

Epigenomic Alterations in Breast Carcinoma from Primary Tumor to Locoregional Recurrences

Matahi Moarii, Alice Pinheiro, Brigitte Sigal-Zafrani, Alain Fourquet, Martial Caly, Nicolas Servant, Véronique Stoven, Jean-Philippe Vert, Fabien Reyat

► **To cite this version:**

Matahi Moarii, Alice Pinheiro, Brigitte Sigal-Zafrani, Alain Fourquet, Martial Caly, et al.. Epigenomic Alterations in Breast Carcinoma from Primary Tumor to Locoregional Recurrences. PLoS ONE, Public Library of Science, 2014, 9 (8), pp.e103986. 10.1371/journal.pone.0103986 . hal-01064211

HAL Id: hal-01064211

<https://hal-mines-paristech.archives-ouvertes.fr/hal-01064211>

Submitted on 15 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Epigenomic Alterations in Breast Carcinoma from Primary Tumor to Locoregional Recurrences

Matahi Moarii^{1,2,3}, Alice Pinheiro^{4,5}, Brigitte Sigal-Zafrani⁶, Alain Fourquet⁷, Martial Caly⁶, Nicolas Servant^{1,2,3}, Véronique Stoven^{1,2,3}, Jean-Philippe Vert^{1,2,3}, Fabien Reyat^{4,5,8*}

1 Centre for Computational Biology, Mines ParisTech, Fontainebleau, France, **2** Institut Curie, Paris, France, **3** U900, INSERM, Paris, France, **4** UMR144, Oncology Molecular Team, Institut Curie, Paris, France, **5** Department of translational research, Residual Tumor and Response to Treatment Team, Institut Curie, Paris, France, **6** Department of Tumor Biology, Institut Curie, Paris, France, **7** Department of Radiotherapy, Institut Curie, Paris, France, **8** Department of Surgery, Institut Curie, Paris, France

Abstract

Introduction: Epigenetic modifications such as aberrant DNA methylation has long been associated with tumorigenesis. Little is known, however, about how these modifications appear in cancer progression. Comparing the methylome of breast carcinomas and locoregional evolutions could shed light on this process.

Methods: The methylome profiles of 48 primary breast carcinomas (PT) and their matched axillary metastases (PT/AM pairs, 20 cases), local recurrences (PT/LR pairs, 17 cases) or contralateral breast carcinomas (PT/CL pairs, 11 cases) were analyzed. Univariate and multivariate analyzes were performed to determine differentially methylated probes (DMPs), and a similarity score was defined to compare methylation profiles. Correlation with copy-number based score was calculated and metastatic-free survival was compared between methods.

Results: 49 DMPs were found for the PT/AM set, but none for the others (FDR < 5%). Hierarchical clustering clustered 75% of the PT/AM, 47% of the PT/LR, and none of the PT/CL pairs together. A methylation-based score (MS) was defined as a clonality measure. The PT/AM set contained a high proportion of clonal pairs while PT/LR pairs were evenly split between high and low MS score, suggesting two groups: true recurrences (TR) and new primary tumors (NP). CL were classified as new tumors. MS score was significantly correlated with copy-number based scores. There was no significant difference between the metastatic-free survival of groups of patients based on different classifications.

Conclusion: Epigenomic alterations are well suited to study clonality and track cancer progression. Methylation-based classification of TR and NP performed as well as clinical and copy-number based methods suggesting that these phenomena are tightly linked.

Citation: Moarii M, Pinheiro A, Sigal-Zafrani B, Fourquet A, Caly M, et al. (2014) Epigenomic Alterations in Breast Carcinoma from Primary Tumor to Locoregional Recurrences. PLoS ONE 9(8): e103986. doi:10.1371/journal.pone.0103986

Editor: Jorg Tost, CEA - Institut de Genomique, France

Received: April 17, 2014; **Accepted:** July 4, 2014; **Published:** August 6, 2014

Copyright: © 2014 Moarii 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.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. Data are available from the GEO database accession number: GSE44870. The url is: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=rfqthuyiegycaxw&acc=GSE44870>.

Funding: MM was supported by the French league against cancer (www.ligue-cancer.net). JPV was supported by the European Research Council (SMAC-ERC-280032, [url: erc.europa.eu](http://erc.europa.eu)). AF was supported by the hospital program for clinical research (AOM 06 149, [url: http://www.sante.gouv.fr/le-programme-hospitalier-de-recherche-clinique-phrc.html](http://www.sante.gouv.fr/le-programme-hospitalier-de-recherche-clinique-phrc.html)). 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.

* Email: fabien.reyat@curie.fr

Introduction

Breast conservative therapy, consisting in a partial mastectomy followed by whole breast irradiation, is the standard treatment for patients with early stage breast cancer. Overall survival is not significantly different from more physically and psychologically aggressive treatments such as mastectomy [1]. However, patients relapse within 10 years in the same breast as the primary tumor (PT) in approximately 6 % of cases [2], and within 5 years in the contralateral breast in approximately 3.5% of cases [3] or more in BRCA1/2 mutation carriers [4]. Moreover, at the time of diagnosis, early stage breast cancers have already spread to axillary lymph nodes in roughly 30 % of cases [5].

These different types of locoregional evolutions have different implications in terms of survival and treatments. Axillary

metastases (AM) is usually predictive of poor survival [6] and is considerably worsen in triple negative breast cancers [7]. Local recurrences (LR) have been tightly linked with a greater risk of distant metastasis [8]. Veronesi et al. [9] distinguished two categories of local recurrences: true recurrences (TR), corresponding to re-growth of resistant cells after initial treatment, and new primary tumors (NP), corresponding to *de novo* cancer. This classification is of potential interest to define adapted treatment scheme, as NP are considered to have an improved survival compared to TR [10]. Contralateral breast cancers (CL) are also an heterogeneous entity depending on the synchronism with the primary tumor. Synchronous bilateral breast cancers are developed at the same time, with the same genetic, environmental and hormonal background as the PT. Metachronous CL are usually

Table 1. Summarized PT/LR Clinical and histological features.

Pair	PT						Local Recurrence						Discordance	
	Cor	Age	Type	Grade	ER	PR	HER2	Type	Grade	ER	PR	HER2	Loc	Loc
1	1	23.3	D	3	-	+	-	D	2	+	+	-	1	ER
2	3	42.9	D	3	+	+	-	D	3	+	+	-	1	
3	11	49.3	L	3	-	-	-	D	3	-	-	+	1	HER2
4	16	48.8	D	2	+	+	-	D	1	+	+	-	1	
5	12	49.3	L	2	+	+	-	L	2	+	-	-	0	PR
6	13	45.4	D	2	+	+	-	D	2	+	+	-	1	
7	15	46.5	D	2	+	+	-	D	2	+	+	-	1	
8	2	42.4	D	2	+	+	-	L	1	+	+	NA	1	
9	4	48.6	L	1	+	+	-	L	2	+	+	-	1	
10	14	44	L	2	+	+	-	L	2	-	+	-	1	ER
11	18	NA	D	3	-	-	NA	D	2	+	+	NA	1	ER/PR
12	20	47.5	D	3	-	-	+	D	3	-	-	+	0	
13	21	46.7	D	2	+	-	NA	D	3	+	-	-	1	
14	23	31	D	2	-	-	-	D	3	-	-	-	1	
15	24	48.1	D	3	-	-	-	D	3	-	-	-	1	
16	25	43.3	D	3	+	+	-	D	3	+	+	-	1	
17	26	30.8	D	3	-	-	-	D	3	-	-	-	1	

Cor (Correspondence): correspondence number with the Bollet/Servant cohort from [16], **Type**: histological type of the tumor (D= ductal, L= lobular), **Grade**: Aggressiveness of the tumor (1 to 3), **ER**: presence of estrogen receptors, **PR**: presence of progesterone receptors, **HER2**: presence of HER2 receptors, **Loc** (Location): 1 if the recurrence was located less than 4cm from the PT. doi:10.1371/journal.pone.0103986.t001

Table 2. Summarized PT/CL Clinical and histological features.

Pair	Age	PT						Contralateral Recurrence						Discordance									
		Type	Grade	ER	PR	HER2	Type	Grade	ER	PR	HER2	Type	Grade	ER	PR	HER2	Type	Grade	ER	PR	HER2		
1	46.6	L	3	+	+	-	NA	NA	+	+	-	NA	NA	+	+	-	NA	NA	+	+	-	NA	NA
2	46.9	D	2	+	+	-	D	2	+	+	-	D	2	+	+	-	D	2	+	+	-	D	2
3	48.4	D	3	+	+	-	D	3	+	+	-	D	3	+	+	-	D	3	+	+	-	D	3
4	42.6	D	2	-	-	-	D-L	2	+	+	-	D-L	2	+	+	-	D-L	2	+	+	-	D-L	2
5	48.5	D	2	+	+	-	D	3	+	+	-	D	3	+	+	-	D	3	+	+	-	D	3
6	44.5	D	2	+	+	-	Med	2	+	+	-	Med	2	+	+	-	Med	2	+	+	-	Med	2
7	46	D	2	+	+	-	D	1	+	+	-	D	1	+	+	-	D	1	+	+	-	D	1
8	48.9	D	3	+	+	-	Meta	3	+	+	-	Meta	3	+	+	-	Meta	3	+	+	-	Meta	3
9	38.9	D	3	-	-	-	D	3	+	+	-	D	3	+	+	-	D	3	+	+	-	D	3
10	31	D	3	-	-	-	D	3	+	+	-	D	3	+	+	-	D	3	+	+	-	D	3

Type: histological type of the tumor (D = ductal, L = lobular, Med = Medullary, Meta = Metaplastic), **Grade:** Aggressiveness of the tumor (1 to 3), **ER:** presence of estrogen receptors, **PR:** presence of progesterone receptors, **HER2:** presence of HER2 receptors. doi:10.1371/journal.pone.0103986.t002

treated as new cancers [11] although a rare portion are considered as metastases. Overall, CL are still associated with a greater risk of metastasis compared to patients without CL [12].

