Whole genome sequencing of breast cancer

Rossing, Maria; Sørensen, Claus Storgaard; Ejlertsen, Bent; Nielsen, Finn Cilius

Published in:
APMIS - Journal of Pathology, Microbiology and Immunology

DOI:
10.1111/apm.12920

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Whole genome sequencing of breast cancer

MARIA ROSSING,1 CLAUS STORGAARD SØRENSEN,2 BENT EJLERTSEN3 and FINN CILIUS NIELSEN1

1Centre for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen; 2Biotech Research and Innovation Centre, University of Copenhagen, Copenhagen; and 3Danish Breast Cancer Cooperative Group & Department of Clinical Oncology Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

Breast cancer was the first to take advantage of targeted therapy using endocrine therapy, and for up to 20% of all breast cancer patients a further significant improvement has been obtained by HER2-targeted therapy. Greater insight in precision medicine is to some extent driven by technical and computational progress, with the first wave of a true technical advancement being the application of transcriptomic analysis. Molecular subtyping further improved our understanding of breast cancer biology and has through a new tumor classification enabled allocation of personalized treatment regimens. The next wave in technical progression must be next-generation-sequencing which is currently providing new and exciting results. Large-scale sequencing data unravel novel somatic and potential targetable mutations as well as allowing the identification of new candidate genes predisposing for familial breast cancer. So far, around 15% of all breast cancer patients are genetically predisposed with most genes being factors in pathways implicated in genome maintenance. This review focuses on whole-genome sequencing and the new possibilities that this technique, together with other high-throughput analytic approaches, provides for a more individualized treatment course of breast cancer patients.

Key words: Breast cancer; whole genome sequencing; genome maintenance.

Cancer is a disease of the genome and enormous efforts are directed towards understanding of this heterogeneous collection of diseases (1). The expansion of our insight in the cancer genomes is mostly driven by the rapid development in sequencing technologies all the way from the early identification of oncogenes and tumor suppressors to the full annotation of the most common cancers resulting in the so called genomic landscape of cancer (2, 3). Determining the genomic landscape of cancers is an ongoing process supported by The Cancer Genome Atlas (TCGA) dataset which comprises genomic data from >10 000 thousands of tumor samples, providing researchers with a comprehensive catalog of the key genomic changes in more than 30 types of cancer (4). The major advances in sequencing technologies followed by the development of computational tools have enabled analyses like whole-exome sequencing (WES), RNA-sequencing (RNA-seq) and whole-genome sequencing (WGS) to be implemented in the routine clinical setting, hereby supporting the emerging clinical relevance of genomics in cancer medicine as well as for other diseases (5). The cancer genome is somewhat dynamic, and each cancer evolves with the accumulation of several types of somatic mutations, copy number alterations, epigenetic factors, and structural variants. These changes can occur in a predisposed genetic background like the hereditary cancers which again cause diverse patterns for the individual tumor genome.

Thus, accepting the fact that cancer is a genomic disease and combining this with the growing insights in targeted therapies, the way for precision oncology is being founded. Precision oncology is based on the theory that the examination of both the patients’ genome and the tumor genome will direct the clinician to the targeted drug, expected to

Received 1 October 2018. Accepted 16 December 2018

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
be effective (6). With the implementation of several FDA-approved HER2-targeted therapies, breast cancer was one of the earliest cancers where implementing targeted therapies have shown a prolonged survival in patients whose tumors are driven by this tyrosine kinase activation (7). Thus, more than 15% of breast cancer patients follow a scheme of targeted therapy, though there is still room for testing and implementation of precision oncology for these patients. In line with this, a recent comprehensive whole-genome-based study of both the primary, the locally relapsed, and the metastatic breast cancer showed that cell clones seeding metastasis or relapse disseminate late from the primary tumors, but continue to acquire mutations, mainly within the same pathway as the primary tumor (8). However, Yates et al., found that most distant metastasis acquired new driver mutations, including clinically actionable alterations and mutations, which is highly relevant for the implementation of genome-driven oncology with the aim of improving the dismal survival of metastatic breast cancer patients.

