

# Candidate Luminal B Breast Cancer Genes Identified by Genome, Gene Expression and DNA Methylation Profiling

Stéphanie Cornen<sup>1,9</sup>, Arnaud Guille<sup>1,9</sup>, José Adélaïde<sup>1,9</sup>, Lynda Addou-Klouche<sup>1,2</sup>, Pascal Finetti<sup>1</sup>, Marie-Rose Saade<sup>1</sup>, Marwa Manai<sup>1,3</sup>, Nadine Carbuca<sup>1</sup>, Ismahane Bekhouche<sup>1</sup>, Anne Letessier<sup>1</sup>, Stéphane Raynaud<sup>†1</sup>, Emmanuelle Charafe-Jauffret<sup>1,4,5</sup>, Jocelyne Jacquemier<sup>1,4</sup>, Salvatore Spicuglia<sup>5,6</sup>, Hugues de The<sup>7</sup>, Patrice Viens<sup>5,8</sup>, François Bertucci<sup>1,5,8</sup>, Daniel Birnbaum<sup>1</sup>, Max Chaffanet<sup>1\*</sup>

**1** Marseille Cancer Research Center, UMR1068 Inserm, Institut Paoli-Calmettes (IPC), Department of Molecular Oncology, Equipe labellisée Ligue Contre le Cancer, Marseille, France, **2** Biotoxicology Laboratory, Djillali Liabes University, Sidi-Bel-Abbès, Algeria, **3** Biochemistry and Molecular Biology Unit, Biology Dpt, Faculty of Sciences El Manar, Tunis, Tunisia, **4** Institut Paoli-Calmettes (IPC), Department of Biopathology, Marseille, France, **5** Aix-Marseille Université, Marseille, France, **6** Inserm U1090, TAGC, Marseille, France, **7** UMR 944/7212 INSERM/CNRS/Université Paris 7, Hôpital St Louis, Paris, France, **8** Institut Paoli-Calmettes (IPC), Department of Medical Oncology, Marseille, France

## Abstract

Breast cancers (BCs) of the luminal B subtype are estrogen receptor-positive (ER+), highly proliferative, resistant to standard therapies and have a poor prognosis. To better understand this subtype we compared DNA copy number aberrations (CNAs), DNA promoter methylation, gene expression profiles, and somatic mutations in nine selected genes, in 32 luminal B tumors with those observed in 156 BCs of the other molecular subtypes. Frequent CNAs included 8p11-p12 and 11q13.1-q13.2 amplifications, 7q11.22-q34, 8q21.12-q24.23, 12p12.3-p13.1, 12q13.11-q24.11, 14q21.1-q23.1, 17q11.1-q25.1, 20q11.23-q13.33 gains and 6q14.1-q24.2, 9p21.3-p24.3, 9q21.2, 18p11.31-p11.32 losses. A total of 237 and 101 luminal B-specific candidate oncogenes and tumor suppressor genes (TSGs) presented a deregulated expression in relation with their CNAs, including 11 genes previously reported associated with endocrine resistance. Interestingly, 88% of the potential TSGs are located within chromosome arm 6q, and seven candidate oncogenes are potential therapeutic targets. A total of 100 candidate oncogenes were validated in a public series of 5,765 BCs and the overexpression of 67 of these was associated with poor survival in luminal tumors. Twenty-four genes presented a deregulated expression in relation with a high DNA methylation level. *FOXO3*, *PIK3CA* and *TP53* were the most frequent mutated genes among the nine tested. In a meta-analysis of next-generation sequencing data in 875 BCs, *KCNB2* mutations were associated with luminal B cases while candidate TSGs *MDN1* (6q15) and *UTRN* (6q24), were mutated in this subtype. In conclusion, we have reported luminal B candidate genes that may play a role in the development and/or hormone resistance of this aggressive subtype.

**Citation:** Cornen S, Guille A, Adélaïde J, Addou-Klouche L, Finetti P, et al. (2014) Candidate Luminal B Breast Cancer Genes Identified by Genome, Gene Expression and DNA Methylation Profiling. PLoS ONE 9(1): e81843. doi:10.1371/journal.pone.0081843

**Editor:** Chad Creighton, Baylor College of Medicine, United States of America

**Received:** June 12, 2013; **Accepted:** October 17, 2013; **Published:** January 9, 2014

**Copyright:** © 2014 Cornen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Financial support for this study came from Inserm, Institut Paoli-Calmettes, and grants from Ligue Nationale Contre le Cancer (Label DB), GEFLUC (MC), Institut National du Cancer (ACI2007, AO2008, AO/R-TRANS09 «identification, caractérisation et validation clinique de nouvelles cibles cellulaires et moléculaires dans les cancers du sein de mauvais pronostic» IVOIRES (INCA)) and DGOS (SIRIC label: grant INCa-DGOS-Inserm 6038). LAK and IB were supported by a fellowship from the Franco-Algerian program of high level formation in France (PROFAS) and the Algerian government, respectively. MM was supported by a fellowship from the AVEROES European program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: chaffanetm@ipc.unicancer.fr

‡ These authors contributed equally to this work.

† Deceased.

## Introduction

Breast cancer (BC) is a complex and heterogeneous disease whose therapeutic approach must be refined in view of recent studies allowing better classification and/or prognosis assessment [1]. DNA microarray-based expression profiling has identified clinically and biologically relevant intrinsic molecular subtypes (luminal A, luminal B, ERBB2, basal, and normal-like) [2–4] and prognostic and/or predictive gene expression signatures [5]. Genomic studies have also suggested a prognostic impact of genomic data [6–8]. Combining expression and genomic data

allowed the identification of candidate BC genes [6,9–19]. The status of DNA methylation may also contribute to improve BC molecular classification [20–24]. The identification of new fusion genes by RNA-seq approaches [25,26] and of driver mutations in cancer genes in various molecular and clinical BC entities [27–31] will also help design targeted treatments.

Luminal B BCs have a poor prognosis [32]. Although they express hormone receptors, their metastatic risk and resistance to hormone therapy and to conventional chemotherapy demand to develop appropriate therapies. Some proteins (e.g. *CITED2*, *NCOR2*) or molecular networks associated with *BCAR* (breast

cancer anti-estrogen resistance) genes are involved in the resistance to anti-estrogen therapy and in the progression of these cancers [32–35]. Luminal B cancers exhibit various mutated genes, *TP53* and *PIK3CA* being the most frequent (29% each) [28].

To further define molecular alterations associated with the luminal B subtype we studied DNA copy number aberrations (CNAs), DNA promoter methylation alterations (DPMAs), gene expression deregulation (EXP), and selected gene mutations in 188 primary BC samples. These analyses identified luminal B-specific candidate genes.

## Materials and Methods

### Ethics statement

The study was approved by our institutional review board: the “Comité d’Orientation Stratégique” of the Institut Paoli Calmettes (IPC) (Marseille, France). Each patient gave a written informed consent for research use.

### Breast cancer samples

Pre-treatment tumor tissues were collected from 188 patients with invasive adenocarcinomas. Patients underwent surgical biopsies or initial surgery at the Institut Paoli-Calmettes between 1987 and 2007. The main histoclinical, biological and subtype characteristics were established for the 188 BCs as described [12,17–19]. They are listed in **Table S1A** and illustrated in **Figure S1**.

### Gene profiling and data analysis

DNA and RNA were extracted as previously described [12,17–19] and controlled on Agilent Bioanalyzer (Agilent Technologies, Massy, France). Genomic profiles of the 188 BCs were established by using array-comparative genomic hybridization (aCGH) onto high-resolution 244K CGH microarrays (Hu-244A, Agilent Technologies, Massy, France). A pool of 13 normal male DNA was used as reference. Gene expression data from the same 188 BCs and 4 normal breast (NB) samples, which represented 1 pool of 4 samples from 4 women, and 3 commercial pools of respectively 1, 2 and 4 normal breast RNA (Clontech, Palo Alto, CA), were obtained using whole-genome DNA microarrays (HG-U133 Plus 2.0, Affymetrix). Both approaches and analysis methods have been used in our previous studies [12,17,18]. All probes for aCGH, gene expression and DNA promoters methylation analyses were mapped according to the hg18/NCBI human genome mapping database (build 36) to homogeneously integrate the data. The aCGH, gene expression, methylation data, as well as the integrated CNA/gene expression and DPMA/gene expression analyses are illustrated by pipelines (**Figures S2 and S3**, respectively).

Validation and prognostic impact of candidate genes were evaluated in a large public series of BC samples. Thirty-six data sets, including a total of 5,765 non-redundant samples, were collected from public database i.e. Gene Expression Omnibus (GEO/NCBI), Array Express (EBI) and authors’ websites (**Table S1B**). Raw data from each study were normalized using quantile normalization, and log<sub>2</sub>-transformed. The intrinsic molecular subtypes of each tumor sets were defined using Hu single sample predictor (SSP) [4]. To be comparable across data sets, each gene expression levels were standardized within each data set using luminal A population as reference.

The data (experiment called “Candidate luminal B breast cancer genes identified by genome, gene expression and DNA methylation profiling”) are publicly available (ArrayExpress repository ref ID: E-MTAB-1861).

### DNA promoter methylation profiling and data analysis

We captured the methylated DNA of 117 (109 tumors+8 NB) samples by using the MethylMiner Methylated DNA Enrichment Kit (Invitrogen). Genome-wide DNA-methylation analysis was done on custom [A-MEXP-2178 (arrayexpress)] human promoter arrays 2×400K (Agilent Technologies, Massy, France) using the MethylMiner Methylated DNA Enrichment Kit (Invitrogen) [36]. Over 414,000 probes cover promoter regions approximately –3 kb to +3 kb relative to transcription start sites (TSSs) with a resolution of 280 bp in average. Scanning was done with Agilent Autofocus Dynamic Scanner (G2565C, Agilent Technologies). Raw data were obtained from Feature extraction 10.7.3 software (Agilent Technologies). Probes not mapped to exact positions in the genome as well as those under the background signal were removed with the control probes. The final dataset contained 326,350 unique probes covering 18,297 promoter regions according to the hg18/NCBI human genome mapping database (build 36). The M (Log<sub>2</sub>Red-Log<sub>2</sub>Green) values of each probe on the array were then obtained and normalized according to their GC content. Then, inter-array quantile normalization was done for the correction of distribution differences among experiments.

To estimate the global methylation level for a given gene in one sample, we computed a methylation score based on the sum of frequency of probes with a M value greater than zero combined with its amplitude and frequency of probes with a M value less than zero combined with its amplitude. Clustering was done with the Cluster program using Pearson correlation as similarity metrics and average linkage clustering. Results were displayed using TreeView program.

### Gene mutations

Polymerase chain reaction (PCR) and direct sequencing were done using standard conditions with gene-specific primers designed to amplify coding sequence of *ARID1A*, *ASXL1*, *FOXO3*, *L3MBTL4*, *MAP2K4*, *PIK3CA*, *RUNX1*, *RUNX3* (**Table S1C**). Most PCR amplifications were done in a total volume of 25 μl PCR mix containing at least 10 ng template DNA, Taq buffer, 200 μmol of each deoxynucleotide triphosphate, 20 pmol of each primer and 1 unit of appropriated Taq polymerase (**Table S1C**). PCR products were purified using Millipore plate MSNU030. The purified PCR products (2 μl) were used for sequencing using the Big Dye terminator v1.1 kit (Applied Biosystems, Courtaboeuf, France). After G50 purification, sequences were loaded on an ABI 3130XL automat (Applied Biosystems). The sequence data files were analyzed using the Seqscape software and all mutations were confirmed on an independent PCR product.

*TP53* mutation status was determined by a yeast functional assay [37,38].

### Statistical analyses

Correlations between sample groups and histoclinical factors were calculated with the Fisher’s exact test for qualitative variables with discrete categories, and the Mann-Whitney test for continuous variables. Metastasis-free survival (MFS) was calculated from the date of diagnosis until the date of first metastatic relapse. Survivals were calculated using the Kaplan-Meier method and curves were compared with the log-rank test. Stratification into high-risk and low-risk groups was based on relative risk defined by Cox model using the natural threshold of 1. Univariate and multivariate survival analyses were done using the Cox regression model (Wald test). Variables tested in univariate analyses included patients’ age at time of diagnosis ( $\leq 50$  years *vs*  $> 50$ ), pathological tumor size (pT: pT1 *vs* pT2-3), pathological axillary lymph node status (pN: negative *vs* positive), pathological grade (I *vs* 2–3),

histological type, delivery of hormone therapy and chemotherapy, immunohistochemical (IHC) ERBB2 status (negative *vs* positive), molecular subtypes and *RECQL4* expression (continuous value). Variables with a *p*-value < 0.05 in univariate analysis were tested in multivariate analysis. All statistical tests were two-sided at the 5% level of significance. Analyses were done using R software (2.14.2) and associated packages. We followed the reporting REcommendations for tumor MARKer prognostic studies (REMARK criteria) [39].

Luminal B CNAs landscape was drawn with the Circos software [40] and significant mutual exclusive and co-occurring CNAs (FDR < 0.05) were identified by DRP analysis [41].

In the meta-analysis of six recent NGS studies including 875 breast tumors, co-occurring and mutually exclusive gene mutations were identified by using a method previously reported [42].

