HHS Public Access

Author manuscript

Cancer Res. Author manuscript; available in PMC 2016 September 15.

Published in final edited form as:

Cancer Res. 2015 September 15; 75(18): 3980–3990. doi:10.1158/0008-5472.CAN-15-0506.

A Molecular Portrait of High-Grade Ductal Carcinoma *In Situ*

Martin C. Abba[1], **Ting Gong**[2], **Yue Lu**[2], **Jaeho Lee**[2], **Yi Zhong**[2], **Ezequiel Lacunza**[1], **Matias Butti**[1], **Yoko Takata**[2], **Sally Gaddis**[2], **Jianjun Shen**[2], **Marcos R. Estecio**[2,3], **Aysegul A. Sahin**[3], and **C. Marcelo Aldaz**[2]

[1]CINIBA, School of Medical Sciences, National University of La Plata, La Plata, Argentina

[2]The University of Texas MD Anderson Cancer Center, Smithville, Texas

[3]The University of Texas MD Anderson Cancer Center, Houston, Texas

Abstract

Ductal carcinoma *in situ* (DCIS) is a noninvasive precursor lesion to invasive breast carcinoma. We still have no understanding on why only some DCIS lesions evolve to invasive cancer whereas others appear not to do so during the life span of the patient. Here, we performed full exome (tumor vs. matching normal), transcriptome, and methylome analysis of 30 pure high-grade DCIS (HG-DCIS) and 10 normal breast epithelial samples. Sixty-two percent of HG-DCIS cases displayed mutations affecting cancer driver genes or potential drivers. Mutations were observed affecting *PIK3CA* (21% of cases), *TP53* (17%), *GATA3* (7%), *MLL3* (7%) and single cases of mutations affecting *CDH1*, *MAP2K4*, *TBX3*, *NF1*, *ATM*, and *ARID1A*. Significantly, 83% of lesions displayed numerous large chromosomal copy number alterations, suggesting they might precede selection of cancer driver mutations. Integrated pathway-based modeling analysis of RNA-seq data allowed us to identify two DCIS subgroups (DCIS-C1 and DCIS-C2) based on their tumor-intrinsic subtypes, proliferative, immune scores, and in the activity of specific signaling pathways. The more aggressive DCIS-C1 (highly proliferative, basal-like, or ERBB2$^+$) displayed signatures characteristic of activated Treg cells (CD4$^+$/CD25$^+$/FOXP3$^+$) and CTLA4$^+$/CD86$^+$ complexes indicative of a tumor-associated immunosuppressive phenotype. Strikingly, all lesions showed evidence of TP53 pathway inactivation. Similarly, ncRNA and methylation profiles reproduce changes observed postinvasion. Among the most significant findings, we observed

Corresponding Author: C. Marcelo Aldaz, The University of Texas MD Anderson Cancer Center, 1808 Park Road 1C, Smithville, TX 78957. Phone: 512-237-9530; Fax: 512-237-2475; maaldaz@mdanderson.org.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A.A. Sahin, C.M. Aldaz

Development of methodology: J. Shen, A.A. Sahin

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Lee, Y. Takata, J. Shen, M.R. Estecio, C.M. Aldaz

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M.C. Abba, T. Gong, Y. Lu, Y. Zhong, E. Lacunza, M. Butti, A.A. Sahin, C.M. Aldaz

Writing, review, and/or revision of the manuscript: M.C. Abba, T. Gong, E. Lacunza, M. Butti, J. Shen, C.M. Aldaz

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Butti, S. Gaddis, C.M. Aldaz

Study supervision: C.M. Aldaz

upregulation of lncRNA *HOTAIR* in DCIS-C1 lesions and hypermethylation of *HOXA5* and *SOX* genes. We conclude that most HG-DCIS lesions, in spite of representing a preinvasive stage of tumor progression, displayed molecular profiles indistinguishable from invasive breast cancer.

Introduction

Ductal carcinoma *in situ* (DCIS) is by definition a precursor lesion to invasive ductal carcinoma (IDC). The routine use of screening mammography has led to a dramatic increase in DCIS detection during the last four decades (1). More than 62,000 cases per year account for about 25% of new breast cancers in the United States (1). Early retrospective studies concluded that women with biopsy proven DCIS have over a 10-fold higher risk for developing invasive carcinoma than women without DCIS history (2). If left untreated, it has been estimated that at least one-third of DCIS cases are likely to progress to IDC during the lifetime of the patient (1, 3). Regardless, it is beyond debate that the vast majority of IDCs results from the progression of precursor DCIS lesions (3). DCIS lesions are usually classified as either high-grade (HG-DCIS) or low-grade (LG-DCIS). Such distinction has limited impact on clinical management, as most patients with DCIS regardless of grade still receive similar treatment, which is either total mastectomy or segmental resection with radiation therapy (4–6). Recent studies argued that it is possible to identify more indolent DCIS lesions based on a small gene expression panel (Oncotype DX DCIS; ref. 7); however, concerns have been raised on the usefulness of such test and further validation is required (8, 9). Nevertheless, there is a great clinical need to identify those patients who can be spared radiation therapy after segmental resection, as it has been argued that many patients do not benefit from this likely overtreatment (10, 11).

Our current understanding of DCIS is that these are heterogeneous lesions and just like invasive carcinoma, DCIS lesions can be separated into different intrinsic subtypes on the basis of gene expression features (12–15). Mutations affecting genes such as *TP53* and *PIK3CA* were also reported in DCIS (16, 17). However, as reviewed by Polyak, molecular signatures predictive of invasive progression have not been defined (18). In other words, we still have no understanding on why only some DCIS lesions evolve to invasive cancer whereas others appear not to do so. Importantly, no study to date has comprehensively delineated the molecular landscape of DCIS at the mutational, transcriptomic, and epigenetic levels in the same lesions and these are in part major goals of the present study.

Materials and Methods

Samples

Thirty fresh-frozen pure HG-DCIS cases with matched normal adjacent breast tissue samples were obtained from the MDACC Breast Tumor Bank. Normal breast tissue is defined as grossly unremarkable breast parenchyma away from the area of DCIS identified by imaging studies and gross evaluation and confirmed by subsequent histologic evaluation. These areas usually selected at least 1 cm and more away from the lesions. Patient tissue samples were collected after proper informed consents were obtained and protocols approved by ethical and institutional review boards. Ten cosmetic normal mammoplasty

specimens were obtained from the Cooperative Human Tissue Network. Normal breast epithelial organoids were freshly isolated from the normal mammoplasty specimens as was previously described by us (19) and used for identification of differentially expressed genes (DEG) and differentially methylated genes (DMG) between normal and HG-DCIS samples.

DNA from 29 of 30 (97%) pure HG-DCIS samples and their paired normal breast tissue samples (total 58 samples) were subjected to exome capture sequencing analysis (Exome-Seq). DNA from 24 of 30 (80%) HG-DCIS samples and 5 normal breast organoids (total 29 samples) were subjected to reduced representation bisulfite sequencing analysis (RRBS). RNAs from 25 of 30 (83%) pure HG-DCIS and 10 normal breast organoids (total 35 samples) were subjected to RNA sequencing analysis (RNASeq; see Supplementary Table S1).

Histopathologic and immunohistochemical analysis

Histopathologic analysis of tumors and lymphocytic infiltrates were performed on hematoxylin and eosin (H&E)-stained sections. Intratumor-infiltrating lymphocytes (iTIL) are defined as intraepithelial mononuclear cells within DCIS cell nests or in direct contact with cells (intralobular stroma) and are reported as the percentage of the tumor epithelial nests that contain infiltrating lymphocytes. Stromal tumor-infiltrating lymphocytes (sTIL) are defined as the percentage of tumor stroma area that contains lymphocytic infiltrates without direct contact to DCIS cells. All samples were analyzed for estrogen receptor (ER)/progesterone receptor (PR) by immunohistochemistry using standard procedures.

Exome-Seq analysis

DNA from 29 DCIS samples and their paired normal breast tissue samples (n = 58 samples) were purified using the DNeasy Blood and Tissue Kit (Qiagen). Only DNA samples with 260/280 ratios greater than 2.0 were processed for library construction using the SPRIworks Fragment Library Kit I (Beckman Coulter). Four libraries were pooled together and processed for exome capture using the NimbleGen SeqCap EZ Human Exome Library v3.0 (HG19, Roche), covering more than 23,000 genes and ~64 Mbp, 76-nt paired-end sequencing was performed using an Illumina HiSeq2000 platform at our Department's NGS Facility. Image analysis, base-calling, and error calibration were performed using Illumina's Genome analysis pipeline. Sequencing was performed reaching an average depth of 40× per sample. Sequenced 76-bp tags were aligned against the human reference genome (hg19) using BWA v0.7.3 and marked for duplicates using Picard v1.88 (http://picard.sourceforge.net/). Recalibration of base quality and indel realignment was performed using the GATK v2.4 (20). Subsequently, single-nucleotide variants (SNV) were identified using MuTect v1.1.4 (21). Identified variants were annotated using ANNOVAR (22), filtered by functional consequence, and only nonsynonymous and splicing variants were selected. In addition, we used Control-FREEC software to detect copy number alterations and LOH regions in DCIS samples on the basis of Exome-Seq data (23).

