Published in final edited form as:

Sci Transl Med. 2011 March 23; 3(75): 75ra25. doi:10.1126/scitranslmed.3001875.

Breast Cancer Methylomes Establish an Epigenomic Foundation for Metastasis

Fang Fang^{1,*}, Sevin Turcan^{1,*}, Andreas Rimmer^{2,*}, Andrew Kaufman¹, Dilip Giri³, Luc G. T. Morris^{1,4}, Ronglai Shen⁵, Venkatraman Seshan⁵, Qianxing Mo⁵, Adriana Heguy¹, Stephen B. Baylin⁶, Nita Ahuja⁶, Agnes Viale⁷, Joan Massague⁸, Larry Norton⁹, Linda T. Vahdat¹⁰, Mary Ellen Moynahan⁹, and Timothy A. Chan^{1,2,†}

¹Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

²Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

³Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

⁴Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

⁵Department of Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

[†]To whom correspondence should be addressed. chant@mskcc.org.

*These authors contributed equally to this work.

Author contributions: T.A.C., S.T., and F.F. designed the experiments. F.F., S.T., A.R., D.G., A.H., A.V., and A.K. performed the experiments. F.F., S.T., T.A.C., L.G.T.M., R.S., V.S., Q.M., and M.E.M. analyzed the data. S.B.B., N.A., J.M., L.N., and L.T.V. contributed new reagents/analytic tools/expertise. T.A.C., F.F., and S.T. wrote the paper.

Competing interests: S.B.B. is a consultant for MDX in regard to MSP (methylation-specific polymerase chain reaction) assays and receives unrestricted grant support from the company. MDX has no known financial interests associated with this work. The other authors declare that they have no competing interests.

Accession numbers: Data sets are deposited in the Gene Expression Omnibus under accession number GSE26349. The Cancer Genome Atlas Project GBM cancer data sets are publicly available at <http://cancergenome.nih.gov>.

SUPPLEMENTARY MATERIAL

www.sciencetranslationalmedicine.org/cgi/content/full/3/75/75ra25/DC1

Fig. S1. Methylome analysis using the Infinium HumanMethylation27 platform.

Fig. S2. Statistical output from K-means consensus clustering analysis on the 39 breast cancers in the discovery set using the most variant probes ($n = 1359$) (most variant 5% based on ranked SD).

Fig. S3. Validation of B-CIMP loci and the methylation alterations in B-CIMP.

Fig. S4. GSEA results of CIMP repression signature on the van't Veer data set.

Fig. S5. GSEA results with Suz12-associated genes.

Fig. S6. GSEA results with polycomb repressive complex 2 (PRC2) target genes.

Fig. S7. GSEA results of developmental transcription factors bound by Suz12.

Table S1. Clinical characteristics and methylome analysis of breast cancers (Excel files).

Table S2. EpiTYPER primers.

Table S3. Sensitivity and specificity for genes used for B-CIMP validation.

Table S4. Multivariate analysis shows CIMP predicts metastatic risk.

Table S5. Validation samples clinical information.

Table S6. Gene ontology analysis of significantly altered genes in B-CIMP.

Table S7. Data sets for concepts mapping of metastasis transcriptomes.

Table S8. References for Oncomine data sets used in concepts analysis.

Table S9. GSEA output for CIMP gene analysis using polycomb occupancy data sets.

Table S10. GSEA enrichment results for B-CIMP genes (CIMP⁺ versus CIMP⁻) using Molecular Signatures Database.

Table S11. Data sets for PRC enrichment analysis.

Table S12. Data for polycomb occupancy analysis of B-, C-, and G-CIMP.

Table S13. Gene summary for polycomb occupancy analysis of B-, C-, and G-CIMP (Excel files).

Table S14. Genes common to CIMP in breast cancer, colon cancer, and glioblastoma.

⁶Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, 1650 Orleans Street, Suite 544, Baltimore, MD 21231, USA

⁷Genomics Core, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

⁸Department of Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

⁹Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

¹⁰Department of Medicine, Weill Cornell Medical Center, 425 East 61st Street, New York, NY 10065, USA

Abstract

Cancer-specific alterations in DNA methylation are hallmarks of human malignancies; however, the nature of the breast cancer epigenome and its effects on metastatic behavior remain obscure. To address this issue, we used genome-wide analysis to characterize the methylomes of breast cancers with diverse metastatic behavior. Groups of breast tumors were characterized by the presence or absence of coordinate hypermethylation at a large number of genes, demonstrating a breast CpG island methylator phenotype (B-CIMP). The B-CIMP provided a distinct epigenomic profile and was a strong determinant of metastatic potential. Specifically, the presence of the B-CIMP in tumors was associated with low metastatic risk and survival, and the absence of the B-CIMP was associated with high metastatic risk and death. B-CIMP loci were highly enriched for genes that make up the metastasis transcriptome. Methylation at B-CIMP genes accounted for much of the transcriptomal diversity between breast cancers of varying prognosis, indicating a fundamental epigenomic contribution to metastasis. Comparison of the loci affected by the B-CIMP with those affected by the hypermethylator phenotype in glioma and colon cancer revealed that the CIMP signature was shared by multiple human malignancies. Our data provide a unifying epigenomic framework linking breast cancers with varying outcome and transcriptomic changes underlying metastasis. These findings significantly enhance our understanding of breast cancer oncogenesis and aid the development of new prognostic biomarkers for this common malignancy.

INTRODUCTION

Breast cancer is one of the most prevalent human malignancies and is a major cause of cancer-related morbidity and mortality. Invasive ductal carcinoma (IDC) of the breast is a phenotypically diverse disease, consisting of tumors with varying pathologic and molecular characteristics (1-5). The primary biological subtypes of IDC include estrogen receptor (ER)- and progesterone receptor (PR)-positive tumors (luminal A and B), tumors that are human epidermal growth factor receptor 2 (HER2)-enriched, and tumors that are ER/PR-negative (basal-like). These molecular determinants have significant effects on metastatic behavior and clinical outcome. For example, ER/PR⁺ tumors are generally associated with better clinical prognosis, whereas basal-like (ER/PR⁻ and HER2⁻, triple-negative) tumors are associated with higher rates of metastasis and death (6-9). The genomic alterations, including both genetic and epigenetic aberrations, underlying these differing metastatic potentials are ill-defined.

Significant effort has been undertaken to more accurately define the molecular alterations underlying breast cancer. For example, it has been shown that hormone receptor (HR) status is prognostic for clinical outcome. Mutations in genes such as *BRCA1*, *PTEN*, and *PIK3CA* help promote breast cancer oncogenesis and are enriched in specific subgroups of IDC (10-12). Genome-wide sequencing surveys have been performed to identify the scope of mutations in breast cancers (13-15). These data demonstrate that there exists substantial biological heterogeneity between and within the ER/PR⁺ and ER/PR⁻ subgroups for which

the molecular foundations remain obscure (13). In addition, gene expression classifiers have been developed to help predict metastatic risk (16-18). Despite their increasing use in the clinic, the genomic root causes of the transcriptome differences that underlie metastatic potential are unclear.

