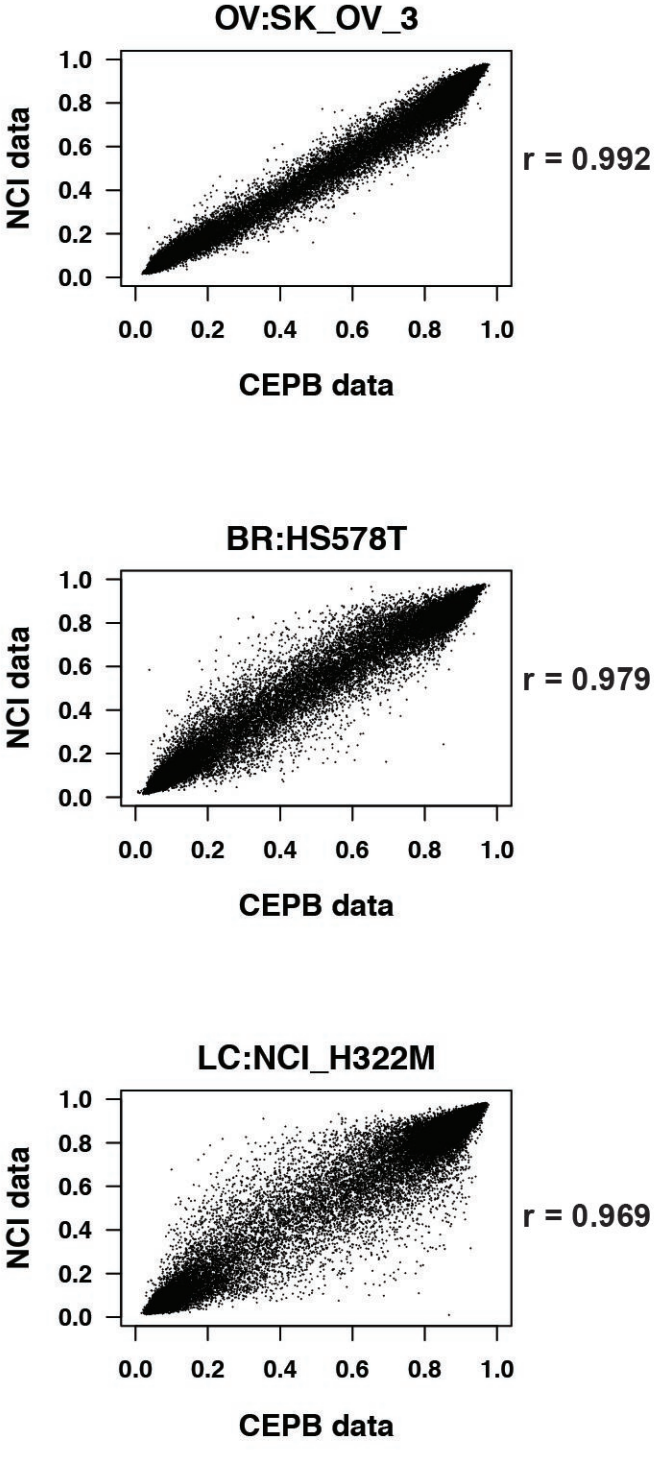
[c] Cancer Chemotherapy National Service Center number.

[d] A7|AlkAg is an alkylating agent at the N-7 position of guanine. T1 is topoisomerase I inhibitor. T2 is topoisomerase II inhibitor. Ds is DNA synthesis inhibitor. AM is antimetabolite. Dr is ribonucleotide reductase inhibitor. Db is DNA binder. PARP is Poly(ADP-ribose)polymerase inhibitor. Tu is tubulin affecting. YK is tyrosine kinase inhibitor. PK is protein kinase inhibitor. EGFR is EGFR inhibitor. MET is MET inhibitor. BRAF is BRAF inhibitor.

[e] FDA is the Food and Drug Administration.

Figure 1

A.


OV:SK_OV_3
NCI data
CEPB data
r = 0.992
BR:HS578T
NCI data
CEPB data
r = 0.979
LC:NCI_H322M
NCI data
CEPB data
r = 0.969

B.

