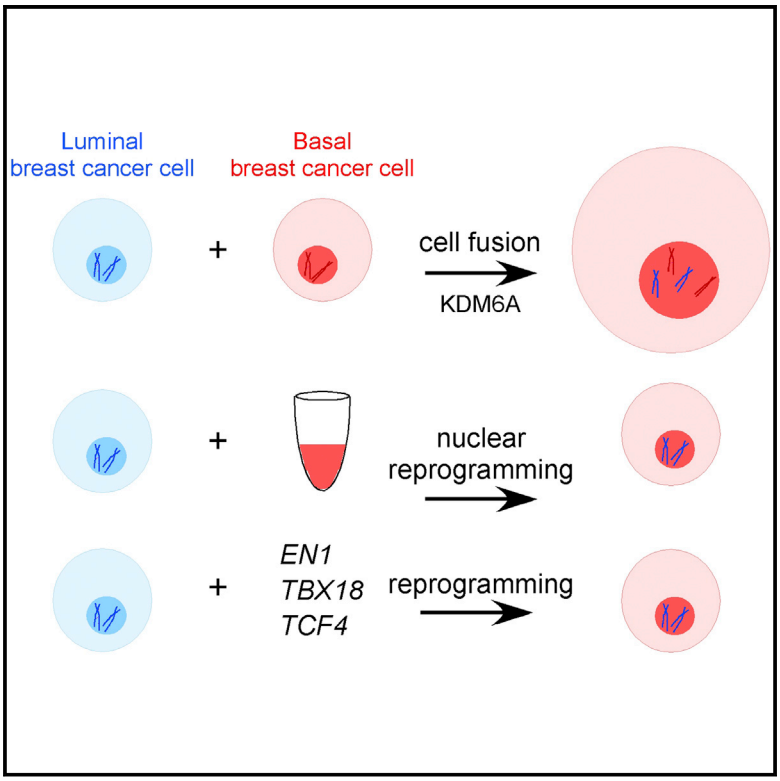
Article

Cell Reports

Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer

Graphical Abstract



Authors

Ying Su, Ashim Subedee, Noga Bloushtain-Qimron, ..., Alexander Gimelbrant, Reo Maruyama, Kornelia Polyak

Correspondence

reomaru@sapmed.ac.jp (R.M.), kornelia_polyak@dfci.harvard.edu (K.P.)

In Brief

Luminal and basal-like breast tumors are clinically distinct. Using somatic cell fusions, Su et al. demonstrate the general dominance of the basal-like phenotype. Epigenomic profiling of parental luminal and basal-like cells and heterofusions define common luminal programs but a high degree of heterogeneity in basal-like cancer cells.

Highlights

- The basal-like trait is generally dominant and defined by epigenetic mechanisms
- Common core epigenetic programs in luminal tumors are defined by *FOXA1* super-enhancers
- Basal tumors show repression of luminal factors and high epigenetic heterogeneity
- Nuclear extracts or transcription factors can reprogram breast cancer cells

Accession Numbers

GSE38548

Su et al., 2015, Cell Reports *11*, 1549–1563
June 16, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.celrep.2015.05.011

Cell Reports
Article

Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer

Ying Su,[1,4,5,14,15] Ashim Subedee,[1,7,14] Noga Bloushtain-Qimron,[1,16] Virginia Savova,[2,6] Marcin Krzystanek,[8] Lewyn Li,[3] Andriy Marusyk,[1,4,5] Doris P. Tabassum,[1,7] Alexander Zak,[2,17] Mary Jo Flacker,[9] Mei Li,[1,18] Jessica J. Lin,[1,4,5] Saraswati Sukumar,[9] Hiromu Suzuki,[10] Henry Long,[3] Zoltan Szallasi,[8,11] Alexander Gimelbrant,[2,6] Reo Maruyama,[1,4,5,10,*] and Kornelia Polyak[1,4,5,7,12,13,*]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute Boston, MA 02215, USA
[2]Department of Cancer Biology, Dana-Farber Cancer Institute Boston, MA 02215, USA
[3]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute Boston, MA 02215, USA
[4]Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
[5]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[6]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[7]BBS Program, Harvard Medical School, Boston, MA 02115, USA
[8]Department of Systems Biology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
[9]Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA
[10]Department of Molecular Biology, Sapporo Medical University School of Medicine, Sapporo 060-8556, Japan
[11]Children's Hospital, Boston, MA 02115, USA
[12]Harvard Stem Cell Institute, Cambridge, MA 02138, USA
[13]The Broad Institute, Cambridge, MA 02138, USA
[14]Co-first author
[15]Present address: Infinity Pharmaceuticals, Inc., Cambridge, MA 02139, USA
[16]Present address: EMEA Site Intelligence and Activation, Tel Aviv 69710, Israel
[17]Present address: Pfizer Research Business Technologies, Cambridge, MA 02140, USA
[18]Present address: Corning China, Pudong Shanghai 201206, China
*Correspondence: reomaru@sapmed.ac.jp (R.M.), kornelia_polyak@dfci.harvard.edu (K.P.)
http://dx.doi.org/10.1016/j.celrep.2015.05.011

SUMMARY

Basal-like and luminal breast tumors have distinct clinical behavior and molecular profiles, yet the underlying mechanisms are poorly defined. To interrogate processes that determine these distinct phenotypes and their inheritance pattern, we generated somatic cell fusions and performed integrated genetic and epigenetic (DNA methylation and chromatin) profiling. We found that the basal-like trait is generally dominant and is largely defined by epigenetic repression of luminal transcription factors. Definition of super-enhancers highlighted a core program common in luminal cells but a high degree of heterogeneity in basal-like breast cancers that correlates with clinical outcome. We also found that protein extracts of basal-like cells are sufficient to induce a luminal-to-basal phenotypic switch, implying a trigger of basal-like autoregulatory circuits. We determined that *KDM6A* might be required for luminal-basal fusions, and we identified *EN1*, *TBX18*, and *TCF4* as candidate transcriptional regulators of the luminal-to-basal switch. Our findings highlight the remarkable epigenetic plasticity of breast cancer cells.

INTRODUCTION

Breast tumors are highly heterogeneous and are classified into ER^+, $HER2^+$, and $ER^-PR^-HER2^-$ (triple-negative breast cancer [TNBC]) tumors based on the expression of estrogen and progesterone receptors (ER and PR) and HER2 or into luminal and basal-like subtypes based on differentiation/epigenetic states (Prat and Perou, 2011). ER^+ and $HER2^+$ tumors typically have luminal features, whereas TNBCs are commonly basal-like, have high propensity to metastasize to distant sites, and currently lack targeted therapies, leading to generally worse clinical outcome (Vaz-Luis et al., 2014). Breast cancer genome sequencing studies failed to identify novel recurrent mutations in basal-like breast tumors (Curtis et al., 2012; Shah et al., 2012; TCGA, 2012), necessitating alternative approaches for targeted therapeutics development (Lehmann et al., 2011). Targeting key transcriptional regulators of lineage dependency and cellular phenotypes has been successful in a number of tumor types, including ER^+ breast cancer, but such approach has not been explored in basal-like breast tumors, in part due to our limited knowledge of these regulators in this cancer type.

Currently, prevailing hypotheses explaining breast tumor subtypes are the distinct cell-of-origin or tumor-subtype-specific transforming events models (Polyak, 2007; Visvader, 2011). According to the cell-of-origin hypothesis, luminal tumors may originate from luminal progenitors, whereas basal-like tumors originate from breast epithelial stem cells. However, based on

relatedness of gene expression profiles, luminal progenitors were also proposed as the cell of origin for basal-like breast cancers in *BRCA1* germline mutation carriers (Lim et al., 2009), which is supported by functional studies in experimental model systems (Liu et al., 2008; Molyneux et al., 2010).

We and others have previously described high degree of intratumor heterogeneity for cellular phenotypes, including stem cell-like and more differentiated luminal features in breast tumors (Al-Hajj et al., 2003; Park et al., 2010a, 2010b; Shipitsin et al., 2007). Interestingly, even in luminal ER^+PR^+ breast tumors, a fraction of cancer cells are ER^-PR^- and $CD44^+CD24^-$ with stem cell-like features, but cells with a luminal phenotype are rarely observed in basal-like breast cancers. Based on these observations, we hypothesized that basal-like breast tumors may contain a factor(s) (i.e., genetic or epigenetic alteration) that blocks luminal epithelial differentiation. Here, we used combined somatic cell genetic and integrative epigenomics approaches to investigate molecular mechanisms that define basal-like and luminal breast cancer cell phenotypes and identify key regulators of these processes. We found that the basal-like phenotype is usually dominant, but, apart from displaying extinguished luminal features, it is not defined by a specific epigenetic state and displays a high degree of heterogeneity.

RESULTS

Somatic Cell Fusions of Luminal and Basal-like Breast Cancer Cells

To investigate the role of hereditary factors in breast tumor subtypes, we generated somatic cell hybrids by the fusion of basal-like and luminal breast cancer cell lines resistant to puromycin and G418, respectively. Homofusions (fusion of the same cell line to itself) and heterofusions of different basal-like cell lines were also generated as controls. Pooled populations of hybrid cells resistant to both puromycin and G418 were used for subsequent experiments. We tested combinations of multiple basal-like and luminal breast cancer cell lines, but not all fusions yielded viable and reasonably stable hybrids (Table S1). To prove the hybrid nature of the cells, first we performed fluorescence-activated cell sorting (FACS) with propidium iodide and confirmed that the nuclear DNA content of the fusions was the sum of the parental cells (Figures 1A and S1A). Fusions judged to be successful by FACS were subjected to STR (short tandem repeat) polymorphism and SKY (spectral karyotyping) analyses to confirm the presence of chromosomes from both parents (Figures 1B and S1B; Table S2). We also performed whole-genome sequencing of selected parental cell lines and their heterofusions, which confirmed the presence of both genomes (Figures 1C and S1C) and mutant alleles of known cancer-driving (CAN) genes (Sjöblom et al., 2006), including *TP53*, *BRCA1*, *GATA3*, and *PIK3CA* (Table S3), in the heterofusions. Most heterofusions contained all CAN genes present in both parental cells, implying lack of selection against specific mutations in the viable fusions.

