



# Somatic mutation and functional polymorphism of a novel regulatory element in the *HGF* gene promoter causes its aberrant expression in human breast cancer

Jihong Ma,<sup>1</sup> Marie C. DeFrances,<sup>1</sup> Chunbin Zou,<sup>1</sup> Carla Johnson,<sup>1</sup>  
Robert Ferrell,<sup>2</sup> and Reza Zarnegar<sup>1</sup>

<sup>1</sup>Division of Experimental Pathology, Department of Pathology, School of Medicine, and

<sup>2</sup>Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.

**The *HGF* gene is transcriptionally silenced in normal differentiated breast epithelial cells, but its repression fails to occur in mammary carcinoma tissues and cell lines. The molecular mechanisms underpinning aberrant *HGF* expression in breast cancer cells are unknown. Here we report the discovery of a DNA element located 750 bp upstream from the transcription start site in the human *HGF* promoter that acts as a transcriptional repressor and is a target of deletion mutagenesis in human breast cancer cells and tissues. This *HGF* promoter element consists of a mononucleotide repeat of 30 deoxyadenosines (30As), which we have termed “deoxyadenosine tract element” (DATE). Functional studies revealed that truncation mutations within DATE have profound local and global effects on the *HGF* promoter region by modulating chromatin structure and DNA-protein interactions, leading to constitutive activation of the *HGF* promoter in human breast carcinoma cell lines. We found that 51% of African Americans and 15% of individuals of mixed European descent with breast cancer harbor a truncated DATE variant (25As or fewer) in their breast tumors and that the truncated allele is associated with cancer incidence and aberrant *HGF* expression. Notably, breast cancer patients with the truncated DATE variant are substantially younger than those with a wild-type genotype. We also suggest that DATE may be used as a potential genetic marker to identify individuals with a higher risk of developing breast cancer.**

## Introduction

Acquisition of an autocrine growth-stimulating circuit is a hallmark of cancer cells (1). HGF is an important paracrine mediator of mesenchymal-epithelial interaction, instructing epithelial cells to undergo processes such as cell proliferation, survival, migration, and morphogenesis (2, 3). HGF exerts its effects via binding to and activating a tyrosine kinase transmembrane cell surface receptor known as Met (2–7). *HGF* gene expression is highly cell type specific and is restricted to mesenchymal cells such as fibroblasts. Normal epithelial cells do not express this gene, although all epithelial cells express the HGF receptor. Studies have established that HGF/Met signaling actively contributes to the process of neoplastic transformation, tumor metastasis, and tumor maintenance (8, 9). The underlying basis of aberrant HGF/Met signaling in different types of cancer has been attributed to various mechanisms such as amplification of the *Met* gene and/or its overexpression as well as the appearance of activating mutations in the cytoplasmic domain of Met (2, 3, 7). In the case of breast adenocarcinomas, activation of autocrine HGF expression in the cancer cells and its overexpression are believed to contribute to neoplasia (8). Clinical studies have reported that high levels of

*HGF* mRNA and protein within the breast carcinoma tissue (as determined by ELISA and quantitative RT-PCR) are associated with poor survival in patients (10–14) and that the levels of HGF are significantly higher in the pleural effusion fluid of patients whose breast cancer has metastasized to the pleura. Studies using *in situ* hybridization uncovered that the *HGF* transcript is expressed in human breast adenocarcinoma cells but not in the normal mammary ductal epithelium (15). Collectively, these data highlight the importance of proper regulation of *HGF* gene expression in homeostasis of normal breast epithelium.

Our laboratory has studied the transcriptional regulation of the *HGF* gene in order to decipher the molecular basis of its cell type-specific and inducible expression. We have identified and characterized some of the regulatory *cis*-acting elements and their cognate transcription factors that govern the transcription of this gene. We previously reported that nuclear receptors such as estrogen receptor, chicken ovalbumin upstream promoter transcription factors (COUP-TFs), and PPAR $\gamma$  as well as transcriptional regulators such as AP2, NF1, SP1, SP3, and the CCAAT/enhancer binding protein  $\beta$  (C/EBP- $\beta$ ) interact with their sites within 1 kb upstream of the *HGF* gene basal promoter and functionally participate in *HGF* gene transcription (16–20). Although a complete understanding of the molecular mechanisms that regulate human *HGF* gene transcription remain elusive, our data combined with those published by others have shown that cell type-specific expression (silencing in normal differentiated epithelial cells) and inducible expression (in stromal cells) is dictated by a complex system involving several regulatory elements located within the DNA 1 kb upstream of the

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Nonstandard abbreviations used:** 30As, 30 deoxyadenosines; AA, African American; 3AB, 3-aminobenzamide; CC, individual of mixed European descent; C/EBP- $\beta$ , CCAAT/enhancer binding protein  $\beta$ ; DATE, deoxyadenosine tract element; PARP, poly (ADP-ribose) polymerase; PR1, protected region 1.

**Citation for this article:** *J. Clin. Invest.* 119:478–491 (2009). doi:10.1172/JCI36640.



*HGF* gene transcription start site, and that chromatin structure and differential transcription factor accessibility to this region play essential roles in this process (16–23). These data highlight the importance of the *cis*-acting elements and their cognate binding factors in determining *HGF* gene transcription.

Given this knowledge and the fact that the *HGF* gene is aberrantly expressed in breast carcinoma cells, we hypothesized that the reason for this may be due to genetic phenomena involving mutation of the 1-kb DNA stretch of the *HGF* promoter region. Therefore, to test this hypothesis, we scanned human carcinoma cell lines that aberrantly expressed the *HGF* gene for promoter mutation. The results led us to the identification of what we believe is a novel 30-bp *cis*-acting DNA element within the proximal promoter region of the human *HGF* gene. The element consists of a tract of 30 deoxyadenosines (30As; poly[dA]) which we have termed “deoxyadenosine tract element” (DATE). Our functional studies revealed that wild-type DATE alters chromatin structure and silences the *HGF* promoter in normal epithelial cells through interactions with several nuclear factors and that DATE deletion mutagenesis (DATE truncation or shortening) distorts this DNA structure, resulting in differential binding of several factors to this element and its adjacent sites. This leads to recruitment of chromatin modifying and transcription factors such as C/EBP- $\beta$  and poly (ADP-ribose) polymerase (PARP), the culmination of which is activation of the *HGF* promoter in the tumor cells. Our studies of human breast and other carcinoma tissues and cell lines revealed that DATE is a target of deletion mutagenesis (DATE truncation) in cancer and that truncated DATE (i.e., DATE with 25As or fewer) significantly associates with transcription factor binding, promoter activation, and high *HGF* gene expression in tumor cells. We found that in breast cancer tissues, truncated DATE occurs at a frequency of 51% in African Americans (AAs) and 15% in individuals of mixed European descent (CCs). Notably, we uncovered that DATE is polymorphic in the general population and that the truncated DATE variant is significantly more prevalent in healthy AAs (26%; 31/119) than in healthy CCs (3%; 15/473). Importantly, we discovered that in normal subjects, *HGF* gene expression is significantly higher in the mammary tissues of individuals with truncated DATE than in tissues of individuals with wild-type DATE. Together, these results not only shed light on our understanding of the genetic basis of human mammary tumorigenesis but also have led to the discovery of a genetic marker that may be used for identification of individuals with a higher risk of developing breast cancer.

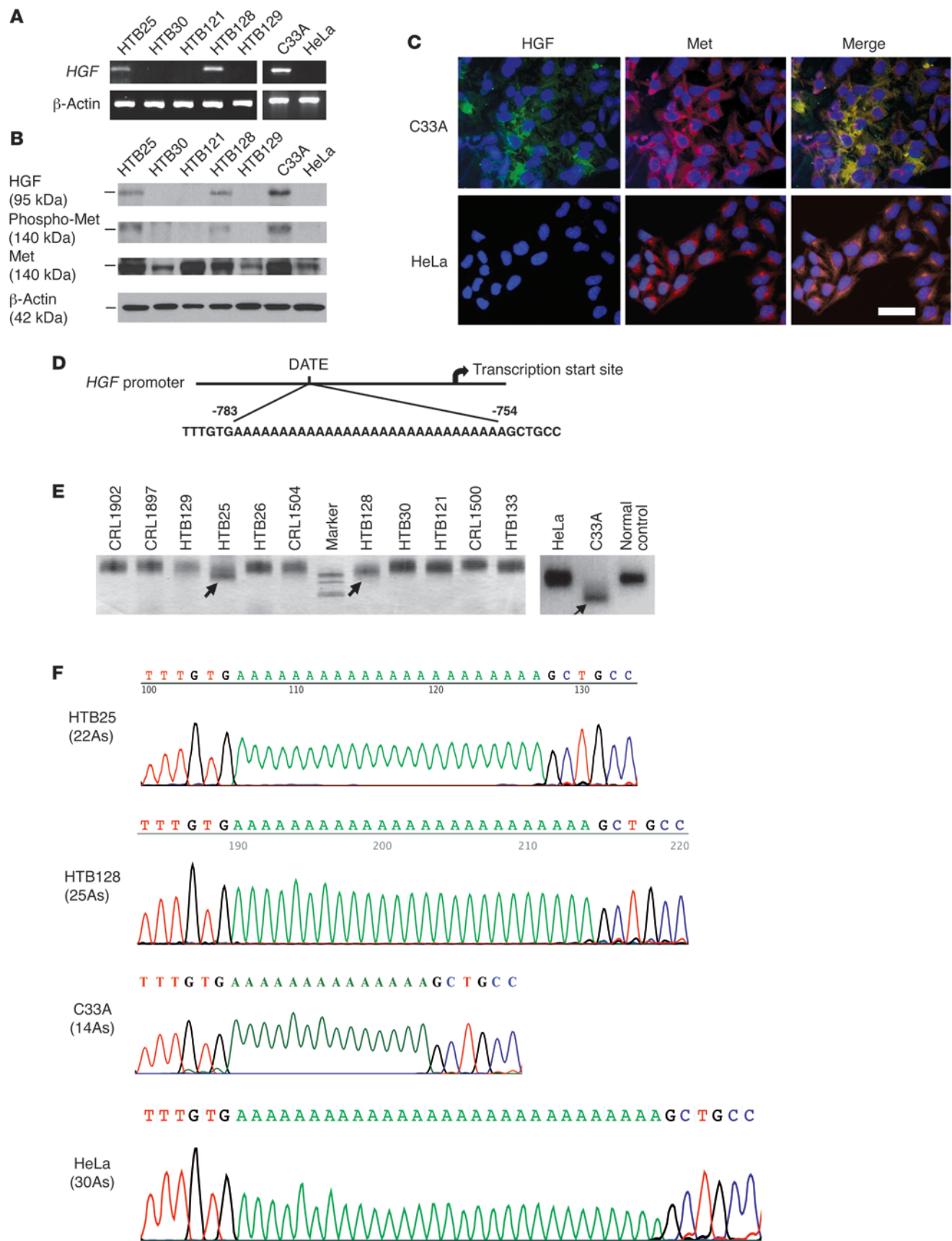
## Results

*Truncation mutation of DATE in the HGF gene promoter region correlates with activation of HGF gene expression in human breast and cervical carcinoma cell lines.* Since previous studies established that *HGF* cell type-specific expression is dictated by the 1-kb upstream promoter region of *HGF* (16–20), we hypothesized that failure of this gene to undergo silencing in carcinoma cells is due to genetic/epigenetic changes of the *HGF* gene promoter. We screened carcinoma cell lines (breast, cervical, and vulvar) for *HGF* gene expression. We found that some breast cancer (HTB25, HTB128) and cervical carcinoma (C33A) cell lines expressed *HGF* and had active Met (phospho-Met), whereas others did not (i.e., HeLa, HTB30, HTB121, and HTB129) (Figure 1, A–C). As expected, all of these carcinoma cell lines expressed Met protein (representative data are shown in Figure 1, B and C). Moreover, we determined that *HGF* protein

secreted by the expressor cells was bioactive as assayed on primary cultures of rat hepatocytes for induction of DNA synthesis and its complete neutralization by anti-*HGF* antibody (data not shown). Sequencing of the 1-kb upstream promoter segment of *HGF* and its comparison to the normal human *HGF* gene deposited in GenBank (accession no. M75967 J05342; GI: 337136) revealed that the only difference between the expressor and nonexpressor cells occurred in an *HGF* promoter region consisting of a poly(dA) mononucleotide tract (located –754 to –783 bp from the transcription start site) (Figure 1D). We termed this poly(dA) tract “DATE.” Specifically, we found that most of the carcinoma cell lines that did not express *HGF* had a stretch of 30As in their DATE identical to the normal human *HGF* promoter sequence, indicating that these cells had a wild-type DATE, whereas those that expressed *HGF* exhibited a truncated mutant version of DATE in which the number of As was reduced by 5 bp or more. For instance, in C33A cells, DATE length was truncated by 16As (thus, these cells had a DATE genotype of 14As in homozygous manner) and HTB25 cells were heterozygous for short alleles of 22As and 25As, while in HTB128 cells, both alleles were truncated by 5 nucleotides (25As) in DATE. On the other hand, HeLa cells that did not express *HGF* had normal a DATE genotype of 30As (see representative genotype data in Figure 1, E and F). Analyses of human breast cancer tissues and surrounding matched normal breast tissue revealed that DATE truncation indeed occurred in tumor tissues (see below for details).

