Europe PMC Funders Group
Author Manuscript
Curr Opin Oncol. Author manuscript; available in PMC 2013 July 17.

Published in final edited form as:
Curr Opin Oncol. 2012 November ; 24(6): 597–604. doi:10.1097/CCO.0b013e328359554e.

Next generation sequencing in breast cancer: first take home messages

Christine Desmedt[1,*], **Thierry Voet**[2,3], **Christos Sotiriou**[1], and **Peter J Campbell**[3]

[1]Breast Cancer Translational Laboratory, Université Libre de Bruxelles, Jules Bordet Institute, Boulevard de Waterloo 121, 1000 Brussels, Belgium

[2]Department of Human Genetics, University of Leuven, Herestraat 49 box 602, B-3000 Leuven, Belgium

[3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus,Hinxton, Cambridgeshire, CB10 1SA

Abstract

Purpose of the review—We are currently on the threshold of a revolution in breast cancer research thanks to the emergence of novel technologies based on next generation sequencing (NGS). In this review, we will describe the different sequencing technologies and platforms, and summarize the main findings from the latest sequencing papers in breast cancer.

Recent findings—First, the sequencing of a few hundreds of breast tumors has revealed new cancer genes. Although these were not *frequently* mutated, mutated genes from different patients could be grouped into the deregulation of similar pathways. Second, NGS allowed further exploration of intratumor heterogeneity and revealed that although subclonal mutations were present in all tumors, there was always a dominant clone which comprised at least 50% of the tumor cells. Finally, tumor-specific DNA rearrangements could be detected in the patient's plasma, suggesting that NGS could be used to personalize the monitoring of the disease.

Summary—The application of NGS to breast cancer has been associated with tremendous advances and promises for increasing the understanding of the disease. However, there still remain many unanswered questions, such as for example the role of structural changes of tumor genomes in cancer progression and treatment response/resistance.

Keywords

breast cancer; next generation sequencing; cancer genes; driver/passenger mutations; tumor heterogeneity

Introduction

A decade ago, 'Sanger' sequencing of the human reference genome and the subsequent development of microarrays revolutionized breast cancer research. Microarrays enabled the interrogation of cancer genomes for DNA-copy number changes and loss-of-heterozygosity events, as well as entire cancer transcriptomes for changes in gene expression level. This in turn led to a better understanding of the biology of breast cancer, to the proposal of a new

*Corresponding author: Christine Desmedt, Institut Jules Bordet, Breast Cancer Translational Research Laboratory JC Heuson, 125 Bld de Waterloo, 1000 Brussels – Belgium, Tel: +32 2 541 31 07/ Fax: +32 2 538 08 58, christine.desmedt@bordet.be.

Conflict of interest
The authors declare no conflict of interest.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

molecular classification system for the disease, to a refinement of breast cancer prognosis and to the identification of predictive markers for response to commonly administrated anti-cancer treatments [1]. Currently, we are at the doorstep of yet another revolution in cancer research based around next generation DNA sequencing (NGS).

Next generation sequencing, a new avenue for cancer research

The NGS or second-generation sequencing technologies have the advantage that enormous numbers of sequencing reactions can be performed in parallel in a time- and cost-effective manner. Genome sequencing projects that previously required decades of Sanger 'chain-termination' sequencing can now be accomplished in a matter of days at 10.000 to 100.000 times cost reduction. Although this review does not aim at providing a detailed description of the different sequencing technologies and platforms (reviewed in [2, 3]), Table 1 provides a concise summary of current prominent commercially available NGS-platforms.

