ARTICLE OPEN

Expression-based analyses indicate a central role for hypoxia in driving tumor plasticity through microenvironment remodeling and chromosomal instability

Anqi Jing[1], Frederick S. Vizeacoumar[2], Sreejit Parameswaran[2], Bjorn Haave[2], Chelsea E. Cunningham[3], Yuliang Wu[3], Roland Arnold[4], Keith Bonham[3,5], Andrew Freywald[2,3], Jie Han[1] and Franco J. Vizeacoumar[2,3,5]

Can transcriptomic alterations drive the evolution of tumors? We asked if changes in gene expression found in all patients arise earlier in tumor development and can be relevant to tumor progression. Our analyses of non-mutated genes from the non-amplified regions of the genome of 158 triple-negative breast cancer (TNBC) cases identified 219 exclusively expression-altered (EEA) genes that may play important role in TNBC. Phylogenetic analyses of these genes predict a "punctuated burst" of multiple gene upregulation events occurring at early stages of tumor development, followed by minimal subsequent changes later in tumor progression. Remarkably, this punctuated burst of expressional changes is instigated by hypoxia-related molecular events, predominantly in two groups of genes that control chromosomal instability (CIN) and those that remodel tumor microenvironment (TME). We conclude that alterations in the transcriptome are not stochastic and that early-stage hypoxia induces CIN and TME remodeling to permit further tumor evolution.

npj Systems Biology and Applications (2018)4:38; doi:10.1038/s41540-018-0074-z

INTRODUCTION

The Darwinian model of clonal selection, where a subset of genetic lesions drives tumor evolution and progression in a stepwise manner,[1–4] is widely accepted as the mode of evolution of malignant cells under therapy or basal conditions.[5–7] However, recent findings in prostate, pancreatic or triple-negative breast cancer (TNBC), challenge this paradigm and question if gradualism is indeed the single mode of evolution.[8–10] It may be instead a punctuated burst of molecular alterations in the early stages of cancer, where changes in biological environment of growing tumors require massive adaptations in the molecular machinery of cancer cells.[11–13] Usually, normal cells respond to stress by deploying repair or resistance tools to maintain their genetic integrity and assure survival.[14,15] In contrast, cancer cells typically do not have intact repair tools, which lead to genetic instability. Chromosomal instability (CIN) is a form of genetic instability that causes changes in both the structure and number of chromosomes.[15–25] For example, mutations in CIN genes like BRCA1/2 increase the number of deletions up to 50 bps, causing multiple defects within the genome.[26] Progressive accumulation of CIN within a tumor allows development of cell populations with heterogeneous properties. Some of these cells will carry selective survival advantages and will be responsible for further tumor progression.[3] Likewise, overexpression of APOBEC3, a member of the cytidine deaminase gene family, may generate frequent C > T base substitutions also leading to tumor heterogeneity and progression along the malignancy pathway.[27] Understanding the sequence of molecular events essential for tumor evolution may not only benefit early detection of malignancies but may also allow the development of more effective treatment and even prevention strategies. While the role of accumulating genetic mutations in tumor progression has been extensively discussed, it is still not clear how alterations in gene expression patterns contribute to tumor evolution.

Changes to gene expression can be brought about by number of factors, including epigenetic modifications, translation regulation, and differences in mRNA, and protein stability.[28] For example, increased activities of growth factor, chemokine and cytokine receptors can set off specific signaling cascades and subsequent changes in gene expression, without any direct involvement of genetic mutations. However, what are the most significant changes that occur within the transcriptome of cancer cells and how they may contribute to tumor evolution is not clear. Here, we use an aggressive malignancy, TNBC, as a model to explore the role of transcriptomic alterations during early stages that are caused not by genomic mutations, but exclusively by differential gene expression. We achieve this by focusing specifically on genes that are heavily upregulated in the non-amplified regions of the genome. We focused specifically on upregulated genes because direct inhibition of these molecules may provide viable cancer treatment/prevention options at early stages of tumor development. Remarkably, our analysis of RNA seq data in 158 TNBC cases revealed that there is indeed a punctuated burst of expressional changes in two major groups of genes controlled by hypoxia-related factors. These two groups included molecules that regulate CIN and remodel tumor microenvironment (TME). This

[1]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G2R3, Canada; [2]Department of Pathology and Laboratory Medicine, College of Medicine, University of Saskatchewan, Saskatoon, SK S7N 5E5, Canada; [3]Department of Biochemistry, Cancer Cluster, College of Medicine, University of Saskatchewan, Saskatoon, SK S7N 5E5, Canada; [4]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, United Kingdom and [5]Cancer Research, Saskatchewan Cancer Agency, Saskatoon, SK S7N 5E5, Canada
Correspondence: Jie Han (jhan8@ualberta.ca) or Franco J. Vizeacoumar (franco.vizeacoumar@usask.ca)

Received: 16 January 2018 Revised: 25 September 2018 Accepted: 27 September 2018
Published online: 24 October 2018

not only reveals new potential targets for TNBC therapy, but also indicates a critical role for hypoxia in very early stages of tumor development.

RESULTS

A multi-step process to identify differentially expressed genes in breast cancer

To identify genes with aberrant expression patterns, we initially curated all the genes that are differentially regulated. We used the breast cancer-specific data from The Cancer Genome Atlas (TCGA) that represents the largest collection of patient samples with information on the mutation status, copy number aberrations (CNA), as well as gene expression patterns at different stages of tumor development. Gene expression in breast tumor samples was compared to the expression of the matching genes in normal samples using fold-change (FC) and false discovery rate (FDR) after Empirical Bayes moderated *t*-test with Benjamini-Hochberg correction. Genes were considered as upregulated genes if FDR ≤ 0.01 and FC ≤ 2. Downregulated genes were selected if FDR ≤ 0.01 and FC ≤ −2. Our initial analyses in overall breast cancer identified 586 genes that were upregulated and 1446 genes that were downregulated at multiple stages of cancer progression (Supplementary Fig. 1a). The overlap between all stages is presented in Supplementary Table 1a. We also ran a complementary analysis to identify differentially regulated genes in specifically in TNBC. We found 1127 genes to be upregulated and 1752 genes downregulated across multiple stages of TNBC (Fig. 1a). The overlap between all stages is shown in Supplementary Table 1b. The Gene Set Enrichment Analysis (GSEA) indicated that the upregulated genes in TNBC are enriched for molecules involved in cell-cycle regulation and chromatin organization ($p < 0.001$) (Fig. 1b). Results of our GSEA analysis of genes differentially upregulated in TNBC tumors correlated well with the previously reported, differentially regulated genes from an independent cohort ($p < 0.001$) (Fig. 1b),[29] which provides an additional support for the relevance of our observations. While we found a higher abundance of downregulated genes, compared to the upregulated genes, no similar significant enrichment was observed within the pool of the downregulated genes. Similar results were obtained for overall breast cancer (Supplementary Fig. 1b). Taken together, these observations indicate that the application of our approach to the analysis of TCGA data allows identifying subsets of genes differentially regulated in TNBC tumors.

Not all differentially expressed genes are equally deregulated across the population of breast cancer patients

While gene expression analysis to identify differentially regulated genes has been a common approach in cancer biology, we attempted to determine, how many of these genes are aberrantly expressed with high frequencies across the population of TNBC patients. We rationalized that common aberrations found in all patients should have arisen earlier in the development of the malignancy, compared to alterations that were found only in a subset of patients. Therefore, we have calculated a frequency of differential expression of each affected gene in TNBC tumors (Fig. 1c and Supplementary Fig. 1c). Throughout this analysis, we maintained a twofold change in the expression level as a minimum requirement for a gene to be considered differentially regulated. The frequency of changes in each differentially expressed gene is calculated as a percentage of patients in whom the gene is up- or downregulated. We found 254 genes were upregulated and 1197 genes were downregulated in almost 70% of the TNBC patients (Supplementary Table 2a, b). Similar results for overall breast cancer are presented Supplementary Table 3a, b. Unfortunately, there were only two patient samples that were available in TNBC-stage IV in TCGA dataset, which was not sufficient to minimize random effects. Therefore, we computed our analyses using the larger number of samples involved in the first three stages of TNBC.

