HHS Public Access

Author manuscript

Clin Cancer Res. Author manuscript; available in PMC 2018 August 01.

Published in final edited form as:

Clin Cancer Res. 2017 August 01; 23(15): 4402–4415. doi:10.1158/1078-0432.CCR-16-3115.

Genetic Heterogeneity in Therapy-Naïve Synchronous Primary Breast Cancers and Their Metastases

Charlotte K.Y. Ng[1,2,*], **Francois-Clement Bidard**[1,3,*], **Salvatore Piscuoglio**[1,2], **Felipe C. Geyer**[1,4], **Raymond S. Lim**[1], **Ino de Bruijn**[1], **Ronglai Shen**[5], **Fresia Pareja**[1], **Samuel H. Berman**[1], **Lu Wang**[1], **Jean-Yves Pierga**[3,6], **Anne Vincent-Salomon**[7], **Agnes Viale**[8], **Larry Norton**[9], **Brigitte Sigal**[7], **Britta Weigelt**[1], **Paul Cottu**[3], and **Jorge S. Reis-Filho**[1,10]

[1]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA [2]Institute of Pathology, University Hospital Basel, Basel, Switzerland [3]Department of Medical Oncology, Institut Curie, PSL Research University, Paris, France [4]Department of Pathology, Hospital Israelita Albert Einstein, Instituto Israelita de Ensino e Pesquisa, São Paulo, Brazil [5]Department of Biostatistics and Epidemiology, Memorial Sloan Kettering Cancer Center, New York, NY, USA [6]University Paris Descartes, Paris, France [7]Department of Pathology, Institut Curie, PSL Research University, Paris, France [8]Integrated Genomics Operations, Memorial Sloan Kettering Cancer Center, New York, NY, USA [9]Breast Medicine Service, Memorial Sloan Kettering Cancer Center, New York, NY, USA [10]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Abstract

Purpose—Paired primary breast cancers and metachronous metastases after adjuvant treatment are reported to differ in their clonal composition and genetic alterations, but it is unclear whether these differences stem from the selective pressures of the metastatic process, the systemic therapies or both. We sought to define the repertoire of genetic alterations in breast cancer patients with *de novo* metastatic disease who had not received local or systemic therapy.

Experimental Design—Up to two anatomically distinct core biopsies of primary breast cancers and synchronous distant metastases from nine patients who presented with metastatic disease were subjected to high-depth whole-exome sequencing. Mutations, copy number alterations and their cancer cell fractions, and mutation signatures were defined using state-of-the-art bioinformatics methods. All mutations identified were validated with orthogonal methods.

Results—Genomic differences were observed between primary and metastatic deposits, with a median of 60% (range 6%–95%) of shared somatic mutations. Whilst mutations in known driver genes including *TP53*, *PIK3CA* and *GATA3* were preferentially clonal in both sites, primary breast cancers and their synchronous metastases displayed spatial intra-tumor heterogeneity. Likely pathogenic mutations affecting epithelial-to-mesenchymal transition-related genes,

Correspondence: Dr. Jorge S. Reis-Filho, MD PhD FRCPath, Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA. reisfilj@mskcc.org; Dr. Larry Norton, MD, Department of Medicine, Memorial Sloan Kettering Cancer Center, 300 East 66th St, New York, NY 10065, USA. nortonl@mskcc.org; Dr. Francois-Clement Bidard, MD PhD, Department of Medical Oncology, Institut Curie, 26 rue d'Ulm, Paris, France, USA. francois-clement.bidard@curie.fr.
*Equally contributed

including *SMAD4*, *TCF7L2* and *TCF4* (*ITF2*), were found to be restricted to or enriched in the metastatic lesions. Mutational signatures of trunk mutations differed from those of mutations enriched in the primary tumor or the metastasis in six cases.

Conclusion—Synchronous primary breast cancers and metastases differ in their repertoire of somatic genetic alterations even in the absence of systemic therapy. Mutational signature shifts might contribute to spatial intra-tumor genetic heterogeneity.

Keywords

Massively parallel sequencing; metastatic breast cancer; somatic mutations; epithelial-mesenchymal transition

INTRODUCTION

Although metastatic breast cancer remains incurable, the survival of patients with advanced disease has shown significant but incremental increases over the last three decades (1). Metastatic dissemination is a complex process and numerous molecular mechanisms have been described, including models where metastatic potential is an intrinsic characteristic of all cancer cells and the emergence of metastasis is a stochastic event (2), whereas others suggested that only subpopulations of cancer cells would be able to colonize other organs (3).

Breast cancers display intra-tumor genetic heterogeneity that vary between cases and may follow spatial patterns (4–7). Moreover, primary and metastatic lesions exhibit phenotypic differences, including changes in estrogen receptor (ER) and HER2 status in 7%–25% of cases (8, 9). Whole-genome analyses in pairs of breast cancers and metachronous metastases after adjuvant treatment revealed genetic differences in their clonal composition, mutational frequencies and structural alterations, suggesting that metastatic disease may emerge from a subclone rather than the dominant clone of the primary tumor (4). Furthermore, analyses of a primary breast cancer and its metastatic lesions revealed that resistance to a PI3K-alpha inhibitor was a convergent phenotype, with resistant metastatic deposits harboring different *PTEN* mutations (10). By contrast, targeted massively parallel sequencing analyses of hotspot mutations or cancer genes demonstrated limited differences between primary tumors and their respective metastases (11–13), and suggested that genotyping the primary tumor may be acceptable to guide systemic treatment if the metastatic sample is not obtainable (12). Crucially, whether the differences are due to the selective pressures imposed by the metastatic process itself, the systemic therapies or both, remains unclear.

Approximately 6%–10% of breast cancers present with metastatic disease at diagnosis (1). Analysis of primary and metastatic lesions from these patients provides a unique opportunity to investigate potential differences in their repertoire of somatic genetic alterations without the effects of the selective pressures imposed by systemic therapies. This is of clinical importance, given that if the driver genetic alterations present in a given primary tumor are distinct from those of its metastatic deposit even without systemic therapy, it would impact on how precision medicine can be implemented and on the selection of samples for genetic analyses (14).

In this study, we subjected up to two anatomically distinct biopsies of primary breast cancers and up to two anatomically distinct biopsies of their synchronous metastatic deposits from patients who presented with *de novo* metastatic disease and had not received any local or systemic therapy to gene copy number analysis and high-depth whole-exome sequencing (WES). All somatic mutations identified by WES were subsequently validated using orthogonal methods. Our aims were to determine the differences in the repertoires of somatic mutations, copy number alterations (CNAs) and mutational signatures (15) between primary breast cancers and their metastases, and to define whether clonal compositions would change from the primary tumor to its metastatic deposit in the absence of any systemic therapy.

MATERIALS AND METHODS

Tissue sample collection

Diagnostic biopsies of primary tumors and synchronous distant metastases were collected from nine patients with stage IV breast cancer at presentation, before any systemic treatment, as part of the routine clinical diagnostic workup. These biopsies were routinely processed, formalin-fixed and paraffin-embedded (FFPE). Second anatomically distinct biopsies from primary tumors and metastases from each patient were collected and flash-frozen as part of the prospective study Changes in Phenotype and Genotype of Breast Cancers During the Metastatic Process and Optimization of Therapeutic Targeting (ESOPE, NCT01956552) at Institut Curie (Paris, France; Supplementary Table S1). All patients gave informed written consent, allowing the collection, storage and genomic analysis of the biopsies. Peripheral blood leukocytes collection for germline DNA sequencing was authorized by a further informed written consent. This study was approved by the Institutional Review Boards of the authors' institutions.

Frozen biopsies from the primary tumor and synchronous metastasis were available for all cases. FFPE biopsies from both the primary and same distant metastatic lesions were also available for Cases 1, 5 and 7–9. For Case 4, a FFPE biopsy was only available from the primary tumor, whereas for Case 6, a FFPE biopsy was only available from the same metastasis. For Cases 2 and 3, no FFPE biopsies were available (Supplementary Table S2). Previous genomics analyses of case 5 were reported elsewhere (16) (Supplementary Methods).

