An ICGC major achievement in breast cancer: a comprehensive catalog of mutations and mutational signatures

François Bertucci[1,2,3], Max Chaffanet[1], Daniel Birnbaum[1]

[1]Centre de Recherche en Cancérologie de Marseille, Laboratoire d'Oncologie Moléculaire, UMR1068 INSERM, UMR725 CNRS, Institut Paoli-Calmettes (IPC), Marseille, France; [2]Département d'Oncologie Médicale, IPC, Marseille, France; [3]Faculté de Médecine, Aix-Marseille Université, Marseille, France

Correspondence to: Prof. François Bertucci, MD, PhD. Department of Medical Oncology, Institut Paoli-Calmettes, 232 Bd de Ste-Marguerite, 13273 Marseille, France. Email: bertuccif@ipc.unicancer.fr.

Breast cancer is the first cause of cancer and death by cancer in women worldwide, with nearly half a million deaths each year (1). Despite progresses in screening and treatment, 15% of patients experience a metastatic relapse during follow-up and die from disease. Current systemic therapies, aimed at the eradication of microscopic and macroscopic metastases, are based on chemotherapy, hormone therapy, and more recently targeted therapies. These latter include several anti-HER2 drugs, the anti-VEGF drug bevacizumab, the mTOR inhibitor everolimus, and very recently the CDK4/6 inhibitor palbociclib. Other promising drugs in development target PARP, PD1, PIK3, AKT, and others (2). But clearly, a better understanding of mammary oncogenesis and the identification of new molecular targets remain crucial for better tackling the molecular heterogeneity and hardly predictable clinical behavior of the disease and improving prevention and treatment.

At the molecular level, breast cancer is a complex disease. The causes of the disease remain unclear: we know some of the likely causes (obesity, radiation exposure, accumulation of random mutations due to ageing), but in half of cases we do not know what causes the gene alterations. Accumulation and combination of those genetic and epigenetic alterations cause tumorigenesis, genetic instability, and acquisition of an increasingly invasive and resistant phenotype. This combinatorial origin, the heterogeneity of malignant cells and the variety of the host background create molecularly distinct tumors endowed with different therapeutic response and clinical outcome. Before the advent of high-throughput molecular analyses ("omics"), conventional biological techniques had successfully elucidated some mechanisms of mammary oncogenesis and identified key and clinically relevant features and alterations: expression of ER and PR, overexpression or amplification of *HER2/ERBB2* (the first successful therapeutic target defined by a genomic aberration), and mutations of *TP53*, *BRCA1* and *BRCA2*. Since the early 2000s, high-throughput molecular analyses ("omics") allowed improving our understanding of breast cancer. Gene expression profiling revealed the extent of molecular heterogeneity of disease (3) and identified biologically and clinically relevant subtypes (4) such as luminal A, luminal B, basal, and HER2-enriched. Multigene prognostic signatures [see (5) for review] were defined to improve treatment decision in early breast cancer and some of them are used in clinics and/or have been recently tested in prospective phase 3 clinical trials (6,7). More recently, next-generation sequencing (NGS) provided the opportunity of sequencing tumor DNA from tens of genes (targeted NGS) to the whole exome (sequencing of only the coding regions) and to the whole genome, allowing for the first time to define the complete repertoire of mutation

Page 2 of 4

Bertucci et al. Whole-genome sequencing of breast cancer

types such as base substitutions, small insertions/deletions (indels), rearrangements and copy number changes (8).

Briefly, two classes of somatic mutations exist in cancer cells. The cancer-causing mutations or "driver mutations", which confer clonal selective advantage to the cell and the sequential acquisition of which is required for oncogenesis, versus the "passenger mutations", stochastic and far more numerous, which result from the increased mutation rate of the cancer genome and, likely, do not have any functional consequence and do not contribute to oncogenesis (9). Identification of driver mutations has become the basis of "precision medicine". In 2012, five NGS studies of whole-exomes established the repertoire of driver gene mutations and copy number alterations in breast cancer (10-14). Nearly 800 cancers representative of all molecular subtypes were studied. Several recurrent new gene mutations were uncovered, often with low frequency inferior to 5%. Mutations in *TP53*, *PIK3CA*, *GATA3* and *PTEN* genes were among the most frequent. Strong inter-patient tumor heterogeneity was highlighted at the mutational level, in addition to the spatial and temporal intra-tumor heterogeneity previously revealed using NGS of paired breast cancer and metastatic samples (15-17). Even if some of these studies (10,11,13,18) already included whole-genome sequencing data in small series (15 to 46 samples), most of NGS data targeted protein-coding exons, leaving unexplored the mutations in untranslated intronic and intergenic regions—which however represent the large majority of genome—as well as the driver genome rearrangements and the mutational processes.