Although each of the second-generation systems is underpinned by different sequencing chemistries of immobilized amplified DNA-molecules (Table 1), they are all able to characterize not only a genome for the full spectrum of DNA mutations but also to dissect the transcriptome of a cancer to unprecedented digital accuracy. Apart from mere nucleotide sequence information at a specific locus, the millions of short DNA sequences that can be generated from a DNA- or RNA-sample can also be applied for DNA-copy number or gene expression level profiling, respectively. This is achieved by counting and normalizing the amount of short sequence reads that map back to selected parts of the human reference genome. Furthermore, it is possible to sequence both ends of each linear DNA-molecule on these second-generation sequencers, referred to as paired-end sequencing (Table 1), and hence, following mapping of the reads to the reference genome, to unveil the linkage between both ends of each DNA-molecule. This enables haplotype reconstruction and consequently the detection of submicroscopic intra- and inter-chromosomal structural rearrangements in a cancer genome, but also the identification of fusion transcripts and splice variants in a cancer transcriptome. Besides full genome or transcriptome sequencing, specific genomic regions can be selected for targeted sequencing by using DNA pull-down or amplicon approaches. For example, all coding exons present in a genome can be cherry-picked from a DNA-pool and sequenced, known as exome sequencing. Since all exons only represent about 1% of the genome, multiple exomes can be sequenced simultaneously in a multiplexed reaction at the same cost and time-span as required for one full genome sequence. Not only can DNA and RNA samples extracted from many cells be used for NGS-analyses, but also amplification products of a single-cell genome or transcriptome can be sequenced to address specific biological questions [4, 5]. However, the interpretation of single-cell sequencing data is complicated by various amplification biases introduced in the cell's DNA or RNA sample and requires dedicated approaches to sift these amplification artefacts from true genetic changes.

While second-generation sequencing methods are commonly applied in cancer genetics studies nowadays, third-generation sequencing methodologies (e.g. Pacific Biosciences, Oxford Nanopore), which are single-molecule sequencers not requiring clonal amplification of DNA-templates, have become available and may prove useful in complementing second-generation sequencing data (Table 1). Apart from the avoidance of DNA-amplification bias and sequencing by polymerase speed, other main advantages include the generation of extremely long kb-sized reads and their putative application in direct measurement of base modifications like cytosine methylation and other DNA-adducts. Their main disadvantages include an order of magnitude higher sequencing error rate when compared to second generation sequencers and still large amounts of DNA or RNA are required for efficient sequence library construction. Last but not least, several vendors of second-generation

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

sequencers have fabricated miniaturized bench-top versions with accelerated sequencing processes at the cost of output (Table 1).

Altogether, the advances in NGS technology have opened up a new avenue for breast cancer research. In this review paper, we will aim at giving an overview of the recent advances and promises of NGS in dissecting the biology of breast cancer, in identifying potential new prognostic and predictive markers, as well as reviewing important large-scale initiatives.

Identification of new cancer genes

By sequencing both the tumor and the germline DNA, NGS may reveal the somatic genetic alterations in a cancer genome. These somatic genetic changes can be classified in two ways. First, according to the role they are playing in the disease, as driver or passenger mutation. Driver mutations contribute to tumor development [6, 7], while the remaining somatic mutations, called passenger mutations, do not and may be the product of the genomic instability of the tumor. However, it is important to note that the distinction between driver and passenger mutations is dynamic and can change throughout the course of the disease. For example a passenger mutation could become a driver mutation after anti-cancer treatment by providing clonal advantage to the resistant clone. Secondly, the mutations can be classified according to the type of genetic alteration: nucleotide substitution mutations, small insertions and deletions (also called “indels”), copy number gains and losses, chromosomal rearrangements (large insertions and deletions, tandem duplications, inversions and translocations), as well as nucleic acids from foreign origin (such as viruses) [8].