RRBS analysis

We analyzed by RRBS a total of 5 normal breast epithelial samples and 24 DCIS samples. DNA libraries were prepared for RRBS at the MDACC's DNA Methylation Core Facility

and subjected to next-generation sequencing (NGS). Short read sequences were mapped to the human reference genome 19 by Bismark, a program to map bisulfate-treated sequencing reads to a genome of interest and perform methylation calls in a single step (24). Raw datasets have been submitted to NCBI GEO database with accession number GSE69994. Promoter regions were calculated on the basis of RefSeq gene annotations, such that the region starts 2-kb upstream of the annotated transcription start site and extends to 500 downstream of the transcription start site (TSS). CpG islands were downloaded from UCSC genome annotation database assembled by NCBI (hg19). CpG shores were defined as 2,000-bp flanking regions on upstream and downstream of a given CGI. To identify differentially methylated sites, first we filtered samples on the basis of read coverage ≥ 20. And the significant differentially methylated CpGi (DMC) sites were identified when the difference of methylation percentages between normal breast epithelia and DCIS were greater than 0.25 and $q < 0.01$.

RNA-Seq analysis

RNA was isolated and purified using TRIzol reagent (Life Technologies) and RNeasy mini kit (Qiagen). RNA concentration and integrity were measured on an Agilent 2100 Bioanalyzer (Agilent Technologies). Only RNA samples with RNA integrity values (RIN) above 8.0 were considered for subsequent analysis. mRNA from normal breast epithelial samples and DCIS samples were processed for directional mRNA-Seq library construction using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre) according to the manufacturer's protocol. We performed 76-nt paired-end sequencing using an Illumina HiSeq2000 platform and obtained about 40 million tags per sample. The short sequenced reads were mapped to the human reference genome (hg19) by the splice junction aligner TopHat (25). We used several R/Bioconductor packages to accurately calculate the gene expression abundance at the whole-genome level using the aligned records (BAM files) and to identify differentially expressed genes between normal and DCIS samples. Briefly, the number of reads mapped to each gene on the basis of the UCSC.hg19.KnownGene database were counted, reported, and annotated using the GenomicFeatures, Rsamtools, and org.Hs.eg.db packages. Raw datasets have been submitted to NCBI GEO database with accession number GSE69994. To identify differentially expressed genes between normal breast epithelium and DCIS samples, we used the edgeR-test on the basis of the normalized number of reads mapped to each gene (26).

Pathway-based analysis was performed using the PARADIGM software at the Five3 Genomics server (default options; discretization bounds of 33%) on the basis of the normalized gene expression profiles of the deregulated transcripts between normal and DCIS samples [false discovery rate (FDR) < 0.01; log fold change (FC) ≥ ±1] expressed in $\log_2$ counts per million (27). PARADIGM produces a data matrix of integrated pathway activities (IPA), this data matrix was used in place of the mRNA expression profiles to identify the topmost variable IPAs among samples. Heatmap visualization of differentially expressed transcripts and IPAs were done with the MultiExperiment Viewer software (MeV v4.9; ref. 28). Intrinsic subtype classification of DCIS samples into luminal-like, basal-like, ERBB2-enriched, and normal-like groups was performed using the 50-gene (PAM50) predictor bioclassifier R script (29). We used the ESTIMATE algorithm (Estimation of

STromal and Immune cells in Malignant Tumors using Expression data) to infer the immune and stromal components from each DCIS sample (30).

Bioinformatic analysis of the The Cancer Genome Atlas-BRCA dataset

To perform a comparative analysis of the mutational, gene expression, methylation, and pathway activities profiles identified in DCIS with invasive stages, we analyzed invasive breast cancer (IBC) datasets obtained from the The Cancer Genome Atlas (TCGA)-BRCA project (31). The clinical (including PAM50 subtypes from RNAseq) and preprocessed data (IBC gene-level mutation, IBC gene expression Illumina HiSeq, IBC DNA methylation 450K and IBC Paradigm RNAseq + CNV profiles) were retrieved from the Cancer Genomics Browser (https://genome-cancer.soe.ucsc.edu/) and cBioPortal (http://www.cbioportal.org/). Data integration and visualization were done with R and MeV software, respectively.

Results and Discussion

Fresh-frozen pure DCIS samples are extremely difficult to obtain, as all tumor materials are submitted for routine histopathologic evaluation with formalin fixation and paraffin embedding. Only cases with available frozen tissue samples are usually those from large palpable DCIS cases, which tend to be HG-DCIS. Nevertheless, we were able to gather from our Institution's breast tumor bank a group of fresh-frozen "pure" HG-DCIS samples (i.e., with no evidence of invasion in the same breast) and matching normal pairs and subjected them to a thorough characterization involving whole Exome-Seq, RRBS, and RNA-Seq. We also generated RNA-Seq and RRBS data from breast epithelial organoids obtained from normal mammoplasty specimens for identification of differentially expressed genes and differential methylated genes between normal and DCIS lesions.

Significantly mutated genes and DNA copy number variations in DCIS

Exome-Seq data on 29 pure DCIS normal matched pairs indicated a median of 75% targeted genome loci having at least 40× coverage. We detected 2,908 single-base substitutions, including 1,324 nonsynonymous (missense) SNVs, 46 stop gain, 2 stop loss, 16 mutations at splicing sites, 151 noncoding RNA (ncRNA) mutations, and 1,369 SNVs encoded synonymous (silent) mutations (Fig. 1A; Supplementary Table S2). Almost 50% of these mutations were C>T:G>A transitions in agreement with the previously described most prevalent mutation-type signature in breast cancer (Fig. 1A; ref. 32). We also detected 24 frameshift deletions, 10 frameshift insertions, and 43 non-frameshift deletion/insertion events. Every single DCIS displayed a significant number of mutations. The total mutation rate was 1.61 mutations per Mbp on average with a range of 0.8 to 3.8 mutations per Mbp (Fig. 1B), indicating that some DCISs have quite higher mutation rates than others. We identified somatic mutations affecting at least 10 genes reported as mutated at a ≥2% rate in IBC (33). In total, 18 of 29 HG-DCIS cases (62%) displayed mutations affecting one or combinations of targets described as cancer driver genes or potential drivers. Among these, we detected mutations affecting *PIK3CA* (21%), *TP53* (17%), *GATA3* (7%), *MLL3* (7%) and single cases of mutations affecting *CDH1*, *MAP2K4*, *TBX3*, *NF1*, *ATM*, and *ARID1A* (Fig. 1B). The comparative frequency of mutations of these genes in pure DCIS and the TCGA

database on IBC (n = 976 cases as per September 2014) is shown in Fig. 1C. Similarly to IBC previous gene mutation data (31, 33), the most frequently mutated genes are *PIK3CA* (6 cases; 4 of them affecting ER/PR-positive cases) and *TP53* (5 cases; 4 of them affecting ER/PR-negative cases). Only one case displayed mutations in both genes. Interestingly, several genes reported as mutated in IBC, such as *MAP3K1*, *PTEN*, *AKT1*, *RUNX1*, *RB1*, and various others so far, have not been detected mutated in DCIS, suggesting that perhaps some of these mutations may be associated with postinvasion events; however, our sample number is still limited. In summary, we observed that 100% of pure HG-DCIS display numerous somatic mutations, but only a subgroup display mutations affecting cancer driver genes or potential drivers. Nevertheless, the mutational profile in pure HG-DCIS is extremely similar to that observed at invasive stages with only a moderate lower frequency of mutation for cancer driver genes. As we performed RNA-Seq in parallel, the identified mutations were validated by allele-specific RNA-Seq data.

