The Journal of Molecular Diagnostics, Vol. 18, No. 4, July 2016

ELSEVIER

the Journal of Molecular Diagnostics

jmd.amjpathol.org

Detection of Germline Mutation in Hereditary Breast and/or Ovarian Cancers by Next-Generation Sequencing on a Four-Gene Panel

Ava Kwong,*†‡ Vivian Y. Shin,* Chun H. Au,§ Fian B.F. Law,‡§ Dona N. Ho,§ Bui K. Ip,§ Anthony T.C. Wong,§ Silvia S. Lau,¶ Rene M.Y. To,§ Gigi Choy,§ James M. Ford,‖ Edmond S.K. Ma,‡§ and Tsun L. Chan‡§

From the Department of Surgery, The University of Hong Kong, Hong Kong, People's Republic of China; the Departments of Surgery,† Molecular Pathology,§ and Medical Physics and Research,¶ Hong Kong Sanatorium & Hospital, Hong Kong, People's Republic of China; the Hong Kong Hereditary Breast Cancer Family Registry,‡ Hong Kong, People's Republic of China; and the Department of Medicine (Oncology),‖ Stanford University School of Medicine, Stanford, California*

Accepted for publication March 21, 2016.

Address correspondence to Ava Kwong, MBBS, Ph.D., Room K1401, Queen Mary Hospital, Pokfulam Road, Pokfulam, Hong Kong, People's Republic of China. E-mail: akwong@asiabreastregistry.com.

Mutation in *BRCA1/BRCA2* genes accounts for 20% of familial breast cancers, 5% to 10% of which may be due to other less penetrant genes which are still incompletely studied. Herein, a four-gene panel was used to examine the prevalence of *BRCA1*, *BRCA2*, *TP53*, and *PTEN* in hereditary breast and ovarian cancers in Southern Chinese population. In this cohort, 948 high-risk breast and/or ovarian patients were recruited for genetic screening by next-generation sequencing (NGS). The performance of our NGS pipeline was evaluated with 80 Sanger-validated known mutations and eight negative cases. With appropriate bioinformatics analysis pipeline, the detection sensitivity of NGS is comparable with Sanger sequencing. The prevalence of *BRCA1/BRCA2* germline mutations was 9.4% in our Chinese cohort, of which 48.8% of the mutations arose from hotspot mutations. With the use of a tailor-made algorithm, HomopolymerQZ, more mutations were detected compared with single mutation detection algorithm. The frequencies of *PTEN* and *TP53* were 0.21% and 0.53%, respectively, in the Southern Chinese patients with breast and/or ovarian cancers. High-throughput NGS approach allows the incorporation of control cohort that provides an ethnicity-specific data for polymorphic variants. Our data suggest that hotspot mutations screening such as SNaPshot could be an effective preliminary screening alternative adopted in a standard clinical laboratory without NGS setup. *(J Mol Diagn 2016, 18: 580—594; http://dx.doi.org/10.1016/j.jmoldx.2016.03.005)*

Breast cancer is the most common diagnosed cancer in women worldwide, with 1.67 million newly diagnosed cases in 2012, which accounts for 25% of all cancers. Besides, it is the fifth common cause of cancer death worldwide and has higher mortality rates in the East than in the West (International Agency for Research on Cancer, *http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx*, last accessed March 1, 2016). Hereditary breast and ovarian cancer (HBOC) syndrome is commonly associated with genetic alterations in *BRCA1* and *BRCA2*, in which the lifetime risk of developing breast cancer by the age of 70 years increased by fourfold to fivefold in *BRCA1/BRCA2* carriers (45% to 85%) compared with the general population (approximately 12.5%).[1,2] The prevalence and risk management of *BRCA1/BRCA2* carriers in patients with breast and/or ovarian cancers are addressed by previous preclinical and clinical studies, resulting in recommendations for screening and risk-reduction intervention for *BRCA1/BRCA2* carriers and their family members.[3] Knowing the germline mutation status will not only help the identification of high-risk patients but also change the screening, cancer risk management, and therapeutic strategies.

Breast cancer mortality to incidence ratio is higher in Asia than in Western countries.[4] This fast growing trend of breast

Supported by the Dr. Ellen Li Charitable Foundation, the Kerry Kuok Foundation, Health and Medical Research Fund grant 1123176 (A.K.), and the Hong Kong Hereditary Breast Cancer Family Registry.

Disclosures: None declared.

http://dx.doi.org/10.1016/j.jmoldx.2016.03.005

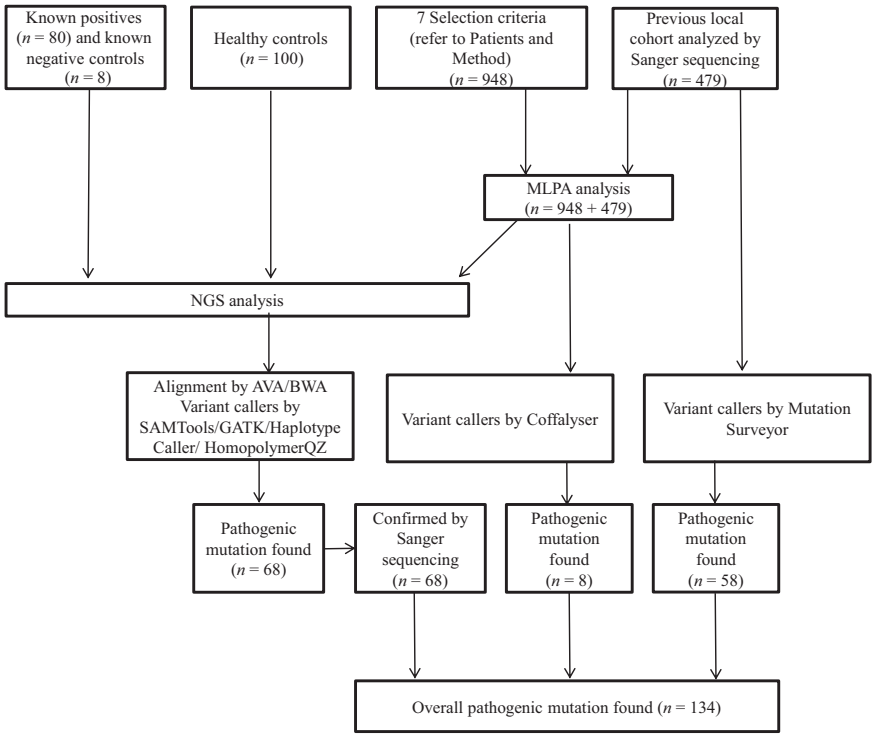

Figure 1 A schematic workflow of the mutation screening in this study. AVA, Amplicon Variant Analyzer; MLPA, multiplex ligation-dependent probe amplification; NGS, next-generation sequencing.

cancer incidence in Asia, especially China and India where the population is large, is anticipated to impose a serious health economic problem. To this end, cost-effective and high-throughput genetic screening in these regions is highly desirable. Nonetheless, genetic testing is rather uncommon in many regions except for a few countries because of cost considerations and availability; hence, there is a need for increased knowledge or data support to implement population-wide preventive health care programs in the community. In view of this, several research groups have joined to form the Asian BRCA Consortium, striving to understand better HBOC syndrome in Asians through research, aiming to make genetic testing as a standard of care for breast cancer similar to that of the Western world.[5,6]

Table 1 Clinicopathologic Data of the Cohort Patients

Characteristic	Current cohort, *n* (%) (*n* = 948)	Cumulative cohort, *n* (%) (*n* = 1427)
Age at diagnosis (years)		
≤45	549 (58)	835 (58)
>45	399 (42)	592 (42)
Bilateral BrCa		
Yes	151 (16)	218 (15)
No	797 (84)	1209 (85)
Triple-negative BrCa		
Yes	124 (13)	213 (15)
No	824 (87)	1214 (85)
Multiple Ca		
Yes	55 (6)	95 (7)
No	893 (94)	1332 (93)
Family history of BrCa/ovarian Ca		
Yes	360 (38)	554 (39)
No	588 (62)	873 (61)
Family history of other Ca		
Yes	691 (73)	1042 (73)
No	257 (27)	385 (27)

BrCa, breast cancer; Ca, cancer.

Our previous studies found that defective *BRCA1* and *BRCA2* genes play important roles in women with early onset of breast cancer in Hong Kong.[7] Moreover, women with triple-negative breast cancer and bilateral breast cancers were prone to genetic predisposition.[8] From these results, patients with these types of breast cancers are now recommended to have genetic counseling and testing.[9] The use of direct PCR-based Sanger sequencing and high resolution melting analysis could identify germline and hotspot mutations, but the turnaround time does not seem to satisfy the needs of the community. Increasing evidence suggests that ovarian cancer patients with *BRCA1/BRCA2* mutations have a better response to olaparib, implicating the need of faster turnaround time of the genetic test for better clinical management.

Furthermore, identification of ethnic-specific founder and hotspot mutations detected in Chinese patients would be beneficial, not only for the patients locally where most are Chinese but also the enormous Chinese population in Mainland China and those residing in Western countries. The development of a genetic testing panel was designed specifically for the population to achieve better risk assessment and preventive measures.

