Review Article · Übersichtsarbeit

Breast Care

Breast Care 2013;8:401–406
DOI: 10.1159/000357461

Published online: December 10, 2013

High-Throughput Gene Expression and Mutation Profiling: Current Methods and Future Perspectives

Thomas Karn

Department of Gynecology and Obstetrics, Goethe University Frankfurt am Main, Germany

Keywords
Gene expression profiling · Next-generation sequencing · Breast cancer, molecular subtypes · Cancer genome · Tumor heterogeneity

Summary
Following the completion of the human genome sequence at the beginning of the new millennium, a series of high-throughput methods have changed cancer research. Using these techniques, global analysis such as expression profiling could be carried out on a genomic scale. In breast cancer they led to the classification of the intrinsic subtypes, and the development of several prognostic and predictive 'genomic tests' for patient stratification. During the last 2 years we have faced a similar dramatic revolution with the introduction of next generation sequencing (NGS). These techniques allow sequencing of the complete human exome or whole genome with a cost reduction in the order of 10,000–100,000 fold. Consequently, the number of known cancer genome sequences exploded with more than 6,000 samples, published between 2011 and 2013. These studies have led to important and surprising discoveries both for basic cancer research and clinical applications. They relate to understanding the development of cancer as well as the heterogeneity of the disease, and how to use this information to guide the development and application of therapies. Although it is foreseeable that the sequencing surveys of neoplasms will soon conclude, their introduction into clinical practice is just beginning.

Schlüsselwörter
Genexpressionsprofiling · «Next-Generation-Sequencing»-Techniken · Mammakarzinom, molekulare Subtypen · Tumorgenom · Tumorheterogenität

Zusammenfassung
Nach Abschluss des humanen Genomprojektes zu Beginn des Jahrtausends führten globale Untersuchungen mittels «Genomics»-Methoden zu vielen neuen Ergebnissen in der Krebsforschung. In der Brustkrebsforschung gehörten dazu die Definition der sogenannten «intrinsischen Subtypen» des Mammakarzinoms sowie die Entwicklung einer ganzen Reihe von molekularen Tests zur Prädiktion von Prognose und Therapieansprechen. In den letzten Jahren hat die Zunahme der Sequenzierungsgeschwindigkeit bei gleichzeitiger immenser Kostenreduktion durch «Next-Generation-Sequencing»-Techniken zu einer weiteren Umwälzung geführt. Die Zahl der komplett entschlüsselten Tumorgenome explodierte in den Jahren 2011–2013 auf mehr als 6000 Proben. Die Untersuchungen führten zu einer Vielzahl von wichtigen und auch überraschenden Entdeckungen in Bereich der Grundlagenforschung aber auch zu möglichen klinischen Anwendungen. So kann «Whole-Genome-Sequencing» einen bedeutenden Beitrag zur Tumorklassifikation liefern und wird wahrscheinlich in überschaubarer Zeit Eingang in die klinische Anwendung finden.

Introduction

The sequencing of the human genome was completed at the end of the millennium. Since then, translational cancer research has been affected by several tidal waves caused by the advancements in high-throughput genomic techniques [1]. The first such wave presented technologies that transformed mRNA expression analysis, comparative genomic hybridization (CGH), and single nucleotide polymorphism (SNP) analyses, pushing these forward into high-throughput microarray

KARGER
Fax +49 761 4 52 07 14
Information@Karger.com
www.karger.com

© 2013 S. Karger GmbH, Freiburg
1661-3791/13/0086-0401$38.00/0

Accessible online at:
www.karger.com/brc

Dr. Thomas Karn
Department of Gynecology and Obstetrics
Goethe University Frankfurt am Main
Theodor-Stern-Kai 7, 60590 Frankfurt, Germany
t.karn@em.uni-frankfurt.de

formats. These array platforms allowed investigations on a global genomic scale, resulting in an enormous flood of data and new bioinformatic and statistical challenges [2, 3]. Subsequently, sequencing techniques were also revolutionized, accompanied by an even more tremendous data torrent during the previous 2 years. In this review we focus on gene expression profiling by microarray methods and mutation profiling through next generation sequencing (NGS). Since a large number of reviews on gene expression signatures in breast cancer and the corresponding genomic tests have recently appeared [4–7], the emphasis here will be on the newer developments in cancer genome sequencing.