Changes in gene expression may not only arise from aberrant expression from an endogenous promoter, but also from accompanying chromosomal amplifications, deletions and other types of mutations. To account for this, we isolated the differentially regulated genes exclusively from the non-amplified/deleted regions of the genome. We identified 77 amplified chromosome regions from the TCGA dataset based on CNA, including several previously reported regions in 1q, 8q, 16p, and 20q (Supplementary Table 4),[30] as presented in the circos plot for TNBC (Fig. 2a) or overall breast cancer (Supplementary Fig. 2a). We further evaluated the concordance of amplification and gene expression by fold-change with FDR and Pearson's correlation. We considered genes likely to be driven by CNA if their Pearson's correlation coefficient between expression and CNA was greater than 0.3, or they show significant differential CNA-associated expression change (Supplementary Fig. 2b, c and Supplementary Table 5a, b). Subsequently, we filtered out from our analysis 20 genes from TNBC patients that were in amplified regions or had strong correlations with chromosome amplification.

We also used somatic mutational analyses of 560 breast cancer whole-genome sequencing database available at COSMIC to eliminate any gene that might be differentially expressed because of a mutation.[31] By also excluding 13 genes whose loci information was ambiguous, we finally identified 219 exclusively expression-altered (EEA) genes that elevated their expression in TNBC (Fig. 2b) and, therefore, may represent good therapeutic targets. Interestingly, we observed multiple distinct patterns of upregulation with varying frequencies across different cancer stages (Fig. 2b). For example, some genes were constitutively upregulated across all stages (PLK1, UBE2C, or KIF4A). Similarly, certain genes were upregulated mostly at later stages (CCNE1, HMGB3, or NUF2). In contrast to this category, some genes were upregulated selectively at early stages but were gradually downregulated through the later stages (MMP1, MMP11, or MMP13). Among the 219 upregulation events, majority of changes occurred in chromosome 1 and 17 (Supplementary Fig. 2d). Surprisingly, although the expression of some initially upregulated genes gradually decreased, we did not observe any instance where their expression returned to normal levels (Fig. 2c). Importantly, while we find that not all upregulated genes are overexpressed in all breast tumors across all cancer stages, our analysis has generated an explicit set of genes that are overexpressed in over 70% of patients at all stages of both all breast cancer and TNBC tumors (Supplementary Table 6a, b).

Lineage analysis of upregulation profiles reveals a punctuated pattern of evolution in early TNBC tumors

Previous studies have used somatic mutations and CNA to understand tumor evolution.[8,10,11,13] However, it is not clear, how alterations in gene expression may affect tumor progression. To further address this, we performed a progression-based analysis on the expression profiles of the newly identified EEA genes to describe how they may influence TNBC progression. First, based on the profile of upregulation status of EEA genes, we partitioned TNBC samples into groups so that the samples within a group have more similar upregulation profiles than other samples in different groups. Hierarchical clustering was applied to group tumor samples into clusters[32]. The cluster structure was graphically represented in Fig. 3a, which revealed that the most distinguishable cluster C1 is diverged at the highest overhang with the highest dissimilarities from the remaining samples. In addition, several distinguishable branches C2, C3, and C4 were also clustered. After identifying distinct tumor clusters, we constructed their distance tree to gain insights into the

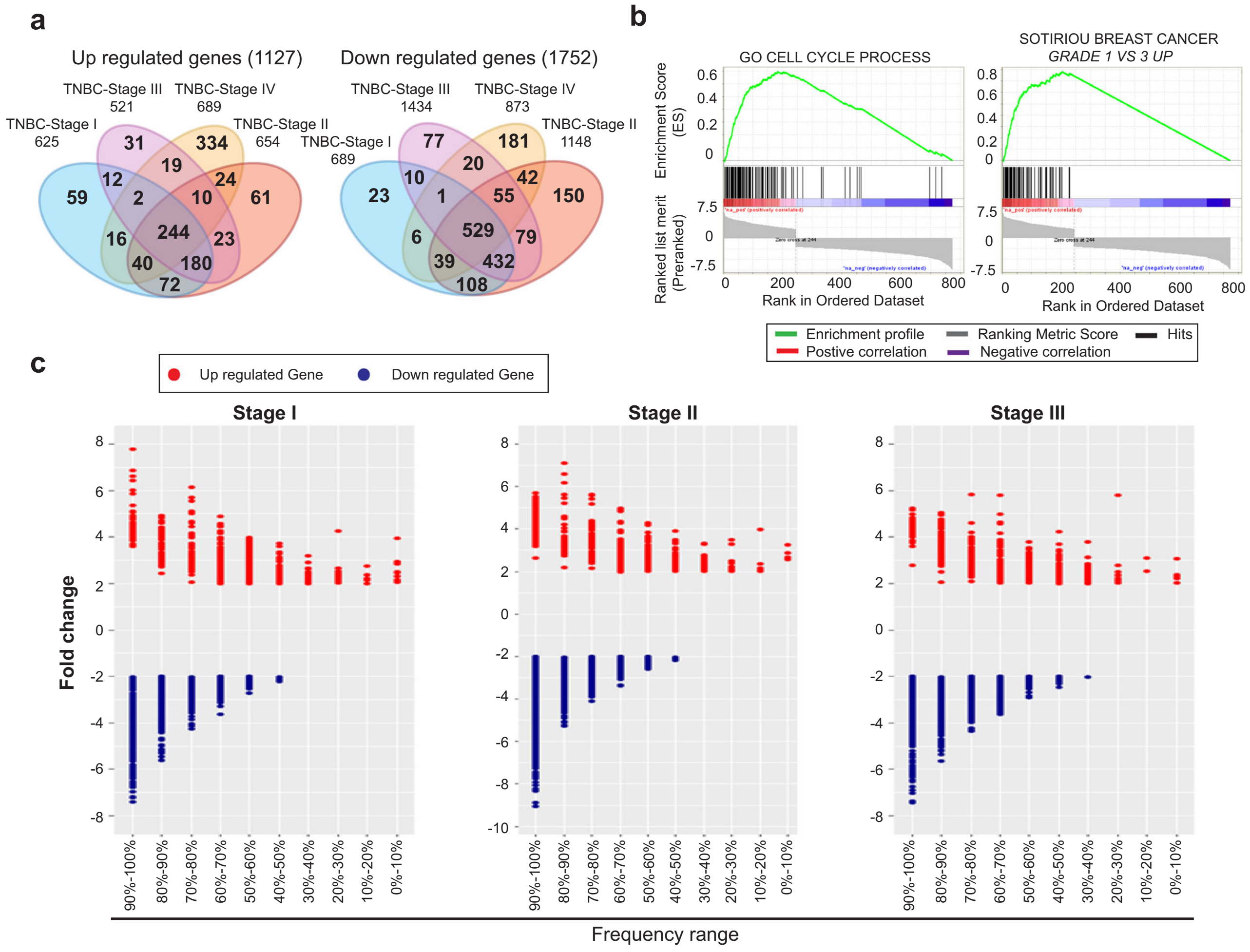

Fig. 1 Identification of upregulated genes in TNBC. **a** Venn-diagram of differentially expressed genes in TNBC-stage-specific tumors. The number of up- and downregulated genes at each stage of tumor and at the intersection between different stages have been represented. **b** Gene set enrichment analysis for up/downregulated genes across all TNBC tumor stages. Gene Set Enrichment Analysis for 244 upregulated genes (left) and 529 downregulated genes (right) across four tumor stages along with previously identified, differentially upregulated genes from Sotiriou et al.[29] **c** Frequency distribution of differential expression in TNBC-stage-specific tumors. Dot plot represents the fold-change and the frequency range of TNBC-stage-specific differentially expressed genes, where the red denotes upregulated gene and the blue denotes downregulated gene

progression path in the context of accumulation of aberrant expression, analogous to a phylogenetic analysis. The construction of the distance tree was based on the neighbor-joining algorithm[33] to display the lineage between the four clusters. Assuming that the tumor is derived from a single or a group of homogenous normal cells and the complexity of upregulation in a tumor increases with time, the history of its progression can be partially inferred by comparing homogeneous groups. We used the vector with 219 elements of all zeros representing no upregulations of EEA genes in normal status as the root vertex, and constructed a tree-like structure by neighbor-joining to evaluate the extent of the accumulation of upregulatory events for each cluster.

The phylogenetic lineage showed that cluster C1 has the shortest distance from normal samples, which suggests that tumors in C1 may be recorded at the early tumor stage and the upregulated genes in C1 may be of importance to tumor initiation and early progression (Fig. 3b). The lineage shows a large distance from normal cell to C1, indicating a large number of upregulatory events are required for successful tumor progression through very early stages. C2, C3, and C4 clusters diverged for relatively small distances from the common ancestor n2, which suggests less dissimilarity from C2 to C3 and C4, indicating that minimal changes in gene expression were required at later stages. By measuring the Cosine similarity between mean upregulation profile and subset vector (See Methods for details), we found that a burst of 83 upregulation events occurred earliest in C1 suggesting that the 83 EEA genes that may act as potential enabling factors within the early tumor evolution (Supplementary Table 7).

