The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity*

Cristina Morales Torres[#1], **Alva Biran**[#2], **Matthew J. Burney**[1], **Harshil Patel**[3], **Tristan Henser-Brownhill**[1], **Ayelet-Hashahar Shapira Cohen**[2], **Yilong Li**[4], **Rotem Ben-Hamo**[5], **Emma Nye**[6], **Bradley Spencer-Dene**[6], **Probir Chakravarty**[3], **Sol Efroni**[5], **Nik Matthews**[7], **Tom Misteli**[8], **Eran Meshorer**[2], and **Paola Scaffidi**[1,9,‡]

[1]Cancer Epigenetics Laboratory, The Francis Crick Institute, Lincoln's Inn Fields Laboratory, London WC2A 3LY, UK

[2]Department of Genetics, The Institute of Life Sciences, and The Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, 91904, Israel

[3]Bioinformatics, The Francis Crick Institute, Lincoln's Inn Fields Laboratory, London WC2A 3LY, UK

[4]Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB101SA, UK

[5]The Mina and Everard Goodman Faculty of Life Science, Bar Ilan University, Ramat-Gan, 52900, Israel

[6]Experimental Histopathology, The Francis Crick Institute, Lincoln's Inn Fields Laboratory, London WC2A 3LY, UK

[7]Advanced sequencing, The Francis Crick Institute, Lincoln's Inn Fields Laboratory, London WC2A 3LY, UK

[8]National Cancer Institute, NIH, Bethesda, MD, 20892, USA

[9]UCL Cancer Institute, University College London, London WC1E 6DD, UK

[#] These authors contributed equally to this work.

Abstract

Tumors comprise functionally diverse subpopulations of cells with distinct proliferative potential. Here, we show that dynamic epigenetic states defined by the linker histone H1.0 determine which cells within a tumor can sustain the long-term cancer growth. Numerous cancer types exhibit high inter- and intratumor heterogeneity of H1.0, with H1.0 levels correlating with tumor differentiation status, patient survival and, at the single-cell level, cancer stem cell markers. Silencing of H1.0

---

promotes maintenance of self-renewing cells by inducing de-repression of megabase-sized gene domains harboring downstream effectors of oncogenic pathways. Self-renewing epigenetic states are not stable and re-expression of H1.0 in subsets of tumor cells establishes transcriptional programs that restrict cancer cell long-term proliferative potential and drive their differentiation. Our results uncover epigenetic determinants of tumor-maintaining cells.

Cancer is a clonal disease originating from a single cell. Yet, most human cancers are characterized by extensive intratumor heterogeneity and comprise various subpopulations of cells with distinct phenotypes and biological properties (1, 2). Intratumor heterogeneity poses major challenges in understanding cancers, managing patients and designing effective treatment strategies. Functional heterogeneity within individual tumors is partly due to intercellular genetic variation, which results in the generation of genetically distinct subclonal cell populations (3, 4). Furthermore, tumors have complex architecture, differing regionally in vessel content, stroma, host infiltrates, and other features which can alter the phenotype of genetically-identical cells (5).

In many cancers, phenotypic and functional heterogeneity can be mapped to distinct differentiation states (5–7), suggesting that epigenetic changes occurring during tumor growth may establish cellular hierarchies within the neoplastic mass, thereby affecting the long-term proliferative potential of cancer cells. In line with this notion, individual tumors have been shown to contain distinct subpopulations of undifferentiated, self-renewing cells, and more differentiated cells, which only have limited proliferative ability (8, 9). Regardless of their cell-of-origin, cancer cells endowed with unlimited proliferative potential can be identified by their ability to propagate the disease when transplanted into immunocompromised mice and are referred to as tumor-propagating cells (TPCs) or cancer stem cells (CSCs) (9). The mechanisms through which epigenetic changes occurring during tumor growth establish differentiation hierarchies and contribute to functional heterogeneity within individual tumors are largely unknown.

We have previously shown that hierarchically-organized human primary tumors containing functionally distinct subsets of cancer cells can be generated in a controlled fashion using *in vitro* transformed cells. Expression of human telomerase (hTERT) and oncogenic HRAS[V12], and concomitant inhibition of p53 and pRB by SV40 T antigens confer tumorigenic potential to various cell types (10–12). We have shown that upon *in vitro* transformation of primary dermal fibroblasts, a subpopulation of cells marked by the CSC marker SSEA1 (13, 14) acquire uncontrolled long-term proliferative potential and multipotency (15). Thus, when injected into mice, SSEA1[+] cells form hierarchically-organized tumors in which a small subset of self-renewing SSEA1[+] cells maintain tumor growth, while generating SSEA1[-] differentiated progeny with only limited proliferative capacity (15). The functional differences that distinguish SSEA1[+] and SSEA1[-] tumor cells – undifferentiated phenotype, long-term self-renewal ability and high tumorigenic potential *versus* differentiated phenotype, limited proliferative potential and low tumorigenicity - are general features associated with differentiation hierarchies in many cancer types, regardless of the tissue of origin.

Taking advantage of our ability to generate hierarchically-organized human tumors in a controlled fashion, here we identify an epigenetic mechanism that establishes intratumor functional heterogeneity. We report that reversible silencing of the linker histone H1.0 (16) affects the differentiation state of cancer cells and contributes to define which cells within a tumor can maintain long-term self-renewal potential and drive tumor growth. H1.0 is one of multiple H1 variants. Unlike replication-dependent H1 variants, which are mainly expressed in proliferating cells, H1.0 is expressed in both dividing and non-dividing cells (16). H1.0 levels are low in pluripotent cells, but accumulate in somatic cells, replacing replication-dependent H1 variants (17). Here, we demonstrate a critical role for histone H1.0 in inhibiting tumor maintenance.

Results

Heterogeneous levels of H1.0 in individual tumors

We have previously used an engineered system to model functional intratumor heterogeneity of human tumors (see introduction) (15). In this system, we have shown by gene expression analysis that self-renewing tumor cells, which are marked by the surface antigen SSEA1 (SSEA1$^+$ cells), are molecularly distinct from their SSEA1$^-$ differentiated progeny (Fig. S1A) (15). Among the differentially expressed genes, *H1F0*, which encodes histone H1.0, showed consistent downregulation in SSEA1$^+$ cells compared to SSEA1$^-$ cells in multiple tumors (Fig. 1A; Fig. S1B). Other H1 variants were either expressed at very low levels in all tumor cells (*HISTH1A*, *HISTH1B*, *HISTH1C*, *HISTH1D*, *HISTH1T*, *H1Foo*, *H1FNT*, *H1FX*) or expressed at comparable levels in SSEA1$^+$ and SSEA1$^-$ cells (*HISTH1E*) (Fig. S1B). Analysis of protein levels by quantitative immunofluorescence microscopy of sorted tumor cells and tumor sections confirmed low levels of H1.0 in self-renewing SSEA1$^+$ cells (Fig. 1B,C; Fig. S1C). In contrast, SSEA1$^-$ cells expressed heterogeneous, but overall higher levels of H1.0 (Fig. 1B,C; Fig. S1C), similarly to hTERT-immortalized, non-transformed fibroblasts, from which SSEA1$^+$ cells were originally derived (Fig. 1B,C). As control, comparable levels of H1.4 (*HIST1H1E* gene product) were detected in SSEA1$^+$ and SSEA1$^-$ cells (Fig. 1B; Fig. S1D). Low abundance of H1.0 in SSEA1$^+$ cells was further confirmed in unsorted tumor cells by imaging flow cytometry, which showed a mutually exclusive relationship between H1.0 and SSEA1 (Fig. 1D,E). H1.0 levels negatively correlated with the presence of the mitotic marker phopho-H3S28 in tumors, confirming an association between H1.0 levels and cell proliferative potential (Fig. S2).

We next examined H1.0 levels in clinical samples from cancer patients. Glioblastoma multiforme (GBM) is an aggressive brain cancer, characterized by highly undifferentiated cells. SSEA1 is an established GBM stem cell marker that enriches for tumor-propagating cells (13, 14). Single-cell analysis of normal brain and GBM tissue sections by quantitative immunofluorescence microscopy showed overall reduced levels of H1.0, but not H1.4, in cancer samples (Fig. 1F,G; Fig. S3A). However, while the SSEA1$^-$ cell population was heterogeneous and included cells expressing H1.0 at comparable levels as normal brain cells (*H1F0*^{high} cells), SSEA1$^+$ cells consistently exhibited low levels of the protein (*H1F0*^{low} cells) (p < 0.01) (Fig. 1F,G; Fig. S3A). Analysis of single-cell RNA-seq datasets from primary GBMs (2) confirmed these results, since self-renewing GBM cells isolated from

multiple tumors (CSC) showed ~10 times lower average levels of *H1F0* compared to the whole tumor cell population and ~4 times lower compared to cells cultured under differentiating conditions (FCS) (Fig. 1H). Although the average *H1F0* levels were variable across GBM samples, the range of *H1F0* levels in individual cells was similar in all analyzed tumors (Fig. S3B), indicating that bulk measurements of the heterogeneous GBM populations mainly reflected the relative abundance of *H1F0*high and *H1F0*low cells within each population. Furthermore, analysis of gene expression datasets from The Cancer Genome Atlas (TCGA) showed significantly higher bulk levels of *H1F0* in lower grade glioma (LGG) compared to grade 4 GBM ($p < 0.001$) (Fig. S3C), suggesting that more aggressive brain tumors contain more *H1F0*low cells than well- and moderately-differentiated tumors.

Intratumor heterogeneity of H1.0 levels was also observed in other types of cancer. In breast cancer (BRCA), the abundance of H1.0low cells correlated with histopathological grade, being lower in well-differentiated cancers (grade I) and higher in moderately- (grade II) and poorly-differentiated (Grade III) tumors (Fig. 1J; Fig. S3D). Levels of H1.0 were particularly low ($p < 0.01$) in cells expressing ITGA6, a surface antigen marking breast CSCs (18), which has been shown to have prognostic value as a single parameter (19) (Fig. 1I,K). As control, H1.4 was not affected in BRCA samples (Fig. S3D). Similarly, analysis of samples from patients affected by stomach, prostate, uterus and ovary cancer revealed heterogeneous H1.0 levels in tumor sections, while homogeneous high levels were detected in the corresponding normal tissues (Fig. S4A). A panel of 20 cell lines from 10 cancer types also showed overall reduced levels of H1.0 (Fig. S4B,C). In culture, H1.0 levels were sensitive to extracellular signaling, as assessed by treatment of cells with various ligands previously implicated in mediating interactions between cancer cells and the tumor microenvironment. Out of 16 proteins tested, including cytokines, activators or inhibitors of stem cell-related pathways and growth factors, 9 increased H1.0 levels while 2 further lowered them ($p < 0.05$) (Fig. S4D). This observation suggests that heterogeneous H1.0 expression patterns within tumors may be the result of differential exposure of cancer cells to extracellular cues.

