**Research Article**

**TP53 Mutation, Epithelial-Mesenchymal Transition, and Stemlike Features in Breast Cancer Subtypes**

Danila Coradini, Marco Fornili, Federico Ambrogi, Patrizia Boracchi, and Elia Biganzoli

Department of Clinical and Community Health Sciences, Medical Statistics, Biometry and Bioinformatics, University of Milan, Via Vanzetti, 5 20133 Milan, Italy

Correspondence should be addressed to Danila Coradini, danila.coradini@yahoo.it

Received 23 March 2012; Revised 11 May 2012; Accepted 14 June 2012

Academic Editor: F. C. Schmitt

Copyright © 2012 Danila Coradini et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Altered p53 protein is prevalently associated with the pathologic class of triple-negative breast cancers and loss of p53 function has recently been linked to the induction of an epithelial-mesenchymal transition (EMT) and acquisition of stemness properties. We explored the association between TP53 mutational status and expression of some genes involved in the canonical TGF-β signaling pathway (the most potent EMT inducer) and in two early EMT associated events: loss of cell polarity and acquisition of stemness-associated features. We used a publicly accessible microarray dataset consisting of 251 p53-sequenced primary breast cancers. Statistical analysis indicated that mutant p53 tumors (especially those harboring a severe mutation) were consistent with the aggressive class of triple-negative cancers and that, differently from cell cultures, surgical tumors underexpressed some TGF-β related transcription factors known as involved in EMT (ID1, ID4, SMAD3, SMAD4, SMAD5, ZEB1). These unexpected findings suggest an interesting relationship between p53 mutation, mammary cell dedifferentiation, and the concomitant acquisition of stemlike properties (as indicated by the overexpression of PROM1 and NOTCH1 genes), which improve tumor cells aggressiveness as indicated by the overexpression of genes associated with cell proliferation (CDK4, CDK6, MKI67) and migration (CXCR4, MMP1).

1. Introduction

TP53 tumor suppressor is the most commonly altered gene in human breast cancer where it is mutated in about 30–40% [1]. TP53 gene mutations result in altered and stable p53 proteins that function as dominant negative with gain-of-function properties, including drug resistance, and contribute to malignant progression with detrimental effects on patient's outcome [2]. In particular, clinical evidence indicated that altered p53 proteins are prevalently associated with the pathologic class of triple-negative breast cancers, that is, tumors characterized by the immunohistochemical expression of basal cytokeratins and epidermal growth factor receptor (EGFR), but negative for estrogen (ER), progesterone receptor (PR), and HER2 expression [3–5]. Recently, triple-negative tumors have been also associated with a new less common subtype, known as claudin-low [6].

Wild-type p53 functions as a sequence-specific DNA binding transcription factor that regulates a plethora of target genes involved in DNA repair, cell cycle control, apoptosis, senescence, angiogenesis, and other fundamental biological process [7]. Therefore, it is not surprising that mutations of TP53 gene or inactivations of its signaling pathway are prerequisite for the development of tumors. Most mutations occur within the central DNA binding domain (exons 5–8) and, in particular, at several specific amino acids required for DNA binding. According to the type of mutation (point mutation, deletion, insertion, or stop codon), p53 protein synthesis may be totally inhibited or may generate functionally altered molecules that affect cell homeostasis in a different manner [8]. In fact, it is well known that not all p53 mutations have equal effects: some of them confer loss of function, others have a dominant negative effect and still others are classified as wild-type-like
protein and represent mutant forms with a limited biological effect [9, 10].

Recently, some excellent papers have provided experimental evidence linking loss of p53 function to the induction of epithelial-mesenchymal transition (EMT) and acquisition of stemness properties in different tumor cell lines [11–13]. EMT is a key program in embryonic development, the aberrant reactivation of which may induce progression, invasion, dissemination, and finally metastasis in cancer cells. The most evident peculiarities of a cell undergoing EMT process are loss of the epithelial phenotype and acquisition of mesenchymal features and abnormal motility capabilities [14, 15]. The most potent inducer of EMT is transforming growth factor-β (TGF-β) that triggers the activity of several transcription factors (ZEB1/TCF8, ZEB2/SIP1, Snail, Slug, Twist, and Ids), which in turn repress the expression of genes coding for epithelial markers and activate the expression of mesenchymal genes [16]. According to recent acquisition, p53 should prevent EMT by repressing ZEB1 and ZEB2 expressions via miRNAs activity. Consequently, p53 loss-of-function should downregulate miRNAs expression, the transactivation of transcription factors promoting EMT, and the emergence of tumor cells with stemlike properties [17–19].

So far, despite the availability of a huge amount of information on the transcript profile from microarray analysis, the interrelations among p53 mutations and genes involved in EMT have not been specifically assessed. Therefore, to investigate the association between TP53 mutational status and EMT process, we interrogated a publicly accessible microarray dataset consisting of 251 p53 sequenced primary breast cancers [20]. Adopting an unconventional approach, we did not use the whole transcript profile but we selected a priori panel of genes experimentally recognized as involved in the canonical TGF-β signaling pathway and in two early events associated with EMT: loss of epithelial cell polarity and acquisition of stemness-associated features. To delineate a more comprehensive picture of the relationship among p53 mutation, EMT, and tumor aggressiveness, we also considered some genes involved in cell proliferation, apoptosis, and metastatic spread.

2. Materials and Methods

2.1. Materials. As reported in the original paper [20], gene expression profile was determined by using the Affymetrix Human Genome HG-U133A and -B GeneChip, and microarray dataset while was available at the ArrayExpress website (http://www.ebi.ac.uk/arrayexpress/), with the accession number E-GEOD-3494. Patients and tumors characteristics were provided as supporting information in the original paper [20].

