Genomic landscape of metastatic breast cancer and its clinical implications

Lindsay Angus[1], **Marcel Smid**[1], **Saskia M. Wilting**[1], **Job van Riet**[1,2,3], **Arne van Hoeck**[4], **Luan Nguyen**[4], **Serena Nik-Zainal**[5], **Tessa G. Steenbruggen**[6], **Vivianne C.G. Tjan-Heijnen**[7], **Mariette Labots**[8], **Johanna M.G.H. van Riel**[9], **Haiko J. Bloemendal**[10,11], **Neeltje Steeghs**[6,11], **Martijn P. Lolkema**[1,11], **Emile E. Voest**[6,11], **Harmen J.G. van de Werken**[2,3], **Agnes Jager**[1], **Edwin Cuppen**[4,12], **Stefan Sleijfer**[1,11], **John W.M. Martens**[1,11,*]

[1]Department of Medical Oncology, Erasmus MC Cancer Institute, Erasmus University, Rotterdam, The Netherlands [2]Cancer Computational Biology Center, Erasmus MC Cancer Institute, Erasmus University Medical Center, Wytemaweg 80, 3015 GD, Rotterdam, The Netherlands [3]Department of Urology, Erasmus MC Cancer Institute, University Medical Center, Wytemaweg 80, 3015 GD, Rotterdam, the Netherlands. [4]Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht, The Netherlands [5]Department of Medical Genetics, The Clinical School, University of Cambridge, Cambridge, UK [6]Department of Medical Oncology, The Netherlands Cancer Institute, Antoni van Leeuwenhoek, Amsterdam, The Netherlands [7]Department of Medical Oncology, GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands [8]Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam, The Netherlands [9]Department of Internal Medicine, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands [10]Department of Medical Oncology, Meander Medical Center, Amersfoort, The Netherlands [11]Center for Personalized Cancer Treatment, Rotterdam, The Netherlands [12]Hartwig Medical Foundation, Amsterdam, The Netherlands

Abstract

Whole genome sequencing (WGS) of prospectively collected tissue biopsies of 442 metastatic breast cancer (mBC) patients reveals that, compared to primary BC, tumour mutational burden (TMB) doubled, relative contributions of mutational signatures shifted, and mutation frequency of

six known driver genes increased in mBC. Significant associations with pre-treatment were observed as well. The contribution of mutational signature 17 was significantly enriched in patients pre-treated with 5-FU, taxanes, platinum and/or eribulin, whereas the here identified de novo mutational signature I was significantly associated with pre-treatment containing platinum-based chemotherapy. Clinically relevant subgroups of tumours were identified exhibiting either homologous recombination deficiency (13%), high TMB (11%) or specific alterations (24%) linked to sensitivity to FDA-approved drugs. This study provides important novel insight into the biology of mBC and identifies clinically useful genomic features for future improvement of patient management.

Keywords

breast cancer; metastasis; whole genome sequencing

Breast cancer (BC) is the most common malignancy among women worldwide[1]. In-depth analyses of primary BC have provided clear evidence of clonal evolution and have resulted in the identification of a heterogeneous repertoire of nearly one hundred disease-causing genes as well as passenger events, both resulting from various underlying mutational processes[2–6] including age-related decay, homologous repair deficiency[7] and APOBEC mutagenesis[8,9].

However, patients do not die from their primary breast tumour but as a consequence of metastases and it is known that, due to tumour evolution and treatment pressure, the genomic alterations in metastatic BC (mBC) can differ substantially from the primary tumour[10–15]. Therefore, thorough genomic characterization of metastases will yield valuable insight into the active molecular processes in metastatic disease, which is crucial to understand the effects of systemic treatment on the tumour genome and ultimately to improve treatment of mBC patients.

To date, in-depth analyses of mBC lesions are limited to studies using either whole exome sequencing[16,17] in relatively small cohorts or targeted sequencing of cancer-associated genes in a larger cohort[18]. These studies have suggested that mBC largely carries the same drivers as seen in primary BC, however also manifest clear differences in the numbers and types of genes that are affected.

To obtain an unbiased and complete picture of the genomic landscape of mBC and its underlying mutational processes, as reflected by mutational signatures, we performed whole genome sequencing (WGS) on a large multicentre, prospective collection of snap frozen metastatic tissue biopsies from 442 breast cancer patients starting a new line of systemic treatment. These data enabled us to investigate the potential of clinical genomics, *i.e.* the drive to gain insight into patient-specific relevant (patterns of) aberrations for subsequent treatment choices. We performed an in-depth characterisation of the genomic landscape of these mBC patients and here report on the presence of genomic alterations, mutational and rearrangement signatures in comparison to a well-characterized cohort of primary BC (BASIS)[6]. Next, the available clinical data allowed us to associate genomic features with clinical information such as prior treatment. Finally, we identified subgroups of patients with

specific and targetable genomic features who might be eligible for established or experimental therapies.

Results

Metastatic biopsies and matched germline DNA (peripheral blood) of 625 patients with mBC were analysed (Fig. 1a). Patients with mBC who were biopsied in their primary tumour ($n = 55$) were excluded from the metastatic analyses, but were used as an additional control group. Metastatic biopsy sites mainly included liver, lymph node, bone and soft tissue (Fig. 1b). Twenty-two percent of all metastatic biopsies was non-evaluable, while lesions obtained from bone metastases had a failure rate of 33% (Supplementary Table S1). BC subtype distribution did not differ between non-evaluable and evaluable biopsies. Metastatic tumour biopsies and paired normal of the remaining 442 patients were sequenced at a median read coverage of 107 (IQR: 98 – 114) and 38 (IQR: 35 – 42), respectively.

The somatic landscape of mBC differs from primary BC

Metastatic lesions showed a median of 7,661 single nucleotide variants (SNVs, interquartile range (IQR): 4,607–14,417), 57 multiple nucleotide variants (MNVs, IQR: 32–106), 689 small insertions and deletions (InDels, IQR: 443–1,084), and 214 structural variants (SV, IQR: 99–392 (Supplementary Fig. 1). ER-negative tumours had a 1.6 fold higher SV count than ER-positive tumours (95% confidence interval (CI) 1.3-2.0, $P < 0.001$) and, HER2-positive tumours had higher SV counts than HER2-negative cases ($P = 0.013$).

Compared to WGS from 560 primary BC samples (BASIS cohort)[6], the median numbers of SNVs, InDels and SVs were significantly higher in mBC: 3,491 SNVs/MNVs (IQR: 2,075–6,911; 2.2x 95%CI 1.9-2.4, $P <$ 1e-5), 204 InDels (IQR: 133–365; 3.3x, 95%CI 3.0-3.6, $P <$ 1e-5), and 85 SVs (IQR: 25–208; 2.4x, 95%CI 2.1-2.8, $P <$ 1e-5). Consequently, the median tumour mutational burden (TMB) of 2.97 per million base pairs (Mb) (IQR 1.84 – 5.44) in mBC was significantly higher than that observed in the BASIS primary BC cohort (Supplementary Table S2) (1.29/Mb; IQR 0.78–2.56; 2.2x 95%CI 2.0–2.5, $P <$ 1e-5). In line with our finding, another cohort of mBC patients (Supplementary Table S2) also reported an elevated median TMB of 3.19/Mb[17]. In our mBC cohort, we did not observe differences in median TMB between BC subtypes or biopsy sites (Supplementary Fig. 2).

To ensure that the higher TMB we observed in our mBC cohort compared to primary disease was not due to methodological differences (Supplementary Table S2), we used the data of the 55 patients in our cohort that were biopsied in their primary tumour (Fig. 1a), including 31 treatment naïve patients (group 1) and 24 pre-treated patients (group 2). We compared the TMB of these primary tumours with the TMB of metastatic biopsies of 61 treatment naïve patients (group 3) and 369 pre-treated patients (group 4) (Supplementary Fig. 3). In a multivariate linear regression model using these four groups, both type of tissue (metastatic/primary) and pre-treatment (yes/no) were associated with TMB ($P <$ 1e-5 for the model, estimated coefficients were 0.3212 ($P =$ 0.02) and 0.3664 ($P =$ 0.001), respectively). After stratifying for ER-status both pre-treatment (0.4404, $P =$ 0.014) and type of tissue (0.5208, $P =$ 0.0003) remained associated with TMB in ER-positive cases, but not in ER-negative cases.

However, low numbers (only 8 pre-treated primary ER-negative tumours), render interpretation of this latter regression result inconclusive. This implicates that, next to the disease course, treatment pressure is a major contributor to TMB.

