1   **ZHX2 Promotes HIF1α Oncogenic Signaling in Triple-Negative Breast Cancer**

2   Wentong Fang,[1,2]* Chengheng Liao, [3]* Rachel Shi, [3] Jeremy M. Simon, [2,4,5] Travis S.

3   Ptacek, [2,5] Giada Zurlo, [3] Youqiong Ye, [6,7] Leng Han, [7] Cheng Fan, [2] Christopher Llynard

4   Ortiz, [8,9,10] Hong-Rui Lin, [8] Ujjawal Manocha, [2] Weibo Luo, [3] William Y. Kim, [2] Lee-Wei Yang,

5   [8,9,11] and Qing Zhang[3]*

6

7   [1] Department of Pharmacy, The First Affiliated Hospital of Nanjing Medical University,

8   Nanjing, Jiangsu 210029, China

9   [2] Lineberger Comprehensive Cancer Center, University of North Carolina School of

10   Medicine, Chapel Hill, NC 27599, USA

11   [3] Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX

12   75390, USA

13   [4] Department of Genetics, Neuroscience Center, University of North Carolina, Chapel Hill,

14   NC 27599, USA

15   [5] UNC Neuroscience Center, Carolina Institute for Developmental Disabilities, University of

16   North Carolina, Chapel Hill, NC 27599, USA

17   [6] Shanghai Institute of Immunology, Faculty of Basic Medicine, Shanghai Jiao Tong

18   University School of Medicine, Shanghai, 200025, China

19   [7] Department of Biochemistry and Molecular Biology, The University of Texas Health

20   Science Center at Houston McGovern Medical School, Houston, TX, 77030, USA

21   [8] Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu

22   300, Taiwan

23   [9] Chemical Biology and Molecular Biophysics Program, Taiwan International Graduate

24   Program, Institute of Chemistry, Academia Sinica, Taipei 115, Taiwan

25   [10] Department of Chemistry, National Tsing-Hua University, Hsinchu 30013, Taiwan

26   [11] Physics Division, National Center for Theoretical Sciences, Hsinchu 30013, Taiwan

27

28   *: These authors contribute equally to this work

29   Corresponding author (Q.Z): Qing.Zhang@UTSouthwestern.edu

30

31

Abstract

Triple-negative breast cancer (TNBC) is an aggressive and highly lethal disease, which warrants the critical need to identify new therapeutic targets. We show that Zinc Fingers And Homeoboxes 2 (ZHX2) is amplified or overexpressed in TNBC cell lines and patients. Functionally, depletion of ZHX2 inhibited TNBC cell growth and invasion in vitro, orthotopic tumor growth and spontaneous lung metastasis in vivo. Mechanistically, ZHX2 bound with hypoxia inducible factor (HIF) family members and positively regulated HIF1α activity in TNBC. Integrated ChIP-Seq and gene expression profiling demonstrated that ZHX2 co-occupied with HIF1α on transcriptionally active promoters marked by H3K4me3 and H3K27ac, thereby promoting gene expression. Furthermore, multiple residues (R491, R581 and R674) on ZHX2 are important in regulating its phenotype, which correspond with their roles on controlling HIF1α activity in TNBC cells. These studies establish that ZHX2 activates oncogenic HIF1α signaling, therefore serving as a potential therapeutic target for TNBC.

Key words: ZHX2, TNBC, HIF1α, VHL

Introduction

Triple-negative breast cancer (TNBC) accounts for 15-20% of all breast cancer (Anders & Carey, 2009). TNBC is associated with a more aggressive clinical history, a higher likelihood of distant metastasis, shorter survival and a higher mortality rate compared to other subtypes of breast cancer (Anders & Carey, 2009). In addition, recent studies illustrate high rates of brain metastasis in TNBC that is associated with poor survival (Heitz *et al*, 2009; Lin *et al*, 2008; Niwinska *et al*, 2010). Since TNBCs do not express estrogen receptor (ER), progesterone receptor (PR) or human epidermal growth factor receptor 2 (HER2), treatment options have historically been limited to chemotherapy(Masui *et al*, 2013), which has significant toxicity and a suboptimal impact on the five-year relapse rate. Therefore, it is critical to identify novel therapeutic targets in TNBC.

The Zinc-fingers and homeoboxes (ZHX) family includes ZHX1, 2 and 3. ZHX2 is located on 8q24.13 and contains two zinc finger domains and five homeodomains (HDs) (Kawata *et al*, 2003). In addition, between amino acids 408-488, it contains a proline rich region (PRR). ZHX2 can form homodimers or can hetero-dimerize with the other two family members ZHX1 or ZHX3 (Kawata *et al.*, 2003). Originally, ZHX2 was found to be a key transcriptional repressor for the alpha-fetoprotein regulator 1 (*Afr1*) (Perincheri *et al*, 2005), which is an important oncogene in liver cancer. From this perspective, ZHX2 was identified and reported to function as a transcriptional repressor (Kawata *et al.*, 2003), where fusion of ZHX2 with a GAL4-DNA binding domain repressed transcription of a GAL4-dependent luciferase reporter. Additionally, ZHX2 was reported to have tumor suppressor activity in hepatocellular carcinoma (HCC), by repressing cyclin A, E and multidrug resistance 1 (MDR1) expression (Ma *et al*, 2015; Yue *et al*, 2012). ZHX2 was also indicated to be a tumor suppressor in Hodgkin lymphoma or myeloma although there is no direct

80 experimental evidence supporting this hypothesis (Armellini *et al*, 2008; Nagel *et al*, 2012;

81 Nagel *et al*, 2011).

82   However, accumulating evidence suggests that ZHX2 may contribute to cancer

83 pathology in other contexts. Tissue microarray and clinicopathological analysis show that

84 ZHX2 protein expression in metastatic HCC is twice as high as in the primary lesions,

85 indicating that ZHX2 expression is associated with metastasis in HCC (Hu *et al*, 2007). In

86 addition, our recent findings through genome-wide screening identify that ZHX2 is a

87 substrate of von Hippel Lindau (gene name *VHL*, protein name pVHL) protein, accumulates

88 in kidney cancer, and promotes oncogenic signaling by at least partially activating NF-κB

89 signaling in clear cell renal cell carcinoma (ccRCC) (Zhang *et al*, 2018). These pieces of

90 evidence suggest that ZHX2 acts as a tumor suppressor or oncogene in a context-

91 dependent manner. It is also important to point out that ZHX2 is located on 8q24, a genomic

92 region that is frequently amplified in various cancers including breast cancer (Guan *et al*,

93 2007). More importantly, the role of ZHX2 in other cancers, such as in TNBC, remains

94 largely unknown.

95   Tumor hypoxia is a characteristic of most solid tumors. Hypoxic cells are known to

96 confer radio- or chemotherapeutic resistance, and therefore are hypothesized to undergo

97 positive selection during cancer development (Brown & Wilson, 2004; Gray *et al*, 1953). The

98 key proteins mediating oxygen sensing in these cells involve two classes of proteins: (1)

99 upstream oxygen sensors, namely the prolyl hydroxylases EglN1-3, responsible for the

100 hydroxylation of various substrates, such as hypoxia inducible factor (HIF), FOXO3a, ADSL,

101 SFMBT1 and TBK1 (Hu *et al*, 2020; Liu *et al*, 2020; Semenza, 2012; Zheng *et al*, 2014;

102 Zurlo *et al*, 2019); (2) the downstream VHL E3 ligase complex. For example, EglN family

103 members (EglN1, primarily *in vivo*) hydroxylate HIF1α on proline 402 and 564 positions,

104 which lead to pVHL binding and HIF1α ubiquitination and degradation (Appelhoff *et al*,

105 2004; Ivan *et al*, 2001; Jaakkola *et al*, 2001). HIF1α has been well established to be an

106 important oncogene in multiple cancers, including breast cancer (Briggs *et al*, 2016;

107 Semenza, 2010). Tumor hypoxia or pVHL loss will lead to the accumulation of HIF1α. As a

108 result of the accumulation and translocation of HIFα factors into the nucleus, HIF1α

109 dimerizes with a constitutively expressed HIF1β subunit (also called ARNT) and

110 transactivates genes that have hypoxia response elements (NCGTG) in promoters or

111 enhancer regions. HIF1-transactivated genes include those involved in angiogenesis (e.g.

112 *VEGF*), glycolysis and glucose transport (e.g. *GLUT1*), and erythropoiesis (e.g. *EPO*)

113 (Semenza, 2012). Besides tumor hypoxia or pVHL loss, other potential regulators of HIF1α

114 may exist, which remains to be investigated.

115   In our current study, we investigated the role of ZHX2 as a new oncogene in TNBC,

116 where it activates HIF1α signaling. In addition, we also provide some evidence for critical

117 residues on ZHX2 that binds with DNA, which contributes to TNBC tumorigenicity.

118

119 **Results**

120 **ZHX2 is Amplified in TNBC and is Potentially Regulated by pVHL**

121 *ZHX2* is located on 8q24, where c-*Myc* resides. Analysis of copy number across different

122 cancer types from The Cancer Genome Atlas (TCGA) showed that *ZHX2* is amplified in

123 various cancers, including ovarian cancer (~40%) and breast cancer (~15%) (Figure 1A).

124 Importantly, *ZHX2* and c-*Myc* share co-amplification in most cancer types observed (Figure

125 1B). Interestingly, *ZHX2* is not amplified in ccRCC (Referred as KIRC in Figure 1A, B),

126 where it is regulated mainly post-transcriptionally by pVHL loss in this cancer (Zhang *et al.*,

127 2018). Detailed analyses of several breast cancer patient datasets also revealed that TNBC

128 had the highest ZHX2 amplification rate in all breast cancer subtypes (Figure 1C; Table

129   Supplement 1). Next, we performed correlation studies to examine the copy number gain of

130   *ZHX2* and its expression in TCGA (Cancer Genome Atlas, 2012) and METABRIC datasets

131   (Curtis *et al*, 2012a) (Figure 1D). We found a significant correlation between copy number

132   and gene expression (Figure 1D), suggesting that *ZHX2* amplification may be at least

133   partially responsible for its overexpression in breast cancer. We further explored the effect

134   of ZHX2 expression on breast cancer patient survival. ZHX2 overexpression (Affymetrix

135   probe 236169_at) correlates with worse survival in TNBC but not in other breast cancer

136   subtypes (Figure 1-figure supplement 1A). We obtained a panel of breast cancer cell lines

137   as well as two immortalized normal breast epithelial cell lines, HMLE and MCF-10A.

138   Interestingly, all breast cancer cell lines displayed relatively higher ZHX2 protein levels

139   compared to HMLE or MCF-10A (Figure 1E). We also obtained 10 pairs of TNBC patient

140   tumors and paired normal tissue. Consistently, ZHX2 was upregulated in the majority of

141   tumors compared to normal (7 out of 10) (Figure 1F).

142   Our previous research established an oncogenic role of ZHX2 in ccRCC as a pVHL

143   substrate (Zhang *et al.*, 2018). However, the role of ZHX2 in other cancers remains largely

144   unclear. In addition, it is unclear whether ZHX2 may also act as a pVHL target in breast

145   cancer. Since TNBC had the highest ZHX2 amplification rate in all breast cancer subtypes

146   (Figure 1C; Table supplement 1), we decided to focus on TNBC for this current study. To

147   this purpose, we first examined the relationship between the expression of ZHX2 and pVHL

148   in several TNBC cell lines (MDA-MB-231, MDA-MB-436, MDA-MB-468, HCC3153 and

149   HCC70) as well as two normal breast epithelial cell lines, HMLE and MCF-10A.

150   Interestingly, all TNBC cell lines displayed relatively lower pVHL protein levels

151   corresponding with higher ZHX2 protein level when compared to normal breast cells (Figure

152   1-figure supplement 1B). Our co-immunoprecipitation (Co-IP) experiments showed that

153   ZHX2 interacts with pVHL in two representative TNBC cell lines (Figure 1G; Figure 1-figure

154   supplement 1C). Next, we aimed to examine whether pVHL can promote the degradation of

155   ZHX2, therefore decreasing ZHX2 protein levels in these cells. First, we overexpressed HA-

156   VHL in two different TNBC cell lines and found that pVHL overexpression leads to

157   decreased ZHX2 protein levels (Figure 1H; Figure 1-figure supplement 1D). Conversely, we

158   also deleted pVHL expression by three independent sgRNAs (#1, 2, 8) in these two cell

159   lines and found that pVHL depletion led to upregulation of ZHX2 protein (Figure 1I and J).

160   Our previous research showed that ZHX2 regulation by pVHL potentially depends on ZHX2

161   prolyl hydroxylation (Zhang *et al.*, 2018). We treated these two cell lines with hypoxia, the

162   pan-prolyl hydroxylase inhibitor DMOG, or the proteasomal inhibitor MG132 and found that

163   ZHX2 was upregulated by these inhibitors (Figure 1K and L), further strengthening the

164   conclusion that ZHX2 is regulated by pVHL for protein stability through potential prolyl

165   hydroxylation in breast cancer. In addition, we obtained 10 additional pairs of TNBC tumors

166   and paired normal tissue to analyze ZHX2 and pVHL protein levels. In accordance with the

167   cell line data, we found that ZHX2 was upregulated in most of tumor tissues compared to

168   normal, coinciding with decreased pVHL protein levels in respective tumor tissues (Figure

169   1M). Our data suggest that ZHX2 may be regulated by pVHL and play an important role in

170   TNBC.

171

172   **ZHX2 is Essential for TNBC Cell Proliferation and Invasion**

173   Next, we examined the potential role of ZHX2 in TNBC cell proliferation and invasion. First,

174   we obtained two previously validated ZHX2 shRNAs (sh43, sh45) (Zhang *et al.*, 2018) and

175   these shRNAs led to efficient ZHX2 protein (Figure 2A) and mRNA (Figure 2B)

176   downregulation in both MDA-MB-231 and MDA-MB-468 cells. Next, we found that ZHX2

177   depletion led to decreased cell proliferation in both cell lines as a function of time, 2-D

178   colony formation as well as 3-D soft agar growth (Figure 2C-F). One important contributor

179   for the poor prognosis in TNBC is the aggressively invasive nature of this subtype.

180   Therefore, we also used the Boyden chamber assay to examine the effect of ZHX2 on cell

181   invasion in TNBC cells. Consistent with the results above, ZHX2 depletion led to decreased

182   cell invasion in several different TNBC cell lines (Figure 2G and H). We also showed that

183   ZHX2 shRNA induced phenotypes in TNBC cells could be completely rescued by shRNA-

184   resistant ZHX2 (Figure 2I-N; Figure 2-figure supplement 2A-G), suggesting that these

185   phenotypes were due to on-target depletion of ZHX2 by shRNAs.

186

187   **ZHX2 is Important for TNBC Cell Proliferation *in Vivo***

188   Next, to examine the ability of ZHX2 to maintain TNBC tumor growth *in vitro* and *in vivo*, we

189   first infected two TNBC cell lines with doxycycline inducible ZHX2 shRNAs (Teton sh43,

190   sh45). Upon doxycycline addition, we achieved efficient depletion of ZHX2 protein levels,

191   corresponding to decreased ZHX2 mRNA levels (Figure 3A; Figure 3-figure supplement 3A

192   and B) in the cells. Next, MTS assays showed that ZHX2 depletion upon doxycycline

193   addition led to decreased TNBC cell proliferation, 2-D growth, 3-D anchorage independent

194   growth (Figure 3B-E; Figure 3-figure supplement 3C-F). Next, we expressed firefly

195   luciferase in the doxycycline-inducible ZHX2 shRNA (sh43 or sh45) or control shRNA cells

196   before injecting these cells orthotopically into the 4th mammary fat pads of NoD SCID

197   Gamma (NSG)-deficient mice. We performed weekly bioluminescence imaging to ensure

198   the successful implantation and growth of TNBC tumor cells. After 12 days post-implantation

199   when palpable tumors were formed, we fed these mice doxycycline chow. We found that

200   ZHX2 depletion by both hairpins significantly decreased tumor growth overtime. Upon

201   necropsy, ZHX2 shRNA-infected TNBC cells displayed reduced tumor burden retrieved from

202   tumor-bearing mice compared to control mice (Figure 3F and G). We also performed a

203   western blot for tumors extracted from the mammary fat pad and found decreased ZHX2

204   protein levels in ZHX2 shRNA-infected groups (Figure 3H), arguing that anti-tumor effect in

205   these groups may result from efficient ZHX2 knockdown in these TNBC cells. In addition,

206   we also measured spontaneous lung metastasis *ex vivo* upon necropsy and found that

207   ZHX2 depletion led to significantly decreased lung metastasis in TNBC (Figure 3I and J).

208   Lastly, we also injected ZHX2 Teton sh45-infected cells into the mammary fat pad, followed

209   by regular chow or doxycycline chow. Consistent with results above, doxycycline chow led

210   to significantly decreased TNBC tumor growth over time (Figure 3K and L) corresponding

211   with lower ZHX2 protein levels in these tumors (Figure 3M). Taken together, our data

212   strongly indicate that ZHX2 is important for TNBC tumorigenesis *in vivo*.

