Hypoxia-induced changes in intragenic DNA methylation correlate with alternative splicing in breast cancer

Deepak Pant[1], **Sathiya Pandi Narayanan**[1,2], **Nagarjun Vijay**[1,*], **Sanjeev Shukla**[1,*]

[1]Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal 462 066, India

Abstract

The tumor microenvironment is marked by gradients in the level of oxygen and nutrients, with oxygen levels reaching a minimum at the core of the tumor, a condition known as tumor hypoxia. Mediated by members of the HIF family of transcription factors, hypoxia leads to a more aggressive tumor phenotype by transactivation of several genes as well as reprogramming of pre-mRNA splicing. Intragenic DNA methylation, which is known to affect alternative splicing in cancer, could be one of several reasons behind the changes in splicing patterns under hypoxia. Here, we have tried to establish a correlation between intragenic DNA methylation and alternative usage of exons in tumor hypoxia. First, we have generated a custom hypoxia signature consisting of 34 genes that are up-regulated under hypoxia and are direct targets of HIF-1α. Using this gene expression signature, we have successfully stratified publicly available breast cancer patient samples into hypoxia positive and hypoxia negative groups followed by mining of differentially spliced isoforms between these groups. The Hypoxia Hallmark signature from MSigDB was also used independently to stratify the same tumor samples into hypoxic and normoxic. We found that 821 genes were showing differential splicing between samples stratified using a custom signature, whereas, 911 genes were showing differential splicing between samples stratified using the MSigDB signature. Finally, we performed multiple correlation tests between the methylation levels (α) of microarray probes located within 1 kilo base pairs of isoform-specific exons using those exons' expression levels in the same patient samples in which the methylation level was recorded. We found that the expression level of one of the exons of DHX32 and BICD2 significantly correlated with the methylation levels, and we were also able to predict patient survival (p-value: 0.02 for DHX32 and 0.0024 for BICD2). Our findings provide new insights into the potential functional role of intragenic DNA methylation in modulating alternative splicing during hypoxia.

Keywords

Hypoxia; alternative splicing; DNA methylation; breast cancer; epigenetics

[*]Corresponding authors (nagarjun@iiserb.ac.in; sanjeevs@iiserb.ac.in).
[2]Present Address: Michigan Center for Translational Pathology, Department of Pathology, University of Michigan, Ann Arbor, MI 48109-0940, USA

1 Introduction

One of the most defining features of the microenvironment of solid tumors is hypoxia caused by an inadequate supply of oxygen owing to the lack of proper vasculature in the core of the tumor (Vaupel and Harrison 2004). Low oxygen levels induce several changes in cancer cells as well as in other associated components of the tumor microenvironment. These changes, which can be mainly attributed to Hypoxia Inducible Factors (HIFs), help in augmenting the metastatic potential of cancer cells (Petrova *et al.* 2018). The major effector of the hypoxia pathway and the most studied member of HIF family of transcription factors is HIF-1, a heterodimer protein consisting of a stable HIF-1β subunit along with HIF-1α, which undergoes ubiquitin-mediated degradation under normal oxygen levels. Upon stabilization, during hypoxia, HIF-1α forms a complex with HIF-1β, which along with cofactors like CBP and p300 activate the transcription of a wide variety of genes involved in processes that are vital for the survival and spread of tumor cells to metastatic sites (Wang *et al.* 1995; Jaakkola *et al.* 2001; Petrova *et al.* 2018).

Like other solid tumors, the activation of the hypoxia pathway in breast cancer has been shown to play an important role by contributing to processes like the formation of new blood vessels (angiogenesis), remodeling of the extracellular matrix, establishment of pre-metastatic niche, invasion and extravasation at metastatic site, among others (Semenza 2016). Apart from the change in the transcriptional activity of genes downstream of HIF-1α, hypoxia has also been shown to induce changes in the alternative splicing of pre-mRNA transcripts, the latter being relatively less understood. Previous studies have reported changes in alternative splicing due to hypoxia in breast cancer cells (Hirschfeld *et al.* 2009; Han *et al.* 2017); however, epigenetic regulation of alternative splicing under hypoxia has not been reported.

There is increasing support for epigenetic mechanisms like DNA methylation as one of the major players that drive cancer progression (Flavahan *et al.* 2017). Moreover, intragenic DNA methylation has also been shown to modulate the alternative splicing of pre-mRNA (Maunakea *et al.* 2013). Earlier, we have reported the role of intragenic methylation in the expression of cancer-specific PKM2 isoform in breast cancer (Singh *et al.* 2017). Also, epigenetics plays a crucial role in cellular response to hypoxia (Watson *et al.* 2010; Wu *et al.* 2015).

Furthermore, hypoxia is shown to result in global changes in DNA methylation due to alterations in DNMTs and TET activities (Watson *et al.* 2012; Thienpont *et al.* 2016). Considering that DNA methylation is known to regulate alternative splicing, it is possible that hypoxia-induced response contributes to cancer-specific, alternatively spliced transcripts, due to changes in DNA methylation. Hence, we hypothesized that alteration in the level of DNA methylation at a global level in tumor hypoxia could be the reason behind widespread changes in alternative splicing that occur in response to hypoxia. In this study, we have identified a gene expression signature indicative of the hypoxia pathway activation. We use this gene expression signature to perform in silico stratification of the publicly available TCGA (The Cancer Genome Atlas) breast cancer RNA-seq gene expression data into hypoxia-positive or hypoxia-negative individuals. This stratification of patient samples

was used to identify differentially used splice isoforms. Finally, to evaluate the possible functional role of intragenic DNA methylation in altering RNA splicing, we looked for statistically significant correlations between exon expression and methylation levels in the genes that were found to be differentially spliced between the hypoxia-positive and hypoxia-negative individuals. The expression level of one of the exons of DHX32 (DEAH-Box Helicase 32) and BICD2 (Protein Bicaudal D Homolog 2) genes showed significant correlation with DNA methylation. Moreover, with the same exons, we were able to predict the overall survival of breast cancer patients, further strengthening our hypothesis.

2   Materials and methods

The complete computational pipeline used in this study is represented in figure 1. Individual steps are described in detail in the following sections. The complete code has been uploaded at https://github.com/erpliiserb/Hypoxia.

2.1   Differential gene expression analysis

Microarray gene expression profiling data of MCF7 cells under hypoxia (cultured at 0% oxygen) was obtained from Gene Expression Omnibus (GSE41491) (Starmans *et al.* 2012) and converted to.CHP format using Expression Console (v 4.1). Expression at 0-hour time-point (normoxia) was compared to expression at hypoxic time-points (1 h to 24 h) using Transcriptome Analysis Console (v 4.0.1.36), as given in table 1.

A fold change   + 1.2 or   − 1.2, with FDR-corrected p-value < 0.05 was considered significant. The list of differentially expressed genes was annotated with the microarray probe locations downloaded using the (Affymetrix HG U133 Plus 2.0) biomaRt (Durinck *et al.* 2009) package in R. To address the issue of a gene being detected by multiple probe sets, we used the jmap function of Jetset (Li *et al.* 2011) package in R to obtain a single, most representative probe set for each gene.

2.2   Identification of HIF-1 α binding sites using ChIP-Seq data

HIF-1α ChIP-sequencing dataset (containing both control and experiment data) was downloaded from GEO (GSE28352) (Schödel *et al.* 2017) and converted to FASTQ format, followed by mapping with Bowtie (v 1.2.2) (Langmead *et al.* 2009). With SAM files generated after mapping, peak calling was performed using 3 peak callers MACS2 (v 2.1.2) (Zhang *et al.* 2008), GEM (v 3.4) (Guo *et al.* 2012) and CCAT (included in Peak Ranger v 1.18) (Xu *et al.* 2010; Feng *et al.* 2011). To detect the presence of the HIF-1α binding site at the promoter region, the positions of transcription start sites (TSS) for all the differentially expressed genes were downloaded from BioMart on Ensembl genome browser 96 (GRCh38.p12 version of the human genome was used). To assign the nearest peak to each of the genes, output of peak calling was intersected with the positions of microarray probes corresponding to differentially expressed genes using closest function of bedtools (v 2.27.1) (Quinlan and Hall 2010). Genes which had HIF-1α ChIP-seq peaks within 2000 base pairs upstream or downstream (promoter region) were considered to be targets of HIF-1α.