Mutational signatures are associated with pre-treatment

To investigate which mutational processes are operative in mBC and to what extent pre-treatment associates with the resulting mutational patterns, we applied the mathematical approach proposed by Alexandrov et al.[2] to categorize mutational signatures. De novo signature calling revealed the presence of 10 signatures in mBC, all of which could be mapped back to the already known Cosmic signatures (cosine similarities ranging from 0.79 to 0.99) (Fig. 2a; Supplementary Fig. 4). Except for de novo signatures I and J, all identified signatures have been previously described in primary BC.

De novo signature J was very similar to Cosmic mutational signature 7, which is likely due to ultraviolet (UV) exposure. Detailed evaluation showed the algorithm only identified this signature due to one patient with a very high contribution (>98%), suggesting that this liver biopsy, containing mostly UV-induced DNA damage, is misclassified as mBC.

De novo signature I (221 patients with >10% contribution, 27 patients >25%) was very similar to Cosmic mutational signatures 4 and 8 (Fig. 2b) and was more frequently observed in patients pre-treated with platinum-based chemotherapy ($P = 0.001$) in our cohort. Signature 4 has been associated with tobacco mutagens[15] and is characterized by C>A substitutions and CC>AA dinucleotide substitutions[5]. The aetiology of signature 8 is still unknown but its presence has been observed in primary BC and recently linked to *BRCA1/2* deficiency[7]. This signature also shows C>A substitutions and has the CC>AA characteristic. Cisplatin mainly forms Pt-d(GpG) diadducts[16] and patients pre-treated with platinum-based chemotherapy showed higher levels of CC>AA substitutions (1.8x, 95%CI 1.2–2.5, $P = 0.0013$) than patients who did not receive prior platinum. Also, patients with at least 10% contribution of de novo signature I had higher levels of CC>AA (2x, 95%CI 1.6–2.2, $P <$ 1e-5, Fig. 2c), but patients with at least 10% contribution of signature 8 did not have elevated CC>AA levels ($P = 0.706$). A previously published cisplatin signature[19] with characteristic C>T conversion peaks, which are absent in de novo signature I, had a higher cosine similarity in patients pre-treated with platinum-based chemotherapy than in patients who did not receive this pre-treatment (1.2x, 95%CI 1.1–1.3, $P = 0.0008$, Fig. 2d). Furthermore, when dichotomising samples in two groups based on their similarity to the cisplatin-signature of Boot et al.,[19] (permutation p<0.05 and p>0.05, respectively), 23 out of 27 samples with at least 25% contribution of de novo signature I had a high similarity to this signature (2.6x, 95%CI 2.4–3.0, $P <$ 1e-5, Fig. 2e). Next, since de novo signature I resembles signature 8, which in turn is linked to BRCA1/2 deficiency[7], we analysed germline *BRCA*-status in this context. A multivariate regression model showed that both germline *BRCA*-status and pre-treatment with platinum-containing drugs were significantly associated with the relative contribution of de novo signature I ($P <$ 1e-5 for the model, estimated coefficients for germline *BRCA*- status 10.36 ($P = 5.4$e-7) and 4.13 ($P = 0.0014$) for pre-treatment with platinum).

Since the observed de novo signatures largely overlapped with known Cosmic signatures, we also determined the contributions of these 30 known signatures to the mutational landscape of our cohort (Supplementary Fig. 5 and 6). Out of the 30 Cosmic signatures, 12 contributed to ≥10% of the observed mutations in at least 5 patients and were therefore defined as a dominant signature (Fig. 3). The most frequently represented signatures in mBC were signature 8 (64% of patients); signature 1 (59%), related to aging; signature 2 (43%) and 13 (36%), related to APOBEC mutagenesis; and signature 3 (41%), associated with homologous recombination deficiency (HRD). Analyses per BC subtype revealed that signature 3 and signature 9 mutations were significantly more often present (2.7x, 95%CI 1.9–3.9 and 1.3x, 95%CI 1.1–1.6, respectively) in ER-negative compared to ER-positive mBC, whereas signature 2 (APOBEC) mutations were significantly more frequent (2.1x, 95%CI 1.5–2.9) in ER-positive mBC (all P < 0.05).

Ten of the 12 Cosmic signatures detected in our mBC cohort, were previously described in the BASIS primary BC cohort[6] (signatures 1-3, 5, 6, 8, 13, 17, 18 and 30), whereas signature 9 (8%) and 16 (14%) were not reported in the BASIS cohort[6]. After re-evaluation of the previously published BASIS data[6] we found that the latter two signatures were actually present, but at relatively low levels (median relative contribution <5%). Subsequently, we compared the absolute and relative contributions of the 12 dominant Cosmic signatures between our mBC cohort and the BASIS primary BC cohort per BC subtype (Fig. 3, see Supplementary Fig. 5 for all relative comparisons of the 30 Cosmic signatures). Irrespective of BC subtype, the median absolute number of mutations was higher in mBC compared to primary BC for almost all signatures (Supplementary Fig. 6), reflecting the significantly higher TMB in mBC and ongoing mutagenic processes.

On a relative scale we found a decrease in signatures 1 and 5 (age) and signature 16 (reported in liver cancer), as well as an increase in signatures 2 and 13 (APOBEC), and signature 17 (unknown aetiology) in ER+/HER2- metastatic disease compared to ER+/HER2- primary BC from the BASIS cohort. In triple negative BC (TNBC) a decrease in signature 3 (HRD) and an increase in signatures 2 and signature 17 was seen in metastatic lesions compared to primary BC. In patients with HER2+ disease, no differences in the relative contributions of the 12 dominant signatures were found between primary and metastatic disease (Fig. 3), irrespective of taking ER-status into account.

To determine whether these differences between primary BC and mBC were driven by disease course or pre-treatment we performed a multivariate linear regression analysis using the previously defined 4 groups of primary and metastatic lesions with and without pre-treatment. This showed a significantly lower (signature 1) and higher (signature 17) contribution in pre-treated patients, irrespective of disease course. Thus, pre-treatment in itself – regardless of treatment type – causes a limited shift in certain signature patterns.

Within mBC lesions we also investigated the potential role of specific pre-treatments on the relative signature contributions for all 12 dominant Cosmic signatures as defined above. Pre-treatment with 5-FU, taxanes, platinum-containing chemotherapy and/or eribulin was associated with significantly higher relative contributions of signature 17 (all FDR P-values <0.05, with 5-FU the most significant at FDR P=2.0e-9) (Supplementary Fig. 7). These

treatments have been given to 40%, 58%, 10%, and 3% of patients, respectively. The large overlap in patients who were pre-treated with all abovementioned therapies hampers further specification of which of these therapies is directly associated with signature 17. Although signature 17 is present in primary BC due to endogenous processes, the fact that signature 17 is mainly characterised by T>G and T>C in a CTT context, might implicate 5-FU inhibiting thymidylate synthase and thus synthesis of thymidine[20] as a likely drug contributing to this pattern. Finally, we investigated the association between the mutational signatures and response to the line of therapy that was initiated directly after sampling tumour material. Patients with progression at first response evaluation after twelve weeks of treatment had a significantly higher relative contribution of signature 17 ($P = 0.0012$). However, we also observed that the number of pre-treatments given is higher in patients with ≥10% signature 17 contribution, rendering it hard to distinguish whether or not signature 17 is truly a biomarker for poor response to therapy or a marker of poor outcome in general.

In conclusion, virtually all mutational processes present in primary BC contribute to the observed increased TMB in mBC. On a relative scale, we do observe a shift from more indolent age-related mutagenesis in primary disease towards more APOBEC-driven processes in mBC. On top of that, previously given lines of therapy can impose specific mutational profiles in BC cells.

Structural variation and homologous recombination deficiency

To evaluate structural variation in metastatic lesions we extracted the six previously described rearrangement signatures[6]. Rearrangement signature 1 and 3 (SV1, SV3) were the least prevalent (both 6% of all rearrangements) in metastatic lesions while SV2, SV4, SV6 contributed 20%, 14% and 19%, respectively. SV5 was most dominant and contributed to 36% of all rearrangements. Compared to primary BC, the relative contribution of *BRCA1*-related SV3 is significantly decreased (2.9x, 95%CI 1.5–7.1, $P <$ 1e-5) while *BRCA2*-related SV5 increased (3.2x, 95%CI 2.7–-3.8, $P <$ 1e-5) in metastatic lesions regardless of BC subtype (Supplementary Fig. 8).

To more specifically investigate the presence of an HRD phenotype based on somatic alterations, we applied the recently developed Classifier of Homologous recombination Deficiency (CHORD) (Nguyen, van Hoeck and Cuppen, manuscript in preparation). This algorithm predicts HRD and assigns the most likely responsible *BRCA* gene, based on a combination of rearrangement signatures (SV1, SV3 and SV5), a specific type of InDels flanked by micro homology and mutational signature 3. In our cohort of 442 patients, 18 were known to have a germline loss of *BRCA1* or *2* (*BRCA1* n=5; *BRCA2* n=13). CHORD identified 39 additional patients carrying a HRD tumour next to all 18 germline *BRCA*-mutation carriers.