213

214   **ZHX2 Regulates HIF Signaling in TNBC**

215   Next, we aimed to determine the molecular mechanism by which ZHX2 contributes to

216   TNBC. For this purpose, we performed RNA-seq analyses in MDA-MB-231 cells with two

217   independent ZHX2 shRNAs (sh43, sh45) and found that two individual shRNAs

218   concordantly altered gene expression patterns (Figure 4A). Pathway enrichment as well as

219   gene set enrichment analyses (GSEA) revealed that genes differentially expressed following

220   ZHX2 depletion were enriched for members of the hypoxia pathway (Figure. 4B; Figure 4-

221   figure supplement 4A-C). Overall, there were 7,690 genes differentially expressed following

222   ZHX2 silencing by shRNA, and 1,849 (24%) of these genes overlapped with genes

223   differentially expressed in a previously published HIF double knockout (HIF1α and HIF2α

224   double knockout, HIF DKO) model (Figure 4C) (Chen *et al*, 2018). Since ZHX2 was similarly

225   reported as an oncogene in ccRCC and preferentially upregulated the transcription of

226   downstream genes (Zhang *et al*., 2018), we focused on the 3,969 ZHX2 positively regulated

227   genes (downregulated following ZHX2 silencing by shRNA) as these genes may be more

228   relevant in breast cancer. Of these, 678 genes (17%) overlapped with downregulated genes

229   in the HIF DKO RNA-seq (Figure 4B-D), representing a significant association (adj. p = 1.07

230   x 10^{-69}). This data strengthens the potential functional link between HIF and ZHX2. The

231   genes positively regulated by ZHX2 were also enriched for other potentially relevant

232   biological pathways such as cell adhesion and cell morphogenesis (Figure 4B). Next, we

233   examined whether ZHX2 depletion affected some of the canonical HIF target genes. Based

234   on our ZHX2 RNA-seq and GSEA results (Figure 4-figure supplement 4C), we chose a few

235   representative HIF target genes and examined their expression levels by qRT-PCR in MDA-

236   MB-231 cells infected with control or ZHX2 shRNA under either normoxia or hypoxia. As

237   expected, hypoxia treatment increased the expression of HIF target genes, and this effect

238   was ameliorated by ZHX2 depletion (Figure 4-figure supplement 4D), suggesting that ZHX2

239   affects HIF activity and regulates HIF target gene expression in TNBC. Consistently, HIF

240   reporter assay found that ZHX2 depletion led to decreased HIF reporter activity either under

241   normoxia or hypoxia (Figure 4-figure supplement 4E).

242   Next, to gain further insight on how ZHX2 affect the HIF signaling. Consider a well

243   characterization on HIF1 function in TNBC tumorigenesis by previous studies (Bos *et al*,

244   2001; Briggs *et al.*, 2016). We first performed Co-IP experiments and showed that ZHX2

245   bound with HIF1α and HIF1β exogenously as well as endogenously (Figure 4E-G). In

246   addition, ZHX2 knockdown led to decreased HIF1α protein levels in two TNBC cell lines

247   under hypoxia condition (Figure 4H and I). Conversely, ZHX2 overexpression led to

248   increased HIF1α protein levels under normoxia (Figure 4J). However, these ZHX2 loss-of-

249   function or gain-of-function manipulations did not grossly change HIF2α protein level (Figure

250   4I and J). Co-IP experiments showed that ZHX2 could not bind with HIF2α (Figure 4K).

251   These results suggest that ZHX2 regulates the hypoxia signaling mainly through HIF1α.

252   Interestingly, qRT-PCR showed that ZHX2 overexpression did not increase HIF1α mRNA

253   level (Figure 4L), suggesting a mechanism of post-transcriptional regulation on HIF1α We

254   then treated the ZHX2 knockdown cells with the proteasome inhibitor MG132 under

255   hypoxia, which could fully rescue the HIF1α protein level in the knockdown cells (Figure

256   4M). Altogether, these data suggested ZHX2 controls HIF1α protein stability by preventing

257   proteasome-mediated degradation. On the other hand, western blot analysis showed that

258   HIF1α knockdown did not affect ZHX2 protein levels (Figure 4-figure supplement 4F). It is

259   important to note that the detailed mechanism on how ZHX2 binds with HIF1 and

260   transactivates HIF1 signaling remains unclear, which awaits future investigation.

261   To identify the direct downstream target genes of ZHX2 and HIF that may be

262   important in TNBC, we performed chromatin immunoprecipitation followed by high-

263   throughput sequencing (ChIP-seq) to assess the genomic binding pattern of ZHX2 in TNBC.

264   We identified 957 binding sites across the genome, of which 94% of them overlap with

265   H3K27ac and 96% of them overlap with H3K4me3 (Figure 5A), indicating these overlapping

266   peaks bound preferentially to active promoters for gene expression (Shlyueva *et al*, 2014).

267   We again focused on the ZHX2 positively regulated genes (downregulated following ZHX2

268   silencing by shRNA) that exhibited ZHX2 binding in the promoter (transcription start site ± 5

269   kb). These promoters (n = 258) demonstrated robust enrichment for HIF1α (Chen *et al.*,

270   2018), as well as H3K4me3 and H3K27ac (Rhie *et al*, 2014) (Figure 5A). We then filtered

271   these genes further, focusing on only those bound and positively regulated by HIF1α and

272   identified seven interesting candidate genes for functional validation and follow-up

273   (*PTGES3L*, *KDM3A*, *WSB1*, *AP2B1*, *OXSR1*, *RUNDC1* and *COX20*). We performed qRT-

274   PCR analysis and found that ZHX2 depletion by shRNAs indeed led to decreased mRNA

275   expression for all of these target genes (Figure 5B). Conversely, we also overexpressed

276   ZHX2 in TNBC cells and found that ZHX2 overexpression led to increased expression of

277   these target genes (Figure 5C), arguing that ZHX2 promotes HIF signaling by at least

278   directly activating these targets in TNBC. To further strengthen whether these are HIF

279   downstream target genes, we also obtained HIF DKO cells and found that HIF depletion led

280   to decreased expression of these target genes under hypoxic conditions (Figure 5D). Taken

281   together, our data suggests that ZHX2 binds with HIF and affects HIF activity in TNBC.

282

283   **Potential Important Sites on ZHX2 that May affect its Function in TNBC**

284   Previous research showed that ZHX2 contributed to ccRCC tumorigenesis by at least

285   partially activating NF-κB signaling (Zhang *et al.*, 2018). In that setting, ZHX2 may act as

286   transcriptional activator by overlapping primarily with H3K4me3 and H3K27Ac epigenetic

287   marks. However, it remains unclear whether there may be critical residues on ZHX2 that

288   may mediate its binding to DNA and exert its transcriptional activity. To this end, we

289   conducted both data-driven analyses and structural simulations to predict the essential DNA

290   binding residues. Detailed methods are described in Supplementary. Briefly, many DNA-

291   bound structures of human HD2, 3 and 4 (PDB ID: 3NAU, 2DMP and 3NAR) were

292   homology-modeled by SWISS-MODEL (23) using as many X-ray/NMR-solved homologous

293   DNA-bound HD proteins as the structural templates. 12, 20 and 15 DNA-bound complexes

294   were modeled for HD2, HD3 and HD4, respectively (Figure 5-figure supplement 5). We then

295   counted the number of DNA-protein contacts at the atomic level for each residue in the

296   DNA-contacting helices in HDs, normalized by the number of DNA-complexed structures

297   used for each HD. The top-ranked DNA-contacting residues in HDs are listed together with

298   their evolutionary conservation in Table supplement 2. We found that Lys485/Arg491 in HD2,

299   Arg581 in HD3 and Arg674 in HD4 are the most contacted residues in DNA binding, where

300   Arg674 receives the highest contact among all the HD proteins. Our MD simulations further

301   revealed that Arg491 in HD2, Arg581 in HD3 and Arg674 in HD4 indeed have the highest

302   affinity with DNA, in terms of MM/PBSA-derived contact potential energies (see Table

303   supplement 3 and 4; Movie Supplement1, 2 and 3), among other residues in the same

304 proteins. Among the top 4 DNA-contacting residues in each of the HD proteins, we consider

305 those with relatively high sequence conservation being important for ZHX2 binding to DNA,

306 which may affect the phenotype of TNBC. These residues are Asp489 (D489), Arg491

307 (R491), Glu579 (E579), Arg581 (R581), Lys582 (K582), Arg674 (R674), Glu678 (E678), and

308 Arg680 (R680) (Table Supplement 2).

309   Given this, we generated a series of TNBC breast cancer cell lines where we

310 depleted endogenous ZHX2 expression and restored with exogenous shRNA-resistant

311 ZHX2 WT or mutant versions (D489A, R491A, E579A, R581A, K582, R674A, E678A, or

312 R680A). First, upon generation of stable cell lines, we performed western blot analyses and

313 confirmed that these cell lines all expressed similar amounts of ZHX2, relatively comparable

314 to endogenous ZHX2 levels in MDA-MB-231 cells (Figure 6A). Next, we performed 2-D cell

315 proliferation MTS assays. Our cell proliferation data showed that consistent with our

316 previous results, ZHX2 depletion led to decreased TNBC cell proliferation, and this

317 phenotype was completely rescued by WT ZHX2. On the other hand, some of mutants

318 (including R581A and R674A) failed to rescue the cell growth defect in ZHX2 shRNA

319 infected MDA-MB-231 cells (Figure 6B).

320   Motivated by our cell proliferation assay results, we further examined cell

321 proliferation phenotypes using long-term 2-D colony formation, 3-D anchorage independent

322 growth and cell invasion assays. ZHX2 R491A, R581A and R674A mutants displayed a

323 defect in cell growth in TNBC cell lines (Figure 6C-F). In summary, our results showed that

324 there may be multiple residues (including R491, R581 and R674) that may be important in

325 regulating the phenotype of ZHX2 in TNBC. Co-IP analysis showed that mutated ZHX2

326 (R581A and R674A) can interact with HIF1α as wild type ZHX2 (Figure 6G). While western

327 blot analysis showed that the ZHX2 mutations did not affect the HIF1α protein levels

328 (Figure 6H), qRT-PCR showed that the ZHX2 mutations indeed led to decreased mRNA

329 expression of HIF1α targeted genes (Figure 6I). Taken together, our data suggests that

330 these residues are essential for mediating the transactivation of ZHX2 on HIF1α target

331 genes.

332

333 **Discussion**

334 In this study, we discover that ZHX2 is an important oncogene in TNBC. Depletion of ZHX2

335 leads to decreased TNBC cell proliferation as well as invasion. By performing gene

336 expression analyses, ZHX2-regulated genes display a significant overlap with HIF-regulated

337 genes in TNBC. ZHX2 binds with HIF1α and HIF1β and regulates HIF1α protein levels and

338 transcriptional activity. By using structural simulation and the re-constitution system, we

339 pinpoint residues (R491, R581 and R674) on ZHX2 that may be important for its DNA

340 binding function as well as tumorigenic potential. Overall, our study establishes an important

341 role of ZHX2 in regulating HIF1α signaling and tumorigenesis in TNBC.

342 From a genomic perspective, it is known that *ZHX2* is located on 8q24, a

343 chromosomal region frequently amplified in cancers. Indeed, *ZHX2* is amplified in several

344 cancers, including breast cancer, ovarian cancer, and prostate cancer. In most cases, ZHX2

345 is co-amplified with another well-established oncogene c-*Myc* (Figure 1A-C). This finding

346 bears several implications. First, it suggests that ZHX2 and c-Myc may act in concert in

347 promoting tumorigenesis. Second, the role of ZHX2 in cancers can be context dependent.

348 Although ZHX2 may be amplified in multiple cancers, its protein levels can be regulated

349 post-transcriptionally. Our previous research showed that in ccRCC, ZHX2 can be regulated

350 by pVHL potentially through hydroxylation on multiple proline residues in the PRR domain

351 (Zhang *et al.*, 2018). Therefore, the presence or absence of factors that mediate prolyl

352 hydroxylation in the same niche as ZHX2 can dictate its regulation and downstream

353 function. Further, it remains uncertain whether ZHX2 interacts with DNA directly or indirectly

354 via other transcription factors to exert its transcriptional regulation on downstream target

355 genes. The repertoire of different co-activators/repressors with which ZHX2 may interact

356 can therefore govern its localization in the genome and thus its downstream function in

357 different cancer settings.

358 Thus far, it remains unclear which residues on ZHX2 are critical for its transcriptional

359 activity as well as its oncogenic role in cancer. We also did MD simulation to confirm the

360 import residues for regulating the phenotype of ZHX2. We found that the top DNA-

361 contacting residues across the three HDs were arginine namely Arg491, Arg581, and

362 Arg674. In terms of interaction, it can be observed that Arg491 and Arg581 of HD 2 and 3,

363 respectively, were seen to be interacting with the DNA's phosphate backbone throughout

364 the 350ns MD simulations (Movie supplement 1-2). However, in the case of HD 4, Arg674

365 was seen to be interacting first with the nucleobases within 4Å distance. Shortly after

366 ~120ns, the DNA slowly drifted away but this was prevented because of an interaction

367 between Arg674 and the DNA's phosphate backbone (Movie supplement 3). These

368 arginine-DNA phosphate backbone interactions could be critical for the stabilization of

369 transcription factor binding with the DNA to either stimulate or repress the transcription of a

370 specific gene. By performing ZHX2 depletion and reconstitution experiments, we found that

371 several residues located in the HDs of ZHX2 may be important for its oncogenic role in

372 TNBC (Figure 6A-F). In addition, these three arginine residues are found to be important on

373 mediating the effect of ZHX2 on transactivating HIF1α activity in TNBC cells (Figure 6I).

374 Further research needs to be performed to determine whether these residues are critical for

375 ZHX2 localization to DNA, either directly or indirectly via recruitment of co-

376 activators/repressors. Lastly, given that we already found three residues important for the

377 function of ZHX2 in TNBC, we can potentially design small peptides to competitively bind

378 ZHX2 and inhibit its localization to DNA. By engineering these peptides to be membrane

379   permeable, we can potentially test whether they can inhibit the oncogenic role of ZHX2 in

380   TNBC. Given that ZHX2 inhibitors are still not available at this time, these peptide inhibitors

381   can be used as a proof-of-principle approach to motivate further development of specific

382   ZHX2 inhibitors in potential cancer therapies.

383

384   **Materials and Methods**

385   **Cell Culture and Reagents.** MDA-MB-231, MDA-MB-436, MCF-7, Hs578T and 293T cells

386   were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (GIBCO 11965118)

387   supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin (Pen

388   Strep). T47D, BT474, HCC1428, HCC3153, HCC1143, HCC70 and MDA-MB-468 cells

389   were cultured in 10% FBS, 1% Pen Strep RPMI 1640 (GIBCO 11875093). Normal breast

390   epithelial cells HMLE and MCF-10A were cultured in MEGM (Lonza CC-3151) containing

391   SingleQuots Supplements (Lonza CC-4136). 293T cells were obtained from UNC Tissue

392   Culture Facility and authenticated by short tandem repeat testing. HMLE and HCC3153 are

393   in-house cell lines. All other cell lines were obtained from ATCC. Mycoplasma detection was

394   routinely performed to ensure cells were not infected with mycoplasma by using MycoAlert

395   Detection kit (Lonza, LT07-218). Cells were maintained at 37°C in a 5% CO_2 incubator.

396   Cells were incubated overnight in 1% O_2 hypoxia chamber for hypoxia treatment.

397   Doxycycline (D9891) was purchased from Sigma-Aldrich, DMOG (D1070-1g) was from

398   Frontier Scientific, and MG132 (IZL-3175-v) was from Peptide International.

399

400   **Amplification status of ZHX2 in different breast cancer subtype.** All data in table S4

401   were got from cBioPortal (https://www.cbioportal.org/) (Gao *et al*, 2013). We searched the

402   percentage of ZHX2 amplification in all breast cancer datasets, and found ZHX2 was mainly

403   amplified in seven datasets: Breast Cancer (METABRIC, Nature 2012(Curtis *et al*, 2012b) &

404   Nat Commun 2016(Pereira *et al*, 2016)), The Metastatic Breast Cancer Project (Provisional,

405   February 2020), Breast Invasive Carcinoma (TCGA, Cell 2015)(Ciriello *et al*, 2015), Breast

406   Invasive Carcinoma (TCGA, Firehose Legacy), Breast Invasive Carcinoma (TCGA,

407   PanCancer Atlas), Metastatic Breast Cancer (INSERM, PLoS Med 2016)(Lefebvre *et al*,

408   2016), and Breast Invasive Carcinoma (TCGA, Nature 2012)(CancerGenomeAtlasNetwork,

409   2012). Two datasets, Breast Invasive Carcinoma (TCGA, PanCancer Atlas) and Metastatic

410   Breast Cancer (INSERM, PLoS Med 2016)(Lefebvre *et al.*, 2016), did not show the ER, PR

411   and HER2 status, and were excluded from our study. In all the datasets, ER and PR status

412   were determined by immunohistochemistry (IHC). Three datasets, Breast Cancer

413   (METABRIC, Nature 2012(Curtis *et al.*, 2012b) & Nat Commun 2016(Pereira *et al.*, 2016)),

414   The Metastatic Breast Cancer Project (Provisional, February 2020), and Breast Invasive

415   Carcinoma (TCGA, Cell 2015)(Ciriello *et al.*, 2015) assigned the HER2 status by the original

416   researchers. Two datasets, Breast Invasive Carcinoma (TCGA, Cell 2015)(Ciriello *et al.*,

417   2015) and Breast Invasive Carcinoma (TCGA, Firehose Legacy) assigned the HER2 status

418   by two standard, IHC and fluorescence in situ hybridization (FISH). In this study, HER2+

419   status in these two datasets, were determined by IHC.

