



Published in final edited form as:

Cell Rep. 2022 January 11; 38(2): 110220. doi:10.1016/j.celrep.2021.110220.

2-Hydroxyglutarate destabilizes chromatin regulatory landscape and lineage fidelity to promote cellular heterogeneity

Meena Kusi¹, Maryam Zand², Li-Ling Lin¹, Meizhen Chen¹, Anthony Lopez¹, Chun-Lin Lin¹, Chiou-Miin Wang¹, Nicholas D. Lucio¹, Nameer B. Kirma¹, Jianhua Ruan², Tim H.-M. Huang^{1,*}, Kohzoh Mitsuya^{1,3,*}

¹Department of Molecular Medicine, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229, USA

²Department of Computer Science, University of Texas at San Antonio, San Antonio, TX 78249, USA

³Lead contact

SUMMARY

The epigenome delineates lineage-specific transcriptional programs and restricts cell plasticity to prevent non-physiological cell fate transitions. Although cell diversification fosters tumor evolution and therapy resistance, upstream mechanisms that regulate the stability and plasticity of the cancer epigenome remain elusive. Here we show that 2-hydroxyglutarate (2HG) not only suppresses DNA repair but also mediates the high-plasticity chromatin landscape. A combination of single-cell epigenomics and multi-omics approaches demonstrates that 2HG disarranges otherwise well-preserved stable nucleosome positioning and promotes cell-to-cell variability. 2HG induces loss of motif accessibility to the luminal-defining transcriptional factors FOXA1, FOXP1, and GATA3 and a shift from luminal to basal-like gene expression. Breast tumors with high 2HG exhibit enhanced heterogeneity with undifferentiated epigenomic signatures linked to adverse prognosis. Further, ascorbate-2-phosphate (A2P) eradicates heterogeneity and impairs growth of high 2HG-producing breast cancer cells. These findings suggest 2HG as a key determinant of cancer plasticity and provide a rational strategy to counteract tumor cell evolution.

In brief

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Correspondence: huangt3@uthscsa.edu (T.H.-M.H.), mitsuya@uthscsa.edu (K.M.).

AUTHOR CONTRIBUTIONS

T.H.-M.H., K.M., and M.K. jointly conceived the project, designed the experiments, interpreted the results, and prepared the manuscript with contributions from all co-authors. M.K. carried out the majority of the experiments and data analyses. J.R., M.Z., and C.-L.L. assisted with scATAC-seq data analysis, computational modeling, and statistical methods. M.K., A.L., and K.M. performed RNA-seq experiments and data analysis. K.M. and M.K. established the methods for chromatin CyTOF. L.-L.L. and M.C. conducted IncuCyte cell proliferation assays and CyTOF single-cell profiling. C.-M.W., N.D.L., and N.B.K. assisted with validation of metal-tagged antibodies.

DECLARATION OF INTERESTS

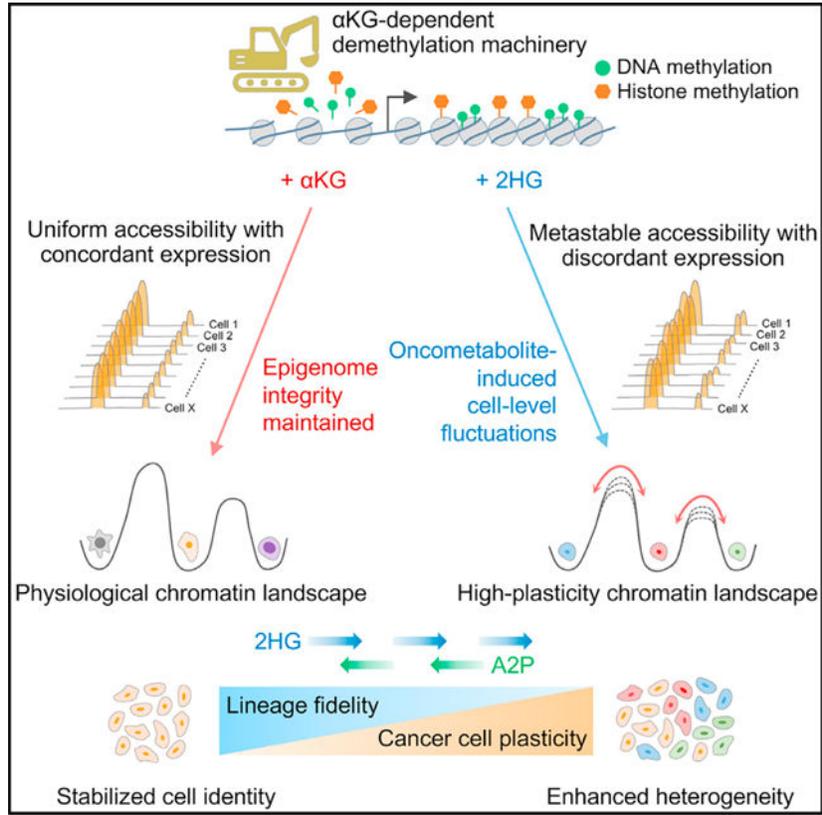
The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2021.110220>.

Kusi et al. show that the oncometabolite 2-hydroxyglutarate (2HG) initiates cell-level epigenome fluctuations in the chromatin regulatory landscape, accompanied by loss of lineage fidelity. Breast tumors with high 2HG accumulation exhibit enhanced cellular heterogeneity with undifferentiated stem-like epigenomic signatures. The findings suggest metabolic derangement as a molecular origin of breast cancer heterogeneity.

Graphical Abstract



INTRODUCTION

Breast cancer is characterized by prominent cell-to-cell variations in molecular makeup, and enhanced heterogeneity predisposes affected individuals to adverse clinical outcomes (Shipitsin et al., 2007; Jackson et al., 2020). Clonal heterogeneity is a universal phenotype across a multitude of cancer types and responsible for a breadth of disease manifestations, including metastatic dissemination and therapy resistance (McGranahan and Swanton, 2017; Dagogo-Jack and Shaw, 2018). Although not overtly tumorigenic, cancer cell plasticity fuels tumor heterogeneity, which is an essential step in tumor potentiation, initiation, and progression rather than being a secondary or passive phenomenon occurring late in tumor cell evolution (Flavahan et al., 2017; Huang, 2021). Because of the dynamic and transient nature of cancer cells, tumor development or progression does not typically follow a fixed sequence of processes, and systemic approaches are vital for developing conceptual

frameworks to integrate the multidimensional tumor cell diversity (Hanahan and Weinberg, 2011; Marusyk et al., 2020).

Although studies on tumor heterogeneity have focused primarily on genetic alterations, non-genetic factors, such as transcriptomic heterogeneity or noise in gene expression, could result in substantial neoplastic variability (Hinohara and Polyak, 2019). Non-genetic priming in cancer cells also contributes to rapid adaptation to therapeutic stress and metastatic progression (Shaffer et al., 2017; Roe et al., 2017). Chromatin architectures are instrumental in stabilizing transcriptional programs as well as in preserving cell type identities, and physiological cell fate transitions are epigenetically restricted via DNA and histone modifications, as depicted in Waddington's epigenetic landscape (Marusyk et al., 2012; Allis and Jenuwein, 2016). It has been suggested that deviation from this "epigenetic restriction" could confer a fitness advantage to cancer cells, allowing a selection process to occur (Flavahan et al., 2017; Pastore et al., 2019). This, in turn, implies that loss of epigenetic constraints could potentiate cancer cell plasticity by increasing phenotypic flexibility, resulting in enhanced tumor heterogeneity. Epigenomes are being recognized as a potential key determinant of oncogenic plasticity (Wainwright and Scaffidi, 2017; Nam et al., 2021). Molecular characterization of non-genetic mechanisms that regulate the plasticity of cellular states may provide a unique opportunity to advance our understanding of and ability to eradicate malignant cell evolution (Dagogo-Jack and Shaw, 2018; Hinohara and Polyak, 2019).

Epigenome regulators are highly sensitive to changes in intracellular metabolism, and one such example is the oncometabolite 2-hydroxyglutarate (2HG) (Faubert et al., 2020). 2HG accumulates to millimolar levels in solid and hematologic malignancies, including breast, kidney, colon, and pancreatic tumors (Dang et al., 2009; Terunuma et al., 2014; Jezek, 2020). 2HG occurs as two enantiomers, and somatic mutations in the isocitrate dehydrogenase genes *IDH1* and *IDH2* result in stereospecific production of the D-enantiomer (D2HG) in glioma and acute myeloid leukemia (AML) (Dang et al., 2009; Figueroa et al., 2010). In contrast, the L-enantiomer (L2HG) is promiscuously generated in the absence of *IDH1/2* mutations under the hypoxic and acidic conditions (Losman et al., 2020) that often coexist in the tumor microenvironment. Both enantiomers structurally resemble α -ketoglutarate (α KG), an intermediary metabolite in the Krebs cycle, and antagonize α KG-dependent dioxygenases, including ten-eleven translocation (TET) DNA hydroxylases and Jumonji domain-containing histone demethylase (JHDM) enzymes, which remove methyl groups from DNA and histones, respectively (Losman et al., 2020). Extensive studies in gliomas and AML have demonstrated that these oncometabolites promote cellular transformation and tumor progression by altering gene expression through local chromatin signaling (Losman et al., 2013; Flavahan et al., 2016; Inoue et al., 2016; Sulkowski et al., 2020); however, the genome-scale effect of 2HG on the stability and plasticity of the cancer epigenome has yet to be explored. This study focuses on the molecular origin of non-genetic heterogeneity and aims to investigate the role of 2HG in regulating chromatin homeostasis and lineage fidelity to uncover upstream mechanisms that promote breast tumor heterogeneity.

RESULTS

The chromatin-transcriptional landscape is reversibly remodeled by 2HG and restored by ascorbate-2-phosphate (A2P)

To determine the regulatory mechanisms involving cellular metabolism that mediate breast cancer development, we first analyzed the intracellular levels of D2HG and L2HG using enantiomer-selective liquid chromatography-mass spectrometry (LC-MS) (Figures S1A and S1B). Relatively higher levels of D2HG were observed in estrogen receptor α (ER α)-positive and -negative breast cancer cell lines compared with benign and primary mammary epithelial cells, whereas elevated L2HG levels were predominant in ER α -negative breast cancer cells (Figure S1C). In line with this, ER α -negative cells had higher levels of the 2HG pool (Figure S1D). We next leveraged the Cancer Cell Line Encyclopedia (CCLE) multi-omics repository (<https://sites.broadinstitute.org/ccle/>) and found that 2HG levels in ER α -negative cancer cells were significantly higher relative to ER α -positive cells and were comparable with those in *IDH*-mutated cells, whereas no distinctive differences were detected in the levels of other structurally similar metabolites that modulate the catalytic activity of α KG-dependent dioxygenases (Figure S1E). Our investigation of somatic mutations using multiple independent datasets, including the CCLE and Sanger Catalogue of Somatic Mutations in Cancer (COSMIC) panels, showed no evidence of gain-of-function mutations in *IDH1* or *IDH2* among 62 breast cancer cell lines, corroborating the finding of infrequent *IDH* mutations but high accumulation of 2HG in primary breast tumors (Figure S1F).

Primary human mammary epithelial cells (HMECs), which exhibited low levels of intracellular 2HG (Figures 1C and 1D), were then exposed for 72 h to 100 μ M cell-permeable 2HG. Either of the two enantiomers led to a decrease in 5-hydroxymethylcytosine (5hmC) and a reciprocal increase in 5-methylcytosine (5mC) at long interspersed element 1 (LINE-1) elements, consistent with competitive inhibition of TET-mediated conversion of 5mC to 5hmC during the process of active DNA demethylation (Figures 1A and S1G). Similarly, immunofluorescence staining with an anti-5hmC antibody showed a dose-dependent loss of 5hmC in response to D2HG (Figure 1B). In contrast, global levels of H3K27me3 histone methylation were elevated in proportion to the increasing concentrations of D2HG (Figures 1B and S1H). These results indicate that, compared with previous studies (Losman et al., 2013; Intlekofer et al., 2018; McBrayer et al., 2018), short-term exposure at a relatively low level of 2HG is sufficient to induce considerable changes in the mammary epithelial epigenome.

In contrast to DNA methylation, lysine residues on core histone undergo varying degrees of methylation, resulting in mono-, di-, or trimethylated states, and we therefore employed a multiplexed LC-MS assay to quantify changes in histone methylation (Figure S1I). Exposure to 2HG enantiomers led to elevated methylation levels in repressive (H3K27me3, H3K9me3, and H4K20me3) and permissive (H3K4me3, H3K36me3 and H3K79me3) histone marks, along with a decrease in the levels of unmethylated (me0) lysine residues (Figures 1C and S1J). L2HG led to preferential accumulation of H3K27me3 on the histone variant H3.3, which is enriched in actively transcribed regions (Ahmad and Henikoff, 2002). Of

note, there was no detectable change in cell cycle phase distribution after adding 2HG enantiomers (Figure 1D). Hence, 2HG-mediated chromatin remodeling cannot simply be attributed to cell cycle differences.

To characterize transcriptional changes imposed by 2HG, we performed RNA sequencing (RNA-seq), comparing HMECs that were treated with vehicle or 2HG enantiomers (Figure 1E). On the basis of the reported genome-scale distribution of histone modifications (Pellacani et al., 2016), more than three-quarters of the downregulated genes (80% and 76% in D2HG- and L2HG-treated cells, respectively) were found to be enriched with permissive H3K4me3 chromatin marks in primary mammary epithelial cells (Figure S1K), suggesting that transcriptional repression may be mediated through altered chromatin modifications, such as *de novo* DNA methylation. Gene set enrichment analysis (GSEA) revealed significant upregulation of gene signatures involved in hypoxia, IL2-JAK-STAT5, and TNF- α signaling, whereas DNA repair-related genes were downregulated, with no measurable effects on apoptosis-related gene signatures (Figure 1F; Table S1). Enrichment analysis using a previously defined panel of 267 genes associated with DNA damage repair (DDR) pathways in The Cancer Genome Atlas (TCGA) pan-cancer cohort (Knijnenburg et al., 2018) demonstrated that 2HG led to transcriptional repression of a broad spectrum of DDR pathway genes involved in homology-dependent recombination (HR), mismatch repair (MMR), nucleotide excision repair (NER), base excision repair (BER), and Fanconi anemia (FA) (Figures 1G, 1H, and S1L; Table S1). 2HG-induced downregulation of DDR pathway proteins was confirmed by immunoassay (Figure 1I). This result complements previous studies showing 2HG-induced impairment of DNA repair mechanisms caused by defective recruitment of DNA repair proteins (Sulkowski et al., 2020), direct enzymatic inhibition of DNA repair proteins (Wang et al., 2015), or inactivation of ATM-dependent DNA damage sensing (Inoue et al., 2016). In addition, pathway-level dysregulation was markedly abolished 4 days after removal of 2HG (Figure 1F), suggesting that this is an essentially reversible epigenetic process and dependent on continued presence of excess 2HG.

Previous studies have shown that vitamin C stimulates the catalytic activity of DNA and histone demethylases by donating an electron to Fe³⁺ to generate Fe²⁺, which is necessary for optimal activity (Wang et al., 2011; Blaschke et al., 2013). Although growing evidence supports the therapeutic benefits of high-dose vitamin C infusion, the development of more appropriate and rigorous clinical trial designs has been hampered by lack of a clear understanding of the mechanism of action (Ngo et al., 2019). To test whether vitamin C can reverse 2HG-induced transcriptional remodeling through epigenetic mechanisms, cells were treated with L2HG in the presence or absence of A2P, an oxidatively stable form of vitamin C, which does not contribute to extracellular hydrogen peroxide formation and is converted into vitamin C during transport across the cell membrane (Takamizawa et al., 2004). A2P supplementation largely restored DDR signaling pathways in the presence of L2HG (Figure S1L; Table S1). Sequence motif analysis of differentially expressed genes revealed that promoter regions of downregulated genes were enriched with transcription factor binding motifs containing CpG dinucleotides, which are target sites for DNA methylation (Figure S1M). In line with this, highly methylated genes in ER α -negative breast cancer cells showed reduced expression upon exposure to L2HG, which was restored by A2P supplementation

(Figure 1J). These results underscore the potent activity of A2P in antagonizing the inhibitory action of the oncometabolites.

2HG dysregulates the lineage-defining chromatin landscape and drives heterogeneity

To delineate the chromatin regulatory landscape imposed by 2HG, we applied a single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq) (Figures S2A–S2C). scATAC-seq reads exhibited the expected periodicity of ~200-bp insert size fragments corresponding to nucleosome occupancies (Figure S2B). Additionally, aggregate scATAC-seq profiles showed high concordance with the ensemble measurement of the accessibility landscape profiled by DNase-seq (Figures 2A and S2C). Principal-component analysis (PCA) and model-based clustering of single-cell DNA accessibility profiles identified five distinct cell subpopulations that were discernible not only between but also within the treatment groups (Figures 2B and 2C). In contrast, replicate samples were distributed similarly despite being processed in different experiments (Figure S2D), supporting the robustness of the method. The first principal component (PC1) broadly separated L2HG-treated cells from D2HG-treated cells (Figure 2B), indicating that the two enantiomers impose discrete chromatin regulatory landscapes. L2HG-treated cells were divided into multiple subpopulations oriented in the opposite direction on the PC2 axis and displayed a more diverse distribution compared with D2HG-treated cells. These results show that 2HG enantiomers enhance molecular heterogeneity in motif accessibility, with L2HG being more potent in modulating chromatin regulatory architectures.