Unsupervised clustering reveals eight distinct genomic clusters in mBC

Based on the above described genomic characteristics of our mBC cohort comprising 442 metastatic lesions, we performed an unsupervised clustering analysis, which revealed eight clusters representing tumours with distinct genomic phenotypes (Fig. 4). Biopsy site and

treatment outcome were evenly distributed among the eight clusters. Cluster A and B are both characterized by mutational signature 3. Cluster A is further characterized by short tandem duplications and by SV3 and cluster B by large deletions and SV5. In addition, these two clusters are enriched for HRD (P < 1e-5) as predicted by the CHORD algorithm. In cluster A, HRD is predicted to be based on *BRCA1* deficiency and in cluster B based on *BRCA2* deficiency. However, clusters A and B also contained one and four patients, respectively, who were predicted HR-proficient. In these patients, we checked for mutated genes that are known in HR (as described in method section), however, none of these genes were homozygously affected.

Clusters C, D and E are respectively characterized by mutational signatures 17, 18 and 16. Cluster F is mainly based on insertions. Cluster G shows a low TMB, few structural variants and a relatively high proportion of mutational signature 5. Finally, cluster H represents tumours predominantly harbouring mutational signatures 2 and 13 related to APOBEC mutagenesis, a relative high TMB and kataegis events. Kataegis was observed in 177 (40%) mBC patients (ranging from 1 – 144 events), with 15 patients exhibiting 10 or more foci. In kataegis foci, mainly APOBEC mediated mutagenesis occurred (P < 0.001, Supplementary Fig. 9). Patients exhibiting kataegis frequently harboured *ATR* mutations (21 out of 25 identified patients with an *ATR* mutation showed kataegis), suggesting that kataegis might be associated with collapsing replication forks in these patients.

Somatic drivers of mBC: SNVs and copy number alterations

Using the ratio of nonsynonymous and synonymous mutations (dN/dScv)[21], we identified 21 potential driver genes, including known key drivers of BC. The top five driver genes were *TP53* (42.8%), *PIK3CA* (42.3%), *ESR1* (14.3%), *GATA3* (11.3%) and *KMT2C* (11.3%) (Supplementary Table 3; Supplementary Fig. 10). With respect to BC subtypes, we observed that *TP53* was enriched in TNBC (P < 1e-5), whereas *ESR1, PIK3CA,* and *GATA3* were more often mutated in ER-positive mBC (all P-values <0.001). *ESR1* mutations were, as expected, more frequently present in patients pre-treated with aromatase inhibitors (AI) (26.9% in pre-treated patients versus 2.7% in patients without AI pre-treatment).

In addition to nonsynonymous mutations, we observed 44 rearrangements involving *ESR1* in 34 patients and deep gains of *ESR1* in 29 patients. Fusions, mutations and deep gains were not mutually exclusive, as were specific for ER-positive BC. No amplifications of cis-acting enhancers of *ESR1* were observed.

We compared the frequency of alterations in our 21 identified potential drivers in mBC with two primary BC cohorts: TCGA[4] and BASIS[6] (Supplementary Table S2). Using an FDR <0.05, six genes, including *ESR1, TP53, NF1, AKT1, KMT2C* and *PTEN* (Supplementary Table S4), were more frequently mutated in ER+/HER2- metastatic lesions than in primary BC. Except for *ESR1,* these genes were not associated with pre-treatment, nor with response. Individual analysis did not reveal mutual exclusivity of these genes, however grouping of MAPK-pathway and ER transcriptional regulator genes (*NF1, TBX3, ERBB2, CTCF, EGFR, KRAS, BRAF, ERBB3, HRAS, MYC*) did show mutual exclusivity with *ESR1*, as shown by Razavi et al[18]. In patients with HER2+ disease (irrespective of

subdivision by ER-status) or TNBC no significant differences were observed. A bootstrap analysis to better estimate the distribution of gene mutation frequencies in primary disease (TCGA and BASIS combined) confirmed that observed enrichments of *ESR1, TP53, NF1, AKT1, KMT2C and PTEN* were unlikely to be explained by sampling bias.

The *d*N/*d*Scv analysis identified an additional potential driver gene: *GPS2* which was not identified as driver in primary BC[6], but was recently described by Martincorena et al.[21] in primary BC. The GPS2 protein forms a complex with NCOR1 and HDAC3. These three genes were virtually mutually exclusively affected; 35 out of 36 patients harbouring mutations in these genes had only one gene affected (CoMet exact test, *P* < 1e-5), indicating that the loss of either gene in this complex is sufficient. Alterations in GPS2-NCOR1-HDAC3 complex are enriched in mBC compared to primary BC (*P* = 0.004), but not associated with a certain prior treatment or BC subtype.

Regarding the primary 93 BC driver genes reported by Nik-Zainal et al.[6] we found that, in addition to the above described differences between ER+/HER2- primary and metastatic disease for *ESR1, NF1* and *TP53, KMT2D* was also more frequently affected in metastatic disease whereas *AX1N1* was less frequently altered compared to primary BC (FDR <0.05) (Supplementary Table 5). Again, no differences for HER2+ (irrespective of subdivision by ER-status) and TNBC were observed.

Copy number analyses identified 51 narrow regions with somatic copy number alterations, including amplification peaks containing known driver genes such as *ERBB2, MYC*, and *CCND1* and deletion peaks containing known tumour suppressor genes such as *PTEN, CKDN2A, RB1*, and *NF1*. Using an FDR <0.05, 29 regions were associated with ER-status, i.e. *MYC, SLC1A2*, and *HOOK3* were more frequently amplified in ER-negative mBC and *PLK2* was more frequently deleted in ER-negative mBC. All amplification and deletion peaks in relation to ER-status are shown in Supplementary Table 6. The total number of copy number alterations within these 51 regions was not associated with metastatic site or prognosis after 12 weeks of treatment. In addition we observed 6 focal amplification peaks (<5kb) in non-coding parts near three known BC driver genes *ZNF217, ZNF703, MYC* and three other genes *LINC00266-1, TRPS1* and *KCNMB2*.

Potential clinical implications of WGS

To evaluate whether WGS may represent a valuable tool to improve treatment choices for future mBC patients we specifically focussed on 1) high TMB/ microsatellite instability (MSI) as a potential biomarker to select patients for immunotherapy, 2) HRD for PARP-inhibitors and/or double strand DNA breaks inducing chemotherapy, and 3) specific genomic alterations for which FDA-approved drugs are already available (Fig. 5).

Using a threshold of ≥10 mutations per Mb in our cohort, previously used to distinguish responding from non-responding lung cancer patients receiving nivolumab plus ipilimumab[22], we identified 50 patients (11%) in our cohort with a high TMB, which in most patients (70%) could be largely attributed to APOBEC-related mutations (≥50% of all mutations). In primary BC APOBEC mutagenesis was previously associated with the

presence of tumour infiltrating lymphocytes further confirming antigenicity of APOBEC mutant cancers[23,24]. High TMB was not associated with BC subtype, suggesting inclusion of future patients into clinical trials on check point inhibitors should potentially be based on their genomic landscape rather than on tumour subtype (Supplementary Fig. 2a). Of note, there were 5 patients with a high TMB and a mutation in either *JAK2* or *STAT3,* which might also be clinically relevant since these mutations could help to evade the native immune response[25]. We also identified seven (1.5%) patients with MSI according to MSIseq[26,27], which is currently not tested in standard care but for which pembrolizumab has been approved for use in all tumour types[28].

Using CHORD, we identified 39 additional HRD patients (9%) that did not harbour germline alterations in BRCA1/2. Based on their HRD-phenotype, these patients might benefit from PARP-inhibitors and/or chemotherapeutics that induce double strand DNA breaks.[29]

Finally, we analysed which patients could be treated with FDA-approved drugs based on the alterations present in their genome using the clinical annotation database OncoKB[30]. 105 patients (24%) had at least one actionable event for which an FDA approved drug is currently available. 67 (15%) of these patients had an *ERBB2* amplification, 7 of which were clinically known as HER2-negative. These patients might benefit from anti-HER2 therapies, which is already approved for BC. Additionally, 47 patients had at least one alteration predicting response to a drug registered for other tumour types than BC. (Fig. 5; **Supplementary Table 7**). In summary, WGS provides us with a valuable tool to determine clinically relevant molecular features for informed treatment choices, such as TMB, HRD, MSI, and actionable mutations in one assay.