Fusion Cell Phenotypes

We began characterizing the fusions by assessing their morphology and the expression of known luminal (e.g., E-cadherin,) and basal (e.g., vimentin) cell-specific markers. All luminal-basal heterofusions derived from the SUM159PT and CAL51 basal-like breast cancer cell lines displayed cellular morphology, vimentin, E-cadherin (Figures 2A and S2A), CD24, CD44, ERBB3, and ITGA5 (Figures S2B and S2C) patterns more similar to those of parental basal-like cells. In contrast, the MDA-MB-231/MCF7 heterofusion showed higher similarity to parental MCF7 luminal cells with respect to morphology and the expression of these markers (Figures 2A, S2B, and S2C). Heterofusion of Hs578T basal and T-47D luminal cell lines also displayed a basal phenotype, but these cells rapidly lost the T-47D cell genome and thus were excluded from further analyses. These results are largely in agreement with two prior studies describing universal loss of E-cadherin expression in luminal-basal fusions of breast epithelial cells (Hajra et al., 1999; MacDougall and Matrisian, 2000).

Next, we evaluated the functional properties of the cells using assays that distinguish basal-like and luminal breast cancer cells such as cell migration, invasion, sensitivity to selected compounds (Marotta et al., 2011), and ability to form tumors (Figures 2B–2E). Compared to parental luminal cells, all luminal-basal heterofusions were more migratory, invasive, and tumorigenic, but they were still distinguishable from parental basal-like cells. The MDA-MB-231/MCF7 fusion was more similar to luminal MCF7 cells in migration and invasion assays (Figure 2B) but displayed an intermediate phenotype in xenograft studies (Figures 2D–2E). Although sensitivity to compounds that most differentially affected luminal and basal-like parental cells (e.g., Taxol) clustered the cells into luminal and basal groups, the sensitivity profiles of most luminal-basal heterofusions were still distinct from both parents (Figure 2C).

Molecular Profiles of Fusion Cells: Gene Expression Patterns

To investigate molecular mechanisms that define the phenotypes of luminal-basal heterofusions, we first analyzed the global gene expression profiles of each parental and fusion cell line. Using PCA based on all transcripts detected, we observed clear luminal and basal clusters, with the CAL51 cell line being more unique. Some of the parental luminal cells were ER^+ and/or $HER2^+$ (Table S1), but this did not appear to influence the molecular classification of the derived fusions. The gene expression profiles of all six SUM159-derived luminal-basal heterofusions showed high similarity to SUM159 cells, whereas the CAL51/MCF7 and MDA-MB-231/MCF7 fusions were intermediate between their parental lines but still closer to luminal cells, suggesting that SUM159 cells have the most dominant basal-like phenotype (Figure 3A). We selected the top 72 genes that were consistently and significantly differentially expressed (log2ratio > 4 and p value < 0.01, t test) between luminal and basal parents (Figure 3B; Table S4). The expression of these 72 genes clustered all cell lines into luminal and basal groups, with three heterofusions (CAL51/MCF7, MDA-MB-231/MCF7, and SUM159/21NT) showing higher relatedness to the luminal than the basal parent. More in-depth analysis of the expression pattern of these 72 genes revealed that the expression of luminal transcription factors (TFs) *FOXA1* and *SPDEF* was preserved in fusions with luminal features (Figure 3B; Table S4). Similarly, the consistent expression of *ETS1*, *AXL*, *CSF1*, and targets

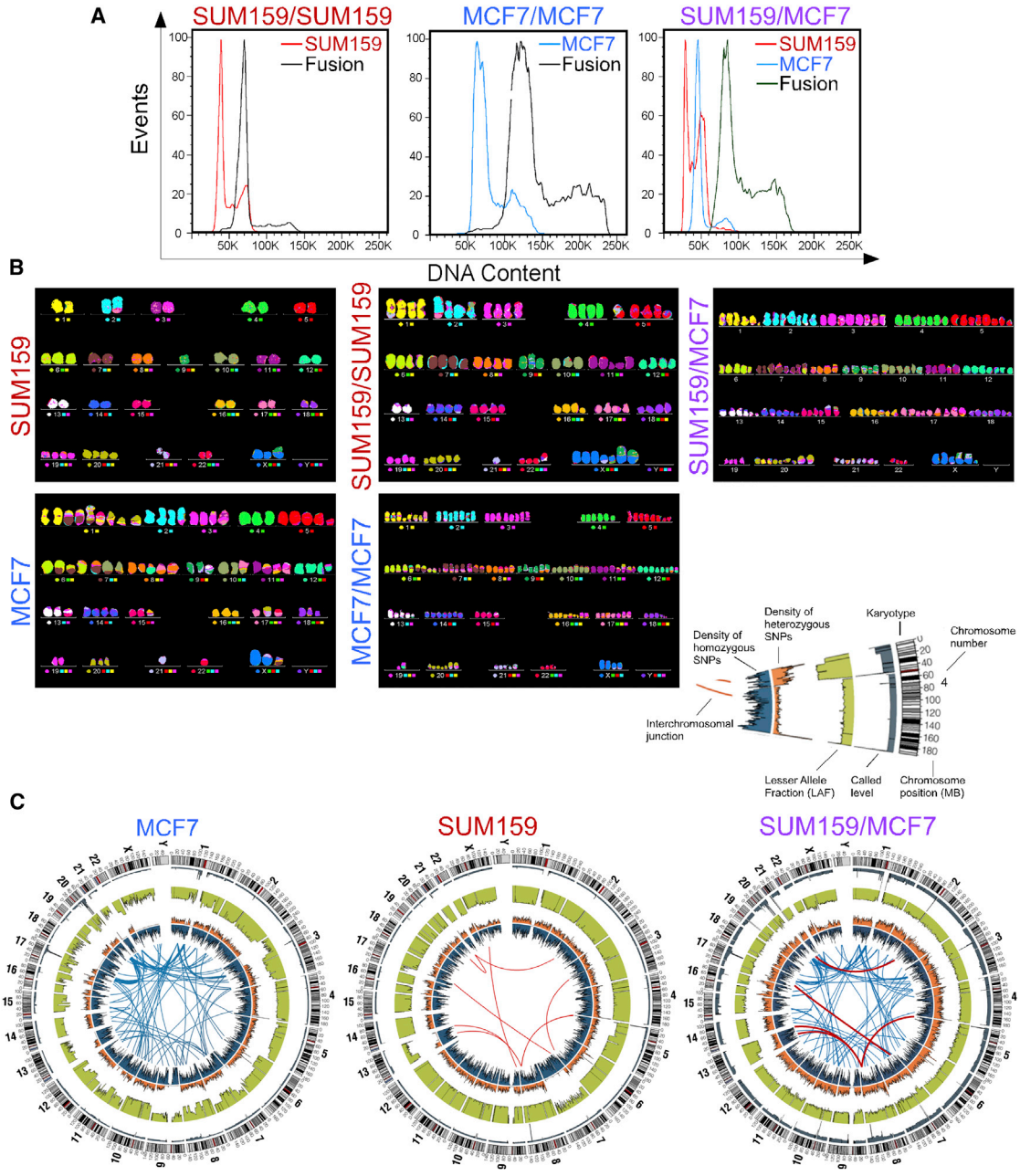


Figure 1. Confirmation and Phenotype of Fusions

(A) Cellular DNA content of basal SUM159 (red) and luminal MCF7 (blue) cell lines and their homo- (black) and heterofusions (dark green) based on FACS analysis of cells stained with propidium iodide.

(B) Representative SKY (spectral karyotyping) analysis of parental SUM159 and MCF7 cell lines and their homo-/heterofusions.

(C) Circos plots depicting genomic alterations in parental SUM159 and MCF7 cell lines and in their heterofusion.

See also Figure S1 and Tables S1, S2, and S3.

of the transforming growth factor β (TGF-β) signaling pathway (e.g., *FN1*) in all fusions with basal-like phenotype suggests their potential roles in basal-like cells. The expression pattern of these 72 genes also classified primary human breast tumors into luminal and basal-like subtypes, supporting their physiologic relevance (data not shown).

Genetic Inheritance Patterns of Heterofusions

Next, we investigated the potential role of genetic factors as determinants of fusion phenotypes by analyzing allelic inheritance based on Affymetrix SNP6.0 array hybridization of DNA from parental and fusion cells. To assist in genotyping of highly aneuploid and genetically complex samples such as fusions, we

