MOLECULAR ONCOLOGY 8 (2014) 1588–1602

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

Integrative genomic and transcriptomic characterization of papillary carcinomas of the breast

Salvatore Piscuoglio[a,1], Charlotte K.Y. Ng[a,1], Luciano G. Martelotto[a], Carey A. Eberle[a], Catherine F. Cowell[a], Rachael Natrajan[b], François-Clement Bidard[a,c], Leticia De Mattos-Arruda[a], Paul M. Wilkerson[b], Odette Mariani[c], Anne Vincent-Salomon[c,**], Britta Weigelt[a], Jorge S. Reis-Filho[a,*]

[a]Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York 10065, NY, USA
[b]The Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London, UK
[c]Institut Curie, Department of Biopathology and INSERM U934, Paris, France

ARTICLE INFO

Article history:
Received 9 March 2014
Received in revised form
8 June 2014
Accepted 17 June 2014
Available online 26 June 2014

Keywords:
Single nucleotide polymorphism arrays
Microarrays
RNA-sequencing
Fusion gene
Mutation

ABSTRACT

Papillary carcinoma (PC) is a rare type of breast cancer, which comprises three histologic subtypes: encapsulated PC (EPC), solid PC (SPC) and invasive PC (IPC). Microarray-based gene expression and Affymetrix SNP 6.0 gene copy number profiling, and RNA-sequencing revealed that PCs are luminal breast cancers that display transcriptomic profiles distinct from those of grade- and estrogen receptor (ER)-matched invasive ductal carcinomas of no special type (IDC-NSTs), and that the papillary histologic pattern is unlikely to be underpinned by a highly recurrent expressed fusion gene or a highly recurrent expressed mutation. Despite displaying similar patterns of gene copy number alterations, significant differences in the transcriptomic profiles of EPCs, SPCs and IPCs were found, and may account for their different histologic features.

1. Introduction

Breast cancer is a heterogeneous disease comprising numerous distinct entities with different biological features and clinical behaviors (Reis-Filho and Pusztai, 2011; Weigelt and Reis-Filho, 2009). Invasive ductal carcinomas of no special type (IDC-NST), the commonest histologic type of breast cancer, have been shown to be heterogeneous at the

Abbreviations: PC, papillary carcinoma; EPC, encapsulated papillary carcinoma; SPC, solid papillary carcinoma; IPC, invasive papillary carcinoma.

* *Corresponding author*. Tel.: +1 212 639 8054; fax: +1 212 639 2502.
** *Corresponding author*. Department of Pathology, Institut Curie, Department of Biopathology and INSERM U934, 26 rue d'ULM, 75248 Paris cédex 05, France. Tel.: + 33 (0)1 443242 15; fax: +33 (0)1 531040 10.
E-mail addresses: anne.salomon@curie.fr (A. Vincent-Salomon), reisfilj@mskcc.org (J.S. Reis-Filho).
[1] These authors contributed equally to this work.
http://dx.doi.org/10.1016/j.molonc.2014.06.011
1574-7891/

transcriptomic and genomic levels, and can be classified into several molecular subtypes (Cancer Genome Atlas, 2012; Reis-Filho and Pusztai, 2011). The 'intrinsic' gene classification is the most widely used molecular taxonomy for IDC-NSTs, and classifies breast cancers into luminal A, luminal B, HER2-enriched and basal-like subtypes (Parker et al., 2009; Perou et al., 2000). These subtypes have been shown not only to be associated with distinct clinical outcomes, but also with different risk factors, clinicopathologic features and repertoires of genetic aberrations (Cancer Genome Atlas, 2012; Parker et al., 2009; Reis-Filho and Pusztai, 2011).

The complexity and diversity of breast cancers have also been documented at the histologic level (Weigelt et al., 2010c; Weigelt and Reis-Filho, 2009), however this information has not been fully explored in molecular subtyping studies, which have primarily focused on IDC-NSTs. Based on the cytological and architectural features of invasive breast cancers, the World Health Organization (WHO) recognizes the existence of 21 special histologic subtypes in addition to IDC-NSTs (Lakhani et al., 2012). Genomic and transcriptomic analyses of special histologic types of breast cancer conducted by our group and others (Bertucci et al., 2008; Duprez et al., 2012; Geyer et al., 2010; Gruel et al., 2010; Horlings et al., 2013; Lacroix-Triki et al., 2010; Lopez-Garcia et al., 2010b; Marchio et al., 2009, 2008; Vincent-Salomon et al., 2007; Weigelt et al., 2009a, 2010b, 2008, 2009b; Wetterskog et al., 2012) have demonstrated that tumors from each of the special histologic types of breast cancer are more homogeneous amongst themselves than IDC-NSTs. In addition, some of the histologic special types have been shown to be driven by recurrent fusion genes resultant of chromosomal translocations. For example, secretory carcinomas of the breast are characterized by the t(12; 15) translocation, which results in the formation of the *ETV6-NTRK3* fusion gene (Tognon et al., 2002), and >90% of adenoid cystic carcinomas of the breast display the t(6; 9) translocation, which leads to the fusion of *MYB* with *NFIB* (Persson et al., 2009; Wetterskog et al., 2012).

Papillary carcinoma (PC) is a rare (<1%) special histologic type of breast cancer that often affects postmenopausal women and has an overall favorable outcome (Grabowski et al., 2008; Pal et al., 2010; Rakha et al., 2011; Solorzano et al., 2002; Weigelt et al., 2010a). PCs comprise a morphologically heterogeneous group of lesions, all of which share a growth pattern characterized by the presence of arborescent fibrovascular stalks lined by a layer of neoplastic epithelial cells devoid of an intervening myoepithelial cell layer, a feature that distinguishes them from benign intraductal papillomas and papillary carcinomas *in situ* (Collins and Schnitt, 2008; Hill and Yeh, 2005; Pal et al., 2010; Weigelt et al., 2010a). Papillary neoplasms of the breast include three histologic subtypes, namely encapsulated papillary carcinoma (EPC), solid papillary carcinoma (SPC) and invasive papillary carcinoma (IPC). EPC is a well-circumscribed lesion where the involved duct is surrounded by a thick fibrous capsule; in EPCs, the neoplastic cells are arranged in papillary fronds in the majority of cases, however areas with cribriform and/or solid patterns are not uncommonly found (Lakhani et al., 2012; Wynveen et al., 2011). SPC is also a well-circumscribed lesion that is densely cellular and composed of expansile nodules of epithelial cells, and IPC comprises papillary lesion with neoplastic cells arranged in finger-like projections with clear invasion into adjacent stroma. Although the classification of EPCs and SPCs as invasive or *in situ* disease still remains a matter of controversy, these tumors have the potential to disseminate to axillary lymph nodes and, albeit rarely, distant metastatic deposits of PCs have been documented (Rakha et al., 2011; Wynveen et al., 2011). Based on these observations, it has been proposed that EPCs and SPCs should be considered forms of invasive breast cancer with excellent outcome (Lakhani et al., 2012; Rakha et al., 2011; Wynveen et al., 2011).

Our group has previously demonstrated that PCs are preferentially estrogen receptor (ER) and progesterone receptor (PR) positive, lack *HER2* gene amplification, and display relatively simple genomes in terms of their repertoires of gene copy number aberrations (Duprez et al., 2012). In addition, the similarities in the gene copy number profiles of PCs and grade- and ER-matched IDC-NSTs have led us to suggest that PCs may be best positioned as part of the spectrum of ER-positive IDC-NSTs, rather than a distinct entity (Duprez et al., 2012). On the other hand, the transcriptomic characteristics of PCs and whether these tumors would differ from IDC-NSTs at the gene expression level remain to be determined. Therefore, the primary aims of this study were i) to investigate whether PCs would constitute a molecular entity distinct from grade- and ER-matched IDC-NSTs at the transcriptomic level, and ii) to define whether PCs would be driven by recurrent fusion genes or pathognomonic mutations. In addition, we carried out an exploratory, hypothesis-generating analysis to investigate whether EPCs, SPCs and IPCs would display distinct transcriptomic and genomic profiles.

2. Materials and methods

2.1. Samples

Nineteen PCs of the breast were retrieved from Institut Curie, Paris, France (from 1995 to 2009). In this study, we included PCs from patients diagnosed and managed in the above institution, whose tumors were <5 cm and who had no clinical and/or radiological evidence of distant metastases. Exclusion criteria were (a) patients with multiple tumors, either ipsi- or contra-lateral, (b) patients who received neoadjuvant chemotherapy, (c) patients for whom all histologic slides and blocks were not available for review, (d) tumors not consistent with the final diagnosis of EPC, SPC or IPC, and (e) tumors whose frozen samples contained <50% of tumor cell content. For sixteen cases, both DNA and RNA of sufficient quality and quantity were available for microarray-based gene expression and copy number profiling; for three cases only RNA could be extracted due to limited frozen tissue availability (Supplementary Table S1). Samples were anonymized prior to analysis and the study was approved by local research ethics committees of the authors' institutions. All cases were independently reviewed by two pathologists (AV-S and JSR-F), who subtyped the tumors into EPC, SPC and IPC following the WHO criteria (Lakhani et al., 2012), and graded the tumors according to the Nottingham grading system (Elston and Ellis, 1991). Histologic grade- and ER-matched IDC-NSTs, whose

frozen samples contained >50% of tumor cells, were retrieved from the files of Institut Curie, Paris, France. Sixteen IDC-NSTs were matched 1:1 according to histologic grade and ER status with the PCs included in this study (Table 1). For gene expression profiling analysis, power calculations employing the approach described by Dobbin et al. (Dobbin et al., 2008) revealed that 16 samples of PCs and IDC-NSTs would result in the identification of genes significantly differentially expressed between the groups with a tolerance = 0.10.

Eight cases of PCs (3 IPCs, 3 EPCs, 2 SPCs), from which sufficient RNA of acceptable quality was available, were subjected to paired-end massively parallel RNA-sequencing (Supplementary Table S1), with the aim of identifying highly recurrent fusion genes and pathognomonic mutations. Power calculations, assuming that if PCs were driven by a recurrent fusion gene in a way akin to secretory carcinomas (which harbor the *ETV6-NTRK3* fusion gene in >95% of cases) (Tognon et al., 2002) or adenoid cystic carcinomas of the breast (which harbor the *MYB-NFIB* fusion gene in >90% of cases) (Persson et al., 2009; Wetterskog et al., 2012), and that a 'pathognomonic' driver event (i.e. pathognomonic fusion gene or mutation) would be present in at least ≥70% of cases (an estimate that is conservative)), revealed that based on a binomial distribution, by sequencing 8 samples we would have been able to identify a recurrent driver event (i.e. in two or more cases) with >95% statistical power. Furthermore, a 'pathognomonic' driver event present in ≥50% of cases would be detectable with >80% power if 8 samples were subjected to RNA-sequencing.