We next examined cell-to-cell variability in chromatin accessibility at transcription factor (TF) binding motifs ($n = 386$) using chromVAR (Schep et al., 2017) (Figures 2D and 2E). Consistent with the observations that cell-level heterogeneity was most evident upon L2HG exposure, motif accessibilities were highly variable in clusters 4 and 5, which were largely composed of L2HG-treated cells (Figures 2C and 2D). Although to a lesser extent, increased cell-to-cell variability in motif accessibility was also detected in cluster 3. These data suggest that 2HG-mediated chromatin remodeling accompanies cell-level fluctuations in chromatin accessibility to gene regulatory regions, and we refer to this admixture of destabilized epigenetic states as a high-plasticity chromatin landscape. Considering that L2HG is produced preferentially in ER α -negative tumor cells, intracellular accumulation of L2HG could contribute to the increased degree of heterogeneity or transcriptional plasticity reported in basal-like tumors (Wang et al., 2014; Garrido-Castro et al., 2019).

Analysis of motif accessibility to lineage-specific TFs involved in mammary epithelial cell differentiation (Pellacani et al., 2016; Dravis et al., 2018; Bi et al., 2020; Kim et al., 2020) revealed considerable changes in transcriptional programs that regulate differentiation states. Cluster 5 displayed diminished motif accessibility to the luminal-specific TFs FOXA1, FOXP1, and GATA3, and this was accompanied by increased motif accessibility to luminal progenitor markers such as ELF1 (Figure 2F). Loss of motif accessibility to mature luminal factors was also detected in clusters 3 and 4 (Figures 2D and 2F). The results show that 2HG dysregulates epigenetic mechanisms that sustain lineage-specific transcriptional profiles, suggesting that 2HG-mediated loss of lineage fidelity could facilitate emergence of alternate and seemingly more perturbable (i.e., metastable) transcriptional programs,

which may result in enhanced heterogeneity in differentiation state. To corroborate the effect of 2HG on cell type fidelity, we analyzed expression of lineage markers and found that 2HG induced a decrease in luminal cytokeratin expression and reciprocal increase in basal cytokeratin expression (Figure S2E). Mass cytometry CyTOF analysis confirmed a slight but distinct shift toward increased basal (CD49f) and decreased luminal (EpCAM) markers in L2HG-treated cells (Figure S2F). These data suggest that 2HG alters lineage fidelity and facilitates non-physiological luminal to basal-like reprogramming of mammary epithelial cells.

2HG destabilizes the chromatin regulatory landscape to promote cell-to-cell variability

To define the underlying mechanisms that drive clonal heterogeneity in differentiation potential, we next investigated the genome-wide occupancy of open, accessible chromatin regions by utilizing the 12-state chromatin segmentation defined in HMECs (Ernst and Kellis, 2017). As expected, ATAC-seq signals in control cells were enriched in active/weak gene promoters and strong enhancers but underrepresented in inactive regions such as heterochromatin and repetitive regions, both of which are associated with nuclear laminae (Ernst et al., 2011; Figure S3A). Following 2HG exposure, chromatin accessibility was reduced in genomic regions enriched with permissive chromatin marks (H3K4me3, H3K27ac, and H3K9ac) (Ernst et al., 2011), such as active/weak promoters and strong/weak enhancers (Figure 3A). In contrast, genomic regions that are largely devoid of permissive chromatin marks displayed a substantial increase in chromatin accessibility. These two seemingly opposing effects on the mammary epigenome (i.e., loss of accessibility in active chromatin and gain of accessibility in repressive chromatin states) may reduce the stability of the epigenetic landscape. A recent study proposed a model in which disrupted coordination of mutually exclusive activating and repressing histone modifications increases cell-to-cell diversity in the epigenome, leading to an admixture of cells with diverging cellular identities (Pastore et al., 2019). 2HG-mediated dysregulation of the coordinated chromatin accessibility landscape may likewise result in erosion of the epigenetic landscape, altering lineage fidelity and, thus, enhancing epithelial cell heterogeneity.

To address the sparse nature of scATAC-seq data, further inspection was carried out using the previously defined annotations of genomic regions (Buenrostro et al., 2015; Schep et al., 2017). Analysis of mammary epithelial enhancer and super-enhancer regions (Hnisz et al., 2013; Jiang et al., 2019) revealed that 2HG rendered accessible enhancer elements less open while inducing less closed chromatin at inaccessible elements (Figures 3B and 3C). Of note, the accessible peaks detected at super-enhancers were found to be diffused upon 2HG exposure, which was most evident in L2HG-treated cells (Figure 3C, top right panel). Decreased and diffused accessibility was also detected at chromatin domains that are bivalently marked by H3K4me3 and H3K27me3 (Figure S3B, top right panel). Given the established role of super-enhancers and bivalent genes in regulating cell identity and lineage differentiation (Hnisz et al., 2013; Feinberg et al., 2016), the findings corroborate that 2HG dysregulates epigenetic mechanisms that maintain lineage fidelity. To further probe how 2HG-mediated chromatin remodeling contributes to the development of epithelial cell heterogeneity, we next investigated accessibility variation across the genome in single-cell measurements. As shown in Figure 3D, where each color represents DNA accessibility

in a single cell, chromatin accessibility at transcriptional start sites (TSSs; $n = 20,242$) was highly concordant in control cells; however, open chromatin peaks were found to be diffused and heterogeneous in 2HG-treated cells. Notably, cell-level epigenetic fluctuations were also evident in enhancer elements, weakly transcribed and heterochromatin regions (Figures 3E–3G). These results suggest that 2HG-mediated attenuation of the demethylation machinery results in genome-scale displacement of nucleosomes in chromatin regulatory modules, disarranging otherwise well-preserved and stable nucleosome positioning.

Because our RNA-seq analysis suggests that 2HG is sufficient to mediate transcriptional repression of genes that are highly methylated in breast cancer (Figure 1J), we sought to investigate whether 2HG can alter promoter accessibility of genes susceptible to DNA methylation in breast cancer. As expected, when a set of highly methylated genes ($n = 150$) in breast tumors (Huang et al., 2015) was examined, a marked reduction in chromatin accessibility was detected (Figures S3C and S3D). Similar results were obtained with genes that were independently found to be hypermethylated in promoter regions ($n = 150$) (Thienpont et al., 2016), indicating that 2HG exposure leads to reduced chromatin accessibility in gene promoters displaying DNA hypermethylation in breast cancer. In contrast, no significant accessibility changes were observed in sets of genes that are highly methylated in other tumor types, including endometrial cancer. These results suggest that 2HG-induced destabilization of the mammary epigenome could enhance cell fate plasticity by disrupting the chromatin regulatory landscape that confers epithelial cell identity.

Chromatin remodeling involves tumor-associated promoter hypermethylation

Given that 2HG imposes a loss of promoter accessibility of genes that are susceptible to DNA hypermethylation in breast cancer, we next sought to assess the effect on different molecular cancer subtypes. To this end, we analyzed RNA-seq profiles of 521 primary breast tumors and 112 adjacent uninvolved tissues from the TCGA cohort (Figure 4A). Upon unsupervised hierarchical clustering, the majority of highly methylated genes appeared to be downregulated in all five PAM50 molecular subtypes compared with control tissues, although no statistical significance was observed in luminal A tumors (Figure 4B). Transcriptional repression was most evident in basal-like breast cancers, 84% of which displayed a triple-negative phenotype. These findings are in agreement with earlier results showing that 2HG levels were elevated predominantly in ER α -negative breast cancer cells. In addition, none of the individuals with breast cancer examined had 2HG-producing mutations in the *IDH1* or *IDH2* gene (Figure 4A). Functional pathway analysis revealed enrichment of genes involved in DDR and hereditary breast cancer signaling (Figures 4C and S4A).

Next we leveraged the reported ATAC-seq accessibility data from 70 TCGA breast tumors (Corces et al., 2018) and found that hypermethylated genes with diminished expression showed a prominent decrease in chromatin accessibility in basal-like tumors (Figures 4D and 4E). To address whether the tumor-associated chromatin landscape could be attributable to 2HG-mediated epigenetic remodeling, we assessed changes in promoter accessibility in response to 2HG enantiomers. Compared with unexposed control cells, 2HG-exposed cells showed decreased or diffused chromatin accessibility at ChromHMM-defined active

promoter regions (Figure 4F). Additionally, whole-genome bisulfite sequencing (WGBS) data indicated that accessible promoter regions were essentially devoid of DNA methylation in control HMECs. These observations prompted us to explore the possibility that 2HG exposure may facilitate tumor-associated promoter hypermethylation.

To differentiate DNA methylation (5mC) from DNA hydroxymethylation (5hmC), we utilized oxidative bisulfite (oxBS) conversion followed by pyrosequencing. As expected, a marked gain in promoter methylation was shown in the majority of genes examined, including the *BRCA1*, *MSH2*, and *MLH1* genes involved in DDR pathways (Figure S4B). Promoter hypermethylation in response to 2HG was confirmed by methylated DNA immunoprecipitation followed by qPCR (MeDIP-qPCR) (Figure S4C). Similar but less pronounced increased DNA methylation was seen in immortalized, non-transformed mammary epithelial cells (hTERT-HME1) following 2HG exposure (Figure S4D). To validate the disease relevance of 2HG-mediated promoter hypermethylation, we analyzed the breast cancer methylomes profiled by methyl-CpG binding domain-based capture sequencing (MBDCap-seq) using our breast cancer cohort, which consisted of 77 primary tumors and 10 uninvolved tissues (Jadhav et al., 2015). A considerable increase in DNA methylation was detected at promoter regions in tumor samples in comparison with their normal counterparts (Figure S4E). These results complement the finding that a methylator phenotype is established in immortalized astrocytes when mutant IDH is ectopically expressed (Turcan et al., 2012). Collectively, the findings suggest that two prominent hallmarks of the cancer epigenome, promoter hypermethylation of tumor suppressor genes and genome-wide depletion of 5hmC, may be driven, at least in part, by defective metabolic fluxes in breast tumors.

2HG induces expression and epigenetic discordancy and enhances phenotypic diversity

To disentangle epigenetically heterogeneous responses to 2HG enantiomers, we performed high-dimensional mass cytometry on viably cryopreserved cell suspensions. The panel of metal-conjugated antibodies designed in this study was able to detect changes in DNA modifications as well as intracellular proteins (Figures S5A–S5C; Table S3). Consistent with our prior mass spectrometry measurements (Figures 1C and S1J), chromatin CyTOF analysis showed that 2HG led to a substantial increase in histone methylation, and the biaxial gating of live single cells indicated a concurrent increase in multiple distinct classes of histone modifications, which was reverted upon 2HG removal (Figures S5D and S5E). The two-dimensional projection using *t*-distributed stochastic neighbor embedding (*t*-SNE) revealed that control and withdrawal groups had similar multidimensional phenotypes and cell population distributions that were clearly distinguishable from those of 2HG-perturbed cells (Figure 5A), corroborating that the chromatin remodeling is essentially reversible and dependent on the presence of 2HG oncometabolites.

Self-organizing maps generated by FlowSOM identified six discrete clusters, and a large population of HMECs (cluster 1) displayed a concurrent increase in all histone markers examined following 2HG exposure (Figures 5B and 5C). We next wanted to determine whether 2HG-mediated chromatin remodeling could alter expression of DDR genes, whose promoter regions exhibited increased DNA methylation in response to 2HG. As expected,

2HG led to diminished expression of BRCA1, MSH2, and MLH1 (Figure 5D). BRCA1 expression was found to be inversely correlated with repressive histone marks on H4K20 residues (Figure S5F), and, to our surprise, we observed a significant association between the expression of BRCA1 and MSH2 genes (Figure S5G). The positive association between BRCA1 and MSH2 expression was also evident in CCLE cancer cell lines as well as in TCGA primary breast tumors (Figures S5H and S5I). Although the precise mechanism is currently unknown, the findings suggest that the DDR genes are controlled by a shared set of transcriptional regulators, possibly via α KG-dependent epigenetic mechanisms. Importantly, all three DDR genes exhibited increased variance in expression levels following 2HG exposure, as indicated by the wider distribution of data points, and these cell-level fluctuations were substantially restored after 2HG withdrawal (Figures 5D and S5G).

Noting that 2HG initiates genome-scale fluctuations in the chromatin regulatory landscape, we sought to determine the effect of 2HG on coordination of gene expression programs at single-cell resolution. Pairwise correlation measures between individual markers in untreated control cells revealed a highly concordant expression of DDR genes, all of which displayed a negative or low association with histone modification levels (Figure 5E). As indicated by reduced correlation between two variables, the coordinated levels of gene expression or histone modifications in single cells were impaired following 2HG exposure and restored upon removal of 2HG. This corrupted coordination of gene regulation supports cell-level epigenome fluctuations imposed by 2HG oncometabolites. Spanning tree progression of density normalized events (SPADE) analysis was performed to investigate changes in cell population structures, and we found that the distribution of cell subpopulations was more even in 2HG-treated groups, representing a relatively heterogeneous structure (Figure 5F). In agreement with this, compared with control and withdrawal groups, 2HG-perturbed cells displayed greater complexity, as indicated by higher values of Simpson and Shannon diversity indices (Risom et al., 2018; Almendro et al., 2014; Figure 5G). Additionally, downregulation of DDR and KRT8/18 luminal markers were particularly evident in cell populations displaying a higher variance in H3K9me3 marks, which are essentially involved in long-term transcriptional repression (Figure 5H). These data suggest that 2HG may undermine molecular mechanisms that mediate the coordinated control of epigenetic programs, thereby decreasing lineage fidelity.

2HG-high tumors display enhanced heterogeneity and an undifferentiated stem-like signature

Because our single-cell profiling using scATAC-seq and CyTOF suggests that 2HG increases cell-level variability in chromatin accessibility to gene regulatory regions and subverts lineage fidelity enhancing epithelial cell heterogeneity, we next set out to assess the effect of 2HG on varying heterogeneity in breast cancer cell lines as well as in primary tumors. Pairwise correlation of DNA methylation levels in CpG islands across the genome revealed that cell lines with high levels of intracellular 2HG had a higher degree of inter-sample variability in the DNA methylation landscape (Figures 6A and S6A). No association was detected between methylation variability and intracellular levels of α KG (Figure S6B). 2HG accumulation was seen in relation to increasing levels of histone methylation (Figure S6C), and DNA methylation heterogeneity inversely correlated with

elevated histone methylation on H3K27 and H3K36 residues (Figure S6D). In addition to DNA methylation heterogeneity, inter-sample variability in transcriptomic and proteomic landscapes was significantly correlated with intracellular 2HG levels (Figures 6A, S6E, and S6F). These results suggest that defective α KG-dependent demethylation could contribute to inter-sample diversity in breast cancer cell lines. To investigate the relationship between differentiation-state heterogeneity and 2HG accumulation, we examined the expression of mammary epithelial lineage markers (Risom et al., 2018), and high expression of luminal and basal/myoepithelial genes was found to be enriched in 2HG-high cell lines, suggesting that 2HG accumulation is one of the key variables (Figure 6B). In agreement with this, variance in gene expression between luminal and myoepithelial gene sets was reduced in 2HG-high cell lines (Figure 6C). To our surprise, 2HG levels had no association with genomic intratumor heterogeneity (gITH) scores (Carter et al., 2012; Figure 6D). These data suggest that 2HG could potentiate differentiation-state plasticity independent of genomic heterogeneity.

Genome-wide pairwise comparison further revealed that 2HG-high primary tumors had elevated levels of inter-sample variability in DNA methylation as well as transcriptomic landscapes (Figures 6E, 6F, S6G, and S6H). Quantitative characterization of cellular variability in transcriptomic landscapes based on the distribution of alternative splice-site usage (Kim et al., 2019) indicated that increasing levels of 2HG correlated with enhanced transcriptomic intratumor heterogeneity (tITH) scores, which predicted shorter survival (Figures 6G and S6I). Consistent with the results from cell lines, 2HG levels showed no direct correlation with gITH scores (Carter et al., 2012; Thorsson et al., 2018; Figure 6H). These data suggest that 2HG-mediated phenotypic heterogeneity could arise through epigenetic mechanisms at least partly independent of genomic alterations. Expression analysis of lineage-specific markers (Risom et al., 2018) demonstrated loss of luminal gene signatures in 2HG-high breast tumors (Figure S6J). Of note, although expression of individual luminal genes was concordantly regulated in 2HG-low tumors, as indicated by lower dispersion scores, 2HG-high tumors showed more variable, discordant expression of luminal genes, corroborating dysregulation of coordinated gene expression in 2HG-high tumors. Similar results were observed in 2HG-high versus 2HG-low breast cancer cell lines as well as in tITH-high versus tITH-low TNBC tumors (Figures S6K and S6L). To extend this observation, we analyzed expression of gene sets representing mature luminal and embryonic stem cell lineages (Lim et al., 2009; Ben-Porath et al., 2008) and found that loss of mature luminal gene expression in 2HG-high or tITH-high TNBC tumors was accompanied by a gain of undifferentiated stem-like transcriptional signatures (Figures 6I and 6J). Progression-free survival analysis demonstrated that loss of mature luminal and concurrent gain of undifferentiated stem-like transcriptional signatures were associated with a poor clinical outcome (Figure 6K). Furthermore, individuals with high DNA methylation in mature luminal and low methylation in stem-like genes had a steeper and prolonged drop in survival (Figure 6L), suggesting a contribution of epigenetic mechanisms to the underlying molecular pathology. These data support a model in which intratumoral 2HG enhances cellular heterogeneity by potentiating differentiation-state plasticity.