Somatic copy number variations (CNV) were estimated using Exome-Seq data to predict regions of genomic alterations such as amplification, gains, and losses (23). Figure 1D summarizes the overall frequency of DNA gains/losses affecting all chromosome arms among DCIS samples. The profile of copy number changes across the genome in HG-DCIS is practically identical to profiles reported in invasive breast lesions (34) and is in agreement with early observations using comparative genome hybridization (CGH) approaches (35). In our approach, however, we were able to define with more precision the genomic regions spanning the CNV (Supplementary Table S2). The most common regions of increased DNA copy number include chr1q, chr8q, chr17q, and chr20q and regions of common copy number loss include regions chr8p, chr11q, chr17p, and chr22q (Fig. 1D; Supplementary Fig. S1 and Supplementary Table S2). Regions of potential focal amplification were also identified, including *ERBB2*, *VEGFA*, *MYC*, *AURKA*, *MDM2*, *FGFR1*, and *CCND1* (Fig. 1B and Supplementary Table S2).

Among the most important conclusions of the CNV analysis is that 83% (24 of 29) of DCIS lesions displayed evidence of large chromosomal alterations; only in 5 samples (T9, T15, T28, T30, and T31), large chromosomal copy number changes were not detected (Supplementary Table S2). Interestingly, multiple cases (T2, T4, T5, T10, T13, T16, T22, and T24) displayed significantly large chromosomal copy number alterations but no mutations in known cancer driver genes, whereas 16 of 18 cases (89%) with mutations in cancer driver genes all show significant chromosomal copy number alterations (Fig. 1B; Supplementary Table S2). Importantly, in only 3 of 29 (10%) HG-DCIS (T9, T15, T28), no large chromosomal abnormalities or mutations in putative cancer driver genes were detected. Nevertheless our findings also suggest that large chromosomal copy number alterations (trisomies, monosomies, large chromosomal duplications, and deletions) occur early in tumor progression, perhaps preceding the selection of alleles with mutations in driver genes.

Gene expression analysis of coding RNAs

Whole-transcriptome unsupervised analysis from RNA-Seq data demonstrates a clear segregation of normal breast epithelium and DCIS samples on the basis of similarity

distances (measured in leading logFC) from dimension 1 in a multidimensional scaling plot (MDS; Fig. 2A). More importantly, DCIS samples were clustered into two subgroups (cluster 1 or DCIS-C1 and cluster 2 or DCIS-C2) according to the positive and negative distances from MDS dimension 2.

We observed that the ER/PR status determined by immunohistochemistry and their intrinsic subtypes predicted by using the 50-gene PAM50 model (29) were significantly associated with each identified DCIS cluster ($P < 0.001$). Sixty-seven percent of the cases in DCIS-C1 were ER/PR double negative and basal-like or ERBB2$^+$ with high proliferative scores. On the other hand, 100% of the cases in DCIS-C2 were ER/PR double positive, luminal-like with low/moderate proliferative scores (Fig. 2B).

Statistical analysis of RNA-Seq data revealed 5,985 transcripts differentially expressed between normal and DCIS samples (FDR < 0.01; logFC ≥ ±1), 95% were coding RNAs and 5% were ncRNAs. Among the deregulated genes, 885 were differentially expressed between both DCIS clusters (FDR < 0.01; logFC ≥ ±1; Supplementary Table S3). To identify bioprocesses that differentiate DCIS-C1 from the DCIS-C2, we performed functional enrichment and pathway activity analyses. Gene ontology (GO) analysis of deregulated transcripts between DCIS clusters revealed specific functional modules characteristic for each of both subgroups: a functional module strongly related to immune ($P < 1.46E{-}35$) and inflammatory ($P < 6.4E{-}19$) responses clearly associated to the DCIS-C1 group and a module related with response to steroid hormones ($P < 9.1E{-}4$) associated to the DCIS-C2 group (Fig. 2C and Supplementary Fig. S2). In addition, functional annotation analysis identified "T-cell differentiation", "B-cell activation", "angiogenesis," "negative regulation of apoptosis," and "ECM remodeling" to be highly associated with DCIS-C1 -modulated genes. Given the aforementioned strong association with immune-related themes, we performed a blind evaluation of lymphocytic infiltrates on H&E sections of all DCIS samples determining iTIL and sTIL scores. Supporting the GO and functional annotation observations, DCIS-C1 samples showed a significantly higher number of iTILs ($P < 0.01$) than the DCIS-C2 counterparts. Similarly, a gene expression signature–based method known as ESTIMATE (30) predicted higher immune scores on DCIS-C1 versus DCIS-C2 ($P < 0.001$; Fig. 2D).

Further analysis of the behavior of the 885 deregulated transcripts that discriminate both DCIS clusters within the TCGA breast cancer dataset demonstrated that a large number of genes (~80%)identified as upregulated in the DCIS-C1 group remained upregulated in basal-like and ERBB2$^+$ breast cancers compared with the luminal A/B subtypes (Fig. 2E). This is in itself very significant, and similarly to the above described mutational and CNV analyses, it indicates that most of the gene expression changes characteristic of IBC are already present at the pure DCIS (preinvasive) stage.

Pathway-based representation analysis (PARADIGM; ref. 27) of deregulated transcripts identified a plethora of signaling pathways that differ in their activity between normal and DCIS samples, such as AP1, TNF, TP53, FOXM1, MYB, and E2F1/DP1 pathways (Fig. 3A; Supplementary Table S3). It was striking to observe that the TP53 signaling pathway was found deactivated in all DCIS samples analyzed regardless of TP53 mutation status,

ER/PR expression status or intrinsic subtype when compared with normal epithelium (Fig. 3B). Since on the basis of mutational evidence is currently believed that TP53 inactivation is mostly a feature of the basal-like intrinsic subtype (31, 33), we were surprised by our finding, thus we decided to perform a pathway-based analysis focused on the TP53 pathway exclusively in the existing TCGA IBC dataset. Interestingly, PARADIGM analysis based on TCGA-BRCA RNA-Seq and CNV profiles predict that TP53 pathway inactivation is a common theme affecting well above 85% of breast cancers regardless of intrinsic subtypes (and as consequence also regardless of *TP53* mutation status; Supplementary Fig. S3). Nevertheless, this novel information on TP53 pathway inactivation in most breast cancers should be taken with caution and requires further validation.

In the DCIS-C2 cases, the activity of the E2/ERa pathway was the predominant signature (Supplementary Fig. S2), whereas in the DCIS-C1 subgroup, the HES1 and HIF1A/ARNT pathways were more frequently activated. Interestingly, the HIF1A/ARNT was previously identified as one of the key regulatory features associated with basal-like carcinomas (31).

More importantly, PARADIGM analysis identified that 80% of DCIS-C1 cases showed activity of Treg cells ($CD4^+/CD25^+/FOXP3^+$) and $CTLA4^+/CD86^+$ complexes (Fig. 3C). These results are revealing, as a growing body of evidence indicates that the outcome of an immune response toward a tumor is largely determined by the specific characteristics of the infiltrating lymphocytes. A tumor-directed immune response involving cytolytic $CD8^+$ T cells, T-helper 1 cells (T_h1), and natural killer (NK) cells appears to protect against tumor development and progression, whereas activation of humoral immunity involving B cells and/or T_h2 response appears associated to protumorigenic phenotypes (36). Suppression of antitumor immune response by inducing T-cell anergy due T_h2-polarized activity and/or expansion of Treg cells ($CD4^+CD25^+FOXP3^+$) with a subsequent loss of T-cell–mediated cytotoxicity, together with the development of angiogenesis and tissue remodeling could be instrumental for promoting the progression of HG-DCIS to the infiltrating stages. Recently, Kristensen and colleagues (37) suggested that the perturbation in the immune response and IL signaling (IL4, IL6, IL12, and IL23) can lead to classification of IBC subclasses with prognostic value. Patients who have basal-like breast carcinomas are characterized by high expression of protumorigenic T_h2/humoral-related genes and a low T_h1/T_h2 ratio. In addition, the immune response and IL signaling identified in IBC appears also prominent in DCIS (37). Importantly, the IPA described in our study indicates that DCIS-C1 lesions are characterized by a tumor-associated immunosuppressive phenotype, suggesting a protumorigenic polarization of the immune response. In other words, DCIS-C1 lesions bear all the immunomodulatory characteristic of invasive breast lesions, thus indicating that the host immune defenses (tumor rejection responses) have been or are in the process of being defeated.