Differences between the PT and either the AM, the LR or the CL have been studied at the genomic, transcriptomic and proteomic levels. Ellsworth et al. [13] showed an overall frequency of allelic imbalance greater in PT than in AM. Weigelt et al. [14] explored the gene expression profile of PT and their matched AM but were not able to identify a subset of genes to discriminate them, while Feng et al. [15] identified a set of 79 genes able to differentiate PT from matched AM. Studies between PT and LR have mainly focused on distinguishing TR and NP. A criterion based on clinical and pathological features was first established but judged insufficiently robust for most clinical applications. Several studies investigated the difference between TR and NP based on pangenomic analyzes of DNA copy number alterations (CNA) [16,17], intratumoral immune responses [18], loss of heterozygosity [19], to p53 analysis [20], or X-chromosome inactivation [21]. Finally, studies of PT and CL highlighted the role of synchronism of the CL. Similarity measures based on DNA copy number profiles [22] or allelic imbalance [23] showed a higher level of similarity between PT and synchronous CL compared to PT and metachronous CL.

Epigenetic modifications in cancer has recently been the topic of many studies. In particular the link between hypermethylation and gene silencing is well known [24–26]. Several studies have then focused to describe cancer as an epigenetic disease. Baylin et al. [27] have shown that aberrant hypermethylation of specific regions, dominantly CpG islands, are linked with the silencing of tumor suppressor genes and that this phenomenon is present in most cancers. Laird [28], Ehrlich [29] and Das [30] suggested that a global hypomethylation phenomenon was also linked with tumorigenesis. Jones [31] made a complete review of the hallmarks of epigenomics associated with cancer. Moreover, DNA methylation is conserved during cell division [26,32] and could serve as a measure for clonality between cells in the classification of LR as either TR or NP.

In this study, epigenetic differences as well as similarities between PTs and either their AMs, LRs or CLs are analyzed. In the first part, univariate and multivariate analyzes are performed between the methylome profiles of primary tumors and their matched recurrences to observe recurrent patterns in cancer progression. Then in the second part, epigenome-wide similarity analyzes on the same samples is performed to observe clonality between tumor cells.

Results

Methylation differences between PT and their matched metastasis or recurrence

A collection of 17 PT/LR pairs, 11 PT/CL pairs, and 20 PT/AM pairs was analyzed. The methylation data are available in the GEO database record number: GSE44870. Tables 1, 2 and 3 detail the summarized clinico-histopathological properties of each sample. Some of the PT/LR samples match in part the cohort studied by Bollet et al. [16], and the corresponding sample numbers from both studies are provided in Table 2. Tables S1, S2 and S3 provide more detailed characteristics.

Within each of the three cohorts, pairs of tumors including a PT and a metastatic or relapse sample can be used to investigate whether particular patterns in methylation profiles can serve as marker for cancer progression.

Within each cohort, investigations were made to detect differences at the methylome level between PT and the

Table 3. Summarized PT/AM Clinical and histological features.

Pair	Age	Type	Grade	ER	PR	HER2
1	45.9	D	3	+	+	-
2	NA	D	3	+	+	-
3	NA	NA	NA	+	+	-
4	48.8	D	1	+	+	-
5	43.6	D	3	-	-	-
6	35.3	D	2	+	+	-
7	45.1	D	3	+	+	-
8	41.9	D	2	+	+	NA
9	43.5	D	1	+	+	-
10	43.7	D	3	+	+	NA
11	44.9	D	2	-	-	-
12	43.6	D	1	+	-	-
13	40.2	D	3	-	-	+
14	32.5	L	3	+	+	+
15	38.5	D	2	-	+	-
16	37.5	D	3	+	+	-
17	39.3	D	3	+	+	-
18	37.6	D	3	-	-	-
19	36.6	D	3	+	+	+
20	35.4	D	3	-	+	-

Age: Age of the patient at diagnosis of the primary tumor in years, **Type:** histological type of the tumor (D = ductal, L = lobular, Meta = Metaplasia), **Grade:** Aggressiveness of the tumor (1 to 3), **ER:** presence of estrogen receptors, **PR:** presence of progesterone receptors, **HER2:** presence of HER2 receptors.
doi:10.1371/journal.pone.0103986.t003

corresponding matched metastasis (AM) or relapse samples (LR or CL). Using a paired Wilcoxon test, 49 probes significantly differentially methylated were found between PT and AM samples (at a 5% FDR level). The top 50 probes ranked by p-value and the corresponding genes are listed in Table 4. This suggests that a general signal characteristic of cancer progression from PT to AM might exist. However, no probe was found significantly differentially methylated between PT and LR, and between PT and CL. This may be due to the lack of cancer progression marker at the methylation level between PT and relapse, to the fact that most relapses may not be biologically related to the PT, or to the small size of the cohort which limits the power of statistical tests. The top 50 probes ranked by p-value then by absolute methylation variation between the primary tumor and its recurrence is also provided in Tables S4 (PTLR) and S5 (PTCL). No overlap existed between the three lists except for one gene (*PI3K5R* between the PT/AM and PT/LR datasets). All the corresponding quantile-quantile plots are available in Figure S1.

On the PT/AM cohort, the SVM model correctly identified the PT and AM in 18 out of 20 held-out pairs (90% success rate, $P\text{-value} = 2.0 \times 10^{-4}$) when considering the whole methylation profile probes. The SVM model obtained after dimensionality reduction by filtering the 22 most significant probes selected according to a Wilcoxon test gave a 100% accuracy. As illustrated in Figure 1, good accuracy was still achieved when considering an increasing number of probes (Accuracy $\sim 90\%$). On the PT/LR and PT/CL cohorts, however, the success rate was respectively 58% (10 out of 17 pairs, $P\text{-value} = 0.31$) and 27% (3 out of 11 pairs, $P\text{-value} = 0.11$) when taking all probes into account. Note that these values are not significantly different from random guess.

Methylation conservation between PT and their matched metastasis or recurrence

Instead of searching for differences between PT and their matched metastasis or recurrence, which may characterize markers for cancer progression, the study also focuses on similarities between methylation profiles, which may be useful for example to characterize clonality between a PT and a recurrence. A hierarchical clustering was first performed for all samples within each cohort to characterize the similarities between real matched pairs compared to unrelated samples. The resulting dendrograms are presented in Figure 2. Interestingly we see that matched pairs of PT and metastasis/recurrence samples are usually closer to each other than to any unrelated tissues in the PT/AM cohort (15 out of 20, 75%), less often in the PT/LR cohort (8 out of 17, 41%), and never in the PT/CL cohort. This observation is consistent with decreasing proportions of real clonal pairs from the PT/AM to the PT/CL set.

Another way to see this phenomenon is to assess statistically, within each cohort, how the methylation distances between matched pairs differ from the methylation distances between unmatched pairs. Figure 3 displays the distributions of methylation distances for different sets of sample pairs compared to the distance between matched sample pairs. We also display in Figure 4 the boxplot of methylation distances by groups. Real matched pairs between a PT and its corresponding metastasis or recurrence are significantly closer in terms of global methylation than a random pair of samples taken from two different individuals, both in the PT/AM cohort ($P\text{-value} = 3.5 \times 10^{-7}$) and in the PT/LR cohort ($P\text{-value} = 1.6 \times 10^{-6}$). This is however not true in the PT/CL cohort, where we detect no differences between correctly and randomly matched pairs ($P\text{-value} = 0.44$).

Table 4. Significantly differentially methylated genes between PT and AM samples.