Next-generation sequencing (NGS), based on clonal amplification coupled with massively parallel sequencing, is a powerful DNA sequencing tool that has been broadly

Table 2 Evaluation of Mutation Detection Performance from 54 Mutation Controls in Local Cohort

Gene	DNA mutation type	Mutation	Allele frequency (%)	Variant callers		
				AVA	SAMtools	HomopolymerQZ
BRCA1	Deletion	c.470_471delCT	56.6	Yes	No	NA
		c.959_960delGA	55.0	Yes	Yes	NA
		c.964delG	55.2	Yes	Yes	NA
		c.981_982delAT	52.6	Yes	No	NA
		c.1016delA*		No	No	Yes (*z* score, 4.56)
		c.3214delC	39.1	Yes	Yes	NA
		c.3333delA	24.3	Yes	No	NA
		c.3342_3345delAGAA	46.3	Yes	Yes	NA
		c.3858_3861delTGAG	48.5	Yes	No	NA
		c.4065_4068delTCAA	50.0	Yes	Yes	NA
		c.4764delT	31.0	Yes	Yes	NA
		c.5406_5406+2delAGT	50.0	Yes	Yes	NA
	Duplication	c.1874_1877dupTAGT	55.6	Yes	Yes	NA
		c.1961dupA*		No	No	Yes (*z* score, 4.67)
		c.4049dupG	50.0	Yes	Yes	NA
		c.4695dupA	50.4	Yes	Yes	NA
	Nonsense SNV	c.34C>T	46.7	Yes	Yes	NA
		c.502A>T	57.1	Yes	Yes	NA
		c.2338C>T	49.2	Yes	Yes	NA
		c.2635G>T	63.3	Yes	Yes	NA
		c.3286C>T	55.6	Yes	Yes	NA
		c.4573C>T	61.1	Yes	Yes	NA
		c.4656C>G	42.9	Yes	Yes	NA
		c.4675G>C	53.0	Yes	Yes	NA
		c.4903G>T	54.7	Yes	Yes	NA
	Splice site SNV	c.212+1G>A	50.8	Yes	Yes	NA
		c.212+1G>T	33.3	Yes	Yes	NA
		c.5193+1G>C	48.5	Yes	Yes	NA
BRCA2	Deletion	c.2595delA	36.7	Yes	Yes	NA
		c.2808_2811delACAA	56.1	Yes	Yes	NA
		c.3202delG	41.5	Yes	Yes	NA
		c.3836delA	54.5	Yes	Yes	NA
		c.4121delA	69.2	Yes	Yes	NA
		c.4563_4564delGT	57.8	Yes	Yes	NA
		c.5218_5223delTTAAGT	61.5	Yes	No	NA
		c.5722_5723delCT	50.0	Yes	Yes	NA
		c.8068_8069delGT	55.4	Yes	Yes	NA
		c.9393delC	56.0	Yes	Yes	NA
	Deletion and Insertion	c.4467_4474delinsTGTTTTT*	53.8	Multiple SNV	Multiple SNV	NA
	Duplication	c.5238dupT	55.4	Yes	Yes	NA
		c.6096dupT	46.4	Yes	Yes	NA
		c.7409dupT	50.0	Yes	Yes	NA
		c.7467dupT	42.9	Yes	Yes	NA
		c.9097dupA*		No	No	Yes (*z* score, 4.34)
	Nonsense SNV	c.250C>T	45.1	Yes	Yes	NA
		c.1261C>T	51.0	Yes	Yes	NA
		c.3109C>T	55.6	Yes	Yes	NA
		c.7471C>T	41.9	Yes	Yes	NA
		c.7878G>A	57.8	Yes	Yes	NA
		c.9294C>G	38.6	Yes	Yes	NA
		c.10150C>T	53.6	Yes	Yes	NA
	Splice site SNV	c.7806-9T>G	48.8	Yes	Yes	NA
		c.7976+1G>A	59.2	Yes	Yes	NA
		c.7976+5G>A	49.4	Yes	Yes	NA

Twenty-six overseas controls are listed in Supplemental Table S1.

*Variants missed by the variant caller.

AVA, Amplicon Variant Analyzer; NA, not applicable for non-homopolymer loci; SNV, single nucleotide variant.

applied in clinical diagnostic use.[10–12] It allows the investigation of multiple genes at a lower cost. However, the accuracy of mutation detection and the appropriate annotation of large numbers of variants of unknown significance (VUS) remain the challenges in data interpretation.[13]

In this study, we used microfluidic access array for PCR amplification. A total of >4600 PCR reactions can be performed in one single reaction with 48 barcoded patients. Besides, we further expanded the gene panel to four genes, including *TP53* and *PTEN*, where there is evidence to show the association with breast cancer and prostate cancer found in HBOC families, respectively.[14,15] We also found that 454 pyrosequencing provides a set of highly traceable and reproducible quantitative/semiquantitative data through different methods of analyses, through which we devised a more accurate variant calling bioinformatics pipeline. Finally, introducing a population-specific normal cohort strengthened the database of single nucleotide polymorphisms (SNPs) from The 1000 Genomes Project and significantly enhanced the analysis of VUS.

Materials and Methods

Ethics Statement

The study was performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants recruited in this study. This study was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority West Cluster and other contributing hospitals in Hong Kong.

Samples and Selection Criteria

A cohort of 1427 clinically high-risk breast and/or ovarian cancer probands was recruited by the Hong Kong Hereditary and High Risk Breast Cancer Program from March 2007 to October 2014. The patient selection criteria were described previously with modification.[16] High-risk patients were recruited if they fulfilled one of the following criteria: i) had at least one first- or second-degree relative with breast and/or ovarian cancer, regardless of age; ii) diagnosed age ≤45 years with cancer; iii) had bilateral breast cancer; iv) had triple-negative or medullary type pathology; v) had at least one relative with cancers other than breast and/or ovarian cancer such as stomach and prostate that are known to be related to *BRCA1/BRCA2* mutations; vi) male breast cancer; or vii) were ovarian cancer patients with a family history of breast/ovarian or HBOC-associated cancer. This cohort comprised 479 probands who were previously characterized by Sanger sequencing and 948 probands that underwent NGS screening (Figure 1).[16] Concurrent with sequencing, all patients were tested for large genomic rearrangement of *BRCA1* and *BRCA2* by multiplex ligation-dependent probe amplification (MLPA). Clinicopathologic characteristics of patient cohorts are shown in Table 1.

To enhance the validation process, two independent control cohorts, namely 88 Sanger-validated controls [positive (n = 80): 54 local Chinese and 26 overseas) and negative (n = 8)] and Sanger naive negative or healthy controls (n = 100) were used. For the evaluation of NGS performance, control samples were selected from previously characterized probands of the same cohort with *BRCA1* and *BRCA2* mutations (based on bidirectional Sanger full gene sequencing). There were 80 (44 *BRCA1* and 36 *BRCA2* from both local and overseas patients) different varieties of Sanger-validated positive samples subjected to 454 GS Junior and MiSeq, respectively, for cross-validation [Table 2 and Supplemental Table S1 (controls provided by J.M.F.)]. Simultaneously, *BRCA1*-negative controls (n = 8) were also included. It was reported that polymorphic alleles from 100 healthy individuals would cover almost 80% of the VUS in the population.[17] Hence, healthy local individuals (n = 100) with no known history of cancer and unknown mutation status also underwent the same NGS sequencing analysis as the probands.

DNA Extraction

Genomic DNA samples for NGS or conventional Sanger sequencing were extracted from peripheral blood samples with the use of QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) or automatically with the use of QIAsymphony DNA Mini Kit (Qiagen) on QIAsymphony SP instrument (Qiagen). RNA samples were also extracted from peripheral blood samples manually with the use of QIAamp RNA Blood Mini Kit (Qiagen). In the case of splicing variant analysis in transcript level, RNA samples were reverse transcribed to cDNA samples with the use of Superscript III First Strand Synthesis System (Invitrogen, Carlsbad, CA).

Microfluidic PCR Amplification of Target Regions

A total of 91 PCR primer pairs were used to target all exons of *BRCA1, BRCA2*, *TP53*, and *PTEN*, and 10-bp intron-exon boundaries (except downstream of *BRCA1* exon 2 where 8 bp was covered) for amplification of genomic DNA samples (Table 3). Access Array System (Fluidigm, South San Francisco, CA) was used to generate separate pools of 91 amplicons from 48 DNA samples per run. A 10-bp barcode nucleotide sequence was incorporated in both ends of each amplicon for sample identification. A total of 96 different barcode sequences were used in rotation to minimize the usage of the same barcode in two consecutive Access Array runs. After Access Array amplification, the amplicon product of each sample and purified pooled samples were subjected to quality check with the use of Agilent DNA 1000 Kit on a 2100 Bioanalyzer system (Agilent Technologies, Santa Clara, CA).