Gene expression analysis of noncoding RNAs

RNA-Seq analysis allowed us to identify 193 long noncoding RNAs (lncRNA) as differentially expressed between normal and DCIS samples ($P < 0.01$, FDR < 0.01), 127 of which were antisense RNAs and 66 were long intergenic noncoding RNAs (lincRNA; Fig. 4A and B). Eighty-seven percent of the identified lncRNAs (168 of 193) were upregulated in

DCIS (Supplementary Table S4). Interestingly, 46 of the 127 antisense RNAs and 18 of the 66 lincRNAs have been recently identified as potentially relevant in breast cancer (38). The Genomic Regions Enrichment of Annotations Tool (GREAT) resource was used to evaluate the correlation between deregulated lncRNAs and their neighboring genes (39). We found that 52% of the mRNA genes in the vicinity of the affected lncRNAs were also deregulated (Supplementary Table S4), and most of them (80%) showed a positive association with the expression of the lncRNAs, which would indicate that deregulation of *cis*-acting lncRNA prevails in DCIS. Among the deregulated lnRNAs in DCIS, we found *HOTAIR* (*HOX transcript antisense RNA*; logFC = +5.03; FDR = 1.92E−8) and *HOTAIRM1* (*HOXA transcript antisense RNA myeloid–specific 1*; logFC = −1.90; FDR = 8.63E−7; Fig. 4C). *HOTAIR* and *HOTAIRM1* expression levels were positively correlated with the expression of their adjacent *HOXC* (located at chr12q13.13) and *HOXA* (located at chr7q15.2) gene clusters, respectively (Fig. 4D). Recent studies have shown that lncRNAs can be associated with enhancer regions, leading to increased activity of neighboring genes (40). One of the best-known lncRNAs, *HOTAIR*, found significantly upregulated in DCIS-C1 lesions, was suggested to promote breast cancer metastasis and shown to be pervasively overexpressed in most human cancers when tumor tissue was compared with adjacent noncancerous tissue (41). *HOTAIR* interacts with and recruits the Polycomb-repressive complex 2 (PRC2) and regulates chromosome occupancy of EZH2 (a subunit of PRC2), which leads to histone H3 lysine 27 trimethylation of the *HOXD* locus and subsequent silencing of this gene cluster. On the other hand, *HOTAIRM1*, found significantly downregulated in many DCIS lesions (particularly DCIS-C1), was shown to play a role in the regulation of gene expression during retinoic acid–induced myeloid differentiation preventing the induction of *HOXA* genes (42). Mechanistically, our data indicate that *HOTAIR* and *HOTAIRM1* may act by modulating *HOX* gene expression in trans and cis during early stages of breast cancer progression. This is the first study demonstrating that these lncRNAs are deregulated "early" at preinvasive stages of breast cancer progression. The remaining lncRNAs (*LINC00277*, *LINC00861*, *LINC00578*, *LINC00426*, and others shown in Supplementary Table S4) would constitute a novel group of lncRNAs deregulated in the transition normal DCIS.

In addition, we were able to identify 36 differentially expressed pri/pre-miRNAs among the deregulated noncoding transcripts between normal and DCIS as well. We detected *MIR3606* (logFC = +6.58), *MIR4728* (logFC = +4.25), and *MIR503HG* (logFC = +2.91) among the most upmodulated small noncoding RNAs in DCIS samples. Interestingly, *MIR4728* was specifically upmodulated in DCIS-C1 samples (logFC = +2.28). *MIR4728* gene is located at chr17q12 and encoded within intron 24 of the *ERBB2* gene (43). Recently, Newie and colleagues demonstrated that *ERBB2* amplification might lead to *ESR1* downregulation through internal seed interaction with *miR-4728-3p* in breast cancer cells (44). We corroborate this observation identifying a positive correlation between *ERBB2* and *MIR4728* expression ($r = 0.90$; $P < 0.01$) in DCIS samples in agreement with the CNV data.

On the other hand, *MIR4260* (logFC = −4.62), *MIRNA221* (logFC = −3.55), *MIR22/MIR22HG* (logFC = −2.34), *MIR3661* (logFC = −2.07), and *MIR17HG* (logFC = −1.94), among others, were detected as downmodulated in DCIS samples. The human *MIR22* gene is located in a minimal LOH region on chromosome 17 close to *TP53*. *MIR22* has been

reported to be downregulated in hepatocellular, lung, colorectal, ovarian and breast cancer, acting as a tumor suppressor (45). Furthermore, *MIR22* overexpression induces growth suppression and senescence-like phenotypes in human breast epithelial and breast cancer cells (46). *MIR17HG* is the host gene for the *MIR17–20* cluster (a group of at least six miRNAs) located at 13q31 that function as a tumor suppressor in human breast cancer by decreasing *AIB1* and cyclin D1 expression (47). The *MIR17–20* cluster is known to inhibit breast cancer cellular proliferation through G_1–S cell-cycle arrest (48). Our study suggests that gain and loss of expression of key regulatory miRNAs in normal breast epithelium might constitute important early procarcinogenic events conducive to overcome the barrier imposed by senescence and limited cell proliferation. However, further studies on the profiling of the mature miRNAs in normal and DCIS samples are needed to corroborate the aforementioned observations.

DNA methylation analysis

RRBS data analysis coupled with RNA-Seq profiling allowed us to identify the most relevant methylation events associated to the development of DCIS lesions. We identified 1,103 DMCs between normal breast epithelia and DCIS samples (Supplementary Table S5), mapping at distal (8% at −4 kb to −2 kb relative to TSS), proximal promoter regions (56% at −2 kb to +500 bp) or the gene body (36% at +500 to +2 kb) of 311 loci (Fig. 5A). Among these 1,103 DMC, 1,029 were hypermethylated and 74 were hypomethylated CpGi sites in DCIS samples. Functional enrichment analysis of the hypermethylated regions in DCIS identified "regulation of transcription/RNA pol II activity," "cell fate commitment/ morphogenesis," and "regulation of cell proliferation and differentiation" as the predominant associated biologic processes (Fig. 5B). Furthermore, we identified 137 DMC for which mRNA abundance was inversely correlated in DCIS samples (Supplementary Table S5). Sixteen genes were highly hypermethylated and downmodulated in DCIS compared with normal epithelium (Fig. 5C). Some of the most hypermethylated CpGi regions were located at the promoters of three *SOX* family members (*SOX10*, *15*, and *17*) and the *HOXA5* gene. SOX10, SOX15, and SOX17 are novel HMG box–containing tumor suppressors involved in a variety of developmental processes that can act as antagonists to the Wnt/β-catenin signaling pathway (49–51). These three *SOX* genes were systematically hypermethylated and transcriptionally downregulated in almost all DCIS samples compared with normal samples. In agreement with these findings, gene expression analysis of these *SOX* genes in the TCGA dataset showed a significant downmodulation in IBC compared with normal samples (Supplementary Fig. S4). In a recent study, it was shown that the SOX10 protein is predominantly expressed in basal-like breast carcinomas compared with the other subtypes (52). However, TCGA data analysis clearly showed that *SOX10* was downmodulated in luminal-like and $ERBB2^+$ breast cancer subtypes compared with normal tissue samples (Supplementary Fig. S4). Fu and colleagues (50) demonstrated that silencing of *SOX17* due to promoter hypermethylation is frequent in invasive carcinomas and may contribute to aberrant activation of Wnt signaling in breast cancer. This evidence raises the possibility that hypermethylation of specific *SOX* family members could constitute common important epigenetic phenomena occurring at early stages of breast cancer development that deserves further study.

Interestingly, in our dataset, *HOXA5* hypermethylation was more frequently associated with the DCIS-C1 group, characterized by the presence of tumors with basal-like/ERBB2 intrinsic subtypes (Fig. 5D). *TP53* gene expression has been shown to be under the control of HOXA5 in breast cancer cells (53). Reduced HOXA5 protein levels correlate with the lack of p53 expression, supporting the concept that HOXA5 may act as a tumor suppressor via activation of p53 expression. HOXA5 has also been shown to be a strong positive regulator of PR (54). It is also important to note that 4 of 5 cases with *TP53* non-synonymous mutations were detected in ER/PR double negative tumors from the DCIS-C1 group (T3, T7, T26, and T29). Overall, these data suggest that *TP53* mutation and *HOXA5* hypermethylation could constitute early events during breast cancer progression that can cooperate to inactivate the TP53 pathway in more prone to progress DCIS lesions. In addition our pathway-based analysis indicated that TP53 pathway inactivation is an extremely common feature in practically all HG-DCIS lesions and *HOXA5* methylation could be one of the more relevant causative mechanisms. CpGi methylation and expression analysis from the TCGA breast cancer dataset demonstrated that *HOXA5* hypermethylation and downregulation can be also identified and is common in more advanced stages of breast cancer progression (Fig. 5E and F).