High Throughput Gene Expression Profiling – the First Wave

One decade ago microarray profiling emerged as a new method allowing the global analysis of gene expression in tumors. The application of these 'transcriptome' techniques improved our understanding of breast cancer biology, leading to new prognostic information [8–10] and a refined molecular system of tumor classification [11, 12].

Results from Unsupervised Methods – Molecular Portraits of Breast Cancer

Unsupervised methods for the analysis of gene expression datasets led to the definition of the intrinsic subtypes of breast cancer [4, 11]. The basic classification of these intrinsic subtypes encompasses the estrogen receptor (ER)-negative 'basal-like' and 'HER2-like' subgroups as well as two different ER-positive 'luminal A' and 'luminal B' subtypes (table 1). The intrinsic subtypes differ both in their prognosis [12] and their response to systemic therapies [5–7]. The basal-like and HER2-like subtypes, for example, display a poor prognosis but an increased response to chemotherapy [13–15]. However, the additional clinical value of molecular classification is limited by its close correspondence with the status of ER, PR, and HER2, along with tumor grade [4]. Recent results of unsupervised analyses demonstrate further heterogeneity within the intrinsic subtypes, showing that additional clinically relevant subgroups can be stratified within the basal-like subtype [16–18].

Supervised Analysis and the Development of Clinical Assays

On the other hand, supervised methods using information from outcome variables such as follow-up or treatment response were applied to identify prognostic or predictive gene signatures [5]. Some of these signatures (e.g. MammaPrint, OncotypeDX, Genomic Grade Index, and EndoPredict) have made their way to clinically usable prognostic tests, as recently reviewed in detail [4–7, 19]. Most of these assays have mainly prognostic value, and can be applied only partially for prediction of treatment response. They appear to quantify mainly tumor grade and proliferation and displayed similar performance when tested in the same datasets. An important limitation of all these assays is that they assign the high-risk category to almost all ER-negative patients. The gene signatures are most useful in ER-positive patients and the expression of cell cycle- and proliferation-related genes drive the performance of most of them [4]. Since conventional chemotherapy targets the proliferating fraction of tumors, the finding that those signatures also predict benefit from conventional chemotherapy regimens was not unexpected. Predictors of response to specific chemotherapy agents derived from empirical analyses of response to neoadjuvant therapy were less successful [4]. Although many predictive signatures were developed, at present there is no validated and commercially available gene signature to predict response to a specific therapeutic agent [5]. The realization that the different subtypes of breast cancer are fundamentally distinct in their mRNA expression profiles led several groups to investigate these subgroups separately, leading to so-called second generation signatures. These studies identified immune infiltration as an important component for prognosis and prediction in ER-negative subtypes [17, 20–24].

Further Array Platforms

Subsequent to the first microarrays for analyzing the transcriptome, additional arrays have followed for studying the microRNA profiles, arrayCGH for studying copy number variations (CNV), and SNP arrays allowing copy number analysis and genome-wide association studies (GWAS). GWAS have emerged as an important tool for discovering regions in the human genome associated with cancer risk, and their current status has recently been reviewed [25, 26]. Other high-throughput methods for epigenetics, proteomics, and metabolomics still lack the comprehensiveness, usability and robustness of the DNA- and RNA-based technologies.