Taken together, these results indicate that several cancer types exhibit highly heterogeneous expression patterns of H1.0 within individual tumors, and that cells characterized by an undifferentiated phenotype and expressing functionally validated CSC markers contain particularly low levels of H1.0.

Dynamic *H1F0* methylation in tumors

Aberrant DNA-methylation patterns contribute to gene downregulation in cancer (8). *H1F0* is unmethylated and highly expressed in most adult tissues (Fig. S3A,D; Fig. S4A) (17) (Fig. 2A; Fig. S5A). To assess whether DNA methylation affects *H1F0* levels within tumors, we performed bisulfite sequencing analysis of the *H1F0* locus, comparing self-renewing SSEA1+ cells with the most differentiated tumor cells, characterized by the absence of SSEA1 and reduced levels of CD166 (SSEA1-/CD166low) (15) (see Fig. 3D). We examined a region spanning ~1.1 Kb around the transcriptional start site (TSS) (from -263 to +858), including a CpG island (CGI) (Fig. 2A). The CGI in the *H1F0* promoter was not methylated

in either subset (Fig. S5B), but a differentially-methylated region was identified in the adjacent CGI shore (20) (Fig. 2A). SSEA1[+]/CD166[high] cells showed considerable methylation in multiple consecutive CpGs, (26% to 94%, average: 50%) whereas significantly reduced methylation (0% to 37%, average: 18%) was detected in differentiated SSEA1[-]/CD166[low] cells (p <0.001) (Fig. 2B). Treatment of isolated SSEA1[+] tumor cells with the DNA methyltransferase inhibitor 5-Aza-2′-Deoxycytidine resulted in demethylation of the CGI shore and a progressive increase of *H1F0* mRNA with similar kinetics, while *H1FX,* as control, remained constant (Fig. 2C; Fig. S5C-D; see Methods). Chromatin immunoprecipitation (ChIP) revealed high levels of H3K27ac in the CGI shore, suggesting the presence of an enhancer element (21) (Fig. S5E,F). In agreement, the *H1F0* CGI shore showed trans-activating potential in reporter assays (Fig. 2D), and its deletion by CRISPR-Cas9-mediated genome editing resulted in decreased levels of endogenous *H1F0* mRNA (Fig. S5G,H). Furthermore, targeted methylation of the enhancer region by CRISPR-Cas9-DNMT3A (22) (~65% methylation) induced a ~45% reduction in *H1F0* mRNA levels (Fig. 2E and Fig. S5I), and *in vitro*-methylated reporter constructs showed a similar reduction in luciferase assays (Fig. 2F; Fig. S5J). We conclude that differential methylation of an enhancer controlling *H1F0* expression induces *H1F0* silencing in self-renewing tumor cells.

We then analyzed expression and methylation datasets from various cancer types generated by TCGA. For each cancer type, correlation between *H1F0* mRNA levels measured by RNA-seq and *H1F0* methylation assessed by 450K Infinium microarrays was examined. Similarly to what we observed in tumors induced by *in vitro*-transformed cells, the *H1F0* promoter (450K probes 1-4) was not methylated in most patients (Fig. 2A,G; Fig. S6), whereas a downstream region centered on the CGI shore (450K probes 5-10) showed variable degrees of methylation, which inversely correlated with *H1F0* mRNA levels (Fig. 2A,G; Fig. S6). Out of the 27 analyzed types of cancer, 26 showed a significant inverse correlation between *H1F0* mRNA levels and DNA methylation (p-value range: p < 2.2 e-16 to p = 0.009) (Fig. 2G; Fig. S6). Remarkably, the location of the methylated region within *H1F0* was the same in all cancers, with the CGI shore (probes 7 and 8) showing the highest correlation with expression levels. In several cohorts, patients characterized by high *H1F0* methylation showed a significantly higher proportion of aggressive tumors, such as triple negative breast cancers and high-grade glioma, kidney cancer and stomach adenocarcinoma (p-value range: p < 0.0001 to p = 0.016) (Table S1). Bisulfite sequencing analysis of multiple cell lines from various cancer types confirmed methylation of the *H1F0* CGI shore in clinically-derived samples expressing low H1.0 levels (Fig. S5K and S4B,C). We conclude that methylation of the *H1F0* CGI shore is associated with silencing of *H1F0* in a large variety of cancers.

Chromatin-mediated inhibition of cancer cell self-renewal

To test whether changes in H1.0 levels play a functional role in generating functional heterogeneity within tumors, we assessed whether restoring high levels of H1.0 in self-renewing SSEA1[+] cells during tumor growth or, conversely, inhibiting re-expression of the protein in their differentiated progeny, affects tumor organization and cancer cell long-term proliferative potential. To do so, we introduced into *in vitro*-transformed cells lentiviral

constructs expressing either H1.0 cDNA or shRNAs targeting *H1F0* (shH1.0) under a Doxycycline (Dox)-inducible promoter. Induction of H1.0 cDNA *in vitro* resulted in a ~4-fold increase in *H1F0* mRNA (Fig. S7A), whereas expression of two distinct shRNAs reduced mRNA levels by ~80% (Fig. S7B-E), leaving other H1 variants unaffected (Fig. S7F,G). Cell lines expressing H1.4 cDNA or an H1.4-targeting shRNA were also generated as controls (Fig. S7A,B). Neither forced expression nor knockdown of H1.0 or H1.4 affected cell viability, but cells expressing exogenous H1.0 showed decreased proliferation rate *in vitro* (Fig. S8 A,B). Although the overall population of H1.0 expressing cells did not undergo immediate cell-cycle arrest and most cells were positive for the proliferation marker Ki67 early upon H1.0 expression, the fraction of Ki67-negative cells increased over time (p = 0.014) (Fig. S8C,D), similarly to what is observed upon cell differentiation (23, 24), suggesting that constitutive expression of H1.0 impairs cell long-term proliferative potential. In agreement, cells expressing high levels of exogenous H1.0 were negatively selected over time in favor of those expressing lower levels (Fig. S8E). Progressive negative selection of Ki67$^-$/H1.0^{high} cells over passages hindered detection of complete arrest of the whole cell population.

To test whether changes in H1.0 levels within established tumors affect tumor organization, cells containing, but not expressing, inducible H1.0 cDNA or H1.0-targeting shRNAs were injected into *NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ* (NSG) mice to induce tumor formation. About 4 weeks after injection, when palpable tumors appear, constitutive expression or knockdown of H1.0 was induced and sustained for ~4 weeks while tumors grew (Fig. 3A). Although analysis of control EGFP-expressing tumors showed efficient *in vivo* induction of the Dox-responsive construct (5/5 fluorescent tumors) (Fig. S8F), when H1.0 was expressed, only 4/11 induced tumors showed higher *H1F0* mRNA levels compared to the uninduced tumors (Fig. 3B,C). This was not due to inefficient Dox delivery to the tumors, since dissociated tumor cells also failed to induce H1.0 *in vitro* (Fig. S8G,H), and suggested that cells expressing constitutively high levels of H1.0 were negatively selected during tumor growth, similarly to what we observed *in vitro*. As control, treatment of tumors with Dox for only 3 days resulted in detectable expression of exogenous H1.0 in 4/4 tumors (Fig. 3B; Supplementary text). In contrast, when H1.0 knockdown was induced, no negative selection was observed and all Dox-treated tumors contained high percentages of cells expressing H1.0-targeting shRNAs (9/9) (Fig. S8I). Thus, constitutively high levels of H1.0 appear to inhibit the long-term proliferative potential of cells within tumors.

To directly assess whether modulation of H1.0 levels affects tumor organization, we first measured the relative abundance of cells expressing markers of an undifferentiated (SSEA1$^+$/CD166^{high}) or a differentiated (SSEA1$^-$/CD166^{low}) phenotype (15) in uninduced and induced tumors (Fig. 3D). Constitutive expression of H1.0 resulted in a significant decrease in the fraction of undifferentiated tumor cells (p = 0.039) and a concomitant increase in differentiated cells (p = 0.005) in the tumors showing increased *H1F0* levels (Fig. 3E). Furthermore, the percentage of self-renewing tumor cells, as measured by clonogenic assays *in vitro*, was ~2-fold lower in induced tumors (p = 0.045) (Fig. 3E,F). As control, expression of H1.4 had no significant effect on tumor organization (Fig. S9A). Conversely, when knock-down of H1.0 was induced to prevent its accumulation in differentiated cells, the percentage of undifferentiated tumor cells significantly increased (p = 0.005), the

fraction of differentiated cells decreased (p = 0.007) and tumors contained ~ 2-fold more self-renewing cells (p = 0.044) (Fig. 3G; Fig. S9D; Supplementary text). In agreement, tumors expressing constitutive H1.0 showed a lower fraction of mitotic cells, while tumors in which H1.0 had been knocked-down had a higher mitotic index (Fig. S9B,C). Limiting dilution transplantation assays for secondary tumor formation confirmed altered self-renewal ability of cancer cells *in vivo*, showing a lower frequency of tumorigenic cells in primary tumors constitutively expressing H1.0, and a higher frequency in tumors where H1.0 had been knocked-down (Fig. 3H; Supplementary text and Table S2). In line with the notion that even moderate changes in the fraction of self-renewing cancer cells strongly affects long-term tumor growth (18, 25, 26) (Supplementary text), transplantation of cells from primary tumors constitutively expressing H1.0 resulted in significantly delayed appearance of secondary tumors compared to control tumors (p < 0.01), whereas faster growth was observed upon transplantation of shH1.0-expressing cells (p < 0.01) (Fig. 3I). Orthotropic transplantation assays using multiple breast cancer cell lines confirmed these results. Forced expression of H1.0 cDNA in MDA-MB-231 or HCC1954 cells specifically impaired tumor growth, while knockdown of the endogenous protein in MDA-MB-231 or HCC-1569 cells resulted in more aggressive tumors (Fig. 3J,K). Furthermore, analysis of Dox-induced tumors showed partial negative selection of H10[high] cells, confirming the negative impact of H1.0 on cell long-term proliferative potential (Fig. S8J,K). We conclude that H1.0 silencing is required for the ability of cancer cells to self-renew, and that its re-expression in subsets of cells during tumor growth contributes to drive their differentiation and restrict their proliferative potential.

De-repression of AT-rich genes sustains cancer cell self-renewal

To understand how H1.0 affects cancer cell self-renewal, we employed two inducible shRNAs to knockdown H1.0 in transformed cultured cells and compared the transcriptional profiles of three cellular states by RNA-seq: uninduced cells expressing H1.0 (NT), cells induced with Dox for 14d, which downregulated H1.0 similarly to self-renewing SSEA1[+] tumor cells (DOX), and cells washed out of Dox for 4 days, which began to re-express H1.0 and mimicked differentiated SSEA1[-] tumor cells (washDOX). Comparison between NT and DOX samples detected 860 differentially expressed genes (DEGs) (FDR < 0.05) upon H1.0 knockdown by both shRNAs (475 upregulated and 385 downregulated) (Table S3; Fig. S10A-C). The extent of gene expression changes was moderate, with only 25 genes showing > 2-fold changes (Table S3), but differences were highly consistent and reversible upon H1.0 re-expression (Fig. S10C,D). qRT-PCR analysis confirmed differential expression of genes detected by RNA-seq and showed that H1.0-sensitive genes did not respond to H1.4 knock-down (Fig. S10E).