2.2. Gene Selection. According to the aim of the study, we selected 147 genes (Table 1). Specifically, the panel was composed of 27 genes recognized as involved in TGF-β-induced EMT, 57 involved in epithelial cell plasticity, 13 coding for stemlike properties and 31 involved in cell proliferation, apoptosis, and metastatic spread. In addition, to describe breast cancer subtypes, 19 genes coding for luminal and basal markers were also considered. The 147 genes corresponded to 352 Affymetrix probe sets, as verified by GeneAnnot system v2.0 (http://bioinfo2.weizmann.ac.il/geneannot/), that additionally provided information about the quality of each probe set in terms of sensitivity and specificity score [21] (Supplementary Table 1 available at doi:10.1155/2012/254085).

2.3. Statistical Analysis. As some genes are recognized by more than a single probe set, each of which characterized by an individual specificity and sensitivity that differently contribute to gene expression value, a gene expression mean value was calculated after weighting each probe set for its own sensitivity and specificity score. Specifically, each expression value (already log2 transformed in the original dataset) was multiplied for the semisum of sensitivity and specificity scores of the corresponding probe set.

Prediction Analysis for Microarray (PAM) analysis was used to identify genes associated with the TP53 mutation status. PAM methodology minimizes the classification error using cross validation. For the selected genes, shrunken centroids across the different mutation groups were plotted. The FDR level was estimated through a permutation method.

To identify the tumors characterized by a similar intrinsic phenotype (tumor subtypes), unsupervised hierarchical cluster analysis was performed using the subset of genes coding for luminal and basal markers, HER-2, and claudins. The choice of the number of clusters to be used was supported by mean silhouette values [22]. PAM methodology was used to detail differential gene expression among clusters [23].

Principal Components Analysis (PCA) and PCA-based biplots were used to assess gene expression among clusters [24]. Moreover, for evaluating the associations among genes, specific subsets, not used to build PCA, were passively projected over the PCA-based biplots of intrinsic phenotypes.

All analyses were performed using open source software R 2.14.1 packages stats, cluster, and HDMD (http://www.R-project.org/).

3. Results and Discussion

As described in the original paper [20], the cases series was composed of 251 tumors, 58 of which characterized by a TP53 mutation. Of the 58 mutant tumors, 37 had point mutations and 21 had “severe” mutations, that is, insertions (n = 3), deletions (n = 11), and stop codons (n = 7), that result in frame shift and truncations with deleterious functional consequences.

To identify genes differentially expressed between mutant and wild-type p53 tumors, we first applied a PAM analysis on the overall cases series (Figure 1). According to the expected loss-of-function, mutant p53 tumors were characterized by the underexpression of genes under p53 control (i.e., CDKN1A, coding for p21, and TP53INP, coding for the p53-inducible nuclear protein 1). In addition, they were showed as underexpressed several genes coding for epithelial cell
Table 1: Genes entered in the study.