## Results

### Genomic characterization of 188 BC samples

We first describe the results on the whole set of 188 tumors before addressing the specific question of the luminal B cases. The 188 samples were first profiled using whole-genome gene expression microarrays. **Figure S1** shows the hierarchical clustering of samples based on the expression of 13,031 probe sets. Samples were sorted into four major clusters, which strongly correlated with histoclinical features (grade, IHC data) and molecular subtypes. The 188 cases (**Table S1A**) included 54 basal, 64 luminal A, 32 luminal B, 16 ERBB2 and 22 normal-like cases [4] (**Table S1D**).

High-resolution aCGH profiles were established for the 188 samples. High and low level CNAs (amplifications or homozygous losses and gains or hemizygous losses, respectively) were identified. As previously reported [8], the three most frequently gained regions were on the 1q, 8q and 17q chromosomal arms whereas the most frequently lost regions were on 8p, 11q and 16q. The median percentage of probe sets displaying a CNA in a sample was 41%, with a great variability between samples (range 0.05–80%). As expected, this percentage was higher in grade 3 tumors (52%) than in grade 1 tumors (17%;  $p = 1.92 \times 10^{-7}$ ; Mann-Whitney test). A total of 52 cases (28%) were “simplex”, 81 (43%) “complex sawtooth”, and 55 (29%) “complex firestorm” (**Table S1A**). As expected, only 11% of grade 3 tumors were “simplex”, whereas 89% were “complex” and conversely, 53% of grade 1 tumors were “simplex” and 47% “complex”. The highest proportion of simplex patterns was observed in luminal A tumors (58%) whereas the highest proportion of complex patterns was found in ERBB2 (100%), luminal B and basal (91% for both) cases (**Fig. S4**).

### Copy number aberrations in luminal B tumors

We next focused on the CNAs found in luminal B tumors. The luminal B subtype shows both common and specific alterations [7,9,16,19]. To identify the latter we did a supervised analysis comparing CNAs observed in luminal B to those found in each of the other subtypes (**Fig. S5** and **Tables S2A–F**).

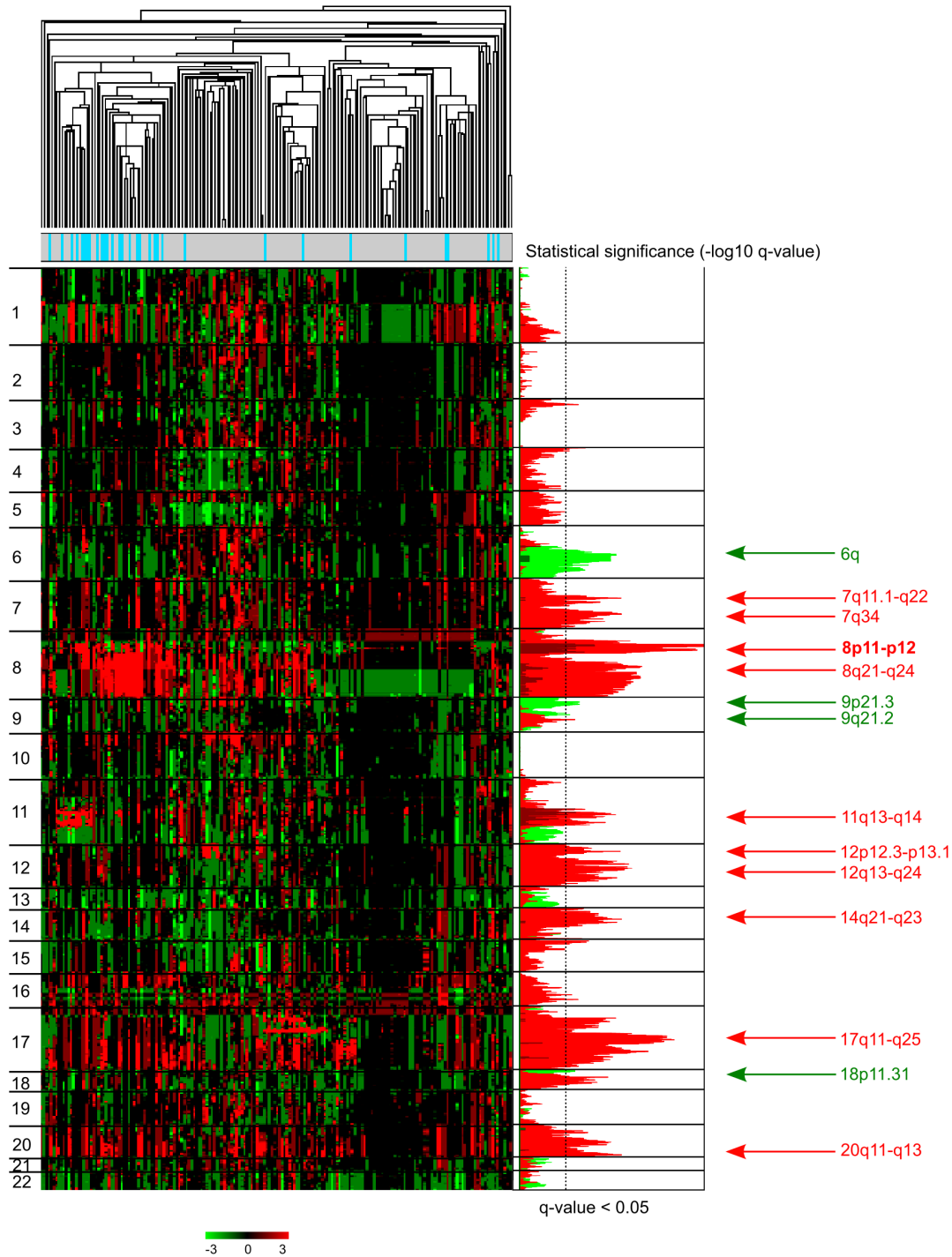
The luminal B/luminal A comparison showed that 8p11-12 amplification and gains of distinct regions (i.e. 7p11.2-22.1, 7q11.21-36.3, 8p11-12, 8q, 11q13-14, 12, 14q12-23, 17q11-25.3, 18q12.1, 20p11.21-12.3 and 20q with a frequency  $\geq 30\%$ ) are more frequent in luminal B samples (Fisher's exact test; FDR < 0.05) (**Fig. S5**, **Tables S2A, B**). CNAs associated with gains or amplification targeted a total of 3,364 genes (**Table S2B**). Genes amplified in luminal B included the 8p12 genes *ZNF703*, *SPFH2*, *PROSC*, *GPR124*, *BRF2*, *RAB11FIP1*, *GOT1L1*, *ADRB3*, *EIF4EBP1*, *ASH2L*, *STAR* and *LSMI*.

The luminal B/basal comparison showed that gains and losses of distinct regions (including 5q11.1, 8p11-12, 8q13.2-21.3, 11q13.3, 12q12-24.23, 14q21.3-24.23, 16p11.2-13.3, 17q, 20q11.23-q13.33 and 6q, 9p22-p24, 13q34, and 18p11.31-11.32 regional gains and losses with a frequency  $\geq 30\%$ , respectively) are more frequent in luminal B samples (Fisher's exact test; FDR < 0.05) (**Fig. S5**, **Tables S2C, D**). Gains and losses targeted 1,091 and 3,339 genes, respectively (**Table S2D**). The comparison with the ERBB2 subtype did not distinguish regions with a different CNA frequency, perhaps because of the low number of samples.

The luminal B/non-luminal B comparison (non-luminal B includes luminal A, ERBB2, basal, normal-like) (**Fig. 1**) showed that 8p11-p12 and 11q13.1-q13.4 amplifications, gains of distinct regions (including 7q11.22-q34, 8p11.21-p12, 8q21.12-q24.23, 11q13.3-q14.1, 12p12.3-p13.1, 12q13.11-q24.11, 14q21.1-q23.1, 17q11.1-q25.1, 20q11.23-q13.33 gains and 6q14.1-q24.2, 9p21.3-p24.3, 9q21.2, 18p11.31-p11.32 losses with a frequency  $\geq 30\%$ , respectively) are more frequent in luminal B samples (Fisher's exact test; FDR < 0.05) (**Fig. 1**, **Tables S2E, F**). Amplification, gains and losses targeted 122, 2,541 and 277 genes, respectively (**Table S2F**). The most significant amplified genes associated with luminal B included again the 8p12 genes *ZNF703*, *SPFH2*, *PROSC*, *GPR124*, *BRF2*, *RAB11FIP1*, *GOT1L1*, *ADRB3*, *EIF4EBP1*, *ASH2L*, *STAR* and *LSMI* (34%; Fisher test,  $p < 9.5 \times 10^{-6}$ , FDR <  $8.1 \times 10^{-3}$ ) and the 11q13 genes *CCND1*, *ORAOV1*, *FGF19*, *FGF4*, *FGF3* (38%; Fisher test,  $p < 1 \times 10^{-4}$ , FDR < 0.05) (**Table S2E**). Many of the 277 lost genes were located in chromosome arm 6q with a frequency comprised between 50% and 66% (**Table S2E**). The most significant losses were associated with *C6ORF167/MMS22L* (6q16.1) and *MCHR2*, *SIMI*, *ASCC3* (6q16.3) (66%; Fisher test,  $p = 7.75 \times 10^{-7}$ , FDR =  $1.93 \times 10^{-3}$ ). They were 3.3 times more frequent in luminal B than in non-luminal B tumors. Genomic profiles showed various 6q regional losses as well as rare homozygous deletions and small deleted regions targeting *MLLT4*, *ARID1B*, *PARK2*, *FOXO3*, *UFL1/NLBP*, *ASCC3* genes (**Fig. S6**), suggesting the existence of several potential TSGs within 6q. Chromosomal regions 9p21.3-24.3 and 9q21.2 were targeted by losses with a frequency comprised between 50% and 56%, respectively (**Fig. S7** and **Table S2E**). These events were twice more frequent in luminal B than in non-luminal B tumors (56%, Fisher test,  $p < 1 \times 10^{-3}$ , FDR < 0.05). *MTAP*, *CDKN2A* and *CDKN2B* were the most frequently 9p deleted genes in luminal B samples. Chromosomal region 18p11.31-11.32 was targeted by copy number losses with a frequency comprised between 53% and 56% (**Table S2E**). These losses were almost twice more frequent in luminal B than in non-luminal B tumors (56%, Fisher test,  $p < 2 \times 10^{-3}$ , FDR < 0.05). Noticeably, genes coding for TP53 repressor *PRDM1* (6q21) and TP53 effector *PERP* (6q23.3) were among the genes with loss frequencies associated with the luminal B subtype (**Table S2E**). Conversely, gains of *TP53INP1* (8q22.1) (81%, Fisher test,  $p < 9 \times 10^{-4}$ , FDR < 0.05), *TP53I13* (17q11.2) (50%, Fisher test,  $p < 5 \times 10^{-4}$ , FDR < 0.05), *TP53INP2* (20q11.22) (56%, Fisher test,  $p < 5 \times 10^{-3}$ , FDR < 0.05) and *TP53RK* (20q13.12) (56%, Fisher test,  $p < 1 \times 10^{-2}$ , FDR < 0.05) were associated with the luminal B subtype (**Table S2E**). This suggests that a TP53-associated pathway may play a role in luminal B tumors.

### Integrated comparative analysis and luminal B candidate genes

We compared the degree of CNA-driven RNA up or downregulation in 32 luminal B (i) vs 64 luminal A, (ii) vs 54 basal, and (iii) vs 156 pooled non-luminal B tumors, by analyzing the 13,127 genes common to the two platforms (aCGH Agilent



**Figure 1. Specific regions targeted by CNAs in luminal B BCs.** Genomic profiles were established for 188 breast tumors (32 luminal B and 156 non luminal B). On the left is shown a hierarchical clustering of genome copy number profiles measured by aCGH on 24,907 probes or groups of probes (without X and Y). Red indicates increased copy number and green indicates decreased copy number. To the left are indicated chromosome locations with chromosome 1pter to the top and 22qter to the bottom. Next on the right, significant copy number amplifications (dark red), gains (red) and losses (green) observed in luminal B compared to non-luminal B tumors (Fisher's exact test), are plotted as a function of chromosome location. Only amplification, gains, and losses associated with luminal B tumors are shown (Fisher's exact test; FDR < 0.05) for each chromosome. In addition to the previously reported 8p11-12, 11q13 amplifications, 17q, 20q gains and 18p losses we previously reported (16,20), other luminal B CNAs include 7q11.22, 7q34, 8p11.21-p12, 8q21.12-q24.23, 11q13.3-q14.1, 12p12.3-p13.1, 12q13.11-q24.11, 14q21.1-q23.1, 17q11.1-q25.1, 20q11.23-q13.33 gains and 6q14.1-q24.2, 9p21.3-p24.3, 9q21.2, 18p11.31-p11.32 losses.  
doi:10.1371/journal.pone.0081843.g001

and Hu233 2.0 plus Affymetrix) and retained after filtering based on the expression level. From these supervised analyses (**Tables S2A–F**), 160, 148 and 307 genes had an expression level that varied according to CNA (Mann-Whitney;  $FDR < 0.05$ ) after luminal B/luminal A (**Table S3A**), luminal B/basal (**Table S3C**), and luminal B/non-luminal B (**Table S3E**) comparison, respectively. Of these, 138, 132, and 251 genes had a deregulated gene expression in relation with CNA specifically associated with the luminal B subtype (t test,  $FDR < 0.05$ ) after luminal B/luminal A (**Table S3B**), luminal B/basal (**Table S3D**), and luminal B/non-luminal B (**Table S3F**) comparison, respectively. These specific luminal B candidate genes were qualified as potential oncogenes or TSGs if they presented up or downregulated expression in relation with amplification/gain or losses, respectively, as previously described [12,17,18]. Overall, out of the 337 candidates identified, 66% (221) were found in the TCGA data [28], deregulated in relation with their copy number alterations. (**Tables S3B, 3D and S3F**, respectively).