Concluding Remarks

Comprehensive characterization of pure HG-DCIS lesions at the genome, transcriptome, and methylome levels allowed us to identify the most relevant changes occurring at a preinvasive stage of breast cancer progression. A comparison of the mutation and chromosomal copy number alteration profiles identified in these pure *in situ* lesions with previous observations reported by the TCGA study on IBC, revealing expected similarities with IBC (e.g., *PIK3CA*, *TP53*, *GATA3*). However, it was striking to observe that practically all known major genomic abnormalities are already present at preinvasive stages at high frequency mutation rates and 83% of lesions displaying very large copy number chromosomal alterations. In only 10% of lesions we did not detect large chromosomal changes or mutations in so-called cancer driver genes. More importantly, at the transcriptional level, pathway-based analysis pointed to TP53 pathway inactivation as extremely common in DCIS regardless of tumor intrinsic subtype. We also observed that deregulation of genes associated with the suppression of the antitumor immune response are hallmarks of a HG-DCIS subgroup (highly proliferative basal-like or ERBB2^{+} tumors), displaying profiles very similar to those found at invasive and metastatic stages. In addition, we identified novel and relevant regulatory circuits significantly deregulated in DCIS that were not previously reported and that involve coding and noncoding transcripts (e.g., *HOTAIR*, *HOTAIRM1*). Our comprehensive catalogue of differentially expressed genes is also consistent with the existence of the most common breast cancer subtypes, but now we show that these important and complex epigenetic changes, such as hypermethylation of *HOXA5* and specific *SOX* genes, are already operating at the *in situ* stage. Our findings clearly indicate that a subgroup of HG-DCIS lesions can be identified displaying more aggressive molecular profiles, more importantly however is that most, if not all, HG-DCIS lesions displayed profiles indistinguishable from IBC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant Support

This study was supported by CPRIT Core Facility Support Grant RP120348 and the MD Anderson's Institutional Tissue Bank, award 2P30CA016672 from the NCI and additionally by grant NIH/NCI R01CA102444 (C.M. Aldaz).

References

1. Kerlikowske K. Epidemiology of ductal carcinoma *in situ*. J Natl Cancer Inst Monogr. 2010; 2010:139–141. [PubMed: 20956818]
2. Page DL, Dupont WD, Rogers LW, Jensen RA, Schuyler PA. Continued local recurrence of carcinoma 15–25 years after a diagnosis of low grade ductal carcinoma *in situ* of the breast treated only by biopsy. Cancer. 1995; 76:1197–1200. [PubMed: 8630897]
3. Allred DC. Ductal carcinoma *in situ*: terminology, classification, and natural history. J Natl Cancer Inst Monogr. 2010; 2010:134–138. [PubMed: 20956817]
4. Fisher ER, Dignam J, Tan-Chiu E, Costantino J, Fisher B, Paik S, et al. Pathologic findings from the National Surgical Adjuvant Breast Project (NSABP) eight-year update of Protocol B-17: intraductal carcinoma. Cancer. 1999; 86:429–438. [PubMed: 10430251]
5. Mokbel K, Cutuli B. Heterogeneity of ductal carcinoma *in situ* and its effects on management. Lancet Oncol. 2006; 7:756–765. [PubMed: 16945771]
6. Bijker N, Meijnen P, Peterse JL, Bogaerts J, Van Hoorebeeck I, Julien JP, et al. Breast-conserving treatment with or without radiotherapy in ductal carcinoma-in-situ: ten-year results of European Organisation for Research and Treatment of Cancer randomized phase III trial 10853–a study by the EORTC Breast Cancer Cooperative Group and EORTC Radiotherapy Group. J Clin Oncol. 2006; 24:3381–3387. [PubMed: 16801628]
7. Solin LJ, Gray R, Baehner FL, Butler SM, Hughes LL, Yoshizawa C, et al. A multigene expression assay to predict local recurrence risk for ductal carcinoma *in situ* of the breast. J Natl Cancer Inst. 2013; 105:701–710. [PubMed: 23641039]
8. Marshall E. Breast cancer. Dare to do less. Science. 2014; 343:1454–1456. [PubMed: 24675949]
9. Lagios MD, Silverstein MJ. Risk of recurrence of ductal carcinoma *in situ* by oncotype Dx technology: some concerns. Cancer. 2014; 120:1085. [PubMed: 24382657]
10. Formenti SC, Arslan AA, Pike MC. Re: Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS. J Natl Cancer Inst. 2011; 103:1723. [PubMed: 21969339]
11. Thorat MA, Parmar V, Nadkarni MS, Badwe RA. Radiation therapy for ductal carcinoma *in situ*: is it really worth it? J Clin Oncol. 2007; 25:461–462. author reply 462. [PubMed: 17264349]
12. Hannemann J, Velds A, Halfwerk JB, Kreike B, Peterse JL, van de Vijver MJ. Classification of ductal carcinoma *in situ* by gene expression profiling. Breast Cancer Res. 2006; 8:R61. [PubMed: 17069663]
13. Vincent-Salomon A, Lucchesi C, Gruel N, Raynal V, Pierron G, Goudefroye R, et al. Integrated genomic and transcriptomic analysis of ductal carcinoma *in situ* of the breast. Clin Cancer Res. 2008; 14:1956–1965. [PubMed: 18381933]
14. Muggerud AA, Hallett M, Johnsen H, Kleivi K, Zhou W, Tahmasebpoor S, et al. Molecular diversity in ductal carcinoma *in situ* (DCIS) and early invasive breast cancer. Mol Oncol. 2010; 4:357–368. [PubMed: 20663721]
15. Clark SE, Warwick J, Carpenter R, Bowen RL, Duffy SW, Jones JL. Molecular subtyping of DCIS: heterogeneity of breast cancer reflected in pre-invasive disease. Br J Cancer. 2011; 104:120–127. [PubMed: 21139586]

16. Done SJ, Eskandarian S, Bull S, Redston M, Andrulis IL. p53 missense mutations in microdissected high-grade ductal carcinoma *in situ* of the breast. J Natl Cancer Inst. 2001; 93:700–704. [PubMed: 11333292]
17. Miron A, Varadi M, Carrasco D, Li H, Luongo L, Kim HJ, et al. PIK3CA mutations in *in situ* and invasive breast carcinomas. Cancer Res. 2010; 70:5674–5678. [PubMed: 20551053]
18. Polyak K. Molecular markers for the diagnosis and management of ductal carcinoma *in situ*. J Natl Cancer Inst Monogr. 2010; 2010:210–213. [PubMed: 20956832]
19. Brenner AJ, Stampfer MR, Aldaz CM. Increased p16 expression with first senescence arrest in human mammary epithelial cells and extended growth capacity with p16 inactivation. Oncogene. 1998; 17:199–205. [PubMed: 9674704]
20. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–498. [PubMed: 21478889]
21. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013; 31:213–219. [PubMed: 23396013]
22. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]
23. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012; 28:423–425. [PubMed: 22155870]
24. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011; 27:1571–1572. [PubMed: 21493656]
25. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–1111. [PubMed: 19289445]
26. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]
27. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics. 2010; 26:i237–i245. [PubMed: 20529912]
28. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 2003; 34:374–378. [PubMed: 12613259]
29. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009; 27:1160–1167. [PubMed: 19204204]
30. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4:2612. [PubMed: 24113773]
31. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]
32. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–421. [PubMed: 23945592]
33. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012; 486:400–404. [PubMed: 22722201]
34. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012; 486:346–352. [PubMed: 22522925]
35. Hwang ES, DeVries S, Chew KL, Moore DH II, Kerlikowske K, Thor A, et al. Patterns of chromosomal alterations in breast ductal carcinoma *in situ*. Clin Cancer Res. 2004; 10:5160–5167. [PubMed: 15297420]