2.2. Immunohistochemistry

Immunohistochemical profiles of the included cases of PCs and grade- and ER-matched IDC-NSTs were assessed on 3 μm thick sections, using antibodies against ER, PR and HER2 as previously described (Duprez et al., 2012). Positive and negative controls were included in each experiment. The results of immunohistochemical analyses were interpreted by two pathologists (AV-S and JSR-F). Antibody clones, dilutions, antigen retrieval methods, scoring systems and cut-offs used are described in Supplementary Table S2.

2.3. Nucleic acid extraction

DNA and RNA were extracted using the DNeasy Blood and Tissue Kit (Qiagen, Hamburg, Germany) and Trizol (Invitrogen, Life Technologies, Paisley, UK) respectively, according to the manufacturers' guidelines. DNA and RNA quantity and quality were analyzed with the Qubit Fluorometer (Invitrogen) and the Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA). Only samples with RNA integrity (RIN) >6 (Bioanalyzer, Agilent, Santa Clara, CA, USA) were used for downstream analyses.

2.4. Copy number analysis

Gene copy number profiling was performed using the SNP Array 6.0 platform (Affymetrix, Santa Clara, CA, USA) as per manufacturer's instructions. Raw data were analyzed using the CRMAv2 algorithm in the aroma.affymetrix package (Bengtsson et al., 2009) and ABSOLUTE (Carter et al., 2012). In all cases, ABSOLUTE defined estimates of tumor cellularity were >50% (median 71%, range 57%–93%). Log_2 ratios obtained from CRMAv2 were then smoothed using the circular binary segmentation (cbs) algorithm in the DNAcopy package (Olshen et al., 2004) in R. cbs-smoothed Log_2 ratios were then used in all subsequent analyses to define gains, losses and amplifications. Homozygous deletions were defined as a cbs-smoothed Log_2 ratio <−1 and using ABSOLUTE; only homozygous deletions identified by both methods were considered valid. Low-level gain was defined as a cbs-smoothed Log_2 ratio of >0.15 and ≤0.5, and high-level gains/amplification were defined as a Log_2 ratio >0.5. To identify statistically significant

Table 1 – Histologic grade and intrinsic molecular subtype of 16 papillary carcinomas and 16 grade- and ER-matched IDC-NSTs.

Papillary carcinomas					Invasive ductal carcinomas of no special type			
Sample ID	Histologic subtype	Histologic grade	ER-status	Molecular subtype	Sample ID	Histologic grade	ER-status	Molecular subtype
PAPI11	IPC	1	Positive	Luminal A	IDC13	1	Positive	Luminal A
PAPI12	EPC	1	Positive	Luminal A	IDC14	1	Positive	Luminal B
PAPI20	EPC	1	Positive	Luminal A	IDC16	1	Positive	Luminal A
PAPI13	EPC	1	Positive	Basal-like	IDC17	1	Positive	Luminal A
PAPI2	IPC	1	Positive	Luminal A	IDC19	1	Positive	Luminal A
PAPI18	EPC	1	Positive	Luminal A	IDC20	1	Positive	Luminal A
PAPI4	EPC	1	Positive	Luminal A	IDC7	1	Positive	Luminal B
PAPI25	IPC	2	Positive	Luminal A	IDC2	2	Positive	Luminal A
PAPI21	SPC	2	Positive	Luminal B	IDC21	2	Positive	Luminal A
PAPI15	SPC	2	Positive	Luminal B	IDC3	2	Positive	Luminal A
PAPI22	EPC	2	Positive	Luminal A	IDC4	2	Positive	Luminal A
PAPI24	IPC	2	Positive	Luminal A	IDC5	2	Positive	Luminal A
PAPI14	IPC	2	Positive	Luminal B	IDC8	2	Positive	Luminal B
PAPI23	SPC	3	Positive	Luminal B	IDC10	3	Positive	HER2
PAPI26	EPC	3	Positive	Luminal A	IDC11	3	Positive	Luminal B
PAPI27	SPC	3	Positive	Luminal B	IDC9	3	Positive	Luminal B

ER, estrogen receptor; IDC-NST, invasive ductal carcinoma of no special type.

differences between the genomic profiles of the different subtypes of PCs, we compared the copy number profiles of EPC, SPC and IPC using categorical data and a multi-Fisher's exact test adjusted by the false discovery rate as previously described (Duprez et al., 2012). Hierarchical clustering analysis was performed with categorical states (i.e. gains, losses, high-level gains/amplifications) using Euclidean distance and the Ward's algorithm.

2.5. Gene expression profiling

Gene expression profiling was performed with the HumanHT-12 v4 Expression BeadChip Kit (Illumina Inc, San Diego, CA, USA) as per manufacturer's instructions. Raw data were robust-spline normalized using the lumi package in R (Du et al., 2008). Illumina transcript probes with detection p-value<0.01 in fewer than 20 samples of both PCs and IDC-NSTs or without unambiguous mapping information were excluded, resulting in a set of 17,296 probes for further analysis.

For intrinsic subtyping using the PAM50 classification (Parker et al., 2009; Weigelt et al., 2010d), the gene expression profiles of PCs and IDC-NSTs were normalized together with 997 breast cancers of the METABRIC discovery cohort, as described by Curtis et al. (Curtis et al., 2012). Samples were then assigned to one of the five subtypes using Spearman's rank correlation and the PAM50 centroids curated for the Illumina platform strictly as described by Curtis et al. (Curtis et al., 2012) and following the University of North Carolina (UNC) guidelines (https://genome.unc.edu/pubsup/breast-GEO/Guide%20to%20Intrinsic%20Subtyping%209-6-10.pdf).

For unsupervised clustering analysis, we included only probes with a median absolute deviation (MAD) >0.6 and excluded replicate probes mapping to the same genes by keeping only the probe with the greatest MAD, resulting in a set of 2310 probes for clustering. Gene expression values for the series of PCs and IDC-NSTs were median centered prior to clustering. Hierarchical clustering was performed using the Ward's or the complete linkage algorithm with Pearson correlation as the distance metric. The stability of the clusters was estimated by performing 1000 iterations of bootstrapping using the R package pvclust (Suzuki and Shimodaira, 2006).

To determine the genes significantly differentially expressed between PCs and IDC-NSTs, and between different subtypes of PCs, we used the significance analysis of microarrays (SAM) software implemented in the samr R package. We included probes with a MAD >0.3 and excluded replicate probes as described above, resulting in a set of 11,108 probes for SAM. Probes were median-centered prior to SAM analysis and 1000 permutations were performed to estimate the false discovery rate (FDR).

The raw and processed data are available on GEO (GSE55640), and the R script and code are available in the Supplementary materials.

2.6. Integration of copy number and expression data

To identify genes whose expression levels correlate with copy number changes, cbs-smoothed Log_2 ratios from SNP 6.0 data were used to assign the SNP 6.0 states for each of the 17,296 probes in the gene expression dataset using the median values for all probes that overlap with the genomic position of each gene. This resulted in a 1:1 matrix of expression ratios and SNP 6.0 cbs values, which were subsequently employed for downstream statistical analysis.

To define genes that were up-regulated when gained or amplified, and down-regulated when lost, we performed a Mann–Whitney U test using categorical gene copy number states (i.e. gain vs. no gain, loss vs. no loss or amplification vs. no amplification) as the grouping variable and the expression of genes as the dependent variable. p-values were corrected using Benjamini and Hochberg adjustment (Benjamini and Hochberg, 1995) to minimize type I or alpha errors, and adjusted p-values<0.05 were considered significant.

2.7. Pathway analysis

Significantly regulated pathways and networks in the gene expression data were determined using the Ingenuity Pathway Analysis (IPA) program (http://www.ingenuity.com) and the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov). For IPA analysis, the fold difference of transcripts differentially expressed identified by SAM was mapped to networks available in the Ingenuity database and ranked by score. The score indicates the likelihood of the genes in a network being found together due to chance. Using a 99% confidence level, scores of ≥ 3 are considered significant (Lopez-Garcia et al., 2010b).

2.8. Paired-end massively parallel RNA-sequencing and fusion gene identification

Paired-end massively parallel RNA-sequencing (2x 54bp cycles) was performed according to the standard Illumina mRNA paired-end library protocol on a Genome Analyzer II (Illumina) as previously described (Natrajan et al., 2014).

Data were aligned to the genome using Bowtie 2 (Langmead and Salzberg, 2012). Chimerascan and deFuse were used to identify mate-pairs supporting novel chimeric transcripts as previously described (Iyer et al., 2011; McPherson et al., 2011; Natrajan et al., 2014). High-confidence chimeric transcripts detected using both algorithms were considered as potential candidates. Nominated in-frame fusion gene candidates identified by RNA-sequencing were prioritized for validation in the index cases by RT-PCR. The split sequences (sequences encompassing the breakage point/fusion junctions) were used as reference to design primer sets for each fusion gene pair (Supplementary Table S3). For this, 100 ng of total RNA was reverse transcribed using Superscript III (Invitrogen) and PCR was performed using the GoTaq® Green Master Mix Kit (Promega, USA, Madison, WI, USA).

2.9. Quantitative reverse transcriptase-PCR (qRT-PCR)

For each sample, triplicate reactions containing 100 ng RNA were subjected to reverse transcription as described previously (Arriola et al., 2008) using Superscript III (Invitrogen). qRT-PCR was performed for *RET, IGFBP5, TMPRSS6, ERP27, NFKBIZ and GBP3* (Assay on Demand IDs: Hs01120030_m1, Hs00181213_m1, Hs00542184_m1, Hs00328370_m1,

Hs00230071_m1, Hs00544385_m1, respectively; Applied Biosystems, Foster City, USA) using TaqMan® chemistry on the StepOnePlus™ Real-Time PCR System (Applied Biosystems), based on the standard curve method (Weigelt et al., 2010b). Target gene expression levels were normalized to the geometric mean of two reference genes (*TFRC* and *MRPL19*; Hs00174609_m1-TFRC, Hs00608522_g1-MRPL19) (Natrajan et al., 2012). All statistical comparisons were performed using a Mann–Whitney *U* test (GraphPad Prism 6).