420

421   **Immunoblotting and Immunoprecipitation Experiments.** EBC buffer (50mM Tris-HCl

422   pH8.0, 120 mM NaCl, 0.5% NP40, 0.1 mM EDTA and 10% glycerol) supplemented with

423   complete protease inhibitor and phosphoSTOP tablets (Roche Applied Bioscience) was

424   used to harvest whole cell lysates at 4°C. Cell lysate concentrations were measured by

425   Protein assay dye (Bio Rad). An equal amount of cell lysates was resolved by SDS-PAGE.

426   For immunoprecipitation, whole-cell lysates were prepared in EBC buffer supplemented with

427   protease inhibitor and phosphatase inhibitor. The lysates were clarified by centrifugation and

428   then incubated with primary antibodies or HA antibody conjugated beads (HA beads, Roche

429   Applied Bioscience) overnight at 4°C. For primary antibody incubation, cell lysates were

430   incubated further with protein G sepharose beads (Roche Applied Bioscience) for 2 hours at

431   4°C. The bound complexes were washed with EBC buffer 5× times and were eluted by

432   boiling in SDS loading buffer. Bound proteins were resolved in SDS-PAGE followed by

433   immunoblotting analysis.

434

435   **Antibodies.** Antibodies used for immunoblotting, immunoprecipitation and IHC staining

436   were as follows: Rabbit anti ZHX2 antibody (Genetex, 112232), Rabbit anti HIF1α (Cell

437   Signaling, 36169), Rabbit anti HIF1β (Cell Signaling, 5537), Rabbit anti VHL (Cell Signaling,

438   68547), rabbit anti HA tag (Cell Signaling, 3724), mouse anti α-Tubulin (Cell Signaling,

439   3873). Peroxidase conjugated goat anti-mouse secondary antibody (31430) and peroxidase

440   conjugated goat anti-rabbit secondary antibody (31460) were from Thermo Scientific.

441

442   **Plasmids**. pBABE HA-VHL, pcDNA-3.1-FLAG-HA-ZHX2(WT), pcDNA-3.1-FLAG-HA-

443   ZHX2(ZHX2sh45 resistant), and pcDNA-3.1- HA-HIF1α were previously described. pcDNA-

444   3.1-FLAG-HA-ZHX2(D489A), pcDNA-3.1-FLAG-HA-ZHX2(R491A), pcDNA-3.1-FLAG-HA-

445   ZHX2(E579A), pcDNA-3.1-FLAG-HA-ZHX2 (R581A), pcDNA-3.1-FLAG-HA-ZHX2 (K582A),

446   pcDNA-3.1-FLAG-HA-ZHX2 (R674A), pcDNA-3.1-FLAG-HA-ZHX2 (E678A), and pcDNA-

447   3.1-FLAG-HA-ZHX2 (R680A) were constructed using standard molecular biology

448   techniques. Quick Change XL Site-Directed Mutagenesis Kit (200516, Agilent Technologies)

449   was used to construct ZHX2 mutants. The GATEWAY Cloning Technology (11789020 and

450   11791019, Invitrogen) was used to recombine plasmids for virus production. All plasmids

451   were sequenced to confirm validity.

452

453   **Lentiviral shRNA, and sgRNA Vectors.** Lentiviral ZHX2 shRNAs (pLKO vector based)

454 were obtained from Broad Institute TRC shRNA library. sgRNAs were cloned into the

455 lentiCRISPR v2 backbone (Addegene Plasmid #52961). Target sequences were as follows:

456 Control shRNA: AACAGTCGCGTTTGCGACTGG

457 ZHX2 shRNA (43): CCCACTAAATACTACCAAATA

458 ZHX2 shRNA (45): CCGTAGCAAGGAAAGCAACAA

459 HIF1α shRNA (3809): CCAGTTATGATTGTGAAGTTA

460 HIF1α shRNA (3810): GTGATGAAAGAATTACCGAAT

461 Control sgRNA: GCGAGGTATTCGGCTCCGCG

462 VHL sgRNA (1): CATACGGGCAGCACGACGCG

463 VHL sgRNA (2): GCGATTGCAGAAGATGACCT

464 VHL sgRNA (8): ACCGAGCGCAGCACGGGCCG

465

466 **Virus Production and Infection.** 293T packaging cell lines were used for lentiviral

467 amplification. Lentiviral infection was carried out as previously described (Zhang *et al.*,

468 2018). Briefly, viruses were collected at 48 h and 72 h post-transfection. After passing

469 through 0.45-µm filters, viruses were used to infect target cells in the presence of 8 µg/mL

470 polybrene. Subsequently, target cell lines underwent appropriate antibiotic selection.

471

472 **Cell Viability Assay**. For MTS assay, cells were seeded in triplicate in 96-well plates (1000

473 cells/well) in appropriate growth medium. At indicated time points, cells were replaced with

474 90 µl fresh growth medium supplemented with 10 µl MTS reagents (Abcam, ab197010),

475 followed by incubation at 37°C for 1-4 hrs. OD absorbance values were measured at 490

476 nm using a 96-well plate reader (BioTek).

477

478 **2-D Cell Proliferation Assay**. For colony formation assays, cells were seeded in duplicate

479 in 6-well plates (2 x 10³ cells/well) in appropriate growth medium. Media was changed every

480 two days. After 7 days, cells were fixed with 4% formaldehyde for 10 minutes at room

481 temperature, stained for 10 minutes with 0.5% crystal violet and then washed several times

482 with distilled water. Once dried, the plates were scanned.

483

484 **3-D Anchorage Independent Soft Agar Growth Assay**. Cells were plated in a top layer at

485 a density of 10,000 cells per ml in complete medium with 0.4% agarose (Life Technologies,

486 BP165-25), onto bottom layers composed of medium with 1% agarose followed by

487 incubation at 4°C for 10 minutes. Afterwards, cells were moved to a 37°C incubator. Every 4

488 days, 200µl of complete media were added onto the plate. After 2-4 weeks, the extra liquid

489 on the plate was aspirated, and 1 ml medium supplemented with 100 µg/ml

490 iodonitrotetrazoliuim chloride solution was added onto each well. After incubating overnight

491 at 37°C, the colonies were captured by an image microscope and quantified after a whole

492 plate scan.

493

494 **Cell invasion assay** MDA-MB-231 and MDA-MB-468 cell invasion assay was performed

495 using BD BioCoat Matrigel Invasion Chamber (354480) according to the manufacturer's

496 instructions. In total, 3×10^4 (for MDA-MB-231) and 3×10^5 (for MDA-MB-468) cells were

497 inoculated into each chamber in triplicate and incubated for 18 h at 37 °C, 5% CO2

498 incubator. The cells on the lower surface of the membrane were stained using Diff-Quick

499 stain kit (B4132-1A) from SIEMENS, and then counted under EVOS XL Core Microscope

500 (Cat# AMEX1000, Thermo Fisher Scientific).

501

502   **RNA-seq Analysis.** Procedures was described previously (Liao *et al*, 2020). Briefly, Total

503   RNA from triplicates was extracted from MDA-MB-231 cells infected with control or ZHX2

504   shRNAs by using RNeasy kit with on column DNase digestion (Qiagen). Library preparation

505   and sequencing were performed by BGI as paired end 50bp reads. Reads were then filtered

506   for adapter contamination using cutadapt (Patro *et al*, 2017) and filtered such that at least

507   90% of bases of each read had a quality score >20. Reads were aligned to the reference

508   genome (hg19) using STAR version 2.5.2b, and only primary alignments were retained

509   (Love *et al*, 2014). Reads overlapping blacklisted regions of the genome were then

510   removed. Transcript abundance was then estimated using salmon (Miller *et al*, 2012), and

511   differential expression was detected using DESeq2 (Bird *et al*, 2010). RNA-seq data are

512   available at GSE175487. Pathway enrichments were calculated using g:Profiler (Reimand *et*

513   *al*, 2019) where the pathway database was supplemented with the list of HIF DKO

514   downregulated genes to obtain the adjusted p-value indicating a significant association.

515   Association with the HALLMARK hypoxia pathway was conducted using GSEA (Reimand *et*

516   *al.*, 2019).

517

518   **Real-Time PCR.** Total RNA was isolated with RNeasy mini kit (Qiagen). First strand cDNA

519   was generated with an iScript cDNA synthesis kit (BioRad). Real-time PCR was performed

520   in triplicate. Real time PCR primer sequences are listed in Table S5, Supplementary.

521

522   **ChIP-seq Analyses**. MDA-MB-231 cells were infected by HA-ZHX2 which is resistant to

523   ZHX2 sh45, and then infected by ZHX2 sh45. ChIP was performed with HA tag (Cell

524   Signaling, 3724). The ChIP-Seq library was prepared using a ChIP-Seq DNA sample

525   preparation kit (Illumina) according to manufacturer's instructions. Samples were sequenced

526   on an Illumina HiSeq2500 with single-end 76 bp reads. Reads were then filtered for adaptor

527 contamination using Cutadapt and filtered such that at least 90% of bases of each read had

528 a quality score > 20. Duplicated sequences were then capped at a maximum of 5

529 occurrences, and reads were aligned to the reference genome (hg19) using STAR (Dobin *et*

530 *al*, 2013) version 2.5.2b retaining only primary alignments. Reads overlapping blacklisted

531 regions of the genome were then removed. Reads were then extended *in silico* to a

532 fragment size of 250 bp, and regions of significant enrichment relative to input control were

533 identified using MACS2 (Zhang *et al*, 2008). A unified set of enriched regions for ZHX2 was

534 obtained by taking the intersection of the two replicates using bedtools (Quinlan & Hall,

535 2010). ChIP-seq data for HIF1α was obtained from GSE108833, and data for H3K4me3 and

536 H3K27ac were obtained from GSE49651. ChIP enrichment heatmaps over promoters were

537 generated using deepTools (Ramirez *et al*, 2016).

538

539 **Orthotopic Tumor Xenograft.** Procedures for animal studies was described previously

540 (Liao *et al.*, 2020). Briefly, six-week-old female NOD SCID Gamma mice (NSG, Jackson

541 lab) were used for xenograft studies. Approximately 1×10^6 viable MDA-MB-231 cells

542 expressing Teton control or Teton ZHX2 shRNAs were resuspended in 1: 1 ratio in 50 µl

543 medium and 50 µl matrigel (Corning, 354234) and injected orthotopically into the fourth

544 mammary fat pad of each mouse. After cell injection and following two consecutive weeks of

545 tumor monitoring to ensure the tumor was successfully implanted, mice were fed Purina

546 rodent chow with doxycycline (Research Diets Inc., #5001). Tumor size was measured twice

547 a week using an electronic caliper. Tumor volumes were calculated with the formula: volume

548 $= (L \times W^2)/2$, where L is the tumor length and W is the tumor width measured in millimeters.

549 The rough mass of tumors was presented as mean ± SEM and evaluated statistically using *t*

550 test. After mice were sacrificed, lung ex vivo imaging was performed immediately to

551 examine tumor metastasis. All animal experiments were in compliance with National

552 Institutes of Health guidelines and were approved by the University of Texas, Southwestern

553 Medical Center Institutional Animal Care and Use Committee.

554

555 **Binding Free Energy Calculations using MM/PBSA.** To calculate the binding Gibbs free

556 energy change between two groups of molecules, we used Molecular Mechanics Poisson–

557 Boltzmann Surface Area continuum solvation (MM/PBSA) approach(Miller *et al.*, 2012) to

558 analyze trajectories out of Molecular Dynamics (MD) simulations (detailed in Supporting

559 Information). In this approach, the binding Gibbs free energy change, ΔG_{bind}, can be

560 expressed in terms of the xyz (ΔG_{MM}), xyz (ΔG_{solv}), and the entropy of the system ($T\Delta S$)

561 as shown in equation (1).

562 $$\Delta G_{bind} = \Delta G_{MM} + \Delta G_{solv} - T\Delta S \qquad (1)$$

563 where $\Delta G_{MM} = \Delta G_{elec} + \Delta G_{VDW}$ and $\Delta G_{solv} = \Delta G_{polar} + \Delta G_{nonpolar}$

564

565 **Statistical Analysis.** All statistical analysis was conducted using Prism 8.0 (GraphPad

566 Software). All graphs depict mean ± SEM unless otherwise indicated. Statistical

567 significances are denoted as n.s. (not significant; $P>0.05$), $*P < 0.05$, $**P < 0.01$, $***P <$

568 0.001, $****P < 0.0001$. The numbers of experiments are noted in Figure legends. To assess

569 the statistical significance of a difference between two conditions, we used unpaired two-tail

570 student's *t*-test. For experiments comparing more than two conditions, differences were

571 tested by a one-way ANOVA followed by Dunnett's or Tukey's multiple comparison tests.

572

573 **Data availability**

574 RNA-Seq and CHIP-Seq data are available GEO175487. All data generated or analysed

575 during this study are included in the manuscript and supporting files.