To confirm that the 83 upregulation events are relevant to breast cancer progression, we next inquired if these changes in gene expression correlate with the loss of expression of known tumor suppressors. Vogelstein and colleagues identified ~70 tumor suppressor genes that when inactivated by intragenic mutations can promote tumorigenesis.[34] We found a strong negative correlation in the expression of the 83 EEA genes and the 74 tumor suppressors (Fig. 3c). In summary, our analysis revealed that a large number of EEA events appear at the earliest stage of tumor development with fewer subsequent events at later stages, reflecting an emerging pressure from rapidly changing biological environment within early progressing tumors.

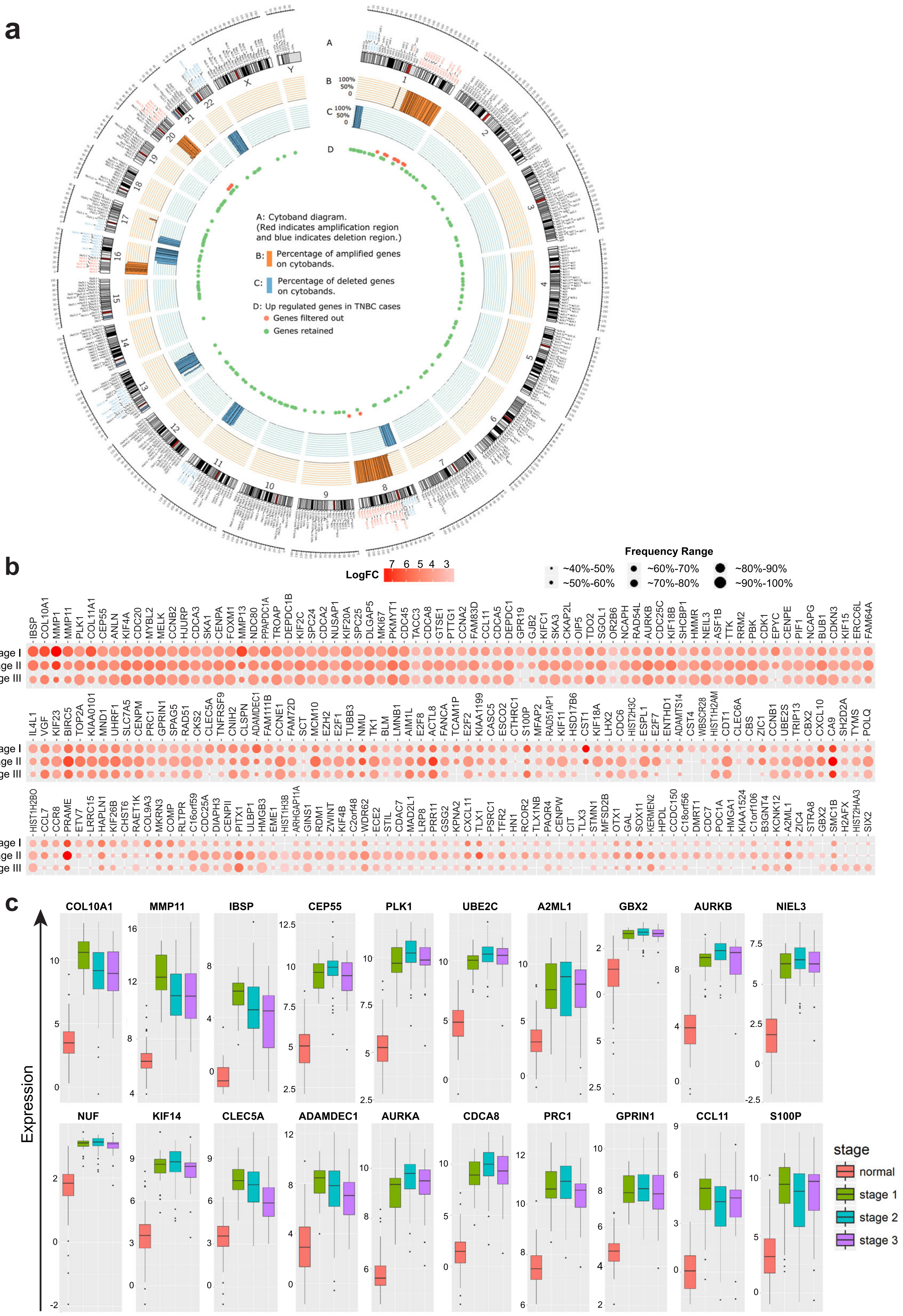

Fig. 2 Elimination of amplified genes to identify 219 upregulated events. **a** Amplified chromosome cytobands and upregulated genes locus. Track A displays the cytoband diagram where the texts in red indicate identified amplified regions. Track B and C displays the frequency of genes showing amplification and deletion, respectively, in at least in 40% of patients in each cytoband. Genes in Fig. 1d were mapped to the Track D. **b** Fold-change and frequency distribution for genes showing upregulation in at least 70% of TNBC patients. Nodes in each column represent upregulated genes with their sizes indicating the frequency of samples and their colors representing the fold-change value in the specific tumor stage. **c** Box plots of Cluster 1 gene expression at various stages of TNBC tumor. The *y*-axis represents log2-transformed gene expression and *x*-axis denotes TNBC stages