*Regulatory effects of DATE on HGF promoter activity.* Given the close proximity of DATE to the transcription start site of the *HGF* promoter (only 750 bp upstream of the start site) in combination with data obtained from cancer cell lines, we hypothesized that DATE regulates *HGF* promoter function and that its deletion (retraction) alters promoter activity. Therefore, we isolated 1.1 kb of the human *HGF* promoter region containing DATE from various human breast cases with different DATE variants (genotypes 30A, 29A, 28A, 27A, 26A, 25A, 20A, or 14A) and subcloned them upstream of the luciferase reporter plasmid (Figure 2A). Next, we transfected them into epithelial cell lines such as HeLa cells and evaluated them for promoter activity. We discovered that truncated DATE significantly activated the luciferase promoter. The promoter activity exhibited dependency on the extent of DATE truncation. DATE with 14As showed more activity than DATE with 20As, and DATE with 20As had more activity than DATE with 25As, as compared with the wild-type 30A DATE (Figure 2B, 2-tailed paired *t* test,  $P < 0.01$ ). The effects of minor loss of deoxyadenosine residues (i.e., DATE having 26As to 29As) on the promoter activity were not distinguishable among individual constructs, although they all had slightly higher activity as compared with the 30A construct (Supplemental Figure 1). These data indicate that wild-type DATE (i.e., DATE 30A) exerts a negative regulatory effect on *HGF* promoter activity and that its shortening relieves this repressive function.

*Truncated DATE affects HGF promoter chromatin structure and DNA-protein interaction.* To determine whether DATE influences promoter activity via modulation of DNA-protein interaction, we first carried out in vivo DNase I hypersensitivity assays. For these assays, nuclei were isolated from matched carcinoma cell lines that had a truncated or wild-type DATE (i.e., 14As vs. 30As) and treated them with DNase I followed by BamHI digestion, to release the 1.7-kb promoter fragment, and Southern blot analysis using a  $^{32}$ P-labeled probe encompassing DATE and its flanking sites (Figure 3A). As depicted in Figure 3B, both cell types had the 1.7-kb promoter





**Figure 1**

Truncation of DATE in the *HGF* gene promoter region correlates with aberrant *HGF* expression in human carcinoma cell lines. **(A)** Detection of *HGF* gene expression by RT-PCR. *HGF* gene expression was noted in HTB25, HTB128, and C33A cells but not in other cells. **(B)** Detection of *HGF*, Met, and phosphorylated Met protein expression by western blot assay. *HGF* and phospho-Met protein expression were observed in HTB25, HTB128, and C33A cells. Met expression was observed in all cells. **(C)** Detection of *HGF* and Met protein expression by immunofluorescence microscopy in C33A and HeLa cells. Scale bar: 40  $\mu$ m. *HGF* protein expression was notable only in C33A cells but not HeLa cells, whereas Met protein expression occurred in both cells. Nuclei were stained blue with VECTASHIELD Mounting medium with DAPI. **(D)** Schematic representation of the human *HGF* promoter and the location and DNA sequence of wild-type DATE (30As). **(E)** Representative PCR genotype analysis of DATE from 11 human breast cancer cell lines and 12 human cervical carcinoma cell lines. The 70-bp PCR product was fractionated on 15% denaturing polyacrylamide gel containing 8 M urea. Markers used were cloned and sequenced cell-derived 70-bp DNA fragments containing 28As, 26As, 21As, 17As, or 14As in DATE. Arrows indicate truncated DATE band. **(F)** Representative DNA sequencing of the major band corresponding to the truncated DATE (arrows in **E**) confirmed that the shorter alleles of truncated DATE in HTB25, HTB128, and C33A cell lines were 22As, 25As, and 14As, respectively, while HeLa cells showed wild-type DATE with 30As.

fragment when they were treated with low amounts of DNase I (15 U/ml). However, when treated with higher doses of DNase I (30 and 45 U/ml), cells harboring the wild-type DATE (and silenced for *HGF* gene expression) had the 1.7-kb band and an additional 0.9-kb band, indicating highly condensed, DNase I-inaccessible DNA, whereas cells having the truncated DATE variant displayed a DNase I-hypersensitive promoter and hence a relaxed and accessible chromatin configuration. To further investigate whether a truncated DATE variant activates the *HGF* gene in vivo, we analyzed polII, acetyl-histone 3, and acetyl-histone 4 recruitment to the *HGF* promoter as surrogates of gene activity using ChIP assays. To this end, PCR was performed with specific primers for the *HGF* gene transcription start sites (–293 to –12 of the human *HGF* gene; Figure 3C). As expected, HeLa cells that contained wild-type DATE did not have these hallmarks of active gene transcription, whereas C33A cells with truncated DATE (i.e., 14As) did have them, supporting the notion that the presence of truncated DATE correlates with *HGF* gene transcriptional activity.

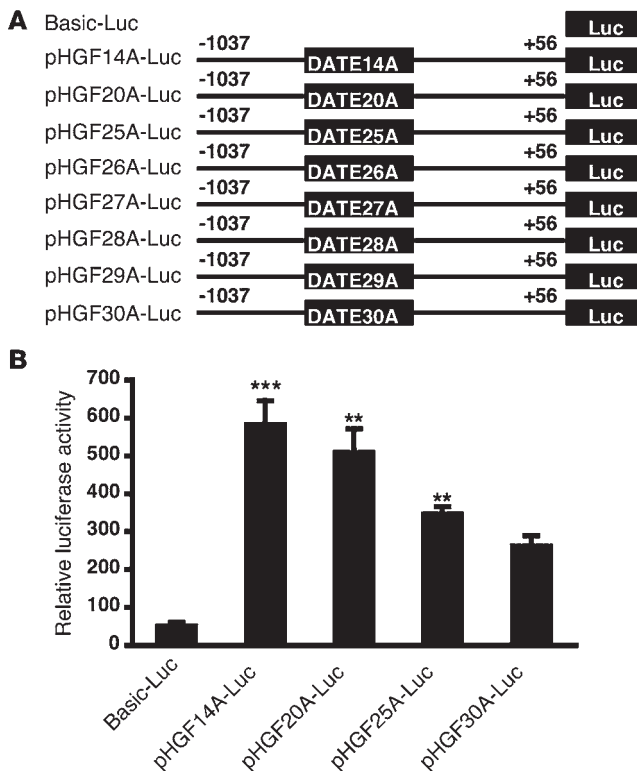
We wanted to prove directly that the truncated DATE variant indeed interferes with DNA-protein interaction. Therefore, a DNase I footprinting assay was carried out using a 300-bp *HGF* promoter fragment extending from –928 to –627 and containing either wild-type (30As) or truncated DATE (14As) as probes in side-by-side experiments with the same nuclear protein extract. The 2 radiolabeled probes were subjected to DNase I digestion and electrophoresis on a DNA sequencing gel according to standard procedures. Interestingly, we identified a different pattern between the probes (Figure 4A). At least 5 interaction sites (or protected regions 1–5 [PR1–PR5]) were noted. With the wild-type probe, we observed PR1, PR2, PR3, and PR5 (Figure 4A). PR1 (–753 to –764 bp) overlapped the 3' end of DATE covering approximately the last 12As, while PR2 encompassed –792 to –800, PR3 encompassed –804 to –816, and PR5 encompassed –831 to –843, located upstream of DATE. We also identified a hypersensitive band between PR2 and PR3 with the wild-type probe but not the trun-

cated probe (asterisk in Figure 4A). On the other hand, the probe harboring truncated DATE exhibited 1 protected region (PR5) shared with the wild-type probe; however, it was found to have a unique protected region labeled as PR4 (Figure 4A). Sequence analysis of these protected regions using the TRANSFAC database identified a potential C/EBP- $\beta$  binding site in PR4 (Figure 4A) and hunchback binding sites in the DATE region. The above results establish that the DNA-protein interactions are different between the wild-type and the truncated variants of the *HGF* promoter in breast carcinoma tissues.

To further expand these findings, we first wanted to determine whether wild-type DATE can bind to nuclear proteins (as evidenced by PR1). Thus, we carried out band shift assays with synthetic double-stranded 30-mer oligonucleotide corresponding to wild-type DATE or its truncated variant (14As) as probes. To ensure the probes were equal length, linkers were added to truncated DATE. We found that DATE with 30As but not 14As strongly and specifically bound to nuclear factors (Figure 4B). Competition assays with excess unlabeled oligos corresponding to HNF3 $\beta$  or hunchback binding sites (i.e., putative factors identified through TRANSFAC analysis that could potentially bind to the poly[dA] stretch) and addition of anti-HNF3 $\beta$  antibody did not alter complex formation with wild-type DATE, indicating that the actual factor(s) binding to DATE are different than these putative factors (data not shown).

*C/EBP- $\beta$  and PARP have high binding affinity for the truncated but not wild-type DATE and are involved in *HGF* gene activation.* The DNase I footprinting studies mentioned above identified gene regulatory factors that differentially bound to the *HGF* promoter region harboring wild-type vs. truncated DATE. This was particularly true with respect to PR4, which contained a potential C/EBP- $\beta$  site. Therefore, we performed additional DNA affinity purification experiments using long synthetic double-stranded oligos that encoded the wild-type and truncated DATE with identical flanking sites (these oligos were 78 and 62 mers, respectively). The eluted proteins were subjected to SDS-PAGE and then to immunoblot analysis using anti-C/EBP- $\beta$  antibody as well as to silver staining and MALDI-TOF tandem mass spectrometry (MS/MS) to identify other potential binding proteins that differentially bind to DATE. In these studies we used biotinylated commercial C/EBP- $\beta$  site oligos and unrelated oligos as positive and negative controls, respectively. Western blot analyses clearly showed that both C/EBP- $\beta$  isoforms specifically and strongly bound to the promoter fragment harboring truncated 14A DATE but not to the fragment containing wild-type 30A DATE (Figure 5A). Densitometric analysis revealed that the ratio of C/EBP- $\beta$ 1 to C/EBP- $\beta$ 2 was higher in the 14A DATE oligo as compared with that obtained from the consensus C/EBP- $\beta$  site (ratios of 1.7 and 0.4, respectively), indicating that DATE truncation favors  $\beta$ 1 over  $\beta$ 2 binding to the *HGF* promoter region. Deletion of the entire 30As from the wild-type probe or its truncation to 14As allowed C/EBP- $\beta$  binding to its cognate binding site located immediately upstream of DATE, proving that the 30A element with its associated binding proteins hinders C/EBP- $\beta$  interaction with this region (Supplemental Figure 2). To confirm that C/EBP- $\beta$  indeed binds in vivo to the *HGF* promoter possessing truncated but not wild type DATE, we performed ChIP using anti-C/EBP- $\beta$  antibody and control IgG and found that C/EBP- $\beta$  occupied only the *HGF* promoter, which had a truncated DATE (in C33A cells; Figure 5B).