Discussion

In the present study we provide the first in-depth whole genome analysis of metastatic lesions from 442 patients with mBC. We have identified differences in mBC compared to primary BC regarding TMB, the frequency in which driver genes are affected, and relative contribution of mutational signatures. Moreover, we show that the use of WGS enables to identify subgroups of patients (42% of all patients with mBC) for personalised treatment. Future clinical trials should therefore focus on treatment stratification based on 'clinical genomics' and incorporate tissue biopsies for sequencing in their study protocols to enable correlations of genomic alterations with response to therapy.

Nevertheless, based on the current knowledge and treatment armamentarium we still have, a substantial number of mBC patients (58%) without currently known targetable genomic features. Expanding the number of samples will allow for better stratified analyses based on patient characteristics such as pre-treatment, BC subtype and line of treatment. Further exploration of large copy number changes, specific combinations of mutated genes and RNA sequencing will potentially unravel new actionable targets or profiles. The development and approval of new drugs which are currently under investigation, such as PI3K inhibitors potentially relevant to a large subset of mBC patients (42% harbours a *PIK3CA* mutation in our cohort), will further increase the targetability of the tumour's genome.

Overall, our study provides significant insight into the biology of mBC and generates useful genomic information for future improvement of patient management.

Online Methods

Patient cohort and study procedures

For our analyses, we selected patients with metastatic breast cancer who were included under the protocol of the Centre for Personalized Cancer Treatment (CPCT) consortium (NCT01855477). A detailed description of the consortium and the whole patient cohort has been described in detail recently[27]. This consortium consists of 49 oncology centres in The Netherlands and aims to analyse the cancer genome of patients with advanced cancer, irrespective of cancer type, to develop predictors for outcome to systemic treatment. Patients of ≥18 years, with incurable locally advanced or metastatic solid tumours of whom a histological biopsy could be safely obtained and for whom systemic treatment with anti-cancer agents was indicated, were eligible for inclusion. All patients gave written informed consent prior to any study procedure. Here we performed an in depth analysis of all included metastatic breast cancer patients. Patients who were biopsied in their primary breast tumour (n=55), were excluded from the metastatic analyses, but were used as an additional control group. Patients with evaluable biopsies were classified according to the oestrogen receptor (ER) status and HER2 status (Supplementary Table S1). Collection and sequencing of samples was performed as described previously[27].

Treatment outcome

Clinical outcome was evaluated according to RECISTv1.1 after 12 weeks of treatment and was defined as stable disease (SD), partial response (PR), complete response (CR) or progressive disease (PD)[31]. To relate outcome to genomic data we defined response to therapy as CR of PR after 12 weeks of treatment and non-response as PD at 12 of weekend.

Detection of somatic changes

Detailed methods on calling if somatic SNVs, MNVs and structural variants were previously described[27]. Additional annotation of somatic variants and heuristic filtering was performed: heuristic filtering removed somatic SNV, InDel and MNV variants based on the following criteria: 1) minimal alternative reads observations ≤3; 2) gnoMAD exome (ALL) allele frequency ≥0.001 (corresponding to ~62 gnoMAD individuals); and 3) gnoMAD genome (ALL) ≥0.005 (~75 gnoMAD individuals)[32]. GnoMAD database v2.0.2 was used. Per gene overlapping a genomic variant, the most deleterious mutation was used to annotate the overlapping gene. Structural variants, with B-allele frequency (BAF) ≥0.1, were further annotated by retrieving overlapping and nearest up-stream and downstream annotations using custom R scripts based on GRCh37 canonical UCSC promoter and gene annotations with respect to their respective upstream or downstream orientation (if known)[33]. Only potential fusions with only two different gene-partners were considered; structural variants with both breakpoints falling within the same gene were simply annotated as structural variant mutations. Fusion annotation from the COSMIC (v85), CGI and CIVIC databases were used to assess known fusions[34–36]. The COSMIC (v85), OncoKB (July 12, 2018),

CIVIC (July 26, 2018), CGI (July 26, 2018) and the list from Martincorena et al. (dN/dS) were used to classify known oncogenic or cancer-associated genes[21,34–36].

Ploidy and copy number analysis

Ploidy and copy number (CN) analysis was performed by a custom pipeline as previously described[27]. Briefly, this pipeline combines (BAF), read depth and structural variants to estimate the purity and CN profile of a tumour sample. Recurrent focal and broad CN alterations were identified by GISTIC2.0 (v2.0.23)[37]. GISTIC2.0 was run with the following parameters: a) genegistic 1; b) gcm extreme; c) maxseg 4000; d) broad 1; e) brlen 0.98; f) conf 0.95; g) rx 0; h) cap 3; i) saveseg 0; j) armpeel 1; k) smallmem 0; l) res 0.01; m) ta 0.1; n) td 0.1; o) savedata 0; p) savegene 1; q) gvt 0.1. Categorization of shallow and deep CN aberration per gene was based on thresholded GISTIC2 calls. Focal peaks detected by GISTIC2 were re-annotated, based on overlapping genomic coordinates, using custom R scripts and UCSC gene annotations. GISTIC2 peaks were annotated with all overlapping canonical UCSC genes within the narrow peak limits. If a narrow GISTIC2 peak overlapped with ≤3 genes, the most-likely targeted gene was selected based on oncogenic or tumour suppressor annotation in the COSMIC (v85), OncoKB (July 12, 2018), CIVIC (July 26, 2018) and CGI (July 26, 2018) lists[21,34–36]. Peaks in gene deserts were annotated with their nearest gene.

Putative enhancer regions (as detected by GISTIC2; focal amplification peaks with a width < 5,000 bp) were retrieved per sample. If regions overlapped multiple distinct copy-number segments, the maximum copy-number value of the overlapping segments was used to represent the region. Samples with gene-to-enhancer ratios deviating >1 studentised residual from equal 1:1 gene-to-enhancer ratios (linear model: log2(copy number of enhancer) - log2(copy number of gene locus) ~ 0) were categorized as gene or enhancer enriched. Based on the direction of the ratio, samples were either denoted as enhancer (if positive ratio) or gene (if negative ratio) enriched.

Estimation of tumour mutational burden

The mutation rate per megabase (Mb) of genomic DNA was calculated as the total genome-wide amount of SNV, MNV and InDels divided over the total amount of mappable nucleotides (ACTG) in the human reference genome (hg19) FASTA sequence file:

$$TMB_{genomic} = \frac{(SNV_g + MNV_g + InDels_g)}{\frac{2858674662}{10^6}} \tag{1}$$

The mutation rate per Mb of coding mutations was calculated as the amount of coding SNV, MNV and InDels divided over the summed lengths of distinct non-overlapping coding regions, as determined on the subset of protein-coding and fully supported (TSL = 1) transcripts in GenCode v28 (hg19)[38]:

$$TMB_{coding} = \frac{(SNV_c + MNV_c + InDels_c)}{\left(\frac{28711682}{10^6}\right)} \tag{2}$$

MSI and HRD prediction

HRD/BRCAness was estimated using the CHORD classifier (Nguyen, van Hoeck and Cuppen, manuscript in preparation). This classifier was based on the HRDetect[7] algorithm, however, redesigned to improve its performance beyond primary BC. The binary prediction score (ranging from zero to one) was used to indicate BRCAness level within sample. A *BRCA1/2* variant was assigned as pathogenic when annotated by ENIGMA[39] (February 26, 2018) or ClinVar[40] (January 28, 2018). We used the following gene-list to check whether HR-related genes were mutated in samples that clustered in cluster A and B and were classified as HR-proficient (Fig. 4). Gene- list: *ATM, BARD1, BLM, BRCA1, BRCA2, BRIP1, EME1, ERCC1, ERCC4, EXO1, GEN1, H2AFX, MRE11A, MUS81, NBN, NSMCE1, NSMCE2, PALB2, PCNA, RAD18, RAD21, RAD50, RAD51, RAD51AP1, RAD51C, RAD51L1, RAD51L3, RAD52, RAD54B, RAD54L, RECQL4, RECQL5, RTEL1, SLX1A, SLX4, TDP1, WRN, XRCC2,* and *XRCC3.*

MSI status was determined using the MSIseq score[26,27]. In short, this validated score classifies a sample based on the number of InDels per million bases occurring in homopolymers of 5 or more bases or dinucleotide, trinucleotide and tetranucleotide sequences of repeat count 4 or more. A sample is considered MSI with a MSIseq score >= 4.

Detection of (onco-)genes under selective pressure

To detect (onco-)genes under tumour-evolutionary mutational selection, we employed a Poisson-based dN/dS model (192 rate parameters; under the full trinucleotide model) by the R package dndscv (v0.0.0.9)[21]. Briefly, this model tests the normalized ratio of non-synonymous (missense, nonsense and splicing) over background (synonymous) mutations whilst correcting for sequence composition and mutational signatures. A global q-value ≤ 0.1 (with and without taking InDels into consideration) was used to identify statistically-significant (novel) driver genes.