Table 1. Molecular subtypes of breast cancer from gene expression profiling [5]

Molecular subtype	Basal-like	HER2-like	Luminal A	Luminal B
ER-positive	0–19%	25–59%	91–100%	91–100%
PR-positive	6–13%	25–30%	70–74%	41–53%
HER2-positive	9–13%	66–71%	8–11%	15–24%
Ki-67	high	high	low	high
Grade G3	88–93%	55–89%	13–30%	41–62%
Prognosis	poor	poor	good	intermediate/
Chemotherapy benefit	high	intermediate	low	poor intermediate

High-Throughput Mutation Profiling – the Next Wave

As with microarrays at the beginning of the millennium, during the last 2 years we have faced a similar revolution through the introduction of NGS techniques [1, 27]. While just 2 genome sequences of breast cancers were published in the years 2009 [28] and 2010 [29], the sample sizes of those studies literally exploded during 2011 and 2012. The total number of published cancer genomes surpassed 6,000 cases in 2013 and still counting (fig. 1). The number of breast cancer genomes reached more than 900 samples in 2012, and there are many more samples in the waiting queue. For example, > 16,000 cancer samples had been genome/exome sequenced by late 2012 at one institution alone (Broad institute) [30]. This dramatic increase in sequencing throughput has been achieved through NGS technologies. In the last 10 years we have witnessed a tremendous increase in sequencing speed paralleled by costs falling dramatically by 10,000–100,000 fold compared to the classical Sanger method. The National Human Genome Research Institute (NHGRI) has tracked the costs associated with DNA sequencing performed at the sequencing centers funded by the Institute. The results presented in figure 2 give the cost of sequencing a human genome, demonstrating the remarkable improvements in DNA sequencing technologies and data-production pipelines in recent years [31]. Genome sequencing projects previously requiring decades can now be accomplished within days [27].

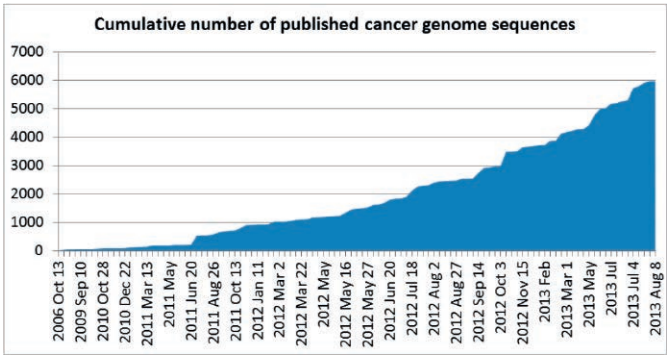

Fig. 1. Increase in the number of published cancer genomes. The cumulative number of cancer genome sequences has been plotted for 109 publications reporting new whole genome or whole exome sequences of cancer samples, sorted according to their publishing date between October 2006 and August 2013.

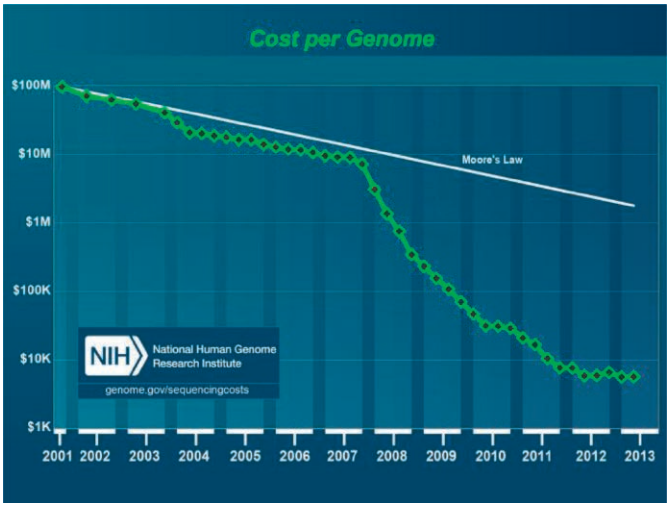

Fig. 2. Decrease in the cost of sequencing a human genome. A logarithmic scale is used on the Y axis of this graph to show the decrease in the sequencing costs since 2001. A straight line represents the hypothetical data reflecting Moore's Law, which describes a long-term trend in the computer hardware industry that involves the doubling of 'compute power' every 2 years. However, there is a sudden and profound outpacing of Moore's Law beginning in January 2008, when the sequencing centers changed from using Sanger-based (dideoxy chain termination sequencing) to 'next-generation' DNA sequencing technologies [31].