We then turned our attention to identifying other proteins that may differentially bind to the DATE region. Silver staining of the fraction from DATE-containing oligo columns revealed a major

**Figure 2**

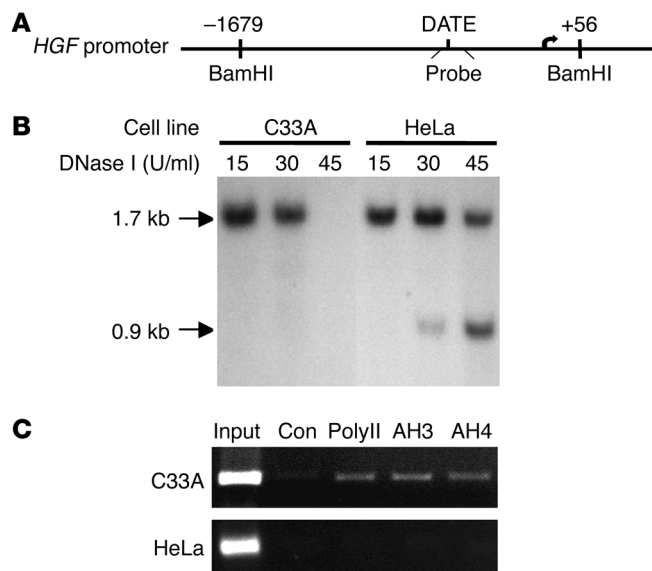
Regulatory effects of DATE on *HGF* promoter activity. **(A)** Schematic representation of pHGF–luciferase reporter plasmids. *HGF* promoter region (1.1 kb; –1,037 to +56) containing wild-type (30As) or shortened DATE (29As, 28As, 27As, 26As, 25As, 20As, and 14As) from normal and different breast cancer tissues, respectively, were inserted upstream of the luciferase report gene in pGL2-basic vector and were termed pHGF30A-Luc, pHGF29A-Luc, pHGF28A-Luc, pHGF27A-Luc, pHGF26A-Luc, pHGF25A-Luc, pHGF20A-Luc, and HGF14A-Luc. pGL2-basic (basic-Luc) was used as negative control vector. **(B)** Luciferase activity profile. The above constructs along with pRL-TK as an internal control vector were transiently transfected into HeLa cells as indicated in the figure. As depicted, luciferase activity in pHGF30A-Luc was significantly lower than those constructs with truncated DATE (30As vs. 14As,  $P = 0.0006$ ; 30As vs. 20As,  $P = 0.0016$ ; 30As vs. 25As,  $P = 0.0025$ ). \*\* $P < 0.01$ , \*\*\* $P < 0.001$  compared with pHGF30A-Luc.

band with a relative molecular mass of 62 kDa, which was only detectable in the chromatographic fraction enriched by truncated DATE but not wild-type DATE (Figure 5C), and subsequent MS/MS analysis identified the band as PARP. This was verified by western blot using an antibody against PARP that recognizes both isoforms of PARP (PARP1 and PARP2; Figure 5D). PARPs are potent activators of gene transcription via their ability to open chromatin by polyribosylation of histones (24). We then investigated the effects of gradual loss of DATE truncation on C/EBP- $\beta$ 1 and PARPs binding to the promoter region to determine the minimum numbers of As that need to be deleted from DATE to enhance factor binding. In agreement with promoter activation results (Figure 2 and Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI36650DS1), the results showed that shortening of DATE by 5 or more nucleotides (i.e., DATE containing 25As or less) enhanced C/EBP- $\beta$ 1 and PARP binding to the promoter region (Figure 5, E and F).

Finally, to test whether C/EBP- $\beta$  and PARP affect *HGF* expression in vivo, we downregulated C/EBP- $\beta$ 1 and PARPs using siRNAs in cells having truncated DATE. We found that *HGF* expression was significantly reduced when both transcription factors were knocked down simultaneously, while downmodulation of either factor individually had little effect, suggesting that these 2 factors cooperate to maximally induce *HGF* expression (Figure 6, A and B). To further confirm the activation function of C/EBP- $\beta$ 1 and PARPs on the *HGF* gene with truncated DATE, we cotransfected the *HGF* promoter-reporter construct containing the truncated 14A DATE with C/EBP- $\beta$ 1 (i.e., full-length C/EBP- $\beta$ ) into HeLa cells and then treated the cells with 3-aminobenzamide (3AB, an inhibitor of PARPs). As shown in Figure 6C, *HGF* promoter activity was induced by C/EBP- $\beta$ 1, and treatment with the PARP inhibitor significantly reduced basal and C/EBP- $\beta$ 1-induced promoter

activity, suggesting that C/EBP- $\beta$ 1 and PARP1/2 cooperate with each other and serve as coactivators of the *HGF* gene promoter.

DATE is a target of deletion mutagenesis in human breast cancer tissue, and its truncated version associates with aberrant *HGF* expression. Armed with the information that DATE is a functional element, that its truncated version occurs in human breast cancer cell lines, and that it confers aberrant *HGF* promoter activity, we next wanted to determine whether DATE is mutated in human breast tumor tissues. We evaluated the DATE genotype in samples from 145 different breast tissues from 95 cases of human breast carcinoma and matched normal adjacent tissue as well as 50 cases of normal breast tissue obtained from reduction mammoplasty. We were intrigued to find that retraction (i.e., truncation) of DATE occurred in breast tumors and that truncated DATE was significantly associated with breast cancer. We found that approximately 51% (19/37) of AA and 15% (8/53) of CC breast cancer tissues possessed the truncated DATE variant (Table 1). To be conservative in these analyses, we designated DATE as a truncated variant if it had 25As or fewer based on the functional promoter activity and factor binding profile we described above (see Figure 2, Supplemental Figure 1, and Figure 5, E and F). The size of truncated DATE among different breast tumor cases ranged from 17As (loss of 13As) to 25As (loss of 5As) (see Table 1). In 21% (6/28) of the cancer cases with shortened DATE, truncated DATE was present only in the tumor and not in the corresponding adjacent normal tissue, indicating that DATE is unstable and prone to deletion mutagenesis in tumor cells. Representative sets of genotyping resulting from gel electrophoresis and DNA sequencing from 3 breast cancer cases and their paired adjacent normal tissues are depicted in Figure 7 (see panels A and D, respectively). The genotype of DATE in these 3 cases was 17As/26As, 19As/29As, and 18As/30As in the tumors, whereas their corresponding adjacent normal tissues were homozygous for 26As, 29As, and 30As, respectively. *HGF* mRNA was highly expressed in cancer tissues with the truncated DATE, since it was readily detectable by 32 cycles of RT-PCR, but little or no *HGF* expression was detected in tissues with wild-type DATE under these conditions (Figure 7, B and C). We performed immunohistochemistry to assess the expression and cellular localization of *HGF* (Figure 8A) and Met (Supplemental Figure 3) in breast cancer cases. These studies revealed that *HGF* expression in breast cancer tissues and in normal adjacent tissues with truncated DATE were significantly higher than those containing wild-type DATE (see Figure 8, A–D, and Supplemental Table 2;  $P = 8.8 \times 10^{-7}$



**Figure 3**

Truncated DATE affects chromatin structure in *HGF* promoter region and associates with *HGF* gene activation. **(A)** Map of the human *HGF* promoter region showing the BamHI site, DATE, and the probe region used. The arrow shows the transcription start site. **(B)** Fresh nuclei were isolated from actively growing C33A and HeLa cells, treated with different amounts of DNase I, and then subjected to Southern blot analysis. As shown, a 1.7-kb DNA fragment in the human *HGF* promoter region (+56 to -1,679) was generated when the genomic DNA was digested by BamHI and was detectable in both HeLa and C33A cells treated with low amounts of DNase I (15 U/ml). This band disappeared in C33A nuclei treated with 30 ng/ml DNase I, indicating an open chromatin configuration, hence a readily accessible promoter. On the other hand, in HeLa cells (in which DATE was 30A), the *HGF* promoter region was condensed and resistant to DNase I; a protected band of 0.9 kb appeared in HeLa cells but not in C33A cells. **(C)** Analysis of *HGF* gene activation by ChIP assays. Con, negative control with no antibody. Poly II, acetyl-histone 3 (AH3) and acetyl-histone 4 were observed in C33A cells but not in HeLa cells.

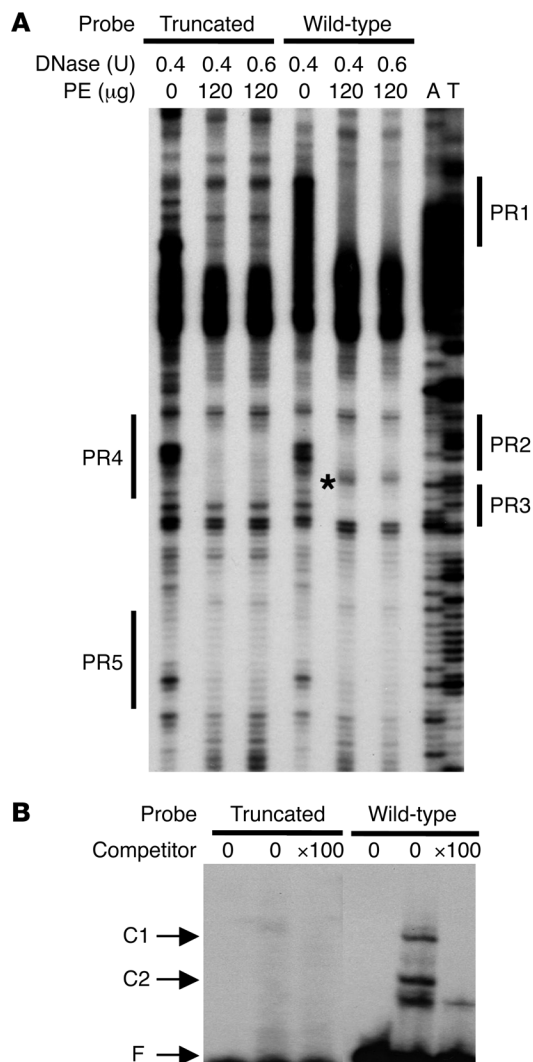
in breast cancer,  $P = 0.003$  in normal adjacent tissues with truncated DATE versus those with wild-type DATE, by 2-tailed unpaired  $t$  test). As expected, Met protein expression was expressed in all breast cancer cases and trended toward overexpression in cancer cases with the truncated DATE genotype (Supplemental Figure 3 and Supplemental Table 2). This finding is noteworthy since it has been shown that HGF exerts a positive feedback loop on Met expression by engaging the MAPK-AP1 axis (18).

