Published in final edited form as:
Sci Transl Med. 2013 January 23; 5(169): 169ra10. doi:10.1126/scitranslmed.3005211.

DNA methylation alterations exhibit intra-individual stability and inter-individual heterogeneity in prostate cancer metastases

Martin J Aryee[1,4], **Wennuan Liu**[3], **Julia C Engelmann**[1,9], **Philipp Nuhn**[1], **Meltem Gurel**[1], **Michael C Haffner**[1], **David Esopi**[1], **Rafael A Irizarry**[5], **Robert H Getzenberg**[1,6,11], **William G Nelson**[1,6], **Jun Luo**[1,6], **Jianfeng Xu**[3], **William B Isaacs**[1,6], **G Steven Bova**[7,10], and **Srinivasan Yegnasubramanian**[1,8,*]

[1]Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, USA

[2]Oncology Department, Division of Biostatistics, Johns Hopkins University, Baltimore, MD, USA

[3]Center for Cancer Genomics, Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem, NC, USA

[5]Biostatistics Department, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

[6]Brady Urological Institute, Johns Hopkins University, Baltimore, MD, USA

[7]Pathology Department, Johns Hopkins University, Baltimore, MD, USA

[8]Johns Hopkins Physical Sciences in Oncology Center, Johns Hopkins University, Baltimore, MD, USA

Abstract

Human cancers nearly ubiquitously harbor epigenetic alterations. While such alterations in epigenetic marks, including DNA methylation, are potentially heritable, they can also be dynamically altered. Given this potential for plasticity, the degree to which epigenetic changes can be subject to selection and act as drivers of neoplasia has been questioned. Here, we carried out genome-scale analyses of DNA methylation alterations in lethal metastatic prostate cancer and created DNA methylation “cityscape” plots to visualize these complex data. We show that somatic

*Correspondence: S.Y., syegnasu@jhmi.edu.
[4]Current affiliation: Department of Pathology, Massachusetts General Hospital, Charlestown, MA
[9]Current affiliation: University of Regensburg, Regensburg, Germany
[10]Current affiliation: Institute of Biomedical Technology, PELICAN-Personalized Cancer Medicine Group, University of Tampere, Biokatu 8, FI-33014 Tampere, Finland
[11]Current affiliation: GTx Inc., Memphis, TN

Author contributions: S.Y., G.S.B., R.H.G., J.L., W.G.N., J.X., and W.B.I. conceptualized the study. G.S.B. directed the PELICAN autopsy study of lethal prostate cancer from which prostate cancer tissue metastasis tissues were obtained. W.B.I assisted in collection of autopsy study tissues. R.H.G. directed collection of normal human tissue specimens from organ donors. S.Y., W.L., M.C.H., P.N., J.L. and D.E. performed experiments. M.J.A., S.Y., M.G., J.E. and R.A.I. designed computational methods and performed data analysis. M.J.A. and S.Y. wrote the manuscript.

Competing interests: S.Y., M.C.H., D.E., W.G.N. and W.B.I. and Johns Hopkins University (JHU) have provisional and/or fully-executed patents relating to DNA methylation biomarkers in prostate cancer (U.S. Patent Number 5,552,277, “Genetic diagnosis of prostate cancer”; Patent Pending, “Epigenetic Tests for Prostate Cancer”; Patent Pending, “DNA methylation biomarkers for prostate cancer”). S.Y. and W.G.N., along with JHU, hold a patent (U.S. Patent Number 7,906,288, “COMPARE-MS: A Novel Technique for Rapid, Sensitive, And Specific Detection of DNA Methylation”) for use of the MBD2-MBD polypeptide for detection of methylated DNA, and this reagent has been made available to the research community via a non-exclusive license with Clontech Inc., which provides royalties to JHU, S.Y. and W.G.N from sale of kits containing this reagent. The authors are pursuing intellectual property protection for the new prostate cancer biomarkers described here.

Data and materials: Raw and normalized data are available from the Gene Expression Omnibus (GEO) with accession number GSE38242.

DNA methylation alterations, despite showing marked inter-individual heterogeneity among men with lethal metastatic prostate cancer, were maintained across all metastases within the same individual. The overall extent of maintenance in DNA methylation changes was comparable to that of genetic copy number alterations. Regions that were frequently hypermethylated across individuals were markedly enriched for cancer and development/differentiation related genes. Additionally, regions exhibiting high consistency of hypermethylation across metastases within individuals, even if variably hypermethylated across individuals, showed enrichment of cancer-related genes. Interestingly, whereas some regions showed intra-individual metastatic tumor heterogeneity in promoter methylation, such methylation alterations were generally not correlated with gene expression. This was despite a general tendency for promoter methylation patterns to be strongly correlated with gene expression, particularly at regions that were variably methylated across individuals. These findings suggest that DNA methylation alterations have the potential for producing selectable driver events in carcinogenesis and disease progression and highlight the possibility of targeting such epigenome alterations for development of longitudinal markers and therapeutic strategies.

INTRODUCTION

Cancer is thought to arise from a series of somatic genome and epigenome defects that allow the cell to evade the rules that control the growth and organization of normal cells (1, 2). In order for genetic and epigenetic somatic genome alterations to drive cancer initiation and progression, the cancer cell would need to maintain those changes in a heritable way throughout disease progression for as long as such changes confer a selective advantage. Genetic alterations are maintained by semi-conservative DNA replication and have been clearly implicated in carcinogenesis and disease progression (3). However, epigenetic processes present a fundamental paradox in this regard: they are by definition potentially heritable across cell divisions and are stable over time (4, 5), but they can also be plastic (5, 6). For instance, recent reports have suggested that the epigenetic process of DNA methylation can be dynamic and reversible in both replication-dependent (e.g. during differentiation and development (6)) and -independent (e.g. cyclical methylation patterns during transcription (7, 8)) processes. Additionally, DNA methylation marks can occur at both copies of a given locus, or occur at only one copy, resulting in allele-specific methylation (ASM) (9–11). Unfortunately, most previous reports on DNA methylation in human cancers have only examined total methylation (TM) at an allele-agnostic level, and little is known about the maintenance of ASM in human neoplasia.

Consequently, it is currently unclear which DNA methylation and other epigenetic alterations can be maintained stably as driver genome alterations fueling cancer initiation and progression. A lack of such evidence has dampened enthusiasm for using DNA methylation alterations, which can be more frequent than genetic alterations (12), as targets for biomarker development and therapeutic intervention. Here, we show that, although there is marked heterogeneity in DNA methylation profiles in men with lethal metastatic prostate cancer, each individual's distinct DNA methylation signature is tightly maintained in disseminated metastases.

RESULTS

Performance of the MBD-SNP approach

We developed and applied a new technology and associated computational methods enabling simultaneous genome-scale analysis of genetic (copy number) and epigenetic (total and allele-specific DNA methylation) alterations. This method, called MBD-SNP (see Figure 1A for overview), features affinity enrichment of methylated genomic DNA

fragments (13) using the methyl-binding domain polypeptide from the MBD2 protein (MBD2-MBD), which was previously shown to preferentially bind methylated DNA with >100 fold selectivity compared to unmethylated DNA (14, 15). The resulting library of methylated DNA fragments and an unenriched total input fraction from the same specimen were then processed and hybridized to Affymetrix SNP 6.0 high-density oligonucleotide microarrays. Comparison of the enriched methylated fraction with the total input using new statistical approaches allowed parallel genome-scale assessment of TM, ASM, and copy number in a rapid and cost-effective manner.

Using a series of control specimens (Figure 1B, Supplementary Materials and Methods), we determined that the MBD-SNP technology allowed accurate point estimates of TM (Figure 1C, area under the receiver operator characteristic curve (AUC) = 0.89) and ASM (Figure 1C, AUC = 0.95) for regions of the genome with ≥ 2.5% CpG density. Therefore, this platform allowed accurate interrogation of total and allele-specific DNA methylation patterns at 51,501 (TM) and 24,498 (ASM) regions (Figure S1), including 7,323 (TM) and 4,295 (ASM) gene promoter regions, 5,766 (TM) and 4,277 (ASM) CpG islands, and 15,210 (TM) and 9,969 (ASM) CpG island shores (Table S1).

Somatic alterations in total and allele-specific DNA methylation patterns in lethal metastatic prostate cancer

Previous studies have used analysis of genetic alterations to examine the clonal evolution of cancer metastases (16–18). Using such a study design, featuring analysis of multiple metastatic deposits as well as matched normal tissues from each subject from a lethal metastatic prostate cancer rapid autopsy cohort, Liu et al. showed that prostate cancer metastases within an individual have monoclonal origins and display subsequent clonal evolution (16). We examined the same specimens from this rapid autopsy cohort (a total of 71 specimens, including 3 to 6 metastases and 1 to 2 normal tissues from each of 13 subjects) to understand whether DNA methylation alterations also showed clonal maintenance and evolution across metastatic dissemination. Additionally, we examined 24 normal prostate tissues from organ donors without evidence of prostate disease as reference samples.

