

Genome wide screening of CAG trinucleotide repeat lengths in breast cancer

Hamdi Jarjanazi^{a,b}, Hong Li^a, Irene L. Andrulis^{a,b,c} and Hilmi Ozcelik^{a,b,*}

^a*F. Litwin Centre for Cancer Genetics, Samuel Lunenfeld Research Institute and Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ont., Canada*

^b*Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ont., Canada*

^c*Molecular and Medical Genetics, University of Toronto, Toronto, Ont., Canada*

Abstract. Trinucleotide repeat sequences are widely present in the human genome. The expansion of CAG repeats have been studied very extensively, and shown to be the causative mechanism of more than 40 neuromuscular and neurodegenerative diseases. In the present study, we performed a genome wide screening of CAG repeat expansions in non-neoplastic tissues of 212 breast cancer cases and 196 healthy population controls using the Repeat Expansion Detection (RED) method. Distribution of CAG repeat lengths in cases was not significantly different from controls. However, dramatically expanded CAG repeats were detected in 2.4% ($n = 5$) of breast cancer cases where no repeats of similar size were detected in any of the healthy population controls. Although this trend shows only borderline significance ($p = 0.06$), this finding suggests a potential involvement of CAG repeat expansion in breast cancer susceptibility. These repeats may potentially affect the function of cancer predisposition genes, with a similar mechanism as in neurodegenerative and neuromuscular disorders.

Keywords: Breast cancer, CAG repeats, rapid expansion detection, trinucleotide repeat expansions

1. Introduction

Breast cancer is one of the most common causes of cancer death in women worldwide [1]. To date, two highly penetrant breast cancer predisposition genes, BRCA1 and BRCA2, have been identified [2–4]. Mutations in these genes are associated with a high risk of developing breast and/or ovarian cancers [5]. Mutations in a number of genes, such as p53 [6], ATM [7], and Chek2 [8] have also been shown to contribute to breast cancer risk in a very small fraction of breast cancer cases. Overall these highly penetrant mutations are only responsible for a small proportion of breast cancer cases, thus the genetic factors associated with the majority of breast cancers remain unknown. Other studies show evidence for the presence of other high-, as well as low- penetrance breast cancer alleles [9,10].

Microsatellites are widely present in the human genome in the form of mono, di, tri, or up to six nucleotide repeats [11]. These sequences are highly polymorphic, and serve as important genetic markers. Defects in the DNA mismatch repair mechanism are associated with the instability of the microsatellite repeats, which has been associated with multiple tumor types [12]. A specific form of repeat instability known as trinucleotide repeat expansion (TRE), has been shown to be associated with more than 40 neurological and neuromuscular disorders [13]. The mechanisms by which TREs disrupt the function of disease related proteins depend on their location in relation to the gene sequence. TRE, in coding regions, result in the alteration of the amino acid composition of the protein, thereby affecting its structure and function. TRE in non-coding regions may affect the regulation of transcription, translation, or posttranslational modification, resulting in defective proteins [14]. Expansion of CAG repeats is more common than any other type of TRE. The majority of known CAG expansions occur in the

*Corresponding author: Hilmi Ozcelik, PhD, 600 University Avenue, Room 992A, Toronto M5G 1X5, Ontario, Canada. Tel.: +1 416 586 4996; Fax: +1 416 586 8869; E-mail: Ozcelik@mshri.on.ca.

coding regions of genes, and only a few are located in the non-coding sequences. Diseases associated with CAG repeat expansions in the coding sequences are known as polyglutamine diseases. The threshold for most CAG expansion diseases is a repeat length of 35–40 triplets and the pathogenic repeat length varies between 26 and up to 1700 triplets [14].

The Repeat Expansion Detection (RED) methodology was first introduced by Schalling et al. in 1993 [15], and since then has been used widely to study the lengths of TRE in the human genome. This method allows the detection of expanded repeats in the genome without prior knowledge of their genomic location. Although microsatellite instability and alterations in specific repeat regions have been widely studied, genome-wide investigations of TRE in cancer have been limited. One of the main limitations of the RED method is the fact that it requires the use of large amounts of genomic DNA for screening and this may have restricted its application to other diseases, specifically when biomaterials were limited. Recently, we have developed a modified version of the RED method, which can be performed using significantly less genomic DNA [16]. To date, TREs have been shown extensively in neurodegenerative and neuromuscular disorders; however, this mechanism is likely to be related to other genetic disorders, such as cancer.