It is well established that widespread changes in DNA methylation patterns occur during oncogenesis and tumor progression (19, 20). Cancer-specific changes in DNA methylation can alter genetic stability, genomic structure, and gene expression (21, 22). Promoter CpG island methylation can result in transcriptional silencing and plays an important role in the oncogenic process (19). A CpG island methylator phenotype (CIMP), which is associated with a strong tendency to hypermethylate specific loci, has been described in a subset of colorectal cancers, and recently in a subgroup of gliomas (2, 23). Aberrations in DNA methylation have been reported in human breast cancer, but the impact of the methylome on metastasis and the presence of a B-CIMP have remained elusive (24-30). To resolve these questions, we conducted a systematic, genome-wide characterization of the breast cancer methylome in breast cancers with diverse metastatic behavior.

RESULTS

Identification and validation of distinct DNA methylome subgroups of breast cancer

The Illumina Infinium HumanMethylation27 platform, which consists of bead arrays that query 27,578 CpG sites each covering 14,495 genes, provides efficient genome-wide interrogation of CpG islands and is both well validated and highly reproducible (mean correlation coefficient = 0.987) (31, 32). Analyses of replicate breast cancer samples by this platform generated highly concordant data (fig. S1). Therefore, we used this platform to analyze a discovery set of IDCs with differing metastatic behavior, including samples with varying ER/PR and HER2 status, from patients with complete clinical follow-up ($n = 39$, table S1).

To identify breast cancer subgroups, we selected the most variant probes and performed consensus clustering (Fig. 1A, fig. S2, A to D, and table S1) and unsupervised hierarchical clustering (Fig. 1B). We identified two robust DNA methylation clusters, one encompassing a portion of the HR⁺ tumors (defined as ER/PR⁺, cluster 2) and one encompassing tumors that were ER/PR⁺ or ER/PR⁻ (cluster 1). Cluster 2 tumors had a highly characteristic DNA methylation profile with high coordinate cancer-specific hypermethylation at a subset of loci (Figs. 1 and 2A and fig. S2), similar to the CIMP phenotype seen in colorectal cancer (for details of hypermethylation definition, see Materials and Methods) (2, 23). Thus, we defined this group (cluster 2) as having a breast CpG island methylator phenotype (B-CIMP).

In our discovery set, 17 of 39 (44%) tumors were B-CIMP⁺. The composition of the B-CIMP⁺ subgroup was confirmed by two independent clustering algorithms [two-dimensional (2D) hierarchical and *K*-means consensus clustering] (Fig. 1), and both approaches identified the same set of tumors as exhibiting a B-CIMP. Moreover, SigClust evaluation of cluster significance found class boundaries to be significant (Fig. 1B) (33). Indeed, a number of genes, such as *ARHGEF7*, *RASGRF2*, and *SOX8*, showed marked differences between CIMP⁺ and CIMP⁻ tumors (examples in Fig. 2A).

Our array results were validated with EpiTYPER (34), a mass spectrometry-based technique, allowing sensitive detection of DNA methylation at base-pair resolution (fig. S3A and table S2). We observed perfect concordance between B-CIMP calls on the Infinium platform and those from EpiTYPER.

These results demonstrate that there exist profound differences across the methylomes of breast cancers. Although it is possible that additional smaller subgroups exist, the robust nature of our consensus modeling here suggests that there are two primary epigenomic subgroups of IDC as defined by genome-wide methylome profiling.

Clinical characterization and validation of B-CIMP in breast cancer

We then defined the relationship among B-CIMP status, clinical co-variables, and known molecular determinants. CIMP⁺ tumors consisted almost entirely of ER/PR⁺ tumors (94%, 16 of 17) (Fig. 2B), whereas CIMP⁻ tumors consisted of similar numbers of HR⁺ (45%, 10 of 22) and HR⁻ tumors (55%, 12 of 22) ($P = 0.001$, χ^2 test) (Fig. 2B). In contrast, there was no significant difference in frequency of HER2 positivity between the CIMP groups (Fig. 2B). As expected, average methylation intensities were significantly different between CIMP⁺ versus CIMP⁻, ER/PR⁺ versus ER/PR⁻, and HR⁺/CIMP⁺ versus HR⁺/CIMP⁻ tumors (Fig. 3, A to D).

There was already a strong trend toward improved distant metastasis-free survival in patients with CIMP⁺ tumors in our initial discovery set ($P = 0.078$) (Fig. 3E). To validate the existence of the B-CIMP, we used EpiTYPER to evaluate an independent cohort of 132 primary breast cancers. We examined methylation of three loci (*ALX4*, *ARHGEF7*, and *RASGRF2*) that were among the most predictive for B-CIMP in our Infinium data. None of these genes are methylated in normal breast (table S1). Our microarray data indicated that methylation at two of three genes from this set correctly identified ~95% of CIMP⁺ tumors (table S3). In this validation set, B-CIMP⁺ tumors demonstrated a significantly lower risk for metastatic relapse ($P = 0.007$) and death ($P = 0.006$) (Fig. 3, F to I).

The presence of the B-CIMP was associated with ER/PR status (Fig. 2B) but determined metastatic risk within the HR⁺ group as well (Fig. 3I). ER/PR⁺ tumors that were CIMP⁻ were associated with significantly worse prognosis than ER/PR⁺ CIMP⁺ tumors. Furthermore, multivariate analysis showed that the B-CIMP was a strong predictor of prognosis ($P = 0.03$; hazard ratio, 0.49) independent of stage, age, nodal status, and HR status (tables S4 and S5). Most of the differentially methylated CpG sites were within known CpG islands, although a portion were not, and most sites were within 500 base pairs (bp) of the transcriptional start site (Fig. 4A).

The methylation landscape of breast cancer and its effects on the metastasis transcriptome

We next sought to define the nature of the methylome differences between the B-CIMP subgroups and characterize the effects of these differences on the breast cancer transcriptome. Probes were filtered for analysis by ranking log₂-transformed β values using decreasing false discovery rate (FDR)-adjusted P values and increasing β -value difference to identify the most differentially hypermethylated genes in the B-CIMP group. Of the 3297 CpG sites that were differentially methylated between CIMP⁺ and CIMP⁻ tumors, 2333 (71%) were hypermethylated (Fig. 4B, table S1, and fig. S3B). There were 2543 unique genes represented within this group, including 1764 that were hypermethylated and 779 that were hypomethylated. Consistent with the results in Fig. 1B, a volcano plot showing differentially methylated genes between B-CIMP⁺ and B-CIMP⁻ tumors was highly asymmetric, with many more hypermethylated genes in CIMP⁺ tumors (fig. S3B).