Applying the MBD-SNP technology and our new computational approaches to the study samples, we computed normalized TM and ASM scores at all informative regions. We confirmed that the approach allowed highly accurate point estimates of TM across the study samples by validating the data in a subset of the study samples using both real-time methylation specific PCR (RT-MSP) assays (19) and the bisulfite-based Illumina HumanMethylation 450k microarray platform (Figure 1D,E; Supplementary Materials and Methods). To assess the ability to identify regions showing ASM across the study samples, we examined ASM signals at known imprinted loci including multiple known allele-specific differentially methylated regions (DMR) at the *IGF2/H19* imprinting control regions. The MBD-SNP derived ASM scores were consistent with imprinting of these regions in all normal samples analyzed (Figure S2). Interestingly, some cancer specimens showed evidence for somatic loss of ASM of these regions, consistent with cancer-specific *IGF2/H19* loss of imprinting (LOI), a known hallmark of many cancers including prostate cancer (20, 21). We observed 1,873 regions that showed gain or loss of ASM in at least one tumor compared to normal tissues (Table S2; Figure S3), with 667 of these showing alterations in at least two subjects.

Focusing on total methylation, we identified a total of 3943 regions that showed no evidence of methylation in any of the normal prostate tissues, but were hypermethylated in at least one prostate cancer specimen (Table S3). The frequency of these hypermethylation events ranged widely, from affecting just a single subject to being hypermethylated in all subjects

analyzed (Figures S4, S5); 1329 regions showed hypermethylation in at least 25% of the metastatic prostate cancer tissues. Among these were several gene promoters known to be frequently hypermethylated in primary and metastatic prostate cancer (19), such as those of the *GSTP1* (100% of tumors) and *APC* (89% of tumors) genes (Table S3). Regions showing hypermethylation in any tumor were highly enriched within gene promoter regions (Figure S6). On average, each subject showed hypermethylation at 611 promoter probes (range: 372 – 1039) representing 498 gene promoter regions (~1500 when extrapolated to the whole genome) (Figures S7, S8). Some subjects showed alteration of >700 gene promoter regions (~2100 when extrapolated to the whole genome), consistent with a CpG island hypermethylator phenotype (CIMP) as has been suggested for colon and other cancer types (22–24). We investigated differences in methylation patterns associated with promoter proximity and found that promoter-associated CGIs are significantly more likely to be hypermethylated than non-promoter CGIs ($p < 1 \times 10^{-10}$). The genes associated with hypermethylated gene promoters were highly enriched for differentiation and development associated Gene Ontology terms (25), as well as the Memorial Sloan Kettering Prostate Cancer Pathways gene set (26) and multiple gene sets from the NCI Cancer Gene Index (27) (Figure S6, Supplementary Materials and Methods). Thus, DNA hypermethylation events may be involved in reprogramming developmental and differentiation states and in activating carcinogenic pathways. In contrast, hypomethylation alterations, defined as those regions that were methylated in all normal prostate tissues but undermethylated in at least one of the metastatic cancer specimens, were more numerous than hypermethylation alterations (12,799 hypomethylated regions occurring in at least 25% of the cancer specimens), but did not show enrichment of promoter regions, any relevant gene ontology terms, or of cancer-related gene sets (Figure S6). Additionally, the hypomethylated regions tended to have lower CpG content than the hypermethylated regions (Figure S9). These results are consistent with previous reports showing global, widespread losses of DNA methylation, particularly at regions with lower CpG density, accompanied by focal gains of DNA hypermethylation at CpG-rich promoter regions in cancer cells (28–33).

Clonal maintenance of DNA methylation alterations across metastatic dissemination

Having established genome-scale measurements of total and allele-specific methylation patterns in each specimen, we examined the degree to which these patterns were maintained across anatomically distinct metastases within each individual. Interestingly, total methylation patterns in metastases from any given individual showed very high pairwise correlations, with much lower pairwise correlations between metastases from different individuals (Figure 2A). In contrast, the normal specimens showed high correlations between individuals. A similar but less pronounced pattern was observed for allele-specific methylation patterns (Figure 2B). Unsupervised hierarchical clustering using Euclidean distance confirmed that there was relatively little tumor heterogeneity in metastases within subjects compared to the significant tumor heterogeneity across subjects, resulting in clustering of tumors by subject, even after rigorously controlling for copy number effects (Figure 2C). Taken together, these data suggest that the tumor/metastasis-initiating clone or sub-clone in each individual has a unique DNA methylation signature that is then closely maintained across metastatic dissemination.

Since genetic alterations in copy number are highly maintained across prostate cancer metastatic dissemination (16), we compared the extent of maintenance of epigenetic alterations in DNA methylation to that of copy number alterations. In order to facilitate comparisons between these genetic and epigenetic datasets, we fit probe-level ANOVA models to estimate the degree of maintenance of each type of somatic alteration normalized to the total variability of that alteration (represented by R^2 from the model). This was done for those probes showing low variability in the normal prostate tissues, but a high degree of

variability across metastases. The resulting R^2 measures have values between 0 and 1, with values near zero indicating high variability across different metastases from each subject and values near 1 indicating nearly perfect consistency of methylation levels across all metastases from each subject. These analyses confirmed that copy number alterations showed a high degree of clonal maintenance. Interestingly, total DNA methylation alterations showed a degree of clonal maintenance that was comparable to that of copy number alterations (Figures 3A, S10, S11, S12, S13, Supplementary Materials and Methods). Interestingly, ASM alterations, showed markedly less clonal maintenance, although approximately 17% of regions exhibited a level of clonal maintenance in ASM comparable to that of copy number alterations.

The observation that copy number and TM alterations were maintained to a similar extent across metastases from each subject suggested that these genetic and epigenetic changes may have developed through parallel clonal evolutionary processes. For instance, subject 21 showed a near-perfect co-evolution of copy number and DNA methylation patterns (Figure 3B). This similarity between the branching patterns from hierarchical clustering dendrograms generated from copy number data and that generated from the DNA methylation data was significant across all subjects (Figure 3C; $p < 0.001$; Supplementary Materials and Methods). Taken together, these analyses indicate that DNA methylation patterns can be as robustly maintained across metastatic dissemination as genetic copy number alterations.

Given the differences in the functional annotations of hyper- and hypomethylation events in the prostate cancer specimens, we examined whether there were differences in the tendency to maintain hypermethylation versus hypomethylation alterations during metastatic dissemination. Hypermethylation alterations showed a higher degree of maintenance (R^2) than hypomethylation alterations (Figure 3D). This difference was evident even after controlling for CpG density (Figure S14). Taken together with the enrichment of differentiation/development and cancer-related gene sets in hypermethylated gene promoters, the high degree of maintenance of hypermethylation changes suggests that these events are enriched for driver alterations.

Association of gene expression patterns with DNA methylation alterations

We next explored associations between DNA methylation alterations and gene expression patterns in order to understand the functional consequences of somatic DNA methylation alterations in the lethal metastatic prostate cancers. We measured genome-wide gene expression patterns for 18 metastases from 5 autopsy cohort subjects and 21 organ donor benign prostate samples, selected as an arbitrary subset of our overall study samples for which high quality RNA was available, using the Agilent whole human genome gene expression microarray platform. Analysis of differential expression between the metastases and normal prostate tissues revealed 235 up-regulated and 1082 down-regulated genes (at thresholds of $|\text{fold change}| > 2$ and $p<0.01$; Table S4), including several previously known prostate cancer differentially expressed genes (e.g. *AMACR, HPN, EZH2, GSTP1*; (34–37)). Interestingly, unsupervised hierarchical sample clustering by gene expression measures of the 500 most variably expressed genes across all samples showed the same patterns of within-subject maintenance and between-subject heterogeneity as was observed for DNA methylation (Figure 4A).

Given this similarity, we examined whether there was a correlation between the DNA methylation and expression patterns. Intersecting the coverage of the MBD-SNP platform with the gene expression microarray platform resulted in a set of 4194 genes for which both expression and promoter methylation data were available. We first examined whether DNA methylation alterations were correlated with gene expression at these genes. There was a

weak but highly significant inverse correlation overall between gene promoter methylation and gene expression measures across all samples and all 4,194 genes (Figure S15). Since we previously found that there were differences in the degree of maintenance (R^2) between the hypermethylated and hypomethylated loci, we now assessed whether there were differences in the correlation between DNA methylation and expression at each of these types of somatic methylation alterations. Regions showing promoter hypermethylation in the metastases were strongly associated with reduced levels of gene expression (average fold decrease = 1.33, $p=5.87 \times 10^{-38}$, Figure 4B), including at genes in the development/differentiation pathways (Figure S16), which were enriched for hypermethylation. Additionally, promoter hypermethylation was significantly more associated with gene expression differences than was promoter hypomethylation ($p = 9.2 \times 10^{-8}$), which generally had negligible association with gene expression.

We next identified the individual loci that showed evidence for significant correlation between gene expression and DNA methylation in the metastatic prostate cancer tissues. For these analyses, we were restricted to the 3,158 loci that were in the top 50th percentile of variability for either gene expression or promoter methylation in order to exclude regions that showed little or no variation in either dataset. We found that 452 out of these 3,158 loci showed a nominally significant correlation between gene expression and DNA methylation at $p < 0.05$, of which the majority showed the expected negative correlation (Figure S16). The positive correlation between DNA methylation and gene expression in a minority of regions (Figure S16), may be due to our definition of promoter regions, which includes portions of gene bodies and insulator regions, or might be due to complex *cis* and *trans* regulation leading to activation of hypermethylated genes (38).