Immunohistochemistry

Immunohistochemistry using antibodies against ER, progesterone receptor (PR) and HER2 was performed as previously described (5) and evaluated according to the American Society of Clinical Oncology/College of American Pathologists guidelines (17, 18) (Supplementary Methods).

Microdissection and DNA extraction

Fifteen representative eight-μm-thick histologic sections of the frozen and FFPE biopsies of the primary tumors and metastases were subjected to microdissection and DNA extraction as previously described (19) (Supplementary Methods).

WES of frozen biopsies

DNA extracted from frozen biopsies of the primary tumors, metastases and germline were subjected to exome capture using SureSelect Human All Exon v4 (Agilent) and to massively parallel sequencing on an Illumina HiSeq 2000 following validated protocols (19) (Supplementary Table S2, Supplementary Methods). WES data have been deposited in the Sequence Read Archive (SRP055001).

WES data processing was performed as previously described (19). Somatic single nucleotide variants (SNVs) were identified using MuTect (v1.0) (20) and somatic small insertions and deletions (indels) were identified using GATK (v2.7.4) (21) and the micro-assembly-based Scalpel (v0.1.1) (22) (Supplementary Methods).

Validation of mutations in frozen samples and discovery in FFPE samples

Orthogonal validation of mutations found by WES in the frozen samples and mutation discovery in the FFPE samples were performed by either targeted capture sequencing using a customized bait set (EzCap, Nimblegen, Roche) on an Illumina HiSeq 2000 or amplicon sequencing using a custom AmpliSeq panel on an Ion Torrent Personal Genome Machine (Supplementary Tables S2, Supplementary Methods).

Gene copy number profiling

FACETS (23) was used to define CNAs for samples subjected to WES. DNA extracted from microdissected FFPE and selected frozen biopsies were subjected to copy number profiling using the OncoScan v3 molecular inversion probe array (Affymetrix) following manufacturer's instructions (16) (Supplementary Table S2, Supplementary Methods).

Segmented Log_2 ratios from FACETS (WES) and Nexus Express for OncoScan (OncoScan) were used as input for ABSOLUTE (v1.0.6) (24) to determine integer copy number and cancer cell fractions (CCFs) of CNAs. CNAs were defined using the modal copy number from ABSOLUTE (Supplementary Methods). Regions of loss of heterozygosity (LOH) were defined using FACETS (WES) and Nexus Express for OncoScan (OncoScan). For six sample with both OncoScan and FACETS results, a substantial to perfect agreement for the CNA profiles was observed (median Cohen's weighted kappa 0.85, range 0.76–0.88; Supplementary Fig. S1A, Supplementary Table S2).

Identification of likely pathogenic mutations

A combination of MutationTaster (25), CHASM (breast) (26) and FATHMM (27) was used to define the potential functional effect of each missense SNV. In-frame indels defined as "neutral" by MutationTaster (25) and PROVEAN (28) were considered likely passengers. The remaining in-frame indels, as well as frameshift, splice-site and nonsense mutations were considered likely pathogenic if they were targeted by loss of the wild-type allele (i.e. LOH) or affected haploinsufficient genes (29). SNVs affecting hotspot residues (30) were considered likely pathogenic. Mutations affecting cancer genes (31–33) were annotated. Mutations that were neither likely pathogenic nor likely passenger were considered of indeterminate pathogenicity (Supplementary Methods).

Classification of trunk and branch mutations, and mutations enriched in the primary or metastatic lesion

The CCF of each validated mutation was inferred using ABSOLUTE (24) (Supplementary Methods). A mutation was considered 'trunk' if it was clonal in all available biopsies in any given patient. We defined mutations 'specific to the metastatic lesion' as those present in at least one biopsy of the metastasis but absent from all biopsies of the primary tumor from the same patient and defined mutations 'enriched in the metastatic lesion' as those associated with an increase in CCF by at least 20% in the metastasis compared to the primary tumor, and *vice versa* for mutations 'specific to the primary tumor' and 'enriched in the primary tumor'.

Analysis of pathways associated with the metastatic process

To identify pathways that may be associated with the metastatic process, we analyzed the genes affected by likely pathogenic mutations specific to, enriched in the metastatic lesion (see above) and those associated with LOH in at least one biopsy of the metastasis and not associated with LOH in any biopsy of the matched primary tumor using Ingenuity Pathway Analysis (http://www.ingenuity.com) and g:Profiler (34) (Supplementary Methods).

Mutational signatures

Mutational signatures were defined as previously described (35), based on the Pearson's correlation coefficient to the 12 breast cancer-associated mutational signatures (15, 36) (Supplementary Methods).

Phylogenetic tree construction

Maximum parsimony trees were built based on the repertoire of somatic mutations, gene amplifications and homozygous deletions as previously described (37) (Supplementary Methods).

Statistical analysis

All statistical analyses were performed in R v3.1.2. Comparisons of continuous and categorical variables were performed using Mann-Whitney U and Fisher's exact tests, respectively. Associations were performed using the Spearman's rank correlation tests. Agreement of CNAs was assessed using Cohen's weighted kappa statistic. All statistical tests were two-tailed and $P<0.05$ was considered statistically significant.

RESULTS

Temporal and spatial genetic heterogeneity in primary breast cancers and synchronous metastatic deposits

Nine treatment-naïve breast cancer patients with stage IV metastatic disease at presentation from the prospective ESOPE study (NCT01956552) were included in this study (Supplementary Table S1). Cases 1–8 were invasive ductal carcinomas of no special type and Case 9 a metaplastic breast carcinoma. Five patients had metastases in multiple anatomical sites and four patients had a single metastasis (Supplementary Table S1). Across

all patients, liver (n=6) and bone (n=5) were the most common metastatic sites. Breast cancers of all clinical ER/HER2 subtypes were included and ER/HER2 status was concordant in primary tumors and their respective metastases in all patients tested. None of the patients received local or systemic therapy prior to the synchronous collection of diagnostic needle biopsies from the primary and metastatic deposits.

We first sought to determine whether the repertoire of somatic genetic alterations in biopsies of primary tumors and their respective metastases in treatment-naïve patients would differ. DNA extracted from two frozen biopsies, one from the primary tumor and one from a distant metastatic lesion (liver, n=5; bone, n=2; skin, n=1, contralateral axillary lymph node, n=1) from these nine patients was subjected to WES to median depths of 202× (range 189x–229x) in primary tumors, 207× (range 124x–277x) in metastases and 208× (range 58x–267x) in matched normal counterparts (Fig. 1A, Supplementary Table S2). Additionally, to define the spatial heterogeneity within the primary and metastatic lesions, we obtained a second, independent FFPE biopsy from the primary tumor of six cases (Cases 1, 4, 5, 7–9), and an independent FFPE biopsy from the same metastatic lesion in six cases (Cases 1, 5–9, Fig. 1A, Supplementary Table S2). We subjected the DNA obtained from the FFPE biopsies of these seven patients to a combination of target capture sequencing and/or amplicon re-sequencing to assess the presence of the somatic mutations found in the WES analysis of either or both frozen biopsies of the respective patients. WES analysis of the frozen biopsies of the primary and the metastatic deposits revealed the presence of 1,041 and 1,115 somatic mutations, all of which were subjected to validation using either targeted sequencing or amplicon re-sequencing and 889 (85%) and 1,049 (94%) of the mutations of the primary tumors and the metastatic deposits, respectively, were validated (Supplementary Methods, Supplementary Tables S2–S3) and were included in subsequent analyses.

Restricting the analyses to the two frozen biopsies of each patient revealed significantly more somatic mutations in the metastatic lesions (median 110, range 31–239) than in the primary tumors (median 78, range 31–204, P=0.0209, paired Mann-Whitney U test, Fig. 1A). When we considered the targeted sequencing of the additional second biopsies of the six primary tumors and the six metastatic lesions, we identified a median of one (range 0–23) and 0 (range 0–2) additional mutations in the primary and metastatic lesions, respectively (Fig. 1A).

Akin to a clinical scenario, where single diagnostic biopsies are employed for sequencing analysis, we compared the repertoire of mutations detected in the single frozen biopsies of the primary and metastatic lesions from each patient, which revealed a median of 60% (range 6%–95%) of somatic mutations in common (Fig. 1B). Importantly, however, substantial heterogeneity was observed, with a median of 12% (0%–44%) and a median of 24% (3%–52%) of mutations restricted to the primary and the metastatic lesions, respectively (Fig. 1B).