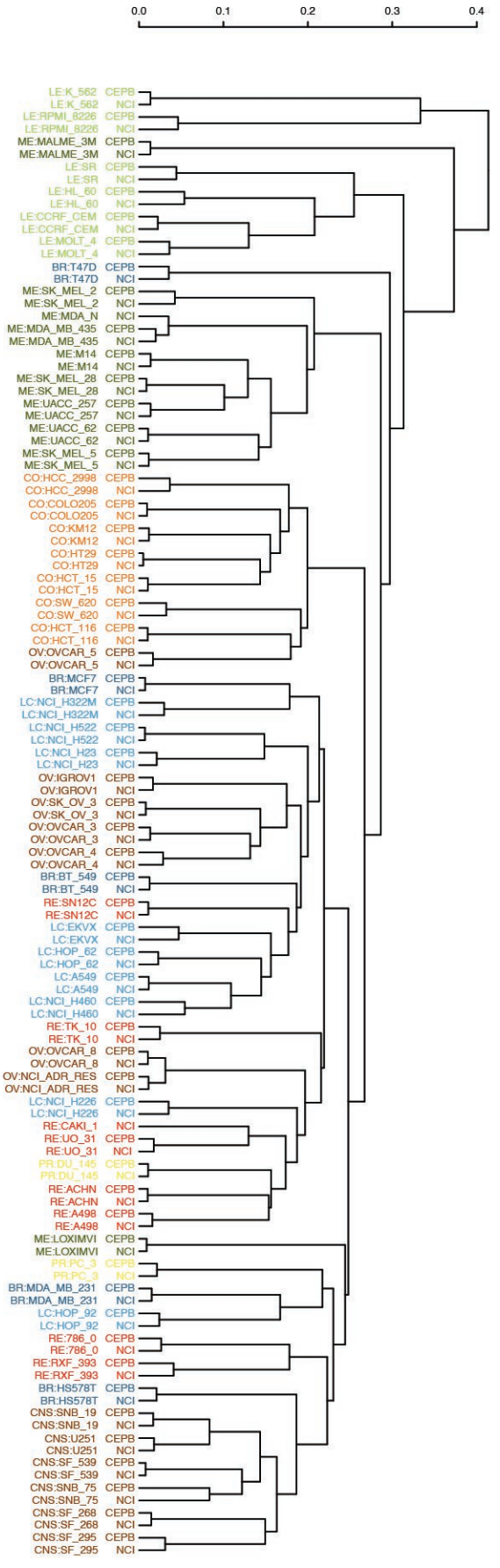

0.0
0.1
0.2
0.3
0.4
LE:K_562 CEPB
LE:K_562 NCI
LE:RPMI_8226 CEPB
LE:RPMI_8226 NCI
ME:MALME_3M CEPB
ME:MALME_3M NCI
LE:SR CEPB
LE:SR NCI
LE:HL_60 CEPB
LE:HL_60 NCI
LE:CCRF_CEM CEPB
LE:CCRF_CEM NCI
LE:MOLT_4 CEPB
LE:MOLT_4 NCI
BR:T47D CEPB
BR:T47D NCI
ME:SK_MEL_2 CEPB
ME:SK_MEL_2 NCI
ME:MDA_N NCI
ME:MDA_MB_435 CEPB
ME:MDA_MB_435 NCI
ME:M14 CEPB
ME:M14 NCI
ME:SK_MEL_28 CEPB
ME:SK_MEL_28 NCI
ME:UACC_257 CEPB
ME:UACC_257 NCI
ME:UACC_62 CEPB
ME:UACC_62 NCI
ME:SK_MEL_5 CEPB
ME:SK_MEL_5 NCI
CO:HCC_2998 CEPB
CO:HCC_2998 NCI
CO:COLO205 CEPB
CO:COLO205 NCI
CO:KM12 CEPB
CO:KM12 NCI
CO:HT29 CEPB
CO:HT29 NCI
CO:HCT_15 CEPB
CO:HCT_15 NCI
CO:SW_620 CEPB
CO:SW_620 NCI
CO:HCT_116 CEPB
CO:HCT_116 NCI
OV:OVCAR_5 CEPB
OV:OVCAR_5 NCI
BR:MCF7 CEPB
BR:MCF7 NCI
LC:NCI_H322M CEPB
LC:NCI_H322M NCI
LC:NCI_H522 CEPB
LC:NCI_H522 NCI
LC:NCI_H23 CEPB
LC:NCI_H23 NCI
OV:IGROV1 CEPB
OV:IGROV1 NCI
OV:SK_OV_3 CEPB
OV:SK_OV_3 NCI
OV:OVCAR_3 CEPB
OV:OVCAR_3 NCI
OV:OVCAR_4 CEPB
OV:OVCAR_4 NCI
BR:BT_549 CEPB
BR:BT_549 NCI
RE:SN12C CEPB
RE:SN12C NCI
LC:EKVX CEPB
LC:EKVX NCI
LC:HOP_62 CEPB
LC:HOP_62 NCI
LC:A549 CEPB
LC:A549 NCI
LC:NCI_H460 CEPB
LC:NCI_H460 NCI
RE:TK_10 CEPB
RE:TK_10 NCI
OV:OVCAR_8 CEPB
OV:OVCAR_8 NCI
OV:NCI_ADR_RES CEPB
OV:NCI_ADR_RES NCI
LC:NCI_H226 CEPB
LC:NCI_H226 NCI
RE:CAKI_1 NCI
RE:UO_31 CEPB
RE:UO_31 NCI
PR:DU_145 CEPB
PR:DU_145 NCI
RE:ACHN CEPB
RE:ACHN NCI
RE:A498 CEPB
RE:A498 NCI
ME:LOXIMVI CEPB
ME:LOXIMVI NCI
PR:PC_3 CEPB
PR:PC_3 NCI
BR:MDA_MB_231 CEPB
BR:MDA_MB_231 NCI
LC:HOP_92 CEPB
LC:HOP_92 NCI
RE:786_0 CEPB
RE:786_0 NCI
RE:RXF_393 CEPB
RE:RXF_393 NCI
BR:HS578T CEPB
BR:HS578T NCI
CNS:SNB_19 CEPB
CNS:SNB_19 NCI
CNS:U251 CEPB
CNS:U251 NCI
CNS:SF_539 CEPB
CNS:SF_539 NCI
CNS:SNB_75 CEPB
CNS:SNB_75 NCI
CNS:SF_268 CEPB
CNS:SF_268 NCI
CNS:SF_295 CEPB
CNS:SF_295 NCI

