AD_____

Award Number: DAMD17-99-1-9302

TITLE: Investigation of Trinucleotide Repeat Expansion in Familial Breast Cancer

PRINCIPAL INVESTIGATOR: Hilmi Ozcelik, Ph.D.

CONTRACTING ORGANIZATION: Mount Sinai Hospital
Toronto, Ontario, Canada M5G 1X5

REPORT DATE: April 2002

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2002	Final (1 Sep 99 - 1 Mar 02)

4. TITLE AND SUBTITLE
Investigation of Trinucleotide Repeat Expansion in Familial Breast Cancer

5. FUNDING NUMBERS
DAMD17-99-1-9302

6. AUTHOR(S)
Hilmi Ozcelik, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Mount Sinai Hospital
Toronto, Ontario, Canada M5G 1X5

E-Mail: Ozcelik@mshri.on.ca

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Reports contains color

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Our approach is based on the hypothesis that trinucleotide repeat expansion occurs in breast cancer. Our objective is to identify putative breast cancer predisposition gene(s) which are associated with expansions of trinucleotide repeats. In this project we have developed a Rapid Expansion Detection method which can be used very efficiently in the detection of repeat expansions using small amount of genomic DNA. In the analysis of population-based breast cancer cases (212) and controls (196), we have shown the allelic distribution of CAG repeats. The distribution of smaller repeats (40-96 repeats) did not vary greatly in cases and controls. However, we have shown that a fraction (2.3%) of breast cancer cases carried expanded CAG repeats (144 and more). Further investigation on the transmission of these CAG repeat expansions suggested a paternal transmission in 2 of the cases. These repeats may be associated with and affect the function of breast cancer predisposition genes. These genes will be identified by cloniong.

Our approach has the potential to allow the rapid identification of novel breast cancer predisposition genes which will provide obvious benefits for families with breast cancer, as well as the potential for insights into the pathobiology of this devastating disease.

14. SUBJECT TERMS
Breast cancer, predisposition, trinucleotide repeats, Rapid Expansion Detection (RED)

15. NUMBER OF PAGES
25

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

Genetic factors have been shown to influence the risk of developing breast cancer. To date only two dominant predisposition genes, BRCA1 and BRCA2, have been identified and shown to confer a high risk of developing breast cancer when mutated in the germline (Miki et al 1994, Wooster et al., 1995, Tavigtian et al., 1996). However, there is now clear evidence that other important predisposition genes contribute to the susceptibility to breast cancer (Serova et al. 1997). Classic genetic approaches to identify susceptibility loci involve whole genome scans by linkage analysis. These studies are labor-intensive, and require extended families for success. Our approach is based on the hypothesis that genetic anticipation occurs in some breast cancer families, and is characterized by a decrease in age of onset of the disease in successive generations of affected family members. Recent investigations in other genetic diseases have revealed that anticipation is the result of intergenerational instability of trinucleotide repeats in disease susceptibility genes (Sanjeeva et. al., 1997, Margolis et al., 1999). Importantly, the typical expansion that occur are readily detected by the Repeat Expansion Detection (RED) system, thus bypassing the traditional search for predisposition genes. Our objective, therefore, is to identify putative breast cancer predisposition gene(s) which have undergone intergenerational expansions of trinucleotide repeats during germline transmission from mother to daughter. We will utilize the BRCA1 and BRCA2 mutation negative families that we have accrued, and study mother-daughter pairs with a significant decrease in the age of breast cancer onset from one generation to the next. Our approach has the potential to allow the rapid identification of novel breast cancer predisposition genes which will provide obvious benefits for families with breast cancer, as well as the potential for insights into the pathobiology of this devastating disease.

Statement of Work:

Technical Objective 1:	**Identification of trinucleotide repeat expansion in mother daughter-pairs with breast cancer.**
Task 1: Months 1-3	The optimization of the RED analysis on control samples.
Task 2: Months 2-14	Digestion of DNA samples with EcoR1 and preparation of gel fragments from mother-daughter pairs as well as from control specimens. The RED analysis will be carried out for CAG/GTC trinucleotide repeat.

Task 3: Months 12-24 RED analysis for the CGG/GCC repeat.

BODY

RESULTS AND DISCUSSION

1. Developing a modified version of RED method

Our initial goal was to develop the most efficient approach for trinucleotide repeat detection before actually screening patient DNA samples. Increasing the efficiency is very important for the proposed RED method since the method utilizes a linear as opposed to exponential amplification as would be in a regular PCR reaction. Unfortunately, this necessitates the use of high quantities of starting genomic template from the patient samples ((Schaling et al., 1993, Lindblad et al., 1996, Vincent et al., 1997, Zander et al., 1998). Since patient DNA material is a limited source, we were specifically interested in developing an efficient approach. The following modifications were introduced to the traditional RED method. In contrast to the traditional protocol (see attached manuscript (proofs) for details of the protocol).

a) previous to ligation reaction, the oligonucleotides are phosphorylated using γ-^{32}P ATP and thus labeled with radioactivity,

b) the gel transfer and hybridization steps are eliminated,

c) the amount of genomic DNA required is reduced to a great extent.

We have demonstrated that the amount of oligonucleotide used in the labeling reaction has a significant influence on the efficiency of the ligation reaction. The use of a large amount of oligonucleotides would reduce the total labeling efficiency, resulting in excess unlabeled oligonucleotides in the labeling reaction. During the ligation reaction, unlabelled oligonucleotides compete with labeled oligonucleotides in binding to the genomic DNA template. An excess of unlabelled oligonucleotides will influence the production of ligation products, specifically those made using the low copy number expanded repeats as a template. Titration for different concentrations of $(CTG)_8$ oligonuleotide has demonstrated that the efficient ligation for large repeats can be performed using oligonucleotide concentrations ranging between 0.025 and 0.5 μM in the labeling reaction (*Figure 1*). As seen in the figure, there is an inverse relation between the oligonucleotide concentration used in the labeling reaction and the

quality of the ligation products obtained. The ligation products were successfully detected at oligonucleotide concentration of 0.025 μM, whereas they are almost absent at 1.0μM.

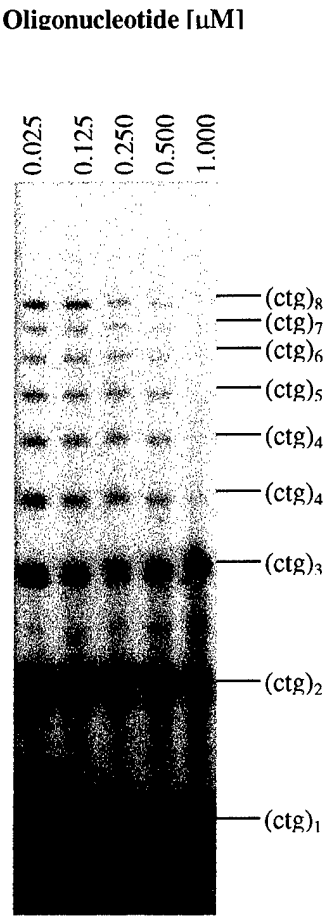

Figure 1: Titration for oligonucleotide concentration in the labeling reaction. The labeling reactions were carried out using different oligonucleotide concentrations ranging from 0.025 to 1.0 μM in the presence of 15 μCi of γ-^{32}P ATP (specific activity of 3000 Ci/mmole) and 0.175 μM ATP. The human leukemia cell line (HL-60) DNA extracted using phenol/chloroform is used as a template. A 10 μl aliquot of the reaction was loaded on each gel lane.