We next sought to define if including an additional anatomically distinct biopsy of the primary tumor and of its metastatic lesion would mitigate the apparent heterogeneity observed between the lesions. In the five cases with two biopsies of both the primary and metastatic lesions, if a single biopsy of the primary tumor and the metastasis were

employed, a median of 22% (3%–50%) of mutations would be restricted to the metastatic lesion. Including the additional anatomically distinct biopsies of the two lesions slightly reduced this proportion to 21% (3%–47%, Fig. 1B), suggesting that heterogeneity between primary tumor and metastases can be only marginally mitigated by the inclusion of second anatomically distinct biopsies.

Distinct repertoire of genetic alterations in primary breast cancers and synchronous metastatic deposits

Previous studies based on the analysis of hotspot mutations or cancer genes have suggested limited genetic differences between primary tumors and their metastases (11–13). These observations suggest that the genetic heterogeneity observed between primary tumors and their synchronous metastases would constitute genetic drift (38) and therefore the heterogeneity would preferentially affect passenger alterations (6). Since having second anatomically distinct biopsies only marginally mitigated the mutational differences, we focused on the frozen biopsies of the synchronous primary breast cancer and the metastases to compare their repertoires of somatic mutations. We identified a median of seven (range 3–21) likely pathogenic mutations per case (primary and/or metastasis), including a median of three (range 0–5) likely pathogenic mutations in cancer genes (31–33) (Fig. 2A, Supplementary Fig. S2A). Six cases each harbored one hotspot mutation, all of them shared between primary and metastasis, namely *PIK3CA* H1047R (Cases 2 and 4), *TP53* R248W/R282W/R196* (Cases 3, 5 and 7, respectively) and *NFE2L2* R34G (Case 9, Supplementary Table S3).

Consistent with the observations that driver mutations tend to be clonal (i.e. present in virtually 100% of tumor cells) (4, 37), 24/28 (86%) of the likely pathogenic mutations affecting cancer genes were found in the frozen biopsies of both the primary and their synchronous metastases and 22/28 (79%) were clonal in both (Fig. 2B). This is in agreement with previous analyses restricted to hotspot mutations and/or mutations affecting known cancer genes (11–13). Indeed, all *TP53* (n=7), *PIK3CA* (n=2) and *GATA3* (n=1) mutations were clonal in both biopsies (Fig. 2A). We also identified clonal likely pathogenic mutations affecting the cancer genes *RAD21*, *BRCA2*, *ERBB3*, *RB1* and *ATR* in single primary-metastasis pairs. In the one case without likely pathogenic mutations affecting cancer genes (Case 1), we identified clonal likely pathogenic mutations affecting *RHOB* and *FLT1,* as well as focal amplifications in 8q24.21 (encompassing *MYC*) and 17q12 (*ERBB2*) in both lesions (Fig. 2A, Supplementary Fig. S1B, Supplementary Table S3).

Despite limited heterogeneity in the repertoire of hotspot mutations or mutations affecting known cancer genes between primary tumors and their respective metastases, in four cases, likely pathogenic mutations affecting cancer genes were found to be restricted to the metastatic lesions, namely, *TCF7L2* S122* in Case 5, *SMAD4* D355G and *PSIP1* I412fs, both associated with LOH, in Case 7, and *FLT4* A835T in Case 8 (Fig. 2A). Notably, these four mutations were not identified in the primary tumors even with a second, anatomically distinct biopsy (Supplementary Fig. S2A).

In agreement with the hypothesis that heterogeneity would preferentially affect passenger alterations (6), mutations of indeterminate pathogenicity, likely passenger and synonymous

mutations were less frequently present in both lesions than likely pathogenic mutations in cancer genes (53%, 56% and 57% vs 86%, and P=0.0007, P=0.0021 and P=0.0023, respectively, Fisher's exact tests, Fig. 2B). Likely pathogenic mutations (affecting any gene) were also more frequently concurrently present or clonal in the frozen biopsies of both lesions than indeterminate/likely passenger/synonymous mutations (75% vs 55% and 66% vs 43%, P=0.0002 and P<0.0001, respectively, Fisher's exact tests, Fig. 2B). Similar results were obtained when the additional FFPE biopsies of the primary tumor and metastases were included, with likely pathogenic mutations (affecting any gene) more frequently present or clonal in all available biopsies compared to indeterminate/likely passenger/synonymous mutations (73% vs 53% and 49% vs 37%, P=0.0003 and P=0.0299, respectively, Fisher's exact tests, Supplementary Fig. S2B).

Copy number analysis revealed that 62% (29/47) of focal amplifications and homozygous deletions were found in both lesions, including amplification affecting *MYC* (Cases 1, 5 and 7), *ERBB2* (Cases 1, 2, 3, 6 and 7), *EGFR* (Case 2), *AKT1* (Case 8), *FGFR1* (Case 7) and *MAP2K3* (Cases 3 and 7) (Fig. 3A, Supplementary Fig. S1B, Supplementary Table S3). Important differences between the primary and the metastasis were also observed; in Case 7, 53% (10/19) and 11% (2/19) of amplifications were restricted to the primary and the metastatic lesions, respectively, and in Case 9, one of two amplifications was restricted to the metastasis (Fig. 3B). Notably, five homozygous deletions were found exclusively in the metastases of Cases 5, 7 and 9 (Fig. 3B and Supplementary Table S3), including 17p12 (encompassing *MAP2K4*) and 19q13.42 (*KIR3DL3*) in Case 5, 17p13.1 (*TNFSF12* and *SENP3*) and 20p13 (*SOX12*) in Case 7 and 9q34.3 (*MAMDC4*) in Case 9, none of which could be detected in the second biopsies of the primary tumors of the respective patients.

Taken together, these results suggest that although the majority of likely pathogenic mutations affecting any (75%) or cancer genes (86%, including all hotspot mutations), as well as 62% of amplifications and homozygous deletions, were present in the frozen biopsies of primary tumors and their synchronous metastases in treatment-naïve breast cancer patients, potential driver mutations (14% of the likely pathogenic mutations affecting cancer genes) were restricted to the metastatic deposit and absent in up to two biopsies of the primary tumor. These results demonstrate that heterogeneity between primary breast cancers and their synchronous metastases in the absence of systemic therapy preferentially, but not exclusively, affected passenger genetic alterations.

Likely pathogenic mutations affecting epithelial-to-mesenchymal transition (EMT)-related genes are enriched for in metastases

Although the heterogeneity between primary tumors and their metastases preferentially affected passenger genetic alterations, we found likely pathogenic mutations in cancer genes *FLT4, PSIP1, SMAD4* and *TCF7L2* restricted to the metastases. We therefore sought to investigate whether the metastatic process was associated with specific biological pathways. Pathway analysis on 22 genes affected by likely pathogenic mutations specific to or enriched in (i.e. increase in CCF by at least 20% compared to the primary lesions (frozen and FFPE whenever available)) or associated with LOH in the metastatic lesions but not the primary lesions (Fig. 4A) revealed an enrichment in genes associated with the regulation of EMT

using Ingenuity Pathway Analysis (P<0.0001, Fig. 4B, Supplementary Table S4). These EMT-associated mutations were identified in four cases, with *JAK3* E1033D mutation associated with LOH in the metastasis but not the primary tumor in Case 4, *SMAD4* D355G mutation associated with LOH in Case 7, *TCF7L2* (Transcription factor 7-like 2, also called *TCF4* or T-cell factor 4, chr10q25.3) S122* mutation in Case 5, and *TCF4* (Transcriptional factor 4, also called *ITF2* or immunoglobulin transcription factor 2, chr18q21.1) splice site c.379–1G>C mutation associated with LOH in Case 9. We further refined the analysis to genes rarely mutated (<1%) in primary breast cancers (39) and of the remaining 21 genes, the enrichment remained significant (P<0.0001). Interrogating the KEGG and Reactome pathways using g:Profiler (34) showed that these 22 and 21 genes were significantly enriched in the KEGG adherens junction pathway (P=0.0001 and P=0.0141, respectively, Supplementary Table S4).