576

597

598 **References**

599 Anders CK, Carey LA (2009) Biology, metastatic patterns, and treatment of patients with
600 triple-negative breast cancer. *Clin Breast Cancer* 9 Suppl 2: S73-81
601 Appelhoff RJ, Tian YM, Raval RR, Turley H, Harris AL, Pugh CW, Ratcliffe PJ, Gleadle
602 JM (2004) Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the

603   regulation of hypoxia-inducible factor. *The Journal of biological chemistry* 279: 38458-
604   38465
605   Armellini A, Sarasquete ME, Garcia-Sanz R, Chillon MC, Balanzategui A, Alcoceba M,
606   Fuertes M, Lopez R, Hernandez JM, Fernandez-Calvo J *et al* (2008) Low expression of
607   ZHX2, but not RCBTB2 or RAN, is associated with poor outcome in multiple myeloma. *Br J*
608   *Haematol* 141: 212-215
609   Bird LE, Ren J, Nettleship JE, Folkers GE, Owens RJ, Stammers DK (2010) Novel structural
610   features in two ZHX homeodomains derived from a systematic study of single and multiple
611   domains. *BMC Struct Biol* 10: 13
612   Bos R, Zhong H, Hanrahan CF, Mommers EC, Semenza GL, Pinedo HM, Abeloff MD,
613   Simons JW, van Diest PJ, van der Wall E (2001) Levels of hypoxia-inducible factor-1 alpha
614   during breast carcinogenesis. *J Natl Cancer Inst* 93: 309-314
615   Briggs KJ, Koivunen P, Cao S, Backus KM, Olenchock BA, Patel H, Zhang Q, Signoretti S,
616   Gerfen GJ, Richardson AL *et al* (2016) Paracrine Induction of HIF by Glutamate in Breast
617   Cancer: EglN1 Senses Cysteine. *Cell* 166: 126-139
618   Brown JM, Wilson WR (2004) Exploiting tumour hypoxia in cancer treatment. *Nature*
619   *reviews Cancer* 4: 437-447
620   Cancer Genome Atlas N (2012) Comprehensive molecular portraits of human breast
621   tumours. *Nature* 490: 61-70
622   CancerGenomeAtlasNetwork (2012) Comprehensive molecular portraits of human breast
623   tumours. *Nature* 490: 61-70
624   Chen Y, Zhang B, Bao L, Jin L, Yang M, Peng Y, Kumar A, Wang JE, Wang C, Zou X *et al*
625   (2018) ZMYND8 acetylation mediates HIF-dependent breast cancer progression and
626   metastasis. *J Clin Invest* 128: 1937-1955
627   Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan
628   M, Yau C, Kandoth C *et al* (2015) Comprehensive Molecular Portraits of Invasive Lobular
629   Breast Cancer. *Cell* 163: 506-519
630   Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG,
631   Samarajiwa S, Yuan Y *et al* (2012a) The genomic and transcriptomic architecture of 2,000
632   breast tumours reveals novel subgroups. *Nature* 486: 346-352
633   Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG,
634   Samarajiwa S, Yuan Y *et al* (2012b) The genomic and transcriptomic architecture of 2,000
635   breast tumours reveals novel subgroups. *Nature* 486: 346-352
636   Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
637   Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21
638   Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha
639   R, Larsson E *et al* (2013) Integrative analysis of complex cancer genomics and clinical
640   profiles using the cBioPortal. *Science signaling* 6: pl1
641   Gray LH, Conger AD, Ebert M, Hornsey S, Scott OC (1953) The concentration of oxygen
642   dissolved in tissues at the time of irradiation as a factor in radiotherapy. *The British journal*
643   *of radiology* 26: 638-648
644   Guan Y, Kuo WL, Stilwell JL, Takano H, Lapuk AV, Fridlyand J, Mao JH, Yu M, Miller
645   MA, Santos JL *et al* (2007) Amplification of PVT1 contributes to the pathophysiology of
646   ovarian and breast cancer. *Clinical cancer research : an official journal of the American*
647   *Association for Cancer Research* 13: 5745-5755

648   Heitz F, Harter P, Lueck HJ, Fissler-Eckhoff A, Lorenz-Salehi F, Scheil-Bertram S, Traut A,
649   du Bois A (2009) Triple-negative and HER2-overexpressing breast cancers exhibit an
650   elevated risk and an earlier occurrence of cerebral metastases. *Eur J Cancer* 45: 2792-2798
651   Hu L, Xie H, Liu X, Potjewyd F, James LI, Wilkerson EM, Herring LE, Xie L, Chen X,
652   Cabrera JC *et al* (2020) TBK1 Is a Synthetic Lethal Target in Cancer with VHL Loss.
653   *Cancer discovery* 10: 460-475
654   Hu S, Zhang M, Lv Z, Bi J, Dong Y, Wen J (2007) Expression of zinc-fingers and
655   homeoboxes 2 in hepatocellular carcinogenesis: a tissue microarray and clinicopathological
656   analysis. *Neoplasma* 54: 207-211
657   Ivan M, Kondo K, Yang H, Kim W, Valiando J, Ohh M, Salic A, Asara JM, Lane WS,
658   Kaelin WG, Jr. (2001) HIFalpha targeted for VHL-mediated destruction by proline
659   hydroxylation: implications for O2 sensing. *Science* 292: 464-468
660   Jaakkola P, Mole DR, Tian YM, Wilson MI, Gielbert J, Gaskell SJ, von Kriegsheim A,
661   Hebestreit HF, Mukherji M, Schofield CJ *et al* (2001) Targeting of HIF-alpha to the von
662   Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* 292:
663   468-472
664   Kawata H, Yamada K, Shou Z, Mizutani T, Yazawa T, Yoshino M, Sekiguchi T, Kajitani T,
665   Miyamoto K (2003) Zinc-fingers and homeoboxes (ZHX) 2, a novel member of the ZHX
666   family, functions as a transcriptional repressor. *The Biochemical journal* 373: 747-757
667   Lefebvre C, Bachelot T, Filleron T, Pedrero M, Campone M, Soria JC, Massard C, Lévy C,
668   Arnedos M, Lacroix-Triki M *et al* (2016) Mutational Profile of Metastatic Breast Cancers: A
669   Retrospective Analysis. *PLoS Med* 13: e1002201
670   Liao C, Zhang Y, Fan C, Herring LE, Liu J, Locasale JW, Takada M, Zhou J, Zurlo G, Hu L
671   *et al* (2020) Identification of BBOX1 as a Therapeutic Target in Triple-Negative Breast
672   Cancer. *Cancer Discov* 10: 1706-1721
673   Lin NU, Claus E, Sohl J, Razzak AR, Arnaout A, Winer EP (2008) Sites of distant
674   recurrence and clinical outcomes in patients with metastatic triple-negative breast cancer:
675   high incidence of central nervous system metastases. *Cancer* 113: 2638-2645
676   Liu X, Simon JM, Xie H, Hu L, Wang J, Zurlo G, Fan C, Ptacek TS, Herring L, Tan X *et al*
677   (2020) Genome-wide Screening Identifies SFMBT1 as an Oncogenic Driver in Cancer with
678   VHL Loss. *Molecular cell* 77: 1294-1306 e1295
679   Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for
680   RNA-seq data with DESeq2. *Genome biology* 15: 550
681   Ma H, Yue X, Gao L, Liang X, Yan W, Zhang Z, Shan H, Zhang H, Spear BT, Ma C (2015)
682   ZHX2 enhances the cytotoxicity of chemotherapeutic drugs in liver tumor cells by repressing
683   MDR1 via interfering with NF-YA. *Oncotarget* 6: 1049-1063
684   Masui K, Gini B, Wykosky J, Zanca C, Mischel PS, Furnari FB, Cavenee WK (2013) A tale
685   of two approaches: complementary mechanisms of cytotoxic and targeted therapy resistance
686   may inform next-generation cancer treatments. *Carcinogenesis* 34: 725-738
687   Miller BR, 3rd, McGee TD, Jr., Swails JM, Homeyer N, Gohlke H, Roitberg AE (2012)
688   MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory*
689   *Comput* 8: 3314-3321
690   Nagel S, Schneider B, Meyer C, Kaufmann M, Drexler HG, Macleod RA (2012)
691   Transcriptional deregulation of homeobox gene ZHX2 in Hodgkin lymphoma. *Leukemia*
692   *research* 36: 646-655

693   Nagel S, Schneider B, Rosenwald A, Meyer C, Kaufmann M, Drexler HG, MacLeod RA
694   (2011) t(4;8)(q27;q24) in Hodgkin lymphoma cells targets phosphodiesterase PDE5A and
695   homeobox gene ZHX2. *Genes Chromosomes Cancer* 50: 996-1009
696   Niwinska A, Murawska M, Pogoda K (2010) Breast cancer brain metastases: differences in
697   survival depending on biological subtype, RPA RTOG prognostic class and systemic
698   treatment after whole-brain radiotherapy (WBRT). *Annals of oncology : official journal of*
699   *the European Society for Medical Oncology / ESMO* 21: 942-948
700   Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and
701   bias-aware quantification of transcript expression. *Nat Methods* 14: 417-419
702   Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, Pugh M, Jones L,
703   Russell R, Sammut SJ *et al* (2016) The somatic mutation profiles of 2,433 breast cancers
704   refines their genomic and transcriptomic landscapes. *Nat Commun* 7: 11479
705   Perincheri S, Dingle RW, Peterson ML, Spear BT (2005) Hereditary persistence of alpha-
706   fetoprotein and H19 expression in liver of BALB/cJ mice is due to a retrovirus insertion in
707   the Zhx2 gene. *Proceedings of the National Academy of Sciences of the United States of*
708   *America* 102: 396-401
709   Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic
710   features. *Bioinformatics* 26: 841-842
711   Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F,
712   Manke T (2016) deepTools2: a next generation web server for deep-sequencing data
713   analysis. *Nucleic Acids Res* 44: W160-165
714   Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L,
715   Meyer M, Wong J, Xu C *et al* (2019) Pathway enrichment analysis and visualization of
716   omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature protocols* 14:
717   482-517
718   Rhie SK, Hazelett DJ, Coetzee SG, Yan CL, Noushmehr H, Coetzee GA (2014) Nucleosome
719   positioning and histone modifications define relationships between regulatory elements and
720   nearby gene expression in breast epithelial cells. *Bmc Genomics* 15
721   Semenza GL (2010) Defining the role of hypoxia-inducible factor 1 in cancer biology and
722   therapeutics. *Oncogene* 29: 625-634
723   Semenza GL (2012) Hypoxia-inducible factors in physiology and medicine. *Cell* 148: 399-
724   408
725   Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to
726   genome-wide predictions. *Nat Rev Genet* 15: 272-286
727   Yue X, Zhang Z, Liang X, Gao L, Zhang X, Zhao D, Liu X, Ma H, Guo M, Spear BT *et al*
728   (2012) Zinc fingers and homeoboxes 2 inhibits hepatocellular carcinoma cell proliferation
729   and represses expression of Cyclins A and E. *Gastroenterology* 142: 1559-1570.e1552
730   Zhang J, Wu T, Simon J, Takada M, Saito R, Fan C, Liu XD, Jonasch E, Xie L, Chen X *et al*
731   (2018) VHL substrate transcription factor ZHX2 as an oncogenic driver in clear cell renal
732   cell carcinoma. *Science (New York, NY)* 361: 290-295
733   Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers
734   RM, Brown M, Li W *et al* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*
735   9: R137
736   Zheng X, Zhai B, Koivunen P, Shin SJ, Lu G, Liu J, Geisen C, Chakraborty AA, Moslehi JJ,
737   Smalley DM *et al* (2014) Prolyl hydroxylation by EglN2 destabilizes FOXO3a by blocking
738   its interaction with the USP9x deubiquitinase. *Genes & development* 28: 1429-1444

739   Zurlo G, Liu X, Takada M, Fan C, Simon JM, Ptacek TS, Rodriguez J, von Kriegsheim A,
740   Liu J, Locasale JW *et al* (2019) Prolyl hydroxylase substrate adenylosuccinate lyase is an
741   oncogenic driver in triple negative breast cancer. *Nat Commun* 10: 5177
742

743   **Figures Legend**

744   **Figure. 1. ZHX2 is amplified in TNBC and is potentially regulated by pVHL.**

745   (**A, B**) The percentage of tumor samples with ZHX2 (top) or MYC (bottom) focal

746   amplification across cancer types (A) or samples with both ZHX2 and MYC (red), ZHX2-

747   specific (magenta), and c-MYC-specific (orange) (B). Asterisk indicates statistical

748   significance for overlap of ZHX2 and MYC focal amplification (Fisher's exact test, $p < 0.05$).

749   (**C**) The percentage of ZHX2 amplification of different breast cancer subtypes in several

750   breast cancer datasets.

751   (**D**) The relation of ZHX2 copy number gain and its expression in TCGA datasets and

752   METABRIC datasets.

753   (**E**) Immunoblots of lysates from different normal breast cell and breast cancer cell lines.

754   (**F**) Immunoblots of lysates from paired TNBC patient-derived non-tumor (N) and tumor (T)

755   breast tissues.

756   (**G**) Immunoprecipitations (IP) of MDA-MB-231 cells infected with either control vector (EV)

757   or FLAG-HA-ZHX2.

758   (**H**) Immunoblots of lysates with indicated antibodies from MDA-MB-231 cells infected with

759   either EV or HA-VHL.

760   (**I, J**) Immunoblot of cell lysates from MDA-MB-231 (I) or MDA-MB-468 (J) infected with

761   lentivirus encoding either VHL sgRNAs (1, 2, or 8) or control sgRNA (Ctrl).

762   (**K, L**) Immunoblots of lysates from MDA-MB-231 (K) or MDA-MB-468 (L) cells treated with

763   indicated inhibitors for 8 h.

764   (**M**) Immunoblots of lysates from another 10 pair of TNBC patient-derived non-tumor (N) and

765   tumor (T) breast tissues.

766   **Figure supplement 1**. ZHX2 overexpression leads worse survival and is potentially

767   regulated by pVHL in breast cancer.

768   **Figure 1—source data.** Uncropped western blot images for Figure 1.

769   **Figure supplement 1—source data.** Uncropped western blot images for Figure

770   supplement 1.

771

772   **Figure. 2. ZHX2 is essential for TNBC cell proliferation and invasion.**

773   (**A-H**) Immunoblot of cell lysates (A), qRT-PCR of RNA (B), cell proliferation assays (C), 2D-

774   clones (D), soft agar growth (E) and quatification (F), invasion (G) and quantification (H) of

775   MDA-MB-231/468 cells infected with lentivirus encoding two individual ZHX2 shRNAs (43,

776   45) or control shRNA (Ctrl).

777   (**I-N**) Immunoblot of cell lysates (I), cell proliferation (J), 2D-clones (K), soft agar growth

778   (upper) and invasion assays (down) (L) as well as quantification (M-N) of MDA-MB-231

779   cells infected with lentivirus encoding ZHX2 sh45-resistant FLAG-HA-ZHX2 or control vector

780   (EV), followed by ZHX2 sh45 or Ctrl.

781   Error bars represent mean ± SEM, unpaired t-test. * denotes p value of <0.05, * * denotes p

782   value of <0.01, * * * denotes p value of <0.005.

783   **Figure supplement 2**. The phenotype of ZHX2 shRNA on cell proliferation and invasion is

784   due to its ontarget effect.

785   **Figure 2—source data.** Uncropped western blot images for Figure 2.

786   **Figure supplement 2—source data.** Uncropped western blot images for Figure

787   supplement 2.

788

789

790 **Figure. 3. ZHX2 is important for TNBC cell proliferation *in vivo*.**

791 (**A-E**) Immunoblot of cell lysates (A), cell proliferation (B), 2D-colony growth (C), soft agar

792 growth (D) and quantification (E) of MDA-MB-231 cells infected with lentivirus encoding

793 Teton-ZHX2 shRNAs (43, 45) or Teton-Ctrl.

794 (**F-G**) Tumor growth (F), and tumor weight (G) of doxycycline-inducible ZHX2 knockdown

795 MDA-MB-231 cells injected orthotopically at the mammary fat pad of NSG mice. Treatment

796 of doxycycline food started as indicated time.

797 (**H**) Immunoblot of tumor lysates from mouse that injected doxycycline-inducible ZHX2

798 knockdown MDA-MB-231 cells.

799 (**I, J**) Images (I) and plot (J) of lung necropsy of doxycycline-inducible ZHX2 knockdown

800 MDA-MB-231 cells injected orthotopically at the mammary fat pad of NSG mice.

801 (**K, L**) Tumor growth (K), and tumor weight (L) of doxycycline induced or not induced ZHX2

802 sh45 MDA-MB-231 cells injected orthotopically at the mammary fat pad of NSG mice.

803 (**M**) Immunoblot of tumor lysates from mouse treated with regular chow or doxycycline

804 chow.

805 Error bars represent mean ± SEM, unpaired t-test. * denotes p value of <0.05, * * denotes p

806 value of <0.01, * * * denotes p value of <0.005.

807 **Figure supplement 3**. ZHX2 is important for maintaining TNBC tumorigenesis in vivo.

808 **Figure 3—source data.** Uncropped western blot images for Figure 3.

809 **Figure supplement 3—source data.** Uncropped western blot images for Figure

810 supplement 3.

811

812 **Figure. 4. ZHX2 regulates HIF signaling in TNBC.**

813   (**A**) Venn diagram showing overlap in downregulated (top) and upregulated (bottom) genes

814   between two different ZHX2 shRNA.

815   (**B**) Pathway analysis of the significantly decreased pathways in ZHX2 depleted MDA-MB-

816   231 cells.

817   (**C, D**) Venn diagram showing overlap in differentially expressed genes (C) and

818   downregulated genes (D) between ZHX2 depletion and HIF double knock-out (DKO)

819   (GSE108833).

820   (**E, F**) Immunoblots of immunoprecipitations (IP) of MDA-MB-231 cells overexpress either

821   FLAG-HA-ZHX2 or HA-HIF1α (E) or both ZHX2 and HIF1α (F).

822   (**G**) Immunoblots of immunoprecipitations (IP) of MDA-MB-231 cells treated with DMOG for

823   8 hours.

824   (**H, I**) Immunoblots of cell lysates from MDA-MB-231 cells (H) and MDA-MB-468 cells (I)

825   infected with lentivirus encoding ZHX2 shRNAs or Ctrl, followed treated with normoxia or

826   hypoxia (1% O_2).

827   (**J, L**) Immunoblots (J) and immunoprecipitations (K) of cell lysates, qRT-PCR of mRNA (L)

828   from MDA-MB-231 cells infected with lentivirus encoding EV or FLAG-HA-ZHX2.

829   (**M**) Immunoblots of cell lysates from MDA-MB-231 cells infected with lentivirus encoding

830   ZHX2 shRNAs or Ctrl treated with MG132 overnight under hypoxia (1% O_2).

831   Error bars represent mean ± SEM, unpaired *t*-test. * denotes p value of <0.05, * * denotes p

832   value of <0.01, * * * denotes p value of <0.005.

833   **Figure supplement 4**. ZHX2 regulates HIF1 signaling in TNBC.