Table 3 Sequences of Microfluidic PCR Primers

Amplicon	Gene	Exon	Forward and reverse primers
1	*BRCA1*	1	5′-TACGGTAGCAGAGACTTGGTCTCCAGTACCCCAGAGCATCAC-3′
			5′-ACACTGACGACATGGTTCTACAGGAAAGAGTGGGGGATTGGG-3′
2		2	5′-ACACTGACGACATGGTTCTACATCCCAAATTAATACACTCTTGTGCTG-3′
			5′-TACGGTAGCAGAGACTTGGTCTTGGAGAAAGGAAAAGACCCAAGG-3′
3		3	5′-ACACTGACGACATGGTTCTACAAGGTGTTTCCTGGGTTATGAAGG-3′
			5′-TACGGTAGCAGAGACTTGGTCTAAAGATGAGATGTGCACCCACAG-3′
4		5	5′-ACACTGACGACATGGTTCTACAACCTGTATAAGGCAGATGTCCCA-3′
			5′-TACGGTAGCAGAGACTTGGTCTGCTTATGCAGCATCCAAAAACAA-3′
5		6	5′-ACACTGACGACATGGTTCTACAAAAAATAGATTACAGATACAGAACTA-AAATTAACCTA-3′
			5′-TACGGTAGCAGAGACTTGGTCTGCTGTGGATTTAGGGCAGTATTATATC-3′
6		7	5′-TACGGTAGCAGAGACTTGGTCTGGAGGACTGCTTCTAGCCTGG-3′
			5′-ACACTGACGACATGGTTCTACACATTGTTGGTGTCTTAGCTTTAGTGA-3′
7		8	5′-ACACTGACGACATGGTTCTACAAGATTCTTCAAGGTGGGAACTGC-3′
			5′-TACGGTAGCAGAGACTTGGTCTGGGCATTTTTTCCAGGCATCATA-3′
8		9	5′-ACACTGACGACATGGTTCTACAACTAAATCCAACAAACAAAAAGGGT-3′
			5′-TACGGTAGCAGAGACTTGGTCTCCACTTCTCTGTATTACATACTAGCTTA-3′
9		10	5′-ACACTGACGACATGGTTCTACATCAGACCATACCACGACATTTGA-3′
			5′-TACGGTAGCAGAGACTTGGTCTCGTGCCCAGCAACCATTTC-3′
10		11	5′-ACACTGACGACATGGTTCTACACCAAGGAACATCTTCAGTATCTCTAGG-3′
			5′-TACGGTAGCAGAGACTTGGTCTTGACAATTCAGTTTTTGAGTACCTTGTT-3′
11			5′-TACGGTAGCAGAGACTTGGTCTAGGCTTGCCTTCTTCCGATA-3′
			5′-ACACTGACGACATGGTTCTACATCACTAAAGACAGAATGAATGTAGAAAAG-3′
12			5′-TACGGTAGCAGAGACTTGGTCTGGATATTTAATTCGAGTTCCATATTGC-3′
			5′-ACACTGACGACATGGTTCTACAATAGACTTACTGGCCAGTGATCCTC-3′
13			5′-TACGGTAGCAGAGACTTGGTCTGATTGACAAATTCTTTAAGTTCACTGG-3′
			5′-ACACTGACGACATGGTTCTACATTCAGAATGAGAAAAATCCTAACCC-3′
14			5′-TACGGTAGCAGAGACTTGGTCTAGCATCAAGTTCACTTTCTTCCAT-3′
			5′-ACACTGACGACATGGTTCTACAACAAGCCAAATGAACAGACAAGTAA-3′
15			5′-TACGGTAGCAGAGACTTGGTCTCCATTTCTCTTTCAGGTGACATTG-3′
			5′-ACACTGACGACATGGTTCTACAAATGACACAGAAGGCTTTAAGTATCC-3′
16			5′-ACACTGACGACATGGTTCTACAAACAGCAGAACTTTCCTTAATGTC-3′
			5′-TACGGTAGCAGAGACTTGGTCTCACCACTTTTTCCCATCAAGTCA-3′
17			5′-ACACTGACGACATGGTTCTACAGGATCCTGGGTGTTTGTATTTGC-3′
			5′-TACGGTAGCAGAGACTTGGTCTCCTGATGACCTGTTAGATGATGGTG-3′
18			5′-ACACTGACGACATGGTTCTACAACAAAATGATTAAATTCCTTGCTTTGGG-3′
			5′-TACGGTAGCAGAGACTTGGTCTCACAGTGCAGTGAATTGGAAGAC-3′
19		12	5′-ACACTGACGACATGGTTCTACATGAGTTCCATCAAGGTGCTTACA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTAAAAGCAAATCCAGGTGTCCCA-3′
20		13	5′-ACACTGACGACATGGTTCTACAGGCTCCATAATTACCCATGTGCT-3′
			5′-TACGGTAGCAGAGACTTGGTCTACTAGGTGATTTCAATTCCTGTGC-3′
21		14	5′-ACACTGACGACATGGTTCTACATGCCTGTATGCAAAAAACTGGAG-3′
			5′-TACGGTAGCAGAGACTTGGTCTCTGTCTGTTGCATTGCTTGTGTT-3′
22		15	5′-ACACTGACGACATGGTTCTACAAGTAAACCTTGATTAACACTTGAGC-3′
			5′-TACGGTAGCAGAGACTTGGTCTGGGCTCTTTTTTGCCAGTCATTT-3′
23		16	5′-ACACTGACGACATGGTTCTACATTCAGGGTCATCAGAGAAGAGGC-3′
			5′-TACGGTAGCAGAGACTTGGTCTCGTGCTCGGCCCATAAACT-3′
24			5′-ACACTGACGACATGGTTCTACAACAGAACTGTGATTGTTTTCTAGATTTCTT-3′
			5′-TACGGTAGCAGAGACTTGGTCTTCCCATTCCTTTCAGAGGGAACC-3′
25		17	5′-TACGGTAGCAGAGACTTGGTCTTTGAGACGGAGCCTTGCTCT-3′
			5′-ACACTGACGACATGGTTCTACAACTTTAAATAGTTCCAGGACACGTGT-3′
26		18	5′-ACACTGACGACATGGTTCTACACATGGATTCCTGCCGACTATT-3′
			5′-TACGGTAGCAGAGACTTGGTCTAGGCTCTTTAGCTTCTTAGGACA-3′
27		19	5′-ACACTGACGACATGGTTCTACAATGATCCTCATTATCATGGAAAATTTGTG-3′
			5′-TACGGTAGCAGAGACTTGGTCTCTCCTAGCAATGTTAGCATATGAGC-3′
28		20	5′-TACGGTAGCAGAGACTTGGTCTAGAGAAAGACACCCCAGTGAAGT-3′
			5′-ACACTGACGACATGGTTCTACATCTGCTCCACTTCCATTGAAG-3′
29		21	5′-ACACTGACGACATGGTTCTACATTTCAGCAATCTGAGGAACCC-3′

(*table continues*)

Table 3 (*continued*)

Amplicon	Gene	Exon	Forward and reverse primers
			5′-TACGGTAGCAGAGACTTGGTCTCATTAGGGAGGAGCCTTGG-3′
30		22	5′-ACACTGACGACATGGTTCTACACACAGGTATGTGGGCAGAGAAG-3′
			5′-TACGGTAGCAGAGACTTGGTCTAGTACATGGCATATCAGTGGCAAA-3′
31		23	5′-TACGGTAGCAGAGACTTGGTCTTTGACCTCCATCTTGACCTCC-3′
			5′-ACACTGACGACATGGTTCTACAAGCTGAGATCACACCACTGCA-3′
32		24	5′-ACACTGACGACATGGTTCTACATTTTGTGCTCATGGCAGATTTCC-3′
			5′-TACGGTAGCAGAGACTTGGTCTCCTGGAGTCGATTGATTAGAGCC-3′
33	*BRCA2*	1	5′-ACACTGACGACATGGTTCTACACACTCCAGCTCCCGCTTTAT-3′
			5′-TACGGTAGCAGAGACTTGGTCTGACAAAAGGGCAAGAAGCCG-3′
34		2	5′-ACACTGACGACATGGTTCTACATAAGCTGTTACCGTTCCAGGAGAT-3′
			5′-TACGGTAGCAGAGACTTGGTCTAACCTGCAAACGATGATTATGTTGTTA-3′
35		3	5′-ACACTGACGACATGGTTCTACATCACTGGTTAAAACTAAGGTGGGA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTCAAAGGAGGGATGAAAGAGAACA-3′
36		4	5′-ACACTGACGACATGGTTCTACAGGGTAATCAGCAAACTGAAAAACCT-3′
			5′-TACGGTAGCAGAGACTTGGTCTAACTATTCTATTAAAGGACTCAACCTAAAAG-3′
37		5,6	5′-ACACTGACGACATGGTTCTACATGTGTTGGCATTTTAAACATCACT-3′
			5′-TACGGTAGCAGAGACTTGGTCTCCTGTATGAGGCAGAATGCTAGG-3′
38		7	5′-ACACTGACGACATGGTTCTACAGCGTTATACCTTTGCCCTGAGAT-3′
			5′-TACGGTAGCAGAGACTTGGTCTTCAGTTACTAACACACTTATCAAAGACA-3′
39		8	5′-ACACTGACGACATGGTTCTACATCACTGTGTTGATTGACCTTTCT-3′
			5′-TACGGTAGCAGAGACTTGGTCTGAGAGACAGCAGAGTTTCACAGG-3′
40		9	5′-ACACTGACGACATGGTTCTACATGTGCATTGAGAGTTTTTATACTAGTGATTT-3′
			5′-TACGGTAGCAGAGACTTGGTCTAGAGGAGCAATCCTTCAATGGTG-3′
41		10	5′-ACACTGACGACATGGTTCTACACTATGAGAAAGGTTGTGAGAATAATATAAAT-3′
			5′-TACGGTAGCAGAGACTTGGTCTGTTAGTTGAGACCATTCACAGGC-3′
42			5′-TACGGTAGCAGAGACTTGGTCTAGAAGCCCTTTGAGAGTGGAAGT-3′
			5′-ACACTGACGACATGGTTCTACAGCAAACAGTATGTATTTCCAGTCCA-3′
43			5′-ACACTGACGACATGGTTCTACATATGACTGATCCAAACTTTAAAAAAGA-3′
			5′-TACGGTAGCAGAGACTTGGTCTCATGTATACAGATGATGCCTAAGATTAA-3′
44		11	5′-ACACTGACGACATGGTTCTACACCTCCCAAAAGTGCTGAGATTA-3′
			5′-TACGGTAGCAGAGACTTGGTCTGGCATTTTCATGATCATATAAAAGA-3′
45			5′-TACGGTAGCAGAGACTTGGTCTAGAAGAGGTCTTGGCTGCAGC-3′
			5′-ACACTGACGACATGGTTCTACATCAAGTCTGTTTCATGAAGTTCCTT-3′
46			5′-ACACTGACGACATGGTTCTACATCCAGACTCTGAAGAACTTTTCTCA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTTCAACACAAGCTAAACTAGTAGGATATT-3′
47			5′-ACACTGACGACATGGTTCTACAGCAGGACTCTTAGGTCCAATTTCA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTTGCTGCTGTCTACCTGACCAAT-3′
48			5′-ACACTGACGACATGGTTCTACAATCAAAAGAAACTGAGCAAGCC-3′
			5′-TACGGTAGCAGAGACTTGGTCTTAAAAGCCCCTAAACCCCACTTC-3′
49			5′-ACACTGACGACATGGTTCTACAGCAGCAAGCAATTTGAAGGTACA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTGATCAGTAAATAGCAAGTCCGT-3′
50			5′-ACACTGACGACATGGTTCTACATCTGCAGAGGTACATCCAATAAGTT-3′
			5′-TACGGTAGCAGAGACTTGGTCTTCCGTTTTAGTAGCAGTTAACTGTTCT-3′
51			5′-ACACTGACGACATGGTTCTACACTCAAGAAGCATGTCATGGTAATACT-3′
			5′-TACGGTAGCAGAGACTTGGTCTCAGCTGTGATCTCAATGGTCTCA-3′
52			5′-ACACTGACGACATGGTTCTACAAATCACCAGTTTTAGCCATCAATG-3′
			5′-TACGGTAGCAGAGACTTGGTCTATCTTGTTTTTCGGAGAGATGATTT-3′
53			5′-ACACTGACGACATGGTTCTACAGCAGATTATGTAGGAAATTATTTGTATGA-3′
			5′-TACGGTAGCAGAGACTTGGTCTGTTTGGCAAATTTTTGATTTATTCTC-3′
54			5′-ACACTGACGACATGGTTCTACACCAGGTATCAGATGCTTCATTACAA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTCAACAAGTGAGACTTTGGTTCCT-3′
55			5′-ACACTGACGACATGGTTCTACAGGTTCTTCAGAAAATAATCACTCTATTAAA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTACTCCCCCAAACTGACTACACAA-3′
56			5′-ACACTGACGACATGGTTCTACATAGGGCCACCTGCATTTAGGATA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTGTTCTGGAGTACGTATAGCAGTATTT-3′
57		12	5′-ACACTGACGACATGGTTCTACAGGCCTTATTGCCAGTAAACCTAGAG-3′
			5′-TACGGTAGCAGAGACTTGGTCTACCATACCTATAGAGGGAGAACAGA-3′

