



High-throughput resequencing in the diagnosis of *BRCA1/2* mutations using oligonucleotide resequencing microarrays

Christopher Schroeder¹, Fanny Stutzmann¹, Bernhard H.F. Weber³, Olaf Riess¹ and Michael Bonin¹

Institutions

1) Dept. of Medical Genetics, University of Tübingen, Calwer Str. 7, 72076 Tübingen

2) Institute of Human Genetics, University of Regensburg, Franz-Josef-Stauß-Allee 11, 93053 Regensburg, Germany

Corresponding author's address:

Michael Bonin, Ph.D.

Dept. of Medical Genetics

University of Tübingen

Calwer Str. 7

72076 Tübingen/Germany

Tel.: +49 (0)7071-29 72295

Fax: +49 (0)7071-29 5172

Email: Michael.bonin@med.uni-tuebingen.de

Abstract

Breast cancer is the most frequent form of carcinoma in European females (incidence 65 per 100,000). In about 10% of all cases pedigree analysis predicts a hereditary breast-ovarian cancer syndrome (HBOC) to be causative for the disease. Frequently, mutations in two genes, *BRCA1* (Chr. 17q21) and *BRCA2* (Chr. 13q12), are associated with HBOC. In females, mutations in these genes result in a lifetime risk of 80–85 % for breast cancer and 54 % (*BRCA1*) or 23 % (*BRCA2*) for ovarian cancer. Current genetic diagnostic tools for *BRCA1* and *BRCA2* remain laborious and expensive. Here we present the first oligonucleotide resequencing microarray covering the complete coding sequence of both genes. In total, 36 previously characterized DNAs were resequenced; all 11 patients with single-nucleotide mutations and, due to a special mutational design, 8 patients with heterozygous deletions were detected correctly. In total, 47 different single-nucleotide variants (SNVs) were found. A newly developed software, SeqC, reduced the number of ambiguous calls with the help of a statistical module comparing the acquired data to an online-database. SeqC improved the average call rate to 99 % (GSeq: 97 %) and reduced time and efforts for manual analysis. SeqC confirmed the results obtained by GSeq and found an additional 33 sequences changes representing 14 SNVs. In total, 945 kb were screened and the overall turnaround time for each patient took approximately 3 days, including analysis.

Keywords

oligonucleotide resequencing microarray, *BRCA1*, *BRCA2*, breast cancer, ovarian cancer, HBOC, HBC, HOC, SeqC

Introduction

Breast cancer is the most common form of carcinoma in females in Europe, with an average incidence of 65 per 100,000 individuals ¹. In most patients the tumor appears sporadic, but it is estimated that at least 10 % involve a genetic predisposition which as a consequence leads multiple manifestation of hereditary breast cancer (HBC) in distinct families ². In the early 1990s, *BRCA1* (Chr. 17q21) and *BRCA2* (Chr. 13q12) (Table 1) were identified as high-risk susceptibility genes for autosomal-dominant breast cancer ^{3,4}. In addition, follow-up studies were able to link both genes to ovarian cancer (OC) and less frequently to several other types of carcinoma (e.g., melanoma colon, pancreatic, and prostate cancer) ⁵. The association of at least 30 % of all HBC to OC led to the term hereditary breast-ovarian cancer syndrome (HBOC). HBOC caused by mutations in *BRCA1* or *BRCA2* is associated with a lifetime risk of 80–85 % for BC and 54 % (*BRCA1*) or 23 % (*BRCA2*) for OC, respectively ⁶. Both BC and OC are potential life-threatening diseases with an average 5-year-survival rate of 80 % (BC) and 43 % (OC) in Europe ¹. As for every tumor disease prognosis, therapy and survival strongly depend on its stage at diagnosis. Genetic screening tests, especially for patients with a known familial risk of ovarian cancer, are of great value since 71 % of clinically suspicious patients are in stage III or IV of the disease, and, as a consequence, their prognosis is very poor ⁷. It is essential to identify these patients and provide genetic counseling to assess their individual disease risk, calculate the probability of mutation-carrier status, and offer genetic analysis. For families with known mutations in *BRCA1* or *BRCA2*, extended preventive screening as well as therapeutic strategies have been established ⁸.

During the past 18 years, 815 mutations were identified in *BRCA1* and 571 in *BRCA2* (Human Gene Mutation Database; www.hgmd.cf.ac.uk). However, genetic screening for mutations in the two genes with a total of 50 exons and approximately 17380 base pairs (including exon-intron-boundaries) remains a major challenge. Traditional techniques are

direct sequencing, denaturing gradient gel electrophoresis (DGGE), and denaturing high-performance liquid chromatography (dHPLC). Direct sequencing is especially laborious and expensive while DGGE and dHPLC are technically complex and the results that are obtained have to be evaluated by direct sequencing.

Recently developed sequencing tools and resequencing-microarrays are more cost- and time-efficient because they allow high-throughput, parallel resequencing of several genes. As proof-of-principle, one of the first oligonucleotide microarrays was designed for a single exon (exon 11) of *BRCA1*⁹. Over the last five years, resequencing oligonucleotide microarrays were established for general mitochondrial mutation analysis¹⁰ and a few diseases like cardiomyopathy^{11,12}, intrahepatic cholestasis¹³, and retinal degeneration¹⁴. Our Breast Cancer Microarray is the first oligonucleotide resequencing microarray spanning the complete sequence of *BRCA1* and *BRCA2*. In a first approach, 36 patients were screened for hereditary breast and ovarian cancer. To speed up the process, all patients were handled using multiplexes. The SeqC-software (www.jsi-medisys.de) was developed to further accelerate and facilitate analysis of oligonucleotide resequencing microarrays.

Material and Methods

Array Design

The principle of custom resequencing oligonucleotide microarrays uses 25-mer oligonucleotide probes of interest synthesized on an array¹⁵. Each position of the sequence of interest is represented as the central position of 8 different 25-mer oligonucleotides (for each position in the sense and antisense directions four probes with all four different nucleotides). Thus, each sequence is represented in sense and antisense direction with one perfect match and three mismatch nucleotides. This design allows detection of all possible single nucleotide

variations. Following the same principle, the most important deletions were designed on the chip with perfectly matched 25-mer oligonucleotide probes.

The design of the Breast Cancer Microarray was based on the commercially available CustomSEQ array (Affymetrix, Santa Clara). Reference sequences for both genes (*BRCA1*: GenBank U14680; *BRCA2*: GenBank U43746) were downloaded from GenBank in FASTA format. The sequence file contained all exons plus 10 bp of each flanking intron. In addition, common deletions and insertions were added to the design (Table 2). The instruction file contained the name of the reference sequence, start and end positions of the fragments, and their first and last 4 nucleotides. The final design contained all 50 exons (*BRCA1* – 23, *BRCA2* – 27) plus 61 selected deletions (Table 2). Each array is able to sequence approximately 26217 bps.