Affymetrix transcriptome data were obtained from the same breast tumors analyzed for methylation to determine which genes demonstrated differential expression and methylation. A total of 279 genes were significantly down-regulated and 238 genes were significantly up-regulated (table S1). Gene ontology (GO) analysis showed that the significantly up-

regulated genes were highly enriched for functional categories involving cell motion, angiogenesis, apoptosis, development, kinase activity, and DNA binding (table S6). The down-regulated genes were enriched for functional categories involved in mitosis, cytokinesis, exocytosis, chromosomal segregation, transcription factor activity, and kinase activity (table S6).

A starburst plot showing the relationship between DNA methylation and expression levels is shown in Fig. 4B. Here, significance levels of methylation (*x* axis) versus expression (*y* axis) differences between CIMP⁺ and CIMP⁻ tumors were plotted. Integration of the normalized gene expression and DNA methylation gene sets identified 102 genes with both significant hypermethylation and down-regulation in B-CIMP⁺ tumors (Fig. 4B and table S1). Among these genes are those that are involved in breast cancer outcome or epithelial-mesenchymal transition (EMT), including *LYN*, *MMP7*, *KLK10*, and *WNT6* (35-38). GO analysis showed that B-CIMP-specific down-regulation of genes (hypermethylated and down-regulated in B-CIMP) was associated with cell motion, development, signaling, and catalytic activity, as some of the most relevant functional categories (table S6).

Although mRNA expression signatures have been developed to help predict the risk of metastatic disease in breast cancer patients, the genomic foundations for these differences in gene expression are incompletely understood (17, 39, 40). Few genetic changes have been shown to be causally related to these transcriptional differences. Because B-CIMP status affects metastatic risk, we wondered whether methylation helps account for the transcriptome diversity underlying common breast cancer prognostic expression signatures. To address this question, we performed concept mapping analysis as previously described (41). The methylated and down-regulated genes that make up the transcriptomic footprint of the B-CIMP (B-CIMP repression signature) were markedly enriched among the most differentially expressed genes defining prognosis in multiple breast cancer cohorts (Fig. 5A). Low expression of genes in the B-CIMP repression signature was seen in tumors that did not metastasize, and high expression of the signature was seen in tumors that metastasized and/or resulted in poor survival (Fig. 5A and tables S7 and S8).

We observed significant associations between B-CIMP genes and breast cancer relapse expression signatures from multiple independent data sets, confirming the validity of our findings. Using the van't Veer cohort, we demonstrated that the presence of the B-CIMP repression signature strongly predicted survival (Fig. 5B). Again, breast cancers in which the CIMP repression signature was present were associated with significantly better survival than tumors lacking the signature. Furthermore, gene set enrichment analysis (GSEA) demonstrated a significant inverse correlation between B-CIMP-repressed genes and genes up-regulated in highly metastatic tumors (fig. S4). These findings indicate that epigenomic alterations associated with the B-CIMP underlie many of the gene expression differences observed in currently used breast cancer prognostic signatures such as MammaPrint.

Distinct gene sets targeted by the B-CIMP

To elucidate the differences in the methylation landscape between the two epigenomic subclasses, we mapped regions of the most significant methylation differences between CIMP⁺ and CIMP⁻ tumors across the genome (Fig. 6). Dense clusters of methylation density were apparent in the arms of a number of chromosomes. GSEA of the differentially methylated genes showed a significant enrichment for polycomb complex 2 (PRC2) targets (Fig. 6, figs. S5 to S7, and table S9) (42). Within the B-CIMP gene set, genes targeted by PRC2-associated marks or proteins were significantly overrepresented; these included H3K27 methylation, Suz12, and EZH2. These genes are noted along the genome, adjacent to the B-CIMP genes, highlighting the similarity between these two gene sets (Fig. 6).

GSEA analysis using the Broad molecular signature database demonstrated that CIMP genes were most significantly enriched in polycomb (PcG) occupancy data sets, although other processes were also implicated, including EMT and Wnt signaling, which are known to contribute to metastasis (tables S10 and S11) (43). It has been shown that the presence of a bivalent chromatin mark involving the key PcG mark, trimethylated H3K27, in stem cells may predispose specific genes to become hypermethylated and silenced in cancer and may be indicative of a contribution of stem cells to the derivation of specific cancers (44, 45). Perhaps this process is active in breast tumors of the B-CIMP subclass.

CIMP across multiple human cancers

Does the CIMP target the same genes in different human tumor types? We compared the CIMP-associated loci from breast cancer, colon cancer, and glioma (publicly available from The Cancer Genome Atlas, <http://cancergenome.nih.gov>). CIMP-associated genes were defined for glioma and colon cancer with the same methodology as above and were consistent with previous data (1, 2). Colon CIMP (C-CIMP) genes were derived from Memorial Sloan-Kettering Cancer Center (MSKCC) tumors ($n = 24$) with hierarchical clustering and confirmed as described in Materials and Methods and by Weisenberger *et al.* (2). Glioma CIMP (G-CIMP) genes were as previously described (1). All data sets were generated with the same Infinium HumanMethylation27 platform and were directly comparable.

We first wished to determine whether CIMP selectively targeted PcG targets not only in breast cancer but in other malignancies as well. All methylated loci (β -value FDR-corrected $P < 0.05$) in the three tumor types were compared with previously generated global PcG target gene sets (46, 47) (tables S12 and S13). Highly significant overlap was observed between CIMP and PcG targets in breast, glioma, and colon cancer (Fig. 7, A and B), potentially indicating that the CIMP may use similar processes across cancer types. Using the 33 most significant common predictors of CIMP, we generated a consensus signature for CIMP positivity across these tumor types (Fig. 7C and table S14). Breast tumors that are identified as B-CIMP⁺ by the 3-gene signature used above are concordant with those positive for the 33-gene signature and thus predicted good prognosis in breast cancer as shown in Fig. 3E. We then verified that this same 33-gene signature was associated with better prognosis in glioblastoma multiforme (GBM) (Fig. 7C). Thus, the CIMP imparted a favorable clinical prognosis in breast cancer, colon cancer (CIMP-hi, microsatellite unstable) (2, 48), and glioma (1), and as such, this epigenomic signature may be useful as an indicator of outcome across multiple human malignancies.

DISCUSSION

Our findings have several important implications for the understanding of breast cancer. First, we have definitively identified distinct epigenomic subtypes of breast cancer and documented the existence of a global CIMP in breast cancer. Aberrant hypermethylation of genes has been described in breast cancer previously (26, 49-51), and the methylation state of specific genes has been linked to outcome (52-54). However, the existence of a global B-CIMP has remained elusive before our study.

Our global approach robustly identified the B-CIMP as a characteristic of a subset of hormone-positive tumors. The ER has a profound effect on transcriptional activation and repression of many genes, and it may be that these changes contribute to patterning the epigenome. However, from our data, it does not appear that ER/PR positivity alone dictates B-CIMP status because our analysis above repeatedly identified a subset of hormone-positive tumors that were not positive for CIMP. Perhaps, other genetic or epigenetic events are important in dictating epigenomic fate in these tumors. Also, there may exist other

epigenomic subgroups of breast cancer within the major methylome subclasses that we describe here. Additional studies will be required to address this possibility.