Among the 251 luminal B candidate genes of the luminal B/non-luminal B comparison (**Table S3F**), 189 and 62 were potential oncogenes and TSGs, respectively. By order of significance, *ZNF703*, *CCND1*, *ORAOV1*, *BRF2*, *RAB11FIP1*, *LSM1*, *PPAPDC1B*, *ASH2L*, *DDHD2*, *AP3M2*, *VDAC3*, *FGFR1*, *EIF4EBP1*, *MYST3*, and *TM2D2* were overexpressed in relation with their amplification (**Fig. S8**). Among the 62 TSG candidates, 53, 6, 2 and 1 were located on 6q (**Fig. 2**), 9p21.3-p24.3, 18p11.31 and 9q21.2 regions, respectively.

A total of 39 luminal B candidate oncogenes were identified from luminal B/normal-like comparisons (**Table S3G**). No luminal B candidate was identified from the luminal B/ERBB2 comparison.

Among luminal B candidates defined from the different comparisons, 7 (*BRCA1*, *CCND1*, *COX6C*, *EZH2*, *FGFR1*, *MSI2*, *RECQL4*) code for potentially druggable proteins [43], and 11 (*BIRC5*, *CITED2*, *FAM82B*, *FOXMI*, *GPR172A*, *NHERF1*, *RECQL4*, *SLC39A4*, *SQLE*, *UBE2C*, *YWHAZ*) are associated with endocrine resistance (**Table S3G**) [33,34,44].

A total of 34, 47 and 2 candidate genes were identified as specific of luminal B tumors when compared to luminal A, basal, and normal-like tumors (**Tables S3H–J**), respectively. These different repertoires of candidate genes could help identify pathways and mechanisms either specifically affected in luminal B tumors or shared by the major subtypes.

Interestingly, 101 luminal B candidate oncogenes were common to both luminal B *vs* luminal A and luminal B *vs* non-luminal B comparisons (**Table S3K**). This “core” of luminal B oncogenes was mainly associated with the cell cycle and proliferation ( $p < 0.05$ ;  $FDR < 0.25$ ) (**Table S3L**). They included 14, 4, 37, 8, 23 and 10 genes from amplified 8p11-p12, 11q13, and gained 8q13-q24.3, 12q, 17q11.2-q25.1 and 20q11-q13 regions. Among them, *AP3M2*, *ASH2L*, *BRF2*, *DDHD2*, *FGFR1*, *LETM2*, *LSM1*, *PPAPDC1B*, *RAB11FIP1*, *ZNF703* (8p12), *TPD52* (8q21.13), and *DCAF7/WDR68* (17q23.3) have been described as potential BC genes [6,12,19,45–47]. In a large public series of 5,765 BCs, 100 (99%) of the candidate oncogenes included in this “luminal B core” were validated at the gene expression level in both comparisons (**Table S3M**). The overexpression of 67 of them was associated with poor metastatic-free survival (MFS) in luminal tumors ( $p < 0.05$ ;  $FDR < 0.25$ ). While several data show that loss of PR expression is an important predictor of poor patient outcome in ER+ BCs [48,49], 94% (61/67) of our luminal B candidates were pertinent predictor independently of the PR status (**Table S3M**).

*RECQL4* was the most significant. Corresponding Kaplan-Meier MFS curves are shown in **Figure 3A** ( $p = 3.80 \times 10^{-9}$ ). Interestingly, concordant results were obtained with another referent molecular classifier i.e. PAM50 SSP [50] (**Table S3M**). Indeed, the definition of luminal B tumors is highly variable due to their molecular heterogeneity, making their assignment in the luminal B subtype non-reproducible by different gene expression signatures [51]. This suggests that the robustness of our candidates is not impacted by the choice of the classifier [4,50].

A high *RECQL4* gene expression level has recently been reported to confer proliferation advantage and survival to breast cancer cells [50]. We observed that *RECQL4* overexpression was associated with *MDM2* overexpression (t-test  $p = 2.75 \times 10^{-2}$ ) as well as with mutated *TP53* or/and overexpressed *MDM2* gene status (t-test  $p = 1.73 \times 10^{-20}$ ) in the independent public TCGA data set [28] (data not shown). This might highlight high genomic instability and DNA repair perturbations in such tumors.

In multivariate analysis including the other histoclinical prognostic features, *RECQL4* expression remained significant for MFS ( $p = 3.4 \times 10^{-2}$ ), with pT and molecular subtypes (**Table 1**). The clinical response study in regard to *RECQL4* gene expression in 924 BCs (included in the large public series, see **Table S1B**) showed as expected that luminal B tumors have a better pathological response to chemotherapy (Fisher,  $p = 3.04 \times 10^{-4}$ ) than luminal A tumors. Moreover, *RECQL4* overexpression was associated with a better pathological response in the group including both luminal A and B tumors (Fisher,  $p = 0.013$ ) (**Figure 3B**) but not within separated luminal subgroups (data not shown).

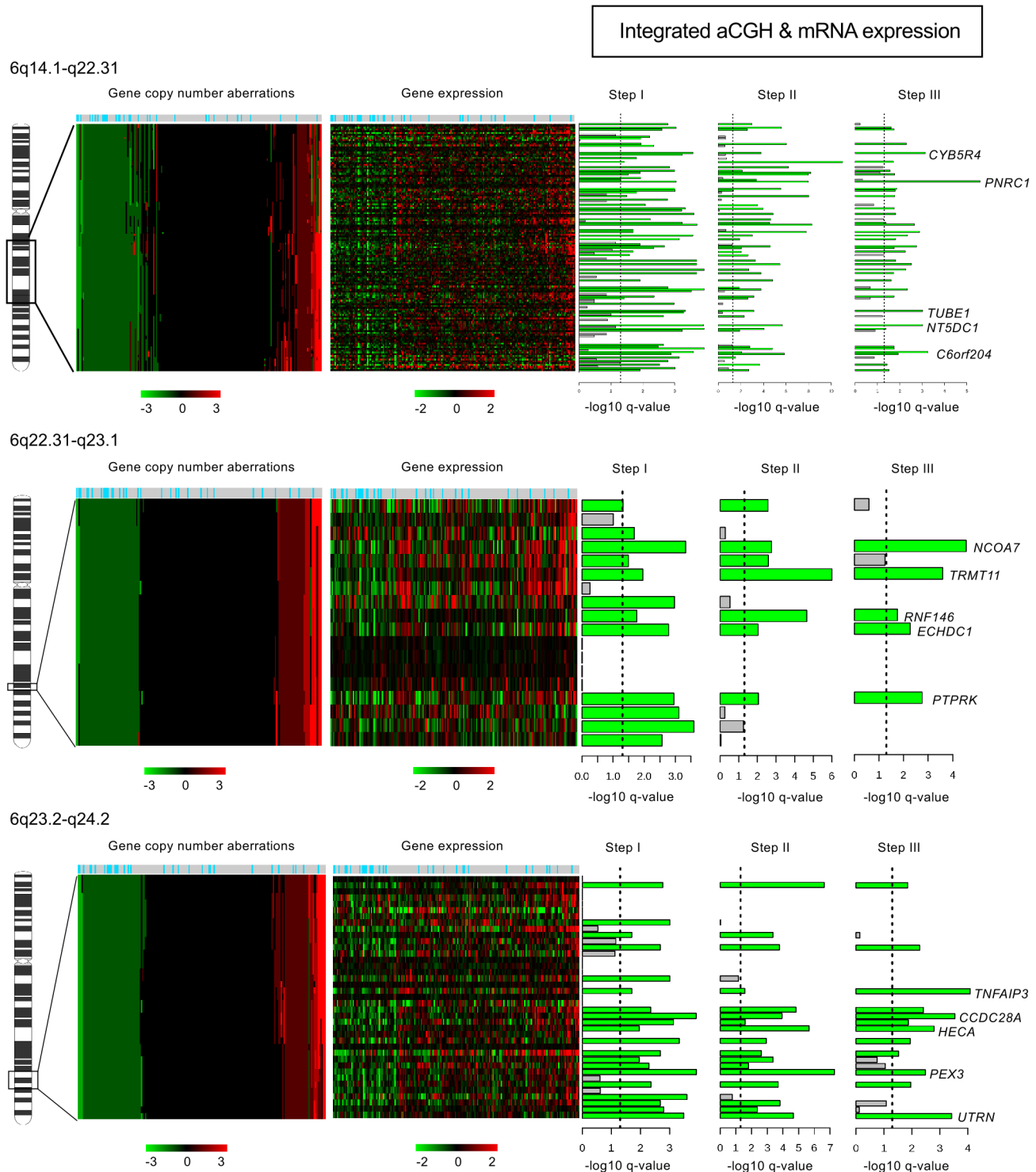
We further identified 4 mutually exclusive CNA pairs (*LYZ* gains/6q25-qter losses, *RUNX1* loss/6q22.1-q24.3 losses, *RUNX1* loss/*RUNX3* loss and *RUNX1* loss/6q14.1-q21 losses) and 215 co-occurring CNAs in luminal B ( $FDR < 0.05$ ) (**Fig. S9, Table S3N**). The mutually exclusive CNA pairs suggest that these altered genes and/or regions could participate in the same pathways.

## DNA methylation profiles of 109 BCs

The hierarchical clustering established with the most variant methylation scores observed in 5,492 promoters ( $SD > 0.3$ ) classified the 109 BC and 8 NB samples into clusters associated with different DNA methylation patterns (**Fig. 4A**) variably associated with ER and PR expression ( $p = 5.6 \times 10^{-7}$  and  $p = 2.26 \times 10^{-7}$ ), SBR grade, molecular subtype ( $p = 5 \times 10^{-4}$ ), and *TP53* status ( $p = 1.2 \times 10^{-3}$ ) (**Table S4A**). The 8 NB samples and most of the ER- tumors were included in cluster I while ER+ tumors were mainly distributed in cluster II. Most of the basal tumors were included in cluster I whereas cluster II contained most of the luminal A and luminal B tumors. ERBB2 tumors were distributed in the two clusters probably because ERBB2 tumors are heterogeneous for ER status [17]. These different profiles were consistent with recently reported BC methylation patterns [21–23], which pointed to a possible relationship between DNA methylation and ER status.

## Validation of methylation results

The supervised analysis comparing methylation score data of ER+ and ER- tumors identified 3,484 gene promoters with methylation differences between the two groups (**Table S4B**) (t test,  $FDR < 0.05$ ). Among them, 1,753 gene promoters (including those of *APC*, *CAVI*, *CCND2*, *CDC47*, *CDH3*, *CDKN2A*, *CDKN2B*, *HEY2*, *RASSF1*, *RECK*) had a DNA methylation level higher in the ER+ group (t-test,  $FDR < 0.05$ ). Conversely, 1,731 gene promoters (including those of *BCL2*, *ESR1*, *HSD17B4*, *PISD*, *WNK4*) had a higher level of methylation in the ER- group ( $FDR < 0.05$ ). These



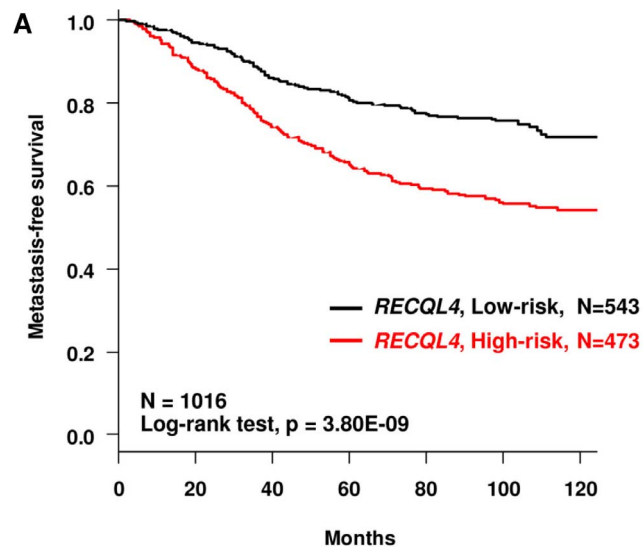
**Figure 2. Correlation between gene expression and genome alterations on the 6q regions.** Genomic and gene expression profiles were established for 188 breast tumors (32 luminal B and 156 non luminal B identified at the top by blue and grey boxes, respectively) in three 6q regions: 6q14.1-q22.31 (top), 6q22.31-q23.1 (middle) and 6q23.2-q24.2 (bottom). For each region, heatmaps for genome copy number and gene expression profiles are consecutively drawn. Genome copy number was measured by aCGH on probes or groups of probes spanning each of these regions. Red indicates increased copy number and green indicates decreased copy number. In the heatmap tumors are organized from the tumor that presented the most copy number losses to the tumor that exhibited the most copy number gains. The next heatmap was established with the expression of the independent genes located on the corresponding 6q region and profiled in the same 188 tumors similarly organized. For gene copy number and gene expression heatmaps, we used color scale limits from  $-3$  to  $+3$  and  $-2$  to  $+2$ , respectively. Next to the right, are plotted genes successively selected by steps I, II and III of the integrated analysis "aCGH & mRNA expression" as defined by the work pipeline (Figure S2). Grey and green lines correspond to rejected and selected genes, respectively. Among genes with an expression level that varied according to CNAs, we retained genes showed significant differences (vertical line) in copy number loss correlated with downregulated expression in luminal B compared to non-luminal B

tumors. They were qualified as potential TSGs. For each region, only the first five most significant are listed. *PNRC1*, *NCOA7* and *TNFAIP3* genes were the most significant candidate TSGs for the 6q14.1-q22.31 (top), 6q22.31-q23.1 (middle) and 6q23.2-q24.2 (bottom) regions, respectively. doi:10.1371/journal.pone.0081843.g002

data are in agreement with those reported on Illumina® platform [20,21,23,24]. The strong correlation between DNA methylation data of *RASSF1* gene promoter established by EpiTyper and promoter array approaches further validated our method (Results and References S1, Fig. S10, S11).