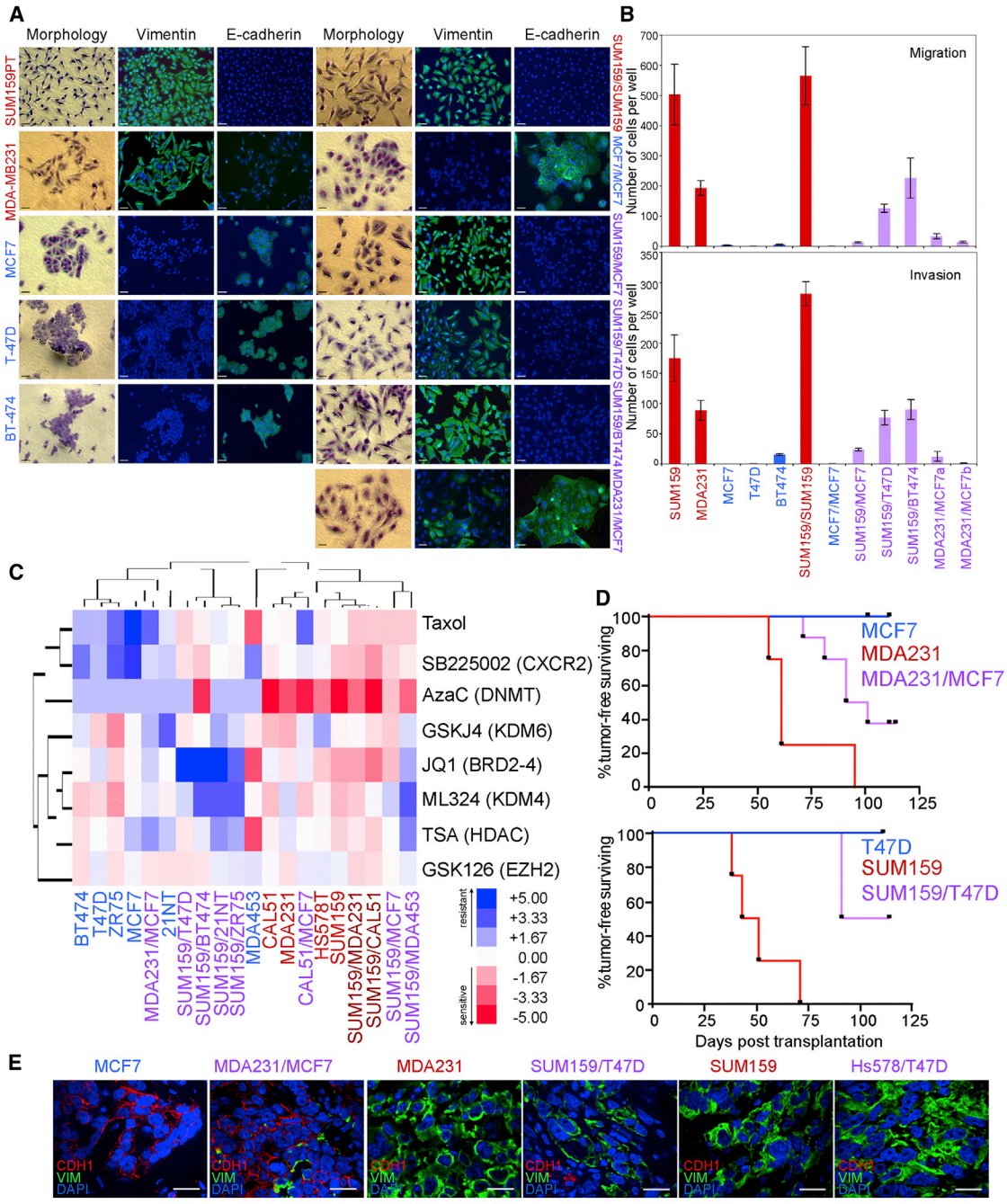


Figure 2. Functional Properties of the Parental Cell Lines and Their Fusions

Colors indicate parental basal (red) and luminal (blue) cell lines or their homofusions and luminal-basal heterofusions (purple) in all panels.

(A) Cellular morphology and expression of basal vimentin and luminal E-cadherin markers in parental basal and luminal cell lines and in their homo- and heterofusions. Scale bar corresponds to 25 μm.

(B) Migration and invasion of parental basal (red) and luminal (blue) cell lines and in their homo- and hetero-fusions. y axis indicates the number of cells per well. Data are shown as mean ± SD of a representative experiment performed in triplicate. The experiment was repeated three times with similar results.

(C) Heatmap depicting cellular viability after 5 days of treatment with the indicated compounds. Color scale indicates higher sensitivity (red) and relative resistance (blue).

(D) Kaplan-Meier plots depicting the tumor-free survival of mice after transplantation with the indicated breast cancer cells.

(E) Immunofluorescence analysis of xenografts derived from parental and fusion cell lines for luminal (CDH1-E-cadherin) and basal (VIM -vimentin) markers. Scale bar corresponds to 25 μm.

See also Figure S2.

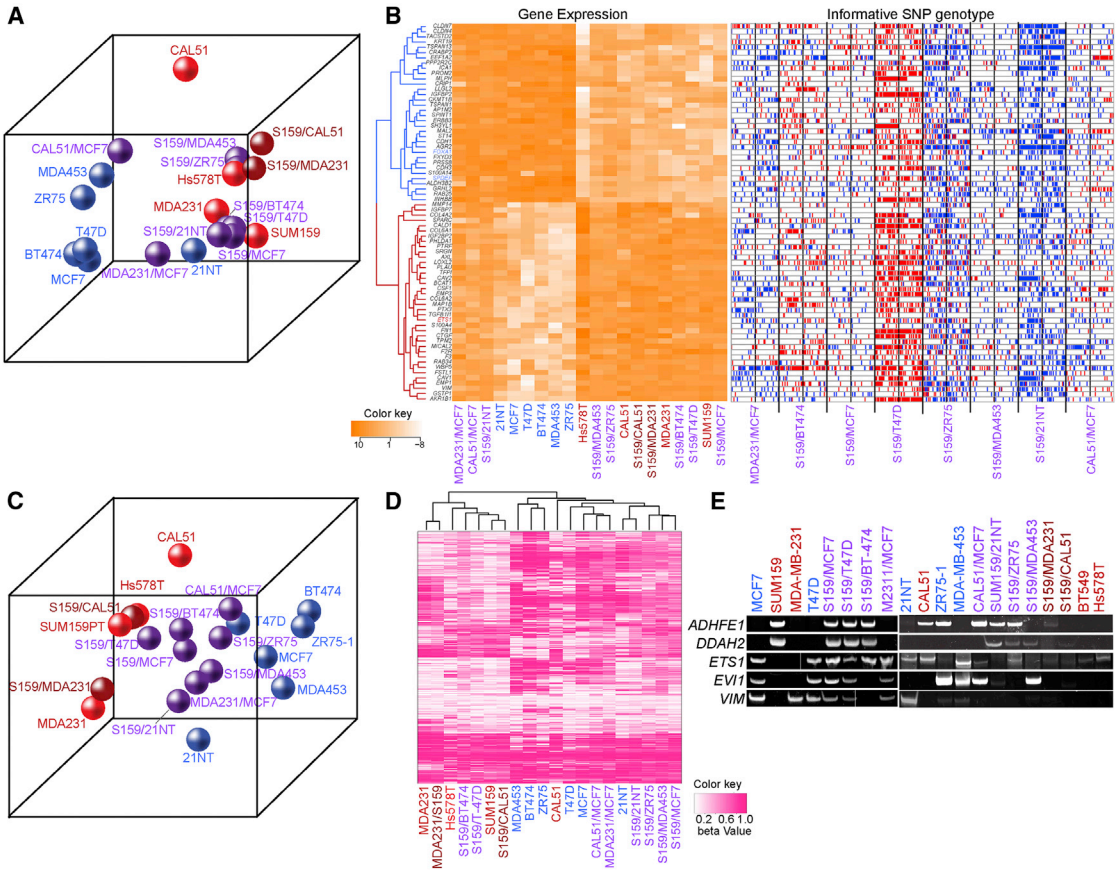


Figure 3. SNP Inheritance, Gene Expression, and DNA Methylation Patterns in Parental Cell Lines and Fusions

Colors indicate parental basal (red) and luminal (blue) cell lines and their luminal basal (purple) and basal-basal (dark red) heterofusions in all panels.

(A) PCA plot depicting the relatedness of cells based on gene expression profiles.

(B) Most significantly differentially expressed genes and informative SNPs within a distance of 10 Mb from their transcription start site in each of the fusions. The colors red and blue in the dendrogram indicate clusters of genes with strong association with basal and luminal phenotype, respectively, whereas genes marked in blue and red are transcription factors with known roles in luminal and basal cells, respectively. Blue and red colors indicate SNPs inherited from luminal and basal parental cell lines, respectively. Scale indicates normalized expression values.

(C) PCA plot depicting the relatedness of cells based on DNA methylation profiles.

(D) Dendrogram depicting similarities and differences among cells based on DNA methylation patterns of the most variable probe in each gene. Color scale indicates beta-values (0–1) defined using Illumina Infinium HumanMethylation450 BeadChip. Variable probes were determined based on the SD of all basal and luminal parental lines.

(E) Validation of selected genes by methylation-specific PCR (MSP).

See also Figures S3 and S4 and Tables S2, S5, S6, S7, and S8.

applied highly stringent filters, leaving only the most robust SNPs based on their behavior in public HapMap data (Frazer et al., 2007) and in our samples (see Supplemental Experimental Procedures for details). After filtering, 22,497–52,119 SNP calls remained, depending on the fusion. These markers were dispersed throughout all autosomes and the X chromosome, with ~100 kb between markers on average, allowing us to assess whole genome at moderate resolution. Of these genotypes, 4.8%–27% were "informative," allowing unequivocal assignment of its parental cell of origin (Figure S3A; Table S5; see Supplemental Experimental Procedures for details). In the five heterofusions with a more basal-like phenotype, a high fraction of informative SNPs were inherited from the SUM159 basal-like parent, with the SUM159/MCF7 fusions displaying the most balanced luminal to basal allele ratio (35.1 to 64.9 of luminal to basal). In contrast, the SUM159/ZR75, SUM159/21NT, and MDA-MB-231/MCF7 fusions had a predominance of luminal parental SNPs (Figures 3B and S3B).

We also analyzed whether any particular chromosome showed skewed inheritance and found a strong bias toward X chromosome inherited from SUM159 basal-like parental cells in all heterofusions derived from it (Figure S3A; Table S6). This result was intriguing, as perturbed X chromosome inactivation has been implicated in BRCA1-associated and sporadic basal-like breast cancers (Ganesan et al., 2002; Richardson et al., 2006). In order to determine if X inactivation status was associated with biased X chromosome inheritance, we analyzed the expression of *XIST* by RT-PCR. We found that *XIST* was

expressed in the SUM159 basal-like cell line and in most fusions derived from it (Figure S3C), indicating that at least one copy of the inactive X from SUM159 was inherited in most cells with basal-like phenotype. However, *XIST* was also expressed in 21NT and ZR75-1 luminal lines, and the expression of X chromosome-linked genes (protein-coding mRNAs) in the fusions did not correlate with basal-like and luminal phenotypes (data not shown). Thus, based on these data, it is unclear what, if any, role the X chromosome could play in defining luminal and basal-like features.

To test whether allelic inheritance correlates with cell-type-specific gene expression patterns, we assessed informative SNPs located near (±1 Mb) the TSS (transcription start site) of differentially expressed genes. Over 80% of the genes differentially expressed between basal-like and luminal breast cancer cells had an informative SNP within 1 Mb of its TSS, and in 60%–90% of these (depending on the fusion), the expression pattern was concordant with allelic inheritance (Figures S3B–S3D; Table S5). However, for the most consistently and highly differentially expressed genes, allelic inheritance of informative SNPs was not associated with the cell type specificity of gene expression (Figures 3B and S3B). For example, in the SUM159/MCF7 fusion the ratio of informative SNPs located near genes with high cell-type-specific expression was inherited with almost the same frequency from the SUM159 basal-like and MCF7 luminal parent regardless of expression pattern. There were an equal number of genes displaying basal expression pattern with alleles inherited from the luminal parent and genes with luminal expression but SNPs with basal origin (Figure 3B and S3B–S3D). Thus, it appears that the phenotype and gene expression profiles of the fusions are likely to be defined by non-genetic factors.