2.10. *Identification of mutations from RNA-sequencing data*

For the identification of mutations from RNA-sequencing data, the reads were aligned using Bowtie 2 (Langmead and Salzberg, 2012). PCR duplicates were removed using samtools (Li et al., 2009). Local re-alignment and base quality recalibration were performed using the Genome Analysis Tool Kit (GATK) (McKenna et al., 2010). Mutations were called using the GATK Haplotype Caller (McKenna et al., 2010), where a minimum of 5 variant reads, 10× total read depth, and a genotype quality (GQ) of 40 were required to call a mutation. To remove germline polymorphisms and sequencer-specific artifacts, we filtered out mutation calls from 13 unrelated, normal female RNA samples, whose RNA-sequencing results were generated on the same sequencer using the same protocols. Only single nucleotide variants (SNVs) or insertions and deletions (indels) supported by at least 2 reads supporting the reference alleles were included, as those devoid of a reference allele were highly enriched for germline variations. Further, we filtered out mutations catalogued in dbSNP as having >1% global population frequency and all germline variants identified in the 1000 Genomes Project. The functional effects of missense mutations were predicted using Mutation Assessor (Reva et al., 2011), CHASM (Carter et al., 2009), FATHMM (Shihab et al., 2013) and Polyphen-2 (Adzhubei et al., 2010)/Mutation Taster (Schwarz et al., 2010). A mutation was considered potentially pathogenic if at least two predictors classified it as pathogenic. MutationTaster and Polyphen-2 were used in combination, given that these two algorithms are underpinned by the same bioinformatic principles and their results, albeit not identical, are remarkably similar. Genes recurrently mutated in PCs and likely to constitute somatic mutations were validated at the genomic level using Sanger sequencing. Briefly, 50 ng tumor DNA was amplified with specific primers (Supplementary Table S4) and sequencing performed using the BigDye Terminator v 1.1 Kit (Applied Biosystems), as previously described (Hernandez et al., 2012).

3. Results

3.1. *Papillary carcinomas are preferentially ER-positive, luminal breast cancers*

Histopathologic review of the 16 PCs subjected to SNP6 and gene expression profiling revealed that 7 (44%), 6 (37%) and 3 (19%) cases were of histologic grades 1, 2 and 3, respectively, and 4 (25%), 7 (44%) and 5 (31%) cases were of SPC, EPC, and IPC subtype, respectively (Table 1). Immunohistochemical analysis revealed that all PCs were ER-positive and lacked HER2 expression in accord with our previous results (Duprez et al., 2012) (Table 1 and Supplementary Table S1); 81% of PCs were ER-positive/PR-positive/HER2-negative and 19% of cases were ER-positive/PR-negative/HER2-negative (Supplementary Table S1). PAM50 subtyping classified 10 (63%), 5 (31%) and 1 (6%) PCs as of luminal A, luminal B and basal-like subtypes, respectively. All but one EPC and all but one IPC were classified as of luminal A subtype, whereas all SPCs were classified as of luminal B (Table 1).

3.2. *Papillary carcinomas display distinct transcriptomic profiles but similar patterns of DNA copy number alterations as compared to grade- and ER-matched IDC-NSTs*

Unsupervised hierarchical clustering of PCs and grade- and ER-matched IDC-NSTs revealed that PCs preferentially clustered together (two-tailed Fisher's exact test, p-value = 0.002; Figure 1), irrespective of the clustering algorithm employed. These data suggest that although PCs and grade- and ER-matched IDC-NSTs are similar at the gene copy number level (Duprez et al., 2012), their global patterns of gene expression are distinct.

SAM analysis further supported this observation, given that 8.3% of the significantly regulated probes (795 unique genes) were differentially expressed between PCs and histologic grade- and ER-matched IDC-NSTs (FDR <10%), of which 780 unique genes were expressed at significantly lower levels in PCs than in grade- and ER-matched IDC-NSTs (Supplementary Table S5). IPA and DAVID gene ontology and pathway analysis employing these 780 genes revealed a significant enrichment for genes involved in cellular growth and proliferation (e.g. *KIT, ALDH1A3, ELF4, CCND2* and *MAP4K4*, right-tailed Fisher exact test p-value<0.0001, Figure 2A, Supplementary Table S5). This is consistent with the observation that PCs are a form of ER-positive breast cancer with an excellent outcome (Rakha et al., 2011), given that the expression levels of proliferation-related genes have been shown to constitute one of the strongest predictors of outcome of ER-positive breast cancers (Reis-Filho and Pusztai, 2011). Other genes expressed at significantly lower levels in PCs than in grade- and ER-matched IDC-NSTs included those involved in cell assembly and organization (e.g. *LAMB1*, *ACTN1* and collagen genes such as *COL1A2*, *COL6A1*, *COL8A1*, COL12A1, right-tailed Fisher exact test p-value<0.0001, Figure 2B) and cellular movement and migration (e.g. *MMP3*, *MMP7* and THBS4, right-tailed Fisher exact test p-value<0.0001, Figure 2C). Consistent with these observations, DAVID KEGG pathway analysis revealed that the "cell adhesion molecules" (CAMs) pathway was the most significantly enriched for genes expressed at lower levels in PCs than in IDC-NSTs (data not shown). By contrast, 15 genes involved in cellular homeostasis (e.g. *QSOX1*) and angiogenesis (e.g. *VEGFA*, *ESM1*, SLC4A11 and *VCAM1*) were expressed at significantly higher levels in PCs than in grade- and ER-matched IDC-NSTs (Supplementary Table S5).

We have previously demonstrated that the patterns of gene copy number alterations found in PCs are relatively simple and

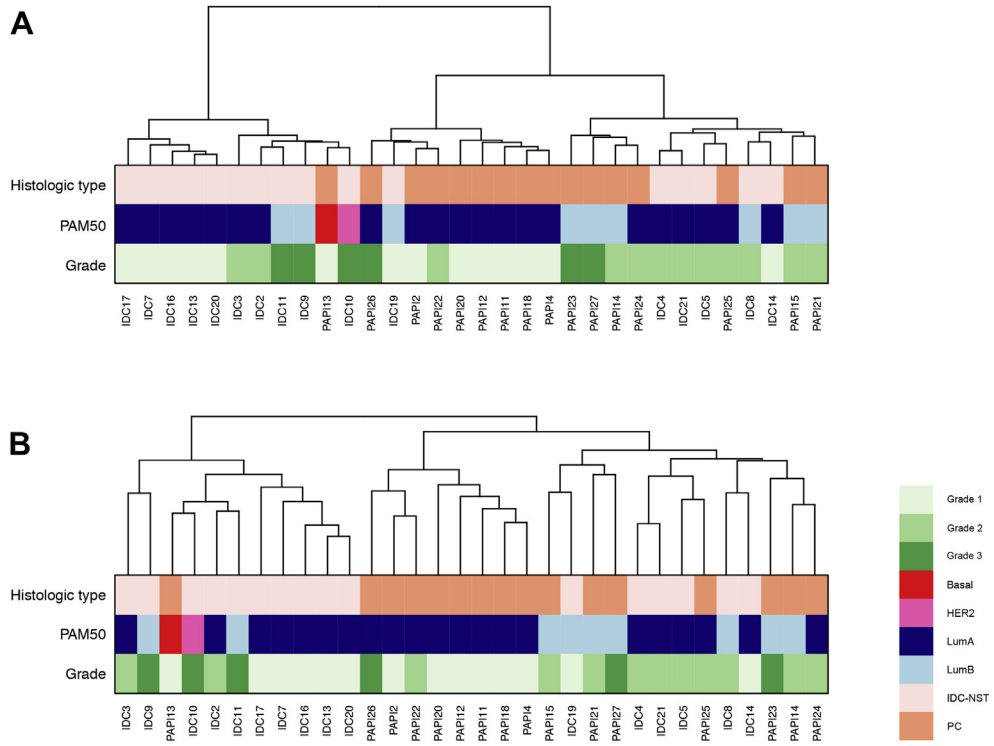

Figure 1 – Unsupervised hierarchical clustering of gene expression data of 16 papillary carcinomas and 16 histologic grade- and ER-matched invasive ductal carcinomas of no special type. A) Clustering was performed with 2310 probes using Ward's algorithm and Pearson correlation as distance metric. B) Clustering was performed with 2310 probes using complete-linkage algorithm and Pearson correlation as distance metric. IDC-NST, invasive ductal carcinoma of no special type; PAM50, intrinsic molecular subtype using the microarray version of the PAM50 approach; PC, papillary carcinoma.

similar to those found in IDC-NSTs of the same histologic grade and ER-status (Duprez et al., 2012). Here, using a platform with a higher resolution, we confirmed our previous observations that PCs displayed genomic features consistent with those of low-grade, ER-positive IDC-NSTs (i.e. 16q losses, 16p gains and 1q gains; Figure 3 and Supplementary Figure S1). After excluding regions mapping to known copy number polymorphisms (http://dgv.tcag.ca/dgv/app/home), we observed recurrent amplifications in 11q13.3 ($n = 2$) and 8p11.23-p11.22 ($n = 2$) loci (Supplementary Figure S1 and Supplementary Table S6). Consistent with our previous observations (Duprez et al., 2012), *HER2* and *ESR1* gene amplification was not observed in the PCs studied.

To define genes that are up-regulated and gained, or down-regulated and lost in PCs, we performed an integrative transcriptomic and genomic analysis. We identified 214 genes that were significantly over-expressed when gained (Mann–Whitney *U* test, adjusted p-value<0.05), and these were significantly enriched for cellular functions including "Cellular Growth and Proliferation" (e.g. *AGO2* and *CAPN2*, right-tailed Fisher's exact test p-value<0.01, IPA) and "Cell Death and Survival" (e.g. *LMNA*, *STK3* and *PTK2*; right-tailed Fisher's exact test p-value<0.01, IPA, and p-value<0.01, DAVID; Supplementary Table S7; Supplementary Figure S2A). On the other hand, 275 genes were found to be significantly down-regulated when lost (Mann–Whitney *U* test, adjusted p-value<0.05). Functional annotation of these genes revealed a significant enrichment for genes involved in "Cell cycle progression" (e.g. *ASAH1 and BIRC2*, right-tailed Fisher's exact test p-value<0.01, IPA) and "Cellular Death and Survival" (e.g. *CDT1 and MAX*, right-tailed Fisher exact test p-value<0.01, IPA; Supplementary Table S7; Supplementary Figure S2B).

We further sought to identify potential amplicon drivers in PCs on the basis that amplicons drivers are frequently overexpressed when amplified. We identified 182 genes expressed at significantly higher levels when amplified in PCs (Mann–Whitney *U* test, p-value<0.05; Supplementary Table S8). This list included genes mapping to the 8p11.2-p12 amplicon (i.e. *PROSC*) and the 11q13 amplicon (i.e. *CTTN* and *FADD*), which have been previously shown to be overexpressed when amplified in IDC-NSTs (Adelaide et al., 2007; Chin et al., 2006, 2007; Hu et al., 2009; Mackay et al., 2009; Natrajan et al., 2009) (Supplementary Figure S3). In addition, *FADD* and *ADAM9* have also been shown to have significant overexpression–amplification correlation in other histologic special types of breast cancer (Horlings et al., 2013). When p-values were corrected for multiple testing, however, only 16 genes were shown to be significantly overexpressed when amplified, including *RSF1*, which plays a role in the RSF chromatin-remodeling complex.