834   **Figure 4—source data.** Uncropped western blot images for Figure 4.

835   **Figure supplement 4—source data.** Uncropped western blot images for Figure

836   supplement 4.

837

838

Figure. 5. Representative ZHX2 and HIF downstream targets and analysis of their chromatin binding motifs.

(**A**) Integrated analyses of ChIP-Seqs (including ZHX2, HIF1α, H3K4me3 and H3K27ac), signals expressed as relative to input control when available. Log2 fold change (LFC) for ZHX2 knock down RNA-Seq and HIF double knockout (HIF DKO) RNA-Seq; Critical target genes were marked on the right.

(**B-D**) qRT-PCR quantification of ZHX2 target genes from MDA-MB-231 cells infected with lentivirus encoding ZHX2 shRNAs (43, 45) (B), EV or FLAG-HA-ZHX2 (C) or HIF double knockout under normoxia (21% O_2) and hypoxia (1% O_2) (D).

848

Figure. 6. Identification of important sites on ZHX2 that may affect its function in TNBC.

(**A-F**) Immunoblot of cell lysates (A), cell proliferation (B), 2D colony (top), 3D soft-agar (middle) and invasion (bottom) (C) as well as quantification 2D colony (D), 3D soft-agar (E) and invasion (F) of MDA-MB-231 cell lines infected with lentivirus encoding either ZHX2 wild type (WT) or mutation, followed by ZHX2 sh45 or Ctrl.

(**G-I**) Immunoprecipitations (G), Immunoblot (H) of cell lysates and qRT-PCR of mRNA (I) of MDA-MB-231 cells infected with lentivirus encoding either ZHX2 WT or mutation.

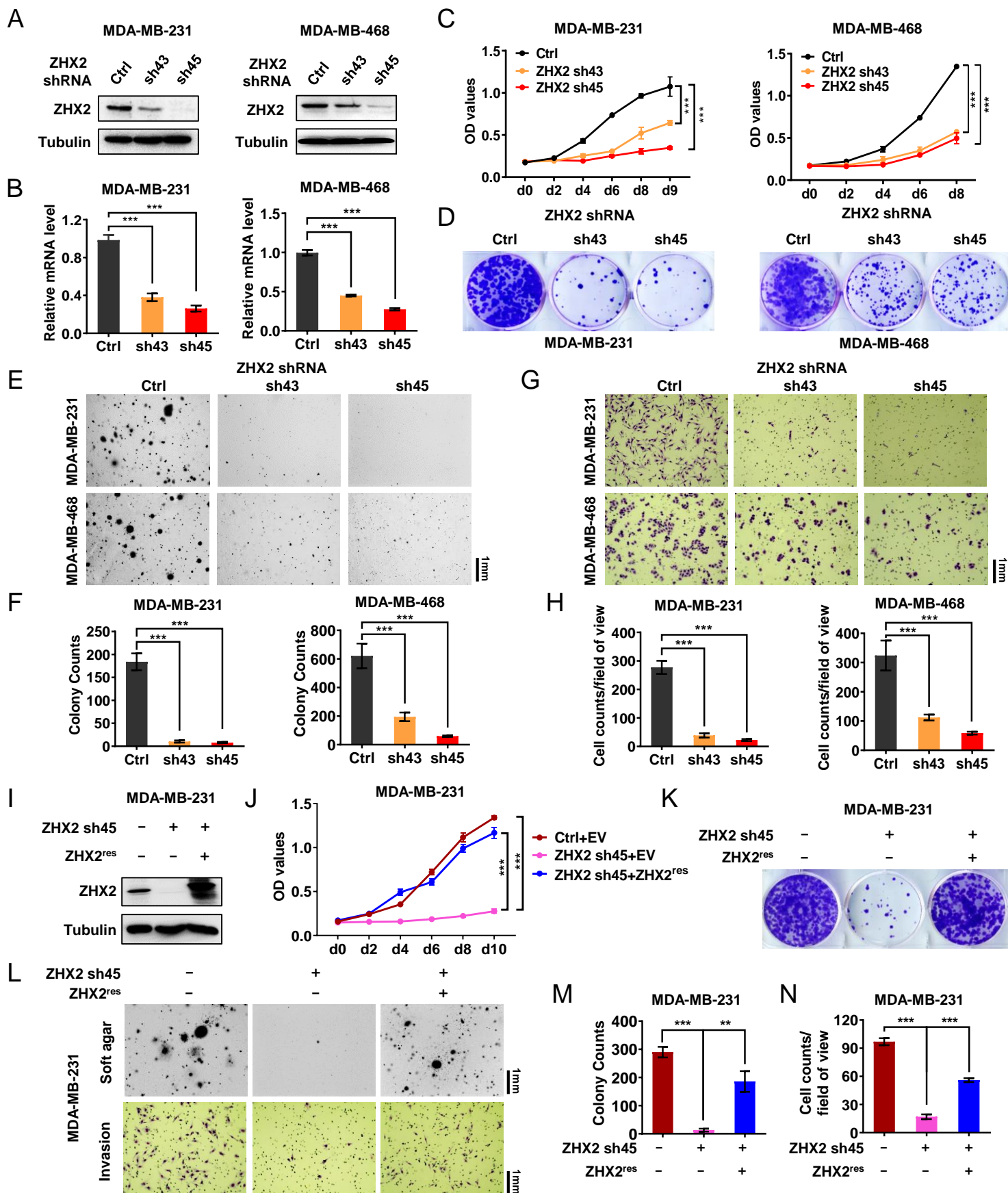Error bars represent mean ± SEM, n=3 replicates per group, unpaired *t*-test. *** denotes p value of <0.005.

Figure supplement 5. Residue based DNA contact analysis derived from ensembles of DNA bound HD proteins.

861 **Figure 6—source data.** Uncropped western blot images for Figure 6.

Figure 1
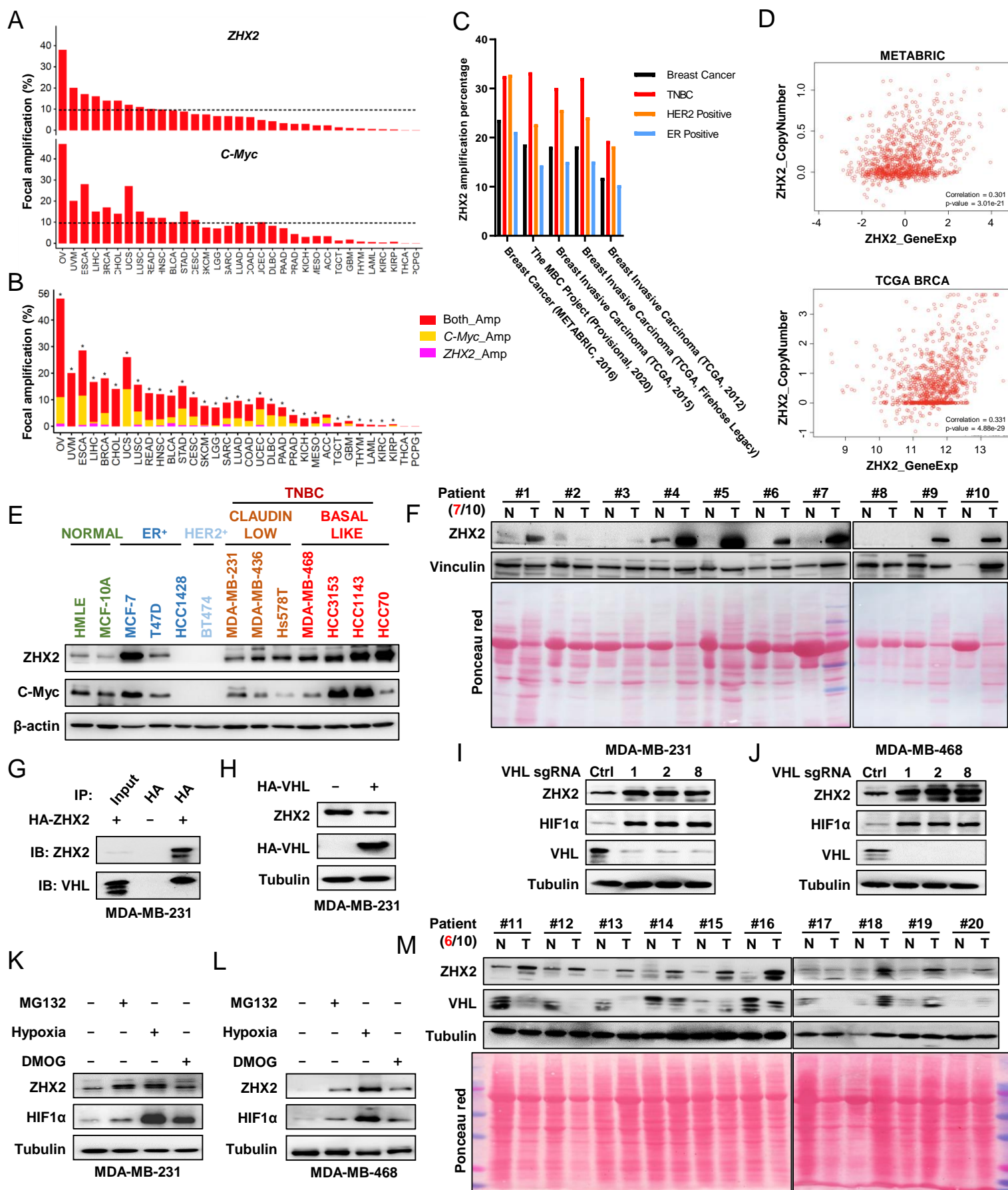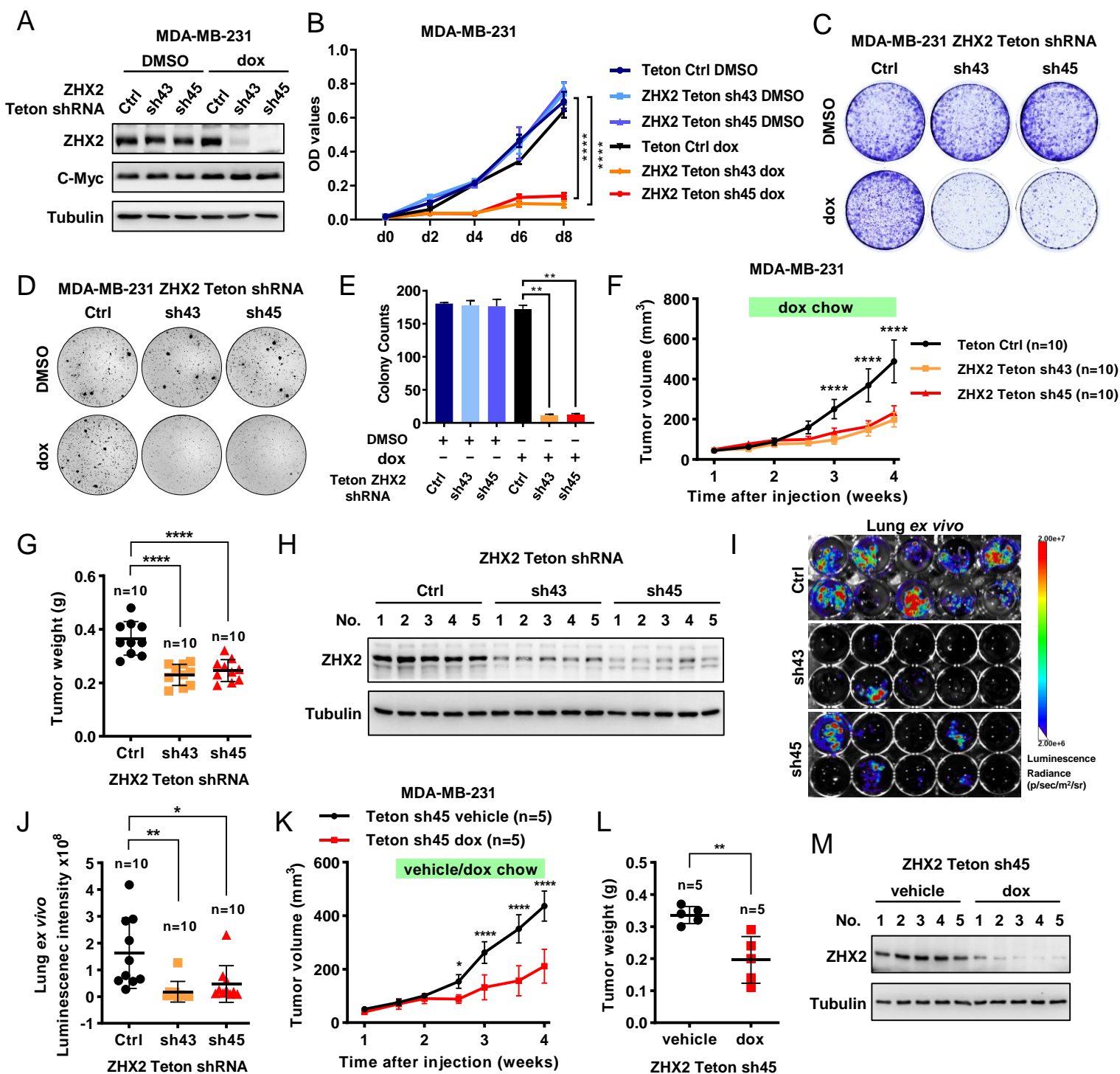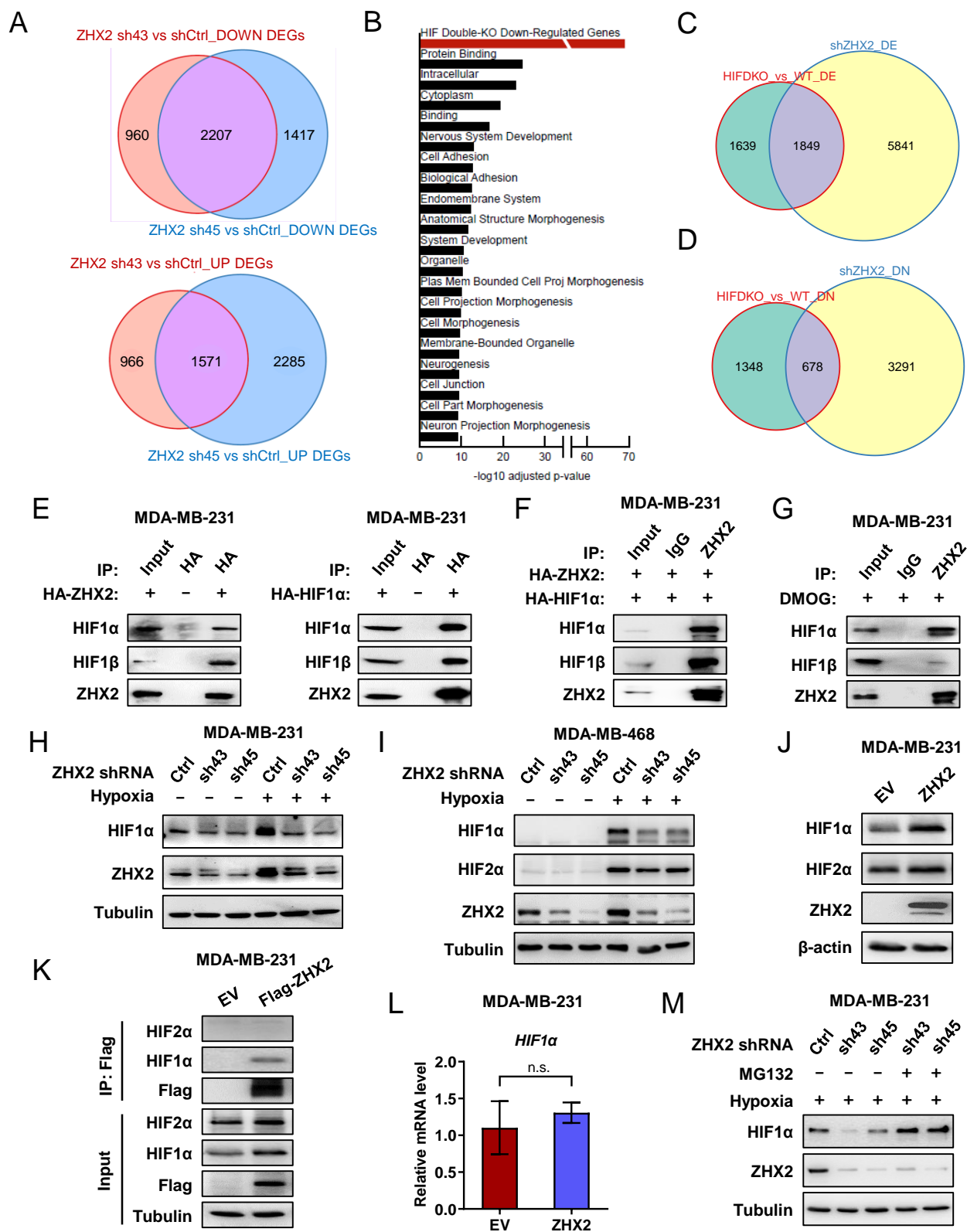
Figure 2