Epigenetic Inheritance Patterns

Next, we investigated the potential role of epigenetic mechanisms by analyzing DNA methylation and histone H3 lysine 27 trimethylation (H3K27me3) and acetylation (H3K27ac) patterns of each parental cell line and fusion. We chose to analyze H3K27me3 patterns, as this has been linked to stem cell-like and more differentiated features in multiple tissue types (Mikkelsen et al., 2007), including normal human breast epithelial cells (Choudhury et al., 2013; Maruyama et al., 2011). Similarly, H3K27ac enrichment identifies super-enhancers that are particularly important for defining cellular identity (Hnisz et al., 2013; Whyte et al., 2013).

In contrast to expression profiles, the DNA methylation patterns of luminal-basal heterofusions showed no clear luminal and basal clustering but an intermediary state (Figures 3C, 3D, S4A, and S4B). Correlating with related data in primary tumors (Fackler et al., 2011), luminal cell lines in general were more methylated than basal-like ones (Figures 3D and S4A). All three heterofusions with some luminal features (MDA-MB-231/MCF7, CAL51/MCF7, and SUM159/21NT) were more similar to their luminal than basal parent. These fusions displayed fairly balanced luminal to basal allele ratio, with a bit more skewing toward luminal alleles (Figures 3B and S3B); however, we did not observe any significant association between allele-specific inheritance and DNA methylation status (Table S7). We also analyzed associations between promoter DNA methylation and gene expression and found that basal-high (highly expressed in basal cells) genes are more likely to be hypermethylated in luminal cells than luminal-high (highly expressed in luminal cells) genes in basal cells (Figures S4C and S4D). These results are in agreement with our prior findings in normal mammary epithelial cells (Choudhury et al., 2013; Maruyama et al., 2011) and imply distinct roles for DNA methylation in the regulation of cell-type-specific expression patterns in luminal and basal cells. We validated the methylation of selected genes that showed significant luminal-basal differences by methylation-specific PCR (Figure 3E).

Next, we performed PCA of the samples based on H3K27me3 enrichment and found a clear clustering and separation of parental luminal cells but a high degree of heterogeneity of parental basal cells and heterofusions (Figure 4A). Virtually all three basal-like parental cell lines were unique. Some of the SUM159-derived heterofusions showed high similarity to SUM159 cells, while others formed a cluster in between luminal and basal parental cells. Similar to gene expression and DNA methylation profiles, we did not observe any association between allele-specific inheritance and H3K27me3 enrichment patterns. For example, the basal-high *ZEB2* gene showed luminal and basal allelic inheritance in the SUM159/MCF7 and SUM159/T47D fusions, respectively, but the H3K27me3 enrichment and expression pattern was similar to the basal parent in both fusions, implying reprogramming of the luminal genomes to basal-like chromatin patterns (Figure S4E).

Clustering of the samples based on H3K27ac enrichment profiles also showed very tight and distinct luminal cluster and high heterogeneity of basal-like parental cells and heterofusions (Figure 4B). We defined super-enhancers based on H3K27ac peak counts to identify key drivers of cellular identity. Clustering of the samples based on super-enhancers depicted an even more striking clustering of the luminal cell lines with high relatedness to each other and very distinct features from the highly heterogeneous basal-like group (Figure 4C). More detailed analysis of super-enhancers demonstrated that the high similarity of luminal cells is likely driven by luminal TFs (*FOXA1*, *GATA3*, *ESR1*, and *SPDEF*) that appear to be among the top-ranking super-enhancers in all luminal cell lines (Figure 4D; Table S8). In contrast, super-enhancers in basal cell lines did not show almost any overlap with the exception of *HIF1A* that was a top super-enhancer in both SUM159 and MDA-MB-231 cells. Interestingly, the super-enhancers of heterofusions seemed to reflect the dominance of basal-like parental lines in each case, even for fusions with more luminal phenotypes. In the CAL51/MCF7 fusion, none of the MCF7 super-enhancers were among the top-ranking ones, whereas in the MDA231/MCF7 fusion, *GATA3*, *FOXA1*, *ELF3*, *NCOA3*, and *CDH1* were detected as low-ranking super-enhancers together with the high-ranking MDA-MB-231-derived ones. All SUM159-derived heterofusions displayed high-ranking basal super-enhancers, with luminal ones (*GATA3*) only detected in SUM159/T47D fusion, again highlighting the strong dominance of SUM159 cells. The allelic inheritance again did not correlate with H3K27ac enrichment patterns, confirming epigenetic reprogramming of the genome (Figure 4E).

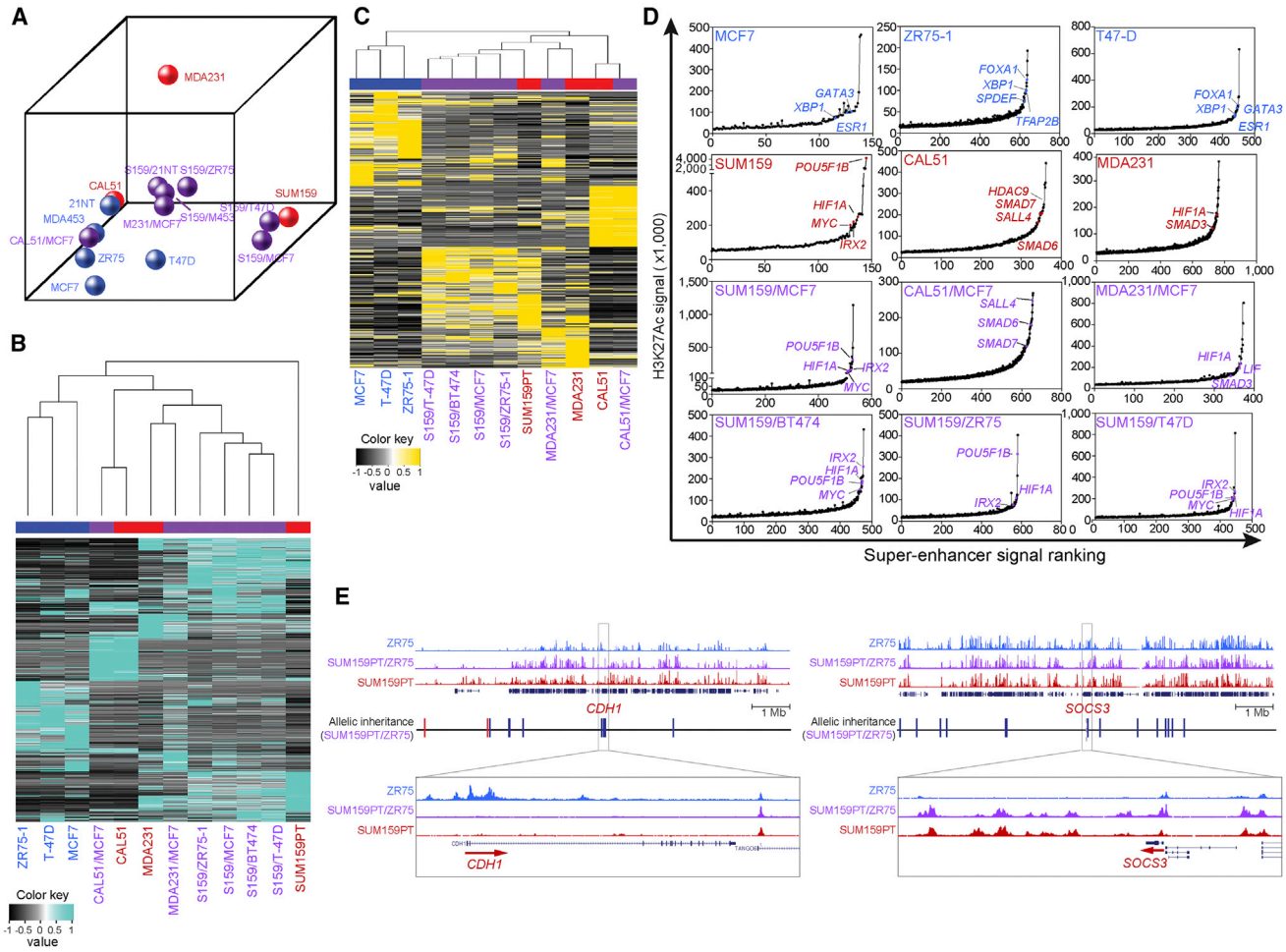


Figure 4. Chromatin Profiles of Parental and Fusion Cells

(A) PCA plot depicting the relatedness of cells based on H3K27me3 patterns.

(B and C) Heatmap and dendrogram from hierarchical clustering analysis depicting similarities and differences among cell lines based on H3K27ac patterns based on top 10% of the most variable peaks considered (B) or super-enhancers (C).

(D) Hockey-stick plots highlighting super-enhancers in each cell line.

(E) Chromatin immunoprecipitation sequencing (ChIP-seq) reads for H3K27ac for *CDH1* and *SOCS3* in parental and fusion cells are shown as tag density along chromosomal position. Height of each track was scaled to total ChIP-seq counts. Allelic inheritance information is indicated as color bar along chromosomal position in the middle.

See also Figure S4 and Tables S4, S7, and S8.

These results highlight the dominance of the basal-like breast cancer epigenome over the luminal one and epigenotype versus genotype, but they also highlight high degree of epigenetic heterogeneity of basal-like breast cancer cell lines and heterofusions that were not evident by expression profiling.