Taken together, our analyses demonstrate that the transcriptomic profiles of PCs differ from those of grade-matched ER-positive IDC-NSTs, with down-regulation of proliferation-related, cell assembly and organization and

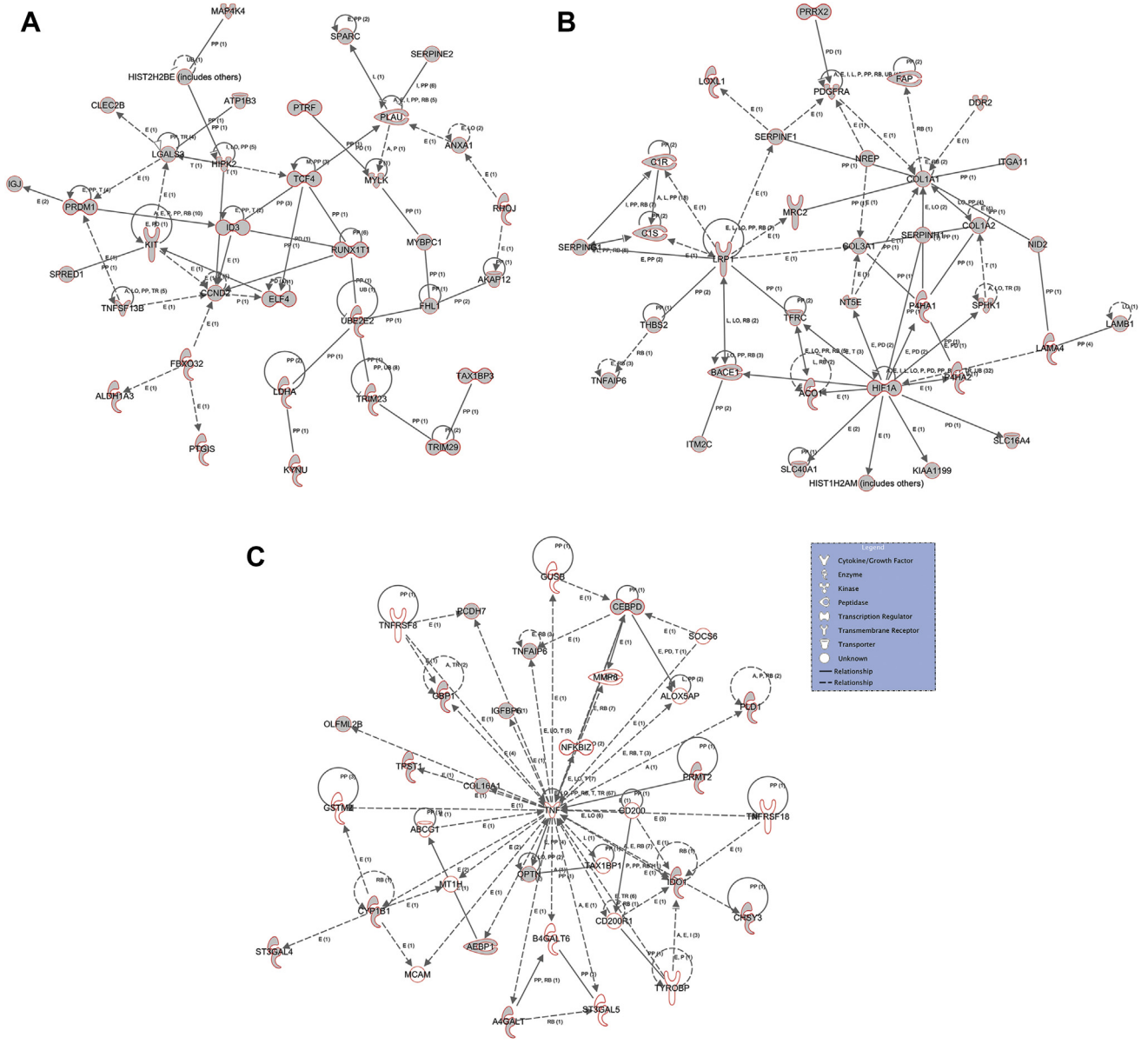

Figure 2 – Ingenuity Pathway Analysis of 780 unique genes expressed at significantly lower levels in 16 papillary carcinomas than in 16 grade- and ER-matched invasive ductal carcinomas of no special type. (A) Genes of the 'Cellular Growth and Proliferation' were down-regulated in PCs in comparison with IDCs-NSTs. (B) Cell assembly and organization genes were down-regulated in PCs versus IDC-NSTs. (C) Genes involved in cellular movement and migration were down-regulated in PCs compared with grade- and ER-matched IDC-NSTs. IDC-NST, invasive ductal carcinoma of no special type; PC, papillary carcinoma.

cellular movement and migration genes, and overexpression of genes related to homeostasis and angiogenesis.

3.3. *Papillary carcinomas are not underpinned by a highly recurrent fusion gene*

To determine whether PCs of the breast are driven by a highly recurrent fusion gene, we subjected 8 PCs to paired-end massively parallel RNA-sequencing. Analysis using deFuse and Chimerascan (Iyer et al., 2011; McPherson et al., 2011) revealed a total of 1156 fusion transcripts (involving 999 unique gene combinations) and 396 fusion transcripts (involving 339 unique gene combinations), respectively, of which 33 unique gene combinations were identified by both deFuse and Chimerascan (Supplementary Table S9).

We next sought to identify potential gain-of-function fusion genes that may underpin PCs. In a way akin to the oncogenic fusion genes identified in special histologic types of breast cancer, leukemias, lymphomas and sarcomas, we prioritized fusion transcripts that were identified by both algorithms, involved open reading frames with known associated functions, and that harbored intact functional domains. Using the above criteria, we identified 14 in-frame fusion transcripts, 12 of which were successfully validated by RT-PCR (Figure 4,

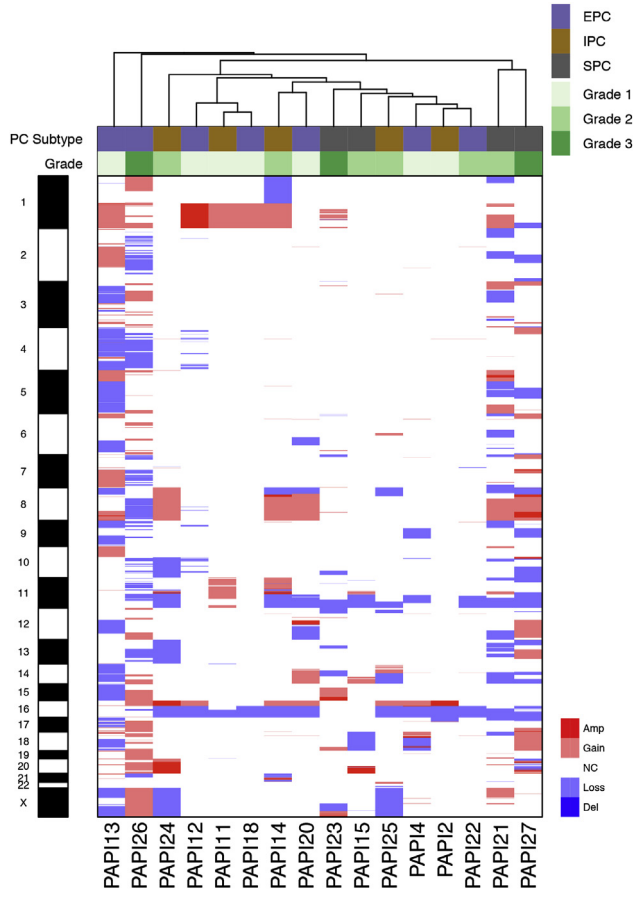

Figure 3 – Unsupervised hierarchical clustering of gene copy number profiles of papillary carcinomas. Clustering was performed using SNP6-derived categorical states (i.e. gains, losses and amplifications) of 16 PCs, using Ward's algorithm and Euclidean distance. Amp, amplification; Del, homozygous deletion; EPC, encapsulated papillary carcinoma; Gain, copy number gain; IPC, invasive papillary carcinoma; Loss, copy number loss; NC, no copy number change; PC, papillary carcinoma; SPC, solid papillary carcinoma.

Supplementary Figure S4; Supplementary Table S10). These validated fusion transcripts did not map to regions harboring amplifications and were only detected and validated in the index cases. Only one of these confirmed fusion transcripts, *MAPKAPK3-HEMK1*, harbored intact functional domains encoded by the first 4 exons of the protein kinase *MAPKAPK3* and the last 7 exons of the methyltransferase *HEMK1* (Figure 4D). In addition to the kinase domain, this chimeric transcript has a conserved DNA methylase domain, which has been reported to methylate the mitochondrial translation release factor *MTRF1L* (Ishizawa et al., 2008). Taken together, our results demonstrate that PCs are not underpinned by a highly recurrent fusion gene, at variance with some other special histological types of breast cancer.

3.4. *RNA-sequencing based mutational analysis*

To assess the constellations of expressed mutations in PCs, mutational analysis of the RNA-sequencing data was performed. Given the challenges in the *de novo* identification of mutations based on the analysis of RNA-sequencing data, we have focused on the 35 significantly mutated genes in common types of breast cancers as reported by TCGA (Cancer Genome Atlas, 2012). This analysis revealed that five of these genes were mutated in PCs (*PIK3CA*, *TP53*, *MAP3K1*, *AKT1*, and *CTCF*; Table 2). The *PIK3CA* (H1047R) mutation (case PAPI13) and *TP53* (R273H) mutation (case PAPI17) are known hotspot mutations (Muller and Vousden, 2013; Samuels et al., 2004). To identify further potentially pathogenic mutations, we defined genes that were recurrently mutated in the 8 PCs analyzed and predicted to be pathogenic using five mutation function predictor algorithms (i.e. Mutation Assessor (Reva et al., 2011), CHASM (Carter et al., 2009), FATHMM (Shihab et al., 2013) and Polyphen-2 (Adzhubei et al., 2010)/Mutation Taster (Schwarz et al., 2010)). After filtering out single nucleotide variants and insertions and deletions present in dbSNP at a >1% global population frequency, and all germline variants identified in the 1000 Genomes Project, we identified 156 potential somatic mutations (Supplementary Table S11). Out of these, *LAMA5* and *MACF1* were found to be recurrently mutated (i.e. in two samples each) in the PCs analyzed and considered to be potentially pathogenic mutations by at least two of the mutation function prediction algorithms employed in this study. These mutations were validated in the index cases by Sanger sequencing (Supplementary Figure S5). Mutations affecting *LAMA5* and *MACF1* were not restricted to papillary carcinomas, however, as also 1.1% and 2.4% of the common types of breast cancers in the TCGA cohort harbored a mutation in these genes, respectively (Gao et al., 2013) (www.cBioPortal.org, accessed March 8th 2014; Table 2). This analysis suggests that the papillary phenotype is not underpinned by a highly recurrent expressed mutation (i.e. present in >50% of samples) affecting a gene expressed in PCs.

