

Gene assembling: a new approach in molecular diagnosis of hereditary breast cancer

VR. Yassaee, MD, PhD.¹, A. Dalton, PhD, MRCPath.²

Genomic Research Center of Shahid Beheshti University, Tehran, Iran.

Abstract

Background: Many disease susceptibility genes are large and consist of many exons in which point mutations are scattered throughout. Scanning each exon individually represents a tedious task which can be time consuming and expensive. There has been increasing demand for rapid and accurate methods for full scanning of unknown point mutations in large multi-exon genes. Gene Assembling is a new method for creating chimeric DNA molecules using a modified PCR reaction that allows maximizing the length of sequence that can be scanned by sensitive downstream technique.

Methods: In the present study assembling of exons 2, 20, 23 and 24 of the BRCA1 gene and their subsequent analysis by direct sequencing is demonstrated. The BRCA1 exons 2 and 20 are hot spot regions that are known to harbor particularly deleterious mutations. In order to avoid missing any mutation in these two exons, the four exons previously mentioned were assembled in the following order of preferences: 23, 20, 2 and 24. However, the order of fragments can be predetermined by primer design.

Results: The order and sequence of the component exons in the gene-assembled products were characterized by direct sequencing as predicted. Gene-assembled products from three previously ascertained heterozygotes for BRCA1 mutations were directly sequenced and gave the same sequence patterns.

Conclusion: This experience suggests that Gene Assembling technique could be applied as a highly sensitive and cost-effective method in identifying mutation in complex genes such as BRCA and CoLA1/2 helping clinical molecular diagnostic laboratories, to fulfill the demand for scanning complex genetic diseases at a lower cost.

Keywords: gene assembling, breast cancer, mutation detection method, BRCA1

Introduction

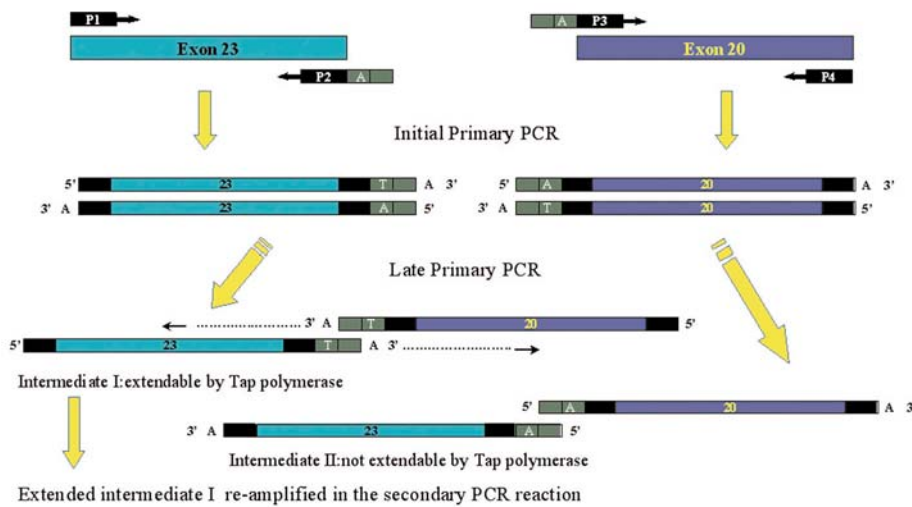
Conventional methods used to scan unknown point mutation rely on either the preparation of cDNA or the isolation of each exon of a gene individually from genomic DNA. For genes that have several exons, often spread over large distances in genomic DNA, preparation of cDNA or fragments of each exon can be a tedious process. In some cases, original tissue is not

available; leaving genomic DNA as the source of genetic material as the most convenient choice for most diagnostic molecular genetic laboratories.

The sheer number of tests involved in scanning whole genes for unknown point mutations means that the more sensitive, labor-intensive methods have severe constraints on throughput. A variety of approaches are possible, some widely used mutation detection methods such

1. **Corresponding author**, Genomic Research Center, Shahid Beheshti University of Medical Sciences, Koodakyar St., Daneshjoo Blvd, Velenjak Ave., Tehran, Iran. P.O. Box: 19395 4719, Tel: +98 21 22439959, +98 9123855076, Fax: +98 21 22439961. Email: vahid_r_yassaee@hotmail.com

2. Head of Molecular Genetics Lab, North Trent Molecular Genetics Service, Sheffield Children's Hospital, Western Bank, Sheffield - S10 2TH, UK. Phone: +44 (0)114 2717004, Fax: +44 (0)114 2737467, email: ann.dalton@sch.nhs.uk



For clarity, amplification and assembly of two exons (23 and 20) are illustrated. Primers P1-P4 are represented as rectangles; the genomic specific portions of each primer are indicated by blackly shaded areas; the linker specific segments of the internal primers P2 and P3 are represented as open areas; the position and function of the adenine (A) spacer nucleotide (incorporated between the genomic and linker segment of P2 and P3 to accommodate the 3' addition of adenine residues by Taq polymerase) is also shown.

