DNA METHYLATION REPELS BINDING OF HIF TRANSCRIPTION FACTORS TO MAINTAIN TUMOUR IMMUNOTOLERANCE

AUTHORS: FLORA D’ANNA§1,2, LAURIEN VAN DYCK§1,2, JIEYI XIONG§1,2, HUI ZHAO§1,2, REBECCA V. BERRENS3,4, JULIE DE SMEDT5, PAWEL BIENIASZ-Krzywiec1,6, VIKAS CHANDRA13, LUC SCHOONJANS1,18,19, JASON MATTHEWS12, LIESBETH MINNOYE1,2, JUNBIN QIAN1,2, RICARDO AMORIM1,6, SEPIDEH Khorasanizadeh13, MARIE DE BORRE15, SAVVAS N. SAVVIDES7,8, CELESTE SIMON16,17, PETER CARMELIET1,18,19, WOLF REIK3,9,10,11, FRAYDOON RASTINEJAD13,14, MASSIMILANO MAZZONE1,6, BERNARD THIENPONT*1,2,15, DIETHER LAMBRECHTS*1,2

AFFILIATIONS:

1. Center for Cancer Biology, VIB, 3000 Leuven, Belgium
2. Laboratory of Translational Genetics, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium
3. Epigenetics Programme, Babraham Institute, Cambridge, CB22 3AT, UK
4. University of Cambridge, The Old Schools, Trinity Lane Cambridge, CB2 1TN, UK
5. Laboratory of Dermatology, Department of Oncology, KU Leuven, 3000 Leuven, Belgium
6. Laboratory of Molecular Oncology and Angiogenesis, Department of Oncology, KU Leuven, 3000 Leuven, Belgium
7. Unit for Structural Biology, Department of Biochemistry and Microbiology, Ghent University, 9052 Ghent, Belgium
8. VIB Center for Inflammation Research, 9052 Ghent, Belgium
9. Centre for Trophoblast Research, University of Cambridge, Cambridge, CB2 3EG, UK
10. Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK
11. Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 3EG, UK
12. Institute of Basic Medical Sciences, University of Oslo, 0372 Oslo, Norway
13. Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK

14. Clinical and Experimental Endocrinology, Department of Chronic Diseases, Metabolism and Ageing, KU Leuven, 3000 Leuven, Belgium

15. Laboratory for Functional Epigenetics, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium

15. Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

16. Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

17. State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, China

18. Laboratory of Angiogenesis and Vascular Metabolism, Department of Oncology, Leuven Cancer Institute, KU Leuven, 3000 Leuven, Belgium

§ These authors contributed equally.

* Correspondence to bernard.thienpont@kuleuven.be or diether.lambrechts@vib-kuleuven.be
**ABSTRACT**

**Background:** Hypoxia is pervasive in cancer and other diseases. Cells sense and adapt to hypoxia by activating hypoxia-inducible transcription factors (HIFs), but it is still an outstanding question why cell types differ in their transcriptional response to hypoxia.

**Results:** Here, we report that HIFs fail to bind CpG dinucleotides that are methylated in their consensus binding sequence, both in *in vitro* biochemical binding assays and *in vivo* studies of differentially methylated isogenic cell lines. Based on in silico structural modelling, we show that 5-methylcytosine indeed causes steric hindrance in the HIF binding pocket. A model wherein cell-type-specific methylation landscapes, as laid-down by the differential expression and binding of other transcription factors under normoxia control cell-type-specific hypoxia responses is observed. We also discover ectopic HIF binding sites in repeat regions which are normally methylated. Genetic and pharmacological DNA demethylation, but also cancer-associated DNA hypomethylation, expose these binding sites, inducing HIF-dependent expression of cryptic transcripts. In line with such cryptic transcripts being more prone to cause double-stranded RNA and viral mimicry, we observe low DNA methylation and high cryptic transcript expression in tumours with high immune checkpoint expression, but not in tumours with low immune checkpoint expression, where they would compromise tumour immunotolerance. In a low-immunogenic tumour model, DNA demethylation upregulates cryptic transcript expression in a HIF-dependent manner, causing immune activation and reducing tumour growth.

**Conclusions:** Our data elucidate the mechanism underlying cell-type specific responses to hypoxia, and suggest DNA methylation and hypoxia to underlie tumour immunotolerance.

**Keywords**

DNA methylation; hypoxia; HIF; cryptic transcripts; immunotherapy; cancer; transcription factor binding
**BACKGROUND**

DNA methylation is central to establishing and maintaining tissue-specific gene expression, and an important contributing factor to oncogenesis. We recently demonstrated that tumour hypoxia drives DNA methylation of tumour suppressor genes by reducing the activity of TET DNA demethylases\(^1\). An outstanding question is, however, if and how DNA methylation changes mechanistically also influence the response of tumours to hypoxia. Indeed, recent evidence suggests that, contrary to traditional concepts, DNA methylation generally does not directly impede transcription factor (TF) binding, but rather acts indirectly by synergizing with other epigenetic marks\(^2\).

The hypoxia response is canonically executed by HIFs, which are heterodimeric TF complexes composed of an O\(_2\)-labile \(\alpha\)-subunit (HIF1\(\alpha\), HIF2\(\alpha\) or HIF3\(\alpha\)) and a stable \(\beta\)-subunit (HIF1\(\beta\)). The constitutively expressed HIF\(\alpha\)-subunits are directly targeted for proteasomal degradation under normal oxygen tension (normoxia), but stabilized under limiting oxygen conditions (hypoxia), when they translocate to the nucleus to induce expression of hypoxia-responsive genes. This induction of hypoxia-responsive genes occurs rapidly, often within minutes following hypoxia\(^3\). In tumours, hypoxia is widespread and leads to transcriptional activation of numerous cancer hallmark genes involved in cell survival, angiogenesis and invasion\(^4\). Interestingly, the impact of hypoxia differs among cell types. For instance, endothelial cells proliferate and migrate towards hypoxic regions, macrophages become immunosuppressive and CD8\(^+\) T-cell activation is enhanced under hypoxia\(^5-8\). Also, tumours affecting different organs exhibit divergent phenotypic responses to hypoxia\(^6\). In line with this, concordance between HIF binding sites in MCF7 breast and 786-O renal cell carcinoma cell lines is only 40–60% (Schödel et al., 2011). This divergence is particularly intriguing because HIF\(\alpha\) paralogues are often expressed at similar levels in different cell types, and because the consensus DNA sequence that binds HIF complexes, i.e., the hypoxia response element (HRE) RCGTG, does not differ between HIF\(\alpha\) paralogues or cell types. Thus, although the concept that genes induced by hypoxia differ dramatically between cancer and cell types is well-
established\textsuperscript{3, 9-11}, the reason for these divergent responses and expression programs is poorly understood.

One possibility is that the underlying cell-type-specific patterns of chromatin determine which HIF target genes are accessible and hence become expressed following acute hypoxia. Interestingly, HIFs are recruited to genes that are already expressed in normoxic cells\textsuperscript{11}, suggesting that perhaps DNA methylation could determine accessibility of the HIF complex to the RCGTG core sequence. HIF binding to the erythropoietin promoter was indeed suggested to be sensitive to DNA methylation\textsuperscript{12}, but this observation relied on gel shift binding assays, which are known to poorly reflect the authentic setting in cells. Indeed, the binding of the transcriptional repressor CTCF also appeared to be methylation-sensitive in gel shift binding assays\textsuperscript{13}, but analyses of its binding preference in living cells mostly failed to reveal methylation sensitivity\textsuperscript{14}. We therefore set out to investigate whether DNA methylation directly repels HIF binding in living cells, and whether cell-type-specific DNA methylation patterns established under normoxic conditions determine genome-wide HIF binding profiles, defining the response to hypoxia.

RESULTS

DNA methylation of HRE sites anti-correlates with HIF binding

To investigate the role of DNA methylation in HIF binding, we stabilized HIFs in MCF7 breast cancer cells by culturing them under hypoxia (0.5% O\textsubscript{2} for 16 hours; Figures S1a and S2, and Methods). We next performed chromatin-immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) for HIF1\textbeta, which is the obligate dimerization partner of HIF1\textalpha, HIF2\textalpha and HIF3\textalpha. Model-based analysis for ChIP-seq (MACS)\textsuperscript{15} revealed 7,153 HIF1\textbeta binding peaks (Figure 1a, Supplementary Table 1). These were high-quality, bona fide HIF binding regions: they were 4.6-fold enriched for the HRE motif (RCGTG), enriched near genes involved in the hypoxia response, >90% overlapping with peaks identified in another HIF1\textbeta ChIP-seq dataset on MCF7 cells and reproducibly detected in independent repeats (Figure S1b-d).
To assess methylation in these 7,153 HIF1β binding peaks, we performed target enrichment-based bisulfite sequencing (BS-seq) on DNA extracted from normoxic MCF7 cells, in which HIF is inactive, obtaining >40× coverage for ~86% of the HIF1β binding peaks identified by ChIP-seq. The methylation level at these peaks was invariably low (4.95±0.15%) compared to average CpG methylation levels detected in the genome (61.6±0.07%, Wilcoxon test P<2.2−16, Figure 1b). Results were confirmed using another whole-genome BS-seq dataset (Figure 1a)16. Also when quantifying methylation across all RCGTG motifs, including those located outside of HIF1β binding peaks, the inverse correlation between DNA methylation and HIF binding was confirmed (Figure 1c). As BS-seq does not discriminate between 5-methyl (5mC) and 5-hydroxymethyl-cytosine17, we confirmed by DNA immunoprecipitation with an antibody recognizing only 5mC (5mC-DIP-seq) that HIF1β binding peaks were 6-fold depleted in 5mC-DIP-seq reads (Figure 1a). Moreover, methylation analysis of normoxic HIF1B-knockout MCF7 cells18 revealed identical methylation patterns (Figure S1e-g), indicating that the unmethylated state of HIF1β binding sites is not due to baseline activities of HIF1β under normoxia. Importantly, identical results were obtained for murine embryonic stem cells (mESCs): the loci corresponding to the 4,794 HIF1β binding sites identified in wild-type ESCs were unmethylated in normoxia, and this both in wild-type and Hif1b-knockout ESCs19 (Figure S1h-j). Since cells were intentionally exposed only briefly to hypoxia (16 hours), which fails to induce pronounced DNA methylation changes5, these data suggest that regions to which HIF1β binds upon hypoxia are devoid of DNA methylation under normoxic conditions.

Cell-type-specific DNA methylation of HREs determines HIF binding

Different cell types respond differently to hypoxia. To assess whether cell-type-specific DNA methylation could underlie this phenomenon, we profiled DNA methylation and HIF1β binding in 2 additional cell lines (RCC4 and SK-MEL-28). 20,613 HIF1β binding peak positions were detected across these cell lines (Supplementary Table 2). For each cell line, HIF1β binding was annotated as ‘present’ if the peak area showed >4-fold enrichment over the local read depth, and as ‘absent’ if it showed <2.5-fold enrichment; intermediate
enrichment was annotated as unclassified (Supplementary Table 2). When comparing cell lines using these criteria, HIF1β binding was shared by all 3 cell lines at 6,152 sites, and unique for an individual cell line at 7,140 sites (437, 1,193 and 5,510 unique sites, respectively for RCC4, MCF7 and SK-MEL-28) (Figure 1d, Figure S1k-l). Crucially, when assessing DNA methylation, HIF1β binding peaks unique to individual cell lines were unmethylated in cells where the binding site was active, while active HIF1β binding peaks shared between all cell lines were unmethylated in all cell lines (Figure 1e-f, Figure S1m-n). This strict correlation suggests that DNA methylation underlies the cell-type-specific response to hypoxia. Differences in DNA methylation and concomitant HIF binding also appeared functional, as transcriptome profiling under normoxic and hypoxic conditions revealed that genes with a flanking HIF1β binding peak unique to one cell line, were more frequently increased in expression under hypoxia in that cell line (Figure 1g).

**DNA methylation determines HIF binding independently of other chromatin marks**

To analyse whether other epigenetic modifications similarly correlate with HIF binding, we analysed public ENCODE data for MCF7 cells (no data are available for RCC4 and SK-MEL-28). Particularly, we investigated marks of heterochromatin (H3K9me3, H3K27me3), active promoters (H3K4me3, H3K9ac, H3K14ac), active enhancers (H3K4me1, H3K27ac), open chromatin (FAIRE) and active transcription (RNA Pol-II). Although some histone marks were enriched in a subset of HIF1β binding peaks, none were consistently found at all active HIF1β binding peaks, especially when looking outside of CpG islands (Figure S3a). The previously reported co-occupancy with RNA polymerase II or open chromatin was also not consistently found at all active HIF1β binding peaks. This was confirmed in linear regression analyses assessing how each mark individually predicts HIF1β binding in MCF7 cells. DNA methylation ($R^2=0.43$) outperformed all other marks, with marks of active chromatin such as RNA polymerase II occupancy, H3K4me3, open chromatin and H3K27ac showing poor correlations ($R^2$ resp. 0.11, 0.11, 0.04 and 0.04). When combining all marks in one model, the total $R^2$ was 0.47, with DNA methylation contributing to 67.5% of the predictive power (partial $R^2=0.32$). In line with this, omitting DNA methylation from the model reduced the total $R^2$ by more than half, to 0.21 (Figure S3b).
We also assessed whether more general differences in chromatin states (using ChromHMM\textsuperscript{21}) underlie differential HIF binding. This revealed that while shared HIF1β binding sites were more frequent in promoters, sites unique to MCF7 were more frequent in enhancers, and sites inactive in MCF7 (but unique to RCC4 or SK-MEL-28) more frequent in MCF7-repressed chromatin (Figure S3c). In line with enrichment at open chromatin, HIF1β binding thus appears exclusive to active enhancers and promoters while depleted in areas of repressed chromatin. Finally, NOMe-seq data from MCF7 cells revealed that, while open chromatin regions were generally unmethylated, a significant fraction of open chromatin (7-19\%) in fact showed methylation (Figure S3d), providing a potential rationale for the relatively small contribution of open chromatin to predict HIF1β binding. Combined, these data show that in normoxia poised HIF binding sites are located in unmethylated regions that consist mostly of active, open chromatin but are not consistently marked by other epigenetic modifications.

\textit{Other TFs determine the methylation landscape to guide HIF binding}

Interestingly, many of the HIF1β binding peaks overlapped with binding sites for other TFs. Specifically, out of the 7,153 HIF1β binding peaks detected in MCF7 cells, 5,903 overlapped with the binding site of at least one TF (83\%), out of a set of 11 TFs for which genome-wide binding site data were available in MCF7 cells\textsuperscript{22} (Figure 1h). This could indicate that these TFs, being already active under normoxic conditions, drive demethylation of HIF1β binding regions\textsuperscript{23}, thus setting the stage for HIF binding upon hypoxia. To further support this notion, we assessed whether these 11 TFs also bind at HIF1β binding peaks identified in RCC4 and SK-MEL-28 cells. Interestingly, TFs expressed by the 3 cell lines (e.g. CTCF or STAG1) co-localize in their binding with the shared HIF1β binding peaks. In contrast, TFs only expressed in MCF7 cells (e.g. ESR1 or GATA3) overlap in their binding sites only with MCF7-specific HIF1β binding peaks. Finally, binding of these 11 TFs in MCF7 did not overlap with HIF1β binding peaks unique to RCC4 or SK-MEL-28 (Figure 1i-j). These data were replicated in an independent cell line (Figure S3e-g).

Differential expression and binding of TFs between different cells is thus likely to shape the DNA methylation landscape and determine subsequent HIF binding.
DNA methylation does not determine differential binding of HIF1α and HIF2α

Comparison of our 7,153 HIF1β peaks to previously published HIF1α and HIF2α ChIP-seq data in MCF7 cells revealed the methylation status of HIF1β binding peaks to be independent of the HIFα binding partner (Figure S3h). Remarkably however, there were differences in the chromatin profiles of HIF1α- and HIF2α-bound regions: HIF1α binding sites showed 1.37-fold higher average levels of the promoter mark H3K4me3, whereas levels of the enhancer mark H3K4me1 were 0.75-fold lower at HIF1α binding sites than at HIF2α sites (Figure S3i). Similarly, chromHMM analysis showed enrichment of HIF1α at promoters and depletion at enhancers relative to HIF2α (Figure S3j). Moreover, other TFs similarly differed in occupancy between HIF1α- and HIF2α-specific sites: HIF2α was enriched at MCF7-specific TF binding sites, (which mostly correspond to cell-type-specific enhancers), and TFs shared between MCF7, RCC4 and SK-MEL-28 showed no enrichment of binding between HIF1α and HIF2α target sites (Figure S3f,k). In conclusion, while HIF1α preferentially binds at promoters and HIF2α at enhancers, DNA methylation differences do not determine their binding specificities.

DNA methylation directly repels HIF binding in cells

To more firmly establish a causal link between DNA methylation and HIF binding, we excluded several confounders. Firstly, since our chromatin state analysis revealed that HIF preferentially binds active enhancers and promoters, which are known to carry low levels of methylation, we performed HIF1β ChIP-bisulfite sequencing (HIF1β ChIP-BS-seq). MCF7 cells were exposed to hypoxia, HIF1β-bound DNA was immunoprecipitated and bisulfite-converted prior to sequencing to uncover its methylation pattern. This revealed that, while methylation levels of input DNA (not immunoprecipitated, bisulfite-converted DNA) were mostly low but with some sites displaying intermediate to high methylation levels, HIF1β-bound DNA was invariably very low in methylation and this at all sites (Figure 2a, Figure S4a).