2HG induces DNA damage accumulation, and A2P eradicates cancer cell heterogeneity

To further elucidate the mechanisms that could contribute to breast tumorigenesis, we assessed the effect of 2HG on homologous recombination deficiency (HRD), which is enriched in breast tumors (Alexandrov et al., 2013; Ma et al., 2018). In addition to HMECs, low 2HG-producing MCF-7 cells were treated with 2HG, and double-strand breaks (DSBs) were detected by phosphorylation of histone H2AX (γ H2AX). As shown in Figures 7A and 7B, 2HG caused an increase in formation of γ H2AX foci in HMECs as well as in MCF-7 cells. Following 2HG withdrawal, there was a substantial decrease in H2AX phosphorylation (Figure 7C). These data indicate that 2HG impairs DNA repair mechanisms, leading to accumulation of DNA damage, and are in accordance with the observation that breast tumors with high 2HG accumulation exhibited increased rates of HRD (Thorsson et al., 2018; Figure 7D). Although epigenetic repression of DDR pathway genes has been reported to confer an HRD phenotype, the cause of this epigenetic alteration remains obscure (Turner, 2017). Our findings thus underscore a potential contribution of oncometabolite accumulation to DDR gene silencing and HRD-associated genome instability in breast tumors.

To gain more relevant insight into treatment of breast tumors with high levels of 2HG accumulation, we evaluated the antitumor effect of A2P supplementation. High 2HG-producing MDA-MB-231 cells were treated with 1 mM A2P for 72 h, and the expression of DDR pathway genes and cancer stem cell (CSC) markers was analyzed using CyTOF. As seen in Figures 7E and S7A, A2P-treated MDA-MB-231 cells exhibited elevated expression of HR repair genes, including BRCA1, RAD51C, MRE11, and XRCC2. In contrast, A2P suppressed expression of CSC-associated genes such as SOX2, NANOG, and CD44 (Figures 7E and S7B). Similar results were obtained when cells were treated with 100 μ M α KG for 72 h. Analysis of cell population structures using single-cell CyTOF profiling revealed that the distribution of cell subpopulations was less even in the A2P-treated group, representing a relatively homogeneous cell population (Figure 7F). Quantification of cellular diversity also demonstrated that A2P caused a significant reduction in the extent of cancer cell heterogeneity, and the results were again similar to α KG-treated cells (Figure 7G). Pairwise correlation analysis of expression levels in single cells indicated that A2P supplementation partially restored coordinated gene regulation (Figure 7H). Further, A2P inhibited cancer cell proliferation in the presence or absence of conventional treatment with tamoxifen (Figure S7C). These observations underscore the effect of A2P in eradicating tumor cell heterogeneity as well as in promoting drug sensitization and suggest that antagonizing 2HG may be a rational strategy to restrict cancer cell plasticity and counteract adaptive evolution of breast tumors.

DISCUSSION

In-depth multiregional tissue sampling and single-cell genome sequencing have revealed that intratumor heterogeneity occurs through evolution of genetically distinct subpopulations during malignant progression (Turajlic et al., 2019), which is potentially challenging to modulate using current technologies (DagogoJack and Shaw, 2018). In contrast, only a few studies have utilized single-cell epigenomic approaches to characterize the intricate

epigenetic makeup of cancer cells (Grosselin et al., 2019; Marjanovic et al., 2020; LaFave et al., 2020), highlighting the role of local chromatin signaling in gene regulation involved in treatment resistance and metastatic potential. In the present study, a combination of two single-cell epigenomic approaches has enabled us to demonstrate that global stability and plasticity of the epigenome are involved in enhancement of cell-to-cell variability, suggesting that breast cancer heterogeneity is mediated at least in part by metabolic reprogramming through epigenetic mechanisms.

In more general terms, our results show that defects in epigenetic machinery could promote cancer cell plasticity, potentially leading to generation of effectively neutral, metastable cell population, allowing non-physiological cell fate transitions. This, in turn, implies that α KG-dependent demethylation machinery is required for stable maintenance of cell identity as a safeguard of the epigenome. Considering the role of α KG-dependent demethylation in chromatin restriction, 2HG may expand cell fate plasticity by blurring the ridge line rather than by lowering the barrier heights of the basins in Waddington's epigenetic landscape. It is of interest in this context that constant active turnover of DNA modifications could contribute to development of cell-level epigenetic variability (Rulands et al., 2018; Parry et al., 2021). The histone H3K4 demethylase KDM5B is also linked with cell-level transcriptional heterogeneity (Hinohara et al., 2018), corroborating the role of chromatin homeostasis in preserving lineage fidelity. These findings suggest that altered metabolic fluxes could fuel the number of possible cellular states with higher adaptive potential via epigenetic mechanisms, which may enable drug-resistant tumor cells to emerge and expand. Although future well-controlled studies will be required to evaluate the quantitative effect of A2P on cancer cell plasticity, upfront or combinatorial treatment targeting α KG-dependent demethylation machinery, which can render tumor cells more homogeneous, may boost the efficacy of the existing high-cost interventions by suppressing the epigenetic mechanisms driving tumor cell evolution involved in therapy resistance (Dagogo-Jack and Shaw, 2018).

The present study suggests that 2HG-mediated inhibition of α KG-dependent demethylation induces high-plasticity chromatin landscape and lineage infidelity to promote epithelial cell heterogeneity. Epigenetic liabilities imposed upon 2HG exposure are found to be dynamic and essentially reversible, substantiating the inherent plasticity of the epigenome that is sensitive to cell-intrinsic and -extrinsic cues, such as metabolic intermediates, cofactors, environmental disruptors, and chromatin modifying compounds. Future studies that use single-cell epigenomic approaches enabling integrated assessment of cancer cell heterogeneity or plasticity, including chromatin mass cytometry as described here, may identify additional regulatory molecules involved in maintaining, disrupting, or restoring chromatin homeostasis, which is required for stable preservation of cell identity.

Limitations of the study

Our study highlights the role of 2HG in regulating epithelial lineage fidelity and DNA repair responses and suggests that altered metabolic fluxes in breast tumors could influence cellular plasticity through epigenetic as well as genetic mechanisms. A limitation of this study is that these experiments do not establish how epigenetic and genetic plasticity are linked. Of note, DNA damage induced by BRCA gene inactivation has been reported to induce luminal-to-

basal/mesenchymal transdifferentiation and result in basal-like tumorigenesis (Wang et al., 2016, 2019). Our findings thus raise the possibility that 2HG-induced DNA damage may likewise facilitate lineage reprogramming in non-BRCA-mutated breast tumors. Additional in-depth characterization of defects in DNA repair signaling is necessary to understand the underlying mechanisms. Another limitation is the lack of the study to evaluate the long-term effect of 2HG accumulation. Although the current study focuses on the role of 2HG in regulating the mammary epithelial epigenome to dissect upstream mechanisms that promote lineage infidelity and cancer cell heterogeneity, it is currently uncertain how enhanced plasticity imposed by the high-plasticity chromatin landscape confers a selective advantage or drug resistance during breast tumor evolution. Future studies in preclinical models will expand our mechanistic and translational understanding of the *in vivo* effect of oncometabolite accumulation and vitamin C supplementation.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kohzoh Mitsuya (mitsuya@uthscsa.edu).

Materials availability—This study did not generate new unique reagents.

Data and code availability

- Raw sequencing files of scATAC-seq and bulk RNA-seq data have been deposited at NCBI GEO (GSE161628) and are publicly available as of the date of publication. This paper also analyzes existing, publicly available data. Accession numbers are listed in the in Key Resources Table.
- There are no original codes generated in this paper.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture—Human breast cancer cell lines BT20, BT474, MCF-7, MDA-MB-157, MDA-MB-231 and MDA-MB-361, and non-malignant immortalized epithelial cells hTERT-HME1, 184B5 and MCF12A were acquired from the American Type Culture Collection (ATCC). Unless otherwise stated, cancer cell lines were maintained in DMEM (Gibco) supplemented with 10% fetal bovine serum (Sigma-Aldrich) and 100 U/ml penicillin plus 100 µg/mL streptomycin (Gibco) as previously reported (Hsu et al., 2013). Primary HMECs were obtained from Invitrogen and maintained at low passage number (below 5). HMECs, hTERT-HME1, 184B5 and MCF12A cells were cultured in mammary epithelial growth medium according to the manufacturer's instructions. Authentication of cell line genomic DNA was performed at ATCC using DNA fingerprint analysis of polymorphic, short tandem repeat sequences. Exposure to cell-permeable 2HG analogues was carried out by supplementing octyl esters of *R*-2-hydroxyglutarate or *S*-2-hydroxyglutarate (Cayman

Chemical) to the culture medium at a final concentration 72 h before harvesting. Vitamin C supplementation was carried out by adding ascorbate-2-phosphate (A2P, Sigma-Aldrich) to the culture medium at the pharmacological concentration of 5 mM. Dimethylxalylglycine (DMOG) and tamoxifen (4-hydroxytamoxifen) were obtained from Sigma-Aldrich. For the withdrawal study, cells were initially exposed to L2HG for 72 h and then cultured in regular medium for additional 4 days.

METHOD DETAILS

Metabolite extraction and quantification by LC-MS—Following dissociation, cells were washed twice with ice-cold phosphate-buffered saline (PBS) and cell pellets were flash-frozen on dry ice. For α KG analysis, metabolites were extracted with 80:20 methanol: water (-80°C) containing stable isotope-labeled internal standard [1,2,3,4- $^{13}\text{C}_4$] α -ketoglutaric acid (Cambridge Isotope Laboratories) and incubated at -80°C for 1 h as described previously (Lin et al., 2015). Extracts were then centrifuged at 13,800g for 10 min and supernatants were transferred to glass autosampler vials for high-performance liquid chromatography-electrospray ionization-mass spectrometry (HPLC-ESI-MS) measurements. For 2HG analysis, cells were processed as mentioned above except that [1,2,3,4- $^{13}\text{C}_4$]L-malic acid (Cambridge Isotope Laboratories) was added as an internal standard and dried extracts were derivatized with diacetyl-L-tartaric anhydride (DATAN, Sigma-Aldrich). HPLC-ESI-MS detection was conducted on a ThermoFisher Q Exactive mass spectrometer with on-line separation by a ThermoFisher Dionex Ultimate 3000 HPLC. HPLC conditions for α KG analysis were: column, Synergi Polar-RP, 4 μm , 2 \times 150 mm (Phenomenex); mobile phase A, 0.1% formic acid in water; mobile phase B, 0.1% formic acid in acetonitrile; flow rate, 250 $\mu\text{L}/\text{min}$; gradient, 1% B to 5% B over 5 min, 5% B to 95% B over 1 min and held at 95% B for 2 min. HPLC conditions for 2HG analysis were: column, Luna NH2, 3 μm , 2 \times 150 mm (Phenomenex); mobile phase A, 5% acetonitrile in water containing 20 mM ammonium acetate and 20 mM ammonium hydroxide, pH 9.45; mobile phase B, acetonitrile; flow rate, 300 $\mu\text{L}/\text{min}$; gradient, 85% B to 1% B over 10 min and held at 1% B for 10 min. The conditions used to selectively quantify D2HG and L2HG were: column, Kinetex C18, 2.6 μm , 2.1 \times 100 mm (Phenomenex); mobile phase, 1% acetonitrile with 125 mg/L ammonium formate, pH 3.6; flow rate, 400 $\mu\text{L}/\text{min}$. Full scan mass spectra were acquired in the orbitrap using negative ion detection over a range of m/z 100–800 at 70,000 resolution (m/z 300). Metabolite identification was based on accurate mass match to the library \pm 5 ppm and agreement with the HPLC retention time of authentic standards. Quantification of metabolites was carried out by integration of extracted ion chromatograms with the corresponding standard curves.

Immunofluorescence staining—Cells were plated in 8-well chamber slides (Falcon) at a density of $1\text{--}2 \times 10^4$ cells/well at least 24 h prior to 2HG exposure. Cells were then fixed with 4% paraformaldehyde (PFA) in PBS for 10 min at room temperature (RT) and permeabilized with 0.2% Triton X-100 in PBS for 10 min at RT. For 5hmC staining, permeabilized cells were treated with 2N HCl for 30 min at RT and neutralized with 100 mM Tris-HCl (pH 8.5). Nonspecific binding was blocked with 10% goat serum in 0.2% Triton X-100 and PBS for 1 h at RT and stained with primary antibodies in PBS with 5% goat serum and 0.2% Triton X-100 overnight at 4°C . After incubation with Alexa

Fluor-conjugated secondary antibodies (Molecular Probes) for 1 h at RT, nuclei were stained for 5 min with DAPI (ThermoFisher). Single optical sections were acquired using a Zeiss LSM710 confocal microscope and image quantification was performed with NIH ImageJ software (version 1.52n). Primary antibodies included rabbit polyclonal anti-5hmC (Active Motif, 39769, 1:500), rabbit monoclonal anti-H3K27me3 (Cell Signaling Technologies, 9733, 1:800) and mouse monoclonal anti-phospho-histone H2A.X (Ser139) (Millipore, 05-636, 1:200).

Tet-assisted bisulfite (TAB) pyrosequencing—TAB pyrosequencing was used to differentiate 5hmC from 5mC (Yu et al., 2012). High molecular weight genomic DNA was extracted using Genra Puregene reagents (QIAGEN), followed by an additional ethanol precipitation and resuspension in low-EDTA TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). RNase A and proteinase K digestion were included in the isolation procedure. UV absorbance was measured on a NanoDrop 2000 (ThermoFisher) and each DNA sample was routinely examined by agarose gel electrophoresis with GelRed staining to ensure the absence of contaminating RNA and degradation of genomic DNA. Isolated genomic DNA was then subjected to Tet-assisted bisulfite (TAB) treatment as we previously described (Mitsuya et al., 2017). After bisulfite conversion using EpiTect Fast Bisulfite Conversion kit (QIAGEN), pyrosequencing was conducted on a PyroMark Q96 MD instrument using CpG LINE-1 assay (QIAGEN, 973043). To monitor bisulfite conversion efficiency, a C outside a CpG site was added within dispensation order for the sequence to be analyzed as a built-in control. The quantitative levels of 5mC and 5hmC for each CpG dinucleotide were determined using PyroMark CpG software (version 1.0, QIAGEN).

Multiplexed chromatin profiling by mass spectrometry—Nuclei were isolated from 2×10^6 cells using Nuclear Isolation Buffer (NIB) composed of 15 mM Tris-HCl (pH 7.5), 60 mM KCl, 15 mM NaCl, 5 mM MgCl₂, 1 mM CaCl₂, 250 mM sucrose, 0.3% NP-40, 1 mM DTT plus 10 mM sodium butyrate added immediately prior to use, for 30 min on ice. Nuclei were pelleted at 600 *g* for 5 min at 4°C and detergent was removed by washing twice with NIB without NP-40. Histones from isolated nuclei were acid extracted with 5 volumes of 0.2 M H₂SO₄ for 1 h at RT. Cellular debris was removed by centrifugation at 4,000 *g* for 5 min. Trichloroacetic acid was added to the supernatant at a final concentration of 20% (v/v) and incubated for 1 h to precipitate histone proteins. Histones were pelleted at 10,000 *g* for 5 min, washed once with 0.1% HCl in acetone, twice with 100% acetone followed by centrifugation at 15,000 *g* for 5 min, and then briefly air-dried. Histones were derivatized, digested and analyzed by targeted LC-MS/MS as described previously (Diebold et al., 2019).

Flow cytometry—Cell cycle phase distribution was analyzed by flow cytometry. Cells were fixed with ice-cold 70% methanol for 1 h on ice. Following centrifugation, cells were washed with PBS and stained with 10 µg/mL propidium iodide (Sigma-Aldrich) solution in PBS containing 1 µg/mL DNase-free RNase A (ThermoFisher) and incubated in the dark on ice for 1 h. Samples were then processed on a BD FACS-Calibur flow cytometer equipped with CellQuest Pro software (version 5.2.1, Becton Dickinson) and data were analyzed using FlowJo software (version 7.6.5, Tree Star).