Fig. 1. Schematic representations of the gene assembling reaction amplification and assembly of two exons 23 and 20.

as single strand conformation polymorphism (SSCP) and heteroduplex analysis (HA) are technically simple but are only capable of scanning small DNA fragments with variable sensitivity [1,2], and protein truncation test (PTT) is capable of locating nonsense mutations in DNA fragments much larger than the average exon [3].

Recently, a new approach has emerged that involves pre-screening of any abnormalities. It is called capillary heteroduplex analysis, dH-PLC [4] that may couple with DNA sequencing.

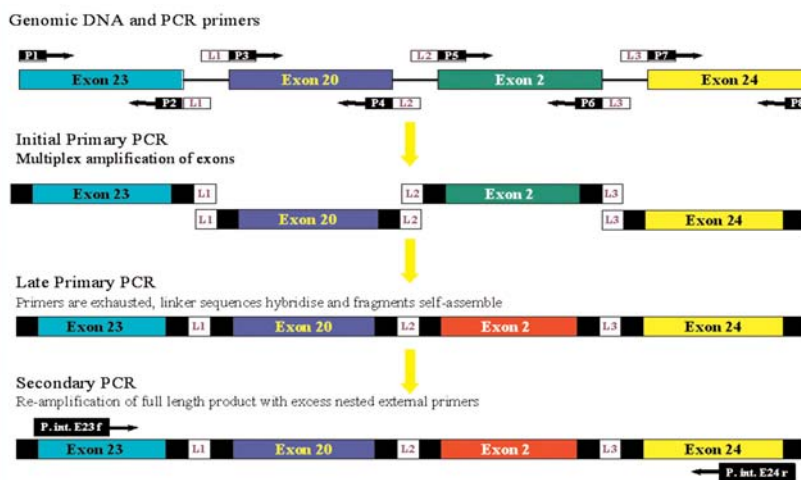
Complex genes such as BRCA1/2, APC and ColA1/2 are multi-exon genes, in which point

mutations are scattered throughout. Identifying unknown point mutations in these genes requires exploiting different techniques but which consume time and resources.

Combining several individual exons together into a single DNA fragment can thus increase throughput by reducing downstream processing and analytical steps.

Thus, we applied an alternative strategy to get rid of pre-screening step and directly sequence the whole region of interest in which led to development of a technique named Gene Assembling (GA).

GA uses genomic DNA as its substrate and



The complementary linker-specific segments of each primer pair are represented by the L1-L3.

Fig. 2. Assembling of four exons.



Number	Name	Length	Sequence 5' → 3'	Tm genomic segment (°C)	Tm Linker (°C)
01	Ex 23F (ext.)	25	atgatgaagtgacagttccagtagt	59.9	—
02	Ex 23R-L1F	35	ctg <i>cgcg</i> ggcg gct <u>A</u> caaaaggaccccatatagcac	59.8	65
03	Ex 20F-L1R	35	agc <i>cg</i> cg cg cag <u>A</u> gtgtctgtccacttcattg	59.3	65
04	Ex20R-L2F	38	tcg <i>gg</i> ccc cg gga <u>A</u> cctgtgtgaaagtatctagcactg	59.7	64
05	Ex2F-L2R	38	tcc <i>g</i> cg gg ccc gga <u>A</u> tgtgttaaagttcattggaacaga	60.3	64
06	Ex 2R-L3F	37	agg <i>cc</i> gg gg cc gct <u>A</u> cccaaattaatacactcttgtgc	59.6	65
07	Ex 24F-L3R	36	agc <i>g</i> gg ccc gg cct <u>A</u> cgattgattagagcctagcca	59.9	65
08	Ex 24R (ext.)	23	agtagccaggacagtagaaggac	59.7	—
09	Ex 23F (int.)	24	TCCTACTTTGACACTTTGAATGCT	61.1	—
10	Ex 24R (int.)	21	GTCCCTGTGGCTCTGTACC TGT	59.8	—

Primers 2 through to 7 are all internal to the respective assembled PCR product and contained one of three pairs of complementary 5' termini, which are highlighted in bold italics. The extra adenine nucleotide incorporated to accommodate the 3' terminal adenine residue added by Taq polymerase to the nascent strand is underlined in bold upper-case letter. The melting temperature (Tm) of the respective genomic or linker portions of each of the primers is also indicated.

Table 1. Primer sequences for amplification of BRCA1 exons 23, 20, 2 and 24 by Gene Assembling assay.

increases the throughput of direct sequencing when exon size is small. This technique was originally described by Wallace et al [5] and called Meta-PCR.