| Official gene symbol | Gene name                                | Entrez gene ID | Ensembl genomic location |
|----------------------|------------------------------------------|----------------|--------------------------|
| **Apical junctional complex** |                                          |                |                          |
| CDH1                 | Cadherin 1, type 1, E-cadherin (epithelial) | 999            | 16q22.1                  |
| CDH2                 | Cadherin 2, type 1, N-cadherin (neuronal)  | 1000           | 18q12.1                  |
| CLDN1                | Claudin 1                                 | 9076           | 3q28                     |
| CLDN2                | Claudin 2                                 | 9075           | Xq22.3                   |
| CLDN3                | Claudin 3                                 | 1365           | 7q11.23                  |
| CLDN4                | Claudin 4                                 | 1364           | 7q11.23                  |
| CLDN5                | Claudin 5                                 | 7122           | 22q11.21                 |
| CLDN6                | Claudin 6                                 | 9074           | 16p13.3                  |
| CLDN7                | Claudin 7                                 | 1366           | 17p13.1                  |
| CLDN8                | Claudin 8                                 | 9073           | 21q22.11                 |
| CLDN9                | Claudin 9                                 | 9080           | 16p13.3                  |
| CLDN10               | Claudin 10                                | 9071           | 13q32.1                  |
| CLDN11               | Claudin 11                                | 5010           | 3q26.2                   |
| CLDN12               | Claudin 12                                | 9069           | 7q21.13                  |
| CLDN14               | Claudin 14                                | 23562          | 21q22.3                  |
| CLDN15               | Claudin 15                                | 24146          | 7q22.1                   |
| CLDN16               | Claudin 16                                | 10686          | 3q28                     |
| CLDN17               | Claudin 17                                | 26285          | 21q22.11                 |
| CLDN18               | Claudin 18                                | 51208          | 3q22.3                   |
| CLDN23               | Claudin 23                                | 137075         | 8p23.1                   |
| CTNNA1               | Catenin (cadherin-associated protein), alpha 1 | 1495          | 5q31.2                   |
| CTNNB1               | Catenin (cadherin-associated protein), beta 1 | 1499          | 3p22.1                   |
| CTNND1               | Catenin (cadherin-associated protein), delta 1 | 1500          | 11q12.1                  |
| JAM2                 | Junctional adhesion molecule 2            | 58494          | 21q21.3                  |
| JAM3                 | Junctional adhesion molecule 3            | 83700          | 11q25                    |
| JUP                  | Junction plakoglobin                      | 3728           | 17q21.2                  |
| MAGI1                | Membrane associated guanylate kinase, WW and PDZ domain containing 1 | 9223          | 3p14.1                   |
| MARVELD2             | MARVEL domain containing 2               | 153562         | 5q13.2                   |
| OCLN                 | Occludin                                 | 4950           | 5q13.2                   |
| PVRL1                | Poliovirus receptor-related 1 (herpesvirus entry mediator C) | 5818          | 11q23.3                  |
| PVRL2                | Poliovirus receptor-related 2 (herpesvirus entry mediator B) | 5819          | 19q13.32                 |
| PVRL3                | Poliovirus receptor-related 3            | 25945          | 3q13.13                  |
| PVRL4                | Poliovirus receptor-related 4            | 81607          | 1q23.3                   |
| SYMPK                | Symplekin                                | 8189           | 19q13.32                 |
| TJP1                 | Tight junction protein 1 (zona occludens 1) | 7082          | 15q13.1                  |
| TJP2                 | Tight junction protein 2 (zona occludens 2) | 9414          | 9q21.11                  |
| TJP3                 | Tight junction protein 3 (zona occludens 3) | 27134         | 19p13.3                  |
| VCL                  | Vinculin                                 | 7414           | 10q22.2                  |
| **Apoptosis**        |                                          |                |                          |
| BAX                  | BCL2-associated X protein                 | 581            | 19q13.33                 |
| BCL2                 | B-cell CLL/lymphoma 2                    | 596            | 18q21.33                 |
| CASP2                | Caspase 2, apoptosis-related cysteine peptidase | 835           | 7q34                     |
| ERBB2                | V-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (avian) | 2064          | 17q12                    |
| MDM2                 | Mdm2 p53 binding protein homolog (mouse) | 4193           | 12q15                    |
| TP5                  | Tumor protein p53                        | 7157           | 17p13.1                  |
| TP53INP1             | Tumor protein p53 inducible nuclear protein 1 | 94241         | 8q22.1                   |
| XIAP                 | X-linked inhibitor of apoptosis           | 331            | Xq25                     |
| **Basal markers**    |                                          |                |                          |
| ACTA1                | Actin, alpha 1, skeletal muscle           | 58             | 1q42.13                  |
| CD44                 | CD44 molecule (Indian blood group)        | 960            | 11p13                    |
| EGFR                 | Epidermal growth factor receptor          | 1956           | 7p11.2                   |
| KRT5                 | Keratin 5                                | 3852           | 12q13.13                 |
| KRT6A                | Keratin 6A                               | 3853           | 12q13.13                 |
| KRT6B                | Keratin 6B                               | 3854           | 12q13.13                 |
| Official gene symbol | Gene name                   | Entrez gene ID | Ensembl genomic location |
|----------------------|-----------------------------|----------------|--------------------------|
| KRT14                | Keratin 14                  | 3861           | 17q21.2                  |
| KRT17                | Keratin 17                  | 3872           | 17q21.2                  |
| TP63                 | Tumor protein p63           | 8626           | 3q28                     |
| VIM                  | Vimentin                    | 7431           | 10p13                    |
| CCNB3                | Cyclin B3                   | 85417          | Xp11.22                  |
| CCND1                | Cyclin D1                   | 595            | 11q13.3                  |
| CCNE2                | Cyclin E2                   | 9134           | 8q22.1                   |
| CDK2                 | Cyclin-dependent kinase 2   | 1017           | 12q13.2                  |
| CDK4                 | Cyclin-dependent kinase 4   | 1019           | 12q14.1                  |
| CDK6                 | Cyclin-dependent kinase 6   | 1021           | 7q21.2                   |
| CDKN1A               | Cyclin-dependent kinase inhibitor 1A (p21, Cip1) | 1026 | 6p21.2                  |
| CDKN1B               | Cyclin-dependent kinase inhibitor 1B (p27, Kip1) | 1027 | 12p13.1                  |
| CDKN2A               | Cyclin-dependent kinase inhibitor 2A (melanoma, p16, CDK4) | 1029 | 9p21.3                  |
| CDKN2B               | Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4) | 1030 | 9p21.3                  |
| CDKN2D               | Cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4) | 1032 | 19p13.2                  |
| GAS1                 | Growth arrest-specific 1    | 2619           | 9q21.33                  |
| MKI67                | Antigen identified by monoclonal antibody Ki-67 | 4288 | 10q26.2                  |
| CRB1                 | Crumbs homolog 1(Drosophila) | 23418          | 1q31.3                   |
| CRB3                 | Crumbs homolog 3 (Drosophila) | 92359       | 19p13.3                  |
| DLG1                 | Discs, large homolog 1 (Drosophila) | 1739 | 3q29                     |
| DLG2                 | Discs, large homolog 2 (Drosophila) | 1740 | 11q14.1                  |
| DLG3                 | Discs, large homolog 3 (Drosophila) | 1741 | Xq13.1                   |
| DLG4                 | Discs, large homolog 4 (Drosophila) | 1742 | 17p13.1                  |
| DLG5                 | Discs, large homolog 5 (Drosophila) | 9231 | 10q22.3                  |
| INADL                | InaD-like (Drosophila)      | 10207          | 1p31.3                   |
| MPP5                 | Membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5) | 64398 | 14q23.3                  |
| LLGL1                | Lethal giant larvae homolog 1 (Drosophila) | 3996 | 17p11.2                  |
| LLGL2                | Lethal giant larvae homolog 2 (Drosophila) | 3993 | 17q25.1                  |
| PARD3                | Par-3 partitioning defective 3 homolog (C. elegans) | 56288 | 10p11.21                 |
| PARD3B               | Par-3 partitioning defective 3 homolog B (C. elegans) | 117583 | 2q33.3                   |
| PARD6A               | Par-6 partitioning defective 6 homolog alpha (C. elegans) | 50855 | 16q22.1                  |
| PARD6B               | Par-6 partitioning defective 6 homolog beta (C. elegans) | 84612 | 20q13.13                 |
| PARD6G               | Par-6 partitioning defective 6 homolog gamma (C. elegans) | 84552 | 18q23                     |
| PRKCI                | Protein kinase C, iota      | 5584           | 3q26.2                   |
| PRKCC                | Protein kinase C, zeta      | 5590           | 1p36.33                  |
| SCRIB                | Scribbled homolog (Drosophila) | 23513 | 8q24.3                   |
| AKT1                 | V-akt murine thymoma viral oncogene homolog 1 | 207            | 14q32.33             |
| AKT2                 | V-akt murine thymoma viral oncogene homolog 2 | 208            | 19q13.2                  |
| BMP1                 | Bone morphogenetic protein 1 | 649           | 8p21.3                   |
| ID1                  | Inhibitor of DNA binding 1, dominant negative helix-loop-helix protein | 3397 | 20q11.21                |
| ID2                  | Inhibitor of DNA binding 2, dominant negative helix-loop-helix protein | 3398 | 2p25.1                   |
| ID3                  | Inhibitor of DNA binding 3, dominant negative helix-loop-helix protein | 3399 | 1p36.12                  |
| ID4                  | Inhibitor of DNA binding 4, dominant negative helix-loop-helix protein | 3400 | 6p22.3                   |
| SMAD2                | SMAD family member 2        | 4087           | 18q21.1                  |
| SMAD3                | SMAD family member 3        | 4088           | 15q22.33                 |
| SMAD4                | SMAD family member 4        | 4089           | 18q21.2                  |
| Official gene symbol | Gene name | Entrez gene ID | Ensembl genomic location |
|----------------------|-----------|---------------|--------------------------|
| SMAD5                | SMAD family member 5 | 4090 | 5q31.1 |
| SMAD6                | SMAD family member 6 | 4091 | 15q22.31 |
| SMAD7                | SMAD family member 7 | 4092 | 18q21.1 |
| SMURF1               | SMAD specific E3 ubiquitin protein ligase 1 | 57154 | 7q22.1 |
| SMURF2               | SMAD specific E3 ubiquitin protein ligase 2 | 64750 | 17q24.1 |
| SNAI1                | Snail homolog 1 (Drosophila) | 6615 | 20q13.13 |
| SNAI2                | Snail homolog 2 (Drosophila) | 6591 | 8q11.21 |
| TCF3                 | Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47) | 6929 | 19p13.3 |
| TGFBI                | Transforming growth factor, beta-induced, 68 kDa | 7045 | 5q31.1 |
| TGFBR1               | Transforming growth factor, beta receptor 1 | 7046 | 9q22.33 |
| TGFBR2               | Transforming growth factor, beta receptor II | 7048 | 3p24.1 |
| TGFBR3               | Transforming growth factor, beta receptor III | 7049 | 1p22.1 |
| TGIF2                | TGFβ-induced factor homeobox 2 | 60436 | 20q11.23 |
| TWIST1               | Twist homolog 1 (Drosophila) | 7291 | 7p21.1 |
| TWIST2               | Twist homolog 2 (Drosophila) | 117581 | 2q37.3 |
| ZEB1                 | Zinc finger E-box binding homeobox 1 | 6935 | 10p11.22 |
| ZEB2                 | Zinc finger E-box binding homeobox 2 | 9839 | 2q22.3 |
| CD24                 | CD24 molecule | 100133941 | 6q21 |
| ESR1                 | Estrogen receptor 1 | 2099 | 6q25.1 |
| GATA3                | GATA binding protein 3 | 2625 | 10p14 |
| KRT7                 | Keratin 7 | 3855 | 12q13.13 |
| KRT8                 | Keratin 8 | 3856 | 12q13.13 |
| KRT18                | Keratin 18 | 3875 | 12q13.13 |
| KRT19                | Keratin 19 | 3880 | 17q21.2 |
| MUC1                 | Mucin 1, cell surface associated | 4582 | 1q22 |
| PGR                  | Progesterone receptor | 5241 | 11q22.1 |
| CXCR4                | Chemokine (C-X-C motif) receptor 4 | 7852 | 2q22.1 |
| CXCR5                | Chemokine (C-X-C motif) receptor 5 | 643 | 11q23.3 |
| MMP1                 | Matrix metallopeptidase 1 (interstitial collagenase) | 4312 | 11q22.2 |
| MMP2                 | Matrix metallopeptidase 2 (gelatinase A, 72 kDa gelatinase, 72 kDa type IV collagenase) | 4313 | 16q12.2 |
| MMP3                 | Matrix metallopeptidase 3 (stromelysin 1, progelatinase) | 4314 | 11q22.2 |
| MTA1                 | Metastasis associated 1 | 9112 | 14q32.33 |
| MTA2                 | Metastasis associated 1 family, member 2 | 9219 | 11q12.3 |
| MTA3                 | Metastasis associated 1 family, member 3 | 57504 | 2p21 |
| PAK6                 | P21 protein (Cdc42/Rac)-activated kinase 6 | 56924 | 15q15.1 |
| TIMP1                | TIMP metallopeptidase inhibitor 1 | 7076 | Xp11.23 |
| ABCG2                | ATP-binding cassette, sub-family G (WHITE), member 2 | 9429 | 4q22.1 |
| ALDH1A1              | Aldehyde dehydrogenase 1 family, member A1 | 216 | 9q21.13 |
| ALDH1A3              | Aldehyde dehydrogenase 1 family, member A3 | 220 | 15q26.3 |
| BMI1                 | BMI1 polycomb ring finger oncogene | 648 | 10p12.2 |
| JAG1                 | Jagged 1 | 182 | 20p12.2 |
| JAG2                 | Jagged 2 | 3714 | 14q32.33 |
| NANO G               | Nanog homeobox | 79923 | 12p13.31 |
| NOTCH1               | Notch 1 | 4851 | 9q34.3 |
| NOTCH2               | Notch homolog 2 (Drosophila) | 4853 | 1p12 |
| NOTCH3               | Notch homolog 3 (Drosophila) | 4854 | 19p13.12 |
| NOTCH4               | Notch homolog 4 (Drosophila) | 4855 | 6p21.32 |
| NUMB                 | Numb homolog (Drosophila) | 8650 | 14q24.3 |
| PROM1                | Prominin 1 | 8842 | 4p15.32 |
Figure 1: Shrunken centroids for wild-type TP53 and mutant TP53 tumors. Left-sided bars indicate lower expression in the subgroups relative to overall centroid; right-sided bars indicate higher expression in subgroups relative to overall centroid.