**Truncated DATE significantly associates with breast cancer.** The data described above indicate that DATE is unstable in human breast cancer and is a target of deletion mutagenesis. Additionally, these data suggest that DATE is most likely polymorphic in nature, since in some cases both tumor and the corresponding normal tissue had the truncated DATE variant. Therefore, to address this possibility, we determined the incidence of truncated DATE using peripheral blood lymphocyte genomic DNA from 119 healthy AA and 473 healthy CC adult human subjects. We observed an incidence of truncated DATE (DATE with 25As or less) in the general population of approximately 26% (31/119) in AA and 3% (15/473) in CC (Table 1). The shortest truncated DATE allele that we observed in the normal population had 18As. Comparison of the prevalence of truncated DATE in breast cancer cases with that of normal subjects indicated that truncated DATE either in AAs or in CCs significantly correlates with breast cancer incidence (51% vs. 26% in AAs,  $P = 0.0080$ ; 15% vs. 3% in CCs, 2-tailed Fisher's exact test,  $P = 0.0009$ ; Table 1). To specifically investigate the occurrence of truncated DATE in normal breast tissue, we determined the frequency of truncated DATE occurrence in 50 normal adult individuals (normal human breast tissues obtained from reduction mammoplasty) (see genotyping results in Supplemental Figure 4). The results revealed that the truncated DATE was present in normal breast tissues at an overall frequency of 28% (7/25) in AAs and 4% (1/25) in CCs, in agreement with the incidence we observed in the peripheral blood DNA of normal subjects (26% in AAs and 3% in CCs; Table 1). No significant difference was detected in the frequency distribution of truncated DATE between the samples of genomic DNA obtained from normal blood and those from normal breast tissues. Stratification of the overall data from normal subjects (blood and normal breast) revealed that the truncated DATE was more frequent in the

AA population than in the CC population (26% in AAs vs. 3% in CCs,  $P < 0.0001$ , Fisher's exact test; Table 1). Further analyses of the occurrence of the truncated DATE variant with different clinicopathological features of breast cancer indicated that the average age of breast cancer patients harboring the truncated DATE variant tended to be younger than those with wild-type DATE (51 yr vs. 58 yr,  $P = 0.018$ , 2-tailed unpaired Student's  $t$  test;  $n = 93$ ).

**Healthy subjects with the truncated DATE variant have high HGF expression in breast tissues.** Since our studies established that DATE is polymorphic in nature in the general population and that truncated DATE has a functional consequence on *HGF* promoter activity and HGF expression and significantly associates with breast cancer, we were interested to determine whether normal healthy individuals carrying the truncated allele of DATE (Supplemental Figure 4) had increased *HGF* gene expression in their breast tissues. Therefore, we assessed DATE genotype status as well as expression of *HGF* mRNA and protein using quantitative RT-PCR, western blot, and ELISA in normal breast tissues obtained from reduction mammoplasty. Quantitative RT-PCR confirmed that samples with truncated DATE did indeed show at least a 5-fold higher abundance of *HGF* mRNA than those with wild-type DATE (Figure 9A). Western immunoblot also revealed higher HGF expression in tissues with truncated DATE compared with tissues with wild-type DATE (Figure 9B). In accordance with these data, ELISA results confirmed that HGF protein abundance was indeed significantly higher in cases with truncated DATE than in those with wild-type DATE (the average level of HGF protein was 162 pg/mg in truncated DATE tissues compared with 79 pg/mg in wild-type DATE, 2-tailed unpaired  $t$  test,  $P = 0.045$ ; Figure 9C). We also determined whether upregulation of HGF correlated with Met activation. Accordingly, using western blot, we assessed the levels of total Met and active Met (with specific antibodies recognizing phosphotyrosine residues in Met, the presence of which indicate autophosphorylation and Met signaling) in normal breast tissues with wild-type or truncated DATE. Although Met was detectable in all samples, phospho-Met expression was higher in cases with truncated DATE compared with cases with wild-type DATE (Figure 9B), providing evidence that a functional HGF-Met autocrine loop is driven by DATE mutation in the *HGF* promoter.





**Figure 4**

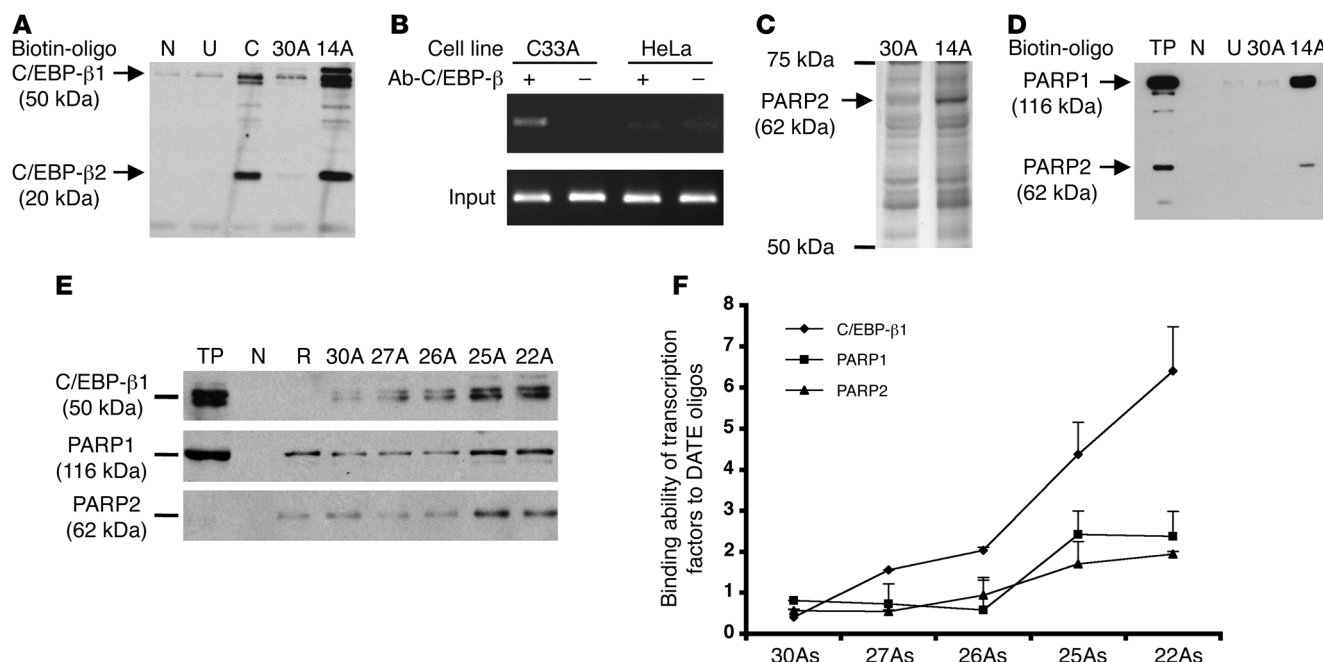
Truncated DATE affects DNA-protein interactions in the *HGF* promoter region. **(A)** Truncated DATE dictates DNA-protein interaction as confirmed by in vitro DNase I footprinting. An approximately 300-bp DNA fragment (from –928 to –627) of the *HGF* promoter region containing either the wild-type DATE (DATE with 30As) or the truncated DATE (DATE with 14As) were labeled and used as probes in side-by-side experiments. In the representative figure shown, protein extract (PE) from a human breast cancer case was incubated with these 2 probes and treated with DNase I. Four protected regions were detectable with the wild-type probe: PR1 (–753 to –764), PR2 (–803 to –810), PR3 (–814 to –821), and PR5 (–833 to –845). Two protection sites were detected with the truncated DATE probe: PR4 (–800 to –815) and PR5. Asterisk indicates DNase I hypersensitive sites (or unprotected sites) in the wild-type promoter region. **(B)** Identification of DNA-protein interactions by gel shift assay. Wild-type but not truncated DATE by itself specifically interacts with nuclear proteins. Wild-type oligo with 30As and truncated DATE (14As) were used as probes in gel shift assays. Unlabeled oligos were used as competitors to show specific binding. Two complexes, C1 and C2, were seen with the wild-type but not the truncated probe. F, free probe.

the shortening of DATE, and that truncated DATE modifies the ability of the *HGF* promoter to bind to various gene regulatory proteins. Subsequently, truncated DATE activates the *HGF* promoter in cancer cells, resulting in the acquisition of an HGF-Met autocrine loop (i.e., a gain-of-function mutation).

A key question stemming from these observations is how DATE mechanistically affects the *HGF* promoter. Several studies reported that poly(dA) promoter elements in general regulate transcription by affecting nucleosome formation (25–29). Studies have shown that long adenosine tracts (31As, 41As, and 69As in length) readily wrap around histone octamers, preventing accessibility to promoter regions, while a short adenosine tract of 17As stimulates transcription by improving accessibility to promoters in vivo (28, 29). Others have shown that a 16A element, when incorporated into nucleosomal DNA, destabilizes histone-DNA interactions and increases the accessibility of DNA target sites (30). The crystal structure of a 16 poly(dA) DNA tract shows a high degree of propeller twist, which causes the double helix to have a deep, narrow minor groove and a wide, shallow major groove, thus providing “access windows” for transcription factors (27). It was reported that the DAT1 protein encoded by the *DAT1* gene is the only known poly(dA)-binding protein in yeast; it specifically binds to the ILV1 poly(dA) DNA element and functions as a trans-activating factor (31). Chemical interference experiments showed that the DAT1 protein interacts with the minor groove of the double helix (31). Collectively, these studies highlight the importance of poly(dA) tract length and its influence on DNA structure and protein interaction. We found that the wild-type DATE acts as a repressor not only in the context of the *HGF* promoter but also out of context, when placed upstream of a heterologous promoter (data not shown), suggesting that wild-type DATE has a direct gene-silencing function. Our footprinting, gel shift, and functional experiments demonstrate that the 30A stretch hinders the recruitment of transcriptional activators like C/EBP-β1 and PARPs to the *HGF* promoter and that its shortening relieves this repression. It should also be mentioned that at this point we do not know whether the *HGF* gene that harbors truncated DATE fails to undergo silencing during the differentiation process from primitive mammary stem/mesenchymal cell to mature mammary

## Discussion

Breast cancer is the second leading cause of cancer mortality in American women. Despite significant efforts being made in the research arena, the molecular genetics of mammary tumorigenesis remain ill defined. It is well accepted that genetic factors play an important role in breast cancer and that yet to be identified susceptibility genes are involved. It is also well established that genomic instability fuels tumorigenesis, but the genes that are consequently affected and drive neoplastic growth have not been elucidated. In this work we show that the *HGF* gene, which encodes the ligand for the *Met* protooncogene (a known culprit in breast tumorigenesis), is a target of mutagenesis in breast cancer. We show that deletion mutation occurs in what we believe is a novel regulatory region located 750 bp upstream from the transcriptional start site in the human *HGF* gene promoter, which consequently results in activation of this otherwise silenced gene in epithelial cells. This DNA element (which we have dubbed “DATE”) in its wild-type form consists of a mononucleotide repeat of 30 adenosines (poly[dA]), which exerts profound local and regional effects on the *HGF* promoter by modulating DNA-protein interactions and chromatin dynamics. We show here that DATE is unstable in breast cancer cells, that loss of adenosine residues results in



**Figure 5**

Identification of C/EBP- $\beta$  binding to the *HGF* promoter by DNA affinity purification assay. **(A)** Biotin-labeled 80-mer oligos containing 30As and 14As as well as an unrelated oligo (U) and a positive control commercial C/EBP- $\beta$  oligo (C) were used in pull-down assays employing streptavidin agarose beads and subjected to western blot using C/EBP- $\beta$  antibody. A sample lacking oligos was used as a negative control (N). The ratio of C/EBP- $\beta$ 1 to C/EBP- $\beta$ 2 was higher in samples treated with 14A oligos (ratio of 1.7) as compared with samples exposed to the positive control C/EBP- $\beta$  oligo containing the canonical C/EBP- $\beta$  site (ratio of 0.4) **(B)** Confirmation of C/EBP- $\beta$  binding to the *HGF* promoter in vivo by ChIP in cells with truncated DATE (C33A) but not wild-type DATE (HeLa). **(C)** Identification of proteins differentially binding to DATE as evidenced by DNA affinity purification and silver staining. PARP2 was identified binding to the truncated DATE (14A). **(D)** Confirmation of PARP binding to the *HGF* promoter by DNA affinity purification assay and western blot. The same oligo probes described in **A** were used. TP, total crude nuclear protein. The arrows point to PARP1/2 preferably binding to 14A. **(E and F)** Gradual loss of deoxyadenosines from DATE promotes binding of C/EBP- $\beta$ 1 and PARPs. **(E)** Biotin-labeled 80-mer oligos containing wild-type (30As) or shortened DATE with incrementally reduced As (27As, 26As, 25As, and 22As) were used in DNA affinity pull-down assay and western blot. R, random oligo. **(F)** Graphical representation of data in **E** using densitometric values (arbitrary units shown on the y axis) as a function of DATE shortening (shown on the x axis).