Figure 2

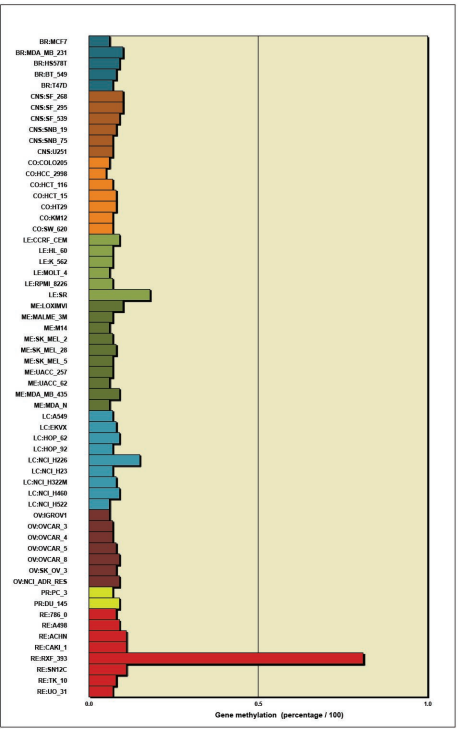
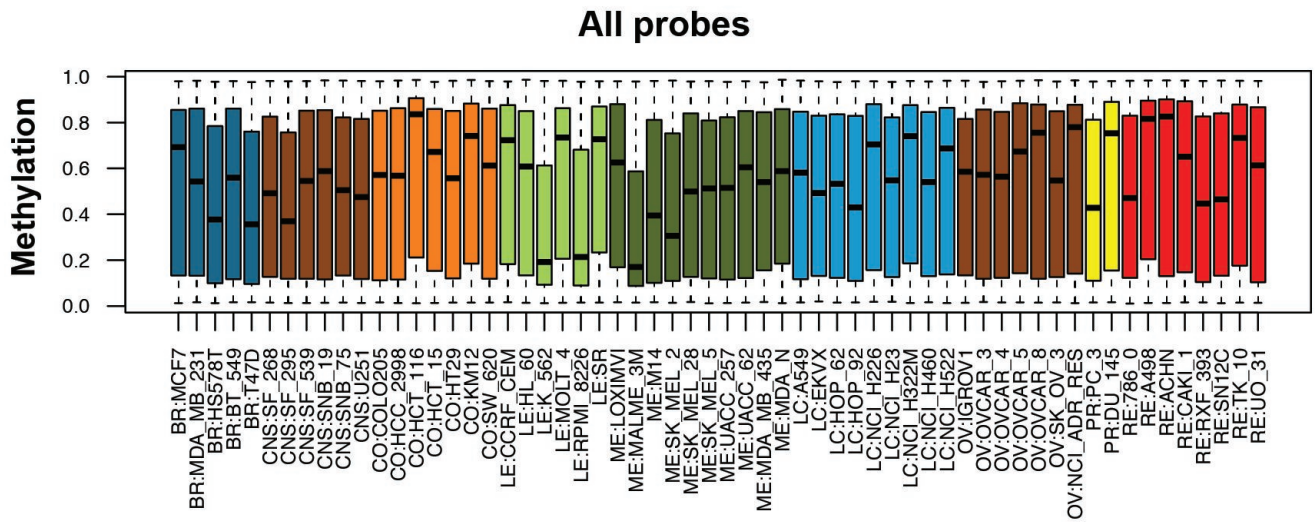
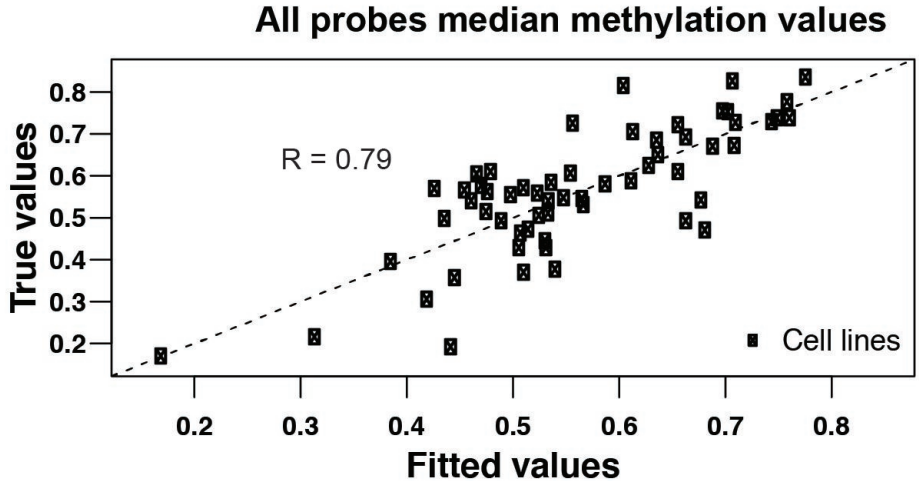