36. DeNardo DG, Coussens LM. Inflammation and breast cancer. Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression. Breast Cancer Res. 2007; 9:212. [PubMed: 17705880]
37. Kristensen VN, Vaske CJ, Ursini-Siegel J, Van Loo P, Nordgard SH, Sachidanandam R, et al. Integrated molecular profiles of invasive breast tumors and ductal carcinoma *in situ* (DCIS) reveal differential vascular and interleukin signaling. Proc Natl Acad Sci U S A. 2012; 109:2802–2807. [PubMed: 21908711]
38. Su X, Malouf GG, Chen Y, Zhang J, Yao H, Valero V, et al. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. Oncotarget. 2014; 5:9864–9876. [PubMed: 25296969]
39. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28:495–501. [PubMed: 20436461]
40. Orom UA, Shiekhattar R. Noncoding RNAs and enhancers: complications of a long-distance relationship. Trends Genet. 2011; 27:433–439. [PubMed: 21831473]
41. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010; 464:1071–1076. [PubMed: 20393566]
42. Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M, et al. A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. Blood. 2009; 113:2526–2534. [PubMed: 19144990]
43. Persson H, Kvist A, Rego N, Staaf J, Vallon-Christersson J, Luts L, et al. Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. Cancer Res. 2011; 71:78–86. [PubMed: 21199797]
44. Newie I, Sokilde R, Persson H, Grabau D, Rego N, Kvist A, et al. The HER2-encoded miR-4728-3p regulates ESR1 through a non-canonical internal seed interaction. PLoS ONE. 2014; 9:e97200. [PubMed: 24828673]
45. Xiong J, Yu D, Wei N, Fu H, Cai T, Huang Y, et al. An estrogen receptor alpha suppressor, microRNA-22, is downregulated in estrogen receptor alpha-positive human breast cancer cell lines and clinical samples. FEBS J. 2010; 277:1684–1694. [PubMed: 20180843]
46. Xu D, Takeshita F, Hino Y, Fukunaga S, Kudo Y, Tamaki A, et al. miR-22 represses cancer progression by inducing cellular senescence. J Cell Biol. 2011; 193:409–424. [PubMed: 21502362]
47. Yu Z, Wang C, Wang M, Li Z, Casimiro MC, Liu M, et al. A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation. J Cell Biol. 2008; 182:509–517. [PubMed: 18695042]
48. Yu Z, Xu Z, Disante G, Wright J, Wang M, Li Y, et al. miR-17/20 sensitization of breast cancer cells to chemotherapy-induced apoptosis requires Akt1. Oncotarget. 2014; 5:1083–1090. [PubMed: 24658544]
49. Tong X, Li L, Li X, Heng L, Zhong L, Su X, et al. SOX10, a novel HMG-box-containing tumor suppressor, inhibits growth and metastasis of digestive cancers by suppressing the Wnt/beta-catenin pathway. Oncotarget. 2014; 5:10571–10583. [PubMed: 25301735]
50. Fu DY, Wang ZM, Li C, Wang BL, Shen ZZ, Huang W, et al. Sox17, the canonical Wnt antagonist, is epigenetically inactivated by promoter methylation in human breast cancer. Breast Cancer Res Treat. 2010; 119:601–612. [PubMed: 19301122]
51. Thu KL, Radulovich N, Becker-Santos DD, Pikor LA, Pusic A, Lockwood WW, et al. SOX15 is a candidate tumor suppressor in pancreatic cancer with a potential role in Wnt/beta-catenin signaling. Oncogene. 2014; 33:279–288. [PubMed: 23318427]
52. Cimino-Mathews A, Subhawong AP, Elwood H, Warzecha HN, Sharma R, Park BH, et al. Neural crest transcription factor Sox10 is preferentially expressed in triple-negative and metaplastic breast carcinomas. Hum Pathol. 2013; 44:959–965. [PubMed: 23260325]
53. Raman V, Martensen SA, Reisman D, Evron E, Odenwald WF, Jaffee E, et al. Compromised HOXA5 function can limit p53 expression in human breast tumours. Nature. 2000; 405:974–978. [PubMed: 10879542]

54. Raman V, Tamori A, Vali M, Zeller K, Korz D, Sukumar S. HOXA5 regulates expression of the progesterone receptor. J Biol Chem. 2000; 275:26551–26555. [PubMed: 10875927]

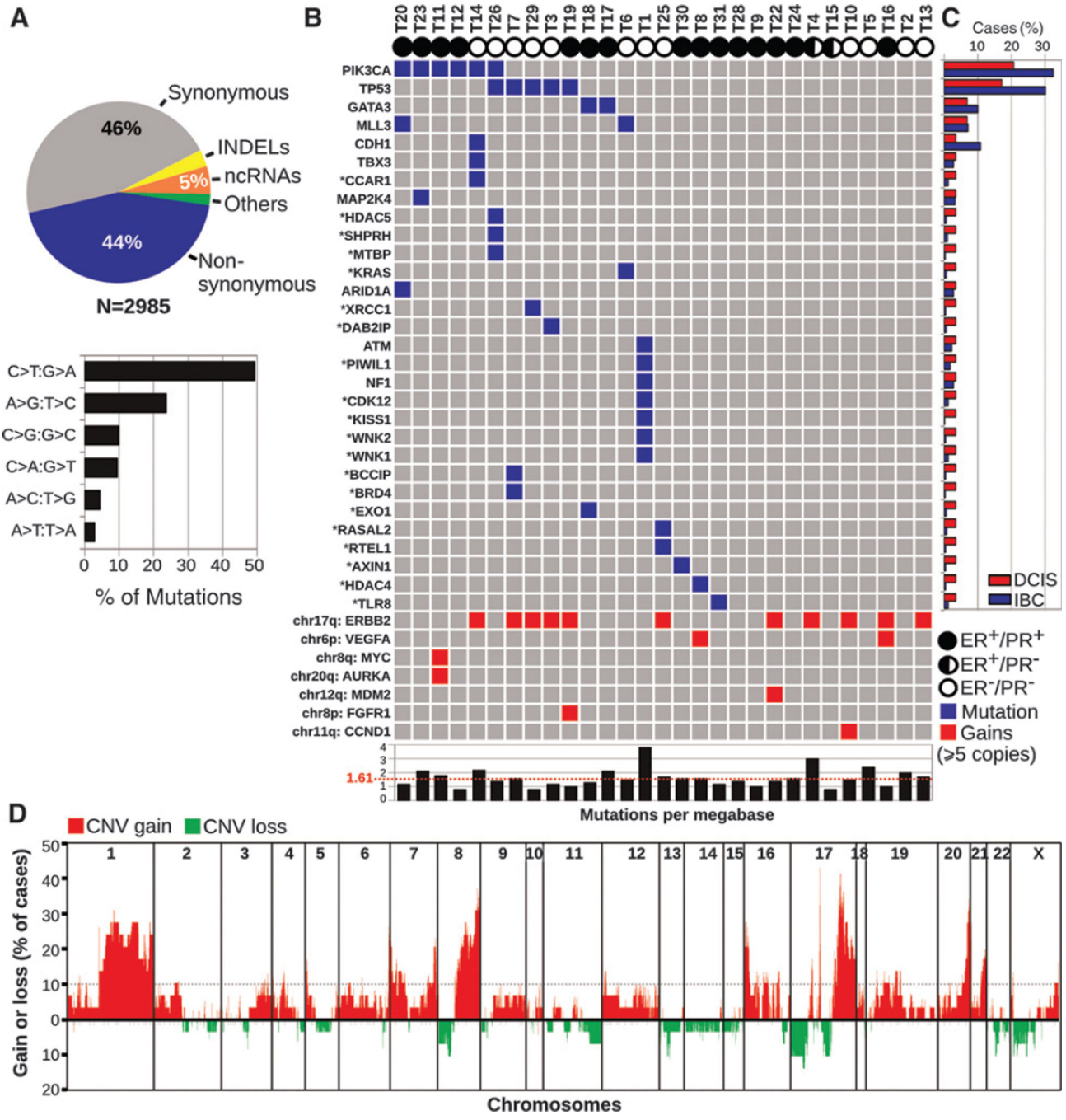

Figure 1.

Mutations and copy number alterations in DCIS samples. A, types of single-base substitutions and InDels identified by Exome-Seq analysis. B, driver and co-driver mutations in DCIS samples. Each column represents one DCIS case and each row represents the number of mutations in each gene per tumor. Blue squares, + mutation. Genes shown with an asterisk are rarely mutated in IBC (≤2%); however, they are noted as possible co-drivers events, as they are found either in the SANGER Cancer Gene Census database or have significant potentially carcinogenic biological functions. C, comparative frequency of

mutations of driver/co-driver mutations in pure DCIS (red bars) and the TCGA database on IBC (blue bars). D, CNVs among DCIS samples. Chromosomal gains (red, up) and losses (green, down) as percentage of occurrence in the respective region are indicated.