### DNA methylation associated with breast cancer molecular subtypes

Using the ANOVA method, we identified 4,545 gene promoters with a DNA methylation level different in at least one subtype among the five major ones (Table S4C). They included several genes previously reported differentially methylated in BC subtypes and particularly in luminal and basal (or ER+ and ER-) tumors (Table S4C). The median methylation levels of these 4,545 subtype-associated genes were highest in the luminal and ERBB2 subtypes and lowest in the basal subtype ( $p = 6.24 \times 10^{-12}$ ) (Fig. 4B).



**Figure 3. Kaplan-Meier MFS curves and pathological clinical response in luminal BCs according to *RECQL4* mRNA expression.** **A** -Kaplan-Meier survival curves were drawn according to *RECQL4* gene expression status of 1,016 luminal BCs established from a large public series of 5,765 BCs (Table S1B). Stratification into high-risk (red curve) and low-risk (black curve) groups were based on relative risk defined by the Cox model using the natural threshold of 1. *RECQL4* gene expression is associated with MFS ( $p = 3.8010^{-9}$ ). **B** - From the same large public series, 924 BCs were informed for the pathological response (Table S1B). The *RECQL4* overexpression was associated with a better pathological response within the group including both luminal A and B tumors (N = 435) (Fisher,  $p = 0.013$ ). doi:10.1371/journal.pone.0081843.g003

Tukey's range test integrating ten supervised analyses was applied to distinguish gene promoters exhibiting a DNA methylation level different in only one subtype as compared to the others ( $FDR \leq 0.05$ ) (Table S4C). A total of 375 and 431 promoters had a methylation level higher and lower in the luminal B tumors than in the other subtypes, respectively. Among them, 265 and 295 had a methylation level higher and lower than in NB tissues (t test,  $FDR < 0.05$ ). This analysis was extended to the other subtypes (Results and References S1).

### Molecular subtypes and specific deregulated gene expression in relation with the DNA methylation level variation

Among the 4,545 promoters with a DNA methylation level different in at least one subtype (Table S4C), only 459 genes presented an associated mRNA deregulation (correlation  $< -0.40$ ) (Table S4D). These genes are listed with their CNA status in Table S4E. The presence of CpG islands was observed in 62.5% of them. In the luminal B cases, among the 560 promoters associated with a significant DNA methylation variation, 46 corresponding genes presented a deregulated expression (24 and 22 were down and upregulated, respectively) (correlation  $< -0.40$ ) (Table S4E). 52% (24 genes) were found significantly deregulated in relation with their level of DNA methylation in the TCGA data [28] (Table S4E).

Among them, high DNA methylation level targeted promoters of *ASS1*, *CITED4*, *DCR1*, *FAM78A*, *FBXO32*, *SAMD9L*, *SP100*, *STAT5A* and *ZFP36L2* genes, previously reported as TSGs or associated with tumor progression (Table S4E). Only lower *ASS1*, *C6ORF145/PXDC1*, *ZFP36L2* and higher *C3ORF67*, *C12ORF60*, *H2AF7*, *RAB11FIP1* gene expression was observed in luminal B compared to the other subtypes (t test,  $p < 0.05$ ) (Fig. 4C and S12A). The other subtypes were also analyzed (Results and References S1, Table S4E, Fig. S12B–S12D1–D4). Overall, out of the 168 candidates identified in our analysis, 54% (90 candidate genes) were found in the TCGA data [28] significantly deregulated in relation with their level of DNA methylation (Table S4E).

### Subtype-specific candidates presenting gene expression deregulation in relation with CNA and with DNA methylation aberrations

We integrated genomic, gene expression and DNA methylation profiles to identify subtype-specific candidate genes presenting expression deregulation in relation with CNA and with DNA promoter methylation aberrations (Table S4F).

In the luminal B cases, no gene was downregulated in relation with both copy number loss and with high methylation (Table S4F). Conversely, 8 genes were upregulated in relation with copy number gain or amplification and with low DNA methylation (Table S4F). *C12ORF60* and two previously reported oncogenes, *H2AF7* and *RAB11FIP1/RCP* were the most significantly upregulated (t-test,  $p < 0.05$ ) (Fig. S12A). Except for *TPD52* and *C12orf60*, 75% were upregulated in relation with copy number gain or amplification and with low DNA methylation in the TCGA data [28]. The other subtypes were also analyzed (Results and References S1, Table S4F, Fig. S12B–S12D1–D4).

**Table 1.** Uni- and multivariate analysis of MFS in the luminal BC public data.

		Univariate			Multivariate		
		N	HR [95CI]	p	N	HR [95CI]	p
Age	>50 vs ≤ 50 y	512	1.03 [0.73–1.46]	0,87			
<b>pT</b>	pT2-3 vs pT1	556	1.76 [1.25–2.48]	<b>1,30E-03</b>	509	1.51 [1.05–2.18]	<b>2,65E-02</b>
<b>pN</b>	pos vs neg	829	1.21 [0.9–1.63]	0,22			
<b>SBR grade</b>	2–3 vs 1	650	2.72 [1.69–4.4]	<b>4,22E-05</b>	509	1.39 [0.82–2.34]	0,22
<b>Histology**</b>	ILC vs IDC	128	1.45 [0.59–3.56]	0,41			
	other vs IDC	128	0.78 [0.18–3.29]	0,73			
<b>Hormone therapy</b>	Yes vs No	746	0.64 [0.45–0.93]	<b>2,01E-02</b>	509	0.68 [0.41–1.12]	0,13
<b>Chemotherapy</b>	Yes vs No	749	1.08 [0.71–1.66]	0,71			
<b>ERBB2 IHC status</b>	pos vs neg	148	1.19 [0.36–3.91]	0,77			
<b>Molecular subtype*</b>	LumB vs lumA	1016	2.07 [1.67–2.57]	<b>3,99E-11</b>	509	2.53 [1.66–3.87]	<b>1,66E-05</b>
<i>RECQL4</i>		1016	1.55 [1.37–1.75]	<b>1,44E-12</b>	509	1.33 [1.02–1.73]	<b>3,40E-02</b>

\*as previously defined [4].

\*\*IDC: infiltrated ductal carcinomas; ILC: infiltrated lobular carcinomas.

doi:10.1371/journal.pone.0081843.t001

### Mutations in candidate genes

Based on both previous studies and our observations of potential driver CNAs and mutations in luminal B BCs [16,27–31,53] (**Table S5A**), we searched for mutations in *ARID1A*, *ASXL1*, *FOXO3*, *L3MBTL4*, *MAP2K4*, *PIK3CA*, *RUNX1*, *RUNX3* and *TP53* genes in a large set of 390 BCs, including a part of our panel (181/188 BCs) (**Tables S5B–C**). Mutations (including frameshift, nonsense, missense mutations) targeted in decreasing frequency order *TP53* (46.3%), *PIK3CA* (23.3%), *RUNX3* (9.1%), *FOXO3* (8.6%), *RUNX1* (4.7%), *ARID1A* (3.7%), *MAP2K4* (3.3%), *L3MBTL4* (2.5%), and *ASXL1* (2.2%) genes (**Table S5C**). *ARID1A* and *L3MBTL4* mutations were previously reported [16,53]. Mutations in *TP53*, *PIK3CA*, *MAP2K4* and *RUNX1* genes were previously reported as driver mutations [27,29–31]. We verified that the p.A147Gfs\*13 and p.S295Cfs\*304 *RUNX1* and p.S588X *FOXO3* mutations were acquired (**Fig. S13, S14**, respectively). We could not confirm that the other missense mutations were similarly somatic. However, they were not included as SNP or missenses in NCBI db SNP build 131.

As illustrated for *RUNX1* (**Fig. S15**), some genes can be targeted by different alterations (CNAs, mutations) with frequencies that vary in the major subtypes (**Table S5C**). In the luminal B cases, we identified mutations of *FOXO3* (23.8%), *TP53* (23.1%), *PIK3CA* (16.7%), *RUNX3* (8.5%), *ASXL1* (7.1%), *MAP2K4* (6.1%), *RUNX1* (4.9%) and *L3MBTL4* (2.1%) genes. We found that only *FOXO3* mutations were specifically associated with this subtype (Fisher test, FDR<0.05 and odd ratio>1) (**Table S5C**). No *ARID1A* mutation was observed. The mutations existed in the luminal B genomic context including *ZN703*, *MYC*, *CCND1* and *ZN7217* gains/amplifications and *FOXO3*, *CDKN2A*, *L3MBTL4* and *ARID1B* losses (Fisher test, FDR<0.05 and odd ratio>1) (**Table S5C**). Only *RUNX3* and *TP53* mutations were significantly found in luminal A and basal tumors, respectively (t-test, FDR<0.05 and odd ratio>1) (**Table S5C**). *PIK3CA*, *EGFR*, *MYC* gains/amplifications, *MAP3K1* losses and *ERBB2* amplification reflected genomic contexts associated with basal and ERBB2

molecular subtypes, respectively. **Figure 5** illustrates the presence of CNAs and/or mutations targeting some of these genes in each of the breast tumor samples constituting our panel and classified by molecular subtype.

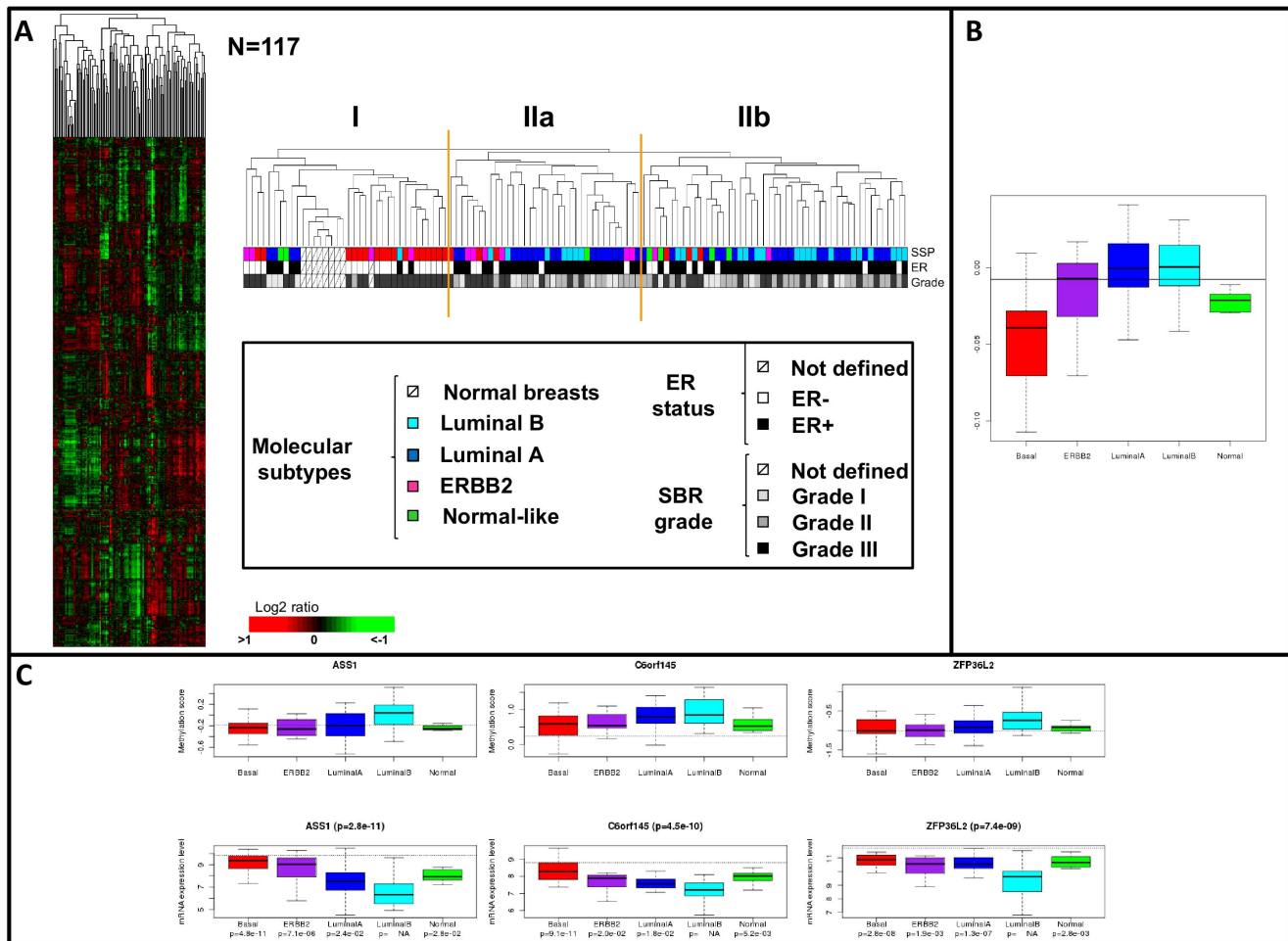
### Meta-analysis of six recent NGS studies in 875 breast tumors

To include our results in a more global context we did a meta-analysis of whole genome/exome sequence data generated from a total of 875 BC samples (“NGS samples”) in six 2012 studies (**Table S6A**). A total of 35,603 somatic mutations (including in/del, nonsense, missense mutations) were identified in 13,793 genes (**Table S6B**). In a decreasing frequency order, *TP53*, *PIK3CA*, *TIN*, *GATA3*, *MLL3*, *MAP3K1*, *MUC16*, *CDH1*, *SYNE1* and *USH2A* were the top ten mutated genes in the NGS samples; the frequencies of somatic mutations in *TP53*, *PIK3CA*, *RUNX1*, *ARID1A*, *MAP2K4*, *L3MBTL4* and *ASXL1* were 34%, 31%, 2.4%, 2.1%, 3.4%, 0.3% and 0.6%, respectively.