CpG	Gene	Pvalue	Methylation Variation
cg20161089	<i>IFI27</i>	0.013	0.238
cg18140857	<i>RDHE2</i>	0.013	0.102
cg04619381	<i>LOC222171</i>	0.013	-0.048
cg23698969	<i>SLC22A18</i>	0.013	0.042
cg16179125	<i>CTSZ</i>	0.020	0.182
cg24959428	<i>GBP6</i>	0.020	0.126
cg22630748	<i>INHBE</i>	0.020	0.100
cg03623878	<i>MCF2L</i>	0.020	-0.050
cg25115460	<i>TP73</i>	0.022	0.109
cg11946165	<i>CTSK</i>	0.022	0.098
cg01318557	<i>LAT2</i>	0.022	0.063
cg13453139	<i>PIK3R5</i>	0.022	-0.063
cg21416237	<i>FKBP10</i>	0.028	0.085
cg19814116	<i>KCNAB2</i>	0.031	-0.217
cg22392666	<i>FXYD7</i>	0.031	-0.217
cg18212039	<i>EXTL1</i>	0.031	0.106
cg03532879	<i>SMAF1</i>	0.031	0.041
cg27149093	<i>SLC41A2</i>	0.032	0.145
cg18946226	<i>MYR8</i>	0.032	0.139
cg15448245	<i>GGTLA1</i>	0.032	0.132
cg15792367	<i>KLK11</i>	0.032	0.111
cg07459489	<i>SLC30A8</i>	0.032	0.085
cg15021292	<i>PIK3R1</i>	0.032	0.075
cg26267561	<i>OXT</i>	0.032	0.072
cg08647446	<i>RASSF6</i>	0.032	-0.070
cg20967028	<i>ART4</i>	0.032	-0.062
cg08550724	<i>C6orf182</i>	0.032	0.052
cg09737668	<i>SLAMF9</i>	0.032	0.049
cg23036025	<i>SLC27A5</i>	0.032	0.047
cg15296858	<i>PPM1G</i>	0.032	0.038
cg04961553	<i>OCIAD2</i>	0.032	-0.018
cg17558126	<i>RASSF5</i>	0.038	-0.137
cg06852652	<i>CYP2C18</i>	0.038	0.055
cg05840031	<i>PAX6</i>	0.038	-0.031
cg15043801	<i>DNMT1</i>	0.038	-0.016
cg05649009	<i>CHRNA1</i>	0.046	0.171
cg16176379	<i>AYTL1</i>	0.046	-0.128
cg05538432	<i>C15</i>	0.046	0.123
cg25151295	<i>RANBP5</i>	0.046	0.121
cg23841186	<i>SOAT2</i>	0.046	0.096
cg05656364	<i>VAMP8</i>	0.046	-0.085
cg14833385	<i>HLA-DMA</i>	0.046	-0.085
cg27655855	<i>CST9L</i>	0.046	0.084
cg27461196	<i>FXYD1</i>	0.046	0.069
cg14106308	<i>VEPH1</i>	0.046	-0.056
cg22857604	<i>RASSF5</i>	0.046	-0.053
cg10891879	<i>CASZ1</i>	0.046	-0.029
cg25042226	<i>PAX8</i>	0.046	-0.025
cg11655418	<i>RPS10</i>	0.046	0.005

CpG: CpG probe name. **Gene:** Associated gene. **Pvalue:** FDR corrected p-value. **Methylation Variation:** Mean variation of methylation from the primary tumor to the axillary metastasis.

doi:10.1371/journal.pone.0103986.t004

In addition, we calculated the distribution of distances between the CL tumors. We performed the same analysis between the PT tumors. We observed that the distribution were not significantly different ($P\text{-value} = 0.52$), as expected. This is in agreement with the assumption we made that CL tumors could be considered as new primary tumors. Finally, we also compared the distribution of distances between the healthy breast tissue i and all the other healthy breast tissues from the cohort to assess the heterogeneity between normal breast tissues.

Clonality detection based on methylation profiles

The above results suggests that methylation profiles tend to be conserved during clonal expansion (such as samples in the PT/AM cohort), but strongly differ between unrelated tumors in a given person (such as samples in the PT/CL cohort). Moreover, methylation seems to be a stable mechanism in normal tissues compared to cancerous ones. It is therefore tempting to use methylation distance as a tool to discriminate true recurrences from new tumors in ambiguous cases, that is, for samples in the PT/LR cohort.

9 out of 17 PT/LR pairs (52%) have a MS score higher than the threshold given by the 95% percentile of the MS score between unrelated pairs ($MS_{Threshold} = 6.6 * 10^{-4}$) as shown in Figure 5; they are therefore considered as clonal pairs from the methylation point of view. The remaining 8 pairs are considered as non-clonal, meaning that the LR may correspond to a new primary tumor. Figure S2 shows how related pairs are similar compared to unrelated pairs for the PT/AM (Panel A) and PT/CL (Panel B) groups.

Comparison between the methylation-based similarity measure MS score with the partial identity score (PIS), a copy-number based similarity measure developed by [16] show a good correlation overall ($\rho = 0.55$, $P\text{-value} = 3.7 * 10^{-5}$, see Figure 6). Table 5 gives a comparison of the outcomes given by methylation-based, copy-number based and clinical-based classification of LR as TR or NP. The methylation-based classification method agreed with the copy-number based PIS classification method on 14 out of 17 pairs (concordance = 82%, $P\text{-value} = 6.3 * 10^{-3}$) and agreed with the clinical-based classification on 14 out of 17 pairs (concordance = 82% , $P\text{-value} = 6.3 * 10^{-3}$).

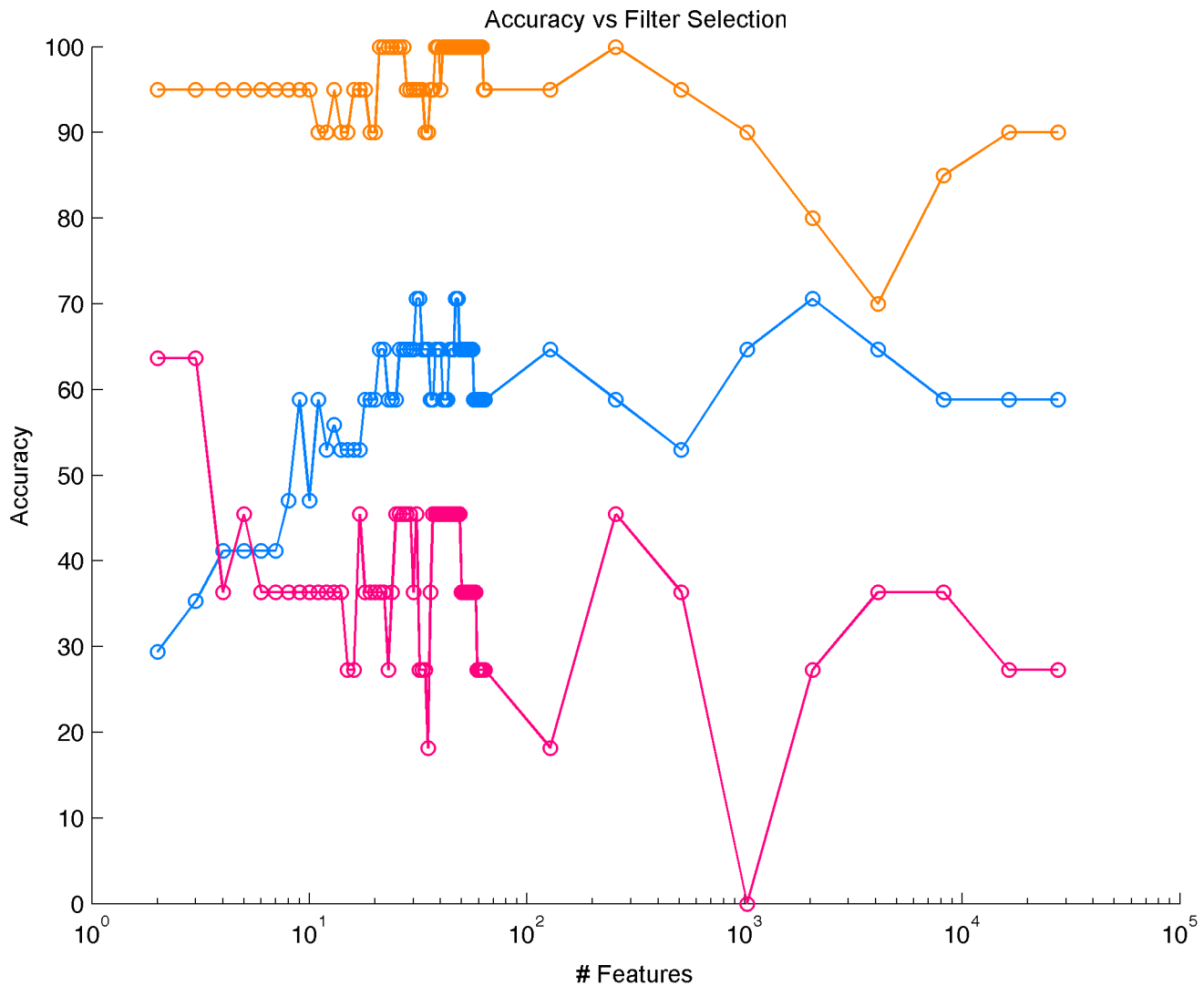


Figure 1. Accuracy of multivariate analysis with respect to feature selection to classify primary tumors from locoregional evolutions. Accuracy to classify PT from AM (resp. LR, resp. CL.) is represented in yellow (resp. blue, resp. pink). doi:10.1371/journal.pone.0103986.g001