polarity complex (*DLG4*, *DLG5*, *INADL*, *PARD3B*, *PARD6B*) and apical junctional components (*CLDN11*, *JAM2*, *JAM3*, *MAGI1*, *OCLN*, *PVRL2*, *TPJ1*, *TPJ3*). Conversely, mutant p53 tumors overexpressed *CDH2*, the gene coding for N-cadherin, and some genes involved in cell proliferation (*CCNE2*, *CDK4*, *CDK6*, *CDKN2A*, *CDKN2D*, *MKI67*) and metastatic spread (*CXCR4*, *MMP1*). An overall similar pattern of expression was found when we interrogated another publicly available microarray dataset where p53 mutational status was known [3] using the same panel of 147 genes. Despite the different microarray platform used (42K cDNA microarrays instead of Affymetrix GeneChip), mutant p53 tumors were associated with a dramatic underexpression of genes coding for apical junctional components (*INADL*, *JAM2*, *JAM3*, *MAGI1*, *PARD6B*, *PVRL2*) coupled with the overexpression of genes coding for N-cadherin (*CDH2*) or involved in cell proliferation and metastatic spread (*CCNE2*, *CDK4*, *CDK6*, *MKI67*, *MTA1*) (Supplementary Figure 1). Notably, in both datasets, p53 mutant tumors were associated with the overexpression of *PROM1*, supporting the experimental evidence indicating the relation among p53 mutation and the reacquisition of some stemlike properties according to an EMT-like process [12, 13].