Secondly, since TFs can drive demethylation of their binding sites both passively and actively, we excluded the possibility that DNA fragments bound by HIF would undergo DNA demethylation upon HIF binding. Indeed, HIF1β has previously been shown to
actively recruit DNA demethylases. However, HIF1β ChIP-BS-seq in hypoxic ESCs deficient for all DNA demethylases (Tet1, Tet2 and Tet3) showed results identical to those observed in wild-type MCF7 cells: HIF1β-bound DNA was unmethylated compared to input DNA subjected to whole-genome BS-seq (Figure 2b, Figure S4b).

Additionally, other (unknown) confounders related to the binding location of HIF, such as chromatin environment or sequence context, may contribute to preferential HIF binding to unmethylated DNA. To exclude this possibility, we generated isogenic murine ES cell lines in which a human HIF1β binding site-encoding DNA fragment was inserted that was either in vitro methylated or not (Figure 2c). Following recombination, the difference in methylation state between both fragments was maintained (Figure S4c-d). HIF1β ChIP-qPCR revealed that methylation was sufficient to induce a 12.4-fold reduction in HIF1β binding in these isogenic cell lines (Figure 2d).

Finally, to directly assess methylation sensitivity of HIF binding to unchromatinized DNA, we employed microscale thermophoresis, and tested the binding of recombinant co-purified HIF1α-HIF1β and HIF2α-HIF1β heterodimers to double-stranded DNA oligonucleotides containing a methylated or unmethylated RCGTG motif. Importantly, HIF1α- and HIF2α-containing heterodimers both showed a 15-fold higher affinity (K_D) for an unmethylated than methylated RCGTG motif, thus confirming that methylation directly repels binding of HIF1α-HIF1β and HIF2α-HIF1β heterodimers (Figure 2e-f).

Indeed, leveraging the crystal structure of the HIF1α-HIF1β and the HIF2α-HIF1β complexes bound to DNA, revealed that both cytosines in the CpG dinucleotide of the HIF binding sequence are snugly accommodated via van der Waals interactions with the guanidine groups of Arg102 in HIF1β and Arg27 in HIF1α or HIF2α, respectively (Figure 2g). Methylation of any of the two cytosines either on the top or bottom strand would in a static model drastically violate the minimum 3.1 Å length of van der Waals radii, and would be poised to cause severe steric clashes with these two functionally important arginine residues in HIF1α or HIF2α (Figure 2h).

**DNA demethylation enables ectopic HIF binding**
Next, we investigated which parts of the genome are protected from HIF binding by DNA methylation. For this, we compared HIF1β binding in hypoxic wild-type murine ESCs versus ESCs deficient for DNA methyltransferases (Dnmt-TKOs), which lack DNA methylation, using HIF1β ChIP-seq (n=4 replicates for each; for data quality assessment see Figure S5). This revealed a marked increase in the number of HIF1β binding peaks, from 7,875 in wild-type to 9,806 in Dnmt-TKO ESCs (Figure 3a). Whole-genome BS-seq further revealed that, while shared binding peaks were unmethylated in both cell lines, Dnmt-TKO-specific HIF1β binding peaks had high methylation levels in wild-type ESCs (Figure 3b).

All shared binding peaks were associated with a similar enrichment of the RCGTG motif (Figure 3c), as well as with genes that were induced upon hypoxia (Figure 3d). However, Dnmt-TKO-specific sites were more often distal to annotated transcription start sites (TSS) or regions of open chromatin, and more frequently in repressed chromatin regions of wild-type ESCs (Figure 3e-g). Gene ontology analysis moreover failed to identify enrichment of hypoxia-related processes for Dnmt-TKO-specific binding peaks, in contrast to shared peaks (Figure 3h). Thus, the majority of these Dnmt-TKO-specific binding peaks represents ectopic binding events.

**DNA methylation represses hypoxia-induced expression of retrotransposons**

Indeed, a substantial fraction of novel Dnmt-TKO-specific HIF1β binding peaks were found in repetitive genomic regions. Particularly, repeat class analysis revealed a 1.65-fold increase in binding peaks near retrotransposons (2,737 of 7,875 (34.8%) shared peaks versus 1,106 of 1,931 (57.3%) Dnmt-TKO-specific peaks; Figure S6a). Although HIF1β binding events were frequently observed at LINEs and SINEs, only binding at long terminal repeats (LTRs) was enriched over a randomisation of HIF1β binding site positions, and this both for all binding events and those distal to TSS (Figure 3i). The bulk of this increase was ascribable to binding at the 5′-end of endogenous retrovirus K (ERVK) LTR sequences (Figure 3j), with 344 of 1,106 (31%) novel repeat-binding peaks being at ERVKs versus only 3% of randomly shuffled HIF1β binding sites. These were mostly at solitary LTRs (Figure S6b-e). Given that ChIP-seq analyses rely on uniquely-mapping reads, which are
inherently depleted at repeat regions, this enrichment is likely to represent an underestimate.

We then assessed whether a similar phenomenon is at play in cancer cell lines, and pharmacologically demethylated MCF7 cells using 5-aza-2′-deoxycytidine (aza), overall reducing DNA methylation by 70.5±5.5% (Figure S6f). HIF1β ChIP-seq revealed that aza exposed 1,236 new HIF1β binding peaks. These were all methylated in untreated MCF7 cells and showed a 2.5-fold reduced methylation in aza-treated cells (Figure 4a-b). While HIF1β binding peaks in retrotransposons were already present in vehicle-treated MCF7 cells, novel aza-specific HIF1β binding peaks were 1.7-fold enriched for retrotransposons (9.7% versus 16.4%, respectively; Figure S6g). Again, these novel HIF1β binding peaks were often distal to TSSs, and binding at LTRs was enriched over a randomisation of HIF1β binding site positions (Figure 4c). Notably, different retrotransposons were affected in human MCF7 cells compared to murine ESCs due to the evolutionary divergent repeat content of these genomes. An analysis of the distribution of HIF1β binding peaks at retrotransposons, however, revealed that HIF1β binding sites were often at the 5’-end of retrotransposon sequences, and that patterns of binding were conserved between mouse and human genomes (Figure 3j versus Figure 4d), suggesting that HIF binding on retrotransposons is not random but functional, inducing their expression.

To confirm the latter, we applied RNA-seq to assess changes in retrotransposon expression after 24 hours of hypoxia with or without aza. Repeat expression was analysed using different bio-informatics pipelines. First, we used RepEnrich\textsuperscript{27}, which combines repeat-associated reads, also those that are non-uniquely mapping, to assess repeat expression for each of the 779 retrotransposon subfamilies annotated in the human genome (these are each members of one of the 25 families that constitute the LTR, LINE and SINE retrotransposon classes). We found that, already under hypoxia alone, 251 of all LTR (44%), 51 of all LINE (32%) and 5 of all SINE (10%) subfamilies were upregulated, while only 16 LTR, 7 LINE and no SINE subfamilies were downregulated (5% FDR; Figure S6h-i). Next, we used SQuiRE, which assigns reads (including non-uniquely-mapping reads) to a specific repeat locus based on an expectation-maximization algorithm\textsuperscript{28}. With
SQuIRE, 72% (n=2,781) of all differentially expressed repeat loci exhibited increased expression under hypoxia (P<10^{-16}; Figure S6j).

Induction of cryptic transcripts by hypoxia

Hypoxia-induced transcripts were, however, often not matching the annotated repeat locus, but extending well beyond their annotated end, with some transcripts encompassing multiple repeat elements. Also, we noticed that for many transcripts, HIF1β binding did not occur in the retrotransposon promoter, while some other transcripts did even not contain a retrotransposon-associated sequence. Similar transcripts were also induced by aza. We therefore refer to these as ‘cryptic transcripts’ (Figure S7a). To more accurately quantify them, we developed a novel analysis pipeline, CREDENToR. CREDENToR first performs a de novo transcriptome assembly to define cryptic transcripts and then assigns uniquely-mapping reads to them to quantify their expression. The cryptic transcripts detected by CREDENToR are poorly conserved, often unspliced transcripts, shorter than lincRNAs but expressed at similar levels (see Methods and Figure S7b-g for benchmarking).

CREDENToR identified that out of 1,389 differentially expressed cryptic transcripts (1% FDR), 67% were upregulated by hypoxia (Figure S6k). As expected, focussing on HIF-bound cryptic transcripts revealed an even stronger enrichment, with 82% and 91% (respectively, at 1% and 0.001% FDR) differentially expressed transcripts being upregulated following hypoxia (Figure 4e). HIF binding was enriched at the promoter of hypoxia-induced cryptic transcripts, but far less in those induced by aza (Figure S6l). Interestingly, significant fractions of cryptic transcripts contained palindromic repeats, or overlapped with other transcripts in the reverse orientation, and could thus produce double-stranded (ds) RNA. HIF-bound cryptic transcripts were twice as likely to generate such dsRNAs (Figure 4f). Together, this suggests HIF binding to leverage cryptic TSS structures within and outside the repeat genome to express dsRNA-generating cryptic transcripts.
Cryptic transcript expression was indeed dependent on HIF, as non-HIF-bound cryptic transcripts failed to show induction following hypoxia (Figure 4g). To confirm this, we assessed expression in HIF1B-knockout MCF7 cells. Here, hypoxia failed to upregulate cryptic transcripts, according to both CREDENToR and RepEnrich (Figure 4g, Figure S6m). As expected, aza-induced overexpression was retained, while hypoxia in HIF1B-knockout MCF7 cells failed to increase the effect of aza. Pharmacological activation of HIF using dimethyloxalylglycine (DMOG), a broad-spectrum inhibitor of 2-oxoglutarate-dependent hydroxylases, affected cryptic transcripts similar to hypoxia (Figure S6m-o). Combined, these data indicate that hypoxia triggers HIF binding to unmethylated repeat regions, inducing HIF-dependent expression of cryptic transcripts, most of which are associated with retrotransposons.

**Hypoxia and repeat transcript expression affect tumour immunotolerance**

Expression of repeat transcripts has been linked to tumour foreignness, interferon (IFN) response and enhanced cytolytic activity, all critical determinants of response to checkpoint immunotherapy. Similar to our own data, such transcripts were shown to increase dsRNA formation. This triggers IFN responses through viral mimicry. Cryptic transcripts induced by HIF could thus contribute to an immune-activated microenvironment.

To study this in more detail, we reanalysed expression and DNA methylation data from The Cancer Genome Atlas (TCGA). We classified 5,193 tumours from 14 tumour types as hypoxic or normoxic using an established hypoxia metagene expression signature. We remapped all RNA-seq reads to determine expression of retrotransposon subfamilies using RepEnrich, and also performed de novo transcript assembly to identify on average 11,654 non-overlapping cryptic transcripts per tumour type using CREDENToR (Figure S8a). While TCGA tumours were not exposed to DNA demethylating agents, they did show variation in DNA methylation at TSS of cryptic transcripts. Indeed, although CpGs cryptic transcript promoters showed mostly high methylation levels (median = 80.7%), there was considerable variability (9.2% standard deviation), and one in 10 tumours displayed median levels below 67.3%. Remarkably, and in line with our in vitro data, there was a
significant interaction between hypoxia and DNA methylation in determining cryptic transcript expression (P=0.0109), with expression being increased in hypoxic tumours having lower methylation at cryptic transcripts (Figure 5a). At least 1,279 cryptic transcripts showed increased expression of 10-fold or higher (FDR<0.01, Figure S8b). A reanalysis of combined single-cell methylome-and-transcriptome sequencing data from colorectal cancer cells\(^{36}\) moreover confirmed that cryptic transcript expression and promoter methylation are inversely correlated, and this more strongly in hypoxic than normoxic cancer cells (P=0.032), suggesting that the observed interactions are cancer cell-intrinsic (Supplementary Table 3).

In TCGA, this interaction was only detected in tumour types known to respond to immunotherapy\(^{37}\) (see Methods; P=0.0031 in responsive versus P=0.69 in non-responsive tumours; Figure 5b-c). As expected, responsive tumour types exhibited an increased tumour mutation burden (TMB), elevated immune checkpoint expression, more CD8\(^+\) T-cells and increased cytolytic activity (Figure S8c)\(^{38}\). Importantly, responsive types also had on average lower methylation at cryptic transcripts and higher cryptic transcript expression than non-responsive types (P<10\(^{-16}\) for both comparisons, Figure 5d). Single-cell RNA-seq analyses (both from 5’ and 3’ end) highlighted that cancer cells show the highest level of cryptic transcript expression compared to stromal cells, indicating they represent the main source of cryptic transcripts expression (Figure S8d). In line with our \textit{in vitro} findings, DNA hypomethylation thus underlies cryptic transcript expression in hypoxic tumours, an effect that was particularly striking in immunotherapy-responsive tumours.

Overall, these observations support a model wherein hypoxia-induced cryptic transcripts are tolerated in high-immunogenic tumours, as these are characterized by high immune checkpoint expression, but not in low-immunogenic tumours where their expression would compromise tumour immunotolerance. This suggests that low-immunogenic tumours may need to maintain high DNA methylation levels in cryptic transcripts to downregulate their expression and avoid the induction of tumour immunogenicity.

\textit{Aza compromises tumour immunotolerance in mice via HIF}
To confirm that in low-immunogenic tumours DNA methylation prohibits cryptic transcript expression, we identified 59 such retrotransposons that correlate in expression with cytolytic activity in immunotherapy-responsive tumour types within TCGA. Remarkably, all of these were upregulated in vitro, by hypoxia alone or hypoxia in combination with aza (P<0.05, Figure 5e), suggesting that hypoxia and DNA demethylation can indeed enhance tumour immunogenicity. To confirm this experimentally, we screened several mouse tumour models for their immunogenicity. The orthotopic 4T1 breast cancer model was identified as low-immunogenic. Indeed, 4T1 tumours exhibited a low TMB, cytolytic activity, number of CD8+ T-cells and expression of immune checkpoints (Pd1, Pdl1) compared to other models (Figure S9a). In line with 4T1 grafts being low-immunogenic tumours, anti-PD1 treatment failed to affect their growth (-8%, P=0.397), while significantly reducing growth of high-immunogenic tumours, as described previously39, 40 (Figure S9b). Importantly, also the expression of cryptic transcripts was lower in 4T1 than in high-immunogenic tumour models (Figure S9c).

Next, we verified in low-immunogenic 4T1 cells whether DNA demethylation upregulates cryptic transcripts in a HIF-dependent manner. In vitro, we observed that, similar to MCF7 cells, both hypoxia and aza independently increased cryptic transcript expression, both using CREDENToR and RepEnrich (Figure 6a; Figure S9d). Likewise, aza increased cryptic transcript expression in vivo (Figure 6b). To confirm that this upregulation was at least partially hypoxia-mediated, we investigated whether tumour hypoxia enhances aza-induced cryptic transcript expression. We compared aza-treated 4T1 tumour-bearing mice injected either with control or anti-VEGFR-2 antibody (DC101). While vehicle-treated 4T1 tumours were hypoxic in ~40% of the tumour, DC101 further reduced blood vessel density (-35%; P<0.05) and increased hypoxic tumour areas (68%; P<0.05; Figure 6c-d). Importantly, this was associated with an increase in cryptic transcripts (+ 9%; P=2.6x10^-16; Figure 6e).

We then explored whether this increase also compromised immunotolerance. As immunogenicity of cryptic transcripts is mediated via dsRNA formation, we first confirmed in vitro by immunostaining the increase in dsRNA after both hypoxia and aza
in 4T1 cells (Figure 6f). *In vivo*, aza reduced growth of 4T1 tumours (-32%; P=3.0x10^{-3}; Figure 6g), but did not reduce cell proliferation marker expression (Figure S9e). In contrast, immune activation was enhanced in tumours treated with aza, as activated T-cell and natural killer cell signatures were upregulated and myeloid-derived suppressor cell signatures downregulated (Figure 6h). Immunofluorescence of CD8+ T-cells confirmed these changes: while T-cell infiltration was unaffected, the number of activated, granzyme B-positive T-cells increased 2.1-fold (P<0.05; Figure S9f).

To verify HIF-dependence of these immunogenic effects, we generated polyclonal 4T1 cells deficient for HIF1β by CRISPR-Cas9 (4T1^{Hif1b-KO}; Figure S9g) and compared these cells to scramble-control 4T1 cells (4T1^{Hif1b-scr}), while treating with aza or vehicle. *In vitro*, loss of HIF1β abrogated hypoxia-induced dsRNA formation and HIF1β-bound cryptic transcript expression both in vehicle and aza-treated cells (Figure S9d and h), similar to what we observed in MCF7 cells. Also *in vivo*, 4T1^{Hif1b-KO} showed reduced cryptic transcript expression compared to 4T1^{Hif1b-scr} grafts, effects that were limited to HIF-bound cryptic transcripts as expected (Figure 6i; Figure S9i). Of note, 4T1^{Hif1b-KO} tumours grew much slower than 4T1^{Hif1b-scr} tumours, presumably because HIF1β also has direct effects on cell proliferation, thus rendering it challenging to disentangle effects on immunogenicity. Nevertheless, aza induced a similar and significant upregulation of cancer testis antigen expression in both cell lines (Figure S2), suggesting similar treatment efficacy. Interestingly, while 4T1^{Hif1b-scr} grafts also showed a significantly reduced size when comparing aza to vehicle (46% reduction; P=3.3x10^{-6}), 4T1^{Hif1b-KO} failed to show as strong a reduction (only 22%; P=7.0x10^{-4}, or 1.8-fold less than the scramble effect; P=0.021; Figure 6j). The differential effect of aza in 4T1^{Hif1b-KO} versus 4T1^{Hif1b-scr} grafts was also highly significant in an interaction analysis (P<0.0001). Moreover, while the number of activated T-cells increased in 4T1^{Hif1b-scr} grafts following aza, 4T1^{Hif1b-KO} grafts failed to show such increase (Figure 6k). This differential effect of aza, depending on the Hif1b background, was similarly significant in an interaction analysis (P=6.3x10^{-3}). Together, these data provide a mechanistic link between HIF1β binding, DNA methylation and immune
activation, highlighting the potential of DNA methylation inhibitors to activate the immune system and render immune-cold tumours immune-hot.