The modification applied to the protocol has also shortened the protocol to a great extent. The introduction of labeled oligonucleotides into the ligation reaction has eliminated the need for two cumbersome steps. These steps involve the transfer of the ligation products from gels to nylon membranes, and hybridization of the nylon membranes using radiolabeled probes. The

elimination of these steps in the modified method successfully reduced the labor and the time involved in performing the traditional RED analysis.

A significant advantage of the modified RED analysis is that it allows the use of small amount of genomic DNA as a starting material compared to the traditional method. In fact, the necessity for large amounts of genomic DNA may have been one of the reasons for the limited application of this technique in diseases other than neurological disorders. In this study, we have shown that the modified RED method can be used to detect genomic repeat expansions using as low as 50-100 ng of genomic DNA (*Figure 2*). This is comparable to the amount of genomic DNA used in a standard PCR reaction. As seen in the figure, the modified RED method was able to detect sufficiently intense ligation signal using 200 ng of genomic DNA. The signal was also visible at lower concentrations (50 and 100 ng), although not as intense.

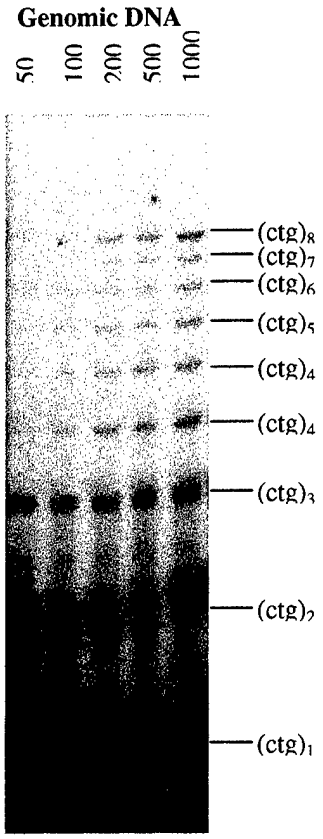

Figure 2: RED analysis using varying amounts of genomic DNA as a template. Genomic DNA from HL-60 cell line ((CTG) $_{80}$) was used to determine the sensitivity of the RED method in terms of DNA concentration. The ligation reaction was carried out in 20 µl volume (as described in the text) using various amounts of starting genomic DNA ranging from 50 ng to 1 µg.

The number of neurological diseases associated with TRE has been increasing extensively. The RED method constitutes a powerful tool to identify other diseases in addition to neurological and neuromuscular disorders, caused by the same mechanism. The application of this method is not restricted to the study of trinucleotide repeats only. Several other diseases have also been reported to be associated with the expansion of other repeat motifs including tetranucleotides, pentanucleotides, hexanucleotids and even dodecanucleotides (Liquori et al., 2001, Matsuura et al., 2000, Sirugo and Kidd, 1998, Lalioti et al., 1997). The modified RED method described here allows a rapid and sensitive screening for TRE which can be efficiently applied to various diseases in studies where limited amounts of patient genomic DNA are available

2. Screening for AAG, CAT, GTT, CCG and CAG repeats

We have optimized the RED method to investigate the role of 5 types of trinucleotide repeats, CAG, CCG, AAG, CAT, GTT, in breast cancer. CAG, CCG, and AAG repeats are selected because they are shown to undergo repeat expansion, leading to neurological and neuromuscular disorders. CAT and GTT repeats are selected since they occur commonly in the genome (Serova et al., 1997, Sanjeeva et al., 1997). The method used to study these repeats is described above. Beside the sequences of the oligonucleotides used, the experimental protocol for each repeat was quite similar.

a) Screening of breast cancer cases and cell lines

In order to determine the distribution of these types of repeats we have initially studied 12 cancer cell lines and 30 breast cancer cases using RED method.

Analysis of CAT and GTT repeats has revealed a maximal repeat size of 18 and 24 repeats respectively (see Figure 3b and 3d). There was no difference in the allelic lengths observed among the samples within breast cancer cases or cell lines. The RED is capable of determining a minimal expansion (difference) of 8-12 trinucleotide repeats (24-36 base pairs), therefore it is not sensitive to detect any change in the allelic size that is smaller than 24-36 base pairs. The RED results clearly indicates that the CAG and GTT repeats in our population do not show large expansions or variation in repeat size that is more than 24-36 base pairs in length. Our findings

regarding the allelic distribution of CAT and GTT repeats were consistent with other studies (Schaling et al., 1993, Hofferbert et al., 1997).

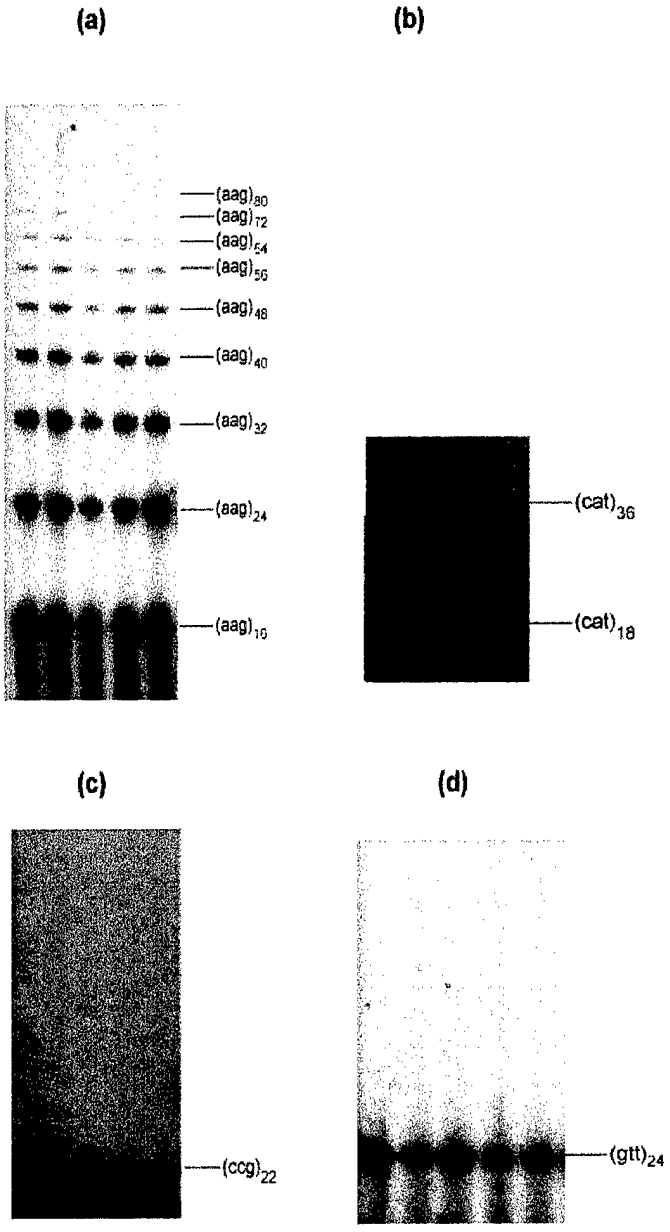

Figure 3: Products of RED reaction using the above described method using the following oligonucleotides in the ligation reaction: $(AAG)_8$ (a), $(CAT)_9$ (b), $(CCG)_{11}$ (c) and $(GTT)_{12}$ (d).