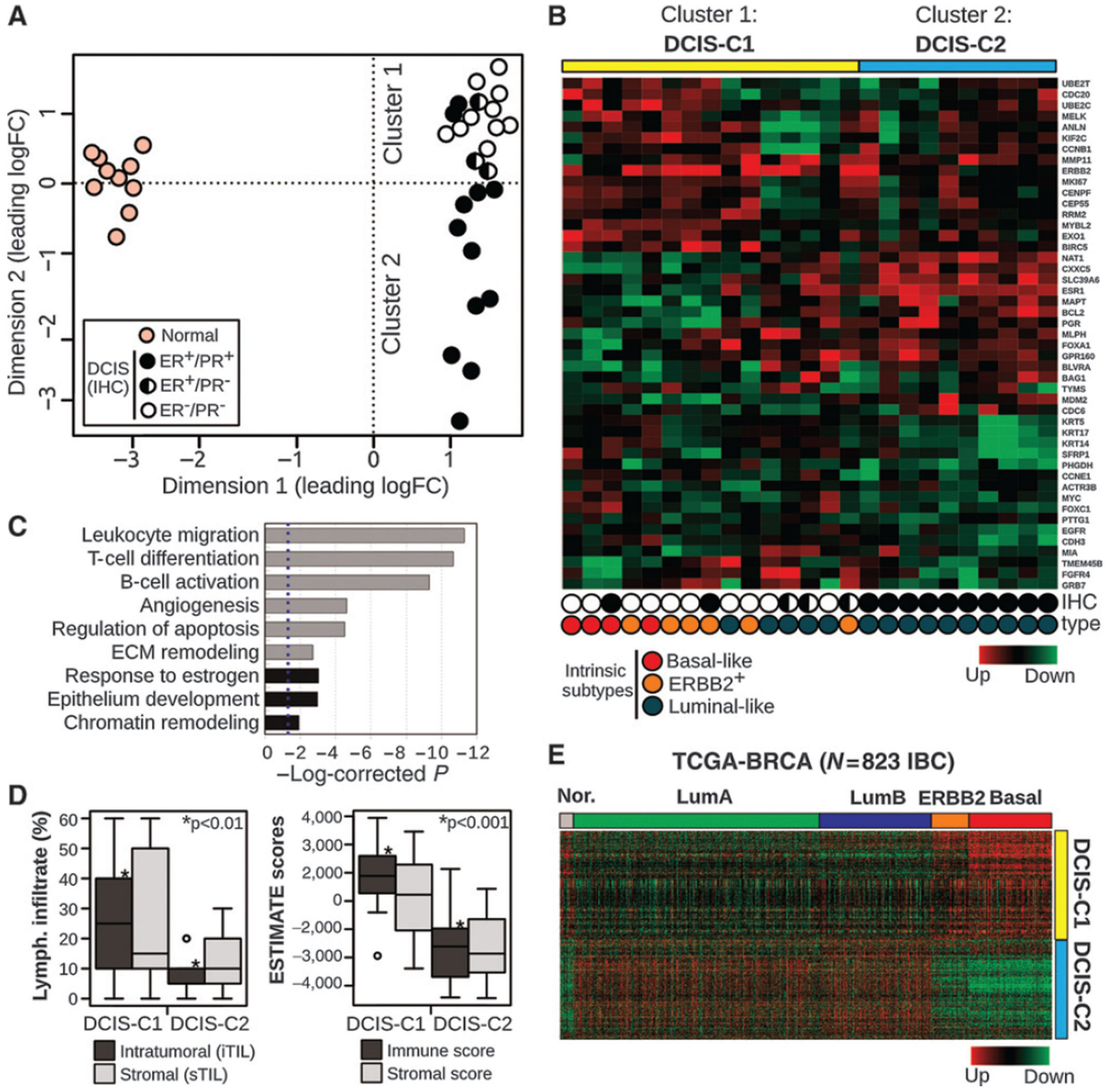

Figure 2.

Transcriptome analysis of normal and DCIS samples. A, multidimensional scaling plot showing the distance of each sample from each other determined by their leading logFC. The leading logFC is a distance metric that represents the average (root mean square) of the largest absolute logFC between each pair of samples. B, prediction of DCIS intrinsic subtypes based on the PAM50 gene model based on RNA-Seq profiles. All normal samples (not included in the heatmap) were classified as the normal-like subtype. C, functional enrichment analysis of the differentially expressed genes between DCIS clusters. D, box and whisker plots display increased iTIL (left) and immune scores determined by the

ESTIMATE (30) algorithm (right) for the DCIS-C1 group compared with DCIS-C2 group and normal samples. E, heatmap of DCIS-C1 and DCIS-C2 differentially expressed transcripts among the IBC subtypes obtained from the TCGA breast cancer database.

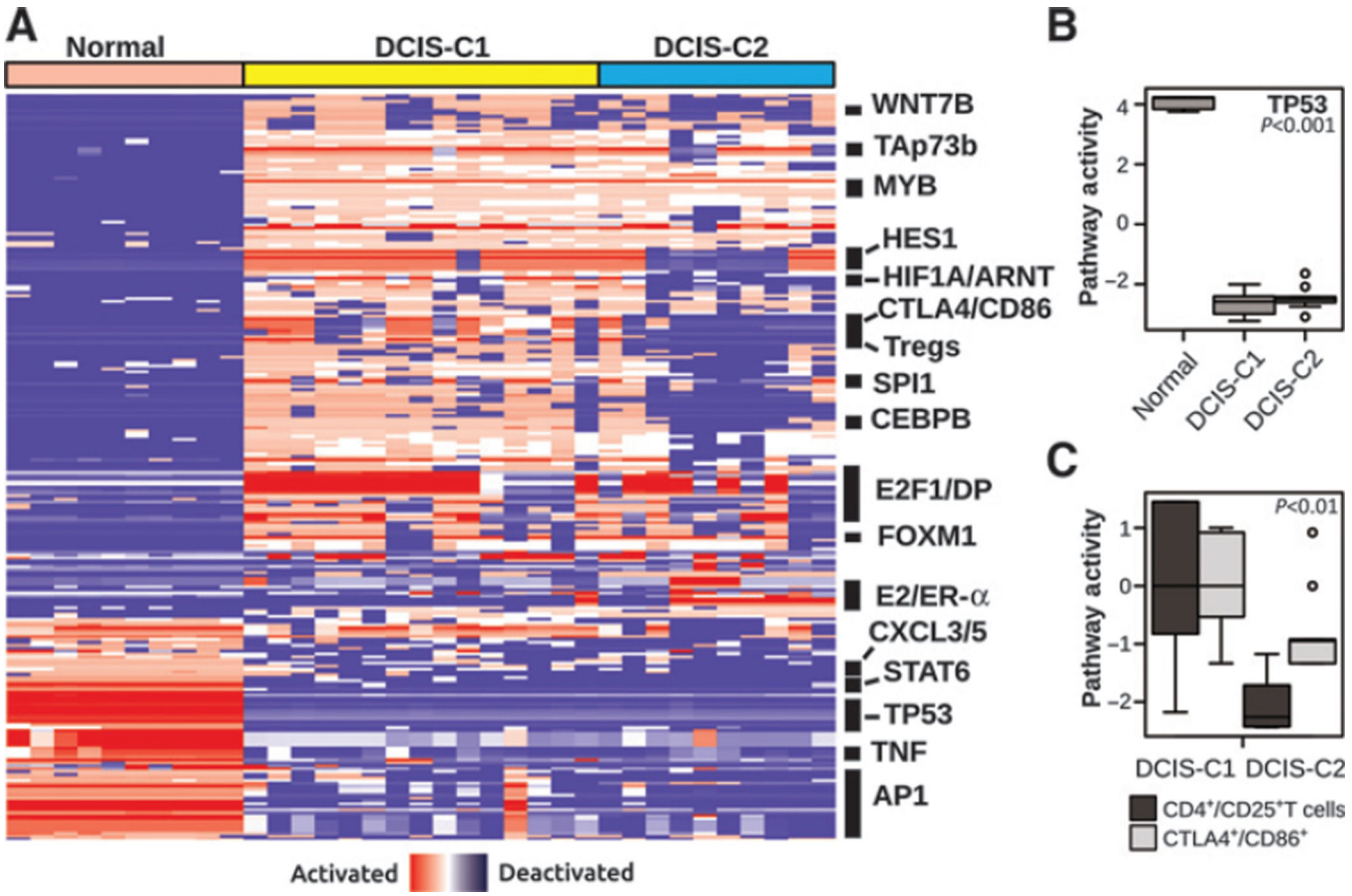

Figure 3.

Pathway activities analysis in DCIS samples versus normal mammary epithelium. A, PARADIGM inferences of the most variable integrated pathway activities using the normalized gene expression profiles of the deregulated transcripts between normal and DCIS samples. B, TP53 pathway activity was found systematically deactivated in both DCIS groups compared with normal epithelium ($P < 0.001$). C, PARADIGM analysis identified increased activities of Treg cells ($CD4^+/CD25^+/FOXP3^+$) and $CTLA4^+/CD86^+$ complexes in DCIS-C1 compared with DCIS-C2 samples ($P < 0.01$).

