Cancer Letters xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

Mini-review

Next generation analysis of breast cancer genomes for precision medicine

Maurizio Previati [a], Marco Manfrini [a], Marco Galasso [a], Carlotta Zerbinati [a], Jeff Palatini [a], Pierluigi Gasparini [b], Stefano Volinia [a,b,c,*]

[a] LTTA, Dept. of Morphology, Surgery and Experimental Medicine, Human Anatomy Branch, University of Ferrara, Italy
[b] Comprehensive Cancer Center, Wexner Medical Center, The Ohio State University, Columbus, OH, United States
[c] Biomedical Informatics, The Ohio State University, Columbus, OH, United States

ARTICLE INFO

Article history:
Received 8 May 2013
Received in revised form 10 July 2013
Accepted 14 July 2013
Available online xxxx

Keywords:
Breast cancer
Next generation sequencing
Epigenomics
Oncogenomics
miRNA

ABSTRACT

For many years breast cancer classification has been based on histology and immune-histochemistry. New techniques, more strictly related to cancer biology, partially succeeded in fractionating patients, correlated to survival and better predicted the patient response to therapy. Nowadays, great expectations arise from massive parallel or high throughput next generation sequencing. Cancer genomics has already revolutionized our knowledge of breast cancer molecular pathology, paving the way to the development of new and more effective clinical protocols. This review is focused on the most recent advances in the field of cancer genomics and epigenomics, including DNA alterations and driver gene mutations, gene fusions, DNA methylation and miRNA expression.

1. Introduction

Cancer is considered to originate from the contribution of both inherited and acquired genomic alterations, the first being dispensable, albeit strongly favorable for cancer induction. These genomic alterations are ultimately the result of the balance between the cell's ability to maintain the integrity of genetic information and the effects of the environmental efforts to alter it. Acquired DNA changes, in addition to inherited predisposition, can push the cell outside the frame which regulates its life, and determinate the timing and shaping of its growth, division, metabolism and death.

In particular, genetic alterations in the form of mutations, deletions and rearrangements can greatly affect the tumor suppressive activity of many genes, while insertions, duplications, translocations and gene fusions can push tumor promoting genes toward gains of function.

The analysis of DNA structure and gene expression using high throughput technologies such as array comparative genomic hybridization (aCGH) and DNA microarrays represented a big advancement in the study of breast cancer (BC). Previously, BC had been classified primarily by its histological appearance, followed by the identification of the breast cancer cells into normal breast tissue (grading) and on the basis of the tumor migration from the primitive site into lymph nodes and more distant tissues such as bone, liver, lung and brain. In addition, alterations in the structure and expression of hormone- or growth factors receptors were of pivotal importance for both subtype diagnosis and targeted therapy because they are strictly linked to the tumor physiology and offering a cue for pharmacological intervention. Although, hormone- or trastuzumab-based therapies can be successful, there are a high number of cases that do not positively respond to those therapies, indicating the coexistence of other factors driving therapy resistance. Gene expression profiling brought a novel classification, with different classes: HER2-enriched, basal-like, luminal A, luminal B, normal-like and claudin-low [1–3]. This new classification, more strictly related to cancer biology, succeeded in fractionating patients and better predicted the patient's response to therapy. Expression profiling and aCGH represented a substantial advancement, but also showed several limitations. Microarray expression data were based on the use of probes, which implied a semi-quantitative determination of RNA and a partial representation of the human genome, limited to selected genomic features chosen a priori.

Today, great expectations arise from a further technical advancement, represented by massive parallel or high throughput next generation sequencing (NGS). This technique is based on deep sequencing, which produces billions of short sequences at a time. NGS is quantitative and scan analyze the entire genome at

* Corresponding author at: LTTA, Dept. of Morphology, Surgery and Experimental Medicine, Human Anatomy Branch, University of Ferrara, Italy.
E-mail address: Stefano.Volinia@osumc.edu (S. Volinia).