In the study we have used GA method coupled with direct sequencing to show chimerization of four segments of BRCA1 gene [6], exons 2, 20, 23 and 24.

Methods

Gene Assembling is a method of creating a synthetic DNA molecule comprising any combination of PCR amplifiable DNA in any order. It consists of two separate steps of PCR reaction in which small PCR amplicons are assembled into a single large chimeric DNA molecule in

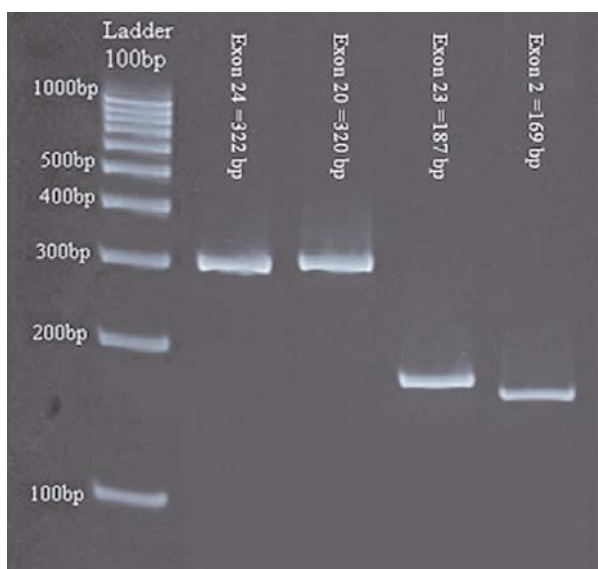


Fig. 3. Products under common reaction conditions of the individual component from BRCA1 exon 23, 20, 2 and 24 that monitored on 15% page.

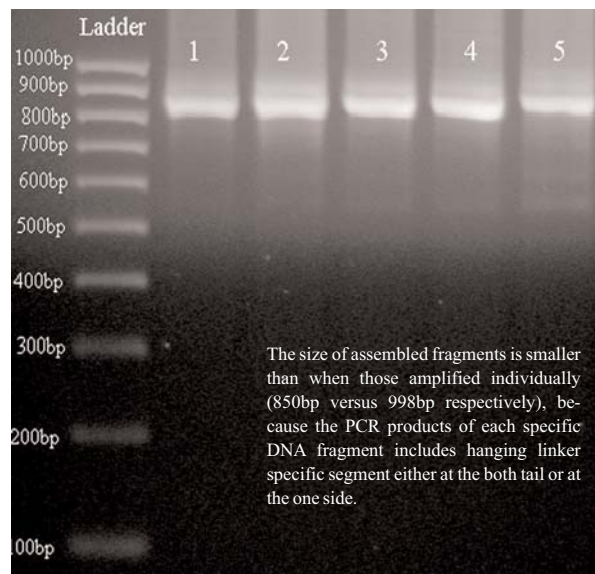


Fig. 4. The secondary PCR products (sizes = 850bp) monitored on 1% agarose. BRCA1 exons 23, 20, 2 and 24.