2.3 Generation of custom hypoxia signature

To find overlap between genes having HIF-1α ChIP-seq peaks in promoter region and genes up-regulated in at least one of the comparisons mentioned in table 1, a Venn diagram was constructed using an online tool InteractiVenn (Heberle *et al.* 2015). The expression level of the overlapping genes, across all time points (0 h to 24 h) (supplementary table 1) obtained from the microarray dataset (GSE41491) (Starmans *et al.* 2012) was visualized using a heat map generated by the online tool Morpheus (https://software.broadinstitute.org/morpheus/). Those genes which showed a consistent increase in gene expression level across all time-points were included in the custom hypoxia signature. Kendall's correlation coefficient (τ) between the expression level of genes included in the custom hypoxia signature and time (in hours) was found to be largely positive (supplementary table 2). Overlap between up-regulated genes, HIF-1α ChIP-seq peak containing genes and genes included in the Hypoxia Hallmark signature from the Molecular Signature Database (MSigDB) collection (Liberzon *et al.* 2015) was visualized using a five-panel Venn diagram (supplementary figure 1).

2.4 Collection of TCGA data

All of the cancer patient data used in this study were obtained from the TCGA PANCAN (Pan-Cancer) (Network *et al.* 2013) dataset provided on https://xenabrowser.net/. The details of the data obtained are provided in table 2. The set of 787 patient samples for which DNA methylation has been measured using the 450K array are also found in the gene expression estimated using RSEM, transcript expression estimated using kallisto, exon expression estimated based on read counts and curated clinical dataset. These patient samples with multiple types of data form the basis of our study.

2.5 Stratification of TCGA breast cancer samples

A subset of PANCAN gene expression data containing only breast cancer patient sample (Primary tumor and metastatic) data was used (n = 1,099) for this analysis. The arithmetic sum of expression levels of genes included in our custom hypoxia signature was calculated for each patient tumor sample to get the total hypoxia activation score. The samples were then stratified on the basis of the hypoxia activation score, similar to a protocol by Yang *et al.* (2018). The tumor samples having higher scores (top 25%, 20% and 10%) were assumed to be hypoxic, whereas those having lower scores (bottom 25%, 20% and 10%) were assumed to be normoxic. Same procedure was repeated with genes of the Hypoxia Hallmark signature from MSigDB (Liberzon *et al.* 2015). To verify the distinctiveness of these two groups of individuals we performed unsupervised clustering of gene expression data using K-means clustering implemented in R (v 3.6). The sensitivity, specificity and accuracy of stratification were calculated using confusionMatrix function of R package caret (Kuhn 2008). This package calculates these parameters using two lists. One list contains clusters generated on the basis of hypoxia activation score ('True Clusters') and the other contains clusters generated after K-means clustering ('Predicted Clusters'). The functioning of this package is explained in table 3.

The sensitivity, specificity and accuracy are thus calculated as follows:

$$Sensitivity = TP/TP + FN,$$
$$Specificity = TN/TN + FP,$$
$$Accuracy = (Sensitivity + Specificity)/2.$$

Clusters were defined at the thresholds of 25%, 20% and 10%. The stratification accuracy, sensitivity and specificity were highest for the top and bottom 10% cluster (table 4). Thus, the top 10% (n = 110 samples) and bottom 10% samples (n = 110 samples), according to the hypoxia activation score, were chosen for further analysis. The clusters identified by K-means clustering were visualized in a Principal Component Analysis (performed using prcomp function in R) of the gene expression values with each individual sample colored as either red (normoxia) or blue (hypoxia). The Receiver Operating Characteristic (ROC) curves were generated using the R package ROCR (Sing *et al.* 2005).

2.6 Differential isoform expression analysis

Transcript expression data from TCGA (see table 2 for details) was used to calculate the Isoform fraction (IF) for all transcripts, for only those samples included in our hypoxic and normoxic clusters. The transcript expression level (est_counts) was converted into raw counts (i) by using the formula

$$antilog2(est\_count) - 1$$

Gene and transcript ID for all human genes (GRCh38.p12) were downloaded from BioMart on Ensembl genome browser 96. Isoform fraction was calculated for all isoforms (i_{1-n}) of each gene as follows:

$$IF = \frac{i_n}{\Sigma_{i1-n}}.$$

The mean of IF for both hypoxic and normoxic individuals was calculated, and the absolute difference in isoform fraction was obtained (|dIF|). Only those genes were considered to have had a switch in isoform usage for which at least 2 isoforms were showing a |dIF| value > 0.2 change between hypoxia and normoxia.

2.7 Correlation test for functional intragenic methylation

Methylation data measured in the form of Beta values (β) corresponding to Illumina 450K methylation probes spanning different positions in the genome for breast cancer patients was obtained from the TCGA database. The isoform-specific exons were identified using bedtools subtract function with the—A flag (i.e., remove entire feature if it overlaps any other exon) with the setting—f 0.99 (i.e., at least 1% of the exon should be isoform specific). Moreover, isoforms with expression less than 1 TPM across all TCGA breast cancer samples were not considered when identifying isoform-specific exons. Using closest function of bedtools (Quinlan and Hall 2010), bed file containing methylation probe positions was overlapped with the file having hypoxia/normoxia isoform-specific exon positions. The 5 closest probes present within 1 kb upstream or downstream to each hypoxia/normoxia

isoform-specific exon were identified. Correlation tests was performed between exon expression and methylation level, both of which were measured in the same patient samples (n = 787 samples). Both Kendall and Pearson's correlation coefficients were calculated for each probe–exon pair using R. The criteria followed to perform correlation test between exon expression and DNA methylation has been explained by a model in figure 2.

2.8   Survival analysis based on isoform-specific exon expression

Curated clinical data for TCGA cancer samples was downloaded from https://pancanatlas.xenahubs.net. The expression level of exons belonging to DHX32 and BICD2 that showed significant correlation with methylation was used in this analysis. The dataset used was Exon Expression RNASeq data (HiSeqV2_exon). The 1,104 samples for which exon expression data is available were sorted based on the expression levels of normoxia-specific exons of DHX32 (ENSE00001447872) and BICD2 (ENSE00001404188) that showed correlation with methylation. The samples with highest exon expression levels (top 25%, n = 276) were assumed to be normoxic and the samples with the lowest exon expression levels (bottom 25%, n = 276) were assumed to be hypoxic. The cohort of these 552 TCGA samples was used to perform survival analysis using survival package in R. The survminer package in R was then used to construct a Kaplan–Meier curve.

3   Results

3.1   Generating a novel gene expression signature to distinguish tumor sample data on the basis of hypoxia pathway activation

Our first objective was to narrow down upon a set of genes that are indicative of hypoxia. The Hypoxia Hallmark signature from MSigDB (Liberzon *et al.* 2015) is a previously published gene set that has been generated using 87 founder gene sets. These 87 founder gene sets vary considerably in their model of study, techniques used for quantifying mRNA expression and tissue of origin. More than 50% of the 87 gene sets consist of genes are involved in various metabolic pathways and represent the healthy state of human cells (for example REACTOME_GLUCOSE_METABOLISM dataset, KEGG_PYRUVATE_METABOLISM dataset, etc.). Out of the 22 datasets originating from cancer cell lines (mouse and human), only 4 are derived from breast cancer (MCF7) (Elvidge *et al.* 2006) (supplementary table 3). A dataset consisting of HIF-1α target genes has also been included as a founder gene set, but it is not derived from breast cancer model (Semenza 2001). Thus, in order to generate a signature that is more specific to our model of study, we came up with a custom hypoxia signature.