Recently, four papers reported results on the largest breast cancer datasets ever published: Shah et al. performed RNA and DNA sequencing on 80 and 65 cases of triple negative breast cancers respectively (i.e. cancers which do not express the estrogen, progesterone and HER2 receptors), Ellis et al. performed DNA sequencing on 77 breast tumors (46 by whole genome sequencing and 31 by whole exome sequencing) which do express the estrogen receptor, and Stephens et al. and Banerji et al. performed DNA sequencing on 100 (all whole exome sequencing) and 108 breast tumors (17 by whole-exome and whole-genome sequencing, 5 by whole-genome sequencing alone, and 86 by whole-exome sequencing alone), respectively, from all breast cancer subtypes [9-12]. The common messages from these papers are the following. First, although new cancer (driver) genes were identified, there were no new *frequently* mutated cancer genes. Indeed, unlike the known cancer genes P53 and PIK3CA which are mutated in >30% of breast cancer patients (P53 preferentially in ER-negative and PIK3CA preferentially in ER-positive tumors), most newly identified cancer genes are mutated in less than 10% of the patients. Second, there is a very large genetic diversity among different breast tumors. Stephens et al. for example demonstrated that among the 100 breast cancers they investigated there were 73 different combination possibilities of mutated cancer genes [12]. However, although breast tumors are heterogeneous with regard to mutated genes, the mutated genes can be grouped into the deregulation of similar pathways. For instance, Stephens et al. demonstrated that 6 cancer genes were acting in the same JUN kinase pathway [12], whereas Shah et al. observed that pathways involving P53, chromatin remodeling, PIK3 and ERBB signaling were over-represented in the mutated genes [11]. This means that although the tumors are genetically different, some could be phenotypically similar due to mutations in the same pathway, which is very important in terms of treatment. Third, in some tumors, no driver mutation was obvious. This may suggest that a different mechanism is driving tumor development in these neoplasms, such as for example DNA methylation. Fourth, some mutations might be associated with the response/resitsance to anti-cancer treatment. Ellis et al. for example

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

suggested that mutant GATA3 correlated with suppression of proliferation upon aromatase inhibitor treatment [10].

Altogether, these findings increase our knowledge of the disease and the discovery of driver events; however functional analyses should complement these structural analyses of the cancer genomes to confirm the biological relevance of the findings.

RNA-seq can do more than quantifying gene expression

Although the vast majority of somatic mutations occur in the tumor epithelial cells, it is known that the tumor microenvironment plays an important role in tumor growth and response or resistance to various anti-cancer treatments. In this respect, quantifying global gene expression in the tumor cells and their surrounding non-epithelial cells can reveal important information on breast cancer prognosis and prediction of the efficacy of chemotherapy, as exemplified by numerous studies which interrogated the transcriptome using gene expression arrays[13, 14]. Although less straightforward in terms of experimental procedures and bio-informatics analyses, RNA-seq (the sequencing of RNA through NGS) has the advantage of having a higher dynamic range compared to gene expression arrays [15]. In other words, RNA-seq can accurately quantify genes expressed either at low or very high levels, whereas analysis of these transcripts are very limited using gene expression arrays. However, to our knowledge, no large breast cancer dataset has been published so far to demonstrate the added value for RNA quantification using RNA-seq compared to microarrays.

As already alluded to in the introduction, RNA-seq can do more than merely providing a digital readout of the gene expression levels. It also has the potential to detect expressed mutations, fusion genes, RNA editing events and splice variants of transcripts. Interestingly, by comparing exome and RNA sequencing data, Shah et al. observed that only 36% of the mutations were expressed [16]. RNA editing and alternative splicing are mechanisms which modify the transfer of genetic information from the nuclear genome to proteins and represent an important additional source of the biological complexity of a cancer. RNA-editing events represent mutations which are present only in the RNA and not in the DNA. Most of these RNA editing events are mediated by the ADAR enzyme. The first paper to report the presence of RNA editing using RNA-seq in breast cancer was by Shah and colleagues, which identified two new RNA-editing events in lobular breast cancer [17]. Interestingly, the authors also observed that the ADAR enzyme belonged to the top 5% of genes expressed. Alternative splicing on the other hand, represents a key molecular event in the gene expression process. It allows for the synthesis of different products from the same gene. Genes contain both exons and introns. While introns must be removed from the initial transcribed RNA (or pre-RNA) by splicing, alternative splicing events can also result in the removal of exons from the pre-RNA. By selectively retaining different sets of exons within the spliced transcript, cells can thus produce multiple isoforms from a single gene, and, if subsequently translated, multiple proteins. Using RNA-seq, Shapiro and colleagues showed that epithelial to mesenchymal transition (EMT), a process which can enhance the migration and invasion of cancer cells, can be driven by alternative splicing [18]. Their initial findings were based on a human mammary epithelial cell line expressing the EMT-associated transcription factor TWIST. However, when they further examined a panel of breast cancer cell lines they observed that aggressive and metastatic cells had more EMT-associated alternative splicing events than cells that were poorly metastatic. These data suggest that alternative splicing may have an important role in directing EMT and also that alternative splicing could maybe be used in the future as prognostic markers for breast cancer.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Although very few papers have used RNA-seq to investigate breast cancer until now, we believe that application of RNA-seq to breast cancers has great promises to deliver better understanding of the complexity of breast cancer disease.