Evolutionary dynamics of somatic mutations and CNAs

Single cell sequencing studies suggested that whilst mutations are acquired gradually, CNAs develop in punctuated bursts of evolution (6, 7). We sought to explore whether somatic mutations and CNAs may display distinct evolutionary patterns using the integrated ABSOLUTE framework that simultaneously quantifies the CCFs of somatic mutations and CNAs (24). A median of 24% (range 4%–42%) and 16% (range 5%–28%) of the mutations in the frozen biopsies of the primary and the metastatic lesions, respectively, were subclonal (Fig. 5). By contrast, there was little intra-tumor CNA heterogeneity, with a median of 100% (range 89%–100%) and 97% (75%–100%) of CNAs (excluding amplifications, see Supplementary Methods) found to be clonal in the primary and the metastatic lesions, respectively. There was no difference or association between the proportion of subclonal mutations or of subclonal CNAs within primary-metastasis pairs (all P>0.05, paired Mann-Whitney U and Spearman's correlation tests). Moreover, the proportion of subclonal mutations and CNAs did not show correlation in either the primary or the metastatic lesions (both P>0.05, Spearman's correlation tests), suggesting that mutations and CNAs likely evolved independently. Crucially, we observed in all cases that the CCFs of somatic mutations formed a continuous spectrum whereas the vast majority of CNAs were found in all cancer cells. An analysis of the proportion of subclonal mutations and CNAs, alone or together, revealed no difference between the primary tumors stratified by ER or HER2 status, triple-negative phenotype, grade (2 vs 3), number of metastatic sites (1 vs >1), the presence of liver metastasis and the presence of bone metastasis, and between the metastatic lesions stratified by ER or HER2 status, triple-negative phenotype, and metastatic site biopsied (liver vs bone, all P>0.05, Mann-Whitney U tests). We further found no correlation between heterogeneity within the primary tumor and its volume or the volume of total metastases (both P>0.05, Spearman's correlation tests). Taken together, our results suggest that the evolutionary trajectories of somatic mutations and CNAs are distinct, with mutations accumulating in a gradualistic manner, whereas CNAs likely occurred in punctuated bursts of evolution.

Shifts in mutational signatures

We hypothesized that, akin to the switch from the smoking-associated mutational signature to that of APOBEC cytidine deaminase activity in lung cancers evolution (40), the

mutational processes that shaped breast cancer genomes may evolve over time. We therefore compared the mutational signatures that underpin i) trunk mutations, ii) mutations specific to or enriched in the primary tumors and iii) mutations specific to or enriched in the metastatic lesions to 12 mutational signatures previously reported in breast cancer (36) for all cases except Case 5, for which there were not sufficient non-trunk mutations to perform this analysis. In six cases, our analysis suggested shifts in the dominant mutational signatures during tumor evolution (Fig. 6A, Supplementary Fig. 3A). The trunk mutations displayed several different signatures, with signatures 1 (associated with aging), 2 and 13 (C>T transitions and C>G transversions, respectively, in the TpCpW context, associated with APOBEC cytidine deaminase activity) being the most frequent, suggesting that the biological processes driving trunk mutations may vary from patient to patient. Interestingly, we observed that mutational signatures were not constant throughout tumor evolution, with shifts in mutational signatures detected between trunk mutations and mutations specific to or enriched in the primary and/or metastatic lesions. Although no consistent pattern in the shifts between trunk and non-trunk mutations was detected, four of eight cases displayed a dominant signature 13 in the non-trunk mutations, with the APOBEC signature frequently found in both the mutations enriched in or specific to primary tumors (Cases 2, 4, 7 and 9) and the mutations enriched in or specific to the metastatic lesions (Cases 2, 4 and 9, Fig. 6B, Supplementary Fig. 3B). Notably, these four cases spanned the entire spectrum of clinical ER/HER2 phenotypes, suggesting the APOBEC signature is not restricted to specific subtypes of breast cancer.

DISCUSSION

As the biopsies for this study were *de facto* diagnostic core biopsies, our study addresses an important clinical question, namely whether sequencing analysis of primary tumors is representative of the metastatic deposits in the absence of systemic therapy. Consistent with previous studies (11, 12), we demonstrated that pathogenic mutations affecting known breast cancer genes were present in both the primary tumors and their respective metastases and that heterogeneity preferentially affected passenger alterations. Importantly, however, likely pathogenic mutations affecting the cancer genes *SMAD4*, *PSIP1*, *TCF7L2* and *FLT4* were found only in the synchronous metastases even with a second, anatomically distinct diagnostic biopsy of the respective primary tumors. In fact, the mutational differences observed between primary tumor and metastases were only marginally mitigated by including second, anatomically distinct diagnostic biopsies of the primary and of the metastatic lesions. Among the likely pathogenic mutations, only one (*ASB15* R528H in Case 7) was ‘rescued’ by a second biopsy of the primary tumor. Thus, the information extracted from a second biopsy of the primary tumor might not be sufficient to overcome the challenges posed by intra-tumor genetic heterogeneity.

Previous sequencing studies of cancer-associated genes revealed that *TP53*, *PTEN*, *KRAS* and *SMAD4* were among the genes most frequently affected by mutations restricted to the metastasis (11, 12, 41, 42). Although *TP53*, *PTEN* and *KRAS* mutations were not found to be restricted to or enriched in the metastases in this study, we have identified a likely pathogenic *SMAD4* mutation associated with LOH, along with likely pathogenic mutations affecting other EMT-related genes, such as *TCF7L2* and *TCF4* (*ITF2*), restricted to or

enriched in the metastases. EMT, a phenomenon by which epithelial cells acquire invasive and migratory properties, has been shown to be key to the metastatic process (43). SMAD4 loss has been implicated in the metastatic spread of breast cancer in preclinical models (44), and *SMAD4* mutations restricted to metastases have been reported in three breast cancers (11, 12) and in two colorectal cancers with synchronous liver metastasis (42). *TCF7L2* loss-of-function mutations have been shown to activate Wnt pathway (45), which plays pivotal roles in EMT, and knock-down of *TCF4 (ITF2)* deregulates TGF-β signaling, EMT and apoptosis (46). Our analyses also resulted in the identification of metastasis-specific secondary LOH events associated with likely pathogenic mutations affecting *JAK3* and *FLT4*. Alterations in both genes have previously been described as metastasis-specific, with a *JAK3* mutation restricted to the liver metastasis of a colorectal cancer (42) and a *FLT4* mutation restricted to the liver metastasis of a breast cancer (47). We did not observe specific patterns in terms of the number of private mutations in the metastases, likely pathogenic mutations enriched in the metastases or mutational signature in the five liver metastases or the two bone metastases, or any significant associations between the number of mutations and histopathologic parameters, such as histologic grade, tubule formation, nuclear pleomorphism and mitotic index (Supplementary Table S1).

The mutational processes that shape cancer genomes leave imprints as mutational signatures, many of which are linked to specific biological phenomena (15, 36). In our study, we found evidence that somatic mutations are acquired in a gradualistic manner and, akin to lung cancer (40), mutational processes are not constant in breast cancers throughout tumor evolution, even in the absence of selective pressures imposed by systemic therapy. The evolution of mutational signatures was observed across all ER/HER2 clinical subtypes and did not follow a specific pattern, however the C>G APOBEC-associated signature 13, which is frequent in breast cancer (36), was found to be enriched in the non-trunk mutations and likely contributed to the genetic diversity of four of eight cancers studied. By contrast, and consistent with observations derived from single-cell sequencing analysis (6, 7), most CNAs were found to be present in all cancer cells, suggesting that these were acquired in punctuated bursts of evolution. Although the number of mutations for the analyses of mutational signatures was relatively small in some sets of mutations, our findings support the notion that the evolution of mutational signatures parallels that of the somatic mutations over time in several cases, and may suggest that, in breast cancer, the C>G APOBEC mutational process is linked to tumor evolution rather than early tumor development, consistent with the findings reported by Lefebvre et al (48), who observed that approximately 59% of the mutations found in hormone receptor-positive/HER2–negative metastatic tumors were consistent with signatures 2 and 13, whereas only 32% of the mutations detected in hormone receptor-positive/HER2–negative primary breast cancers displayed this pattern. Furthermore, the lack of correlation between the extent of subclonal mutations and CNAs supports the hypothesis that somatic mutations and CNAs evolve independently (6, 7).