Capillary-based Western immunoassay—Cells were washed with ice-cold PBS and lysed on ice for 30 min with lysis buffer containing 20 mM Tris-HCl, 1% NP-40, 150 mM NaCl, 10% glycerol, and protease inhibitor cocktail (Thermo Fisher Scientific). After clearing by centrifugation, protein concentration was determined by Bradford protein assay kit (Pierce). Protein separation and detection were performed using an automated capillary electrophoresis system and 12–230 kDa separation module (ProteinSimple). Anti-rabbit or anti-mouse detection module according to the manufacturer’s instructions. Primary antibodies used were as follows: PNCA (Cell Signaling Technologies, 13110), BRCA1 (Cell Signaling Technologies, 9010), RAD51C (Novus Biologicals, NB100–177), BRIP1 (Novus Biologicals, NBP1–31883), MSH2 (Novus Biologicals, NBP3–07211), FANCD2 (Novus Biologicals, NB100–182), FEN1 (Novus Biologicals, NB100–150), XRCC2 (Abcam, ab180752), and β -actin (R&D Systems, MAB8929). Data were analyzed and displayed in Compass for SW (Version 3.1.7).

Live cell proliferation assay—Cell growth was evaluated using the IncuCyte ZOOM live-cell analysis system (Essen Bioscience). Cells were seeded in 96-well plates at a density of 1,000–2,000 cells per well. Approximately 24 h later, 96-well plates were placed in IncuCyte instrument and images were taken every 12 h over the course of five days. Growth curves were generated with quadruplicate replicates. The real-time kinetic data of proliferation was calculated as phase object confluence and analyzed by IncuCyte software (version 2016B).

RNA isolation and library preparation—Cells were collected, lysed and homogenized in TRI reagent (Zymo Research). Total RNA was extracted in triplicates from each condition according to Direct-zol RNA miniprep plus (Zymo Research) protocol. The quantity and quality of isolated RNA were assessed by NanoDrop (2000) (ThermoFisher) and Agilent 5200 Fragment Analyzer system (Agilent). RNA quality scores ranged from 9.2 to 10.0. Sequencing libraries were prepared with 500 ng of total RNA using Zymo RiboFree Total RNA Library Kit (Zymo Research). Following first strand cDNA synthesis, rRNA depletion was carried out to enrich mRNA as per the protocol. A second strand of DNA was synthesized and adapter ligation was completed to obtain dual-indexed sequencing libraries. ERCC RNA spike-in control mix (Ambion) was also added to each sample (1:10 dilution). The concentration and quality of libraries were evaluated using Qubit dsDNA assay on Qubit 2.0 fluorometer (Invitrogen) and High Sensitivity DNA Kit on Agilent 2100 Bioanalyzer system (Agilent).

RNA sequencing and data analysis—RNA-seq libraries were sequenced on HiSeq 3000 platform using 50-bp, single-read Illumina chemistry with a coverage depth of 25–30 million reads per sample. Raw sequence reads were checked for initial quality control using FastQC tools (version 0.11.9) and summarized using MultiQC tool (version 0.8). Adapter sequences were trimmed using Cutadapt (version 1.18). The trimmed reads were then aligned to the human reference genome (GRCh38/hg38) using STAR (Dobin et al., 2013) alignment tool (version 2.7.5a) with standard input parameters. Transcript counts were normalized and differentially expressed genes were estimated by using DESeq2 package

(version 1.26.0). A log₂ fold change of 1 and p value < 0.05 were considered for differential gene expression analysis.

Gene set enrichment analysis (GSEA)—Gene set enrichment analysis (GSEA) was evaluated using the GSEA (Liberzon et al., 2015) software (version 4.1.0). The normalized read matrix generated using DESeq2 was used as input. A catalog of annotated genesets from Molecular Signature Database (MSigDB, version 7.2) was used for enrichment analysis. For enrichment of DNA repair genes, the geneset was created based on published signature genes (Knijnenburg et al., 2018). False discovery rate (FDR) was assessed using 1,000 random permutations of the geneset with a signal-to-noise ratio for ranking genes. A FDR-corrected value of $q < 0.1$ was considered to be statistically significant.

Single-cell ATAC-seq library preparation—Single-cell ATAC-seq libraries were prepared on a Fluidigm C1 workstation using ‘ATAC SeqCell Load and Stain Rev C’ script as previously described (Buenrostro et al., 2015) with modifications. Briefly, cells were passed through a 20 μ m cell strainer (CellTrics, Sysmex) to remove debris and remaining cell aggregates, and mixed at a ratio of 7:3 with C1 suspension reagent. The resulting single-cell suspension was loaded on C1 Single-Cell Open App IFC chip (1862x, 10–17 μ m, Fluidigm) at a concentration of 350 cells/ μ l. Captured cells were stained with 2 μ M green-fluorescent calcein-AM and 4 μ M red-fluorescent ethidium homodimer-1 (Invitrogen) and visualized under an EVOS FL cell imaging station (Life Technologies) to ensure successful capture and to determine cell viability. The single-cell capture rates were typically >80%, and >90% of captured single cells were alive. After cell lysis and Tn5 transposition, 8 cycles of pre-amplification were run on IFC chip. Pre-amplified PCR products were transferred to 96-well plates and further amplified for an additional 13 cycles using custom Nextera dual-index primers and NEBNext High-Fidelity 2X PCR master mix (New England Biolabs). Individually barcoded libraries were pooled and purified on a single MinElute column (QIAGEN). The quality and size distribution of pooled libraries were evaluated on an Agilent 2100 Bioanalyzer using High Sensitivity DNA reagents (Agilent).

Single-cell ATAC-seq data analysis—scATAC-seq libraries were sequenced on a NextSeq 500 platform with High Output reagents (Illumina) using paired-end 75-bp reads. All scATAC-seq data were preprocessed as essentially described (Buenrostro et al., 2015). In short, adapter and primer sequences were trimmed and initial quality control checks were performed using FastQC tools (version 0.11.9). Sequencing reads were aligned to the GRCh37/hg19 assembly of the human genome using Bowtie2 with the parameter ‘-X2000’ to ensure paired reads were within 2 kb of one another. PCR duplicates were eliminated using Picard tools (version 2.9.2, <http://broadinstitute.github.io/picard/>) and alignments with mapping quality less than 30 were subsequently removed by samtools. Reads mapped to the mitochondria and unmapped contigs were filtered out and excluded from further analysis. PCA projections of scATAC-seq profiles were performed using SCRAT (Ji et al., 2017), and gene feature was applied to aggregate sequencing reads from each cell, in which 3,000 bp upstream to 1,000 bp downstream of TSS is regarded as the region of interest for each gene. After aggregation, the signals for each feature were normalized to adjust for library size and model-based clustering (mclust) module was utilized to identify cell subpopulations. Peak

calling was performed using MACS2 with the following settings: `-nomodel -nolambda -keep-dup all -call-summits`. Artifact signals were excluded using ENCODE blacklist. Bias-corrected accessibility deviations at transcriptional factor (TF) motifs were assessed using chromVAR (Schep et al., 2017) (version 3.12) with JASPAR's motif set.

Breast cancer cohorts, resources and data analysis—Level 3 TCGA Breast Invasive Carcinoma (BRCA) data of tumor and normal samples were accessed from the Broad GDAC Firehose (<http://gdac.broadinstitute.org>) and RSEM-normalized RNA-seq values were log₂ transformed before analysis. Unsupervised hierarchical clustering was utilized to distinguish mRNA expression profiles among different genes and heatmaps were generated using heatmap.2 function implemented in gplots package of R statistical program. Clinical data including PAM50 intrinsic subtypes, ER/PR/HER2 expression and *IDH* mutation status were retrieved using the Cancer Genomics cBioPortal (<https://www.cbioportal.org/>) and were integrated into RNA-seq heatmap. TCGA DNA methylation data generated using Infinium Human Methylation 450K (HM450K) BeadChip array were retrieved from the cBioPortal repository. Normalized methylation scores at each CpG dinucleotide are expressed as β values, representing a continuous measurement from 0 (completely unmethylated) to 1 (completely methylated). In the event of multiple CpG probes per gene, the most negatively correlated with mRNA expression was selected. Chromatin accessibility data of TCGA primary tumor tissue samples were extracted from the UCSC Xena browser (<https://xenabrowser.net/>). After Z-scale normalization of ATAC-seq signals, open chromatin occupancies at promoter regions were correlated with PAM50 gene signature to evaluate DNA accessibility profiles across breast cancer subtypes. Our breast cancer methylome data generated using MBDCap-seq are available at The Cancer Methylome System (<http://cbbiweb.uthscsa.edu/KMethylomes/>). Global chromatin profiling, metabolomics and high-throughput sequencing datasets for breast cancer cell lines were retrieved from the Broad Institute CCLE repository (<https://portals.broadinstitute.org/ccle>). Homologous recombination deficiency (HRD) scores for TCGA patients were obtained from Thorsson et al. (Thorsson et al., 2018). Briefly, HRD score was quantitatively assessed by taking a sum of 3 separate metrics of genomic scarring: loss of heterozygosity (LOH), large-scale state transitions, and subtelomeric regions with allelic imbalance.

Oxidative bisulfite (oxBS) pyrosequencing—To selectively detect 5mC modification, genomic DNA was subjected to oxBS conversion (Booth et al., 2012) using TrueMethyl oxBS module (NuGEN Technologies) as per the manufacturer's recommendations. In short, genomic DNA was affinity-purified using 80% acetonitrile (Fisher Scientific) and TrueMethyl magnetic beads to eliminate potential contaminating compounds. After the denaturation step, genomic DNA was oxidized to convert 5-hydroxymethylcytosine to 5-formylcytosine. Bisulfite treatment was then carried out to convert 5-formylcytosine to uracil, leaving 5-methylcytosine intact. Following desulfonation and purification, converted DNA was quantified using Qubit ssDNA assay (Invitrogen). PCR amplification of oxBS converted DNA was carried out with biotin-labeled primers. Primer design was carried out using PyroMark Assay Design software (version 2.0, QIAGEN). Pyrosequencing of biotinylated PCR products was performed using PyroMark Q48 Advanced CpG reagents (QIAGEN) on a PyroMark Q48 Autoprep apparatus (-QIAGEN) following the

manufacturer's protocol. 5mC levels at CpG sites were determined using PyroMark Q48 Autoprep software (version 2.4.2, QIAGEN) in CpG Assay mode. All samples were prepared, amplified and sequenced in triplicates. PCR and pyrosequencing primers are listed in Table S2.

Methylated DNA immunoprecipitation qPCR (MeDIP-qPCR)—Prior to the 5mC immune-capture procedure, genomic DNA was fragmented to an average length of 200–600 bp using a Covaris S220 system (Covaris). MeDIP was performed using MeDIP reagents (Active Motif) as per the manufacturer's instructions. In brief, fragmented DNA was heat-denatured and immunoprecipitated with anti-5mC antibody (39649, Active Motif). An additional quantity of fragmented DNA equivalent to 10% of DNA being used in the immunoprecipitation reaction was also denatured and saved as input DNA. Immunoprecipitated DNA and input DNA were then purified with phenol/chloroform extraction and amplified using GenoMatrix Whole Genome Amplification kit (Active Motif). Quantitative PCR was performed using PowerUP SYBR Green master mix on an ABI StepOnePlus real-time PCR instrument (Applied Biosystems). All PCR reactions were run in triplicates. The relative enrichment of target sequences after MeDIP was evaluated by calculating the ratios of the signals in immunoprecipitated DNA versus input DNA. Locus-specific primers were designed with NCBI Primer-BLAST and synthesized by Integrated DNA Technologies. Primer sequences are provided in Table S2.

Panel design and heavy-metal labeling of antibodies—Prior to antibody conjugation, the antibody panel was designed by allocating targets to specific heavy-metal isotopes depending on the sensitivity of the mass cytometer, e.g., assigning low abundance targets to high sensitivity channels in order to minimize potential spectral overlap (Takahashi et al., 2017). Subsequently, in-house conjugation of antibodies was performed using Maxpar X8 antibody labeling reagents (Fluidigm) with some modifications. Briefly, up to 100 µg of carrier-free IgG antibody was subjected to buffer exchange by washing with R-buffer using a 50 kDa Amicon filter (Millipore) that was pre-soaked with R-buffer. Antibodies were then partially reduced with 4mM TCEP (ThermoFisher) for 30 min at RT followed by washing with C-buffer. In parallel, metal chelation was carried out by adding lanthanide metal solutions (Fluidigm) to chelating polymers (Fluidigm) in L-buffer. Metal-loaded polymers were then washed with L-buffer and concentrated on a 3 kDa Amicon filter (Millipore). Partially reduced antibodies were incubated with metal-loaded polymers for 90 min at RT followed by washing with W-buffer. Following conjugation, antibody concentration was determined by spectrometry with a NanoDrop 2000 (ThermoFisher). Metal-conjugated antibodies were stored in antibody stabilization solution (Candor Bioscience) supplemented with 0.05% sodium azide at 4°C. The panel of metal-conjugated antibodies is provided in Table S3.

Multidimensional chromatin profiling by mass cytometry—Cell suspensions were prepared at a concentration of 1×10^7 cells/ml in serum-free, protein-free medium and stained with 1 µM cisplatin (195Pt) for 5 min at RT to determine cell viability. After quenching with CyTOF buffer composed of PBS with 1% BSA (Invitrogen), 2mM EDTA (Ambion) and 0.05% sodium azide (Teknova), staining with lanthanide-conjugated

antibodies was performed as previously described (Cheung et al., 2018), but with the following modifications. In brief, following extracellular marker staining, cells were fixed with 1.6% PFA (Electron Microscopy Sciences) for 15 min at RT and permeabilized with ice-cold methanol (Fisher Scientific) for 30 min at 4°C. For 5hmC and 5mC staining, cells were treated with 2N HCl for 30 min at RT followed by neutralization with 100 mM Tris-HCl buffer (pH 8.5). After adding Fc receptor blocker (BioLegend), cells were labeled overnight at 4°C with a cocktail of antibodies recognizing chromatin modifications or intracellular components. On the next day, excess of antibodies was washed off with CyTOF buffer and cells were stained with 250 nM 191/193Ir-containing DNA intercalator (Fluidigm) in PBS with 1.6% PFA for 30 min at RT. After resuspending in double-deionized water, samples were kept on ice. Immediately prior to acquisition, cells were prepared at a concentration of $0.2\text{--}1.0 \times 10^6$ cells/ml in 0.1X EQ bead solution containing four element calibration beads (Fluidigm) and filtered through a 20 μm cell strainer (CellTrics, Sysmex) to remove any potential aggregates. Cells were then acquired at a rate of 300–500 events/s using a Helios mass cytometer (Fluidigm) and CyTOF software (version 6.7) with noise reduction, a lower convolution threshold of 400, event length limits of 10–150 pushes, a sigma value of 3 and a flow rate of 0.030 mL/min.

Mass cytometry data analysis—Data analysis was conducted using the cloud-based platform Cytobank (<https://www.cytobank.org/>) and the statistical programming environment R. Following data acquisition, mass cytometry data were normalized using EQ calibration beads. Bead-normalized data were then uploaded onto Cytobank platform to carry out initial gating and population identification using the indicated gating schemes (Figure S5B). For downstream analysis, live single cells were identified based on 140Ce bead, event length, DNA content (191Ir) and live/dead (195Pt) channels. Histograms and two-dimensional contour plots were generated to assess the global levels of chromatin modifications across the samples. Using an equal number of randomly selected live singlets from each sample, dimensionality reduction was implemented by *t*-SNE analysis with the following settings: perplexity = 60, theta = 0.5, iteration = 1,000. FlowSOM clustering was carried out on the same data using the standard parameters to quantify changes in cell subsets in an unbiased manner. The 2D coordinates of the *t*-SNE map were fed to FlowSOM analysis for population identification based on hierarchical consensus clustering. Comparisons of chromatin modifications among the samples in each cluster were performed by generating heatmaps in R using gplots package and median signal intensities extracted from Cytobank. Spanning-tree Progression Analysis of Density-normalized Events (SPADE) analysis was performed to construct minimum spanning trees (MST) using live singlets from each treatment conditions as described in Qiu et al. (Qiu et al., 2011). SPADE clustering was conducted to generate unified trees based on the expression of 16 markers using the following main parameters: percent downsampling = 10% and number of clusters = 40. Cellular abundance of each node, and intensity and co-efficient of variation of each marker in every node were exported for further analysis.

Calculating heterogeneity—Shannon and Simpson diversity indices (Risom et al., 2018; Almendro et al., 2014) were used as metrics of phenotypic heterogeneity. The frequency distribution of events/cells across the clusters were calculated from SPADE analysis as

described above. The mean diversity indices for each condition were used to compare the degree of heterogeneity induced by oncometabolite exposure.

Single sample gene set scoring—Single sample gene set scoring was performed using the R-package *Singscore* (version 3.12) as described previously (Foroutan et al., 2018). This method utilizes rank-based statistics and generates scores for expression activities of genesets at a single-sample level. A high geneset score indicates that the gene expression pattern in a sample is concordant with the pattern exhibited by the gene-expression signature. Genesets were extracted and compiled from a study examining cell-state heterogeneity in basal-like breast cancer (Risom et al., 2018), a study analyzing expression profiles of tumor cells with regards to embryonic stem cell identity (Ben-Porath et al., 2008), and a study examining preneoplastic and normal mammary tissue in breast tumor development (Lim et al., 2009).