Interestingly, while there was an overall strong pattern of within-subject stability in gene expression (see Figure 4A), this intra-individual stability (R^2) was strongest for those 452 genes whose expression level was associated with promoter DNA methylation (Figure 4C). Similarly, those promoter methylation loci that were associated with gene expression changes were more stably maintained within subjects than those whose methylation did not show significant correlation with gene expression (Figure S17). These observations suggest that DNA methylation alterations that are associated with phenotypic changes in gene expression have a greater tendency to be maintained within individuals, perhaps due to selection of those phenotypes, leading to maintenance of both the DNA methylation and gene expression patterns.

A careful examination of Figure 2C shows that although intra-individual heterogeneity in DNA methylation is much smaller than inter-individual heterogeneity, different metastases within each individual show clear patterns of clonal evolution and tumor heterogeneity in their methylation patterns. We next explored whether this intra-individual tumor heterogeneity in DNA methylation was correlated with phenotypic changes in gene expression. At a general level, we found that although there was a strong correlation between DNA methylation and gene expression at the inter-individual level, this general correlation was absent at the intra-individual level (Figure S18). To focus this analysis further, we assessed the correlation between methylation and expression at specific regions that showed evidence of significant intra-individual tumor heterogeneity in DNA methylation across multiple subjects. We identified 74 hypermethylated loci (1.9% of all hypermethylated regions), including 30 gene promoters, and 1,255 hypomethylated loci (5.6% of all hypomethylated regions), including 115 promoters, that showed such a pattern of recurrent intra individual tumor heterogeneity in DNA methylation (See Supplementary Materials and Methods and Figures S19, S20). Interestingly, we found no correlation between DNA methylation and gene expression at the intra-individual level even at these loci showing significant and recurrent DNA methylation heterogeneity in different

metastases within individuals. Furthermore, such regions were not enriched for any GO gene sets. Taken together, these analyses suggest that DNA methylation alterations showing within-subject tumor heterogeneity are not significantly correlated with gene expression patterns across different metastases within the same individual.

Genomic "Cityscapes" of DNA methylation alterations in lethal metastatic prostate cancer

To visualize both frequency and maintenance of DNA methylation alterations by position across the genome, we constructed "cityscapes" of DNA methylation changes in lethal metastatic prostate cancer. Such "cityscape" plots were constructed for regions showing gains and losses in TM (hyper- and hypo- methylated regions; Figure 5) and in ASM (Figure S21) in the metastases compared to normal prostate tissues. Within each "cityscape", chromosomes were folded into neighborhoods along a Hilbert curve (39); the area of each chromosome neighborhood was proportional to the number of informative probes. Within these neighborhoods, each "address" represents a single locus in the genome (Tables S2, S3 list the location and annotation of all interrogated addresses for ASM and TM respectively). The height of each structure in the cityscape represents the number of tumors showing a somatic DNA methylation alteration. The color of each structure represents the degree of maintenance of methylation across all metastases within each individual as measured by R^2 from our ANOVA model, with red indicating a high degree of somatic alteration maintenance (high R^2) and white indicating a low degree of maintenance (low R^2) relative to total variability. Note that the R^2 maintenance metric is not meaningful at loci where overall variability is negligible due to all tumors being fully methylated (e.g. the *GSTP1* promoter).

For DNA hypermethylation alterations, several regions appeared as "skyscrapers" in the cityscape, indicating regions that were frequently hypermethylated (Figures 5A, S22). Such skyscrapers were highly enriched for the Memorial Sloan Kettering Prostate Cancer Pathways gene set (26) and several sets from the NCI Cancer Gene Index (27) (Figure S23). The cityscape contains several densely populated neighborhoods with clustered skyscrapers, suggesting contiguous chromosomal segments frequently prone to hypermethylation (Figure 5A). This observation is consistent with previous findings of long-range epigenetic silencing in large chromosomal tracts (40). We also observed several "low- and mid-rises" in the cityscape, indicating regions that were hypermethylated in only one or a few subjects. Interestingly, among these low- to mid-rise regions we found that those that were red, indicating high maintenance of hypermethylation, showed enrichment for cancer related genes (Figure S23) relative to white-yellow regions. This "cityscape" of hypermethylation alterations in lethal metastatic prostate cancer revealed an unexpected importance of low frequency but highly maintained DNA methylation alterations as potential driver epigenome alterations.

The hypomethylation cityscape (Figure 5B) differed from the hypermethylation cityscape in two major ways: i) it contained many more structures, representing extensive regions of frequent hypomethylation; and ii) it showed a much higher fraction of white and yellow structures, where intra-individual variation represented a much greater fraction of overall tumor heterogeneity. These findings are consistent with those from a recent report showing large blocks of highly variable hypomethylation in human cancers (41). Such widespread regions of hypomethylation may contribute to genomic instability by multiple mechanisms, including insertion of transposable elements (42). Finally, somatic ASM alterations were much less numerous than total methylation alterations, and were significantly less consistently maintained compared to hypermethylation alterations (Figure S21; Table S2).

DISCUSSION

Given that epigenetic alterations can be labile, it has been questioned whether DNA methylation alterations can even be stable enough to be subject to selection during the clonal expansion events occurring during carcinogenesis, disease progression, and metastatic dissemination. With our study design of examining multiple metastases with a monoclonal origin within the same individual, we were able to examine distinct clonal expansion events within individuals (each metastasis) to assess the extent to which DNA methylation alterations were maintained across these metastases. We found that, overall, epigenetic alterations in DNA methylation were maintained to a similar extent as genetic alterations in copy number, suggesting that they have a similar potential as genetic alterations in serving as selectable driver events during clonal expansion/metastatic dissemination. This suggests that DNA methylation alterations could serve as a valuable source of targets for development of markers for cancer detection and prognosis and for development of new therapeutic strategies. However, this striking stability of DNA methylation alterations also implies that it will be important to distinguish between driver and passenger DNA methylation alterations, just as it has been important to do so for genetic alterations.

In this regard, our ability to survey different types of DNA methylation alterations, including DNA hypermethylation, DNA hypomethylation, and allele-specific methylation, has provided several new insights. First, we found a general tendency for widespread, but somewhat variable loss of methylation at normally methylated regions in metastases within individuals (see the large number of hypomethylation events in the hypomethylation cityscape in Figure 5B). These hypomethylation events were not strongly correlated with any functional gene sets, or with *cis* activation of gene expression. Therefore, if DNA hypomethylation played a driver role, it would likely be through promotion of genetic instability, for example through promotion of retrotransposition (42), rather than through direct *cis* regulation of specific genes. In contrast, somatic acquisition of DNA hypermethylation at regions that were normally unmethylated (see hypermethylation cityscape) were less numerous, but showed a greater tendency to stay methylated across all metastases. This staunch maintenance of newly acquired hypermethylation events in the metastases against the backdrop of a tendency to lose methylation at normally methylated regions in a widespread fashion across the genome, suggests that most of the hypermethylation events were likely subject to specific selection across metastatic dissemination and expansion. This observation, combined with the strong correlation with *cis* regulation of gene expression and enrichment of hypermethylation at development/differentiation and cancer genes, suggests that DNA hypermethylation events may be highly enriched for driver epigenetic events.

Given the parallel evolution of DNA methylation and copy number alterations, it is possible that the DNA methylation alterations may be caused by genetic alterations in the cancers, or vice versa. However, even in the former case, the DNA methylation alterations may still be part of the causal chain in cancer progression – e.g. genetic alterations lead to epigenetic alterations which are required for carcinogenesis or metastatic dissemination. While this would need to be investigated in future functional studies, there is emerging evidence to implicate this chain of causation involving epigenetic alterations. For example, in the *Apc Min* mouse model (43), which typically develops dozens of intestinal polyps by 3 to 6 months of age, disruption of DNMT1 or MBD2, key mediators of DNA methylation-induced gene silencing, leads to pronounced reduction of polyp formation (44–46). Additionally, malignant transformation via activation of a variety of oncogenes often involves widespread epigenome alterations that have also been implicated in the causal chain (23, 47–49). Such somatic epigenetic alterations resulting from genetic mutations may

be of particular interest since they may be more targetable/reversible through pharmacological manipulation than the upstream genetic alterations.

Our studies also reveal important insights on prostate tumor heterogeneity. There is a considerable amount of inter-individual tumor heterogeneity at both the genetic (16, 26, 50, 51) and epigenetic levels. This inter-individual heterogeneity challenges “one-size-fits-all” approaches for cancer management, and highlights the need for individualized medicine approaches. Second, while the amount of intra-individual heterogeneity across metastases is considerably less than the inter-individual variability for both genetic and epigenetic alterations, there is clonal evolution leading to appreciable intra-individual metastatic tumor heterogeneity in DNA methylation patterns. However, despite the strong relationship between heterogeneity in promoter methylation, particularly hypermethylation, and gene expression at the inter-individual level, there was essentially no correlation between DNA methylation and gene expression at the intra-individual level. Based on these results, we can speculate that DNA methylation heterogeneity between different metastases within individuals arises in a largely stochastic manner, without much impact on *cis* regulation of gene expression phenotypes. It is therefore possible that lethal metastatic prostate cancer arises after passing through a very narrow, but individual-specific clonal gate, with very little functional heterogeneity developing afterwards. In a similar vein, a recent whole genome analysis of primary and metastatic renal carcinoma showed that the degree of heterogeneity across different metastases within the same individual was much lower than the degree of heterogeneity across different portions of the primary tumor from the same individual (52). On an optimistic note, this striking intra-individual homogeneity across the lethal metastatic clonal gate, now observed at both the genetic (16) and epigenetic levels, may therefore represent a window of opportunity for effectively treating the lethal metastatic prostate cancer cell clone systemically. Studies such as the ones presented here could potentially focus target selection to the most promising genomic loci, exhibiting consistent somatic genome alterations across all metastases in affected individuals.