To identify H1.0-sensitive genes important for cancer cell self-renewal, we compared the identified DEGs with a gene signature (132 genes) defined by comparing self-renewing SSEA1[+] and differentiated SSEA1[-] cells from multiple tumors (Table S4) (15). 33% of genes expressed at higher levels in SSEA1[+] cells underwent upregulation upon H1.0 knockdown (p < 10 e-14), whereas 10% of genes downregulated in SSEA1[+] cells showed lower levels in cells lacking H1.0 (p = 1.5 e-5), indicating that a significant fraction of genes defining the subset of self-renewing tumor cells, particularly those upregulated, are regulated

by H1.0 (Fig. 4A). Gene set enrichment analysis (GSEA) confirmed that genes upregulated upon H1.0 knockdown were particularly relevant for cancer cell self-renewal ability (Fig. 4B). Several oncogenic gene signatures, such as active KRAS, EGFR and RAF, and PTEN loss-of-function signatures, were enriched in the subset of upregulated genes, as well as Polycomb-related signatures, such as EZH2, EDD, BMI and SUZ12, which are important for stem cell maintenance (27) (Fig. 4B). Based on GSEA, downregulated genes did not seem to substantially contribute to the self-renewing phenotype. Thus, loss of H1.0 results in activation of downstream effectors of multiple oncogenic and self-renewal cellular pathways. Interestingly, these genes are not randomly distributed in the genome but are enriched in AT-rich regions (Fig. S11A).

GSEA also revealed linear proximity of H1.0-sensitive genes along chromosomes. Of 296 positional gene sets, corresponding to genes located in chromosomal cytogenetic sub-bands, 75 were enriched in DOX samples, using either H1.0-targeting shRNA, indicating overall upregulation of genes located in those bands upon H1.0 knockdown, whereas 37 bands negatively correlated with the DOX samples, indicating overall gene downregulation (Fig. 4C; Fig. S11 B,C; Table S5). In agreement, analysis of the distribution of gene expression fold-changes between NT and DOX samples (Methods) showed that domains of upregulated genes alternated with downregulated regions along chromosomes (Fig. 4D). The differentially-expressed domains contained up to 365 genes over several megabases (average genes/domain = 30) (Fig. S11E). A similar profile was observed 24h after Dox induction, when H1.0 levels began to decrease, indicating that transcriptional changes occur rapidly upon H1.0 loss (Fig. S11F,G). Re-expression of H1.0 in washDOX samples, mimicking what happens in differentiated tumor cells, restored transcriptional profiles similar to those detected in uninduced NT samples (Fig. 4E; Fig. S11F,G), indicating that gene expression changes induced by H1.0 silencing are reversible.

Upregulated domains were enriched in AT-rich chromosomes, such as chr.4, chr.5, chr.18 (average GC content <0.40), whereas downregulated domains mainly localized in GC-rich chromosomes such as chr.17, chr.19, chr.22 (average GC content > 0.45) (28) (Fig. 4D). In agreement, individual upregulated genes had significantly lower GC content compared to all RefSeq genes (median %GC: 0.40 Vs 0.45, p < 10 e-16) or downregulated genes (median %GC: 0.40 Vs 0.53, p < 10 e-16) (Fig. 4F), even though the two groups of H1.0-sensitive genes contained a similar proportion of CGI-containing genes (Fig. S11D). We conclude that cells lacking H1.0 undergo genome-wide changes in gene expression, affecting large gene domains in a coordinated manner. In particular, upregulation of sets of neighboring AT-rich genes mediating oncogenic cellular responses and stem cell maintenance contributes to the activation and maintenance of transcriptional programs that support uncontrolled self-renewal. These effects are reversible and re-expression of H1.0 re-establishes gene expression profiles that restrict cell tumorigenic potential.

Destabilized nucleosome-DNA interactions in self-renewing cancer cells

To map genomic regions bound by H1.0 prior to its downregulation in self-renewing tumor cells, we performed ChIP-seq analysis of transformed cultured cells. Approximately 120,000 H1.0 peaks were consistently detected in two biological replicates (Fig. S12A-C; Table S6),

using an H1.0-specifc antibody (Fig. S7C-E). ChIP-qPCR confirmed accurate detection of H1.0 peaks (Fig. S12E). H1.0 binding sites positively correlated with the histone mark H3K27me3 (Fig. 5A; Fig. S12B,D). No significant correlation was observed between H1.0 binding profiles and maps of topological associating domains (TADs) (29) (Fig. S12D). H1.0 binding sites were enriched in GC-rich chromosomes and at GC-rich genes (Fig. S12F,G) and H1.0 peak density showed a tight linear relationship with GC content in regions with > 40% GC (R = 0.86) (Fig. 5B), indicating that GC-content is a major determinant of H1.0 binding sites. Megabase-scale density plots of H1.0 binding sites along chromosomes allowed visualization of the correlation between H1.0 occupancy, GC content and cytogenetic bands (Fig. 5C; Fig. S13A). In line with the notion that nucleosomes are overall enriched in GC-rich genomic regions, H1.0 occupancy also correlated with nucleosome density (30, 31) (Fig. S12H). We conclude that large-scale H1.0 binding profiles are mainly determined by DNA sequence, which affects the broad distribution of nucleosomes. Of note, AT-rich regions characterized by low H1.0 peak density contain the self-renewal-related genes upregulated upon H1.0 knockdown (Fig. 5C; Fig. S13A).

To gain insights into how H1.0 loss results in upregulation of AT-rich self-renewal-related genes, we examined H1.0 occupancy around their TSS. Although upregulated genes were bound by H1.0 at overall low density (Fig.5 D,E), similarly to the pattern observed at other AT-rich regions, H1.0 binding sites were highly enriched around the TSS (Fig. 5F,G; Fig. S12J-L). In contrast, all RefSeq genes and genes downregulated upon H1.0 knockdown showed an average depletion of H1.0 binding sites around the TSS (Fig. 5F; Fig. S12J-L). Similar patterns, with enhanced differences, were observed when analyzing the positional gene sets identified as upregulated by GSEA (Fig.13B-D). Together with the notion that AT-rich regions may be thermodynamically unstable when wrapped around nucleosomes (32, 33), this observation suggests that H1.0 may repress the AT-rich, self-renewal-related genes by stabilizing nucleosomes at their promoters. To test this possibility, we assessed changes in genome-wide nucleosome occupancy induced by H1.0 loss by FAIRE-seq (34), which allows detection of nucleosome-depleted regions irrespectively of DNA GC-content (35, 36) (Fig. S14A). Upregulated genes specifically showed increased FAIRE signal around the TSS upon H1.0 knockdown using two independent H1.0-targeting shRNAs, indicating decreased nucleosome occupancy (Fig. 5H; Fig. S13E; Fig. S14B-C). Similar patterns were observed at distal regulatory regions of H1.0-sensitive genes, identified by comparing an H3K27ac map with an atlas of human enhancers (21) (Methods) (Fig. S14F,G). As control, knock-down of H1.4, which was depleted at H1.0-sensitive genes (Fig. S12I), only induced minimal changes in promoter nucleosome occupancy (Fig. S14H), which did not translate into changes in gene expression (Fig. S10E). Nucleosome remodeling in the absence of H1.0 did not correlate with increased occupancy of other nucleosome-binding architectural proteins such as HMGA1, HMGN2 and PARP1 (Fig. S15).

In agreement with H1.0 distribution, genome-wide analysis of differential FAIRE peaks indicated that altered nucleosome occupancy in the absence of H1.0 depends on GC content, both at a global and local scale. In GC-rich domains (average GC content/500bp > 0.45) most detected nucleosome-depleted regions were stable and unaffected by H1.0 knockdown (constitutive FAIRE peaks, Table S6), and only ~25% either appeared or disappeared as a consequence of altered nucleosome occupancy (Fig. 5I). In contrast, in AT-rich domains

(average GC content/500bp < 0.45) ~50% of detected nucleosome-depleted regions showed altered nucleosome occupancy, indicating that nucleosome-DNA interactions in AT-rich regions become particularly unstable in the absence of H1.0 (Fig. 5I). Furthermore, nucleosome-depleted regions appeared in DNA regions with local minimal GC content, whereas they disappeared from regions of local maximal GC content, suggesting that nucleosomes forced to occupy regions with higher AT content, in the absence of H1.0 moved to locations with higher GC-content (Fig. 5J; Fig. S14D).

Mapping of H3K27ac and H3K27me3 by ChIP-seq indicated that altered nucleosome occupancy induced by H1.0 loss was uncoupled from changes in histone modifications. Most appearing or disappearing FAIRE peaks were located within non-acetylated chromatin (Fig. S14E), suggesting that alterations in nucleosome occupancy were not a consequence of transcription. Furthermore, upregulated self-renewal-related genes did not show significant differences in H3K27ac and H3K27me3 at their promoters upon H1.0 knockdown compared to downregulated genes ($p > 0.05$) (Fig. 5K), indicating that changes in gene expression were a direct consequence of altered nucleosome occupancy. Taken together, these results suggest that loss of H1.0 destabilizes nucleosome-DNA interactions in AT-rich genomic regions and increases accessibility to regulatory elements important for cellular self-renewal.

Low *H1F0* levels predict negative patient outcome in multiple cancer types

To assess the clinical relevance of H1.0 alterations, we examined whether *H1F0* levels stratify cancer patients. Kaplan-Meier and multivariate analysis of three GBM datasets showed a significant correlation between low *H1F0* levels and negative patient outcome ($p < 0.05$) (Fig. 6A; Fig. S16C), revealing a prognostic value for *H1F0*. Low *H1F0* levels also predicted poor patient prognosis in three BRCA datasets (Fig. 6B). Importantly, *H1F0* correlated with BRCA patient survival independently of ER, PR, Her2 and lymph node status and effectively stratified patients typically associated with negative outcome (ER or PR negative and lymph node positive patients) ($p<0.05$) (Fig. S16A,C). Furthermore, *H1F0* expression showed no significant association with breast cancer subtype, which strongly affects patient outcome (Fig. S16B). Analysis of 17 other types of cancers using TCGA datasets showed a significant correlation between low *H1F0* levels and poor patients prognosis also in melanoma (SKMC), liver cancer (LIHC), kidney cancer (KIRP) and low-grade glioma (LGG), independently of other clinically-relevant features (Fig. 6C; Fig. S16C). As controls, other H1 variants either did not stratify patients or correlated with patient survival in an opposite manner compared with *H1F0* (Fig. S17). Thus, alterations in H1.0 levels are clinically relevant in multiple cancer types.