In contrast to CAT and GTT repeats, the screen for AAG repeats demonstrated large maximal repeat sizes in the range of 80 and above (*Figure 3*). The length of AAG repeats were not distinguishable among the breast cancer cases, cell lines and a positive control DNA specimen obtained from a patient with Frederick's Ataxia (with an expansion in AAG repeat length). Since the RED method is not capable of detecting exact differences among large repeats, we were not able to observe if there was variability in the allelic lengths in breast cancer cases. The observation regarding large repeat lengths for AAG repeats was also made previously (Hofferbert et al., 1997, Vincent et al., 1998).

The RED analysis of CCG repeats has revealed a maximal of 22 trinucleotide repeat lengths. We can not conclude whether this finding is correct since the same RED application did not detect larger CCG repeats in a positive control DNA specimen obtained from a Fragile X patient (with CCG repeat expansion). This is probably due to secondary structure formation or other processes that would inhibit oligonucleotide binding and ligation at these repeat regions. The similar observations on difficulty in detecting CCG expansions were also made by other investigators (Hofferbert et al., 1997, Pearson et al., 1996).

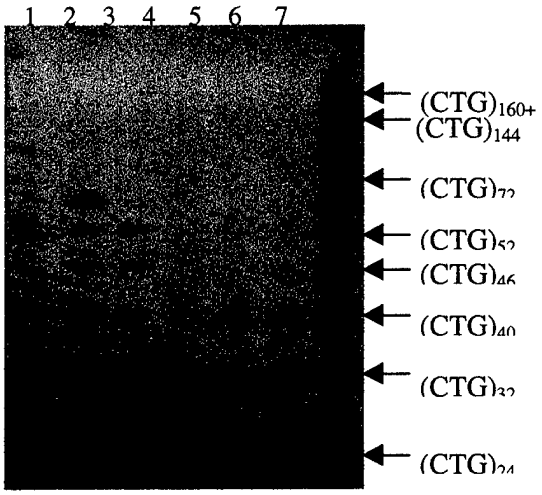

Figure 4: RED analysis of breast cancer cases. Lanes 1, 4-6 represent patient RED results of differing CAG-repeats in the short repeat range (24-72-CAG-repeat). Sample in lane 7 shows a CAG expansion of at least 144 CAG-repeats. During RED reaction the ^{32}P-labelled CAG-oligos (24-mer) were ligated at the presence of a Thermastable Ligase. The DNA samples from the cases with large CAG- repeats resulted in larger ligation products as seen in lane 7.

Allelic variance in the distribution of trinucleotide repeats was only observed for CAG repeats (*Figure 4*). Analysis of CAG repeats have demonstrated a spectrum of alleles, majority of, which ranged from 24 to 96, repeats. In one of the cases we were able to detect a large expansion that was in the range of 144 and more repeats (*Figure 4*) indicating that large expansions of CAG repeats may be associated with breast cancer. An extensive analysis to determine the significance of this finding has been carried out using a larger panel of breast cancer and population control cases (see below).

b) Screening of breast cancer (mother-daughter) pairs

As a part of our objective in this proposal, we have compared the lengths of AAG, CAT, GTT, CCG and CAG repeats between 20 pairs of cases of mothers and daughters both diagnosed with breast cancer. Similar to observations in the panel of breast cancer cases and the cell lines, with the exception of CAG repeats, the lengths for each repeat were same among all breast cancer pairs. However, as expected, we have detected a spectrum of alleles for CAG repeats in this panel (*Figure 5*).

Among 20 mother-daughter pairs studied, one (5.0%) has shown a CAG expansion that was detected in daughter but not in mother. The magnitude of the CAG expansion in this case was at least 144 CAG repeats, whereas the shorter allele detected in mother was 40 CAG repeats. In this family (4050), the age of diagnosis for breast cancer was, 50 and 79 for daughter and mother, respectively. Additional to maternal side of the family (maternal grandmother with breast cancer) there was also breast cancer history in the paternal site (paternal aunt and a stepsister from another marriage father). The father of the daughter with expansion was also diagnosed with colon cancer at the age of 79. In this family, using RED analysis, we have shown that the expanded repeat was not inherited from the mother, suggesting that it may be inherited from the father. However, we could not confirm this since DNA sample was not available from the father.

In order to understand the inheritance of expanded repeats, we have also studied additional 50 breast cancer cases where DNA samples were available from their parents. In this group, RED analysis has demonstrated an additional breast cancer case with a large CAG expansion ($(CAG)_{144+}$). The age of diagnosis for breast cancer in this individual was 39. There was no breast cancer family history in either maternal or paternal side of the family. Her mother was

diagnosed with basal cell carcinoma at the age of 62, and her father was diagnosed with bladder cancer at the age of 74. Red analysis of the parents has revealed a 40-CAG repeat allele in mother and an expanded 144+ CAG repeat in the father, demonstrating a paternal transmission for the expanded repeat in this family.

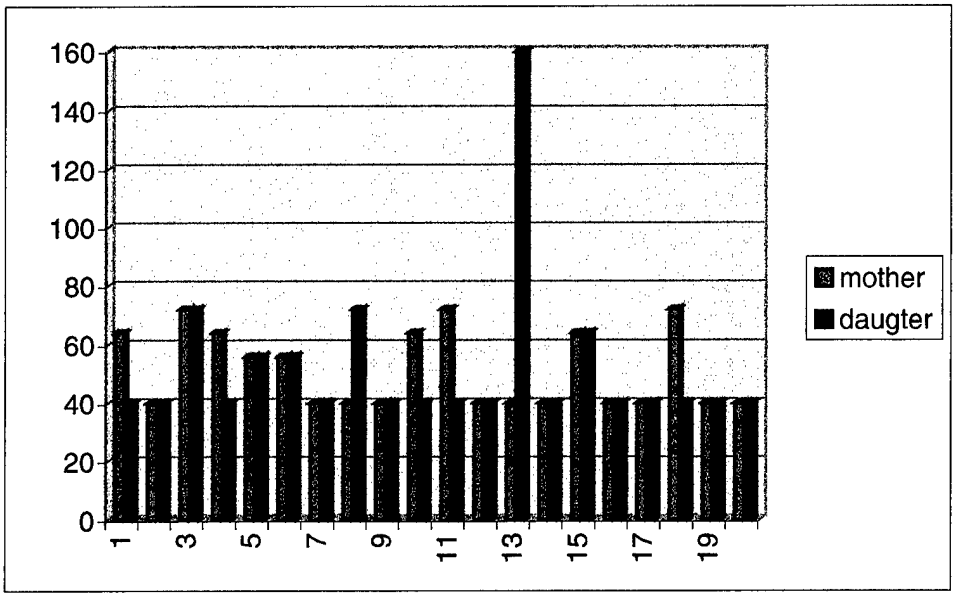

Figure 5: CAG-repeat length comparison in mother and daughter pairs using RED analysis. In 1 of the 20 pairs analyzed, an expansion of a CAG-repeat have been observed (pair 13). In this pair the CAG expansion was observed in the daughter, and not in mother.