**DISCUSSION**

Here, we show that DNA methylation directly repels binding of HIF transcription factors and that cell-type-specific DNA methylation patterns established under normoxic conditions underlie the differential hypoxic response between cell types. Furthermore, ectopic HIF binding sites in repeat elements are normally masked by DNA methylation but become accessible to HIF upon DNA demethylation, leading to expression of cryptic transcripts which enhance tumour immunogenicity.

Our findings are important for a number of reasons. Firstly, an instructive role of DNA methylation in gene expression regulation, as originally proposed by Holliday and Pugh and by Riggs\(^{41, 42}\), has remained controversial. Indeed, in many instances it is unclear whether DNA methylation changes are a direct or indirect cause, or rather a consequence of TF binding or gene expression\(^{2}\). Our findings in murine and human both differentiated and undifferentiated cells align well with a recent study showing methylation-dependence of NRF1 binding in mESCs. By demonstrating that DNA methylation directly repels HIF binding, we thus highlight the importance of DNA methylation profiling, especially in poorly oxygenated tissues. Since tumour hypoxia has long been associated with increased malignancy, poor prognosis and resistance to radio- and chemotherapy\(^{6}\), DNA methylation could especially provide insights in the processes underlying therapeutic resistance. For instance, Vanharanta and colleagues recently showed an association between DNA methylation near *CYTIP* and the survival of disseminating cancer cells\(^{43}\). Combined with our observations that DNA methylation directly repels HIF binding, this suggests remethylation of the *CYTIP* promoter as a viable avenue for decreasing cancer dissemination.

Secondly, it has been challenging to identify a guiding principle as to why specific genes are induced upon hypoxia in one, but not the other cell type\(^{9}\). Our findings suggest that cell-type-specific TF binding under normoxia causes differences in DNA methylation,
which subsequently determine HIF binding under hypoxia and predict the cell-type-specific hypoxia response. Importantly, we also confirmed earlier observations that HIF1β binding peaks are characterized by an active, open chromatin structure\textsuperscript{11}. This additional requirement for functional HIF1β binding peaks probably explains why each of the RCGTG consensus sequences in the genome cannot serve as an equal HIF binding substrate in normal cells, or upon genetic or pharmacological demethylation. Similar observations were made for other TFs, such as CTCF, for which binding was similarly limited to sites containing a permissive chromatin structure\textsuperscript{14, 23}. Importantly, binding specificities for HIF1α versus HIF2α are independent of DNA methylation, but appear to be influenced by chromatin context. This is in line with the identical structure of DNA binding domains of HIF1α and HIF2α; swapping DNA binding domains between both proteins has no influence on their binding profile\textsuperscript{44}. Instead, the transactivation domain appears to endow specificity, suggesting that accessory chromatin binding partners govern the differential binding of HIF1α and HIF2α\textsuperscript{44}.

Thirdly, several publications by now reported how 5-aza-2′-deoxycytidine initiates cryptic TSSs in the repeat genome, leading to expression of cryptic transcripts\textsuperscript{31, 45}. Our data add to these findings by demonstrating that cryptic transcript expression is at least partly HIF-dependent, while more importantly, hypoxia alone is also capable of inducing their expression. Based on single-cell analyses, we observed this effect to be cancer cell-autonomous, consistent with cancer cells being hypomethylated. Our findings reinforce a growing body of evidence that highlights how during evolution transposable elements have copied and amplified regulatory regions throughout the genome\textsuperscript{22, 46-50}. Most likely, transposable elements hijacked the transcriptional apparatus of their host to support their germline propagation\textsuperscript{51}. In doing so, they copied the associated TF binding site and seeded it at the site of insertion. Transposable elements having binding sites for TFs that are active in the germline, are more likely to hijack these and transpose. Accordingly, HIF is activated in early development, when DNA methylation levels are also low\textsuperscript{50, 52}; ancestral cooption of HIF binding sites by cryptic transcripts to increase their expression is thus plausible. In line with specific TFs preferentially acting on particular
retrotransposon subfamilies, we observe enrichment of HIF binding and activation at LTRs, particularly at the LTR of ERVKs.

Finally, we uncover an intriguing opportunity for cancer immunotherapy. Chiapinelli et al. already demonstrated that aza-induced cryptic transcripts are highly immunogenic and can sensitize tumours to checkpoint immunotherapy\textsuperscript{31}, while Sheng et al. showed that also histone demethylase LSD1-ablation increases cryptic transcripts, thereby enabling checkpoint blockade\textsuperscript{53}. The mechanism underlying immunogenicity likely depends on the formation of dsRNA, which via a viral mimicry-mediated process activates the immune system\textsuperscript{33, 45, 53}. In addition, some of these transcripts contain open-reading frames, which could translate into abnormal proteins that can be antigenic\textsuperscript{45}. Importantly, hypoxia is endemic to most solid tumours, and therefore could have a more widespread impact than aza. Indeed, in hypoxic tumours with high checkpoint expression, DNA methylation at TSS of cryptic transcripts was reduced and consequently, cryptic transcript expression increased. Since tumours with high checkpoint expression often respond to checkpoint immunotherapy, and as cryptic transcripts could sensitize tumours to checkpoint blockade\textsuperscript{31, 33}, this suggests hypoxia-induced cryptic transcripts to play an important role in mediating the therapeutic effects exerted by checkpoint inhibitors. In contrast, immune-cold tumours characterized by low immune checkpoint expression were much less tolerant to cryptic transcript expression, showing high methylation at retrotransposon promoters. In light of our findings that methylation directly repels HIF binding, this suggests DNA methylation to block hypoxia-induced cryptic transcript expression in immune-cold tumours to maintain immunotolerance. Pharmacological demethylation of immune-cold 4T1 tumours indeed increased cryptic transcription, enhanced immunogenicity and reduced tumour growth in a HIF-dependent manner. By showing that low-immunogenic, hypoxic tumours can be rendered immunogenic through DNA methylation inhibitors, we thus highlight a novel treatment strategy for tumours otherwise refractory to immunotherapies.
METHODS

Materials

All materials were molecular biology grade. Unless noted otherwise, all were from Sigma (Diegem, Belgium).

Cell lines

MCF7, RCC4, SK-MEL-28, A549, 4T1, MC38 and CT26 cell lines were obtained from the American Type Culture Collection and their identity was not further authenticated. None of these cell lines are listed in the database of commonly misidentified cell lines maintained by ICLAC. MCF7 HIF1B-knockout cells were previously described. MCF7, RCC4, A549, MC38 and 4T1 cells were cultured at 37 °C in Dulbecco’s modified Eagle medium (DMEM) with 10% fetal bovine serum (FBS), 5 mL of 100 U/mL Penicillin-Streptomycin (Pen-Strep, Life Technologies) and 5 mL of L-Glutamine 200 mM. SK-MEL-28 and CT26 cell lines were cultured at 37 °C in Roswell Park Memorial Institute 1640 Medium (RPMI) with 10% FBS 1% Pen-Strep and 1% L-Glutamine.

Murine embryonic stem cells (mESCs) that were triple-knockout for Dnmt1, Dnmt3a and Dnmt3b (Dnmt-TKO), triple-knockout for Tet1, Tet2 and Tet3 (Tet-TKO) and their appropriate wild-type (WT) control mESCs were obtained from Dr. Masaki Okano and Dr. Guoliang Xu respectively. mESCs that were knockout for Hif1b (Hif1b-KO) and their WT control mESCs were previously described. Dnmt-TKO, Tet-TKO, Tet-WT, Hif1b-WT and Hif1b-KO mESCs were cultured feeder-free in fibroblast-conditioned medium (DMEM with 4,500 mg/L glucose, 2 mM L-glutamine, 1 mM sodium pyruvate, 15% FBS, 1% Pen-Strep, 0.1 mM of non-essential amino acids, 0.1 mM β-mercaptoethanol) on 0.1% gelatine-coated plates. mESCs from the 159 background used for the recombinase-mediated cassette exchange reaction were kindly provided by Prof. Dirk Schubeler (Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland) and grown in ESC medium (DMEM with 4,500 mg/L glucose, 2 mM L-glutamine, 1 mM sodium pyruvate, 15% FBS, 1% Pen-Strep, 0.1 mM of non-essential amino acids, 0.1 mM β-mercaptoethanol, 10³ U LIF ESGRO (Millipore)) containing 25 μg/mL hygromycin (50μl of...
5mg/mL stock per 10 mL medium) for at least 10 days. 4T1 cells that were knockout for Hif1b (Hif1b-KO) and their WT control cells were cultured at 37 °C in DMEM with 10% FBS, 5 mL of 100 U/mL Pen-Strep, 10 µg/mL of blasticidin (ant-bl-05, Invivogen), puromycin (P9620, Sigma-Aldrich) 1.5 µg/mL medium and 5 mL of L-Glutamine 200 mM.

All cell cultures were confirmed to be mycoplasma-free every month.

Cell line treatment conditions

Cell cultures were grown under atmospheric (21%) oxygen concentrations in the presence of 5% CO₂, or rendered hypoxic by incubating them under 0.5% oxygen (5% CO₂ and 94.5% N₂). For ChIP-seq experiments, hypoxia was induced during 16 hours, whereas 24 hours of exposure were used when assessing effects of hypoxia on gene or protein expression level. Where indicated, cells were pre-treated with 5-aza-2’-deoxycytidine (aza, 1 µM) for 3 days by adding the required volume to fresh culture medium. Equal volumes of the carrier (DMSO) were used as control. This was followed by another day of exposure to aza in hypoxia or normoxia, bringing the total aza exposure time for experiments to 4 days.

2mM of DMOG (dimethylxalylglycine, Sigma) was added to culture medium for 24 hours where indicated. Cells were always plated at a density tailored to reach 80-95% confluence at the end of the treatment. Fresh medium was added to the cells just prior to hypoxia. To prove that the extent to which cells were exposed to hypoxia was similar across experiments, we assessed that induction of hypoxia marker genes (BNIP3, EGLN, ALDOA, CA9) but not HIF1A occurred in each experiment (Figure S2). For experiments involving exposure to aza, we assessed the expression of cancer testis antigens as a positive control (Figure S2).

LC-ESI-MS/MS of DNA to measure 5mC

DNA was extracted and processed for LC-ESI-MS/MS to determine 5mC concentrations exactly as described previously¹.

Western blot

To assess HIF1α protein stabilization, proteins were extracted from cultured cells as follows: cells were placed on ice, washed twice with ice-cold PBS and lysed in protein
DNA methylation repels HIF binding  

D’Anna et al

07/02/2020  
Page 23 of 64

extraction buffer (50 mM Tris HCl, 150 mM NaCl, 1% Triton X-100, 0.5% Na-deoxycholate, 0.1% SDS and 1× protease inhibitor cocktail (Roche)). Protein concentrations were determined using a bicinchoninic acid protein assay (BCA, Thermo Scientific) following the manufacture’s protocol. An estimated 60 µg of protein was loaded per well on a NuPAGE Novex 3-8 % Tris-Acetate Protein gel (Life Technologies), separated by electrophoresis and blotted on polyvinylidene fluoride membranes. Membranes were activated with methanol and washed with Tris-buffered saline (TBS; 50 mM Tris HCl, 150 mM NaCl) with 0.1% Tween 20, and incubated with rabbit α-tubulin (2144S, Cell Signaling), rabbit β-actin (4967, Cell Signaling), rabbit HIF-1β/ARNT (D28F3) XP® (5537, Cell Signaling) at 1:1,000 dilution and rabbit HIF-1α (C-Term) Polyclonal Antibody (Cayman Chemical Item 10006421) 1:3,000. Incubation with the secondary antibodies and detection were performed according to routine laboratory practices. Western blotting was done on 3 independent biological replicates.

Analysis of HIF1β target genes using ChIP-seq

20-25x10⁶ cells were incubated in hypoxic conditions for 16 hours. Cultured cells were subsequently immediately fixed by adding 1% Formaldehyde (16% Formaldehyde (w/v), Methanol-free, Thermo Scientific) directly in the medium and incubating for 8 min on a flat-bed shaker at room temperature (RT). Fixed cells were incubated with 150 mM of glycine for 5 min to revert the cross-links, washed twice with ice-cold PBS 0.5% Triton-X100, scraped and collected by centrifugation (1000 xG, 5 min at 4 °C). The pellet was resuspended in 1,400 µL of RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA pH 8, 1% Triton-X100, 0.5% Sodium deoxycholate, 1% SDS, 1% protease inhibitor) and transferred to a new Eppendorf tube. The lysate was homogenized by passing through an insulin syringe, and incubated on ice for 10 min. The chromatin was sonicated for 3 min by using a Branson 250 Digital Sonifier with 0.7 s 'On' and 1.3 s 'Off' pulses at 40% power amplitude, yielding predominantly fragment sizes between 100 and 500 bps. The sample was kept ice-cold at all times during the sonication. Next, samples were centrifuged (10 min at 16,000 × G at 4 °C) and supernatant transferred in a new Eppendorf tube. Protein concentration was assessed using a BCA. 50 µL of sheared chromatin was used as “input”
and 1.4 μg of primary ARNT/HIF1β monoclonal antibody (NB100C124, Novus) per 1 mg of protein was added to the remainder of the chromatin and incubated overnight at 4 °C in a rotator. Next, Pierce Protein A/G Magnetic Beads (Life Technologies) were added to the samples in a volume that is 4× the volume of the primary antibody and incubated at 4 °C for at least 5 hours. A/G Magnetic Beads were collected and washed 5 times with washing buffer (50 mM Tris-HCl, 200 mM LiCl, 2 mM EDTA, pH 8, 1% Triton, 0.5% Sodium deoxycholate, 0.1% SDS, 1% protease inhibitor), and twice with TE buffer. The A/G magnetic beads were resuspended in 50 μL of TE buffer, and 1.5 μL of RNase A (200 units, NEB, Ipswich, MA, USA) was added to the A/G beads samples and to the input, incubated for 30 min at 37 °C. After addition of 1.5 μL of Proteinase K (200 units, NEB) and overnight incubation at 65 °C on a stirrer, the beads were removed from the solution using a magnet and DNA was purified using 1.8× volume of Agencourt AMPure XP (Beckman Coulter) according to the manufacturer’s instructions. DNA was eluted in 20 μL of TE buffer. The input DNA was quantified on NanoDrop. Next, 1 μg of the input and all the immunoprecipitated DNA was converted into sequencing libraries using the NEBNext DNA library prep master mix set (NEB) following manufacturer’s instructions.

A single end of these libraries was sequenced for 50 bases on a HiSeq, either HiSeq2500 or HiSeq4000 (Illumina), mapped using Bowtie and extended for the average insert size (250 bases). ChIP peaks were called by Model-based Analysis for ChIP-Seq (MACS)\(^{15}\), with standard settings and using read counts from an input sample as baseline.

HIF1β binding peak positions in the human cell lines MCF7 (both vehicle- and aza-treated), RCC4, A549 and SK-MEL-28 were defined by using the stringent threshold P<10\(^{-15}\). A threshold equal to P<10\(^{-10}\) was used to define HIF1β binding peaks in murine cell lines (4T1, Dnmt-WT and Dnmt-TKO ESCs).

To compare HIF1β binding peaks between human cell lines (MCF7, RCC4, A549 and SK-MEL-28), HIF1β binding peaks were called as present if the average coverage at the 200 bps centered on the summit was >4-fold bigger than the local background, and as absent if it was <2.5-fold smaller than the local background, with local background being defined...
as the read depth across regions 1.5-5 kb up- and downstream of the peak. Intermediate coverage was annotated as unclassified. To compare HIF1β binding peaks between murine Dnmt-WT and -TKO ESCs, the HIF1β binding peak was called as present if the average coverage at the 200 bps centered on the summit was >4-fold bigger than the background, and as absent if it was <4-fold smaller than the background. To compare efficiency between experiments, scatter plots of reads counts at peak regions of HIF1β binding regions were generated per cell line in a pairwise fashion.

**Annotation of genomic features**

Human sequences were mapped to genome build hg19 and murine sequences to genome build mm10. Putative HIF binding sites were detected genome-wide by screening the whole genome for RCGTG motifs using the regular expression search tool dreg (www.bioinformatics.nl/cgi-bin/emboss/help/dreg). The frequency per bp of RCGTG motifs inside HIF1β binding peaks and in the rest of the genome was calculated, and enrichment of RCGTG motifs at HIF1β binding peaks quantified by overlapping RCGTG positions in the genome with the HIF1β binding peak positions as defined by MACS.