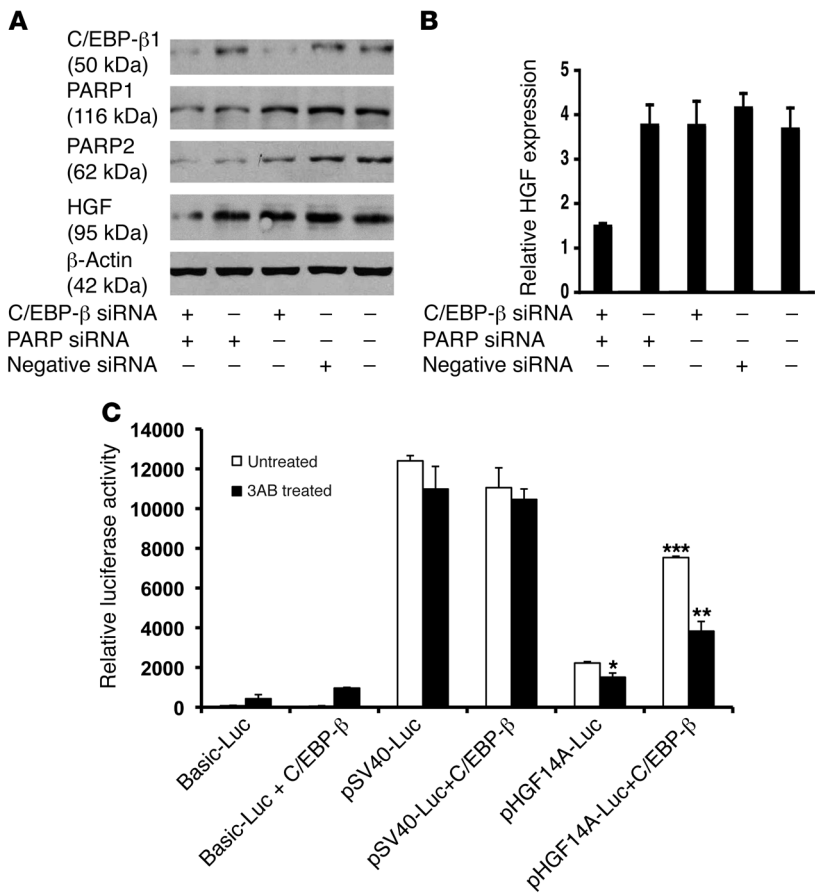
epithelial cell to breast cancer cell, or whether the *HGF* gene with truncated DATE becomes reactivated during the tumorigenic process (i.e., via de-differentiation). These possibilities should be the focus of future studies.

C/EBP- $\beta$ 1 is a transcriptional activator of a variety of genes involved in regulating not only normal growth and differentiation of mammary tissues, but also breast cancer and breast cancer-associated inflammation (32, 33). On the other hand, C/EBP- $\beta$ 2, a truncated version of C/EBP- $\beta$ 1 that is produced due to use of an in-frame internal alternative translation initiation codon, generates a C/EBP- $\beta$  isoform that retains DNA binding property but lacks a transactivation domain. Hence it can act as a dominant-negative factor by making a defective heterodimer with C/EBP- $\beta$ 1 and/or competing with C/EBP- $\beta$ 1 for DNA binding (33). Notably, we found that the C/EBP- $\beta$  binding site adjacent to DATE (having the sequence ACTTTGGGTAAATGTGTGGTATTTCTGTG) binds C/EBP- $\beta$ 1 better than the canonical C/EBP binding site (which has the sequence TGCAGATTGCGCAATCTGCA; see Supplemental Figure 2). PARP1 and PARP2 proteins were first characterized as DNA repair factors that mainly respond to DNA damage and maintain the chromatin structure and integrity. Recent studies have established that PARPs serve as transcriptional enhancers and coactivators (24,

34, 35). Our functional experiments support the notion that C/EBP- $\beta$ 1 and PARPs cooperate with each other and activate the *HGF* gene promoter (Figure 6, A-C).

It is well known that mutations (additions or deletions) of repetitive sequences (microsatellites) including the poly(dA) tracts are generated by slippage of DNA polymerases during DNA replication and that the resultant “protruding” DNA is excised by the DNA mismatch repair system. Enhanced slippage or a faulty mismatch repair system could alter this dynamic equilibrium and result in microsatellite instability (MSI) (36). We do not know the molecular mechanisms through which DATE undergoes deletion mutagenesis in tumor cells, but it is possible that, similar to microsatellites (defined as mono- or dinucleotide simple repeats), which undergo deletion mutagenesis in cancer (36–38), DATE deletion mutagenesis in cancer cells also involves defective DNA mismatch repair. It should be noted that MSI does not appear to be prevalent in breast cancer, although some report MSI to be frequent in advanced breast cancer (39). These researchers suggest that MSI is an indicator of chemotherapy resistance and is associated with poor prognosis in breast cancer patients treated with high-dose chemotherapy/autologous stem cell transplantation (HDCT/ASCT) (39). Clearly, future studies are warranted to investigate the molecular basis of DATE instability in breast cancer cells.





**Figure 6**  
C/EBP-β1 and PARPs are involved in *HGF* gene activation. (A and B) *HGF* protein expression was down-regulated in C/EBP-β1 and PARP knockdown cells. C/EBP-β1 and PARP siRNAs were transfected singly or doubly into *HGF*-expressing cells to inhibit the expression of C/EBP-β1 and PARPs. *HGF* expression was reduced by 50% when both C/EBP-β1 and PARPs were knocked down. (C) C/EBP-β1 and PARPs functioned as activators of the *HGF* gene promoter. pHGF14A-Luc promoter construct along with promoter-less and SV40 promoter-driven Luc plasmids (used as control vectors) were cotransfected with the full-length C/EBP-β expression plasmid into HeLa cells and treated without or with 10 mM 3AB. Luciferase activity of the pHGF14A-Luc but not SV40-Luc was significantly stimulated in the presence of C/EBP-β. The basal and C/EBP-β-induced activity of *HGF*, but not the SV40-Luc promoter, was significantly reduced with PARP inhibitor (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ).

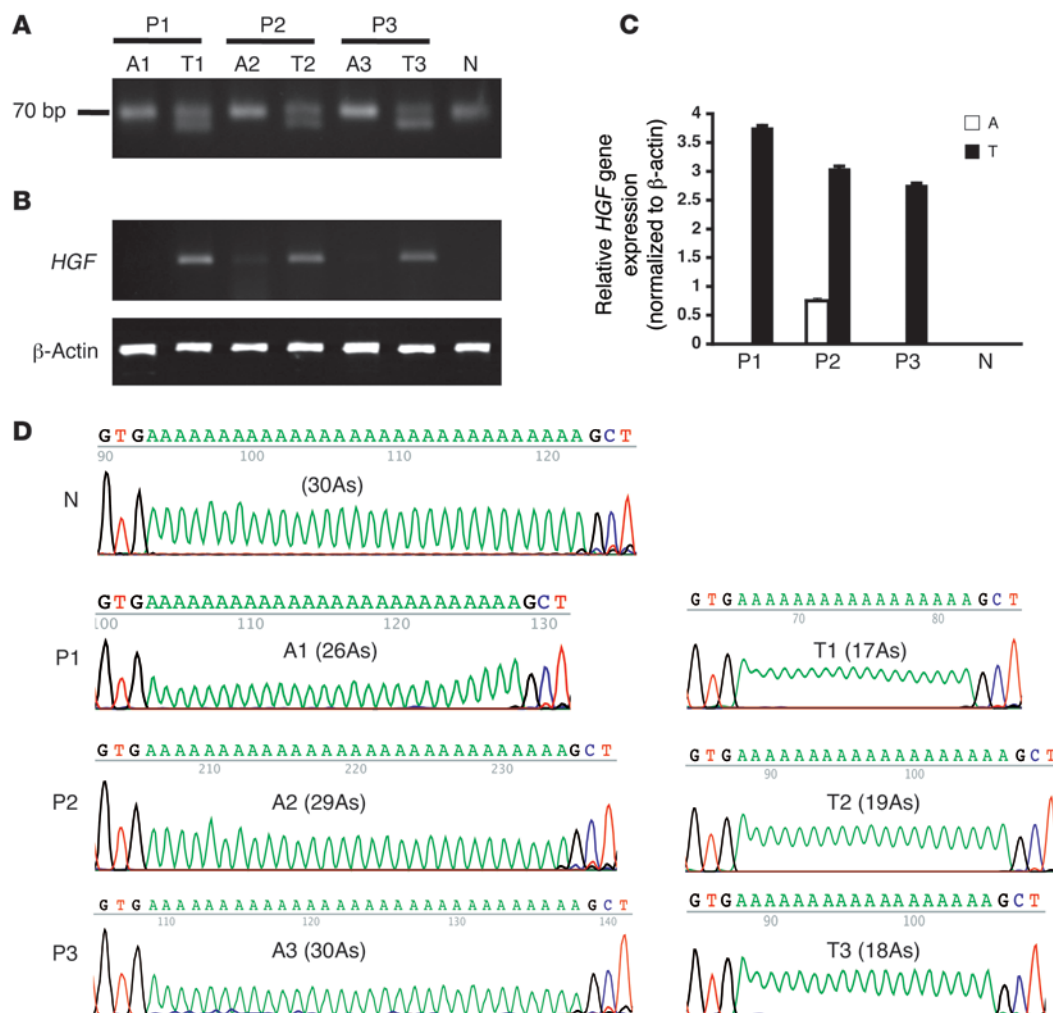
We found that DATE truncation occurs in breast cancer tissues and that truncated DATE significantly associates with human breast cancer. Another salient feature of our studies is that truncated DATE occurs in normal subjects, indicating that DATE is polymorphic and that it may be a heritable disease predisposition modifier. Our data establish that truncated DATE is carried in the normal population at a frequency of 26% (31/119) in AAs and 3% (15/473) in CCs. Our quantitative investigations of *HGF* expression in the normal breast tissues from apparently normal individuals uncovered that subjects who have truncated DATE express higher levels of *HGF* in the breast as compared with those that have wild-type DATE. These provocative data indicate that high *HGF* expression in these individuals may promote the process of breast tumorigenesis and provide a mechanistic explanation as to why truncated DATE tends to segregate with breast cancer. Interestingly, we found a very significant difference in the prevalence of truncated DATE in normal, healthy AAs versus normal CCs (26% vs. 3%, respectively;  $P < 0.0001$ ). It remains controversial as to why AA women have a higher rate of breast cancer at a younger age than do CC women (40, 41). Our results showing that truncated DATE is present more frequently in normal AAs as compared with CCs may be one reason. We should emphasize that in our analyses we were conservative in our classification of truncated DATE as a variant with 25As or fewer (Supplemental Table 1). This assignment was in

part based on the modest effects of DATE with 26As to 29As on *HGF* promoter activity and factor binding as compared with DATE with 25As or fewer (for example, see Figure 2, Figure 5, E and F, and Supplemental Figure 1). In other words, DATE with extensive deletion of 30As (i.e., loss of 5As or more) had a dramatic effect on *HGF* expression, while DATE with deletion of 4 or fewer As (i.e., 26As to 28As) had a marginal effect on *HGF* gene expression. We envision that DATEs harboring 26As, 27As, or 28As are perhaps progressing toward becoming shorter, given that mononuclear repeats are known to be highly unstable and hypermutable. In support of this contention, a comprehensive population study using matched-blood and sperm DNA addressed the issue of mononucleotide repeat stability in the human population. The authors used as a paradigm the poly(dA) repeat of 40As named “BAT40” (which is

**Table 1**  
Frequency of truncated DATE in breast cancer patients and normal subjects

DATE	Normal blood <sup>A</sup>		Normal breast		Breast cancer	
	AA	CC	AA	CC	AA <sup>B</sup>	CC <sup>C</sup>
Wild type	88	458	18	24	18	45
Truncated variant	31	15	7	1	19	8
Total	119	473	25	25	37	53

<sup>A</sup>AA vs. CC in normal blood,  $P < 0.0001$ . <sup>B</sup>Breast cancer vs. normal blood in AAs,  $P = 0.0080$  (odds ratio, 3.0; 95% confidence interval: 1.4–6.4). <sup>C</sup>Breast cancer vs. normal blood in CCs,  $P = 0.0009$  (odds ratio, 3.9; 95% confidence interval: 2.1–7.3). Analysis was performed by 2-tailed Fisher’s exact test using GraphPad and Instat 2.1 software.