The conceptual breakthrough of NGS relates to the abandonment of gel electrophoresis and a massive increase in parallelization. All methods are principally based on clonal amplification and immobilization of numerous short fragments of DNA on surfaces. Afterwards all these spatially separated clones are sequenced simultaneously in situ. The subsequent analysis is performed by digitally counting the short sequence reads after aligning them to a reference genome sequence. In addition to the detection of somatic mutations and germ-line variants, this quantitative analysis also allows the determination of DNA copy number alterations [32]. Moreover, by increasing the coverage of the analysis to several 100× ('ultra deep sequencing'), it is possible to detect small amounts of mutated DNA from mixed populations of cells. Sequencing of RNA through NGS (RNA-seq) even allows transcriptome expression profiling through digital counting of transcripts, the identification of expressed fusion genes, and altered splicing events [27].

Applications of NGS

Several applications of NGS in cancer research exist (table 2). The substantial increase in sequencing power facilitates studies on germ-line mutations in cancer, e.g. the identification of predisposing factors in hereditary forms of cancer or pharmacogenomics studies and GWAS approaches [25, 26]. In this review, however, we focus mainly on somatic mutations in cancer. Whole genome or whole exome sequencing provides the opportunity to identify new 'driver' mutations in cancer. Such mutations are defined as conferring a selective growth advantage (which has been estimated in the range of 0.4%) [33]. Although this definition is simple in principle, it is more difficult to identify which somatic mutations are drivers and which are 'passengers' [34]. Passenger mutations encompass, for example, all those neutral mutations that have been accumulated in the founder cell during normal development before the oncogenic event occurred [35]. These passengers seem to account for roughly half of the mutations found in a typical breast cancer [36], and a large part of the remaining mutations would also be passengers acquired after the tumor-initiating event [33]. Individual genes can contain both driver

mutations and passenger mutations. Thus, the term 'Mut-driver genes' has been coined to categorize genes suspected of increasing the selective growth advantage of tumor cells [34]. Since the rate of mutation varies dramatically both between tumors and also within the human genome [37], statistical methods based on mutation frequencies can only prioritize genes as belonging to this group. However, a simple classification has been proposed requiring a prevalence (> 20%) of either missense mutations at recurrent positions for an oncogene or inactivating mutations for a tumor suppressor gene. This '20/20 rule' is far surpassed by all well-documented cancer genes, and identified a total of 125 Mut-driver genes from 294,881 mutations of genome sequencing of 3,284 tumors [34]. Although it is conceivable that further cancer genome sequencing will unveil additional Mut-driver genes, the current data suggest that a plateau is being reached [30, 34]. Even when many of the identified Mut-driver genes had been known previously, cancer genome sequencing projects have uncovered several surprises (table 3). Nearly half of the newly discovered genes are implicated in chromatin regulation. Furthermore, alterations were detected in genes implicated in mRNA splicing and genes related to tumor metabolism. Will we get further driver events from cancer genome sequencing? In fact we do not know, but it has been estimated that for each tumor type about 2,000 samples are needed to assemble the catalogue of coding mutations present in at least 2% of tumors of a given type. Moreover, many more cancer drivers may be lurking in the so-called 'dark matter' of the genome. Today's 'tier 1' projects focus on the 1% of the genome containing coding sequences (the exome). What currently cannot be readily interpreted are 'tiers 2–4', the noncoding (e.g. regulatory) regions, including copy number alterations and chromosomal rearrangements [30, 34, 38].