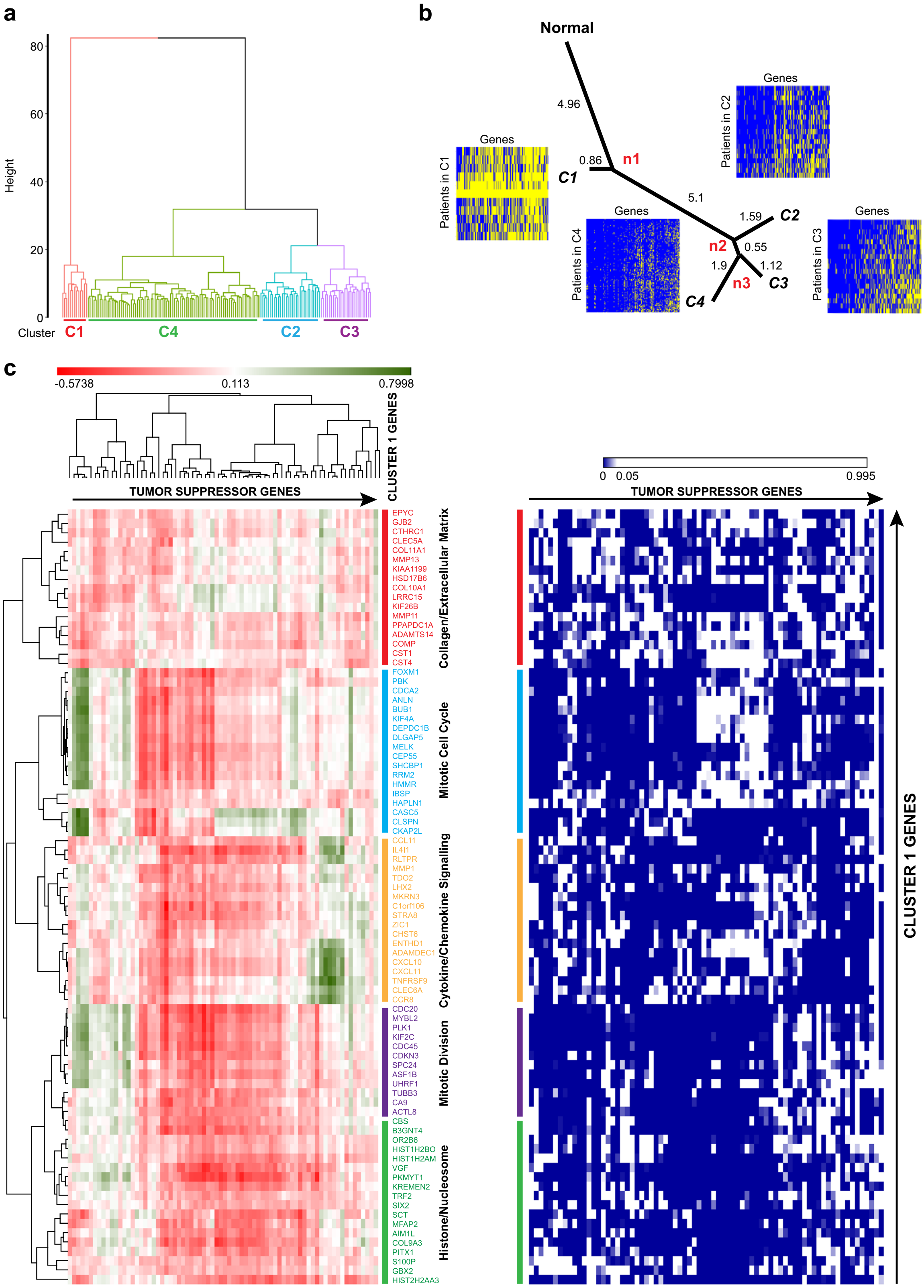

Fig. 3 Identification of 83 upregulation events that occur in early stages. **a** Hierarchical clustering of 219 EEA genes in TNBC patients. Different colors show TNBC patients clustered into four clusters represented as Red for Cluster 1 (C1), Purple for Cluster 2 (C2), Blue for Cluster 3 (C3), and Green for Cluster 4 (C4). **b** The progression of gene upregulations in different TNBC clusters as shown by the phylogenetic tree. The figure shows the lineage of progression of gene upregulations from the normal to distinct subpopulations. Heat maps with genes in columns and TNBC samples in rows display the upregulation status (yellow: no upregulation; blue: upregulation) for different TNBC clusters. **c** Correlation clustergram of cluster 1 genes compared to known tumor suppressors. Red indicates negative correlation and green indicates positive correlation. The panel on the right represents, the significance of the correlation as a heat map. Blue indicates significance (<0.05) and white indicates lack of significance (>0.05)

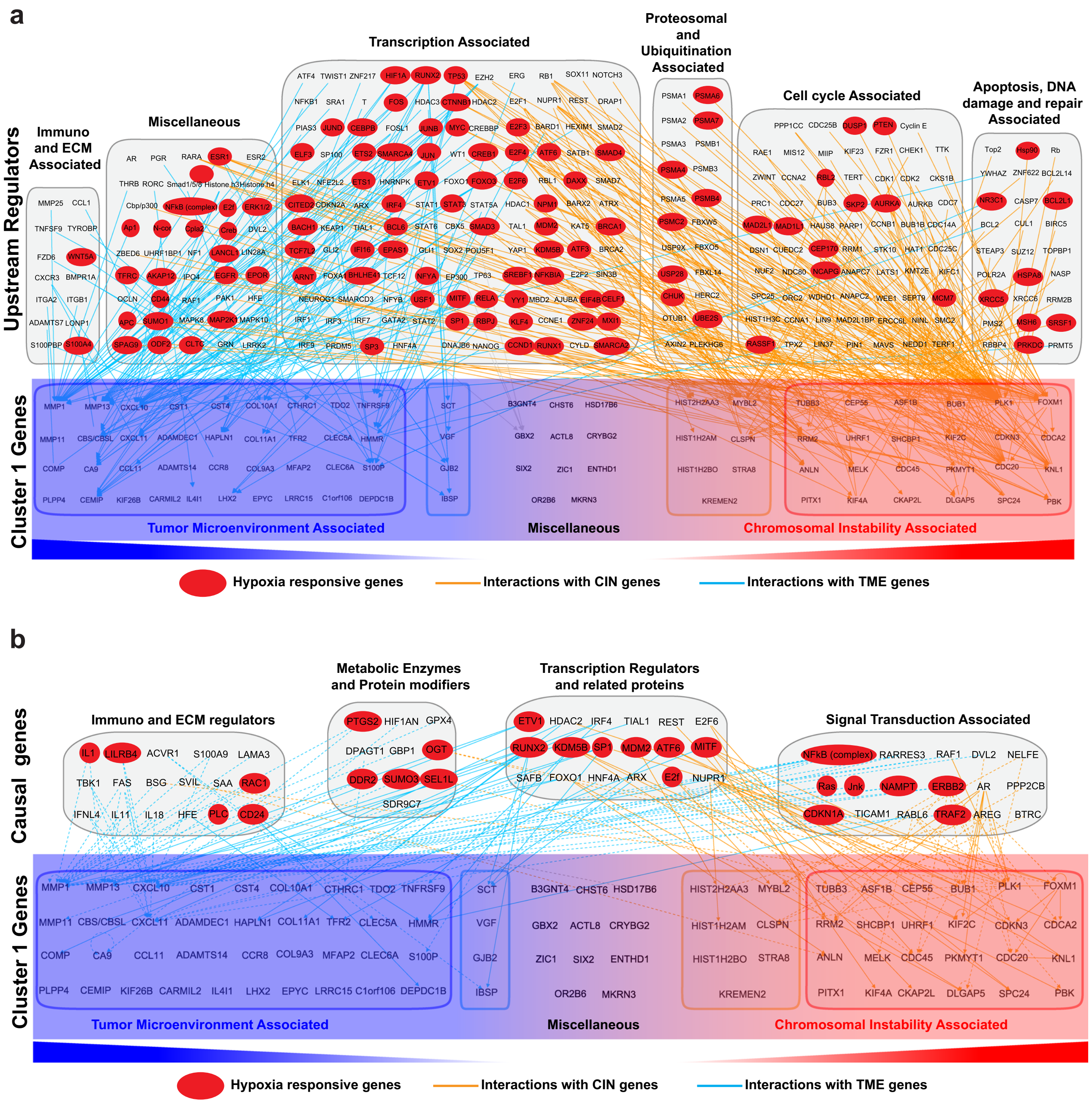
a
Upstream Regulators
Immuno and ECM Associated
MMP25 CCL1 TNFSF9 TYROBP FZD6 WNT5A CXCR3 BMPR1A ITGA2 ITGB1 ADAMTS7 LONP1 S100PBP S100A4
Miscellaneous
AR PGR RARA ESR1 ESR2 THRB RORC Smad1/5/8 Histone h3 Histone h4 Cbp/p300 NFkB (complex) E2f ERK1/2 Ap1 N-cor Cpla2 Creb DVL2 ZBED6 UHRF1BP1 NF1 LANCL1 LIN28A TFRC AKAP12 IPO4 EGFR EPOR OCLN CD44 RAF1 PAK1 HFE APC SUMO1 MAPK8 MAP2K1 MAPK10 SPAG9 OGF2 CLTC GRN LRRK2
Transcription Associated
ATF4 TWIST1 ZNF217 HIF1A RUNX2 TP53 EZH2 ERG RB1 SOX11 NOTCH3 NFKB1 SRA1 T FOS HDAC3 CTNNB1 HDAC2 E2F1 NUPR1 REST DRAP1 PIAS3 JUND CEBPB FOSL1 JUNB MYC CREBBP E2F3 BARD1 HEXIM1 SMAD2 ELF3 SP100 ETS2 SMARCA4 JUN WT1 CREB1 E2F4 ATF6 SATB1 SMAD4 ELK1 NFE2L2 ETS1 HNRNPK ETV1 FOXO1 FOXO3 E2F6 RBL1 DAXX SMAD7 CITED2 CDKN2A ARX IRF4 STAT1 STAT3 STAT5A HDAC1 NPM1 BARX2 ATRX BACH1 KEAP1 TIAL1 BCL6 STAT6 CBX5 SMAD3 TAL1 MDM2 KAT5 BRCA1 TCF7L2 GLI2 IFI16 EPAS1 GLI1 SOX2 POU5F1 YAP1 KDM5B ATF3 BRCA2 ARNT FOXA1 BHLHE41 TCF12 NFYA EP300 TP63 SREBF1 NFKBIA E2F2 SIN3B NEUROG1 SMARCD3 NFYB USF1 MITF RELA YY1 MBD2 AJUBA EIF4B CELF1 IRF1 IRF3 IRF7 GATA2 STAT2 SP1 RBPJ KLF4 CCNE1 ZNF24 MXI1 IRF8 PRDM5 SP3 HNF4A DNAJB6 NANOG CCND1 RUNX1 CYLD SMARCA2
Proteosomal and Ubiquitination Associated
PSMA1 PSMA6 PSMA2 PSMA7 PSMA3 PSMB1 PSMA4 PSMB3 PSMA5 PSMB4 PSMC2 FBXW5 USP9X FBXO5 USP28 FBXL14 CHUK HERC2 OTUB1 UBE2S AXIN2 PLEKHG6
Cell cycle Associated
PPP1CC CDC25B DUSP1 PTEN Cyclin E RAE1 MIS12 MIIP KIF23 FZR1 CHEK1 TTK ZWINT CCNA2 RBL2 TERT CDK1 CDK2 CKS1B PRC1 CDC27 BUB3 SKP2 AURKA AURKB CDC7 MAD2L1 MAD1L1 HAUS8 PARP1 CCNB1 BUB1B CDC14A DSN1 CUEDC2 CEP170 RRM1 STK10 HAT1 CDC25C NUF2 NDC80 NCAPG ANAPC7 LATS1 KMT2E KIFC1 SPC25 ORC2 WDHD1 ANAPC2 WEE1 SEPT9 MCM7 HIST1H3C CCNA1 LIN9 MAD2L1BP ERCC6L NINL SMC2 RASSF1 TPX2 LIN37 PIN1 MAVS NEDD1 TERF1
Apoptosis, DNA damage and repair Associated
Top2 Hsp90 Rb YWHAZ ZNF622 BCL2L14 NR3C1 CASP7 BCL2L1 BCL2 CUL1 BIRC5 STEAP3 SUZ12 TOPBP1 POLR2A HSPA8 NASP XRCC5 XRCC6 RRM2B PMS2 MSH6 SRSF1 RBBP4 PRKDC PRMT5
Cluster 1 Genes
MMP1 MMP13 CXCL10 CST1 CST4 COL10A1 CTHRC1 TDO2 TNFRSF9 MMP11 CBS/CBSL CXCL11 ADAMDEC1 HAPLN1 COL11A1 TFR2 CLEC5A HMMR COMP CA9 CCL11 ADAMTS14 CCR8 COL9A3 MFAP2 CLEC6A S100P PLPP4 CEMIP KIF26B CARMIL2 IL4I1 LHX2 EPYC LRRC15 C1orf106 DEPDC1B
SCT VGF GJB2 IBSP
B3GNT4 CHST6 HSD17B6 GBX2 ACTL8 CRYBG2 SIX2 ZIC1 ENTHD1 OR2B6 MKRN3
HIST2H2AA3 MYBL2 HIST1H2AM CLSPN HIST1H2BO STRA8 KREMEN2
TUBB3 CEP55 ASF1B BUB1 PLK1 FOXM1 RRM2 UHRF1 SHCBP1 KIF2C CDKN3 CDCA2 ANLN MELK CDC45 PKMYT1 CDC20 KNL1 PITX1 KIF4A CKAP2L DLGAP5 SPC24 PBK
Tumor Microenvironment Associated
Miscellaneous
Chromosomal Instability Associated
Hypoxia responsive genes
Interactions with CIN genes
Interactions with TME genes
b
Causal genes
Immuno and ECM regulators
IL1 LILRB4 ACVR1 S100A9 LAMA3 TBK1 FAS BSG SVIL SAA RAC1 IFNL4 IL11 IL18 HFE PLC CD24
Metabolic Enzymes and Protein modifiers
PTGS2 HIF1AN GPX4 DPAGT1 GBP1 OGT DDR2 SUMO3 SEL1L SDR9C7
Transcription Regulators and related proteins
ETV1 HDAC2 IRF4 TIAL1 REST E2F6 RUNX2 KDM5B SP1 MDM2 ATF6 MITF SAFB FOXO1 HNF4A ARX E2f NUPR1
Signal Transduction Associated
NFkB (complex) RARRES3 RAF1 DVL2 NELFE Ras Jnk NAMPT ERBB2 AR PPP2CB CDKN1A TICAM1 RABL6 TRAF2 AREG BTRC
Cluster 1 Genes
MMP1 MMP13 CXCL10 CST1 CST4 COL10A1 CTHRC1 TDO2 TNFRSF9 MMP11 CBS/CBSL CXCL11 ADAMDEC1 HAPLN1 COL11A1 TFR2 CLEC5A HMMR COMP CA9 CCL11 ADAMTS14 CCR8 COL9A3 MFAP2 CLEC6A S100P PLPP4 CEMIP KIF26B CARMIL2 IL4I1 LHX2 EPYC LRRC15 C1orf106 DEPDC1B
SCT VGF GJB2 IBSP
B3GNT4 CHST6 HSD17B6 GBX2 ACTL8 CRYBG2 ZIC1 SIX2 ENTHD1 OR2B6 MKRN3
HIST2H2AA3 MYBL2 HIST1H2AM CLSPN HIST1H2BO STRA8 KREMEN2
TUBB3 ASF1B CEP55 BUB1 PLK1 FOXM1 RRM2 SHCBP1 UHRF1 KIF2C CDKN3 CDCA2 ANLN MELK CDC45 PKMYT1 CDC20 KNL1 PITX1 KIF4A CKAP2L DLGAP5 SPC24 PBK
Tumor Microenvironment Associated
Miscellaneous
Chromosomal Instability Associated
Hypoxia responsive genes
Interactions with CIN genes
Interactions with TME genes