0304-3835/$ - see front matter © 2013 Published by Elsevier Ireland Ltd.
http://dx.doi.org/10.1016/j.canlet.2013.07.018

base-pair (bp) resolution without the limitations of microarrays. Low input requirements allow genomes from a single cell to be resolved with a pre-amplification, effectively coupling molecular and cellular investigation. NGS benefits biomedical research in multiple ways by interrogating whole or partially targeted genomes, transcriptomes and epigenomes, including non-coding RNAs (ncRNAs) and microRNAs (miRs). Pairing high-throughput genotyping and copy number aberrations or somatic rearrangements at bp resolution, NGS promises to have significant utility in breast cancer research and treatment.

An example of this approach is offered in the study managed by the Breast Cancer International Cancer Genome Consortium (http://www.icgc.org), which has the goal of completely sequencing the genomes of 1500 breast cancers. Tumors would be then classified according to genetic aberrations, affected molecular networks, and ultimately, to the therapies they are responsive to. The 1000 Genomes Project (1000 Genomes Project, http://www.1000genomes.org) provides a catalogue of SNPs, copy number polymorphisms, and short insertion and deletion polymorphisms in the general population, thus helping with the discovery of pathogenic mutations in cancer. The Cancer Genome Atlas (TCGA) was initially set up in 2006, as a combined effort by the National Cancer Institute and the National Human Genome Research Institute, to characterize more than 20 tumor types, including breast lobular and ductal invasive cancer. As of July 2012 data have been available from the TCGA portal for more than 800 samples for breast cancer (The Cancer Genome Atlas Data Portal [http://tcga-data.nci.nih.gov/tcga/]) and the landmark paper has been recently published [4]. This and other genome-wide approaches are already providing a boost to our knowledge of BC molecular pathology, and will pave the way for the development of new and more effective clinical protocols.

1.1. Driver genes

Somatic mutations are a feature of breast cancer. A subset of these somatic mutations, known as driver mutations, can confer selective advantages to cancer cells, and drive the tumor through the multistep path of oncogenesis. The majority of somatic mutations instead are non-pathogenic passenger alterations which complicate the identification of causative events. Stephens and coauthors [5] sequenced the coding exons of 21,416 protein coding genes and 1664 microRNAs and examined copy number changes in 100 primary breast cancers, 79 of which were estrogen receptor positive (ER+). In the attempt to individuate breast cancer driver genes, they searched for non-random clustering of somatic mutations in each of the 21,416 protein-coding genes and sequenced a subset of genes. They found driver substitutions in cancer genes previously implicated in breast cancer development, including *AKT1*, *BRCA1*, *CDH1*, *GATA3*, *PIK3CA*, *PTEN*, *RB1* and *TP53*. Among others, *ARID1B*, *CASP8*, *MAP3K1*, *MAP3K13*, *NCOR1*, *SMARCD1* and *CDKN1B* were potentially recessive, while *AKT2* was probably an activated, dominantly-acting cancer gene. Banerji et al. [6] also sequenced the whole exome in tumors and matched-normal DNAs of 103 breast cancer patients from Mexico and Vietnam, together with whole-genome sequences of 22 breast cancer/normal pairs. *AKT1, PIK3CA* (mutations in these two genes were mutually exclusive in the cohort), *GATA3*, *TP53*, and *MAP3K1* were mutated with significant recurrence. In addition they discovered, for the first time in an epithelial cancer, some recurrent mutations in the *CBFB* transcription factor gene and deletions of its partner *RUNX1*. Oncogenic rearrangements of *RUNX1* or *CBFB* are common in acute myeloid leukaemia. In 2012, the TCGA network reported its comprehensive investigation describing NGS of BC tumors from more than 800 patients. They identified nearly all the mutations in genes previously implicated in breast cancer (*PIK3CA*, *PTEN*, *AKT1*, *TP53*, *GATA3*, *CDH1*, *RB1*, *MLL3*, *MAP3K1* and *CDKN1B*). In addition, the cancer genes group included *TBX3* (involved in mammary gland development); *CTCF*, *FOXA1, RUNX1* and *CBFB* (transcription factors already known to be involved in AML); *AFF2*, *CCND3, PIK3R1* and *NF1*; *PTPN22* and *PTPRD* (two protein tyrosine phosphatases, the latter involved in lung adenocarcinoma); *SF3B1* (a splicing factor involved in myelodysplastic syndromes and CLL) [4]. Shah and colleagues [7] precisely characterized all somatic coding mutations that occurred during the development and progression of individual cancers, by re-sequencing the DNA from a metastatic lobular breast cancer specimen (89% tumor cellularity) at approximately 40-fold haploid reference genome coverage (using a paired-end sequencing method). They identified 32 somatic non-synonymous coding mutations in the metastasis, and measured the frequency of these somatic mutations in DNA from the primary tumor of the same patient, which arose 9 years earlier. Five of the 32 mutations (in ABCB11, HAUS3, SLC24A4, SNX4 and PALB2) were prevalent in the DNA of the primary tumor removed at diagnosis 9 years earlier, six (in KIF1C, USP28, MYH8, MORC1, KIAA1468 and RNASEH2A) were present at lower frequencies (1–13%), while 19 were not detected in the primary tumor.