Table 2. Applications of next-generation sequencing in cancer research

Germline mutations
GWAS
Hereditary forms of cancer
Predisposing of protective mutations
Pharmacogenomics
Somatic mutations
Identification of new 'driver' mutations for cancer
Tumor classification
Prediction of response based on mutations conferring drug resistance
Disease monitoring
Analysis of inter- and intra-tumoral heterogeneity
RNA-sequencing
Digital expression profiling
Identification of fusion genes
Altered splicing

The complete mutational profile of cancer samples may be used to classify tumors based on driver mutations and copy number alterations with the goal, for example, of refining the current breast cancer classification [32, 39, 40]. In addition, identified mutations could allow prediction of response to therapy [41], with the ultimate aim of personalized cancer diagnostics [42]. Moreover, personalized cancer sequencing could lead to specific individual genomic markers that are suited for highly sensitive non-invasive disease monitoring [43].

One important result from cancer genome sequencing studies is the enormous heterogeneity both between and within tumors. First, the number of genes with frequent alter-

Table 3. Genes identified from cancer genome sequencing and corresponding cellular processes [30]

Cancer genes discovered or extended to new cancer types	Cellular process
EGFR, ERBB2, MET, ALK, JAK2, RET, ROS, FGFR1, FGFR2, PDGFRA, CRKL	RTK signaling
KRAS, NRAS, BRAF, MAP2K1	MAPK signaling (oncogenes)
NF1	MAPK signaling (tumor suppressor gene, TSG)
PIK3CA, AKT1, AKT3	PI3K signaling (oncogenes)
PTEN, PIK3R1	PI3K signaling (TSG)
NOTCH1, NOTCH2, NOTCH3	Notch signaling (oncogene or TSG)
STK11,TSC1,TSC2	TOR signaling (TSG)
APC, CTNNB1	Wnt/β-catenin signaling (TSG)
SMAD2, SMAD4, TGFBR2	TGF-β signaling (TSG)
MYD88	NF-κB signaling
RAC1, RAC2, CDC42, KEAP1, MAP3K1, MAP2K4, ROBO1, ROBO2, SLIT2, SEMA3A, SEMA3E, ELMO1, DOCK2	Other signaling
DNMT3A	Epigenetics DNA methylation
TET2	Epigenetics DNA hydroxymethylation
MLL, MLL2, MLL3, EZH2, NSD1, NSD3	Chromatin histone methyltransferases
JARID1A, UTX, KDM5A, KDM5C	Chromatin histone demethylases
CREBP, EP300	Chromatin histone acetyltransferases
SMARCA1, SMARCA4, ARID1A, ARID2, ARID1B, PBRM1	Chromatin SWI/SNF complex
CHD1, CHD2, CHD4	Chromatin other
MITF, NKX2–1, SOX-2, ERG, ETV1, CDX2	Transcription factor lineage dependency or oncogene
MYC, RUNX1, GATA3, FOXA1, NKX3.1, SOX9, NFE2L2, MED12	Transcription factor other
SF3B1, U2AF1, SFRS1, SFRS7, SF3A1, ZRSR2, SRSF2, U2AF2, PRPF40B	Splicing
DIS3	RNA abundance
SPOP, FBXW7, WWP1, FAM46C, XBP1	Translation/protein homeostasis/ubiquitination
IDH1, IDH2	Metabolism
TP53, MDM2, MSH, MLH, ATM	Genome integrity
TERT promoter mutations	Telomere stability
CCND1, CCNE1	Cell cycle (oncogene)
CDKN2A, CDKN2B, CDKN1B	Cell cycle (TSG)
MCL1, BCL2A1, BCL2L1	Apoptosis regulation

ations in cancers is rather low [44]. In breast cancer, only 3 genes (PIK3A, TP53, GATA3) were recurrently found mutated in at least 10% of all patients [45–47]. 3 additional genes were found mutated in at least 5% of the patients. However, the majority of the 20,000 detected somatic mutations in 500 patients were only sporadically observed [40]. It appears that there are virtually no two tumors with a similar mutational pattern. Nevertheless, further analysis of the genetic changes seem to suggest that different mutational events may be grouped to common oncogenic pathways, somewhat reducing this complexity [30, 34, 40, 48].