**BREAST CANCER**

Breast cancer is the most common malignancy among females, and together with lung and colorectal cancers, among the three most common cancers worldwide (9). In 2012 approximately 1.7 million women were diagnosed with breast cancer worldwide, and breast cancer accounts for nearly a third (29%) of all new cancers in women. In women under the age of 50, breast cancer is the leading cause of cancer death, but in elder women it is surpassed by lung cancer as the most frequent cause of death. Overall mortality rates are decreasing although the rates vary greatly worldwide and are the leading cause of cancer death in the less developed parts of the world. The incidence of breast cancer in Denmark is around 4700 new cases per year, and all though mortality rate has decreased mostly attributable to early detection via screening programs and advanced and efficient therapies, breast cancer still causes 1100 deaths per year; being the second cause of cancer death also in Danish women (10).

The diagnostic procedures of breast cancer include clinical examination, breast imaging usually comprising mammography and ultrasound, core-needle tumor biopsy for histopathology with biomarker assessment, and assignment of intrinsic subtype by molecular genetic analysis (see separate paragraph on molecular subtypes; 11). Based on a clinical risk assessment, presence of comorbidities and patient preferences, a treatment recommendation is for most patients reached by the multidisciplinary team using the biomarker tumor profile of each patient. Algorithms in combination with risk factors have been developed for clinical risk assessment of recurrence and have for Danish breast cancer patients been extended to allow calculation of a prognostic standard mortality rate (SMR) index (PSI) algorithm based on a comprehensive study of >6500 postmenopausal patients with ER positive high-risk breast cancer (12, 13). Early breast cancer without distant metastasis (Stage I + II) is a possible curative disease with breast conserving surgery in combination with systematic therapy being the standard of care (14). In brief, luminal-like, ER+ - patients are treated with either endocrine therapy alone, or in combination with chemotherapy. If tumors are HER2+, trastuzumab is added. For patients with triple-negative tumors, standard systemic chemotherapy consists of anthracyclines and taxanes (15).

Neoadjuvant chemotherapy (NACT) – e.g. chemotherapy initiated before breast cancer surgery – is widely used as a standard of care to reduce surgical morbidity of the breast and axilla (16, 17). A recent meta-analysis substantiated that NACT results in higher rates of breast-conserving therapy without comprising the risk distant recurrence, breast cancer survival, or overall survival (18). However, uncertainty remains regarding the extend of axillary lymph node assessment and target of radiotherapy after NACT. Targeted therapy directed by tumor profile is a cornerstone in the NACT – setting where a complete pathological response and long-term outcome is superior to the HER2-positive or the ER-negative and HER2-normal tumors in comparison to the luminal breast cancers for which the response to NACT is more unclear (19).

**MOLECULAR SUBTYPES**

Transcriptome analysis enables classification of breast tumors into intrinsic molecular subtypes which are biologically distinct entities with specific prognostic and therapeutic features (20–25). The pivotal studies proposed five subclasses: (i) the ER-receptor positive and human epidermal growth factor receptor 2 (HER2)-receptor negative tumors i.e., luminal A (lumA), luminal B (lumB) and normal breast-like subclass, (ii) the HER2-receptor positive tumors: HER2-like subclass, and (iii) the ER- and HER2-receptor negative tumors called the basal-like (basL) subclass. Four of the subclasses can be distinguished by a 50-gene molecular
WHOLE GENOME SEQUENCING OF BREAST CANCER