(*table continues*)

Table 3 (*continued*)

Amplicon	Gene	Exon	Forward and reverse primers
58		13	5′-ACACTGACGACATGGTTCTACAACATTTATTGAGCATCTGTTACATTCA-3′
			5′-TACGGTAGCAGAGACTTGGTCTATATGAACAGCACTATAAAATACTACCAAAG-3′
59		14	5′-ACACTGACGACATGGTTCTACATGTAGCAAATGAGGGTCTGCAA-3′
			5′-TACGGTAGCAGAGACTTGGTCTACTGAAAGGCAAAAATTCATCACACA-3′
60		15	5′-ACACTGACGACATGGTTCTACAGACCTCAGATGATCTGCCCG-3′
			5′-TACGGTAGCAGAGACTTGGTCTGGATGAGGCAGGCCTAGCAA-3′
61		16	5′-ACACTGACGACATGGTTCTACATTTTGGTAAATTCAGTTTTGGTTTGTT-3′
			5′-TACGGTAGCAGAGACTTGGTCTCCAGTTCATTAAACCCCAGGACA-3′
62		17	5′-ACACTGACGACATGGTTCTACATGTACAGAGAATAGTTGTAGTTGTTGA-3′
			5′-TACGGTAGCAGAGACTTGGTCTGCAAGCAATGCATGAATGTGTGA-3′
63		18	5′-ACACTGACGACATGGTTCTACATTATTCAGTGACTTGTTTAAACAGTGG-3′
			5′-TACGGTAGCAGAGACTTGGTCTGAAATTGAGCATCCTTAGTAAGCAT-3′
64		19	5′-ACACTGACGACATGGTTCTACAATCGAAGTTCCTTTTATCTGTTTTCT-3′
			5′-TACGGTAGCAGAGACTTGGTCTACCTGTATAGGGTATGCTCTTTGA-3′
65			5′-ACACTGACGACATGGTTCTACATTTTCCTCTGCCCTTATCATCGC-3′
			5′-TACGGTAGCAGAGACTTGGTCTTGTTACACACACCAAAAAAGTCACA-3′
66		20	5′-ACACTGACGACATGGTTCTACACCACTGTGCCTGGCCTGATA-3′
			5′-TACGGTAGCAGAGACTTGGTCTCCTGATATTTCTGTCCCTTGTTGC-3′
67		21	5′-ACACTGACGACATGGTTCTACATAAATCTCCCTTCTTTGGGTGTT-3′
			5′-TACGGTAGCAGAGACTTGGTCTAGTTGAAAACATACCACCACACT-3′
68		22	5′-ACACTGACGACATGGTTCTACATTAACCACACCCTTAAGATGAGC-3′
			5′-TACGGTAGCAGAGACTTGGTCTTCTTTGTGGGCATTAGTAGTGGA-3′
69		23	5′-ACACTGACGACATGGTTCTACAGTTTATATCAGAGAAGCAAAATCCACT-3′
			5′-TACGGTAGCAGAGACTTGGTCTCCCGTGGCTGGTAAATCTGAA-3′
70		24	5′-ACACTGACGACATGGTTCTACAAAGTAAATCTGAAAGAGCTAACATACAG-3′
			5′-TACGGTAGCAGAGACTTGGTCTGCCAACTGGTAGCTCCAACTAAT-3′
71		25	5′-ACACTGACGACATGGTTCTACAGCACTGTAAGCAACAGGTCATT-3′
			5′-TACGGTAGCAGAGACTTGGTCTCCAAAATGTGTGGTGATGCTGAA-3′
72		26	5′-ACACTGACGACATGGTTCTACATTCCTTCAGATACACATACTTTTTCTCT-3′
			5′-TACGGTAGCAGAGACTTGGTCTTAAAATTTCTAAATAACTATACTTACAGGAGC-3′
73		27	5′-ACACTGACGACATGGTTCTACACCTATTAGGAGTTAGGGGAGGG-3′
			5′-TACGGTAGCAGAGACTTGGTCTTGCAAGTTCTTCGTCAGCTATTG-3′
74			5′-ACACTGACGACATGGTTCTACATACCTCCACCTGTTAGTCCCATT-3′
			5′-TACGGTAGCAGAGACTTGGTCTAAACGATTTTTGCCCGATACACA-3′
75	*PTEN*	1	5′-ACACTGACGACATGGTTCTACAAGTCGCTGCAACCATCCAG-3′
			5′-TACGGTAGCAGAGACTTGGTCTAAAGCAATCGGTGGCTTGAC-3′
76		2	5′-ACACTGACGACATGGTTCTACAAACCGTGAGTTTCTGTTTTTCTCA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTCCCCCTGAAGTCCATTAGGT-3′
77		3	5′-TACGGTAGCAGAGACTTGGTCTCCCATAGAAGGGGTATTTGTTGG-3′
			5′-ACACTGACGACATGGTTCTACATGCCAAGTCATTTAAAGCAGATCA-3′
78		4	5′-ACACTGACGACATGGTTCTACATGGCATCACAAGTTTTTAAGCA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTGCTGCACTTTAGTCTTCCTGA-3′
79		5	5′-ACACTGACGACATGGTTCTACACCTGAATAAAATGGGGGAAAAT-3′
			5′-TACGGTAGCAGAGACTTGGTCTTCAGATCCAGGAAGAGGAAAGG-3′
80		6	5′-ACACTGACGACATGGTTCTACATGGCTACGACCCAGTTACCA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTTTGGCTTCGACTACATTAAAACC-3′
81		7	5′-ACACTGACGACATGGTTCTACATGCTTGAGATCAAGATTGCAGA-3′
			5′-TACGGTAGCAGAGACTTGGTCTGGCCTTTTCCTTCAAACAGG-3′
82		8	5′-ACACTGACGACATGGTTCTACACATGTGAATGAAAATGCAACAGA-3′
			5′-TACGGTAGCAGAGACTTGGTCTTGTCAAGCAAGTTCTTCATCAGC-3′
83		9	5′-ACACTGACGACATGGTTCTACAGATGAGTCATATTTGTGGGTTTTCA-3′
			5′-TACGGTAGCAGAGACTTGGTCTAAAACGGGAAAGTGCCATCT-3′
84	*TP53*	1	5′-TACGGTAGCAGAGACTTGGTCTTACGCTTTAGGCCGGCTCAA-3′
			5′-ACACTGACGACATGGTTCTACACCTTACTTGTCATGGCGACTGT-3′
85		2,3	5′-TACGGTAGCAGAGACTTGGTCTGGACAGCATCAAATCATCCATTG-3′
			5′-ACACTGACGACATGGTTCTACATTGGAAGTGTCTCATGCTGGATC-3′
86		4	5′-TACGGTAGCAGAGACTTGGTCTGGACAGGAGTCAGAGATCACACATT-3′

(*table continues*)

Table 3 (*continued*)