Patients and amplification of target sequences

In total, 36 DNAs (33 patients + 3 reference DNAs) were analyzed. The study was approved by the local ethics committee and all patients gave their informed consent before starting standard diagnostic techniques. In order to validate our Breast Cancer microarray all 36 DNAs were reanalyzed by means of this extended diagnostic approach. Mutations had been previously characterized by another method (sequencing or dHPLC + sequencing). Five patients were included as controls although obvious disease-causing mutations were not identified by conventional analysis as they carried neutral polymorphisms. A total of 12 patients had disease-causing single nucleotide mutations, while another 16 patients had small deletions of 1 to 11 base pairs (bp). DNA was isolated from lymphocytes using Roche's Magpure Compact instrument (Roche, Penzberg, Germany) following the manufacturer's protocol.

Primers were designed with Primer3 using default settings (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi, Supplementary Table S1). Three PCR-approaches were used

to generate all 50 amplicons ranging from 0.190 to 5.12 kbp (PCR-protocols Supplementary Table S1). Nine multiplexes with four to five amplicons each (Table 3), five single-PCRs, and three long-range PCRs were established to amplify the coding regions of both genes. A total of 1.0 µg DNA per patient was needed for amplification. The quality of all PCR products was determined using an Agilent Bioanalyzer 2100 (multiplexes) and gel electrophoresis (single-plex).

Chip processing and analysis

According to the protocols provided by Affymetrix, all amplicons were pooled and DNA was purified using a QIAquick PCR purification Kit (Qiagen, Hilden). The following steps (fragmentation, labeling, and hybridization) were performed as recommended by the manufacturer (GeneChip CustomSeq Resequencing Array Protocol Version 5.0, Affymetrix). All arrays were stained, washed using automated fluidic stations, and scanned with a GeneChip 3000 Scanner (Affymetrix, Santa Clara). Affymetrix GCOS Vers. 1.4 and GSEQ 4.0 Software (default settings) were used to process raw data and analyze the nucleotide sequences. SeqC Vers. 3.2.1.5 (JSI-medisys, www.jsi-medisys.de), was used to re-analyze the acquired datasets. SeqC is a module of the SeqPilot software and was originally used to analyze sequencing electropherograms from capillary sequencing. For further analysis, the generated data was compared to the Breast Cancer Mutation Database (<http://research.nhgri.nih.gov/projects/bic/>) and Human Genome Mutation Database (www.hgmd.cf.ac.uk/). In cases of contradicting database entries, the Breast Cancer Mutation Database annotation was finally used.

Results

The developed Breast Cancer Microarray was evaluated with 36 previously characterized DNAs (Table 4). In total, about 945 kbp (26217 bp per patient) were amplified, processed, and analyzed. Using the newly developed software, SeqC (www.jsi-medisys.de), an average call rate of 99 % was reached (GSeq – 97 %). The call rate improved with the rising number of chip experiments.

Amplification of target sequences

BRCA1 and *BRCA2* together consist of 50 exons. To facilitate and hasten the amplification, PCR-multiplexes were successfully established with four to five amplicons each. For detection reasons, every multiplex consisted of amplicons with a difference of at least 20 bp. Sufficient DNA quality was ensured using a Bioanalyzer (Affymetrix). A total of 9 multiplexes, 5 single PCRs, and 3 long-range PCRs were used and proved to be stable for all 36 DNAs. A list of all multiplexes can be found in Table 3. As a consequence, for each patient, 17 PCRs were necessary and, overall, 612 PCR-reactions were performed in this study. Due to the use of multiplexes, the total number of PCRs and duration of pipetting was reduced to about one-third, saving chemicals and hands-on-time.

New software

Data analysis was managed as a two-step process. In the first run, all patients were analyzed using GSeq (Affymetrix, Santa Clara), while in the second run, they were re-analyzed using SeqC (www.jsi-medisys.de). Samples marked with asterisk (s. Table 4) were excluded from the following statistical analysis. GSeq reached an average call rate (CR) of 97 %. About 212 loci per patient were Ns, since standard software parameters were not able to resolve the designated bases. These Ns must be evaluated manually and the corresponding fluorescence signals have to be screened for sequence changes. In our patients and with GSeq, 75 Ns in *BRCA1* and 137 Ns in *BRCA2* had to be evaluated manually. On the basis of the GSeq result

files, SeqC was able to generate an average CR of 99 % with its own basecaller and consequently reduced the overall number of Ns. Using SeqC, only 15 different loci for both genes (7 % of Ns in GSeq) were found and had to be evaluated manually. The average patient had 5 Ns in *BRCA1* and 9 Ns in *BRCA2*. Due to the length of *BRCA1* Exon 10 and *BRCA2* Exon 10 + 11, both SeqC and GSeq found most of the ambiguous calls in these exons. The average of 15 Ns plus an average of 12 polymorphisms per patient led to approximately 27 SNVs per patient.

Analysis with GSeq and evaluating ambiguous calls by comparing fluorescence intensities is especially laborious. In contrast to GSeq, SeqC offers a simple table with all Ns. Each N is directly linked to the corresponding position in the electropherogram and to a statistical module. The electropherogram helps to control the hybridization quality at that specific locus plus the nucleotides close to it. The statistical module compares each selected base with already archived chips, showing the medium relative intensity and the standard deviation. To maintain comparable chips, each chip had to be assigned to a PCR arrangement and only chips of the same arrangement were compared. To further facilitate analysis of oligonucleotide resequencing microarrays, SeqC was able to compare sequence changes with an online database (e.g., Ensembl, www.ensembl.org) and use the official nomenclature for base changes as well as the amino-acid-changes.

Due to the different underlying reference sequences in SeqC and in GSeq, SeqC found three additional SNPs (rs206075, rs206076, rs169547) that were homozygous in all patients and one SNP (rs543304) was detected in the complementary genotype. The newly discovered homozygous SNPs and rs543304 were not considered in the comparison of GSeq and SeqC. In total, SeqC was able to confirm the results of GSeq and added another 35 changes of 16 SNVs which would have been missed using GSeq alone (Figure 1). Of these different 16 SNVs, 3 were only detected in SeqC with one occurrence. One result of this study is the

confirmation that the different approach used by SeqC proved to be very efficient and reduced the duration of analysis to about one-tenth of the time.