### Figure 7

DATE is unstable and a target of deletion mutagenesis in human breast cancer tissue and associates with high *HGF* expression. **(A)** Representative DATE genotypes as determined by PCR and agarose gel electrophoresis. The 70-bp PCR product was analyzed on 3% agarose gel. The 3 cases (P1, P2, and P3) show the mutant allele with truncated DATE in the cancer tissues. **(B and C)** Representative RT-PCR (using 32 cycles) results for *HGF* gene expression in tumors with truncated DATE and their surrounding normal tissues. High *HGF* gene expression was readily notable in tumor tissues (T) compared with the corresponding adjacent (A) normal tissue and with normal (N) breast tissue with wild-type DATE. **(D)** Confirmation of the DATE genotype from cases described in **A** by DNA sequencing. Individual bands were cut out and sequenced. N denotes sequence from breast tissue of a normal subject showing 30As. T1, T2, and T3 denote the truncated DATE sequence from breast cancer tissues; A1, A2, and A3 denote the DATE sequence from the corresponding normal adjacent breast tissues, respectively. The DNA sequencing results showed that all tumor tissues had truncated DATE as well as long DATE: T1 had 17As and 26As, T2 had 19As and 29As, and T3 had 18As and 30As.

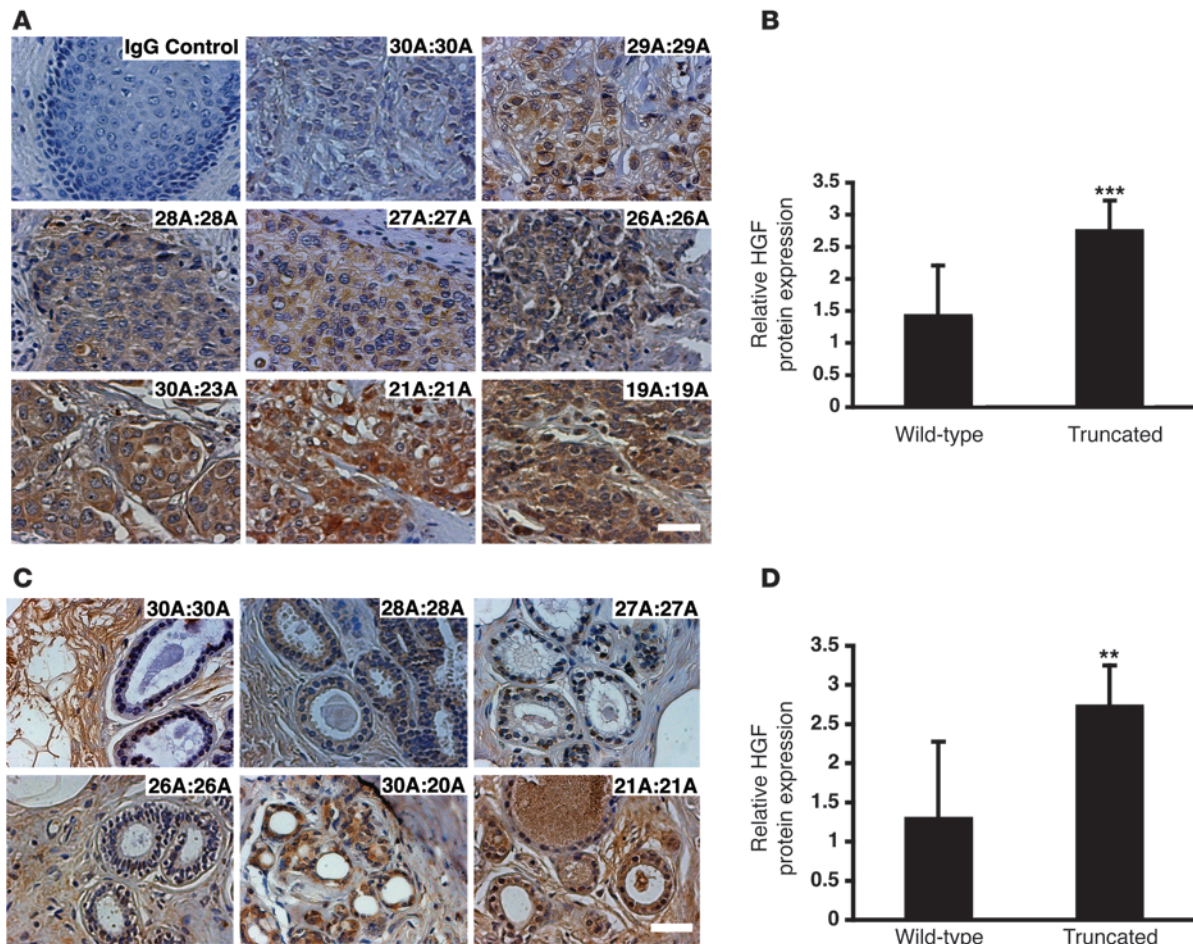
known to be a target of deletion mutagenesis in colorectal cancers, especially those with DNA mismatch repair deficiency) and found that BAT40 is polymorphic and undergoes germline mutation at a high frequency, leading the authors to conclude that long A repeats are unstable in general (42).

In summary, our data show that DATE has a functional role in governing *HGF* promoter activity and that its deletion has major effects on the *HGF* gene promoter region by aberrantly activating *HGF* expression in epithelial tumor cells such as breast carcinomas (Figure 10). Additionally, we show that the truncated DATE variant occurs in the normal population, associates with high HGF expression in normal breast tissue, and significantly segregates with breast cancer. Given our data, population studies are warranted to assess whether the truncated DATE variant correlates

with other forms of cancers known to overexpress HGF such as those derived from the colon, stomach, pancreas, endometrium, and cervix as well as gliomas and sarcomas, to name a few.

## Methods

**Cell lines and tissues.** Human breast cancer and cervical carcinoma cell lines (MCF7, HTB25, HTB30, HTB121, HTB129, HeLa, C33A, etc.) were all purchased from ATCC and cultured following the supplier's instructions. Human breast cancer and paired normal adjacent breast tissues from 95 patients as well as normal breast tissues from 50 unrelated healthy subjects undergoing reduction mammoplasty were obtained from Magee Women's Hospital, University of Pittsburgh according to protocol approved by the University of Pittsburgh institutional review board. Genomic DNA from peripheral blood of 119 healthy AAs and 473 healthy CCs was also obtained



**Figure 8**

Truncated DATE is associated with high HGF expression in human breast cancer tissues. (**A** and **B**) Representative immunohistochemistry results depicting HGF protein expression in human breast cancer cases with or without truncated DATE. Significantly higher (+++) HGF protein expression was noted in tumors with truncated DATE compared with those with wild-type DATE, which exhibited mostly moderate (++) to weak (+) expression ( $P = 8.8 \times 10^{-7}$  using unpaired 2-tailed  $t$  test,  $n = 42$ ;  $***P < 0.001$ ) (see also Supplemental Table 2). (**C** and **D**) Adjacent normal breast tissues with truncated DATE also exhibit significantly higher HGF expression as compared with those with wild-type DATE ( $P = 0.003$ , unpaired 2-tailed  $t$  test;  $n = 42$  (Supplemental Table 2).  $**P < 0.01$ . Notably, HGF was localized to epithelium in cases with truncated DATE. Scale bars: 40  $\mu$ M.

according to protocol approved by the University of Pittsburgh institutional review board. Informed consent was obtained from all subjects.

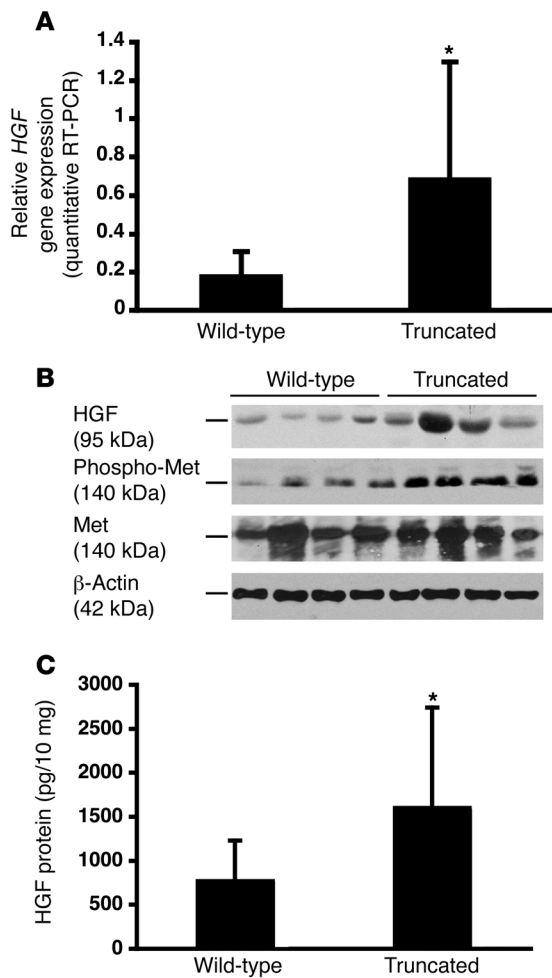
**PCR screening and quantitative RT-PCR analysis.** Genomic DNA and RNA were prepared using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Nested PCR was used to screen the poly(dA) tract mutations in the *HGF* promoter region. The outer primers (forward primer: 5'-GGGACAGGCTATGGACAATGACTGTTTCTTGG-3'; reverse primer: 5'-GGGTGTGGTATTGTGGGGCCAAATAAG-3') generated a 247-bp product, while the inner primers (forward primer: 5'-CGTGAGTTTGGCAGTTTGTG-3'; reverse primer: 5'-ATGGGCTCAGAGCAGGCAGC-3') generated a 70-bp product. The PCR products were analyzed on 3% agarose gel and 15% denaturing polyacrylamide gel containing 8 M urea. The mutant samples were subcloned into PCR 2.1 vector (Invitrogen), and then 2–5 clones were chosen for automated sequencing by the University of Pittsburgh Sequencing Core facility. *HGF* mRNA expression was detected by RT-PCR and quantitative RT-PCR analysis. Total RNA was reverse transcribed using 2 mg of total RNA with AMV Reverse Transcriptase (Promega Corp.) and Random Primer (Roche Applied Sciences) in a volume of 40  $\mu$ l according to standard

procedures. The forward primer was from exon 11 (5'-CTCATTCCTTGGGATTATTGCCC-3'), and the reverse primer was from exon 14 (5'-CCTTTATCAATGCTCCGAG-3'). A 1:5 dilution of the reverse transcriptase product was used to carry out quantitative PCR analysis for 40 cycles using the ABI Prism 7000 Sequence Detection System with reagents and primers specific to human *HGF* supplied by Applied Biosystems. As an internal control, the 18S rRNA primer was included to determine the quality of the sample, and diethylpyrocarbonate-treated water was also included as a negative control for contamination. Human *HGF* cDNA diluted to  $1 \times 10^5$ ,  $1 \times 10^4$ , and  $1 \times 10^3$  copies was subjected to quantitative PCR analysis for quantification purposes.

**Preparation of nuclei, nuclear protein, and protein extracts.** The nuclei and nuclear protein extracts were prepared by methods described previously (16–19). For preparation of protein extracts, 500 mg of breast cancer tissue in 500  $\mu$ l of RIPA buffer was homogenized. The cell pellets were subjected to 3 freeze-thaw cycles and the cell lysates were cleared by centrifuging at 11,000  $g$  for 20 minutes.