In two cases with CAG repeat expansion, we have shown that the transmission of the expanded allele was inherited paternally (confirmed in one and suggested in the other one). The mechanisms for trinucleotide repeat transmission is not fully understood, however several models have been suggested (Usdin and Grabczyk, 2000). In some loci expansion of trinucleotide repeat have been found to occur during paternal or maternal transmission (Takiyama et al, 1997). In order to determine the transmission mode of expanded CAG repeats, specimens (daughter and parents) from a larger group of families with CAG-repeat expansions are required.

3. Extensive Screening for CAG/CTG repeats

In order to determine the distribution of CAG repeats more extensively, we have carried out a case-control study, investigating DNA samples from 212 breast cancer cases and 196 population controls using RED analysis. The cases and controls studied were selected from the population-based Ontario Familial Breast Cancer Registry (OFBCR).

a) Selection of cases and controls

i) OFBCR Design and Methods

All Ontario women aged 20-54, 35% of women aged 55-69, and all men aged 20-79 diagnosed with invasive breast cancer, with pathologic confirmation, between 1996 and 1998 were identified through the population-based OCR. After receiving permission from physicians (for about 90% of cases), a family history questionnaire was mailed to the patients. When the family history was received (response rate is about 68%), the initially identified individual with breast cancer (proband) was classified according to set criteria as being at potentially increased genetic risk (familial) or likely sporadic. All those who are considered familial and a 25% random sample of those who are considered sporadic were invited to participate further in the OFBCR and about 90% agreed. Sporadic probands were treated in the same manner as familial probands. Further participation included filling out epidemiological risk factor and diet questionnaires and providing a blood sample, at which time probands receive genetic counseling.

ii) Breast Cancer Cases

About 90% of probands who respond to the family history questionnaire identify themselves as Caucasian. Since the prevalence of SNPs varies greatly among different population we limited case selection to those who are Caucasian. Although we do not have the initial response rate of Caucasian ethnicity alone (overall is 68%), we know from the other phases of the registry that response rate of Caucasians is higher compared to non-Caucasian populations. Therefore, participants in this ethnic group are more likely to be representative. Very few (~1%) probands are male. Therefore, we will only include female breast cancer cases. Because of sampling on the basis of age and family history, the cases are no longer representative of all women aged 20-69 with incident breast cancer in Ontario. Women under age 55 and those with a family history are over-represented. As can be seen in Table3, a greater number of women can be included in a

representative sample of women under 55. Younger women are also more likely to have a genetic component to their disease. Therefore we focused on this group in this study. In order to re-create a representative sample of women under age 55 with breast cancer, we have sampled 75% non-familial and 25% of those who meet familial criteria in the Registry. This allowed us to create a representative sample for this age group in spite of the sampling by genetic risk criteria.

iii) Population controls

Population controls frequency-matched by five-year age groups to female cases in the OFBCR are being recruited into the registry. Population controls, have been recruited by calling randomly selected residential telephone numbers. Eligible women who agreed to participate were mailed a package including the family history questionnaire, epidemiologic questionnaire, and diet questionnaire that are used in the OFBCR. Approximately 75% of those responding to the questionnaires expressed a willingness to provide a blood sample, and samples are currently being collected.

b) Distribution of CAG repeats in cases and controls

The distribution of CAG repeats in 212 breast cancer cases and 196 population controls were given in *Figure 6*. We have observed that there was no difference in distribution of short, $(CAG)_{40}$, and medium length repeats, $(CAG)_{48}$ to $(CAG)_{96}$, between the breast cancer cases and controls. However, there was a fraction of breast cancer cases (2.4%) which revealed expansions larger than $(CAG)_{144}$ (*Table 1*). No expansion of this magnitude has been detected in 196 population controls samples.

It has been published that two loci, ERDA1 (Nakamoto et al. 1997) and SEF2-1 gene (Breschel et al., 1997), have been shown to account for most of the large CAG alleles detected by RED method. The repeat allelic distribution of these sites have been determined to be between 90-120 repeats in various ethnic populations (Breschel et al., 1997, Nakamoto et al., 1997, Benson et al., 1998, Ikeuchi et al., 1998, Deka et al., 1999). Therefore the RED results of 144 CAG-repeat and up is not due to the alleles of these sites. The allelic distribution of the trinucleotide repeats may also vary due to the differences in the ethnic background (Sirugo et al.,

1997, Gyurus et al., 1999). However in our study all of the cases and the controls were of Caucasian background mostly of Northern European descent.

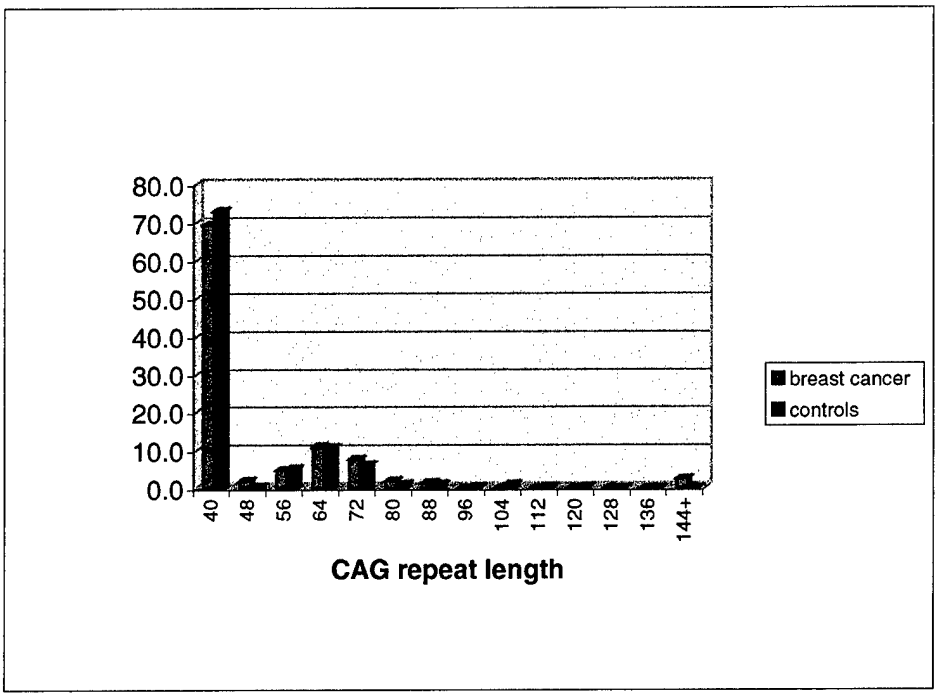

Figure 6: Distribution of CAG repeat alleles in 212 breast cancer cases and 196 population controls. A fraction of breast cancer cases has shown to carry large CAG repeats (144+) compared to population controls.