As regards the genes related to the canonical TGF-β signaling pathway, mutant p53 tumors showed the down-regulation of many genes coding for pivotal elements of this pathway (*ID1*, *ID4*, *SMAD3*, *SMAD4*, *SMAD5*, *TGFBR2*, *TGFBR3*, *ZEB1*) coupled with the overexpression of *SMURF2*, coding for a SMAD-specific E3 ubiquitin protein ligase, and *TGIF2*, coding for a transcriptional repressor interacting with TGF-β activated SMAD proteins (Figure 2). A similar pattern of expression (i.e., downregulation of *SMAD2*, *SMAD3*, *SMAD5*, *SMAD7*, *TGFBR2*, *TGFBR3*, *ZEB1*, and overexpression of *SMURF2* and *TGIF2*) was found in Langerod dataset [3] (Supplementary Figure 2). This unexpected finding could be explained taking into account that TGF-β is a multifunctional cytokine and a powerful tumor suppressor that governs many aspects of mammary epithelial cells physiology and homeostasis [25].

Consistent with the notion that estrogen receptor and TGF-β signaling pathways are major regulators during mammary
Figure 2: Boxplots of the genes associated with EMT in wild-type (WT) or mutant (mut) TP53 tumors with respect to overall case series (all).

gland development [26], it is not surprising that p53 mutant tumors concomitantly underexpressed ESR1 (coding for ER\(\alpha\)), TGFBR2, TGFBR3 (coding for TGF-\(\beta\) receptors) and ID1, ID4, SMAD3, SMAD4, SMAD5, ZEB1 (coding for key elements of the pathway).

When severe and missense mutations were considered separately, PAM analysis provided evidence that severe TP53 mutations were responsible for the differential gene expression observed in mutant with respect to wild-type p53 tumors, even though some important alterations were already present in missense TP53 mutations, as, for example, the downregulation of some apical junctional components or the overexpression of genes related to cell proliferation and invasion. As shown in Figure 3, in addition to the expected decrease in the expression of TP53, CDKN1A, and TP53INP1, tumors harboring severe mutations were characterized by a dramatic overexpression of genes associated with proliferation (CDK4, CDK6, MKI67) and metastatic spread (CXCR4, MMP1), and by underexpression of several genes involved in epithelial cell identity (DLG5, INADL, JAM2, JAM3, MAGI1, OCLN, PARD3B, PARD6B, PVRL2, SCRIB, TJP1 and TJP3).

As regards the association between missense or severe TP53 mutation and EMT-related genes, Figure 4 indicates that, with respect to tumors harboring a missense mutation, those with severe mutations were characterized by the overexpression of SMURF2, SNAI1, and TGIF2 genes. Of particular interest is the overexpression of SNAI1 gene because of the concomitant overexpression of NOTCH1 pointed out by PAM analysis in tumors with severe TP53 mutation (Figure 3). Indeed, Notch signalling pathway, which is implicated as an important contributor to EMT in tumorigenesis, has been recently suggested to play a direct role on the expression of the Snail transcription factor [27].