Discussion

Heterogeneity among cancer cells within individual tumors has emerged as a general feature of cancer, with critical implications for cancer diagnosis and treatment (4, 9). Increasing evidence points to the existence of both genetic and non-genetic sources of intratumor heterogeneity (4, 5, 14, 37, 38), but the molecular features underlying functionally diverse cellular phenotypes have been elusive. Cell-to-cell signaling such as Wnt, Tgfβ and Notch pathways, have been shown to play critical roles in driving functional heterogeneity within

tumors (38), but very little is known about the downstream, cell-intrinsic mechanisms that translate signaling into differential cell function. Here, we show that distinct epigenetic states determined by an integral component of chromatin, define cellular subpopulations that differentially contribute to tumor maintenance. The observation that experimental modulation of epigenetic states affects the balance between self-renewing and differentiated cancer cells demonstrates that chromatin-based mechanisms mediating differentiation programs play a key role in specifying tumor organization and affecting tumor maintenance. In line with the notion that epigenetic mechanisms acting during tumor growth may be dominant over genetic alterations that initiate the disease (37, 39), we show that subsets of cells which stably silence H1.0 preserve their ability to proliferate indefinitely, whereas cells that re-express the protein and, as a consequence, repress oncogenic gene networks, acquire a differentiated phenotype characterized by limited proliferative potential. Reversible changes in H1.0 levels may be due to interactions of cancer cells with the tumor microenvironment (38). In line with this notion, H1.0 expression responds to a variety of extracellular cues, many of which are associated with cellular differentiation (40, 41) (Fig. S4D). It is likely that at least in the early stages of tumor development, cancer cells may be exposed to differentiation stimuli that support normal tissue homeostasis, and subsets of cells may respond to such stimuli, changing their epigenetic landscape, partly through H1.0, and losing self-renewal potential. Thus, H1.0 may act as a downstream effector of extracellular signaling inhibiting cancer cell-self renewal.

Our results suggest that numerous cancer types may share similar epigenetic heterogeneity. Numerous solid tumors show heterogeneity of H1.0 and a regulatory region within the *H1F0* gene shows variable degrees of DNA methylation in 26 types of cancers, correlating with *H1F0* expression levels. Furthermore, low *H1F0* levels are independent predictors of poor patient outcome in 6 prevalent cancer types. Importantly, the molecular mechanism through which H1.0 restricts cell proliferative potential is largely dependent on DNA sequence and GC content, suggesting that changes in H1.0 levels may have similar consequences in various cell types. We show that silencing of *H1F0* supports cancer cell self-renewal by inducing simultaneous de-repression of downstream effectors of oncogenic and stem cell-related pathways located in AT-rich genomic regions. AT-rich regions are known to poorly incorporate into nucleosomes, partly because of their inherent relative rigidity and difficulty to bend around the core histone particle (30–33). Thus, it is conceivable that AT-rich regions may be particularly dependent on the linker histone to stabilize nucleosome positioning at unfavorable sequences, especially when nucleosomes are forced to occupy promoters of genes that must be repressed. Indeed, we find lower nucleosome occupancy at AT-rich gene promoters in the absence of H1.0, correlating with enhanced transcription.

In conclusion, our results uncover epigenetic determinants of tumor-maintaining cells and identify an integral component of chromatin as an important regulator of cell differentiation states within tumors. We propose that only cells insensitive to extracellular differentiation cues, capable of permanently silencing H1.0, can act as self-renewing tumor-maintaining cells, and that such a mechanism supports maintenance of several types of cancer. Our results suggest that intervention aimed at restoring high levels of H1.0 in all cancer cells may enhance the differentiation process that naturally occurs during tumor growth and may be beneficial for therapeutic purposes.

Material and Methods

Cell lines and constructs

Growth conditions of all cell lines used in the study are listed in Table S7. Inducible cell lines were generated by introducing lentiviral constructs (original or modified pTRIPZ, Open Biosystems) expressing specific cDNAs, shRNAs or control plasmids into *in vitro*-transformed fibroblasts (15) or breast cancer cells lines. All cancer cell lines were sourced from the Crick Institute common repository, authenticated by STR profiling and tested for mycoplasma. Cell proliferation rate was measured using the 96 Aqueous One Solution kit (Promega). For details on cell lines generation, see the supplementary Material and Methods (SM).

Patient samples

Normal brain and GBM (grade 4 astrocytoma) tissue sections were obtained from the Whittington tissue bank (UCL license number: 12055, 3 normal and 4 tumors) and from US Biomax (tissue microarray GL806b). All other tissue sections were obtained from US Biomax (tissue microarrays: BR1053b, BCN962). Single-cell RNA-Seq data of GMB tumors was downloaded from GEO series GSE57872 (2). DNA methylation and survival analysis were performed using TCGA datasets. For details on patients used in the study, see SM.

Protein immunodetection

Western blot analysis, flow cytometric analysis, cell sorting and immunofluorescence microscopy of cultured or sorted tumor cells were performed as previously described (15) using anti-H1.0 (Millipore, raised in mouse, clone 3H9, 1:300, originally produced in M. Bustin's laboratory) (42), and other antibodies detailed in SM. Recombinant proteins were purchased from Peprotech. Quantification of the fluorescent signal was performed using Metamorph software or with HCS Studio Cell Analysis Software. For details about staining procedures, see SM.

Tumorigenicity assays

Tumor formation in NSG mice, tumor dissociation, limiting dilution transplantation assays, soft agar assays and proliferation assays were performed as previously described (15). For experiments requiring *in vivo* modulation of H1 variants in established tumors, cells infected with inducible constructs were injected intradermally into NSG mice (2,000 transformed cells and 100,000 hTERT-fibroblasts as carrier cells) in the uninduced state to induce tumor formation (10 injections per condition). When tumors first became palpable, typically after 4 weeks, doxycycline treatment (2μg/ml Dox in drinking water supplemented with 1% sucrose, changed every 2-3 days) was started and sustained for about a month until tumors were collected and analyzed as detailed in SM. Animal studies were conducted in agreement with the approved NCI animal protocol LRBGE-007 and the Crick project license PPL 70/8167.

DNA methylation analysis

For bisulfite sequencing analysis, genomic DNA was treated for bisulfite conversion using EZ DNA Methylation-Direct Kit (Zymo research) according to the manufacturer's instructions. Three regions of *H1F0* (CGI_1, CGI_2 and CGIshore) were amplified by PCR (primer sequences in Table S8) from bisulfite-treated DNA and cloned into pCR 2.1 Topo vector. 15-20 colonies were sequenced for each region. For 5-Aza-2'-deoxycytidine (5-Aza) treatment, SSEA1[+] cells isolated from a tumor were treated with 5nM 5-Aza (Sigma) or DMSO as control (Fisher Chemical) for up to 14 days. Higher concentrations of 5-Aza were toxic to the SSEA1+ cells. For analysis of TCGA samples, gene expression (Illumina HiSeq RNA-seq) and DNA methylation (Illumina 450K Infinium analysis) datasets from individual cancers downloaded from the UCSC Cancer browser https://genome-cancer.ucsc.edu were analyzed as detailed in SM.

CGI shore functional assays

Luciferase assays were performed using the Nano-Glo Dual-Luciferase reporter assay system (Promega) comparing the *H1F0* CGI shore and two previously published DNA fragments showing no transactivation potential (21, 43). Deletion of the endogenous *H1F0* CGI shore was carried out by CRISPR-Cas9-mediated genome editing using two sgRNAs flanking the CGI shore (Table S8). Clones with homozygous deletions of the *H1F0* CGI shore were analyzed by qRT-PCR using primers located downstream of the deletion. For heterozygous clones, allele-specific primers probing either the deleted or the full-length mRNA were used (Table S8). To induce targeted methylation of the endogenous *H1F0* CGI shore, cells were transiently transfected with plasmids encoding either WT or ANV mutant Cas9-DNMT3A (22) and a pool of 4 sgRNAs targeting *H1F0* CGI shore (Table S8). For details on luciferase assays and CRISPR-mediated alterations of the *H1F0* CGI shore, see SM.

RNA-seq analysis and GSEA

RNA libraries prepared using TruSeq Stranded Total RNA-Seq Library Prep (Illumina) were sequenced on a HiSeq 2500 sequencer (Table 6). Differentially-expressed genes were defined as those showing statistically significant differences (FDR <0.05) using both H1.0-targeting shRNAs. Genome-wide expression plots were generated using similar methods to those described in (44). Deviation from random distribution was calculated on raw, unsmoothed data, by performing a Montecarlo simulation in which genes where randomly shuffled along chromosomes (1000 iterations). Gene set enrichment analysis (GSEA) was performed using GSEA software (version 2.1.0, Broad Institute). FPKM values for NT and DOX samples were provided to the algorithm and tested for enrichment of oncogenic signatures from the MSigDB or custom-made positional gene sets generated based on RefSeq genes and cytogenetic sub-bands coordinates downloaded from the UCSC genome browser. For details on RNA-seq analysis, GSEA and validation by qRT-PCR, see SM.

ChIP-seq analysis

Chromatin immunoprecipitation (ChIP) was performed on *in vitro*-transformed fibroblasts using an anti-H1.0 antibody (Millipore, 05-629, originally generated in M. Bustin's

laboratory) (42), and other antibodies detailed in SM. Multiplexed libraries were sequenced on a HiSeq2500 using either 50bp single end or 101bp paired read runs (Table S6). Peak detection for H1.0 was performed with a custom-made algorithm as detailed in SM.

FAIRE-seq analysis

Cells were processed and analyzed as previously described (48). FAIRE peaks were called using MACS-2.0.10 in paired-end mode using broad settings with a bandwidth of 300bp, normalizing each sample to its own background as described (48). The top 100,000 peaks (based on adjusted p-value) were selected for each replicate and those common to both (with a 20% reciprocal overlap) were selected for downstream analysis. Differential FAIRE peaks analysis was performed using DiffReps (49). For details on FAIRE-seq analysis, see SM.

Survival analysis

Survival analysis was performed on all TCGA datasets containing more than 200 patients. Patients were ranked based on *H1F0* expression and the top and bottom tertile were compared with respect to vital status by Kaplan-Meier analysis. Log-rank test was used to assess statistical significance of the differences between patient groups. Additional datasets for GBM and BRCA were analyzed as validation. Multivariate analysis and chi-square test were performed to assess the relationship between H1.0 and other clinically-relevant features. For details on the survival analysis, see SM.