Amplicon	Gene	Exon	Forward and reverse primers
			5′-ACACTGACGACATGGTTCTACAGGTAAGGACAAGGGTTGGGC-3′
87		5,6	5′-TACGGTAGCAGAGACTTGGTCTGGAGGGCCACTGACAACCAC-3′
			5′-ACACTGACGACATGGTTCTACAGAGGTGCTTACGCATGTTTGTT-3′
88		7	5′-TACGGTAGCAGAGACTTGGTCTAGAGGCAGTAAGGAAATCAGGTC-3′
			5′-ACACTGACGACATGGTTCTACACTTGCCACAGGTCTCCCCAA-3′
89		8,9	5′-TACGGTAGCAGAGACTTGGTCTATTGTCTTTGAGGCATCACTGC-3′
			5′-ACACTGACGACATGGTTCTACACAGAAAGGACAAGGGTGGTTG-3′
90		10	5′-TACGGTAGCAGAGACTTGGTCTGAAGGCAGGATGAGAATGGAA-3′
			5′-ACACTGACGACATGGTTCTACAAGCCAAGATTGCACCATTGC-3′
91		11	5′-TACGGTAGCAGAGACTTGGTCTAACCCTTAACTGCAAGAACATTTCT-3′
			5′-ACACTGACGACATGGTTCTACATGTGGCCACCATCTTGATTTG-3′

Massively Parallel DNA Sequencing

Resulting amplicon pools with an average of 12 patients were further pooled for sequencing library preparation according to the protocols of the manufacturer. Each library was subjected to 454 GS Junior system (Roche, Basel, Switzerland) or MiSeq (Illumina, San Diego, CA) so that each nucleotide position in target regions is sequenced at a minimum depth of 30-fold which was also equivalent to the average depth of at least 150-fold in GS Junior. For MiSeq (Illumina) sequencing, the minimum depth was 50-fold and average depth of 300. Two different sets of signal processing filters are used in parallel for reads and base calling with the use of GS Run Processor (Roche): factory default amplicon signal processing filters, and the same sets of filters except i) doValleyFilterTrimBack = true, ii) vfLastFlowToTest = 480, and iii) vfScanAllFlows = false. Sequencing reads returned from the two processing methods were analyzed separately with the same computational variant detection and annotation pipeline described in the next section. Positions of insufficient depth were subjected to Sanger sequencing for verification. For MiSeq, BWA-MEM, and variant callers, SAMTools and GATK HaplotypeCaller were adopted. In all circumstances, the interpretation of the control results was blinded between different platforms. The bioinformatic analysis was performed on a Cray XC30 supercomputer (Cray, Seattle, WA).

Variant Detection and Annotation

Sequencing reads were assigned to corresponding samples according to sequence barcodes with the use of Amplicon Variant Analyzer (AVA) version 2.7 (Roche). Reads were then aligned to reference sequences (*BRCA1*: NM_007294.3; *BRCA2*: NM_000059.3; *TP53*: NM_000546.4; *PTEN*: NM_000314.4) by AVA with the use of default settings. Each patient-specific BAM sequence alignment[14] was processed by the following three variant callers: i) AVA with default settings, ii) SAMtools version 0.1.[18] with customized settings (samtools mpileup −BQ0; vcfutils.pl varFilter -D 1000000 -1 1e-999 -2 1e-999 -4 1e-999), and iii) HomopolymerQZ (described in the next section). Variant effect on corresponding protein coding genes was annotated with Ensembl Variant Effect Predictor version 2.4.[19] Annotated variants were automatically prioritized for manual review according to criteria, including i) the variant prevalence among previous patients, ii) whether the variant is detected in both signal processing methods, and iii) the predicted effect on the protein coding genes. Variants were designated according to the recommendations from Human Genome Variation Society.[20]

Putative mutations were validated by bidirectional Sanger sequencing on Applied Biosystems 3130xl or 3730 Genetic Analyzers (Life Technologies, Carlsbad, CA) from different PCR and sequencing primers other than microfluidic PCR. Sanger sequencing data were analyzed by Mutation Surveyor version 4.0.6 (Softgenetics, State College, PA), followed by manual review. Selected hotspot pathogenic mutations were subjected to haplotype analysis as described previously[12] from the probands, their family members, and 54 Sanger-validated controls. Mutations potentially affecting pre-mRNA processing were subjected to cDNA sequencing analysis.

VUS were further characterized by their frequencies among 2504 individuals from The 1000 Genomes Project[21] and 100 healthy local individuals from this study, and sequence-alignment−based *in silico* prediction algorithms: SIFT,[22] PolyPhen-2,[23] and Align-GVGD.[24]

Homopolymer Variant Detection

In the target regions of *BRCA1*, *BRCA2*, *TP53*, and *PTEN*, there were 1330 homopolymer loci of lengths from 3 to 15 bp. Quantitative pyrosequencing signal intensities of the homopolymer loci were retrieved from the Standard Flowgram Format files of 54 control samples with known mutation, which are wild-type at these loci, and were used to determine the normal reference ranges of the 1330 homopolymer loci. Pyrosequencing signal intensities of each locus (with the bin size of 0.5) consistent cluster with a bell shape and peak at slightly shorter length, for example, a normal A_8 homopolymer returns with the signal of $A_{7.5 \pm 0.5}$

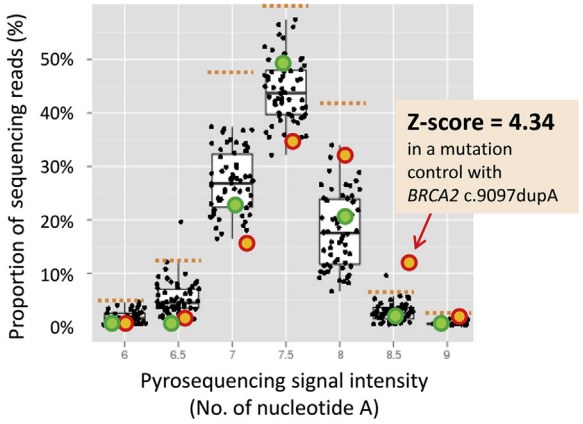

Figure 2 Distribution of pyrosequencing signal intensity. Representative homopolymer at *BRCA2* c.9097 is shown as an example. A positive control sample with *BRCA2* c.9097dupA mutation was represented by **red circles** (A9/A8). A wild-type sample is represented by **green circles** (A8/A8). Other wild-type samples (**black dots**) are also shown. **Orange dashed lines** represent 1.5 interquartile ranges above third quartile. Each sample is represented by one dot in each signal intensity group.

(Supplemental Figure S1). The distribution frequencies of every signal were plotted against the variable size of each repeating unit.

HomopolymerQZ was a software suite developed to detect variants in homopolymer loci from pyrosequencing signal intensity (PSI) instead of base-called sequences. A Perl script parse_signalDist_v2.pl was developed to extract a set of PSI values for a given DNA sample, corresponding to every sequencing read aligned to every homopolymer locus from the AVA (Roche) alignment results. The extracted PSI values are quantitative measurements of number of nucleotide to the first decimal place. PSI of reverse sequencing reads (relative to the reference sequence) were reverse complemented as if they are forward sequencing reads. For example, a homopolymer locus of nucleotide A (based on reference sequence) sequenced at 100-fold (sequencing of A by 50 forward sequencing reads plus sequencing of T by 50 reverse sequencing reads) will return a total of 100 PSI values. Although PSI value is accurate up to the first decimal place, the average 454 sequencing depth is 150-fold in this study so that some possible PSI values are not represented by any reads. PSI values are therefore categorized as bins b, rounding to 0.5 nucleotide (eg, 3.0, 3.5, 4.0, 4.5, …), with the use of awk script flowstat_assign_flowvalue_group.awk. Those with first decimal places 0.0 to 0.4 are grouped into a bin and 0.5 to 0.9 are grouped into another bin. For example, PSI of 10 sequencing reads {3.1, 3.1, 3.2, 3.3, 3.5, 3.6, 3.8, 3.9, 3.9, 3.9} are represented as $b_{3.0} = 4$ and $b_{3.5} = 6$. These PSI values are stored in a SQLite database and further processed by an R script classify_homopolymer.R. To enable comparison of PSI between samples with different sequencing depths, PSI are then normalized as a fraction of sequencing read count as $f_{hb} = n_{hb}/N_h$, where h is a given homopolymer locus, b is PSI value bin as described above, n_{hb} is the number of sequencing reads in the bin b for homopolymer h, and N_h is the total number of sequencing reads covering the homopolymer h regardless of PSI value. Summation of f_{hb} of all bins in a given locus h equals to 1. In this study, there are 54 control DNA samples and 1330 homopolymer loci of interest (length, 3 to 15 bp). After PSI extraction and transformation, there are $54 \times 1330 = 71{,}820$ sets of f_{hb} values serving as control values. When f_{hb} values of a given testing DNA sample s (f_{hbs}) are significantly different from those of control samples c (f_{hbc}), possibly there are variants in the corresponding homopolymer loci. Outlier loci with significantly different f_{hb} values are defined as $f_{hbs} > 1.5(\mathrm{IQR}_{hbc}) + Q_{3_{hbc}}$, where IQR_{hbc} and Q_{3hbc} are the interquartile range and the third quartile of f_{hbc} values of control samples, respectively. Those outlier loci are ranked by their z scores as follows:

$$z_{hbs} = \frac{f_{hbs} - \overline{f_{hbc}}}{\mathrm{SD}_{hbc}},$$

where SD_{hbc} is the SD of f_{hbc} values of control samples. Source code and documentation of HomopolymerQZ are available for download at *http://github.com/tommyau/homopolymerqz* (Hong Kong Sanitorium and Hospital).

MLPA Assay

All patient samples were subjected to MLPA for detection of large genomic rearrangement of *BRCA1* and *BRCA2*. SALSA MLPA *BRCA1* and *BRCA2* probemix (MRC-Holland, Amsterdam, the Netherlands) was used according to the manufacturer's protocol. Data analysis was performed by Coffalyser version 7 (MRC-Holland).