classifier (PAM50) which has been developed as a commercial FDA approved platform (ProsigmaVR; 26). This was followed by the 70-gene signature; MammaPrint, which together with a clinical risk evaluation, showed promising results for predicting the low-risk cancers that may omit chemotherapy (27). In line with this, a comprehensive, multicenter analysis, TAILORx, freshly showed that the 21-gene recurrence-score assay (Oncotype DX, Genomic Health) providing prognostic information in hormone-receptor–positive breast cancer in combination with the 21-gene-signature, can direct patients with either a very low score to omit chemotherapy or patients with a high score to enroll for chemotherapy (28). However, for most patients (71%) who were assigned with a mid-range recurrence score, the TAILORx-study could not clarify whether disease-free survival or overall survival was correlated was patients who underwent treatment with chemotherapy or not; this based on 9 years of follow-up. Recent taxonomies optimized the subclasses by applying integrative genomic analysis and Guedj et al. refined the subclasses by introducing six stable molecular subtypes based on genomic rearrangement and the expression of 256 transcripts (29, 30). However, a comprehensive genomic study integrating both genetic and epigenetic alterations concluded that breast cancers, in addition to the intrinsic subtypes and clinical heterogeneity, can be explained by structural variants defining subclusters within the subtypes (31). Subsequent studies into defining subclusters of subtypes has resulted in several new signatures for identifying specific somatic and pathway-based subclusters among the intrinsic subtypes (32–36). For example, Lehmann et al., identified six subclusters among the triple-negative breast cancers; two basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), a mesenchymal stem-like (MSL), and a luminal androgen receptor (LAR) subtype. In addition, 17 pathway-driven subclusters were suggested with Gatza et al., hereby linking the heterogeneity to tentative therapeutic strategies. In many institutions genomic analysis has in recent years become a part of the standard of care for most breast cancer patients. It is well-established that the biological hallmark of luminal A subtype is low proliferation, high expression of the ESR1 gene and a favorable clinical outcome (37). Since 2011, the St Gallen international expert consensus panel has recommended merely endocrine therapy in patients with luminal A disease (15, 38). Attempts have been made indirectly to approximate luminal A – like subtype using IHC biomarkers – ER- and PGR-positive, HER2-negative and low Ki67 protein staining (39, 40). However, classification with only four biomarkers does not entirely recapitulate the intrinsic subtype of breast cancer (37). From 2017, DBCG recommended molecular subtyping for all patients at intermediate risk, and a recent study showed the benefits of applying up-front routine subtyping of all early breast cancers, hereby identifying both high and low risk patients (41).

FROM SANGER TO WHOLE-GENOME SEQUENCING

Sequencing of nucleic acids is a method for determining the exact order of each nucleotide (C, G, T and A for DNA sequence and C, G, U and A for RNA sequence). The order of these nucleotides in a polynucleotide string is what finally gives the hereditary and biological codes for all human life. Human DNA consists of about 3 billion nucleotides, and more than 99% of those are identical in all people. Thus, science, including medical human research, is conditioned by the ability to determine the order of nucleotides by sequencing. Or as the double Nobel laureate Frederick Sanger has said it once in a more modest quote: ‘…knowledge of sequences could contribute much to our understanding of living matter.’ (42). In 1953 Watson and Crick determined the three-dimensional structure of the DNA double helix which revolutionized our understanding of science (43). However, the ability to decode the order of nucleotides or sequence were time-consuming and labor-intensive until 1977, where the establishment of the Sanger sequencing method took place; Selective incorporation of chain-terminating deoxynucleotides by DNA polymerase during in vitro DNA replication (44, 45). Like previous sequencing techniques based on step-by-step lengthening of sequencing and labeling of the final nucleotide, Sanger sequencing outperformed them all due to the precision, robustness and usability of the chain-termination method, which made the Sanger technique the most widely used method in more than two and a half decades. Automation of the Sanger sequencing method became a reality in 1986 when Applied Biosystems launched the automated DNA sequencing instrument where, although still based on the Sanger method, each nucleotide was now labeled with a fluorescent dye and allowing a read out of specific nucleotides by color (46). Also, the Applied sequencing instrument enabled 24 samples to run at the same time and generated sequencing data faster and cheaper than ever before. This automated Sanger technique, also known as ‘first-generation sequencing’, allowed one of the largest projects in
human science; ‘The Human Genome project’. The Human Genome project set out to sequence every human chromosome with the objective to advance knowledge of human biology and improve medicine. After years of planning, the project was formally launched in 1990, and after a 13-year-long endeavor the project completed in 2003, 2 years ahead of schedule (47, 48). It still remains the world’s largest collaborative biological project with a budget of $3 billion (49).