In the present study, we have used the modified version of the RED method to perform a genome wide screening for CAG repeat expansions in the peripheral blood samples of 212 breast cancer patients and 196 matching healthy population control samples.

2. Methods

2.1. Study population

The breast cancer cases and population control samples were selected from the population-based Ontario Familial Breast Cancer Registry (OFBCR) [17,18]. The case control design included breast cancer patients under the age of 55 with differing familial histories of breast cancer who were selected from the registry to represent the normal distribution of breast cancer in the population. These women had been diagnosed with breast cancer between the years of 1996 and 1998, and identified through the population-based Ontario Cancer Registry. Approximately 90% of the breast cancer patients were invited to enroll in the study, and approximately 68% responded with a completed family

history questionnaire. Further participation included completing epidemiological risk factor and diet questionnaires, and providing a blood sample with the opportunity for genetic counseling. Approximately 90% of probands who responded to the family history questionnaire identified themselves as Caucasian. Since the prevalence of genetic variants may differ greatly among different populations, we have limited case selection to those who were Caucasian. Among breast cancer cases studied in this report, 77.8% (165/212) did not have any breast cancer history in first-degree relatives, whereas the remaining 22.2% (47/212) did.

Healthy population controls frequency-matched by five-year age groups to female cases in the OFBCR were recruited into the registry. Population controls were recruited by calling randomly selected residential telephone numbers. Eligible women who agreed to participate were mailed a package that included standard OFBCR questionnaires for family history, epidemiology and diet. Approximately 75% of those responding to the questionnaires expressed a willingness to provide a blood sample.

2.2. RED analysis

In the framework of this study, we modified the RED technique that allowed the use of a small amount of starting genomic DNA from the case and the control samples. RED analyses of cases and controls have been performed blindly to avoid any systematic error or bias in the results obtained from two study groups. RED analysis was performed using page purified (CAG)₈ probe (0.125 μ M) (Invitrogen Canada, Burlington, Ontario), end-labeled in 20 μ l volume, containing 1xT4 kinase buffer, 15 μ Ci of ³²P λ -ATP (3000 Ci/mmol) (Perkin Elmer Life Sciences, Boston, MA), 0.175 μ M rATP, and 10 units of T4 kinase (Invitrogen Canada, Burlington, Ontario), incubated at 37°C for 30 minutes. The ligation reaction was carried out in a 20 μ l total volume containing 0.5 μ g of genomic DNA template, 5 μ l of the labeling mixture, 1X ligase buffer, and 15 units of Ampligase (Epicentre technology, Madison, WI). The reaction was performed in a PTC-100 MJ thermocycler (MJ research, Waltham, MA) by applying an initial denaturation for 5 minutes at 95°C, then 500 cycles of ligation at 65°C for 30 seconds, and denaturing at 95°C for 10 seconds. Ligation products were run on 6% denaturing polyacrylamide gel (1mm thick) at 500 Volts for two hours. Gels were dried and exposed to X-ray film for 12–24 hours.