Analysis of *BRCA1/2* oligonucleotide resequencing microarrays

All 12 out of 12 patients with single-base mutations were confirmed (Figure 2) by our resequencing microarray analysis, and the special design to detect known mutations recognized all 8 theoretically detectable deletions (Figure 3). In total, 48 different single-nucleotide variants (SNVs) with 415 occurrences were found among all 36 DNAs using SeqC with standard software parameters (Table 4). All sequence variants were compared to the breast cancer mutation database and the human gene mutation database. The clinical relevance of all mutations (single-nucleotide and deletions) was confirmed (Table 5). The clinical relevance of 15 SNVs was marked as “unknown” in at least one of the two databases. Two occurrences of variants exclusively detected with SeqC were of unknown clinical importance: Patient 6 – rs4987046 and rs28897762. Three SNVs (rs28897706, rs144848, rs11571833) led to completely contradicting results in both online databases (Table 5). These three polymorphisms were designated as recommended by the Breast Cancer Mutation Database as being of “no clinical relevance”. In total, 13 different SNVs marked as “unknown clinical relevance” were detected following the breast cancer mutation database. Two of them (rs1801499, rs28897762) result in a synonymous amino-acid change and are therefore not likely to be deleterious. The remaining 11 SNVs of “unknown clinical relevance” may contribute to HBOC. In total, 19 SNPs did not seem to be related to HBOC and another 5 SNPs could not be found. Two SNVs (*BRCA1* S784S, *BRCA2* C>T -143) were found neither in Ensembl (www.ensembl.org) nor both mutation databases. Due to the synonymous amino acid change it is likely that S784S is a new, neutral polymorphism described in this study. All eight detectable heterozygous deletions were properly detected manually (Table 4). In contrast to SeqC, GSeq is not able to evaluate the special perfect-match design for mutations.

Though all perfect-match deletion sequences (Table 2) were analyzed manually. An elevated relative fluorescence signal indicates a deletion in a patient. Figure 3 illustrates an example of manual deletion detection with patient 30.

Accuracy, specificity, and sensitivity of CustomSeq resequencing arrays have been proven by correctly recognizing 25 occurrences of previously analyzed SNVs within our positive controls. To confirm reproducibility, a total of six chips were amplified and hybridized at least twice with equal results using SeqC. Detecting hemizygous deletions remained laborious; we were not able to safely detect hemizygous deletions that were not designed on the chip.

Using SeqC and multiplexes, each patient took approximately three days, including analysis. PCR amplification and pooling took place on the first day. On the second day, the sample was processed and hybridized to the chip, then on the third day, staining, scanning, and analysis was carried out. All ambiguous calls were evaluated manually and therefore a very dense map with high informative value for all SNVs was generated.

Discussion

Microarray technology is an important tool to analyze cancer genomes and to understand the causes of cancer¹⁶. Sequencing microarrays have been appreciated as a valid tool for parallel and accurate resequencing of genes. In the last few years, the total number of different custom oligonucleotide resequencing microarrays has increased and this technique has been successfully applied in mitochondrial mutation analysis¹⁰, cardiomyopathy^{11,12}, intrahepatic cholestasis¹³, and retinal degeneration¹⁴. This study focused on two genes, *BRCA1* and *BRCA2*, that are causative for HBOC⁵. Altogether, 33 positive controls and 3 standard DNAs with a total of 945 kb were analyzed and all previously known SNVs were detected with high accuracy (100 %, 25/25). In addition to this, we were able to confirm high reproducibility by

repeating different experiments. Previous studies showed comparable results concerning accuracy (99.9 %-100 %) and reproducibility ¹⁰⁻¹⁴. These studies also reported call rates ranging from 93.5 % to 97.6 %.

A new program, SeqC (www.jsi-medisys.de), was used in our study to analyze the data. With the help of this tool, the average call rate could be raised from 97 % using the standard analysis-tool GSeq (www.affymetrix.com) to 99 %. While comparing the results of GSeq and SeqC, no loss of information was detected and known single nucleotide mutations could be evaluated in both programs. Though most of the SNVs were detected with both GSeq and SeqC, we were able to find 35 additional occurrences of 16 SNVs with SeqC (Figure 1), and three SNVs were exclusively found in SeqC.

Additional features of SeqC facilitated the analysis of the data. In the end, an average of 15 ambiguous calls (Ns) had to be evaluated manually for both genes. A statistical module compared each nucleotide with already archived arrays and visualized the average fluorescence intensity and standard deviation at each position. An electropherogram indicated current signal intensities and their course. The most time-consuming factor, data analysis, was effectively reduced, saving hands-on time and money. In addition, we successfully established multiplexes (Table 9) and were able to further reduce pipetting-time and use of chemicals.

Two approaches for detection of deletions and insertions are theoretically possible. One is to design special perfect match (PM) oligonucleotides for deleted or inserted sequences ^{17,18}. These oligonucleotides led to an increase in hybridization signal for every patient with that specific deletion compared to wild type sequence. This approach has the advantage of both detecting the presence and identifying the sequence change, but is limited by the size of the microarray and the total number of mutation-specific oligonucleotides. The second possibility is to detect sequence changes based on a loss of hybridization signal within the wild type sequence. Consequently, the sequence variation is only detected but then needs further

identification (for example by sequencing). This approach has been described in the literature to be effective with deletions of more than 50 bp¹⁶. Since insertions and deletions are important in *BRCA1* and *BRCA2*, perfectly match oligonucleotides were designed on the chip for the most important deletions and insertions (Table 2). With the help of these oligonucleotides, all eight detectable heterozygous deletions were detected with respect to the hybridization signal (Figure 3), however, *de novo* detection of heterozygous deletions due to the loss of hybridization signal remains difficult. Large genomic heterozygous deletions, which may occur in *BRCA1*, can theoretically be identified using special designed oligonucleotide probes at the end or the beginning of a specific deletion. Since both, standard Sanger-Sequencing and oligonucleotide resequencing microarrays, have difficulties in detecting large heterozygous deletions a second (screening) method such as multiplex ligation-dependent probe amplification may be used before (re)sequencing¹⁹. In contrast to heterozygous deletions, *de novo* homozygous deletions can be detected easily by the total loss of hybridization signal.

Sequence changes in *BRCA1* and *BRCA2* can be subdivided by their clinical relevance (yes/no/unknown, see <http://research.nhgri.nih.gov/projects/bic/>). For sequence changes with questionable relationships to HBOC, recent studies have tried to develop and apply models to predict their clinical relevance²⁰⁻²². Though all examined patients in this study were previously characterized, screening for mutations ended whenever a deleterious sequence change was found. With the help of the Breast Cancer Microarray, we were able to construct a complete map of single-nucleotide variations for each patient. With this additional information, a total of 19 SNPs that were not thought to be relevant in HBOC and another 5 SNPs that did not appear in the BIC database could be identified. Thirteen SNVs of unknown relevance in regards to HBOC were found and further studies are needed to prove their clinical relevance.