Following the complete sequencing and mapping of the 24 human chromosomes, the demand for sequencing increased rapidly, thus calling for a more low-cost and high-throughput technique. This calling motivated the development of second-generation sequencing, also known as next-generation sequencing (NGS), where pyrosequencing plays a major role. The pyrosequencing technique, pioneered by Pal Nyren and colleagues (50, 51) was still, like the Sanger’s dieoxy, a sequence-by synthesis’ (SBS) technique, as they both require a direct action of DNA polymerase to produce the reading output. The pyrosequencing method is based on detecting the activity of the DNA polymerase with luminescence, hence allowing sequencing a single strand of DNA by synthesizing the complementary strand, one base pair at a time. Each nucleotide is sequentially washed through the system over the template DNA affixed to a solid phase. Light is produced only when the nucleotide solution complements the first unpaired base of the template and the order of solutions which produce the luminescence permits the determination of the sequence of the template. The first major successful commercial NGS platform, the 454 sequencer was based on the concept of pyrosequencing technique and mass parallelization of sequencing reactions, greatly increasing the amount of DNA that can be sequenced in one run (52). Several parallel sequencing techniques followed upon the success of the 454 sequencer where the Solexa method of sequencing, acquired by Illumina in 2006, is the most important and ‘short read’ sequencing platform (53). In short, template DNA is fragmented and end-repaired with poly-A tails to ensure the ligation of adaptors. The adapter oligonucleotides are complementary to the flow-cell anchors. Adapter-modified, single-stranded DNA is added to the flow cell and immobilized by hybridization. Bridge amplification generates clonally amplified clusters that are denatured and cleaved, and sequencing can be initiated with addition of primers, polymerase and the four reversible dye terminators, followed by imaging and recording of fluorescence. When recording of the one cycle is completed, fluorescence and terminators are removed, and the next cycle of synthesis is directly initiated.

Targeted sequencing and gene-panels are widely used when predefined genes are to be sequenced and if coverage need to be high, for instance to identify low-frequency tumor-drives. The actual sequencing process is identical to the one described above, however the region of interest is captured by initial hybridization of a library of biotinylated RNA oligos predesigned to specific genomic regions (e.g. gene-panel) using magnetic beads, flowed by PCR amplification. It is often argued that targeted sequencing is preferable if the suspected disease or condition has already been identified, due to affordable costs and higher coverage yield as well as reduced sequencing time (54). However, it may be reasonable to question this presumption, since it holds some frauds; for instance, sequencing is often used to confirm a tentative diagnosis, and often, variants in the suspected genes associated with the most likely diagnosis are not identified, resulting in another targeted gene panel or a broader screening method like WES or WGS. Thus, choosing targeted gene-panel may indeed increase the time until a causative genetic diagnosis is made, as well as the costs, which would have been reduced if the genomic approach had been the first and only platform chosen (55–57). Other arguments for working towards reducing the use of gene-panels and converting to WGS as the primary choice, is that the sequencing data can be remapped and reanalyzed at a later stage, when new genetic associations are identified; moreover, the data together with the patients phenotype enables research in a specific disease and will lead to the identification of new disease associated candidate genes without further laboratory costs. Interestingly, when comparing sample preparation time in the laboratory, the pre-capture-step necessary for gene-panels exceeds the non-capture technique used for WGS, both in time and hands-on (58).

Whole-exome-sequencing is also a result of a pre-capture-step, this time of all coding exons of the genome, allowing a sequential screen of the exome. In contrast, WGS does not include a capturing step, since it is the entire genome that is intended for sequencing; both coding and non-coding. This fundamental difference which may best be illustrated by a raw sequence output (Fig. 1), greatly affects the downstream applications. WGS provides an incomparable complete coverage of the exome thus, WGS is simply the better WES (59). As such, and with costs declining and with the appropriate in silico panels, WGS has the potential to entirely replace WES and other techniques that involves capturing of target sequences. In addition, a wide range of WGS data applications, like insertion-deletions, copy-number variation (CNV),
intrinsic deletions, structural variation and repetitive DNA element, substantiates why this sequencing platform is superior to WES. It is so far not possible to interpret the significance of most of the additional findings from WGS but then only through continuous sequencing studies, the knowledge database will expand and enable implementation of precision medicine in a clinical setting. However, since average coverage of a WGS is around 30–40×, WGS is not optimal for identification of mosaicism or other low-frequent variants e.g. tumor-drivers. For these purposes, a targeted gene-panel with a minimum of 500× is more suitable. The cost of a WGS opposed to WES and gene-panels is a major concern. The price of sequencing is constantly declining and at present time a WGS costs less than $1.5K (60). The need for data storage, curation and bioinformatic processing is substantial if WGS-analysis is implemented as a routine genetic test (61, 62). Consequently, most western countries are launching national initiatives for large-scale whole-genome-sequencing projects where essential computer- and man-power is centered. Obviously, the long-term impact on health economics is not yet understood. Discussions about how to manage the risk of identifying incidental genomic findings have emerged as one of the more contentious issues in the clinical application of genomic sequencing (63, 64). The subject is vast to review in the present context, however the key elements that the discussions has brought is the transition of incidental to secondary findings and the patients right to 'option out' on secondary findings. The American College of Medical Genetics and Genomics have recently updated their recommendations on how to address secondary findings in a clinical setting, hereby enabling consensus amongst sequencing laboratories (65). The basic laboratory workflow and downstream data pipeline and various applications of WGS will be addressed in the following.