MATERIALS AND METHODS

Prostate specimens

Tissues from multiple anatomically distinct prostate cancer metastases and matched normal tissues from non-prostate tissues were obtained through the Project to Eliminate Lethal Prostate Cancer (PELICAN) rapid autopsy program at the Johns Hopkins Autopsy Study of lethal Prostate Cancer, as previously described (16). Organ donor benign prostate tissues were obtained from 24 brain dead organ donors with no evidence of prostate cancer. Tissue samples were snap-frozen, microdissected with a cryostat, and subjected to DNA isolation as described previously (19). Subject and sample data are provided in Supplementary Table S5, S6.

MBD-SNP approach

The methyl-binding domain of the MBD2 protein (MBD2-MBD) can bind methylated DNA fragments with exquisite selectivity and has been used to efficiently enrich methylated DNA fragments from genomic DNA (14, 15). Analyzing the resulting methylated DNA library with real-time PCR, tiling microarrays, and next generation sequencing has allowed gene-specific, chromosome-wide, and genome-wide DNA methylation analysis previously (14, 15, 53). In the MBD-SNP assay described here, we use the MBD2-MBD polypeptide to isolate methylated DNA fragments from genomic DNA samples followed by analysis with Affymetrix SNP 6.0 high density oligonucleotide microarrays. Comparison with an unenriched total input fraction then allows genome-scale determination of total and allele-specific methylation and copy number in an integrated fashion for each specimen. Briefly,

each genomic DNA specimen (1 μg) was divided into two equal fractions: i) an enriched methylated fraction (EM) and, ii) a total input fraction (TI). Each of these fractions was further divided into two equal reactions, each of which was either digested with the NspI or StyI restriction enzymes in separate reactions. Therefore, each fraction (EM or TI) and restriction enzyme digest (NspI or StyI) received 250 ng of genomic DNA. The resulting genomic DNA fragments were then ligated with Affymetrix SNP 6.0 assay adaptors. These restriction digest and adaptor ligation steps were carried out following the Affymetrix SNP 6.0 assay protocols. Up to this point, the EM and TI fractions were treated identically. After adaptor ligation, the TI fraction was brought to a total volume of 100 uL with water and set aside on ice; the EM fraction was subject to enrichment for methylated DNA fragments using MBD2-MBD polypeptides immobilized on magnetic beads as previously described (14, 15) except that the final DNA was eluted in 45 uL of EB1 buffer (0.2X NEBuffer 1 (New England Biolabs, Ipswich, MA), 0.2X BSA (NEB), 0.25X T4 DNA ligase Buffer (NEB) in water) for the DNA previously digested by NspI and 35 uL of EB2 buffer (0.2X NEBuffer3, 0.2X BSA (NEB), 0.25X T4 DNA ligase buffer (NEB) in water) for the DNA previously digested by StyI. These elution buffers were formulated so that the DNA from the EM fractions would be in the same buffers as the DNA from the TI fraction. For the EM and TI fractions previously digested with NspI, four 10 uL aliquots of DNA were amplified in 4 separate 100 uL one primer amplification reactions (30 cycles); for the EM and TI fractions previously digested with StyI, three 10 uL aliquots of DNA were amplified in 3 separate one primer amplification reactions (30 cycles), according to the Affymetrix SNP 6.0 microarray protocol. The seven amplification reactions for each fraction (EM and TI) were then pooled, and subjected to clean-up, labeling, hybridization to Affymetrix SNP 6.0 microarrays, washing, and scanning according to the manufacturer's protocols.

Affymetrix SNP 6.0 microarray and MBD-SNP probe selection

The Affymetrix SNP 6.0 microarray contains copy number probes at ~900,000 non-polymorphic loci, and an additional ~900,000 SNP probe sets at polymorphic loci. Our assay allows estimation of allele-specific methylation at polymorphic loci, and total methylation estimation at both polymorphic and non-polymorphic loci. We restricted analysis to probes in regions with a CpG density ≥ 2.5%. The CpG density for a given probe was calculated as the average of the CpG densities of the NspI and StyI fragments containing the probe location. NspI and StyI fragments that were not within the size selected range of 100 to 2500 bp were excluded from the calculation of CpG density. The CpG density cutoff was chosen based on preliminary analysis of the fully in-vitro methylated control sample that determined that these regions allow for robust detection of methylation signals. With this filter, 7,323 genes had at least one MBD-SNP total methylation probe within 5kb upstream and 2kb downstream of the transcription start site. Of these, 4,295 genes had at least one MBD-SNP allele-specific methylation probe within the same region.

MBD-SNP Total and Allele-Specific Methylation estimates

For a given sample, let x_{iA}^{E} and x_{iB}^{E} denote the enriched methylated fraction (E) signal intensity recorded at probe location i for allele A and B respectively. Similarly, x_{iA}^{T} and x_{iB}^{T} represent intensity values from the total input fraction (T) array. Taking the intensity ratio of enriched DNA to total DNA resulted in methylation estimates that are normalized for copy-number and probe effects. ASM estimates were restricted to loci with heterozygous genotype calls.

The methylation signal is most directly assessed at non-polymorphic (copy number probe) loci where it is given by $m_i = \log_2 x_i^{E}/x_i^{T}$. The non-polymorphic probe signals are quantile normalized between samples. Quantile normalization is typically inappropriate for methylation data as there can be significant differences in total methylation levels between

samples. In this case however, we take advantage of the fact that the majority of probes on the array are in low CpG-density regions that are below the robust detection limit of the MBD assay. These probes, which dominate the signal distribution, are therefore expected to behave similarly across all samples in accordance with the quantile normalization assumption of equal between-sample signal intensity.

SNP loci methylation estimates are obtained by combining the signal from the two alleles: $m_i=\log_2[(x^E_{iA} + x^E_{iB}) / (x^T_{iA} + x^T_{iB})]$. The polymorphic and non-polymorphic probes are roughly evenly interspersed throughout the genome and, as a result, the methylation distributions of these two sets of probes are expected to be the same. We take advantage of this fact by quantile normalizing the polymorphic signal distribution to a target distribution defined by the non-polymorphic probes, putting both types of probes on the same scale.

Probes on CpG-free restriction fragments were used as unmethylated control loci. As is common in many microarray applications, the unnormalized methylation values displayed a bias related to the probe GC content. This bias was corrected by adjusting values such that the GC-stratified, median control-probe methylation value was set to zero.

The raw allele-specific methylation (ASM) signal at informative (heterozygous (A/B) genotype) loci was calculated as $\log_2[(x_{iA}^E/ x_{iA}^T) / (x_{iB}^E/ x_{iB}^T)]$. As it is reasonable to assume that overall distribution of allele-specific methylation is similar between samples, we quantile normalized these ASM ratios. Since each SNP is represented by 3 replicate probes for the two alleles, the final SNP ASM ratio was calculated as the median of these ASM ratios.

Classification of MBD-SNP methylation status

The total methylation signal distribution had two clear modes, likely representing unmethylated and highly methylated loci and could be modeled as a two-component normal mixture model (Figure S24). This model was used to classify loci as methylated or unmethylated. Allele-specific methylation was similarly classified as being present or absent using a normal mixture model.

Identification of hyper-/hypo-methylation and gain/loss of ASM

Hypermethylated loci were defined as being unmethylated in all organ donor normal samples, and methylated in at least one prostate cancer metastasis. Hypomethylated loci were defined similarly, with all organ donor normal prostates showing methylation and at least one tumor showing lack of methylation. Regions of gain of ASM were defined as those that were classified as not having ASM in any of the organ donor normal prostate tissues, and having ASM in at least one tumor sample. Regions of loss of ASM were defined as those that were classified as having ASM in all of the organ donor normal prostate tissues, but classified as not having ASM in at least one tumor sample. To assess the number of alterations per subject, we restricted analysis to three randomly selected tumors per subject. This allowed comparison of number of alterations across subjects without bias to differences in the number of tumors available for a given subject.

Correlation analysis and hierarchical clustering by DNA methylation measures

Between-sample similarity was computed using the Pearson correlation coefficient. Average linkage Euclidean distance hierarchical clustering was carried out using 71 tissue samples from the 13 patient subjects using the 500 probes/probesets with greatest variance across samples. Copy number had a minimal effect on methylation estimates since both total and allele-specific methylation estimates were calculated as the ratio of methylated DNA to total DNA. However, to exclude the possibility that observed methylation patterns were driven by

residual copy-number effects, we carried out a further two step procedure prior to clustering. First, we restricted our analysis to probes in regions with a copy number of two as determined by Partek Genomic Suite (v6.4). Second, to account for any remaining subject-specific copy number variation, we fit probe level models to adjust for continuous copy-number estimates from CRLMM (v1.10.0).

Genotyping and copy number

Partek Genomic Suite (v6.4) was used to determine regions with gain or loss of copy number. The R/Bioconductor CRLMM package (v1.10.0) was used for genotyping and to generate raw (non-integer) copy number estimates.