The distances of HIF1β peaks to the nearest RCGTG motif (cumulative frequency), TSS and open chromatin (frequency) were calculated by overlapping each genomic feature with HIF1β peak positions using BedTools in R. Protein-coding genes were annotated as per Ensembl version 92. Promoter regions were annotated as being 2 kb upstream or 500 bp downstream of the start site of each gene. Chromatin state annotation of MCF7 and mESCs was as described21,56. HIF1β binding peaks were annotated with these features and overlapped with the repeat genome using BedTools. To assess enrichment of HIF binding at repeats, HIF1β binding peaks were 10,000 times either randomly distributed throughout the genome, or randomly distributed while matching the distal binding peak distribution. Next, the frequency of repeat binding in a random distribution was compared to that in the observed distribution. Peaks randomly assigned to poorly mapping regions were discarded.

**Genome distribution of 5mC: BS-seq, SeqCapEpi BS-seq and mDIP-seq**
BS-seq, SeqCapEpi BS-seq and mDIP-seq were performed as described previously\(^1\). To quantify DNA methylation inside HIF1\(\beta\) binding peaks, SeqCapEpi probes with >40\(\times\) coverage were overlapped with HIF1\(\beta\) binding peaks as defined by MACS. Methylation levels at the probes overlapping and non-overlapping (rest of the genome) HIF1\(\beta\) binding peaks were calculated using Seqmonk.

ChIP-BS-seq was done as ChIP-seq, except that methylated adaptors (NEB) were ligated, and DNA libraries were bisulfite-converted using the EZ DNA Methylation-Lightning™ kit (Zymo) prior to library amplification using HiFi Uracil+ (KAPA). Reads were mapped using Bismark as described\(^1\).

**RNA-seq**

To assess the impact of HIF binding at gene promoters on their expression, strand-specific RNA-seq was performed in human cell lines and murine Dnmt-WT and Dnmt-TKO ESCs. Briefly, total RNA was extracted using TRIzol (Invitrogen), and remaining DNA contaminants in 17-20 \(\mu\)g of RNA were removed using Turbo DNase (Ambion) according to the manufacturer’s instructions. RNA was repurified using the RNeasy Mini Kit (Qiagen). For total RNA-seq, ribosomal RNA present was depleted from 5 \(\mu\)g of total RNA using the RiboMinus Eukaryote System (Life technologies). cDNA synthesis was performed using the SuperScript® III Reverse Transcriptase kit (Invitrogen). 3 \(\mu\)g of random Primers (Invitrogen), 8 \(\mu\)L of 5\(\times\) First-Strand Buffer and 10 \(\mu\)L of RNA mix were incubated at 94 °C for 3 min and then at 4 °C for 1 min. Next, 2 \(\mu\)L of 10 mM dNTP Mix (Invitrogen), 4 \(\mu\)L of 0.1 M DTT, 2 \(\mu\)L of SUPERase• In™ RNase Inhibitor 20U/ \(\mu\)L (Ambion), 2 \(\mu\)L of SuperScript™ III RT (200 units/\(\mu\)L) and 8 \(\mu\)L of Actinomycin D (1 \(\mu\)g/\(\mu\)L) were added and the mix was incubated 5 min at 25 °C, 60 min at 50 °C and 15 min at 70 °C to heat-inactivate the reaction. The cDNA was purified using 80 \(\mu\)L (2\(\times\) volume) of Agencourt AMPure XP and eluted in 50 \(\mu\)L of the following mix: 5 \(\mu\)L of 10\(\times\) NEBuffer 2, 1.5 \(\mu\)L of 10 mM dNTP mix (10 mM dATP, dCTP, dGTP, dUTP, Sigma), 0.1 \(\mu\)L of RNaseH (10 U/\(\mu\)L, Ambion), 2.5 \(\mu\)L of DNA Polymerase I Klenov (10 U/\(\mu\)L, NEB) and water until 50 \(\mu\)L. The eluted cDNA was incubated for 30 min at 16 °C, purified by Agencourt AMPure XP and eluted in 30 \(\mu\)L of dA-Tailing mix (2 \(\mu\)L of Klenow Fragment, 3 \(\mu\)L of 10\(\times\) NEBNext dA-Tailing mix).
Reaction Buffer and 25 µL of water). After 30 min incubation at 37 °C, the DNA was purified by Agencourt AMPure XP, eluted in TE buffer and quantified on NanoDrop. Subsequent library preparation was done using the DNA library prep master mix set and sequencing was performed as described for ChIP-seq.

mRNA capture and stranded library preparation of RNA from MCF7 cells, mouse cell lines and tumours for the purpose of retrotransposon and cryptic transcript expression analysis was performed using the KAPA Stranded mRNA-Seq Kit according to the provided protocol (Illumina). For expression analysis of coding genes, RNA-seq reads were mapped to the human or murine genome reference (hg19 or mm10) using Tophat2. Gene read numbers were counted using HTSeq and normalized to the sum of the mapped expression counts. Gene expression was presented as transcript per million (TPM), 0.01 offset. Differential gene expression was tested using edgeR.

Expression of cancer testis antigens was annotated according to all entries listed in the CTDatabase (www.cta.incc.br/modelo.php). Cytolytic activity was quantified as the log average (geometric mean in TPM) of the RNA expression of 2 key cytolytic enzymes: granzyme A (GZMA) and perforin 1 (PRF1).

**RepEnrich analyses**

RNA-seq data were expressed in TPM with an offset of 0.01. Expression read counts of retrotransposons are calculated using the RepEnrich tool (https://github.com/nerettilab/RepEnrich) and normalized to the total mappable read depth. The repeat genome of the human reference genome hg19 was downloaded from the RepEnrich website. Human retrotransposon classes (LINE, SINE, LTR) contain 16 families and 779 subfamilies. The repeat genome of the mouse genome mm10 was built using the RepeatMasker track from the UCSC genome browser. Mouse retrotransposon classes (LINE, SINE, LTR) contain 24 families and 906 subfamilies.

**CREDENToR analysis**

The overall strategy of CREDENToR is to perform *de novo* assembly of all reads and based on this define all cryptic transcripts. CREDENToR will consider transcripts encompassing
more than one repeat element as one cryptic transcript and quantify gene expression for each of them. To achieve this, fastq files of RNA-seq data were first aligned to the human (build GRCh38) or the murine genome (build mm10) using STAR (version 2.5) with a tolerance of two mismatches. Transcriptomes were subsequently assembled using StringTie\(^57\) (version 1.3.4d), under guidance of the transcript annotation tool Ensembl 92. All de novo assembled transcription annotations from the same set of tumour samples (i.e., MCF7 or 4T1 cell lines, or each of the 14 tumour types downloaded from TCGA) were merged using “StringTie --merge”. HTSeq-counts\(^58\) (version 0.11.2) were used to count the read numbers of known and novel genes. Non-coding transcripts (transcripts not overlapping annotated coding genes) in the merged transcription annotations were assigned as cryptic transcripts when any of their exons overlapped with a retrotransposon repeat annotation (LTR, LINE, or SINE, based on RepeatMasker annotation from UCSC). If a transcript overlapped with >1 annotated repeat, the retrotransposon with the highest overlap was assigned to this cryptic transcript.

For the analysis of MCF7 data, the assembled annotations from all experimental conditions involving MCF7 cells assessed in vitro were merged before read counting. For the analysis of 4T1 data, the assembled annotations from in vitro and in vivo samples were merged together. Cryptic transcripts were considered to be HIF-associated if a HIF binding summit was detected within the transcript promoter (i.e. 2,000 bp upstream and 500 bp downstream of the transcription start site). Per set of experiments (24 samples), we further required that the read number per cryptic transcript exceeds 10 in at least 1 sample, and that the reads per kb per million reads (RPKM) exceeds 1 in at least one sample. For these cryptic transcripts, DESeq (version 1.30.0) was used to test the differences between each pair of conditions. For the TCGA data, we merged assembled transcription annotations for each tumour type separately. Cryptic transcripts were calculated using the total cryptic transcript read count divided by the total coding gene read count. In volcano plots, individual cryptic transcripts were plotted, but in violin plots, where we compare effects to the cryptic transcripts obtained by RepEnrich, we summed cryptic transcript counts into retrotransposon subfamilies, log-transferred counts-per-
million (normalized to total read counts) and considered those as expression values. Violin plots invariably show >95% of data points. P values were corrected for multiple testing following Benjamini and Hochberg correction. The CREDENToR pipeline has been made available on GitHub (https://github.com/Jieyi-DiLaKUeLeuven/CREDENToR).

**Gene Ontology Analysis**

Genes were associated to ontologies as annotated in BioMART (Ensembl GRCh37 release 84), and enrichment of ontologies was analysed using TopGo (version 1.0) in R\textsuperscript{59}, using the *classic* algorithm, contrasting to all protein-coding genes.

**Structural modeling of DNA methylation**

The crystal structure of HIF2α:HIF1β in complex with DNA containing the RCGTG core sequence 5'-ACGTG-3' (\textsuperscript{25}, PDB code 4ZPK) was used as a template for introducing and analyzing the structural consequences of methyl groups at position 5 of the cytosines using the programs PyMOL (Schrodinger, LLC) and Chimera\textsuperscript{60}.

**Microscale thermophoresis (MST) binding assay**

MST measurements were performed in triplicate using the NanoTemper Monolith NT.115 instrument. The two protein complexes (HIF1α-HIF1β and HIF2α-HIF1β) were purified as described earlier\textsuperscript{25}. They were both labeled using Monolith NT Protein labeling kit RED-NHS (Nano Temper technologies). Oligonucleotides were from IDT. In brief, 25 nM of each labeled protein was mixed in 16 serial dilutions of 1:1 with different DNA concentrations starting from a concentration of 25 μM. The experiment was carried out in 20 mM phosphate buffer, 75 mM NaCl, 5 mM DTT, 0.05 % Tween-20, pH 7.4. Samples were incubated for 20 min on ice prior to loading 5 μL of each sample into the standard treated capillaries. MST measurements were carried out at 25 °C at 20 % LED power and medium MST power. Data was normalized to % fraction bound and the values for the equilibrium dissociation constant (K\textsubscript{D}) were calculated by fitting the curves in GraphPad Prism 7.

**Generation of mESCs containing a methylated or unmethylated HIF binding region**

A DNA fragment (human chr16:30,065,212-30,065,711) containing five CGTG motives was selected based on high HIF1β ChIP-enrichment in MCF7, RCC4 and SK-MEL-28 cells.
Oligonucleotides were designed to amplify the target region (AGGTGCAATTGTCCTCGCCTCCCTTAC and AAGGGCAATTGCGAGCTTTTTCCTTTACGA), and used for PCR amplification of the target region using the Q5® Hot Start High-Fidelity 2× Master Mix (NEB), followed by evaluation of the PCR products by gel electrophoresis and purification with the Qiaquick PCR purification kit (28104, Qiagen). These PCR primers were evaluated for specificity in human (MCF7, RCC4, SK-MEL-28) but not in mouse genomic DNA, while MfeI restriction sites were added to the ends of the primer pairs. The purified amplicon was digested with MfeI and cloned into the L1-poly-L1 plasmid (provided by Prof. Dirk Schubeler, Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland), containing a multiple cloning site flanked by two inverted L1 Lox sites. Correct insertion and sequence identity were verified by Sanger sequencing. This plasmid was in vitro methylated using M.SssI (NEB) according to the manufacturer’s instructions, and purified using isopropanol precipitation. Successful and complete in vitro methylation was confirmed by bisulfite-conversion (EZ DNA Methylation-Lightning Kit, D5031, Laborimpex), PCR amplification using the MegaMix Gold 2× Mastermix (Microzone) and Sanger sequencing. 10 μg of pIC-CRE plasmid and 25 μg of (un)methylated plasmid were electroporated in murine ES 159 cells containing an L1-flanked thymidine kinase expression cassette (provided by Prof. Dirk Schubeler, Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland). After electroporation, cells were plated and maintained in nonselective ES medium for 1 day, and from the second day onwards cultured in ES medium containing 10 μM ganciclovir. After 10 to 12 days, individual clones of the surviving cells were picked and transferred to ES medium in 96-well plates, then gradually expanded and, following DNA extraction, assessed for occurrence of successful insertion events using PCR (using the following oligonucleotides: AGGTGCAATTGTCCTCGCCTCCCTTAC and AAGGGCAATTGCGAGCTTTTTCCTTTACGA) and gel electrophoresis.

To verify maintenance of the methylation levels of the cloned HIF binding site, genomic DNA was extracted from a positive clone. 500 ng of DNA was bisulfite-converted using the EZ DNA Methylation-Lightning Kit (D5031, Laborimpex) and amplified using the MegaMix
Gold 2\times mastermix and validated primer pairs for the target locus (Forward: G\texttt{TGTGTGATAGGGGTGT}, Reverse: A\texttt{AAACCCTCCCTTACTCCTTCC}). Per sample, PCR product sizes were verified by gel electrophoresis, and amplicons converted into sequencing libraries using the NEBNext DNA library prep master mix set (E6040L, Bioke). These were next sequenced to a depth exceeding 500\times, and mapped and analyzed as described higher.

Positive colonies were expanded into 10 cm dishes and subjected to ChIP as described above. qPCR was performed with the SYBR GreenER™ qPCR SuperMix Universal (11762500, Life Technologies) on a Quantstudio 12K (Applied Biosystems), by using specific primers for the cloned locus (oligonucleotides T\texttt{CGTTTCCGACTTTTCCATC} and C\texttt{AGCCAGAATGTTGGCAAT}) and an independent murine genomic region for background quantification (oligonucleotides C\texttt{ACTTGCTGAATAATTGGGTGT} and C\texttt{TGTTGTCCAGTTTTCTTCG}). Enrichment was calculated as fold enrichment over background.

**TCGA samples and data analysis**

From the TCGA server, we selected 5,193 tumours from 14 tumour types: 413 bladder urothelial carcinoma (BLCA), 664 breast cancer (BRCA), 303 cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), 201 colon adenocarcinoma (COAD), 497 head and neck squamous cell carcinoma (HNSC), 269 kidney renal papillary cell carcinoma (KIRP), 372 liver hepatocellular carcinoma (LIHC), 460 lung adenocarcinoma (LUAD), 368 lung squamous cell carcinoma (LUSC), 175 pancreatic adenocarcinoma (PAAD), 498 prostate adenocarcinoma (PRAD), 465 skin cutaneous melanoma (SKCM), 338 stomach adenocarcinoma (STAD) and 171 uterine corpus endometrial carcinoma (UCEC) for which RNA data were available. The corresponding RNA-seq read counts were downloaded. DNA methylation data from Infinium HumanMethylation450 BeadChip arrays available were downloaded for the same samples.

Tumour types were classified as responsive or non-responsive to checkpoint immunotherapy following to the classification described by Turajlik and colleagues\textsuperscript{37}, with
3 notable exceptions. Firstly, kidney renal papillary cell carcinoma (KIRP) was classified as non-responsive as no study has yet demonstrated responsiveness of this tumour type. Secondly, clear cell kidney carcinoma (KIRC) tumours were not analysed, as the HIF-constitutive activation in these tumours (due to VHL-loss) precludes their classification into a hypoxic and normoxic subset. Finally, also microsatellite-instable COAD tumours were discarded from all analyses, as these tend to show constitutive hypermethylation, precluding their stratification in high and low methylation subgroups. Importantly, the tumour types that we define as responsive will still contain tumours that fail to respond, whereas the non-responsive tumour types will also contain a minority of tumours that do respond to immunotherapy. For instance, a subset of triple-negative breast tumours responds to checkpoint immunotherapy. Likewise, there is evidence that a small subset of LIHC tumours also responds, and that microsatellite-instable tumours also occur in UCEC tumours. Cryptic transcription loads were calculated using the total cryptic transcript read count divided by the total coding gene read count.

For the methylation stratification, the beta values of HM450K methylation microarray data were downloaded from TCGA. Probes overlapping the cryptic transcript promoter (i.e. 2,000 bp upstream and 500 bp downstream of the transcription start site) were regarded as cryptic transcription-associated probes. For each tumour, its promoter methylation level was calculated as the mean beta value of all cryptic transcription-associated probes. All tumours were then classified into high and low methylation groups based on the median value of methylation levels.

To identify which of these tumour samples were hypoxic or normoxic, we performed unsupervised hierarchical clustering based on a modification of the Ward error sum of squares hierarchical clustering method (Ward.D of the clusth function in R’s stats package) on normalized log-transformed RNA-seq read counts for 15 genes that make up the hypoxia metagene signature (ALDOA, MIF, TUBB6, P4HA1, SLC2A1, PGAM1, ENO1, LDHA, CDKN3, TPI1, NDRG1, VEGFA, ACOT7, CDKN3 and ADM). In each case the top 2 subclusters identified were annotated as normoxic and hypoxic.
To test the interaction between hypoxia and DNA methylation, we assessed read counts for cryptic transcripts in two negative-binomial generalized linear models with both oxygenation (hypoxic and normoxic; encoded as 0 and 1) and methylation (low and high methylation; encoded as 0 or 1), with or without an interaction term. Both models were compared to each other using DESeq. A positive interaction coefficient represents a cooperative enhancement of cryptic transcript expression in low-methylation, hypoxic tumours. To further enrich for tumours that are prone to respond to checkpoint immunotherapy, we stratified all tumour types into high PDL1 mRNA expressing and low PDL1 mRNA expressing tumours, and into tumours with a high or low tumour mutation burden (TMB). Stratification was done on the third decile in both cases. TMB was estimated based on the number of substitutions identified by TCGA in each tumour sample. All substitutions were considered, except for those also present in non-malignant samples (i.e., exclusion of germ-line variants) or those clustering within and across different samples (and therefore most likely representing sequencing or mapping errors).