Results

Validation of NGS-Based Mutation Detection Performance

The NGS-based screening workflow was evaluated from the analysis of *BRCA1* and *BRCA2* mutation-positive ($n = 80$) and -negative ($n = 8$) validated control samples. The control samples were previously characterized by bidirectional Sanger sequencing of *BRCA1* and *BRCA2* coding exons.[7] Five types of DNA mutations in *BRCA1* and *BRCA2* were analyzed, including i) 34 different deletions of length 1 to 11 bp, ii) 13 different duplications of length 1 to 4 bp, iii) 21 different nonsense single nucleotide variants (SNVs) in exons, iv) 11 different SNVs in splice sites, and v) one complex mutation that involved 8-bp deletion and 7-bp insertion on the same allele (Table 2). All 80 mutations, including local Chinese and overseas controls, were detected by NGS.

Among them, 76 mutations were readily detected by the combination of two variant callers, AVA and SAMtools,

as shown in the annotated variant reports. The sensitivity of the test is 95% (95% CI, 93.9−96.1; $n = 80$) and specificity of the test is 100% (95% CI, 95.8−104.2; $n = 8$). These readily detectable mutations include deletions, duplication, nonsense SNVs, and splice site SNVs. The remaining four mutations were the complex mutation *BRCA2* c.4467_4474delinsTGTTTTT and three different homopolymer-length variants *BRCA1* c.1016delA (1-bp deletion from 7-bp homopolymer), *BRCA1* c.1961dupA (1-bp duplication from 8-bp homopolymer), and *BRCA2* c.9097dupA (1-bp duplication from 8-bp homopolymer).

Although the exact *BRCA2* deletion-insertion mutation was not called by both variant callers, there were actually multiple SNVs called along the deleted positions. The splitting of the complex mutation into multiple SNVs is possibly because of the limitation of the variant calling algorithm in handling such complex deletion-insertion situations. However, the run of SNVs was considered sufficient to trigger manual review and follow-up Sanger sequencing validation. Thus, this complex mutation was regarded as detectable, but the mutation naming would require manual correction.

Novel Algorithm HomopolymerQZ for Accurate Variant Detection of Homopolymers

Because of the high false-positive rates in homopolymer loci, including *BRCA1* c.1016 and c.1961 together with *BRCA2* c.9097, in all samples and difficulty in calling insertion-deletion mutation at homopolymer regions, a bioinformatics algorithm HomopolymerQZ was designed to detect variants in homopolymer loci (length, ≥3 bp). Signals of homopolymer loci among the four target genes ($n = 948$) were collected from 54 validated control samples that were wild-type at the loci. HomopolymerQZ successfully detected mutations *BRCA2* c.9097dupA (Figure 2), *BRCA1* c.1016delA and c.1961dupA with high z scores (4.34, 4.56, and 4.67) from the validated positive controls, respectively, without high false-positive rates in other controls (Table 2). Variants in prospective cases can also be detected by HomopolymerQZ, for example, *BRCA1* c.280C>T (p.Gln94*; z score 54.58) and *PTEN* c.1202C>A (p.Thr401Lys; z score 6.88). Overall, with the incorporation of homopolymer-specific variant caller HomopolymerQZ, all 80 mutations from the validated controls were detectable by the NGS workflow.

Mutation Spectrum of 948 Patients

The NGS analysis workflow was then applied to screen 948 uncharacterized probands from an independent cohort with a shorter turnaround time compared with conventional bidirectional Sanger sequencing. Most pathogenic mutations were identified in *BRCA1* ($n = 27$) and *BRCA2* ($n = 41$), but *TP53* ($n = 5$) and *PTEN* ($n = 2$) mutations were also identified (Table 4).[7,16,25] Seven mutations (*BRCA1* c.964delG, c.3333delA, *BRCA2* c.2595delA, c.5164_5165delAG, c.7007G>T, c.7878G>A, and c.9294C>G) were observed in at least two unrelated probands; thus, they were regarded as new hotspot mutations in the cohort. Together with our previously known hotspot mutations,[16] hotspot mutations account for 48.8% of all mutations identified from the probands in this study. Haplotype analysis of two families and 50 unrelated controls supported that the new hotspot mutation *BRCA2* c.5164_5165delAG is a novel founder mutation observed in this cohort (Supplemental Table S2). In addition, the overall incidence of deleterious *BRCA1/BRCA2* mutation of hereditary breast and/or ovarian cancers was 8.8% in our Southern Chinese high-risk families ($n = 1427$) from combining data sets of PCR-based Sanger sequencing and NGS. The overall clinical pathologic variables in *BRCA1/BRCA2* carriers detected by Sanger and NGS are illustrated in Table 5. With the inclusion of MLPA-detected mutations, the overall germline mutation frequency was 9.4%.

Five different mutations that affect pre-mRNA processing were identified in six probands (Table 4). Notably, the *BRCA2* deletion c.-39-1_-39delGA is located across the splice site upstream to translation initiation codon in exon 2. The mutation allele was predicted to be a null allele in production of functional *BRCA2* protein. In another proband, *BRCA2* SNV c.8023A>G was observed and computationally annotated as a missense SNV. However, *in vitro* cDNA sequencing showed that the mutation gave rise to a strong donor site and resulted in aberrant splicing of exon 18. The cDNA-based experimental results of this mutation were consistent with a previous study.[26]

VUS Spectrum of 1427 Patients and 100 Healthy Local Individuals

Among the VUS identified from the probands, missense SNVs were further prioritized according to five selection criteria (namely absence in The 1000 Genomes Project, absence in 100 normal controls, and scoring of SIFT, PolyPhen-2, and Align-GVGD), which were based on allele frequency of populations and *in silico* sequence-alignment−based prediction algorithms. From the 1427 patients, 104 missense VUS were found to be absent among 2504 individuals from worldwide populations of The 1000 Genomes Project. In parallel, 100 healthy local individuals were also screened by the same workflow as the probands. By filtering VUS that were also present in the 100 local individuals (Supplemental Table S3), 92 missense VUS were left for prioritization with the use of *in silico* predictions. Finally, 46 VUS were predicted as probably damaging from three different sequence-alignment−based algorithms that the corresponding protein function will likely be affected (SIFT deleterious, PolyPhen-2 probably damaging, and Align-GVGD most likely deleterious). Further co-segregation analysis on the relevant families is in progress.

Table 4 Spectrum of 75 Pathogenic Mutations Identified by the NGS Method

Proband no.	Gene	Exon/Intron	Mutation detected	Amino acid change
TWH001	*BRCA1*	Exon 2	c.66dupA	p.Glu23Argfs*18
TWH002		Exon 5	c.212+3A>T	p.Cys64* p.Phe46_Arg71del26
TWH003		Exon 6	c.220C>T	p.Gln74*
TWH004		Exon 6	c.280C>T	p.Gln94*
TWH005		Exon 8	c.514C>T	p.Gln172*
QMH001		Exon 8	c.470_471delCT	p.Ser157*[†]
QMH002		Exon 11	c.964delG	p.Ala322Leufs*19[‡]
TWH006		Exon 11	c.2166delC	p.Asn723Ilefs*13
TWH007		Exon 11	c.2253_2254delGT	p.Met751Ilefs*10
KWH001		Exon 11	c.2577_2578insTT	p.Thr860Leufs*34
TWH008		Exon 11	c.2635G>T	p.Glu879*
TWH009		Exon 11	c.2695delG	p.Val899Serfs*101
TWH010		Exon 11	c.3333delA	p.Glu1112Asnfs*5[‡]
KWH002		Exon 11	c.3607C>T	p.Arg1203*
TWH011		Exon 11	c.3756_3759delGTCT	p.Ser1253Argfs*10
TWH012		Exon 11	c.3767_3768delCA	p.Thr1256Argfs*10
TWH013		Exon 11	c.4065_4068delTCAA	p.Asn1355Lysfs*10
KWH003		Exon 12	c.4148C>G	p.Ser1383*
TMH001		Exon 14	c.4372C>T	p.Gln1458*[†]
TWH014		Exon 14	c.4372C>T	p.Gln1458*
TWH015		Exon 14	c.4372C>T	p.Gln1458*
TWH016		Exon 16	c.4780_4793del14	p.Pro1594Cysfs*23
TWH017		Exon 22	c.5353C>T	p.Gln1785*
TWH018		IVS22	c.5406+1_5406+3delGTA	p.Asp1778Glyfs*27[†§¶]
PMH001		IVS22	c.5406+7A>G	p.Asp1778Glyfs*27[§]
HKSH001		Exon 24	c.5503C>T	p.Arg1835*
TWH019	*BRCA2*	IVS1/Exon 2	c.-39-1_-39delGA	p.0?[§]
TWH020		Exon 9	c.755delA	p.Asp252Alafs*25
TWH021		Exon 10	c.1888dupA	p.Thr630Asnfs*6
TWH022		Exon 11	c.2442delC	p.Met815Trpfs*10
KWH004		Exon 11	c.2339C>G	p.Ser780*
QMH003		Exon 11	c.2595delA	p.Glu866Lysfs*8[‡]
TWH023		Exon 11	c.2808_2811delACAA;	p.Ala938Profs*21
TMH002		Exon 11	c.3109C>T	p.Gln1037*[†¶]
TWH024		Exon 11	c.3109C>T	p.Gln1037*
KWH005		Exon 11	c.3109C>T	p.Gln1037*
TWH025		Exon 11	c.3109C>T	p.Gln1037*
KWH006		Exon 11	c.3109C>T	p.Gln1037*
TWH026		Exon 11	c.3109C>T	p.Gln1037*
TWH027		Exon 11	c.3109C>T	p.Gln1037*
TWH028		Exon 11	c.3109C>T	p.Gln1037*
TWH029		Exon 11	c.3760G>T	p.Glu1254*
TWH030		Exon 11	c.4372C>T	p.Gln1458*
TWH031		Exon 11	c.5164_5165delAG	p.Ser1722Tyrfs*4[‡‖]
TWH032		Exon 11	c.5164_5165delAG	p.Ser1722Tyrfs*4
TWH033		Exon 11	c.5164_5165delAG	p.Ser1722Tyrfs*4
QMH004		Exon 11	c.6240dupA	p.Glu2081Argfs*4
QMH005		Exon 11	c.6256_6259delATCA	p.Ile2086Glufs*32
QMH006		Exon 11	c.6678delA	p.Ala2227Glnfs*2
TWH034		Exon 13	c.7007G>T	p.Gly2313Alafs*31[‡§]
TMH003		Exon 13	c.7007G>T	p.Gly2313Alafs*31
TWH035		Exon 13	c.7007G>T	p.Gly2313Alafs*31
KWH007		Exon 14	c.7133C>G	p.Ser2378*
KWH008		Exon 15	c.7490dupA	p.Lys2498Glufs*41
TWH036		Exon 15	c.7558C>T;	p.2520*
TWH037		Exon 17	c.7878G>A	p.Trp2626*[‡]
TWH038		Exon 17	c.7878G>A	p.Trp2626*