Our study has important limitations. First, we studied patients with stage IV breast cancers, which account for a minority of breast cancers in the developed world (1), and it has been hypothesized that these breast cancers may have a different biology (3). Notably, three of the patients here had an apparent history of neglected breast cancer, but the findings from these

cases were consistent with those of the remaining patients. Second, none of the cases analyzed displayed a change in ER/HER2 status from the primary tumor to the metastatic lesion, as reported in a subset of cases (8, 9), and the genetic heterogeneity between paired primary and metastatic tumors with discrepant ER/HER2 status are likely to be more conspicuous. Third, targeted sequencing panels were performed in the DNA extracted from the FFPE biopsies; it is plausible that WES would reveal more profound intra-tumor heterogeneity. Fourth, although the presence of mutations restricted to the metastatic lesions does not prove that these mutations play a causative role in the development of metastasis and no genetic alteration was identified as a common denominator of breast cancer metastasis, our findings are consistent with the notion that the metastatic ability of cancer cells likely constitutes a convergent phenotype (49).

Here we report that primary breast cancers and their distant metastatic outgrowths differ in their repertoire of somatic genetic alterations in therapy-naïve patients. Our results suggest that mutational signature shifts may contribute to intra-tumor genetic heterogeneity and that metastatic spread in breast cancer is coupled to an enrichment of mutations in EMT-related genes. As most *bona fide* and currently actionable drivers of the disease are truncal, sequencing of the primary lesions alone even in therapy-naïve metastatic breast cancer patients may be clinically acceptable until additional alterations or specific pathways become actionable. In fact, therapeutic agents targeting the Wnt pathway, which plays pivotal roles in EMT, have entered clinical trials (50). WES of metastases from these patients may enable the discovery of new late drivers of breast cancer and of the metastatic process in the research setting. Given the distinct repertoire of genetic alterations in synchronous primary breast cancers and metastases in therapy-naïve patients, caution should be exercised with the use of agents targeted to mutations identified in a primary tumor to treat metastatic disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: The ESOPE (NCT01956552) study was funded by the INCa - French Ministry of Health (PHRC 2009 AOM 09 267). F.-C. Bidard was supported by a grant from the Nuovo-Soldati Foundation for Cancer Research and Institut Curie SiRIC (INCa-DGOS N° 4654). S. Piscuoglio is funded by Swiss National Science Foundation (Ambizione grant number PZ00P3_168165). J. S. Reis-Filho is funded in part by the Breast Cancer Research Foundation (BCRF). Research reported in this publication was supported in part by a Cancer Center Support Grant of the National Institutes of Health/National Cancer Institute (Grant No. P30CA008748). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors would like to thank Odette Mariani for her assistance at the Institut Curie Tumor Biobank.

Abbreviations

CCF	cancer cell fraction
CNA	copy number alteration

EMT	epithelial-to-mesenchymal transition
ER	estrogen receptor
FFPE	formalin-fixed paraffin-embedded
Indel	small insertion and deletion
LOH	loss of heterozygosity
PR	progesterone receptor
SNV	single nucleotide variant
WES	whole-exome sequencing

References

1. Andre F, Slimane K, Bachelot T, Dunant A, Namer M, Barrelier A, et al. Breast cancer with synchronous metastases: trends in survival during a 14-year period. J Clin Oncol. 2004; 22:3302–8. [PubMed: 15310773]
2. Bernards R, Weinberg RA. A progression puzzle. Nature. 2002; 418:823. [PubMed: 12192390]
3. Klein CA. Selection and adaptation during metastatic cancer progression. Nature. 2013; 501:365–72. [PubMed: 24048069]
4. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. Nat Med. 2015; 21:751–9. [PubMed: 26099045]
5. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci Transl Med. 2015; 7:302ra133.
6. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature. 2014; 512:155–60. [PubMed: 25079324]
7. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. Nat Genet. 2016; 48:1119–30. [PubMed: 27526321]
8. Amir E, Miller N, Geddie W, Freedman O, Kassam F, Simmons C, et al. Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. J Clin Oncol. 2012; 30:587–92. [PubMed: 22124102]
9. Niikura N, Liu J, Hayashi N, Mittendorf EA, Gong Y, Palla SL, et al. Loss of human epidermal growth factor receptor 2 (HER2) expression in metastatic sites of HER2-overexpressing primary breast tumors. J Clin Oncol. 2012; 30:593–9. [PubMed: 22124109]
10. Juric D, Castel P, Griffith M, Griffith OL, Won HH, Ellis H, et al. Convergent loss of PTEN leads to clinical resistance to a PI(3)Kalpha inhibitor. Nature. 2015; 518:240–4. [PubMed: 25409150]
11. Goswami RS, Patel KP, Singh RR, Meric-Bernstam F, Kopetz ES, Subbiah V, et al. Hotspot mutation panel testing reveals clonal evolution in a study of 265 paired primary and metastatic tumors. Clin Cancer Res. 2015; 21:2644–51. [PubMed: 25695693]
12. Bertucci F, Finetti P, Guille A, Adelaide J, Garnier S, Carbuccia N, et al. Comparative genomic analysis of primary tumors and metastases in breast cancer. Oncotarget. 2016; 7:27208–19. [PubMed: 27028851]
13. Meric-Bernstam F, Frampton GM, Ferrer-Lozano J, Yelensky R, Perez-Fidalgo JA, Wang Y, et al. Concordance of genomic alterations between primary and recurrent breast cancer. Mol Cancer Ther. 2014; 13:1382–9. [PubMed: 24608573]
14. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. Nature. 2013; 501:355–64. [PubMed: 24048068]
15. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–21. [PubMed: 23945592]

16. Bidard FC, Ng CK, Cottu P, Piscuoglio S, Escalup L, Sakr RA, et al. Response to dual HER2 blockade in a patient with HER3-mutant metastatic breast cancer. Ann Oncol. 2015; 26:1704–9. [PubMed: 25953157]
17. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol. 2010; 28:2784–95. [PubMed: 20404251]
18. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. 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. 2013; 31:3997–4013. [PubMed: 24101045]
19. Weinreb I, Piscuoglio S, Martelotto LG, Waggott D, Ng CK, Perez-Ordonez B, et al. Hotspot activating PRKD1 somatic mutations in polymorphous low-grade adenocarcinomas of the salivary glands. Nat Genet. 2014; 46:1166–9. [PubMed: 25240283]
20. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013; 31:213–9. [PubMed: 23396013]
21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–303. [PubMed: 20644199]
22. Narzisi G, O'Rawe JA, Iossifov I, Fang H, Lee YH, Wang Z, et al. Accurate de novo and transmitted indel detection in exome-capture data using microassembly. Nat Methods. 2014; 11:1033–6. [PubMed: 25128977]
23. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res. 2016; 44:e131. [PubMed: 27270079]
24. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012; 30:413–21. [PubMed: 22544022]
25. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010; 7:575–6. [PubMed: 20676075]
26. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. Cancer Res. 2009; 69:6660–7. [PubMed: 19654296]
27. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat. 2013; 34:57–65. [PubMed: 23033316]
28. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 2012; 7:e46688. [PubMed: 23056405]
29. Dang VT, Kassahn KS, Marcos AE, Ragan MA. Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. Eur J Hum Genet. 2008; 16:1350–7. [PubMed: 18523451]
30. Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol. 2016; 34:155–63. [PubMed: 26619011]
31. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer. 2004; 4:177–83. [PubMed: 14993899]
32. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013; 502:333–9. [PubMed: 24132290]
33. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014; 505:495–501. [PubMed: 24390350]
34. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, et al. g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016; 44:W83–9. [PubMed: 27098042]