The results of this study, as well as those in the literature with oligonucleotide resequencing microarrays, are comparable to sequencing techniques like the standard Sanger sequencing and so-called second-generation sequencing tools ²³. Oligonucleotide resequencing microarrays are equal in sensitivity and specificity to both techniques. An advantage of oligonucleotide microarrays is its cost-effectiveness; about 70 % less compared to standard Sanger sequencing. For second-generation sequencing tools, further evaluation is needed and costs for routine usage must decrease prior to broad application.

Mutations in *BRCA1* and *BRCA2* are the cause for the hereditary breast-ovarian cancer syndrome (HBOC). HBOC is the underlying syndrome of about one-third of all hereditary breast cancer cases ⁵ and accounts for 15 % of all cases of ovarian cancer, the majority of hereditary ovarian cancers ⁷. It has been well established that patients with a family risk of breast or ovarian cancer need to be included in special, extended screening programs, which are based on general screening techniques like monthly breast self examination, clinical breast examination (every six months), and imaging (annual mammography or magnet resonance imaging), which will help to detect and treat breast cancer at the early stages ²⁴. Despite this secondary prevention, all patients with a family risk for HBOC should undergo genetic counseling and genetic testing ²⁴. For patients with known HBOC, effective broad-spectrum prophylactic treatment can be offered, ranging from chemopreventive medication to prophylactic surgery (bilateral mastectomy, salpingo-oophorectomy). Primary prevention activities should also be carried out with a multidisciplinary team with geneticists and psycho-oncologists ²⁴. Since an effective cancer screening and prophylactic treatment is possible, appropriate risk assessment and genetic testing is essential. At this point the high-throughput oligonucleotide resequencing microarray presented in this study, spanning the complete sequence of *BRCA1* and *BRCA2*, has proven to be safe, fast, and affordable. Future applications of this tool might comprise assessment of modifying sequence changes and

resequencing of other genes known to contribute to breast cancer, e.g. TP53, PTEN, STK11 and CDH1²⁵.

Conclusions

This study successfully establishes the first high-throughput oligonucleotide resequencing microarray for the total sequence of *BRCA1* and *BRCA2* to supply a safe, fast, and affordable tool to diagnose SNVs, homozygous and selected heterozygous deletions in patients at high risk for HBOC. In addition a new software tool, SeqC, facilitated and fastened the previous laborious analysis of chip data by reducing the absolute number of Ns and clearly arranging the remaining data with the help of online databases.

Acknowledgements

We thank JSI medical systems and especially Joachim Strub and Dr. Volker Horejschi for assistance during data analysis. The analysis of BRCA patients was supported by a multi-center grant from the Deutsche Krebshilfe, Bonn, Germany.

References:

- 1 Karim-Kos HE, de Vries E, Soerjomataram I, Lemmens V, Siesling S, Coebergh JW (2008) Recent trends of cancer in Europe: a combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. *Eur J Cancer* 44:1345-1389
- 2 McPherson K, Steel CM, Dixon JM (2000) ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ* 321:624-628
- 3 Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250:1684-1689
- 4 Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D, et al. (1994) Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* 265:2088-2090
- 5 Lynch HT, Silva E, Snyder C, Lynch JF (2008) Hereditary breast cancer: part I. Diagnosing hereditary breast cancer syndromes. *Breast J* 14:3-13
- 6 King MC, Marks JH, Mandell JB (2003) Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* 302:643-646
- 7 Pal T, Permut-Wey J, Betts JA, Krischer JP, Fiorica J, Arango H, LaPolla J, Hoffman M, Martino MA, Wakeley K, Wilbanks G, Nicosia S, Cantor A, Sutphen R (2005) BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. *Cancer* 104:2807-2816
- 8 Bermejo-Perez MJ, Marquez-Calderon S, Llanos-Mendez A (2007) Effectiveness of preventive interventions in BRCA1/2 gene mutation carriers: a systematic review. *Int J Cancer* 121:225-231
- 9 Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS (1996) Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat Genet* 14:441-447
- 10 Maitra A, Cohen Y, Gillespie SE, Mambo E, Fukushima N, Hoque MO, Shah N, Goggins M, Califano J, Sidransky D, Chakravarti A (2004) The Human MitoChip: a high-throughput sequencing microarray for mitochondrial mutation detection. *Genome Res* 14:812-819
- 11 Fokstuen S, Lyle R, Munoz A, Gehrig C, Lerch R, Perrot A, Osterziel KJ, Geier C, Beghetti M, Mach F, Sztajzel J, Sigwart U, Antonarakis SE, Blouin JL (2008) A DNA resequencing array for pathogenic mutation detection in hypertrophic cardiomyopathy. *Hum Mutat* 29:879-885
- 12 Waldmüller S, Müller M, Rackebrandt K, Binner P, Poths S, Bonin M, Scheffold T (2008) Array-based resequencing assay for mutations causing hypertrophic cardiomyopathy. *Clin Chem* 54:682-687
- 13 Liu C, Aronow BJ, Jegga AG, Wang N, Miethke A, Mourya R, Bezerra JA (2007) Novel resequencing chip customized to diagnose mutations in patients with inherited syndromes of intrahepatic cholestasis. *Gastroenterology* 132:119-126
- 14 Mandal MN, Heckenlively JR, Burch T, Chen L, Vasireddy V, Koeneke RK, Sieving PA, Ayyagari R (2005) Sequencing arrays for screening multiple genes associated with early-onset human retinal degenerations on a high-throughput platform. *Invest Ophthalmol Vis Sci* 46:3355-3362
- 15 Hacia JG (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet* 21:42-47
- 16 Cowell JK, Hawthorn L (2007) The application of microarray technology to the analysis of the cancer genome. *Curr Mol Med* 7:103-120