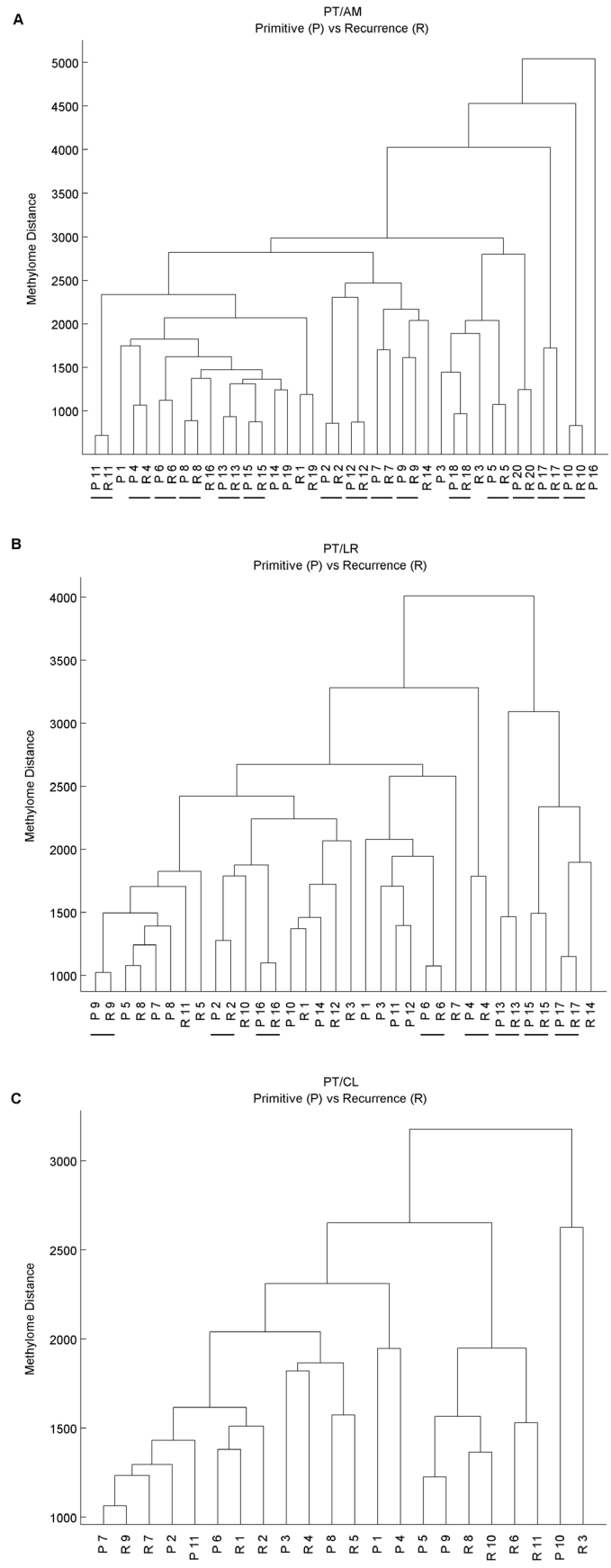


Figure 2. Study of similarity between matched primary tumors and recurrences by hierarchical clustering. Hierarchical clustering based on the manhattan distance between methylome profiles with complete linkage was performed. Real pairs that are closer to each other than to any other samples are underlined. Panel A (resp. B, resp. C) represents the PT/AM (resp. PT/LR, resp. PT/CL) set. doi:10.1371/journal.pone.0103986.g002

Finally, the different classifications of LR as TR or NP were correlated with time-to-recurrence and metastasis-free survivals. The differences in time-to-recurrence for the two groups defined by methylation-based classification or the clinical and histological classification were not statistically significant (P-value = 0.83 and P-value = 0.12). It was however significant using the partial identity score (P-value = 0.03) (Figure S3). This is interesting in the sense that one of the main criteria to distinguish TR and NP is the time-to-recurrence. Therefore, methylation-based classification is based on more information than time only.

The difference in metastasis-free survival of patients with TR and NP was not significant based on methylation (P-value = 0.52, Hazard-Ratio = 3.7, 5 year metastasis-free survival = 75% for NP), copy-number (P-value = 0.15, Hazard-Ratio = 16.9, 5 year metastasis-free survival = 86% for NP) or clinical features (P-value = 0.17, Hazard-Ratio = 6.3, 5 year metastasis-free survival = 86% for NP) (Figure 7). Adjusting for age, grade and ER status did not yield more significant results except for copy-number based classification (P-value = 0.46, Table S6).

Discussion

We studied alterations of methylation profiles from primary breast carcinomas and different types of recurrences, namely, axillary metastases, local recurrences and contralateral breast carcinomas. For this particular dataset, we observed significant methylation differences for 49 CpG probes, which characterizes the progression between a PT and its AM. Consistent with this result, a multivariate analysis with a linear SVM classifier using a small subset of probes perfectly distinguished PTs from AMs with a 100% accuracy. Several significantly differentially methylated probes correspond to genes involved in cancer-related mechanisms such as cell death (*MCF2L*, *RASSF5*, *RASSF6*, *CASZ1*, *SLC22A18*, *IFI27*), tumorigenesis (*CTSZ*, *TP73*, *CTSK*, *PIK3R1*, *KLK11*), cell cycle (*PPM1G*, *RANBP5*, *VAMP8*) and cell differentiation (*SMAF1*, *PAX6*, *PAX8*). On the contrary, for the PT/LR and PT/CL sets, univariate analyzes were not able to find significantly differentially methylated probes. This absence of specific epigenetic alterations between the primary tumors and the

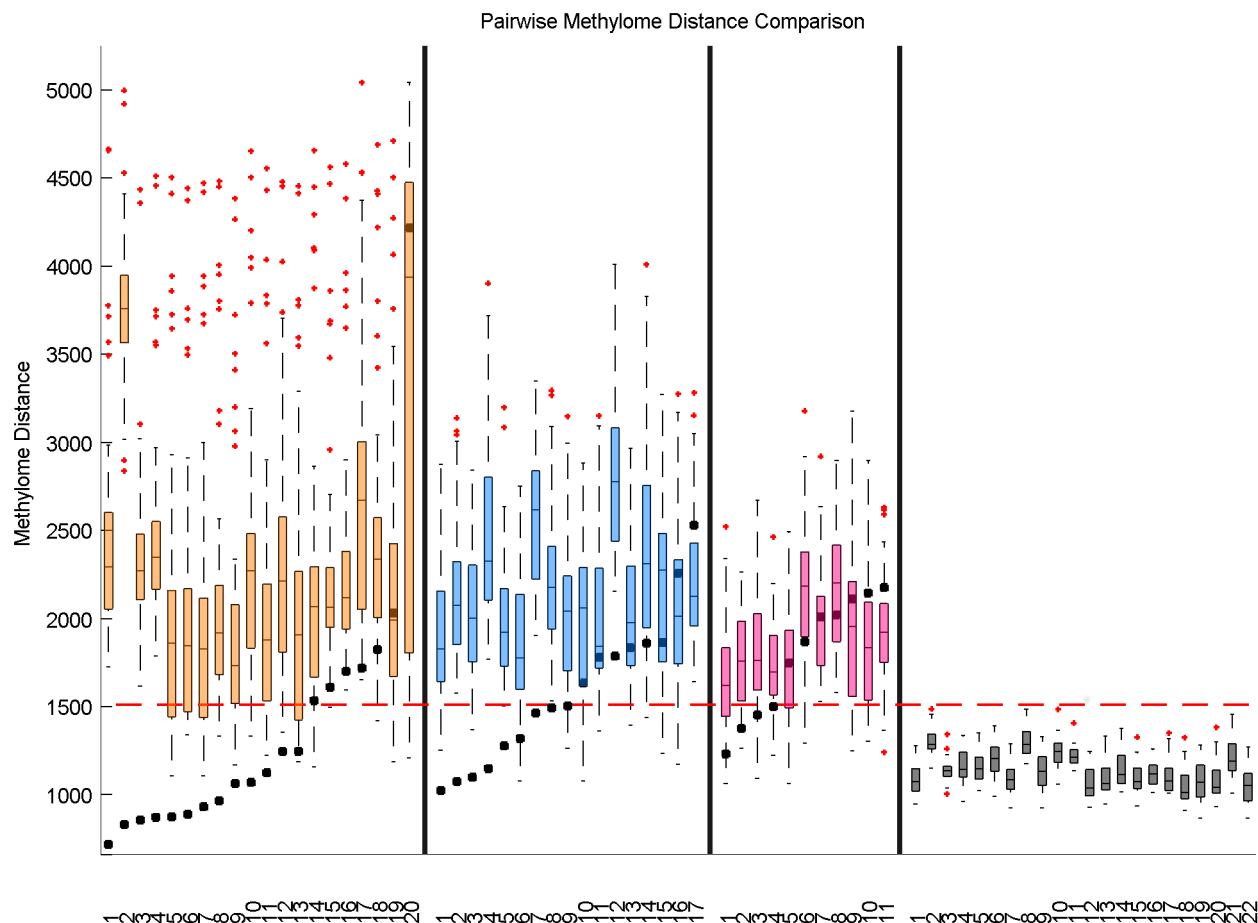


Figure 3. Distribution of methylation distances between different samples pairs for each groups. **Real:** boxplot of methylome distances for all matched pairs that is a PT and its corresponding metastasis or recurrence. **Artificial:** boxplot of methylome distances for all unmatched pairs that is a PT and an unrelated metastasis or recurrence. **Primary:** boxplot of methylome distances to distances between two PT of two different individuals. **Recurrence:** boxplot of methylome distances between two metastasis or recurrence samples of two different individuals. doi:10.1371/journal.pone.0103986.g003