Single-cell analysis

We used CREDENToR to map cryptic transcript expression in each individual cell from a public single-cell RNA-seq dataset. The cryptic transcript annotation was obtained from the CREDENToR analysis of lung TCGA tumours (LUAD and LUSC). CellRanger (version 1.1.0) was used as the mapping tool. The annotation of each individual cell is as previously-defined in Lambrechts et al. To study the effect of hypoxia and DNA methylation on cryptic transcript expression, we used a public single-cell dataset GSE97693. Single-cell RNA-seq reads were downloaded and mapped using STAR (version 2.5). Cancer cells for which the number of uniquely mapped reads exceeded 30% were stratified into hypoxic cells and normoxic cells as described higher. The cryptic transcript annotation was obtained from COAD tumours in TCGA. We selected 458 cryptic transcripts expressed in at least 20% of cells. Methylation was defined as the number of methylated CpGs over the total number of CpGs in a region 2 kb downstream and 500 bp upstream of the cryptic transcript transcription start site.
**4T1 Hif1b-knockout**

Four gRNAs targeting two different exons in the Hif1b locus of the mouse genome and one non-targeting gRNA (scramble) were designed with the appropriate restriction sites for the receiver plasmid using the online Crispor tool (http://crispor.tefor.net). Oligonucleotides corresponding to gRNAs were synthesized by IDT, and forward and reverse oligonucleotides were annealed in the CutSmart buffer (B7240S, NEB) before cloning into the LentiGuide-Puro plasmid (Plasmid 52963, Addgene). Positive colonies were screened by PCR and validated by Sanger sequencing. LentiGuide-Puro plasmids containing GFP were used as positive control to evaluate the transfection- and transduction-efficacy.

A transformation mix containing viral particles, TE, CaCl₂, H₂O and LentiGuide-Puro plasmid was added to HEK 293T cells when reaching 70% confluency. Four plasmids containing the different gRNAs for Hif1b and one plasmid containing the scramble gRNA were used, together with plasmids containing GFP as positive control. Medium was renewed after 14-16 hours and transfection efficiency was evaluated based on GFP expression. After 36 h, supernatant containing the concentrated virus was collected by ultracentrifugation. Virus was dissolved in clean PBS and stored at -150 °C.

4T1 cells were transduced with a lentiviral vector expressing a doxycycline-inducible Cas9 nuclease (Cat # CAS11229, Dharmacon) for a tight regulation of the Cas9 expression and gene editing. An infection rate of 30% was used to ensure that the majority of transduced cells harbour a single copy of the vector. These 4T1 cells were always kept in selection medium containing 10 µg/mL of blasticidin (ant-bl-05, Invivogen). When reaching 70% confluency, cells were transduced with one titer of virus. After 24h, the virus was removed and transduction efficacy evaluated based on GFP expression. After 48h, puromycin (P9620, Sigma-Aldrich) 1.5 µg/mL medium was added to the medium. Cells were kept in the presence of blasticidin and puromycin for the remaining experimental procedures. After 3-5 days, Cas9 expression was induced by adding doxorubicin (D2975000, Sigma-Aldrich) 0.5 µg/mL medium for 3 days. Cells were kept one day without doxorubicin before injection in the mice or further experimental procedures. 4T1 cells transduced
with the four gRNAs targeting Hif1b were expanded and proteins were extracted to test the efficacy of the knockout by Western blot. The most efficient gRNA was used to perform the further experiments (F: CACCGTGAATAAGCGCAGCGA and R: AAACCTGAGTTAGCTCTGGTAGTGC; non-targeting F: CACCGCAGACAGCTAACTCAG and non-targeting R: AAACCTGAGTTAGCTCTGGTAGTGC). Stability of knockout in these polyclonal 4T1 cells after two weeks was confirmed by Western blot.

**Mouse tumour model**

All the experimental procedures were approved by the Institutional Animal Care and Research Advisory Committee of the KU Leuven. 1x10^6 4T1 cells, 4T1 Hif1b-knockout or wild-type 4T1 cells (scramble) were injected orthotopically in the mammary gland of 10 weeks old Balb/c mice, and 1x10^6 CT26 or MC38 cells were injected subcutaneously in 10 weeks old Balb/c or C57BL/6J mice respectively. When the tumour was palpable (starting volume 100 mm^3), the mice were injected intraperitoneally with 0.8mg/kg of 5-aza-2'-deoxycytidine (aza) or PBS, 40mg/kg DC101 antibody (BE0089, InVivoMab) or IgG (BE0060, InVivoMab) or 10mg/kg anti-PD1 antibody (BE0146, InVivoMab) or IgG (BE0089, InVivoMab) according to the following schedules: DC101 three times per week; anti-PD1 every other day, starting when the tumour size was around 200 mm^3; aza was administered in 2 cycles with 2-days rest in between until the control tumours reached the endpoint. Tumour volumes were monitored every two to three days by a calliper, and mice were culled before tumour volumes exceeded 2,000 mm^3. When over 20% of mice were culled, the experiment was terminated (all arms). *In vivo* experiments in 4T1, CT26 and MC38 treated with aza or anti-PD1 antibody were performed three times, with at least 6 mice per treatment group in each experiment.

**Neo-epitope burden**

To assess neo-epitope burden, we mapped RNA sequencing data of isogenic 4T1, B16 and CT26 tumour models, removed duplicate reads from individual samples and merged per tumour model all samples into a single file. In this file, variants were called according to GATK best practices, using GATK3.4. Briefly, reads were split into exon segments and
sequences overhanging the non-exonic regions were hard-clipped using split’n’trim. Next, local indel realignment and base recalibration was performed, followed by variant calling with GATK’s HaplotypeCaller. After quality filtering for minimal Fisher strand values (30) and minimal read depth (10-fold), we removed SNPs reported in the Sanger Mouse project (rsIDdbSNPv137). Remaining variants were annotated by Annovar (version 2.17.0), and only variants in coding regions were retained. Finally, the neo-epitope burden was expressed as the number of non-SNP variants in coding sequences, normalized to the number of coding sequences that were expressed, the latter being defined as having a minimal read depth of 10.

**Immunofluorescence analysis of tumour sections**

Different protocols were applied depending on the epitope of interest: hypoxia (pimonidazole) staining was combined with blood vessel (CD31) staining, and cytotoxic T-cell activity (granzyme B) with T-cell infiltration (CD8a) staining. General (CD45) and cytotoxic (CD8a) T-cell infiltration stainings were performed separately. Tumours were harvested, fixed in formaldehyde and embedded in paraffin using standard procedures. Slides were deparafinated and rehydrated in 2 xylene baths (5 min), followed by 5 times 3 min in EtOH baths at decreasing concentrations (100%, 96%, 70%, 50% and water) and a 3 min Tris-buffered saline (TBS; 50 mM Tris, 150 mM NaCl, pH 7.6) bath. Antigen retrieval was done using AgR (DAKO) at 100 °C for 20 min, followed by cooling for 20 min. Slides were washes in TBS for 5 min, endogenous peroxidase activity was quenched using H₂O₂ (0.3% in MeOH), followed by three 5 min washes in TBS. Slides were blocked using pre-immune goat serum (X0907, Dako) or pre-immune rabbit serum (for pimonidazole, X090210, Dako) 20% in TNB. Binding of primary antibodies: FITC-conjugated mouse anti-pimonidazole (HP2-100, Hydroxyprobe), rabbit anti-Gzmb (ab4059, abcam) and rat anti-CD45 (553076, BD Biosciences) all 1:100 in TNB was allowed to proceed overnight. Slides were washed 3 times in TNT (0.5% Triton-X100 in TBS) for 5 min, after which secondary antibodies: peroxidase-conjugated rabbit anti-FITC (PA1-26804, Pierce), Alexa fluor 488-conjugated goat anti-rabbit (A-11034, Thermo Fisher) and biotinylated goat anti-rat (559286, BD biosciences) all 1:100 in TNB with 10% pre-immune goat serum were allowed
DNA methylation repels HIF binding

D'Anna et al

07/02/2020

Page 37 of 64

to bind for 1 h. Slides were washed 3 times for 5 min in TNT, after which signal amplification was done by 30 min incubation with peroxidase-conjugated streptavidin 1:100 in TNB (for all besides pimonidazole) accompanied by nuclear staining with Hoechst (H3570, Thermo Fisher) 1:500 in TNB only for the single (CD45 or CD8a) stainings, washing (3 times 5 min in TNT) and 8 min incubation using Fluorescein Tyramide (for pimonidazole NEL701A001KT, perkin Elmer) or Cy3 (NEL704A001KT, Perkin Elmer) 1:50 in amplification diluent.

Slides stained for pimonidazole and Gzmb required co-staining for CD31 and CD8a respectively and were subjected to a second indirect staining for the latter epitopes. After 5 min of TNT and 5 min of TBS, slides were quenched again for peroxidase activity using H$_2$O$_2$ and blocked using pre-immune goat or rabbit (CD31) serum, prior to a second overnight round of primary antibody binding: rat anti-CD31 (557355, BD Biosciences) or rat anti-CD8a (14-0808-82, Thermo Fisher) 1:100 in TNB. The next day, 3 times 5 min washes with TNT were followed by a 1 h incubation with biotinylated goat anti-rat (559286, BD biosciences) 1:100 in TNB, again 3 times 5 min washes with TNT, a 30-min incubation with peroxidase-conjugated streptavidine 1:100 in TNB accompanied by nuclear staining with Hoechst (H3570, Thermo Fisher) 1:500 in TNB, 3 times 5 min washes with TNT and signal amplification for 8 min using Cy3 (NEL704A001KT, Perkin Elmer) 1:50 in amplification diluent. Finally, slides were washed 3 times for 5 min with TNT and mounted with Prolong Gold (P36930, Life Technologies).

For immunofluorescence analysis on 4T1 wild-type tumours, slides were imaged on an infraMouse Leica DM5500 microscope. 4 sections from different treatment groups were stained per slide while 6 pictures from different tumour areas were used for processing with Image J. More specifically, nuclei were identified using the Hoechst signal, and signal intensities for Fluorescein (pimonidazole), alexa fluor 488 (Gzmb) and Cy3 (CD45, CD8a and CD31) were used to detect Gzmb+, CD45+ and/or CD8a+ cells. Analyses were exclusively performed on slide regions showing a regular density and shape of nuclei, in order to avoid inclusion of acellular or necrotic areas. Gzmb+ CD8a+ cells were counted directly, allowing the precise quantification of the number of active cytotoxic T-cells per
tumour. The number of CD45+ cells was used to normalize the number of CD8a+ cells, as such calculating the number of infiltrating cytotoxic T-cells compared to the total immune infiltration. CD31-positive regions were quantified manually using Image J. The pimonidazole signal was used together with the Hoechst signal to quantify the percentage of hypoxia per tumour area in each picture and stratify tumours as hypoxic (pimo-high) or normoxic (pimo-low).

For immunohistofluorescence on 4T1 Hif1b-knockout or -scramble grafts, tumours were harvested and snap frozen in liquid nitrogen before temporary storage at -80 °C. Thawed tumours were embedded in paraffin and sectioned using standard procedures (5 µm of thickness). In a Leica Autostainer (30 min), slides were deparafinated and rehydrated in 2 xylene baths for 5 min, followed by 5 min in ethanol baths at decreasing concentrations (100 %, 96 %, 70 %, 50 % and water). Slides were fixed in 10 % neutral buffered formalin for 10 min and rinsed twice in double distilled water. Antigen retrieval proceeded in AR6 buffer (AR600, PerkinElmer) at 100 °C for 23 min in a pressure cooker, followed by cooling in double distilled water for 20 min. Slides were washes in TBST (TBS with 0.5% Tween 20) for 3 min, and blocked using blocking buffer (pre-immune goat serum (X0907, Dako) 10 %, 1 % BSA (126575, Millipore) in TBS)) for 30 min. The primary antibody (rabbit anti-Gzmb) 1:1,000 in dilution buffer (1 % BSA in TBS) was applied for 30 min at RT, followed by 3 washes of 2 min in TBST at RT. Slides were next incubated with the secondary antibody (EnVision™+/HRP goat anti-rabbit (K4003, Dako)) for 10 min at RT, and washed 3 times for 2 min in TBST at RT. The OPAL 570 fluorophore (fp1488, PerkinElmer) 10 % in amplification diluent (FP1498, PerkinElmer) was applied for 10 min at RT followed by 3 washes of 2 min in TBST at RT. Slides were stripped by heating in AR6 buffer just below boiling point and cooled down in double distilled water, followed by rinsing in TBST. These steps were repeated starting from blocking for the second staining with primary antibody rat anti-CD8a 1:300, secondary antibody goat anti-rat (MP-7444, Vector) and opal 690 (fp1497, PerkinElmer) and the third staining with rat anti-CD45 1:1,000, secondary antibody goat anti-rat and Opal 520 (fp1487, PerkinElmer). After the third staining, slides were incubated with spectral DAPI (fp1490, PerkinElmer) 10 % in TBST for 5 min at RT,
washed for 2 min in TBST at RT and mounted with ProLong Diamond Antifade Mountant (P36961 Invitrogen). Images were acquired on a Zeiss Axio Scan.Z1 using a ×20 objective and ZEN 2 software (Zeiss) with exposure times between 10-50 ms. Image processing was done using QuPath (version 0.1.2). Specifically, following visual inspection of the staining results, cells were first automatically detected using the DAPI channel (cell size constrained between 5 and 400 µm²). Next, a cell classifier was generated using QuPath. Specifically, for 1 slide out of all slides, 5 sets of cells were selected: one set that was positive for CD45, one set that was negative for CD45, and three sets of CD45+ cells positive for CD8, Gzmb and CD8, or Gzmb alone. Using these 5 sets of cells, a random trees classifier was generated. Cell classification was visually verified to have occurred correctly. Next, in each tumour section, a representative region was selected, containing at least 1,000 cells. On these cells, the random trees classifier was subsequently applied. This process was reiterated for all other tumour sections stained for the same set of markers. The resulting cell identities were then exported, and processed in R. For each tumour, average cell frequencies were generated, which were summarized using boxplots.

Immunofluorescence analysis of 4T1 cells

For the dsRNA staining on 4T1 cells, 12,000 cells were seeded on gelatin-coated glass slips in 12-well plates on day 0. After attaching for 6h, cells were treated with aza or control (DMSO) for 3 days, with renewal of the medium after 48 hours. After 72h, the medium was refreshed and cells were incubated in hypoxia or normoxia for 16 hours. After washing 3 times with PBS, cells were fixed in 1ml of ice-cold methanol for 15 min at -20 °C. Cells were washed 3 times with PBS and blocked for 1 h with blocking buffer (0.1% triton X-100 with 5% goat serum in PBS). Primary antibody (1:50 dilution in blocking buffer; clone J2, Scicons) was applied overnight at 4 °C. Cells were washed 3 times for 10 min with washing buffer (0.1% triton X-100 in PBS) and secondary antibody (1:500 in secondary antibody buffer (0.1% triton X-100 with 2% goat serum in PBS)). Goat-anti-mouse IgG coupled to alexa 488, Life Technologies) was applied for 1 hour in the dark. Cells were washed 3 times for 10 min with washing buffer and mounted with DAPI stain
on cover glasses. Slides were imaged on an infraMouse Leica DM5500 microscope. 3 slides from different treatment groups were stained in triplicate (biological replicates), while 3 pictures from different slides were used for processing with Image J. More specifically, nuclei were identified using the DAPI signal, and nucleated cells were further selected based on particle size. Signal intensities for alexa fluor 488 in the selected cells were used to detect dsRNA+ cells. Analyses were exclusively performed on slide regions showing a regular density and shape of nuclei, in order to avoid inclusion of acellular or necrotic areas. Mean dsRNA expression was calculated per experiment, normalised to the background signal (secondary antibody only) and expressed as mean pixel intensity relative to the control group (normoxia + DMSO).

Published data sets

Published data sets were obtained from GEO under the following accession numbers: HIF1β, HIF1α, HIF2α and isotype ChIP-seq in MCF7: GSM700947, GSM700944, GSM700945, GSM700948; WGBS of MCF7: GSM1328112, and WGBS of WT mESC: GSM1127953; CTCF, FOXA1 and GATA3 ChIP-seq in MCF7: GSM1003581, GSM1010727, wgEncodeEH002293; transcription factors in MCF7: GSE41561; NOMe-seq: GSE57498; single-cell RNA-seq data for lung tumours; single-cell RNA-seq and methylation data for colorectal tumours: GSE97693.
**DECLARATIONS**

**ETHICS APPROVAL**

All the experimental procedures on animals were approved by the Institutional Animal Care and Research Advisory Committee of the KU Leuven.

**AVAILABLE OF DATA AND MATERIALS**

Sequencing data generated for this study have been deposited in GEO and will be made accessible upon publication.

**COMPETING INTERESTS**

The authors declare that they have no competing interests.

**FUNDING**

This work was supported by grants from the ERC (CHAMELEON 617595 to D.L.; CHAMELEO 334420 to B.T.), Stichting tegen Kanker (‘Foundation contre le Cancer’): FAF-C/2016/876 and FWO: G070615N. Intensive computation was provided by the Flemish Supercomputer Center (VSC). H.Z. and B.T. were supported by FWO-F postdoctoral fellowships and L.V.D by an aspirant FWO fellowship.