To identify co-occurring and mutually exclusive gene mutations, we retained genes targeted by more than 15 mutations in the 875 NGS samples. From 70 such genes, we identified 315 co-occurring and 8 mutually exclusive gene mutations (p<0.05) (**Tables S6C**). Concomitant mutations of *MAP3K1/PIK3CA*, *CDH1/PIK3CA*, *MAP2K4/PIK3CA* were among the most frequently observed (>15 mutations). *MAP3K1/TP53*, *GATA3/TP53*, *CDH1/TP53*, *AKT1/PIK3CA*, *MAP2K4/TP53* and *CDH1/GATA3* mutations were mutually exclusive.

In these NGS studies, only 602 samples were subtyped (**Table S6A**) including 248 luminal A, 134 luminal B, 132 basal, 71 ERBB2 and 17 normal-like BCs. A total of 517 genes exhibited more than 5 mutations in the 602 tumors. For each of them, the frequency of mutation was established and we identified gene mutations associated with a specific subtype by comparing the number of mutations with those observed in the others (Fisher, FDR<0.25 with odds ratio>1 dark grey colored in **Table S6D**). The first 15 most frequently mutated genes are reported in





**Figure 4. DNA Methylation promoter profiles in breast cancers.** **A** - Hierarchical clustering established with the most variant methylation scores observed in 5,492 gene promoters ( $SD > 0.3$ ) in 109 tumors samples and 8 normal breast tissues samples. Each row of the data matrix represents a gene promoter and each column represents a sample. DNA methylation variations are depicted according to the color scale shown at the bottom. Red indicates increased DNA methylation score and green indicates decreased DNA methylation score. The dendrogram of samples (above matrixes) represents overall similarities in DNA methylation profiles and is zoomed to the right. Three groups of tumor samples (I, IIa and IIb) are associated with various DNA methylation patterns and delimited by orange vertical lines. Below the dendrogram are some histoclinical and molecular features of the samples: from top to bottom, intrinsic molecular subtypes<sup>4</sup>, IHC ER status and SBR grade. Color legends for the various features are illustrated below. **B** - The median methylation levels of the 4,545 subtype-associated genes were highest in the luminal and ERBB2 subtypes and lowest in the basal subtype ( $p = 6.24 \times 10^{-13}$ ). **C** - Compared to the other molecular subtypes, the DNA methylation levels (three top panels) of *ASS1*, *C6orf145* and *ZFP36L2* gene promoters and their mRNA expression (three bottom panels) were higher (**Table S4E**) and lower in the luminal B BCs, respectively.

doi:10.1371/journal.pone.0081843.g004

**Table 2** for each molecular subtype. *KCNB2* gene mutations were associated with luminal B tumors (Fisher,  $FDR < 0.25$  with odds ratio  $> 1$ ).

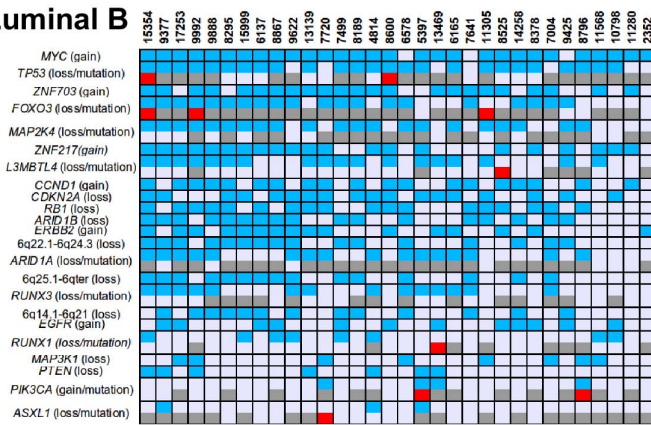
In an NGS subset [29], *TP53* mutations were enriched in luminal B tumors ( $p = 0.04$ ) and in high grade tumors ( $p = 0.02$ ). In our meta-analysis of all NGS samples, *TP53* mutations were enriched in only basal ( $FDR = 4 \times 10^{-18}$ ) and ERBB2 ( $FDR = 5 \times 10^{-6}$ ) tumors. The proportion of analyzed tumors and their heterogeneities could influence the results.

To identify co-occurring and mutually exclusive subtype-specific gene mutations, we retained in the 602 samples, the genes targeted by more than 3 mutations in each subtype. From the 103, 76, 98, 31 retained genes, we identified 74, 91, 121 and 52 co-occurring mutations, and 3 (*AKT1/PIK3CA*, *MAP3K1/TP53*, *MAP3K1/TTN*), 1 (*GATA3/PIK3CA*), 0 and 0 mutually exclusive gene

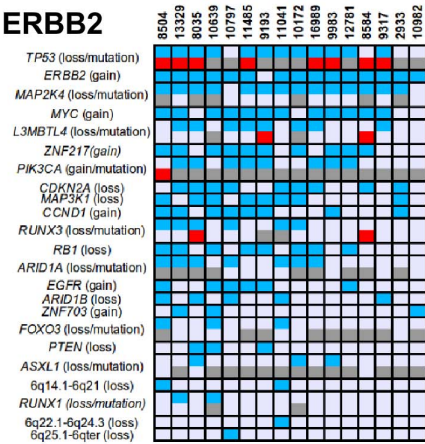
mutations ( $p < 0.05$ ) in luminal A, luminal B, basal and ERBB2 tumors, respectively (**Tables S6E–S6H**, respectively).

To identify key luminal B genes that could be altered by several mechanisms, we crossed in **Table S6I** information between genes found in the meta-analysis as mutated in luminal B cases and those identified in our study as significantly altered in luminal B tumors (**Tables S2A–S2F**) or as potential luminal B candidates (**Table S3G**). *KCNB2* mutations were associated with luminal B tumors (Fisher,  $FDR < 0.25$  with odds ratio  $> 1$ ), and the gene was more frequently gained in luminal B than in the other subtypes (Fisher's exact test;  $FDR < 0.05$ ) (**Tables S2A–S2F**). Among luminal B candidate genes, only *CIT* (12q24) (3%), *CHD6* (20q12) (2%), *MDN1* (6q15), *SRGAP1* (12q14.2) (1.5%), *UTRN* (6q24), *BRCAl* (17q21) and *EVPL* (17q25) (less than 1%) were also targeted by mutations.

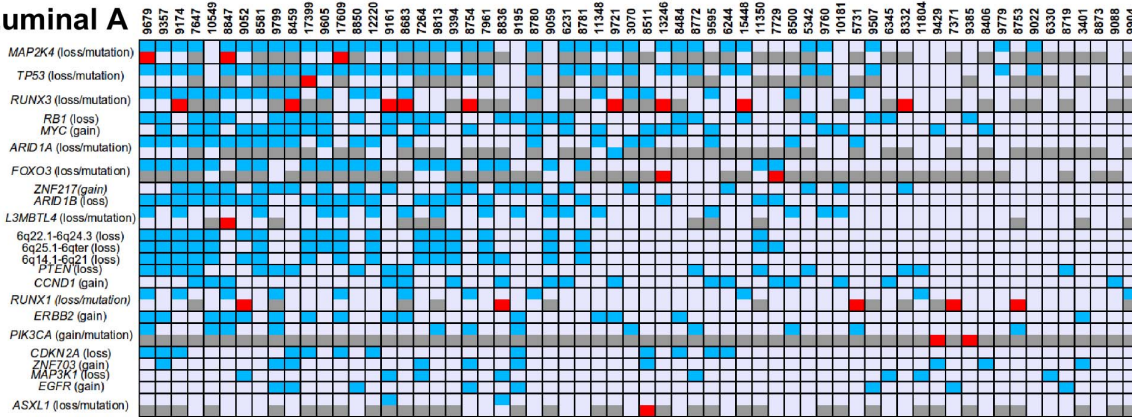
### Luminal B



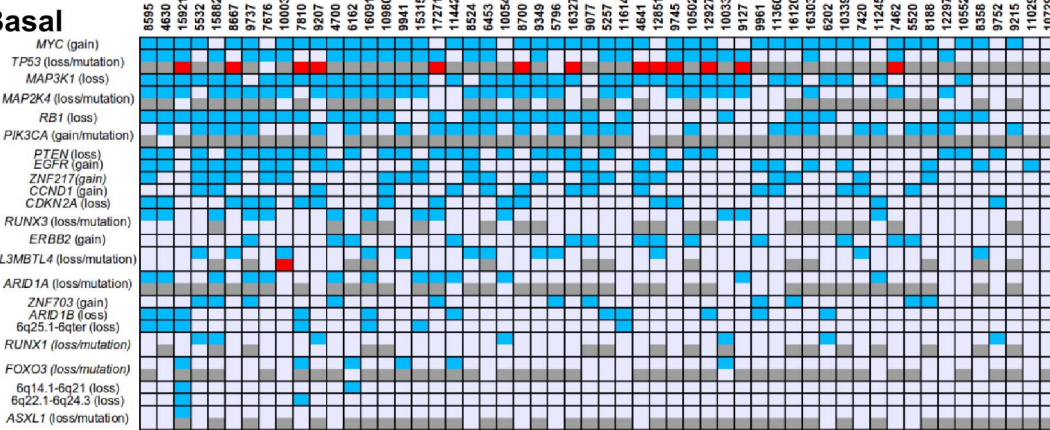
### ERBB2



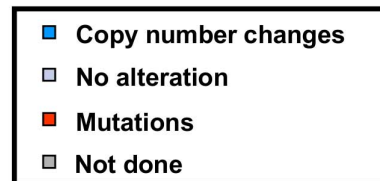
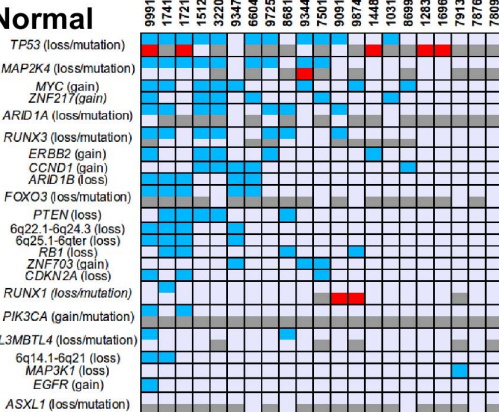
### Luminal A



### Basal



### Normal



**Figure 5. Landscape of specific regions/genes in the 188 breast tumors.** The most significant amplifications/gains, losses and mutations in four major subtypes are shown.

doi:10.1371/journal.pone.0081843.g005

## Discussion

Luminal B breast tumors have aggressive clinical and biological features [32]. A better definition of molecular alterations could improve our understanding of this tumor phenotype and allow the identification of new diagnostic and/or therapeutic targets. Here, we have first established and compared the genomic and gene expression profiles of 32 luminal B tumors and 156 tumors from other subtypes to constitute a repertoire of luminal B candidate genes. Second, in a subset of 109 breast tumors, we have compared the different methylated DNA landscapes associated with each subtype. Third, we have surveyed molecular alterations in some cancer genes and compared the results with a meta-analysis done with NGS data on 875 BCs.