On the other hand, a large degree of intra-tumoral heterogeneity has been detected through 'ultra deep sequencing'. This highly redundant sequencing of the genome allows the relative proportion of specifically mutated DNA molecules to be counted digitally, thereby establishing the frequency of different genetically distinct subclones within the tumor. Such sequencing has been extensively performed for hematological cancers [35, 49] but data on breast cancer are also available [46, 50]. Based on a 'molecular clock' of mutations, it is possible to calculate an ancestral tree of the subclones to describe their evolution within the tumor. The obtained results corroborate data from earlier studies [36, 51] and demonstrate waves of subclonal evolution within the tumor, adding further complexity to the disease [43, 46, 52–54].

The large amount of sequencing data on cancer genomes also spurred recent studies on the mechanisms of cancer development. The patterns of 'mutation signatures' observed in cancer genomes have given new clues on the mutational processes shaping human tumors [37, 55–57]. They also highlighted skewed mutation frequencies between different genes and regions in the genome, underscoring the critical role of the bioinformatic algorithms used to identify mutated genes [37].

Future Perspectives and Challenges

Given the long-tailed distribution of cancer genes and the variable background mutation rates, the currently needed studies to finish compiling the catalogue of significantly mutated genes will require thousands of tumor-normal pairs. However, the current speed of progress suggests that accomplishing this goal will probably not take long [38, 30]. Moreover, the analyses must expand beyond the exome and cover the whole genome, including translocations and the transcriptome. Combining genomics, transcriptomics, and epigenomics has already proven to lead to important insights in breast cancer [39, 58]. For example, in the breast cancer TCGA study both SNP and CGH arrays, DNA methylation analysis and transcriptome, proteome, and microRNA expression analyses were included [40, 44]. Many of the 'comprehensive' insights of this study were enabled through integrative analysis across platforms, and a new genome-driven integrated classification of breast cancer, which includes DNA copy number changes, has been proposed [32]. As a next step, genome sequencing must expand beyond primary tumors to preneoplastic lesions, metastases from different sites, and tumors that show different types of responses to therapies [45, 49]. Carefully considered sample selection according to a specific clinical question will be the critical point in such studies.

New treatment options may be developed for novel targets identified by whole genome sequencing, even though this can be challenging. First, many of the identified genes have no enzymatic activity, in contrast to e.g. protein kinases (against which all the currently clinically approved drugs that target products of genetically altered genes are directed) [34]. Secondly inactivating mutations predominate over activating mutations in most common solid tumors. The observed heterogeneity of tumors presents us with an additional challenge. It is expected that mutations conferring resistance to any targeted drug would pre-exist in at least 1 cell of a typical tumor. Thus, simultaneous treatment with 2 or more drugs will likely be necessary to circumvent the otherwise-inevitable development of resistance (as learned for example from HIV therapy) [34]. However, the paucity of individual oncogene alterations presents a considerable problem to this strategy. Nevertheless, it is certain that treatment according to pathways could lead to an appropriate solution [30, 34, 48].

There are also different exciting possibilities for integrating NGS into clinical practice. One approach will be targeted re-sequencing of mutations with therapeutic importance. Bench-top versions of genome sequencers have already arrived in the clinical laboratory [59]. Even though the throughput of these systems has not been adapted for high coverage genome sequencing projects, they are well suited, for example, for targeted re-sequencing of gene panels. Even ultra deep sequencing of such panels can be performed to detect rare subclones, coping with the problem of tumor heterogeneity. Thus, personalized tumor profiling should be feasible in a clinical setting, ultimately translating genome sequencing from bench to bedside [1, 38]. Pilot studies have already shown that it is possible to analyze the complete genome of patients' tumors in a cost-effective and clinically relevant time frame [42]. Whole genome sequencing data further suggest that each breast cancer has at least one DNA rearrangement. Therefore, personal tumor sequencing could be used for the development of highly sensitive PCR assays for an individual tumor, allowing personal monitoring of disease through specific detection of tumor DNA in peripheral blood [43, 60]. In summary, even though we should avoid unjustified over optimism with respect to a new technology, which will undoubtedly also come with new problems; it is an exciting time in the fascinating field of cancer genomics. Hopefully, some of its promises will ultimately make their way to the patient.