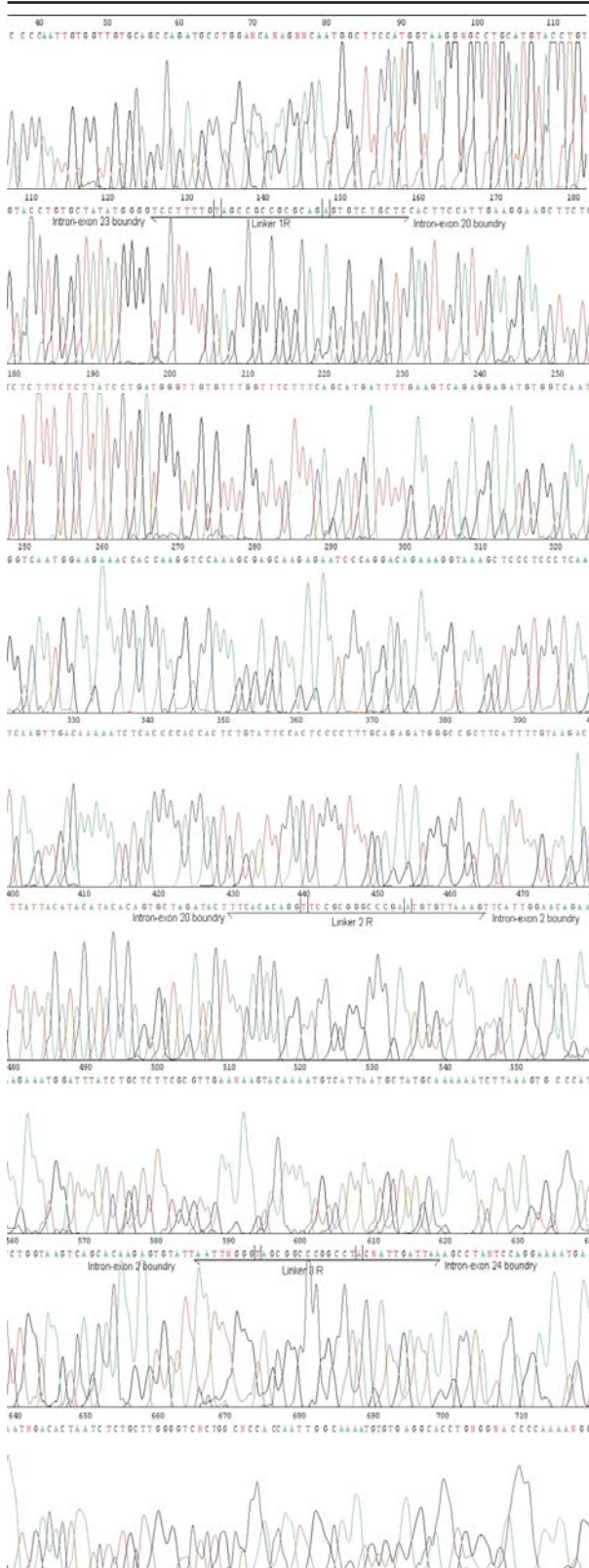


Fig. 5. Sequencing electropherograms from the 850-bp BRCA1 exons 23, 20, 2 and 24 produced by Gene Assembling technique (full-length forward orientation using primer 09).

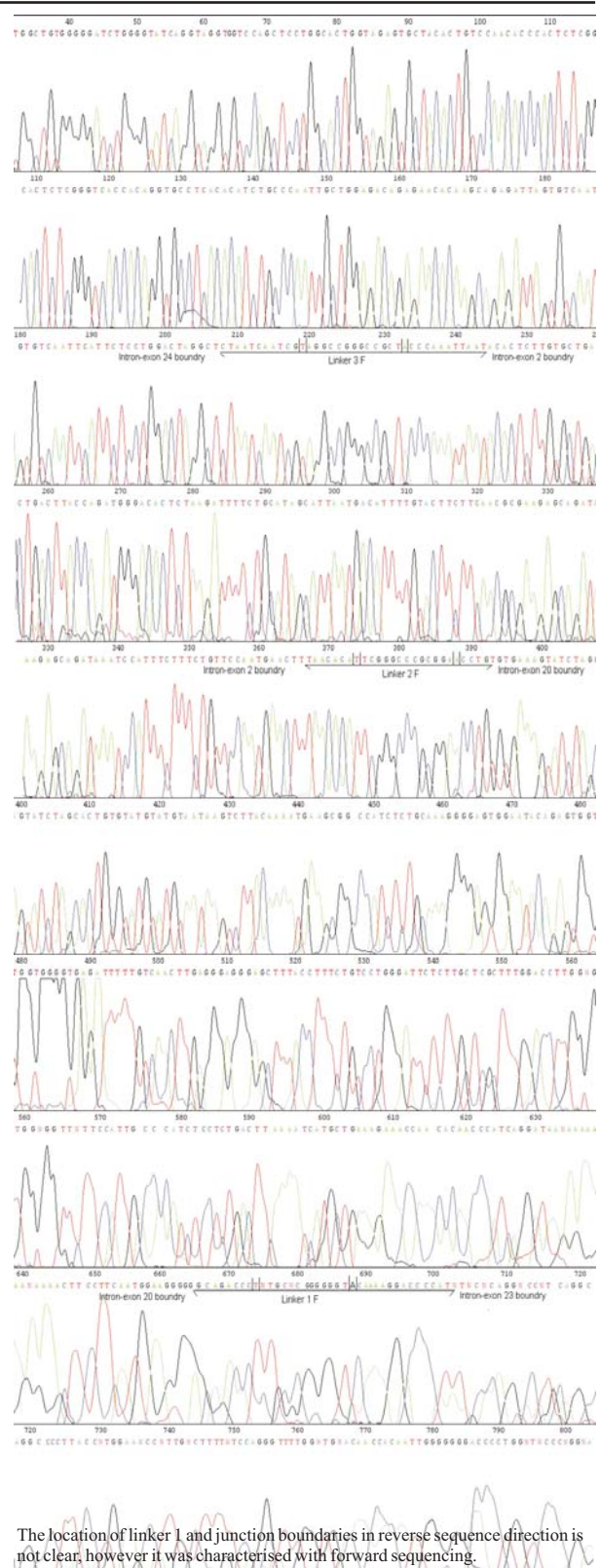
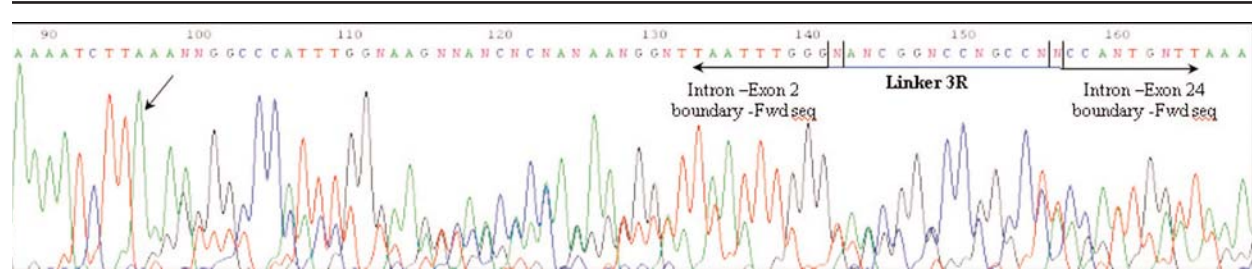