Identification of hypermutated foci (kataegis)

Putative kataegis events were detected using a dynamic programming algorithm which determines a globally optimal fit of a piecewise constant expression profile along genomic coordinates as described by Huber et al. and implemented in the tilingarray R package (v1.56.0)[41]. Only SNVs were used in detecting kataegis. Each chromosome was assessed separately and the maximum number of segmental breakpoints was based on a maximum of five consecutive SNVs (max. 5,000 segments per chromosome). Fitting was performed on $\log_{10}$-transformed intermutational distances. Per segment, it was assessed if the mean intermutational distance was ≤2,000 bp and at least five SNVs were used in the generation of the segment. Samples with >200 distinct observed events were set to zero observed events as these were found to be hypermutated throughout the entire genome rather than locally. Kataegis was visualized using the R package karyoploteR (v1.4.1)[42].

Mutational and structural rearrangement signatures analysis

Mutational signatures analysis using the MutationalPatterns R package (v1.4.2) was performed as previously described[43]. The thirty Cosmic mutational signatures, as established by Alexandrov et. al, (matrix Sij; $i = 96$; number of trinucleotide motifs; $j =$

number of signatures) were downloaded from COSMIC (as visited on 23-05-2018)[2]. For de novo signature calling, between two and twenty signatures were assessed using the NMF package (v0.21.0) with 500 iterations.[44] By comparing the cophenetic correlation coefficient over the range of possible signatures, we opted to assign ten de novo signatures. We used the cosine similarity metric to compare the de novo signatures with the Cosmic signatures. Structural rearrangement signatures were established as previously described.[6] In brief, Structural Variants were called using Manta(v1.0.3)[45] and default parameters, after which additional filters were applied.[27] The reported tandem duplications, deletions, inversions, insertions and translocations were then categorised by size (<10kb, 10-100kb, 100kb-1Mb, 1-10Mb and >10Mb). Inter-rearrangement distances were calculated and rearrangements were labelled as clustered if the average inter-rearrangement distance of a segment was at least 10 times less than the whole-genome average for a patient sample. The segments were determined using a piecewise constant fitting function ('exactPcf' from the 'copynumber' R package) using a minimum of 10 events in a segment (K_{min}) and a γ of 25 (smoothness of segmentation). To assign an *a posteriori* probability for each substitution from which of the signatures it was most likely caused by, we implemented the previously described method[46]. In short, this method uses the contribution of each signature in each sample in conjunction with the probability of a signature to generate the particular substitution in its trinucleotide context.

Unsupervised clustering of mBC WGS characteristics

Samples were clustered using Pearson correlation coefficient (1-*r*) as distance metric and Ward.D hierarchical clustering based the following whole genome characteristics; number of SNV, InDel and MNV per Mb, total number and numbers by type of structural variants and the relative frequencies of the mutational signatures. Data were scaled but not centered (root mean square) prior to calculating Pearson correlation coefficients. After clustering, optimal leaf ordering (OLO) was performed using the seriation package (v1.2.3)[47]. The gap-statistic method was employed to determine optimal number of discriminating clusters.

Comparison with primary breast cancer

BASIS cohort—Somatic mutations for the BASIS cohort were extracted from the European-Genome Phenome Archive, under accession code EGAS00001001178. This cohort of whole genomes of 560 primary breast cancers and paired non-neoplastic tissue as reference, consists of 320 ER-positive/HER2-negative, 46 ER-positive/HER2-positive, 27 ER-negative/HER2-positive and 167 triple negative BC patients[6]. To allow comparison of mutational loads, mutational signatures and somatic mutations to be fair between BASIS and our cohort, we compared whether the calling from both pipelines yielded comparable results for eight patients from the BASIS cohort. As the mutational load (linear regression R^2=0.9987), mutational signatures (average similarity of 0.90 (SD 0.08), which is significantly higher (one-sample t-test, P = 1.57e-5) than the similarity between non-matching samples) and detected driver genes were very similar between both pipelines, we considered the results from the pipelines to be comparable.

TCGA cohort—Breast cancer data (n=805) were downloaded from cbioportal.org (April 2018). Synonymous mutations were removed and multiple mutations in the same gene/

patient were combined. For the Copy Number data, a -2 call was used as Deletion, +2 Amplification. This cohort consists of 143 triple negative, 496 ER-positive/HER2-negative, 39 ER-positive/HER2-negative and 127 ER-positive/HER2-positive patients.

Selection of cohort per analysis—For the most optimal comparison of TMB, and absolute and relative contributions of Cosmic signatures between primary BC and mBC, we selected the BASIS cohort, since this cohort also used WGS making it the most suitable dataset for these comparisons. For the comparison of driver genes, which are located in the coding parts of the genome, we decided to use both cohorts (TCGA and BASIS) to increase power.

Bootstrapping of primary cohort—To investigate whether the enrichment of driver genes in mBC compared to primary BC was influenced by population differences or sampling bias, we performed a bootstrap analysis to better estimate the distribution of gene mutation frequencies in primary BC (TCGA[4] and BASIS[6] combined). For each of the driver genes, a bootstrap analysis was performed by taking the actual mutated frequency in primary BC within a subtype, randomly selected 80% of cases using sampling with replacement and counted the number of times a sample was selected that was mutated for that gene. This was repeated 100,000 times to obtain an estimated distribution for a gene in primary breast cancer. Then, we determined whether the mutation frequency of that gene in the metastatic cohort in the same subtype fell outside the 99% percentile of the estimated primary BC distribution.

Statistics

Pearson's Chi square test or Fisher's exact test (in case of too few expected events) was used to evaluate categorical data (for example, prior treatment versus the occurrence of a certain mutation). To compare continuous variables (for example the relative contribution of mutational signatures versus breast cancer subtype or RECIST1.1 response category (CR/PR or PD)) a Mann-Whitney U-test or a Kruskal-Wallis H-test was performed. Where suitable, effect sizes and confidence intervals were estimated using Hodges-Lehmann's method[48,49]. All statistical tests were two-sided and considered statistically significant when $P < 0.05$. Stata 13.0 software, R version 3.4.4. or SPSS version 24 were used for the statistical analyses. We used the Hochberg procedure to correct p-values for multiple hypothesis testing when appropriate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data availability

WGS data and corresponding clinical data have been requested from Hartwig Medical Foundation and provided under data request number DR-026. The clinical data provided by CPCT have been locked at 1st of June 2018. Both WGS and clinical data are freely available for academic use from the Hartwig Medical Foundation through standardized procedures and request forms can be found at https://www.hartwigmedicalfoundation.nl [27].

Code availability

Full codes are available at https://github.com/hartwigmedical/ and https://bitbucket.org/ccbc/r2ccbc.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018; 68:7–30. [PubMed: 29313949]

2. Alexandrov LB. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–421. [PubMed: 23945592]

3. Nik-Zainal S, Morganella S. Mutational Signatures in Breast Cancer: The Problem at the DNA Level. Clin Cancer Res. 2017; 23:2617–2629. [PubMed: 28572256]

4. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]

5. Nik-Zainal S, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012; 149:979–993. [PubMed: 22608084]

6. Nik-Zainal S, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016; 534:47–54. [PubMed: 27135926]

7. Davies H, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017; 23:517–525. [PubMed: 28288110]

8. Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. Cancer Discov. 2015; 5:704–712. [PubMed: 26091828]

9. Burns MB, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013; 494:366–370. [PubMed: 23389445]

10. Brown D, et al. Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. Nat Commun. 2017; 8:14944. [PubMed: 28429735]

11. Brastianos PK, et al. Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. Cancer Discov. 2015; 5:1164–1177. [PubMed: 26410082]

12. Savas P, et al. The Subclonal Architecture of Metastatic Breast Cancer: Results from a Prospective Community-Based Rapid Autopsy Program "CASCADE". PLoS Med. 2016; 13:e1002204. [PubMed: 28027312]

13. Fumagalli D, et al. Somatic mutation, copy number and transcriptomic profiles of primary and matched metastatic estrogen receptor-positive breast cancers. Ann Oncol. 2016; 27:1860–1866. [PubMed: 27672107]

14. Ng CKY, et al. Genetic Heterogeneity in Therapy-Naive Synchronous Primary Breast Cancers and Their Metastases. Clin Cancer Res. 2017; 23:4402–4415. [PubMed: 28351929]

15. Schrijver W, et al. Mutation Profiling of Key Cancer Genes in Primary Breast Cancers and Their Distant Metastases. Cancer Res. 2018; 78:3112–3121. [PubMed: 29615433]

16. Lefebvre C, et al. Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. PLoS Med. 2016; 13:e1002201. [PubMed: 28027327]

17. Robinson DR, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017; 548:297–303. [PubMed: 28783718]

18. Razavi P, et al. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. Cancer Cell. 2018; 34:427–438 e426. [PubMed: 30205045]

19. Boot A, et al. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. Genome Res. 2018; 28:654–665. [PubMed: 29632087]

20. Wyatt MD, Wilson DM 3rd. Participation of DNA repair in the response to 5-fluorouracil. Cell Mol Life Sci. 2009; 66:788–799. [PubMed: 18979208]

21. Martincorena I, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2018; 173:1823.

22. Ramalingam SS, et al. Abstract CT078: Tumor mutational burden (TMB) as a biomarker for clinical benefit from dual immune checkpoint blockade with nivolumab (nivo) + ipilimumab (ipi) in first-line (1L) non-small cell lung cancer (NSCLC): identification of TMB cutoff from CheckMate 568. Cancer Research. 2018; 78:CT078–CT078. DOI: 10.1158/1538-7445.am2018-ct078

23. Smid M, et al. Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. Nat Commun. 2016; 7

24. Pitt JJ, et al. Characterization of Nigerian breast cancer reveals prevalent homologous recombination deficiency and aggressive molecular features. Nat Commun. 2019; 9:4181.