Intratumor heterogeneity

Although intratumor genetic heterogeneity is known for years, to date, only few groups have the expertise to explore intratumor genetic heterogeneity in primary breast cancer by NGS-analysis of a DNA-sample extracted from many tumor cells [11, 19, 20]. Nik Zainal et al. recently charted the subclonal architecture of 21 breast cancers by NGS. Important messages emerged from this study. First, in all cases studied, they observed both clonal and subclonal mutations, i.e. mutations present in all and a subset of tumor cells respectively. Interestingly, clonal mutations, which are the ones appearing early in tumor development, comprised mutations in many cancer genes such as PIK3CA and P53, as well as all HER2, MYC and CCND1 amplifications and the somatic loss of the wild-type BRCA1 and BRCA2 alleles in hereditary breast cancers. Surprisingly, although many cancer gene mutations were found to be fully clonal, still a tremendous amount of mutations occur sub-clonally. Indeed, for most cases the observed number of subclonal mutations is more than the number of fully clonal mutations. Second, in all tumors a dominant clone was found which comprised 50 to 95% of the tumor cells. The authors postulated that it is the expansion of this dominant clone which triggered diagnosis. Lastly, they identified different mutational signatures which could differentiate early versus late mutations. For example, early mutations were enriched in C>T mutations. Also, they identified a mutational signature that was associated with inherited breast cancer. In their study, Shah et al. investigated the clonal heterogeneity of triple negative breast cancer. They demonstrated that these tumors vary widely in their clonal frequencies, with some tumors having only very few subclones whereas other tumors consisted of more than 10 subclones. In contrast with the results observed by Nik-Zainal et al., although mutations in the P53, PIK3CA and PTEN cancer genes were present in most samples in the largest clone, these mutations were not always fully clonal, i.e. present in all tumor cells.

Altogether, these works depict a dynamic picture of on-going genetic evolution and clonal expansion in breast cancer. Studies which interrogate the implications of the intratumor heterogeneity in terms of response/resistance to different anti-cancer treatment and in terms of metastatic progression are ongoing and should provide both biologically and clinically relevant information in the near future.

Personalized monitoring of the disease

Breast tumors have at least one DNA rearrangement that can be detected by low coverage DNA-sequencing and that is not present in normal DNA [21]. We and others have shown that it is possible to find the genomic rearrangements in breast tumors and further use PCR assays to detect these rearrangements in the plasma, with an excellent sensitivity since the assay could detect one single copy of the tumor genome in the plasma [22, 23]. The advantages of this approach are considerable since it could be used as an accurate non-invasive indicator of the tumor burden and provide an early indication of treatment efficacy or disease recurrence. From a technical perspective the advantage of using rearrangements as a clinical assay is that it should not have any false positive rearrangements. Indeed, rearrangements bring together two DNA sequences which normally should be far away from each other. Since the detection of these rearrangements is based on PCR, there will be no PCR amplification in non-rearranged, normal samples with carefully chosen PCR primers. One limitation however of this approach, is that the clonal selection or evolution of the disease (see previous section) would modify the landscape of genomic rearrangements

observed in the peripheral blood. This limitation could be (at least partially) circumvented by interrogating several rearrangements. Although these results are still preliminary since they are based on a very few number of breast cancer patients (4 patients), we are currently investigating this promising approach in a prospective study involving 100 breast cancer patients. The first results are expected by 2013.