Clonal maintenance R^2

Loci with heterogeneous somatic alterations were identified by choosing probes with low variability among organ donor normal prostate samples (lowest 75%), but high variability among tumor samples (top 500 and top 5%). For copy number estimates, probes were excluded if the mean estimate among organ donor normal was outside the range (1.5, 2.5). Methylation estimates with a single informative subject were excluded. Methylation estimates were adjusted for copy-number effects, as described above in 'Hierarchical clustering by methylation status'. To quantify the fraction of ASM probes with R^2 values comparable to copy number, we calculated the copy number mean R^2 minus one standard deviation and determined the fraction of ASM probes with an R^2 value greater than this threshold.

To ensure that the difference in number of probes available for copy number (1,852,215) and total methylation (51,501) did not drive the observed similarity in maintenance, we repeated the analysis using the top 5% most variable probes and obtained the same result (Figure S10). To exclude the possibility that the observed difference in R^2 between hypermethylated and hypomethylated loci is related to differences in variability between large and small methylation log-ratios we discretized methylation values into unmethylated, partially methylated or fully methylated categories. R^2 values calculated using discretized methylation were very similar to those obtained from continuous methylation estimates.

Gene expression microarray data and analysis of correlation between DNA methylation and gene expression

Samples used for gene expression profiling included 18 metastases from 5 autopsy subjects, processed as described previously (16), and 21 normal prostate specimens from organ donors (54). Total RNA was extracted from cryostat sections and evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) as described previously (55). Gene expression profiling was performed according to the guidelines provided by the Agilent Whole Genome Expression Microarray system (Agilent Technologies). Briefly, each of the 39 RNA samples was linearly amplified and labeled with Cy5, and cohybridized with a common reference RNA sample derived from benign prostatic hyperplasia that was similarly amplified but labeled with Cy3. For each sample, expression ratios of Cy5/Cy3 for each probe constituted the raw gene expression measure for the corresponding gene. Raw data were pre-processed with the R/Bioconductor limma package using within-sample standard locally weighted least squares regression (lowess) normalization and between-sample quantile normalization. Values from replicate probes were averaged. The raw and normalized data is available from the Gene Expression Omnibus (GEO) with accession number GSE38241. Probes differentially expressed between prostate cancer metastases and normal prostate tissues were identified by a linear mixed-effects model that accounts for within-subject correlation between tumor samples. The top 500 most variably expressed probes across all tissues were identified and subjected to average linkage Euclidean distance

hierarchical clustering. For correlation analysis between DNA methylation and gene expression, methylation probes were assigned to genes if they were located within a 5kb upstream to 2kb downstream window around transcriptional start sites. In the case where multiple methylation probes were available for a given gene, one was selected at random. Gene-level linear regression models were used to assess statistical significance of the expression-methylation relationship. When assessing the strength of the intra-individual gene expression – methylation relationship, a subject-specific term was added to the model. R^2 values for $\log_2$ gene expression values were calculated as described for DNA methylation.

DNA methylation cityscapes

Genomic cityscape plots were created to display regions with altered TM or ASM in the metastatic prostate cancer tissues compared to the organ donor normal prostate tissues. Within each "cityscape", genomic loci were folded into neighborhoods in order of chromosomes along a Hilbert curve (39). Each address in the cityscape generally represents a single region of the genome that was interrogatable by the MBD-SNP approach. In rare circumstances, the position of adjacent structures was swapped when this improved visibility of a labeled structure. Due to dimensional constraints on the Hilbert curve layout, some addresses represent the maximal signal from two adjacent genomic loci. Each structure in the cityscape represents a region in which all of the organ donor normal prostate specimens conformed to the appropriate base state (e.g. all classified as unmethylated for the hypermethylation cityscape or all classified as methylated for the hypomethylation cityscape) AND at least one metastasis was altered in methylation state compared to the base state. The height of each structure in the cityscape indicates the fraction of tumors with a DNA methylation alteration. The tallest structures thus represent loci at which 100% of all tumors were classified as methylated and none of the organ donors were methylated (e.g. *GSTP1* in the hypermethylation cityscape). The color of each structure represents the somatic alteration maintenance metric (R^2). In general, when multiple promoter-associated probes were available, all were used for plotting, but only the one with highest alteration frequency was labeled. In the case where multiple probes show the same alteration frequency, the one with the highest R^2 was selected for labeling. For example, in the cityscape, all probes for *GSTP1* are plotted, but only the probe showing the highest frequency (SNP_A-4242162) is labeled as *GSTP1*. Cityscape plots were created using the Processing programming language.

Analysis software

R 2.14 (56), Bioconductor 2.8 (57) and Partek Genomic Suite 6.4 were used for all analyses. All code is available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the family members and friends of the autopsy study participants and organ donor subjects, the Grove Foundation, and John and Kathe Dyson, who provided support to the PELICAN autopsy study.

Funding: This work was supported by funding from the Department of Defense Prostate Cancer Research Program (PC073533/W81XH-08-1-0049), National Institutes of Health (CA58236, CA070196, CA113374, CA135008, GM083084), the Prostate Cancer Foundation Creativity and Challenge awards, the Patrick C. Walsh Prostate Cancer Research Fund/Dr. and Mrs. Peter S. Bing Scholarship (to S.Y.), the V Foundation for Cancer Research Martin D. Abeloff V Scholar Award (to S.Y.), the German Research Foundation (DFG) Research Fellowship (to P.N.), the Finnish Academy of Sciences Finnish Distinguished Professor Award (to G.S.B.), and from generous

support from Mr. David H. Koch and The Irving A. Hansen Memorial Foundation. We also acknowledge the support from the CapCURE foundation (to W.B.I. and G.S.B.), which made the initiation of the PELICAN prostate cancer autopsy study possible.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–674. [PubMed: 21376230]
2. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007; 128:683–692. [PubMed: 17320506]
3. Vogelstein, B.; Kinzler, KW., editors. The Genetic Basis of Human Cancer. McGraw-Hill, Medical Pub. Division; New York: 2002.
4. Probst AV, Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol. 2009; 10:192–206. [PubMed: 19234478]
5. Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris TB, Launer L, Gudnason V, Fallin MD. Personalized epigenomic signatures that are stable over time and covary with body mass index. Sci Transl Med. 2010; 2:49ra67.
6. Hemberger M, Dean W, Reik W. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nat Rev Mol Cell Biol. 2009; 10:526–537. [PubMed: 19603040]
7. Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F, Reid G. Transient cyclical methylation of promoter DNA. Nature. 2008; 452:112–115. [PubMed: 18322535]
8. Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G. Cyclical DNA methylation of a transcriptionally active promoter. Nature. 2008; 452:45–50. [PubMed: 18322525]
9. Rainier S, Feinberg AP. Genomic imprinting, DNA methylation, and cancer. J Natl Cancer Inst. 1994; 86:753–759. [PubMed: 8169973]
10. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993; 366:362–365. [PubMed: 8247133]
11. Tycko B. Allele-specific DNA methylation: beyond imprinting. Human Molecular Genetics. 2010; 19:R210. [PubMed: 20855472]
12. Schuebel KE, Chen W, Cope L, Glockner SC, Suzuki H, Yi JM, Chan TA, Van Neste L, Van Criekinge W, van den Bosch S, van Engeland M, Ting AH, Jair K, Yu W, Toyota M, Imai K, Ahuja N, Herman JG, Baylin SB. Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. PLoS Genet. 2007; 3:1709–1723. [PubMed: 17892325]
13. Cross SH, Charlton JA, Nan X, Bird AP. Purification of CpG islands using a methylated DNA binding column. Nat Genet. 1994; 6:236–244. [PubMed: 8012384]
14. Yegnasubramanian S, Lin X, Haffner MC, DeMarzo AM, Nelson WG. Combination of methylated-DNA precipitation and methylation-sensitive restriction enzymes (COMPARE-MS) for the rapid, sensitive and quantitative detection of DNA methylation. Nucleic Acids Res. 2006; 34:e19. [PubMed: 16473842]
15. Yegnasubramanian S, Wu Z, Haffner MC, Esopi D, Aryee MJ, Badrinath R, He TL, Morgan JD, Carvalho B, Zheng Q, De Marzo AM, Irizarry RA, Nelson WG. Chromosome-wide mapping of DNA methylation patterns in normal and malignant prostate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. BMC Genomics. 2011; 12:313. [PubMed: 21669002]
16. Liu W, Laitinen S, Khan S, Vihinen M, Kowalski J, Yu G, Chen L, Ewing CM, Eisenberger MA, Carducci MA, Nelson WG, Yegnasubramanian S, Luo J, Wang Y, Xu J, Isaacs WB, Visakorpi T, Bova GS. Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. Nat Med. 2009; 15:559–565. [PubMed: 19363497]
17. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA, Velculescu VE, Kinzler KW, Vogelstein B, Iacobuzio-Donahue CA. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010; 467:1114–1117. [PubMed: 20981102]
18. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, Morsberger LA, Latimer C, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal SA, Leroy C, Jia M, Menzies

A, Butler AP, Teague JW, Griffin CA, Burton J, Swerdlow H, Quail MA, Stratton MR, Iacobuzio-Donahue C, Futreal PA. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature. 2010; 467:1109–1113. [PubMed: 20981101]