35. Schultheis AM, Ng CK, De Filippo MR, Piscuoglio S, Macedo GS, Gatius S, et al. Massively Parallel Sequencing-Based Clonality Analysis of Synchronous Endometrioid Endometrial and Ovarian Carcinomas. J Natl Cancer Inst. 2016; 108:djv427. [PubMed: 26832770]
36. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016; 534:47–54. [PubMed: 27135926]
37. Murugaesu N, Wilson GA, Birkbak NJ, Watkins TB, McGranahan N, Kumar S, et al. Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. Cancer Discov. 2015; 5:821–31. [PubMed: 26003801]
38. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor evolution across cancer types. Nat Genet. 2016; 48:238–44. [PubMed: 26780609]
39. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]
40. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science. 2014; 346:251–6. [PubMed: 25301630]
41. Brannon AR, Vakiani E, Sylvester BE, Scott SN, McDermott G, Shah RH, et al. Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. Genome Biol. 2014; 15:454. [PubMed: 25164765]
42. Kovaleva V, Geissler AL, Lutz L, Fritsch R, Makowiec F, Wiesemann S, et al. Spatio-temporal mutation profiles of case-matched colorectal carcinomas and their metastases reveal unique de novo mutations in metachronous lung metastases by targeted next generation sequencing. Mol Cancer. 2016; 15:63. [PubMed: 27756406]
43. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. J Clin Invest. 2009; 119:1420–8. [PubMed: 19487818]
44. Deckers M, van Dinther M, Buijs J, Que I, Lowik C, van der Pluijm G, et al. The tumor suppressor Smad4 is required for transforming growth factor beta-induced epithelial to mesenchymal transition and bone metastasis of breast cancer cells. Cancer Res. 2006; 66:2202–9. [PubMed: 16489022]
45. Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. Nat Rev Cancer. 2013; 13:11–26. [PubMed: 23258168]
46. Forrest MP, Waite AJ, Martin-Rendon E, Blake DJ. Knockdown of human TCF4 affects multiple signaling pathways involved in cell survival, epithelial to mesenchymal transition and neuronal differentiation. PLoS One. 2013; 8:e73169. [PubMed: 24058414]
47. De Mattos-Arruda L, Weigelt B, Cortes J, Won HH, Ng CK, Nuciforo P, et al. Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. Ann Oncol. 2014; 25:1729–35. [PubMed: 25009010]
48. Lefebvre C, Bachelot T, Filleron T, Pedrero M, Campone M, Soria JC, et al. Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. PLoS Med. 2016; 13:e1002201. [PubMed: 28027327]
49. Weigelt B, Reis-Filho JS. Epistatic interactions and drug response. J Pathol. 2014; 232:255–63. [PubMed: 24105606]
50. Zhang X, Hao J. Development of anticancer agents targeting the Wnt/beta-catenin signaling. Am J Cancer Res. 2015; 5:2344–60. [PubMed: 26396911]

STATEMENT OF TRANSLATIONAL RELEVANCE

Despite the incremental increases in the survival of patients with metastatic disease over the past decades, metastatic breast cancer remains incurable. Analyses of primary breast cancers and their metastases after adjuvant treatment have revealed limited but important genetic differences. Whole-exome sequencing analysis of primary and metastatic breast cancers from therapy-naïve patients with stage IV disease, revealed that mutations affecting cancer genes were largely present and clonal in both primary tumors and their respective metastases. Importantly, however, genetic differences were observed between primary tumors and their metastases and were associated with shifts in mutational signatures and an enrichment for mutations affecting epithelial-to-mesenchymal transition (EMT)-related genes in the metastatic lesions. With the increasing number of actionable alterations, and of agents targeting components of the EMT pathway being tested in clinical trials, sequencing metastatic lesions in addition to primary tumors may help realize precision medicine in therapy-naïve metastatic breast cancer patients.

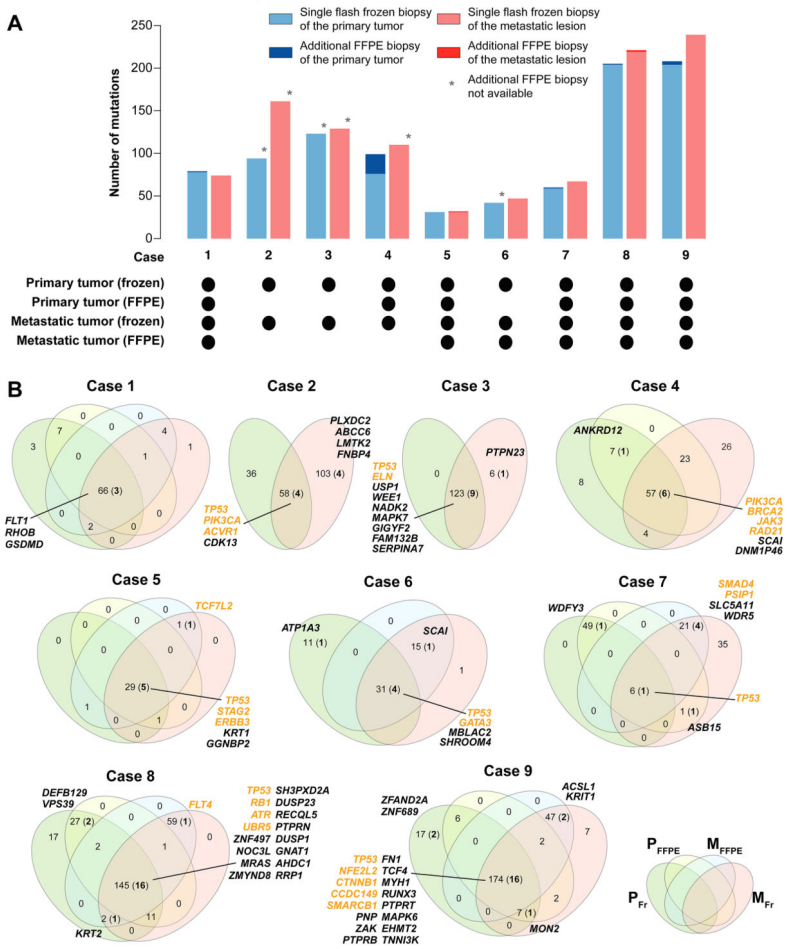

Fig. 1. Spatial and temporal heterogeneity in primary and metastatic lesions in treatment-naïve synchronous metastatic breast cancer

(A) Barplots depict the number of somatic mutations identified in single frozen biopsies of the primary tumors and metastatic lesions, and for seven patients, the additional mutations identified by analyzing an additional FFPE biopsy of the corresponding tumor. Available biopsy samples are indicated by black dots below the barplots. **(B)** Venn diagrams illustrate the number of somatic mutations and the number of likely pathogenic mutations (in bold) present in each of the biopsies. Genes affected by likely pathogenic mutations are listed, with those affecting cancer genes (31–33) labeled in orange. P_{Fr}: frozen biopsy of primary; P_{FFPE}: FFPE biopsy of primary; M_{Fr}: frozen biopsy of metastasis; M_{FFPE}: FFPE biopsy of metastasis. FFPE, formalin-fixed paraffin-embedded.

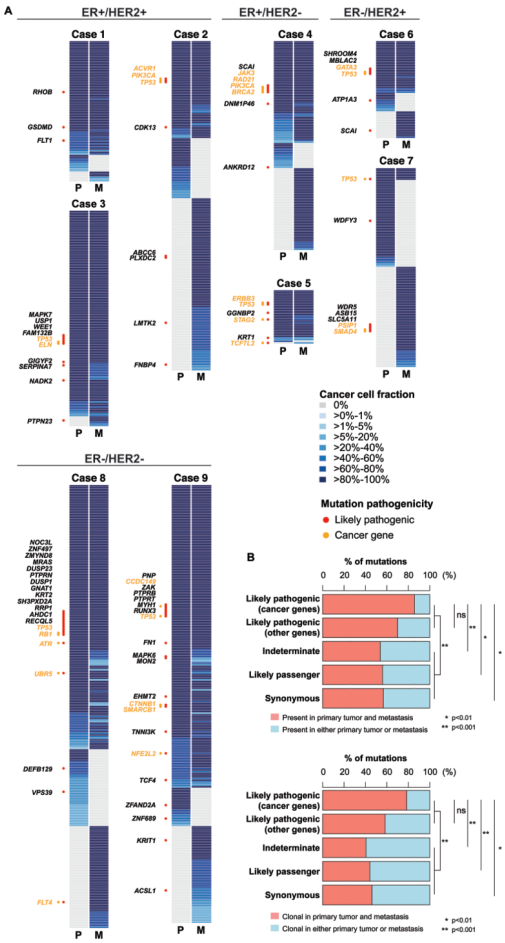

Fig. 2. Distinct repertoire of somatic genetic alterations in treatment-naïve primary breast cancers and synchronous metastatic deposits