Large scale sequencing resources

Two large worldwide collaborative efforts, the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) are cataloguing the genomic landscape of thousands of high quality cancer samples across many cancer types, including breast cancer [24]. All the data generated in these projects is already or otherwise will be made publicly available. However, since NGS data may provide the means to identify patients, most raw and pre-processed data are and will be subjected to controlled-access policy. This means that access to the data is only granted to qualified researchers with a specific scientific proposal which is compliant with the TCGA/ICGC access policy. However, it should be noted that given the size of NGS data, downloading raw data from these sources remains a technically and logistically challenging task which requires an adapted network infrastructure. The main characteristics of TCGA and ICGC are compared in Table 2.

In addition to the two above mentioned initiatives, the open access COSMIC database (Catalog of Somatic Mutations in Cancer) provides the most comprehensive source of curated analyzed somatic mutation data in cancers. This database has been developed and is maintained by the Cancer Genome Project at the Wellcome Trust Sanger Institute (Hinxton, UK) and contains >233,000 somatic mutations of >698,000 tumor samples (www.sanger.ac.uk/genetics/CGP/cosmic;[25]).

Conclusions and perspectives

The application of NGS to cancer research has been linked with incredible challenges and promises. Challenges involve the collection of high quality tumor (and matched normal) samples, the inherent challenges of tumor samples (contamination with other cell types, tumor heterogeneity, aneuploidy), the interpretation of the data, which requires adapted informatics infrastructure and computational tools and the necessary budget. Indeed, although prices are dropping, generating NGS data remains a relatively expensive exercise for most laboratories. Also, an important consideration is that the actual cost for NGS experiments most of the time does not include the costs for data storage and analysis. In that context, large-scale international collaborative efforts, such as TCGA and ICGC, will definitely help to make a gigantic step forward in the characterization of the breast cancer by providing precious sequencing resources. Nonetheless, small-scale focused sequencing efforts with a well-defined scientific or clinical question will undoubtedly also further increase our knowledge of breast cancer.

There are different exiting possibilities for integrating NGS in clinical practice. Targeted genetic sequencing could for example help to detect mutations in genes of therapeutic importance. However, pre-requisites to this are the ability to rapidly screen numerous genes at a small price. This has become possible now with the recent development of bench-top NGS platforms (see [26]for an excellent overview, Table 1). Also, an important technical challenge will be to use DNA from formalin fixed paraffin embedded (FFPE) tissues. Until now, NGS have mainly been working on frozen tissues, which unfortunately are not available for each single patient in contrast to FFPE tissues, especially in smaller hospitals. Although different some results suggest that using FFPE DNA for NGS should be feasible [27], caution is still warranted because sequencing accuracy may depend on the quality of FFPE DNA, which can be highly variable, and also the ability of NGS to distinguish

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

sequence artifacts from low frequency mutations from these samples has to be validated. Besides the possibility of using NGS to perform a rapid gene screen, NGS could also have a role in clinical practice for the personalized monitoring of the disease through the detection of tumor DNA in the peripheral blood of the patients, as described above.

Altogether, the promises for NGS in breast cancer are tremendous for increasing our understanding of the disease, to identify new treatment targets and to move towards genome-informed personalized medicine (Figure 1). However, one should always keep in mind that although characterization of structural changes in the cancer genomes by NGS will provide important pieces of knowledge, epigenetic changes, contributions from the tumor microenvironment and germline genetic variation will also have to be taken into account to have the full picture of the disease.

Acknowledgments

CD and CS are supported by the Fonds National de la Recherche Scientifique (FNRS), Les Amis de l'Institut Bordet and the MEDIC Foundation.