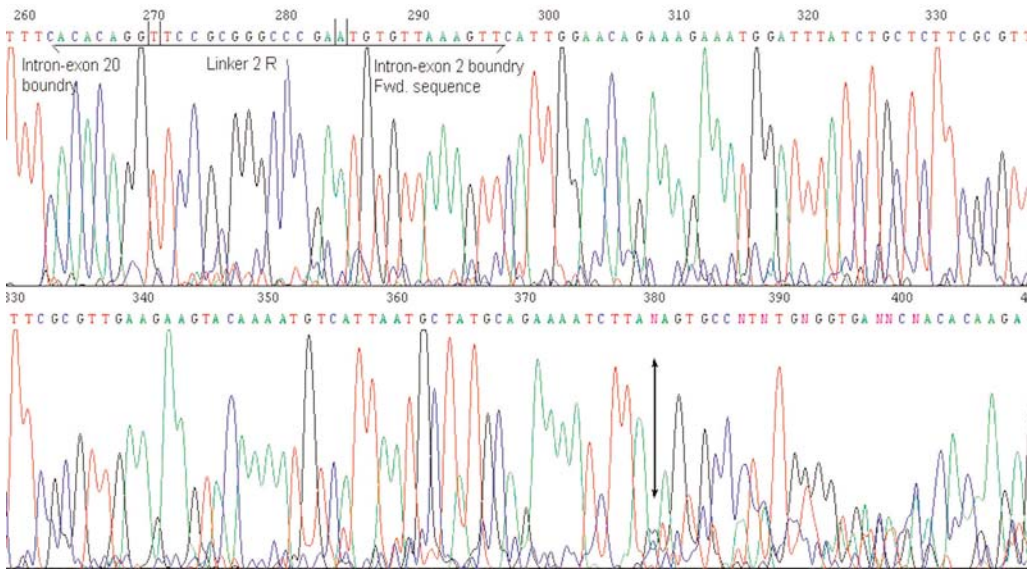
Fig. 6. Sequencing electropherograms from the 850-bp BRCA1 exons 23, 20, 2 and 24 produced by Gene Assembling technique (full-length reverse orientation using primer 10).

The location of linker 1 and junction boundaries in reverse sequence direction is not clear, however it was characterised with forward sequencing.



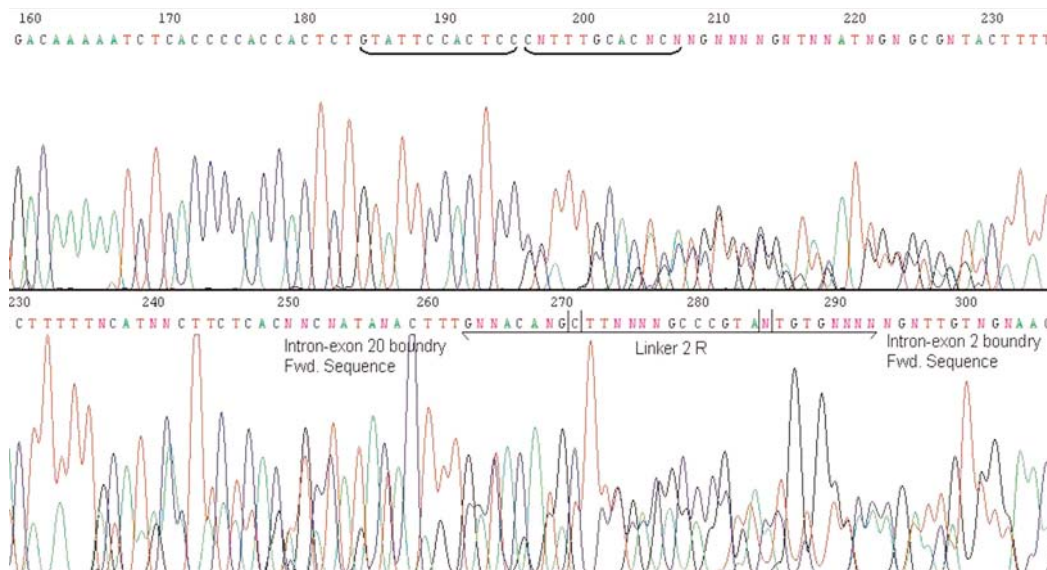
Forward sequences show the mutant heterozygote of BRCA1 exon2 (185-186 ins A) within the assembled BRCA1 exon2 to exon24 by Gene assembling technique. Arrow depicts the location of [A] insertion and frameshift sequence consequently.