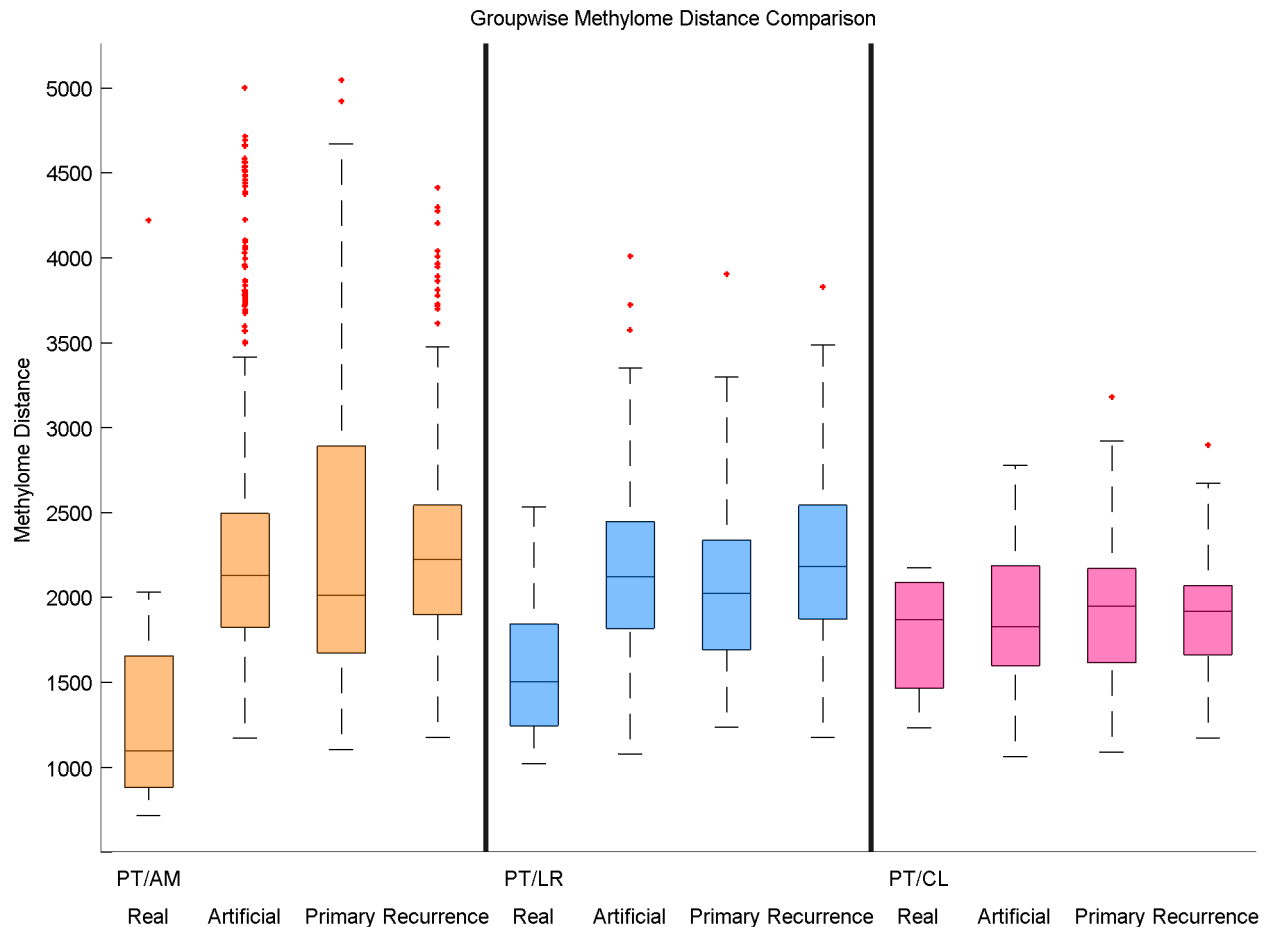


Figure 4. Pairwise methyome distance for each samples. Each boxplot represents the Manhattan distance between primary tumor i and an unrelated locoregional evolution, or the Manhattan distance between locoregional evolution i and an unrelated primary tumor. The black square represent the Manhattan distance between the matched primary tumor and locoregional evolution from sample i . The yellow (resp. blue, resp. pink) panel represents the PT/AM (resp. PT/LR, resp. PT/CL) set. The last panel represents the distribution of distances between the healthy breast tissue i and all the other healthy breast tissues from the cohort. doi:10.1371/journal.pone.0103986.g004

local recurrences or the contralateral breast recurrences was confirmed by the poor performances of linear classifiers, unable to separate PT from LR nor PT from CL significantly better than random guesses. Nevertheless, the absence of methylation markers in the PT/LR and the PT/CL groups does not necessarily mean that the primary tumor and the recurrence are independent. We cannot rule out the possibility that the recurrence arises from a specific subclone which does not match the major subclone of the primary tumor. One could for example analyze the methylation profiles of several microdissections samples of the primary tumor to study potential heterogeneity.

The second part of the study focused on observing stability in methylation profiles. It is interesting to note that although PTs and AMs were significantly differentiable using a subset of probes, they also have overall very similar methylation profiles indicating that the tumors might actually be clones with specific alterations characteristic of the lymph node status. The subset of genes determined in the first part, if confirmed, could be associated with bad prognosis. On the other part, although the LRs and the CLs were not significantly different from their primary tumors, they tend to have overall different methyome profiles especially for the CLs. The overall different methyome profiles for the PT/CL set

was expected since CLs are usually considered to be independent tumors.

The results above suggested to use global methylation analysis as a measure of clonality to tackle the subclonal populations in the local recurrences as proposed by Veronesi et al. [9]. A methylation-based classification was proposed to distinguish LRs as either true recurrences of the first PT or new PT [10]. A comparison with both clinical and copy-number based classifications on the same cohorts agreed on 14 out of 17 samples (82% concordance, $P\text{-value} = 6.410^{-3}$) for both methods, although comparisons on larger cohorts are needed to assess the performance of methylation-based classification. Moreover, a good correlation between the methylation-based similarity score and the copy-number based similarity score seems to indicate a link between modifications at the genomic and epigenomic levels. Although the role of methylation in gene expression has thoroughly been studied [24–26], the relationship between methylation and copy-number still remains unclear. Houseman et al. [33] note that there is a negative bias of methylation when one or both alleles are lost but none in case of gains. Several other studies have reported correlation between the two mechanisms in different types of cells. Strong associations have been reported in urothelial carcinoma [34], head and neck squamous cell carcinoma

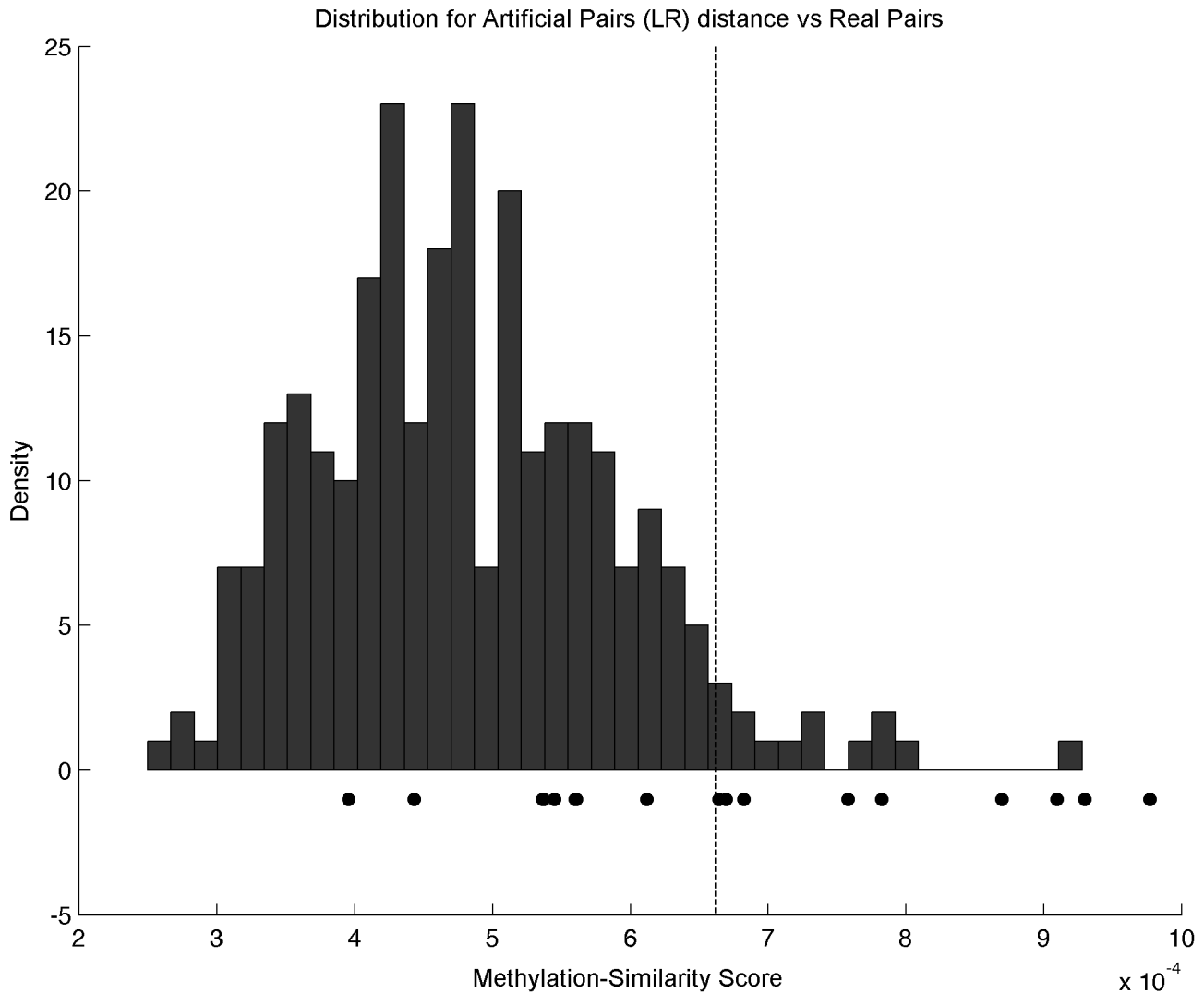


Figure 5. Histogram of the distribution of methylation-similarity score (MS) between unrelated PT/LR pairs. MS score for matched pairs is represented by circles. The vertical dashed line corresponds to the 95% quantile of the distribution of the MS scores for the unrelated pairs, used as a threshold to define clonal pairs ($MS_{Threshold} = 6.6 \times 10^{-4}$). doi:10.1371/journal.pone.0103986.g005

mas [35], and mesothelioma [36]. Our study provides new evidence for association between methylation and copy-number on a global scale.