REFERENCES

1. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. N Engl J Med. 2009; 360:790–800. [PubMed: 19228622]
2. Chin L, Hahn WC, Getz G, et al. Making sense of cancer genomic data. Genes Dev. 2011; 25:534–55. [PubMed: 21406553]
3. Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev. 2011; 32:177–95. [PubMed: 22147957]
4. Xu X, Hou Y, Yin X, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell. 2012; 148:886–95. [PubMed: 22385958] 95.* In these two independent papers by the same group, a novel approach was used: sequencing whole exomes from single cells.
5. Hou Y, Song L, Zhu P, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell. 2012; 148:873–85. [PubMed: 22385957] *Cfr comment above.
6. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000; 100:57–70. [PubMed: 10647931]
7. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–74. [PubMed: 21376230] **This paper describes the common traits ("hallmarks") that govern the transformation of normal cells to cancer cells.
8. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009; 458:719–24. [PubMed: 19360079]
9. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012; 486:405–9. [PubMed: 22722202] **This paper which reports DNA sequencing results from 108 breast cancers of diverse subtypes confirmed known recurrent somatic mutations but identified new ones as well as a recurrent fusion gene enriched in triple-negative breast cancer.
10. Ellis MJ, Ding L, Shen D, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature. 2012; 486:353–60. [PubMed: 22722193] ** This paper reports DNA sequencing of 77 oestrogen-receptor-positive tumours of patients treated with aromatase inhibitors and shows that distinct phenotypes are associated with specific patterns of somatic mutations
11. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012**This paper interrogates the genomes of over 100 cases of triple-negative breast cancers to reveal widely differing mutational spectra within individual cases of this disease.
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. [PubMed: 22722201] **This paper reveals that

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

infrequently mutated genes account for almost half of all driving genetic events in breast cancer and that these cancers are caused by combinations of at least 40 different driver mutations.

13. Ignatiadis M, Singhal SK, Desmedt C, et al. Gene modules and response to neoadjuvant chemotherapy in breast cancer subtypes: a pooled analysis. J Clin Oncol. 2012; 30:1996–2004. [PubMed: 22508827]
14. Desmedt C, Haibe-Kains B, Wirapati P, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res. 2008; 14:5158–65. [PubMed: 18698033]
15. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10:57–63. [PubMed: 19015660]
16. Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012
17. 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. [PubMed: 19812674]
18. Shapiro IM, Cheng AW, Flytzanis NC, et al. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. PLoS Genet. 2011; 7:e1002218. [PubMed: 21876675]
19. Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational Processes Molding the Genomes of 21 Breast Cancers. Cell. 2012; 149:979–993. [PubMed: 22608084] **The authors found that different mutational processes act at different times in the lifespan of a breast tumour and also identified one particular mutational signature which was indicative of inherited breast cancer.
20. Nik-Zainal S, Van Loo P, Wedge DC, et al. The life history of 21 breast cancers. Cell. 2012; 149:994–1007. [PubMed: 22608083] ** The authors charted the subclonal architecture of 21 breast cancer by whole-genome sequencing and reconstructed the life histories of these breast cancers, depicting a dynamic picture of on-going genetic evolution and clonal expansion in breast cancer.
21. Stephens PJ, McBride DJ, Lin ML, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. 2009; 462:1005–10. [PubMed: 20033038]
22. Leary RJ, Kinde I, Diehl F, et al. Development of personalized tumor biomarkers using massively parallel sequencing. Sci Transl Med. 2010; 2:20ra14.
23. McBride DJ, Orpana AK, Sotiriou C, et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. Genes Chromosomes Cancer. 2010; 49:1062–9. [PubMed: 20725990]
24. Hudson TJ, Anderson W, Artez A, et al. International network of cancer genome projects. Nature. 2010; 464:993–8. [PubMed: 20393554]
25. Forbes SA, Tang G, Bindal N, et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res. 2009; 38:D652–7. [PubMed: 19906727]
26. Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012; 30:562.
27. Lipson D, Capelletti M, Yelensky R, et al. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nat Med. 2012; 18:382–4. [PubMed: 22327622]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

KEY BULLET POINTS

- The sequencing of a few hundreds of breast tumors has revealed new cancer genes which are not frequently mutated.
- The exploration of the intratumor heterogeneity using next generation sequencing revealed subclonal mutations in all tumors and a dominant clone comprising at least 50% of the tumor cells.
- Next generation sequencing could be used to personalize the monitoring of the disease by interrogating tumor-specific DNA rearrangements in the patient's plasma.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