19. Yegnasubramanian S, Kowalski J, Gonzalgo ML, Zahurak M, Piantadosi S, Walsh PC, Bova GS, De Marzo AM, Isaacs WB, Nelson WG. Hypermethylation of CpG islands in primary and metastatic human prostate cancer. Cancer Res. 2004; 64:1975–1986. [PubMed: 15026333]
20. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. Nat Rev Genet. 2006; 7:21–33. [PubMed: 16369569]
21. Jarrard DF, Bussemakers MJ, Bova GS, Isaacs WB. Regional loss of imprinting of the insulin-like growth factor II gene occurs in human prostate tissues. Clin Cancer Res. 1995; 1:1471–1478. [PubMed: 9815946]
22. Issa JP. CpG island methylator phenotype in cancer. Nature reviews. Cancer. 2004; 4:988–993.
23. 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. Nature Genet. 2006; 38:787–793. [PubMed: 16804544]
24. 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. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer cell. 2010; 17:510–522. [PubMed: 20399149]
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene Ontology: tool for the unification of biology. Nat Genet. 2000; 25:25–29. [PubMed: 10802651]
26. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, Antipin Y, Mitsiades N, Landers T, Dolgalev I, Major JE, Wilson M, Socci ND, Lash AE, Heguy A, Eastham JA, Scher HI, Reuter VE, Scardino PT, Sander C, Sawyers CL, Gerald WL. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010; 18:11–22. [PubMed: 20579941]
27. NCI Cancer Gene Index. National Cancer Institute; 2009. <https://wiki.nci.nih.gov/display/cageneindex/>
28. Yegnasubramanian S, Haffner MC, Zhang Y, Gurel B, Cornish TC, Wu Z, Irizarry RA, Morgan J, Hicks J, DeWeese TL, Isaacs WB, Bova GS, De Marzo AM, Nelson WG. DNA hypomethylation arises later in prostate cancer progression than CpG island hypermethylation and contributes to metastatic tumor heterogeneity. Cancer Res. 2008; 68:8954–8967. [PubMed: 18974140]
29. Esteller M. Epigenetics in cancer. N Engl J Med. 2008; 358:1148–1159. [PubMed: 18337604]
30. Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med. 2003; 349:2042–2054. [PubMed: 14627790]
31. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983; 301:89–92. [PubMed: 6185846]
32. Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP. Hypomethylation of DNA from benign and malignant human colon neoplasms. Science. 1985; 228:187–190. [PubMed: 2579435]
33. Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW, Ehrlich M. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. 1983; 11:6883–6894. [PubMed: 6314264]
34. Luo J, Zha S, Gage WR, Dunn TA, Hicks JL, Bennett CJ, Ewing CM, Platz EA, Ferdinandusse S, Wanders RJ, Trent JM, Isaacs WB, De Marzo AM. Alpha-methylacyl-CoA racemase: a new molecular marker for prostate cancer. Cancer Res. 2002; 62:2220–2226. [PubMed: 11956072]
35. Luo J, Duggan DJ, Chen Y, Sauvageot J, Ewing CM, Bittner ML, Trent JM, Isaacs WB. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. Cancer Res. 2001; 61:4683–4688. [PubMed: 11406537]

36. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, Rubin MA, Chinnaiyan AM. The polycomb group protein EZH2 isinvolved in progression of prostate cancer. Nature. 2002; 419:624–629. [PubMed: 12374981]
37. Lee WH, Morton RA, Epstein JI, Brooks JD, Campbell PA, Bova GS, Hsieh WS, Isaacs WB, Nelson WG. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. Proc Natl Acad Sci U S A. 1994; 91:11733–11737. [PubMed: 7972132]
38. Bert SA, Robinson MD, Strbenac D, Statham AL, Song JZ, Hulf T, Sutherland RL, Coolen MW, Stirzaker C, Clark SJ. Regional Activation of the Cancer Genome by Long-Range Epigenetic Remodeling. Cancer Cell. 2012
39. Anders S. Visualization of genomic data with the Hilbert curve. Bioinformatics. 2009; 25:1231–1235. [PubMed: 19297348]
40. Coolen MW, Stirzaker C, Song JZ, Statham AL, Kassir Z, Moreno CS, Young AN, Varma V, Speed TP, Cowley M, Lacaze P, Kaplan W, Robinson MD, Clark SJ. Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing(LRES) reduces transcriptional plasticity. Nat Cell Biol. 2010; 12:235–246. [PubMed: 20173741]
41. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, Briem E, Zhang K, Irizarry RA, Feinberg AP. Increased methylation variation in epigenetic domains across cancer types. Nature genetics. 2011; 43:768–775. [PubMed: 21706001]
42. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, Lohr JG, Harris CC, Ding L, Wilson RK, Wheeler DA, Gibbs RA, Kucherlapati R, Lee C, Kharchenko PV, Park PJ. Landscape of somatic retrotransposition in human cancers. Science. 2012; 337:967–971. [PubMed: 22745252]
43. Su LK, Kinzler KW, Vogelstein B, Preisinger AC, Moser AR, Luongo C, Gould KA, Dove WF. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science. 1992; 256:668–670. [PubMed: 1350108]
44. Laird PW, Jackson-Grusby L, Fazeli A, Dickinson SL, Jung WE, Li E, Weinberg RA, Jaenisch R. Suppression of intestinalneoplasia by DNA hypomethylation. Cell. 1995; 81:197–205. [PubMed: 7537636]
45. Eads CA, Nickel AE, Laird PW. Complete genetic suppression of polyp formation and reduction of CpG-island hypermethylation in Apc(Min/+) Dnmt1-hypomorphic Mice. Cancer Res. 2002; 62:1296–1299. [PubMed: 11888894]
46. Sansom OJ, Berger J, Bishop SM, Hendrich B, Bird A, Clarke AR. Deficiency of Mbd2 suppresses intestinal tumorigenesis. Nat Genet. 2003; 34:145–147. [PubMed: 12730693]
47. Bakin AV, Curran T. Role of DNA 5-methylcytosine transferase in cell transformation by fos. Science. 1999; 283:387–390. [PubMed: 9888853]
48. Peli J, Schroter M, Rudaz C, Hahne M, Meyer C, Reichmann E, Tschopp J. Oncogenic Ras inhibits Fas ligand-mediated apoptosis by downregulating the expression of Fas. Embo J. 1999; 18:1824–1831. [PubMed: 10202146]
49. Gazin C, Wajapeyee N, Gobeil S, Virbasius CM, Green MR. An elaborate pathway required for Ras-mediated epigenetic silencing. Nature. 2007; 449:1073–1077. [PubMed: 17960246]
50. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, Onofrio R, Carter SL, Park K, Habegger L, Ambrogio L, Fennell T, Parkin M, Saksena G, Voet D, Ramos AH, Pugh TJ, Wilkinson J, Fisher S, Winckler W, Mahan S, Ardlie K, Baldwin J, Simons JW, Kitabayashi N, MacDonald TY, Kantoff PW, Chin L, Gabriel SB, Gerstein MB, Golub TR, Meyerson M, Tewari A, Lander ES, Getz G, Rubin MA, Garraway LA. The genomic complexity of primary human prostate cancer. Nature. 2011; 470:214–220. [PubMed: 21307934]
51. Rubin MA, Maher CA, Chinnaiyan AM. Common Gene Rearrangements in Prostate Cancer. J Clin Oncol. 2011
52. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C. Intratumor heterogeneity and

branched evolution revealed by multiregion sequencing. N Engl J Med. 2012; 366:883–892. [PubMed: 22397650]

53. Serre D, Lee BH, Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res. 2010; 38:391–399. [PubMed: 19906696]
54. Prakash K, Pirozzi G, Elashoff M, Munger W, Waga I, Dhir R, Kakehi Y, Getzenberg RH. Symptomatic and asymptomatic benign prostatic hyperplasia: molecular differentiation by using microarrays. Proc Natl Acad Sci U S A. 2002; 99:7598–7603. [PubMed: 12032329]
55. Dunn TA, Chen S, Faith DA, Hicks JL, Platz EA, Chen Y, Ewing CM, Sauvageot J, Isaacs WB, De Marzo AM, Luo J. A novel role of myosin VI in human prostate cancer. Am J Pathol. 2006; 169:1843–1854. [PubMed: 17071605]
56. R. D. C. Team. R Foundation for Statistical Computing; Vienna, Austria: 2011.
57. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004; 5:R80. [PubMed: 15461798]
58. Kobayashi Y, Absher DM, Gulzar ZG, Young SR, McKenney JK, Peehl DM, Brooks JD, Myers RM, Sherlock G. DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. Genome Res. 2011; 21:1017–1027. [PubMed: 21521786]