Disclosure Statement

The author declares no conflict of interest.

References

1 Green ED, Guyer MS, Manolio TA, et al.: Charting a course for genomic medicine from base pairs to bedside. Nature 2011;470:204–213.
2 Simon R, Radmacher MD, Dobbin K, et al.: Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst 2003;95:14–18.
3 Ioannidis JPA, Khoury MJ: Improving validation practices in 'omics' research. Science 2011;334:1230–1232.
4 Sotiriou C, Pusztai L: Gene-expression signatures in breast cancer. N Engl J Med 2009;360:790–800.
5 Reis-Filho JS, Pusztai L: Gene expression profiling in breast cancer: classification, prognostication, and prediction. Lancet 2011;378:1812–1823.
6 Kaufmann M, Pusztai L: Use of standard markers and incorporation of molecular markers into breast cancer therapy: Consensus recommendations from an International Expert Panel. Cancer 2011;117:1575–1582.
7 Prat A, Ellis MJ, Perou CM: Practical implications of gene-expression-based assays for breast oncologists. Nat Rev Clin Oncol 2012;9:48–57.
8 van ,t Veer LJ, Dai H, van de Vijver MJ, et al.: Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–536.
9 Ahr A, Karn T, Solbach C, et al.: Identification of high risk breast-cancer patients by gene expression profiling. Lancet 2002;359:131–132.
10 van de Vijver MJ, He YD, van't Veer LJ, et al.R: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
11 Perou CM, Sørlie T, Eisen MB, et al.: Molecular portraits of human breast tumours. Nature 2000;406:747–752.
12 Sørlie T, Perou CM, Tibshirani R, et al.: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001;98:10869–10874.
13 Rouzier R, Perou CM, Symmans WF, et al.: Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 2005;11:5678–5685.
14 Rody A, Karn T, Solbach C, et al.: The erbB2+ cluster of the intrinsic gene set predicts tumor response of breast cancer patients receiving neoadjuvant chemotherapy with docetaxel, doxorubicin and cyclophosphamide within the GEPARTRIO trial. Breast 2007;16:235–240.
15 von Minckwitz G, Untch M, Blohmer J, et al.: Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol 2012;30:1796–1804.
16 Prat A, Parker JS, Karginova O, et al.: Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 2010;12:R68.
17 Rody A, Karn T, Liedtke C, et al.: A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Res 2011;13:R97.
18 Lehmann BD, Bauer JA, Chen X, et al.: Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121:2750–2767.
19 Azim HA, Michiels S, Zagouri F, et al.: Utility of prognostic genomic tests in breast cancer practice: The IMPAKT 2012 Working Group Consensus Statement. Ann Oncol 2013;24:647–654.
20 Teschendorff AE, Miremadi A, Pinder SE, et al.: An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol 2007;8:R157.
21 Schmidt M, Böhm D, von Törne C, et al.: The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res 2008;68:5405–5413.
22 Rody A, Holtrich U, Pusztai L, et al.: T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. Breast Cancer Res 2009;11:R15.
23 Bianchini G, Qi Y, Alvarez RH, et al.: Molecular anatomy of breast cancer stroma and its prognostic value in estrogen receptor-positive and -negative cancers. J Clin Oncol 2010;28:4316–4323.
24 Denkert C, Loibl S, Noske A, et al.: Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J Clin Oncol 2010;28:105–113.
25 Chung CC, Chanock SJ: Current status of genome-wide association studies in cancer. Hum Genet 2011;130:59–78.
26 Hindorff LA, Gillanders EM, Manolio TA: Genetic architecture of cancer and other complex diseases: Lessons learned and future directions. Carcinogenesis 2011;32:945–954.
27 Desmedt C, Voet T, Sotiriou C, et al.: Next-generation sequencing in breast cancer: First take home messages. Curr Opin Oncol 2012;24:597–604.