(*table continues*)

Table 4 (*continued*)

Proband no.	Gene	Exon/Intron	Mutation detected	Amino acid change
TWH039		Exon 17	c.7878G>A	p.Trp2626*
TWH040		Exon 17	c.7878G>A	p.Trp2626*
PMH002		Exon 18	c.8023A>G	p.Met2676_Ile2778del[§]
TWH041		Exon 18	c.8068_8069delGT	p.Val2690Phefs*2
KWH009		Exon 19	c.8400_8402delinsAAAA	p.Phe2801Lysfs*11
KWH010		Exon 19	c.8470A>T	p.Arg2824*
HKSH002		Exon 25	c.9294C>G	p.Tyr3098*[‡]
TWH041		Exon 25	c.9294C>G	p.Tyr3098*
TMH004		Exon 25	c.9382C>T	p.Arg3128*
TWH042		Exon 25	c.9401delG;	p.Gly3134AlafsX29
KWH011		Exon 25	c.9409dupA	p.Thr3137Asnfs*13
QMH007	*PTEN*	Exon 1	c.1A>C	p.?
PMH003		Exon 5	c.388C>T	p.Arg130*
TWH043	*TP53*	Exon 2	c.96+1G>T; r.75_96del	p.Leu26Profs*11
TWH044		Exon 5	c.473G>A	p.Arg158His[‖]
TWH045		Exon 5	c.541C>T	p.Arg181Cys
TWH046		Exon 7	c.743G>A	p.Arg248Gln[‖]
TWH047		Exon 8	c.916C>T	p.Arg306*

[†]Known recurrent mutation in the Registry.
[‡]Novel recurrent mutation in the Registry.
[§]Splicing variant.
[¶]Known founder mutation.[7,16]
[‖]Associated with Li-Fraumeni syndrome[25] or Li-Fraumeni–like syndrome.

Discussion

From our experience in genetic testing with different platforms, the overall incidence of deleterious *BRCA1/BRCA2* mutation of hereditary breast and/or ovarian cancers was 9.4% in Southern Chinese high-risk families ($n = 1427$) by using the combination of direct PCR-based Sanger sequencing, NGS, and MLPA. We have demonstrated that the NGS platform is able to detect all of the Sanger-validated positive controls.

In addition, we have identified 46 new mutations in the cohort by NGS method. The mutation spectrum of the NGS platform was expanded by 57.5% (an increase of detectable mutation from 80 to 126 unique mutations). It greatly enhanced the turnaround time for mutation screening, the possibility of expanding the gene panels, and the feasibility of more extensive investigation of the control groups which is not practical for Sanger Sequencing or high resolution melting analysis. Notably, we are able to identify a mutation by NGS that was previously missed by high resolution melting. To the best of our knowledge, this is the largest cohort of genetic predisposition screening of breast cancer reported in the Chinese population. Overall, 126 pathogenic mutations were identified in *BRCA1*, *BRCA2* genes; in addition, *TP53* and *PTEN* mutations were detected in five and two families, respectively. Detailed analysis of the mutation spectrum revealed that 17 varieties of hotspot mutation (ie, identified in ≥2 families) were seen in our cohort, which contributed to 48.8% of all detected mutations. This suggested that screening these 17 hotspot mutations can pick up approximately one-half of the Southern Chinese mutations. This conclusion implied an enormous impact on the design of HBOC screening for Chinese. Because Hong Kong was once a city with a population of approximately 125,000 in mid-1800 and reached >7 million in 2010, most residents are immigrants from Southern China where one of its many provinces such as Guangdong has a population size of >100 million. Therefore, more patients with hotspot mutations could be detected effectively by a simple, rapid, and cost-effective routine molecular diagnostic method that is easily available in most diagnostic laboratories. Among the hotspot mutations, six were confirmed as founder mutations by identification of common disease haplotype. Interestingly, large deletions were detected by MLPA in seven families. The overall MLPA detectable mutation frequency, including a hotspot mutation of deletion of exons 15 and 16, is 0.56%. Thus, analysis of large duplication or deletion in these genes should be included in all diagnostic platforms, to obtain a more comprehensive spectrum of the population.[27]

It is important to have a group of healthy individuals for VUS classification. Although VUS data from The 1000 Genomes Project are available for reference, we showed that nine population-specific SNPs from the normal cohort, including eight in *BRCA2* and one in *BRCA1*, could be found and useful in classification of VUS (Supplemental Table S3). It suggested that population-specific study of a cohort of healthy individuals should be included in the genetic diagnosis of familial breast cancer. With the use of a

Table 5 Patient Demographic Characteristics in *BRCA* Carriers

Characteristic	*BRCA1*	*BRCA2*
Age, mean (SD) (years)	41.73 (9.85)	46.68 (9.56)
Histologic type		
DCIS	1	14
IDC	52	62
ILC	0	1
Mixed (ILC+IDC)	0	0
Others	5	4
ND	3	3
Histologic tumor grade		
1	1	3
2	10	33
3	34	34
ND	13	11
Lymph node		
N0 [including N(itc)]	42	46
N1	10	20
N2	1	11
N3	3	1
ND	5	6
Distant metastasis		
Before surgery		
No	59	82
Yes	1	2
ND	1	0
After surgery		
No	57	78
Yes	4	6
Stage		
0	1	13
I	25	18
II	28	32
III	4	13
IV	1	2
ND	2	6
Tumor stage		
T0 (including Tis)	1	14
T1 (including T1mi)	30	29
T2	25	30
T3	0	4
T4	1	1
ND	4	6
Bilateral cancer		
No	11	6
Yes	2	6

BRCA1 carriers: 61 tumors from 48 patients. *BRCA2* carriers: 84 tumors from 73 patients.

DCIS, ductal carcinoma *in situ*; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; itc, isolated tumor cells; ND, not defined; NGS, next-generation sequencing; T1mi, T1 with microinvasion; Tis, carcinoma *in situ*.

cohort of 100 individuals, it is able to identify roughly 80% of SNPs in the same population.[17]

From our seven selection criteria, the overall mutation prevalence is 9.4%. Notwithstanding, patients with triple-negative breast cancer (17.6%) and with the history of breast and/or ovarian cancers (16%) showed higher detection rates. It was reported that triple-negative breast cancer is a cost-effective indicator for genetic screening.[9] In this study, there were only eight patients who had ovarian cancer with family history, and six of them were *BRCA1/BRCA2* carriers. A larger cohort is necessary to confirm whether this is a high-risk factor for harboring a germline mutation. This information would greatly improve the genetic counseling and interpretation of the test results. Patients with multiple risk factors increased the likelihood to carry mutations. Hence, their family members should also undergo genetic testing and undertake preventive measures.

The quality of the NGS reads is critical in the clinical application of NGS. However, because of the limitation of bioinformatic personnel in most clinical laboratories, most of the analysis relies on packaged software.[28] In this study, we have demonstrated that pyrosequencing output data could be amenable to systematic quantitative analysis and pinpointed the drawback of data analysis with only one algorithm. NGS data usually contain a lot of nonspecific signals/errors; therefore, it is important to distinguish systematic error from real variations. The exploration of the usage of pyrosequencing signal intensity to assist variant calling has been described but limited to few loci.[10] The presence of homopolymers seems to be the most challenging problem in pyrosequencing platform; a novel in-house bioinformatic analysis, HomopolymerQZ, was therefore developed for automatic homopolymer mutation calling. With the use of the unique characteristic of 454 platform and application of signal reference range to indicate that mutation associated with homopolymer greater than or equal to three could be detected by the unique quantitative signal generated by 454 pyrosequencer. In addition, the signal released by pyruvase in every single homopolymer in our set of amplicons in a 454 sequencer are relatively constant among cases (including controls), although it might not necessarily be equivalent to the empirical signal intensity. With this characteristic, we could calculate both the distribution frequencies and z score of the specific reads against the same locus of a group of normal controls. This characteristic could be explored in a different set of gene-specific data from different diagnostic laboratories. This novel HomopolymerQZ could potentially serve as a standard laboratory monitoring system, which might not be easily available on other platforms.

In this study, we have covered all intron-exon boundaries, including coding or noncoding. We have identified *BRCA2* c.-39-1_-39delGA mutation in one of the families. It is particularly interesting because the mutation would be missed if variant detection is only limited to the coding sequences. Further cDNA analysis also indicated that such mutation induced skipping of exon 2 in the *BRCA2* transcript; thus, the translation start site was deleted in this allele. However, heterozygosity frequency indicated the accuracy of the test. However, a nonsense mutation in *BRCA1* c.280C>T with the variant allele frequency ranged from 16% to 24% in different independent NGS runs.