Integrated Molecular Profiles

Next, we integrated gene expression, DNA methylation, and H3K27me3 and H3K27ac enrichment patterns to investigate potential associations and define which one may be a better predictor of cellular phenotypes. First, we analyzed the epigenetic patterns of the 72 top differentially expressed genes (Figure S5A). Approximately half (40%–51%, depending on comparison) of the 72 most significantly differentially expressed genes also showed differential H3K27me3 and H3K27ac enrichment and differences in DNA methylation (Table S7), suggesting their regulation by epigenetic mechanisms. The subset of genes (17 genes) with differential H3K27me3 marks between the luminal and basal parent showed significantly ($p = 0.009895$) lower expression levels than the subset (55 genes) without a differential H3K27me3 mark, whereas subsets of genes with and without differential H3K27ac marks showed no difference in expression and DNA methylation. In contrast to gene expression, the epigenetic marks associated with these 72 genes did not show clear luminal and basal clustering of the samples, although they still separated the parental luminal cells from all others (Figure S5A). Analysis of the genes most differentially methylated between parental luminal and basal lines gave a clearer picture and

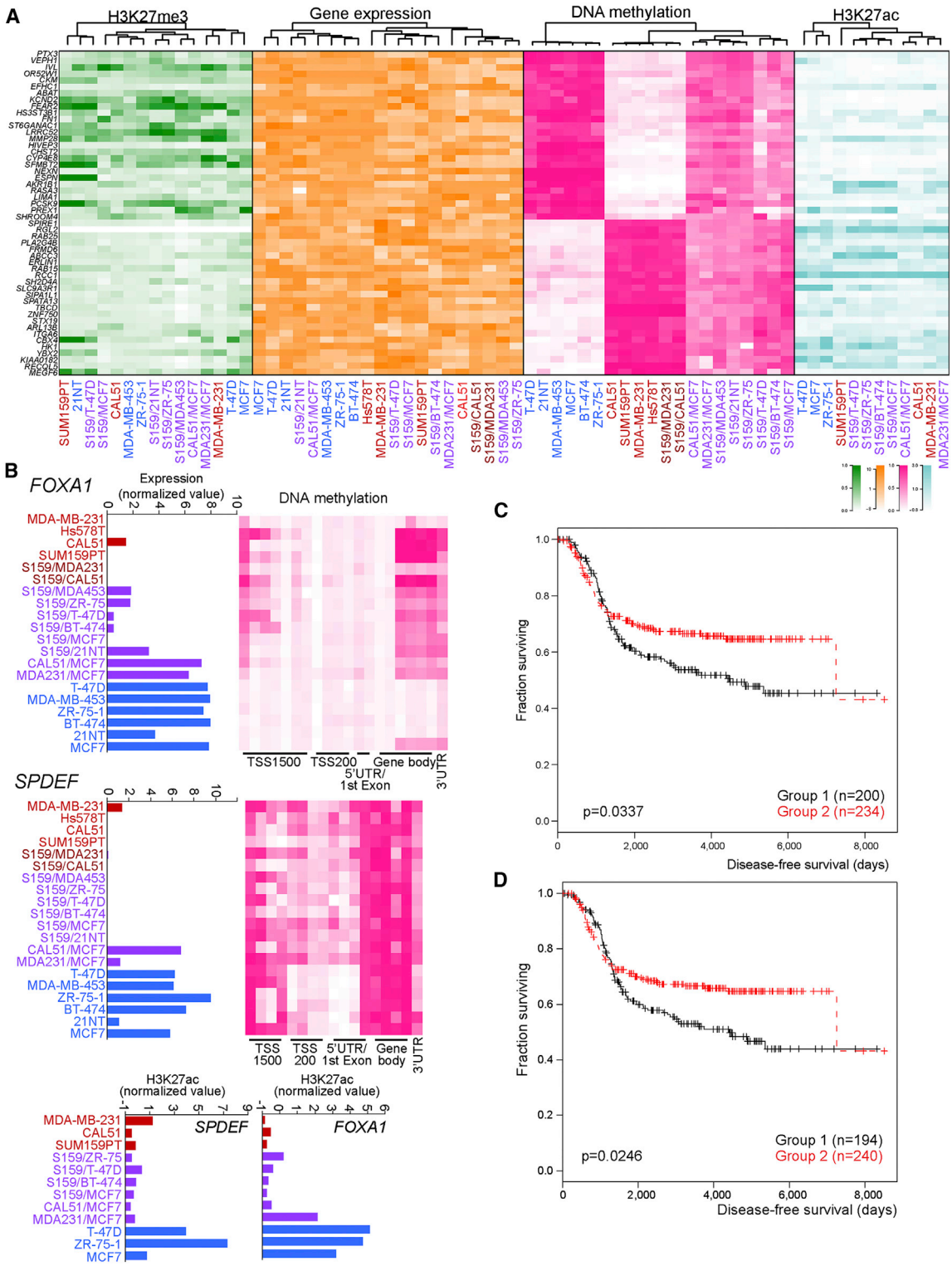


Figure 5. Integrated View of Gene Expression, DNA Methylation, H3K27me3, and H3K27ac Patterns

(A) Heatmap depicting clustering of samples based on gene expression, DNA methylation, H3K27me3, and H3K27ac profiles of the top 50 most differentially methylated genes between parental luminal and basal breast cancer cells.

(B) Gene expression, DNA methylation, and H3K27ac enrichment patterns of luminal transcription factors.

(legend continued on next page)

segregated all heterofusions into a distinct intermediary group (Figure 5A). These results imply that epigenetic patterns may provide a better description of cellular states than gene expression profiles and that the DNA methylation profiles in the fusions are not completely reprogrammed but rather reflect a mixture of luminal and basal features.

Because TFs are key determinants of cellular states, we analyzed their expression and epigenetic patterns in further detail. We identified 123 TFs (34 luminal and 89 basal) that showed significant differences in expression between cells with luminal and basal phenotypes (Table S4). The basal TFs included several members of the ETS family (e.g., *ETS1*), *ZEB1*, and *ZEB2*, whereas top luminal TFs include *FOXA1*, *GATA3*, *SPDEF*, and *ESR1*. None of these basal TFs were among the top basal super-enhancers, and *HIF1A*, which was a top basal super-enhancer, was not among the top significantly differentially expressed genes. In contrast, all top differentially expressed luminal TFs were also top luminal super-enhancers common among all luminal lines (Figure 4D). These key luminal TFs were not expressed and their promoter region was hypermethylated in cells with basal phenotype (Figures 5B and S5B), which is particularly interesting, since most luminal-high genes were not hypermethylated in basal cells (Figures S4C and S4D), implying preferential epigenetic regulation of TFs.

These results further highlight the relative homogeneity of the luminal breast cancer cell state and TFs defining it, which is in contrast to the high heterogeneity of basal-like breast cancer cell state, with lack of luminal TFs being the main commonality.

Clinical Relevance of Epigenetic Heterogeneity in TNBC

To investigate the potential clinical relevance of epigenetic heterogeneity in basal-like breast cancer, we first defined epigenetic differences between SUM159 cells that have the most dominant basal-like trait based on our fusion studies and MDA-MB-231 and CAL-51 cells that were more permissive to luminal differentiation. Then we classified TNBCs in the TCGA cohort (TCGA, 2012) based on probes that displayed DNA hypomethylation in SUM159 relative to the CAL-51 and MDA-MB-231 cell lines (beta < 0.1 in SUM159, beta > 0.9 in the other two lines) and analyzed differences in clinical outcome between groups. Although there was a trend for more methylated cases being associated with shorter overall survival, this did not reach statistical significance (Figure S5C). Thus, we explored the METABRIC cohort (Curtis et al., 2012) that has a larger number of cases; however, it lacks epigenetic data. Therefore, we investigated differences in survival based on the expression of genes that were enriched for H3K27ac and associated with super-enhancers in any of the three basal cell lines (Figure 5C) or only in SUM159 cells (Figure 5D). We identified 717 and 163 such genes, respectively, and their expression classified ER$^-$ breast tumors into two groups with statistically significant differences in disease-free survival and tumors, with higher expression of these genes having shorter survival. Thus, heterogeneity for histone modification profiles appears to classify ER$^-$ breast tumors into clinically relevant groups. Because disease-free and overall survival is mainly determined by progression to distant metastatic disease, our results imply that the risk of metastatic progression might be defined by epigenetic programs, and therefore, it is a modifiable risk.

Nuclear Reprogramming of Luminal Breast Cancer Cells

To further investigate if basal-like and luminal breast cancer cell phenotypes are defined by non-genetic factors, we performed in vitro reprogramming experiments (Collas and Gammelsaeter, 2007; Collas and Taranger, 2006). Specifically, we exposed streptolysin-O permeabilized luminal cells (e.g., MCF7) to whole-cell extracts from basal-like (e.g., SUM159 and MDA-MB-231) breast cancer cells for 1 hr followed by resealing the cells and cultivation for 2–6 weeks (Figure 6A). Reprogrammed cells were analyzed for CD24 and CD44 as luminal and basal cell-surface markers, respectively. A small (<2%) but reproducible percentage of MCF7 cells exposed to extracts from SUM159, but not MDA-MB-231, cells switched from the luminal CD24$^+$CD44$^-$ to basal CD44$^+$CD24$^-$ phenotype, implying successful reprogramming (Figure 6B). The frequency of these cells is in agreement with results of similar assays performed in other cell types, including embryonic stem cells (Egli et al., 2008; Hanna et al., 2010). The results using Dnase-I-treated extracts were essentially the same (data not shown), indicating that the transmission of DNA is not required. We characterized these CD44$^+$CD24$^-$ cells in further detail following their enrichment by FACS. The cells displayed spindle-shape morphology and high expression of vimentin, a basal-specific marker (Figure 6C). We also analyzed the expression of several other basal-like (e.g., *VIM*, *ITGA5*, *IGFBP7*, and *CD44*) and luminal (e.g., *CDH1* and *ERBB3*) markers by qRT-PCR in control and reprogrammed MCF7 cells. The reprogrammed CD44$^+$ cells expressed basal-like markers comparable to the parental SUM159 cells but still maintained the expression of the luminal markers analyzed (Figures 6D–6E), implying an incomplete phenotypic switch. Correlating with this, analysis of the DNA methylation status of a set of cell-type-specifically expressed and methylated genes by MSP (methylation-specific PCR) revealed an intermediary pattern in these CD44$^+$ cells (Figure 6F). We tried to expand these CD44$^+$CD24$^-$ cells to further characterize their molecular and functional features, but the majority of cells lost their phenotype during repeated passage and reverted back to CD44$^-$CD24$^+$ cells. We also tried reprogramming basal-like cells (SUM159 and CAL51) with nuclear extracts from luminal cells (MCF7 and 21NT), but we were not able to detect any luminal features in basal-like cells exposed to luminal extracts (data not shown). These results are consistent with our somatic cell fusion studies and further emphasize the general

(C) Kaplan-Meier plot depicting disease-free survival of breast cancer patients in the METABRIC cohort with ER negative (ER$^-$) tumors (440 cases) classified into groups 1 and 2 based on the expression of super-enhancer-associated genes (717 genes) in any of the three parental basal cell lines analyzed.
(D) Kaplan-Meier plot depicting disease-free survival of breast cancer patients in the METABRIC cohort with ER negative (ER$^-$) tumors (440 cases) classified into groups 1 and 2 based on the expression of 186 probes (163 genes) positive for super-enhancers in SUM159 cell line within 20 kb of the TSS.
See also Figure S5 and Tables S4, S7, and S8.