25. Yates LR, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell. 2017; 32:169–184 e167. [PubMed: 28810143]

26. Huang MN, et al. MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. Sci Rep. 2015; 5

27. Priestley P, et al. Pan-cancer whole genome analyses of metastatic solid tumors. bioRxiv. 2018; doi: 10.1101/415133

28. Brahmer JR, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. N Engl J Med. 2012; 366:2455–2465. [PubMed: 22658128]

29. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science. 2017; 355:1152–1158. DOI: 10.1126/science.aam7344 [PubMed: 28302823]

30. Chakravarty D, et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol. 2017; 2017

31. Eisenhauer EA, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009; 45:228–247. [PubMed: 19097774]

32. Lek M, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285–291. [PubMed: 27535533]

33. Casper J, et al. The UCSC Genome Browser database: 2018 update. Nucleic Acids Res. 2018; 46:D762–D769. [PubMed: 29106570]

34. Forbes SA, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017; 45:D777–D783. [PubMed: 27899578]

35. Tamborero D, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. 2018; 10:25. [PubMed: 29592813]

36. Griffith M, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet. 2017; 49:170–174. [PubMed: 28138153]

37. Mermel CH, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011; 12:R41. [PubMed: 21527027]

38. Harrow J, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012; 22:1760–1774. [PubMed: 22955987]

39. Spurdle AB, et al. ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. Hum Mutat. 2012; 33:2–7. [PubMed: 21990146]

40. Landrum MJ, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016; 44:D862–868. [PubMed: 26582918]

41. Huber W, Toedling J, Steinmetz LM. Transcript mapping with high-density oligonucleotide tiling arrays. Bioinformatics. 2006; 22:1963–1970. [PubMed: 16787969]

42. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics. 2017; 33:3088–3090. [PubMed: 28575171]

43. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Medicine. 2018; 10:33.doi: 10.1186/s13073-018-0539-0 [PubMed: 29695279]

44. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010; 11:367. [PubMed: 20598126]

45. Chen X, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016; 32:1220–1222. [PubMed: 26647377]

46. Morganella S, et al. The topography of mutational processes in breast cancer genomes. Nat Commun. 2016; 7

47. Hahsler M, Hornik K, Buchta C. Getting Things in Order: An Introduction to the R Package seriation. 2008; 25:34.doi: 10.18637/jss.v025.i03

48. Hodges JL, Lehmann EL. Estimates of Location Based on Rank Tests. The Annals of Mathematical Statistics. 1963; 34:598–611.

49. Lehmann EL. Nonparametric Confidence Intervals for a Shift Parameter. 1963; 34:1507–1512. DOI: 10.1214/aoms/1177703882
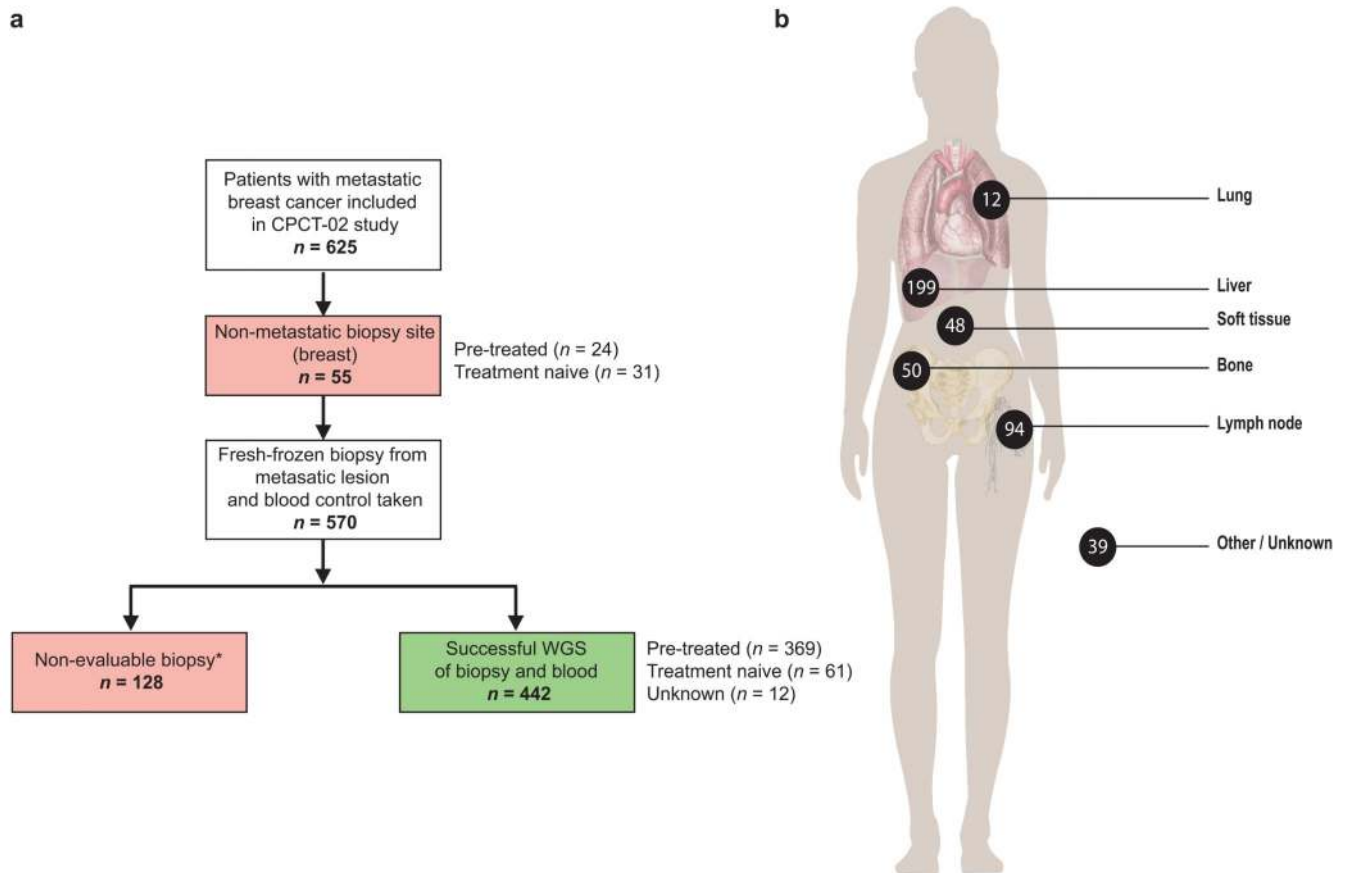
a