Cases	Age of dx	Father (dx, age)	Mother (dx, age)	Paternal side (brca fhx)	Maternal side (brca fhx)	BRCA1/2
14046-001	54	Pancreas 72	Breast 50	No	No	negative
16135-001	45	-	Skin 44 Uterus 48	No	Yes	negative
00097-001	47	-	Cervix 44	No	No	N/A
02193-001	53	-	-	No	No	negative
02418-001	48	-	Basal cell 65	No	No	N/A

Table1: Personal and family history of breast cancer cases with CAG expansions

Since the total number of cases with CAG repeat expansions is small (*Table 1*) it is difficult to derive phenotypic similarities that may be associated with having a CAG expansion.

Our study population (212 cases) consisted of families with varying degrees (from weak to strong) of familial risk. Interestingly, the CAG expansions were observed mostly in families with moderate to weak familial history of breast cancer. The lack of breast cancer in the majority of the mothers (4/5) weakens the possibility that these hypothesized disease-causing expansions are transmitted maternally. Our evidence for paternal transmission in this proposal strengthens this suggestion. We are not able to provide further evidence for this since the DNA specimens from parents of these cases were not available.

RED analysis does not provide information on the location of expansions in the genome, therefore we are far from predicting the mechanism of their functional involvement in breast cancer at this point. The extensive knowledge of how repeat expansions may disrupt the function of some genes comes from experience with neuromuscular and neurological diseases (Serova et al., 1997, Sanjeeva et al., 1997). The location of expanded trinucleotide repeat relative to gene structure may vary in different disorders. They can be found in exonic, intronic, 5' or 3' untranslated regions, affecting function or intrinsic properties of the genes. The molecular basis of the association of repeat expansion and disease is not completely clear and several hypothesis has been suggested (Usdin and Grabczyk et al., 2000).

Expansions of CAG repeats located within protein coding regions are currently known to be cause eight disorders (Margolis et al., 1999). The mechanism is similar in the majority of these diseases. CAG encodes the amino acid glutamine, and the mutant proteins contain extended stretches of glutamine residues. This long glutamine stretches are believed to result in toxic gain of function of proteins which than contributes to the phenotypic outcomes. It is highly possible that CAG repeat expansions detected in breast cancer may act through similar mechanism affecting genes that are crucial for breast cancer development. It will be a great value to isolate these genes, evaluating the influence of these expansions on their function and the intrinsic properties.

4. Future Work

We have received funding from U.S. Army Medical Research and Materiel Command (Fiscal year 2002) for a project titled "Cloning and Characterization of Expanded CAG-Repeat Containing Sequence(s): Identification of Candidate Breast Cancer Predisposition Gene(s)". The aim of this study is to identify breast cancer predisposition genes, functions and/or intrinsic

properties of which are affected by CAG-repeat expansions. In this proposal we are planing to develop a cloning strategy to identify genes from cases shown to have large CAG-repeat expansion in current study. The cloned sequences will be used to identify complete genes. The properties of these genes and their interaction with the CAG-repeat expansions will be studied.

Identification of genes with unstable repeat expansions will open new avenues to the study of the molecular genetics of breast cancer.

5. Future Vision

A. Molecular Epidemiological Studies

It is crucial to design and conduct molecular epidemiological studies to establish a phenotype-genotype correlation. Studies of various populations and determination of the associated breast cancer risk for carriers of these genetic alterations will determine their utility in genetic testing. This information, then, will be used to develop molecular diagnostic protocols and genetic counseling strategies for breast cancer families.

B. Segregation Studies

The identification of families with CAG-repeat expansion and the recruitment of the family members is a necessary step for understanding the inheritance of these repeats. Segregation analysis will be applied to study the transmission of the disease.

C. Functional Studies

Functional tests will be needed to study the effect of the expanded CAG-repeat on the function and the intrinsic properties of genes containing or surrounding the CAG-repeat expansions.

KEY RESEARCH ACCOMPLISHMENTS

- Development of an efficient Rapid Expansion Detection Method
- Detection of large CAG repeat expansions in breast cancer cases compared to population controls
- Providing evidence of paternal transmission of large CAG repeats in families
- Suggestion of the presence of breast cancer predisposition genes associated with CAG repeat expansion

REPORTABLE OUTCOMES

1. Manuscript in press (*Biotechniques*), titled "A modified rapid expansion detection (RED) method to analyze the CAG/CTG repeat expansions (see attached manuscript (proofs)).
2. New funding received from U.S. Army Medical Research and Materiel Command to identify the genes that were suggested in the current proposal to be involved with CAG expansions
3. Manuscript in preparation to report the presence of large CAG repeat alleles in breast cancer cases compared to population controls.
4. Poster presentation in Oncogenomics Conference organized by Nature Genetics (January 25-27, 2001, Tuscon, AZ). Title "Trinucleotide repeat expansion in sporadic and familial breast cancer".

CONCLUSION

In this project have suggested the association of CAG repeat expansions with breast cancer. We have developed a Rapid Expansion Detection method, which can be used very efficiently in the detection of repeat expansions using small amount of genomic DNA. In the analysis of population-based breast cancer cases (212) and controls (196), we have shown the allelic distribution of CAG repeats. The distribution of smaller repeats (40-96 repeats) did not vary greatly in cases and controls. However, we have shown that a fraction (2.4%) of breast cancer cases carried expanded CAG repeats (144 and more). Further investigation on the transmission of these CAG repeat expansions suggested a paternal transmission in two of the cases. These repeats may be associated with and affect the function of breast cancer predisposition genes. We have received further funding to identify and characterize the genes that may be involved with these CAG repeats.

Our approach has the potential to allow the rapid identification of novel breast cancer predisposition genes which will provide obvious benefits for families with breast cancer, as well as the potential for insights into the pathobiology of this devastating disease.

REFERENCES

Benson KF, Horwitz M, Wolff J, Friend K, Thompson E, White S, Richards RI, Raskind WH, Bird TD. 1998. CAG repeat expansion in autosomal dominant familial spastic paraparesis: novel expansion in a subset of patients. Hum Mol Genet 1998 7(11):1779-86.

Breschel TS, McInnis MG, Margolis RL, Sirugo G, Corneliussen B, Simpson SG, McMahon FJ, MacKinnon DF, Xu JF, Pleasant N, Huo Y, Ashworth RG, Grundstrom C, Grundstrom T, Kidd KK, DePaulo JR, Ross CA. 1997. A novel, heritable, expanding CTG repeat in an intron of the SEF2-1 gene on chromosome 18q21.1. Hum Mol Genet 6(11):1855-63.