**AUTHOR CONTRIBUTIONS**

Funding Acquisition, Supervision, Conceptualization, B.T., M.M. and D.L.; Methodology, B.T., F.D.A, L.V.D., J.X., L.M., S.N.S., P.B.-K., L.S., J.M., J. Q., R. A., M.D.B., C.S., F.R., V.C., S.K., J.D.S., P.C.; Software, B.T., F.D.A., J.X. and H.Z.; Formal Analysis B.T., F.D.A., L.V.D., H.Z., R.V.B, L.M., S.N.S., F.D., V. C., S.K.; Investigation B.T., F.D.A., H.Z., L.V.D., J.X.; Writing Original Draft, B.T., W.R., S.N.S., F.D.A and D.L.

**ACKNOWLEDGEMENTS**

We thank Masaki Okano and Guoliang Xu for murine Dnmt-TKO and Tet-TKO ESCs and the corresponding matching WT ESCs.
REFERENCES

1. Thienpont, B. et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* **537**, 63-68 (2016).

2. Héberlé, É. & Bardet, A.F. Sensitivity of transcription factors to DNA methylation. *Essays in biochemistry* (2019).

3. Ivan, M. & Kaelin Jr, W.G. The EGLN-HIF O2-sensing system: multiple inputs and feedbacks. *Molecular cell* **66**, 772-779 (2017).

4. Semenza, G.L. Targeting HIF-1 for cancer therapy. *Nat Rev Cancer* **3**, 721-732 (2003).

5. Casazza, A. et al. Tumor stroma: a complexity dictated by the hypoxic tumor microenvironment. *Oncogene* **33**, 1743-1754 (2014).

6. Keith, B., Johnson, R.S. & Simon, M.C. HIF1alpha and HIF2alpha: sibling rivalry in hypoxic tumour growth and progression. *Nat Rev Cancer* **12**, 9-22 (2012).

7. Li, W. et al. Hypoxia-induced endothelial proliferation requires both mTORC1 and mTORC2. *Circulation research* **100**, 79-87 (2007).

8. Palazon, A. et al. An HIF-1α/VEGF-A axis in cytotoxic T cells regulates tumor progression. *Cancer Cell* **32**, 669-683. e665 (2017).

9. Chi, J.T. et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med* **3**, e47 (2006).

10. Sung, F.L. et al. Genome-wide expression analysis using microarray identified complex signaling pathways modulated by hypoxia in nasopharyngeal carcinoma. *Cancer Lett* **253**, 74-88 (2007).

11. Xia, X. & Kung, A.L. Preferential binding of HIF-1 to transcriptionally active loci determines cell-type specific response to hypoxia. *Genome Biol* **10**, R113 (2009).

12. Wenger, R.H., Kvietikova, I., Rolfs, A., Camenisch, G. & Gassmann, M. Oxygen-regulated erythropoietin gene expression is dependent on a CpG methylation-free hypoxia-inducible factor-1 DNA-binding site. *Eur J Biochem* **253**, 771-777 (1998).

13. Bell, A.C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-485 (2000).
14. Maurano, M.T. et al. Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep 12, 1184-1195 (2015).

15. Feng, J., Liu, T. & Zhang, Y. Using MACS to identify peaks from ChIP-Seq data. Curr Protoc Bioinformatics Chapter 2, Unit 2 14 (2011).

16. Menafra, R. et al. Genome-wide binding of MBD2 reveals strong preference for highly methylated loci. PLoS One 9, e99603 (2014).

17. Huang, Y. et al. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One 5, e8888 (2010).

18. Ahmed, S., Wang, A., Celius, T. & Matthews, J. Zinc Finger Nuclease–Mediated Knockout of AHR or ARNT in Human Breast Cancer Cells Abolishes Basal and Ligand-Dependent Regulation of CYP1B1 and Differentially Affects Estrogen Receptor α Transactivation. toxicological sciences 138, 89-103 (2013).

19. Maltepe, E., Schmidt, J.V., Baunoch, D., Bradfield, C.A. & Simon, M.C. Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT. Nature 386, 403 (1997).

20. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74 (2012).

21. Taberlay, P.C., Statham, A.L., Kelly, T.K., Clark, S.J. & Jones, P.A. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Res 24, 1421-1432 (2014).

22. Griffon, A. et al. Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. Nucleic Acids Res 43, e27 (2015).

23. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. Nature Reviews Genetics 18, 517 (2017).

24. Mariani, C.J. et al. TET1-mediated hydroxymethylation facilitates hypoxic gene induction in neuroblastoma. Cell Rep 7, 1343-1352 (2014).

25. Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. Structural integration in hypoxia-inducible factors. Nature 524, 303-308 (2015).
26. Tsumura, A. et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805-814 (2006).

27. Criscione, S.W., Zhang, Y., Thompson, W., Sedivy, J.M. & Neretti, N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* **15**, 583 (2014).

28. Yang, W.R., Ardeljan, D., Pacyna, C.N., Payer, L.M. & Burns, K.H. SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic acids research* **(2019)**.

29. Elvidge, G.P. et al. Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition the role of HIF-1α, HIF-2α, and other pathways. *Journal of biological chemistry* **281**, 15215-15226 (2006).

30. Blank, C.U., Haanen, J.B., Ribas, A. & Schumacher, T.N. CANCER IMMUNOLOGY. The "cancer immunogram". *Science* **352**, 658-660 (2016).

31. Chiappinelli, K.B. et al. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162**, 974-986 (2015).

32. Zitvogel, L., Galluzzi, L., Kepp, O., Smyth, M.J. & Kroemer, G. Type I interferons in anticancer immunity. *Nat Rev Immunol* **15**, 405-414 (2015).

33. Leruste, A. et al. Clonally expanded T cells reveal immunogenicity of rhabdoid tumors. *Cancer cell* **(2019)**.

34. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61 (2015).

35. Buffa, F.M., Harris, A.L., West, C.M. & Miller, C.J. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br J Cancer* **102**, 428-435 (2010).

36. Bian, S. et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060-1063 (2018).
37. Turajlic, S. et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol* **18**, 1009-1021 (2017).

38. Roulois, D. et al. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **162**, 961-973 (2015).

39. Kim, K. et al. Eradication of metastatic mouse cancers resistant to immune checkpoint blockade by suppression of myeloid-derived cells. *Journal for Immunotherapy of cancer* **2**, P267 (2014).

40. Ngiow, S.F. et al. A threshold level of intratumor CD8+ T cell PD1 expression dictates therapeutic response to anti-PD1. *Cancer research*, canres. 1082.2015 (2015).

41. Riggs, A.D. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* **14**, 9-25 (1975).

42. Holliday, R. & Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226-232 (1975).

43. Vanharanta, S. et al. Epigenetic expansion of VHL-HIF signal output drives multiorgan metastasis in renal cancer. *Nature medicine* **19**, 50-56 (2013).

44. Partch, C.L. & Gardner, K.H. Coactivators necessary for transcriptional output of the hypoxia inducible factor, HIF, are directly recruited by ARNT PAS-B. *Proc Natl Acad Sci U S A* **108**, 7739-7744 (2011).

45. Brocks, D. et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nature genetics* **49**, 1052 (2017).

46. Trizzino, M. et al. Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res* **27**, 1623-1633 (2017).

47. Simonti, C.N., Capra, J.A. & Pavličev, M. Transposable Element Exaptation into Regulatory Regions Is Rare, Influenced by Evolutionary Age, and Subject to Pleiotropic Constraints. *Molecular Biology and Evolution* **34**, 2856-2869 (2017).

48. Navarro, C. The Mobile World of Transposable Elements. *Trends Genet* **33**, 771-772 (2017).
49. Trizzino, M., Kapusta, A. & Brown, C.D. Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genomics* **19**, 468 (2018).

50. Simon, M.C. & Keith, B. The role of oxygen availability in embryonic development and stem cell function. *Nature reviews Molecular cell biology* **9**, 285 (2008).

51. Rodriguez-Terrones, D. & Torres-Padilla, M.-E. Nimble and ready to mingle: transposon outbursts of early development. *Trends in Genetics* (2018).

52. Reik, W., Dean, W. & Walter, J. Epigenetic Reprogramming in Mammalian Development. *Science* **293**, 1089-1093 (2001).

53. Sheng, W. *et al.* LSD1 Ablation Stimulates Anti-tumor Immunity and Enables Checkpoint Blockade. *Cell* (2018).

54. Hu, X. *et al.* Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **14**, 512-522 (2014).

55. Sakaue, M. *et al.* DNA methylation is dispensable for the growth and survival of the extraembryonic lineages. *Curr Biol* **20**, 1452-1457 (2010).

56. Bogu, G.K. *et al.* Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Mol Cell Biol* **36**, 809-819 (2015).

57. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33**, 290 (2015).

58. Anders, S., Pyl, P.T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).

59. Alexa, A. & Rahnenfuhrer, J. topGO: enrichment analysis for gene ontology. *R package version 2.12.0* (2010).

60. Pettersen, E.F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).

61. Zhao, H. *et al.* Mismatch repair deficiency endows tumors with a unique mutation signature and sensitivity to DNA double-strand breaks. *Elife* **3**, e02725 (2014).

62. Lambrechts, D. *et al.* Phenotype molding of stromal cells in the lung tumor microenvironment. *Nature medicine* **24**, 1277-1289 (2018).
63. Habibi, E. et al. Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. *Cell Stem Cell* **13**, 360-369.

64. Charoentong, P. et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* **18**, 248-262 (2017).

65. Kong, S.L., Li, G., Loh, S.L., Sung, W.K. & Liu, E.T. Cellular reprogramming by the conjoint action of ERalpha, FOXA1, and GATA3 to a ligand-inducible growth state. *Mol Syst Biol* **7**, 526 (2011).

66. Ponnaluri, V.C. et al. NicE-seq: high resolution open chromatin profiling. *Genome biology* **18**, 122 (2017).
**Figure Legends**

**Figure 1 | Methylation at HIF1β binding sites**

(a) Heatmaps of HIF1β binding and DNA methylation for 7,153 regions (identified using a stringent threshold of $P<10^{-15}$ in MACS) surrounding the HIF1β ChIP-seq peak summit (±5 kb). Heatmaps depict reads per kb per million reads (RPKM) of HIF1β ChIP-seq and of 5mC DNA IP-seq (mDIP), and % DNA methylation as estimated by SeqCapEpi BS-seq or whole-genome BS-seq (respectively, SeqCapEpi and WGBS). HIF1β binding was assessed after 16 hours of 0.5% O$_2$ (hypoxia) and DNA methylation under 21% O$_2$ (normoxia). (b) Violin plots of the methylation level inside and outside HIF1β binding peaks, as estimated by SeqCapEpi BS-seq. (c) Sequencing read depth of HIF1β ChIP and its input, at all RCGTG sequences in MCF7 cells, stratified for methylation at the CG in the core RCGTG sequence. Shown are boxplots for all RCGTG’s in the human genome for which $>10	imes$ coverage was obtained after SeqCapEpi BS-seq, with dark red dots denoting averages. See Figure S5 for additional QC of ChIP-seq data and Supplementary Table 1 for more details about HIF1β binding peak locations. (d) Venn diagram of 20,613 shared and unique HIF1β binding sites detected across 3 cell lines. Only stringent binding sites ($P<10^{-15}$) are shown. Binding sites showing intermediate levels of HIF1β ChIP-seq enrichment in 1 or 2 cell lines are unclassified and not shown here ($n=445, 2,812$ and 887 peaks, detected in SK-MEL-28, RCC4 and MCF7 respectively). (e) Heatmaps of HIF1β binding (red) and DNA methylation as estimated using SeqCapEpi BS-seq (blue) at regions flanking the HIF1β ChIP-seq peak summit (±5 kb). (top) HIF1β binding peaks shared between the 3 cell lines. (bottom) HIF1β binding peaks unique to each cell line. Heatmaps depict RPKM of HIF1β ChIP-seq and % DNA methylation. HIF1β binding was assessed after 16 hours of 0.5% O$_2$ (hypoxia) and DNA methylation under 21% O$_2$ (normoxia). (f) Quantification of the methylation level at HIF1β binding peak summits ±100 bps, for peaks that are shared between or unique to one of the 3 cell lines. (g) Enrichment of gene expression (observed/expected) upon hypoxia per cell line, for genes associated with HIF1β binding sites (within 50 kb) that are shared between or unique to one of the 3 cell lines, as labelled on the X-axis. (h) Fraction of HIF1β peaks overlapping with the binding peaks of individual transcription factors$^{22}$, or
with any of the 11 transcription factors profiled in MCF7 cells ("combined"). (i-j) (i) Overlap between HIF1β binding peaks and other transcription factor binding sites detected in MCF7 cells. Shown are fractions of HIF1β binding peaks shared between (grey) or unique for a cell line (coloured). (j) mRNA expression level of transcription factors in each cell line, as determined using RNA-seq. Transcription factors expressed in all 3 cell lines are highlighted as "shared TFs" with a light grey box.

**Figure 2 | DNA methylation directly repels HIF1β binding**

(a-b) Boxplot (left) and scatter plot (right) of methylation levels of HIF1β-bound immunoprecipitated DNA fragments obtained by ChIP-BS-seq (ChIP-BS) compared to input by SeqCapEpi BS-seq (SeqCapEpi) in MCF7 cells (a), or of HIF1β-bound immunoprecipitated DNA fragments obtained by ChIP-BS compared to input by whole-genome BS-seq (WGBS) in mouse Tet-triple-knockout (Tet-TKO) ESCs (b). The red dotted line in the scatter plot indicates the theoretical value of equal methylation in immunoprecipitated and input DNA. P values by t-test. (c-d) Recombination-mediated cassette exchange. (c) A human HIF binding site (chr16: 30,065,212-30,065,711 on hg38) was cloned between 2 L1 Lox sites and in vitro methylated (blue) or not (red) prior to co-transfection with a CRE recombinase-encoding plasmid into mESCs transformed to contain an L1 Lox-flanked thymidine kinase (TK). (d) Following successful cassette exchange, these ESCs were incubated in hypoxia (0.5% O2 for 16 hours) and probed using HIF1β ChIP-qPCR for HIF binding at the differentially methylated cassette. Shown is the fold enrichment over background (n=3 independent ChIP pairs; *P<0.05 by t-test). (e-f) Microscale thermophoresis-based assessment of sensitivity of HIF1α-HIF1β (e) and HIF2α-HIF1β (f) heteroduplexes to methylation at HIF binding sites in physiological buffer (PBS). RCGTG sequences in the double-stranded DNA oligonucleotides were either absent (grey), methylated (blue) or unmethylated (red) at the CpG site. Calculated K_D values are shown under each graph. (g) Excerpt from the crystal structure of HIF2α-HIF1β in complex with a DNA duplex containing the core HIF binding sequence 5'-ACGTG-3' (PDB code 4ZPK)25. (h) Modelling of methylation of CpG cytosines in ACGTG reveals severe steric hindrance. The two views show hard-sphere models of methylated cytosines modelled at
position 5 (including bonding hydrogen atoms) and how they severely violate the van der Waals envelopes (2.5 Å width) of Arg27 in HIF2α (left) and Arg102 in HIF1β (right).

**Figure 3 | DNA demethylation uncovers new HIF1β binding sites**

(a) Heatmaps of HIF1β binding (RPKM) and DNA methylation as determined using WGBS at regions flanking the summit of HIF1β binding peaks (± 5 kb) either shared with WT or TKO-specific ESCs. (b) % methylation at shared and TKO-specific HIF1β binding sites in WT ESCs. See Figure S5 for scatter plots illustrating the correlations between HIF1β ChIP-seq replicates in *Dnmt-WT versus Dnmt-TKO* ESCs. (c) Cumulative frequency of distance to the nearest RCGTG motif for shared, TKO-specific and randomized HIF1β binding peaks. (d) Observed/expected frequency of upregulated genes associated with shared and TKO-specific HIF1β binding peaks in WT and *Dnmt*-TKO ESCs exposed to 24 hours of hypoxia (0.5% O₂). (e) Distance of shared and TKO-specific HIF1β binding peaks in ESCs to the nearest TSS. A bimodal peak was detected indicating proximal and distal binding events. (f) Functional genome annotation using ChromHMM of shared and TKO-specific HIF1β binding peaks in ESCs. (g) Distance of shared and TKO-specific HIF1β binding peaks to open chromatin regions in ESCs. A bimodal peak was detected indicating proximal and distal binding events. (h) Ontology analysis of genes associated with shared and TKO-specific HIF1β binding peaks in ESCs. (i) HIF1β binding sites in LINEs, LTRs and SINEs after 10,000 random permutations and as observed by HIF1β ChIP-seq (actual HIF1β binding) for all HIF1β sites (*top panel*) and only for distal HIF1β sites (*bottom panel*). *** P<0.001 by Fisher’s exact test. (j) Distribution of HIF1β binding peaks detected in murine *Dnmt*-TKO ESCs for the retrotransposon families, colour-coded by retrotransposon class (green: LTR; violet: LINE; yellow: SINE).