Using expression levels of the genes included in our signature, we must be able to stratify tumor samples in silico into hypoxia positive or hypoxia negative. For this, we wanted to prioritize those genes that are both up-regulated under hypoxia and also have a HIF-1α binding site in their promoter region. We used publicly available microarray transcriptome profile (Starmans *et al.* 2012) (GSE41491) of MCF7 cells under several time-points after hypoxic exposure. After performing differential gene expression analysis, we found that 3,682 genes were getting differentially expressed under hypoxia, out of which 1,818 were getting up-regulated. To check whether the genes differentially expressed in our analysis are

having HIF-1α binding site, we analyzed the ChIP-sequencing dataset for HIF-1α (GSE28352) (Schödel *et al.* 2017). For identifying HIF-1α binding sites, we performed ChIP-seq peak calling using 3 different peak callers (MACS2, GEM and CCAT). Only those peaks that were present within 2 kilo base pairs upstream or downstream of the transcription start sites of differentially expressed genes were considered. The number of differentially expressed genes having putative binding site for HIF-1α at promoter region was 67, 74, and 153 for MACS2, GEM, and CCAT, respectively. Out of the 1818 upregulated genes, 43 were also having HIF-1α binding site (identified as having a HIF-1α peak by all 3 peak callers) (figure 3A). The trend of gene expression for these 43 genes was visually evaluated to shortlist 34 genes based on a consistent trend of increase in gene expression across all time-points. These 34 genes were selected to be a part of our custom hypoxia signature (figure 3B, supplementary table 1). On comparison with the Hypoxia Hallmark signature from MSigDB, we found that only 17 genes show up-regulation upon hypoxia treatment, have HIF-1α binding site in their promoter regions and are also contained in the MSigDB signature (supplementary figure 1). This could be due to the fact that many genes included in the signature from MSigDB are not direct targets of HIF-1α.

3.2 Stratification of TCGA tumor samples

Gene expression data of breast cancer patients from The Cancer Genome Atlas was used to stratify patient samples into normoxic or hypoxic separately using the custom hypoxia signature and Hypoxia Hallmark signature from MSigDB. After getting the hypoxia activation scores for each of the genes included in our signature, the patient samples with highest scores (top 10%) were considered hypoxia positive and the samples with lowest activation score (bottom 10%) were considered hypoxia negative. We were able to stratify a total of 220 samples of invasive breast carcinoma into hypoxic (n = 110 samples) and normoxic clusters (n = 110 samples) (supplementary table 4). K-means clustering revealed that the two clusters are partitioning distinctively (figure 4A). Similarly, for the Hypoxia Hallmark signature from MSigDB, we stratified the same number of samples into hypoxic and normoxic clusters on the basis of hypoxia activation scores calculated using the 200 genes included in this signature (supplementary table 4). Again, on performing K-means clustering, these two set of samples partitioned into distinct clusters (figure 4C). The accuracy, sensitivity and specificity of the stratification were 1.0, as indicated by the ROC curves (figure 4B and D).

3.3 Global changes in alternative splicing under hypoxia

After distinguishing samples on the basis of activation of the hypoxia pathway, we now checked the differences in isoform usage at a global level between hypoxic and normoxic samples. On calculating the difference between mean Isoform Fraction for all isoforms across hypoxic and normoxic samples (|dIF|), we found that a total of 821 genes had at least 2 transcripts that were showing large changes between the hypoxic and normoxic samples (supplementary table 5). One of the genes showing such a change was BARD1 (BRCA1 Associated RING Domain 1) (supplementary figure 2A and B). It is known that BARD1 can be both oncogenic and tumor suppressor, depending on which of the several transcripts is getting expressed (Cimmino *et al.* 2017). The full-length transcript acts as a tumor

suppressor in breast and ovarian cancer, whereas isoforms that lack the RING domain act as protooncogene.

Similarly, for samples stratified using the MSigDB signature 911 genes showed at least 2 isoforms with significant changes between hypoxia and normoxia (supplementary table 6). One of the genes showing change was CPEB1 (cytoplasmic polyadenylation element binding protein 1) (supplementary figure 2C, 2D). CPEB1 is known to suppress metastasis in breast cancer by post-transcriptional modification of mRNAs of metastasis associated genes (Nagaoka *et al.* 2015). Another gene Lethal 3 Malignant Brain Tumor-Like Protein 4 (L3MBTL4), had 3 isoforms, which were showing differential expression between hypoxia and normoxia (supplementary figure 3A and B). L3MBTL4 acts as a tumor suppressor and is known to have deletions in a protein coding region in breast cancer cell lines and tissue (Addou-Klouche *et al.* 2010).

3.4 Correlation between intragenic DNA methylation and exon expression level

While isoform expression is biologically more intuitive than exon expression level, estimation of isoform expression level using short read sequencing data has been shown to be error prone (Soneson *et al.* 2016). Moreover, we have hypothesized that intragenic methylation would affect the inclusion/exclusion of nearby exons resulting in alteration of isoform expression level. Hence, comparison of exon expression level with the methylation level of nearby probes is better than comparing the isoform expression level with methylation level of all the probes located within the gene. Based on this rationale we performed correlation test between DNA methylation and expression of isoform-specific exons belonging to genes that show differential splicing between samples stratified as hypoxic and normoxic (figure 2).

In order to evaluate the potential role of intragenic DNA methylation in modulating alternative splicing we correlated the expression levels of hypoxia-/normoxia-specific exons with the DNA methylation levels of microarray probes located within 1 kb of these exons. The correlation tests between exon-microarray probe pairs were restricted to isoform-specific exons from only those transcripts that showed a large change in the isoform usage pattern between hypoxic and normoxic patient samples. A negative correlation between exon expression and DNA methylation indicates that fewer transcripts with that exon occur when the adjoining (within 1 kb) DNA is hyper-methylated. Whereas, a positive correlation between exon expression and DNA methylation indicates more transcripts with that exon occur when the adjoining DNA is hypermethylated. The correlation test can only be performed if a methylation array probe is located within 1 kb of an isoform-specific exon. However, we found that the distribution of methylation probes and presence of isoform-specific exons is variable across the genome. This results in many genes lacking probes and isoform-specific exons. Even if probes and isoform-specific probes exist within the same gene, they might not be located within 1 kb. In all these cases, it is not possible to test for a correlation between exon expression level and DNA methylation. The complete set of differentially spliced isoforms and the presence of probe pairs are provided in supplementary table 7. The summary of how many genes lack methylation array probes, isoform-specific exons and exon-probe pairs is provided in table 5.

We used the criteria of both Pearson and Kendall rank correlation coefficients less than – 0.1 or more than 0.1, after FDR correction with a threshold of q-value < 0.01. While considering normoxia-specific exon expression and DNA methylation, 60 exon–probe pairs corresponding to 16 genes were identified using custom signature based stratification of samples (supplementary table 8). Similarly, for the exons that were hypoxia-specific, we found significant correlations for 158 exon–probe pairs corresponding to 48 genes. One of the genes for which a normoxia-specific exon showed positive correlation was DHX32 (DEAH Asp-Glu-Ala-His Box Polypeptide 32). The 1st exon of the normoxic-specific exon of DHX32 was showing positive correlation with DNA methylation at a probe located within 1 kb of the exon (figure 5A and B). DHX32 is an RNA helicase and belongs to DEAD/H-box family of proteins. It is up-regulated in colorectal cancer and is known to induce expression of Vascular Endothelial Growth Factor A (VEGFA) in a β-catenin dependent mechanism, contributing to angiogenesis, which is a hypoxia dependent process (Lin *et al.* 2017).

The correlation tests were also repeated for isoform-specific exons belonging to differentially used isoforms in samples stratified using MSigDB signature. A total of 116 exon–probe pairs belonging to normoxia-specific exons corresponding to 40 genes were showing significant correlation. For hypoxia-specific exons in MSigDB set, 202 exon–probe pairs belonging to 86 genes were showing significant correlation between exon expression and DNA methylation (supplementary table 8). One of the genes from this set was BICD2 (Protein Bicaudal D Homolog 2). The expression of 7th exon belonging to the normoxia-specific isoform of BICD2 negatively correlated with DNA methylation levels (figure 5C and D). BICD2 is a motor adaptor protein. In a recent study it was found that BICD2 mediates the translocation of HIF-1α into the nucleus under hypoxia in umbilical cord blood-derived mesenchymal stem cells. Depletion of this protein caused increased ROS (reactive oxygen species) accumulation and apoptosis establishing its role as a major factor involved in adaptation towards hypoxia (Lee *et al.* 2019). Interestingly, in the CELSR2 gene we found seven exon–probe pairs that showed negative correlation between the expression level of hypoxia-specific exons and DNA methylation of nearby probes (supplementary figure 4). However, the hypoxia-specific isoform of CELSR2 was protein coding and the normoxia-specific isoform does not code for any protein.