Fig. 7. Sequencing electropherogram of the assembled BRCA1 exon2 to exon24, PCR products from mutant heterozygote of BRCA1 exon2 (185-186 ins A).



Forward electropherogram shows the nucleotide sequence of the mutant heterozygote of BRCA1 exon2 (185-186 delAG) within the assembled BRCA1 exon20 to exon2 by Gene Assembling technique. Arrow depicts the location of [AG] deletion and frameshift sequence consequently.

Fig. 8. Sequencing electropherogram of the assembled BRCA1 exon20 to exon2, PCR products from mutant heterozygote of BRCA1 exon2 (185del AG).



Forward electropherogram shows the nucleotide sequence of the mutant heterozygote of BRCA1, IVS 20+48 [12bp duplication gtattccactcc] within the assembled BRCA1 exon20 to exon2 by Gene Assembling technique. Down brackets depict the location of 12bp duplication gtattccactcc and frameshift sequence consequently occurred.

Fig. 9. Sequencing electropherogram of the assembled BRCA1 exon20 to exon2, PCR products from mutant heterozygote of BRCA1, IVS20+48 [12bp duplication gtattccactcc].

two coupled stages (Fig 1).

In the first step, it behaves like a multiplex PCR reaction, which uses pairs of primers at limiting concentrations ($0.04\mu\text{M}$ each) that have complementary 5' linkers attached via an adenine base to specific 3' genomic sequences. The extreme 5' and 3' primers of the GA complex are a standard pair of primers without linker sequences (Fig 1). The complementary linkers on the internal primers are all designed to have no internal secondary structure, to share a common annealing temperature which slightly exceeds that of the genomic primers and to have no significant homology with repetitive human DNA sequences.

During the initial stages of the first step of PCR individual fragments are solely amplified. During this phase the complementary linkers become incorporated into the fragments.

Towards the end of the first step the primers become exhausted and overlap extension may occur, thus producing small quantities of the full length synthetic product (Fig 2).

The second step of GA comprises a second round of PCR amplification using a 1/10 volume of the first step PCR products as template. A single pair of internally nested PCR primers re-amplifies the small quantity of full length linked products which has arisen at the end of the primary reaction (Fig 2).

Primer design

In the work presented here, direct sequencing approach as a downstream analysis was selected in Gene Assembling assay. Choosing this approach, allows one to be flexible in the positioning of primers, and make it relatively easy to determine the most suitable annealing temperatures (T_m) which is critical for the assembly of DNA fragments.

GA primers are divided into two segments, a 5' linker sequence separated from a 3' genomic specific sequence by an adenine base. A supplementary unmatched adenine residue was used to accommodate the 3' adenine overhangs

added by Taq polymerase to the nascent DNA strand. Adding a supplementary unmatched adenine (A) residue at the sense sequence causes an insertion of a supplementary thymine (T) residue at the 3' end in the antisense DNA fragments assembled into PCR products.

Linker sequences are designed to all have closely matched melting temperatures (T_{ms}), have little or no internal secondary structure and show no homology with repetitive human DNA sequences. In this study three linkers (each 13bp) were used to produce a chimeric DNA molecule from BRCA1 exons 2, 20, 23 and 24 (Fig.2). Two nested primers were designed to produce sufficient products for direct sequencing analysis and prime few residues upstream and downstream of external primers (Table 1).

The main modification in our strategy since the original paper by Wallace et al. was published is related to the design of linkers which consists of a new sequence and size which differs from those initially employed Wallace et al (1999). In the original paper mean size of linkers was about 23bp, whilst our modification diminished their size to 13bp. Using shorter linker has a significant effect on improvement of product yield.