With respect to missense TP53 mutations, severe ones were also characterized by an increased expression of PROM1, the gene encoding for prominin, a pentaspan transmembrane glycoprotein (CD133) often overexpressed on cancer cells, where it is thought to function in maintaining stem cell properties by suppressing differentiation. This finding is in agreement with the hypothesis that basal cancers, which have been proposed to have a stem cell origin, are virtually all TP53 mutants and express high levels of PROM1 transcript and protein [28]. Unfortunately, since in Langerod dataset [3] severe mutations accounted for only three cases, we were unable to verify all these observations in an independent dataset.
One of the aims of the study was to explore the relationship among p53 mutation, EMT, and tumor aggressiveness, a peculiar characteristic of certain breast cancer subtypes (especially basal-like phenotype). To this specific aim, we performed an unsupervised hierarchical cluster analysis, using the subset of genes coding for luminal and basal markers, HER-2, and claudins, and we looked at the distribution of p53 mutations according to tumor subtype. The analysis indicated that mutant p53 tumors distributed into three main clusters (Figure 5). Of the 58 mutant p53 tumors, 23 were included in Cluster 1, 17 in Cluster 2, and 18 in Cluster 3. However, looking at the relative percentage, we found that, on the total number of tumors in each cluster, only 17% (23/133) of Cluster 1 and 19% (18/95) of Cluster 3 tumors had p53 mutations, whereas 74% (17/23) of Cluster 2 tumors did have. Notably, 10 of these 17 mutations were severe mutations.

Conversely, Cluster 3 tumors were positively associated with genes related to luminal phenotype (ESR1, GATA3, MUC1, PGR, KRT18) and with a different panel of claudin-coding genes (CLDN3, CLDN4, CLDN7) and negatively associated with genes related to basal phenotype. Cluster 1 tumors showed a less clear-cut phenotype according to the more heterogeneous nature of this cluster, even though they appeared prevalently, associated with genes related to basal phenotype (KRT5, KRT6B, KRT14, KRT17, TP63). Remarkably, Cluster 2 tumors also showed the concomitant underexpression of ERBB2 gene providing evidence that these tumors had a gene expression profile consistent with the pathologic class of triple-negative tumors (Supplementary Figure 3), which are characterized by the expression of basal cytokeratins (mainly Krt5) and EGFR, but do not express estrogen and progesterone receptors, and HER2.

When we looked at the expression of EMT-associated genes according to clusters (Figure 7), we found that Cluster 2, consistent with the pathologic class of triple-negative cancers, showed a gene expression profile similar to that of tumors harboring a severe p53 mutation. Conversely, Cluster 3, consistent with the luminal-like phenotype, had a pattern of expression similar to that of wild-type p53 tumors whereas the phenotypically heterogenous Cluster 1 looks like the group of tumors with a missense p53 mutation. In particular,
Cluster 2 (consistent with the pathologic class of triple-negative cancers and akin to severe p53 mutated tumors) was characterized by the underexpression of SMAD2, SMAD5, ZEB1, and TGFBR3 and the overexpression of SMURF2, TGIF2, and SNAI1, in agreement with the gene profile observed in tumors harboring a severe TP53 mutation (Figure 4).

Notably, when the subset of genes related to stemness properties was passively projected over the PCA-based biplots provided in Figure 6, Cluster 2 tumors were positively associated with PROM1 and NOTCH1, and negatively associated with ALDH1A1, BMI1, NUMB (Figure 8). Similar to the latter but opposite in the sign, was the pattern of association shown by Cluster 3 tumors.
The imbalance in Numb/Notch pathway observed in Cluster 2 tumors, associated with the overexpression of SNAIL1, is of particular interest because the involvement of this pathway in differentiation program and epithelial cancer progression and metastasis. Numb is an evolutionary conserved protein that plays a critical role in cell fate determination, including control of asymmetric cell division, endocytosis, cell adhesion, cell migration, and ubiquitination of specific substrates as p53. Loss of Numb causes increased activity of the oncogene Notch1 and for this reason, low expression of Numb and high levels of Notch1 have been associated with tumor progression and used as markers of tumor aggressiveness, especially in basal-like breast cancer [29]. The aggressiveness of this group of tumors was corroborated by the observation that Cluster 2 tumors were prevalently poorly differentiated (17/23 tumors were Grade III) with respect to Cluster 1 and Cluster 3 tumors, and by the positive association with genes promoting cell proliferation and metastatic spread.

In this context, it can be viewed the overexpression of MMP1 and CXCR4, and the downregulation of TIMP1. Indeed, MMP1 encodes for a matrix metalloproteinase family member (specifically, a collagenase) involved in the breakdown of extracellular matrix whereas TIMP1 encodes for a specific tissue inhibitor of metalloproteases, including MMP-1. Because MMP1 is a target gene for wild-type p53 activity, the functional inactivation of the protein results in a gene overexpression that allows tumor cell migration after degradation of basement membrane and cell detachment [30, 31]. The concomitant overexpression of CXCR4 due to a gain-of-function mutant p53 [32, 33], further contributes to enhance tumor cell migration and metastatic spread [34]. In fact, CXCR4 encodes a C-X-C motif chemokine receptor specific for stromal cell-derived factor-1 (SDF-1/CXCL12), a member of the family of chemoattractant molecules, physiologically involved in the migration of immune cells. The CXCL12/CXCR4 signaling axis is also known to be important for tumor cell migration: CXCR4 expressed on tumor cells, provides a means of homing for metastatic cells to target organs [35]. Due to its implication with tumor dissemination, CXCR4 overexpression has been linked to a poor prognosis in breast cancer patients [35].