An overview of the laboratory workflow for generating WGS sequencing results is shown in Fig. 2 and is described only for the commonly applied Illumina protocol. Genomic DNA purified from either whole blood or tumor is prepared for library by using Nextera DNA Flex Library Preparation Kit with dual indexes. In brief, 10–500 ng of genomic DNA is used as input for the on-bead tagmentation step, followed by a limited PCR amplification and bead clean-up of the WGS DNA sequencing library. Sequencing can be readily performed as 2 × 150 bp paired-end sequencing on a NovaSeq6000 instrument. Raw fastq files are mapped to the hg19/GRCh37 human reference genome using BWA-MEM v0.7.12 (Li, 2013) software. Quality thresholds for the sequencing are >30-fold average sequencing depth and >98% of the genome sequenced at least 10-fold. Alignment file pre-processing and germline variant calling is performed by GATK v3.8.0 using Best Practices guidelines (66). For clinical implications and from sample to clinical reporting, the entire laboratory workflow and data processing can be completed in less than a week, depending on the downstream manual and medical variant classification. Preprocessed WGS data can be used to identify for structural variants e.g. CNVs and fusion genes or CNVs (59, 67, 68). Furthermore, WGS data enables de novo assembly, which is an alternative way of variant calling, since it does not map to a reference genome, in other words, the variant caller does not assume any normal positions; thus, de novo assembly allows ‘wrong’ reads which may reflect insertions, inversions or translocations (69, 70). It should be noted, that if DNA originates from tumor tissue the standard algorithm for data processing and variant calling does not explicit consider tumor impurity or intra-tumor-heterogeneity. So far, estimation of tumor content and identification of clonal somatic aberrations require high-coverage sequencing, which is not the case for WGS data where coverage is limited (30–40×). Comparative studies clearly show that tols developed for high-sequence coverage data are not suitable for WGS data (71). Accordingly, bioinformatic tools for deconvolution of genomic low-coverage data are emerging (61, 72).

HEREDITARY BREAST CANCER

One out of every seven women diagnosed with invasive breast cancer will have a close relative (mother, sister, or daughter) also diagnosed with
Pathogenic variants in the two major breast cancer susceptibility genes \textit{BRCA1} and \textit{BRCA2} may explain 15\% of increased risk of breast cancer among female relatives of breast cancer patients. Germline variants in several other susceptibility genes also confer a high risk of breast cancer, including pathogenic variants in \textit{ATM}, \textit{BARD1}, \textit{BLM}, \textit{BRIP1}, \textit{CDH1}, \textit{CHEK2}, \textit{PALB2}, \textit{PMS2}, \textit{FAM175A}, \textit{FANCC/-M}, and \textit{RAD51B/-C/-D} as well as the more syndromic predisposing genes; \textit{TP53} (Li–Fraumeni syndrome), \textit{PTEN} (Cowden disease), \textit{STK11} (Peutz-Jeghers syndrome), \textit{NF1} (Neurofibromatosis), and \textit{CDH1} (Hereditary diffuse gastric cancer syndrome). Large panels of breast cancer susceptibility genes have become widely available (73). Identification of pathogenic \textit{BRCA1} and \textit{BRCA2} variants is routinely used to predict risk of breast- and ovarian cancer and guide the use of risk reducing surgery, thus upfront genetic screening of all new breast cancer patients enables a more tailored surgical procedure (74–77). Hence, the meta-analysis by Li et al., recently substantiated the significant decrease in overall mortality for patients with pathogenic \textit{BRCA1} or \textit{BRCA2} variants undergoing contralateral prophylactic mastectomy as part of the standard breast cancer intervention. Adjuvant chemotherapy is not recommended to all patients with ER-positive and HER2-normal breast cancer but may in women with germline pathogenic \textit{BRCA1} or \textit{BRCA2} variants confer a distinct survival benefit (75, 78). Furthermore, platinum-based chemotherapy has in the neoadjuvant and metastatic setting been superior to conventional anthracycline and taxane based regimens (79). Finally, targeting impaired \textit{BRCA1} or \textit{BRCA2} by poly ADP ribose polymerase-inhibitors (PARPi) is approved for metastatic breast cancer in women with germline pathogenic \textit{BRCA1} or \textit{BRCA2} variants (80, 81). Accordingly, up-front screening for \textit{BRCA1} and \textit{BRCA2} variants is increasingly offered to patients at the diagnosis of breast cancer and may be completed in less than a week (41). National consensus on breast cancer panel testing is emerging and may result in clinical utility of a wider range of germline variants (82). It is evident,
that more breast cancer predisposing genes will occur and for this purpose, an international screening program was established to assemble the sequencing data and collaborating on identifying and validating new candidate breast cancer genes (83).