Statistical analysis

Unless otherwise stated in figure legends, data are presented either as individual samples or as mean ± standard error of the mean (SEM) of multiple replicates, with N indicated in the figure legend. In box plots, the boundary of the box closest to zero indicates the 25th percentile, a line within the box marks the median, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers (error bars) above and below the box indicate the 95th and 5th percentiles. Outliers are plotted as dots. The statistical test used for each comparison, whether adjustment for multiple corrections was performed and the p-value are indicated in the corresponding figure legends.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References and Notes

1. Heppner GH. Tumor heterogeneity. Cancer research. 1984; 44:2259–2265. [PubMed: 6372991]

2. Patel AP, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014; 344:1396–1401. [PubMed: 24925914]

3. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976; 194:23–28. [PubMed: 959840]

4. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013; 501:338–345. [PubMed: 24048066]

5. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nature reviews Cancer. 2012; 12:323–334. [PubMed: 22513401]

6. Illmensee K, Mintz B. Totipotency and normal differentiation of single teratocarcinoma cells cloned by injection into blastocysts. Proceedings of the National Academy of Sciences of the United States of America. 1976; 73:549–553. [PubMed: 1061157]

7. Barabe F, Kennedy JA, Hope KJ, Dick JE. Modeling the initiation and progression of human acute leukemia in mice. Science. 2007; 316:600–604. [PubMed: 17463288]

8. Easwaran H, Tsai HC, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. Molecular cell. 2014; 54:716–727. [PubMed: 24905005]

9. Kreso A, Dick JE. Evolution of the cancer stem cell model. Cell stem cell. 2014; 14:275–291. [PubMed: 24607403]

10. Hahn WC, et al. Creation of human tumour cells with defined genetic elements. Nature. 1999; 400:464–468. [PubMed: 10440377]

11. Elenbaas B, et al. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev. 2001; 15:50–65. [PubMed: 11156605]

12. Rich JN, et al. A genetically tractable model of human glioma formation. Cancer research. 2001; 61:3556–3560. [PubMed: 11325817]

13. Son MJ, Woolard K, Nam DH, Lee J, Fine HA. SSEA-1 is an enrichment marker for tumor-initiating cells in human glioblastoma. Cell stem cell. 2009; 4:440–452. [PubMed: 19427293]

14. Suva ML, et al. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. Cell. 2014; 157:580–594. [PubMed: 24726434]

15. Scaffidi P, Misteli T. In vitro generation of human cells with cancer stem cell properties. Nature cell biology. 2011; 13:1051–1061. [PubMed: 21857669]

16. Izzo A, Kamieniarz K, Schneider R. The histone H1 family: specific members, specific functions? Biological chemistry. 2008; 389:333–343. [PubMed: 18208346]

17. Doenecke D, Alonso A. Organization and expression of the developmentally regulated H1(o) histone gene in vertebrates. The International journal of developmental biology. 1996; 40:395–401. [PubMed: 8735954]

18. Pece S, et al. Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. Cell. 2010; 140:62–73. [PubMed: 20074520]

19. Ali HR, et al. Cancer stem cell markers in breast cancer: pathological, clinical and prognostic significance. Breast cancer research: BCR. 2011; 13:R118. [PubMed: 22112299]

20. Irizarry RA, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009; 41:178–186. [PubMed: 19151715]

21. Andersson R, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014; 507:455–461. [PubMed: 24670763]

22. Vojta A, et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic acids research. 2016

23. Jarvinen E, Angers-Loustau A, Osiceanu AM, Wartiovaara K. Timing of the cell cycle exit of differentiating hippocampal neural stem cells. International journal of stem cells. 2010; 3:46–53. [PubMed: 24855540]

24. Snir M, et al. Assessment of the ultrastructural and proliferative properties of human embryonic stem cell-derived cardiomyocytes. American journal of physiology Heart and circulatory physiology. 2003; 285:H2355–2363. [PubMed: 14613910]

25. Prost S, et al. Erosion of the chronic myeloid leukaemia stem cell pool by PPARgamma agonists. Nature. 2015; 525:380–383. [PubMed: 26331539]

26. Wang Y, et al. The long noncoding RNA lncTCF7 promotes self-renewal of human liver cancer stem cells through activation of Wnt signaling. Cell stem cell. 2015; 16:413–425. [PubMed: 25842979]

27. Aloia L, Di Stefano B, Di Croce L. Polycomb complexes in stem cells and embryonic development. Development. 2013; 140:2525–2534. [PubMed: 23715546]

28. Costantini M, Clay O, Auletta F, Bernardi G. An isochore map of human chromosomes. Genome research. 2006; 16:536–541. [PubMed: 16597586]

29. Dixon JR, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–380. [PubMed: 22495300]

30. Valouev A, et al. Determinants of nucleosome organization in primary human cells. Nature. 2011; 474:516–520. [PubMed: 21602827]

31. Tillo D, Hughes TR. G+C content dominates intrinsic nucleosome occupancy. BMC bioinformatics. 2009; 10:442. [PubMed: 20028554]

32. Anderson JD, Widom J. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. Molecular and cellular biology. 2001; 21:3830–3839. [PubMed: 11340174]

33. Vinogradov AE. Bendable genes of warm-blooded vertebrates. Molecular biology and evolution. 2001; 18:2195–2200. [PubMed: 11719569]

34. Hogan GJ, Lee CK, Lieb JD. Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. PLoS genetics. 2006; 2:e158. [PubMed: 17002501]

35. Koohy H, Down TA, Hubbard TJ. Chromatin accessibility data sets show bias due to sequence specificity of the DNase I enzyme. PloS one. 2013; 8:e69853. [PubMed: 23922824]

36. Dingwall C, Lomonossoff GP, Laskey RA. High sequence specificity of micrococcal nuclease. Nucleic acids research. 1981; 9:2659–2673. [PubMed: 6269057]

37. Dalerba P, et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nature biotechnology. 2011; 29:1120–1127.

38. Kise K, Kinugasa-Katayama Y, Takakura N. Tumor microenvironment for cancer stem cells. Adv Drug Deliv Rev. 2016; 99:197–205. [PubMed: 26362921]

39. Shipitsin M, et al. Molecular definition of breast tumor heterogeneity. Cancer cell. 2007; 11:259–273. [PubMed: 17349583]

40. Alonso A, Breuer B, Bouterfa H, Doenecke D. Early increase in histone H1(0) mRNA during differentiation of F9 cells to parietal endoderm. The EMBO journal. 1988; 7:3003–3008. [PubMed: 2846273]

41. Terme JM, et al. Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. The Journal of biological chemistry. 2011; 286:35347–35357. [PubMed: 21852237]

42. Mendelson E, Bustin M. Monoclonal antibodies against distinct determinants of histone H5 bind to chromatin. Biochemistry. 1984; 23:3459–3466. [PubMed: 6205690]

43. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–283. [PubMed: 21160473]

44. Letourneau A, et al. Domains of genome-wide gene expression dysregulation in Down's syndrome. Nature. 2014; 508:345–350. [PubMed: 24740065]

45. Blecher-Gonen R, et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nature protocols. 2013; 8:539–554. [PubMed: 23429716]

46. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic acids research. 2012; 40:e72. [PubMed: 22323520]

47. Zhu J, et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell. 2013; 152:642–654. [PubMed: 23333102]

48. Simon JM, Giresi PG, Davis IJ, Lieb JD. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nature protocols. 2012; 7:256–267. [PubMed: 22262007]

49. Shen L, et al. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. PloS one. 2013; 8:e65598. [PubMed: 23762400]

50. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. [PubMed: 23582322]
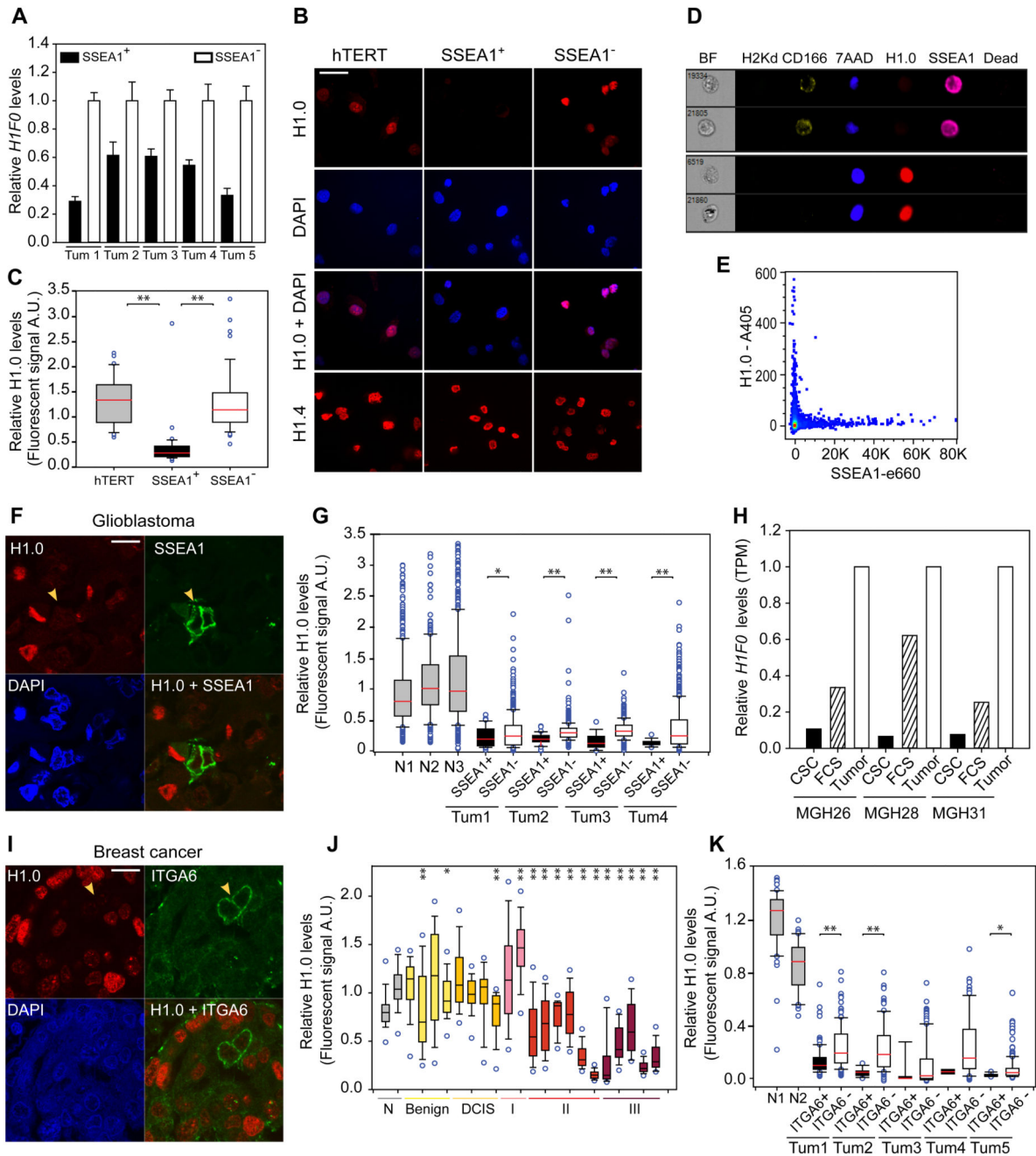
Figure 1. Intratumor heterogeneity of H1.0 levels.