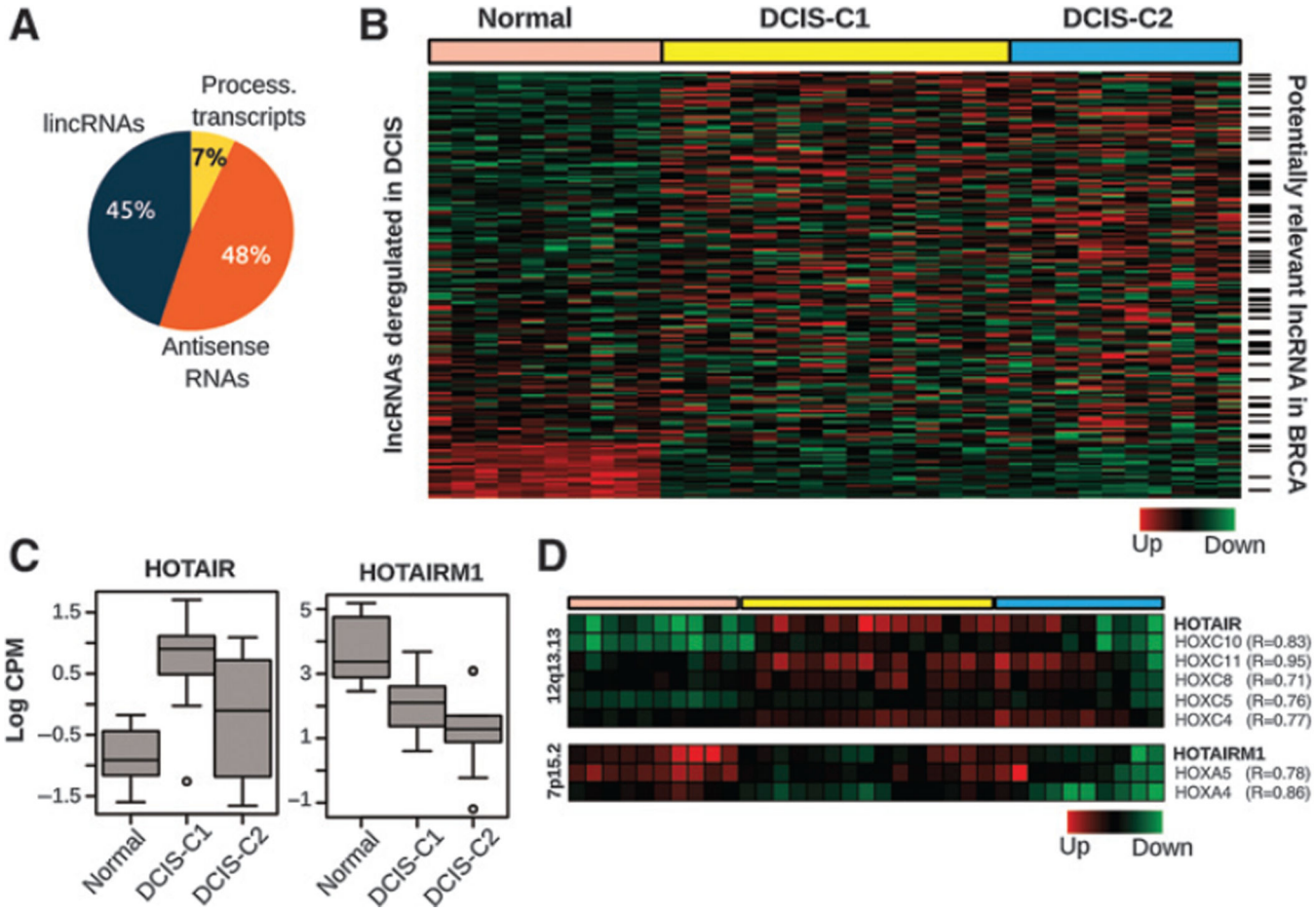

Figure 4.

Gene expression profile of deregulated lncRNAs. A, pie chart displaying the types of lncRNAs differentially expressed between normal and DCIS samples. B, heatmap of deregulated lncRNAs where black lines on the right indicate potentially relevant transcripts in IBC as was determined by Su and colleagues (38). C, box and whisker plots representing *HOTAIR* and *HOTAIRM1* expression levels among normal and DCIS samples. D, transcriptomic coexpression analysis using *HOTAIR* and *HOTAIRM1* as templates identified significant positive correlations with the expression of their neighboring *HOXC* (located at chr12q13.13) and *HOXA* (located at chr7q15.2) gene clusters, respectively ($r > 0.7$ and $P < 0.0001$).

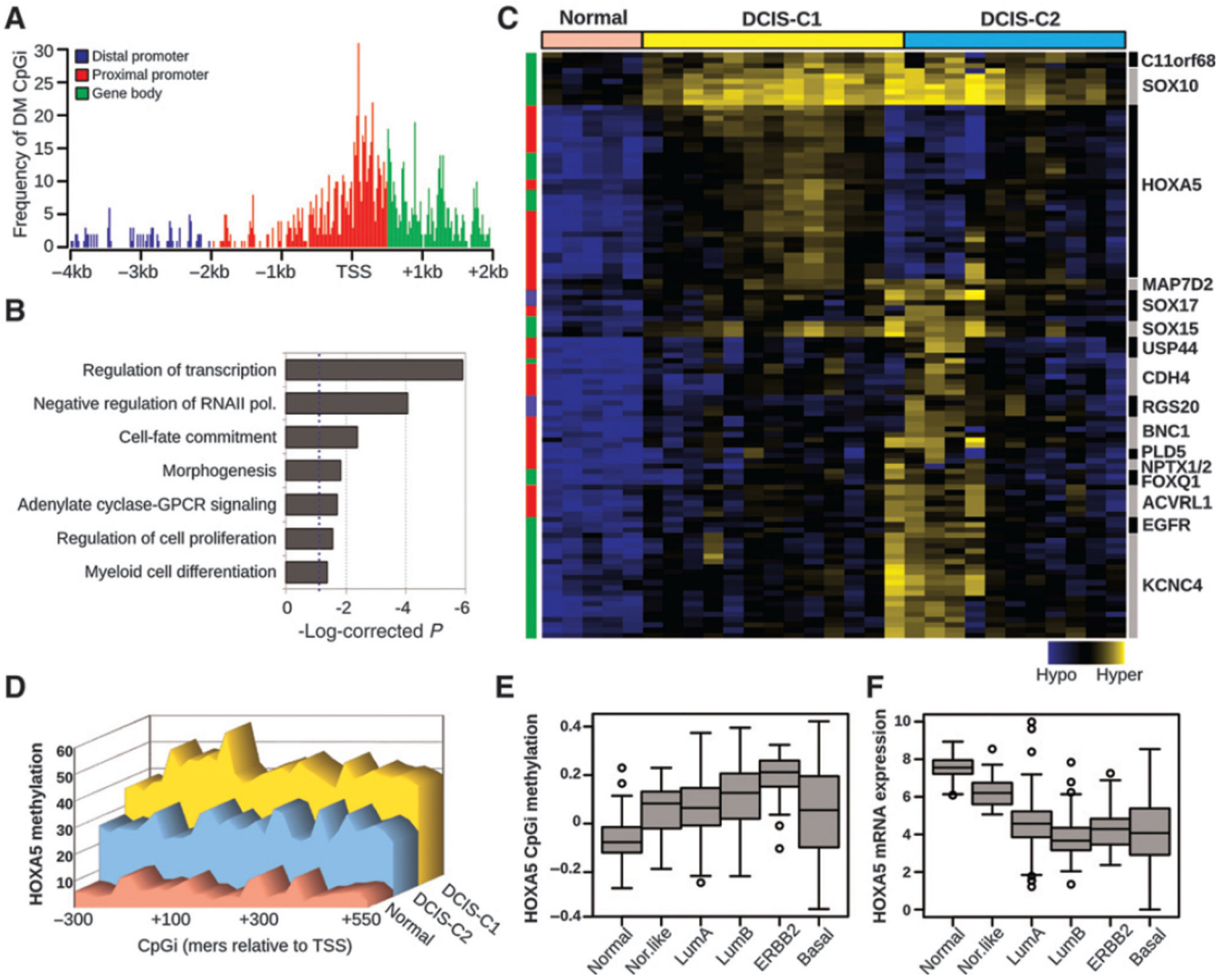

Figure 5.

Analysis of differentially methylated CpGi among normal and DCIS samples determined by RRBS. A, distribution of 1,103 differentially methylated CpGi regions between normal and DCIS samples mapping at distal (−4 kb to −2 kb relative to TSS), proximal promoter regions (−2 kb to +500 bp), or the gene body (+500 to +2 kb). B, functional enrichment analysis of differentially methylated CpGi regions. C, heatmap of CpGi hypermethylated and downregulated genes in DCIS samples. D, *HOXA5* CpGi methylation profile in normal and DCIS groups in a promoter region spanning from −285 to 560 bp relative to TSS. *HOXA5* methylation (E) and expression profiles (F) among normal and IBC from the TCGA breast cancer dataset.