QUANTIFICATION AND STATISTICAL ANALYSIS

Pairwise comparisons were carried out with a two-tailed unpaired Student's t test and multiple comparisons were assessed using a one-way ANOVA followed by Dunnett's multiple comparison post-hoc test unless otherwise indicated in the figure legends. Correlation analyses were performed using Spearman's rank-based coefficient. For Kaplan-Meier survival analysis, expression or methylation values were classified as high or low by using the median as a cutoff value and progression-free survival data was used to measure prognosis. Log-rank (Mantel-Cox) test was used to evaluate statistical differences and hazard ratio was reported with 95% confidence interval. Statistical analyses were performed using GraphPad Prism program (version 8.1). For all statistical analyses, differences of $p < 0.05$ were considered statistically significant. All quantitative data are presented as mean \pm SEM unless specified otherwise.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank all laboratory members for helpful discussions and technical assistance. We are grateful to the BioAnalytics and Single-Cell Core (BASiC) for single-cell analysis, the Mass Spectrometry Core for metabolite mass spectrometry, the Optical Imaging Facility for confocal imaging, the Flow Cytometry Shared Resource Facility at the University of Texas Health Science Center at San Antonio for flow cytometry, and the Northwestern Proteomics Core Facility for proteomics analyses. Sequencing data were generated in the Genome Sequencing Facility, which is supported by UT Health San Antonio, NIH-NCI P30 CA054174 (Mays Cancer Center at UT Health San Antonio), NIH Shared Instrument grant 1S10OD021805-01 (S10 grant), and Cancer Prevention and Research Institute of Texas (CPRIT) Core Facility Award RP160732. This study was supported by NIH grants U54 CA217297 and P30 CA054174 and CPRIT grant RP150600. J.R. acknowledges funding from the US National Science Foundation (ABI 1565076). M.K. is a recipient of the CPRIT predoctoral fellowship (RP170345).

REFERENCES

Ahmad K, and Henikoff S (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* 9, 1191–1200. [PubMed: 12086617]

- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421. [PubMed: 23945592]
- Allis CD, and Jenuwein T (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet* 17, 487–500. [PubMed: 27346641]
- Almendo V, Kim HJ, Cheng YK, Gonen M, Itzkovitz S, Argani P, van Oudenaarden A, Sukumar S, Michor F, and Polyak K (2014). Genetic and phenotypic diversity in breast tumor metastases. *Cancer Res.* 74, 1338–1348. [PubMed: 24448237]
- Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, and Weinberg RA (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet* 40, 499–507. [PubMed: 18443585]
- Bi M, Zhang Z, Jiang YZ, Xue P, Wang H, Lai Z, Fu X, De Angelis C, Gong Y, Gao Z, et al. (2020). Enhancer reprogramming driven by high-order assemblies of transcription factors promotes phenotypic plasticity and breast cancer endocrine resistance. *Nat. Cell Biol* 22, 701–715. [PubMed: 32424275]
- Blaschke K, Ebata KT, Karimi MM, Zepeda-Martinez JA, Goyal P, Mahapatra S, Tam A, Laird DJ, Hirst M, Rao A, et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 500, 222–226. [PubMed: 23812591]
- Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, and Balasubramanian S (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336, 934–937. [PubMed: 22539555]
- Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, and Greenleaf WJ (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490. [PubMed: 26083756]
- Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol* 30, 413–421. [PubMed: 22544022]
- Cheung P, Vallania F, Warsinske HC, Donato M, Schaffert S, Chang SE, Dvorak M, Dekker CL, Davis MM, Utz PJ, et al. (2018). Single-cell chromatin modification profiling reveals increased epigenetic variations with aging. *Cell* 173, 1385–1397. [PubMed: 29706550]
- Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, Silva TC, Groeneveld C, Wong CK, Cho SW, et al. (2018). The chromatin accessibility landscape of primary human cancers. *Science* 362, eaav1898. [PubMed: 30361341]
- Dagogo-Jack I, and Shaw AT (2018). Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15, 81–94. [PubMed: 29115304]
- Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, Fantin VR, Jang HG, Jin S, Keenan MC, et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739–744. [PubMed: 19935646]
- Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, et al. (2018). The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* 46, D794–D801. [PubMed: 29126249]
- Diebold LP, Gil HJ, Gao P, Martinez CA, Weinberg SE, and Chandel NS (2019). Mitochondrial complex III is necessary for endothelial cell proliferation during angiogenesis. *Nat. Metab* 1, 158–171. [PubMed: 31106291]
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. [PubMed: 23104886]
- Dravis C, Chung CY, Lytle NK, Herrera-Valdez J, Luna G, Trejo CL, Reya T, and Wahl GM (2018). Epigenetic and transcriptomic profiling of mammary gland development and tumor models disclose regulators of cell state plasticity. *Cancer Cell* 34, 466–482. [PubMed: 30174241]
- Ernst J, and Kellis M (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc* 12, 2478–2492. [PubMed: 29120462]

- Ernst J, Kheradpour P, Mikkelson TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49. [PubMed: 21441907]
- Ewels P, Magnusson M, Lundin S, and Kaller M (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. [PubMed: 27312411]
- Faubert B, Solmonson A, and DeBerardinis RJ (2020). Metabolic reprogramming and cancer progression. *Science* 368, eaaw5473. [PubMed: 32273439]
- Feinberg AP, Koldobskiy MA, and Gondor A (2016). Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat. Rev. Genet* 17, 284–299. [PubMed: 26972587]
- Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567. [PubMed: 21130701]
- Flavahan WA, Drier Y, Liao BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suva ML, and Bernstein BE (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110–114. [PubMed: 26700815]
- Flavahan WA, Gaskell E, and Bernstein BE (2017). Epigenetic plasticity and the hallmarks of cancer. *Science* 357, eaal2380. [PubMed: 28729483]
- Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, and Davis MJ (2018). Single sample scoring of molecular phenotypes. *BMC Bioinformatics* 19, 404. [PubMed: 30400809]
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* 6, p11. [PubMed: 23550210]
- Garrido-Castro AC, Lin NU, and Polyak K (2019). Insights into molecular classifications of triple-negative breast cancer: Improving patient selection for treatment. *Cancer Discov.* 9, 176–198. [PubMed: 30679171]
- Grosselin K, Durand A, Marsolier J, Poitou A, Marangoni E, Nemati F, Dahmani A, Lameiras S, Reyat F, Frenoy O, et al. (2019). High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet* 51, 1060–1066. [PubMed: 31152164]
- Hanahan D, and Weinberg RA (2011). Hallmarks of cancer: The next generation. *Cell* 144, 646–674. [PubMed: 21376230]
- Hinohara K, and Polyak K (2019). Intratumoral heterogeneity: More than just mutations. *Trends Cell Biol* 29, 569–579. [PubMed: 30987806]
- Hinohara K, Wu HJ, Vigneau S, McDonald TO, Igarashi KJ, Yamamoto KN, Madsen T, Fassl A, Egri SB, Papanastasiou M, et al. (2018). KDM5 histone demethylase activity links cellular transcriptomic heterogeneity to therapeutic resistance. *Cancer Cell* 34, 939–953. [PubMed: 30472020]
- Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, and Young RA (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947. [PubMed: 24119843]
- Hsu PY, Hsu HK, Lan X, Juan L, Yan PS, Labanowska J, Heerema N, Hsiao TH, Chiu YC, Chen Y, et al. (2013). Amplification of distant estrogen response elements deregulates target genes associated with tamoxifen resistance in breast cancer. *Cancer Cell* 24, 197–212. [PubMed: 23948299]
- Huang S (2021). Reconciling non-genetic plasticity with somatic evolution in cancer. *Trends Cancer* 7, 309–322. [PubMed: 33536158]
- Huang WY, Hsu SD, Huang HY, Sun YM, Chou CH, Weng SL, and Huang HD (2015). MethHC: A database of DNA methylation and gene expression in human cancer. *Nucleic Acids Res.* 43, D856–D861. [PubMed: 25398901]
- Inoue S, Li WY, Tseng A, Beerman I, Elia AJ, Bendall SC, Lemonnier F, Kron KJ, Cescon DW, Hao Z, et al. (2016). Mutant IDH1 downregulates ATM and alters DNA repair and sensitivity to DNA damage independent of TET2. *Cancer Cell* 30, 337–348. [PubMed: 27424808]
- Intlekofer AM, Shih AH, Wang B, Nazir A, Rustenburg AS, Albanese SK, Patel M, Famulare C, Correa FM, Takemoto N, et al. (2018). Acquired resistance to IDH inhibition through trans or cis dimer-interface mutations. *Nature* 559, 125–129. [PubMed: 29950729]

- Jackson HW, Fischer JR, Zanotelli VRT, Ali HR, Mechera R, Soysal SD, Moch H, Muenst S, Varga Z, Weber WP, et al. (2020). The single-cell pathology landscape of breast cancer. *Nature* 578, 615–620. [PubMed: 31959985]
- Jadhav RR, Ye Z, Huang RL, Liu J, Hsu PY, Huang YW, Rangel LB, Lai HC, Roa JC, Kirma NB, et al. (2015). Genome-wide DNA methylation analysis reveals estrogen-mediated epigenetic repression of metallothionein-1 gene cluster in breast cancer. *Clin. Epigenetics* 7, 13. [PubMed: 25763113]
- Jezek P (2020). 2-Hydroxyglutarate in cancer cells. *Antioxid. Redox Signal* 33, 903–926. [PubMed: 31847543]
- Ji Z, Zhou W, and Ji H (2017). Single-cell regulome data analysis by SCRAT. *Bioinformatics* 33, 2930–2932. [PubMed: 28505247]
- Jiang Y, Qian F, Bai X, Liu Y, Wang Q, Ai B, Han X, Shi S, Zhang J, Li X, et al. (2019). SEdb: A comprehensive human super-enhancer database. *Nucleic Acids Res.* 47, D235–D243. [PubMed: 30371817]
- Kim M, Lee S, Lim S, and Kim S (2019). SpliceHetero: An information theoretic approach for measuring spliceomic intratumor heterogeneity from bulk tumor RNA-seq. *PLoS One* 14, e0223520. [PubMed: 31644551]
- Kim MR, Wu MJ, Zhang Y, Yang JY, and Chang CJ (2020). TET2 directs mammary luminal cell differentiation and endocrine response. *Nat. Commun* 11, 4642. [PubMed: 32934200]
- Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, Fan H, Shen H, Way GP, Greene CS, et al. (2018). Genomic and molecular landscape of DNA damage repair deficiency across the Cancer Genome Atlas. *Cell Rep* 23, 239–254. [PubMed: 29617664]
- LaFave LM, Kartha VK, Ma S, Meli K, Del Priore I, Lareau C, Naranjo S, Westcott PMK, Duarte FM, Sankar V, et al. (2020). Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* 38, 212–228. [PubMed: 32707078]
- Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. [PubMed: 22388286]
- Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, and Tamayo P (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 1, 417–425. [PubMed: 26771021]
- Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, et al. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med* 15, 907–913. [PubMed: 19648928]
- Lin AP, Abbas S, Kim SW, Ortega M, Bouamar H, Escobedo Y, Varadarajan P, Qin Y, Sudderth J, Schulz E, et al. (2015). D2HGDH regulates alpha-ketoglutarate levels and dioxygenase function by modulating IDH2. *Nat. Commun* 6, 7768. [PubMed: 26178471]
- Losman JA, Koivunen P, and Kaelin WG Jr. (2020). 2-Oxoglutarate-dependent dioxygenases in cancer. *Nat. Rev. Cancer* 20, 710–726. [PubMed: 33087883]
- Losman JA, Looper RE, Koivunen P, Lee S, Schneider RK, McMahon C, Cowley GS, Root DE, Ebert BL, and Kaelin WG Jr. (2013). (R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science* 339, 1621–1625. [PubMed: 23393090]
- Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. [PubMed: 25516281]
- Ma J, Setton J, Lee NY, Riaz N, and Powell SN (2018). The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. *Nat. Commun* 9, 3292. [PubMed: 30120226]
- Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M, Hartmann GG, Smith OC, Kim JY, Evans KV, et al. (2020). Emergence of a high-plasticity cell state during lung cancer evolution. *Cancer Cell* 38, 229–246. [PubMed: 32707077]
- Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12.
- Marusyk A, Almendro V, and Polyak K (2012). Intra-tumour heterogeneity: A looking glass for cancer? *Nat. Rev. Cancer* 12, 323–334. [PubMed: 22513401]
- Marusyk A, Janiszewska M, and Polyak K (2020). Intratumor heterogeneity: The Rosetta stone of therapy resistance. *Cancer Cell* 37, 471–484. [PubMed: 32289271]

- McBrayer SK, Mayers JR, DiNatale GJ, Shi DD, Khanal J, Chakraborty AA, Sarosiek KA, Briggs KJ, Robbins AK, Sewastianik T, et al. (2018). Transaminase inhibition by 2-hydroxyglutarate impairs glutamate biosynthesis and redox homeostasis in glioma. *Cell* 175, 101–116. [PubMed: 30220459]
- McGranahan N, and Swanton C (2017). Clonal heterogeneity and tumor evolution: Past, present, and the future. *Cell* 168, 613–628. [PubMed: 28187284]
- Mitsuya K, Parker AN, Liu L, Ruan J, Vissers MCM, and Myatt L (2017). Alterations in the placental methylome with maternal obesity and evidence for metabolic regulation. *PLoS One* 12, e0186115. [PubMed: 29045485]
- Nam AS, Chaligne R, and Landau DA (2021). Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat. Rev. Genet* 22, 3–18. [PubMed: 32807900]
- Ngo B, Van Riper JM, Cantley LC, and Yun J (2019). Targeting cancer vulnerabilities with high-dose vitamin C. *Nat. Rev. Cancer* 19, 271–282. [PubMed: 30967651]
- Parry A, Rulands S, and Reik W (2021). Active turnover of DNA methylation during cell fate decisions. *Nat. Rev. Genet* 22, 59–66. [PubMed: 33024290]
- Pastore A, Gaiti F, Lu SX, Brand RM, Kulm S, Chaligne R, Gu H, Huang KY, Stamenova EK, Beguelin W, et al. (2019). Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in CLL. *Nat. Commun* 10, 1874. [PubMed: 31015400]
- Pellacani D, Bilenky M, Kannan N, Heravi-Moussavi A, Knapp D, Gakkhar S, Moksa M, Carles A, Moore R, Mungall AJ, et al. (2016). Analysis of normal human mammary epigenomes reveals cell-specific active enhancer states and associated transcription factor networks. *Cell Rep* 17, 2060–2074. [PubMed: 27851968]
- Qiu P, Simonds EF, Bendall SC, Gibbs KD Jr., Bruggner RV, Linderman MD, Sachs K, Nolan GP, and Plevritis SK (2011). Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol* 29, 886–891. [PubMed: 21964415]
- Risom T, Langer EM, Chapman MP, Rantala J, Fields AJ, Boniface C, Alvarez MJ, Kendsersky ND, Pelz CR, Johnson-Camacho K, et al. (2018). Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. *Nat. Commun* 9, 3815. [PubMed: 30232459]
- Roe JS, Hwang CI, Somerville TDD, Milazzo JP, Lee EJ, Da Silva B, Maiorino L, Tiriach H, Young CM, Miyabayashi K, et al. (2017). Enhancer reprogramming promotes pancreatic cancer metastasis. *Cell* 170, 875–888. [PubMed: 28757253]
- Rulands S, Lee HJ, Clark SJ, Angermueller C, Smallwood SA, Krueger F, Mohammed H, Dean W, Nichols J, Rugg-Gunn P, et al. (2018). Genome-scale oscillations in DNA methylation during exit from pluripotency. *Cell Syst* 7, 63–76. [PubMed: 30031774]
- Schep AN, Wu B, Buenrostro JD, and Greenleaf WJ (2017). chromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* 14, 975–978. [PubMed: 28825706]
- Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, Beqiri M, Sproesser K, Brafford PA, Xiao M, et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* 546, 431–435. [PubMed: 28607484]
- Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, Nikolskaya T, Serebryiskaya T, Beroukhim R, Hu M, et al. (2007). Molecular definition of breast tumor heterogeneity. *Cancer Cell* 11, 259–273. [PubMed: 17349583]
- Sulkowski PL, Oeck S, Dow J, Economos NG, Mirfakhraie L, Liu Y, Noronha K, Bao X, Li J, Shuch BM, et al. (2020). Oncometabolites suppress DNA repair by disrupting local chromatin signalling. *Nature* 582, 586–591. [PubMed: 32494005]
- Takahashi C, Au-Yeung A, Fuh F, Ramirez-Montagut T, Bolen C, Mathews W, and O’Gorman WE (2017). Mass cytometry panel optimization through the designed distribution of signal interference. *Cytometry A* 91, 39–47. [PubMed: 27632576]
- Takamizawa S, Maehata Y, Imai K, Senoo H, Sato S, and Hata R (2004). Effects of ascorbic acid and ascorbic acid 2-phosphate, a long-acting vitamin C derivative, on the proliferation and differentiation of human osteoblast-like cells. *Cell Biol. Int* 28, 255–265. [PubMed: 15109981]