The discordances between the methylation-based classification method and the usual clinical method are discussed here for the samples 7, 8 and 14, although no actual method is a gold standard for classifying TR from NP. Sample 8 filled almost all the requirements for clinical classification as TR (location, receptor status) but failed in aggressiveness and type of tumor (PT was ductal type 2 and LR was lobular type 1). A decrease of aggressiveness of the recurrence could be explained by the use of neoadjuvant therapies. For the change of type, Fisher et al. showed that a mixing of ductal and lobular breast carcinoma was a possibility in 6% of the patients [37] which could explain the change in type. Sample 7 was classified as TR by clinical classification and as NP by both methylation and copy-number based classifications. This suggests some limitations to methods based only on clinical features.

An interesting question for clinical applications would have been to predict whether a primary tumor would relapse (either as

AM, LR or CL) or not. However, the patient cohort used in this study does not allow to address this question. Indeed, one would require to compare the methylation profiles of patients who did not display any relapse (AM, LR and CL) to those of the current study.

Materials and Methods

Patients Selection

The patients were 49 years old or younger at diagnosis of the initial tumor; all patients were premenopausal; and had no previous history of cancer, except for one nonmelanoma skin cancer. The patients' PT was either ductal or lobular invasive breast carcinoma. However, both types of tumors did not display significantly differentially methylated probes and were thus all included in this study (min P-value ≥ 0.89).

Specimens from patients with primary breast cancers and breast cancer recurrences were selected from freshly frozen samples of the Institut Curie tissue bank according to the following criteria: all patients had been treated at the Institut Curie by breast-

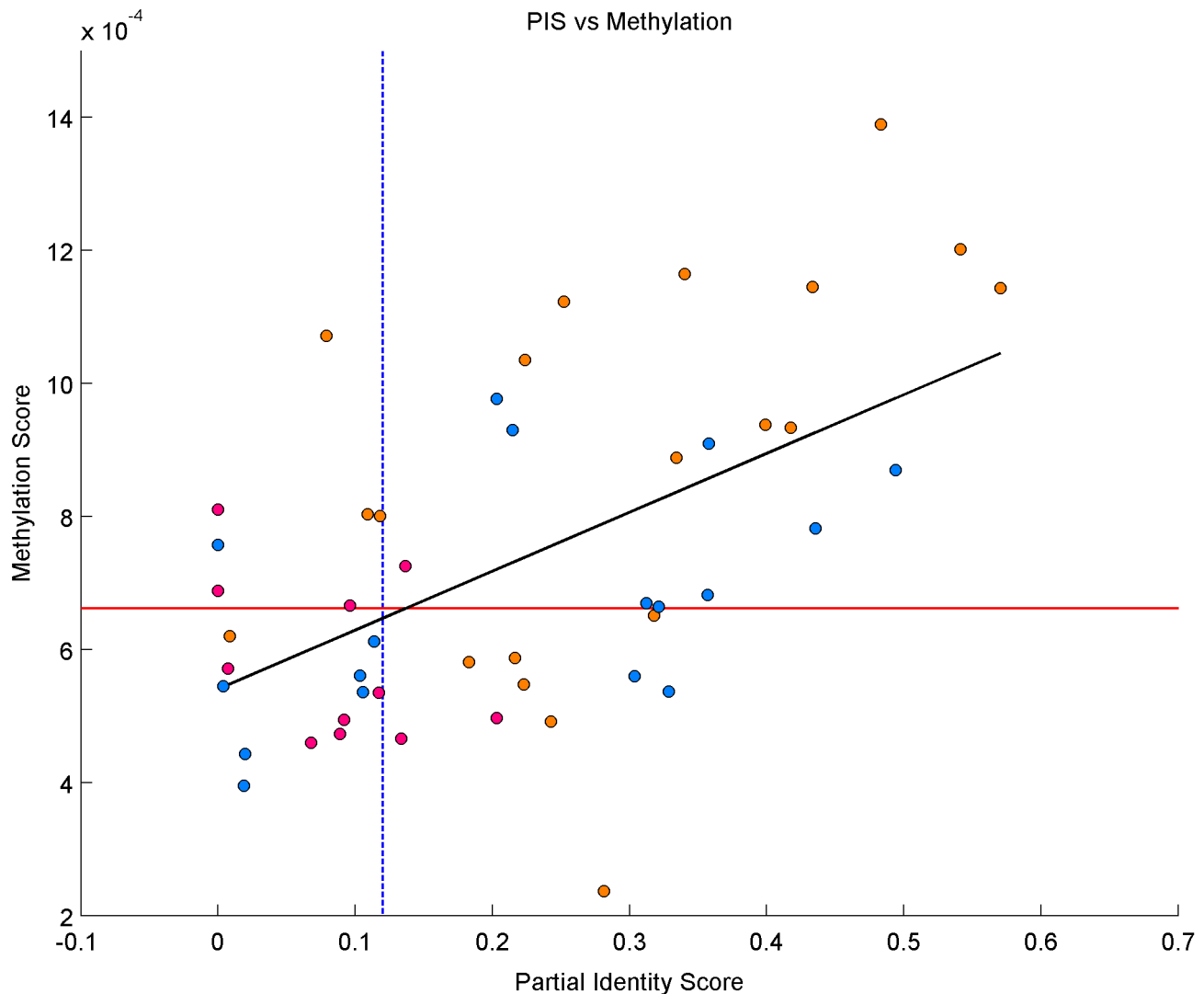


Figure 6. Correlation between methylation and copy-number scores. The horizontal red line (resp. vertical dashed blue line) corresponds to the 95% quantile of the distribution of the methylation-scores (resp. partial identity scores) for the unrelated pairs: $MS_{Threshold} = 6.6 * 10^{-4}$ (resp. $PIS_{Threshold} = 0.12$). PT/AM (resp. PT/LR, resp. PT/CL) pairs are colored in yellow (resp. blue, resp. pink). The black line corresponds to the linear regression between methylation and copy-number scores for all the datasets.
doi:10.1371/journal.pone.0103986.g006

conserving surgery, including dissection of the axillary lymph nodes in most patients, followed by radiotherapy to the breast with or without a boost to the tumor bed (external beam radiotherapy or brachytherapy) and/or to the regional lymph node-bearing areas if indicated and, when required, systemic treatment as part of their initial management. Methylation profiles did not significantly differ depending on either ER, PR, HER2 and grade characteristics (min adjusted P-value = 0.09).

To ensure that the data would be informative, genomic analyzes were restricted to tumors (primary and recurrences) in which at least 50% of cancer cells had been assessed by hematoxylin, eosin, and saffron staining of sections from snap-frozen samples. All the therapies were performed posterior to the biopsies of the primary tumors. Therefore, the studied methylation profiles are not modified by any potential effect of the treatments.

The 22 healthy breast tissues are taken from healthy women who underwent cosmetic plastic surgery at the Institut Curie. Part

of the PT/AM cohort is identical to the cohort studied by Bollet et al. [16].

All experiments were performed retrospectively and in accordance with the French Bioethics Law 2004–800, the French National Institute of Cancer (INCa) Ethics Charter and after approval by the Institut Curie review board and ethics committee (Comit de Pilotage of the Groupe Sein). In the French legal context, our institutional review board waived the need for written informed consent from the participants. Moreover, women were informed of the research use of their tissues and did not declare any opposition for such researches. Data were analyzed anonymously.

Methylation profiling

For each sample the methylation status at 27,578 positions in the genome was measured with the HumanMethylation27 BeadChip of Infinium technology [38] using the standard Illumina

Table 5. Comparison of classification methods for clonality between pairs in the PT/LR cohort.

Panel A		Scores			Classification			
Pair	Cor	PIS	MS ($\times 10^{-4}$)	Time (Years)	PIS	MS	Clinical	Divergence
1	1	0.019	4.42	6.5	NP	NP	NP	
2	3	0.435	7.82	3.2	TR	TR	TR	
3	11	0.018	3.95	6.4	NP	NP	NP	
4	16	0.303	5.59	3.8	TR	NP	NP	PIS
5	12	0.113	6.11	3.4	NP	NP	NP	
6	13	0.214	9.29	4.6	TR	TR	TR	
7	15	0.105	5.36	3.2	NP	NP	TR	Clinical
8	2	0	7.57	5.2	NP	TR	NP	MS
9	4	0.203	9.76	3.5	TR	TR	TR	
10	14	0.321	6.64	2.4	TR	TR	TR	
11	18	0.003	5.44	2.2	NP	NP	NP	
12	20	0.103	5.60	1.4	NP	NP	NP	
13	21	0.356	6.82	4.2	TR	TR	TR	
14	23	0.328	5.37	0.9	TR	NP	TR	MS
15	24	0.312	6.69	1.4	TR	TR	TR	
16	25	0.357	9.09	2.7	TR	TR	TR	
17	26	0.493	8.69	2.0	TR	TR	TR	

Cor (Correspondence): correspondence number with the Bollet/Servant cohort. **scores**: scores obtained with partial identity (PIS) or methylation (MS). **Time**: time elapsed between diagnosis of the PT and diagnosis of the recurrence. **Classification**: classification of the recurrence based on copy number (PIS), methylation (MS) or clinical features (clinical). **Divergence**: which method deviated from the others.
doi:10.1371/journal.pone.0103986.t005