**BRCA1**

Pathogenic variants in **BRCA1** and **BRCA2** predispose to hereditary breast and ovarian cancer (HBOC), but only 15–25% of HBOC cases can be ascribed to either of the genes. Recently, exome sequencing has uncovered a substantial locus ascribed to either of the genes. Recently, exome (HBOC), but only 15–25% of HBOC cases can be ascribed to either of the genes. Recently, exome sequencing has uncovered a substantial locus ascribed to either of the genes.

**BRCA2**

Pathogenic variants in **BRCA1** and **BRCA2** predispose to hereditary breast and ovarian cancer (HBOC), but only 15–25% of HBOC cases can be ascribed to either of the genes. Recently, exome sequencing has uncovered a substantial locus ascribed to either of the genes. Recently, exome (HBOC), but only 15–25% of HBOC cases can be ascribed to either of the genes.

**BREAST CANCER AND GENOME MAINTENANCE**

Pathogenic variants in **BRCA1** and **BRCA2** predispose to hereditary breast and ovarian cancer (HBOC), but only 15–25% of HBOC cases can be ascribed to either of the genes. Recently, exome sequencing has uncovered a substantial locus ascribed to either of the genes. Recently, exome (HBOC), but only 15–25% of HBOC cases can be ascribed to either of the genes. Recently, exome (HBOC), but only 15–25% of HBOC cases can be ascribed to either of the genes.

**BRIP1**

Amongst the DNA repair pathways, mismatch repair (MMR) is also involved though less frequently than HRR. The MMR genes **MLH1**, **MSH2**, **MSH6**, and **PMS2** that were originally implicated in Lynch syndrome or hereditary non-polyposis colorectal cancer (HNPCC), also play a role in breast and ovarian cancer since haploinsufficiency of **MLH1**, **MSH2**, and **PMS2** appears to increase the risk for ovarian cancer and to a minor extent breast cancer (90, 91). The MMR system recognizes and repairs misincorporation of nucleotides and defects in MMR results in the accumulation of thousands of single nucleotide variants and microsatellite instability (MSI) in the tumors (92). Tumor suppressor genes, such as **MRE11** and **RAD50** genes, moreover, harbor microsatellites so impaired MMR may indirectly affect the MRN complex (93).

**RAD51C** and **RAD51D** variants are mainly found in families with ovarian cancer or breast and ovarian cancer. **BRIP1** and **FAM175A** are both **BRCA1** interacting proteins involved in recruitment of **BRCA1** to DNA DSBs (84). The **BRIP1** gene was originally suggested to be a low-penetrant breast cancer susceptibility gene, but, later studies suggested that the risk was elevated for ovarian cancer. Pathogenic variants in **FAM175A** (Abraxas) have been identified in both breast and ovarian cancer patients, however, currently the lifetime risk for breast- and ovarian cancer is unknown (84). The Mre11-Rad50-Nbs1 (MRN) complex is also necessary for detection and signaling of DNA DSBs (88). So far, only **MRE11** and **NBN** variants have been clinically associated to breast and ovarian cancer. The MRN complex activates the ATM (Ataxia Telangiectasia-Mutated) serine/threonine protein kinase that phosphorylates factors in the DNA damage response including TP53, CHK2, and CtIP (**RBBP8**) (89). A number of rare HBOC factors such as **BLM**, **RECQL**, **FANCC**, and **FANCM** may also contribute to HBOC by protecting DNA replication forks and suppressing DNA replication stress (84).