(A) Heatmaps indicate the cancer cell fraction of somatic mutations as determined by ABSOLUTE (24) (blue, see color key) or their absence (grey) in each frozen biopsy. Likely pathogenic mutations are indicated by red dots and the affected genes are shown to the left of the heatmap. Likely pathogenic mutations affecting cancer genes (31–33) are further indicated by orange dots and gene names labeled in orange. Cases are grouped according to their ER/HER2 status. P: primary tumor; M: metastatic tumor. **(B)** Barplots of the distribution of mutations (top) present and (bottom) clonal in frozen biopsies of the paired primary tumors and the metastatic lesions, classified as likely pathogenic (cancer genes), likely pathogenic (other genes), of indeterminate pathogenicity, likely passenger and synonymous mutations from all patients. Comparisons between the groups of mutations of different pathogenicity were performed using Fisher's exact tests. *: $p < 0.01$, **: $p < 0.001$, ns: not significant.

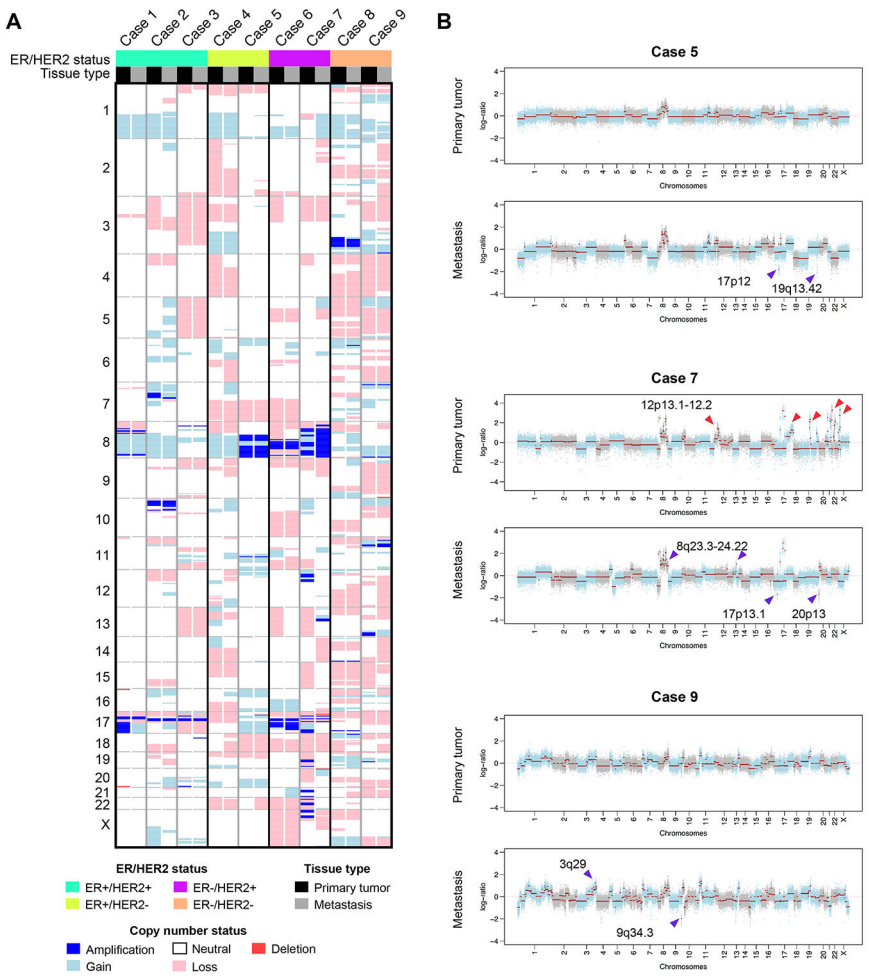

Fig. 3. Copy number alterations in treatment-naïve primary and synchronous metastatic lesions

(A) Heatmap illustrating the copy number alterations, where samples are presented on the X axis (columns) and chromosomal positions are presented on the Y axis (rows). ER/HER2 status of the tumor samples and sample types are indicated in the phenobar. Dark blue: amplification, light blue: copy number gain; white: neutral; light red: copy number loss; dark red: homozygous deletion. **(B)** Genome plots of the primary tumor and the distant metastatic lesion for cases 5, 7 and 9. In the genome plots, segmented Log_2 ratios (y-axis) were plotted according to their genomic positions (x-axis). Alternating blue and grey demarcate the chromosomes. The amplifications and homozygous deletions restricted to the primary tumors or the distant metastatic lesions are highlighted with red and purple arrows, respectively.

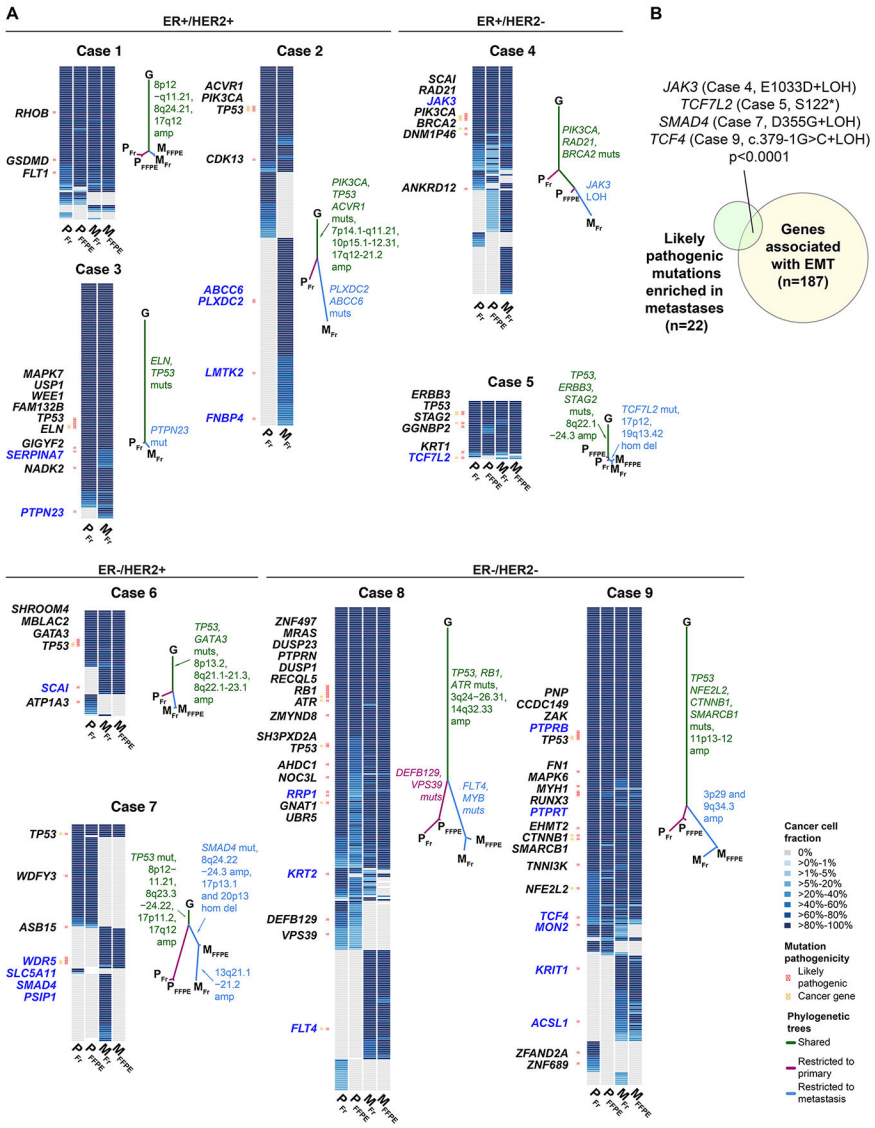

Fig. 4. Likely pathogenic mutations enriched in the metastases affect epithelial-mesenchymal transition-related genes