Fig. 4 IPA analyses showing extensive interaction between hypoxia responsive genes with members of cluster 1 genes. **a** Upstream regulator analysis was performed with IPA for the cluster 1 genes and all the interactions retrieved are presented. Cluster 1 genes are classified into those that are associated with CIN or TME. The upstream genes that are hypoxia responsive, are highlighted in red. **b** Causal network analysis was performed with IPA for the cluster 1 genes and all the interactions retrieved are presented. Cluster 1 genes are classified into those that are associated with CIN or TME. The upstream genes that are hypoxia responsive, are highlighted in red

TME remodeling and CIN cooperatively drive TNBC

Since our phylogenetic analysis indicated that the 83 upregulated EEA genes are crucial early events in early tumorigenesis, we next explored the functionalities of these genes. Interestingly, we found a large subset of genes that are known to be involved in remodeling TME, including metalloproteinases (MMP1, MMP11, MMP13, ADAMDEC1, ADAMTS14), chemokine receptors and ligands (CXCL11, CXCL10, CCL11, CCR8), protease inhibitors (CST4, CST1), pH maintenance factors (CAIX), and different collagens (COL9A3, COL10A1). This emphasizes the critical role of extracellular matrix and TME remodeling at the early-stage of tumor progression. Similarly, we also identified several of the EEA genes including, FOXM1, PLK1, BUB1, KIF2C, CDCA2, CDC20, CDKN3, KNL1 to name a few, that are known for their role in CIN and tumor development.[35–43] This may reflect a selective pressure for additional genetic alterations in early tumors that would allow their further evolution. In addition, cluster 1 included genes like DEPDC1B and HMMR that have known roles in both TME remodeling as well as CIN associated functions.[44–48] Overall, our identification of cluster 1 genes indicates that a punctuated burst of expressional changes occurs simultaneously in both CIN and TME remodeling genes very early in tumor development.