3.5. *Transcriptomic and genomic analysis of papillary carcinoma histologic subtypes*

We next carried out a hypothesis-generating, comparative exploratory analysis of the transcriptomic profiles of PC subtypes to ascertain whether histologic subtypes of PCs (i.e. EPC, SPC and IPC) would have distinctive transcriptomic profiles. Unsupervised hierarchical clustering demonstrated that EPCs preferentially clustered together (Fisher's exact test, $p = 0.0406$, Figure 5A), irrespective of the clustering algorithm employed (Figure 5B). It should be noted, however, that all EPCs were of histologic grade 1, and a significant enrichment for tumors of histologic grade 1 (Fisher's exact test, $p = 0.04$) was also observed in the cluster enriched for EPCs.

SAM analysis using 11,108 probes revealed that less than 0.2% of genes were differentially expressed between EPCs and SPCs/IPCs (FDR <10%; Supplementary Table S12). Functional annotation of these genes with both DAVID and IPA revealed that several genes involved in cell migration, such as *PLAT* and *CTSF*, were down-regulated in EPCs in comparison with SPCs and IPCs. This finding is in line with a less invasive phenotype observed in EPCs. Furthermore a comparison between SPCs and EPCs suggested that, consistent with the reported neuroendocrine phenotype of SPCs, genes related to neuroendocrine differentiation in human cancers,

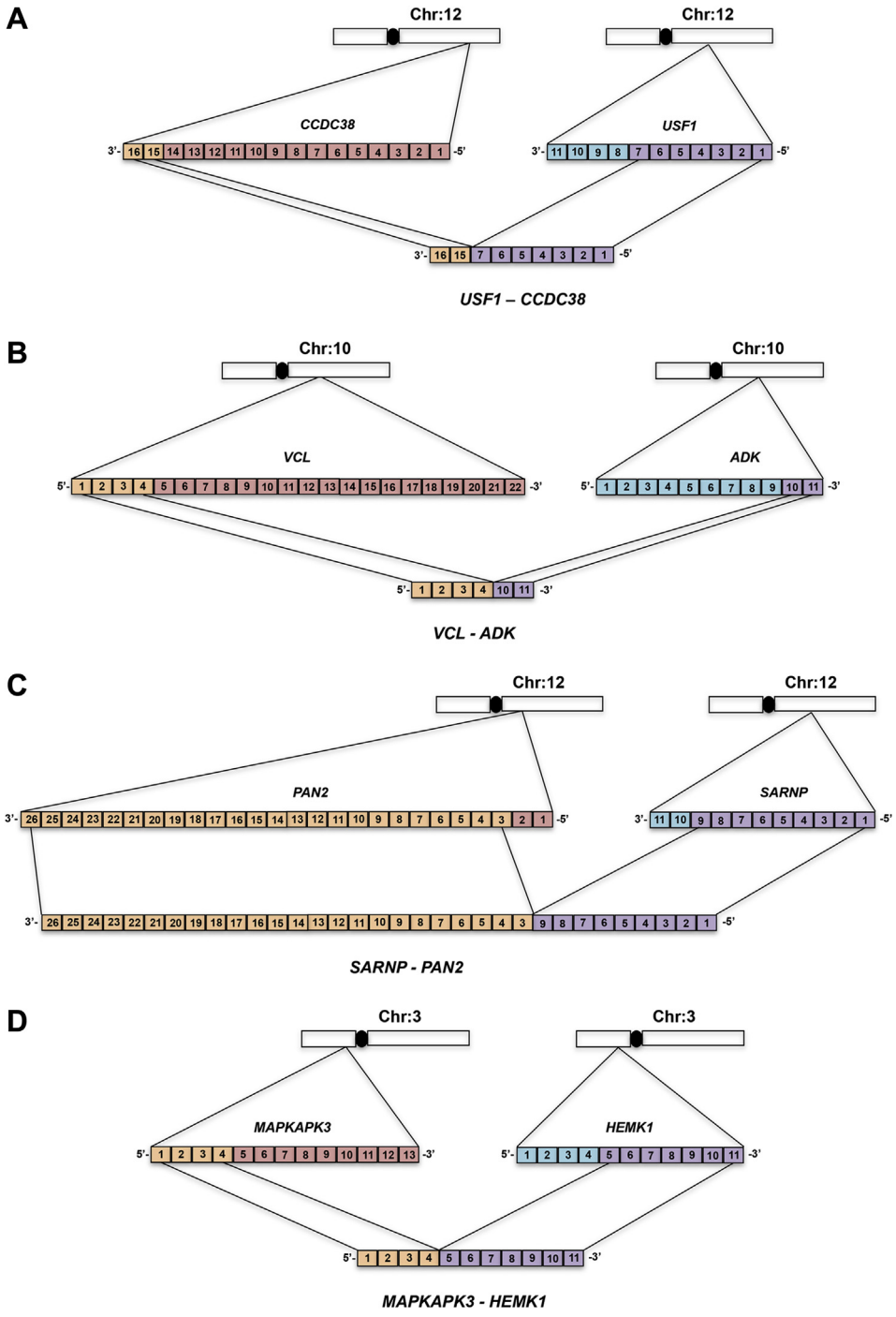

Figure 4 – Schematic representation of 4 validated fusion genes in papillary carcinomas of the breast. (A) *USF1-CCDC38*. (B) *VLC-ADK*. (C) *SARNP-PAN2*. (D) *MAPKAPK3-HEMK1*.

including *RET*, *ASCL1* and *DOK7*, were expressed at significantly higher levels in SPCs than in EPCs (Mann–Whitney *U* test p-value<0.01, Supplementary Table S13). Consistent with these findings, immunohistochemical analysis revealed that three out of four SPCs displayed expression of synaptophysin, whereas only one out of seven EPCs focally displayed synaptophysin expression ($p = 0.0878$, Fisher's exact test; Supplementary Figure S6). Probably due to the small sample size of the groups analyzed, no differentially expressed genes were identified in the other comparisons performed between the different histologic subtypes of PC (data not shown).

To validate the findings obtained by microarray-based gene expression profiling, the expression of genes identified to be significantly differentially expressed between PC histologic subtypes (i.e. *RET*, *IGFBP5*, *TMPRSS6*, *ERP27*, *NFKBIZ and GBP3*) was assessed using qRT-PCR in the 16 PCs included in this study. This analysis confirmed that *ERP27*, *NFKBIZ*, *GBP3*

Table 2 – Significantly mutated genes in The Cancer Genome Atlas (TCGA) breast cancer dataset found to be mutated in the papillary carcinomas of the breast subjected to RNA-sequencing.

Gene	PAPI13 (EPC)	PAPI14 (IPC)	PAPI15 (SPC)	PAPI16 (IPC)	PAPI18 (EPC)	PAPI17 (IPC)	PAPI19 (SPC)	PAPI20 (EPC)	TCGA cases ($n = 507$)#	Luminal A/B TCGA cases ($n = 321$)#
PIK3CA	H1047R								35.1%	42.7%
TP53						R273H			36.9%	18.7%
MAP3K1		K814R							7.7%	11.2%
AKT1					L52H				2.4%	3.4%
CTCF			Q33*, L482P						2.6%	3.1%
LAMA5		G3685R						A907T	1.4%	0.9%
MACF1	G5679D				R3956W				2.2%	1.6%

Of the 35 significantly mutated genes reported in the TCGA breast cancer dataset (Cancer Genome Atlas, 2012), 5 were also targeted by mutations in 5/8 PC cases analyzed here. In addition, *LAMA5* and *MACF1* were identified to be recurrently mutated in the PCs studied (bottom two rows). The last two columns show the frequency of the gene mutations in the TCGA cohort. # data from www.cBioPortal.org, accessed March 8th 2014.

and *IGFBP5* were expressed at significantly higher levels in EPCs than in SPCs and IPCs (Supplementary Figure S7; Mann–Whitney *U* test *p*-values, 0.0002, 0.0008, 0.0035 and 0.0003, respectively), and that *RET and TMPRSS6* were expressed at significantly higher levels in SPCs than in EPCs and IPCs (Supplementary Figure S7; Mann–Whitney *U* test *p*-values: 0.0059 and 0.02, respectively).

Genomic analysis of EPCs, SPCs and IPCs revealed that these histologic subtypes of PCs have similar patterns of copy number losses, gains and amplifications (Figures 3 and 6), consistent with our previous observations (Duprez et al., 2012).

Taken together, although the distinct histologic subtypes of PCs display rather similar patterns of gene copy number alterations, EPCs differ from other PC subtypes in regards to the expression of genes related to cell migration, whereas SPCs display transcriptomic features consistent with neuroendocrine differentiation.

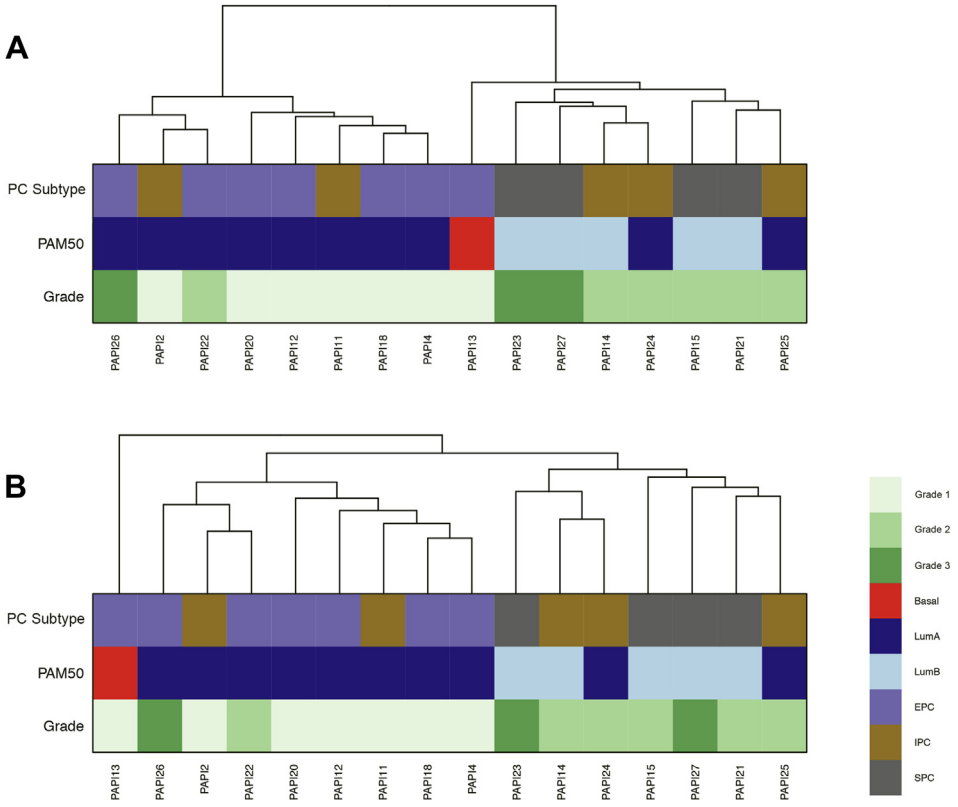

Figure 5 – Unsupervised hierarchical clustering of gene expression data of histologic subtypes of papillary carcinomas. A) Clustering was performed on 2310 probes using Ward's algorithm and Pearson correlation as distance metric. B) Clustering was performed on 2310 probes using complete linkage algorithm and Pearson correlation as distance metric. EPC, encapsulated papillary carcinoma; IPC, invasive papillary carcinoma; SPC, solid papillary carcinoma.