Exons 2 and 20 are hot spot regions that are known to harbor particularly deleterious mutations. In order to avoid missing any mutation in these two exons, the above four exons were assembled in the following order of preferences: 23, 20, 2 and 24. However, the order of fragments can be predetermined by primer design; no limitation in order of preferences of segments was found. The genomic specific segments were selected to have closely equivalent T_m values for the positioning of the primers.

Table 1 shows sequences and T_m values of the 10 primers used in BRCA1 exons 2, 20, 23, 24 order of the exons.

PCR amplification

Genomic DNA was used as a template for amplification. Primers were synthesized on a 0.2 μ M scale and were purified by high performance liquid chromatography (HPLC) at Oswel Ltd (Southampton-UK).

The first and second steps of PCR amplifications are best carried out in different amplification buffers. The first step of amplification is carried out in a high Mg²⁺ buffer suited to multiplexing whereas the secondary reactions are carried out in a low Mg²⁺ low salt buffer leading to high specificity.

The first step PCR reactions performed in 20 μ l volumes using 100ng of genomic DNA, primers (40nM/L) (Table 1), a mixture of dNTP (5mM), Mg²⁺ (35mM), 10x PCR Buffer B [20mM Tris-HCl (pH 8.0 at 25°C), 100mM KCl, 0.1mM EDTA, 1mM DTT, 50% glycerol, 0.5% Tween 20 and 0.5% Nonidet-P40], 1 μ l Taq DNA polymerase (Promega, Southampton, UK; catalogue no. M1665) in accordance with manufacturer's recommendations, and distilled water up to a final volume.

Amplification was carried out on an Applied Biosystems DNA thermal cycler 9700, (Applied Biosystems Europe BV, Cheshire, UK) using the following parameters; initial denaturation 94°C (3 min), 30 cycles of 94°C (1 min), 60°C (1 min), 72°C (2 min) followed by a final synthesis of 72°C (10 min).

The secondary PCR was performed in separate 20ml volumes using 3-5ml of the primary PCR products, internal (nested) primers (5mM/L), a mixture of dNTP (2.5mM), Mg²⁺ (25mM), 10x PCR Buffer B (as above), 1 μ l Taq DNA polymerase (Promega, Southampton, UK; Catalogue No. M1665) in accordance with manufacturer's recommendations, and distilled water up to a final volume.

The buffers as custom mastermixes are now commercially available from ABgene Company (part no. for the first step PCR is 'CM129' and that for the secondary PCR is 'CM130').

The thermal cycling conditions were identi-

cal to those used for the primary amplification. On completion of the secondary PCR, a 3ml portion of the PCR products were checked on 1% agarose gel to confirm the presence of the expected 850bp fragments (Fig. 3).

Post PCR purification

The secondary assembled PCR products were recovered from agarose gel slices using a Jetsorba gel extraction kit (Genomed Inc.) according to the manufacturer's instruction. A total of 60-80ng of recovered PCR products were sequenced in both orientations using only two primers (table 1, primers numbers 09 and 10) to characterize the entire chimeric DNA molecule in two sequencing reactions instead of applying eight primers to analyse individual segments separately (Table 1, primers numbers 01- 08).

Applied Biosystems sequencer machine (model 377) and BigDye terminator cycle sequencing kit (Applied Biosystems) were used. The manufacturer's instruction were followed with the exception that the annealing temperature for the cycle sequencing reaction was increased from 50°C to 55°C to reflect the high T_m of internal primers.

Results

PCR amplification of each exon under common amplification conditions produces a single fragment of the expected size, confirming the integrity of the primer syntheses and the compatibility of the primer pairs (Fig. 3).

The secondary PCR products

The products of secondary PCR assays consistently yielded fragments of the expected size when was performed on a range of genomic DNA samples. The reaction yields were generally high with low backgrounds. For downstream analysis by direct sequencing the secondary PCR products were monitored on a 1% agarose gel confirming the integrity of the PCR products (Fig. 4).

Determination of the nucleotide sequence of wild type -PCR products

Direct sequencing confirmed that assembled PCR products comprised the expected exons in the correct orientation.

Figures 5 and 6 depict the full length of wild type assembled PCR products sequenced from both termini in single reaction. No degradation of sequence was observed across the transitions from one component to the next.

Detection of three previously ascertained heterozygous BRCA1 mutations from PCR products

To test the capability of the gene assembling technique in identifying mutant samples, three previously determined heterozygous BRCA1 mutations in exon 2 and exon 20 [7], were directly sequenced to confirm that the genotype present in genomic DNA was correctly represented in the assembled PCR products. The three mutant heterozygotes were BRCA1; 185-186 insA, 185 delAG and a 12bp duplication [BRCA1, IVS20+48]. These were clearly visible in the heterozygotes in figures 7-9.