This study depicts for the first time a luminal B genomic and epigenomic landscape including frequent *FOXO3*, *PIK3CA* and *TP53* mutation, with significant *ZNF703*, *MYC*, *CCND1*, *ZNF217* gains/amplifications, *FOXO3*, *CDKN2A*, *L3MBTL4*, various 6q regions, *ARID1B* losses and *FOXO3* mutations. In this luminal B context, (i) 237 and 101 luminal B-specific candidate oncogenes and TSGs presented a deregulated expression in relation with their CNAs; as well as (ii) low *ASS1*, *C6ORF145/PXDC1*, *ZFP36L2* and high *C3ORF67*, *C12ORF60*, *H2AFJ*, *RAB11FIP1* gene expression in relation with their DNA methylation levels. From the meta-analysis, only *KCNB2* gene mutation was associated with this subtype. Moreover, in a very large tumor set, we confirmed the luminal B specificity of 100 candidate oncogenes and showed for the first time that the overexpression of 61 of them are pertinent predictor of poorer patient outcome in luminal breast carcinomas independently of PR status. Compared to previous studies, our analysis point to 6q deletions and *FOXO3* mutations as important actors of luminal B genesis.

## Specific gains and amplifications target oncogene candidates in luminal B tumors

The high proportion of complex “sawtooth and firestorm” genomic profiles observed in luminal B tumors suggests a high level of genomic instability. Expectedly, among the most common specific alterations in luminal B tumors ( $\geq 30\%$ ) were the amplified 8p11-p12 and 11q13.1-q13 regions. In these regions *ZNF703*, *FGFR1* and *CCND1* have already been found associated with luminal B BCs [9,19,46,51]. The comparison between luminal B and luminal A genomic profiles showed only differences in the frequency of these, which may contribute to phenotypic differences. *ZNF703* interacts with *WDR68/DCAF7*, *PHB2* and *HSP60* and induces transcriptional repression [19]. Here, we found that *WDR68/DCAF7* (17q23.3) is indeed a candidate luminal B oncogene. Genes of the 8p11 region were coamplified with *ZNF703*, supporting the idea that the 8p amplicon carries multiple genes, such as *RAB11FIP1/RCP* [47], *FGFR1* [54] and *PPAPDC1B* [45], which contribute to the luminal B phenotype.

In the 8p12/11q13 coamplified regions [55], *EIF4EBP1* and *RPS6KB2* could be co-targeted and play a synergistic role associated with the development and cancer progression as AKT/MTOR activators [56]. The recurrent 8p12/11q13 co-amplification was often accompanied by gain of 17q25.1 containing *RPS6KB1*, a paralog of *RPS6KB2*, suggesting again the involvement of the AKT/MTOR pathway in luminal B oncogenesis.

## Specific losses target TSG candidates in luminal B tumors

Chromosome arm 6q showed frequent deletions in luminal B tumors. Several 6q regions were lost and rare homozygous or focal deletions of *ARID1B*, *ASCC3*, *FOXO3*, *PARK2*, *MLLT4/Afadin* and

**Table 2.** NGS meta-analysis: first 15 most frequently mutated genes in each molecular subtype.

Basal (N = 132)		ERBB2 (N = 71)		Luminal A (N = 248)		Luminal B (N = 134)	
mutated genes	frequency (%)	mutated genes	frequency (%)	mutated genes	frequency (%)	mutated genes	frequency (%)
<i>TP53*</i>	71%	<i>TP53*</i>	68%	<i>PIK3CA*</i>	43%	<i>PIK3CA</i>	31%
<i>TTN</i>	19%	<i>PIK3CA</i>	37%	<i>MAP3K1*</i>	13%	<i>TP53</i>	29%
<i>USH2A*</i>	11%	<i>MUC16</i>	14%	<i>GATA3</i>	13%	<i>GATA3</i>	13%
<i>FLG</i>	7%	<i>LRP1*</i>	8%	<i>TP53</i>	11%	<i>TTN</i>	12%
<i>MUC16</i>	7%	<i>ERBB3*</i>	8%	<i>TTN</i>	9%	<i>RYR2</i>	7%
<i>PIK3CA</i>	7%	<i>DNAH11*</i>	8%	<i>CDH1</i>	9%	<i>RELN</i>	5%
<i>MUC17</i>	6%	<i>LRP2</i>	8%	<i>MLL3</i>	7%	<i>FAT3</i>	5%
<i>DNAH7*</i>	5%	<i>TTN</i>	8%	<i>MAP2K4</i>	6%	<i>MLL3</i>	5%
<i>FAT3</i>	5%	<i>ATP1A4*</i>	7%	<i>NCOR1</i>	5%	<i>MUC16</i>	5%
<i>SYNE1</i>	5%	<i>KIAA1109*</i>	7%	<i>AKT1</i>	4%	<i>KCNB2*</i>	4%
<i>DST</i>	5%	<i>CACNA1E</i>	7%	<i>RUNX1</i>	4%	<i>CENPE</i>	4%
<i>F5</i>	5%	<i>FLG</i>	7%	<i>PTEN</i>	4%	<i>DMD</i>	4%
<i>COL12A1</i>	5%	<i>MLL3</i>	7%	<i>MUC16</i>	4%	<i>MUC17</i>	4%
<i>COL6A3</i>	5%	<i>ASPM*</i>	6%	<i>CTCF</i>	4%	<i>CDH1</i>	4%
<i>APOB</i>	5%	<i>DCC*</i>	6%	<i>NEB</i>	4%	<i>MAP3K1</i>	4%

\*Gene mutation associated with the corresponding molecular subtype.

doi:10.1371/journal.pone.0081843.t002

*UFL1/NLBP* genes were observed. Deletion of 6q is also frequent in a variety of cancers [57]. Several other chromosomal regions were targeted by losses suggesting the existence of several potential TSGs in luminal B cancers (**Table S7**). We identified luminal B candidate TSGs *PNRC1* (6q15), *PTPRK* (6q22.33) and *L3MBTL4* (13p11.31). We previously reported *L3MBTL4* as a potential BC TSG. *PNRC1* is a coactivator of nuclear receptors such as ER. *PTPRK* negatively regulates the transcriptional activity of  $\beta$ -catenin in cancer cells.

### Gene expression deregulation of luminal B candidate genes could perturb epigenetic regulation

Eleven of the luminal B candidate genes we found have been associated with endocrine resistance [33,34,44]. Upregulation of *FOXMI* [58] could explain in part the DNA methylated landscape characterizing luminal B tumors. The deregulation of *ARID1B*, *ASHL2*, *CHD6*, *KDM4C/JMJ2C*, *L3MBTL4* and *ZFP161* suggests an important perturbation of the epigenetic regulation in luminal B tumors.

BC subtypes have specific methylation profiles [20,21]. Luminal B and basal BCs were reported as the most and least frequently methylated, respectively [21]. Our data are in agreement with these observations; the median methylation level of the 4,545 subtype-associated promoters was the highest in the luminal and *ERRB2* subtypes and the lowest in the basal subtype. High DNA methylation level associated with the luminal B subtype targeted *CITED4*, *SPI100*, *SAMD9L*, *DCR1*, *FBXO32*, *ASS1*, *FAM78A* and *STAT5A* genes previously reported as TSGs or associated with tumor progression (**Table S7**). None of them is located in 6q. The two specific luminal B candidates, *ASS1* and *ZFP36L2* downregulated in relation with DNA methylation have been reported targeted in cancers. DNA methylation of the *ASS1* promoter leading to its downregulation was associated with poor prognosis and chemoresistance in various cancers. Several phase I/II clinical trials of the arginine-lowering drug, pegylated arginine deiminase, showed encouraging evidence of clinical benefit and low toxicity in patients with *ASS1*-negative tumors that might be extended to luminal B tumors.

Finally, we further noted that multiple alternatively spliced transcript variants have been described for candidate genes *ASB13*, *CIT*, *CPSF1*, *EPN2*, *LSM1*, *PDCD4*, *RNF146*, *TAF4*, *VTCN1*. Transcripts of *CIT*, *CPSF1*, *PDCD4* and *TAF4* are overlapped by *MIR1178*, *MIR939* and *MIR1234* (both), *MIR4680* and *MIR1257*, respectively [59]. Changes in *MIR* expression may also contribute to luminal B tumorigenesis by modulating the functional expression of critical genes involved in cancer growth and metastasis.

### Mutation status

As expected, *TP53* and *PIK3CA* were among the most frequently mutated genes in our luminal B series. *FOXO3* and *RUNX3* genes were lost and mutated. A possible role for *RUNX3* as a tumor suppressor in ER+ BCs has been suggested [60,61]. *FOXO3* functions as a tumor suppressor in both ER+ and ER- BCs [62,63]. Its nuclear localization and subsequent transcriptional activity is a marker of good prognosis among breast cancer patients [64].

Our meta-analysis of NGS samples showed the frequent mutations of *PIK3CA*, *TP53* and *GATA3* in luminal B tumors but also that *KCNB2* gene mutations were, for an unexplained reason, associated with this aggressive subtype. A total of 91 co-occurring mutations and *GATA3/PIK3CA* gene mutual exclusive mutations completed the landscape associated with luminal B tumors. Surprisingly, *RUNX3* and *FOXO3* mutations were not

found in the NGS studies. This may be due the limited coverage and low depth of these early analyses, or to BC heterogeneity and the large number of alterations that could lead to similar deregulations of particular pathways.

### Potential targeted therapy in luminal B breast cancer

Hormonal therapy is the preferred treatment for about two-thirds of all BC patients whose tumor expresses ER. ER+ BCs are commonly treated with anti-estrogens or aromatase inhibitors [65]. Luminal B tumors respond less than luminal A tumors to such treatments. Seven of the luminal B candidate oncogenes we have identified could be targeted [43]. Several phase I/II clinical trials targeting IGF, FGF and PI3K/AKT pathways in luminal B BC have been reported [32]. Except for the IGF pathway, our data are in agreement with the potential activation of the FGF, PI3K/AKT, PIK3/MTOR pathways subsequent to the overexpression of candidate oncogenes such as *RPS6KB1*, *RPS6KB2*, *EIF4EBP1*, *FGFR1*, *FRS2*, and *RAB11FIP1* and the high frequency of *PIK3CA* mutations. Interestingly, targeting the PI3K/AKT/mTOR pathway is one of the most-promising therapeutic approaches to reversing endocrine resistance for ER positive breast cancer (for review, [66]). Therapeutic strategies combining endocrine treatment and signaling pathways inhibition (such as targeting growth factor receptors (*ERBB2*) PI3K/AKT/mTOR, *CDK4* and *CDK6*, *MDM2-TP53* interaction, *FGFR* pathway aberrations and histone deacetylases) might be pertinent [66]. However, the luminal B molecular heterogeneity depicted by its complex genomic and epigenomic landscape strongly suggests that other potential targets could also exist and should be identified for each patient. The candidate oncogene *RECQL4*, which might be involved in endocrine resistance [44], could be also a potential therapeutic target [43] as recently shown in breast cancer cells [52].

The targeting of candidates such as *YWHAZ* and its coregulated proteins such as *FOXMI*, or *NERF1*, could also restore endocrine sensitivity and reduce the risk of BC recurrence [32,34,44,67,68]. *FOXMI* has a significant role in endocrine therapy resistance [67]. In endocrine therapy-resistant BCs, high *FOXMI* expression could result from the loss of *FOXO* function as transcriptional repressor.

In conclusion, this refined molecular dissection of luminal B BCs has pointed to both new and well-known specific candidates. Some code for proteins that participate to the same signaling pathways including those known to cross-talk with ER signaling pathways. Many of the candidate genes have not been previously reported in breast cancer, and deserve further functional validation. Similarly, further characterization of the 6q TSGs is an important goal. This should help better understand pathways and mechanisms affected, and find new therapeutic targets.

### Supporting Information

**Figure S1** Whole-genome expression profiling of 188BCs. **A)** Hierarchical clustering of 188 samples and 13,031 probe sets with significant variation ( $sd > 0.5$ ) in mRNA expression level across the samples. Each row of the data matrix represents a gene and each column represents a sample. Expression levels are depicted according to the color scale shown at the bottom. Red and green indicate expression levels respectively above and below the median. The magnitude of deviation from the median is represented by the color saturation. The dendrogram of samples (above matrix) represents overall similarities in gene expression profiles and is zoomed in B. Colored bars to the right indicate the locations of 6 gene clusters of interest (ECM means extra-cellular

matrix). **B**) Dendrograms of samples. *Top*, four groups of tumor samples (designated I to IV) are evidenced and delimited by orange vertical lines. Below the dendrogram, are some histoclinical and molecular features of the samples: from top to bottom, intrinsic molecular subtypes (dark blue for luminal A, light blue for luminal B, red for basal, pink for ERBB2-overexpressing, and green for normal-like)<sup>4</sup>, GC (genome complexity) (green for simplex, orange for complex saw tooth and brown for complex firestorm), SBR grade (white for grade I, grey for II, and black for III), and IHC ER, ERBB2, and P53 status (white for negative, and black for positive). Crossed white boxes mean not assigned samples.

(TIFF)

**Figure S2** Integrated comparative analysis of luminal B vs non luminal B BCs. The depicted pipeline integrates genomic and gene expression data analyses. The three successive steps are numbered 1, 2 and 3.

(TIF)

**Figure S3** Integrated comparative DNA methylation and gene expression analysis associated with breast cancer molecular subtypes. The depicted pipeline integrates DNA methylation and gene expression data analyses.