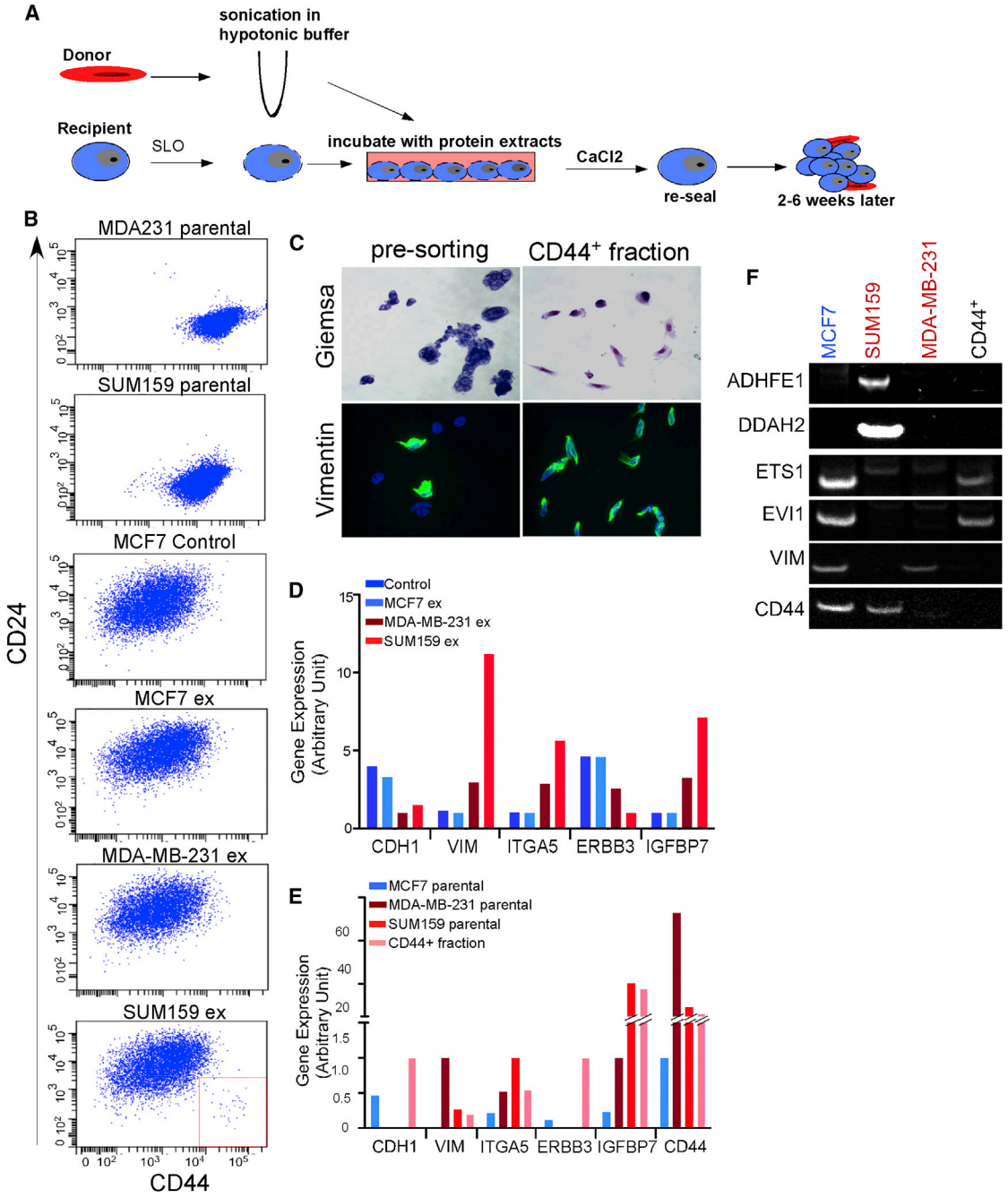


Figure 6. Reprogramming of Luminal Breast Cancer Cells Using Nuclear Extracts

(A) Schematic outline of the in vitro reprogramming assay.

(B) FACS analysis of CD24 luminal and CD44 basal markers 6 weeks after in vitro reprogramming.

(C) Morphological and immunofluorescence analysis of CD44$^+$ fractions of the reprogrammed MCF7 cells.

(D) Gene expression analysis of MCF7 cells reprogrammed with protein extracts from MDA-MB-231 (MDA-MB-231 ex), SUM159 (SUM159 ex), and MCF7 (MCF7 ex) cells. Control represents the cells permeabilized by Streptolysin "o" (SLO) alone.

(E) Gene expression analysis of parental MCF7, MDA-MB-231, and SUM159 cell lines, and CD44$^+$ fraction of the reprogrammed MCF7 cells.

(F) Methylation-specific PCR analysis of MCF7, MDA-MB-231, and SUM159 parental cell lines, and CD44$^+$ fraction of the reprogrammed MCF7 cells.

dominance of the basal-like trait and the role of non-genetic mechanisms in the induction of the luminal to basal-like phenotypic switch.

Transcriptional Regulators of the Basal-like Phenotype

To identify factors that may mediate the luminal to basal-like reprogramming of breast cancer cells, we focused on TFs that are

epigenetically regulated in a cell-type-specific manner, since these are strong candidates for playing key roles in defining cellular phenotypes. We identified 61 TFs with higher expression in basal-like parental cells and heterofusions with basal-like phenotype compared to luminal parental cells and heterofusions with mixed luminal-basal phenotypes (Table S4). We found that the promoter region of 8 of these TFs (*CREB5*, *EN1*, *FOXC2*, *MSC*, *TBX18*, *TCF4*, *TSHZ2*, and *ZBTB16*) is enriched for H3K27me3 in luminal parental and heterofusion cells with low expression of these genes, implying potential epigenetic regulation. Among these 8 TFs, the expression of *EN1* and *TBX18* displayed the most significant association between luminal and basal features in both parental and fusion cells (Figure 7A). Next, we tested whether the expression of these 8 and 19 other top-candidate TFs (total 27) alone or in combination would be able to induce a luminal-to-basal phenotypic switch in MCF7 cells as assessed by FACS analysis for the loss of luminal markers such as CD24. We found that MCF7 cells expressing *EN1*, *TBX18*, or *TCF4* alone or in combination consistently showed a significant increase in the $CD24^-$ cell population, suggesting loss of luminal features (Figure 7B), although we did not detect a gain of basal markers analyzed (e.g., CD44), implying incomplete reprogramming.

Because our somatic cell fusion studies suggested that luminal to basal reprogramming is determined by epigenetic mechanisms, we also investigated whether modulating the activity of epigenetic regulators would influence the success of reprogramming induced by the expression of TFs or by cell fusion. Luminal lines overall had higher level of DNA methylation than basal ones (Figure 3D) and preferentially silenced basal-high genes by DNA methylation (Figures S4C and S4D). Thus, we first tested whether the expression of known DNA demethylases *TET1*, *TET2*, *TET3, AICDA, GADD45a*, *GADD45b*, *TDG*, and *MBD4* in luminal MCF7 cells, alone or in combination, would be sufficient to induce a luminal-to-basal switch, but we did not detect loss of luminal marks in any of the combinations tested (data not shown). Similarly, pre-treating SUM159 or MCF7 cells with inhibitors of DNMTs, EZH2 (H3K27 histone methyltransferase), and KDM4 (H3K9 and H3K36 histone demethylase) prior to cell fusion did not have any effect on the number of fusion cells and their phenotype, despite significant effects on histone modification patterns (Figures S6A–S6C). In contrast, transient pre-treatment of SUM159 cells with KDM6A/B H3K27me3 histone demethylase inhibitor GSK-J4 (Kruidenier et al., 2012) prior to fusion resulted in a significant reduction in the number of viable heterofusion cells (Figure 7C). This decrease was not due to the inhibitory effects of GSK-J4 on cell growth, as treatment of SUM159 cells or SUM159/MCF7 fusions did not reveal a significant decrease in cell numbers (Figures S6D and S6E). Transient downregulation of *KDM6A*, but not *KDM6B*, by small interfering RNAs (siRNAs) (Figure S6F) had similar effects on the frequency of viable heterofusions (Figure 7D), although we also observed a slight decrease in cell growth (Figure S6G). Both GSK-J4 treatment and downregulation of *KDM6A* led to increased H3K27me3 levels (Figures 7E and 7F), and neither pharmacologic inhibition nor downregulation of KDM6A changed the expression of luminal- and basal-specific genes in the SUM159/MCF7 heterofusions (Figures 7G and 7H). These results imply that, similar to what was observed for the generation of induced pluripotent stem cells (iPSCs), the activity of KDM6A/UTX may be required for the epigenetic reprogramming of breast cancer cells (Mansour et al., 2012).

DISCUSSION

Here, we present multiple lines of evidence that the basal-like phenotype in breast cancer is generally a dominant trait defined by epigenetic mechanisms. This conclusion is supported by our integrated genomic analyses of somatic cell fusions generated from luminal and basal-like breast cancer cells and nuclear reprogramming and TF studies. While our data cannot fully rule out the possible contribution of genetic factors, even if such factors exist, they must operate via epigenetic mechanisms. Our results also highlight significant epigenetic heterogeneity in basal-like breast cancer cell lines, with suppression of transcription programs driving luminal differentiation being the main commonality.