b

Figure 1. Overview of study design and biopsy sites (*n* = 442)
(a) **Flowchart of patient inclusion.** From the CPCT-02 cohort, patients with mBC were selected. Patients were excluded if the only available biopsy was from the primary lesion. *Non-evaluable biopsies were defined as no biopsy taken at all, <30% tumour cells or too low DNA yield for WGS. (b) **Overview of biopsy sites.** Number of biopsies per metastatic site analysed with WGS.
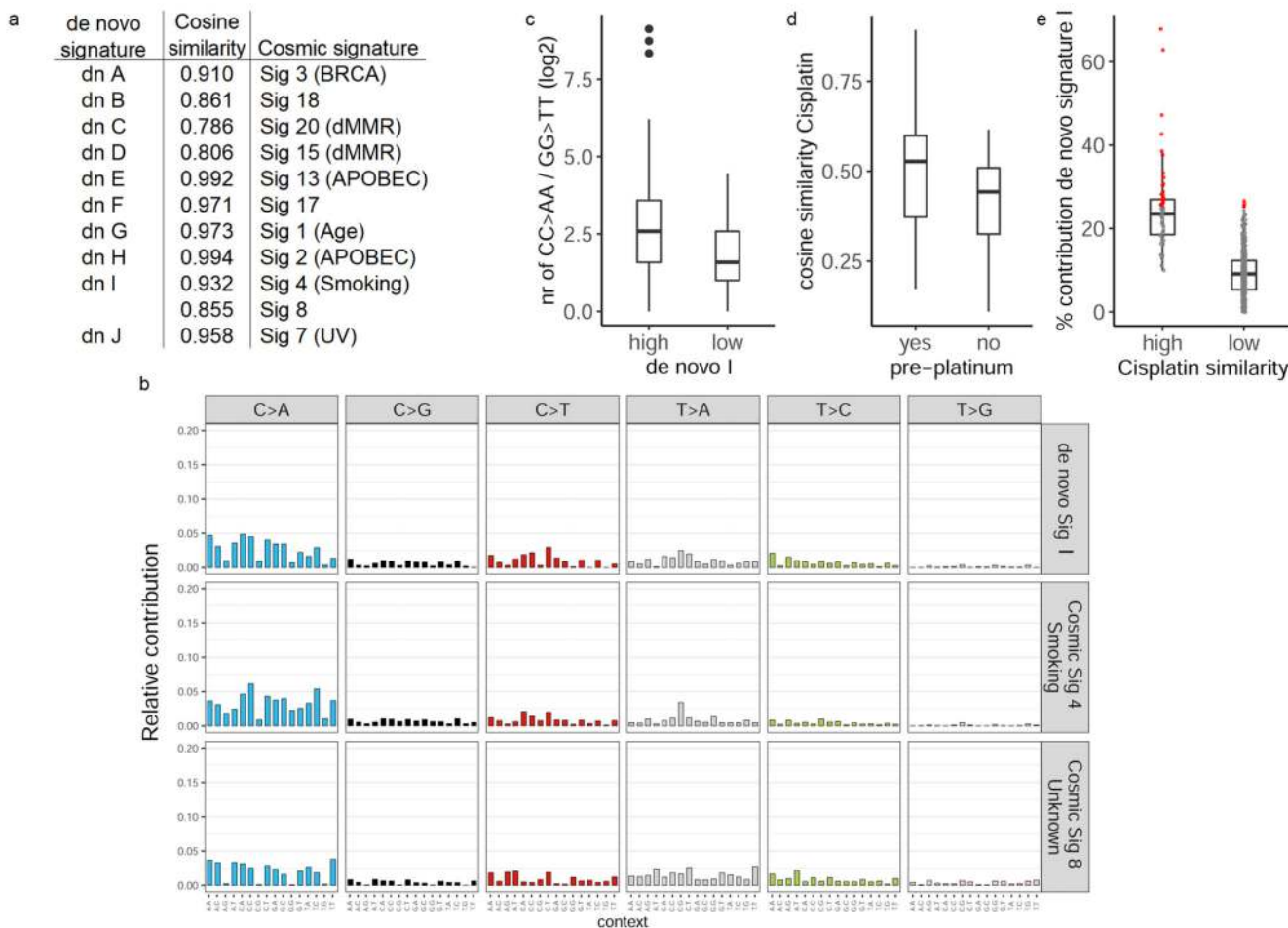
Figure 2. De novo signature I is associated with prior platinum-based chemotherapy

(a) De novo signature calling revealed 10 mutational processes operative in mBC. These de novo mutational signatures have high cosine similarities with known Cosmic signatures.

(b) The mutational spectrum of de novo signature I and Cosmic signatures 4 and 8.

(c) The number of CC>AA or GG>TT mutations in patients with a low (<10%) or high (≥10%) relative contribution of de novo signature I.

(d) Boxplot of the cosine similarity of the cisplatin signature defined by Boot et al.[19] and samples of patients who did or did not receive prior treatment with platinum-based chemotherapy.

(e) Boxplot of the contribution of mutational signature I and samples with a high (permutation p<0.05) or low (permutation p>0.05) similarity to the cisplatin signature of Boot et al. [19]
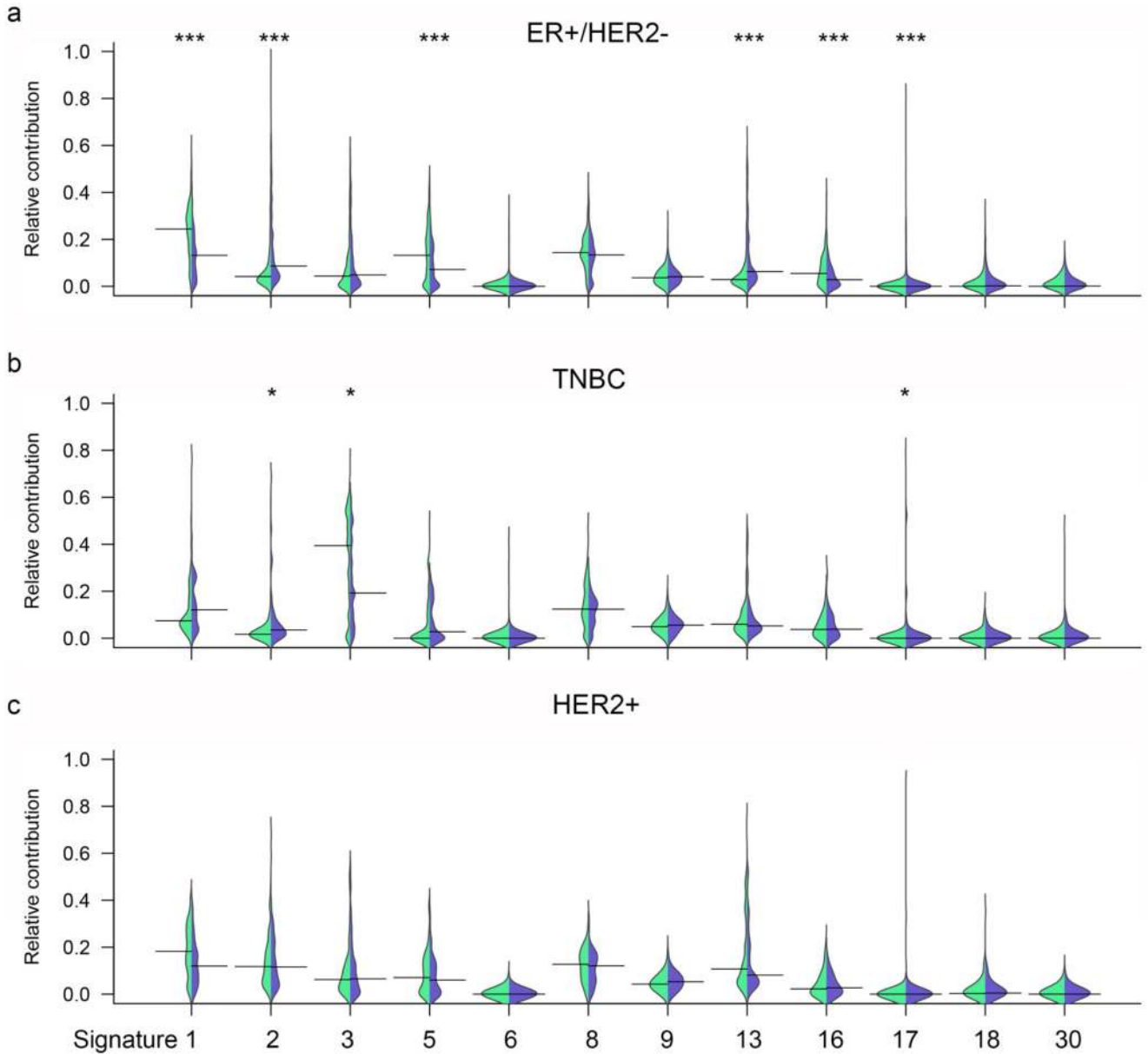
Figure 3. Mutational signatures: mBC versus primary BC