A. qRT-PCR analysis of *H1F0* mRNA levels in SSEA1[+] and SSEA1[-] cells isolated from 5 tumors induced by *in vitro*-transformed fibroblasts (15). Values indicate average ± SEM of three technical replicates. **B-C**. Quantitative immunodetection of H1.0 and H1.4 (red) by immunofluorescence microscopy in the indicated sorted tumor cells and telomerase-immortalized parental cells (hTERT). Scale bar: 20 μm. Two asterisks indicate p < 0.01 (Student t-test). 37 < N < 60 (**C**). **D-E**. Quantitative immunodetection of H1.0 by imaging flow cytometry in unsorted tumor cells. SSEA1[+]: self-renewing cells, CD166[low]: highly

differentiated cells, 7'AAD: nuclei, live/dead$^+$: dead cells, H2Kd$^+$: host mouse cells excluded from analysis. Scale bar: 10 μm. Scatter plot of H1.0 and SSEA1 levels in individual cells (**E**). **F-G**. Quantitative immunodetection of H1.0 and SSEA1 in GBM samples by immunofluorescence microscopy. Scale bar 10 μm. Quantification of H1.0 levels in 3 normal (N1-N3) and 4 tumor (Tum1-Tum4) samples (**G**). One and two asterisks indicate p < 0.05 and p < 0.01, respectively (Student t-test). 15 < N < 910. **H**. Quantification of *H1F0* mRNA levels in cells from three GBM samples by RNA-seq. CSC: Cells grown as neurospheres FCS: cells differentiated *in vitro* by FCS addition, Tumor: whole tumor population. **I-K**. Quantitative immunodetection of H1.0 and ITGA6 in breast cancer samples by immunofluorescence microscopy. Scale bar 10 μm. Quantification of H1.0 levels in 2 normal breast tissues (N) and tumor samples of different histological grade. One and two asterisks indicate p < 0.05 and p < 0.01, respectively, compared to N2 (Studen t-test). 38 < N <134 (**J**). Quantification of H1.0 levels in 2 normal breast tissues (N1, N2) and 5 tumor samples (Tum 1-5) (**K**). One and two asterisks indicate p < 0.05 and p < 0.01, respectively (Student t-test). 7 < N <131 except ITGA6+ Tum 4 for which N =2 (**K**).
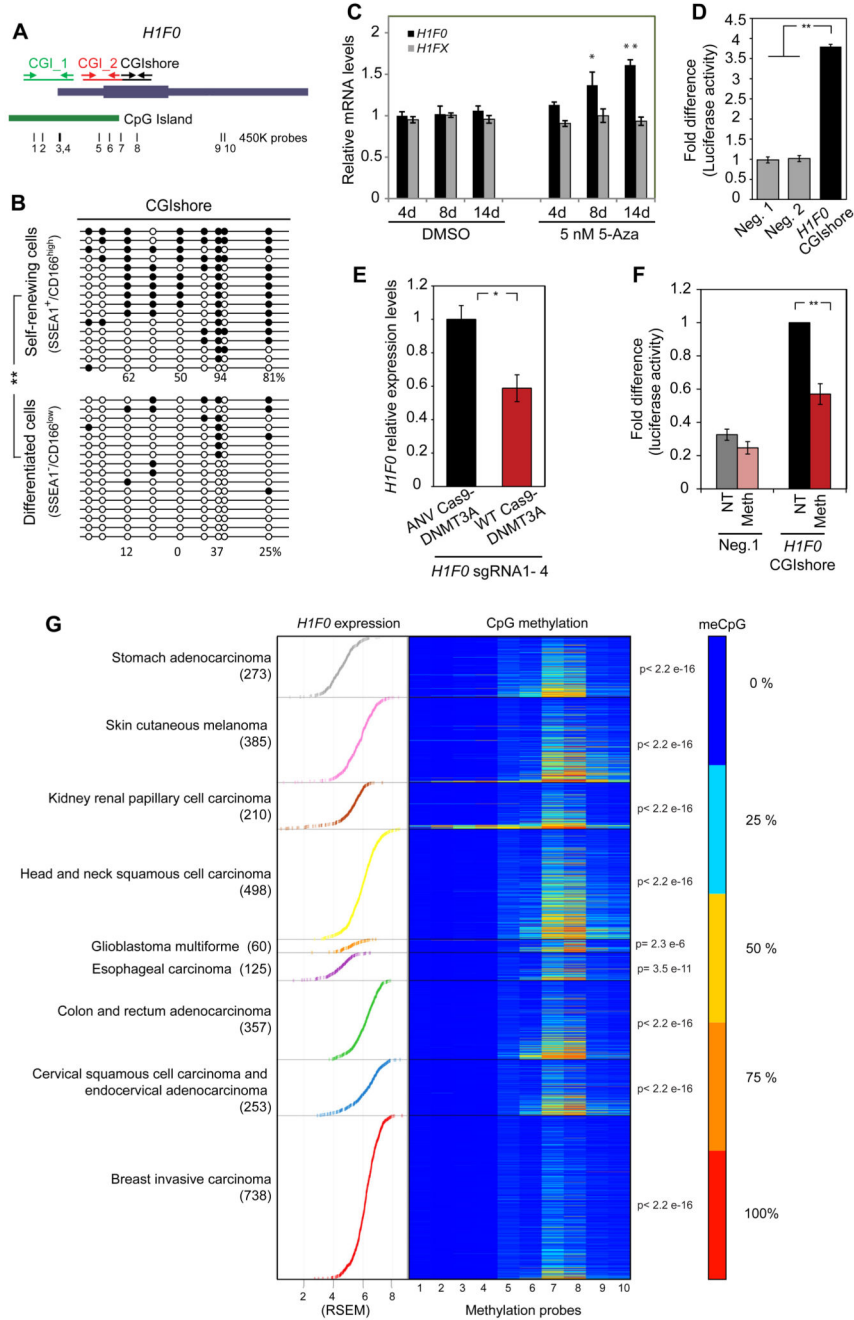
Figure 2. Dynamic methylation of an enhancer element regulates *H1F0* expression in cancer.
A. Representation of the *H1F0* regions probed by bisulfite sequencing analysis (Arrows indicate PCR-amplified regions) and 450K Infinium arrays. **B**. Bisulfite sequencing analysis comparing the *H1F0* CGI shore methylation status in the indicated subsets of sorted tumor cells. Lines represent individual sequenced molecules. White and black circles represent unmethylated and methylated CpGs, respectively. The percentage of methylation of selected CpGs is indicated. Two asterisks indicate p <0.001 (Two-way ANOVA). **C**. qRT-PCR of SSEA1[+] tumor cells treated with 5-Aza-2'-deoxycytidine (5-Aza) or with DMSO as control.

Values represent average ± SEM from three technical replicates. Two independent experiments gave similar results. One and two asterisks indicate p <0.05 and p < 0.01, respectively, compared to DMSO (One-way ANOVA and Tukey Kramer test). **D**. Luciferase reporter assay. Normalized luciferase activity comparing the transactivation potential of two negative control DNA fragments (Neg. 1 and 2 (21, 43)) and the *H1F0* CGI shore. Values are average ± SEM from three experiments. Two asterisks indicate p < 0.01 (Student t-test) **E**. qRT-PCR comparing *H1F0* expression levels in cells expressing *H1F0*-targeting sgRNAs and Cas9 fused to either wild type (WT) or a catalytically dead (ANV) DNMT3A. Values are average ± SEM from three biological replicates. One asterisk indicates p < 0.05 (Student t-test). **F.** Normalized luciferase activity comparing the transactivation potential of untreated (NT) or *in vitro* methylated (Meth) *H1F0* CGI shore, and a negative control fragment (Neg. 1). Values are average ± SEM from five biological replicates. Two asterisks indicate p < 0.01 (Student t-test). **G**. Analysis of *H1F0* methylation in TCGA samples. Patients are sorted based on *H1F0* expression levels (RSEM) and the corresponding DNA methylation levels are visualized as a heatmap. Each row corresponds to a patient and the number of patients for each cohort is indicated. P-values of the Spearman's rank correlation between *H1F0* mRNA levels and methylation of CpG 7 and 8 is indicated (see Fig. S6). Expression levels are normalized across tissues.
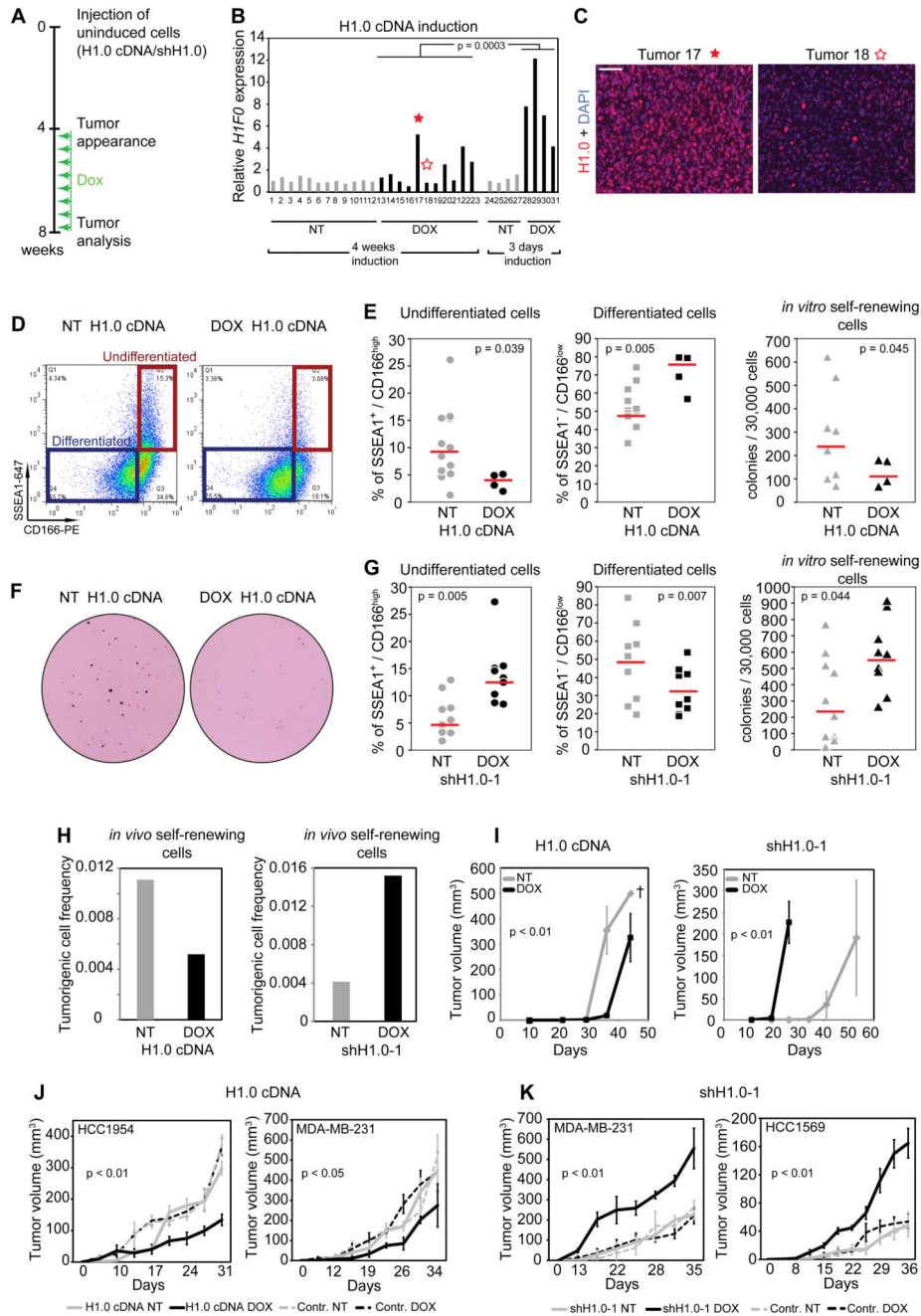
Figure 3. H1.0 inhibits cancer cell self-renewal and drives differentiation.
A. Protocol used to modulate H1.0 levels in established tumors. **B**. Quantification of *in vivo* induction efficiency of H1.0 cDNA by qRT-PCR. Every number indicates a tumor, either uninduced (NT) or induced (DOX) for either 4 weeks or 3 days. Tumors shown in **C** are marked by stars. P-value from Student t-test **C**. Immunodetection of H1.0 in the indicated tumors analyzed by qRT-PCR in **B** (stars). Scale bar: 50 μm. **D**. Flow cytometry analysis of one uninduced and one induced tumor generated by cells containing Dox-responsive H1.0 cDNA constructs. The gates used to measure the fraction of undifferentiated (SSEA1+/