28 Shah SP, Morin RD, Khattra J, et al.: Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 2009;461:809–813.
29 Ding L, Ellis MJ, Li S, et al.: Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 2010;464:999–1005.
30 Garraway LA, Lander ES: Lessons from the cancer genome. Cell 2013;153:17–37.
31 Wetterstrand KA: DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). *http://www.genome.gov/sequencingcosts/*. Accessed 2 September 2013.
32 Dawson S, Rueda OM, Aparicio S, et al.: A new genome-driven integrated classification of breast cancer and its implications. EMBO J 2013;32:617–628.
33 Bozic I, Antal T, Ohtsuki H, et al.: Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci USA 2010;107:18545–18550.
34 Vogelstein B, Papadopoulos N, Velculescu VE, et al.: Cancer genome landscapes. Science 2013;339:1546–1558.
35 Shibata D: Cancer. Heterogeneity and tumor history. Science 2012;336:304–305.
36 Jones S, Chen W, Parmigiani G, et al.: Comparative lesion sequencing provides insights into tumor evolution. Proc Natl Acad Sci USA 2008;105:4283–4288.
37 Lawrence MS, Stojanov P, Polak P, et al.: Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013;499:214–218.
38 Steensma DP: The beginning of the end of the beginning in cancer genomics. N Engl J Med 2013;368:2138–2140.
39 Curtis C, Shah SP, Chin S, et al.: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346–352.
40 Koboldt DC, Fulton RS, McLellan MD, et al.: Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.
41 Iyer G, Hanrahan AJ, Milowsky MI, et al.: Genome sequencing identifies a basis for everolimus sensitivity. Science 2012;338:221.
42 Corless CL: Medicine. Personalized cancer diagnostics. Science 2011;334:1217–1218.
43 Aparicio S, Caldas C: The implications of clonal genome evolution for cancer medicine. N Engl J Med 2013;368:842–851.
44 Gray J, Druker B: Genomics: The breast cancer landscape. Nature 2012;486:328–329.
45 Ellis MJ, Ding L, Shen D, et al.: Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 2012;486:353–360.
46 Shah SP, Roth A, Goya R, et al.: The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486:395–399.
47 Stephens PJ, Tarpey PS, Davies H, et al.: The landscape of cancer genes and mutational processes in breast cancer. Nature 2012;486:400–404.
48 Hanahan D, Weinberg RA: Hallmarks of cancer: The next generation. Cell 2011;144:646–674.
49 Landau DA, Carter SL, Stojanov P, et al.: Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 2013;152:714–726.
50 Nik-Zainal S, van Loo P, Wedge DC, et al.: The life history of 21 breast cancers. Cell 2012;149:994–1007.
51 Yachida S, Jones S, Bozic I, et al.: Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 2010;467:1114–1117.
52 Gerlinger M, Rowan AJ, Horswell S, et al.: Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366:883–892.
53 Diaz LA, Williams RT, Wu J, et al.: The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature 2012;486:537–540.
54 Yap TA, Gerlinger M, Futreal PA, et al.: Intratumor heterogeneity: seeing the wood for the trees. Sci Transl Med 2012;4:127ps10.
55 Alexandrov LB, Nik-Zainal S, Wedge DC, et al.: Signatures of mutational processes in human cancer. Nature 2013;500:415–421.
56 Nik-Zainal S, Alexandrov LB, Wedge DC, et al.: Mutational processes molding the genomes of 21 breast cancers. Cell 2012;149:979–993.
57 Stephens PJ, Greenman CD, Fu B, et al.: Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 2011;144:27–40.
58 Banerji S, Cibulskis K, Rangel-Escareno C, et al.: Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 2012;486:405–409.
59 Loman NJ, Misra RV, Dallman TJ, et al.: Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 2012;30:562.
60 Leary RJ, Kinde I, Diehl F, et al.: Development of personalized tumor biomarkers using massively parallel sequencing. Sci Transl Med 2010;2:20ra14.