Different mutational processes exist in human cancers, including endogenous and exogenous carcinogen exposures, aberrant DNA editing, replication errors and defective DNA repair (18). Each process generates a signature on the genome that is a specific combination of mutation types, which may be related to the mechanisms of DNA damage and repair involved and may give clues about the causes of disease. Indeed, understanding the involved processes may help identify the etiology of mutations and define targets for prevention and treatment. The power of whole-genome sequencing to improve such understanding in breast cancer was first illustrated in 2012 in a small series of 21 samples (18). Mathematical modeling led to the extraction of five mutational signatures. In most tumors, more than one signature was represented in different proportions, meaning that more than one process had been operative and accounted for the heterogeneity of the 21 mutational patterns. Foci of localized substitution hypermutation,

termed kataegis, were detected, often in the vicinity of genomic rearrangements. Based upon the substitution types and sequence context involved in kataegis and signature B, the authors proposed that the APOBEC enzymes (cytidine deaminases) might be implicated. APOBEC family members, which help fight off viral infection, were then considered as fueling subclonal expansions and intratumour heterogeneity (19) and might represent a new class of therapeutic target aimed at limiting disease progression, adaptation, and drug resistance. A subsequent study of 4,938,362 mutations from 7,042 cancers from 30 different types revealed a landscape of 21 mutational processes and showed that APOBEC mutational signatures are enriched in tumor subclones (20).

Recently, breast cancer sequencing has made another leap forward. In a paper published in *Nature* (21), the same team extended these observations by analyzing the whole-genome sequences of 560 breast cancers. The objective was to identify all of the genetic changes that cause breast cancer. Molecular analyses also included RNA sequencing, microRNA expression profiling, array-comparative genomic hybridization and DNA methylation profiling for subsets of cases. In addition, this study sprouted other articles reporting on specific aspects or subsets of these data (22,23). Samples were from patients across USA, Europe and Asia. Cancer samples were pre-therapeutic primary tumors in the vast majority of cases, most often non-metastatic. Gynecological history (parity, age at first child, oral contraception exposure and duration, menopausal status and age at menopause, hormone replacement therapy and duration) and smoking history were available. All pathological subtypes were represented, largely dominated by the ductal subtype, as well as all molecular subtypes. *BRCA1* and *BRCA2* mutation and methylation statutes were available (90 tumors had inactivating *BRCA1* or *BRCA2* mutation or *BRCA1* promoter methylation), as well as the homologous-recombination (HR) deficiency score. Normal samples were blood, adjacent breast tissue or skin. This impressive series of the Breast Cancer Working group of the International Cancer Genome Consortium (ICGC) represents by far the largest cohort of cancer genomes of a single tissue type to date. Whole-genome sequencing gave for the first time in such a large series the view of the rest of the genome, allowing to address the questions unexplored by the whole-exome sequencing, namely the mutations in untranslated genome regions, the genome rearrangements and the mutational processes.

Sequencing of 560 cancer samples detected a total

of 3,479,652 somatic base substitutions, 371,993 small indels, and 77,695 rearrangements, with substantial variation between individual samples. Many thousands of mutations were present in each of the 560 cancer genomes, suggesting genomes profoundly remodeled. The search for substitutions and indels mutations in protein-coding regions concerned the 560 samples associated to 772 samples previously sequenced using whole-exome in other studies. A total of 1,628 alterations were retained, including substitutions (36%), indels (21%), rearrangements (10%) and copy number alterations (34%). These alterations concerned 93 candidate genes, of which at least one was altered in 95% of samples. The ten most often altered genes were *TP53*, *PIK3CA*, *MYC*, *CCND1*, *PTEN*, *ERBB2*, *ZNF703/FGFR1* locus, *GATA3*, *RB1* and *MAP3K1*, and accounted for 62% of driver alterations. This was not a surprise as they have been already described in breast cancer. Interestingly, five new candidate breast cancer genes were uncovered: *MED23*, *FOXP1*, *MLLT4*, *XBP1*, and *ZFP36L1*. MLLT4/afadin had been shown to be altered in breast cancer (24) and XBP1 to play a role in basal subtype (25). Some non-coding regions showed high mutation rates, but most had distinctive structural features likely causing hypermutability rather than being the focus of driver alterations. Most of the driver genes and the biological pathways involved represent potential therapeutic targets.