- Terunuma A, Putluri N, Mishra P, Mathe EA, Dorsey TH, Yi M, Wallace TA, Issaq HJ, Zhou M, Killian JK, et al. (2014). MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *J. Clin. Invest* 124, 398–412. [PubMed: 24316975]
- Thienpont B, Steinbacher J, Zhao H, D'Anna F, Kuchnio A, Ploumakis A, Ghesquiere B, Van Dyck L, Boeckx B, Schoonjans L, et al. (2016). Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* 537, 63–68. [PubMed: 27533040]
- Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, et al. (2018). The immune landscape of cancer. *Immunity* 48, 812–830. [PubMed: 29628290]
- Turajlic S, Sottoriva A, Graham T, and Swanton C (2019). Resolving genetic heterogeneity in cancer. *Nat. Rev. Genet* 20, 404–416. [PubMed: 30918367]
- Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, Campos C, Fabius AW, Lu C, Ward PS, et al. (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* 483, 479–483. [PubMed: 22343889]
- Turner NC (2017). Signatures of DNA-repair deficiencies in breast cancer. *N. Engl. J. Med* 377, 2490–2492. [PubMed: 29262283]
- Wainwright EN, and Scaffidi P (2017). Epigenetics and cancer stem cells: Unleashing, hijacking, and restricting cellular plasticity. *Trends Cancer* 3, 372–386. [PubMed: 28718414]
- Wang H, Bieri B, Li AG, Pathania S, Toomire K, Dimitrov SD, Liu B, Gelman R, Giobbie-Hurder A, Feunteun J, et al. (2016). BRCA1/FANCD2/BRG1-driven DNA repair stabilizes the differentiation state of human mammary epithelial cells. *Mol. Cell* 63, 277–292. [PubMed: 27373334]
- Wang H, Xiang D, Liu B, He A, Randle HJ, Zhang KX, Dongre A, Sachs N, Clark AP, Tao L, et al. (2019). Inadequate DNA damage repair promotes mammary transdifferentiation, leading to BRCA1 breast cancer. *Cell* 178, 135–151. [PubMed: 31251913]
- Wang P, Wu J, Ma S, Zhang L, Yao J, Hoadley KA, Wilkerson MD, Perou CM, Guan KL, Ye D, et al. (2015). Oncometabolite d-2-hydroxyglutarate inhibits ALKBH DNA repair enzymes and sensitizes IDH mutant cells to alkylating agents. *Cell Rep* 13, 2353–2361. [PubMed: 26686626]
- Wang T, Chen K, Zeng X, Yang J, Wu Y, Shi X, Qin B, Zeng L, Esteban MA, Pan G, et al. (2011). The histone demethylases Jhdmla/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* 9, 575–587. [PubMed: 22100412]
- Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, et al. (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512, 155–160. [PubMed: 25079324]
- Yu M, Hon GC, Szulwach KE, Song CX, Jin P, Ren B, and He C (2012). Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protoc* 7, 2159–2170. [PubMed: 23196972]

Highlights

- Inhibition of α KG-dependent demethylation destabilizes chromatin regulatory landscape
- 2HG subverts lineage fidelity and increases cell-level variability in motif accessibility
- 2HG-high breast tumors display enhanced cellular heterogeneity
- A2P eradicates heterogeneity in high 2HG-producing basal-like cancer cells

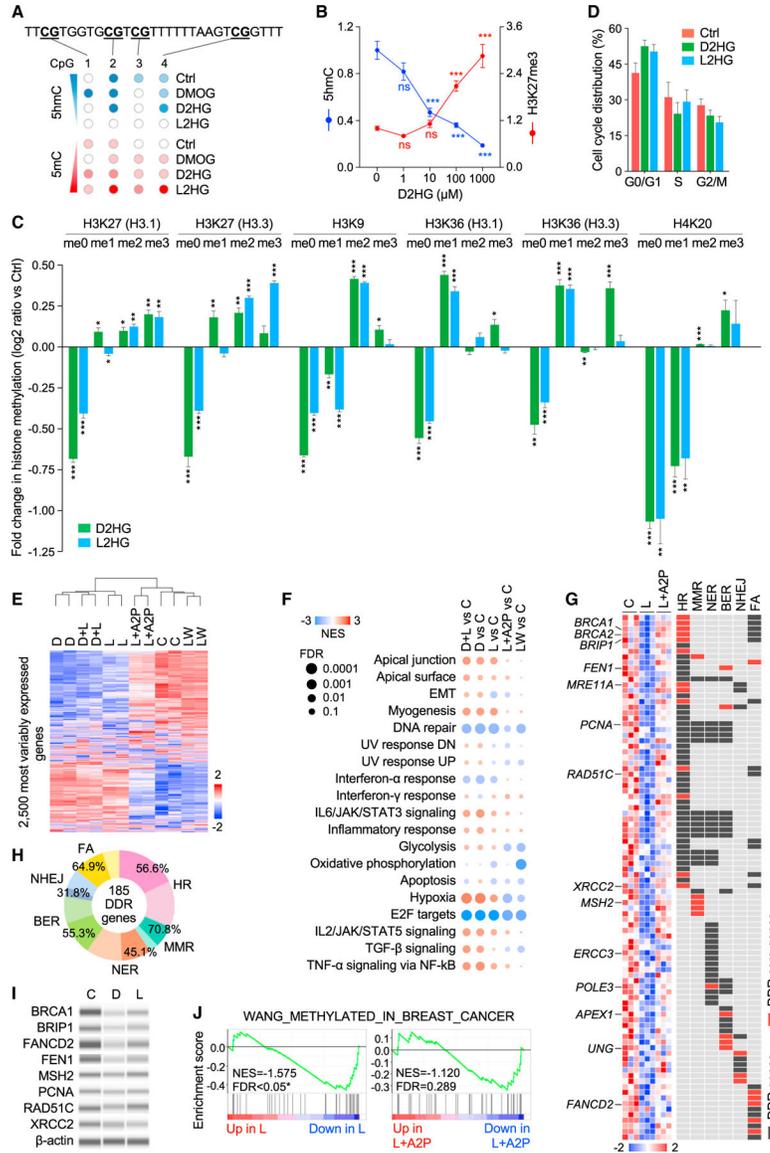


Figure 1. 2HG-mediated remodeling of the chromatin-transcriptional landscape is restored by A2P in the presence of 2HG

(A) 5hmC and 5mC levels at long interspersed element-1 (LINE-1) sequences. HMECs were treated for 72 h with 1 mM of D2HG, L2HG, or dimethylxalylglycine (DMOG) and subjected to TET-assisted bisulfite (TAB) pyrosequencing to differentiate 5hmC from 5mC (see STAR Methods for further details). DMOG is an analog of α KG and competes for binding at the active center of the enzyme.

(B) Global levels of 5hmC and H3K27me3 detected by immunostaining. *** $p < 0.001$ versus control by one-way ANOVA with Dunnett’s multiple comparison test ($n = 8$ images per condition). ns, not significant. At least 60 nuclei were examined, and signal intensities were normalized to DAPI nuclear counterstain. Mean \pm SEM is presented.

(C) Fold change in different types of histone methylation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ versus control ($n = 3$ independent replicates).

- (D) Cell cycle analysis of HMECs exposed for 72 h to 100 μ M of D2HG or L2HG. No statistical significance was observed using one-way ANOVA. Results are from three independent experiments.
- (E) Hierarchical clustering of RNA-seq profiles. Cells were treated for 72 h with 100 μ M of D2HG (D), L2HG (L), D2HG plus L2HG (D + L) or L2HG in the presence of 1 mM A2P (L + A2P), or received no treatment (C). L2HG treatment was followed by 4-day withdrawal (LW).
- (F) Dot plot of GSEA hallmark pathways. Color indicates normalized enrichment scores (NESs) of positively and negatively enriched gene sets relative to control, and circle size corresponds to false discovery rate (FDR).
- (G) Heatmap depicting changes in expression of DDR genes. The corresponding DNA repair pathways are shown on the right.
- (H) Pie chart representing the proportion of DDR pathway genes. Shaded fractions correspond to genes that are downregulated by 2HG.
- (I) Expression levels of DDR proteins detected by capillary-based immunoassay.
- (J) GSEA plots evaluating enrichment for methylated genes in breast cancer cells. See also Figure S1 and Table S1.

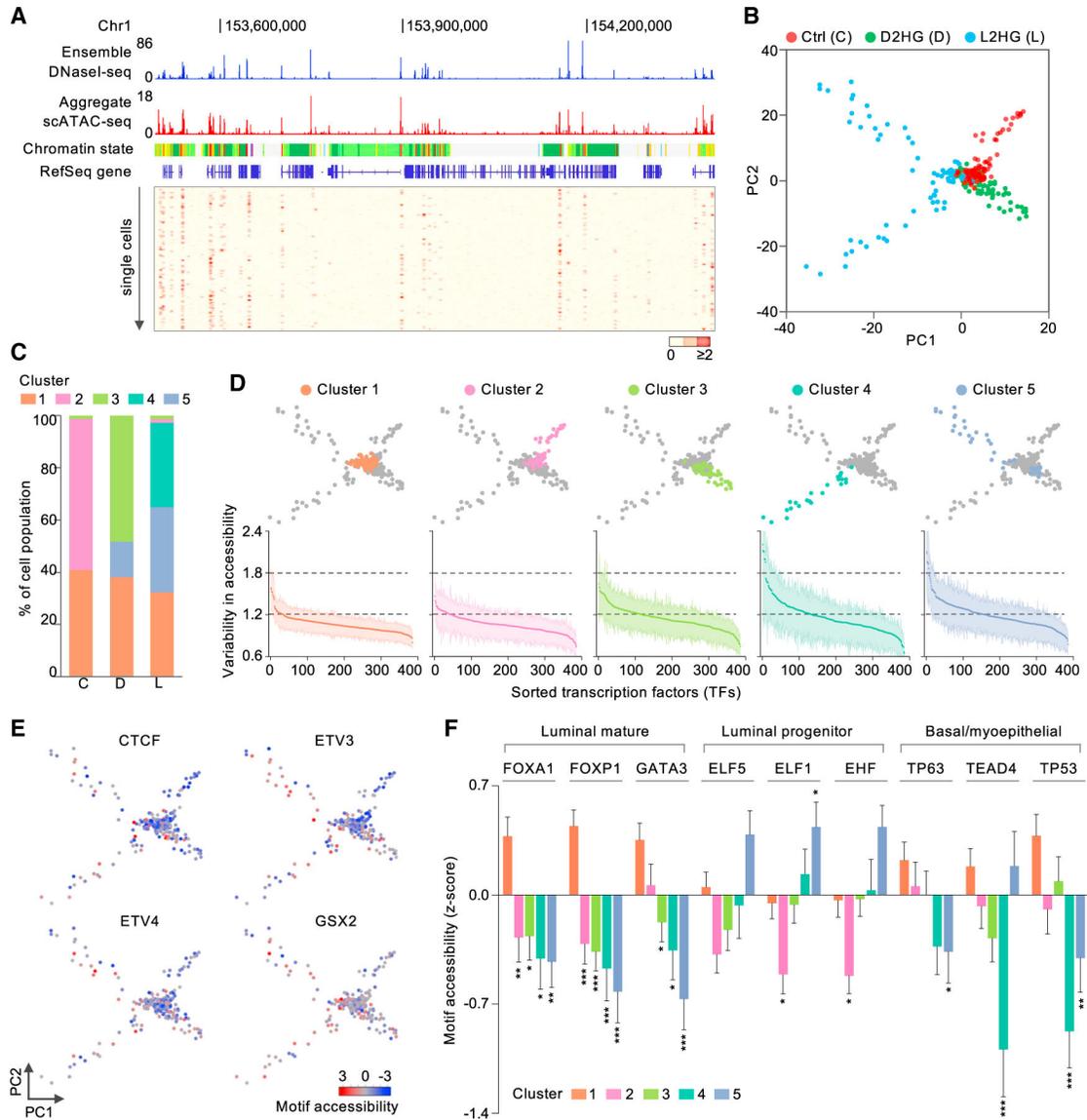


Figure 2. 2HG enhances molecular heterogeneity in gene regulatory dynamics

(A) Genome tracks showing open chromatin regions detected by scATAC-seq. scATAC-seq profiling is highly consistent with DNase hypersensitive sites (DHSs) detected by bulk DNaseI-seq (GSE29692).

(B) PCA projection showing distinct cell subpopulations based on the single-cell chromatin accessibility landscape.

(C) Proportions of cell subpopulations identified by model-based clustering.

(D) PCA plots of five distinct cell clusters identified by model-based clustering (top). Bottom plots indicate variability in chromatin accessibility at consensus TF binding motifs ($n = 386$) across single cells in each cluster.

(E) Representative DNA accessibility at TF binding motifs.

(F) Chromatin accessibility at lineage-specific TF binding motifs. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ versus cluster 1.

See also Figure S2.

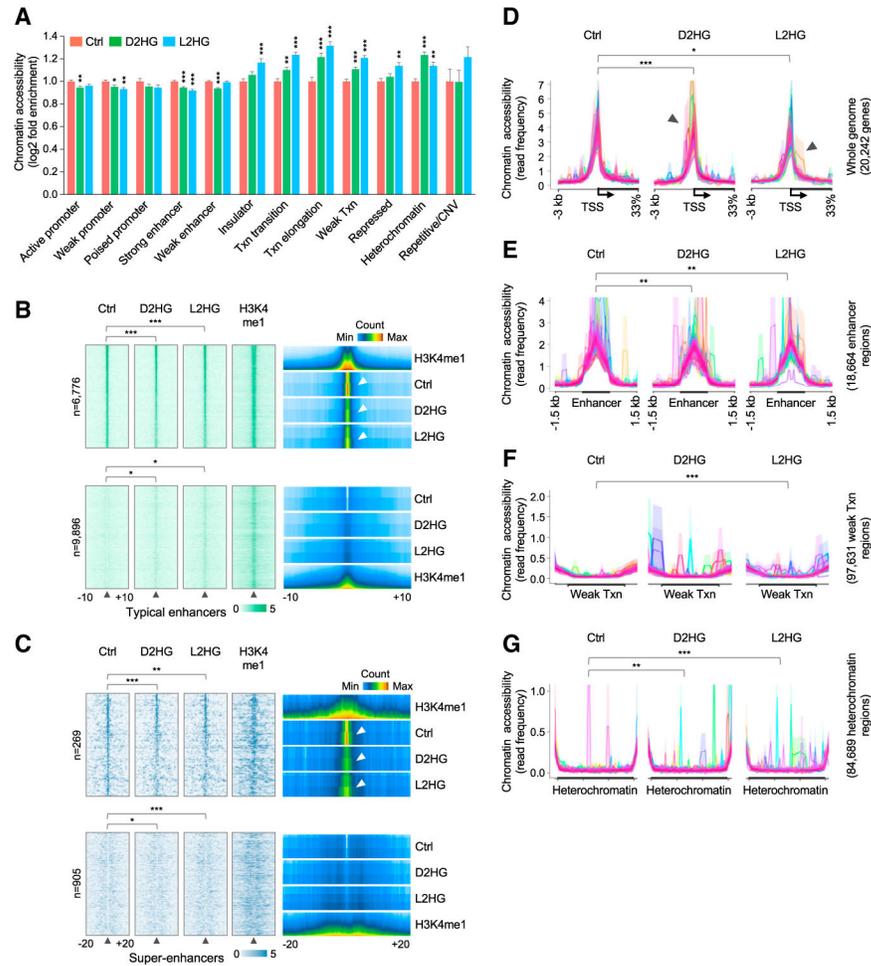


Figure 3. 2HG initiates cell-level fluctuations in the mammary epithelial epigenome

(A) ATAC-seq signal enrichment across ChromHMM-annotated genomic regions. Txn, transcription; CNV, copy number variation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ versus control. (B and C) Heatmaps showing enrichment of scATAC-seq signals at typical enhancers (B) and super-enhancers (C). Each row represents one individual element, and color represents the intensity of chromatin accessibility in the left panels. The corresponding density plots are shown on the right.

(D–G) Single-cell accessibility landscape across the genome centered on TSSs (D), enhancers (E), and weakly transcribed (F) and heterochromatin regions (G). Colors represent individual cells. Solid lines indicate mean values, and semi-transparent shade around the mean curve shows SEM across the region.

See also Figure S3.