- 17 Karaman MW, Groshen S, Lee CC, Pike BL, Hacia JG (2005) Comparisons of substitution, insertion and deletion probes for resequencing and mutational analysis using oligonucleotide microarrays. *Nucleic Acids Res* 33:e33
- 18 Cutler DJ, Zwick ME, Carrasquillo MM, Yohn CT, Tobin KP, Kashuk C, Mathews DJ, Shah NA, Eichler EE, Warrington JA, Chakravarti A (2001) High-throughput variation detection and genotyping using microarrays. *Genome Res* 11:1913-1925
- 19 Bunyan DJ, Eccles DM, Sillibourne J, Wilkins E, Thomas NS, Shea-Simonds J, Duncan PJ, Curtis CE, Robinson DO, Harvey JF, Cross NC (2004) Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification. *Br J Cancer* 91:1155-1159
- 20 Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian SV, Monteiro AN, Iversen ES, Couch FJ, Goldgar DE (2007) A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *Am J Hum Genet* 81:873-883
- 21 Lee E, McKean-Cowdin R, Ma H, Chen Z, Van Den Berg D, Henderson BE, Bernstein L, Ursin G (2008) Evaluation of unclassified variants in the breast cancer susceptibility genes BRCA1 and BRCA2 using five methods: results from a population-based study of young breast cancer patients. *Breast Cancer Res* 10:R19
- 22 Spearman AD, Sweet K, Zhou XP, McLennan J, Couch FJ, Toland AE (2008) Clinically Applicable Models to Characterize BRCA1 and BRCA2 Variants of Uncertain Significance. *J Clin Oncol*
- 23 Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135-1145
- 24 Silva E, Gatalica Z, Snyder C, Vranic S, Lynch JF, Lynch HT (2008) Hereditary breast cancer: part II. Management of hereditary breast cancer: implications of molecular genetics and pathology. *Breast J* 14:14-24
- 25 Campeau PM, Foulkes WD, Tischkowitz MD (2008) Hereditary breast cancer: new genetic developments, new therapeutic avenues. *Hum Genet* 124:31-42

Figure Legends:

- **Figure 1:** Illustration of two SNPs missed by GSeq but automatically detected using SeqC.

The electropherogram shown in this figure is used for an additional manual evaluation. Heterozygous SNVs are presented as elevated fluorescence-signals for two bases at the same time (A – Patient 32), whereas homozygous SNVs are characterized by a single-base, normal fluorescence-signal at the centre position and a lack of fluorescence-signal around the centre position due to a mismatch (B – Patient 11).

- **Figure 2:** Detection of mutations.

The detection of mutations corresponds to the detection of SNVs. This figure shows two mutations that were automatically detected by SeqC as well as GSeq. As part of the analysis, SeqC indicates the correct cDNA position and amino acid change in the protein. Both mutations (A – Patient 18, B – Patient 19) shown in this image lead to a frame shift and result in early termination of the protein.

- **Figure 3:** Detection of deletions.

Most important deletions were designed as perfect match probes on the oligonucleotide resequencing microarray. Samples with that special deletion resulted in a perfect match and normal hybridization signal. DNA without the mutation is theoretically not able to hybridize to the deletion-specific sequence and results in little or no signal. The signal intensity of heterozygous mutation carriers will decrease but is still maintained at significantly elevated level. Since software algorithms cannot consider perfect match deletion sequences, each probe has to be checked manually for all patients. For illustration purposes, the quotient of relative signal-intensities mutated / wildtype for a deletion-specific sequence (*BRCA1* 4184delTCAA) was plotted on the y-axis. The data shown is based on Patient 30, heterozygous for *BRCA1*

4184delTCAA, and one control (patient 3). The increase of signal in relation to the wild type sequence at the central position indicates the presence of the deletion.

Table 1: Overview on genes associated with a high risk for hereditary breast cancer (HBC)

24

Gene	Locus	Ensembl ID	Protein
<i>BRCA1</i>	17q21	ENSG00000012048	Breast cancer type 1 susceptibility protein
<i>BRCA2</i>	13q12	ENSG00000139618	Breast cancer type 2 susceptibility protein
<i>TP53</i>	17p13	ENSG00000141510	Tumor suppressor p53
<i>PTEN</i>	10q23	ENSG00000171862	Phosphatase tensin homolog on chromosome ten
<i>STK11</i>	19p13	ENSG00000118046	Serine/threonine-protein kinase 11
<i>CDH1</i>	16q22	ENSG00000039068	Epithelial cadherin Precursor

Table 2: Specific deletions designed on the Breast Cancer Microarray (annotation following BIC –<http://research.nhgri.nih.gov>).

Gene	Exon	Deletion	Gene	Exon	Deletion
<i>BRCA1</i>	2	185delAG	<i>BRCA2</i>	2	IVS2-7delT
	6	IVS6-2delA		2	L15ins
	11	D369del		3	N55del
	11	I482del		3	EL97del
	11	S616del		5	V145del
	11	S1140del		7	802delAT
	11	Q1281del		8	E215del
	11	1675delA		8	886delGT
	11	2576delC		9	R239ins
	11	2594delC		9	983del4
	11	2800delAA		10	1983del5
	11	2804delAA		10	2041insA
	11	3450del4		11	5301insA
	11	3600del11		11	T1302del
	11	3819del5		11	S976del
	11	3875del4		11	K1286del
	11	S1297del		11	E1382del
	11	4154delA		11	6503delTT
	11	4184del4		11	6174delT
	17	V1688del		11	5950delCT
17	H1673del	11		5804del4	
18	A1693del	11		5578delAA	
19	Q1721del	11		4706delAAAAG	
20	G1738del	11		4075delGT	
		11		3036del4	
		11	2157delG		
		14	7297delCT		
		15	K2498del		
		18	K2750del		
		18	8525delC		
		22	9132delC		
		25	9663delGT		

Table 3: PCR-Design for Multiplex-Reactions

Multiplex	Gene	Exon
1	<i>BRCA1</i>	1, 4, 9, 14

	<i>BRCA2</i>	21
2	<i>BRCA1</i>	3, 8, 15
	<i>BRCA2</i>	13
3	<i>BRCA1</i>	5, 16, 17, 23
4	<i>BRCA1</i>	13, 18
	<i>BRCA2</i>	1, 3
5	<i>BRCA1</i>	11, 12, 20
	<i>BRCA2</i>	4, 7
6	<i>BRCA1</i>	21, 22
	<i>BRCA2</i>	2, 9, 16
7	<i>BRCA2</i>	12, 14, 18, 23
8	<i>BRCA2</i>	17, 22, 24, 25
9	<i>BRCA1</i>	19
	<i>BRCA2</i>	5, 6, 8, 20, 26

Table 4: Overview of all identified sequence changes including mutations, single-nucleotide variants, chip call-rate and total amount of ambiguous calls (Ns) for each patient. Mutations in brackets were not designed on the chip. Asterisk-marked DNAs were amongst the first microarrays and are therefore of lower quality. Though these arrays were included to demonstrate the noise reduction of SeqC in comparison to GSeq.