Amongst the DNA repair pathways, mismatch repair (MMR) is also involved though less frequently than HRR. The MMR genes **MLH1**, **MSH2**, **MSH6**, and **PMS2** that were originally implicated in Lynch syndrome or hereditary non-polyposis colorectal cancer (HNPCC), also play a role in breast and ovarian cancer since haploinsufficiency of **MLH1**, **MSH2**, and **PMS2** appears to increase the risk for ovarian cancer and to a minor extent breast cancer (90, 91). The MMR system recognizes and repairs misincorporation of nucleotides and defects in MMR results in the accumulation of thousands of single nucleotide variants and microsatellite instability (MSI) in the tumors (92). Tumor suppressor genes, such as **MRE11** and **RAD50** genes, moreover, harbor microsatellites so impaired MMR may indirectly affect the MRN complex (93).

**DNA repair** is undoubtedly important in suppressing HBOC, but nevertheless, additional pathways related to genome stability and DNA damage response are also relevant. In this regard, cell cycle checkpoints and cell death pathways may also be implicated in HBOC. These pathways normally eliminate cells with damaged DNA and haploinsufficiency of check-point regulators genes leads to accumulation of mutated cells. The archetypical check-point regulator is TP53 which coordinates several genome stability factors. ATM, CHK2 (encoded by **CHEK2**), TP53, **BRCA1**–**BRCA2** can block the cell cycle in G1, S and in G2 phase...
Constitutive loss of TP53 leads to the autosomal dominant Li–Fraumeni syndrome (LFS) with breast cancer, sarcoma, brain tumors, and adrenocortical carcinoma (96, 97). TP53 is activated by the ATM kinase that phosphorylates CHK2, and in an amplifying step CHK2 subsequently further targets TP53 thereby promoting the actions of ATM (89, 98). The clinical significance of the mechanism is illustrated by the CHEK2 c.1100delC variant that is associated with approximately three-fold increased risk of breast cancer (99, 100). Taken together, cells with impaired DNA repair and cell cycle checkpoints are ultimately likely to gain selective proliferative advantages.

The recent progress in next generation sequencing makes it possible to identify predisposing genetic variants in a fast and cost-effective way. We, however, face a situation where cohort and cosegregation data may not be available for rare variants, and it is important to investigate at the protein level, for example, by employing structural and functional analysis for variant classification. Since the extensive locus heterogeneity appears to...
converge on a relatively small number of genome maintenance pathways, we predict that such analysis may become a reality within a relatively short period of time. Unfounded classification of genetic variants is obviously harmful to the patient, and great care should be taken to generate common protocols and accreditation of the analysis to meet clinical standards.

Whole-genome sequencing can be used to detect an impairment in two of the major pathways in the genome-maintenance system; HRR and MMR. For detection of homolog recombination deficiency (HRD) by WGS, Davies et al. recently generated a predictor for BRCA1 and BRCA2-deficient tumor samples. This WGS-based predictor may be well-suited for the clinical setting since the study showed a doubling in the detection rate of HRD tumors as well as potential actionable targets from high-coverage NGS-panel; Molecular subtype, receptor status and proliferative index from RNA based on RNA-sequencing. Histopathology includes immunohistochemical type, morphological characteristics and receptor status.

Whole-genome sequencing can be used to detect an impairment in two of the major pathways in the genome-maintenance system; HRR and MMR. For detection of homolog recombination deficiency (HRD) by WGS, Davies et al. recently generated a predictor for BRCA1 and BRCA2-deficient tumor samples. This WGS-based predictor may be well-suited for the clinical setting since the study showed a doubling in the detection rate of HRD tumors as well as potential actionable targets from high-coverage NGS-panel; Molecular subtype, receptor status and proliferative index from RNA based on RNA-sequencing. Histopathology includes immunohistochemical type, morphological characteristics and receptor status.

inactivation of BRCA1 and BRCA2 is clinically extremely relevant since these tumors are selectively sensitive to PARPi (102, 103). For identification of MMR-impaired breast cancer tumors, another WGS based mutational signature has recently been described (104). The mutational signature is sought to be a direct pathophysiological reflection of MMR pathway abrogation that may outperform current biomarkers and hereby increasing the sensitivity to immune therapies. Thus, WGS should soon be required for up-front routine clinical analysis and we suggest an optimized diagnostic workflow which allows the performance of precision medicine, Fig. 4.

REFERENCES

35. 2019 The Authors. APMIS published by John Wiley & Sons Ltd on behalf of Scandinavian Societies for Medical Microbiology and Pathology