We also found some fundamental differences in the epigenetic programs of luminal and basal-like breast cancer cells that may reflect the distinct cell of origin of these tumors or tumor subtype-specific transforming events. Specifically, basal-like breast tumors have an overall lower level of DNA methylation than luminal ones, and luminal lineage markers are preferentially silenced by H3K27me3 in basal-like cells. In contrast, in luminal breast cancer cells, stem/basal cell markers are frequently silenced by DNA methylation. Similar methylation patterns were observed in normal $CD24^+$ luminal and basal ($CD24^-CD10^+$ or $CD24^-CD10^-CD44^+$) cells (Maruyama et al., 2011), implying that DNA methylation may be a feature of luminal cells and may be required for the requisition of a differentiated state. One limitation of our study is the use of Illumina Beadchips for assessing DNA methylation, which does not provide information at a single-nucleotide and genome-wide scale; thus, it remains to be seen if these hypotheses are supported by higher-resolution DNA methylation data.

Our most surprising finding is that even a transient exposure of MCF7 luminal cells to nuclear extracts of SUM159 basal cells was sufficient to induce a transient phenotypic switch in a subset of cells. This result implies that some key proteins could initiate a cascade of events, likely involving positive feedback loops and cross-regulatory networks of TFs, which could suppress luminal differentiation programs. Our data indicate the lack of specific TFs that would be common in all basal-like breast cancer and suggest that epigenetic repression of luminal factors is the main commonality. Nevertheless, we identified three TFs (*TBX18*, *EN1*, and *TCF4*), the overexpression of which was able to induce the repression of some luminal features in MCF7 luminal breast cancer cells, although this was not a stable phenotypic switch. None of these TFs have been characterized in detail in breast cancer. T-box 18 (*TBX18*) has not been implicated in any cancer type, but it plays essential roles in embryonic development and is required for the epithelial to mesenchymal transition induced by TGF-β in epicardial cells during heart development (Takeichi et al., 2013). Engrailed 1 (*EN1*) encodes a homeobox protein that is required for mid-hindbrain development. *EN1* is hypermethylated in a number of cancer types,

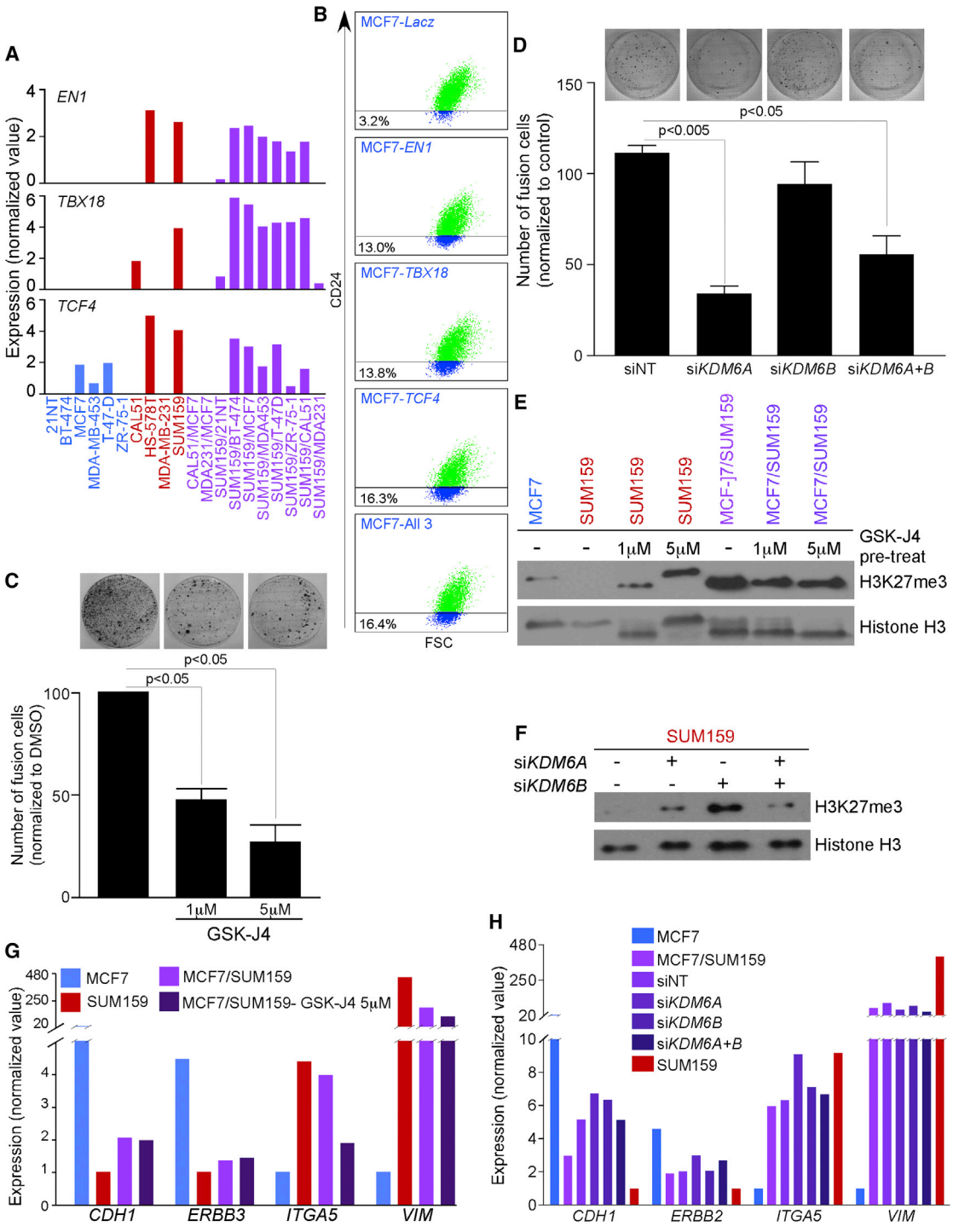


Figure 7. Epigenetic Reprogramming of Breast Cancer Cells

(A) Expression of *EN1*, *TBX18*, and *TCF4* in parental and fusion cell lines.

(B) FACS analysis of CD24 luminal marker in MCF7 cells overexpressing basal-specific transcription factors *EN1*, *TBX18*, and *TCF4* alone or in combination; *LacZ* was used as control.

(C) Representative plates (top) and bar graphs (bottom) of colony growth assays assessing the frequency of colonies resulting from successful SUM159/MCF7 fusions derived from SUM159 cells pre-treated with 1 μM or 5 μM GSK-J4 or DMSO control ($p < 0.05$ when indicated, unpaired t test).

(D) Representative plates (top) and bar graphs (bottom) of colony growth assays assessing the frequency of colonies resulting from successful SUM159/MCF7 fusions derived from SUM159 cells transfected with non-targeting (NT) and siRNA-targeting *KDM6A* or *KDM6B* or both genes ($p < 0.005$ or $p < 0.05$ when indicated, unpaired t test).

(legend continued on next page)

and it was recently reported as a prosurvival factor in basal-like breast cancer cells (Beltran et al., 2014). *TCF4* is a key transcriptional mediator of the WNT/β-catenin pathway and plays a role in a wide-range of human malignancies, including breast cancer (Polakis, 2007). We also found that transient pharmacologic or genetic inhibition of KDM6A in SUM159 cells reduced the number of viable SUM159/MCF7 heterofusions without any clear effects on their phenotype. This is similar to what was described for the generation of iPSCs from human fibroblasts, where KDM6A catalytic activity is required for the induction, but not the maintenance, of pluripotency during in vitro reprogramming (Mansour et al., 2012).

Although spontaneous somatic cell fusions and nuclear reprogramming are not likely to occur in vivo, the transfer of proteins, mRNAs, and microRNAs by exosomes could induce similar phenotypic switch within tumors (Valadi et al., 2007). Correlating with this, a recent study described the enhancement of pro-metastatic phenotype in bone marrow cells by melanoma-derived exosomes (Peinado et al., 2012). Future studies are required to address if similar mechanisms may also contribute to metastatic progression and therapeutic resistance in heterogeneous tumors composed of mixtures of basal-like and luminal breast cancer cells.

In summary, our results imply that basal-like breast cancer phenotype is a generally dominant trait likely defined by epigenetic rather than genetic mechanisms. We also found that basal-like breast cancer cell lines and tumors are epigenetically highly heterogeneous. As a consequence, these phenotypes may change during disease progression and may be modifiable therapeutically to improve treatment outcomes.

EXPERIMENTAL PROCEDURES

For further details, see the Supplemental Experimental Procedures.

Generation of Somatic Cell Fusions

First, we established derivatives of parental luminal and basal cell lines resistant to puromycin or G418 by infecting the cells with pBabe retrovirus expressing the appropriate resistance gene followed by selection using the appropriate drug. Stable pools of drug resistance cell lines were used to generate somatic cell hybrids by the modification of the PEG fusion protocol (Davidson and Gerald, 1976) as previously described (Polyak et al., 1996). For cell lines that we could not derive viable fusions using the PEG procedure, we also tried electrofusion (performed by Mayo Clinic Medical Laboratories) without success. The identity of each cell line was confirmed by STR and SNP array analyses. For fusion with epigenetic inhibitors, the IC_{50} (inhibitory concentration 50%) for each of the inhibitors was determined using standard method. Briefly, 1,000–2,000 cells were plated in each well of 96-well plate, and increasing concentrations of inhibitors were added the next day. Five days after addition of inhibitors, cell viability was measured using the CTG method and IC_{50} determined. SUM159 cells were pre-treated with GSK126, GSKJ4, and ML324 or DMSO control. The effect of the inhibitors was confirmed by checking the protein expression levels for H3K27me3, H3K9me3, and H3K26me3. 4 million SUM159 cells pre-treated with either the inhibitors or DMSO control were fused with 4 million MCF7 cells using the same protocol described previously. After 2 weeks of selection, fusion was confirmed by measuring DNA content and the colonies were stained with 0.5% crystal violet solution, and fusion cells were counted. To determine the effects of inhibitors on cell growth, inhibitor-treated cells used for fusion were grown for 10 days and colonies stained and the number of cells counted. Fusions with siRNAs for KDM6A/B:siRNAs against *KDM6A* and *KDM6B* were purchased from Dharmacon in the form of ONTARGET-plus SMARTpool reagents. Transfections of siRNA (25 nM) were performed using DharmaFECT transfection reagents (Dharmacon) in accordance with the manufacturer's protocol. Non-targeting siRNA was used as a control. Three days after transfection, 4 million SUM159 cells, SUM159 cells transfected with non-targeting siRNA control and siRNA against *KDM6A* and *KDM6B*, were fused with 4 million MCF7 cells using the same protocol described previously.