**Plasmids, DNA transient transfection and luciferase assay.** The human *HGF* promoter region (1.1 kb; –1037 to +56) containing wild-type or DATE



**Figure 9**

Normal breast tissues with the truncated DATE variant show high HGF gene expression. **(A)** Quantification of *HGF* mRNA expression in normal breast tissues (obtained from reduction mammoplasty) by quantitative RT-PCR. *HGF* gene expression was significantly higher in normal breast tissues with truncated DATE compared with tissues with wild-type DATE. \* $P < 0.05$ , 2-tailed unpaired  $t$  test. **(B)** Representative western blot analysis of HGF, Met, and phospho-Met protein expression in normal breast tissues. **(C)** Quantification of HGF protein level by ELISA in normal breast tissues. HGF ELISA kit (R&D Systems) was used to quantify HGF protein in 48 different normal breast tissues with wild-type and truncated genotypes. The mean level of HGF protein was 162 ng/mg in cases with the truncated DATE, compared with 79 ng/mg in cases with wild-type DATE. \* $P < 0.05$ , 2-tailed unpaired  $t$  test.

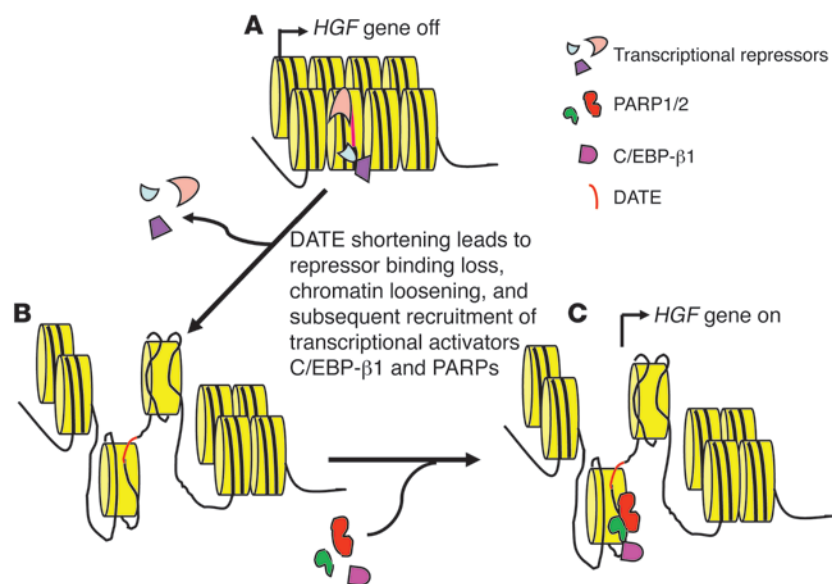
at 37°C. Genomic DNA was then extracted from nuclei and digested by BamHI. Equivalent amounts of DNA were used for Southern blots using a 247-bp probe (–902 to –655) in the DATE region. Molecular size standards ( $\phi$ X174 DNA marker) were used to estimate the size of the DNase I hypersensitive fragment.

**In vitro DNase I footprinting.** The *HGF* promoter region between –928 and –627 (300 bp) containing truncated DATE with 14As or wild-type DATE (DATE with 30As) was produced by PCR and labeled at the 5' end of the primers using the T4 Labeling Kit from ICN Radiochemical and [ $\gamma$ - $^{32}$ P]–dATP. Footprinting was carried out in a total of 200  $\mu$ l of binding reaction mixture containing 200 ng (600 fmol) of the double-strand probes with or without 120  $\mu$ g of nuclear protein extracts and, as indicated in Figure 4A, with whole-protein extracts from human breast cancer tissues in the presence of 18  $\mu$ g of poly(dI-dC) (Pharmacia) in 1 $\times$  binding buffer (50 mM Tris-HCl, 100 mM KCl, 12.5 mM MgCl<sub>2</sub>, 1 mM EDTA, 20% glycerol, 1 mM DTT). After incubation at 37°C for 30 min, 0.4–0.6 U of DNase I (Promega Corp.) were added to the mixture for 1 minute at room temperature. Digestion was stopped by adding stop buffer (12.5 mM EDTA, 12.5 mg/ml proteinase K, 125 mg/ml yeast transfer RNA, and 0.1% SDS). The DNA fragments were isolated after incubation at 37°C for 10 minutes and analyzed on 6% polyacrylamide containing 7 M urea sequence gels, then subjected to autoradiography. Sequencing of the same [ $\gamma$ - $^{32}$ P]–dATP-labeled probes was performed as described by DNA cycle sequencing system (Promega Corp.), and the sequencing reaction mixture was loaded adjacent to the samples analyzed by DNase I footprinting.

**Gel shift assay.** Synthetic double-stranded 40 mers of oligonucleotides containing 30As (wild type, forward strand: 5'-CGCGTTGGCAGTTTGTGAAAAAAAAAAAAAAAAAAAAAAAAAGCTGCCTGCTCTGAGCC-3'; reverse strand: 5'-GATCAACCGTCAAA-CACTTTTTTTTTTTTTTTTTTTTTTTTTTTCGACGGACGAGACTCGG-3') and 30 mers of oligonucleotides containing 14As (truncated, forward strand: 5'-CGCGTTGGCAGTTTGTGAAAAAAAAAAAAAAAAAGCTGCCTGCTCTGAGCC-3'; reverse strand: 5'-GATCAACCGTCAAA-CACTTTTTTTTTTTTTTTCGACGGACGAGACTCGG-3') were synthesized by the core facility at the University of Pittsburgh. The probes were labeled with [ $\gamma$ - $^{32}$ P]–dCTP (ICN Radiochemical) by fill-in technique; poly(dI-dC) was used as the nonspecific competitor. To a 10- $\mu$ l reaction, 4 mg of nuclear extract, 0.3–0.4 ng of probes, and 4 mg of poly(dI-dC) were added. The binding reaction was carried out at room temperature for 15 minutes before loading on 5% native polyacrylamide (19:1) gels. For competition experiments, 50- to 100-fold molar excess of unlabeled oligonucleotides were included in the reaction mixture. Gels were run in 0.5 $\times$  TBE buffer (45 mM Tris-borate, 1 mM EDTA) at a constant voltage of 190 V and dried, followed by autoradiography in the presence of intensifying screen.

of different sizes (30A, 29A, 28A, 27A, 26A, 25A, 20A, and 14A) from different breast cancer patients was amplified by PCR. One restriction enzyme site (BamHI) was incorporated into the forward primer at the 5' end. Following digestion with BamHI, the DNA fragments were inserted upstream of the luciferase reporter gene in the pGL2-basic vector (Promega Corp.) at the BglII site, resulting in the following constructs: pHGF30A-Luc, pHGF29A-Luc, pHGF28A-Luc, pHGF27A-Luc, pHGF26A-Luc, pHGF25A-Luc, pHGF20A-Luc, and HGF14A-Luc. HeLa cells were seeded in 12-well culture plates ( $1 \times 10^5$  cells/well) and cultured in DMEM medium for 24 hours before transfection. The renilla luciferase-encoding plasmid (pRL-TK; expression of renilla luciferase activity was used as an internal standard for transfection efficiency; Promega Corp.) and pHGF-Luc plasmids were transiently cotransfected into HeLa cells using Lipofectamine 2000 (Invitrogen) to detect the promoter activity in the context of different lengths of DATE. For C/EBP- $\beta$  transcriptional activation assays, C/EBP- $\beta$ 1 expression vector, pRL-TK plasmid, and pHGF-Luc plasmids were cotransfected into HeLa cells. The cells were treated with or without 10 mM 3AB to inhibit PARPs. The promoter activities were measured using the Dual Luciferase Reporter Assay System (Promega). After 24 hours transfected cells were collected and luciferase activity was measured using a luminometer (EG & G Berthold).

**In vivo DNase I hypersensitivity assay.** Freshly isolated nuclei from logarithmically grown cultured cells were subjected to DNase I digestion for 2 minutes with DNase I. The reaction was stopped by adding 0.5% SDS, 10 mM EDTA, and 50 mg/ml proteinase K and incubated for 10 minutes

**Figure 10**

Proposed model of *HGF* gene regulation by DATE. **(A)** Wild-type DATE keeps chromatin in a highly condensed conformation, preventing transcriptional activators from accessing the promoter and leading to silencing of the gene. **(B and C)** Truncated DATE causes repressor release, chromatin remodeling, and recruitment of transcriptional activators C/EBP-β1 and PARP1/2 to the promoter, leading to *HGF* gene activation.

**Formaldehyde cross-linking and immunoprecipitation of ChIP assay.** HeLa and C33A cell lines were used for ChIP assays. Formaldehyde (1% final concentration; Fisher Scientific) was added directly to cell culture for fixation at room temperature for 10 minutes. Glycine (125 mM final concentration) was added to stop the reaction at room temperature for 5 minutes. Cell lysates were sonicated 5 times at 15-second pulses, shearing the chromatin into lengths of approximately 200–500 bp. After that, cell lysates were pre-cleared by incubating 2 ml of the samples with 80 µl of protein A sepharose for 1 hour at 4°C. Then the DNA-protein complexes were incubated with 10 mg/ml BSA to reduce background and immunoprecipitated with 1 µg of C/EBP-β (Santa Cruz Biotechnology Inc.), poly II (Abcam Inc.), or acetyl-histone 3 or acetyl-histone 4 (Millipore) antibodies. No antibody was added as the negative control. The mixtures were then incubated overnight at 4°C with rocking. The next day, 2 ml of the IP mixture was incubated with 60 µl of protein A sepharose and 20 µl of 100 mg/ml yeast transfer RNA at 4°C for 1 hour. The protein-DNA complexes were then eluted after washing. After adding 1 µl of 10 mg/ml RNase A, they were incubated at 65°C for 5 hours to reverse the formaldehyde cross-links, followed by precipitation at –20°C overnight. Then, the complexes were incubated in 100 µl 1× proteinase K digesting buffer with 1.5 µl of 20 mg/ml proteinase K at 42°C for 2 hours. DNA was purified using phenol-chloroform. PCR was used (forward primer: 5'-GGGACAGGCTATGGACAATGACTGTTTCTTGG-3'; reverse primer: 5'-GGGTGTGGTATTGTGGGGC-CAAAATAAG-3') to amplify the 247-bp DNA fragment from –902 to –655 for C/EBP-β or the 282-bp DNA fragment from –12 to –293 (forward primer: 5'-AGCTCCAGCTTCCAAATTGA-3'; reverse primer: 5'-CGGCTCTGAGCTGCTTTTA-3') for poly II, acetyl-histone 3, and acetyl-histone 4.

**Immunohistochemistry and immunofluorescence.** Archival breast tissue sections were first deparaffinized with xylene and ethanol and then incubated with polyclonal antibodies specific to human HGF (R&D Systems) and Met (C-28; Santa Cruz Biotechnology Inc.) overnight at 4°C. Non-immune IgG primary antibody acted as a negative control. After incubation with HRP-conjugated secondary antibody for 30 minutes at room temperature, tissues were developed using the ABC kit from Cell Signaling Technology Inc. HGF or Met immunoreactivity was assessed by employing an arbitrary score of 0 to +++, in which 0 indicated no staining, “+” indicated faint staining with 25% of cells positive, “++” indicated 50% of cells staining with moderate intensity, and “+++” indicated intense staining in

which at least 80% of cells were positive for HGF or Met. Random fields totaling more than 1,000 cells were assessed by light microscopy to obtain unbiased data. For immunofluorescence staining for HGF and Met, cells cultured on coverslips were fixed with methanol, incubated with primary antibody at approximately 1:50 dilution, washed, and then incubated with biotinylated secondary antibody (Vector Laboratories), indocarbocyanine, Cy2 (goat, green 1:100), and Cy3 (rabbit, red 1:200) for 30 minutes. Slides were rinsed using a wash buffer (PBS containing 0.1% Tween-20) for 3× for 10 minutes each time. The nuclei were stained blue by VECTASHIELD Mounting Medium with DAPI (Vector Laboratories).