Surprisingly, on the contrary, it should be the negative association, pointed out by PCA-based biplots, between Cluster 2 tumors and BMI1 expression. That, because the role of BMI1 gene in self-renewal of stem cells and as an oncogene in many human cancers where it induces EMT. Although Bmi1 overexpression has been correlated with poor prognosis in several tumor types, a recent study has indicated that, in breast cancer, high Bmi1 expression is limited to the luminal subtype and that it is associated with a good outcome [36]. Under this light, the positive association that we observed between BMI1 expression and Cluster 3 tumors, consistent with the luminal-like phenotype, seems to provide a transcriptomic support to this clinical evidence.
Figure 7: Boxplots of the genes associated with EMT in the three main clusters identified by unsupervised hierarchical cluster analysis using the subset of genes coding for luminal and basal markers, HER-2 and claudins.

Cluster 1 tumors, which are prevalently p53 wild-type, are more difficult to categorize. Dissimilarly from basal-like and luminal-like, these tumors had an indefinite phenotype characterized by the coexpression of luminal and basal cytokeratins. In addition, the overexpression of several transcription factors (ID2, ID4, SNAI2, TWIST1, ZEB2), known to be under TGF-β control, and the concomitant overexpression of some genes coding for stemlike properties (ABCG2, JAG1, JAG2, NANOG, NOTCH4) makes it difficult to have a correct interpretation of the results. Indeed, it is not easy to establish whether such a phenotypical heterogeneity represents an intermediate step of an EMT-like process, in which tumor cells gain characteristics of mesenchymal cells but have not completely lost epithelial characteristics, or it is simply due to the individual heterogeneity of the tumors forming the cluster.

4. Conclusions

Aim of this in silico study was to investigate the association between TP53 mutational status and expression of a panel of genes related to TGF-β induced EMT and stemlike features, using a publicly accessible microarray dataset consisting of 251 p53-sequenced primary breast cancers. According to recent experimental evidence linking loss of p53 function, induction of EMT and acquisition of stemness properties in different tumor cell lines [11–13], we expected an evident positive association between EMT-related genes and p53 mutations, in particular with severe p53 mutations. In addition, since clinical evidence indicates that p53 mutations are prevalently associated with the pathologic class of triple-negative breast cancers, we expected an overexpression of EMT-related genes in this specific subset of tumors. Our analysis supports the notion that mutant p53 tumors (especially those harboring a severe p53 mutation) were consistent with the aggressive clinic class of triple-negative cancers, but it clearly indicates that, differently from cell cultures [11–13], surgical tumors did not overexpress TGF-β-related transcription factors. Taking into account the physiological role of TGF-β in mammary gland differentiation [25, 26], these unexpected findings seem to suggest an interesting relationship between p53 mutation, mammary cell dedifferentiation, and the concomitant acquisition of stemlike properties which improve tumor cells aggressiveness.
References

[1] M. Olivier, A. Langerod, P. Carriero et al., “The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer,” *Clinical Cancer Research*, vol. 12, no. 4, pp. 1157–1167, 2006.

[2] A. Di Leo, M. Tanner, C. Desmedt et al., “p-53 gene mutations as a predictive marker in a population of advanced breast cancer patients randomly treated with doxorubicin or docetaxel in the context of a phase III clinical trial,” *Annals of Oncology*, vol. 16, no. 6, pp. 997–1003, 2007.

[3] A. Langerod, H. Zhao, O. Borgan et al., “TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer,” *Breast Cancer Research*, vol. 9, no. 3, article R30, 2007.

[4] B. J. Chae, J. S. Bae, A. Lee et al., “p53 as a specific prognostic factor in triple-negative breast cancer,” *Japanese Journal of Clinical Oncology*, vol. 39, no. 4, pp. 217–224, 2009.

[5] E. Biganzoli, D. Coradini, F. Ambrogi et al., “p53 status identifies two subgroups of triple-negative breast cancers with distinct biological features,” *Japanese Journal of Clinical Oncology*, vol. 41, no. 2, pp. 172–179, 2011.

[6] A. Prat, J. S. Parker, O. Karginova et al., “Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer,” *Breast Cancer Research*, vol. 12, no. 5, article R68, 2010.

[7] K. H. Vousden and C. Prives, “Blinded by the light: the growing complexity of p53,” *Cell*, vol. 137, no. 3, pp. 413–431, 2009.

[8] A. J. Levine and M. Oren, “The first 30 years of p53: growing ever more complex,” *Nature Reviews Cancer*, vol. 9, no. 10, pp. 749–758, 2009.

[9] P. Monti, P. Campomenosi, Y. Ceribilli et al., “Tumour p53 mutations exhibit promoter selective dominance over wild type p53,” *Oncogene*, vol. 21, no. 11, pp. 1641–1648, 2002.

[10] J. J. Jordan, A. Inga, K. Conway et al., “Altered-function p53 missense mutations identified in breast cancers can have subtle effects on transactivation,” *Molecular Cancer Research*, vol. 8, no. 5, pp. 701–716, 2010.

[11] T. Kim, A. Veronese, F. Pichiorri et al., “p53 regulates epithelial-mesenchymal transition through microRNAs targeting ZEB1 and ZEB2,” *Journal of Experimental Medicine*, vol. 208, no. 5, pp. 875–883, 2011.

[12] C.-J. Chang, C.-H. Chao, W. Xia et al., “P53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs,” *Nature Cell Biology*, vol. 13, no. 3, pp. 317–323, 2011.