Studies of ER+ BC genomes in patients enrolled in neoadjuvant therapy trials, offered the possibility to correlate cancer phenotype and somatic alterations. Ellis and colleagues [8] performed NGS on biopsies from two neoadjuvant aromatase inhibitor clinical trials. Eighteen mutated genes were identified, some of them already known as mutated in breast cancer (*PIK3CA*, *TP53*, *GATA3*, *CDH1*, *RB1*, *MLL3*, *MAP3K1* and *CDKN1B*) as well as *TBX3*, *RUNX1*, *LDLRAP1*, *STNM2*, *MYH9*, *AGTR2*, *STMN2*, *SF3B1* and *CBFB*. Ellis and colleagues correlated mutations with clinical data in 240 additional cases and identified pathways relevant to aromatase inhibitor response. Several pathways were enriched in the aromatase-inhibitor-resistant group, including the *TP53* signaling pathway, DNA replication, and mismatch repair. The relationship between *MAP3K1* mutation, luminal A subtype, low tumor grade and low Ki67 proliferation index was also identified. On this basis, patients with luminal A tumors and *MAP3K1* mutation could be reasonably treated with neoadjuvant aromatase inhibitor, whereas tumors with *TP53* mutations, mostly aromatase inhibitor resistant, would need to be treated differently. These findings showed that an integrated approach, based on gene expression and pathways analysis, can supplement important clinical decisions. *BRCA1* and *BRCA2* are two major BReast CAncer susceptibility genes that encode proteins involved in a common pathway for the protection of genome from double-strand DNA damage [9]. Germ-line mutations of either *BRCA1* or *BRCA2* can lead to breast and ovarian cancer syndrome (HBOC), an autosomal dominant hereditary disease. The individuals with HBOC syndrome have a lifetime risk of developing breast cancer at 50–80% and 30–50% for ovarian cancer [10]. Few reports and data on the prevalence of *BRCA1* and *BRCA2* germ-line mutations in sporadic breast and ovarian cancer patients have been available until recently. De Leeneer et al. investigated the complete coding region of *BRCA1/2* in 193 breast and/or ovarian cancer patients, without evidence for a family history of the disease [11]. In 17 (8.8%), a deleterious germ-line *BRCA1/2* mutation was identified. The authors identified a *BRCA1* mutation in a patient, which was absent in both parents. This was considered a potential de novo mutation [12]. The mutation was detected in 43% of the reads in the patient at 157-fold coverage, but was absent in the maternal and paternal samples both covered at over 1000 folds, consistent with a de novo occurrence. A prevalence of a variant in 43% of all reads is consistent with heterozygosity as heterozygous mutations can be present in 25–60% of the reads [13]. This result may reflect the criteria for mutation analysis of *BRCA1/2*; usually genetic testing is often based on the family history, and for sporadic patients, they are often limited to patients with early age at onset, bilateral-

Please cite this article in press as: M. Previati et al., Next generation analysis of breast cancer genomes for precision medicine, Cancer Lett. (2013), http://dx.doi.org/10.1016/j.canlet.2013.07.018