VHL
BR:MCF7
BR:MDA_MB_231
BR:HS578T
BR:BT_549
BR:T47D
CNS:SF_268
CNS:SF_295
CNS:SF_539
CNS:SNB_19
CNS:SNB_75
CNS:U251
CO:COLO205
CO:HCC_2998
CO:HCT_116
CO:HCT_15
CO:HT29
CO:KM12
CO:SW_620
LE:CCRF_CEM
LE:HL_60
LE:K_562
LE:MOLT_4
LE:RPMI_8226
LE:SR
ME:LOXIMVI
ME:MALME_3M
ME:M14
ME:SK_MEL_2
ME:SK_MEL_28
ME:SK_MEL_5
ME:UACC_257
ME:UACC_62
ME:MDA_MB_435
ME:MDA_N
LC:A549
LC:EKVX
LC:HOP_62
LC:HOP_92
LC:NCI_H226
LC:NCI_H23
LC:NCI_H322M
LC:NCI_H460
LC:NCI_H522
OV:IGROV1
OV:OVCAR_3
OV:OVCAR_4
OV:OVCAR_5
OV:OVCAR_8
OV:SK_OV_3
OV:NCI_ADR_RES
PR:PC_3
PR:DU_145
RE:786_0
RE:A498
RE:ACHN
RE:CAKI_1
RE:RXF_393
RE:SN12C
RE:TK_10
RE:UO_31
0.0
0.5
1.0
Gene methylation (percentage / 100)
SLFN11
BR:MCF7
BR:MDA_MB_231
BR:HS578T
BR:BT_549
BR:T47D
CNS:SF_268
CNS:SF_295
CNS:SF_539
CNS:SNB_19
CNS:SNB_75
CNS:U251
CO:COLO205
CO:HCC_2998
CO:HCT_116
CO:HCT_15
CO:HT29
CO:KM12
CO:SW_620
LE:CCRF_CEM
LE:HL_60
LE:K_562
LE:MOLT_4
LE:RPMI_8226
LE:SR
ME:LOXIMVI
ME:MALME_3M
ME:M14
ME:SK_MEL_2
ME:SK_MEL_28
ME:SK_MEL_5
ME:UACC_257
ME:UACC_62
ME:MDA_MB_435
ME:MDA_N
LC:A549
LC:EKVX
LC:HOP_62
LC:HOP_92
LC:NCI_H226
LC:NCI_H23
LC:NCI_H322M
LC:NCI_H460
LC:NCI_H522
OV:IGROV1
OV:OVCAR_3
OV:OVCAR_4
OV:OVCAR_5
OV:OVCAR_8
OV:SK_OV_3
OV:NCI_ADR_RES
PR:PC_3
PR:DU_145
RE:786_0
RE:A498
RE:ACHN
RE:CAKI_1
RE:RXF_393
RE:SN12C
RE:TK_10
RE:UO_31
0.0
0.5
1.0
Gene methylation (percentage / 100)
IRF6
BR:MCF7
BR:MDA_MB_231
BR:HS578T
BR:BT_549
BR:T47D
CNS:SF_268
CNS:SF_295
CNS:SF_539
CNS:SNB_19
CNS:SNB_75
CNS:U251
CO:COLO205
CO:HCC_2998
CO:HCT_116
CO:HCT_15
CO:HT29
CO:KM12
CO:SW_620
LE:CCRF_CEM
LE:HL_60
LE:K_562
LE:MOLT_4
LE:RPMI_8226
LE:SR
ME:LOXIMVI
ME:MALME_3M
ME:M14
ME:SK_MEL_2
ME:SK_MEL_28
ME:SK_MEL_5
ME:UACC_257
ME:UACC_62
ME:MDA_MB_435
ME:MDA_N
LC:A549
LC:EKVX
LC:HOP_62
LC:HOP_92
LC:NCI_H226
LC:NCI_H23
LC:NCI_H322M
LC:NCI_H460
LC:NCI_H522
OV:IGROV1
OV:OVCAR_3
OV:OVCAR_4
OV:OVCAR_5
OV:OVCAR_8
OV:SK_OV_3
OV:NCI_ADR_RES
PR:PC_3
PR:DU_145
RE:786_0
RE:A498
RE:ACHN
RE:CAKI_1
RE:RXF_393
RE:SN12C
RE:TK_10
RE:UO_31
0.0
0.5
1.0
Gene methylation (percentage / 100)