4. Discussion

Here we demonstrate that PCs of the breast are ER-positive, HER2-negative breast cancers, that largely display a luminal phenotype and gene copy number aberrations consistent with those of low-grade ER-positive breast cancers (i.e., 16q losses, 16p gains, and 1q gains) (Lopez-Garcia et al., 2010a). Despite the genomic similarities between PCs and IDC-NSTs, PCs were shown to have gene expression profiles distinct from those of IDC-NSTs of the same histologic grade and ER-status and to display lower levels of expression of proliferation-related, cell assembly and organization and cellular movement and migration genes. We have also demonstrated that PCs, at variance with some other special histologic types of breast cancer, are not underpinned by a highly recurrent expressed fusion gene or highly recurrent expressed mutation (i.e. present in >50% of samples). In addition, a hypothesis generating, exploratory analysis of the transcriptomic profiles of subtypes of PCs revealed that EPCs express a subset of genes involved in cell migration at significantly lower levels than SPCs and IPCs, and that SPCs display transcriptomic and immunohistochemical features consistent with those associated with neuroendocrine differentiation in human cancers.

Previous genomic analyses of other special histologic types of breast cancer have revealed that micropapillary (Marchio et al., 2009, 2008; Natrajan et al., 2014), mucinous (Horlings et al., 2013; Lacroix-Triki et al., 2010), and adenoid cystic (Wetterskog et al., 2012) carcinomas display different patterns of genomic alterations when compared to grade- and ER-matched IDC-NSTs. Here we have confirmed, employing a platform with a higher resolution, our previous observations (Duprez et al., 2012) that although PCs harbor fewer gene copy number alterations than IDC-NSTs of the same histologic grade and ER-status, the patterns of gene copy number aberrations found in PCs are similar to those observed in IDC-NSTs. Despite the genomic similarities between PCs and IDC-NSTs, PCs expressed proliferation-related genes, including *CCND2*, *KIT* and *MAP4K4*, at significantly lower levels than grade- and ER-matched IDC-NSTs. Given that the number and complexity of gene copy number aberrations (Ciriello et al., 2013; Reyal et al., 2012; Vincent-Salomon et al., 2013) and the levels of expression of proliferation-related genes are prognostic in ER-positive luminal cancers (Reis-Filho and Pusztai, 2011), our findings may provide a biological rationale for the more indolent clinical behavior of PCs.

Functional annotation of the genes expressed at significantly lower levels in PCs also revealed an enrichment for pathways and networks related to cell adhesion and cellular movement and migration, including several laminins, integrins, collagens and matrix metalloproteinases. The genes found to be differentially expressed in this study may also partially explain the differences in clinical behavior (i.e. indolent behavior, favorable prognosis and limited invasiveness) between PCs and IDC-NSTs of similar histologic grade. In particular, PCs displayed lower levels of expression of genes related to cell adhesion, migration and movement, including *MMP2* and *MMP7* and genes that are found in the basement membrane (e.g. laminins, integrins, collagens). These findings are in agreement with those reported by Rakha et al. (Rakha et al., 2012) in that PCs display *MMP2* and *MMP7* expression levels lower than those found in IDC-NSTs (Rakha et al., 2012), and that PCs, in particular EPCs, exhibit an expression pattern of invasion-associated markers intermediate between DCIS and IDC-NSTs (Rakha et al., 2012). The transcriptomic analysis performed in this study has

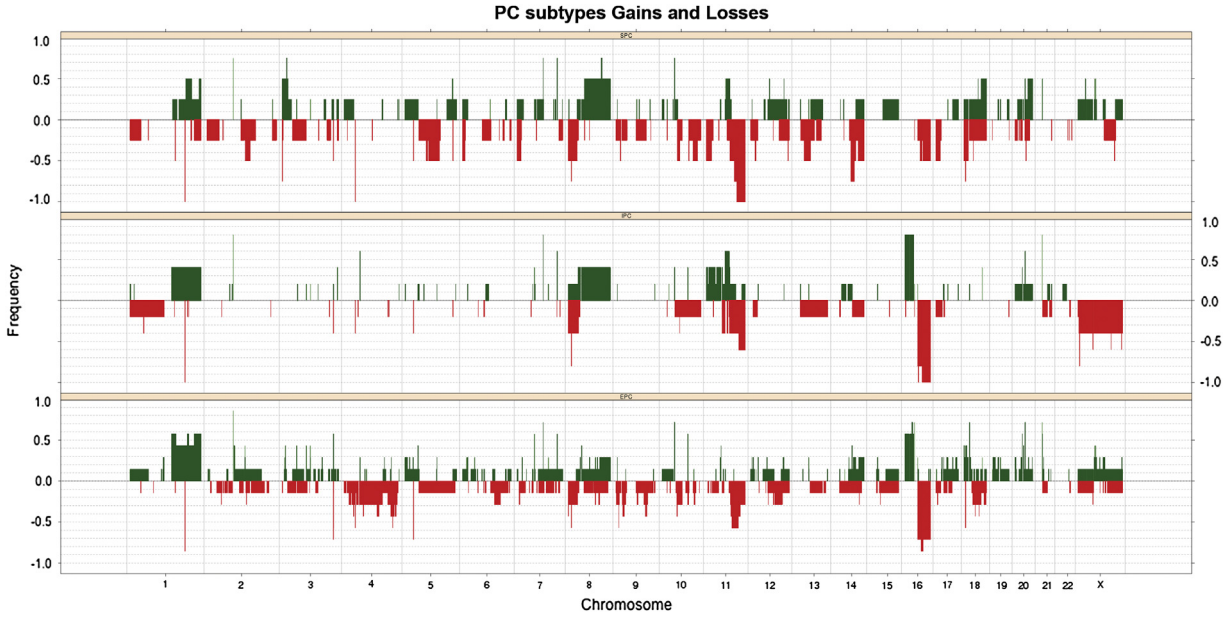

Figure 6 – Comparative genomic profiling of encapsulated, solid and invasive papillary carcinomas. Frequency plots and multi-Fisher's exact comparisons of chromosomal gains and losses in SPCs (top), IPCs (middle) and EPCs (bottom). The proportion of tumors in which each probe is gained (green bars) or lost (red bars) is plotted (*y* axis) for each probe, according to its genomic position (*x* axis). EPC, encapsulated papillary carcinoma; IPC, invasive papillary carcinoma; SPC, solid papillary carcinoma.

also revealed that PCs display higher levels of expression of genes related to angiogenesis (e.g. *VEGFA*, *ESM1* and *SLC4A11*) than grade- and ER-matched IDC-NSTs. This observation is consistent with the notion that PCs are well-vascularized lesions (Wagner et al., 2004) and that histologically these tumors are characterized by the presence of neoplastic cells arranged in papillary fronds containing fibrovascular cores.

RNA-sequencing analysis of 8 PCs (3 EPCs, 2 SPCs, and 3 IPCs) revealed that PCs are neither underpinned by a highly recurrent fusion gene or a highly recurrent mutation. Of the 35 significantly mutated genes present in the TCGA analysis of breast cancers (Cancer Genome Atlas, 2012), five (*PIK3CA, TP53, MAP3K1, AKT1, CTCF*) were also found to be mutated in the PCs subjected to RNA-sequencing. Although potentially pathogenic mutations in *LAMA5* and *MACF1* were found to be recurrent in PCs, they were present in only two out of 8 PCs tested, and could also be found in a subset of the common-type breast cancers analyzed by the TCGA, suggesting that PCs are unlikely to be driven by a highly recurrent mutation in a way akin to lobular carcinomas of the breast, which harbor *CDH1* gene mutations in >60% of cases (Cancer Genome Atlas, 2012).

This study has several limitations. First, the small number of cases subjected to gene expression profiling could have resulted in a type II or β error in the class comparison analysis of the subtypes of PCs; therefore, this comparative analysis should be perceived as exploratory and hypothesis-generating. Although only 8 PCs were subjected to RNA-sequencing for the discovery of a highly recurrent expressed fusion gene or mutation, a type II or β error is unlikely, given that with 8 samples we should have been able to identify a recurrent fusion gene at a prevalence similar to that of *ETV6-NTRK3* in secretory carcinomas (Tognon et al., 2002) or *MYB-NFIB* in adenoid cystic carcinomas (Wetterskog et al., 2012), or a recurrent mutation found at a prevalence similar to *FOXL2* mutations in granulosa cell tumors of the ovary (Shah et al., 2009) with >95% statistical power. It should be noted that if different subtypes of papillary carcinomas would be driven by distinct fusion genes or mutations, we would not be sufficiently powered to detect these events with the eight cases (i.e. three IPCs, three EPCs and two SPCs) subjected to massively parallel sequencing analysis in this study. Given that RNA-sequencing was only performed in the neoplastic tissues, only expressed mutations could be identified and the *de novo* discovery of somatic mutations was not trivial; however, our analysis revealed mutations in PCs that have also been found in ER-positive luminal IDC-NSTs (Cancer Genome Atlas, 2012). Finally, given the retrospective nature of this cohort of PCs, a detailed correlation between the genomic or transcriptomic features of PCs identified in this study with the outcome of PC patients could not be performed. In fact, given that PCs have an indolent clinical course and rarely metastasize, future analyses investigating the impact of the transcriptomic differences between PCs and IDC-NSTs of the same histologic grade and ER-status identified here on the outcome of breast cancer patients are warranted.