Deka R, Guangyun S, Wiest J, Smelser D, Chunhua S, Zhong Y, Chakraborty R. 1999. Patterns of instability of expanded CAG repeats at the ERDA1 locus in general populations. Am J Hum Genet 65(1):192-8.

Gyurus P, Molnar J, Melegh B, Toth G, Morava E, Kosztolanyi G, Mehes K. 1999. Trinucleotide repeat polymorphism at five disease loci in mixed Hungarian population. Am J Med Genet 187(3):245-50.

Hofferbert, S., Schanen, N.C., Chehab, F., Francke, U. 1997. Trinucleotide repeats in the human genome: size distributions for all possible triplets and detection of expanded disease alleles in a group of Huntington disease individuals by the repeat expansion detection method. Hum Mol Genet *6*:77-83

Ikeuchi T, Sanpei K, Takano H, Sasaki H, Tashiro K, Cancel G, Brice A, Bird TD, Schellenberg GD, Pericak-Vance MA, Welsh-Bohmer KA, Clark LN, Wilhelmsen K, Tsuji S. 1998. A novel long and unstable CAG/CTG trinucleotide repeat on chromosome 17q. Genomics 15;49(2):321-6.

Lalioti, M.D., H.S. Scott, C. Buresi, C. Rossier, A. Bottani, M.A. Morris, A. Malafosse and S.E. Antonarakis. 1997. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature *386*:847-851.

Lindblad, K., M.L. Savontaus, G. Stevanin, M. Holmberg, K. Digre, C. Zander, H. Ehrsson, G. David, A. Benomar, E. Nikoskelainen, Y. Trottier, G. Holmgren, L.J. Ptacek, A. Anttinen, A. Brice and M. Schalling. 1996. An expanded CAG repeat sequence in spinocerebellar ataxia type 7. Genome Res *6*:965-971.

Liquori, C.L., K. Ricker, M.L. Moseley, J.F. Jacobsen, W. Kress, S.L. Naylor, J.W. Day and L.P. Ranum. 2001. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science *293*:864-867.

Margolis RL, McInnis MG, Rosenblatt A, Ross CA. 1999. Trinucleotide repeat expansion and neuropsychiatric disease. Arch Gen Psychiatry 56(11):1019-31.

Matsuura, T., T. Yamagata, D.L. Burgess, A. Rasmussen, R.P. Grewal, K. Watase, M. Khajavi,

A.E. McCall, C.F. Davis, L. Zu, M. Achari, S.M. Pulst, E. Alonso, J.L. Noebels, D.L. Nelson, H.Y. Zoghbi and T. Ashizawa. 2000. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet *26*:191-194.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavigain S, Liu Q, Cochran C, Bennet LM, Ding W, Bell R, Rosenthal J, Hussey C, Tran T, McClure M, Frye C, Hattier T, Phelps R, Haugen-Strano A, Katcher H, Yakumo K, Gholami Z, Shaffer D, Stone S, Bayer S, Wray C, Bogden R, Dayananth P, Ward J, Tonin P, Narod S, Bristow PK, Norris FH, Helvering L, Morrison P, Rosteck P, Lai M, Barret C, Lewis C, Neuhausen S, Cannon-Albright L, Goldgar D, Wiseman R, Kamb A, and Skolnick M. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266: 66-71.

Nakamoto M, Takebayashi H, Kawaguchi Y, Narumiya S, Taniwaki M, Nakamura Y, Ishikawa Y, Akiguchi I, Kimura J and Kakizuka A 1997. A CAG/CTG expansion in the normal population. Nature genetics 17:385-386.

Pearson CE, Sinden RR. 1996. Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. Biochemistry 35(15):5041-53.

Sanjeeva P and Housman DE. 1997. The complex pathology of trinucleotide repeats. Current Opinion in Cell Biology 9:364-372.

Schalling, M., T.J. Hudson, K.H. Buetow and D.E. Housman. 1993. Direct detection of novel expanded trinucleotide repeats in the human genome. Nat Genet *4*:135-139.

Serova OM, Mazoyer S, Puget N, Dubois V, Tonin P, Shugart YY, Goldgar D, Narod SA, Lynch HT, Lenoir G. 1997. Mutations in BRCA1 and BRCA2 in breast cancer families: are there more breast cancer-susceptibility genes? American Journal of Human Genetics 60(3):486-495.

Sirugo, G. and K.K. Kidd. 1998. Repeat expansion-detection analysis of telomeric uninterrupted (TTAGGG)n arrays. Am J Hum Genet *63*:648-651.

Sirugo, G., A.S. Deinard, J.R. Kidd and K.K. Kidd. 1997. Survey of maximum CTG/CAG repeat lengths in humans and non-human primates: total genome scan in populations using the Repeat Expansion Detection method. Hum Mol Genet *6*:403-408.

Takiyama Y, Sakoe K, Amaike M, Soutome M, Ogawa T, Nakano I, Nishizawa M. 1999. Single sperm analysis of the CAG repeats in the gene for dentatorubral-pallidoluysian atrophy (DRPLA): the instability of the CAG repeats in the DRPLA gene is prominent among the CAG repeat diseases. Hum Mol Genet, 8:453-7.

Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, Thorlacius S, Offit K, Stoppa-Lyonnet D, Belanger C, Bell R, Berry S, Bogden R, Chen Q, Davis T, Dumont M, Frye C, Hattier T, Jammulapati S, Janecki T, Jiang P, Kehrer R, Leblanc JF, Goldgar DE, et al. 1996. The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. Nature Genetics 12(3):333-337.

Usdin, K. and E. Grabczyk. 2000. DNA repeat expansions and human disease. Cell Mol Life Sci *57*:914-931.

Vincent JB, Kalsi G, Klempan T, Tatuch Y, Sherrington RP, Breschel T, McInnis MG, Brynjolfsson J, Petursson H, Gurling HM, Gottesman II, Torrey EF, Petronis A, Kennedy JL. 1998. No evidence of expansion of CAG or GAA repeats in schizophrenia families and monozygotic twins. Hum Genet, 103:41-7.

Vincent, J.B., A.D. Paterson, E. Strong, A. Petronis and J.L. Kennedy. 2000. The unstable trinucleotide repeat story of major psychosis. Am J Med Genet *97*:77-97.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. 1995. Identification of the breast cancer susceptibility gene BRCA2. Nature 21:789-792.

Zander CJ, Thelaus K, Lindblad M, Karlsson K, Sjoberg and M Schalling. 1998. Multivariate analysis of factors influencing repeat expansion detection. Genome Res *8*:1085-1094.

List of Personnel

1. Hamdi Jarjanazi
   Postdoctoral Fellow

2. Hong Li
   Research Laboratory Technician

Jarjour 324.QXD AH 3/5

Benchmarks 4806

Modified Rapid Expansion Detection Method to Analyze CAG/CTG Repeat Expansions

BioTechniques 32:__-__ (May 2002)

Trinucleotide repeat expansions (TREs) have been associated with several genetic neurological and neuromuscular disorders including Huntington disease, Fragile X syndrome, myotonic dystrophy, and Friedreich ataxia (1,11,12). Among trinucleotide repeats, the expansion of CAG/CTG has been studied most extensively because the expansion of this repeat is found to be associated very frequently in neurological and neuromuscular disorders.