The comparison of the 560 sequences showed that breast cancer genomes are highly individual. Mathematical analysis of the 560 mutational profiles identified 20 mutational signatures, including 13 new, that might influence breast cancer development: 12 base-substitution signatures, two indel signatures and six rearrangement signatures. In each tumor, more than one signature was represented and their proportion varied between tumors. Among the 12 base-substitution signatures, two are correlated with patients' age at diagnosis, two are APOBEC-related, three are associated with mismatch-repair deficiency, two with HR deficiency, and three have unknown etiology. Three rearrangement signatures, characterized by tandem duplications or deletions, are associated with defective HR-based DNA repair as measured by the HR deficiency score for each tumor: one signature is associated with deficient BRCA1 function, another with deficient BRCA1 or BRCA2 function, and the third one has unknown cause. Hierarchical clustering of the 560 tumors based on the proportion of the six rearrangement signatures in each sample showed that tumors with BRCA1 or BRCA2 deficiency, known to

have an increased risk of developing breast and ovarian cancer, had whole-cancer genome profiles very different from most other tumors. However, some tumors without identifiable BRCA1/2 inactivating alteration co-segregated with BRCA1/2-inactivated tumors in the clustering. This observation could be used to classify patients more accurately for treatment. Indeed, because breast cancers with BRCA1/2 inactivation are particularly sensitive to some DNA damaging agents such as platinum salts and inhibitors of DNA repair such as PARP inhibitors (26,27), these co-segregating tumors might also be candidates for these treatments. From the point of view of pathogenesis, such similarity of mutational profiles between *BRCA1-* and *BRCA2*-mutated breast cancers suggests, better than the cellular phenotype (known to be different between these cancers), similarity regarding the underlying biological defect. Several important aspects such as the relationship between mutational processes and genome architecture (23), their impact on immune response (28) or the mechanisms of gene amplification (22) are reported in related articles. This cohort of 560 ICGC genomes is a treasure chest in which many other studies will dig.

Even if other infrequent molecular alterations remain to be uncovered this huge study expands our knowledge of the breast cancer genome and calls for rapid functional and clinical validation of the results because of the major potential applications. Better understanding the causes of breast cancer should improve disease prevention. Cell-based research methods are ongoing by the Nik-Zainal lab to characterize the mutational signatures in human cells through exposure to environmental compounds, editing of genes that code for DNA repair and replication, and through derivation of induced pluripotent stem cells from patients with DNA repair defects. The key mutations of 93 driver genes and the mutational signatures reflecting for example a HR deficiency provide new potential therapeutical targets and predictive biomarkers. In this context, the sequencing of samples from prospective clinical trials should be strongly encouraged for testing eventual correlations with therapeutic response. More generally, this study confirms the potential of whole-genome sequencing for more personalized cancer treatment and the importance of analyzing the non-protein-coding regions of the genome.

Page 4 of 4

Bertucci et al. Whole-genome sequencing of breast cancer

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87-108.
2. Chamberlin MD, Bernhardt EB, Miller TW. Clinical Implementation of Novel Targeted Therapeutics in Advanced Breast Cancer. J Cell Biochem 2016;117:2454-63.
3. Bertucci F, Houlgatte R, Nguyen C, et al. Gene expression profiling of cancer by use of DNA arrays: how far from the clinic? Lancet Oncol 2001;2:674-82.
4. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52.
5. Sabatier R, Gonçalves A, Bertucci F. Personalized medicine: present and future of breast cancer management. Crit Rev Oncol Hematol 2014;91:223-33.
6. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. J Clin Oncol 2008;26:721-8.
7. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. N Engl J Med 2016;375:717-29.
8. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 2010;11:685-96.
9. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009;458:719-24.
10. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486:395-9.
11. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 2012;486:405-9.
12. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature 2012;486:400-4.
13. Ellis MJ, Ding L, Shen D, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 2012;486:353-60.
14. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61-70.
15. Ding L, Ellis MJ, Li S, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 2010;464:999-1005.
16. Shah SP, Morin RD, Khattra J, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 2009;461:809-13.
17. Wang Y, Waters J, Leung ML, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature 2014;512:155-60.
18. Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational processes molding the genomes of 21 breast cancers. Cell 2012;149:979-93.
19. Swanton C, McGranahan N, Starrett GJ, et al. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. Cancer Discov 2015;5:704-12.
20. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415-21.
21. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534:47-54.
22. Ferrari A, Vincent-Salomon A, Pivot X, et al. A whole-genome sequence and transcriptome perspective on HER2-positive breast cancers. Nat Commun 2016;7:12222.
23. Morganella S, Alexandrov LB, Glodzik D, et al. The topography of mutational processes in breast cancer genomes. Nat Commun 2016;7:11383.
24. Fournier G, Cabaud O, Josselin E, et al. Loss of AF6/afadin, a marker of poor outcome in breast cancer, induces cell migration, invasiveness and tumor growth. Oncogene 2011;30:3862-74.
25. Chen X, Iliopoulos D, Zhang Q, et al. XBP1 promotes triple-negative breast cancer by controlling the HIF1α pathway. Nature 2014;508:103-7.
26. Fong PC, Boss DS, Yap TA, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med 2009;361:123-34.
27. Turner N, Tutt A, Ashworth A. Targeting the DNA repair defect of BRCA tumours. Curr Opin Pharmacol 2005;5:388-93.
28. Smid M, Rodríguez-González FG, Sieuwerts AM, et al. Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. Nat Commun 2016;7:12910.