Figure 3

A.



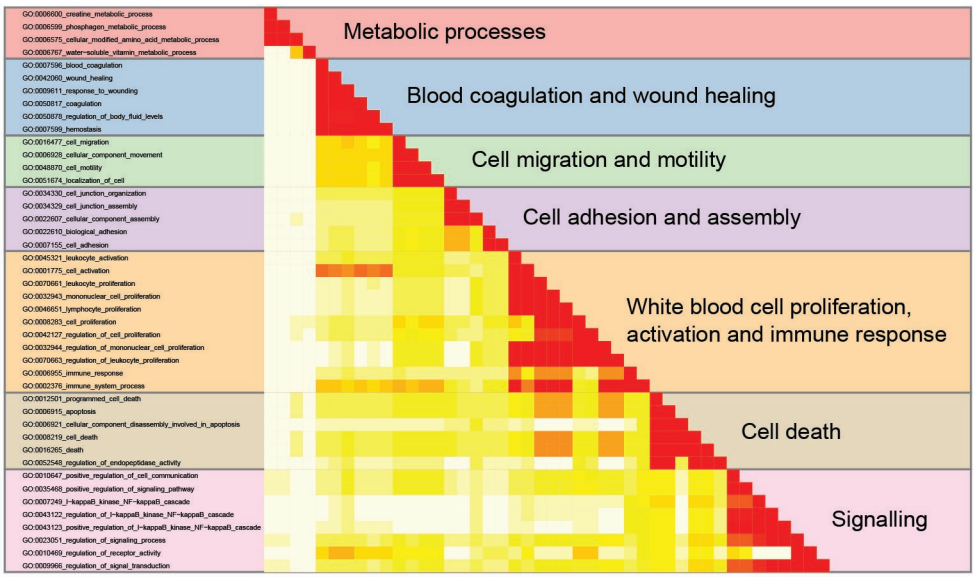
B.



Model:
PARP1 protein expression
MTA1 transcript expression
EP300 protein expression
KDM5C transcript expression
SMARCB1 genetic variants
UHRF1 transcript expression

Figure 4

A.


GO:0006600_creatine_metabolic_process
GO:0006599_phosphagen_metabolic_process
GO:0006575_cellular_modified_amino_acid_metabolic_process
GO:0006767_water-soluble_vitamin_metabolic_process
GO:0007596_blood_coagulation
GO:0042060_wound_healing
GO:0009611_response_to_wounding
GO:0050817_coagulation
GO:0050878_regulation_of_body_fluid_levels
GO:0007599_hemostasis
GO:0016477_cell_migration
GO:0006928_cellular_component_movement
GO:0048870_cell_motility
GO:0051674_localization_of_cell
GO:0034330_cell_junction_organization
GO:0034329_cell_junction_assembly
GO:0022607_cellular_component_assembly
GO:0022610_biological_adhesion
GO:0007155_cell_adhesion
GO:0045321_leukocyte_activation
GO:0001775_cell_activation
GO:0070661_leukocyte_proliferation
GO:0032943_mononuclear_cell_proliferation
GO:0046651_lymphocyte_proliferation
GO:0008283_cell_proliferation
GO:0042127_regulation_of_cell_proliferation
GO:0032944_regulation_of_mononuclear_cell_proliferation
GO:0070663_regulation_of_leukocyte_proliferation
GO:0006955_immune_response
GO:0002376_immune_system_process
GO:0012501_programmed_cell_death
GO:0006915_apoptosis
GO:0006921_cellular_component_disassembly_involved_in_apoptosis
GO:0008219_cell_death
GO:0016265_death
GO:0052548_regulation_of_endopeptidase_activity
GO:0010647_positive_regulation_of_cell_communication
GO:0035468_positive_regulation_of_signaling_pathway
GO:0007249_I-kappaB_kinase_NF-kappaB_cascade
GO:0043122_regulation_of_I-kappaB_kinase_NF-kappaB_cascade
GO:0043123_positive_regulation_of_I-kappaB_kinase_NF-kappaB_cascade
GO:0023051_regulation_of_signaling_process
GO:0010469_regulation_of_receptor_activity
GO:0009966_regulation_of_signal_transduction
Metabolic processes
Blood coagulation and wound healing
Cell migration and motility
Cell adhesion and assembly
White blood cell proliferation, activation and immune response
Cell death
Signalling
1.00
0.75
0.50
0.25
0.00
# of genes in common
# of genes in smaller category

B.