(TIF)

**Figure S4** Genomic characterization of 188 BCs in regard to their molecular subtype. CNA frequencies and the repartition of genomic patterns are described for each molecular subtype. Amplification, gains, homozygous and hemizygous losses CNAs are distinguished by different colors.

(TIFF)

**Figure S5** Specific regions targeted by CNAs in luminal B BCs. Results of supervised analysis comparing CNA frequencies between luminal B and luminal A, luminal B and basal, luminal B and ERBB2 and luminal B and non-luminal B tumors. Only amplification, gains, homozygous and hemizygous losses associated with luminal B tumors are shown (Fisher's exact test; FDR < 0.05) for each chromosome. They are distinguished by illustrated colors. The successive comparisons show distinct regional CNAs significantly associated with luminal B tumors compared to those found in each of the other major molecular subtypes.

(TIFF)

**Figure S6** Examples of chromosome 6 aCGH profiles in luminal B tumors. Tumors exhibit various 6q regional losses as well as rare homozygous deletions and small deleted regions targeting *MLLT4*, *ARID1B*, *PARK2*, *FOXO3A*, *UFL1/NLBP*, *ASCC3* genes.

(TIFF)

**Figure S7** Examples of chromosome 9 aCGH profiles in luminal B tumors. Tumors exhibit various 9p regional losses as well as rare focused deletions and short deleted regions. **(A)** Tumor T17253 on the left does not present any CNA. Tumors T13139, T11305, T13694, T15354, T13148, T10798, T13728, T2362, and T13469 exhibit various copy number losses along the short arm of chromosome 9 suggesting at least two common lost regions involved in the luminal B tumors (bold lines to the right). **(B)** The focused deletions observed in tumors T8189 and T6137-2 target the *MTAP-CDKN2B-CDKN2A* (9p21.3) and *BNC2* (9p23.2) genes, respectively. **(C)** The genomic profiles observed in tumors T2362, and T13469 show a short common deleted region spanning from centromere to telomere, *C9ORF93*, *PSIP1*, *C9ORF59*, *FREM1*, *CERF1*, *ZDHHC21* and *NFIB* genes.

(TIFF)

**Figure S8** Correlation between gene expression and genome alterations on the 8p11.1-p12 and 11q13.1-q13.4 regions. Genomic and gene expression profiles were established for 188 breast tumors (32 luminal B and 156 non luminal B identified at the top by blue and grey boxes, respectively) in the 8p11.1-p12 (top), and 11q13.1-q13.4 (bottom) regions. For each region, heatmaps for genome copy number and gene expression profiles are consecutively drawn. Genome copy number was measured by aCGH on probes or groups of probes spanning each of these regions. Red indicates increased copy number and green indicates decreased copy number. In the heatmap tumors are organized from the tumor that presented the highest copy number gains and amplification to the tumor that exhibited the most copy number losses. The next heatmap was established with the expression of the independent genes located on the corresponding region and profiled in the same 188 tumors similarly organized. For gene copy number and gene expression heatmaps, we used colour scale limits from -3 to +3 and -2 to +2, respectively. Next to the right, are plotted genes successively selected by steps I, II and III of the integrated analysis "aCGH & mRNA expression" as defined by the work pipeline (Figure S2). Grey and red lines correspond to rejected and selected genes, respectively. Among genes with an expression level that varied according to CNAs, we retained genes showed significant differences (vertical line) in copy number gains correlated with upregulated expression in luminal B compared to non-luminal B tumors. They were qualified as potential oncogenes.. For each region, only the first five most significant are listed. *ZNF703* and *CCND1* genes were the most significant candidate oncogenes for the 8p11.1-p12 (top) and 11q13.1-q13.4 (bottom) regions, respectively.

(TIFF)

**Figure S9** Luminal B candidates and gene CNAs landscape. The Circos diagram presents from outside to inside, luminal B candidate genes, luminal B altered chromosomes, luminal B regional CNAs colored in red and green for significant gains and losses, respectively. Oranges and grey arcs indicate respectively genes/regions that present significant mutually exclusive and co-occurring luminal B CNAs (FDR < 0.05) as identified in **Table S3N**.

(TIF)

**Figure S10** Primers design for EpiTyper analysis of *RASSF1* promoter gene. *RASSF1* primers were designed for EpiTYPER<sup>TM</sup> Mass-ARRAY<sup>®</sup> system approach (SEQUENOM<sup>®</sup>, USA), to compare *RASSF1* DNA methylation data obtained by two independent methods. **(A)** *RASSF1* DNA methylation profiles of 15 ER+ and 33 ER- breast tumors established with their median normalized M values obtained for each *RASSF1* oligonucleotide present on the human promoter array (Agilent Technologies). **(B)** For the detection and quantitative analysis of DNA methylation, the EpiTyper approach used eight amplicons spanning the chr3:50,349,000–50,352,780 region including the *RASSF1* promoter and was covered by nine *RASSF1* oligonucleotides in the promoter array.

(TIF)

**Figure S11** Comparison of *RASSF1* DNA methylation data obtained by two independent methods. **(A)** Hierarchical clustering established for 48 breast tumors samples with the methylation data of the 95 CpG within the *RASSF1* promoter and measured by the by EpiTyper method. Each row of the data matrix represents a CpG and each column represents a sample. DNA methylation variations are depicted according to the color scale shown at the bottom. Red indicates increased DNA methylation level and green indicates decreased DNA methylation level. The dendrogram

(above matrixes) of samples represents overall similarities in DNA methylation profiles and is zoomed in the upper right part. The hierarchical clustering distinguished ER+ and ER- tumors (Fisher,  $p = 4.3 \cdot 10^{-3}$ ). **(B)** For each sample, median methylation ratios (EpiTYPER) were calculated, with four informative amplicons overlapping oligoprobes (SQ00001\_RASF1\_07, SQ00001\_RASF1\_12, SQ00001\_RASF1\_25, SQ00001\_RASF1\_32 mentioned by arrows) used to calculate Methylation Score. We observed a strong correlation between median methylation ratios and methylation score (Pearson correlation = 0.66,  $p = 4.6 \cdot 10^{-7}$ ) calculated from data established by EpiTYPER and promoter array approaches, respectively. (TIF)

**Figure S12** Genes exhibiting a molecular subtype specific deregulated expression in relation with a molecular subtype specific methylation level variation. For luminal B **(A)**, luminal A **(B)**, ERBB2 **(C)** and basal **(D1–D4)** molecular subtypes are represented genes exhibiting a significant (i) DNA methylation level variation of their gene promoter (ANOVA, upper part) and (ii) gene expression deregulation compared to the other molecular subtypes (ANOVA and p-value showing significant difference, lower part). (TIF)

**Figure S13** Examples of mutation of *RUNX1* in breast cancer. **a.** Sequence profile of the mutated *RUNX1* allele, demonstrating base change in the forward sequence at the position indicated by an arrow. The corresponding sequence is shown above. **b.** Genomic organization of *RUNX1* gene and RUNX1 protein. Located at 21q22.12 chromosomal band, the *RUNX1* gene spans the chr21:36,160,098–36,421,595 region. The gene map established within Mb scale was extracted from the build GRCh37/hg19 from NCBI (February 2009 version) while its sequence (Ensembl Transcript ID ENST00000300305) was extracted from Ensembl database ([http://www.ensembl.org/Homo\\_sapiens/](http://www.ensembl.org/Homo_sapiens/)), which is based on the Ensembl release 48 - Dec 2007 assembly of the human genome. Functional (i.e. RUNT and RUNX1 [for RUNX Inhibitor domain], as defined by PFAM accession numbers PF00853 and PF08504, respectively) and motifs of the RUNX1 protein were positioned according to the SMART program (<http://smart.embl-heidelberg.de/>). Nucleotide (cDNA level) and deduced aminoacid sequences of the RUNX1 protein are positioned above and below the corresponding protein, respectively. **c.** The mutations observed in tumor samples are located with respect to the modified aminoacid of the RUNX1 protein. (TIF)

**Figure S14** Examples of mutation of *FOXO3* in breast cancer. From top to the bottom, in the 6q21 chromosomal band, the *FOXO3* gene spans the chr6:108,881,026–109,005,971 region. Exons 2 and 3 of *FOXO3* gene were analyzed for mutations. Sequence profile of the mutated *FOXO3* allele, demonstrating base change in the forward sequence at the position indicated by an arrow. The corresponding sequence is shown above. (TIF)

**Figure S15** Examples of aCGH profiles showing *RUNX1* losses. **(A)** From left to right, aCGH profiles of chromosome 21 in cases T6744 (no apparent CNA), T9345, T8056, T50115, T9934, T11568, T15120, and T9207. Arrow shows *RUNX1* location on each genomic profile. Results show that *RUNX1* is targeted by potential breaks in T8056, T9207 and T15120, but also by regional deletions (samples T50115, T9934, T11568, T15120, and T9207). The genomic profiles of T11568 **(B)**, T15120 **(C)**, and T9207 **(D)** BCs presenting the smallest regional deletions were

established with CGH analytics® software (Agilent Technologies), from centromere to telomere, within the genomic intervals [34.0–36.4 Mb] of the long arm of the chromosome 21. The common smallest deleted region observed in T9207 **(D)** involves *RUNX1*. (TIF)

**Results and References S1** In this section, results about: validation of our methylation approach; DNA methylation level associated with the other breast cancer molecular subtypes; and specific deregulated gene expression in relation with the DNA methylation level variation associated with the other breast cancer molecular subtypes; are presented with supplementary references. (DOCX)

**Table S1 A:** Clinical and histological features of the 188 tumors. **B:** Description of the breast cancer data sets. **C:** FOXO3, MAP2K4, MAP3K1, PIK3CA, RUNX3 primers and PCR conditions used before sequencing analysis. **D:** Histoclinical correlations of the 188 BCs with the five SSP groups. (XLSX)

**Table S2 A:** Comparison of CNA frequencies between luminal B and luminal A BCs. **B:** Identification of luminal B specific regions significantly targeted by CNAs after comparison of CNA frequencies between luminal B and luminal A BCs. **C:** Comparison of CNA frequencies between luminal B and basal BCs. **D:** Identification of luminal B specific regions significantly targeted by CNA after comparison of CNA frequencies between luminal B and basal BCs. **E:** Comparison of CNA frequencies between luminal B and non-luminal B BCs. **F:** Identification of luminal B specific regions significantly targeted by CNAs after comparison of CNA frequencies between luminal B and non-luminal B BCs. (XLSX)

**Table S3 A:** Comparison of expression levels according to CNAs (Mann-Whitney ; FDR<0.05) in luminal B vs luminal A BCs. **B:** Identification of specific luminal B candidate genes from comparison between luminal B and luminal A BCs. A comparison with TCGA data [28] was done. **C:** Comparison of expression levels according to CNAs (Mann-Whitney ; FDR<0.05) in luminal B vs basal BCs. **D:** Identification of specific luminal B candidate genes from comparison between luminal B and basal BCs. A comparison with TCGA data [28] was done. **E:** Comparison of expression levels according to CNAs (Mann-Whitney ; FDR<0.05) in luminal B vs non luminal B BCs. **F:** Identification of specific luminal B candidate genes from comparison between luminal B and non-luminal B BCs. A comparison with TCGA data [28] was done. **G:** Summary of specific luminal B candidate genes from different comparisons between luminal B and the other molecular subtypes. **H:** 34 genes are exclusively deregulated in the comparison luminal B vs luminal A. **I:** 47 genes are exclusively deregulated in the comparison luminal B vs basal. **J:** 2 genes are exclusively deregulated in the comparison luminal B vs normal-like. **K:** 101 candidate oncogenes in the luminal B core. **L:** DAVID ontology analysis on the 101 genes of the luminal B core. **M:** The overexpression of 67 “luminal B core” candidate oncogenes is associated with poor MFS. A data comparison was done with those obtained for tumors classified with PAM50 SSP [48]. **N:** Mutually exclusive and co-occurring luminal B CNAs. (XLSX)

**Table S4 A:** Histoclinical characteristics of the 117 samples (109 BC+8 NB) distributed in the three clusters defined by the unsupervised DNA methylation analysis. **B:** Supervised DNA methylation analysis comparing ER+ and ER- BCs. **C:** DNA methylation data analysis – 4,545 gene promoters exhibited a

DNA methylation level different in at least one molecular subtype. **D:** 459 genes exhibited mRNA expression deregulation in relation with DNA methylation level. **E:** Genes exhibiting a mRNA expression deregulation in relation with DNA methylation level in the various molecular subtypes. A comparison with TCGA data [28] was done. **F:** Molecular subtype specific candidates presenting mRNA expression deregulation in relation with CNA and with DNA promoter methylation aberrations. (XLSX)

**Table S5 A:** Criteria for our selection of the 9 sequenced genes. **B:** Gene mutation analysis of *ARID1A*, *ASXL1*, *FOXO3*, *L3MBTL4*, *MAP2K4*, *PIK3CA*, *RUNX1*, *RUNX3* and *TP53* in a large set of BCs including our panel of 188 BCs. **C:** Frequency analysis of genomic alterations targeting specific regions or genes in the five molecular breast cancer subtypes. (XLSX)

**Table S6 A:** References of the six recent NGS studies used in the meta-analysis (dataset Meta-analysis). **B:** Frequency of somatic mutations identified in the 875 NGS breast tumors. **C:** Co-occurring and mutually exclusive mutations in the 875 NGS BCs (without molecular subtype distinction). **D:** Frequency of somatic mutations and association with molecular subtypes. **E:** Co-occurring and mutually exclusive gene mutations associated with

luminal A subtype. **F:** Co-occurring and mutually exclusive gene mutations associated with luminal B subtype. **G:** Co-occurring and mutually exclusive gene mutations associated with basal subtype. **H:** Co-occurring and mutually exclusive gene mutations associated with ERBB2 subtype. **I:** genes targeted by several alterations mechanisms in luminal B subtype.