(A) Heatmaps indicate the cancer cell fractions of the somatic mutations as defined by ABSOLUTE (24) (blue boxes, see color key) or their absence (grey) in each biopsy. Red dots indicate likely pathogenic mutations and those affecting cancer genes (31–33) are indicated by orange dots. Genes affected by likely pathogenic mutations are labeled and those affected by likely pathogenic mutations specific to or enriched in the metastatic lesions, or were associated with the loss of the wild-type allele in the metastatic lesions but not the primary tumors labeled in blue. Phylogenetic trees depicting the evolution of the biopsies were constructed using the maximum parsimony method. The colored branches represent each of the subclones identified, and selected somatic genetic alterations that define a given clone are illustrated along the branches. The length of the branches is representative of the number of somatic mutations and copy number alterations (amplifications and homozygous deletions) that distinguishes a given clone from its

ancestral clone. P_{Fr}: frozen biopsy of primary; P_{FFPE}: FFPE biopsy of primary; M_{Fr}: frozen biopsy of metastasis; M_{FFPE}: FFPE biopsy of metastasis. **(B)** Venn diagram illustrates an enrichment of genes involved in EMT amongst the likely pathogenic mutations specific to, enriched in the distant metastasis or associated with the loss of the wild-type allele in the distant metastasis in the nine patients using Ingenuity Pathway Analysis. EMT, epithelial-mesenchymal transition.

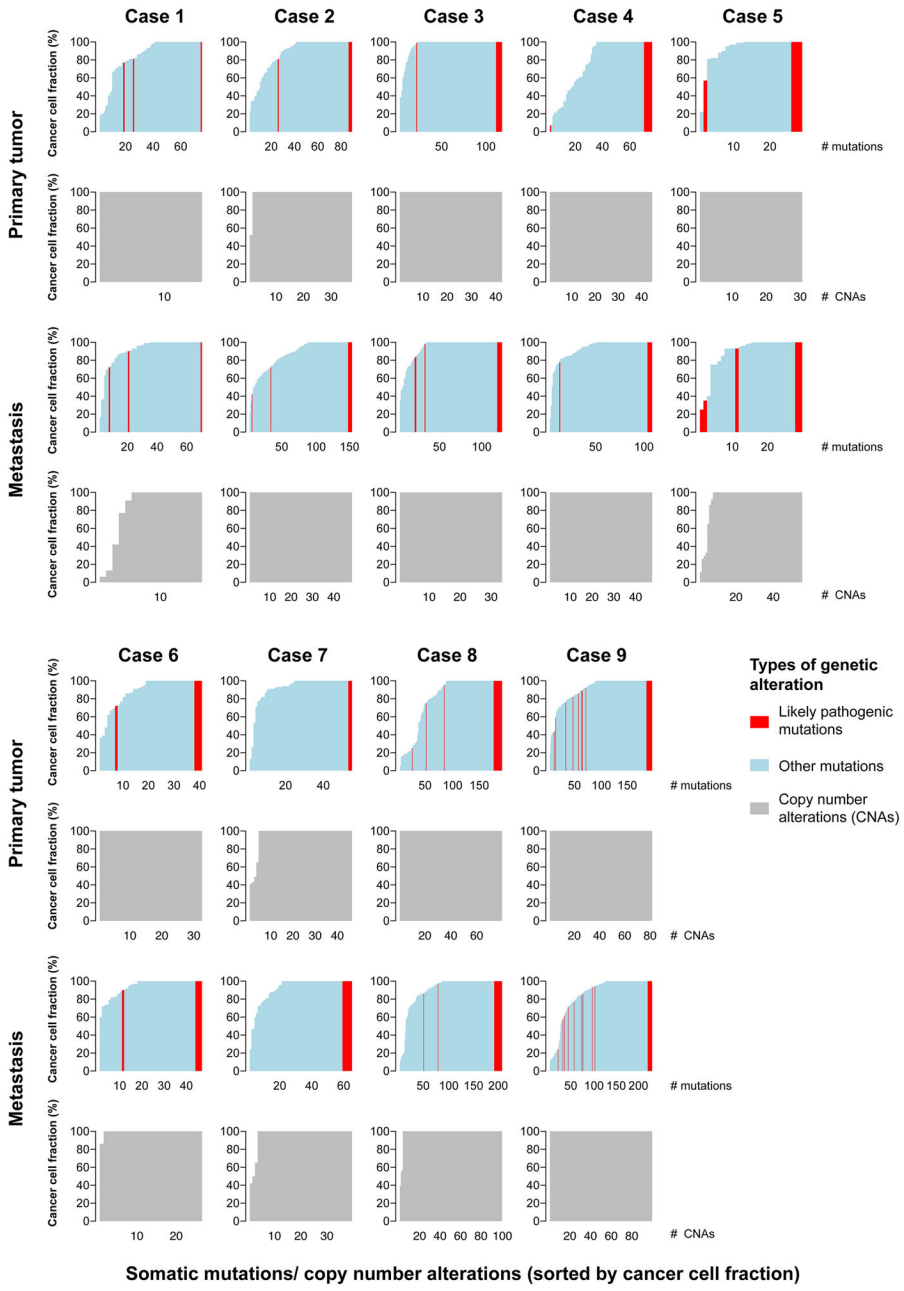

Fig. 5. Evolutionary dynamics of somatic mutations and copy number alterations

Barplots illustrate the cancer cell fractions of the somatic mutations (red and light blue bars) and of the copy number alterations (grey bars, excluding amplifications, see Supplementary Methods). Cancer cell fractions of the somatic mutations and the copy number alterations are sorted in increasing order. Bars for somatic mutations are color-coded based on the pathogenicity according to the color key.

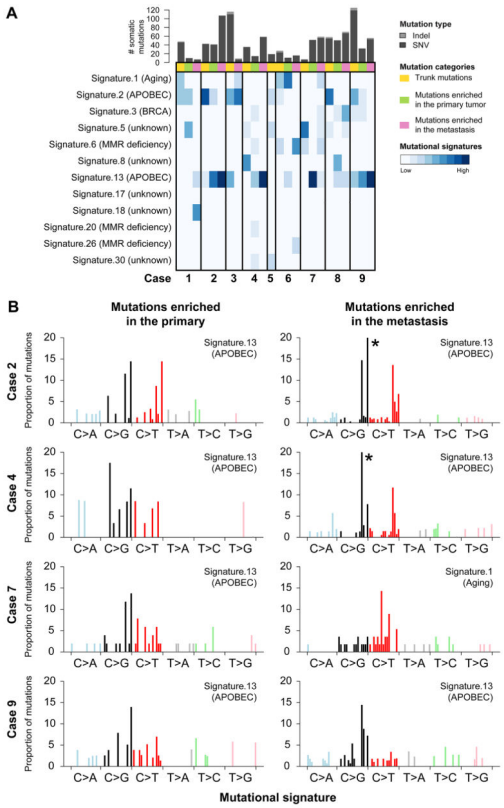

Fig. 6. Evolution of mutational signatures in treatment-naïve patients with *de novo* synchronous metastatic breast cancer

(A) Heatmaps depicting the mutational signatures that shaped the genomes (15, 36) of the tumor samples analyzed, separately as trunk mutations (yellow), mutations enriched in the primary tumor (green) and mutations enriched in the metastatic lesion (pink). The similarity of each mutational signature to the breast cancer-associated signatures (36) is indicated in blue according to the color key. **(B)** Barplots illustrating the mutational signatures of the mutations enriched in the primary tumor and the metastatic lesion of Cases 2, 4, 7 and 9. In each panel, the colored barplot illustrates each mutational signature according to the 96 substitution classification defined by the substitution classes (C>A, C>G, C>T, T>A, T>C and T>G bins) and the 5′ and 3′ sequence context, normalized using the observed trinucleotide frequency in the human exome to that in the human genome. The bars are ordered first by mutation class (C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G, T>G/A >C), then by the 5′ flanking base (A, C, G, T) and then by the 3′ flanking base (A, C, G, T). *: >20%.