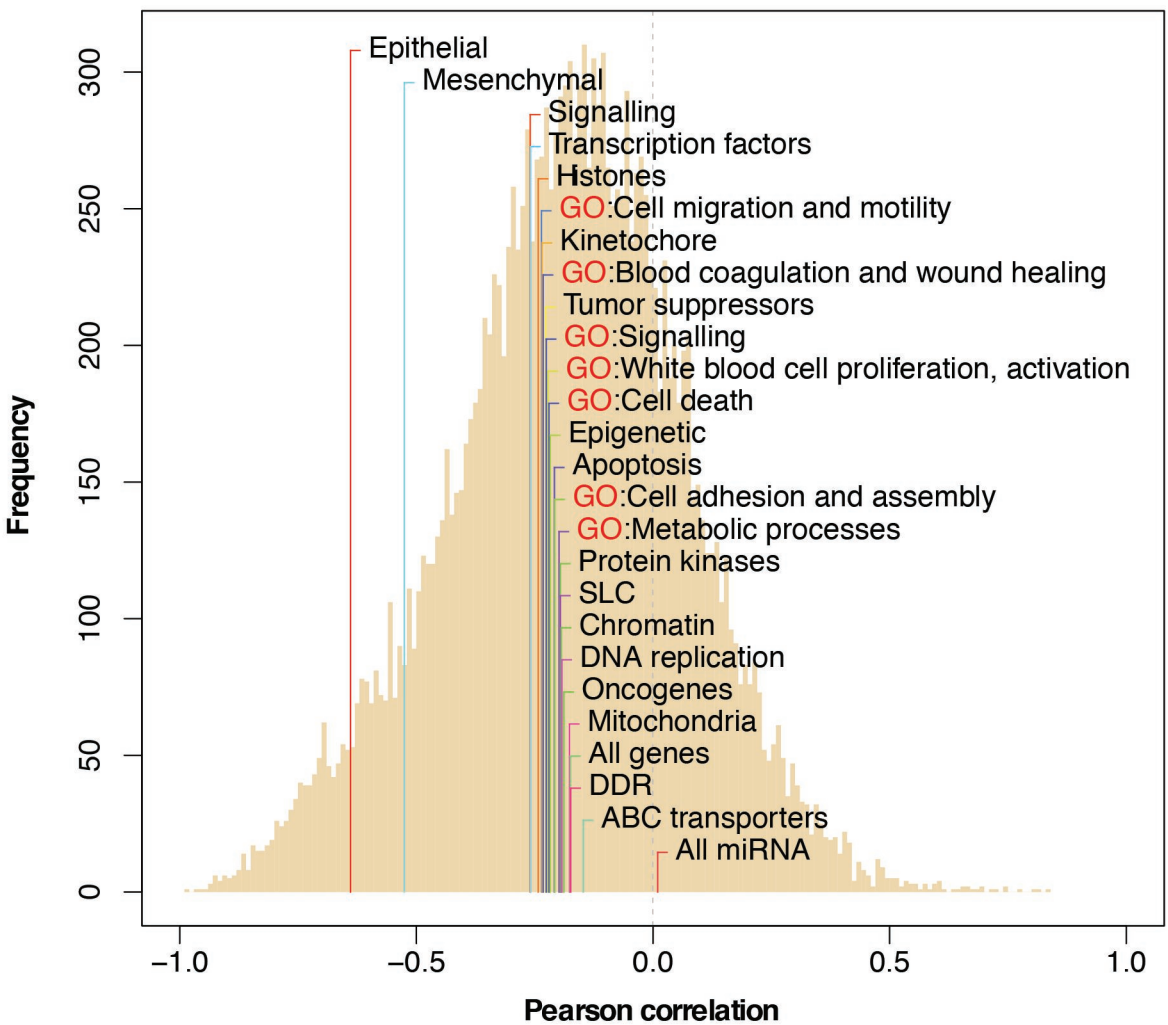

Epithelial
Mesenchymal
Signalling
Transcription factors
Histones
GO:Cell migration and motility
Kinetochore
GO:Blood coagulation and wound healing
Tumor suppressors
GO:Signalling
GO:White blood cell proliferation, activation
GO:Cell death
Epigenetic
Apoptosis
GO:Cell adhesion and assembly
GO:Metabolic processes
Protein kinases
SLC
Chromatin
DNA replication
Oncogenes
Mitochondria
All genes
DDR
ABC transporters
All miRNA
Frequency
0
50
100
150
200
250
300
−1.0
−0.5
0.0
0.5
1.0
Pearson correlation