(XLSX)

#### Table S7

Genes commented in discussion with supplemental references.

(XLSX)

## Acknowledgments

SR was a PhD student who has contributed to the analysis of genomic and expression profiles in several breast cancer samples included in this study. He passed away before the completion of his thesis.

## Author Contributions

Conceived and designed the experiments: SS HdT PV FB DB MC. Performed the experiments: SC JA LAK PF MRS MM NC IB SR. Analyzed the data: AG PF AL SS HdT FB MC. Contributed reagents/materials/analysis tools: AG PF ECJ JJ. Wrote the paper: DB MC.

## References

- Dunn L, Demichele A (2009) Genomic predictors of outcome and treatment response in breast cancer. *Mol Diagn Ther* 13: 73–90.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. (2000) Molecular portraits of human breast tumours. *Nature* 406: 747–752.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98:10869–10874.
- Hu Z, Fan C, Oh DS, Marron JS, He X, et al. (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 7: 96.
- Sotiriou C, Pusztai L (2009) Gene-expression signatures in breast cancer. *N Engl J Med* 360: 790–800.
- Gelsi-Boyer V, Orsetti B, Cervera N, Finetti P, Sircoulomb F, et al. (2005) Comprehensive profiling of 8p11–12 amplification in breast cancer. *Molecular Cancer Research* 3: 655–667.
- Bergamaschi A, Kim YH, Wang P, Sorlie T, Hernandez-Boussard T, et al. (2006) Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* 45: 1033–1040.
- Andre F, Job B, Dessen P, Tordai A, Michiels S, et al. (2009) Molecular characterization of breast cancer with high-resolution oligonucleotide comparative genomic hybridization array. *Clin Cancer Res* 15: 441–451.
- Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, et al. (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* 10: 529–541.
- Sransky N, Vallot C, Reyat F, Bernard-Pierrot I, de Medina SG, et al. (2006) Regional copy number-independent deregulation of transcription in cancer. *Nat Genet* 38: 1386–1396.
- Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 10: 515–527.
- Adelaide J, Finetti P, Bekhouche I, Repellini L, Geneix J, et al. (2007) Integrated profiling of basal and luminal breast cancers. *Cancer Res* 67: 11565–11575.
- Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, et al. (2007) High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol* 8: R215.
- Vincent-Salomon A, Lucchesi C, Gruel N, Raynal V, Pierron G, et al. (2008) breast cancer study group of the Institut Curie. Integrated genomic and transcriptomic analysis of ductal carcinoma in situ of the breast. *Clin Cancer Res* 14:1956–1965.
- Addou-Klouche L, Adelaide J, Finetti P, Carbuca N, Ferrari A, et al. (2010) Loss, mutation and deregulation of *L3MBTL4* in breast cancers. *Mol Cancer* 9: 213.
- Horlings HM, Lai C, Nuyten DS, Halfwerk H, Kristel P, et al. (2010) Integration of DNA copy number alterations and prognostic gene expression signatures in breast cancer patients. *Clin Cancer Res* 16: 651–663.
- Sircoulomb F, Bekhouche I, Finetti P, Adelaide J, Ben Hamida A, et al. (2010) Genome profiling of ERBB2-amplified breast cancers. *BMC Cancer* 10: 539.
- Bekhouche I, Finetti P, Adelaide J, Ferrari A, Tarpin C, et al. (2011) High-resolution comparative genomic hybridization of inflammatory breast cancer and identification of candidate genes. *PLoS ONE* 6: e16950.
- Sircoulomb F, Nicolas N, Ferrari A, Finetti P, Bekhouche I, Rousset E, et al. (2011) ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. *EMBO Mol Med* 3: 153–166.
- Bediaga NG, Acha-Sagredo A, Guerra I, Viguri A, Albaina C, et al. (2010) DNA methylation epigenotypes in breast cancer molecular subtypes. *Breast Cancer Res* 12: R77.
- Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jönsson G, et al. (2010) Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res* 12: R36.
- van der Auwera I, Yu W, Suo L, Van Neste L, van Dam P, et al. (2010) Array-based DNA methylation profiling for breast cancer subtype discrimination. *PLoS One* 5: e12616–12625.
- Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, 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.
- Hill VK, Ricketts C, Bieche I, Vacher S, Gentle D, et al. (2010) Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity. *Cancer Res* 71: 2988–2999.
- Edgren H, Murumagi A, Kangaspeka S, Nicorici D, Hongisto V, et al. (2011) Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol* 12: R6.
- Ha KC, Lalonde E, Li L, Cavallone L, Natrajan R, et al. (2011) Identification of gene fusion transcripts by transcriptome sequencing in BRCA1-mutated breast cancers and cell lines. *BMC Med Genomics* 4: 75.
- Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, et al. (2012) Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 486: 405–409.
- Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490: 61–70.
- Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, et al. (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486: 353–360.
- Shah SP, Roth A, Goya R, Oloumi A, Ha G, et al. (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486: 395–399.
- Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, et al. (2012) The landscape of cancer genes and mutational processes in breast cancer. *Nature* 20486: 400–404.
- Tran B and Bédard PL (2011) Luminal-B breast cancer and novel therapeutic targets. *Breast Cancer Res* 13: 221.
- van Aghoven T, Sieuwerts AM, Veldscholte J, Meijer-van Gelder ME, et al. (2009) CITED2 and NCOR2 in anti-oestrogen resistance and progression of breast cancer. *Br J Cancer* 2009; 101:1824–1832.
- Karn T, Ruckhäberle E, Hanker L, Müller V, Schmidt M, et al. (2011) Gene expression profiling of luminal B breast cancers reveals NHERF1 as a new marker of endocrine resistance. *Breast Cancer Res Treat* 130: 409–420.

35. van Agthoven T, Godinho MF, Wulfkühle JD, Petricoin EF 3rd, Dorschers LC (2012) Protein pathway activation mapping reveals molecular networks associated with antiestrogen resistance in breast cancer cell lines. *Int J Cancer* 131: 1998–2007.
36. Dadi S, Le Noir S, Payet-Bornet D, Lhermitte L, Zacarias-Cabeza J, et al. (2012) TLX homeodomain oncogenes mediate T cell maturation arrest in T-ALL via interaction with ETS1 and suppression of TCR $\alpha$  gene expression. *Cancer Cell* 21:563–576.
37. Flaman JM, Frebourg T, Moreau V, Charbonnier F, Martin C, et al. (1995) A simple p53 functional assay for screening cell lines, blood, and tumors. *Proc Natl Acad Sci U S A* 92: 3963–3967.
38. Bertheau P, Turpin E, Rickman DS, Espié M, de Reyniès A, et al. (2007) Exquisite sensitivity of TP53 mutant and basal breast cancers to a dose-dense epirubicin-cyclophosphamide regimen. *PLoS Med*:e90.
39. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, et al. (2006) Reporting recommendations for tumor MARKer prognostic studies (REMARK). *Breast Cancer Res Treat* 100: 229–235.
40. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19: 1639–1645.
41. Deng N, Goh LK, Wang H, Das K, Tao J, et al. (2012) A comprehensive survey of genomic alterations in gastric cancer reveals systematic patterns of molecular exclusivity and co-occurrence among distinct therapeutic targets. *Gut* 61:673–684.
42. Yeang CH, McCormick F, Levine A (2008) Combinatorial patterns of somatic gene mutations in cancer. *Faseb J* 22: 2605–2622.
43. Patel MN, Halling-Brown MD, Tym JE, Workman P, Al-Lazikani B (2013) Objective assessment of cancer genes for drug discovery. *Nat Rev Drug Discov* 12: 35–50.
44. Bergamaschi A, Christensen BL, Katzenellenbogen BS (2011) Reversal of endocrine resistance in breast cancer: interrelationships among 14-3-3 $\zeta$ , FOXM1, and a gene signature associated with mitosis. *Breast Cancer Res* 13: R70.
45. Bernard-Pierrot I, Gruel N, Stransky N, Vincent-Salomon A, Reyat F, et al. (2008) Characterization of the recurrent 8p11-12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer. *Cancer Res* 68: 7165–7175.
46. Holland DG, Burleigh A, Git A, Goldgraben MA, Perez-Mancera PA, et al. (2011) ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. *EMBO Mol Med* 3: 167–180.
47. Zhang J, Liu X, Datta A, Govindarajan K, Tam WL, et al. (2009) RCP is a human breast cancer-promoting gene with Ras-activating function. *J Clin Invest* 119: 2171–2183.
48. Braun L, Mietzsch F, Scibold P, Schneeweiss A, Schirmacher P, et al. (2013) Intrinsic breast cancer subtypes defined by estrogen receptor signalling-prognostic relevance of progesterone receptor loss. *Mod Pathol* 26: 1161–1171.
49. Cancellò G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, et al. (2013) Progesterone receptor loss identifies Luminal B breast cancer subgroups at higher risk of relapse. *Ann Oncol* 24: 661–668.
50. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, et al. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27: 1160–1167.
51. Weigelt B, Mackay A, A'hern R, Natrajan R, Tan DS, et al. (2010). Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol* 11: 339–349.
52. Fang H, Nie L, Chi Z, Liu J, Guo D, et al. (2013) RecQL4 Helicase Amplification Is Involved in Human Breast Tumorigenesis. *PLoS One* 8:e69600.
53. Cornen S, Adelaide J, Bertucci F, Finetti P, A Guille A., et al. (2012) Mutations and deletions of ARID1A in breast tumors. *Oncogene* 31: 4255–4256.
54. Turner N, Pearson A, Sharpe R, Lambros M, Geyer F, et al. (2010) FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. *Cancer Res* 70: 2085–2094.
55. Kwek SS, Roy R, Zhou H, Climent J, Martinez-Climent JA, et al. (2009) Co-amplified genes at 8p12 and 11q13 in breast tumors cooperate with two major pathways in oncogenesis. *Oncogene* 28:1892–1903.
56. Karlsson E, Waltersson MA, Bostner J, Pérez-Tenorio G, Olsson B, et al. (2011) High-resolution genomic analysis of the 11q13 amplicon in breast cancers identifies synergy with 8p12 amplification, involving the mTOR targets S6K2 and 4EBP1. *Genes Chromosomes Cancer* 50: 775–787.
57. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2013). Mitelman F, Johansson B and Mertens F (Eds.), <http://cgap.nci.nih.gov/Chromosomes/Mitelman>. Accessed 2013 Dec 6.
58. Teh MT, Gemenetzidis E, Patel D, Tariq R, Nadir A, et al. (2012) FOXM1 induces a global methylation signature that mimics the cancer epigenome in head and neck squamous cell carcinoma. *PLoS One* 2012; 7: e34329.
59. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*34 (Database issue): D140–144. miRBase is available at <http://microrna.sanger.ac.uk/>.
60. Lau QC, Raja E, Salto-Tellez M, Liu Q, Ito K, et al. (2006) RUNX3 is frequently inactivated by dual mechanisms of protein mislocalization and promoter hypermethylation in breast cancer. *Cancer Res* 66: 6512–6520.
61. Huang B, Qu Z, Ong CW, Tsang YH, Xiao G, et al. (2012) RUNX3 acts as a tumor suppressor in breast cancer by targeting estrogen receptor  $\alpha$ . *Oncogene* 31: 527–534.
62. Accili D, Arden KC (2004) FoxOs at the crossroads of cellular metabolism, differentiation, and transformation. *Cell* 117 : 421–426.
63. Zou Y, Tsai W-B, Cheng C-J, Hsu C, Chung YM, et al. (2008) Forkhead box transcription factor FOXO3a suppresses estrogen-dependent breast cancer cell proliferation and tumorigenesis. *Breast Cancer Research* 10: R21.
64. Habashy HO, Rakha EA, Aleskandarany M, Ahmed MAH, Green AR, et al. (2011) FOXO3a nuclear localisation is associated with good prognosis in luminal-like breast cancer. *Breast Cancer Research and Treatment* 129: 11–21.
65. Knight 3rd WA, Osborne CK, Yochmowitz MG, McGuire WL (1980) Steroid hormone receptors in the management of human breast cancer, *Ann Clin Res* 12: 202–207.
66. Ignatiadis M, Sotiriou C (2013) Luminal breast cancer: from biology to treatment. *Nat Rev Clin Oncol* 10: 494–506.
67. Koo CY, Muir KW, Lam EW (2012) FOXM1: From cancer initiation to progression and treatment. *Biochim Biophys Acta* 1819: 28–37.
68. Wang M, Gartel AL (2011) Micelle-encapsulated thiostrepton as an effective nanomedicine for inhibiting tumor growth and for suppressing FOXM1 in human xenografts. *Mol Cancer Ther* 10: 2287–2297.