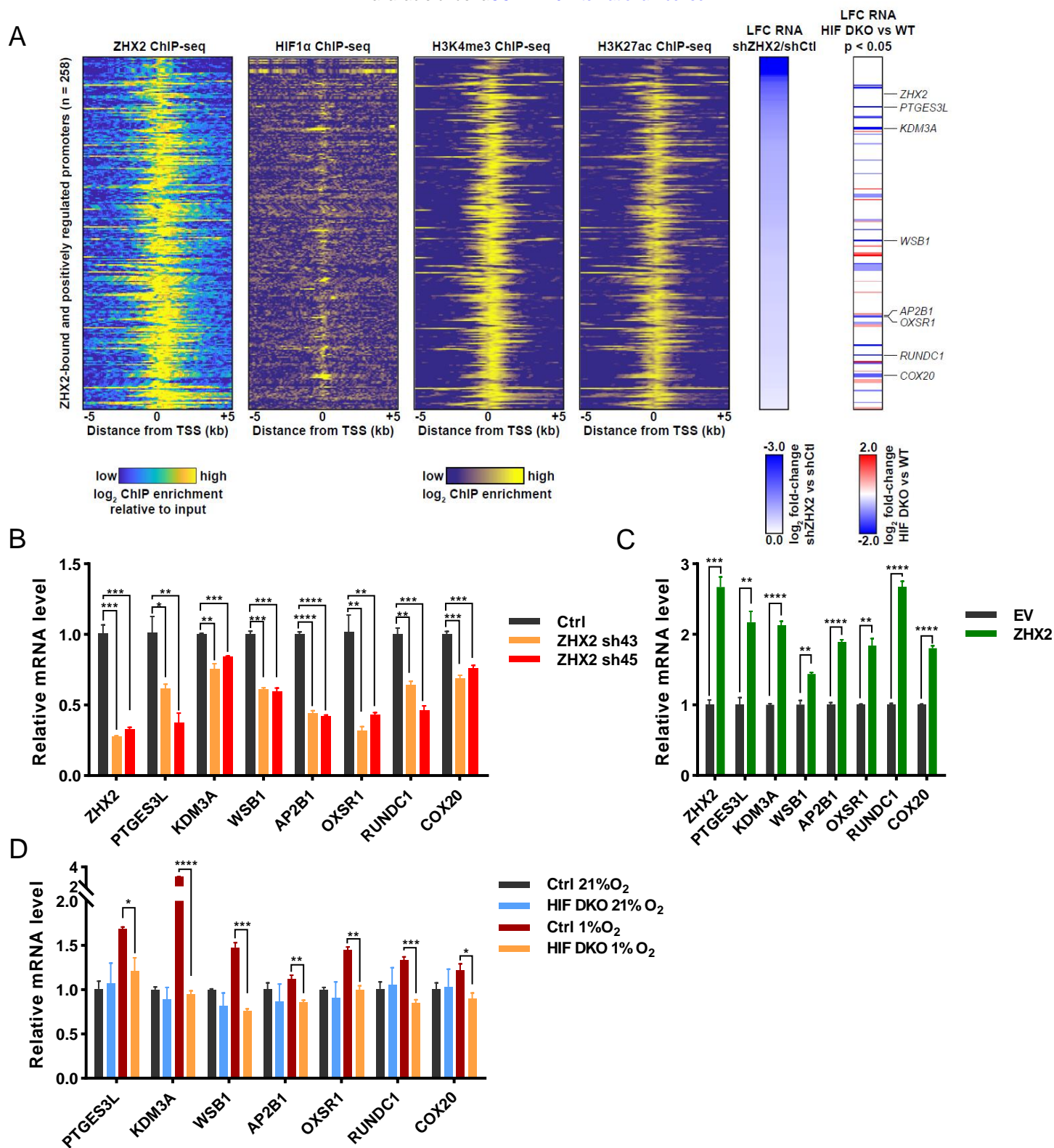**Figure 3**

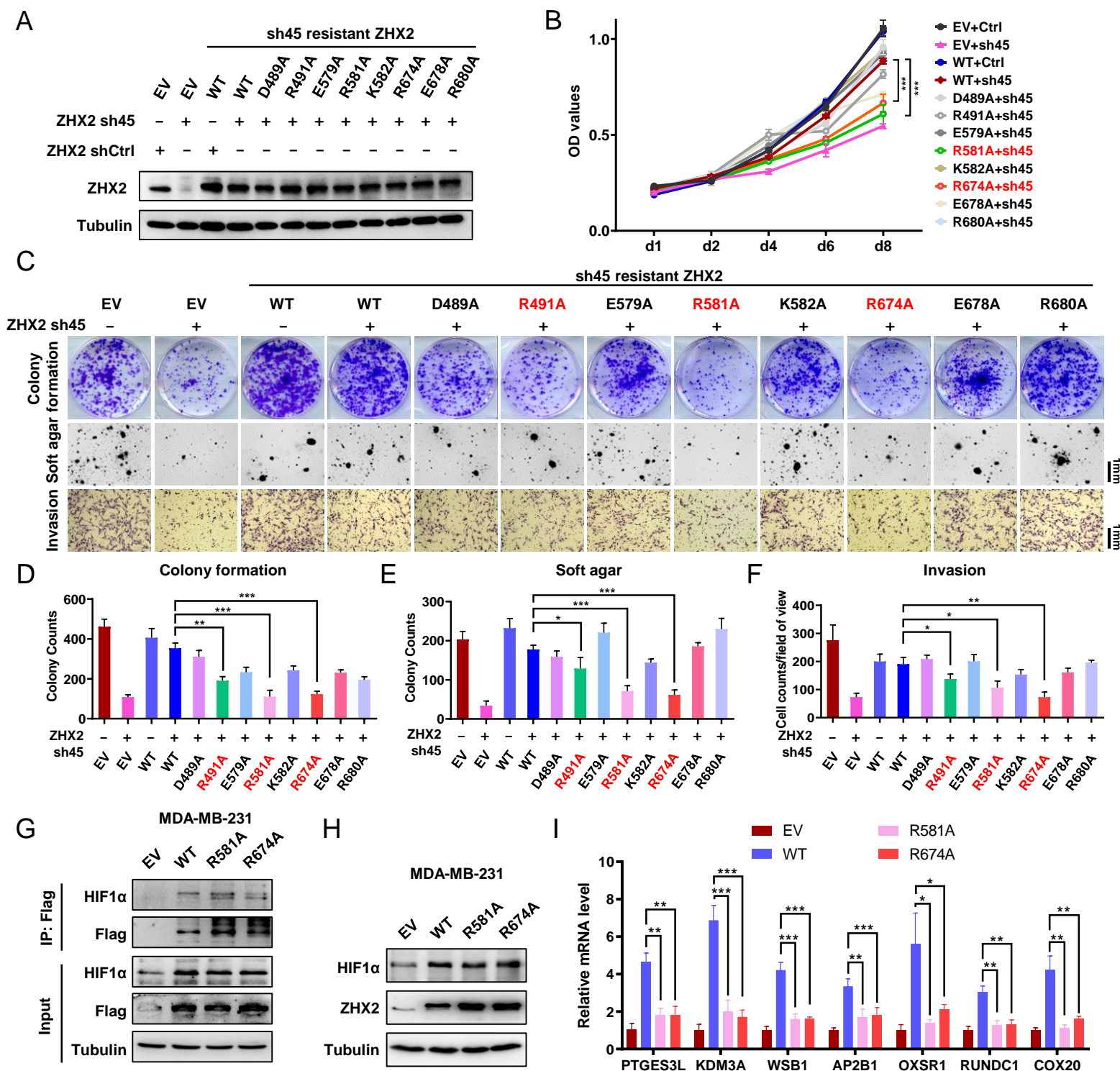**Figure 4**

Figure 5

Figure 6

1   **Supplemental Methods**

2   **Survival analysis.**   The K-M plots were got from https://kmplot.com(9). We chose TNBC

3   patients as follow: ER status-IHC: ER negative, ER status-array: ER negative, PR status-IHC:

4   PR negative, HER2 status-array: HER2 negative. Finally, 153 TNBC patients were included in

5   the overall survival (OS) analysis. ZHX2 overexpression were chosen as upper tertile

6   expression.

7

8   **Luciferase reporter assay.** For HIF transcription assay, sub-confluent MDA-MB-231 cells

9   (200,000 cells/24-well plate) were transiently transfected with 30 ng pCMV-Renilla and100 ng of

10 HRE-Luci reporter. Forty-eight hours after transfection, luciferase assays were performed by

11 Dual-Luciferase® Reporter Assay System (Promega, E1960). The experiments were repeated

12 in triplicate with similar results.

13

14 **DNA-protein contact analysis from structural bioinformatics data.** Sequences in the apo-

15 form human homeodomains 2,3 and 4 (HD2/3/4) (PDB ID: 3NAU, 2DMP and 3NAR) (Bird et al.,

16 2010) (Bank, 2020) were BLASTed against the sequences in the Protein Data Bank

17 (PDB)(Berman et al., 2000). Among the resolved structures, 12, 20 and 15 DNA-bound

18 complexes with sequence identity higher than 30% (Rost, 1999) for HD2, HD3, and HD4 were

19 identified. We then carried out the homology-modeling, using SWISS-MODEL (Waterhouse et

20 al., 2018), to structurally model the HD2/3/4 using their corresponding bound-forms as the

21 templates, so that their sequences assume the protein structures in the DNA-complexed forms.

22 For instance, HD2 sequence could therefore adapt 12 different bound-form protein structures.

23 With this method, we found every HD protein contact the DNA with its last (C-terminal) helix

24 (Supplemental Figure 5). We then count the number of DNA-protein contacts at the atomic level

25 for every residue in the C-terminal helix. The top-ranked residues in HD2/3/4, in terms of their

26   DNA contact frequency normalized by the number of bound-forms used, are listed together with

27   their evolutionary conservation in Supplemental Table 1.

28   **MD simulations**

29   *Homology Modeling to create DNA-bound HD complexes for simulations*

30   Because there are no experimentally solved DNA-complexed structures for human HD2, 3 and

31   4, in order to simulate the human HD-DNA interaction, our goal is to find structurally solved

32   DNA-bound forms whose DNA sequence could have the highest chance to stably interact with

33   human HD2/3/4. We aimed to ensure the highest likelihood of stable interaction between the

34   selected DNA and HD2/3/4 as well as to have a fair comparison of binding ability for HD2/3/4

35   and their DNA-binding residues. To this end, we searched the DNA-bound HD proteins

36   containing a DNA-binding helical stretch that has the highest sequence identity with the C-

37   terminal helices (the main DNA-binding helix; see Supplemental Figure 1 and Video 1) in HD2,

38   3 and 4, respectively. Among the bound-form proteins that have the top 2 highest sequence

39   homology in the C-terminal helices with those in human HD2/3/4, by homology modeling using

40   SWISS-MODEL (Waterhouse et al., 2018), a NMR-resolved structural ensemble of VND/NK-2

41   homeodomain-DNA complex (PDB ID: 1NK2, where the first model is taken) (Gruschus, Tsao,

42   Wang, Nirenberg, & Ferretti, 1997) was chosen to build the DNA-bound form of HD2, an X-ray

43   resolved structure of Yeast MATα2 homeodomain/MCM1 transcription factor/DNA complex

44   (PDB ID: 1MNM;) (Tan & Richmond, 1998)was chosen to build the DNA-bound form of HD3,

45   and an X-ray resolved structure of Oct-1 Transcription factor DNA complex (PDB ID:1HF0;)

46   (Remenyi et al., 2001) was chosen to build the DNA-bound form of HD4, respectively. The

47   homology of C-terminal helices between HD2/3/4 and their corresponding bound-form templates

48   are 54.55%, 53.85%, and 100% respectively.

49   *System Setup and Energy Minimization*

50   Prior to solvation and addition of ions, protonation state and the net charge of HD2/3/4-dsDNA

51   complexes at pH 7.0 were calculated using PDB2PQR(Dolinsky et al., 2007). The starting

52   structure was prepared using ff14SB (Maier et al., 2015)force-fields for proteins, bsc1 (Ivani et

53   al., 2016) force fields for the DNA, TIP3P water model, and monovalent ion parameters (Joung

54   & Cheatham, 2009) through tLeap(Daoudi et al., 2019) from AmberTools18. To neutralize the

55   charge of each system, 24 Na+ were added into hb2-dsDNA and hb3-dsDNA complexes and 23

56   Na+ were added into hb4-dsDNA complex systems. In addition, 23 Na^+ and 23 Cl^- ions were

57   added to each system to reach 100 mM salt concentration. Each system was prepared in a

58   water box measuring 78Å on all sides.

59   Energy minimization for each of the systems was done in two stages. In the first stage, a

60   harmonic restraint of 100 kcal/mol/$Å^2$ was applied on all heavy atoms of both protein and

61   dsDNA. In the second stage, the harmonic restraints for protein's CA atoms were relaxed to 2

62   kcal/mol/$Å^2$ while all the DNA's heavy atoms were still subject to a 100 kcal/mol/$Å^2$ restraint.

63   *Equilibration and Explicit Solvent Production MD Simulations*

64   Each energy-minimized system was gradually heated from 50K to 320K and cooled down to

65   310K in a canonical (NVT) ensemble, using Langevin thermostat (Pastor, Brooks, & Szabo,

66   1988) with a collision frequency of 2 ps^{-1}, for 25 ps while applying harmonic restraints of 10

67   kcal/mol/$Å^2$ on dsDNA's C2, C4', and P atoms and 2 kcal/mol/$Å^2$ on protein's CA atoms. Each of

68   the systems was equilibrated first in a canonical ensemble at 310K for 15 ns. This was followed

69   by an isothermal-isobaric ensemble for 20 ns at 310K applying harmonic restraints of 2

70   kcal/mol/$Å^2$ on dsDNA's C2, C4', and P atoms and 1 kcal/mol/$Å^2$ on protein's CA atoms. Further

71   equilibration isothermal-isobaric ensemble (NPT), where a constant pressure was maintained by

72   Berendsen barostat (H. J. C. Berendsen, 1998) at 1 atm and 310K, was done for 40 ns, while

73   harmonic restraints of 2 kcal/mol/$Å^2$ on dsDNA's C2, C4', and P atoms and 0.1 kcal/mol/$Å^2$ on

74   protein's CA atoms were applied. This was followed by a 350 ns production run at 2 fs time step

75   applying the SHAKE constraint algorithm (Hopkins, Le Grand, Walker, & Roitberg, 2015) to

76   hydrogen atoms in isothermal-isobaric ensemble at 310K and 1 atm. All the simulations were

77   carried out by the AMBER18 software package (Daoudi et al., 2019) with long-range

78   electrostatic forces being calculated using Particle Mesh Ewald method (Tom Darden, 1993) at

79   a 10Å cutoff distance.

80

81   **Legend for Figure-source data**

82   **Figure 1-source data**. Uncropped western blot images for Figure 1.

83   **Figure 2-source data**. Uncropped western blot images for Figure 2.

84   **Figure 3-source data**. Uncropped western blot images for Figure 3.

85   **Figure 4-source data**. Uncropped western blot images for Figure 4.

86   **Figure 6-source data**. Uncropped western blot images for Figure 6.

87   **Figure supplement 1-source data**. Uncropped western blot images for Figure supplement 1.

88   **Figure supplement 2-source data**. Uncropped western blot images for Figure supplement 2.

89   **Figure supplement 3-source data**. Uncropped western blot images for Figure supplement 3.

90   **Figure supplement 4-source data**. Uncropped western blot images for Figure supplement 4.