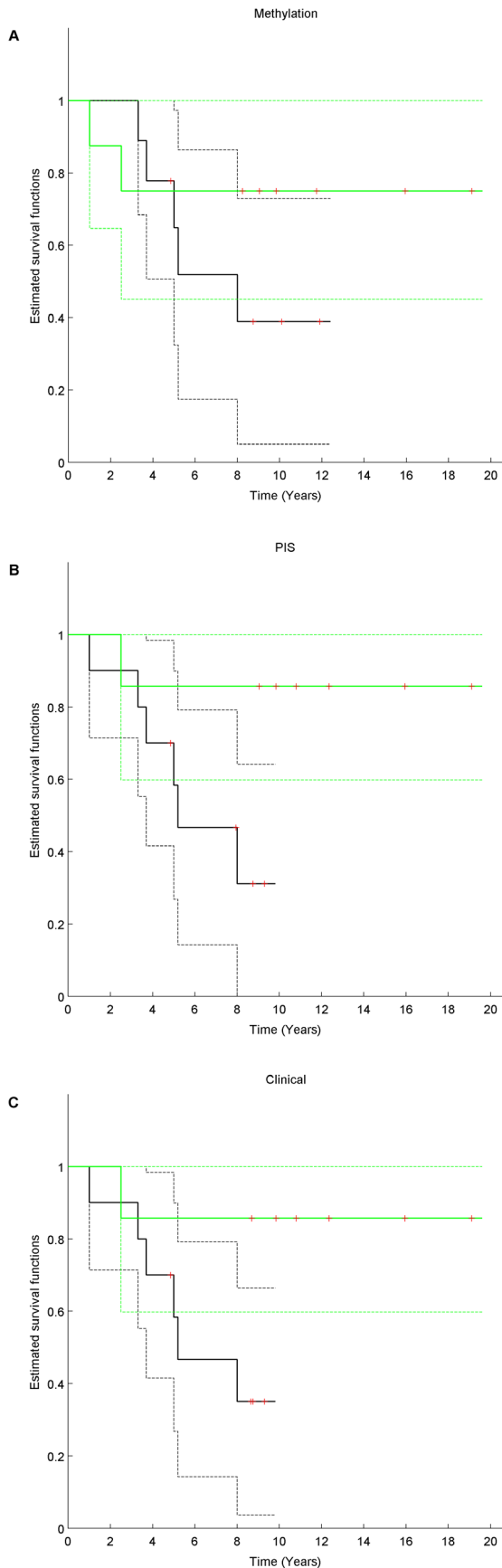


Figure 7. Kaplan-Meier estimates of the metastasis-free survival between TR and NP for the different classification methods. The full black (resp. green) line corresponds to the survival for samples classified as TR (resp. NP) and the corresponding dashed lines correspond to upper and lower 95% CI. The red crosses represent censored data. Panel A (resp. B, resp. C) represent the methylation-based (resp. copy-number based, resp. clinical based) classification. doi:10.1371/journal.pone.0103986.g007

protocol. Quality control was assessed using in-built Illumina technology.

Copy number based classification

The PIS score, based on copy number alterations similarities between the primary tumor and its recurrence, was retrieved from [16] for the same population.

Clinical Classification

Histopathologic characteristics were reviewed by a single pathologist. The histological and biological properties of each sample was determined by subjecting tissue sections to immunohistochemical analysis for the estrogen receptor (clone 6F11, 1:200 dilution; Novocastra, Newcastle Upon Tyne, England) and progesterone receptor (clone 1A6, 1: 200 dilution; Novocastra) antibodies. Tumors were considered to be positive for these receptors if at least 10% of the invasive tumor cells in a section showed nuclear staining [39,40]. The HER2 analysis was performed using the standard ASCO guidelines [41]. In accordance with theories of the clonal evolution of tumor cell populations, LR were clinically defined as TR if they had the same histologic subtype (ductal or lobular) and a similar or increased growth rate, similar estradiol, progesterone and HER2 receptor statuses, and similar or decreased differentiation as the initial tumor [10]. TR also had to share with their PT the same breast quadrant. Thus, new PT were clinically defined as such when the LR had occurred in a different location, had a distinct histologic type, or had less aggressiveness features (lower grade, presence of hormonal receptors) than the initial tumor.

Data analysis

A spatial normalization process was applied to all profiles [42]. Among the 27,578 probes measured on each sample, 5 probes were removed due to missing values for some individuals, and all subsequent analysis was performed on the 27,573 remaining probes.

Differentially methylated probes between PT and their matched AM, LR and CL are obtained using two-sided paired and unpaired Wilcoxon tests, correcting the p-values for multiple testing with the methods of Benjamini and Hochberg [43]. Multivariate analysis was performed using a linear support vector machine (SVM) multidimensional classifier on either the complete methylation profile or after dimensional reduction by considering only the most significant probes based on the Wilcoxon test. A p-value was calculated to assess the significance of the predictor accuracy compared to a predictor that would predict classes randomly. Unsupervised classifications were performed with complete linkage agglomerative clustering using the MATLAB bioinformatics toolbox, while the support vector machine implemented in LIBSVM [44] was computed with a linear kernel and nested leave-one-out cross validation for parameter selection for supervised classification.

The similarity between two copy number profiles is assessed with the partial identity score (PIS) as defined by Bollet et al. [16], which is based on the quantity of shared breakpoints between the two profiles and their frequencies. Following [16], a recurrence

from a matched PT/LR pair was considered TR based on copy numbers when the PIS between the PT and LR profiles was above the 95% quantile of the empirical PIS distribution between unrelated sample pairs. Similarly, a Methylation-Similarity score (MS) is defined based on the methylation profiles of a PT and its matched LR as the inverse of the Manhattan distance between their methylation profiles considered as 27,573-dimensional vectors. LR are then classified as TR of its matched PT when the MS score is above the 95% quantile of the empirical MS distribution between unrelated pairs. As a baseline, these results were compared to the Manhattan distance between unrelated normal breast tissues.

Metastasis-free survival was estimated by the Kaplan-Meier Method [45] and compared between the group of patients who were diagnosed as TR and the group diagnosed as NP using the log-rank test. The confidence interval of the hazard ratio was obtained using a semi-parametric Cox model [46]. Computation was done using MATLAB packages Logrank [47] and KMPlot [48].

Supporting Information

Figure S1 Quantile-quantile plot of the Wilcoxon test statistics for each groups. Plot of the data quantiles (black dots) against normal theoretical quantiles. The red line is $y = x$. (TIF)

Figure S2 Histograms of the distribution of Methylation-Similarity score (MS) between unrelated PT/AM and PT/CL pairs. MS score for matched pairs is represented by crosses for the PT/AM pairs (Panel A) and by stars for the PT/CL pairs (Panel B). The vertical dashed line corresponds to the 95% quantile of the distribution of the MS scores for the unrelated pairs. (TIF)

Figure S3 Correlation between time to recurrence and classification of the recurrence. Boxplots of time between the primary tumor and the local recurrence depending on the classification as true recurrence (TR) or new primary tumor (NP) according to the methylation-based, copy-number based (PIS) and clinical based classification. (TIF)

Table S1 Complete PT/LR Clinical and histological features. **Cor** (Correspondence): correspondence number with the Bollet/Servant cohort from [16], **Type**: histological type of the tumor (D = ductal, L = lobular), **Grade**: Aggressiveness of the tumor (1 to 3), **ER**: percentage of estrogen receptors, **PR**:

percentage of progesterone receptors present, **HER2**: presence of HER2 receptors, **Loc** (Location): 1 if the recurrence was located less than 4cm from the PT.

(TIF)

Table S2 Complete PT/CL Clinical and histological features. **Type**: histological type of the tumor (D = ductal, L = lobular, Med = Medullary, Meta = Metaplastic), **Grade**: Aggressiveness of the tumor (1 to 3), **ER**: percentage of estrogen receptors present, **PR**: percentage of progesterone receptors present, **HER2**: presence of HER2 receptors.

(TIF)

Table S3 Complete PT/AM Clinical and histological features. **Age**: Age of the patient at diagnosis of the primary tumor in years, **Type**: histological type of the tumor (D = ductal, L = lobular, Meta = Metaplasia), **Grade**: Aggressiveness of the tumor (1 to 3), **ER**: percentage of estrogen receptors present, **PR**: percentage of progesterone receptor present, **HER2**: presence of HER2 receptors.

(TIF)

Table S4 Top 50 CpG loci between PT and LR samples. **CpG**: CpG probe name. **Gene**: Associated gene. **Pvalue**: FDR corrected p-value. **Methylation Variation**: Mean variation of methylation from the primary tumor to the local recurrence.

(TIF)

Table S5 Top 50 probes between PT and CL samples. **CpG**: CpG probe name. **Gene**: Associated gene. **Pvalue**: FDR corrected p-value. **Methylation Variation**: Mean variation of methylation from the primary tumor to the contralateral recurrence.

(TIF)

Table S6 Predictive impact of the classification methods on survival in breast cancer. **Variables**: variable considered for predictive impact adjusted for the other variables present in the table. **Coef**: Associated coefficient in the Cox regression. **lower/upper .95**: lower and upper 95% confidence interval. **Pvalue**: P-value associated with the predictive impact on survival.

(TIF)

Author Contributions

Conceived and designed the experiments: AF JPV FR. Performed the experiments: AP MC BSZ. Analyzed the data: MM NS VS. Contributed reagents/materials/analysis tools: MM JPV. Contributed to the writing of the manuscript: MM NS VS JPV FR.