CD166^{high}) or differentiated (SSEA1$^-$/CD166^{low}) cells are indicated. **E,G**. Quantification of the indicated subsets of cells by FACS (undifferentiated and differentiated cells) or by soft agar assay (*in vitro* self-renewing cells) in tumors generated by cells containing the indicated constructs. . P-value from Student t-test. **F**. Representative images of clonogenic soft agar assay. **H**. Limiting dilution transplantation assay for secondary tumor formation using cells from uninduced or induced primary tumors containing the indicated constructs. See also Table S2 and Supplementary text. **I**. Growth of secondary tumors induced by 5000 cells from uninduced or induced primary tumors containing the indicated constructs. Tumor volume values represent mean ± SEM from four tumors each. The value indicated by a dagger corresponds to 2 tumors, due to earlier culling of other animals bearing large tumors. P-value from Student t-test based on the last time point. **J-K**. Growth of secondary tumors induced by 1(**K**) or 2 (**J**) million cells from the indicated breast cancer cell lines containing the indicated constructs. Tumor volume values represent mean ± SEM from five tumors. Significance of the differences between uninduced and induced H1.0 cDNA or shH1.0 at the last time point is indicated (Student t-test). Differences between induced and uninduced control tumors (contr.), which express TurboRFP and the mir30 cassette from empty pTRIPZ, are not significant.
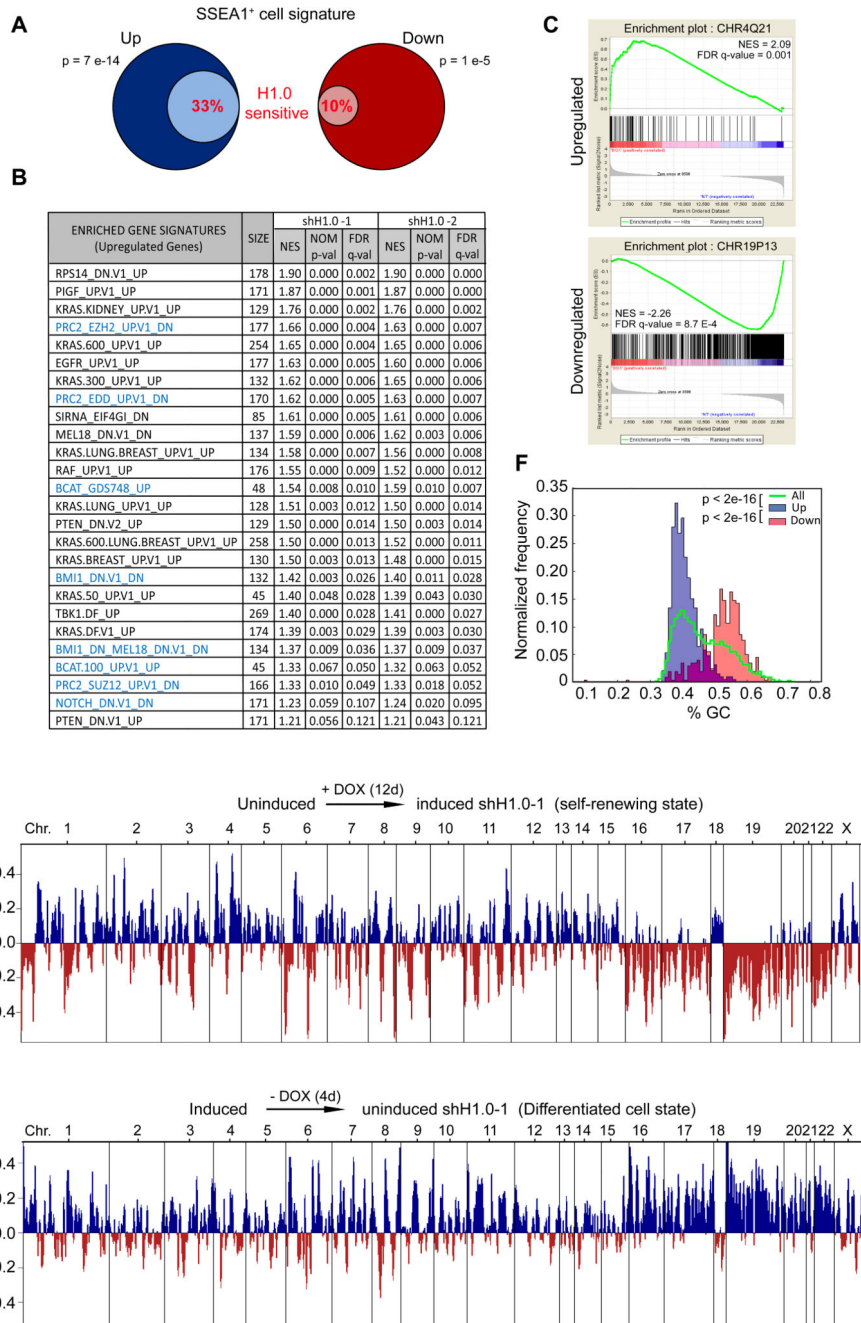
A. SSEA1+ cell signature — Up (p = 7 e-14), 33% H1.0 sensitive, Down (p = 1 e-5), 10%

B.

ENRICHED GENE SIGNATURES (Upregulated Genes)	SIZE	shH1.0 -1			shH1.0 -2		
		NES	NOM p-val	FDR q-val	NES	NOM p-val	FDR q-val
RPS14_DN.V1_UP	178	1.90	0.000	0.002	1.90	0.000	0.000
PIGF_UP.V1_UP	171	1.87	0.000	0.001	1.87	0.000	0.000
KRAS.KIDNEY_UP.V1_UP	129	1.76	0.000	0.002	1.76	0.000	0.002
PRC2_EZH2_UP.V1_DN	177	1.66	0.000	0.004	1.63	0.000	0.007
KRAS.600_UP.V1_UP	254	1.65	0.000	0.004	1.65	0.000	0.006
EGFR_UP.V1_UP	177	1.63	0.000	0.005	1.60	0.000	0.006
KRAS.300_UP.V1_UP	132	1.62	0.000	0.006	1.65	0.000	0.006
PRC2_EDD_UP.V1_DN	170	1.62	0.000	0.005	1.63	0.000	0.007
SIRNA_EIF4GI_DN	85	1.61	0.000	0.005	1.61	0.000	0.006
MEL18_DN.V1_DN	137	1.59	0.000	0.006	1.62	0.003	0.006
KRAS.LUNG.BREAST_UP.V1_UP	134	1.58	0.000	0.007	1.56	0.000	0.008
RAF_UP.V1_UP	176	1.55	0.000	0.009	1.52	0.000	0.012
BCAT_GDS748_UP	48	1.54	0.008	0.010	1.59	0.010	0.007
KRAS.LUNG_UP.V1_UP	128	1.51	0.003	0.012	1.50	0.000	0.014
PTEN_DN.V2_UP	129	1.50	0.000	0.014	1.50	0.003	0.014
KRAS.600.LUNG.BREAST_UP.V1_UP	258	1.50	0.000	0.013	1.52	0.000	0.011
KRAS.BREAST_UP.V1_UP	130	1.50	0.003	0.013	1.48	0.000	0.015
BMI1_DN.V1_DN	132	1.42	0.003	0.026	1.40	0.011	0.028
KRAS.50_UP.V1_UP	45	1.40	0.048	0.028	1.39	0.043	0.030
TBK1.DF_UP	269	1.40	0.000	0.028	1.41	0.000	0.027
KRAS.DF.V1_UP	174	1.39	0.003	0.029	1.39	0.003	0.030
BMI1_DN_MEL18_DN.V1_DN	134	1.37	0.009	0.036	1.37	0.009	0.037
BCAT.100_UP.V1_UP	45	1.33	0.067	0.050	1.32	0.063	0.052
PRC2_SUZ12_UP.V1_DN	166	1.33	0.010	0.049	1.33	0.018	0.052
NOTCH_DN.V1_DN	171	1.23	0.059	0.107	1.24	0.020	0.095
PTEN_DN.V1_UP	171	1.21	0.056	0.121	1.21	0.043	0.121

C. Enrichment plot : CHR4Q21 (Upregulated) NES = 2.09, FDR q-value = 0.001; Enrichment plot : CHR19P13 (Downregulated) NES = -2.26, FDR q-value = 8.7 E-4

F. Normalized frequency vs % GC. p < 2e-16, p < 2e-16. All, Up, Down

D. Uninduced → + DOX (12d) → induced shH1.0-1 (self-renewing state). DOX / NT log2[FC]

E. Induced → - DOX (4d) → uninduced shH1.0-1 (Differentiated cell state). washDOX / DOX log2[FC]

Figure 4. Activation of transcriptional programs supporting oncogenic self-renewal via upregulation of large gene domains.