Repeat expansion detection (RED) has been used widely to identify and locate TREs in the human genome. The RED technique was introduced by Schalling et al. (8) and modified by other investigators (5,13,14). This method allows the detection of expanded repeats without prior knowledge of the location of the repeats or the flanking sequences. In a RED reaction, adjacent phosphorylated short oligonucleotides that anneal to TREs containing genomic DNA template are ligated with a thermostable DNA ligase (Ampligase®; Epicentre, Madison, WI, USA). These ligated oligonucleotides are then electrophoresed on a gel, transferred to nylon membrane, and visualized by hybridization with a radiolabeled probe. The details of the published protocol are as follows. Phosphorylation of oligonucleotides is carried out using ATP in the presence of T4 polynucleotide kinase. The ligation reactions are performed in 400–500 cycles of 20 s ligation at 65°C–75°C, according to the length of oligonucleotides used, and 10 s denaturing at 95°C. This ligation reaction is linear compared to an exponential PCR. In each cycle of the ligation reaction, only one copy of the ligated oligonucleotides is produced; thus, the RED product yield is very low. Therefore, compared to PCR-based methods, a large amount of starting genomic DNA template is required in a RED analysis. Published protocols suggest the use of 1–10 µg genomic DNA (2,3,8,10,14). After performing the ligation reaction, the product is electrophoresed on a denaturing 6% polyacrylamide (19:1 acrylamide bis-acrylamide), transferred onto a nylon membrane by capillary action, and hybridized with radiolabeled repeat-oligonucleotide that is complementary to the oligonucleotide used in ligation reaction. The membrane is then exposed to X-ray film.

One of the main limitations of the RED method has been the need for large quantities of genomic DNA for each analysis. The amount of available patient specimen is a limiting factor for several research areas, including cancer genetics. For example, most tissue samples obtained from cancer patients are small and yield very small quantities of DNA, which in turn limits their use for research. To increase the efficiency of the RED method (by using smaller quantities of DNA), we have modified the protocol as described below. For optimization, we have used the human leukemia cell line, HL-60 (ATCC, Manassas, VA, USA) as the source for DNA template. In a previous screen using the RED method, we have shown that HL-60 carries a relatively large CAG repeat (approximately 80 CAGs) compared to other cell lines.

The following changes were made to the traditional protocol: (*i*) before the ligation reaction, the oligonucleotides are phosphorylated using γ-^{32}P ATP, (*ii*) the gel transfer and hybridization steps are eliminated, and (*iii*) the amount of genomic DNA required is reduced to a great extent. The labeling reaction is carried out in a 20-µL total volume containing 0.125 µM $(CTG)_8$ oligonucleotide, purified by PAGE (Invitrogen Canada, Burlington, Ontario, Canada), 15 µCi γ-^{32}P ATP, 3000 Ci/mmol (Perkin Elmer Life Sciences, Boston, MA, USA), 0.175 µM ATP, 1× T4 polynucleotide kinase buffer and 10 U T4 polynucleotide kinase (Invitrogen Canada). The mixture is incubated at 37°C for 30 min, and the reaction is then stopped by heating to 65°C for 5 min. The labeled oligonucleotides can be used immediately or stored at -20°C for future use. The ligation reaction is carried out in a 20-µL total volume containing 0.1–0.5 µg genomic DNA template, 5 µL labeling mixture, 1× ligase buffer, and 15 U Ampligase. The reaction is performed in a PTC-100™ thermal cycler (MJ Research, Waltham, MA, USA) by applying an initial denaturation for 5 min at 95°C, 500 cycles of ligation at 65°C for 30 s, and denaturing at 95°C for 10 s. Ligation products are mixed with a 1:1 ratio of loading buffer containing 98% formamide,10 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol. A 10-µL aliquot from each ligation reaction is run on a 6% denaturing polyacrylamide gel (1 mm thick) at 500 V for 2 h. The gel is dried and exposed to X-ray film for 12–24 h.

The amount of oligonucleotide used in the labeling reaction has a significant influence on the efficiency of the ligation reaction. The use of a large amount of oligonucleotide would reduce the total labeling efficiency, resulting in excess unlabeled oligonucleotide in the

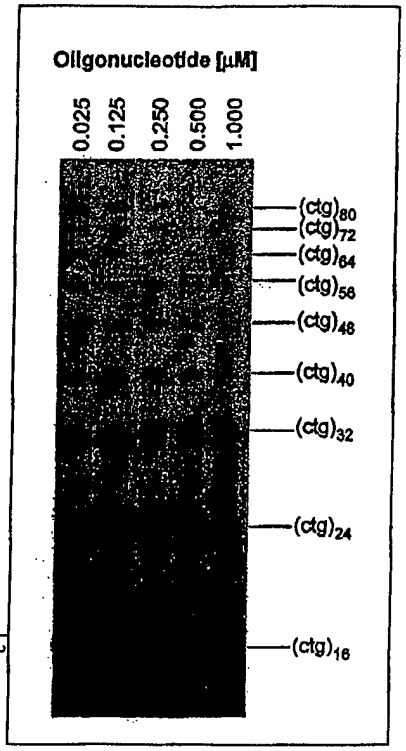

Figure 1. Titration of oligonucleotide concentration in the labeling reaction. The labeling reactions were carried out using different oligonucleotide concentrations ranging from 0.025 to 1.0 µM. A 10-µL aliquot of the reaction was loaded in each lane.

labeling reaction. During the ligation reaction, unlabeled oligonucleotides compete with labeled oligonucleotides in binding to the genomic DNA template. An excess of unlabeled oligonucleotides will influence the production of ligation products, specifically those made using the low copy number expanded repeats as a template. Titration of different concentrations of $(CTG)_8$ oligonucleotide has demonstrated that the efficient ligation for large repeats can be performed using oligonucleotide concentrations between 0.025 and 0.5 µM in the labeling reaction (Figure 1). As seen in Figure 1, there is an inverse relationship between the oligonucleotide concentration used in the labeling reaction and the quality of the ligation products obtained. The ligation products were successfully detected at an oligonucleotide concentration of 0.025 µM, whereas they are almost undetectable at 1.0 µM.

This modification has also shortened the protocol to a great extent. The introduction of labeled oligonucleotides into the ligation reaction has eliminated the need for two cumbersome steps. These steps involve the transfer of the ligation products on the gels to nylon membranes and hybridization of the nylon membranes using radiolabeled probes. The elimination of these steps in the modified method successfully reduced the labor and the time involved in performing the traditional RED analysis.

A significant advantage of the modified RED analysis is that it allows the use of a small amount of genomic DNA as a starting material compared to the traditional method. In fact, the necessity for large amounts of genomic DNA may have been one of the reasons for the limited application of this technique in diseases other than neurological disorders. In this study, we have shown that the modified RED method can be used to detect genomic repeat expansions using as little as 50–100 ng genomic DNA (Figure 2). This is comparable to the amount of genomic DNA used in standard PCR. As seen in Figure 2, the modified RED method was able to detect sufficiently an intense ligation signal using 200 ng genomic DNA. The signal was also visible at lower concentrations (50 and 100 ng), although it was not as intense.