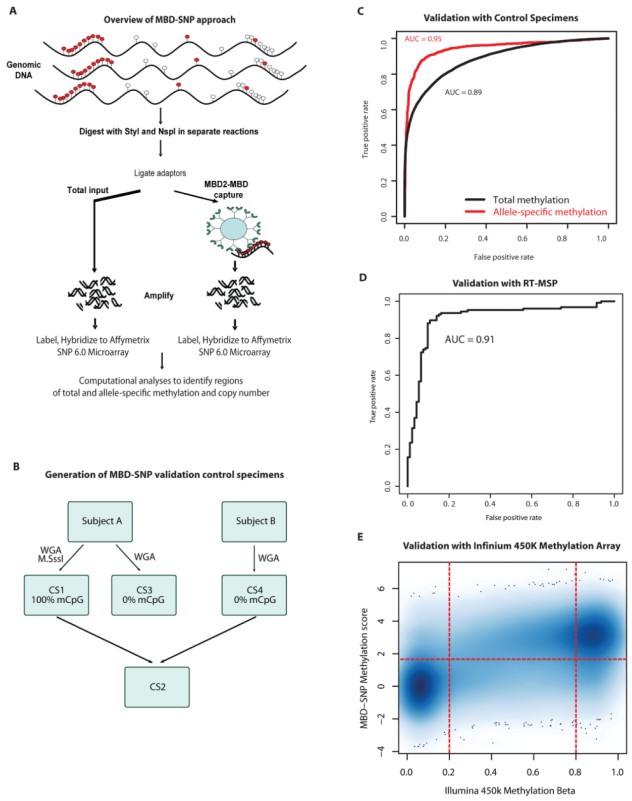

Figure 1. Overview and performance of the MBD-SNP method

(A) Overview of MBD-SNP workflow. Genomic DNA is fragmented by NspI and StyI in two separate reactions, and digested ends are ligated to adaptors. These products are then divided into a total input and enriched methylated fraction, the latter of which is subjected to enrichment of methylated DNA fragments by binding to MBD2-MBD immobilized magnetic beads. Both the total and enriched methylated fractions are then subjected to amplification, labeling, and hybridization to Affymetrix SNP 6.0 microarrays. Subsequent computational analyses comparing the enriched methylated and total input fractions allow assessment of total and allele-specific methylation. **(B)** Schematic showing generation of control specimens for testing MBD-SNP performance. **(C)** Receiver operator characteristic (ROC) curves for classification of total (allele-agnostic) and allele-specific methylation, generated by using CS1 (100% methylated), CS3 (0% methylated) and CS2 (50% methylated in individual specific fashion). **(D)** ROC curve for array performance as benchmarked against 5 loci across 44 samples verified by real-time methylation specific PCR (RT-MSP). **(E)** Concordance between MBD-SNP and Illumina 450k methylation estimates. There are 13,426 MBD-SNP methylation probes with an Illumina 450k probe located within 150bp. The MBD-SNP methylation score is plotted against the Illumina 450k methylation measure for each of the 13,426 probes and each of 12 specimens analyzed on both platforms. Among sites classified as unmethylated (Beta < 0.2) or highly methylated (Beta > 0.8) in the Infinium platform, 86.9% were concordant by the MBD-SNP mixture-model based classification of methylation status ($p \ll 1 \times 10^{-10}$).

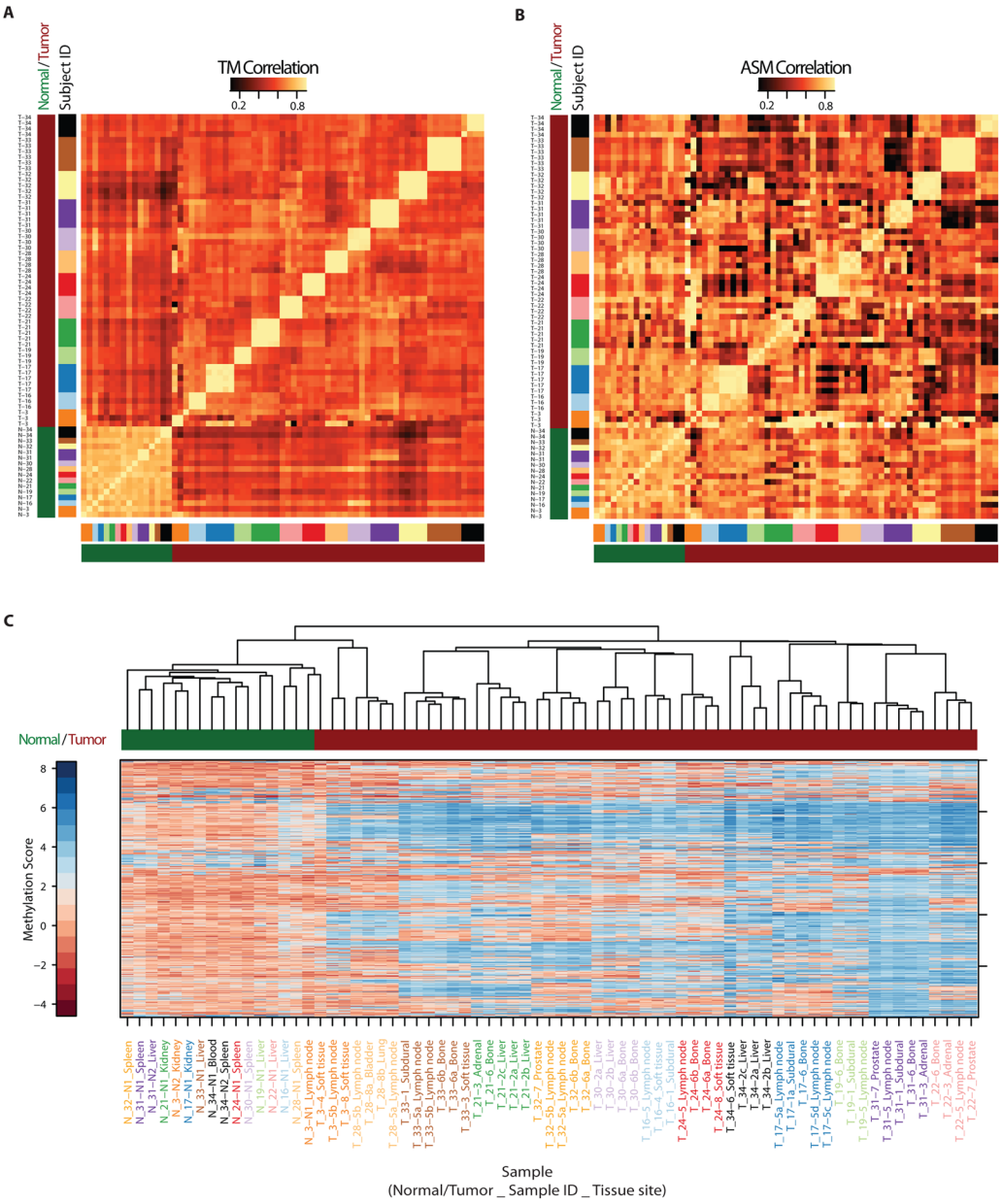

Figure 2. DNA methylation alterations are heterogeneous across individuals, but closely maintained in the metastases within each individual

Between-sample correlation for **(A)** total DNA methylation (TM), and **(B)** allele-specific DNA methylation (ASM). Axis labels indicate tissue type (tumor vs. normal) and subject identifier. The light colored block diagonal structure shows high inter-individual correlation for normal tissues (bottom left) and high intra-individual correlation and inter-individual heterogeneity for metastases. **(C)** DNA methylation hierarchical clustering showing high within-subject similarity and between-subject heterogeneity of metastases. These analyses were done using the 500 probes showing greatest variability across all samples.

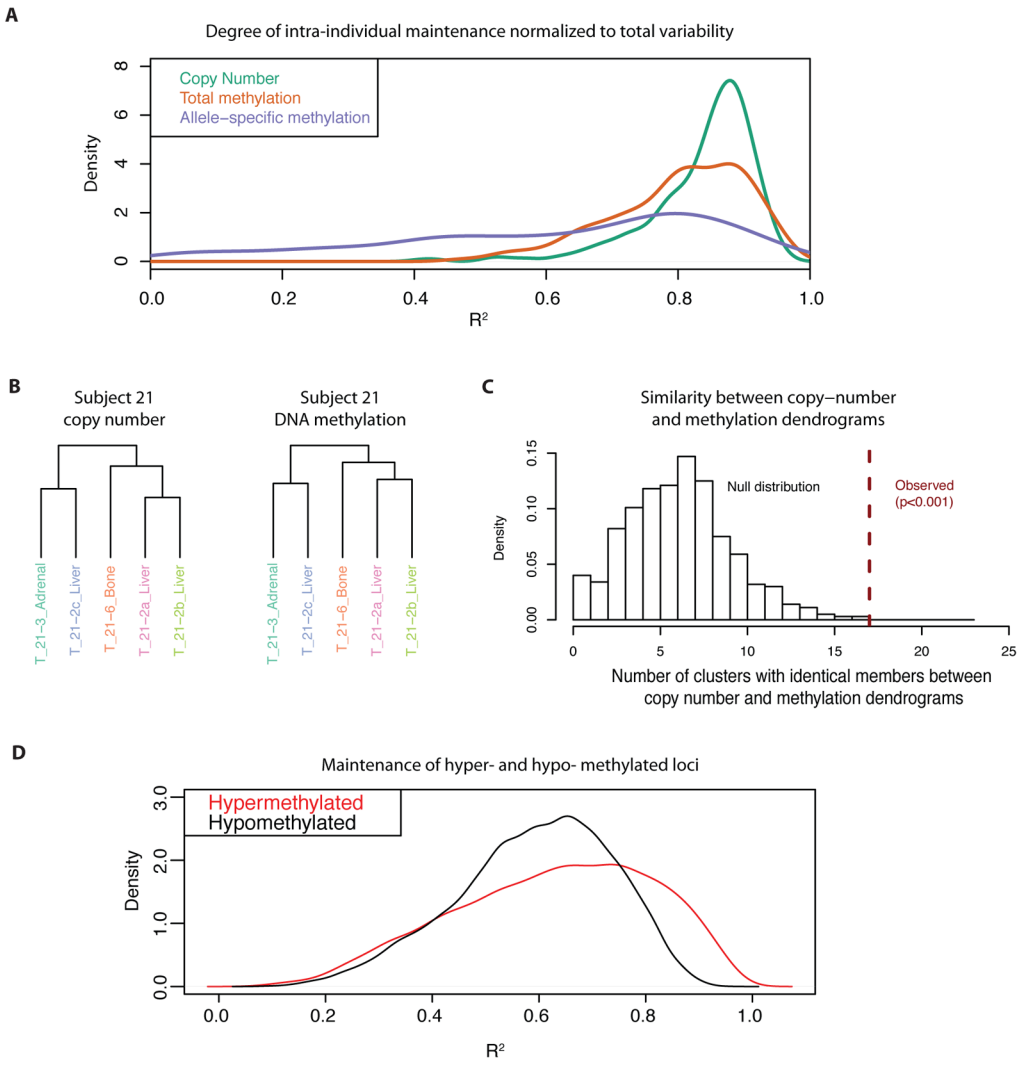

Figure 3. Epigenetic DNA methylation changes are maintained to an extent comparable to genetic copy number alterations

(A) Degree of maintenance within subjects normalized to the overall variability (R^2) for somatic copy number alterations, total DNA methylation alterations, and allele-specific DNA methylation alterations. The 500 probes with highest overall tumor variability were analyzed. **(B)** A subject showing near-perfect similarity between copy-number and DNA methylation hierarchical clustering dendrograms. **(C)** Degree of similarity between within-subject hierarchical clustering of copy-number and DNA methylation profiles is significantly greater than would be expected by random chance. **(D)** R^2 distribution for somatically altered regions. The degree of maintenance normalized to overall variability is higher for somatically hypermethylated regions than hypomethylated regions.