Bean plots showing the relative contribution of 12 Cosmic signatures which are dominantly contributing to the total number of SNVs in the metastatic cohort. Relative contributions were compared between mBC and primary BC samples from the BASIS cohort and shown per breast cancer subtype: ER+/HER2- (**a**), TNBC (**b**), HER2+ (**c**). Per graph, left of centre (green) indicate the distribution of primary tumours from the BASIS cohort, right of centre (purple) metastatic biopsy. Mann-Whitney U: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.
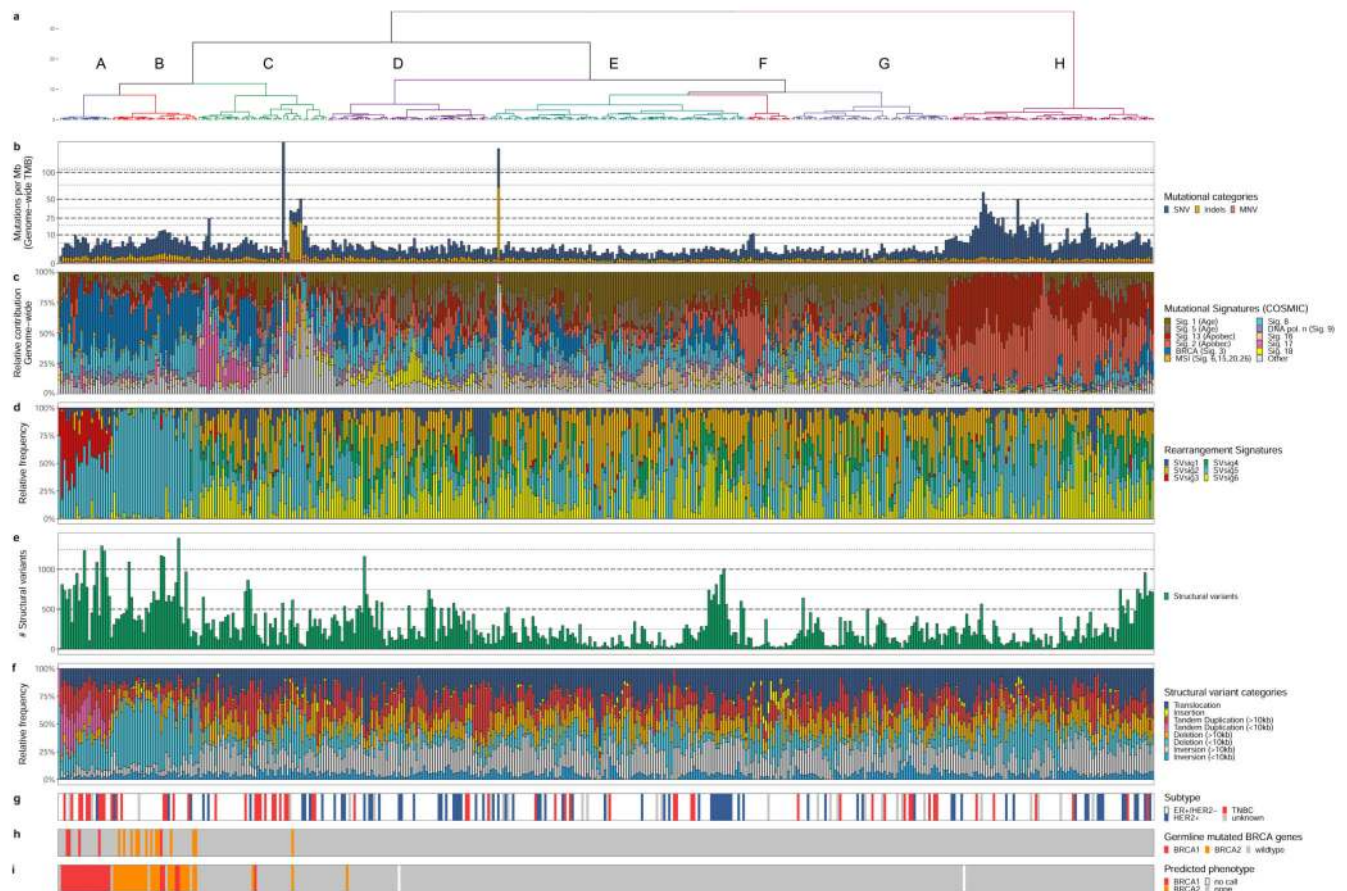
Figure 4. Unsupervised clustering reveals distinct genomic phenotypes in mBC

(a) Dendrogram of unsupervised clustering. The top eight clusters are denoted A to H. The Y-axis displays clustering distance (Pearson; ward.D).

(b) Number of genomic mutations per Mb (TMB) divided into mutational categories SNV, InDels and MNV. All genome-wide somatic mutations were taken in to consideration.

(c) Relative contribution of Cosmic mutational signatures.

(d) Relative contribution of rearrangement signatures.

(e) Absolute number of unique structural variants per sample.

(f) Relative frequency per structural variant category, tandem duplications, deletions and inversions are subdivided into <10kb and >10kb categories.

(g) Breast cancer subtype subdivided in ER+/HER2-, HER2+ and triple negative and unknown at time of analysis.

(h) Germline *BRCA*1/2 mutational status.

(i) HR-deficient score as assessed by CHORD. Predicted phenotypes *BRCA1* deficiency and *BRCA2* deficiency are depicted.

Figure 5. Actionability
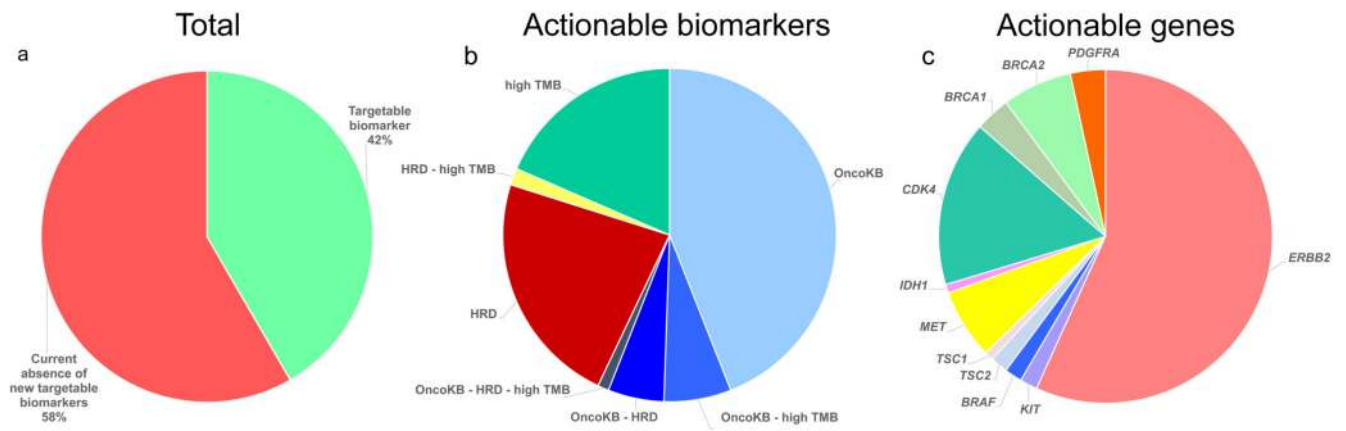
(a) Percentage of patients with and without an actionable target for treatment.

(b) Actionable targets by type: HR deficiency (HRD), high TMB (≥10 mutations/Mb) and/or targetable alterations for which an FDA approved drug is available (OncoKB).

(c) Genes indicated by OncoKB for which targeted drugs are FDA approved (*ERBB2* for breast cancer, all other genes for other cancer types).

Table 1

Frequency of affected driver genes and type of genomic alteration in mBC

	ER+/HER2-				HER2+				TNBC			
	Total	Gain	Deletion	SNV/InDels/SVs	Total	Gain	Deletion	SNV/InDels/SVs	Total	Gain	Deletion	SNV/InDels/SVs
TP53	88	0	0	88	45	0	0	45	46	0	0	46
ESR1	53	6	0	47	10	0	0	10	0	0	0	0
CDH1	34	0	0	34	3	0	0	3	5	0	1	4
MAP3K1	22	0	1	21	6	1	0	5	1	0	0	1
GATA3	38	1	0	37	9	2	0	7	2	2	0	0
CBFB	8	0	0	8	1	0	0	1	0	0	0	0
ARID1A	23	0	1	22	7	0	0	7	3	0	0	3
ERBB2	17	2	0	15	51	46	0	5	4	1	0	3
RUNX1	5	0	0	5	0	0	0	0	1	0	0	1
MAP2K4	22	0	6	16	8	2	2	4	1	0	0	1
GPS2	6	0	0	6	2	0	0	2	0	0	0	0
FOXA1	17	2	0	15	6	1	0	5	2	0	0	2
TBX3	21	1	1	19	2	0	0	2	0	0	0	0
NCOR1	17	1	0	16	5	2	0	3	4	0	0	4
PTEN	39	0	7	32	4	0	0	4	8	0	3	5
PIK3CA	128	0	0	128	35	1	0	34	11	0	0	11
KMT2C	30	1	0	29	9	0	0	9	6	0	0	6
RB1	10	0	1	9	3	0	1	2	5	0	0	5
AKT1	20	2	0	18	3	2	0	1	1	0	0	1
CDKN1B	12	1	1	10	0	0	0	0	4	2	0	2
NF1	31	2	1	28	6	3	0	3	5	1	0	4