**Figure 4 | DNA methylation represses hypoxia-induced cryptic transcript activation**

(a) Heatmaps of HIF1β binding (RPKM) and DNA methylation as determined using SeqCapEpi BS-seq at regions flanking the summit of HIF1β binding peaks (± 5 kb). Shown are HIF1β binding peaks that are shared between vehicle-treated and aza-treated MCF7...
cells, or that are specific to aza-treated cells. 12,782 HIF1β binding peak positions were detected across vehicle- and aza-treated MCF7 using a P<10^{-15} threshold. (b) Violin plots of methylation detected by SeqCapEpi BS-seq at HIF1β binding peaks that are shared between vehicle-treated and aza-treated MCF7 cells, or that are specific to aza-treated MCF7 cells. (c) HIF1β binding sites in LINES, LTRs and SINEs after 10,000 random permutations and as observed by HIF1β ChIP-seq (actual HIF1β binding) for HIF1β binding peaks that are shared between vehicle- and aza-treated MCF7 cells, or specific to aza-treated MCF7 cells for all HIF1β sites (top panel) and only for distal HIF1β sites (bottom panel). (d) Distribution of HIF1β binding peaks detected in aza-treated MCF7 cells at retrotransposon families, colour-coded by retrotransposon class (green: LTR; violet: LINE; yellow: SINE). (e) Violcano plots showing differential expression of HIF-bound cryptic transcripts, as determined by CREDENToR in MCF7 cells exposed to vehicle (DMSO) or 5-aza-2'-deoxycytidine (aza; 1 μM) for 4 days, hypoxia (0.5% oxygen, 1 day) or normoxia. Significantly upregulated and downregulated transcripts are highlighted in red and blue, respectively. The associated numbers refer to how many transcripts are up- or downregulated at a 1% FDR and a 0.001% FDR, as indicated by the horizontal line. (f) dsRNA formation potential of all cryptic transcripts (grey) and of HIF-bound cryptic transcripts (red). Shown are the fraction of all RNAs for which transcription overlaps with a transcript expressed from the complementary strand (“sense-antisense”), and RNAs containing the same retrotransposon repeat element in sense and antisense orientation (“palindromic”). P values from chi-square test. (g-h) Expression of HIF-bound (g) and non-HIF-bound (h) cryptic transcripts relative to vehicle-treated controls (vehicle normoxia) in MCF7 cells wild-type (WT) (g upper panels and h) or knockout (KO) (g bottom panels) for HIF1B. Shown are expression changes as assessed using CREDENToR (g left panels and h) and RepEnrich (g right panels), with error bars indicating geometric mean ± s.e.m. n.s. not significant, ***P<0.001 by t-test.

**Figure 5 | Cryptic transcript expression in tumours**

(a-c) Cryptic transcript expression in tumours characterized by TCGA. Shown is cryptic transcript expression in tumours with high or low methylation of cryptic transcript
promoter regions (blue or red; > or ≤ the median methylation level of each tumour type),
and in normoxic or hypoxic (light or dark colour) tumours. Data are shown for (a) all
tumour types combined, (b) stratified into those that are responding or non-responding
to immunotherapy following the classification described by Turajlik and colleagues37, and
(c) for each tumour type separately. P values by t-test, red values indicating inverse
correlations. (d) DNA methylation levels at cryptic transcript promoters (left) and cryptic
transcript expression (right) in tumours profiled in TCGA, stratified into tumour types that
are responsive (n=2,280) or non-responsive (n=2,214) to checkpoint immunotherapy.
***P<0.001 by t-test. (e) Heatmap showing the expression (Z score, blue to red) of the 59
cryptic transcripts associated with cytolytic activity in tumours responsive to
immunotherapy from TCGA. The boxplot on the right depicts the log fold change in
expression of the same 59 cryptic transcripts in hypoxic versus normoxic MCF7 cells (24
hours, 0.5% O2), and of MCF7 cells after 4-day exposure to aza versus vehicle-treated
hypoxic MCF7 cells (P<0.05 for all cryptic transcripts, either for hypoxia versus vehicle, or
for hypoxia plus aza versus aza alone). At the bottom, cytolytic activity of each TCGA
sample is depicted. LUAD; lung adenocarcinoma; LUSC, lung squamous cell carcinoma;
HNSC, head and neck squamous cell carcinoma; BLCA, bladder urothelial carcinoma;
CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; SKCM, skin
cutaneous melanoma.

Figure 6 | Aza treatment increases tumour immunogenicity HIF-dependently

(a) Expression in 4T1 cells of cryptic transcripts (CREDENToR, left) or retrotransposon
subfamilies (RepEnrich, right) bound by HIF1β in hypoxic 4T1 cells, following exposure to
vehicle (DMSO) or 5-aza-2′-deoxycytidine (aza; 1 μM) for 4 days, hypoxia (0.5% oxygen, 1
day) or normoxia. Difference in the distribution of expression is expressed as fold change
of counts per million over control 4T1 cells. Error bars indicate geometric mean ± s.e.m.
(b) Expression of cryptic transcripts (CREDENToR, left) or retrotransposon subfamilies
(RepEnrich, right) bound by HIF1β in 4T1 cells in vehicle- and aza-treated 4T1 tumours
(n=6 per treatment condition). Difference in the distribution of expression is expressed as
fold change of counts per million over control 4T1 tumours. Error bars indicate geometric
mean ± s.e.m. (c-d) Quantification of the number of blood vessels (CD31 staining; c) and percentage of hypoxia (pimonidazole staining; d) in 4T1 tumours from mice injected with DC101 or control IgG (with at least 4 mice per treatment condition, see Methods). (e) Expression of cryptic transcripts and retrotransposon subfamilies (bound by HIF1β in 4T1 cells) as determined by CREDENToR (left) and RepEnrich (right) in control antibody- and DC101-treated 4T1 tumours (n=6 per treatment condition). Difference in the distribution of expression is expressed as fold change of counts per million over control 4T1 tumours. Error bars indicate geometric mean ± s.e.m. (f) (left) Signal intensity of dsRNA staining in 4T1 cells treated with aza or PBS, and incubated in hypoxia or normoxia for 24 hours. (right) Immunofluorescence of dsRNA using a dsRNA antibody (clone J2, green) in 4T1 cells treated with aza or vehicle (PBS), and by a 24h incubation in hypoxic (0.5% O₂) or normoxic conditions (scale 40 µm). A representative image is shown for each condition. (g) Barplot showing the tumour weight of vehicle- and aza-treated 4T1 tumours (n=6 per treatment condition). ***P<0.001 by paired t-test. (h) Association between aza treatment and immune cell infiltration estimates in 4T1 tumours from mice treated with either aza or PBS, as calculated by GSVA on PanCancer immune metagenes and visualized by their T value (at least 6 mice per treatment condition were sequenced). Red bars indicate significant associations (P<0.05). (i) Expression of cryptic transcripts (CREDENToR, top) and retrotransposon subfamilies (RepEnrich, bottom) bound by HIF1β in 4T1 cells in 4T1 tumours WT or KO for Hif1b implanted in mice treated with vehicle or aza (see Methods, at least 6 tumours per treatment condition were sequenced). Difference in the distribution of expression is expressed as fold change of counts per million over Hif1b-WT vehicle-treated 4T1 tumours. Error bars indicate geometric mean ± s.e.m. (j) Growth of tumours generated by grafting mice orthotopically with 4T1 cells wild-type (4T1\(^{Hif1b-scr}\)) or KO for Hif1b (4T1\(^{Hif1b-KO}\)). Mice were treated with aza or vehicle (PBS) on the days indicated with an arrow (see Methods). Data represent mean and s.e.m. from independent experiments each with at least n=6 mice per group. *P<0.05 by t-test. A genotype-by-treatment interaction as assessed by ANOVA was P<0.001. (k) Quantification of CD8\(^+\) and granzyme b (Gzmb)\(^+\) cells, depicted as percentage of CD8\(^+\)
cells, from 4T1 cells WT for Hif1b (4T1^{Hif1b-WT}) or KO for Hif1b (4T1^{Hif1b-KO}) and treated with aza or vehicle (PBS) (n=6 per group; see Methods). *P<0.05 by t-test.
Figure 1

(a) 7,153 HIF1β binding peaks from SeqCapEpi WGBS mDIP

(b) % methylation of RCGTG motifs

(c) Methylation - HIF1β binding across all RCGTG motifs

(d) Venn diagram showing the overlap of HIF1β peaks among RCC4, MCF7, and SK-MEL-28

(e) Heatmaps showing the methylation profiles of RCC4, MCF7, and SK-MEL-28

(f) Enrichment of overexpressed genes within 50 kb of HIF1β binding sites

(g) RNA expression (reads per million) of overexpressed genes

(h) Heatmap showing the fraction of overlap with 7,153 HIF1β peaks

(i) Heatmap showing the fraction of HIF1β peak summits overlapping binding peaks of other TFs in MCF7

(j) RNA expression (reads per million) of unique TFs
Figure 2

(a) 

% CpG methylation

P=2.3×10^{-14}

SeqCapEpi | ChiP-BS

(b) 

% CpG methylation

P=2.2×10^{-16}

WgBS | ChiP-BS

(c) 

mES cells

or

M. SssI-methylated plasmid

unmethylated plasmid

(d) 

HIF1β ChiP-qPCR

fold enrichment

unmethylated
methylated

(e) 

HIF1α-HIF1β

fraction bound (%)

DNA conc (log_{10}[nM])

0 1 2 3

not methylated: 27.7 nM
methylated: 425.2 nM
not RCGTG: 2195.0 nM

(f) 

HIF2α-HIF1β

fraction bound (%)

DNA conc (log_{10}[nM])

0 1 2

not methylated: 35.4 nM
methylated: 543.5 nM
not RCGTG: 3918.0 nM

(g) 

5'-ACGTG

TGCAC-5'

R27

R102

HRE

HIF2α

HIF1β

(h) 

5'-ACGTG

TGCAC-5'

R27

R102

R27

R102

M. C

M. C

2.5 Å
Figure 3

(a) ESC WT ESC TKO ESC WT WGBS

(b) % methylation

(c) observed/expected

(d) genes upregulated upon hypoxia in TKO

(e) distance of HIF1β peak to nearest transcription start site

(f) distance of HIF1β peak to nearest region of open chromatin

(g) all HIF1β binding sites

(h) P value

(i) % of HIF1β binding sites

10,000 random permutations: actual HIF1β binding:
Figure 4

(a) 10,000 random permutations:
- shared peaks (n=11,546)
- aza-specific peaks (n=1,236)

(b) Methylation at HIF1β ChIP-seq summit ± 100 bp

(c) All HIF1β binding sites
- % of HIF1β peaks:
  - LINE
  - LTR
  - SINE
- 10,000 random permutations:
  - actual HIF1β binding:
    - shared peaks
    - aza-specific peaks

(d) Human genome
- LINE (n=158)
- LTR (n=169)
- SINE (n=100)
- SINE (n=79)
- MIR (n=79)

(e) Hypoxia versus normoxia (vehicle)
- Hypoxia versus normoxia (aza)

(f) HIF1B-WT
- RepEnrich
- REDENToR

(g) Non-HIF-bound
Figure 6

**a** - RepEnrich and REDENToR fold change vs vehicle

**b** - RepEnrich fold change vs vehicle

**c** - CD31+

**d** - pimonidazole

**e** - RepEnrich and REDENToR fold change

**f** - relative dsRNA signal intensity

**g** - Tumour weight (g)

**h** - Immune cell signature (aza vs control)

**i** - RepEnrich and REDENToR fold change vs vehicle

**j** - Tumour volume (mm$^3$)

**k** - Genotype x treatment interaction ANOVA P value < 0.001
**SUPPLEMENTARY FIGURE LEGENDS**

**Figure S1 | HIF1β peaks in MCF7 cells under 0.5% O₂**

(a) (top) Immunoblot analysis of HIF1α in MCF7 cells exposed for 24 hours to 21% or 0.5% O₂. (b) Comparison of the 7,153 HIF1β peaks detected in this study to a previously published dataset of ChIP-seq for HIF1β. (c) Frequency per basepair of the RCGTG motif inside and outside HIF1β peaks. On average, 4.6 HIF consensus sequences (RCGTG) are found in each HIF1β peak (average length 1,088 bp), whereas in the rest of the genome, 0.98 HIF consensus sequences (RCGTG) are found every 1,088 bp (Fisher exact test P<2.2×10⁻¹⁶). (d) Ontology analysis of genes associated with the 7,153 HIF1β peaks detected in MCF7 cells by ChIP-seq for HIF1β. (e) Immunoblot analysis of HIF1β in MCF7 cells that are wild-type (WT) or knockout (KO) for HIF1β. (f-g) Heatmaps of HIF1β binding for WT MCF7 cells and of DNA methylation for WT and HIF1β-KO MCF7 cells, at regions surrounding the HIF1β ChIP-seq peak summit in WT MCF7 cells (±5 kb) (f), and violin and boxplots of methylation at the HIF1β ChIP-seq peak summit (±100 bps) for WT and HIF1β-KO MCF7 cells (g). Depicted are the data from the 4,794 regions having >30-fold methylation coverage for both genotypes (WT and KO) in the 100 bps flanking the binding summit, as assessed using SeqCapEpi BS-seq. HIF1β binding was assessed after 16 hours of 0.5% O₂ (hypoxia), and DNA methylation under 21% O₂ (normoxia). (h) Immunoblot analysis of HIF1β in mESCs that are wild-type (WT) or knockout (KO) for Hif1b. (i, j) Heatmaps of HIF1β binding for ESCs and of DNA methylation for murine WT and Hif1b-KO ESCs, at regions surrounding the HIF1β ChIP-seq peak summit in WT ESCs (±5 kb) (i), and violin and boxplots of methylation at the HIF1β ChIP-seq peak summit (±100 bps) for WT and HIF1β-KO ESCs (j). Depicted are the data from the 1,644 regions having >30-fold methylation-coverage for both genotypes (WT and KO) in the 100 bps flanking the binding summit, as assessed using whole-genome BS-seq. HIF1β binding was assessed after 16 hours of 0.5% O₂ (hypoxia), and DNA methylation under 21% O₂ (normoxia). (k) Ontology analysis of genes associated with HIF1β peaks in 3 cell lines (SK-MEL-28, green; RCC4, yellow; MCF7, purple). (l) Cumulative frequency of distance to the nearest RCGTG motif, for HIF1β
DNA methylation repels HIF binding  

D’Anna et al

(91x744)

Page dimensions: 612.0x792.0

binding peak summits detected in 3 cell lines and for a random set of genomic positions. (m-n) Sequencing read depth of HIF1β ChIP and its input, at all RCGTG sequences in RCC4 (m) and SK-MEL-28 (n) cells, stratified for methylation at the CG in the core RCGTG sequence. Shown are data for all RCGTG sequences in the human genome for which >10× coverage was obtained after SeqCapEpi BS-seq.

Figure S2 | Expression of hypoxia genes and cancer testis antigens

(a) mRNA expression of hypoxia-induced genes in human and mouse cell lines and in 4T1 tumours. Shown is the average fold change of transcript per million over the appropriate control. (b) mRNA expression of cancer testis antigens in human cell lines, in 4T1 cells and mESCs exposed to hypoxia, DMOG, aza or deficient for the indicated gene (TKO for Dnmt). Shown are boxplots of log2 transcripts per million normalized over the proper control.

Figure S3 | Cell-type-specific HIF1β binding

(a) Patterns of DNA methylation and of the indicated chromatin modifications in MCF7 cells, at HIF1β binding sites detected in MCF7 cells at CpG islands (top), not at CpG islands (middle) or in SK-MEL-28 and/or RCC4 cells but not MCF7 (bottom). DNA methylation was determined using SeqCapEpi BS-seq or WGBS; occupancy of chromatin modifications was determined using the ENCODE project. (b) Estimates of the contribution of individual epigenetic marks in predicting the presence of HIF1β binding peaks, as determined by partial R² analysis of linear regression models. Shown are the results of a linear model using all available marks, a linear model using all available marks but not DNA methylation, and the results of linear models using each modification in an individual analysis. All 3 available FAIRE-seq datasets show similar results. (c) Functional genome annotation of MCF7 cells using ChromHMM, at shared and cell-type-specific HIF1β binding peaks. (d) Fraction of regions in MCF7 cells showing high versus low methylation in regions of open chromatin. 10,805 regions of open chromatin that showed high GpC methylation levels were assessed (>10-fold coverage) for their CpG methylation. Indicated are the fraction of these 10,805 regions showing at least 25%, 50% or 80% CpG...
methylation. GCG trinucleotides were not considered in these analyses. (e) Fraction of HIF1β binding peaks, shared between or unique for a cell line, that overlaps with other TF binding sites detected in MCF7 cells. Note how HIF1β binding peaks unique to MCF7 cells (purple) predominantly overlap with TFs uniquely expressed in MCF7 cells (TFASP2C, ESR1, ...), while HIF1β binding peaks shared between cell lines (dark grey) more frequently overlap with commonly expressed TFs (NR2F2, STAG1, ...). (f) mRNA expression level of indicated transcription factors in each cell line, as determined using RNA-seq. These data enable stratification of TFs into those that are shared in expression (light grey box) and those that are uniquely expressed (white box). (g) Fraction of HIF1β binding peaks, shared between or unique for a cell line, that overlaps with other transcription factor binding sites detected in A549 cells. Note how HIF1β binding peaks unique to A549 cells (blue) predominantly overlap with TFs expressed in A549 cells (GATA3 and FOXA1), while HIF1β binding peaks shared between cell lines (dark grey) more frequently overlap with binding of CTCF, a commonly expressed TF. (h) Comparison of the HIF1β peaks detected to a previously published dataset of ChIP-seq for HIF1β, HIF1α and HIF2α. (i) ChIP-seq signal for the indicated epitopes at the 200 bps surrounding the HIF1β peak summits that are uniquely co-bound by HIF1α or HIF2α. (j) Functional genome annotation of MCF7 cells using ChromHMM, at HIF1β binding peaks that are uniquely co-bound by HIF1α or HIF2α. (k) Fraction of HIF1β binding peaks, uniquely co-bound by HIF1α or HIF2α, that overlaps with other transcription factor binding sites detected in MCF7 cells.