In our study, B-CIMP⁺ tumors demonstrated a lower propensity for metastasis and a better clinical outcome than B-CIMP⁻ tumors. The association of better clinical outcome with CIMP⁺ tumors could be seen across multiple malignancies (breast, colon, and glioma) (2, 23). In these tumors, it may be that the epigenomic defects causing the CIMP initially helped promote neoplastic transformation but inactivate genes that facilitate tumor aggressiveness in later stages of cancer progression. It is important to note, however, that the association of methylation at CIMP genes with good clinical outcome is not universally applicable to methylation at all genes. Methylation of specific candidate genes or groups of genes has been associated with poorer prognosis, and these genes may have an effect on tumor aggressiveness independent of the CIMP (27, 53, 55-57). Genes such as these, including *CDKN2A*, *PTPRD*, and *BRCA1*, were not included among the B-CIMP loci.

The genomic basis of prognostic transcriptional signatures is unclear. Our data demonstrate that aberrations in the DNA methylome explain many of the mRNA expression differences that underlie these signatures. The tight association of these changes with a genome-wide concerted hypermethylation phenotype and their enrichment for PcG targets argues against the inactivation of these genes as being sporadic events. Rather, the B-CIMP phenotype is consistent with a global, systematic derangement in epigenetic regulation. The methylome profiles we have derived and the associated CIMP repression signature provide a previously unknown mechanistic link between breast cancers with differing metastatic behavior and transcriptional signatures that predict metastatic relapse. However, although we show that methylation-associated gene silencing underlies many metastasis-associated gene expression changes, genetic changes are undoubtedly important as well. Indeed, mutations of a number of genes such as *BRCA1*, *PTEN*, and *ERBB2* have been shown to be associated with an increased risk of metastasis (5, 58, 59). The relationship between these mutations and the B-CIMP phenotype remains to be elucidated, and it is likely that both genetic and epigenetic alterations contribute to the metastatic phenotype. *BRCA1* has recently been shown to up-regulate *DNMT1*, which may help explain the association between *BRCA1* mutation, basal-type tumors, and the lack of methylation we have observed in our study among HR⁻ breast cancers (60). Future studies will be required to define any potential causal relationship between mutations and derangements in the epigenomic landscape.

Our data show a large-scale consensus between CIMP genes from different human cancers. The CIMP in these different malignancies target many of the same genes, which are PcG targets. We speculate that these similarities reflect common mechanistic foundations. Despite their similarities, differences do exist between the PcG targets that comprise the B-, C-, and G-CIMP, which may reflect a degree of tissue or organ specificity.

In summary, these data provide an epigenomic framework for understanding the transcriptomic signatures present in breast cancers with differing metastatic behaviors. Our findings may enable the development of new molecular diagnostics that more accurately reflect the epigenomic underpinnings of breast cancer prognosis. These diagnostics may help further refine our ability to implement personalized medicine for breast cancer patients.

MATERIALS AND METHODS

Tumor samples

Breast tumors (discovery set, $n = 39$; validation set, $n = 132$) from the MSKCC were obtained after patient consent and with institutional review board (IRB) approval. For the breast tumor data, tissues from primary breast cancers were obtained from therapeutic

procedures performed as part of routine clinical management. The samples that make up the discovery set were chosen to include tumors reflecting the spectrum of different types of breast cancers in representative frequencies (ER/PR⁺, 67%; HER2 status, 50%; distant failures, 36%). The validation set was chosen on the basis of available material and in order for the set to be representative of known frequencies of ER status, HER2 status, stage, and distant failure status (8). Notably, both ER/PR⁺ and ER/PR⁻ tumors, HER2⁺ and HER2⁻ tumors, and tumors that resulted in metastases and those that did not were included. For the validation set, of 132 cases with an estimated event rate of 40%, we expected a power of >95% ($\alpha = 0.05$). Source DNAs or RNAs were extracted from frozen or paraffin-embedded primary tumors for the methylation and expression studies. Frozen samples were “snap-frozen” in liquid nitrogen and stored at -80°C . Each sample was examined histologically with hematoxylin and eosin (H&E)-stained sections. Regions were microdissected from the slides to provide a consistent tumor cell content of more than 70% in tissues used for analysis. Genomic DNA was extracted with the QIAamp DNA Mini kit or the QIAamp DNA FFPE Tissue kit (Qiagen) using the manufacturer’s instructions. RNA was extracted with Trizol (Invitrogen) according to the manufacturer’s directions. Nucleic acid quality was determined with the Agilent 2100 Bioanalyzer. Nucleic acids from the discovery set were used for methylation and expression analysis as described below.

Methylome analysis using the Infinium platform

Bisulfite-converted genomic DNA was analyzed with the Infinium HumanMethylation27 Beadchip Kit (Illumina, WG-311-1202) by the MSKCC Genomics Core. Processing of the array was per the manufacturer’s protocol. Methylation analysis controls included *in vitro*-methylated DNA (positive control) (61) and human HCT116 DKO DNA [DNA methyltransferase double-knockout cells (DNMT1 and DNMT3b)] (62).

Gene expression profiling

Methods for RNA extraction, labeling, and hybridization for DNA microarray analysis have been described previously (39). All gene expression analyses were performed with the Affymetrix Human Genome U133A 2.0 microarray.

Data analysis for genomics

For expression analysis, the Affymetrix data were imported into the Partek Genomics Suite (Partek Inc.). Data were normalized, log-transformed, and median-centered for analysis. Analysis of variance (ANOVA) followed by false discovery correction (FDR) (63, 64) was used to identify genes that were differentially expressed between the CIMP groups. Hierarchical clustering was performed with either Euclidean distance or Pearson correlation. SigClust significance as implemented in the R package SigClust was used as described (33). For Gene Ontology analysis, functional analysis of gene lists was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) (65, 66) and the PANTHER functional annotation classes. PANTHER categories with adjusted *P* values (FDR-corrected with Benjamini-Hochberg) of <0.05 were considered as significantly overrepresented in our gene lists.

For methylation analysis, Illumina data were imported into Partek with custom software. β Values were logit-transformed and mean-centered before analysis. ANOVA with false discovery correction (FDR) (63, 64) was used to identify genes that were differentially methylated between the CIMP groups. Significant changes were defined as genes having an FDR-corrected *P* value of <0.05 . Probes with an adjusted *P* value below 0.05 were considered significantly differentially methylated between the two sets of tumors. The hierarchical clustering of the methylation data was performed as above with the top 5% most variant probes across the samples (defined by SD). *K*-means consensus clustering was

performed with the R statistical package. Consensus clustering was performed with K -means clustering ($K_{\max} = 9$) with Euclidean distance and average linkage over 1000 resampling iterations with random restart (as implemented in GenePattern v3.2.3) (67).