[13] A. V. Pinho, I. Rooman, and F. X. Real, “p53-dependent regulation of growth, epithelial-mesenchymal transition and stemness in normal pancreatic epithelial cells,” *Cell Cycle*, vol. 10, no. 8, pp. 1312–1321, 2011.

[14] J. Yang and R. A. Weinberg, “Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis,” *Developmental Cell*, vol. 14, no. 6, pp. 818–829, 2008.

[15] D. S. Micalizzi, S. M. Farabaugh, and H. L. Ford, “Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression,” *Journal of Mammary Gland Biology and Neoplasia*, vol. 15, no. 2, pp. 117–134, 2010.

[16] J. Zavadil and E. P. Böttinger, “TGF-β and epithelial-to-mesenchymal transitions,” *Oncogene*, vol. 24, no. 37, pp. 5764–5774, 2005.

[17] J. Schubert and T. Braibatzis, “P53 spreads out further: suppression of EMT and stemness by activating miR-200c expression,” *Cell Research*, vol. 21, no. 5, pp. 705–707, 2011.
[18] N. H. Kim, H. S. Kim, X. Y. Li et al., “A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition,” *Journal of Cell Biology*, vol. 195, pp. 417–433, 2011.

[19] C. J. Braun, X. Zhang, I. Savelyeva et al., “p53-responsive microRNAs 192 and 215 are capable of inducing cell cycle arrest,” *Cancer Research*, vol. 68, no. 24, pp. 10094–10104, 2008.

[20] L. D. Miller, J. Smeds, J. George et al., “An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 38, pp. 13550–13555, 2005.

[21] F. Ferrari, S. Bortoluzzi, A. Coppe et al., “Novel definition files for human GeneChips based on GeneAnnot,” *BMC Bioinformatics*, vol. 8, article 446, 2007.

[22] P. J. Kaufman and L. Rousseeuw, *Finding Groups in Data—An Introduction To Cluster Analysis*, John Wiley & Sons, New York, NY, USA, 1990.

[23] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu, “Diagnosis of multiple cancer types by shrunked centroids of gene expression,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 10, pp. 6567–6572, 2002.

[24] T. Jolliffe, *Principal Components Analysis*, Springer, New York, NY, USA, 2nd edition, 2002.

[25] R. Derynck, R. J. Akhurst, and A. Balmain, “TGF-β signaling in tumor suppression and cancer progression,” *Nature Genetics*, vol. 29, no. 2, pp. 117–129, 2001.

[26] A. M. Band and M. Laiho, “Crosstalk of TGF-β and estrogen receptor signaling in breast cancer,” *Journal of Mammary Gland Biology and Neoplasia*, vol. 16, no. 2, pp. 109–115, 2011.

[27] S. Saad, S. R. Stanners, R. Yong, O. Tang, and C. A. Pollock, “Notch mediated epithelial to mesenchymal transformation is associated with increased expression of the Snail transcription factor,” *International Journal of Biochemistry and Cell Biology*, vol. 42, no. 7, pp. 1115–1122, 2010.

[28] P. Bertheau, E. Turpin, D. S. Rickman et al., “Exquisite sensitivity of TP53 mutant and basal breast cancers to a dose-dense epirubicin—cyclophosphamide regimen,” *PLoS Medicine*, vol. 4, no. 3, article e90, 2007.

[29] K. Rennstam, N. McMichael, P. Berglund et al., “Numb protein expression correlates with a basal-like phenotype and cancer stem cell markers in primary breast cancer,” *Breast Cancer Research and Treatment*, vol. 122, no. 2, pp. 315–324, 2010.

[30] Y. Sun, X. R. Zeng, L. Wenger, G. S. Firestein, and H. S. Cheung, “p53 down-regulates matrix metalloproteinase-1 by targeting the communications between AP-1 and the basal transcription complex,” *Journal of Cellular Biochemistry*, vol. 92, no. 2, pp. 258–269, 2004.

[31] S. Cheng, M. Tada, Y. Hida et al., “High MMP-1 mRNA expression is a risk factor for disease-free and overall survivals in patients with invasive breast carcinoma,” *Journal of Surgical Research*, vol. 146, no. 1, pp. 104–109, 2008.

[32] W. A. Yedall, C. A. Vaughan, H. Miyazaki et al., “Gain-of-function mutant p53 upregulates CXC chemokines and enhances cell migration,” *Carcinogenesis*, vol. 33, pp. 442–451, 2012.

[33] S. A. Mehta, K. W. Christopherson, P. Bhat-Nakshatri et al., “Negative regulation of chemokine receptor CXCR4 by tumor suppressor p53 in breast cancer cells: implications of p53 mutation or isoform expression on breast cancer cell invasion,” *Oncogene*, vol. 26, no. 23, pp. 3329–3337, 2007.

[34] S. Liekens, D. Schols, and S. Hatse, “CXCL12-CXCR4 axis in angiogenesis, metastasis and stem cell mobilization,” *Current Pharmaceutical Design*, vol. 16, no. 35, pp. 3903–3920, 2010.

[35] Q. D. Chu, N. T. Holm, P. Madumere, L. W. Johnson, F. Abreo, and B. D. L. Li, “Chemokine receptor CXCR4 overexpression predicts recurrence for hormone receptor-positive, node-negative breast cancer patients,” *Surgery*, vol. 149, no. 2, pp. 193–199, 2011.

[36] A. M. Pietersen, H. M. Horlings, M. Hauptmann et al., “EZH2 and BMI1 inversely correlate with prognosis and TP53 mutation in breast cancer,” *Breast Cancer Research*, vol. 10, no. 6, article R109, 2008.