3.5 Prediction of patient survival based on isoform-specific exon expression level

The association between the expression levels of differentially expressed isoform-specific exons corresponding to all TCGA breast cancer samples (n = 1,104 samples) and overall patient survival was evaluated. The normoxia-specific exons of DHX32 (1st exon) and BICD2 (7th exon) that showed significant correlation with methylation were used in this analysis. This was done to assess the prognostic significance of changes in alternative splicing under hypoxia in breast cancer. It was found that the 1st exon of normoxia-specific isoform of DHX32 that showed positive correlation with methylation was significantly associated with patient survival (p-value = 0.02) (figure 6A). This implies that the higher expression of this exon due to increased methylation in the downstream region is more favorable for the survival of patients of breast cancer. Whereas, decreased level of this exon leading to increased usage of the longer isoform of DHX32 is associated with increased

mortality in breast cancer patients. Similarly, the 7th exon of normoxia-specific isoform of BICD2 was also significantly associated with patient survival (figure 6B). This exon showed significant negative correlation with methylation indicating that its decreased expression in hypoxic patient samples is associated with higher mortality. On the other hand, the patients stratified as having hypoxia-negative tumors had better chances of survival.

4 Discussion

Hypoxia is one of the most distinctive features of breast cancer that contributes to increased invasion and metastasis. Hypoxia can also reduce the responsiveness of tumor towards chemotherapeutic drugs by decreasing susceptibility to DNA damage, which subsequently leads to less favorable patient prognosis (Rundqvist and Johnson 2013). Apart from the transcriptional activation of genes, changes in alternative splicing events also contribute to an aggressive phenotype of breast cancer under hypoxia. The rewiring of alternative splicing under hypoxia and its underlying causes are under explored. In the present study, we have used publicly available breast cancer data and stratified patient samples into hypoxic and normoxic, independently, using two hypoxia signatures. Hypoxic and normoxic conditions can vary within a patient and is known to exhibit cell to cell heterogeneity (Sutherland *et al.* 1996). Thus, identification of a hypoxic niche at the cellular level would require data resolution at the single cell level for a large number of cells. We have used the hypoxia signature to stratify the patients into hypoxic/normoxic groups as an alternative to analysis of gene expression of hypoxic/normoxic regions of the same tumor. Hence, our hypoxia-signature-based stratification of patients into hypoxic/normoxic groups reflects the magnitude of hypoxia activation. Following the stratification of samples, we also investigated changes in the usage of differentially spliced isoforms between hypoxic and normoxic samples. Finally, we found frequent correlation between differential splicing and the change in intragenic DNA methylation levels.

We began our analysis by generating a custom hypoxia signature by shortlisting a set of genes that have HIF-1α binding site at the promoter and, also, show up-regulation in our analysis. We have also taken into consideration another set of genes from the Molecular Signature Database (MSigDB) (Liberzon *et al.* 2015) to stratify our tumor samples. None of TCGA breast cancer samples classified as hypoxic by either of the signatures was found to be normoxic by the other signature. Similarly, none of the samples classified as normoxic by either of the signatures was found to be hypoxic by the other signature. However, 29 of the 110 samples were found to be hypoxic by both signatures, and 27 of the 110 samples were found to be normoxic by both signatures. The hypoxia gene signature from MSigDB consists of 200 genes that include both direct and indirect targets of HIF-1α, as evident from our data (supplementary figure 1). The advantage of using the custom hypoxia signature in our analysis is that it only includes direct targets of HIF-1α, which makes our signature more reliable. Moreover, our signature is derived from datasets that are based on breast cancer cell line (MCF7). On the other hand, the signature from MSigDB has been generated using several founder datasets, out of which only a few are derived from breast cancer cells (supplementary table 3). However, both signatures are able to demonstrate a correlation between exon expression level and DNA methylation.

To evaluate the proposed role of DNA methylation in modulating mRNA splicing, we correlated estimates of exon expression level obtained from RNA-seq data with the methylation level of micro-array probes located within 1 kb of the differentially spliced exons. Both, the methylation and gene expression levels, were measured in the same patient samples. Since we rely upon a micro-array dataset to measure the methylation level, the entirety of the genome cannot be covered. However, we found that a majority of the differentially spliced isoform-specific exons were located within 1 kb of micro-array probes (table 5). Future efforts that provide genome-wide methylation level measurements using sequencing-based approaches will be able to overcome this shortcoming. Performing such pairwise correlations for all exon and micro-array probes is computationally very expensive and will lead to a very large number of statistical tests. By focusing only on those exons that are unique to isoforms that show a striking change in the pattern of isoform usage between hypoxic and normoxic cancer samples, we reduce the number of statistical tests that need to be performed. This focused approach resulting in increased statistical power is motivated by our hypothesis that functional intragenic DNA methylation is involved in modulating splicing during hypoxia.

In conclusion, we utilize both publicly available microarray gene expression estimates at different time points after hypoxia induction and HIF-1α ChIP-seq data to identify a custom hypoxia signature. Based on this custom signature, we assign a subset of the TCGA breast cancer cohort into hypoxic and normoxic samples. Differentially spliced exons between the normoxic and hypoxic patient samples are identified based on changes in isoform fraction. Correlations between the expression levels of isoform-specific exons and DNA methylation level of microarray probes within 1 kb are seen in 16 genes (for normoxia-specific exons) and 48 genes (for hypoxia-specific exons) while using the custom signature. While using the MSigDB signature for patient sample stratification, normoxia-specific exons from 40 genes and hypoxia-specific exons from 86 genes show correlations between exon expression level and DNA methylation. Our results strongly support a potentially functional role for intragenic DNA methylation in modulating alternative splicing during hypoxia. Candidates identified in our genome-wide screen can form the basis of future studies to understand the mechanism of intragenic DNA-methylation mediated alternative splicing under hypoxia.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Addou-Klouche L, Adélaïde J, Finetti P, Cervera N, Ferrari A, Bekhouche I, Sircoulomb F, Sotiriou C, Viens P, Moulessehoul S, Bertucci F, et al. Loss, mutation and deregulation of L3MBTL4 in breast cancers. Mol Cancer. 2010; 9:213. [PubMed: 20698951]

Cimmino F, Formicola D, Capasso M. Dualistic role of BARD1 in cancer. Genes. 2017; 8:375.

Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009; 4:1184–1191. [PubMed: 19617889]

Elvidge GP, Glenny L, Appelhoff RJ, Ratcliffe PJ, Ragoussis J, Gleadle JM. Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition the role of HIF-1α, HIF-21α, and other pathways. J Biol Chem. 2006; 281:15215–15226. [PubMed: 16565084]

Feng X, Grossman R, Stein L. PeakRanger: A cloud-enabled peak caller for ChIP-seq data. BMC Bioinform. 2011; 12:139.

Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. Science. 2017; 357:eaal2380. [PubMed: 28729483]

Guo Y, Mahony S, Gifford DK. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. PLOS Comput Biol. 2012; 8:e1002638. [PubMed: 22912568]

Han J, Li J, Ho JC, Chia GS, Kato H, Jha S, Yang H, Poellinger L, Lee KL. Hypoxia is a key driver of alternative splicing in human breast cancer cells. Sci Rep. 2017; 7

Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinform. 2015; 16:169.

Hirschfeld M, zur Hausen A, Bettendorf H, Jäger M, Stickeler E. Alternative splicing of Cyr61 is regulated by hypoxia and significantly changed in breast cancer. Cancer Res. 2009; 69:2082–2090. [PubMed: 19244129]

Jaakkola P, Mole DR, Tian Y-M, Wilson MI, Gielbert J, Gaskell SJ, von Kriegsheim A, Hebestreit HF, Mukherji M, Schofield CJ, Maxwell PH, et al. Targeting of HIF-α to the von Hippel-Lindau ubiquitylation complex by O_2-regulated prolyl hydroxylation. Science. 2001; 292:468–472. [PubMed: 11292861]

Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008; 28:1–26. [PubMed: 27774042]

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]

Lee HJ, Jung YH, Oh JY, Choi GE, Chae CW, Kim JS, Lim JR, Kim SY, Lee S-J, Seong JK, Han HJ. BICD1 mediates HIF1α nuclear translocation in mesenchymal stem cells during hypoxia adaptation. Cell Death Differ. 2019; 26:1716–1734. [PubMed: 30464225]

Li Q, Birkbak NJ, Gyorffy B, Szallasi Z, Eklund AC. Jetset: selecting the optimal microarray probe set to represent a gene. BMC Bioinform. 2011; 12:474.

Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. Cell Syst. 2015; 1:417–425. [PubMed: 26771021]

Lin H, Fang Z, Su Y, Li P, Wang J, Liao H, Hu Q, Ye C, Fang Y, Luo Q, Lin Z, et al. DHX32 promotes angiogenesis in colorectal cancer through augmenting β-catenin signaling to induce expression of VEGFA. EBioMedicine. 2017; 18:62–72. [PubMed: 28330603]

Maunakea AK, Chepelev I, Cui K, Zhao K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. 2013; 23:1256–1269. [PubMed: 23938295]

Nagaoka K, Fujii K, Zhang H, Usuda K, Watanabe G, Ivshina M, Richter JD. CPEB1 mediates epithelial-to-mesenchyme transition and breast cancer metastasis. Oncogene. 2015; 35:2893. [PubMed: 26411364]

Network CGAR. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The cancer genome atlas pan-cancer analysis project. Nat Genet. 2013; 45:1113–1120. [PubMed: 24071849]

Petrova V, Annicchiarico-Petruzzelli M, Melino G, Amelio I. The hypoxic tumour microenvironment. Oncogenesis. 2018; 7:10. [PubMed: 29362402]

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]

Rundqvist H, Johnson RS. Tumour oxygenation: Implications for breast cancer prognosis. J Intern Med. 2013; 274:105–112. [PubMed: 23844914]

Schödel J, Oikonomopoulos S, Ragoussis J, Pugh CW, Ratcliffe PJ, Mole DR. High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. Blood. 2017; 117:207–218.

Semenza GL. The hypoxic tumor microenvironment: A driving force for breast cancer progression. Biochim Biophys Acta Mol Cell Res. 2016; 1863:382–391.

Semenza GL. Hypoxia-inducible factor 1 oxygen homeostasis and disease pathophysiology. Trends Mol Med. 2001; 7:345–350. [PubMed: 11516994]

Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005; 21:3940–3941. [PubMed: 16096348]

Singh S, Narayanan SP, Biswas K, Gupta A, Ahuja N, Yadav S, Panday RK, Samaiya A, Sharan SK, Shukla S. Intragenic DNA methylation and BORIS-mediated cancer-specific splicing contribute to the Warburg effect. Proc Natl Acad Sci. 2017; 43:11440–11445.

Soneson C, Matthes KL, Nowicka M, Law CW, Robinson MD. Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage. Genome Biol. 2016; 17:12. [PubMed: 26813113]

Starmans MHW, Chu KC, Haider S, Nguyen F, Seigneuric R, Magagnin MG, Koritzinsky M, Kasprzyk A, Boutros PC, Wouters BG, Lambin P. The prognostic value of temporal in vitro and in vivo derived hypoxia gene-expression signatures in breast cancer. Radiother Oncol. 2012; 102:436–443. [PubMed: 22356756]

Sutherland RM, Ausserer WA, Murphy BJ, Laderoute KR. Tumor hypoxia and heterogeneity: Challenges and opportunities for the future. Semin Radiat Oncol. 1996; 6:59–70. [PubMed: 10717162]

Thienpont B, Steinbacher J, Zhao H, D'Anna F, Kuchnio A, Ploumakis A, Ghesquiere B, Van Dyck L, Boeckx B, Schoonjans L, Hermans E, et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. Nature. 2016; 537:63–68. [PubMed: 27533040]

Vaupel P, Harrison L. Tumor hypoxia: causative factors, compensatory mechanisms, and cellular response. Oncologist. 2004; 9:4–9.

Wang GL, Jiang BH, Rue EA, Semenza GL. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O_2 tension. Proc Natl Acad Sci. 1995; 92:5510–5514. [PubMed: 7539918]

Watson CJ, Neary R, Collier P, Ledwidge M, McDonald K, Baugh J. Hypoxia alters the DNA methylation profile of cardiac fibroblasts via HIF-1α regulation of DNA methyltransferase. Heart. 2012; 98:A7.

Watson JA, Watson CJ, McCann A, Baugh J. Epigenetics, the epicenter of the hypoxic response. Epigenetics. 2010; 5:293–296. [PubMed: 20418669]

Wu M-Z, Chen S-F, Nieh S, Benner C, Ger L-P, Jan C-I, Ma L, Chen C-H, Hishida T, Chang H-T, Lin Y-S, et al. Hypoxia drives breast tumor malignancy through a TET–TNFα–p38–MAPK signaling axis. Cancer Res. 2015; 75:3912–3924. [PubMed: 26294212]

Xu H, Handoko L, Wei X, Ye C, Sheng J, Wei C-L, Lin F, Sung W-K. A signal–noise model for significance analysis of ChIP-seq with negative control. Bioinformatics. 2010; 26:1199–1204. [PubMed: 20371496]