Detailed investigation showed that this variation is the cause of a SNP located in the primer region in the disease allele. Relocation of the primer could effectively avoid this phenomenon, and the frequency of this nonsense mutation falls back into the range of 40% to 60% heterozygosity rate. Therefore, one should be cautious during the analysis of variants if the heterozygosity falls <50%. Another important issue that needs to be addressed is the mutation that leads to generation of a novel splice site; thus, any downstream analysis should consider fetching the association of synonymous/VUS and splicing defect. Finally, we have included 100 apparently normal individuals as a SNPs indicator. It was shown that screening 300 individuals can cover as much as >95% of polymorphisms at approximately 1% frequency in the corresponding population. By combining the polymorphism spectrum of the Chinese populations from The 1000 Genomes Project (Han Chinese and Southern Han Chinese) and also the local population, we could easily eliminate a group of VUS and minimize the report of VUS effectively (Supplemental Table S3).

Our cohort consisted mainly of Chinese who reside in Hong Kong. From our current findings, development of a breast cancer genetic screening strategy for the Chinese population is feasible and also applicable to Mainland Chinese, and Chinese who reside in other parts of the world, most of whom immigrated from Hong Kong and Mainland China. Genetic diagnostic testing and clinical management of hereditary breast cancer are still at a developmental stage in Mainland China.[29] From the experience with colorectal cancer screening, if we could design a cost-effective genetic test, 1.35 billion individuals will likely benefit from this new approach and improve cancer risk assessment and management.

Taken together, we have harnessed the advantages of NGS pyrosequencing platform as follows: i) the quality of the NGS reads from 454 pyrosequencer/MiSeq were important data and could be manipulated appropriately for clinical application; ii) a novel algorithm, HomopolymerQZ, potentially could be used as an automated homopolymer mutation detector (although 454 technology is fading, alternative technology such as Ion Proton rely on comparable method may benefit from the algorithm); iii) criteria for primer design and mutation detection strategies should be addressed probably for routine genetic diagnosis of hereditary breast cancer. Finally, this multigene panel is a reliable diagnostic platform when stringent validation has been implemented.

Acknowledgments

We thank Wing Pan Luk and Ling Hiu Fung for assisting in statistical analysis; Drs. Dacita Suen, Clement Chen, Kwok K. Ma, Lorraine Chow, Annie Chu, and Isabella Cheuk, as well as Desiree Tse, Winner Cheng, and Wong Ling (The University of Hong Kong), and the breast cancer research groups from Caritas Medical Centre, Kwong Wah Hospital, North District Hospital, Pamela Youde Nethersole Eastern Hospital, Princess Margaret Hospital, Queen Elizabeth Hospital, Queen Mary Hospital, Tuen Mun Hospital, Tung Wah Hospital, and United Christian Hospital for patient recruitment.

Supplemental Data

Supplemental material for this article can be found at *http://dx.doi.org/10.1016/j.jmoldx.2016.03.005*.

References

1. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjäkoski K, Kallioniemi OP, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF: Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003, 72:1117–1130
2. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, et al: Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. Am J Hum Genet 1998, 62:676–689
3. Garcia C, Lyon L, Littell RD, Powell CB: Comparison of risk management strategies between women testing positive for a BRCA variant of unknown significance and women with known BRCA deleterious mutations. Genet Med 2014, 16:896–902
4. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD: The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. Cancer Epidemiol 2012, 36:237–248
5. Kwong A, Shin VY, Ho JC, Kang E, Nakamura S, Teo SH, Lee AS, Sng JH, Ginsburg OM, Kurian AW, Weitzel JN, Siu MT, Law FB, Chan TL, Narod SA, Ford JM, Ma ES, Kim SW: Comprehensive spectrum of BRCA1 and BRCA2 deleterious mutations in breast cancer in Asian countries. J Med Genet 2016, 53:15–23
6. Nakamura S, Kwong A, Kim SW, Iau P, Patmasiriwat P, Dofitas R, Aryandono T, Hu Z, Huang CS, Ginsburg O, Rashid MS, Sarin R, Teo SW: Current status of the management of hereditary breast and ovarian cancer in Asia: first report by the Asian BRCA Consortium. Public Health Genomics 2016, 19:53–60
7. Kwong A, Wong LP, Wong HN, Law FB, Ng EK, Tang YH, Chan WK, Ho LS, Kwan KH, Poon M, Wong TT, Leung FC, Chan SW, Ying MW, Ma ES, Ford JM: A BRCA2 founder mutation and seven novel deleterious BRCA mutations in southern Chinese women with breast and ovarian cancer. Breast Cancer Res Treat 2009, 117:683–686
8. Foulkes WD, Stefansson IM, Chappuis PO, Bégin LR, Goffin JR, Wong N, Trudel M, Akslen LA: Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003, 95:1482–1485
9. Kwon JS, Daniels MS, Sun CC, Lu KH: Preventing future cancers by testing women with ovarian cancer for BRCA mutations. J Clin Oncol 2010, 28:675–682
10. Michils G, Hollants S, Dehaspe L, Van Houdt J, Bidet Y, Uhrhammer N, Bignon YJ, Vermeesch JR, Cuppens H, Matthijs G: Molecular analysis of the breast cancer genes BRCA1 and BRCA2 using amplicon-based massive parallel pyrosequencing. J Mol Diagn 2012, 14:623–630

11. Chan M, Ji SM, Yeo ZX, Gan L, Yap E, Yap YS, Ng R, Tan PH, Ho GH, Ang P, Lee AS: Development of a next-generation sequencing method for BRCA mutation screening: a comparison between a high-throughput and a benchtop platform. J Mol Diagn 2012, 14:602–612
12. Trujillano D, Weiss ME, Schneider J, Köster J, Papachristos EB, Saviouk V, Zakharkina T, Nahavandi N, Kovacevic L, Rolfs A: Next-generation sequencing of the BRCA1 and BRCA2 genes for the genetic diagnostics of hereditary breast and/or ovarian cancer. J Mol Diagn 2015, 17:162–170
13. Feliubadaló L, Lopez-Doriga A, Castellsagué E, del Valle J, Menéndez M, Tornero E, Montes E, Cuesta R, Gómez C, Campos O, Pineda M, González S, Moreno V, Brunet J, Blanco I, Serra E, Capellá G, Lázaro C: Next-generation sequencing meets genetic diagnostics: development of a comprehensive workflow for the analysis of BRCA1 and BRCA2 genes. Eur J Hum Genet 2013, 21:864–870
14. Walsh T, Casadei S, Coats KH, Swisher E, Stray SM, Higgins J, Roach KC, Mandell J, Lee MK, Ciernikova S, Foretova L, Soucek P, King MC: Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. JAMA 2006, 295:1379–1388
15. Ngeow J, Stanuch K, Mester JL, Barnholtz-Sloan JS, Eng C: Second malignant neoplasms in patients with Cowden syndrome with underlying germline PTEN mutations. J Clin Oncol 2014, 32:1818–1824
16. Kwong A, Ng EK, Wong CL, Law FB, Au T, Wong HN, Kurian AW, West DW, Ford JM, Ma ES: Identification of BRCA1/2 founder mutations in Southern Chinese breast cancer patients using gene sequencing and high resolution DNA melting analysis. PLOS ONE 2012, 7:e43994
17. Collins JS, Schwartz CE: Detecting polymorphisms and mutations in candidate genes. Am J Hum Genet 2002, 71:1251–1252
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup: The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009, 25:2078–2079
19. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F: Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 2010, 26:2069–2070
20. den Dunnen JT, Antonarakis SE: Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 2000, 15:7–12
21. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: An integrated map of genetic variation from 1,092 human genomes. Nature 2012, 491:56–65
22. Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009, 4:1073–1081
23. Adzhubei I, Jordan DM, Sunyaev SR: Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013. Chapter 7:Unit7.20
24. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A: Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. Hum Mutat 2008, 29: 1342–1354
25. Raymond VM, Else T, Everett JN, Long JM, Gruber SB, Hammer GD: Prevalence of germline TP53 mutations in a prospective series of unselected patients with adrenocortical carcinoma. J Clin Endocrinol Metab 2013, 98:E119–125
26. Bonnet C, Krieger S, Vezain M, Rousselin A, Tournier I, Martins A, Berthet P, Chevrier A, Dugast C, Layet V, Rossi A, Lidereau R, Frébourg T, Hardouin A, Tosi M: Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. J Med Genet 2008, 45:438–446
27. Kwong A, Chen J, Shin VY, Ho JC, Law FB, Au CH, Chan TL, Ma ES, Ford JM: The importance of analysis of long-range rearrangement of BRCA1 and BRCA2 in genetic diagnosis of familial breast cancer. Cancer Genet 2015, 208:448–454
28. Ruiz A, Llort G, Yagüe C, Baena N, Viñas M, Torra M, Brunet A, Seguí MA, Saigí E, Guitart M: Genetic testing in hereditary breast and ovarian cancer using massive parallel sequencing. Biomed Res Int 2014, 2014:542541
29. Yang X, Wu J, Lu J, Liu G, Di G, Chen C, Hou Y, Sun M, Yang W, Xu X, Zhao Y, Hu X, Li D, Cao Z, Zhou X, Huang X, Liu Z, Chen H, Gu Y, Chi Y, Yan X, Han Q, Shen Z, Shao Z, Hu Z: Identification of a comprehensive spectrum of genetic factors for hereditary breast cancer in a Chinese population by next-generation sequencing. PLOS ONE 2015, 10:e0125571