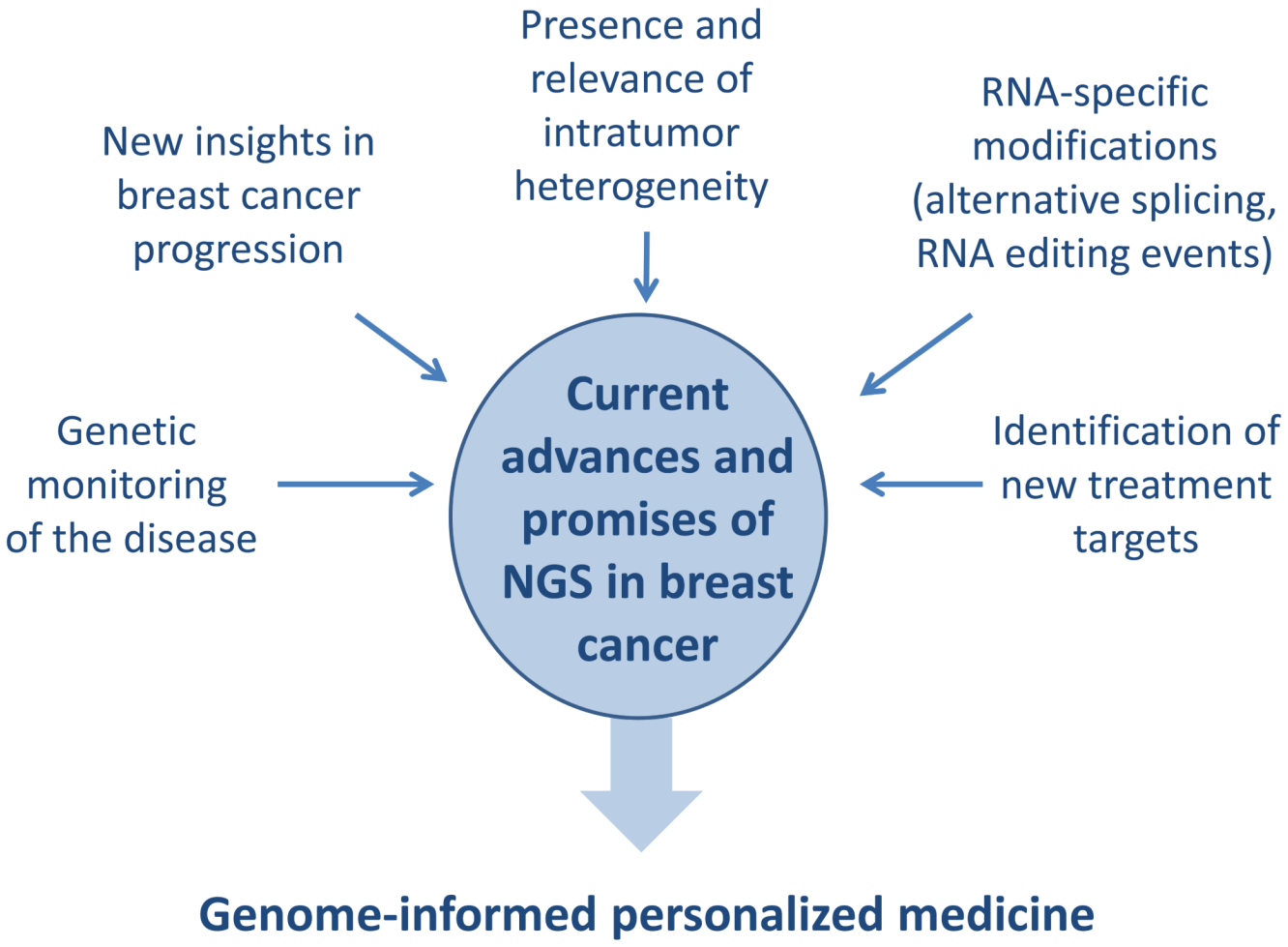

Figure 1.
Current advances and promises of next generation sequencing in breast cancer