Despite these limitations, our study confirms that PCs of the breast preferentially display a luminal phenotype, have relatively simple genomic profiles, and are unlikely to be underpinned by a highly recurrent fusion gene or a pathognomonic expressed mutation. In addition, our findings offer insights into the molecular basis of the relatively indolent behavior and distinct morphological features of the three subtypes of PCs. When compared to that of IDC-NSTs of the same grade and histologic type, PCs express proliferation-related genes at lower levels, features consistent with those of ER-positive cancers with an indolent clinical behavior, and display lower levels of expression of genes related to cell adhesion, migration and movement, suggestive of a less invasive phenotype.

Author contributions

BW, AV-S and JSR-F conceived the study. SP, CKYN and CAE interpreted results and drafted the manuscript; SP, CKYN, LGM, CFC, RN, F-CB and PMW carried out experiments and analyzed data; OM and AV-S provided samples and interpreted data. All authors reviewed and approved the final version of the manuscript.

Acknowledgements

SP is funded by The Swiss National Science Foundation (SNSF; Fellowships for prospective researchers; PBBSP3_146957). RN is funded by a Breast Cancer Campaign Career Development Fellowship (2011MaySF01). AV-S is supported by an 'INTERFACE' grant from INSERM.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.molonc.2014.06.011.

REFERENCES

Adelaide, J., Finetti, P., Bekhouche, I., Repellini, L., Geneix, J., Sircoulomb, F., Charafe-Jauffret, E., Cervera, N., Desplans, J., Parzy, D., Schoenmakers, E., Viens, P., Jacquemier, J., Birnbaum, D., Bertucci, F., Chaffanet, M., 2007. Integrated profiling of basal and luminal breast cancers. Cancer Res. 67, 11565–11575.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249.

Arriola, E., Marchio, C., Tan, D.S., Drury, S.C., Lambros, M.B., Natrajan, R., Rodriguez-Pinilla, S.M., Mackay, A., Tamber, N., Fenwick, K., Jones, C., Dowsett, M., Ashworth, A., Reis-Filho, J.S., 2008. Genomic analysis of the HER2/TOP2A amplicon in breast cancer and breast cancer cell lines. Lab Invest. 88, 491–503.

Bengtsson, H., Wirapati, P., Speed, T.P., 2009. A single-array preprocessing method for estimating full-resolution raw copy

numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6. Bioinformatics 25, 2149–2156.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B 57, 289–300.
Bertucci, F., Orsetti, B., Negre, V., Finetti, P., Rouge, C., Ahomadegbe, J.C., Bibeau, F., Mathieu, M.C., Treilleux, I., Jacquemier, J., Ursule, L., Martinec, A., Wang, Q., Benard, J., Puisieux, A., Birnbaum, D., Theillet, C., 2008. Lobular and ductal carcinomas of the breast have distinct genomic and expression profiles. Oncogene 27, 5359–5372.
Carter, H., Chen, S., Isik, L., Tyekucheva, S., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Karchin, R., 2009. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. Cancer Res. 69, 6660–6667.
Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., Beroukhim, R., Pellman, D., Levine, D.A., Lander, E.S., Meyerson, M., Getz, G., 2012. Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. 30, 413–421.
Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., Chen, F., Feiler, H., Tokuyasu, T., Kingsley, C., Dairkee, S., Meng, Z., Chew, K., Pinkel, D., Jain, A., Ljung, B.M., Esserman, L., Albertson, D.G., Waldman, F.M., Gray, J.W., 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10, 529–541.
Chin, S.F., Teschendorff, A.E., Marioni, J.C., Wang, Y., Barbosa-Morais, N.L., Thorne, N.P., Costa, J.L., Pinder, S.E., van de Wiel, M.A., Green, A.R., Ellis, I.O., Porter, P.L., Tavare, S., Brenton, J.D., Ylstra, B., Caldas, C., 2007. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol. 8, R215.
Ciriello, G., Sinha, R., Hoadley, K.A., Jacobsen, A.S., Reva, B., Perou, C.M., Sander, C., Schultz, N., 2013. The molecular diversity of Luminal A breast tumors. Breast Cancer Res. Treat. 141, 409–420.
Collins, L.C., Schnitt, S.J., 2008. Papillary lesions of the breast: selected diagnostic and management issues. Histopathology 52, 20–29.
Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Graf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Group, M., Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Borresen-Dale, A.L., Brenton, J.D., Tavare, S., Caldas, C., Aparicio, S., 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346–352.
Dobbin, K.K., Zhao, Y., Simon, R.M., 2008. How large a training set is needed to develop a classifier for microarray data? Clin. Cancer Res. 14, 108–114.
Du, P., Kibbe, W.A., Lin, S.M., 2008. lumi: a pipeline for processing Illumina microarray. Bioinformatics 24, 1547–1548.
Duprez, R., Wilkerson, P.M., Lacroix-Triki, M., Lambros, M.B., Mackay, A., Hern, R.A., Gauthier, A., Pawar, V., Colombo, P.E., Daley, F., Natrajan, R., Ward, E., Macgrogan, G., Arbion, F., Michenet, P., Weigelt, B., Vincent-Salomon, A., Reis-Filho, J.S., 2012. Immunophenotypic and genomic characterization of papillary carcinomas of the breast. J. Pathol. 226, 427–441.
Elston, C.W., Ellis, I.O., 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19, 403–410.
Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., Schultz, N., 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal 6, l1.
Geyer, F.C., Weigelt, B., Natrajan, R., Lambros, M.B., de Biase, D., Vatcheva, R., Savage, K., Mackay, A., Ashworth, A., Reis-Filho, J.S., 2010. Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas. J. Pathol. 220, 562–573.
Grabowski, J., Salzstein, S.L., Sadler, G.R., Blair, S., 2008. Intracystic papillary carcinoma: a review of 917 cases. Cancer 113, 916–920.
Gruel, N., Lucchesi, C., Raynal, V., Rodrigues, M.J., Pierron, G., Goudefroye, R., Cottu, P., Reyal, F., Sastre-Garau, X., Fourquet, A., Delattre, O., Vincent-Salomon, A., 2010. Lobular invasive carcinoma of the breast is a molecular entity distinct from luminal invasive ductal carcinoma. Eur. J. Cancer 46, 2399–2407.
Hernandez, L., Wilkerson, P.M., Lambros, M.B., Campion-Flora, A., Rodrigues, D.N., Gauthier, A., Cabral, C., Pawar, V., Mackay, A., A'Hern, R., Marchio, C., Palacios, J., Natrajan, R., Weigelt, B., Reis-Filho, J.S., 2012. Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive breast cancers reveals intra-tumour genetic heterogeneity and clonal selection. J. Pathol. 227, 42–52.
Hill, C.B., Yeh, I.T., 2005. Myoepithelial cell staining patterns of papillary breast lesions: from intraductal papillomas to invasive papillary carcinomas. Am. J. Clin. Pathol. 123, 36–44.
Horlings, H.M., Weigelt, B., Anderson, E.M., Lambros, M.B., Mackay, A., Natrajan, R., Ng, C.K., Geyer, F.C., van de Vijver, M.J., Reis-Filho, J.S., 2013. Genomic profiling of histological special types of breast cancer. Breast Cancer Res. Treat. 142, 257–269.
Hu, X., Stern, H.M., Ge, L., O'Brien, C., Haydu, L., Honchell, C.D., Haverty, P.M., Peters, B.A., Wu, T.D., Amler, L.C., Chant, J., Stokoe, D., Lackner, M.R., Cavet, G., 2009. Genetic alterations and oncogenic pathways associated with breast cancer subtypes. Mol. Cancer Res. 7, 511–522.
Ishizawa, T., Nozaki, Y., Ueda, T., Takeuchi, N., 2008. The human mitochondrial translation release factor HMRF1L is methylated in the GGQ motif by the methyltransferase HMPrmC. Biochem. Biophys. Res. Commun. 373, 99–103.
Iyer, M.K., Chinnaiyan, A.M., Maher, C.A., 2011. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics 27, 2903–2904.
Lacroix-Triki, M., Suarez, P.H., MacKay, A., Lambros, M.B., Natrajan, R., Savage, K., Geyer, F.C., Weigelt, B., Ashworth, A., Reis-Filho, J.S., 2010. Mucinous carcinoma of the breast is genomically distinct from invasive ductal carcinomas of no special type. J. Pathol. 222, 282–298.
Lakhani, S.R., O, E.I., Schnitt, S.J., Tan, P.H., van de Vijver, M.J., 2012. WHO Classification of Breast Tumors. IARC, Lyon.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.Genome Project Data Processing, S, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.
Lopez-Garcia, M.A., Geyer, F.C., Lacroix-Triki, M., Marchio, C., Reis-Filho, J.S., 2010a. Breast cancer precursors revisited: molecular features and progression pathways. Histopathology 57, 171–192.
Lopez-Garcia, M.A., Geyer, F.C., Natrajan, R., Kreike, B., Mackay, A., Grigoriadis, A., Reis-Filho, J.S., Weigelt, B., 2010b. Transcriptomic analysis of tubular carcinomas of the breast reveals similarities and differences with molecular subtype-matched ductal and lobular carcinomas. J. Pathol. 222, 64–75.
Mackay, A., Tamber, N., Fenwick, K., Iravani, M., Grigoriadis, A., Dexter, T., Lord, C.J., Reis-Filho, J.S., Ashworth, A., 2009. A

high-resolution integrated analysis of genetic and expression profiles of breast cancer cell lines. Breast Cancer Res. Treat. 118, 481–498.

Marchio, C., Iravani, M., Natrajan, R., Lambros, M.B., Geyer, F.C., Savage, K., Parry, S., Tamber, N., Fenwick, K., Mackay, A., Schmitt, F.C., Bussolati, G., Ellis, I., Ashworth, A., Sapino, A., Reis-Filho, J.S., 2009. Mixed micropapillary-ductal carcinomas of the breast: a genomic and immunohistochemical analysis of morphologically distinct components. J. Pathol. 218, 301–315.

Marchio, C., Iravani, M., Natrajan, R., Lambros, M.B., Savage, K., Tamber, N., Fenwick, K., Mackay, A., Senetta, R., Di Palma, S., Schmitt, F.C., Bussolati, G., Ellis, L.O., Ashworth, A., Sapino, A., Reis-Filho, J.S., 2008. Genomic and immunophenotypical characterization of pure micropapillary carcinomas of the breast. J. Pathol. 215, 398–410.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

McPherson, A., Hormozdiari, F., Zayed, A., Giuliany, R., Ha, G., Sun, M.G., Griffith, M., Heravi Moussavi, A., Senz, J., Melnyk, N., Pacheco, M., Marra, M.A., Hirst, M., Nielsen, T.O., Sahinalp, S.C., Huntsman, D., Shah, S.P., 2011. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput. Biol. 7, e1001138.