91

92

93

94 **Supplemental Reference**

95 Bank, R. (2020). RCSB PDB - 2DMP: Solution structure of the third homeobox domain of Zinc
96   fingers and homeoboxes protein 2. . *Rcsb.org.* Retrieved from
97   https://www.rcsb.org/structure/2dmp
98 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E.
99   (2000). The Protein Data Bank. *Nucleic Acids Res, 28*(1), 235-242.
100   doi:10.1093/nar/28.1.235
101 Bird, L. E., Ren, J., Nettleship, J. E., Folkers, G. E., Owens, R. J., & Stammers, D. K. (2010).
102   Novel structural features in two ZHX homeodomains derived from a systematic study of
103   single and multiple domains. *BMC Struct Biol, 10*, 13. doi:10.1186/1472-6807-10-13
104 Daoudi, S., Semmeq, A., Badawi, M., Assfeld, X., Arfaoui, Y., & Pastore, M. (2019). Electronic
105   structure and optical properties of isolated and TiO2 -grafted free base porphyrins for
106   water oxidation: A challenging test case for DFT and TD-DFT. *J Comput Chem, 40*(29),
107   2530-2538. doi:10.1002/jcc.26027
108 Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., & Baker, N. A.
109   (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular
110   structures for molecular simulations. *Nucleic Acids Res, 35*(Web Server issue), W522-
111   525. doi:10.1093/nar/gkm276
112 Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M., & Ferretti, J. A. (1997). Interactions of
113   the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy:
114   basis of binding specificity. *Biochemistry, 36*(18), 5372-5380. doi:10.1021/bi9620060
115 H. J. C. Berendsen, J. P. M. P., W. F. van Gunsteren, A. DiNola, and J. R. Haak. (1998).
116   Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics,*
117   *81*(8).
118 Hopkins, C. W., Le Grand, S., Walker, R. C., & Roitberg, A. E. (2015). Long-Time-Step
119   Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput,*
120   *11*(4), 1864-1874. doi:10.1021/ct5010406
121 Ivani, I., Dans, P. D., Noy, A., Perez, A., Faustino, I., Hospital, A., . . . Orozco, M. (2016).
122   Parmbsc1: a refined force field for DNA simulations. *Nat Methods, 13*(1), 55-58.
123   doi:10.1038/nmeth.3658
124 Joung, I. S., & Cheatham, T. E., 3rd. (2009). Molecular dynamics simulations of the dynamic
125   and energetic properties of alkali and halide ions using water-model-specific ion
126   parameters. *J Phys Chem B, 113*(40), 13279-13290. doi:10.1021/jp902584c
127 Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C.
128   (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone
129   Parameters from ff99SB. *J Chem Theory Comput, 11*(8), 3696-3713.
130   doi:10.1021/acs.jctc.5b00255
131 Pastor, R. W., Brooks, B. R., & Szabo, A. (1988). An analysis of the accuracy of Langevin and
132   molecular dynamics algorithms. *Molecular Physics, 65*(6), 1409-1419.
133   doi:10.1080/00268978800101881
134 Remenyi, A., Tomilin, A., Pohl, E., Lins, K., Philippsen, A., Reinbold, R., . . . Wilmanns, M.
135   (2001). Differential dimer activities of the transcription factor Oct-1 by DNA-induced
136   interface swapping. *Mol Cell, 8*(3), 569-580. doi:10.1016/s1097-2765(01)00336-7
137 Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng, 12*(2), 85-94.
138   doi:10.1093/protein/12.2.85
139 Tan, S., & Richmond, T. J. (1998). Crystal structure of the yeast MATalpha2/MCM1/DNA
140   ternary complex. *Nature, 391*(6668), 660-666. doi:10.1038/35563
141 Tom Darden, D. Y., Lee Pedersen. (1993). Particle mesh Ewald: An N·log(N) method for Ewald
142   sums in large systems *The Journal of Chemical Physics 98*(12), 4.

143   Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., . . . Schwede,
144   T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes.
145   *Nucleic Acids Res, 46*(W1), W296-w303. doi:10.1093/nar/gky427
146
147

Supplemental Figures
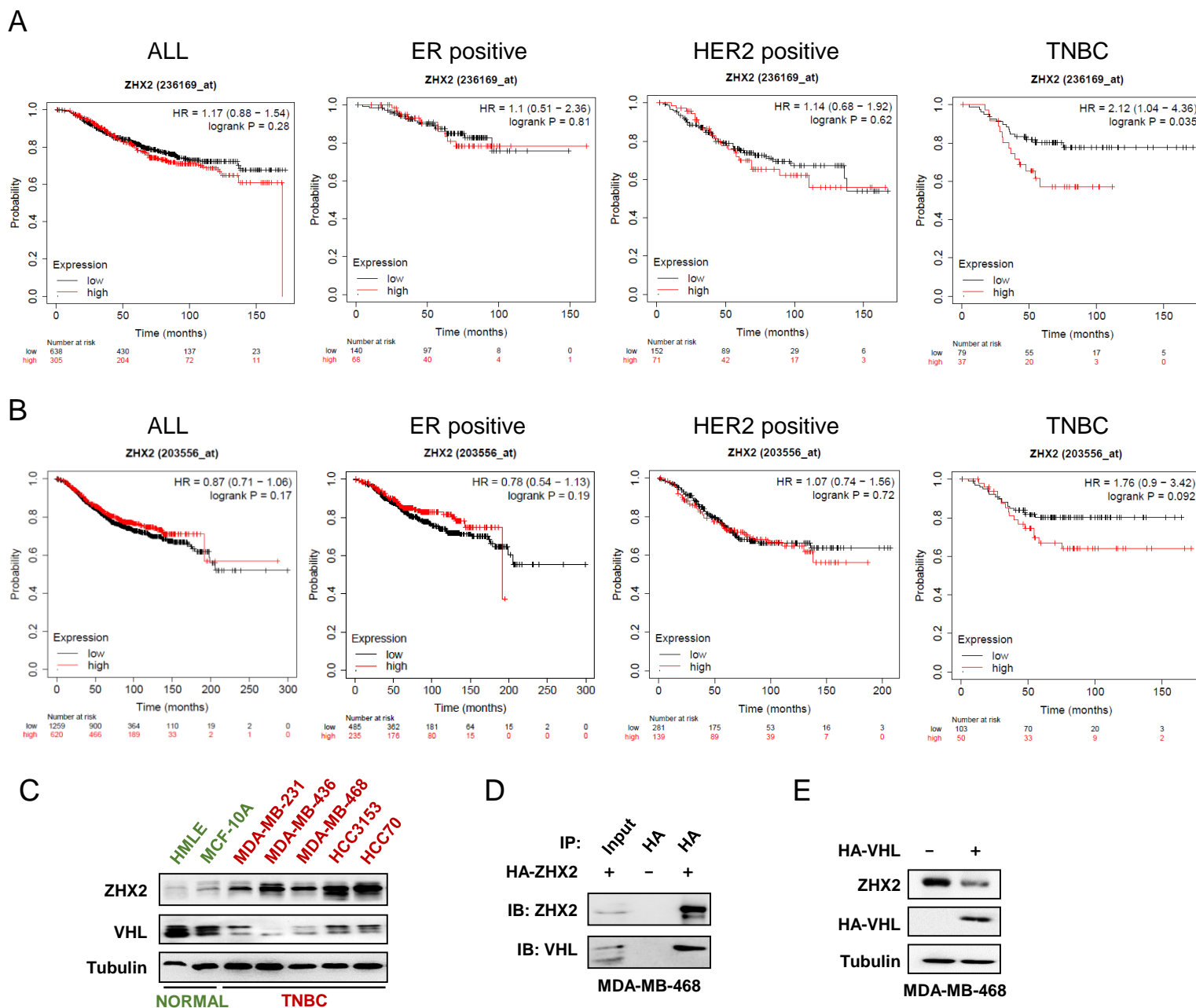
Figure S1. **ZHX2 overexpression leads worse survival and is potentially regulated by pVHL in breast cancer.** (**A-B**) The overall survival of breast cancer patients with high or low ZHX2 expression in different breast cancer subtypes. The K-M plots were generated from https://kmplot.com using two Affymetrix probe, ZHX2: 236169_at (**A**) and 203556_at (**B**). (**C**) Immunoblots of lysates from normal breast epithelial cell and TNBC cell lines. (**D**) Immunoprecipitations of MDA-MB-468 cells expressing either control vector or HA-ZHX2. (**E**) Immunoblots of lysates from MDA-MB-468 cells infected with either control vector or HA-VHL.
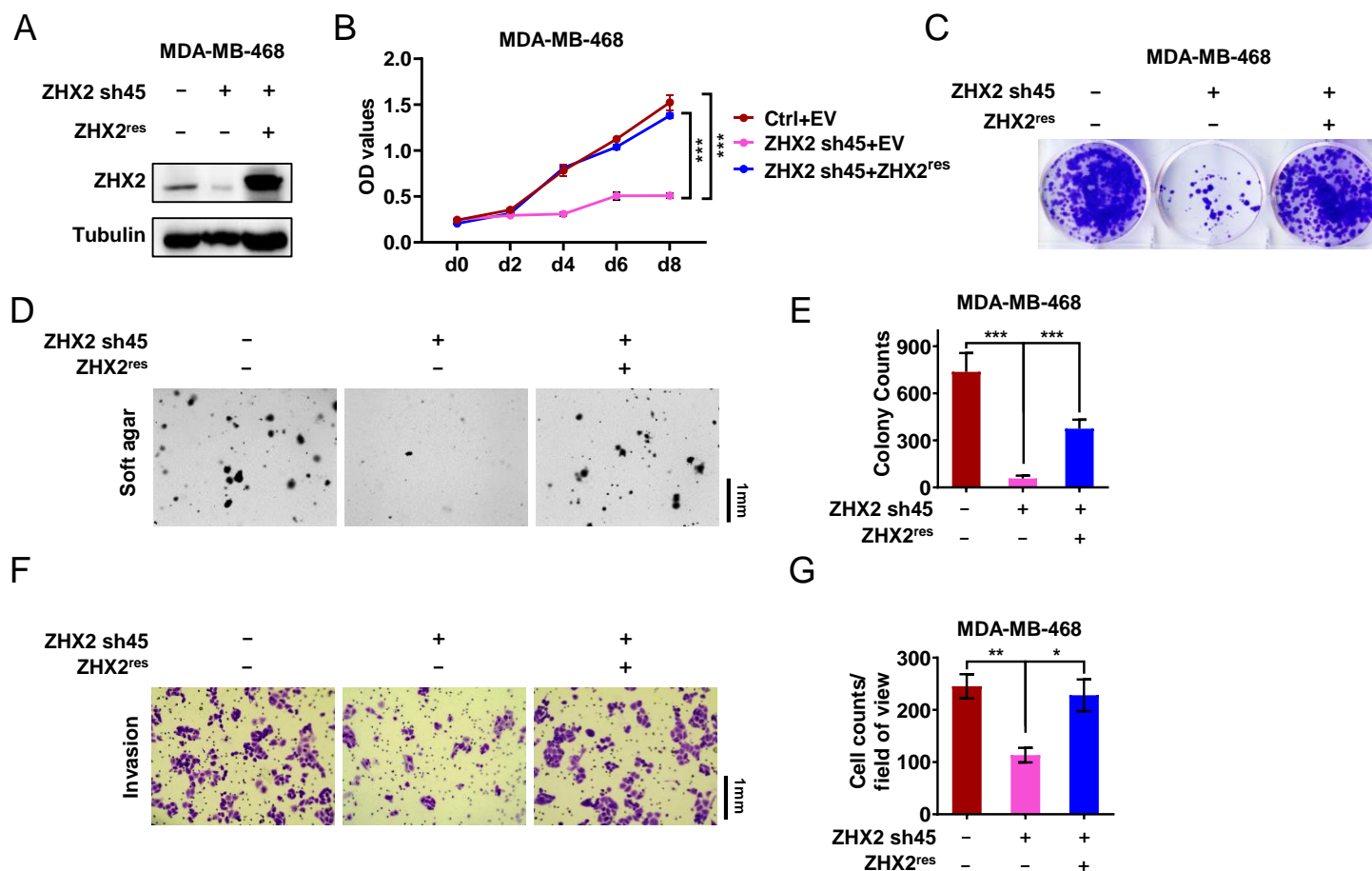
Figure S2. The phenotype of ZHX2 shRNA on cell proliferation and invasion is due to its on-target effect. (**A**) Immunoblot of cell lysates of MDA-MB-468 cells transfected with ZHX2 sh45-resistant ZHX2 (ZHX2^{res}) or empty (EV) vector, followed by ZHX2 sh45 or control (Ctrl) shRNA infection. (**B**) Cell proliferation assays of MDA-MB-468 cells transfected with sh45-resistant ZHX2 (ZHX2^{res}) or empty (EV) vector, followed by ZHX2 sh45 or control (Ctrl) shRNA infection. (**C**) 2-D colony formation assay of MDA-MB-468 cells transfected with sh45-resistant ZHX2 (ZHX2^{res}) or empty (EV) vector, followed by ZHX2 sh45 or control (Ctrl) shRNA infection. (**D-E**) Representative soft agar colony (**D**) and quantification (**E**) of MDA-MB-468 cells transfected with ZHX2 sh45-resistant ZHX2 (ZHX2^{res}) or empty (EV) vector, followed by ZHX2 sh45 or control (Ctrl) shRNA infection. (**F-G**) Invasion assays (**F**) and quantification (**G**) of MDA-MB-468 cells transfected with ZHX2 sh45-resistant ZHX2 (ZHX2^{res}) or empty (EV) vector, followed by ZHX2 sh45 or control (Ctrl) shRNA infection.
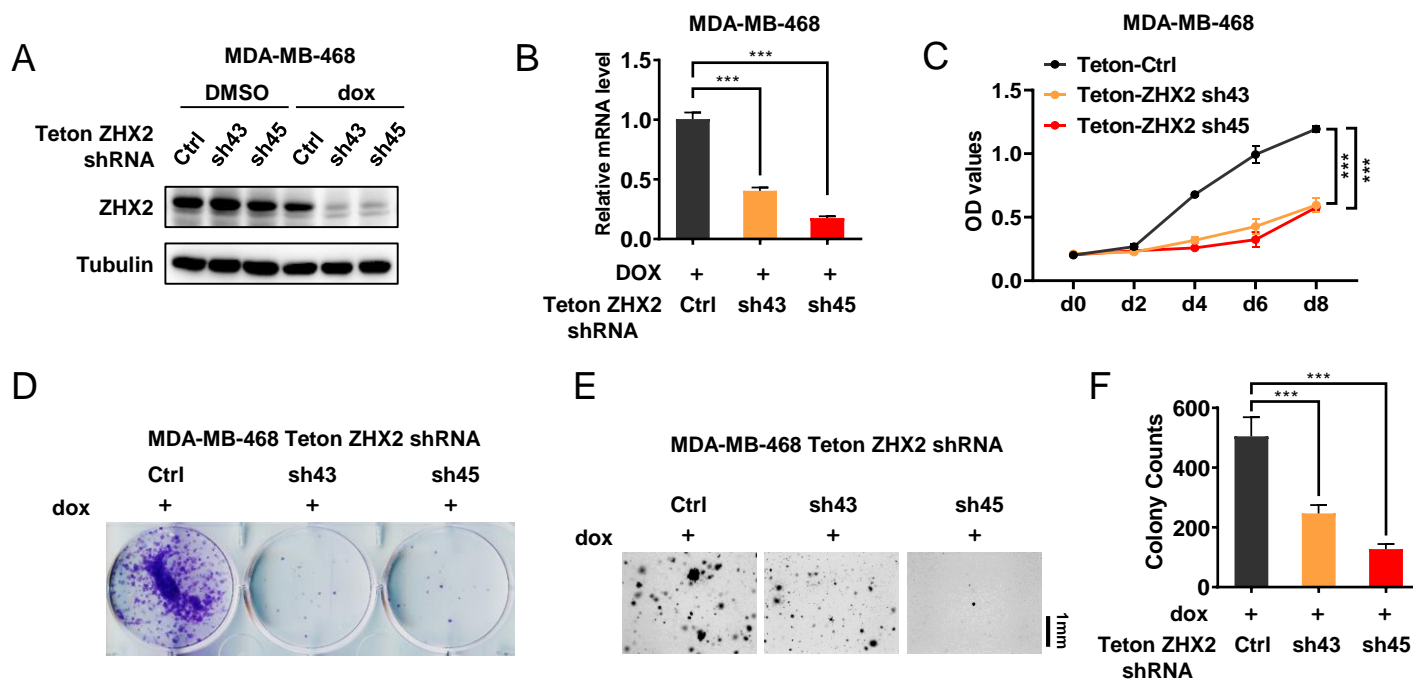
Figure S3. ZHX2 is important for maintaining TNBC tumorigenesis *in vivo*. (A-B) Immunoblot (**A**) and qRT-PCR (**B**) of MDA-MB-468 cells infected with lentivirus encoding either Teton-ZHX2 shRNA (43, 45) or Teton-Ctrl. (**C-D**) Cell proliferation assay (**C**) and 2-D colony formation assay (**D**) of MDA-MB-468 cells infected with lentivirus encoding either Teton-ZHX2 shRNA (43, 45) or Teton-Ctrl. (**E-F**) Representative soft agar colony (**E**) and quantification (**F**) of MDA-MB-468 cell lines infected with lentivirus encoding either Teton-ZHX2 shRNA (43, 45) or Teton-Ctrl.