For identification of CIMP genes in colon cancer and glioblastoma, analysis was performed as follows. Methylation data for GBM were downloaded from The Cancer Genome Atlas data portal. Methylation data for colon cancer ($n = 24$) were generated from MSKCC primary tumors with the Illumina HumanMethylation27 array. Hierarchical clustering was performed as described above with the breast cancer data using the top 5% most variant probes. Iterations using the top 3 to 20% did not significantly alter the clustering results. The cluster results were confirmed with the methylation β values of the five-gene panel described by Weisenberger *et al.* to identify CIMP⁺ tumors in colorectal cancers (2). The cluster of samples that exhibited hypermethylation of these marker genes was selected as CIMP⁺ and used for further analyses. These corresponded to the cluster with high coordinate hypermethylation derived by hierarchical clustering. The glioblastoma CIMP genes were identified as described (1). The Cancer Genome Atlas Project GBM cancer data sets are publicly available at <http://cancergenome.nih.gov>.

Concepts module mapping and GSEA

Concepts module mapping was performed as follows. The methylation signature identified from our analysis (table S1) was imported into Oncomine (<http://www.oncomine.org>) to search for associations with molecular concepts signatures derived from independent cancer profiling studies. We report statistically significant overlaps of our methylation gene signature with the top-ranking gene expression signatures of clinical outcome using percentile cutoffs (10%). Q value is calculated as previously described (41).

GSEA was performed with GSEA software v2.0.7 (68) and MSigDB database v2.5 (68). We assessed the significance of the gene sets with the following parameters: number of permutations = 1000 and permutation_type = phenotype with an FDR Q -value cutoff of 25%. The most differentially expressed genes from statistically significant gene sets were identified with the “leading edge subset” that consists of genes with the most contribution to the enrichment score of a particular gene set. Enrichment of gene sets downloaded from the literature (as referenced in tables S9 and S10) was analyzed together with the curated gene sets (MSigDB collection c2) or within each other.

Quantitative DNA methylation analysis using mass spectrometry

DNA methylation analysis was performed with the EpiTYPER system (Sequenom). The EpiTYPER assay is a tool for the detection and quantitative analysis of DNA methylation with base-specific cleavage of bisulfite-treated DNA and matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS) (69). For primer sequences, target chromosomal sequence, and EpiTYPER-specific tags, see table S2. SpectroCHiPs were analyzed with a Bruker Biflex III MALDI-TOF mass spectrometer (SpectroREADER, Sequenom). Results were analyzed with the EpiTYPER Analyzer software and manually inspected for spectra quality and peak quantification. CIMP positivity was defined as a mean methylated allelic frequency of >50% or a twofold increase over normal breast tissue and the CIMP[−] state.

Statistical analysis and CIMP repression signature

The 295-sample van't Veer microarray data set (NKI295) was downloaded from the Rosetta InPharmatics Web site (17). Seventy of 102 genes of our methylation signature were represented in the NKI295 and were used to test for prognostic significance. An average expression value was calculated for our “hypermethylated and down-regulated in B-CIMP”

gene set across each sample of NKI295. A two-way classifier was developed by separating the patients into two groups on the basis of the average expression value of our methylation signature: CIMP repression signature up-regulated if the average expression value was >0 and CIMP repression signature down-regulated otherwise. Kaplan-Meier curves comparing survival of patient subgroups were generated with SPSS statistical software.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge Y. Zhao, J. Wongvipat, K. Huberman, I. Dolgalev, S. Thomas, and the MSKCC Genomics Core for technical expertise.

Funding: This work was supported in part by The Cancer Genome Atlas Project (S.B.B.); the Flight Attendant's Medical Research Institute (T.A.C.); the Louis Gerstner Foundation (T.A.C.); the Elsa U. Pardee Foundation (T.A.C.); the AVON Foundation (T.A.C.); the STARR Cancer Consortium (T.A.C. and L.T.V.); the Susan G. Komen Foundation (N.A.); the Mary K. Ash Foundation (N.A.); the Howard Hughes Medical Institute (J.M.); a National Cancer Institute training grant (T32CA009685, L.G.T.M.); and the Metastasis Research Center (T.A.C.).

REFERENCES AND NOTES

1. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP, Verhaak RG, Hoadley KA, Hayes DN, Perou CM, Schmidt HK, Ding L, Wilson RK, Van Den Berg D, Shen H, Bengtsson H, Neuvial P, Cope LM, Buckley J, Herman JG, Baylin SB, Laird PW, Aldape K. Cancer Genome Atlas Research Network, Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*. 2010; 17:510–522. [PubMed: 20399149]
2. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, Koh H, Simms L, Barker M, Leggett B, Levine J, Kim M, French AJ, Thibodeau SN, Jass J, Haile R, Laird PW. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with *BRAF* mutation in colorectal cancer. *Nat Genet*. 2006; 38:787–793. [PubMed: 16804544]
3. Perou CM, Sørli T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. *Nature*. 2000; 406:747–752. [PubMed: 10963602]
4. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ. Cancer statistics, 2008. *CA Cancer J Clin*. 2008; 58:71–96. [PubMed: 18287387]
5. Saal LH, Gruvberger-Saal SK, Persson C, Lövgren K, Juppunen M, Staaf J, Jönsson G, Pires MM, Maurer M, Holm K, Koujak S, Subramaniam S, Vallon-Christersson J, Olsson H, Su T, Memeo L, Ludwig T, Ethier SP, Krogh M, Szabolcs M, Murty VV, Isola J, Hibshoosh H, Parsons R, Borg A. Recurrent gross mutations of the *PTEN* tumor suppressor gene in breast cancers with deficient DSB repair. *Nat Genet*. 2008; 40:102–107. [PubMed: 18066063]
6. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006; 295:2492–2502. [PubMed: 16757721]
7. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, Glas AM, Perou CM, Van't Veer LJ. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res*. 2005; 65:9155–9158. [PubMed: 16230372]
8. Muss HB, Woolf S, Berry D, Cirincione C, Weiss RB, Budman D, Wood WC, Henderson IC, Hudis C, Winer E, Cohen H, Wheeler J, Norton L. Cancer and Leukemia Group B, Adjuvant chemotherapy in older and younger women with lymph node-positive breast cancer. *JAMA*. 2005; 293:1073–1081. [PubMed: 15741529]

9. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, Gianni L, Baselga J, Bell R, Jackisch C, Cameron D, Dowsett M, Barrios CH, Steger G, Huang CS, Andersson M, Inbar M, Lichinitser M, Láng I, Nitz U, Iwata H, Thomssen C, Lohrisch C, Suter TM, Ruschoff J, Suto T, Gatrex V, Ward C, Strahle C, McFadden E, Dolci MS, Gelber RD. Herceptin Adjuvant (HERA) Trial Study Team, Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med*. 2005; 353:1659–1672. [PubMed: 16236737]
10. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003; 100:8418–8423. [PubMed: 12829800]
11. Olopade OI, Grushko T. Gene-expression profiles in hereditary breast cancer. *N Engl J Med*. 2001; 344:2028–2029. [PubMed: 11430337]
12. Kalinsky K, Jacks LM, Heguy A, Patil S, Drobnjak M, Bhanot UK, Hedvat CV, Traina TA, Solit D, Gerald W, Moynahan ME. PIK3CA mutation associates with improved outcome in breast cancer. *Clin Cancer Res*. 2009; 15:5049–5059. [PubMed: 19671852]
13. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B. The genomic landscapes of human breast and colorectal cancers. *Science*. 2007; 318:1108–1113. [PubMed: 17932254]
14. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. The consensus coding sequences of human breast and colorectal cancers. *Science*. 2006; 314:268–274. [PubMed: 16959974]
15. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, Zhang Q, Chen L, Lin L, Wendl MC, McMichael JF, Magrini VJ, Cook L, McGrath SD, Vickery TL, Appelbaum E, Deschryver K, Davies S, Guintoli T, Lin L, Crowder R, Tao Y, Snider JE, Smith SM, Dukes AF, Sanderson GE, Pohl CS, Delehaunty KD, Fronick CC, Pape KA, Reed JS, Robinson JS, Hodges JS, Schierding W, Dees ND, Shen D, Locke DP, Wiechert ME, Eldred JM, Peck JB, Oberkfell BJ, Lolofie JT, Du F, Hawkins AE, O’Laughlin MD, Bernard KE, Cunningham M, Elliott G, Mason MD, Thompson DM Jr, Ivanovich JL, Goodfellow PJ, Perou CM, Weinstock GM, Aft R, Watson M, Ley TJ, Wilson RK, Mardis ER. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*. 2010; 464:999–1005. [PubMed: 20393555]
16. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004; 351:2817–2826. [PubMed: 15591335]
17. van’t Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002; 415:530–536. [PubMed: 11823860]
18. Minn AJ, Gupta GP, Padua D, Bos P, Nguyen DX, Nuyten D, Kreike B, Zhang Y, Wang Y, Ishwaran H, Foekens JA, van de Vijver M, Massagué J. Lung metastasis genes couple breast tumor size and metastatic spread. *Proc Natl Acad Sci U S A*. 2007; 104:6740–6745. [PubMed: 17420468]
19. Jones PA, Baylin SB. The epigenomics of cancer. *Cell*. 2007; 128:683–692. [PubMed: 17320506]
20. Costello JF, Frühwald MC, Smiraglia DJ, Rush LJ, Robertson GP, Gao X, Wright FA, Feramisco JD, Peltomäki P, Lang JC, Schuller DE, Yu L, Bloomfield CD, Caligiuri MA, Yates A, Nishikawa R, Su Huang H, Petrelli NJ, Zhang X, O’Dorisio MS, Held WA, Cavenee WK, Plass C. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet*. 2000; 24:132–138. [PubMed: 10655057]

21. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*. 1983; 301:89–92. [PubMed: 6185846]
22. Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med*. 2003; 349:2042–2054. [PubMed: 14627790]
23. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A*. 1999; 96:8681–8686. [PubMed: 10411935]
24. Bae YK, Brown A, Garrett E, Bornman D, Fackler MJ, Sukumar S, Herman JG, Gabrielson E. Hypermethylation in histologically distinct classes of breast cancer. *Clin Cancer Res*. 2004; 10:5998–6005. [PubMed: 15447983]
25. Lee JS, Fackler MJ, Lee JH, Choi C, Park MH, Yoon JH, Zhang Z, Sukumar S. Basal-like breast cancer displays distinct patterns of promoter methylation. *Cancer Biol Ther*. 2010; 9:1017–1024. [PubMed: 20505321]
26. Novak P, Jensen T, Oshiro MM, Watts GS, Kim CJ, Futscher BW. Agglomerative epigenetic aberrations are a common event in human breast cancer. *Cancer Res*. 2008; 68:8616–8625. [PubMed: 18922938]
27. Chan TA, Glockner S, Yi JM, Chen W, Van Neste L, Cope L, Herman JG, Velculescu V, Schuebel KE, Ahuja N, Baylin SB. Convergence of mutation and epigenetic alterations identifies common genes in cancer that predict for poor prognosis. *PLoS Med*. 2008; 5:e114. [PubMed: 18507500]
28. Van der Auwera I, Bovie C, Svensson C, Trinh XB, Limame R, van Dam P, van Laere SJ, van Marck EA, Dirix LY, Vermeulen PB. Quantitative methylation profiling in tumor and matched morphologically normal tissues from breast cancer patients. *BMC Cancer*. 2010; 10:97. [PubMed: 20226036]
29. Gaudet MM, Campan M, Figueroa JD, Yang XR, Lissowska J, Peplonska B, Brinton LA, Rimm DL, Laird PW, Garcia-Closas M, Sherman ME. DNA hypermethylation of *ESR1* and *PGR* in breast cancer: Pathologic and epidemiologic associations. *Cancer Epidemiol Biomarkers Prev*. 2009; 18:3036–3043. [PubMed: 19861523]
30. Campan M, Weisenberger DJ, Laird PW. DNA methylation profiles of female steroid hormone-driven human malignancies. *Curr Top Microbiol Immunol*. 2006; 310:141–178. [PubMed: 16909910]
31. Ehrlich M, Woods CB, Yu MC, Dubeau L, Yang F, Campan M, Weisenberger DJ, Long T, Youn B, Fiala ES, Laird PW. Quantitative analysis of associations between DNA hypermethylation, hypomethylation, and DNMT RNA levels in ovarian tumors. *Oncogene*. 2006; 25:2636–2645. [PubMed: 16532039]
32. Estécio MR, Yan PS, Ibrahim AE, Tellez CS, Shen L, Huang TH, Issa JP. High-throughput methylation profiling by MCA coupled to CpG island microarray. *Genome Res*. 2007; 17:1529–1536. [PubMed: 17785535]
33. Liu Y, Hayes DN, Nobel A, Marron JS. Statistical significance of clustering for high-dimension, low-sample size data. *J Am Stat Assoc*. 2008; 103:1281–1293.
34. Ehrlich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, van den Boom D. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A*. 2005; 102:15785–15790. [PubMed: 16243968]
35. Choi YL, Bocanegra M, Kwon MJ, Shin YK, Nam SJ, Yang JH, Kao J, Godwin AK, Pollack JR. LYN is a mediator of epithelial-mesenchymal transition and a target of dasatinib in breast cancer. *Cancer Res*. 2010; 70:2296–2306. [PubMed: 20215510]
36. Tapper W, Hammond V, Gerty S, Ennis S, Simmonds P, Collins A. Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) Steering Group, D. Eccles. The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. *Breast Cancer Res*. 2008; 10:R108. [PubMed: 19094228]
37. Milovanovic T, Planutis K, Nguyen A, Marsh JL, Lin F, Hope C, Holcombe RF. Expression of Wnt genes and frizzled 1 and 2 receptors in normal breast epithelium and infiltrating breast carcinoma. *Int J Oncol*. 2004; 25:1337–1342. [PubMed: 15492823]