Yang L, Roberts D, Takhar M, Erho N, Bibby BAS, Thiruthaneeswaran N, Bhandari V, Cheng W-C, Haider S, McCorry AMB, McArt D, et al. Development and validation of a 28-gene hypoxia-related prognostic signature for localized prostate cancer. EBioMedicine. 2018; 31:182–189. [PubMed: 29729848]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]
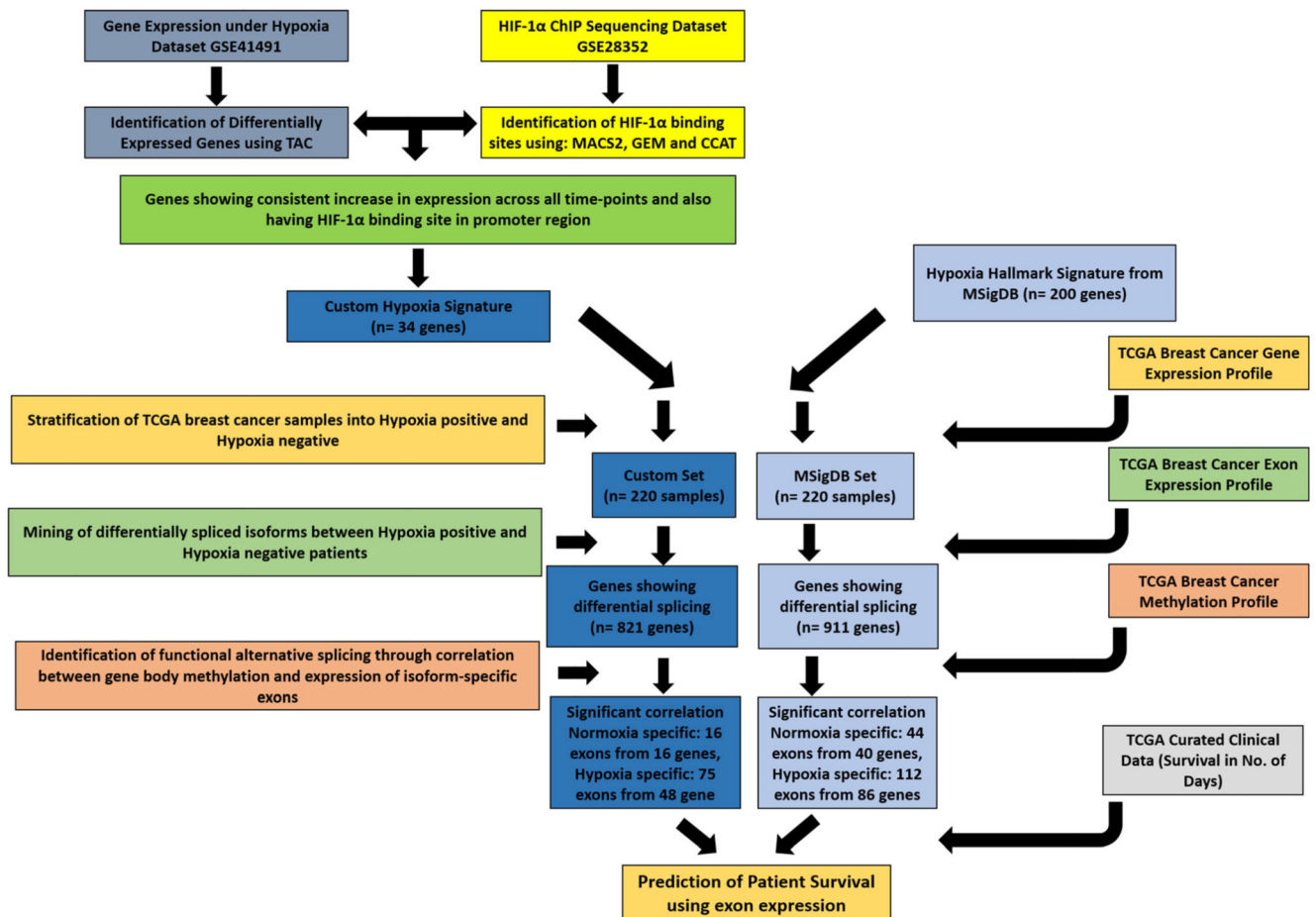
Figure 1.
Overview of the computational pipeline. The genes showing consistent increase in expression in hypoxia and having HIF-1α binding site in the promoter region were shortlisted to be a part of a custom hypoxia signature. The gene expression profile of genes included in this custom signature was used to stratify TCGA breast cancer samples into hypoxia positive or hypoxia negative. Hypoxia Hallmark signature from MSigDB was also used independently for stratification. Isoform-level expression data was used to mine differentially used isoforms between hypoxic and normoxic samples. This was followed by a correlation test between intragenic methylation levels and expression of isoform-specific exons belonging to transcripts differentially spliced in normoxic vs. hypoxic samples. Finally, survival analysis was performed with the exons showing significant correlation between expression and DNA methylation levels. (TAC: Transcriptome analysis console, MACS2: Model-based Analysis of ChIP-Seq, GEM: Genome-wide event finding and motif discovery, CCAT: Control- based ChIP-Seq analysis tools).
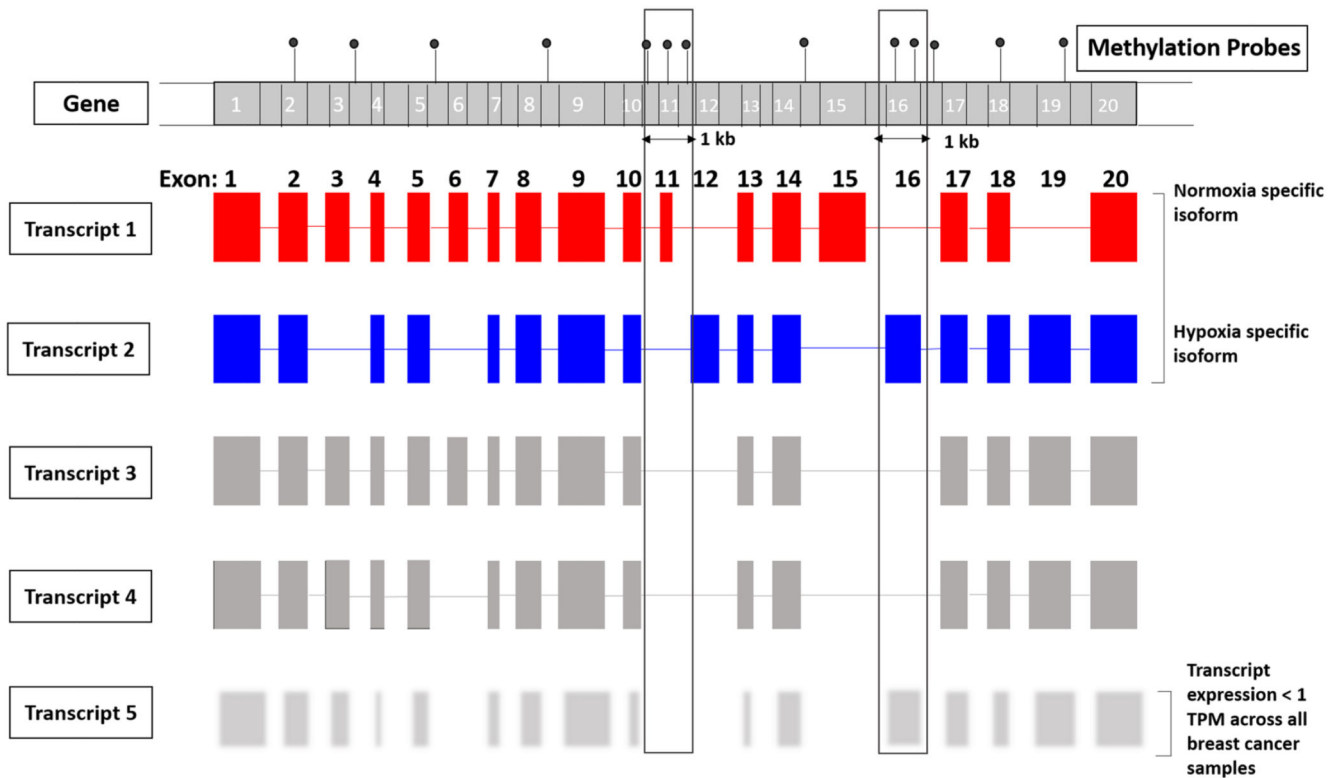
Figure 2.
Hypothetical gene and transcript structure showing criteria followed for calculating correlation between exon expression and DNA methylation. Exons exclusively present in one of the transcripts that were found to be differentially expressed in hypoxic and normoxic samples were considered. The isoforms that had an expression level of less than 1 TPM across all breast cancer samples were excluded from this analysis. Further, only those exons having methylation probes within 1 kb were selected for the correlation test.
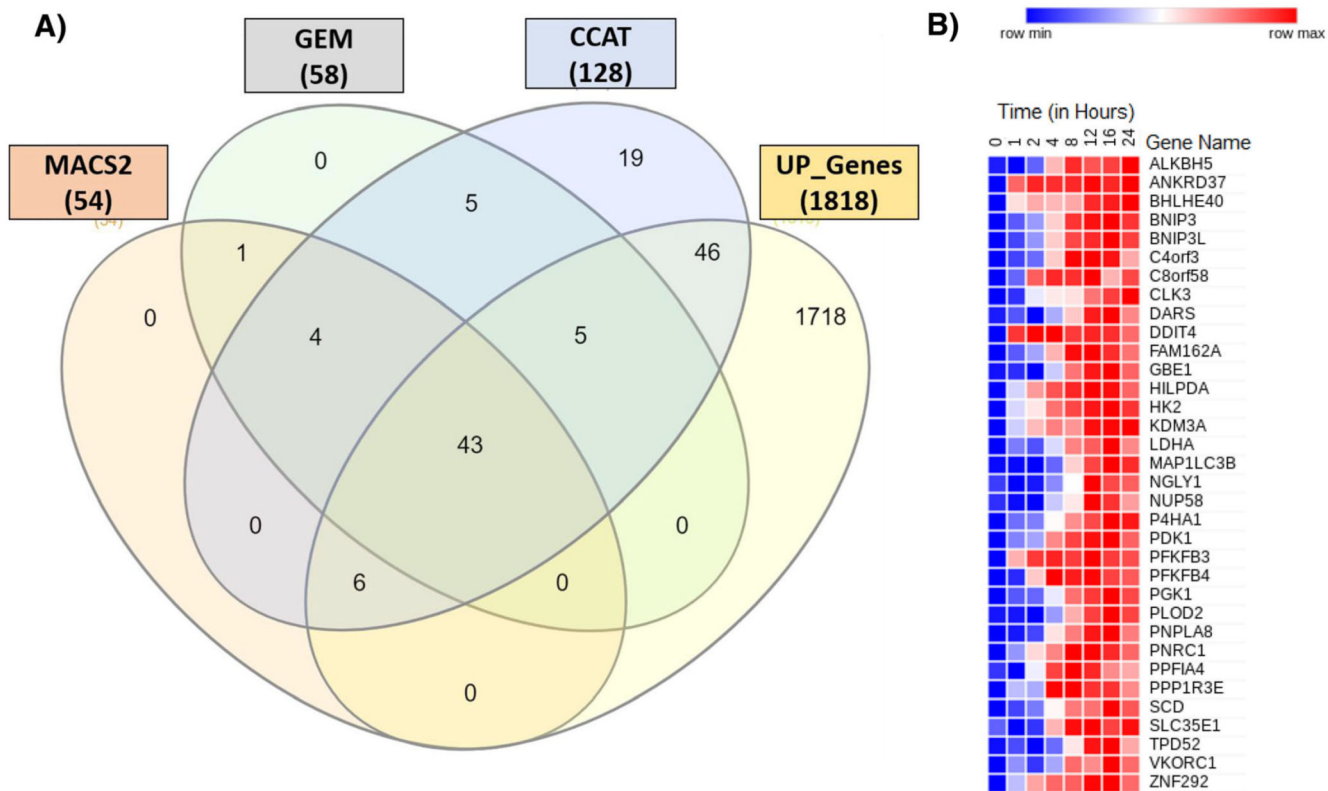
Figure 3.
Generation of Custom Hypoxia Signature. (A) Venn diagram showing common genes which have HIF-1α binding site and, also, show up-regulation under hypoxia. The three panels for MACS2, GEM and CCAT represent those genes that show HIF-1α ChIP-seq peaks in the promoter region for these three respective peak callers. Panel named UP_Genes include those genes which are up-regulated in hypoxia. (B) Heatmap showing mean of expression level (normalized probe intensities) across all probes corresponding to genes included in custom hypoxia signature at various time-points under hypoxia.
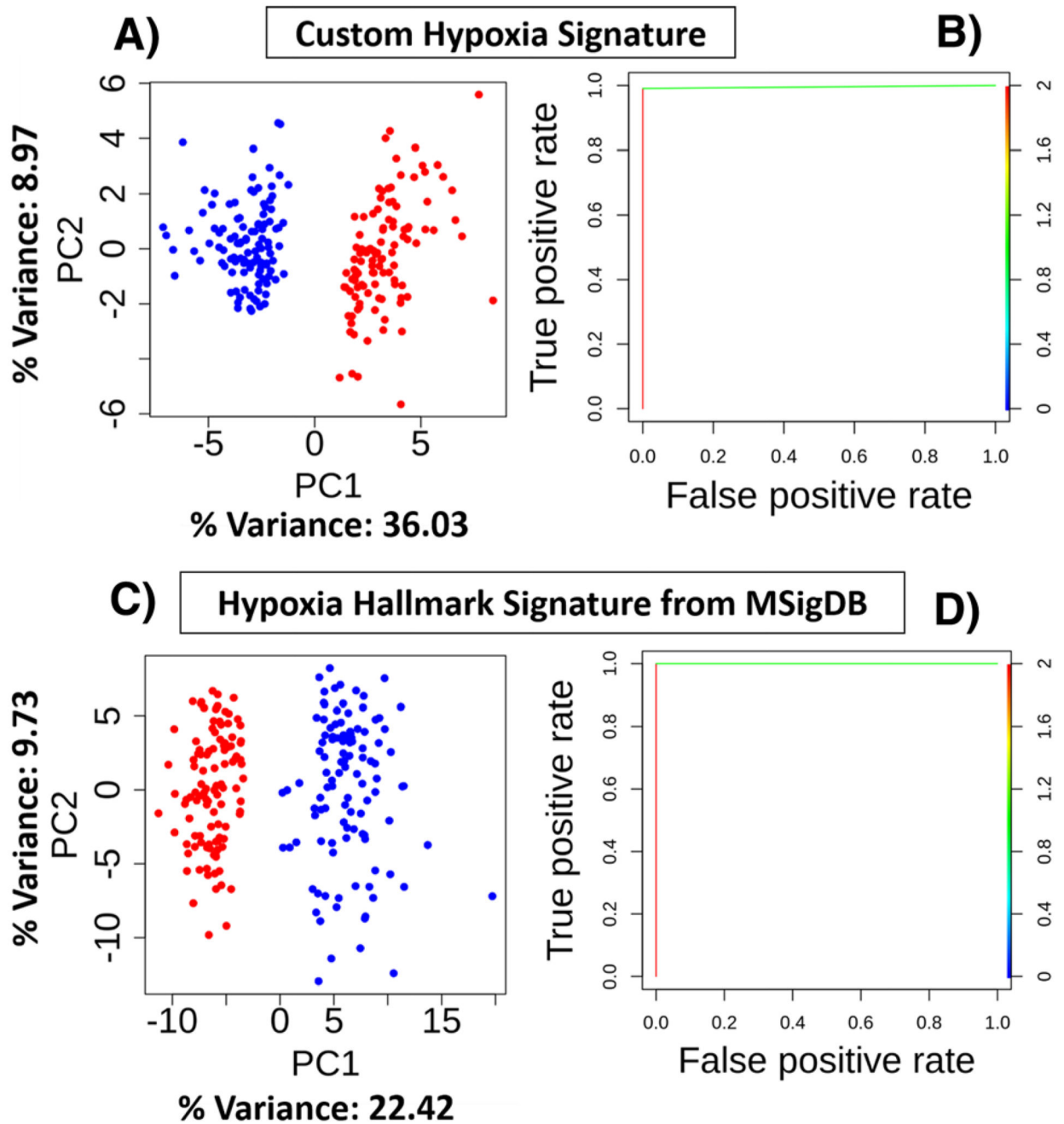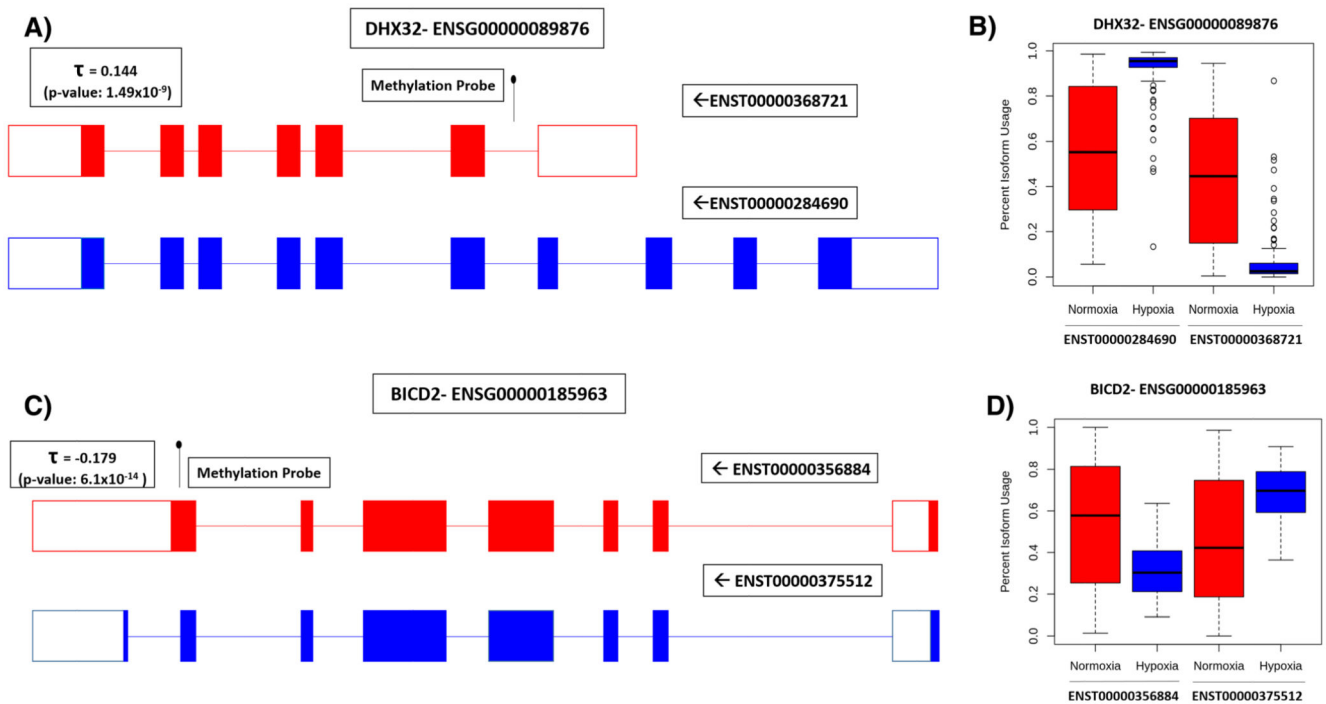
Figure 4.
Stratification of breast cancer samples into hypoxia pathway positive or hypoxia pathway negative. (A, C) Scatter plot of principal component 1 (PC1) vs. principal component 2 (PC2) showing the presence of two clusters each of which correspond to either hypoxic (blue) or normoxic (red) samples. Stratification was done using either custom hypoxia signature (A) or Hypoxia Hallmark signature from MSigDB (C). (B, D) ROC (Receiver operating characteristic) curve showing accuracy of stratification done using custom hypoxia signature (B) and Hypoxia Hallmark signature from MSigDB (D).