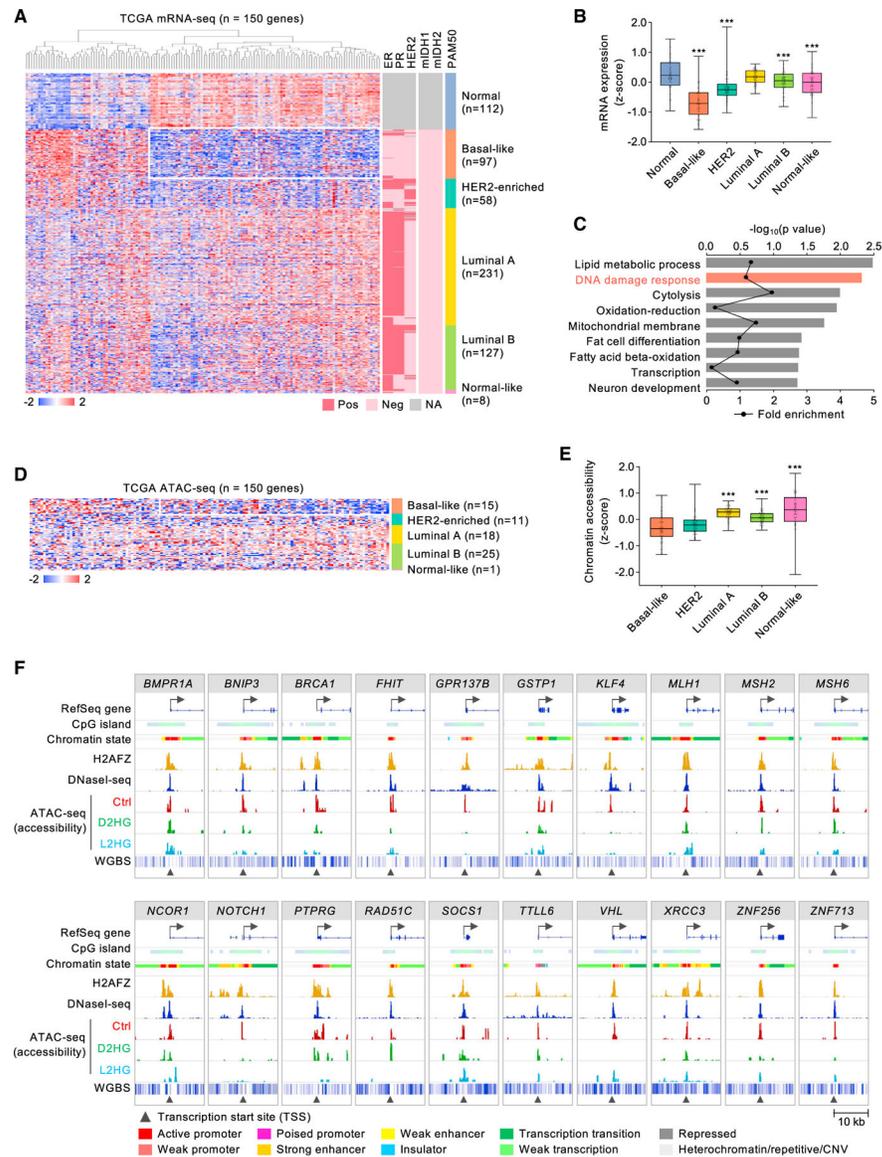


Figure 4. Chromatin remodeling involves tumor-associated promoter hypermethylation
 (A) Heatmap depicting mRNA expression of highly methylated genes (n = 150) whose promoter regions showed diminished chromatin accessibility following 2HG exposure. The rectangle outlined in white represents genes that are downregulated in basal-like tumors.
 (B) Box-and-whisker plot showing mRNA expression of genes that are indicated with white outline in (A). ***p < 0.001 versus adjacent normal tissue.
 (C) Gene Ontology (GO) biological processes identified by DAVID pathway enrichment analysis of 150 genes that are hypermethylated in breast tumors.
 (D) Heatmap showing chromatin accessibility of hypermethylated genes (n = 150) in breast tumors. The rectangle outlined in white represents genes exhibiting decreased expression in (A).
 (E) Box-and-whisker plot showing chromatin accessibility of genes that are indicated with white outline in (D). ***p < 0.001 versus basal-like tumors.

(F) Genome tracks showing diminished chromatin accessibility at gene promoters following 2HG exposure. ChromHMM chromatin states (GSE38163), DNaseI-seq (GSE29692), H2AFZ chromatin immunoprecipitation sequencing (ChIP-seq) (GSE29611), and WGBS profiles (GSE86732) from HMECs are shown. See also Figure S4 and Table S2.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

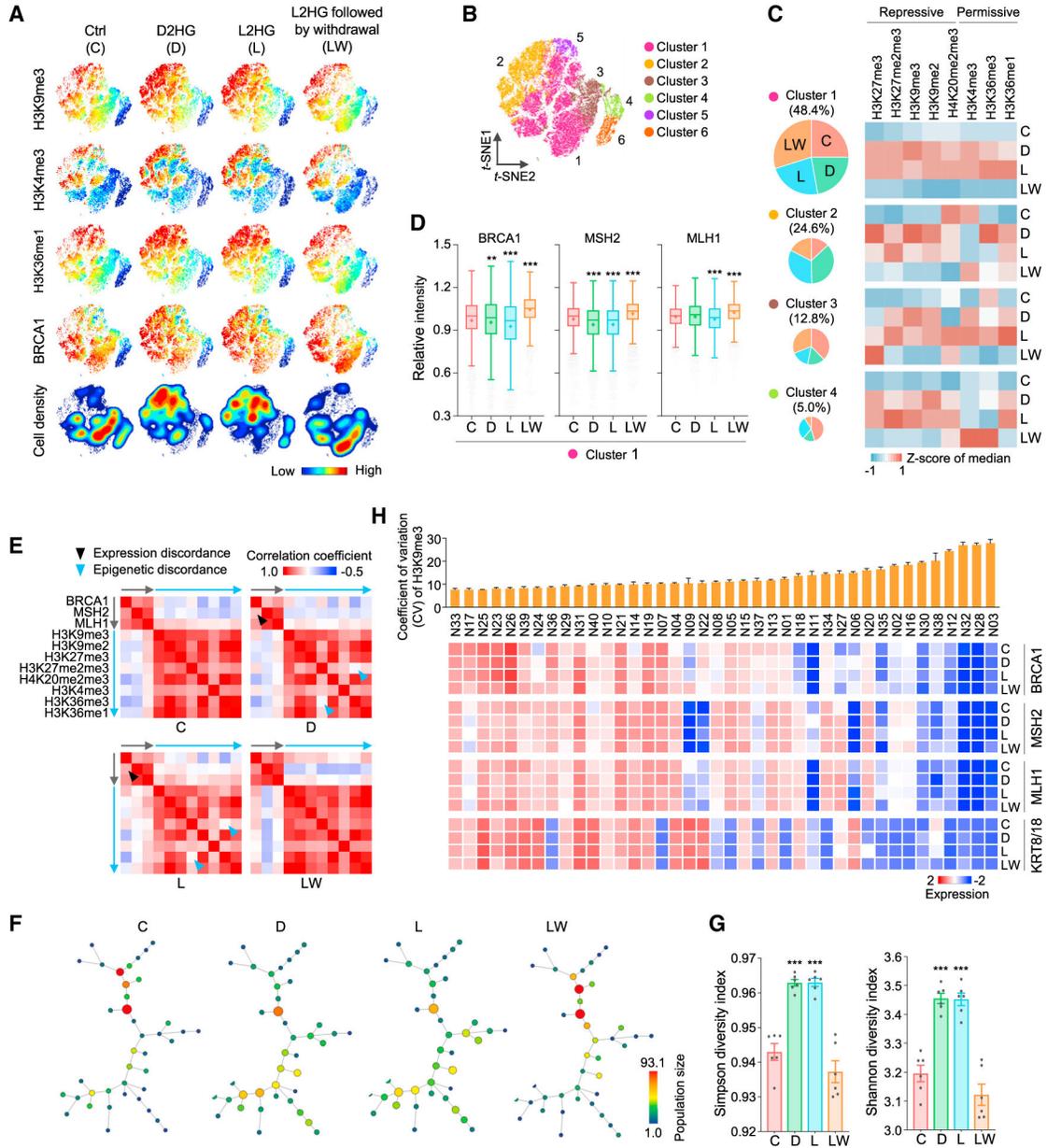


Figure 5. 2HG induces epigenetic discordancy and enhances phenotypic diversity
 (A) Heatmaps of selected markers used for chromatin CyTOF, which simultaneously detects epigenetic modification levels as well as marker expression in single cells. HMECs were treated for 72 h with 100 μ M of D2HG (D) or L2HG (L), or received no treatment (C). L2HG treatment was followed by 4-day withdrawal (LW). Each data point on the *t*-SNE maps represents an individual cell, and color corresponds to cellular levels of each marker.
 (B) *t*-SNE projection of epigenetically distinct cell subsets defined by consensus hierarchical clustering.
 (C) Heatmaps depicting changes in histone modifications. Normalized median values of signal intensities are shown for each cluster (right). Pie charts (left) indicate the proportion of cells from different experimental groups in each cluster.
 (D) Relative intensity of BRCA1, MSH2, and MLH1.
 (E) Heatmaps of expression discordance (red) and epigenetic discordance (blue) for BRCA1, MSH2, MLH1, and various histone modifications across conditions C, D, L, and LW.
 (F) Dendrograms showing population size for each cluster across conditions C, D, L, and LW.
 (G) Simpson and Shannon diversity indices for each cluster across conditions C, D, L, and LW.

(D) Expression levels of DDR genes in cluster 1. ** $p < 0.01$, *** $p < 0.001$ versus control.

(E) Heatmaps of pairwise Spearman correlations between marker levels in single cells.

Black and blue arrowheads indicate reduced concordancy in gene expression and epigenetic modifications, respectively.

(F) Projection of minimum spanning trees (MSTs) obtained by SPADE analysis. The size and color of nodes represent the number of cells in each cluster, allowing visualization of the extent of cellular heterogeneity in treatment groups.

(G) Quantification of cellular heterogeneity using Simpson and Shannon entropy indices. *** $p < 0.001$ versus control.

(H) Correlation between cell-to-cell variance in H3K9me3 and marker expression levels in each node identified by SPADE analysis.

See also Figure S5 and Table S3.

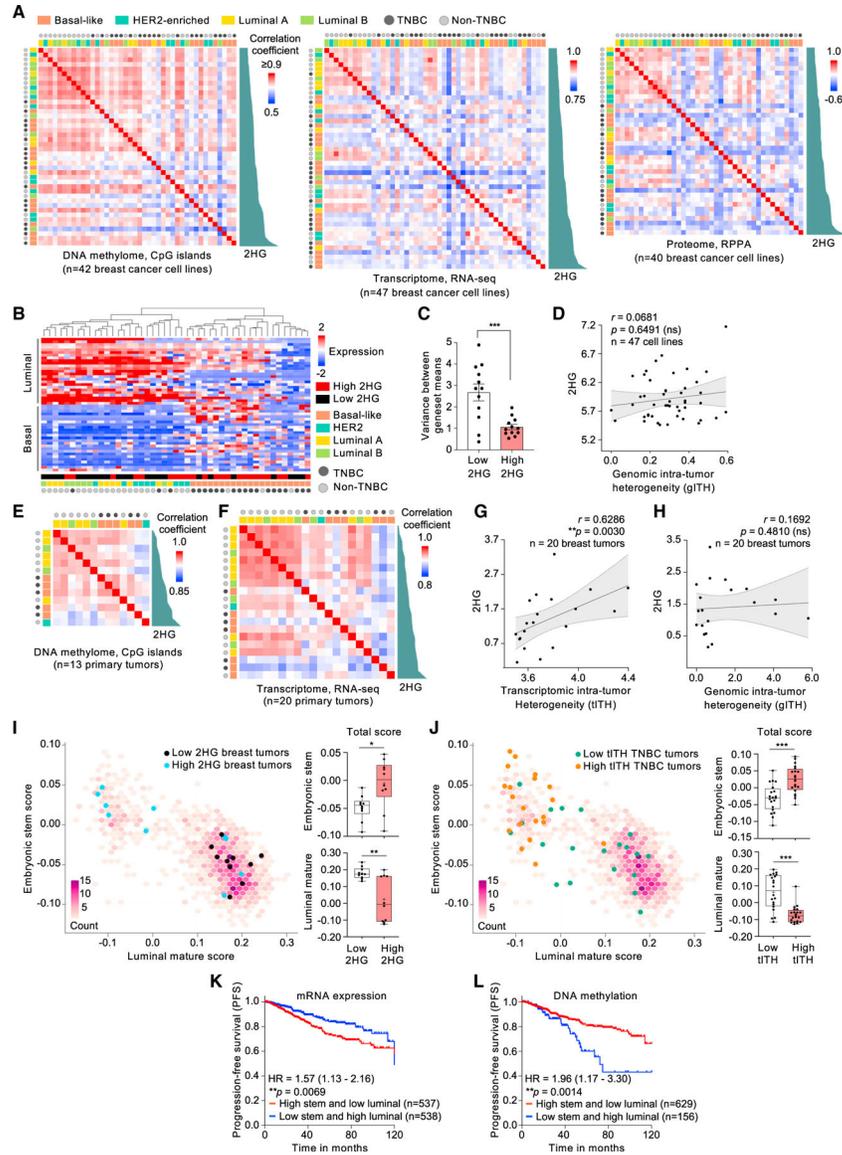


Figure 6. Enhanced cellular diversity and undifferentiated stem-like signatures correlate with 2HG production

(A) Pairwise correlation matrix of breast cancer cell lines. Spearman’s correlation coefficients between samples were assessed based on DNA methylation profiles of 3,000 CpG islands (left), RNA-seq profiles of 18,319 protein-coding transcripts (center), and reverse-phase protein microarray (RPPA) profiles of 213 proteins (right). Green area plots indicate intracellular levels of 2HG.

(B) Heatmap depicting expression of luminal (n = 25) and myoepithelial (n = 25) genes in breast cancer cell lines (n = 47).

(C) Variance in mean expression levels between luminal and myoepithelial gene sets in 2HG-low (n = 12) and 2HG-high (n = 12) cell lines. ***p < 0.001 by unpaired Student’s t test.

(D) Scatterplot showing the relationship between 2HG accumulation and genomic intratumor heterogeneity (gITH) in breast cancer cell lines (n = 47). ns, not significant.

(E and F) Pairwise correlation matrix of TCGA primary tumors.

(G and H) Scatterplots showing the relationship of 2HG accumulation with transcriptomic intratumor heterogeneity (tITH) and genomic intratumor heterogeneity (gITH) in primary breast tumors (n = 20).

(I and J) Mature luminal and undifferentiated stem cell-like scores obtained by single-sample scoring analysis using luminal mature (n = 50) and embryonic stem (n = 40) gene sets. Hexagonal density plots represent all breast tumor samples available in the TCGA cohort (n = 1,100). *p < 0.05, **p < 0.01, ***p < 0.001 by unpaired Student's t test.

(K and L) Prognostic significance of undifferentiated stem-like signatures in individuals listed in the TCGA. Progression-free survival was evaluated using Kaplan-Meier analysis based on mRNA expression (K) and DNA methylation (L) of embryonic stem (n = 40) and luminal mature (n = 50) gene sets.

See also Figure S6.

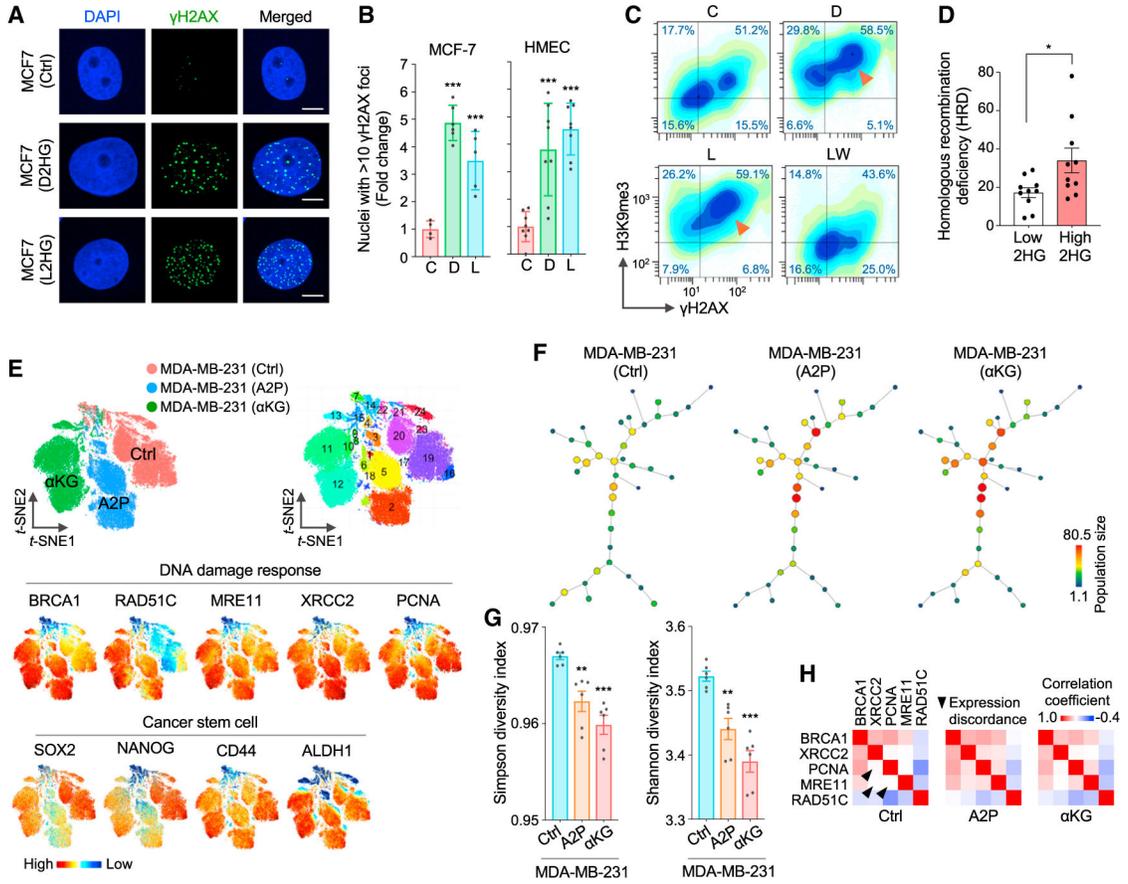


Figure 7. 2HG induces DNA damage accumulation, and A2P eradicates cancer cell heterogeneity

(A) Immunofluorescence staining of γ H2AX. MCF-7 cells were treated with 2HG under the same conditions (100 μ M for 72 h) as those used in the RNA-seq and ATAC-seq experiments. Scale bar, 10 μ m.