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Current prominent NGS platforms

Company	Platform	Sequencing	Amplification	Read Length	Max Output	Run time	Pros/Cons
Illumina	HiSeq 2500 (2nd GS)	Sequence by synthesis using fluorescently labelled nucleotides	Bridge PCR	100bp × 100bp (150bp × 150bp in fast run mode)	600Gb (120Gb in fast run mode)	11days (27 hours in fast run mode)	Pro: High output, platform has a fast run mode Con: Short reads
Illumina	HiSeq 2000 (2nd GS)	Sequence by synthesis using fluorescently labelled nucleotides	Bridge PCR	100bp × 100bp	600Gb	11days	Pro: Output Con: Short reads
Life Tech	SOLiD 5500 (2nd GS)	**S**upported **O**ligonucleotide **L**igation and **D**etection	Emulsion PCR	75bp × 35bp (paired-end), 60bp × 60bp (mate-paired)	~80Gb	<7days	Pro: Low sequencing error-rate Con: Short reads, low output
Roche	454 (2nd GS)	Pyrosequencing	Emulsion PCR	Paired end reads averaging 140-200 bases	~0.6Gb	23 hours	Pro: Longer reads, fast Con: homopolymer-associated indel errors are frequent, low output
Pacific Biosciences	PacBio RS (3rd GS)	Single-molecule sequencing in SMRT cells containing 150K ZMW-wells	none	Average lengths of 3 kb (max 15 kb)	~60Mb/SMRT cell	< 2 hours	Pro: Very long reads, very fast Con: High sequencing error rate, low output
Illumina	MiSeq (*)	Sequence by synthesis using fluorescently labelled nucleotides	Bridge PCR	150bp × 150bp (expected: 250bp × 250bp)	2Gb (expected: 7Gb)	24 hours	Pro**: Low error-rate, high output per run Con**: shorter read lengths
Life Tech	Ion Torrent PGM (*)	Ion semiconductor sequencing	Emulsion PCR	Up to 400bp reads (expected 200bp × 200bp paired-end)	Up to 1Gb	2 hours	Pro**: Long reads Con**: homopolymer-associated indel errors are frequent
Roche	454 GS Junior (*)	Pyrosequencing	Emulsion PCR	~400bp	~35Mb	10 hours	Pro**: Long reads Con**: homopolymer-associated indel errors are frequent, low output

* + bench-top fast sequencing platforms in last 3 rows

** Pro/cons amongst bench-top sequencers

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 2

Details for the breast cancer initiatives of The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC)*

	The Cancer Genome Atlas (TCGA)	The International Cancer Genome Consortium (ICGC)
Breast cancer types	Ductal (collection completed) and Lobular (collection ongoing)	There are 6 breast cancer projects. Three of those (Triple-negative, estrogen-receptor positive and HER2-positive breast cancer) are run together under the umbrella of the Breast Cancer Working Group led by the Wellcome Trust Sanger Institute. The remaining projects are led by Mexico, South Korea and the USA (TCGA).
Accepted source of germline DNA	Blood, saliva, skin	Blood, adjacent normal tissue
Central pathology review	Yes	Yes
Pathology criteria for breast tumors	Minimum 60% tumor cells Maximum 20% necrosis	Minimum 70% tumor cells Maximum 20% necrosis
Data sharing	Open access for the interpreted and summarized data and controlled access for the raw and processed data through the dbGAP (www.ncbi.nlm.nih.gov/gap)	Open access for the interpreted and summarized data and controlled access for the raw and processed data through the Data Access Compliance Office (www.icgc.org/daco)
Web page	http://cancergenome.nih.gov/	**www.icgc.org/**

*It should be mentioned that the relationship between TCGA and ICGC is synergistic to ensure that the datasets produced by the two initiatives are compatible; TCGA is actually also part of ICGC.