References

- Van Dongen JA, Voogd AC, Fentiman IS, Legrand C, Sylvester R, et al. (2000) Long-Term Results of a Randomized Trial Comparing Breast-Conserving Therapy with Mastectomy: European Organization for Research and Treatment of Cancer 10801 Trial. *J Natl Cancer Inst* 92: 1143–50.
- Bartlink H, Horiot JC, Poortmans PM, Van den Bogaert W, Fourquet A, et al. (2007) impact of a higher radiation dose on local control and survival in breast-conserving therapy of early breast cancer: 10-year results of the randomized boost versus no boost EORTC 22881–10882 trial. *J Clin Oncol* 25: 3259–65.
- Vichapat V, Garmo H, Holmqvist M, Liljegren G, Wärnberg F, et al. (2012) Tumor Stage Affects Risk and Prognosis of Contralateral Breast Cancer: Results From a Large Swedish Population Based Study. *J Clin Oncol* 30: 3478–3485.
- Metcalfe K, Lynch HT, Ghadirian P, Tung N, Olivetto I, et al. (2004) Contralateral breast cancer in BRCA1 and BRCA2 mutation carriers. *J Clin Oncol* 22: 2328–2335.
- Jatoi I (1999) Management of the axilla in primary breast cancer. *Surg Clin North Am* 79: 1061–1073.
- Carter CL, Allen C, Henson DE (1989) Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* 63: 181–7.
- Borg A, Tandon AK, Sigurdsson H, Clark GM, Fernö M, et al. (1990) HER2/neu Amplification predicts poor survival in Node-positive Breast Cancer. *Cancer Res* 50: 4332.
- Haffty BF, Reiss M, Beinfield M, Fischer D, Ward B, et al. (1996) Ipsilateral breast tumor recurrence as a predictor of distant disease: implications for systemic therapy at the time of local relapse. *J Clin Oncol* 14: 52–57.
- Veronesi U, Marubini E, Del Vecchio M, Manzari A, Greco M, et al. (1995) Local Recurrences and Distant Metastases After Conservative Breast Cancer Treatments: Partly Independent Events. *J Natl Cancer Inst* 87: 19–27.
- Smith TE, Lee D, Turner BC, Carter D, Haffty BG (1999) True recurrence vs. new primary ipsilateral breast tumor relapse: An analysis of clinical and pathologic differences and their implications in natural history, prognoses, and therapeutic management. *Int J Radiat Oncol* 48: 1281–1289.
- Dawson L, Chow E, Goss PE (1998) Evolving perspectives in contralateral breast cancer. *Eur J Cancer* 34: 2000–2009.
- Healey EA, Cook EF, Orav EJ, Schnitt SJ, Connolly JL, et al. (1993) Contralateral breast cancer: clinical characteristics and impact on prognosis. *J Clin Oncol* 11: 1545–1552.

13. Ellsworth RE, Ellsworth DL, Neatrou DM, Denyarmin B, Lubert SM, et al. (2005) Allelic Imbalance in Primary Breast Carcinomas and Metastatic Tumors of the Axillary Lymph Nodes. *Mole Can Res* 3: 71–77.
14. Weigelt B, Wessels LF, Bosma AJ, Glas AM, Nuyten DS, et al. (2005) No common denominator for breast cancer lymph node metastasis. *Br J Cancer* 93: 924–932.
15. Feng Y, Sun B, Li X, Zhang L, Niu Y, et al. (2007) Differentially expressed genes between primary cancer and paired lymph node metastases predict clinical outcome of node-positive breast cancer patients. *Breast Cancer Res Treat* 103: 319–329.
16. Bollet M, Servant N, Neuvial P, Decraene C, Lebigot I, et al. (2008) High-Resolution Mapping of DNA Breakpoints to Define True Recurrences among Ipsilateral Breast Cancers. *J Natl Cancer Inst* 100: 48–58.
17. Ostrovskaya I, Olshen AB, Seshan VE, Orlov I, Albertson DG, et al. (2010) A Metastasis or a Second Independent Cancer? Evaluating the clonal origin of tumors using array copy number data. *Stat Med* 29: 1608–1621.
18. West NR, Panet-Raymond V, Truong PT, Alexander C, Babinsky S, et al. (2011) Intratumoral Immune Responses can Distinguish New Primary and True Recurrence Types of Ipsilateral Breast Tumor Recurrences (IBTR). *Breast Cancer (Auckl)* 5: 105–115.
19. Vicini FA, Antonucci JV, Goldstein N, Wallace M, Kestin L, et al. (2007) The Use of Molecular Assays to Establish Definitively the Clonality of Ipsilateral Breast Tumor Recurrences and Patterns of In-breast Failure in Patients with Early-stage Breast Cancer Treated with Breast-conserving Therapy. *Cancer* 109: 1264–72.
20. Van Der Sijp JR, van Meerbeek JP, Maat AP, Zondervan PE, Sleddens HF, et al. (2002) Determination of the molecular relationship between multiple tumors within one patient is of clinical importance. *J Clin Oncol* 20: 1105–1114.
21. Shibata A, Tsai YC, Press MF, Henderson BE, Jones PA, et al. (1996) Clonal Analysis of bilateral breast cancer. *Cancer Res* 56: 743–748.
22. Brommsson S, Jönsson G, Strand C, Grabau D, Malmström P, et al. (2008) Tiling array-CGH for the assessment of genomic similarities among synchronous unilateral and bilateral invasive breast cancer tumor pairs. *BMC Clin Pathol* 8: 6.
23. Imyanitov EN, Suspitsin EN, Grigoriev MY, Togo AV, Belogubova EV, et al. (2002) Concordance of Allelic Imbalance Profiles in Synchronous and Metachronous Bilateral Breast Carcinomas. *Int J Cancer* 100: 557–564.
24. Razin A, Riggs AD (1980) DNA Methylation and Gene Function. *Science* 210: 604–10.
25. Tate PH, Bird AP (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev* 3: 226–31.
26. Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16: 6–21.
27. Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, et al. (2001) Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet* 10: 683–692.
28. Laird PW, Jaenisch R (1994) DNA Methylation and Cancer. *Hum Mol Genet* 3: 1487–95.
29. Ehrlich M (2002) DNA methylation in cancer: too much, but also too little. *Oncogene* 21: 5400–5413.
30. Das PM, Singal R (2004) DNA Methylation and cancer. *J Clin Oncol* 22: 4632–4642.
31. Jones PA, Baylin SB (2007) The Epigenomics of Cancer. *Cell* 128: 683–692.
32. Schermelleh L, Haemmer A, Spada F, Rösing N, Meilinger D, et al. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res* 35: 4301–4312.
33. Houseman EA, Christensen BC, Karagas MR, Wrensch MR, Nelson HH, et al. (2009) Copy number variation has little impact on bead-array-based measures of DNA methylation. *Bioinformatics* 25: 1999–2005.
34. Lauss M, Aine M, Sjö Dahl G, Veerla S, Patschan O, et al. (2012) DNA methylation analyses of urothelial carcinoma reveal distinct epigenetic subtypes and an association between gene copy number and methylation status. *Epigenetics* 7: 858–867.
35. Poage GM, Christensen BC, Houseman EA, McClean MD, Wiencke JK, et al. (2010) Genetic and epigenetic somatic alterations in head and neck squamous cell carcinomas are globally coordinated but not locally targeted. *PLoS One* 5: e9651.
36. Christensen BC, Houseman EA, Poage GM, Godleski JJ, Bueno R, et al. (2010) Integrated profiling reveals a global correlation between epigenetic and genetic alterations in mesothelioma. *Cancer Research* 70: 5686–5694.
37. Fisher ER, Gregorio RM, Fisher B, Redmond C, Vellios F, et al. (1975) The pathology of invasive breast cancer. A syllabus derived from findings of the National Surgical Adjuvant Breast Project (protocol no. 4). *Cancer* 36: 1–85.
38. Weisenberger DJ, Van Den Berg D, Pan F, Berman BP, Laird P (2008) Comprehensive DNA methylation analysis on the illumina Infinium assay platform. Technical report.
39. Balaton A, Baviera E, Galet B, Vauray P, Vuong P (1995) Immunohistochemical evaluation of estrogen and progesterone receptors on paraffin sections of breast carcinomas. Practical thoughts based on the study of 368 cases. *Arch Anat Cytol Pathol* 43: 93–100.
40. Balaton A, Coindre J, Collin F, Ettore F, Fiche M, et al. (1996) Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update. *Ann Pathol* 16: 144–148.
41. Wolff AC, Hammond EH, Hicks DG, Dowsett M, McShane LM, et al. (2013) Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update. *J Clin Oncol* 31: 3997–4013.
42. Sabbah C, Mazo G, Paccard C, Reyat F, Hupe P (2011) SMETHILLIUM: Spatial normalisation METHod for ILLumina InfiniUM HumanMethylation BeadChip. *Bioinformatics* 27: 1693–5.
43. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1: 289–300.
44. Chang CC, Lin CH (2011) LIBSVM: a library for support vector machines. *ACM TIST* 2: 27: 1–27: 27.
45. Kaplan EL, Meier D (1958) Nonparametric estimation from incomplete observation. *J Am Statist* 58: 457–481.
46. Cox DR, Oakes D (1984) *Analysis of Survival Data*. London: Chapman and Hall.
47. Cardillo G. Logrank. mathworks website. Available: <http://www.mathworks.com/matlabcentral/fileexchange/22317-logrank>. Accessed 2014 Jul 15th.
48. Cardillo G. Kmpplot. mathworks website. Available: <http://www.mathworks.com/matlabcentral/fileexchange/22293>. Accessed 2014 Jul 15th.