(B) Quantification of cells with γ H2AX focus-positive nuclei. For each treatment, 200–300 cells were examined, and cells with at least four γ H2AX foci in nuclei were counted as positive. *** $p < 0.001$ versus control.

(C) CyTOF analysis of γ H2AX levels. HMECs were treated for 72 h with 100 μ M of D2HG (D) or L2HG (L) or received no treatment (C). L2HG treatment was followed by 4-day withdrawal (LW).

(D) HRD scores in TCGA breast tumors. * $p < 0.05$ by unpaired Student’s t test.

(E) *t*-SNE visualization of CyTOF profiling. MDA-MB-231 cells were treated with 1 mM A2P or 100 μ M α KG for 72 h or received no treatment.

(F) Projection of SPADE trees. The size and color of nodes represent the number of cells in each cluster, allowing visualization of the extent of cellular heterogeneity in treatment groups.

(G) Quantification of cellular heterogeneity using Simpson and Shannon diversity indices. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ versus control.

(H) Heatmaps of pairwise Spearman correlations between marker levels in single cells. Arrowheads indicate reduced concordancy in gene expression.

See also Figure S7 and Table S3.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-5hmC	Active Motif	Cat# 39769
Rabbit monoclonal anti-trimethyl-Histone H3 (Lys27) (Clone C36B11)	Cell Signaling Technology	Cat# 9733
Mouse monoclonal anti-phospho-Histone H2A.X (Ser139) (Clone JBW301)	MilliporeSigma	Cat# 05-636
Goat anti-rabbit IgG (H+L), Alexa Fluor 488	Invitrogen	Cat# A-11008
Goat anti-mouse IgG (H+L), Alexa Fluor 488	Invitrogen	Cat# A-11001
Mouse monoclonal anti-PCNA (Clone D3H8P)	Cell Signaling Technology	Cat# 13110
Rabbit polyclonal anti-BRCA1	Cell Signaling Technology	Cat# 9010
Rabbit polyclonal anti-XRCC2	Abcam	Cat# ab180752
Mouse monoclonal anti-RAD51C (Clone 2H11/6)	Novus Biologicals	Cat# NB100-177
Mouse monoclonal anti-MSH2 (Clone MSH2/2622)	Novus Biologicals	Cat# NBP3-07211
Rabbit polyclonal anti-FANCD2	Novus Biologicals	Cat# NB100-182
Mouse monoclonal anti-FEN-1 (Clone 4E7)	Novus Biologicals	Cat# NB100-150
Rabbit polyclonal anti-BRIP1	Novus Biologicals	Cat# NBP1-31883
Mouse monoclonal anti-beta-actin (Clone 937215)	R&D Systems	Cat# MAB8929
Rat monoclonal anti-phospho-Histone H3 (Ser28) (Clone HTA28)	Fluidigm	Cat# 3175012A
Mouse monoclonal anti-pH2AX [S139] (Clone N1-431)	Fluidigm	Cat# 3165036D
Mouse monoclonal anti-Ki67 (Clone B56)	Fluidigm	Cat# 3168007B
Mouse monoclonal anti-EpCAM (Clone 9C4)	Fluidigm	Cat# 3141006B
Mouse monoclonal anti-KRT8/18 (Clone C51)	Fluidigm	Cat# 3174014A
Rat monoclonal anti-CD49F (Clone GoH3)	Fluidigm	Cat# 3164006B
Mouse monoclonal anti-SOX2 (Clone O30-678)	Fluidigm	Cat# 3150019B
Mouse monoclonal anti-NANOG (Clone N31-355)	Fluidigm	Cat# 3169014A
Rat monoclonal anti-CD44 (Clone IM7)	Fluidigm	Cat# 3171003B
Mouse monoclonal anti-ALDH (Clone 44/ALDH)	Fluidigm	Cat# 3147015B
Mouse monoclonal anti-5mC (Clone 33D3)	Active Motif	Cat# 39649
Rat monoclonal anti-5hmC (Clone AB3/63.3)	Novus Biologicals	Cat# NBP2-50099
Mouse monoclonal anti-trimethyl-Histone H3 (Lys27) (Clone MABI 0323)	Active Motif	Cat# 61017
Mouse monoclonal anti-di, trimethyl-Histone H3 (Lys27) (Clone mAbcam 6147)	Abcam	Cat# ab6147
Mouse monoclonal anti-trimethyl-Histone H3 (Lys9) (Clone MABI 0319)	Active Motif	Cat# 61013
Rabbit polyclonal anti-dimethyl-Histone H3 (Lys9)	Invitrogen	Cat# PA5-24987
Mouse monoclonal anti-di, trimethyl-Histone H4 (Lys20) (Clone 6F8-D9)	Abcam	Cat# ab78517
Mouse monoclonal anti-trimethyl-Histone H3 (Lys4) (Clone mAbcam12209)	Abcam	Cat# ab12209
Mouse monoclonal anti-trimethyl-Histone H3 (Lys36) (Clone MABI 0333)	Active Motif	Cat# 61021

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rabbit polyclonal anti-monomethyl-Histone H3 (Lys36)	Boster Biological Technology	Cat# CI1064
Rabbit polyclonal anti-trimethyl-Histone H3 (Lys79)	Boster Biological Technology	Cat# CI1055
Rabbit polyclonal anti-dimethyl-Histone H3 (Lys79)	Boster Biological Technology	Cat# CI1046
Mouse monoclonal anti-BRCA1 (Clone MS110)	Abcam	Cat# ab16780
Mouse monoclonal anti-MSH2 (Clone 3A2B8C)	Abcam	Cat# ab52266
Mouse monoclonal anti-XRCC2 (Clone EPR5149)	Abcam	Cat# ab248049
Mouse monoclonal anti-PCNA (Clone PC10)	Abcam	Cat# ab264494
Mouse monoclonal anti-RAD51C (Clone 2H11/6)	Novus Biologicals	Cat# NB100-177
Mouse monoclonal anti-MRE11 (Clone 12D7)	Novus Biologicals	Cat# NB100-473
Mouse monoclonal anti-MLH1 (Clone 4C9C7)	Cell Signaling Technology	Cat# 3515S
Chemicals, peptides, and recombinant proteins		
(2R)-Octyl- α -hydroxyglutarate	Cayman Chemical	Cat# 16366; CAS# 1391194-67-4
(2S)-Octyl- α -hydroxyglutarate	Cayman Chemical	Cat# 16367; CAS# 1391194-64-1
Dimethylxalylglycine (DMOG)	Sigma-Aldrich	Cat# D3695; CAS# 89464-63-1
Ascorbate-2-phosphate (A2P)	Sigma-Aldrich	Ca# A8960; CAS# 113170-55-1
4-hydroxytamoxifen (4-OHT)	Sigma-Aldrich	Ca# H7904; CAS# 68047-06-3
[1,2,3,4- ¹³ C ₄] α -Ketoglutaric acid	Cambridge Isotope Laboratories	Cat# CLM-4442; CAS# 305-72-6
[1,2,3,4- ¹³ C ₄]L-Malic acid	Cambridge Isotope Laboratories	Cat# CLM-8065; CAS# 97-67-6
Diacetyl-L-tartaric anhydride (DATAN)	Sigma-Aldrich	Cat# 358924; CAS# 6283-74-5
Triton-X-100	ThermoFisher	Cat# BP151-500
Hydrochloric acid	ThermoFisher	Cat# SA56-500
DAPI	ThermoFisher	Cat# 62248
RNase A	ThermoFisher	Cat# AM2269
Proteinase K	QIAGEN	Cat# 158920
GelRed	Biotium	Cat# 41003
Methanol	Fisher Scientific	Cat# A412-500
Propium iodide	Sigma-Aldrich	Cat# P4864; CAS# 25535-16-4
TRI Reagent	Zymo Research	Cat# R2050-1-200
ERCC RNA spike-in control mixes	Ambion	Ca# 4456740
Ethidium homodimer-1	Life Technologies	Cat# L-3224
Calcein AM	Life Technologies	Cat# L-3224
NP40	Life Technologies	Cat# 28324
Tn5 Transposase	Illumina	Cat# FC-121-1030
TD	Illumina	Cat# FC-121-1030
Nuclease-Free Water	Life Technologies	Cat# AM9937
EDTA	Life Technologies	Cat# AM9260G
Tris-HCl, pH 8.0	Life Technologies	Cat# AM9855G
MgCl ₂	Life Technologies	Cat# AM9530G
C1 Blocking Reagent	Fluidigm	Cat# 100-5316

REAGENT or RESOURCE	SOURCE	IDENTIFIER
C1 Harvest Reagent	Fluidigm	Cat# 100-6248
C1 Preloading Reagent	Fluidigm	Cat# 100-5311
Suspension Reagent	Fluidigm	Cat# 100-5315
Cell Wash Buffer	Fluidigm	Cat# 100-5314
C1 No Salt Loading Reagent	Fluidigm	Cat# 101-0133
Bond-Breaker TCEP Solution, Neutral pH	ThermoFisher	Cat# 77720
Antibody Stabilizer	CANDOR Bioscience	Cat# 131050
Sodium azide	Teknova	Cat# S0209
Cell-ID Cisplatin	Fluidigm	Cat# 201064
Cell-ID Intercalator-Ir	Fluidigm	Cat# 201192A
Maxpar PBS	Fluidigm	Cat# 201058
Maxpar Water	Fluidigm	Cat# 201069
EQ Four Element Calibration Beads	Fluidigm	Cat# 201078
Paraformaldehyde	Electron Microscopy Sciences	Cat# 15710
Fc Receptor Blocking Solution	BioLegend	Cat# 422302
Critical commercial assays		
Gentra Puregene Tissue Kit	QIAGEN	Cat# 158667
EpiTect Fast Bisulfite Conversion Kit	QIAGEN	Cat# 59826
PyroMark Q96 CpG LINE-1	QIAGEN	Cat# 973043
NEBNext High-Fidelity 2X PCR Master Mix	New England Biolabs	Cat# M0541
MinElute PCR Purification Kit	QIAGEN	Cat# 28006
High Sensitivity DNA Kit	Agilent	Cat# 5067-4626
NextSeq 500/550 High Output Kit	Illumina	Cat# TG-160-2005
TrueMethyl oxBS Module	NuGEN Technologies	Cat# 0414-32
Qubit ssDNA Assay Kit	Invitrogen	Cat# Q10212
PyroMark Q48 Advanced CpG reagents	QIAGEN	Cat# 974022
GenoMatrix Whole Genome Amplification kit	Active Motif	Cat# 58001
PowerUP SYBR Green Master Mix	ThermoFisher	Cat# A25743
RiboFree Total RNA Library Kit	Zymo Research	Ca# R3003
Maxpar X8 Multimetal Labeling Kit	Fluidigm	Cat# 201300
Maxpar Nuclear Antigen Staining Buffer Set	Fluidigm	Cat# 201063
12-230 kDa Wes Separation Module, capillary cartridges	ProteinSimple	Cat# SM-W004
Deposited data		
Raw single-cell ATAC-seq data	This paper	SRA: SRP217510 SuperSeries: GSE161628
Processed single-cell ATAC-seq data	This paper	SuperSeries: GSE161628
Raw RNA-seq data	This paper	SRA: SRP292912 SuperSeries: GSE161628
ChromHMM-defined chromatin state	ENCODE	GEO: GSE38163
DNaseI-seq data	ENCODE	GEO: GSE29692

REAGENT or RESOURCE	SOURCE	IDENTIFIER
H2AFZ/H2A.Z ChIP-seq data	ENCODE	GEO: GSE29611
Whole-genome bisulfite sequencing (WGBS) data	ENCODE	GEO: GSE86732
The Cancer Methylome	University of Texas Health Science Center at San Antonio	http://cbbiweb.uthscsa.edu/KMethylomes
Cancer Cell Line Encyclopedia (CCLE)	Broad Institute	https://portals.broadinstitute.org/ccle
Encyclopedia of DNA Elements (ENCODE)	Davis et al. (2018)	https://www.encodeproject.org
Roadmap Epigenomics	Ernst et al. (2011)	http://www.roadmapepigenomics.org
The Cancer Genome Atlas (TCGA)	Cancer Genome Atlas Network	https://www.cancer.gov/tcga
Experimental models: Cell lines		
Human: BT20	ATCC	HTB-19
Human: BT474	ATCC	HTB-20
Human: MCF7	ATCC	HTB-22
Human: MDA-MB-157	ATCC	HTB-24
Human: MDA-MB-231	ATCC	HTB-26
Human: MDA-MB-361	ATCC	HTB-27
Human: 184B5	ATCC	CRL-8799
Human: MCF12A	ATCC	CRL-10782
Human: hTERT-HME1	ATCC	CRL-4010
Human: Human Mammary Epithelial Cell (HMEC)	Invitrogen	A10565
Oligonucleotides		
Nextera dual-index primers	Buenostro et al. (2015)	N/A
CpG assays for oxBS pyrosequencing, see Table S2	This paper	N/A
Primers for pyrosequencing, see Table S2	This paper	N/A
Primers for MeDIP validation, see Table S2	This paper	N/A
Software and algorithms		
NIH ImageJ software	NIH	https://imagej.nih.gov/ij/
PyroMark CpG software	QIAGEN	N/A
CellQuest Pro software	Becton Dickinson	N/A
FlowJo software	Tree Star	N/A
FastQC tools	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MultiQC tools	Ewels et al. (2016)	https://multiqc.info/
Cutadapt	Martin (2011)	https://cutadapt.readthedocs.io/en/stable/
STAR	Dobin et al. (2013)	https://github.com/alexdobin/STAR/
DESeq2	Love et al. (2014)	https://github.com/mikelove/DESeq2/blob/master/vignettes/DESeq2.Rmd
GSEA software	Broad Institute	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bowtie2	Langmead and Salzberg (2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Picard tools	Broad Institute	http://broadinstitute.github.io/picard/
SAMtools	Sanger Institute Broad Institute	http://samtools.sourceforge.net/
BEDtools	University of Utah	https://bedtools.readthedocs.io/en/latest/
MACS2	Dana-Farber Cancer Institute	http://liulab.dfci.harvard.edu/MACS/
SCRAT	Jietal. (2017)	https://zhiji.shinyapps.io/scrat/
Integrative Genomics Viewer (IGV)	Broad Institute and the Regents of the University of California	https://software.broadinstitute.org/software/igv/
Circos tools	Canada's Michael Smith Genome Sciences Centre	http://circos.ca
Broad GDAC Firehose	Broad Institute	http://gdac.broadinstitute.org
R-statistical program	R Foundation	https://www.r-project.org/
cBioPortal	Gao et al. (2013)	https://www.cbioportal.org/
UCSC Xena browser	University of California, Santa Cruz	https://xenabrowser.net/
Xcalibur software	ThermoFisher	N/A
ZEN software	ZEISS	N/A
StepOne software	ThermoFisher	N/A
PyroMark Assay Design software	QIAGEN	N/A
PyroMark Q48 Autoprep software	QIAGEN	N/A
CyTOF software	Fluidigm	N/A
Cytobank	Cytobank Inc	https://cytobank.org
Singscore	Foroutan et al. (2018)	https://davislaboratory.github.io/singscore/articles/singscore.html
Compass software	ProteinSimple	https://www.proteinsimple.com/compass/downloads/
GraphPad Prism	GraphPad Software	N/A
Other		
8-well culture slide	Corning	Cat# 354118
ProLong Gold Antifade Mountant	Invitrogen	Cat# P36934
Slip-Rite cover glass	ThermoFisher	Cat# 152455
CellTrics filters, 20 mm	Sysmex	Cat# 04-004-2325
C1 Single-Cell Open App IFC, 10-17 mm	Fluidigm	Cat# 100-8134
Amicon Ultra-0.5 Centrifugal Filter Unit, 3 kDa	Millipore	Cat# UFC500396
Amicon Ultra-0.5 Centrifugal Filter Unit, 50 kDa	Millipore	Cat# UFC505096