DNA	Mutation (cDNA)	Call rate Chip SeqC / GSeq	SNVs SeqC / GSeq	Polymorphisms	Mutation	Ns (SeqC)
				(SeqC)	(SeqC)	
1	-	99 % / 99 %	20 / 115	17	-	3
2	-	99 % / 98 %	27 / 504	12	-	15
3	-	99 % / 98 %	27 / 440	14	-	13
4	-	99 % / 97 %	28 / 399	14	-	14
5	-	99 % / 96 %	159 / 441	15	-	*
6	-	99 % / 97 %	20 / 297	17	-	3
7	-	99 % / 98 %	17 / 100	13	-	4
8	-	99 % / 98 %	19 / 161	14	-	5
9	(<i>BRCA2</i> 8358delT)	99 % / 98 %	23 / 205	11	-	12
10	(<i>BRCA2</i> 1727delG)	99 % / 98 %	33 / 165	6	-	27
11	(<i>BRCA2</i> 3972delTGAG)	99 % / 97 %	18 / 176	13	-	5
12	(<i>BRCA2</i> 1257delA)	99 % / 97 %	16 / 487	6	-	10
13	(<i>BRCA2</i> 6503delTT)	99 % / 96 %	240 / 3062	5	-	*
14	(<i>BRCA1</i> 3040insT)	99 % / 97 %	63 / 335	13	-	50
15	(<i>BRCA1</i> 3874delTGTC)	99 % / 96 %	172 / 468	14	-	*
16	(<i>BRCA1</i> 5382insC)	99 % / 97 %	88 / 293	14	-	74
17	<i>BRCA1</i> T>G 5228	99 % / 96 %	166 / 886	4	1	*
18	<i>BRCA1</i> C>T 297	99 % / 97 %	38 / 209	13	1	24
19	<i>BRCA1</i> C>G 4808	99 % / 94 %	247 / 2676	5	1	*
20	<i>BRCA1</i> G>A 5629	99 % / 95 %	219 / 629	14	1	*
21	<i>BRCA1</i> IVS 21 G>A -1	99 % / 98 %	24 / 174	6	1	17
22	<i>BRCA1</i> C>T 297	99 % / 97 %	45 / 442	15	1	29
23	<i>BRCA1</i> C>T 4302	99 % / 98 %	29 / 219	14	1	14
24	<i>BRCA1</i> C>A 2428	99 % / 99 %	10 / 82	6	1	3
25	<i>BRCA1</i> G>A 4304	99 % / 98 %	15 / 100	12	1	2
26	<i>BRCA1</i> C>T 4341	99 % / 98 %	19 / 136	15	1	3
27	<i>BRCA1</i> C>G 4808	99 % / 98 %	9 / 94	6	1	2
28	<i>BRCA2</i> A>T 6265	99 % / 98 %	14 / 182	10	1	3
29	<i>BRCA1</i> 4184delTCAA	99 % / 97 %	33 / 223	14	-	19
30	<i>BRCA1</i> 4184delTCAA	99 % / 96 %	94 / 362	11	-	*

31	<i>BRCA2</i> 4075delGT	99 % / 98 %	26 / 171	5	-	21
32	<i>BRCA1</i> 3874delTGTC	99 % / 98 %	28 / 200	14	-	14
33	<i>BRCA1</i> 3600delGAAGATACTAG	99 % / 98 %	38 / 233	15	-	23
34	<i>BRCA1</i> 3600delGAAGATACTAG	99 % / 98 %	18 / 157	11	-	7
35	<i>BRCA1</i> 4184delTCAA	99 % / 98 %	6 / 95	4	-	2
36	<i>BRCA2</i> 3036delACAA	99 % / 98 %	12 / 102	8	-	4

Table 5: Single-nucleotide variants (SNV) found among all screened DNAs and their clinical importance following HGMD and BIC (=not listed; ?=unknown / variant of unknown significance; yes=clinical important; no=clinical unimportant, Mut.=mutation). Known deletions are marked in gray.

Gene	Exon	cDNA	Amino Acid	Accession	Frequency	HGMD	BIC
<i>BRCA1</i>	5	C>T 297	Q60X	Mut.	2	-	yes
<i>BRCA1</i>	5	T>G 300	C61G	Mut., rs28897672	1	yes	yes
<i>BRCA1</i>	5	IVS 5 +3 A>G	3' UTR	Mut.	1	yes	?
<i>BRCA1</i>	11	A>G 1186	Q356R	rs1799950	4	?	?
<i>BRCA1</i>	11	G>A 2196	D693N	Polymorphism	2	?	no
<i>BRCA1</i>	11	C>T 2201	S694S	rs1799949	21	-	no
<i>BRCA1</i>	11	C>A 2428	S770X	Mut.	1	yes	yes
<i>BRCA1</i>	11	T>C 2430	L771L	rs16940	25	no	no
<i>BRCA1</i>	11	G>A 2471	S784S		1	-	-
<i>BRCA1</i>	11	C>T 2731	P871L	rs799917	24	-	no
<i>BRCA1</i>	11	A>G 3232	E1038G	rs16941	25	?	no
<i>BRCA1</i>	11	G>A 3238	S1040N	rs4986852	1	?	?
<i>BRCA1</i>	11	A>G 3667	K1183R	rs16942	25	-	no
<i>BRCA1</i>	11	A>G 4158	R1347G	rs28897689	2	?	?
<i>BRCA1</i>	12	C>T 4302	Q1395X	Mut.	1	yes	yes
<i>BRCA1</i>	12	G>A 4304	Q1395Q	Mut.	1	-	yes
<i>BRCA1</i>	13	C>T 4341	Q1408X	Mut.	1	yes	yes
<i>BRCA1</i>	13	T>C 4427	S1436S	rs1060915	25	-	no
<i>BRCA1</i>	16	C>G 4808	Y1563X	Mut.	2	yes	yes
<i>BRCA1</i>	16	A>G 4956	S1613G	rs1799966	25	-	no
<i>BRCA1</i>	16	G>A 5075	M1652I	rs1799967	2	?	?
<i>BRCA1</i>	18	T>G 5228	Y1703X	Mut.	1	-	yes
<i>BRCA1</i>	21	IVS 21 G>A -1	5' UTR	Mut.	1	-	yes
<i>BRCA1</i>	24	G>A 5629	W1837X	Mut.	1	yes	yes
<i>BRCA2</i>	1	C>T -143	5' UTR		1	-	-
<i>BRCA2</i>	1	A>G -52	5' UTR	rs206118	13	-	-
<i>BRCA2</i>	2	G>A -26	5' UTR	rs1799943	17	-	-
<i>BRCA2</i>	3	A>G125	Y42C	rs4987046	1	?	?
<i>BRCA2</i>	10	A>C 1093	N289H	rs766173	1	-	no
<i>BRCA2</i>	10	C>A 1206	S326R	rs28897706	1	yes	no
<i>BRCA2</i>	10	C>A 1342	H372N	rs144848	14	yes	no
<i>BRCA2</i>	10	A>G 1593	S455S	rs1801439	1	-	no
<i>BRCA2</i>	11	T>C 2457	H743H	rs1801499	1	-	?
<i>BRCA2</i>	11	A>G 3199	N991D	rs1799944	1	-	no
<i>BRCA2</i>	11	A>G 3624	K1132K	rs1801406	20	-	no
<i>BRCA2</i>	11	T>C 4035	V1269V	rs543304	16	-	no
<i>BRCA2</i>	11	A>G 4791	L1521L	rs206075	36	-	no
<i>BRCA2</i>	11	C>T 5427	S1733S	rs28897734	2	-	no
<i>BRCA2</i>	11	T>C 5972	M1915I	rs4987117	2	?	?