Figure S4 | DNA methylation directly repels HIF1β binding
(a) Example plots of regions on chromosomes 6 and 16, showing HIF1β ChIP-seq read depths, and methylation levels as determined by whole-genome BS-seq and HIF1β ChIP-BS-seq in MCF7 cells. Coordinates: human genome build hg19. Grey boxes in HIF1β ChIP-BS-seq tracks indicate regions where insufficient sequences were recovered. (b) As in (a), but in murine Tet-TKO ESCs, and for regions on chromosomes 5 and 3. Coordinates follow genome build Mm10. In each example plot, HIF1β peaks (fragment per million) are shown on top; percentage of DNA methylation (calculated in bulk normoxic cells by WGBS) at
the same position of the HIF1β peaks is shown in the middle; percentage of DNA methylation of the immunoprecipitated DNA fragment from HIF1β ChIP-BS-seq is shown at the bottom. (c-d) Methylation levels of a human DNA fragment inserted in mESCs, as detected using amplicon bisulfite sequencing. A human HIF binding site (chr16: 30,065,212-30,065,711 on hg38) was cloned between 2 L1 Lox sites and in vitro methylated (d) or not (c) prior to insertion into mESCs transformed to contain an L1 Lox-flanked thymidine kinase.

**Figure S5 | Quality control of the ChIP-seq replicates**

(a) Smooth-scatter heatmaps for each experimental HIF1β ChIP-seq repeat, showing pairwise comparisons of the read depth ($\log_2$(RPKM+1)) at a peak set representing the union of peaks detected in WT and Dnmt-TKO cells. (b) Hierarchically clustered heatmap of Pearson correlations between the $\log_2$ read counts of all replicates for a peak set representing the union of all peaks in WT and Dnmt-TKO cells. Note how Pearson correlations between replicates of the same genetic background are higher than the correlations across the two genetic background. (c) A merged peak set (union of all peaks in WT and Dnmt-TKO cells) was generated, and the difference between the ChIP-seq signals ($\log_2$(RPKM+1) of these peaks of Dnmt-TKO and WT), was calculated for each of the 4 repeats. These repeats are here compared with a smooth-scatter heatmap. Also indicated are the associated Pearson Correlation Coefficients (PCC), as well as the number of peaks showing >2-fold change between both repeats. (d) Results of EdgeR analysis of differences in ChIP-seq signal strength between the 4 HIF1β ChIP-seq replicates. Shown are the significant differences between the 4 wild-type and 4 Dnmt-TKO (TKO, red) replicates. Dots colored in red and blue indicate peaks showing respectively decreased and increased binding in Dnmt-TKO cells versus WT cells. (e) Heatmaps showing the relative signal of the 4 HIF1β ChIP-seq data replicates, ranked by their mean fold change (quantified below). Shown are 4 wild-type (WT, blue) and 4 Dnmt-TKO (TKO, red) replicates.
Figure S6 I HIF binds retrotransposons in demethylated genomes

(a) % of shared or Dnmt-TKO-specific (TKO-specific) HIF1β binding peaks in ESCs that overlap retrotransposons, grouped by retrotransposon class and family. Families bound less than 3 times by HIF1β (Dong-R4, Jockey, Penelope, RTE-X, Gypsy, SS-Deu-L2, tRNA, tRNA-Deu) or containing unclassifiable subfamilies are not shown. The number of retrotransposon subfamilies involved is indicated in brackets. (b) HIF binding at long terminal repeats (LTRs, red) and internal sequences (blue). A solo-LTRs arise through non-allelic homologous recombination of the 5’ and 3’ LTRs (bottom). Show are (top) the HIF binding sites in murine ERVK LTRs that are at the 5’ or 3’ end of the ERVK-internal sequence, or that are solitary (“solo-LTR”). As can be appreciated, these have a similar distribution of HIF binding sites. (c) Fraction of HIF binding sites in 5’, 3’ or solo LTRs or in internal sequences, and this in in murine ES cells (left) and MCF7 cells (right). (d-e) Distribution of HIF binding at ERV internal sequences (all combined, blue) and at ERV1, ERVK, ERVL and ERVL-MaLR LTRs (red) in murine ES cells (d) and MCF7 cells (e). LTRs are grouped by subfamily, irrespective of whether they are 5’, 3’ or solo LTRs. As the total number of HIF binding sites in the internal sequence was too low for reliably generating density plots, we merged all subfamilies. (f) 5mC content of DNA from WT and HIF1B-KO MCF7 cells, exposed to 5-aza-dC (aza; 1 μM) or DMSO (control) for 3 cell cycles, as determined using LC/MS. (g) As in (a), but for HIF1β binding peaks that are shared between vehicle- and aza-treated MCF7 cells, or specific to aza-treated MCF7 cells. (h) Heatmap showing expression of 176 retrotransposon subfamilies bound by HIF1β in hypoxic and normoxic MCF7 cells exposed to vehicle (DMSO) or aza. Expression was determined using RepEnrich (n=6 per treatment condition). (i) Differential expression of 13 retrotransposon families bound by HIF1β. Boxplots show average read counts from RNA-seq analyses of vehicle- and aza-treated MCF7 cells exposed for 24 hours to normoxia or hypoxia (resp. 21% and 0.5% O₂), as determined using RepEnrich (n=6 per condition). Families not bound by HIF1β or containing unclassifiable subfamilies are not shown. (j) Volcano plots showing differential expression of retrotransposon loci, as determined by SQuIRE analysis of MCF7 cells exposed for 4 days to vehicle or aza (1 μM),
and for 1 day to hypoxia (0.5% oxygen) or normoxia. Significantly upregulated and
downregulated loci (1% FDR, horizontal line) are highlighted in red and blue, respectively,
with the number of loci being indicated. (k) as in (j) but for all cryptic transcripts, also
those not bound by HIF1β, using an analysis performed with CREDENToR. (l) HIF binding
sites detected at the TSS and transcription end site (TES) of cryptic transcripts
overexpressed (1% FDR) in MCF7 cells exposed to hypoxia and/or aza. Shown are
percentages of 250 bp bins wherein a HIF binding peak was detected. (m-n) Heatmap
showing the expression of 176 retrotransposon subfamilies bound by HIF1β in HIF1B-KO
MCF7 cells exposed to hypoxia (0.5% O₂) for 24 hours and treated with vehicle or aza (m)
and in WT (HIF1B-WT) MCF7 cells exposed to dimethylxalylglycine (DMOG; 2 mM) for 24
hours and treated with vehicle or aza (n). Expression was determined using RepEnrich
(n=6 per treatment condition). (o) Expression of HIF-bound cryptic transcripts
(CREDENToR, left) and retrotransposon subfamilies (RepEnrich, right) relative to vehicle-
treated control (vehicle vehicle) in MCF7 cells wild-type (WT) for HIF1B. *P<0.05,
***P<0.001 by t-test.

Figure S7 I Examples of cryptic transcripts upregulated by HIF1β
(a) Examples of HIF1β binding near cryptic transcripts (top 4 tracks) and the associated
induction of RNA expression (bottom 8 tracks). Shown are 3 regions of 30 kb surrounding
a HIF1β binding peak that is shared between vehicle-treated and aza-treated MCF7 cells
(top) and 3 30 kb regions surrounding a HIF1β binding peak that is specifically detected in
aza-treated MCF7 cells (bottom). Peaks of interest are indicated with an arrow, quantified
expression levels of the cryptic transcripts (RPM; upper right corner), and the associated
fold changes in expression (red arrow) are shown. As RNA-seq was directional, reads
mapping to the top strand (red) or the bottom strand (blue) are shown. These reads are
derived from the cDNA, and thus represent RNA transcripts derived from the bottom (red)
or top (blue) strand respectively. Overlapping reads from both strands are coloured in
dark grey. ChIP-seq was not directional. MCF7 cell genotypes (HIF1B-WT or -KO) and
treatments (aza or vehicle; hypoxia or normoxia) are indicated on the left. At least 3 of
the observed cryptic transcripts consist of multiple repeat elements and are thus characterized by read-through of the individual repeat. X-axis: chromosome positions are annotated in kb according to human genome build hg19; Y-axis: reads per kb per million reads. (b) Reproducibility of CREDENToR in determining transcript levels (left) and hypoxia-induced changes (right) in 6 replicates of MCF7 cells exposed to 1 day of hypoxia. PCC: Pearson Correlation Coefficient. (c) Comparison between repeat transcript expression estimates obtained by CREDENToR and SQuIRE. (d-g) Characterization of key features of the cryptic transcripts detected using CREDENToR, in comparison to annotated long intergenic non-coding RNAs (lincRNAs) and protein coding RNAs. Shown are the transcript lengths (d), the fraction of transcripts that are spliced (e), the RNA expression (f) and the conservation estimated using the average phastCon score per transcript (d). RPKM: reads per million reads and per kb.

**Figure S8 | Cryptic transcript expression in TCGA tumours**

(a) Recurrent detection of cryptic transcripts in different cancer types profiled in TCGA. Shown are the number of cryptic transcripts detected in 1 cancer type or in different cancer types, as per the legend. (b) Volcano plot showing the effect of the interaction between cryptic transcript DNA methylation and tumour oxygenation on the expression of individual cryptic transcripts, tested by DESeq. Positive coefficients represent the cooperative enhancement of cryptic transcript expression in low-methylation, hypoxic conditions. Highlighted in red and blue are significantly positive and negative coefficients (5% FDR) respectively. (c) Immunogenicity estimates for TCGA tumours responsive or non-responsive to immunotherapy as described by Turajlik and colleagues\(^37\) (red and white). Shown are number of somatic mutations extracted from the TCGA database, mRNA expression of PDL1, PD1 and LAG3 as log\(_2\) RPKM, cytolytic activity (CYT) defined as the log\(_2\)-average (geometric mean) of GZMA and PRF1 expression in RPKM, cell fraction for CD8\(^+\) T-cells estimated with respect to the total cells in the sample as defined by quanTIseq (The Cancer Immunome Atlas, TCIA\(^64\)), and cryptic transcript expression as log\(_{10}\) counts per million. **P<0.01 and ***P<0.001 by t-test. (d) The number of cryptic
transcripts expressed in cancer cells, T-cells and other stromal cells isolated from non-small cell lung tumours (left, n=5) or from breast tumours (right, n=22) and profiled using single-cell RNA-sequencing. Note that this expression dataset was generated using unique molecular identifiers at the 3’-end or 5’-end of poly-adenylated transcripts (resp. left and right panel), enabling the counting of transcripts per cell.

**Figure S9 | Aza treatment increases immunogenicity**

(a) Immunogenicity estimates in cell line grafted (4T1, B16, CT26 and MC38) mouse tumour models. Shown are TMB expressed as the number of mutations per Mb of expressed coding DNA sequence (CDS), mRNA expression (transcripts per million, TPM) of *Pdl1* and *Pd1* immune checkpoint molecules and activated CD8+ T-cell enrichment estimated through gene set variance analysis (GSVA) of immune metagenes. (b) Growth of tumours generated by grafting mice subcutaneously with CT26 and MC38, or orthotopically with 4T1 cells. Mice were treated with anti-PD1 or control IgG antibody (see Methods). Data represent estimated mean and s.e.m. from 3 independent experiments, each with at least n=6 per group. Arrows indicate treatment times. * P<0.05 by repeated measurement analysis. (c) Cryptic transcript expression (average counts per million + s.e.m.) in cell line grafted (4T1, B16, CT26 and MC38) mouse tumour models. (d) Heatmap (top) and violin plots (bottom) showing the expression of cryptic transcripts (bound by HIF1β in 4T1 cells) in 4T1 cells WT (*Hif1b*-WT) or deficient (*Hif1b*-KO) for *Hif1b* exposed to hypoxia for 24 hours (0.5% O2) and treated with vehicle or aza; expression of retrotransposon subfamilies using RepEnrich (top violin plot) and of cryptic transcripts was determined using CREDENToR (bottom violin plot) (at least n=5 per treatment condition). Difference in the distribution of cryptic transcript expression is denoted as fold change of counts per million over vehicle-treated and normoxic 4T1 cells. ***P<0.001 by paired t-test. (e) Heatmap showing the expression of a proliferation gene signature1 in 4T1 tumours from mice treated with either aza or PBS (at least 6 mice per treatment condition were sequenced). Samples and genes were unsupervisedly clustered. Aza-treated tumours do not appear to form a more proliferative subset of samples. (f)
Quantification of CD8+ and granzyme b (Gzmb)+ cells, depicted as percentages of CD45+ cells and CD8+ cells, from vehicle- or aza-treated 4T1 tumours bearing mice and representative immunofluorescence images (scale 50 µm). **P<0.01 by t-test. (g)

Immunoblot analysis (left) and mRNA expression (log2 fold change versus WT) (right) of HIF1β in polyclonal 4T1 cells that are WT (Hif1b-WT) or KO (Hif1b-KO) for Hif1b (see Methods). (h) Violin plot showing expression of cryptic transcripts unbound by HIF1β in 4T1 cells. Shown is expression as determined using CREDENToR in 4T1 cells WT (Hif1b-WT) or deficient (Hif1b-KO) for Hif1b exposed to hypoxia for 24 hours (0.5% O2) and treated with vehicle or aza. (i) Expression of cryptic transcripts significantly downregulated in vitro (1% FDR) in hypoxic Hif1b-KO 4T1 cells versus hypoxic WT 4T1 cells, and stratified into being HIF-bound and non-HIF-bound. Expression changes are shown for 4T1 cells grown in vitro (left panel), and for 4T1 tumours orthotopically grafted in mice (right panel, in vivo). ***P<0.001; **P<0.01; *P<0.05 by t-test.
**Figure S1**

(a) Q<sub>3</sub>(%) of HIF1α and α-Tubulin.
(b) Distribution of distance from HIF1β binding peak (kb) for Schödel, HIF1β, isotype SeqCapEpi, and input-seq.
(c) Cumulative frequency of distance to RCGTG (bp) and reads per million at RCGTG motifs.
(d) Enrichment of different cellular responses and processes.
(e) Western blots showing HIF1B WT and KO, HIF1β, and β-actin.
(f) Heat maps of % CpG methylation in HIF1B-WT and HIF1B-KO.
(g) % CpG methylation of RCGTG motifs in HIF1B-WT and HIF1B-KO.
(h) Western blots showing Hif1b WT and KO, HIF1β, and α-Tubulin.
(i) Heat maps of % CpG methylation in Hif1b-WT and Hif1b-KO.
(j) % CpG methylation of RCGTG motifs in Hif1b-WT and Hif1b-KO.
(k) Cumulative frequency of enrichment for different cellular responses.
(l) Cumulative frequency of distance to RCGTG (bp) for SK-MEL-28.
(m) % methylation of RCGTG motifs in RCC4.
(n) % methylation of RCGTG motifs in SK-MEL-28.
Figure S2

a) Treatment effects on HIF1B, EGLN3, BNIP3, ALDOA, and CA9 expression in hypoxic conditions.

b) Cancer testis antigen expression (log2 fold change over vehicle control or HIF1B-WT) under different treatments.
Figure S5

(a) Scatter plots showing the relationship between Dnmt-TKO and Dnmt-WT in different conditions.

(b) Heatmap representing Pearson correlation coefficients between Dnmt-TKO and Dnmt-WT.

(c) Scatter plots with significance levels (PCC) for different conditions.

(d) Heatmap showing HIF binding in Dnmt-TKO versus Dnmt-WT.

(e) Heatmap with Z scores for fold change in expression levels.
Figure S8

(a) Cryptic transcript number (x1,000)

- Specific to 1 cancer type
- Shared by 2 cancer types
- Shared by 3-5 cancer types
- Shared by 6-13 cancer types
- Shared by all 14 cancer types

(b) log2 coefficient (hypoxic versus normoxic)

(c) Log2(mutation number) versus log2(RPKM)

- CTLA4
- PDL1

(d) Transcripts per cell

- PD1
- LAG3
- CD8+T cell

- Non-responding to immunotherapy
- Responding to immunotherapy

- Cancer cell
- T-cell
- Other stromal cell
**SUPPLEMENTARY TABLES**

Supplementary Table 1 | HIF1β binding peaks detected using MACS at $P<10^{-15}$ in MCF7 cell line

Supplementary Table 2 | HIF1β binding peaks detected using MACS at $P<10^{-15}$ across RCC4, MCF7 and SK-MEL-28 cell lines

For each cell line, HIF1β binding was annotated as ‘present’ if the peak area showed >4-fold enrichment over the local read depth, and as ‘absent’ if it showed <2.5-fold enrichment; intermediate enrichment scores were annotated as ‘unclassified’.

Supplementary Table 3 | Cryptic transcript promoter DNA methylation and expression levels

Shown are results from an analysis of combined single-cell methylome-and-transcriptome sequencing of colorectal cancer cells, as generated by Bian and colleagues.36