Figure 5

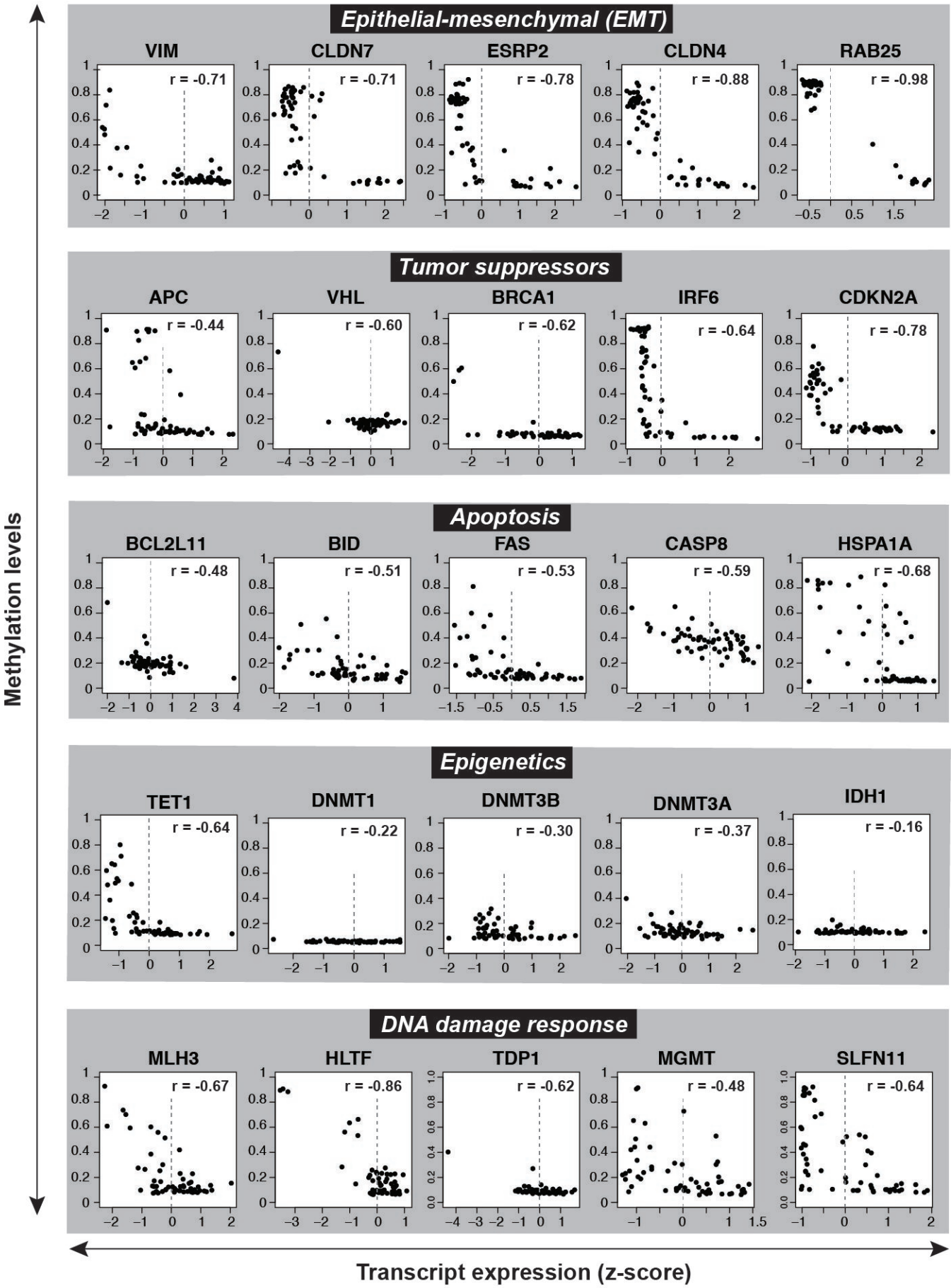
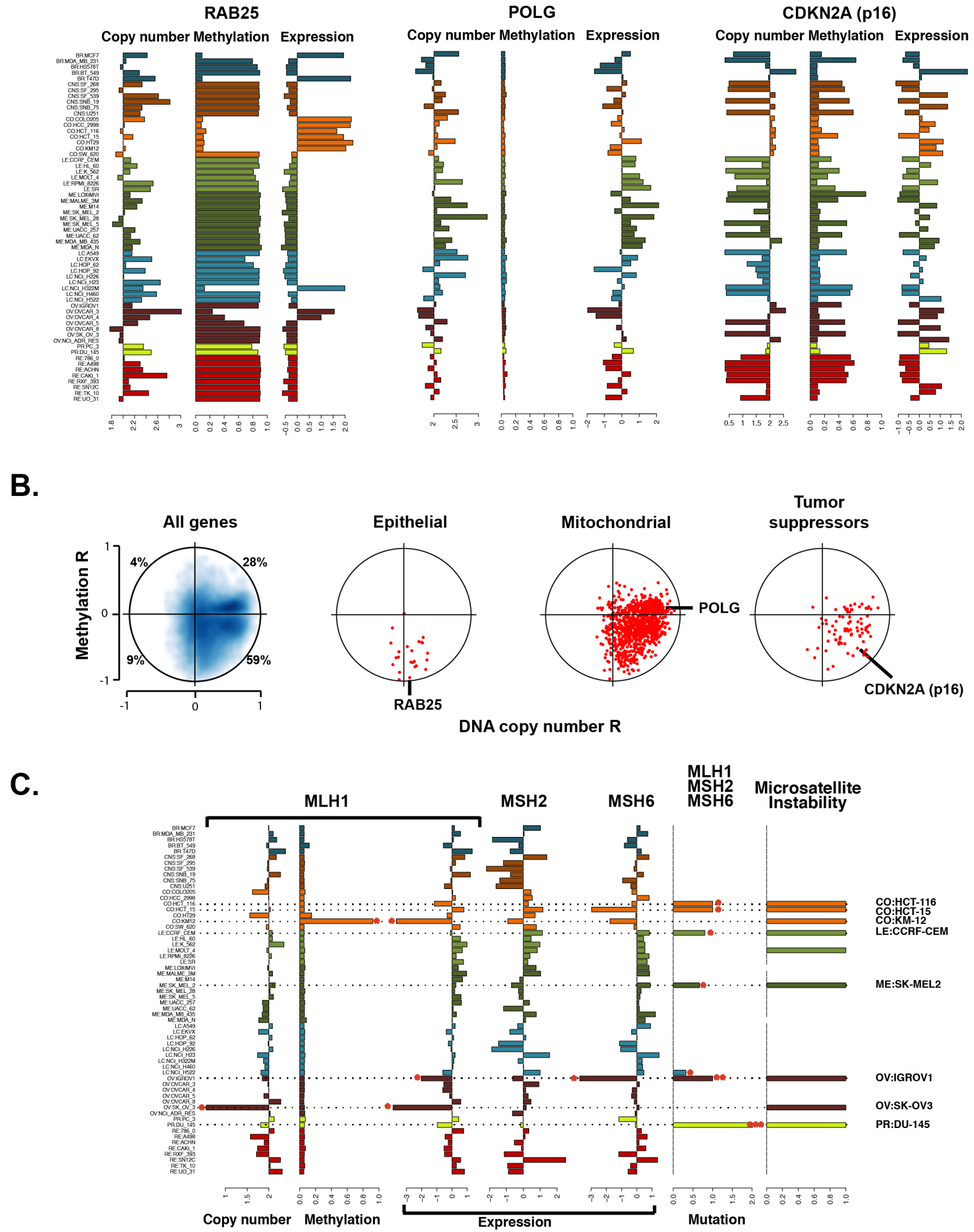