<i>BRCA2</i>	11	A>T 6265	K2013X	Mut.	1	yes	yes
<i>BRCA2</i>	11	C>T 6328	R2034C	rs1799954	1	?	?
<i>BRCA2</i>	11	G>C 6741	V2171V	rs206076	36	-	no
<i>BRCA2</i>	14	G>A -26	5' UTR	rs1799943	1	-	-
<i>BRCA2</i>	14	A>G 7470	S2414S	rs1799955	15	-	no
<i>BRCA2</i>	14	T>C 7625	V2466A	rs169547	36	no	no
<i>BRCA2</i>	22	G>T 9078	K2950N	rs28897754	1	yes	?
<i>BRCA2</i>	27	G>A10110	R3370R	rs28897762	1	-	?
<i>BRCA2</i>	27	A>T 10204	K3326X	rs11571833	1	yes	no

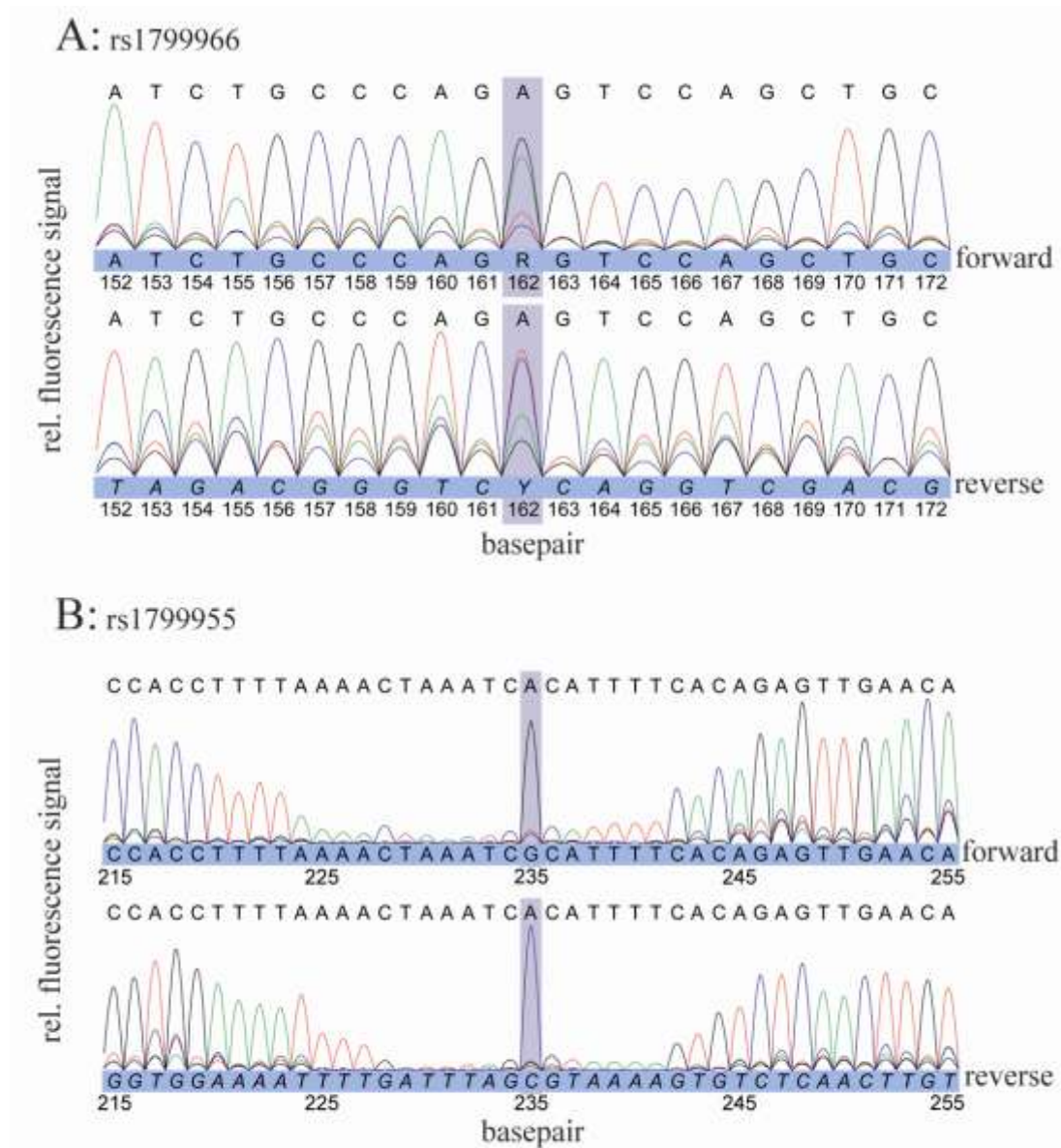


Figure 1

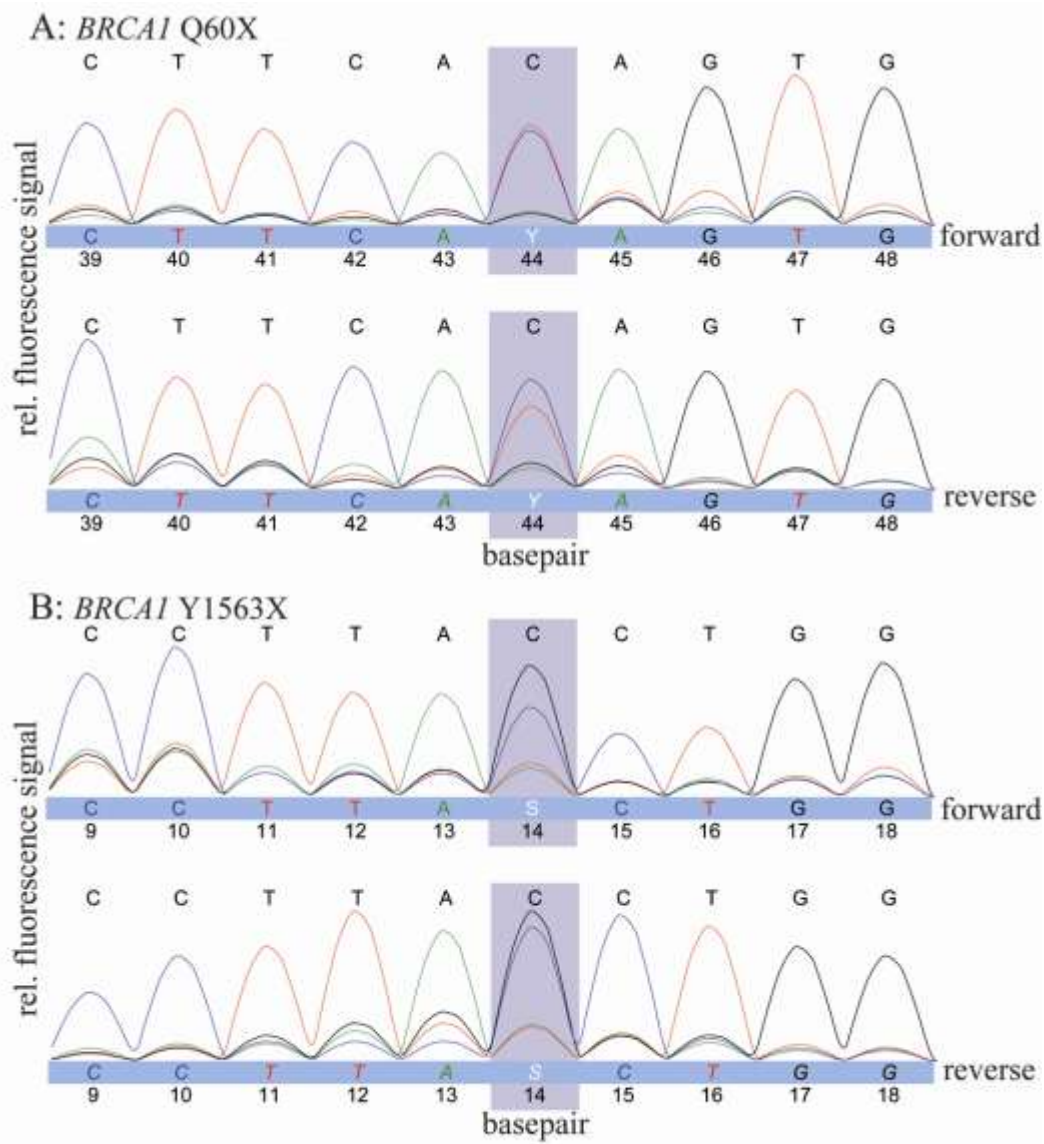


Figure 2

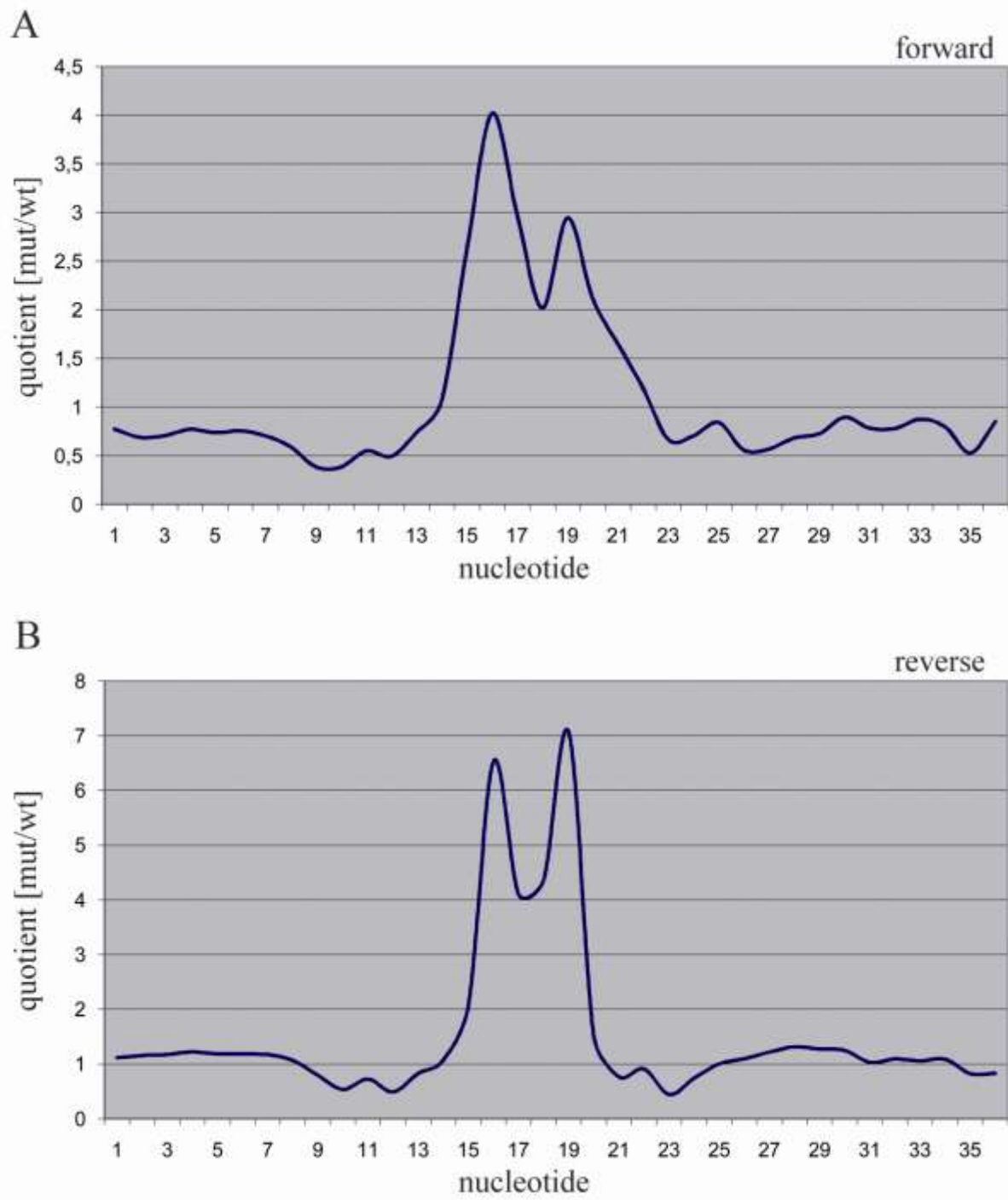


Figure 3