38. Papageorgis P, Lambert AW, Ozturk S, Gao F, Pan H, Manne U, Alekseyev YO, Thiagalingam A, Abdolmaleky HM, Lenburg M, Thiagalingam S. Smad signaling is required to maintain epigenetic silencing during breast cancer progression. *Cancer Res.* 2010; 70:968–978. [PubMed: 20086175]
39. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB, Gerald WL, Massagué J. Genes that mediate breast cancer metastasis to lung. *Nature.* 2005; 436:518–524. [PubMed: 16049480]
40. Nguyen DX, Massagué J. Genetic determinants of cancer metastasis. *Nat Rev Genet.* 2007; 8:341–352. [PubMed: 17440531]
41. Rhodes DR, Kalyana-Sundaram S, Tomlins SA, Mahavisno V, Kasper N, Varambally R, Barrette TR, Ghosh D, Varambally S, Chinnaiyan AM. Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia.* 2007; 9:443–454. [PubMed: 17534450]
42. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 2008; 4:e1000242. [PubMed: 18974828]
43. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005; 102:15545–15550. [PubMed: 16199517]
44. Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins DN, Herman JG, Baylin SB. A stem cell–like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet.* 2007; 39:237–242. [PubMed: 17211412]
45. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nat Genet.* 2007; 39:232–236. [PubMed: 17200670]
46. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolzheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell.* 2006; 125:301–313. [PubMed: 16630818]
47. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* 2006; 20:1123–1136. [PubMed: 16618801]
48. Issa JP. Colon cancer: It's CIN or CIMP. *Clin Cancer Res.* 2008; 14:5939–5940. [PubMed: 18829469]
49. Birgisdottir V, Stefansson OA, Bodvarsdottir SK, Hilmarisdottir H, Jonasson JG, Eyfjord JE. Epigenetic silencing and deletion of the *BRCA1* gene in sporadic breast cancer. *Breast Cancer Res.* 2006; 8:R38. [PubMed: 16846527]
50. Lo PK, Lee JS, Liang X, Han L, Mori T, Fackler MJ, Sadik H, Argani P, Pandita TK, Sukumar S. Epigenetic inactivation of the potential tumor suppressor gene *FOXF1* in breast cancer. *Cancer Res.* 2010; 70:6047–6058. [PubMed: 20587515]
51. Christensen BC, Kelsey KT, Zheng S, Houseman EA, Marsit CJ, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Kushi LH, Kwan ML, Wiencke JK. Breast cancer DNA methylation profiles are associated with tumor size and alcohol and folate intake. *PLoS Genet.* 2010; 6:e1001043. [PubMed: 20686660]
52. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet.* 2002; 3:415–428. [PubMed: 12042769]
53. Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, Glöckner S, Piantadosi S, Gabrielson E, Pridham G, Pelosky K, Belinsky SA, Yang SC, Baylin SB, Herman JG. DNA methylation markers and early recurrence in stage I lung cancer. *N Engl J Med.* 2008; 358:1118–1128. [PubMed: 18337602]

54. Martens JW, Margossian AL, Schmitt M, Foekens J, Harbeck N. DNA methylation as a biomarker in breast cancer. *Future Oncol.* 2009; 5:1245–1256. [PubMed: 19852739]
55. Nimmrich I, Sieuwerts AM, Meijer-van Gelder ME, Schwöpe I, Bolt-de Vries J, Harbeck N, Koenig T, Hartmann O, Kluth A, Dietrich D, Magdolen V, Portengen H, Look MP, Klijn JG, Lesche R, Schmitt M, Maier S, Foekens JA, Martens JW. DNA hypermethylation of *PITX2* is a marker of poor prognosis in untreated lymph node-negative hormone receptor-positive breast cancer patients. *Breast Cancer Res Treat.* 2008; 111:429–437. [PubMed: 17965955]
56. Dominguez G, Silva J, Garcia JM, Silva JM, Rodriguez R, Muñoz C, Chacón I, Sanchez R, Carballido J, Colás A, España P, Bonilla F. Prevalence of aberrant methylation of *p14ARF* over *p16INK4a* in some human primary tumors. *Mutat Res.* 2003; 530:9–17. [PubMed: 14563526]
57. Wendt MK, Cooper AN, Dwinell MB. Epigenetic silencing of CXCL12 increases the metastatic potential of mammary carcinoma cells. *Oncogene.* 2008; 27:1461–1471. [PubMed: 17724466]
58. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science.* 1987; 235:177–182. [PubMed: 3798106]
59. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med.* 2001; 344:783–792. [PubMed: 11248153]
60. Shukla V, Coumoul X, Lahusen T, Wang RH, Xu X, Vassilopoulos A, Xiao C, Lee MH, Man YG, Ouchi M, Ouchi T, Deng CX. BRCA1 affects global DNA methylation through regulation of *DNMT1*. *Cell Res.* 2010; 20:1201–1215. [PubMed: 20820192]
61. Esteller M, Toyota M, Sanchez-Cespedes M, Capella G, Peinado MA, Watkins DN, Issa JP, Sidransky D, Baylin SB, Herman JG. Inactivation of the DNA repair gene *O⁶-methylguanine-DNA methyltransferase* by promoter hypermethylation is associated with G to A mutations in *K-ras* in colorectal tumorigenesis. *Cancer Res.* 2000; 60:2368–2371. [PubMed: 10811111]
62. Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, Schuebel KE, Cui H, Feinberg AP, Lengauer C, Kinzler KW, Baylin SB, Vogelstein B. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature.* 2002; 416:552–556. [PubMed: 11932749]
63. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 2003; 100:9440–9445. [PubMed: 12883005]
64. Storey JD. A direct approach to false discovery rates. *J R Statist Soc B.* 2002; 64:479–498.
65. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4:44–57. [PubMed: 19131956]
66. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003; 4:P3. [PubMed: 12734009]
67. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet.* 2006; 38:500–501. [PubMed: 16642009]
68. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003; 34:267–273. [PubMed: 12808457]
69. Docherty SJ, Davis OS, Haworth CM, Plomin R, Mill J. Bisulfite-based epityping on pooled genomic DNA provides an accurate estimate of average group DNA methylation. *Epigenetics Chromatin.* 2009; 2:3. [PubMed: 19284538]

Fig. 1.