DNA Content, STR Polymorphism, and SKY Analyses

Cells were harvested, washed in PBS, and fixed in ice-cold 70% ethanol at 4°C overnight. Fixed cells were re-suspended in a solution containing 100 μg/ml RNase and 40 μg/ml propidium iodide (Sigma) and incubated for 30 min at 37°C with agitation. The DNA content of 10,000 cells was determined with FACS Canto II (BD Biosciences). STR fingerprinting and SKY were performed by the Molecular Genetics Laboratory (Dana-Farber Cancer Institute) and DF/HCC Cytogenetics Core Facility, respectively, following standard procedures.

ACCESSION NUMBERS

All raw genomic data have been deposited to the NCBI GEO under the accession number GEO:GSE38548.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.05.011.

AUTHOR CONTRIBUTIONS

Y.S. generated cell fusions and performed chromatin immunoprecipitation sequencing (ChIP-seq), RNA sequencing (RNA-seq), nuclear reprogramming, and other experiments. A.S. generated cell fusions, defined reprogramming factors, and performed other experiments. N.B.-Q., M.L., and J.K. generated cell fusions and performed other experiments. V.S., A.Z., and A.G. performed SNP array and statistical analyses. D.P.T. helped with cell culture and WGS data acquisition. M.K. and Z.SZ. analyzed WGS data. R.M. performed ChIP-seq and RNA-seq, analyzed genomic data, and performed statistical analyses. L.X. and H.L. analyzed ChIP-seq data. M.J.F and S.S. performed DNA methylation profiling. H.S. assisted with ChIP-seq and genomic data analyses. K.P. supervised the overall study. All authors helped to design the study and write the manuscript.

ACKNOWLEDGMENTS

We thank members of our laboratories for their critical reading of this manuscript and useful discussions and the Cytogenetics Core of Dana-Farber Harvard Cancer Center (P30 CA006516) for the SKY studies. This work was supported by the US Army Congressionally Directed Research Program BC084006 and

(E) Immunoblot analysis of H3K27me3 levels in MCF7, SUM159, and MCF7/SUM159 fusion cells with or without GSK-J4 pre-treatment. Histone H3 was used as control.

(F) Immunoblot analysis of H3K27me3 levels in SUM159 cells transfected with non-targeting (NT) and siRNA targeting *KDM6A* or *KDM6B* or both genes. Histone H3 was used as control.

(G and H) mRNA expression of luminal (*CDH1*, *ERBB3*) and basal (*ITGA5*, *VIM*) genes in MCF7, SUM159, and MCF7/SUM159 fusion cells with SUM159 pre-treated with DMSO or 5 μM GSKJ4 (G) or derived from SUM159 cells transfected with non-targeting (NT) and siRNA targeting *KDM6A* or *KDM6B* or both genes (H). Error bars in (C) and (D) denote SEM. See also Figure S6.

BC134001 (K.P.) and BC087579 (A.M.), Kræftens Bekæmpelse, the Danish Cancer Society (M.K.), and the Susan G. Komen Foundation (R.M. and Y.S.).

Received: March 10, 2015
Revised: April 28, 2015
Accepted: May 6, 2015
Published: June 4, 2015

REFERENCES

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. USA *100*, 3983–3988.

Beltran, A.S., Graves, L.M., and Blancafort, P. (2014). Novel role of Engrailed 1 as a prosurvival transcription factor in basal-like breast cancer and engineering of interference peptides block its oncogenic function. Oncogene *33*, 4767–4777.

Choudhury, S., Almendro, V., Merino, V.F., Wu, Z., Maruyama, R., Su, Y., Martins, F.C., Fackler, M.J., Bessarabova, M., Kowalczyk, A., et al. (2013). Molecular profiling of human mammary gland links breast cancer risk to a p27(+) cell population with progenitor characteristics. Cell Stem Cell *13*, 117–130.

Collas, P., and Gammelsaeter, R. (2007). Novel approaches to epigenetic reprogramming of somatic cells. Cloning Stem Cells *9*, 26–32.

Collas, P., and Taranger, C.K. (2006). Epigenetic reprogramming of nuclei using cell extracts. Stem Cell Rev. *2*, 309–317.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.; METABRIC Group (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

Davidson, R.L., and Gerald, P.S. (1976). Improved techniques for the induction of mammalian cell hybridization by polyethylene glycol. Somatic Cell Genet. *2*, 165–176.

Egli, D., Birkhoff, G., and Eggan, K. (2008). Mediators of reprogramming: transcription factors and transitions through mitosis. Nat. Rev. Mol. Cell Biol. *9*, 505–516.

Fackler, M.J., Umbricht, C.B., Williams, D., Argani, P., Cruz, L.A., Merino, V.F., Teo, W.W., Zhang, Z., Huang, P., Visvananthan, K., et al. (2011). Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. Cancer Res. *71*, 6195–6207.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al.; International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.

Ganesan, S., Silver, D.P., Greenberg, R.A., Avni, D., Drapkin, R., Miron, A., Mok, S.C., Randrianarison, V., Brodie, S., Salstrom, J., et al. (2002). BRCA1 supports XIST RNA concentration on the inactive X chromosome. Cell *111*, 393–405.

Hajra, K.M., Ji, X., and Fearon, E.R. (1999). Extinction of E-cadherin expression in breast cancer via a dominant repression pathway acting on proximal promoter elements. Oncogene *18*, 7274–7279.

Hanna, J.H., Saha, K., and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell *143*, 508–525.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Kruidenier, L., Chung, C.W., Cheng, Z., Liddle, J., Che, K., Joberty, G., Bantscheff, M., Bountra, C., Bridges, A., Diallo, H., et al. (2012). A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. Nature *488*, 404–408.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. *121*, 2750–2767.

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A., et al.; kConFab (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. *15*, 907–913.

Liu, S., Ginestier, C., Charafe-Jauffret, E., Foco, H., Kleer, C.G., Merajver, S.D., Dontu, G., and Wicha, M.S. (2008). BRCA1 regulates human mammary stem/progenitor cell fate. Proc. Natl. Acad. Sci. USA *105*, 1680–1685.

MacDougall, J.R., and Matrisian, L.M. (2000). Targets of extinction: identification of genes whose expression is repressed as a consequence of somatic fusion between cells representing basal and luminal mammary epithelial phenotypes. J. Cell Sci. *113*, 409–423.

Mansour, A.A., Gafni, O., Weinberger, L., Zviran, A., Ayyash, M., Rais, Y., Krupalnik, V., Zerbib, M., Amann-Zalcenstein, D., Maza, I., et al. (2012). The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. Nature *488*, 409–413.

Marotta, L.L., Almendro, V., Marusyk, A., Shipitsin, M., Schemme, J., Walker, S.R., Bloushtain-Qimron, N., Kim, J.J., Choudhury, S.A., Maruyama, R., et al. (2011). The JAK2/STAT3 signaling pathway is required for growth of $CD44^{+}CD24^{-}$ stem cell-like breast cancer cells in human tumors. J. Clin. Invest. *121*, 2723–2735.

Maruyama, R., Choudhury, S., Kowalczyk, A., Bessarabova, M., Beresford-Smith, B., Conway, T., Kaspi, A., Wu, Z., Nikolskaya, T., Merino, V.F., et al. (2011). Epigenetic regulation of cell type-specific expression patterns in the human mammary epithelium. PLoS Genet. *7*, e1001369.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Molyneux, G., Geyer, F.C., Magnay, F.A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell *7*, 403–417.

Park, S.Y., Gönen, M., Kim, H.J., Michor, F., and Polyak, K. (2010a). Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. J. Clin. Invest. *120*, 636–644.

Park, S.Y., Lee, H.E., Li, H., Shipitsin, M., Gelman, R., and Polyak, K. (2010b). Heterogeneity for stem cell-related markers according to tumor subtype and histologic stage in breast cancer. Clin. Cancer Res. *16*, 876–887.

Peinado, H., Alečković, M., Lavotshkin, S., Matei, I., Costa-Silva, B., Moreno-Bueno, G., Hergueta-Redondo, M., Williams, C., García-Santos, G., Ghajar, C., et al. (2012). Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. Nat. Med. *18*, 883–891.

Polakis, P. (2007). The many ways of Wnt in cancer. Curr. Opin. Genet. Dev. *17*, 45–51.

Polyak, K. (2007). Breast cancer: origins and evolution. J. Clin. Invest. *117*, 3155–3163.

Polyak, K., Waldman, T., He, T.C., Kinzler, K.W., and Vogelstein, B. (1996). Genetic determinants of p53-induced apoptosis and growth arrest. Genes Dev. *10*, 1945–1952.

Prat, A., and Perou, C.M. (2011). Deconstructing the molecular portraits of breast cancer. Mol. Oncol. *5*, 5–23.

Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J.D., Livingston, D.M., and Ganesan, S. (2006). X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell *9*, 121–132.

Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., et al. (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature *486*, 395–399.

Shipitsin, M., Campbell, L.L., Argani, P., Weremowicz, S., Bloushtain-Qimron, N., Yao, J., Nikolskaya, T., Serebryiskaya, T., Beroukhim, R., Hu, M., et al. (2007). Molecular definition of breast tumor heterogeneity. Cancer Cell *11*, 259–273.

Sjöblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., et al. (2006). The consensus

coding sequences of human breast and colorectal cancers. Science *314*, 268–274.

Takeichi, M., Nimura, K., Mori, M., Nakagami, H., and Kaneda, Y. (2013). The transcription factors Tbx18 and Wt1 control the epicardial epithelial-mesenchymal transition through bi-directional regulation of Slug in murine primary epicardial cells. PLoS ONE *8*, e57829.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J.J., and Lötvall, J.O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat. Cell Biol. *9*, 654–659.

Vaz-Luis, I., Ottesen, R.A., Hughes, M.E., Mamet, R., Burstein, H.J., Edge, S.B., Gonzalez-Angulo, A.M., Moy, B., Rugo, H.S., Theriault, R.L., et al. (2014). Outcomes by tumor subtype and treatment pattern in women with small, node-negative breast cancer: a multi-institutional study. J. Clin. Oncol. *32*, 2142–2150.

Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314–322.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.