A

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX
HALLMARK_HYPOXIA	58	0.5568272	2.5719726	0	0	0	912
HALLMARK_KRAS_SIGNALING_UP	78	0.44155815	2.154921	0	8.73E-04	0.002	697
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	64	0.44942394	2.132924	0	0.001757428	0.006	937
HALLMARK_APICAL_JUNCTION	46	0.46112463	2.0707603	0	0.002172801	0.01	402
HALLMARK_UV_RESPONSE_DN	32	0.46017453	1.884963	0.001248439	0.008784941	0.05	627
HALLMARK_TNFA_SIGNALING_VIA_NFKB	64	0.39964557	1.8777542	0	0.007865197	0.054	1049
HALLMARK_COAGULATION	42	0.42355058	1.8189241	0.002375297	0.013317258	0.101	701
HALLMARK_COMPLEMENT	58	0.38593316	1.7986828	0.004581902	0.014482699	0.122	701
HALLMARK_ALLOGRAFT_REJECTION	47	0.35706216	1.5930052	0.02183908	0.06974829	0.513	976
HALLMARK_GLYCOLYSIS	48	0.33867225	1.5142627	0.048292108	0.108907565	0.722	957
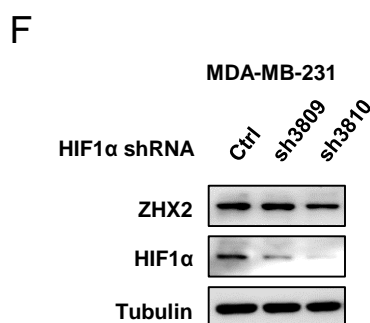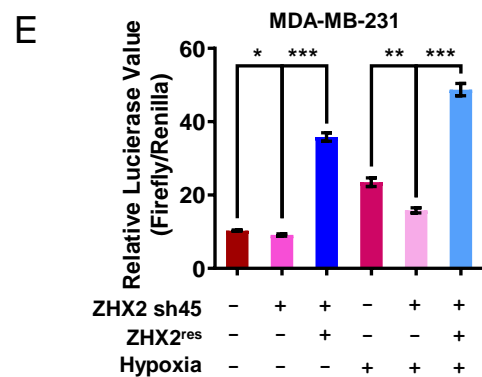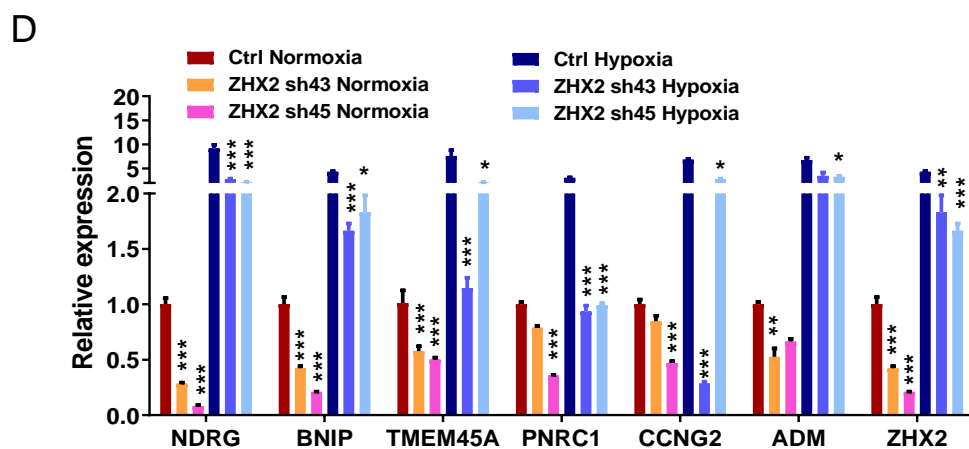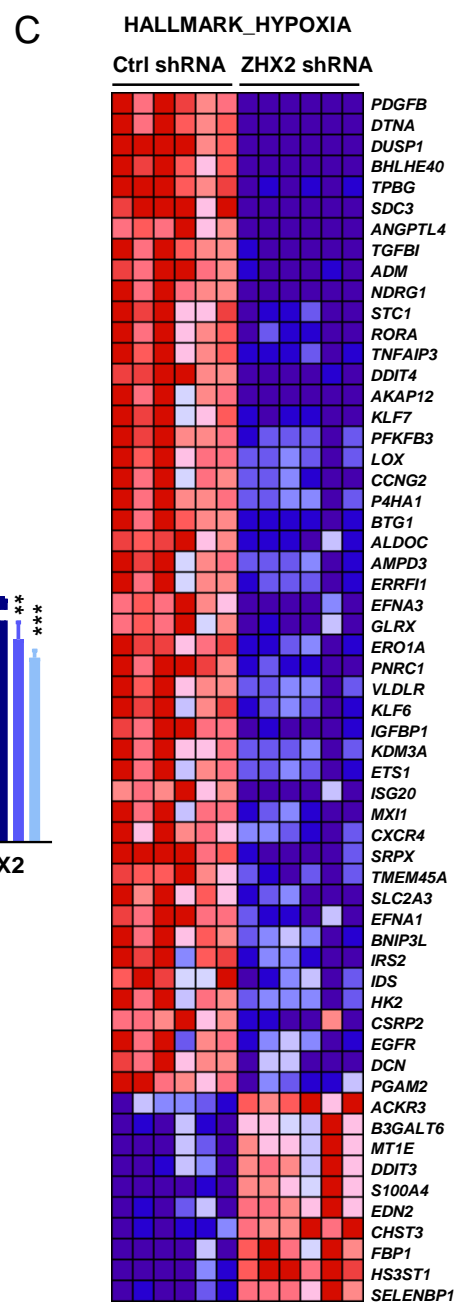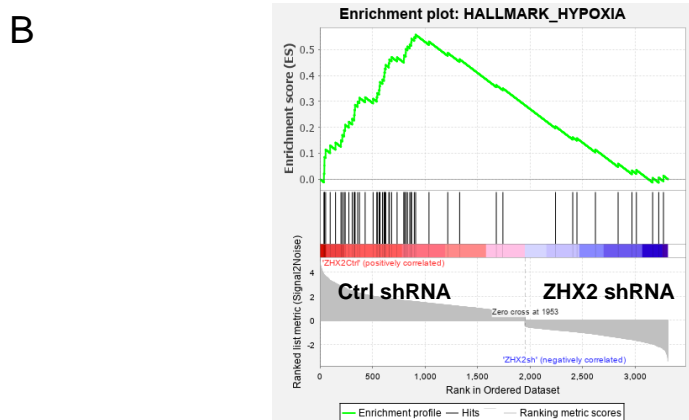
Figure S4. ZHX2 regulates HIF1 signaling in TNBC. (**A**) Gene set enrichment analysis (GSEA) of the significantly decreased pathways in ZHX2 depleted MDA-MB-231 cells. (**B-C**) GSEA plot (**B**) and gene set heatmap (**C**) suggest hypoxia pathway is significantly downregulated in ZHX2 depleted MDA-MB-231 cells. (**D**) qRT-PCR quantification of relative mRNA expression of hypoxia target genes from MDA-MB-231 cells infected with ZHX2 shRNA 43, 45 or Ctrl under normoxia or hypoxia conditions. (**E**) HRE double luciferase gene assay of MDA-MB-231 cells infected with ZHX2 sh45, sh45-resistant ZHX2 (ZHX2^{res}) or Ctrl under normoxia or hypoxia condition. (**F**) Immunoblot of MDA-MB 231 cells infected with lentivirus encoding either HIF1α shRNA 3809, 3810 or Ctrl.
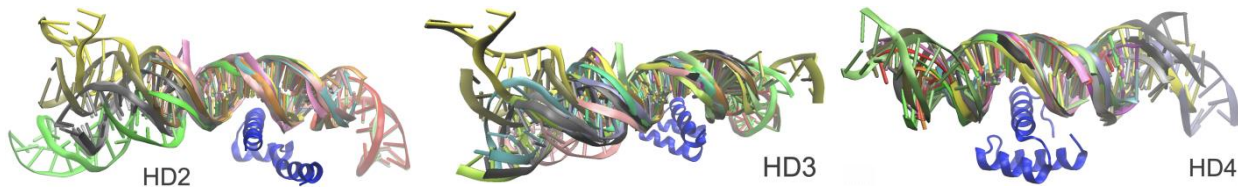
Figure S5. Residue-based DNA contact analysis derived from ensembles of DNA-bound HD proteins.

Supplemental Tables

Table S1. Amplification status of ZHX2 in different breast cancer subtype.

Dataset	Breast cancer patients				TNBC patients				ER positive				HER2 positive			
	Patient	Sample	ZHX2 Amp	Percentage	Patient	Sample	ZHX2 Amp	Percentage	Patient	Sample	ZHX2 Amp	Percentage	Patient	Sample	ZHX2 Amp	Percentage
Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)	2509	2509	512	23.6	320	320	104	32.5	1825	1825	341	21.1	247	247	81	32.8
The Metastatic Breast Cancer Project (Provisional, February 2020)	180	237	44	18.6	5	6	2	33.3	94	126	18	14.3	37	44	10	22.7
Breast Invasive Carcinoma (TCGA, Cell 2015)	817	818	148	18.1	83	83	25	30.1	601	601	90	15.0	121	121	31	25.6
Breast Invasive Carcinoma (TCGA, Firehose Legacy)	1101	1108	197	18.2	116	117	36	32.1	808	814	120	15.1	164	164	39	24.1
Breast Invasive Carcinoma (TCGA, Nature 2012)	825	825	92	11.8	123	123	23	19.3	601	601	58	10.3	114	114	20	18.2

Table S2. Top DNA-contacting residues in HD2/3/4 along with their evolutionary conservation.

Homeobox 2			Homeobox 3			Homeobox 4		
Residue	Mean Total Contacts	Conserv-ation Score*	Residue	Mean Total Contacts	Conserv-ation Score*	Residue	Mean Total Contacts	Conserv-ation Score*
LYS 485	52.50	5	**GLU 579**	31.05	8	**ARG 674**	129.20	6
ARG 491	36.33	9	**ARG 581**	24.80	6	**GLU 678**	58.73	8
PHE 463	32.92	2	SER 575	23.20	5	GLU 671	45.80	7
ARG 493	25.08	5	**LYS 582**	23.00	7	**ARG 680**	28.33	9
ASP 489	22.42	8	PHE 553	13.00	1	LYS 635	27.73	9
GLU 482	14.25	6	GLU 572	9.55	8	LYS 684	25.13	7
LYS 484	11.42	7	ARG 570	9.25	7	LYS 677	23.47	7
TYR 492	9.58	5	TRP 576	7.55	9	TRP 652	15.53	4
ARG 480	8.67	8	ASP 585	4.70	7	TRP 675	10.13	9
TRP 486	7.42	9	SER 578	4.70	5	CYS 681	6.33	7

*obtained using ConSurf server (9 – most conserved, 1 – most diverse); bold-faced residues were chosen for the mutagenesis study on TNBC suppression. DNA-protein contact is defined by heavy atom contact within 4Å, while the same DNA atoms contacted by different protein atoms are considered as separate contacts.

Table S3. MM/PBSA Calculations of the Free Energy (ΔH-TΔS) for Homeobox 2-4 dsDNA complexes.

	ΔH (kcal/mol)	(TΔS) (kcal/mol)	ΔG (kcal/mol)
HD2+DNA	-60.25 ± 0.50	-48.05 ± 0.97	**-12.20 ± 1.09**
HD3+DNA	-56.93 ± 0.46	-45.53 ± 2.39	**-11.40 ± 2.43**
HD4+DNA	-39.00 ± 0.50	-31.23 ± 2.34	**-7.77 ± 2.39**

Table S4. Top C-terminal helix residues in the Homeobox 2-4 contributing the most DNA binding enthalpy.

Homeobox 2		Homeobox 3		Homeobox 4	
Residue number**	ΔH_{total} (kcal/mol)*	Residue number**	ΔH_{total} (kcal/mol)*	Residue number**	ΔH_{total} (kcal/mol)*
ARG 491	-8.75 ± 0.17	ARG 581	-12.51 ± 0.15	ARG 674	-6.81 ± 0.12
ARG 480	-7.98 ± 0.08	ARG 571	-7.54 ± 0.09	LYS 684	-3.46 ± 0.08
ARG 493	-7.41 ± 0.09	LYS 582	-6.98 ± 0.10	LYS 677	-3.18 ± 0.07
LYS 484	-6.49 ± 0.08	ARG 580	-4.43 ± 0.02	ARG 680	-2.41 ± 0.01
LYS 485	-5.00 ± 0.11	ARG 570	-2.66 ± 0.02	ARG 669	-1.78 ± 0.01
TYR 492	-2.88 ± 0.03	ARG 584	-2.24 ± 0.01	TRP 675	-1.45 ± 0.05
ARG 496	-2.12 ± 0.11	SER 578	-0.71 ± 0.06	LEU 682	-0.79 ± 0.02
GLN 495	-1.86 ± 0.07	PHE 577	-0.35 ± 0.01	CYS 681	-0.36 ± 0.02
TRP 486	-0.96 ± 0.06	LEU 583	-0.22 ± 0.00	THR 670	-0.36 ± 0.04
HID 490	-0.64 ± 0.01			ASN 679	-0.24 ± 0.02
SER 481	-0.49 ± 0.03				
PHE 487	-0.28 ± 0.01				
CYS 494	-0.21 ± 0.00				

* ΔH_{total}: enthalpic binding energy derived from MM/PBSA, **this numbering is based on the Uniprot database where the sequence was obtained, highlighted in blue represents the mutated amino acid in the experimental results

Table S5. Real-time PCR primers used in this study.

Genes	Forward Primer	Reverse Primers
β-Actin	AGAAAATCTGGCACCACACC	GGGGTGTTGAAGGTCTCAAA
ZHX2	GATCAGATAGCTGGAGTCAGGC	CACAGCAGTTCTAACAGACTTCC
TMEM45A	TCCTCTCCTTCTCGCCACTT	TGTGTTGGATGGGATCTGGC
PNRC1	TGTTCCGCGATCTTCTCAGG	GCTAGGAAGCTTGTCGCTCA
CCNG2	AGTGATTCCAGAGTGAGCCTT	AAGGCACAGATGCCAAACCTA
ALDOC	CTGCAGCCTCATCTGTTTGC	CATGGTGACAGCTCCCTGTG
AKAP12	CGAGCGCGTCTCCTTCATT	GGGCAAGAGCCAAAAGACG
ADM	ATGAAGCTGGTTCCCGTAGC	TCCACGACTTAGAGCCCACT
NDRG1	CTGCACCTGTTCATCAATGC	AGAGAAGTGACGCTGGAACC
BNIP3	CGCAGACACCACAAGATACCA AC	GCCAGCAAATGAGAGAGCAGC
PTGES3L	GTGTTGAGGACAGCACCGAT	ACACTGGCTTGGAGTTCACTT
KDM3A	GTGCTCACGCTCGGAGAAA	GTGGGAAACAGCTCGAATGGT
WSB1	GGTGTCAGCTTCAAGAGACAAA	AGTCAGGAGAGAATGCACAGC
AP2B1	CTCTTTCCAGACGTAGTGAACTG	GGAGCGGCTCACAGAGATATT
OXSR1	AGGGACGATTACGAGCTGC	TCCGTTTGATTGCCACTTTCTC
RUNDC1	AAGAGGGCAGTTATGACTCGC	GCTGGGTTGACGATCTGAGC
COX20	TAGGATCTGTTGTGGCTGGC	CCAGCATCCCAAAGTCACCA
HIF1α	TATGAGCCAGAAGAACTTTTAGGC	CACCTCTTTTGGCAAGCATCCTG

Movie S1.

The 350-ns MD simulation trajectory of 1NK2-based HB2-dsDNA complex. The HB2-dsDNA complex is in secondary structure cartoon representation. The top C-terminal helix residue that most contacts the DNA, which is ARG491, and the DNA atoms within 4Å distance from it are in licorice representation.

Movie S2.

The 350-ns MD simulation trajectory of 1MNM-based HB3-dsDNA complex. The HB3-dsDNA complex is in secondary structure cartoon representation (violet color: HB3's helix; grey color: dsDNA). The top C-terminal helix residue that most contacts the DNA, which is ARG581, and the DNA atoms within 4Å distance from it are in licorice representation.

Movie S3.

The 350-ns MD simulation trajectory of 1HF0-based HB4-dsDNA complex. The HB4-dsDNA complex is in secondary structure cartoon representation (violet color: HB4's helix; grey color: dsDNA). The top C-terminal helix residue that most contacts the DNA, which is ARG674, and the DNA atoms within 4Å distance from it are in licorice representation.