Identification of a CIMP by clustering of breast cancer samples. Unsupervised clustering was performed with the Infinium DNA methylation probes whose β values varied most across the breast tumor samples (top 5% most variant probes). (A) *K*-means consensus clustering ($K = 2$). Tumors ($n = 39$) are listed in the same order along the x and y axes. Clusters 1 and 2 are noted in purple and green, respectively, and B-CIMP status is noted on the black and white bars. Consensus index values range from 0 to 1, with 0 being highly dissimilar and 1 being highly similar. (B) 2D hierarchical clustering of the tumors using the same methylation probes as in (A). Each row represents a tumor and each column represents a probe. The level of DNA methylation (normalized and transformed β value) is represented with a color scale (red, methylated; blue, nonmethylated). Data for normal breast ($n = 3$) and DKO (a cell line with global loss of methylation) are shown but did not contribute to the clustering analysis (62). HR, HER2, and distant recurrence status is shown along the left side of the heat map. Failed, developed distant metastases. Horizontal bars along the bottom of plot denote genes that have decreased or increased methylation in B-CIMP.

Fig. 2. Characterization of B-CIMP and clinical co-variables. (A) Selected genes with differential methylation between CIMP groups in breast tumors. Average β values for the genes are shown on the y axis. Data are presented for CIMP⁺ (red) and CIMP⁻ (blue) tumors. Whisker box plots summarizing each data set are shown (mean $\pm$ SD). The boxes delineate the 25th to 75th percentile range. (B) B-CIMP⁺ tumors are highly associated with HR positivity but not with HER2 status (χ^2 test).

Fig. 3.

B-CIMP and metastatic risk. (A) Relative methylation (normalized and transformed β value) of the genes analyzed in Fig. 1 in CIMP⁺ versus CIMP⁻ tumors. Whisker box plots summarizing each data set are shown (mean $\pm$ SD). The boxes delineate the 25th to 75th percentile range. The P value indicates significance determined using ANOVA. (B) Relative methylation of the genes in ER/PR⁺ versus ER/PR⁻ tumors. Whisker box plots summarizing each data set are shown (mean $\pm$ SD). Significance determined using ANOVA. (C) Relative methylation of the genes in hormone⁺, CIMP⁺ versus hormone⁺, CIMP⁻ tumors. Whisker box plots summarizing each data set are shown (mean $\pm$ SD). Significance determined using ANOVA. (D) Relative methylation of the genes in CIMP⁺ versus CIMP⁻ tumors (all probes). Whisker box plots summarizing each data set are shown (mean $\pm$ SD). Significance determined using ANOVA. (E) Kaplan-Meier curve for distant metastasis-free survival for B-CIMP⁺ and B-CIMP⁻ subtypes. P value calculated by log-rank analysis. Data from discovery set tumors (Fig. 1). (F) Validation of B-CIMP and its impact on metastatic risk in an independent set of breast tumors ($n = 132$). EpiTYPER assays were developed for three of the most predictive genes for B-CIMP as indicated in the panel. Samples were analyzed and categorized as CIMP⁺ if at least two of three genes were methylated, as shown in the heat map. Red, methylated; blue, unmethylated; purple, B-CIMP⁺. (G and H) Kaplan-Meier curves for (G) distant metastasis-free survival and (H) overall survival by B-CIMP status. P value calculated by log-rank analysis. (I) B-CIMP predicted metastatic risk in ER/PR⁺

breast cancers. Kaplan-Meier curve for distant metastasis-free survival for B-CIMP⁺ and B-CIMP⁻ subtypes. *P* value calculated by log-rank analysis.

Fig. 4.

Methylation landscape of breast cancer. **(A)** Characteristics of the hypermethylated sites in the B-CIMP. The plot on the left [distance to transcriptional start site (tss)] shows the proportion of probes that are within the indicated distances from the tss. The plot on the right (composition) shows the proportion of hypermethylated probes that are located in CpG islands, in shores, or in neither CpG islands nor shores. **(B)** Starburst plot for comparison of DNA methylation and gene expression. $\log_{10}(\text{FDR-corrected } P \text{ value})$ is plotted for DNA methylation (x axis) and gene expression (y axis) for each gene. Along the y axis, data for fold expression <0 are $\log_{10}(\text{FDR-corrected } P \text{ value})$ and data for fold expression ≥ 0 are $-\log_{10}(\text{FDR-corrected } P \text{ value})$. Along the x axis, data for β value <0 are $\log_{10}(\text{FDR-corrected } P \text{ value})$ and data for β value ≥ 0 are $-\log_{10}(\text{FDR-corrected } P \text{ value})$. The black line indicates the FDR-adjusted P value threshold of 0.05. Data points other than red indicate genes that are significantly up-regulated (green) or down-regulated (blue and purple) and also significantly hypo- or hypermethylated in B-CIMP tumors. Points in blue indicate genes that are significantly down-regulated and hypermethylated in B-CIMP⁺ tumors versus B-CIMP⁻ tumors.

Fig. 5. B-CIMP and the metastasis transcriptome. **(A)** Differentially expressed B-CIMP genes in breast cancer metastasis transcriptomes. Concepts mapping of B-CIMP-repressed genes across multiple data sets associated with metastasis. Each row shows individual gene sets from which a breast prognostic expression signature has been described. The top 10% of the most overexpressed genes from these gene sets were used for the concept mapping. Genes that are significantly hypermethylated and down-regulated in B-CIMP⁺ tumors ($n = 102$) correspond to genes whose overexpression is predictive of metastasis (left panel). Red, matching gene between B-CIMP gene and gene predictive of metastatic behavior. Q value calculated as in Materials and Methods (right panel). **(B)** Kaplan-Meier survival curve showing that the CIMP repression signature (hypermethylated and down-regulated in B-CIMP tumors) predicts survival in the van't Veer cohort (17). P value calculated by log-rank.

Fig. 6.

Consensus between the PRC2 and methylome landscapes in B-CIMP. Significance and locations of methylation enrichment in B-CIMP⁺ tumors are plotted across the genome and indicated by the blue bars. Significance and locations of PRC2 enrichment are shown by the red bars. The horizontal axis indicates level of significance (FDR-corrected, ANOVA).

Fig. 7.

Consensus target genes of the CIMP across multiple human cancers. **(A)** Venn diagram showing common gene targets between the PRC2 targets described in (47) and CIMP in the three indicated cancers. Numbers in parentheses indicate number of genes in common between PRC2 target genes and CIMP targets in each cancer type. The table next to the diagram shows the level of significance between these overlapping gene lists (P value, hypergeometric distribution). The numbers in the Venn diagram show the number of CIMP/PRC2 common targets that are shared between the cancer types. **(B)** Same as in (A), except PcG targets are from the Suz12 targets described in (46). **(C)** The left diagram shows the 33 most significant CIMP targets common to all three malignancies. Significance [$-\log_{10}(\text{FDR-corrected } P \text{ value})$] was plotted for DNA methylation along the y axis, and genomic location or chromosome number was plotted along the x axis. Genes are shown in a separate color for each tumor type according to the legend. The right diagram shows a Kaplan-Meier curve depicting the survival of patients with GBM tumors in which the 33-gene CIMP signature was present versus tumors in which the signature was absent. P value calculated by log-rank. TCGA, The Cancer Genome Atlas.