A. Venn diagrams showing the percentage of genes upregulated (Up) or downregulated (Down) in SSEA1+ cells compared to SSEA1- cells that are affected by H1.0 knockdown (See also Table S4). The significance of the overlap is indicated (hypergeometric test). **B**. Oncogenic gene signatures positively correlating with DOX samples. NES: normalized enrichment score, NOM p-val: nominal p-value, FDR q-val: false discovery rate q-value. Blue: stem cell-related gene signatures. **C**. GSEA plots of positional gene sets positively

(CHR4Q21) or negatively (CHR19P13) correlating with DOX samples. **D-E**. Smoothed $\log_2$ fold change of gene expression between DOX and NT (**D**) or washDOX and DOX samples (**E**) (shH1.0-1) along the human genome. Similar plots were obtained with shH1.0-2. Numbers indicate the chromosomes delimited with vertical lines. Only expressed genes (TPM > 0) are plotted. Blue: upregulated domains, red: downregulated domains. **F**. Distribution of GC content in RefSeq genes (all) and in the subsets of upregulated (Up) or downregulated (Down) DEGs. P-value from Student t-test.
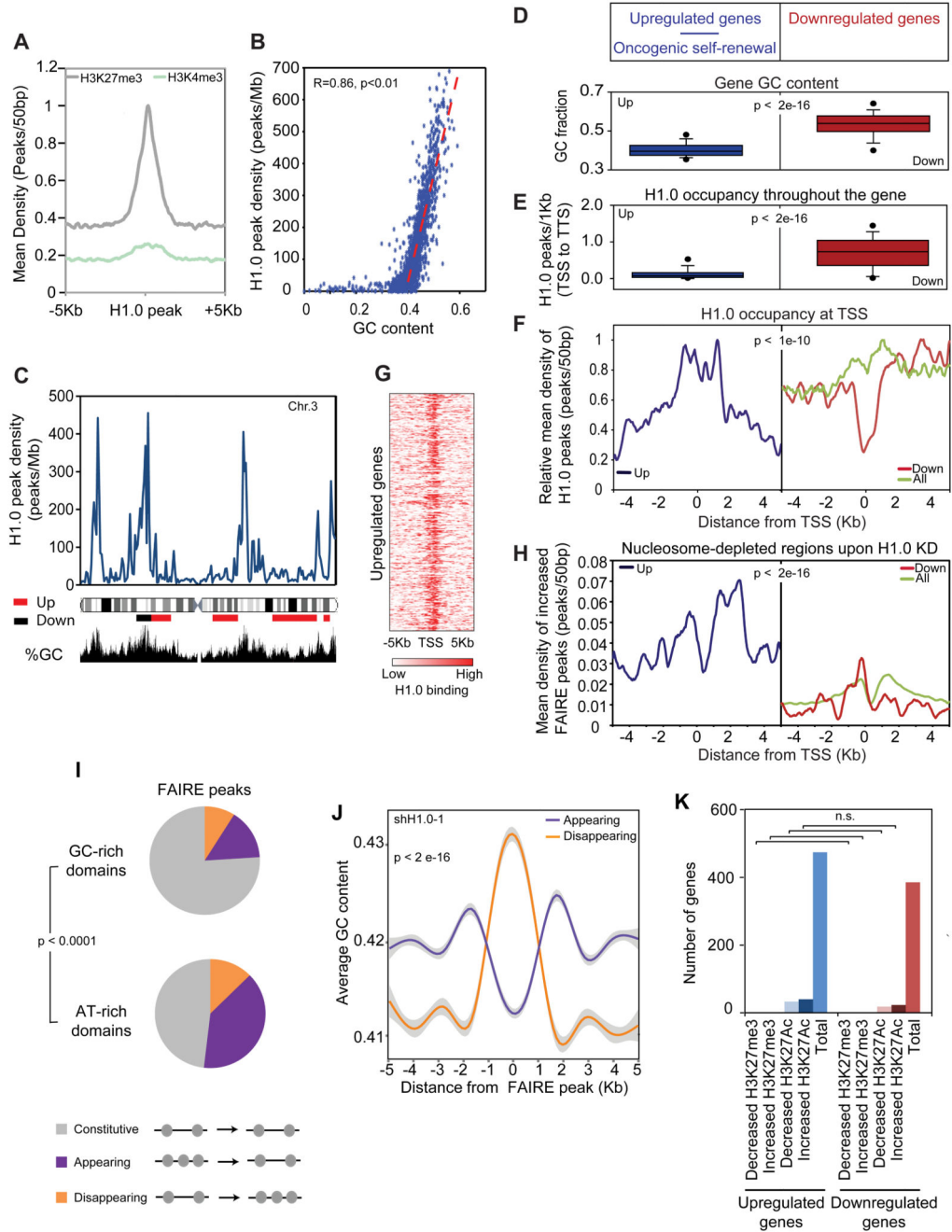
Figure 5. Destabilization of nucleosome-DNA interactions in AT-rich regions in the absence of H1.0.

A. Average peak density profiles of H3K4me3 and H3K27me3 centered on H1.0 binding sites. **B**. Correlation between H1.0 peak density and DNA GC content. The best fit line of the experimental values for GC content > 0.4, the correlation coefficient R and statistical accuracy of the fit are indicated. **C**. H1.0 peak density along chromosome 3. The approximate location of cytogenetic bands, upregulated (Up) or downregulated (Down) positional gene sets identified by GSEA and the corresponding DNA GC content are shown.

D-F, H. Comparison between genes upregulated (Up) or downregulated (Down) in response to H1.0 knock-down, with respect to the indicated features. Smoothed average density profile of H1.0 peaks (**F**) and FAIRE peaks (**H**) show enrichment of H1.0 and increased FAIRE signal (decreased nucleosome occupancy) upon H1.0 knock-down around the transcriptional start site (TSS) of upregulated genes. RefSeq genes (All) are shown as reference. P-value from paired t-test, with Benjamini-Hochberg adjustment for **F** and **H**. **G.** Heatmap showing tag density of H1.0 ChIP-seq around the TSS of genes upregulated upon H1.0 knock-down. Each line represents a gene. **I**. Relative abundance of the indicated types of FAIRE peaks in GC-rich and AT-rich domains. Nucleosome occupancy corresponding to the different types of FAIRE peaks is schematized next to the legend. Black line: DNA, gray circles: nucleosomes. Results from shH1.0-1 are shown. Similar results were obtained with shH1.0-2. P-value from Fisher's exact test for constitutive peaks. **J.** Average DNA GC content centered on FAIRE peaks that appear or disappear upon H1.0 knock-down. See also Fig. S14D. P-value from paired t-test with Benjamini-Hochberg adjustment. Grey area: 95% confidence interval of the best fit after smoothing. **K**. Number of upregulated or downregulated H1.0-sensitive genes showing altered acetylated or methylated H3K27 at TSS upon H1.0 knockdown. Differences are not significant (n.s., Fisher's exact test).
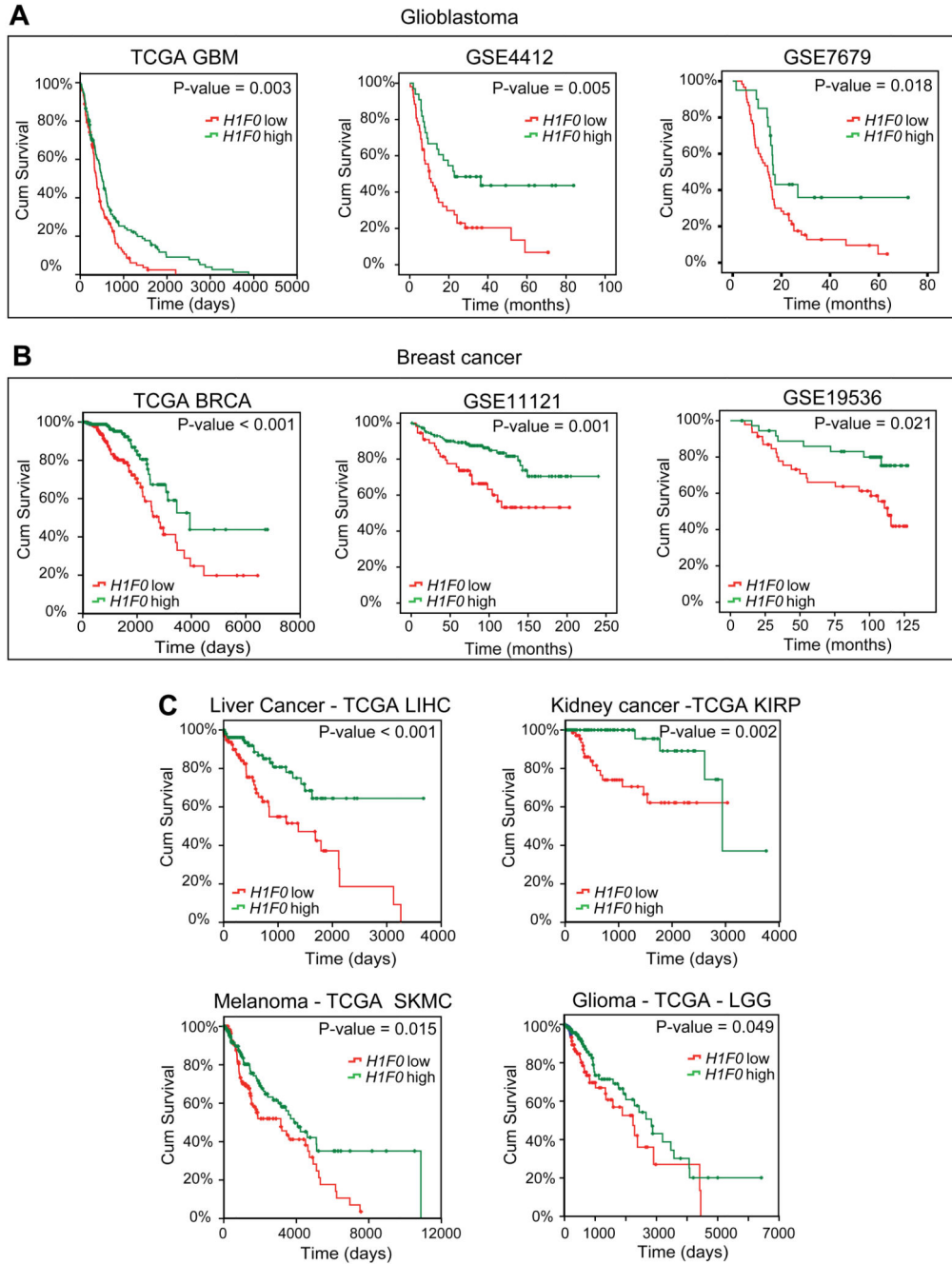
Figure 6. Low *H1F0* levels correlate with low patient survival in multiple cancer types.
A-C. Kaplan-Meier analysis of the indicated datasets showing significant correlation between *H1F0* levels and patient survival. P-value from Log-rank test. Multivariate analysis is shown in Fig. S16.