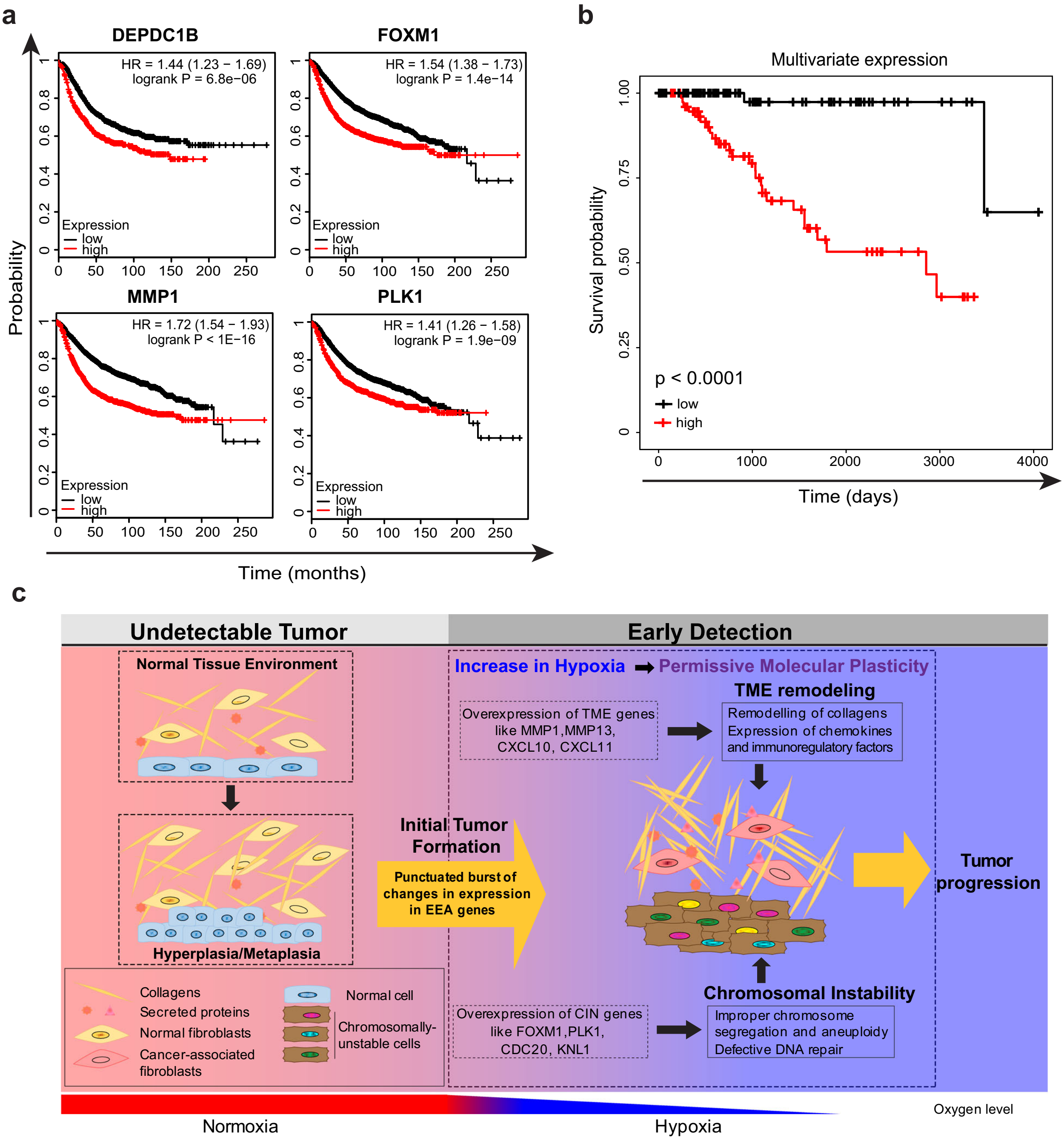

Fig. 5 Survival plot and a model describing the role of cluster 1 genes in tumor evolution. **a** Representative relapse-free survival plots of breast cancer patients with low and high expression of cluster 1 genes. **b** Survival plots of TNBC patients with high and low multivariate gene expression score. **c** Schematic model showing the effect of simultaneous burst of CIN and TME-associated genes in response to hypoxia during early stages of cancer initiation

Supplementary Table 8a lists literature evidence for the role of cluster 1 genes in TME and CIN.

If both CIN and TME remodeling ensue simultaneously, we should ask what possible factors could drive such punctuated burst. To address this, we used recently published causal analyses tools[49] available in the Ingenuity Pathway Analysis. In particular, we performed Upstream Regulator Analysis, and Causal Network Analysis to curate all interactions of cluster 1 genes (Fig. 4a, b). Interestingly, a large subset of direct upstream interactions as well as causal interactions of both the CIN and TME genes (cluster 1), are hypoxia responsive genes[50] (Fig. 4a, b and Supplementary Table 8b, c). Invariably, almost 50% of the cluster 1 genes are also associated with poor prognosis (Fig. 5a and Supplementary Fig. 3). A multivariate expression score was derived from these 50% of the cluster 1 genes by the multivariate survival analysis to further evaluate their simultaneously effect on survival. TNBC patients were divided into two groups based on the multivariate expression score, one with high value multivariate expression and the other one with low expression. Survival probability curves of the two groups compared by the Kaplan–Meier method indicates a poor prognosis (Fig. 5b). This strongly suggests that very early in the course of tumor progression gradually increasing hypoxic conditions induce both CIN and TME remodeling to permit survival of cancer cells and their further evolution at later stages of malignancy.

DISCUSSION

Differential gene expression analyses have been traditionally used to examine fluctuations within the transcriptome in a given context for decades. This has been a powerful strategy to identify biomarkers and drug targets.[51–54] However, tumor genome sequencing has provided new opportunities to re-examine these fluctuations in the context of tumor evolution. We rationalize that common aberrations detected across all patients arose earlier in the development of the malignancy compared to alterations that were found only in a subset of patients. Based on this, our strategy in this work is to explore the frequency of changes in the expression pattern of genes at different stages of TNBC progression. This is similar to previous studies that explored dynamic changes in mutations or CNA for a given patient at a given stage.[8–10,55,56]

Changes in gene expression, unless constitutively observed, are often ignored as stochastic noise, specifically those that arise from variations in transcriptional regulation or biochemical modifications within cells. Our analyses deliver several important observations. First, compared to mutational changes, alterations within the transcriptome are more common and occur at high frequency. For example, the highly significant mutations in genes like PIK3CA or KRAS are observed in ~30% of breast cancer patients. In contrast, overexpression of PLK1 or FOXM1 genes is observed in over 90% of patients (Supplementary Fig. 4). Second, more genes are downregulated compared to upregulated genes. Third, during tumor evolution, changes in expression pattern occur as punctuated bursts, where the initial singular burst results in simultaneous accumulation of overexpression of multiple EEA genes. Fourth, early changes in the expression of EEA molecules occur in genes that remodel TME and maintain chromosomal stability. This is most likely because survival within the progressively changing biological landscape during early stages requires cancer cells to both actively adjust to their microenvironment for their needs and to enhance CIN to facilitate their plasticity and adapt. Indeed, our unbiased genome-wide investigation reveals a strong functional connection between these two mechanisms and a crucial role of their coordinated effort in establishing early tumors. Interestingly, some TME genes, including MMP1, MMP11, and MMP13 proved to be upregulated at early stages and gradually downregulated through the later stages, although never achieving their normal levels. This suggests that their activities are essential at all stages of cancer progression, but their higher activity is required in early tumors, where the TME is not adjusted yet to the needs of malignant cells. Fifth, while we know that hypoxic TME can trigger tumor metastasis and invasion at later stages of cancer progression,[57,58] our causal network analyses suggest that increasing hypoxia may be responsible for the cooperative induction of CIN and TME remodeling much earlier than previously appreciated (Fig. 5c). As hypoxic environment is also known to promote the propagation of tumor initiating cells (TICs),[59,60] we suspect that the expressional changes of EEA genes may facilitate this process.

Although, CIN is nearly ubiquitous in cancer cells, and is considered as an important factor in tumor development,[61] our findings indicate that hypoxic TME of early tumor may function as a trigger of genetic instability. This model is consistent with previous observations, showing that repeated cycles of hypoxia, can downregulate a number of DNA repair pathways in cancer cells, ultimately leading to genetic instability.[62,63] In regards to this, the Glazer group has provided one of the first quantitative assessments of how genetic instability can be instigated by TME.[64] Interestingly, several of the core EEA genes that maintain genome stability were experimentally shown to be involved in tumor development.[35–43] Although some of these examples might be indicative of a direct role for CIN genes in tumorigenesis, in the context of our analyses, we suggest that overexpression of these genes may have enabled cancer cells to acquire properties that allowed them to survive at the early-stage of cancer and thus, to develop detectable tumors (Fig. 5c). This is particularly interesting in the light of apparent disagreements between the somatic mutation theory (SMT), where mutations are argued to be the primary cause of tumor evolution and tissue organization field theory (TOFT), that proposes a direct role for TME and its surrounding tissues in tumor development.[65,66] In this context, our data agrees with a unified model, where tumor initiation is triggered by accumulated driver mutations, while its evolution depends on the interactions of cancer cells with TME (Fig. 5c).

In summary, our unbiased comprehensive analyses of the transcriptome directly link the early onset of hypoxia to the collective burst of CIN and TME remodeling factors in the initial stage of tumor progression, which highlights a therapeutic potential of targeting these molecules in TNBC tumors in their earliest detectable stage.

METHODS

Number of patient samples analyzed

We collected data of breast cancer samples from the Cancer Genome Atlas (TCGA) with information on the copy number aberration, gene expression as well as tumor information. According to the stage information, 1078 samples were classified into four tumor stages; from stage I to stage IV with tumors in stage V not being considered in this study. According to the IHC markers, 158 samples were classified as TNBC tumors in which the ER, PR, and HER2 were all negative. With the tumor stage information, we classified TNBC tumors into TNBC-stage I, TNBC-stage II, TNBC-stage III, and TNBC-stage IV. Similarly, TNBC tumors in stage IV were excluded. Moreover, 114 normal samples were collected from TCGA for comparison with tumor sample data. The numbers of samples in each stage and TNBC-stage-specific samples are displayed in the Supplemental Table 9. No human or animal experiments were performed, which required prior approval.