Figure 5.
Results of correlation test. (A and C) Diagrammatic representation of isoforms of DHX32 (A) and BICD2 (C) showing differential expression between samples stratified as hypoxic and normoxic using custom and MSigDB signatures for DHX32 and BICD2, respectively. Isoform with exons colored in red is normoxia-specific, while the one with blue-colored exons is hypoxia-specific. Positive correlation was seen between methylation level and expression of 7th exon of normoxia-specific DHX32, whereas, negative correlation was seen with expression of 1st exon of normoxia-specific isoform of BICD2. (B and D) Boxplot showing differential isoform usage for DHX32 (B) and BICD2 (D) between samples stratified as normoxic and hypoxic (τ: Kendall's rank correlation coefficient).
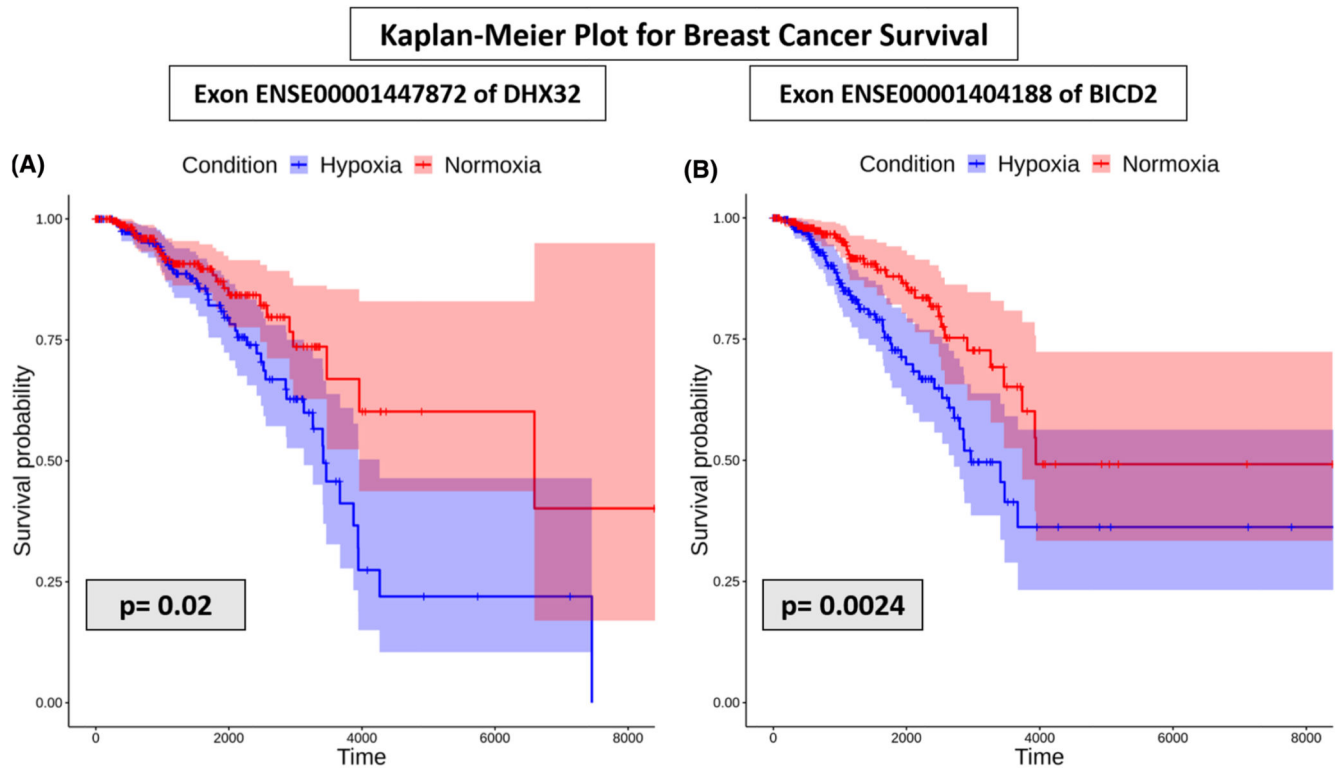
Figure 6.
Kaplan–Meier curve showing association of expression level of exons (that showed significant correlation with DNA methylation) belonging to DHX32 (A) and BICD2 (B), with patient survival (*p*-value: 0.02 for DHX32 and 0.0024 for BICD2). All TCGA breast cancer samples were sorted based on the expression level of normoxia-specific exons of DHX32 and BICD2. The top 25% samples were considered normoxic and the bottom 25% were considered hypoxic, followed by evaluation of survival predictability between these groups.