Epithelial-mesenchymal (EMT)
VIM
r = -0.71
CLDN7
r = -0.71
ESRP2
r = -0.78
CLDN4
r = -0.88
RAB25
r = -0.98
Tumor suppressors
APC
r = -0.44
VHL
r = -0.60
BRCA1
r = -0.62
IRF6
r = -0.64
CDKN2A
r = -0.78
Apoptosis
BCL2L11
r = -0.48
BID
r = -0.51
FAS
r = -0.53
CASP8
r = -0.59
HSPA1A
r = -0.68
Epigenetics
TET1
r = -0.64
DNMT1
r = -0.22
DNMT3B
r = -0.30
DNMT3A
r = -0.37
IDH1
r = -0.16
DNA damage response
MLH3
r = -0.67
HLTF
r = -0.86
TDP1
r = -0.62
MGMT
r = -0.48
SLFN11
r = -0.64
Methylation levels
Transcript expression (z-score)

Figure 6


A.
RAB25
POLG
CDKN2A (p16)
Copy number
Methylation
Expression
B.
All genes
Epithelial
Mitochondrial
Tumor suppressors
Methylation R
4%
28%
9%
59%
POLG
RAB25
CDKN2A (p16)
DNA copy number R
C.
MLH1
MSH2
MSH6
MLH1 MSH2 MSH6
Microsatellite Instability
CO:HCT-116
CO:HCT-15
CO:KM-12
LE:CCRF-CEM
ME:SK-MEL2
OV:IGROV1
OV:SK-OV3
PR:DU-145
Copy number
Methylation
Expression
Mutation