**RNA interference.** The siRNA for human C/EBP-β and PARP1 were purchased from Ambion, and knockdown experiments were performed using the silencer siRNA transfection II kit. In brief, C33A cells were plated at 1 × 10<sup>5</sup>/ml in 6 wells, and 3.75–5 ml of 20 µM siRNA were utilized to transfect C33A cells in 2.5 ml final transfection volume. Cell lysates were assessed by western blotting with specific antibodies after 72 hours.

**DNA affinity purification and western blot assay.** The 5' biotinylated oligos were synthesized by Invitrogen. The sequence of wild-type 30A oligo was 5'-ACTTTGGGTAATGTGTGGTATTTCGTGAGTTTGGCAGTTTGTG-AAAAAAAAAAAAAAAAAAAAAAAAAAGCTG-3'. The truncated DATE oligos (27As, 26As, 25As, 22As, and 14As) contained an identical surrounding sequence and the indicated length of As. The –827 oligo consisted of a 44-bp oligo lying upstream of the DATE region. The biotin-labeled oligos and pre-cleared nuclear protein extracts were mixed with streptavidin agarose beads (Millipore) at 4°C overnight with rocking, then beads were added for an additional 2 hours at 4°C. After washing the beads 4 times with RIPA buffer, the immunoprecipitated proteins were eluted using 1× loading buffer by boiling for 5 minutes. Proteins were analyzed by silver stain (Bio-Rad) or by western blot assay. PARP antibody (BIOMOL International Inc.) and C/EBP-β antibody (Santa Cruz Biotechnology Inc.) were used in western blots.

**HGF quantification by ELISA.** HGF concentration was determined using a human HGF ELISA kit (R&D Systems) as previously described (24). Briefly, 0.5 g of frozen breast tissue was homogenized in 1 ml of 1 M NaCl to extract HGF. The lysate was then diluted by water to a final concentration of 0.2 M NaCl (low-salt buffer) and incubated with 0.5 ml of heparin agarose beads (Sigma-Aldrich) in low-salt buffer for 2 hours to purify HGF by batch adsorption. The agarose beads were washed with 0.2 M NaCl solution 3 times, and bound HGF was eluted with 1 M NaCl. The eluate was diluted, and 10-µl



fractions were subjected to western blot and ELISA using a prepackaged HGF ELISA plate according to the manufacturer's instructions. The final concentration was calculated according to the standards provided by the kit.

**Statistics.** Data were analyzed by Fisher's exact test and Student's *t* test using InStat 2.1 and GraphPad software. A *P* value of 0.05 or less was considered significant.

## Acknowledgments

This work was supported by RO1 grants from NIH (5R01ES06109-14 and 2R01CA095782-06, to R. Zarnegar). We thank Xinping Tan

for his technical assistance with the immunohistochemistry and immunofluorescence studies.

Received for publication April 21, 2008, and accepted in revised form December 10, 2008.

Address correspondence to: Reza Zarnegar, Department of Pathology, Room S-411A Biomedical Science Tower, 200 Lothrop Street, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA. Phone: (412) 648-8657; Fax: (412) 648-1916; E-mail: rezazar@pitt.edu.

- Hanahan, D., and Weinberg, R.A. 2000. The hallmarks of cancer. *Cell*. **100**:57-70.
- Trusolino, L., and Comoglio, P.M. 2002. Scatter-factor and semaphorin receptors: Cell signaling for invasion growth. *Nat. Rev. Cancer*. **2**:289-300.
- Birchmeier, C., Birchmeier, W., Gherardi, E., and Vande Woude, G.F. 2003. Met, metastasis, motility and more. *Nat. Rev. Mol. Cell Biol.* **4**:915-925.
- Wang, X., et al. 2002. A mechanism of cell survival: sequestration of Fas by the HGF receptor Met. *Mol. Cell*. **9**:411-421.
- Zhang, Y.-W., and Vande Woude, G.F. 2003. HGF/SF-met signaling in the control of branching morphogenesis and invasion. *J. Cell Biochem.* **88**:408-417.
- Bell, A., Chen, Q., DeFrances, M.C., Michalopoulos, G.K., and Zarnegar, R. 1999. The five amino acid-deleted isoform of hepatocyte growth factor promotes carcinogenesis in transgenic mice. *Oncogene*. **18**:887-895.
- Corso, S., Comoglio, P.M., and Giordano, S. 2005. Cancer therapy: can the challenge be MET? *Trends Mol. Med.* **11**:284-292.
- Gallego, M.I., Brier, B., and Hennighausen, L. 2003. Targeted expression of HGF/SF in mouse mammary epithelium leads to metastatic adenocarcinomas through the activation of multiple signal transduction pathways. *Oncogene*. **22**:8498-8508.
- Bogenrieder, T., and Meenhard, H. 2003. Axis of evil: molecular mechanisms of cancer metastasis. *Oncogene*. **22**:6524-6536.
- Yamashita, J., et al. 1994. Immunoreactive hepatocyte growth factor is a strong and independent predictor of recurrence and survival in human breast cancer. *Cancer Res.* **54**:1630-1633.
- Toi, M., et al. 1998. Significance of circulating hepatocyte growth factor level as a prognostic indicator in primary breast cancer. *Clin. Cancer Res.* **4**:659-664.
- Taniguchi, T., et al. 1995. Serum concentrations of hepatocyte growth factor in breast cancer patients. *Clin. Cancer Res.* **1**:1031-1034.
- Kang, J.Y., et al. 2003. Tissue microarray analysis of hepatocyte growth factor/Met pathway components reveals a role for Met, triptase, and hepatocyte growth factor activator inhibitor 1 in the progression of node-negative breast cancer. *Cancer Res.* **63**:1101-1105.
- Parr, C., Watkins, G., Mansel, R.E., and Jiang, W.G. 2004. The hepatocyte growth factor regulatory factors in human breast cancer. *Clin. Cancer Res.* **10**:202-211.
- Tuck, A.B., Park, M., Sterns, E.E., Boag, A., and Elliott, B.E. 1996. Coexpression of hepatocyte growth factor and receptor (Met) in human breast carcinoma. *Am. J. Pathol.* **148**:225-232.
- Jiang, J.G., Johnson, C., and Zarnegar, R. 2001. Peroxisome proliferator-activated receptor gamma-mediated transcriptional up-regulation of the hepatocyte growth factor gene promoter via a novel composite cis-acting element. *J. Biol. Chem.* **276**:25049-25056.
- Jiang, J.G., Bell, A., Liu, Y., and Zarnegar, R. 1997. Transcriptional regulation of the hepatocyte growth factor gene by the nuclear receptors chicken ovalbumin upstream promoter transcription factor and estrogen receptor. *J. Biol. Chem.* **272**:3928-3934.
- Seol, D.-W., Chen, Q., and Zarnegar, R. 2000. Transcriptional activation of the Hepatocyte Growth Factor receptor (*c-met*) gene by its ligand (Hepatocyte Growth Factor) is mediated through AP-1. *Oncogene*. **19**:1132-1137.
- Jiang, J.G., and Zarnegar, R. 1997. A novel transcriptional regulatory region within the core promoter of the hepatocyte growth factor gene is responsible for its inducibility by cytokines via the C/EBP family of transcription factors. *Mol. Cell Biol.* **17**:5758-5770.
- Bell, A.W., Jiang, J.G., Chen, Q., Liu, Y., and Zarnegar, R. 1998. The upstream regulatory regions of the hepatocyte growth factor gene promoter are essential for its expression in transgenic mice. *J. Biol. Chem.* **273**:6900-6908.
- Elliott, B.E., Hung, W.L., Boag, A.H., and Tuck, A.B. 2002. The role of hepatocyte growth factor (scatter factor) in epithelial-mesenchymal transition and breast cancer. *Can. J. Physiol. Pharmacol.* **80**:91-102.
- Wojcik, E.J., et al. 2006. A novel activating function of c-Src and Stat3 on HGF transcription in mammary carcinoma cells. *Oncogene*. **25**:2773-2784.
- Price, M., Fivaz, J., Jotteran, A., and Mirkovitch, J. 1998. Tissue-specific chromatin structure at the hepatocyte growth factor/scatter factor gene promoter. *Gene*. **211**:141-150.
- Zou, C., et al. 2007. Lack of Fas antagonism by Met in human fatty liver disease. *Nat. Med.* **13**:1078-1085.
- Shimizu, M., Mori, T., Sakurai, T., and Shindo, H. 2000. Destabilization of nucleosomes by poly(dA)poly(dT) tracts in vivo. *EMBO J.* **19**:3358-3365.
- Koch, K.A., and Thiele, D.J. 1999. Functional analysis of a homopolymeric (dA-dT) element that provides nucleosomal access to yeast and mammalian transcription factors. *J. Biol. Chem.* **274**:23752-23760.
- Fox, K.R. 1992. Wrapping of genomic poly dA.poly dT tracts around nucleosome core particles. *Nucleic Acid Res.* **20**:1235-1242.
- Iyer, V., and Struhl, K. 1995. Poly (dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J.* **14**:2570-2579.
- Anderson, J.D., and Widom, J. 2001. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA targets. *Mol. Cell Biol.* **21**:3830-3839.
- Reardon, B.J., Winters, R.S., Gordon, D., and Winters, E. 1993. A peptide motif that recognizes AT tracts in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **90**:11327-11331.
- Zahnaw, C., Younes, P., Laucirica, R., and Rosen, J.M. 1997. Overexpression of C/EBPβ-LIP, a naturally occurring dominant negative transcription factor in human breast cancer. *J. Natl. Cancer Inst.* **89**:1887-1891.
- Yokota, T., et al. 2007. Differential regulation of elafin in normal and tumor-derived mammary epithelial cells is mediated by CCAAT/enhancer binding protein beta. *Cancer Res.* **67**:11272-11283.
- Cervellera, M.N., and Sala, A. 2000. Poly (ADP-ribose) polymerase is a B-MYB coactivator. *J. Biol. Chem.* **275**:10692-10696.
- Kraus, W.L., and Lis, J.T. 2003. PARP goes transcription. *Cell*. **113**:677-683.
- Elser, M., et al. 2008. Poly(ADP-ribose) polymerase 1 promotes tumor cell survival by coactivating hypoxia-inducible factor-1-dependent gene expression. *Mol. Cancer Res.* **6**:282-290.
- Loeb, L.A. 2001. A mutator phenotype in cancer. *Cancer Res.* **61**:3230-3239.
- Kim, K.M., et al. 2002. PolyA deletions in hereditary nonpolyposis colorectal cancer: mutations before a gatekeeper. *Am. J. Pathol.* **160**:1503-1506.
- Samowitz, W.S., Slattery, M.L., Potter, J.D., and Leppert, M.F. 1999. BAT-26 and BAT-40 instability in colorectal adenomas and carcinomas and germline polymorphisms. *Am. J. Pathol.* **154**:1637-1641.
- Wild, P.J., et al. 2004. Microsatellite instability predicts poor short-term survival in patients with advanced breast cancer after high-dose chemotherapy and autologous stem-cell transplantation. *Clin. Cancer Res.* **10**:556-564.
- Soler-Vila, H., Kasl, S.V., and Jones, B.A. 2003. Prognostic significance of psychosocial factors in African-American and white breast cancer patients: a population-based study. *Cancer*. **98**:1299-1308.
- Ademuyiwa, F.O., and Olopade, O.I. 2003. Racial differences in genetic factors associated with breast cancer. *Cancer Metastasis Rev.* **22**:47-53.
- Bacon, A.L., Dunlop, M.G., and Farrington, S.M. 2001. Hypermutability at a poly (A/T) tract in the human germline. *Nucleic Acids Res.* **29**:4405-4413.