Table 1

Comparative analysis of differential expression between hypoxic and normoxic time-points

Normoxic timepoint (in hours)	Hypoxic timepoint (in hours)
0	1 and 2
0	2 and 4
0	4 and 8
0	8 and 12
0	12 and 16
0	16 and 24

Table 2

Summary of the TCGA datasets used in the study

Type of Data	Dataset ID	Version	Unit	Total no. of samples	No. of samples used
Gene Expression RNAseq	tcga_RSEM_gene_fpkm	2016-09-01	log2(fpkm +0.001)	10,535	1,099
Transcript Expression RNAseq	tcga_Kallisto_est_counts	2016-02-29	log2(est_counts +1)	10,535	1,099
Exon Expression RNAseq	TCGA.PANCAN.sampleMap/HiSeqV2_exon	2016-08-16	log2(rpkm+1)	10,459	1,099
DNA Methylation	jhu-usc.edu_PANCAN_Human Methylation450.betaValue_whitelisted.tsv.synapse_download_5096262.xena	2016-12-29	Beta value	9639	787 (Only Illumina 450 K Methylation array data used)
Curated Clinical Data	Survival_Supplemental Table_S1_20171025_xena_sp	2018-09-13	Days	12,591	1,099

Table 3

Definition of confusion matrix in the R package caret

	True hypoxic	**True normoxic**
Predicted hypoxic	True positive (TP)	True negative (TN)
Predicted normoxic	False negative (FN)	False positive (FP)

Table 4

Stratification evaluation parameters used for choosing custom and MSigDB clusters

Signature used	Threshold used	Accuracy	Sensitivity	Specificity
Custom	25%	0.9945	0.9964	0.9927
	20%	0.9977	1.0	0.9955
	10%	1.0	1.0	1.0
MSigDB	25%	0.9891	1.0	0.9782
	20%	0.9977	1.0	0.9955
	10%	1.0	1.0	1.0

Table 5

Overview of correlation tests performed

	No. of genes having at least 1 probe	No. of genes having at least 1 transcript specific exon	No. of genes having at least 1 exon-probe pair	No. of genes showing significant correlation after FDR correction
Custom: Test for Normoxia-specific isoforms	808 (out of 821)	152	60	16 (16 exons)
Custom: Test for Hypoxia-specific isoforms	808 (out of 821)	221	158	48 (75 exons)
MSigDB: Test for Normoxia-specific isoforms	894 (out of 911)	189	116	40 (44 exons)
MSigDB: Tests for Hypoxia-specific isoforms	894 (out of 911)	245	202	86 (112 exons)

The actual number of genes that were found to contain methylation probes, isoform-specific exons, exon-probe pairs and genes showing significant correlation after FDR correction are provided.