Muller, P.A., Vousden, K.H., 2013. p53 mutations in cancer. Nat. Cell Biol. 15, 2–8.

Natrajan, R., Lambros, M.B., Rodriguez-Pinilla, S.M., Moreno-Bueno, G., Tan, D.S., Marchio, C., Vatcheva, R., Rayter, S., Mahler-Araujo, B., Fulford, L.G., Hungermann, D., Mackay, A., Grigoriadis, A., Fenwick, K., Tamber, N., Hardisson, D., Tutt, A., Palacios, J., Lord, C.J., Buerger, H., Ashworth, A., Reis-Filho, J.S., 2009. Tiling path genomic profiling of grade 3 invasive ductal breast cancers. Clin. Cancer Res. 15, 2711–2722.

Natrajan, R., Mackay, A., Wilkerson, P.M., Lambros, M.B., Wetterskog, D., Arnedos, M., Shiu, K.K., Geyer, F.C., Langerod, A., Kreike, B., Reyal, F., Horlings, H.M., van de Vijver, M.J., Palacios, J., Weigelt, B., Reis-Filho, J.S., 2012. Functional characterization of the 19q12 amplicon in grade III breast cancers. Breast Cancer Res. 14, R53.

Natrajan, R., Wilkerson, P.M., Marchio, C., Piscuoglio, S., Ng, C.K., Wai, P., Lambros, M.B., Samartzis, E.P., Dedes, K.J., Frankum, J., Bajrami, I., Kopec, A., Mackay, A., A'Hern, R., Fenwick, K., Kozarewa, I., Hakas, J., Mitsopoulos, C., Hardisson, D., Lord, C.J., Kumar-Sinha, C., Ashworth, A., Weigelt, B., Sapino, A., Chinnaiyan, A.M., Maher, C.A., Reis-Filho, J.S., 2014. Characterization of the genomic features and expressed fusion genes in micropapillary carcinomas of the breast. J. Pathol. 232, 553–565.

Olshen, A.B., Venkatraman, E.S., Lucito, R., Wigler, M., 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5, 557–572.

Pal, S.K., Lau, S.K., Kruper, L., Nwoye, U., Garberoglio, C., Gupta, R.K., Paz, B., Vora, L., Guzman, E., Artinyan, A., Somlo, G., 2010. Papillary carcinoma of the breast: an overview. Breast Cancer Res. Treat. 122, 637–645.

Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J.F., Stijleman, I.J., Palazzo, J., Marron, J.S., Nobel, A.B., Mardis, E., Nielsen, T.O., Ellis, M.J., Perou, C.M., Bernard, P.S., 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. 27, 1160–1167.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lonning, P.E., Borresen-Dale, A.L., Brown, P.O., Botstein, D., 2000. Molecular portraits of human breast tumours. Nature 406, 747–752.

Persson, M., Andren, Y., Mark, J., Horlings, H.M., Persson, F., Stenman, G., 2009. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. Proc. Natl. Acad. Sci. U S A 106, 18740–18744.

Rakha, E.A., Gandhi, N., Climent, F., van Deurzen, C.H., Haider, S.A., Dunk, L., Lee, A.H., Macmillan, D., Ellis, I.O., 2011. Encapsulated papillary carcinoma of the breast: an invasive tumor with excellent prognosis. Am. J. Surg. Pathol. 35, 1093–1103.

Rakha, E.A., Tun, M., Junainah, E., Ellis, I.O., Green, A., 2012. Encapsulated papillary carcinoma of the breast: a study of invasion associated markers. J. Clin. Pathol. 65, 710–714.

Reis-Filho, J.S., Pusztai, L., 2011. Gene expression profiling in breast cancer: classification, prognostication, and prediction. Lancet 378, 1812–1823.

Reva, B., Antipin, Y., Sander, C., 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 39, e118.

Reyal, F., Bollet, M.A., Caly, M., Gentien, D., Carpentier, S., Peyro-Saint-Paul, H., Pierga, J.Y., Cottu, P., Dieras, V., Sigal-Zafrani, B., Vincent-Salomon, A., Sastre-Garau, X., 2012. Respective prognostic value of genomic grade and histological proliferation markers in early stage (pN0) breast carcinoma. PLoS One 7, e35184.

Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., Yan, H., Gazdar, A., Powell, S.M., Riggins, G.J., Willson, J.K., Markowitz, S., Kinzler, K.W., Vogelstein, B., Velculescu, V.E., 2004. High frequency of mutations of the PIK3CA gene in human cancers. Science 304, 554.

Schwarz, J.M., Rodelsperger, C., Schuelke, M., Seelow, D., 2010. MutationTaster evaluates disease-causing potential of sequence alterations. Nat. Methods 7, 575–576.

Shah, S.P., Kobel, M., Senz, J., Morin, R.D., Clarke, B.A., Wiegand, K.C., Leung, G., Zayed, A., Mehl, E., Kalloger, S.E., Sun, M., Giuliany, R., Yorida, E., Jones, S., Varhol, R., Swenerton, K.D., Miller, D., Clement, P.B., Crane, C., Madore, J., Provencher, D., Leung, P., DeFazio, A., Khattra, J., Turashvili, G., Zhao, Y., Zeng, T., Glover, J.N., Vanderhyden, B., Zhao, C., Parkinson, C.A., Jimenez-Linan, M., Bowtell, D.D., Mes-Masson, A.M., Brenton, J.D., Aparicio, S.A., Boyd, N., Hirst, M., Gilks, C.B., Marra, M., Huntsman, D.G., 2009. Mutation of FOXL2 in granulosa-cell tumors of the ovary. N. Engl. J. Med. 360, 2719–2729.

Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L., Edwards, K.J., Day, I.N., Gaunt, T.R., 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum. Mutat. 34, 57–65.

Solorzano, C.C., Middleton, L.P., Hunt, K.K., Mirza, N., Meric, F., Kuerer, H.M., Ross, M.I., Ames, F.C., Feig, B.W., Pollock, R.E., Singletary, S.E., Babiera, G., 2002. Treatment and outcome of patients with intracystic papillary carcinoma of the breast. Am. J. Surg. 184, 364–368.

Suzuki, R., Shimodaira, H., 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22, 1540–1542.

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

Tognon, C., Knezevich, S.R., Huntsman, D., Roskelley, C.D., Melnyk, N., Mathers, J.A., Becker, L., Carneiro, F., MacPherson, N., Horsman, D., Poremba, C., Sorensen, P.H., 2002. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. Cancer Cell 2, 367–376.

Vincent-Salomon, A., Benhamo, V., Gravier, E., Rigaill, G., Gruel, N., Robin, S., de Rycke, Y., Mariani, O., Pierron, G., Gentien, D., Reyal, F., Cottu, P., Fourquet, A., Rouzier, R., Sastre-Garau, X., Delattre, O., 2013. Genomic instability: a stronger prognostic marker than proliferation for early stage luminal breast carcinomas. PLoS One 8, e76496.

Vincent-Salomon, A., Gruel, N., Lucchesi, C., MacGrogan, G., Dendale, R., Sigal-Zafrani, B., Longy, M., Raynal, V., Pierron, G., de Mascarel, I., Taris, C., Stoppa-Lyonnet, D., Pierga, J.Y., Salmon, R., Sastre-Garau, X., Fourquet, A., Delattre, O., de Cremoux, P., Aurias, A., 2007. Identification of typical medullary breast carcinoma as a genomic sub-group of basal-like carcinomas, a heterogeneous new molecular entity. Breast Cancer Res. 9, R24.

Wagner, A.E., Middleton, L.P., Whitman, G.J., 2004. Intracystic papillary carcinoma of the breast with invasion. AJR Am. J. Roentgenol. 183, 1516.

Weigelt, B., Baehner, F.L., Reis-Filho, J.S., 2010a. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. J. Pathol. 220, 263–280.

Weigelt, B., Geyer, F.C., Horlings, H.M., Kreike, B., Halfwerk, H., Reis-Filho, J.S., 2009a. Mucinous and neuroendocrine breast carcinomas are transcriptionally distinct from invasive ductal carcinomas of no special type. Mod. Pathol. 22, 1401–1414.

Weigelt, B., Geyer, F.C., Natrajan, R., Lopez-Garcia, M.A., Ahmad, A.S., Savage, K., Kreike, B., Reis-Filho, J.S., 2010b. The molecular underpinning of lobular histological growth pattern: a genome-wide transcriptomic analysis of invasive lobular carcinomas and grade- and molecular subtype-matched invasive ductal carcinomas of no special type. J. Pathol. 220, 45–57.

Weigelt, B., Geyer, F.C., Reis-Filho, J.S., 2010c. Histological types of breast cancer: how special are they? Mol. Oncol. 4, 192–208.

Weigelt, B., Horlings, H.M., Kreike, B., Hayes, M.M., Hauptmann, M., Wessels, L.F., de Jong, D., Van de Vijver, M.J., Van't Veer, L.J., Peterse, J.L., 2008. Refinement of breast cancer classification by molecular characterization of histological special types. J. Pathol. 216, 141–150.

Weigelt, B., Kreike, B., Reis-Filho, J.S., 2009b. Metaplastic breast carcinomas are basal-like breast cancers: a genomic profiling analysis. Breast Cancer Res. Treat. 117, 273–280.

Weigelt, B., Mackay, A., A'Hern, R., Natrajan, R., Tan, D.S., Dowsett, M., Ashworth, A., Reis-Filho, J.S., 2010d. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. Lancet Oncol. 11, 339–349.

Weigelt, B., Reis-Filho, J.S., 2009. Histological and molecular types of breast cancer: is there a unifying taxonomy? Nat. Rev. Clin. Oncol. 6, 718–730.

Wetterskog, D., Lopez-Garcia, M.A., Lambros, M.B., A'Hern, R., Geyer, F.C., Milanezi, F., Cabral, M.C., Natrajan, R., Gauthier, A., Shiu, K.K., Orr, N., Shousha, S., Gatalica, Z., Mackay, A., Palacios, J., Reis-Filho, J.S., Weigelt, B., 2012. Adenoid cystic carcinomas constitute a genomically distinct subgroup of triple-negative and basal-like breast cancers. J. Pathol. 226, 84–96.

Wynveen, C.A., Nehhozina, T., Akram, M., Hassan, M., Norton, L., Van Zee, K.J., Brogi, E., 2011. Intracystic papillary carcinoma of the breast: an in situ or invasive tumor? Results of immunohistochemical analysis and clinical follow-up. Am. J. Surg. Pathol. 35, 1–14.