The number of neurological diseases associated with TREs has been increasing extensively. The RED method constitutes a powerful tool to identify other diseases in addition to neurological and neuromuscular disorders, caused by the same mechanism. The application of this method is not restricted to the study of trinucleotide repeats only. Several other diseases have also been reported to be associated with the expansion of other repeat motifs, including tetranucleotides, pentanucleotides, hexanucleotids, and even dodecanucleotides (4,6,7,9). The modified RED method described here allows a rapid and sensitive screening for TREs that can be efficiently applied to various diseases in studies where limited amounts of patient genomic DNA are available.

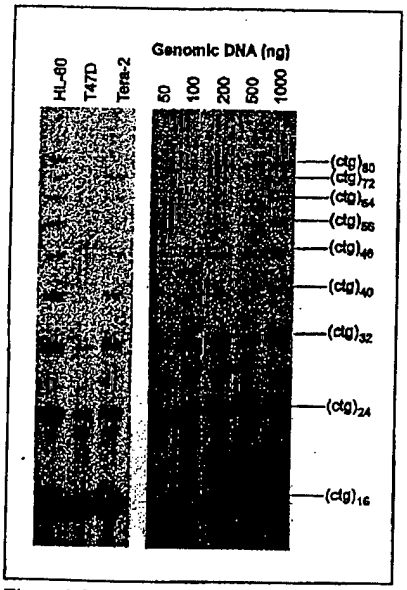

Figure 2. RED analysis using varying amounts of genomic DNA as a template. RED results were obtained using 500 ng genomic DNA from different cell lines with different CAG repeat lengths: Tera-2-$(CTG)_{92}$, T47D-$(CTG)_{32}$, and HL60-$(CTG)_{80}$ (left panel). Genomic DNA from HL-60 cell line was used to determine the sensitivity of the RED method in terms of DNA concentration. The ligation reaction was carried out in a 20-µL volume (as described in the text) using various amounts of starting genomic DNA ranging from 50 ng to 1 µg (right panel).

REFERENCES

1. **Cummings, C.J. and H.Y. Zoghbi.** 2000. Fourteen and counting: unraveling trinucleotide repeat diseases. Hum. Mol. Genet. *9*:909-916.

2. **Hofferbert, S., N.C. Schanen, F. Chehab, and U. Francke.** 1997. Trinucleotide repeats in the human genome: size distributions for all possible triplets and detection of expanded disease alleles in a group of Huntington disease individuals by the repeat expansion detection method. Hum. Mol. Genet. *6*:77-83.

3. **King, B.L., H.Q. Peng, P. Goss, S. Huan, D. Bronson, B.M. Kacinski, and D. Hogg.** 1997. Repeat expansion detection analysis of $(CAG)_n$ tracts in tumor cell lines, testicular tumors, and testicular cancer families. Cancer Res. *57*:209-214.

4. **Lalioti, M.D., H.S. Scott, C. Buresi, C. Rossier, A. Bottani, M.A. Morris, A. Malafosse, and S.E. Antonarakis.** 1997. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature *386*:847-851.

5. **Lindblad, K., M.L. Savontaus, G. Stevanin, M. Holmberg, K. Digre, C. Zander, H. Ehrsson, G. David et al.** 1996. An expanded CAG repeat sequence in spinocerebellar ataxia type 7. Genome Res. *6*:965-971.

6. **Liquori, C.L., K. Ricker, M.L. Moseley, J.F. Jacobsen, W. Kress, S.L. Naylor, J.W. Day, and L.P. Ranum.** 2001. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science *293*:864-867.

7. **Matsuura, T., T. Yamagata, D.L. Burgess, A. Rasmussen, R.P. Grewal, K. Watase, M. Khajavi, A.E. McCall et al.** 2000. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat. Genet. *26*:191-194.

8. **Schalling, M., T.J. Hudson, K.H. Buetow, and D.E. Housman.** 1993. Direct detection of novel expanded trinucleotide repeats in the human genome. Nat. Genet. *4*:135-139.

9. **Sirugo, G. and K.K. Kidd.** 1998. Repeat expansion-detection analysis of telomeric uninterrupted $(TTAGGG)_n$ arrays. Am. J. Hum. Genet. *63*:648-651.

10. **Sirugo, G., A.S. Deinard, J.R. Kidd, and K.K. Kidd.** 1997. Survey of maximum CTG/CAG repeat lengths in humans and nonhuman primates: total genome scan in populations using the repeat expansion detection method. Hum. Mol. Genet. *6*:403-408.

11. **Usdin, K. and E. Grabczyk.** 2000. DNA repeat expansions and human disease. Cell Mol. Life Sci. *57*:914-931.

12. **Vincent, J.B., A.D. Paterson, E. Strong, A. Petronis, and J.L. Kennedy.** 2000. The unstable trinucleotide repeat story of major psychosis. Am. J. Med. Genet. *97*:77-97.

13. **Vincent, J.B., T. Klempan, S.S. Parikh, T. Sasaki, H.Y. Meltzer, G. Sirugo, P. Cola, A. Petronis, and J.L. Kennedy.** 1996. Frequency analysis of large CAG/CTG trinucleotide repeats in schizophrenia and bipolar affective disorder. Mol. Psychiatry *1*:141-148.

14. **Zander, C., J. Thelaus, K. Lindblad, M. Karlsson, K. Sjoberg, and M. Schalling.** 1998. Multivariate analysis of factors influencing repeat expansion detection. Genome Res. *8*:1085-1094.

Benchmarks

This research under Award Number DAMD17-99-1-9302 was supported by the Department of Defense Breast Cancer Research Program, which is managed by the U.S. Army Medical Research and Materiel Command. *Address correspondence to Dr. Hilmi Ozcelik, Mount Sinai Hospital, 600 University Avenue, Room 992, Toronto, Ontario M5G 1X5 Canada. e-mail: ozcelik @mshri.on.ca*

Received 7 January 2002; accepted 12 February 2002.

Hamdi Jarjanazi and Hilmi Ozcelik
Mount Sinai Hospital and University of Toronto
Toronto, Ontario, Canada

PLEASE CHECK GALLEY PROOFS VERY CAREFULLY
AUTHORS ASSUME FULL RESPONSIBILITY FOR CORRECTIONS
PROOFS ARE OK AS IS ☒ OR WITH CORRECTIONS INSERTED ☐
AUTHOR'S INITIALS: HO DATE: March 5, 2002

PLEASE CHECK GALLEY PROOFS VERY CAREFULLY
AUTHORS ASSUME FULL RESPONSIBILITY FOR CORRECTIONS
PROOFS ARE OK AS IS ☒ OR WITH CORRECTIONS INSERTED ☐
AUTHOR'S INITIALS: HJ DATE: 01-5-2002