Conclusion

Some genetic disorders/cancers, osteogenesis imperfecta and breast cancer for instance display heterogeneity that can be caused by mutation(s) scattered throughout almost randomly in the essential known predisposing genes. Gene Assembling could be particularly useful for genetic diseases with several mutational hot spots in more than one gene such as BRCA genes thereby allowing a rapid, targeted mutation scanning strategy.

The advantages to diagnostic laboratories already using techniques like DGGE, CSGE dH-PLC, or PTT of coupling Gene Assembling to their mutation scanning method is considerable time and cost saving over other methods. For instance, dHPLC is generating increasing interest in clinical genetics as a reliable tool for the

analysis of genetic alterations. The technique requires extensive optimization of conditions for each exon/PCR fragment and a high initial investment for the machine. Although the analyses are done sequentially, throughput is very high and fully automated (96 well format). Sensitivity also approaches 100% and all kinds of sequence variants are detected. Gene Assembling could be applied in combination.

Moreover, advances in sequencing technologies are making read lengths of over 1000 bases possible [8] even mutation scanning by direct sequencing could benefit from a Gene Assembling strategy. In direct sequencing approach, using shorter linker specific primer with higher GC content caused a further improvement in product yield. Notably, combination of two or more different downstream analysis methods may necessitate for scanning the entire target gene. Gene structure, splice site sequences (AT rich regions), nature of mutations, presence of multi small and large coding region, etc. are the factors that may influence in the selection of downstream analysis technique.

Do the size and/or the segment numbers matter?

In our experiments, four, three and six segments have been assembled with different approaches direct sequencing and PTT respectively. To date, up to five PCR amplifiable fragments can be combined to form a single linear amplicon [5]. Increasing the number of assembled segments would be an advantageous. Although assembling of more segments may require further PCR optimisation.

Limitation in readability of DNA sequencing, may exclude some exons to have benefit by GA approach including exon 11 in BRCA1/2 gene.

Success in assembling the segments critically depend on, designing the linker specific primer, choosing the right buffer composition, limiting of cross-contamination, and the purity of synthesized primers are the most significant parameters, which should not be overlooked.

However the size and/or number limit for segment assembly as the major question still remains.

Although this work should be valuable for laboratories wishing to replicate our system, component parts (e.g. primer sequences and other methods) may also be useful to laboratories who are implementing other strategies.

In summary Gene Assembling is a technique that can be practical for construction of a new gene for many purposes and mutation detection is one of its applications.

It is a simple, versatile and powerful method for creating chimeric DNA molecules for the analysis of multi-exon genes using a combination of sensitive techniques. It should find widespread applications in molecular genetics and should increase the productivity of mutation scanning methods, permitting many of them to fully realize their potential to analyze several small exons including hot spots in heterogeneous diseases.

Acknowledgement

The authors wish to thank Dr. Andrew Wallace (Regional Molecular Genetics Lab., St. Mary's Hospital, Manchester-UK) for his excellent advice on the primer design.

References

1. White MB, Carvalho M, Derese D, O'Brien SJ, Dean M. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* 1992; 12:301-306.
2. Sheffield VC, Beck JS, Kwitek AE, Sandstorm DW, Stone EM. The sensitivity of single-strand conformational polymorphism analysis for the detection of single base substitution. *Genomics* 1993; 16:325-332.
3. Roest P, Roberts R, Sugino S, Van Ommen GJB, Den Dunnen JT. Protein truncation test (PTT) for rapid detection of translation-terminating mutations. *Hum Mol Genet* 1993;2:1719-1721.
4. Oefner PJ, Underhill PA. Comparative DNA sequencing by denaturing high performance liquid chromatography (DHPLC). *Amer J Hum Genet* 1995;57

(suppl), pp. A266.

5. Wallace AJ, Wu CL, Elles RG. Meta-PCR: a novel method for creating chimaeric DNA molecules and increasing the productivity of mutation scanning techniques. *Genetic Testing* 1999;3(2):173-83.

6. Miki Y, Swensen J, Shattuck-Eidens D, Futreal AP, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994; 266: 66-71.

7. Yassaee VR, Zeinali S, Harirchi I, Jarvandi S, Mo-hagheghi MA, Hornby DP, Dalton A. Novel mutations in the BRCA1 and BRCA2 gene in Iranian women with early onset Breast Cancer. *Breast Cancer Research Journal* 2002; 4 (4):4): R6. Epub 2002 Apr 16. [<http://www.ncbi.nlm.nih.gov/sites/entrez?db=journal>]

8. Carrilho E, Ruiz-Martinez MC, Berka J, Smirnov I, Goetzinger W, Miller AW, et al. Rapid DNA sequencing of more than 1000 bases per run by capillary electrophoresis using replaceable linear polyacrylamide solution. *Anal Chem* 1996; 68: 3305-3313.