Fold-change and FDR calculation

We applied two criteria, fold-change and FDR calculation on the selection of differentially expressed genes. Fold-change is a biological assessment of changes in gene expression that is estimated by log2 (ratio), as represented in Equation 1, where the ratio of average expression of gene i in patients to the average expression in normal samples is calculated.

$$\mathrm{Fold}-\mathrm{change}_{\mathrm{gene}\,i}=\log_2\frac{\mathrm{ave}\left(E^{\mathrm{patients}}_{\mathrm{gene}\,i}\right)}{\mathrm{ave}\left(E^{\mathrm{normal}}_{\mathrm{gene}\,i}\right)} \tag{1}$$

Empirical Bayes moderated t-test was applied to assess the statistical significance of differential expression. False discover rate (FDR) was obtained after Benjamini and Hochberg correction. We employed the Limma package[67] to derive the two assessments of differentially expressed genes.

Computing frequency of differential expression in stage-specific patients

After the identification of up/downregulated genes in each stage and TNBC-stage patients, next we aimed to evaluate the frequency of identified differential expression in stage-specific patients. For each tumor stage, we calculated the fold-change of identified up- or downregulated gene i by comparing the expression in patient j to the average expression in normal samples (Equation 2). Following this, the frequency of patients in which the fold-change of gene i is greater than 2 or less than −2 was calculated.

$$\mathrm{Fold}-\mathrm{change}^{\mathrm{patient}\,j}_{\mathrm{gene}\,i}=\log 2\frac{E^{\mathrm{patient}\,j}_{\mathrm{gene}\,i}}{\mathrm{ave}\left(E^{\mathrm{normal}}_{\mathrm{gene}\,i}\right)} \tag{2}$$

Evaluating the concordance between copy number amplification and upregulated gene expression

As changes in gene expression may arise from the chromosomal amplification, here we aimed to evaluate the associations between copy number amplification (CNA) and upregulations in gene expression and identify the upregulations that are driven by CNA. We generated the CNA profile with patients (in rows) and genes (in columns) from the data obtained from TCGA. Only the data of patients whose CNA profile and upregulation status are available were considered for this study. Using a scoring system, genes getting amplified in a patient were represented as 1 or otherwise 0 and the amplified genes were grouped based on the profile scoring. To avoid patient heterogeneities, only genes showing amplification over 40% of patients were considered as cancer relevant amplified genes. CNA regions were identified by calculating the percentage of amplification of genes on each chromosome region, and regions with at least 40% of amplification genes were identified as CNA regions. Then we analyzed the concordance between CNA and upregulations in gene expression by two evaluation ways. Primarily, for each upregulated gene *i* at CNA regions, according to the gene *i*'s amplification status, the patient set $\boldsymbol{S}$ were grouped into two sets $\{S_i^1, S_i^2\}$, where S_i^1 and S_i^2 denotes patients with and without gene $\boldsymbol{i}$ getting amplified, respectively. The fold CNA-associated change was calculated by comparing the difference between the log2 of mean expression of gene $\boldsymbol{i}$ in S_i^1 and S_i^2. Meanwhile empirical Bayes moderated *t*-test is applied on the two groups. To correct for multiple hypothesis testing, the *p*-value was converted to FDR by Benjamini-Hochberg correction. If gene $\boldsymbol{i}$ showed at least 1 positive fold CNA-associated change and FDR smaller than 0.01, it is considered to be associated strongly with CNA. Secondly, Pearson's correlation coefficient was calculated to quantify the correlation between CNA and gene expression. If a gene showed the value of Pearson's correlation coefficient larger than 0.3, it is considered as CNA-driven genes as well.

TNBC clusters and neighboring-joining algorithm analyses

To identify the distinguishable TNBC subpopulations, which reveal similar upregulation profiles, we performed clustering analysis within TNBC patients. The binary matrix was generated with upregulation status in rows and patient in columns. If a gene is upregulated in a patient, it was indicated by 1, otherwise 0 using the scoring system described in the previous section. Then, we calculated pairwise Euclidian distance between patients and performed hierarchical clustering that clustered TNBC patients into distinguishable clusters. The mean upregulation profile for each cluster was generated to represent the upregulation status. Assuming that no gene upregulations appeared before the initiation of cancer, the profile with all zeros was generated to represent normal samples. The distance tree was constructed based on profiles in normal and mean profiles by Euclidian distance and neighbor-joining algorithm.

Identifying genes in different clusters

To determine the appearance of upregulations in various subsets of the four subpopulations, we generated the four-dimensional binary vectors to represent each of the fifteen possible subsets of the four subpopulations, from (0, 0, 0, 1) to (1, 1, 1, 1). Four each gene *i*, the cosine similarities between mean profile p_i and each subset vector v_i is calculated by the Equation 3.

$$\text{similarity}(p_i, v_j) = \cos(\theta) = \frac{p_i \cdot v_j}{\|p_i\|_2 \|v_j\|_2} \tag{3}$$

The gene was assigned to the subset vector with the maximum similarity to its mean profile.[68] For example, if a gene has the maximum similarity with the subset vector (0, 1, 1, 1), it means it getting upregulated in subpopulations 2, 3, and 4.

Multivariate survival analysis

We derived a multivariate expression score by multivariate Cox regression model, which evaluates the simultaneously effect of the multiple factors on survival. The regression coefficients in Cox regression relate to prognosis effect, thus we calculated the multivariate expression score as the sum of multiplications of the gene expression values by the respective regression coefficients. Based on the multivariate expression score, patients were divided into two groups using the median as the cutoff, one with high multivariate expression score and the other with low score. Survival analysis on the multivariate expression score was performed by Kaplan–Meier method.

IPA and hypoxia analysis

IPA analysis was performed on the genes from cluster 1 as described in Kramer et al.[49] The gene list was first annotated, and the dataset underwent various analyses including for core expression to study the interactions. The gene interactions were explored, built and different overlays including pathways, disease and function and molecule activity prediction were applied to obtain the required outputs. Comparison analysis was also performed among the different subpopulations (referred as clusters). Hypoxia analyses were performed using the hypoxia database (http://www.hypoxiadb.com). This database includes 72,000 manually curated entries taken on 3500 proteins extracted from 73 peer-reviewed publications selected from PubMed. As described in Khurana et al.[50], it provides manually curated literature references to support the inclusion of the protein in the database and establish its association with hypoxia.

DATA AVAILABILITY

The datasets analyzed during the current study are available in the TCGA databases. The authors declare that all the data generated during this study are included in this published article and its supplementary information files.

ACKNOWLEDGEMENTS

We thank members of the Vizeacoumar, Han, and Freywald laboratories for their feedback and comments. C.E.C. is supported by Lisa Rendall Breast Cancer fellowship. S.P. is supported by SCA post-doctoral fellowship. This work is supported by the operating grants from the Saskatchewan Cancer Agency operating grants to F.J.V and K.B., the SHRF Establishment grant (SHRF-3538), NSERC Discovery grant (RGPIN-2014-04110) to F.J.V.

AUTHOR CONTRIBUTIONS

A.J., F.S.V., S.P., A.F., J.H., and F.J.V. developed the concept and designed the study; A. J., F.S.V., S.P., B.H., C.E.C., and R.A. collected and analyzed data and interpreted results; A.J., F.S.V., S.P., and F.J.V. prepared the manuscript; Y.W., R.A., K.B., A.F., J.H., and F.J.V. gave technical support and edited the manuscript. All authors discussed the results and implications and commented on the manuscript.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Systems Biology and Applications* website (https://doi.org/10.1038/s41540-018-0074-z).

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
2. Burrell, R. A. & Swanton, C. The evolution of the unstable cancer genome. *Curr. Opin. Genet. Dev.* **24**, 61–67 (2014).
3. McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell.* **27**, 15–26 (2015).
4. Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer Res.* **72**, 4875–4882 (2012).
5. Armitage, P. & Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**, 1–12 (1954).
6. Nordling, C. O. A new theory on cancer-inducing mechanism. *Br. J. Cancer* **7**, 68–72 (1953).
7. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
8. Gao, R. et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet.* **48**, 1119–1130 (2016).
9. Notta, F. et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016).

10. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
11. Genomic Analysis of Pancreatic Cancers Reveals Punctuated Evolution. *Cancer Discov.* **6**, OF17 (2016).
12. Cross, W., Graham, T. A. & Wright, N. A. New paradigms in clonal evolution: punctuated equilibrium in cancer. *J. Pathol.* **240**, 126–136 (2016).
13. Davis, A., Gao, R. & Navin, N. Tumor evolution: Linear, branching, neutral or punctuated? *Biochim. Biophys. Acta* **1867**, 151–161 (2017).
14. Jallepalli, P. V. & Lengauer, C. Chromosome segregation and cancer: cutting through the mystery. *Nat. Rev. Cancer* **1**, 109–117 (2001).
15. Nowak, M. A. et al. The role of chromosomal instability in tumor initiation. *Proc. Natl Acad. Sci. USA* **99**, 16226–16231 (2002).
16. Roschke, A. V. & Rozenblum, E. Multi-layered cancer chromosomal instability phenotype. *Front. Oncol.* **3**, 302 (2013).
17. Schvartzman, J. M., Sotillo, R. & Benezra, R. Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nat. Rev. Cancer* **10**, 102–115 (2010).
18. Holland, A. J. & Cleveland, D. W. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat. Rev. Mol. Cell Biol.* **10**, 478–487 (2009).
19. Thompson, S. L. & Compton, D. A. Examining the link between chromosomal instability and aneuploidy in human cells. *J. Cell Biol.* **180**, 665–672 (2008).
20. Geigl, J. B., Obenauf, A. C., Schwarzbraun, T. & Speicher, M. R. Defining 'chromosomal instability'. *Trends Genet.* **24**, 64–69 (2008).
21. Barber, T. D. et al. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proc. Natl Acad. Sci. USA* **105**, 3443–3448 (2008).
22. Weaver, B. A. & Cleveland, D. W. Does aneuploidy cause cancer? *Curr. Opin. Cell Biol.* **18**, 658–667 (2006).
23. Gollin, S. M. Mechanisms leading to chromosomal instability. *Semin. Cancer Biol.* **15**, 33–42 (2005).
24. Grigorova, M., Staines, J. M., Ozdag, H., Caldas, C. & Edwards, P. A. Possible causes of chromosome instability: comparison of chromosomal abnormalities in cancer cell lines with mutations in BRCA1, BRCA2, CHK2 and BUB1. *Cytogenet. Genome Res.* **104**, 333–340 (2004).
25. Sieber, O. M., Heinimann, K. & Tomlinson, I. P. Genomic instability--the engine of tumorigenesis? *Nat. Rev. Cancer* **3**, 701–708 (2003).
26. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
27. Harris, R. S. Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Res.* **17**, 8 (2015).
28. Arcondeguy, T., Lacazette, E., Millevoi, S., Prats, H. & Touriol, C. VEGF-A mRNA processing, stability and translation: a paradigm for intricate regulation of gene expression at the post-transcriptional level. *Nucl. Acids Res.* **41**, 7997–8010 (2013).
29. Sotiriou, C. et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* **98**, 262–272 (2006).
30. Guenthoer, J. et al. Assessment of palindromes as platforms for DNA amplification in breast cancer. *Genome Res.* **22**, 232–245 (2012).
31. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
32. Johnson, S. C. Hierarchical clustering schemes. *Psychometrika* **32**, 241–254 (1967).
33. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
34. Wood, L. D. et al. The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).
35. Wang, I. C. et al. Forkhead box M1 regulates the transcriptional network of genes essential for mitotic progression and genes encoding the SCF (Skp2-Cks1) ubiquitin ligase. *Mol. Cell. Biol.* **25**, 10875–10894 (2005).
36. Yu, G. et al. FoxM1 promotes breast tumorigenesis by activating PDGF-A and forming a positive feedback loop with the PDGF/AKT signaling pathway. *Oncotarget* **6**, 11281–11294 (2015).
37. Smith, M. R. et al. Malignant transformation of mammalian cells initiated by constitutive expression of the polo-like kinase. *Biochem. Biophys. Res. Commun.* **234**, 397–405 (1997).
38. Li, Z. et al. Polo-like kinase 1 (Plk1) overexpression enhances ionizing radiation-induced cancer formation in mice. *J. Biol. Chem.* **292**, 17461–17472 (2017).
39. Ricke, R. M., Jeganathan, K. B. & van Deursen, J. M. Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation. *J. Cell Biol.* **193**, 1049–1064 (2011).
40. de Carcer, G. & Malumbres, M. A centrosomal route for cancer genome instability. *Nat. Cell Biol.* **16**, 504–506 (2014).
41. Nam, H. J. & van Deursen, J. M. Cyclin B2 and p53 control proper timing of centrosome separation. *Nat. Cell Biol.* **16**, 538–549 (2014).
42. Li, M., Fang, X., Wei, Z., York, J. P. & Zhang, P. Loss of spindle assembly checkpoint-mediated inhibition of Cdc20 promotes tumorigenesis in mice. *J. Cell Biol.* **185**, 983–994 (2009).
43. Urata, Y. N., Takeshita, F., Tanaka, H., Ochiya, T. & Takimoto, M. Targeted Knockdown of the Kinetochore Protein D40/Knl-1 Inhibits Human Cancer in a p53 Status-Independent Manner. *Sci. Rep.* **5**, 13676 (2015).
44. Marchesi, S. et al. DEPDC1B coordinates de-adhesion events and cell-cycle progression at mitosis. *Dev. Cell* **31**, 420–433 (2014).
45. Yang, Y. et al. DEPDC1B enhances migration and invasion of non-small cell lung cancer cells via activating Wnt/beta-catenin signaling. *Biochem. Biophys. Res. Commun.* **450**, 899–905 (2014).
46. Chen, D. et al. Phosphorylation of DEPDC1 at Ser110 is required to maintain centrosome organization during mitosis. *Exp. Cell Res.* **358**, 101–110 (2017).
47. Connell, M. et al. HMMR acts in the PLK1-dependent spindle positioning pathway and supports neural development. *eLife* **6**, pii: e28672 (2017).
48. Veiseh, M. & Turley, E. A. Hyaluronan metabolism in remodeling extracellular matrix: probes for imaging and therapy of breast cancer. *Integr. Biol. (Camb.)* **3**, 304–315 (2011).
49. Kramer, A., Green, J., Pollard, J. Jr & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014).
50. Khurana, P., Sugadev, R., Jain, J. & Singh, S. B. HypoxiaDB: a database of hypoxia-regulated proteins. *Database (Oxf.)* **2013**, bat074 (2013).
51. Glas, A. M. et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genom.* **7**, 278 (2006).
52. Ludwig, J. A. & Weinstein, J. N. Biomarkers in cancer staging, prognosis and treatment selection. *Nat. Rev. Cancer* **5**, 845–856 (2005).
53. Scherf, U. et al. A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.* **24**, 236–244 (2000).
54. Frank, R. & Hargreaves, R. Clinical biomarkers in drug discovery and development. *Nat. Rev. Drug Discov.* **2**, 566–580 (2003).
55. Sun, Y., Yao, J., Nowak, N. J. & Goodison, S. Cancer progression modeling using static sample data. *Genome Biol.* **15**, 440 (2014).
56. Navin, N. et al. Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20**, 68–80 (2010).
57. Rankin, E. B. & Giaccia, A. J. Hypoxic control of metastasis. *Science* **352**, 175–180 (2016).
58. Schito, L. & Rey, S. Hypoxic pathobiology of breast cancer metastasis. *Biochim. Biophys. Acta* **239-245**, 2017 (**1868**).
59. Lin, Q. & Yun, Z. Impact of the hypoxic tumor microenvironment on the regulation of cancer stem cell characteristics. *Cancer Biol. Ther.* **9**, 949–956 (2010).
60. Carnero, A. & Lleonart, M. The hypoxic microenvironment: A determinant of cancer stem cell evolution. *Bioessays* **38**(Suppl 1), S65–S74 (2016).
61. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
62. Klein, T. J. & Glazer, P. M. The tumor microenvironment and DNA repair. *Semin. Radiat. Oncol.* **20**, 282–287 (2010).
63. Yuan, J., Narayanan, L., Rockwell, S. & Glazer, P. M. Diminished DNA repair and elevated mutagenesis in mammalian cells exposed to hypoxia and low pH. *Cancer Res.* **60**, 4372–4376 (2000).
64. Reynolds, T. Y., Rockwell, S. & Glazer, P. M. Genetic instability induced by the tumor microenvironment. *Cancer Res.* **56**, 5754–5757 (1996).
65. Soto, A. M. & Sonnenschein, C. The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory. *Bioessays* **33**, 332–340 (2011).
66. Sonnenschein, C. & Soto, A. M. Carcinogenesis explained within the context of a theory of organisms. *Prog. Biophys. Mol. Biol.* **122**, 70–76 (2016).
67. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl. Acids Res.* **43**, e47 (2015).
68. Navin, N. et al. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).