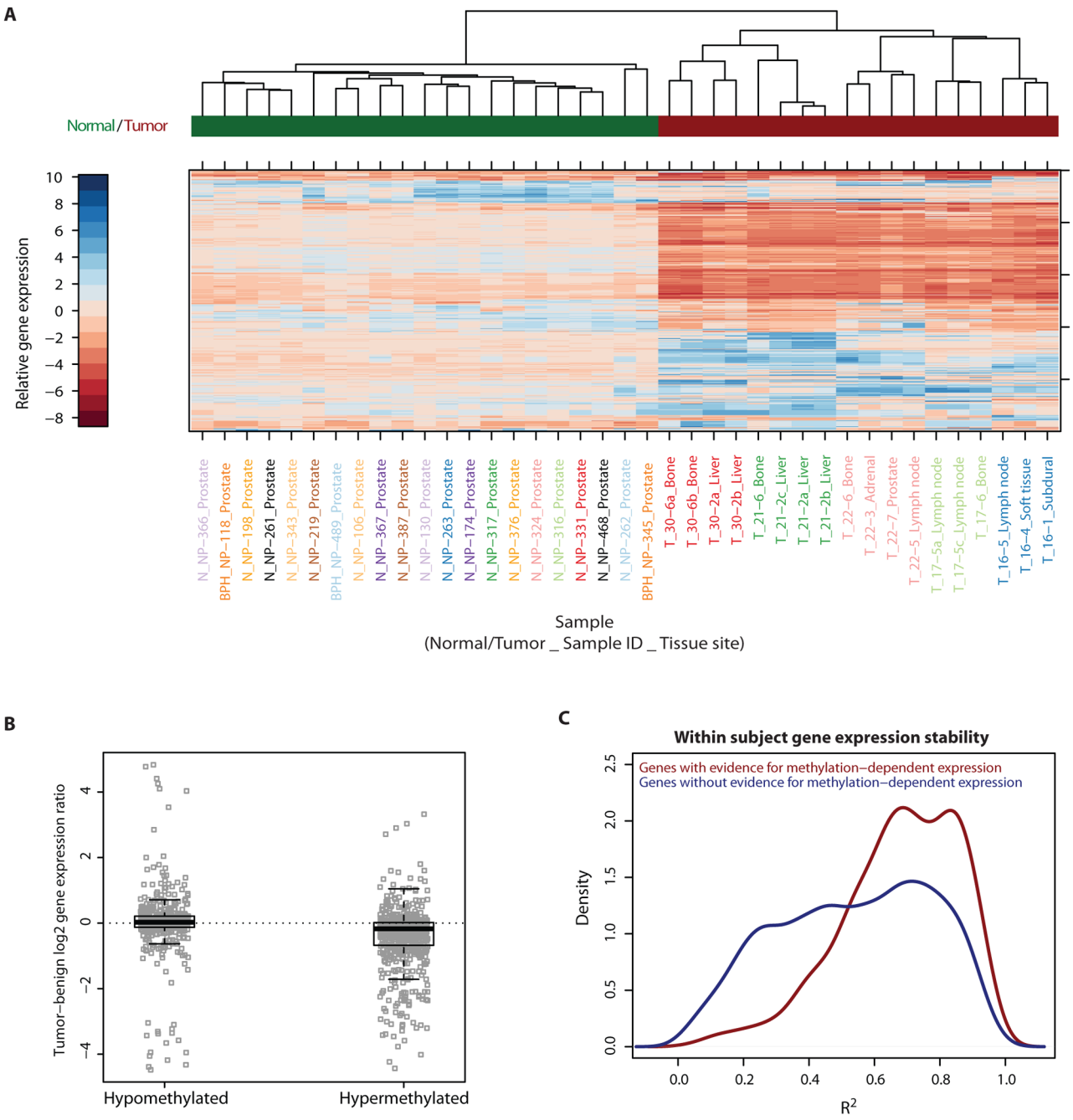

Figure 4. Gene expression patterns show high within-subject maintenance and correlation with DNA hypermethylation

(A) Gene expression hierarchical clustering showing high within-subject similarity and between-subject heterogeneity of metastases. These analyses were done using the 500 genes showing greatest variability in gene expression across all 21 benign and 18 metastasis samples with gene expression microarray data. **(B)** DNA promoter hypermethylation shows significant correlation with downregulation of gene expression (average fold decrease = 1.33, $p=5.87 \times 10^{-38}$) and is more strongly correlated with gene expression than promoter hypomethylation ($p=9.2 \times 10^{-8}$). **(C)** Within-subject maintenance (R^2) of gene expression patterns for genes with significant correlation between DNA methylation and gene

expression was stronger than for genes having no evidence for methylation-expression correlation.

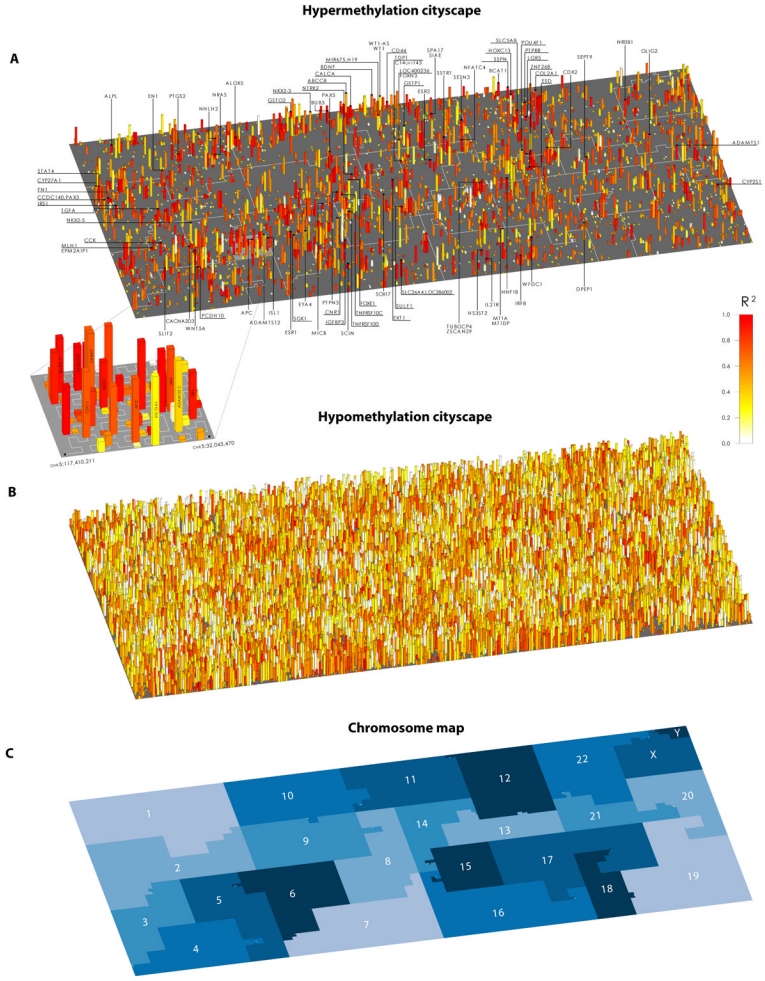

Figure 5. DNA methylation "Cityscape" plots of lethal metastatic prostate cancer highlight frequent and highly maintained alterations

Genomic cityscapes of somatic **(A)** hypermethylation and **(B)** hypomethylation. Each chromosome is folded into neighborhoods along a Hilbert curve as shown in **(C)**. Each structure represents a region showing alteration in TM compared to the normal prostate tissues. The height of each structure indicates the number of tumors showing alteration. The color scale represents the degree of maintenance of these alterations across metastases within individuals normalized to the overall variability (R^2). Hypermethylated promoter regions of genes from the NCI Cancer Gene Index that fell in the top tenth percentile of frequency of alteration or R^2 are labeled. The magnified region in **(A)** illustrates a representative chromosomal segment showing clustering of frequently hypermethylated regions (skyscrapers). The white path shows the Hilbert curve "folding" of this genomic segment.