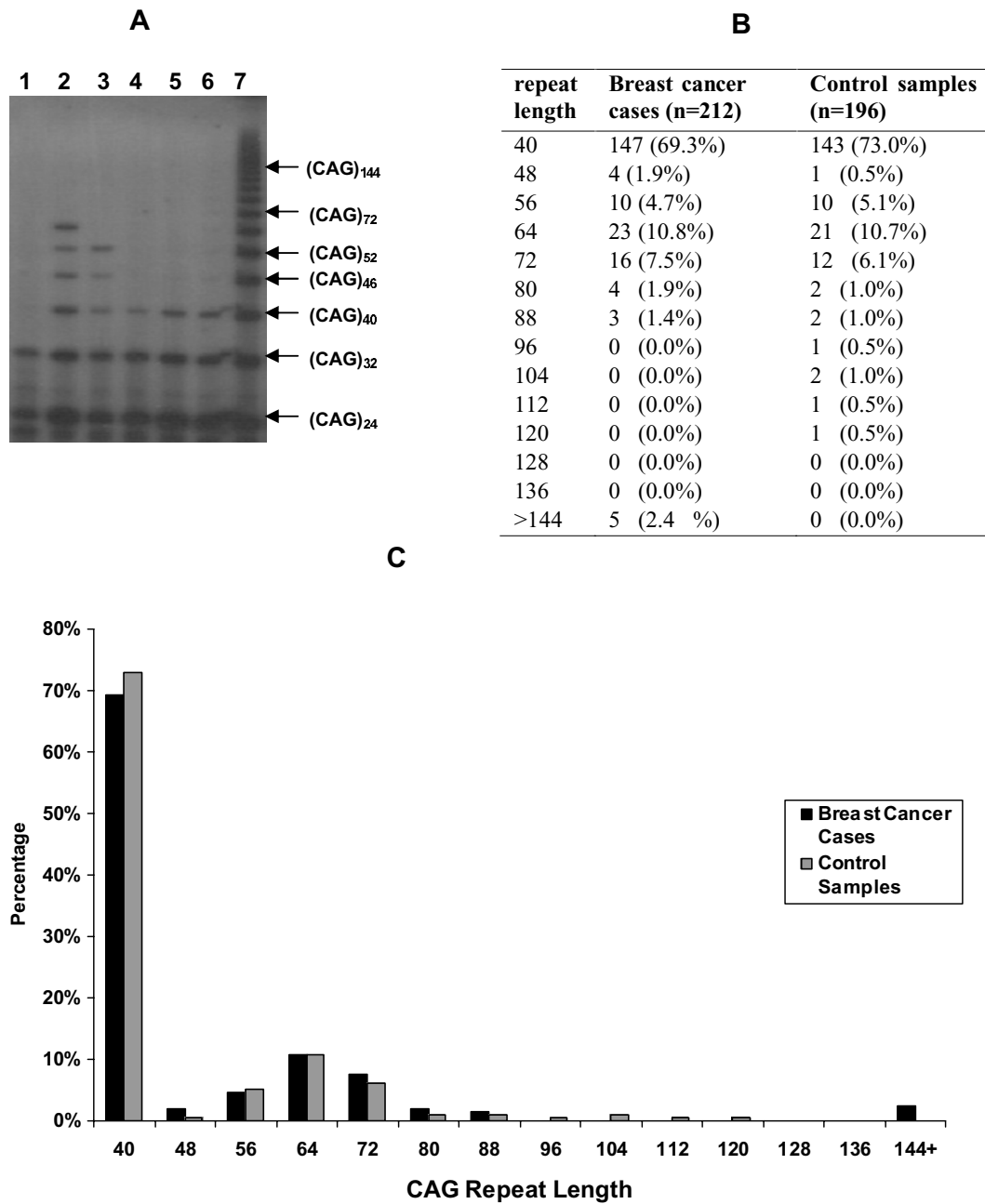


Fig. 1. (A) RED analysis of breast cancer cases. Patient samples in lanes 1–7 represent CAG repeats ranging in size from 40 to more than 144 repeats, (B) Distributions of CAG repeat length in breast cancer cases and healthy control samples and (C) Plot of percent distribution of CAG repeat alleles in 212 breast cancer cases and 196 population controls.

2.3. Analysis of specific CAG repeat regions

The ERDA-1 and SEF2-1a loci were genotyped using the primers ERDA1-1 and ERDA1-2 [19], and SEF2-1a and SEF2-1b [20] respectively. The CAG locus at 13q21 was genotyped using the primers 7aCAG

and 7aCTG [21]. PCR reactions were carried out in 25 μ l reaction volume containing 100 ng of genomic DNA, 1 mM MgCl₂, 200 μ M dNTPs, 0.25 μ M of each primer with one primer end labeled with [γ -³²P] ATP (3000 Ci/mmol) (Perkin Elmer Life Sciences, Boston, MA), 1 U of Platinum Taq polymerase (Invitrogen,

Canada, Burlington, Ontario), and 5% glycerol. The reactions were carried out in the MJ-Dyad thermocycler (MJ research, Waltham, MA) for 35 cycles (30 seconds denaturation at 95°C, 30 seconds annealing at 57°C for ERDA-1 locus and 53°C for SEF2-1 and 13q21 loci, and 30 second elongation at 72°C). PCR products were run on a 6% denaturing polyacrylamide gel for about 2 hours at 1500 V. The gel was dried and exposed to an X-ray film for 12–24 hours. In order to determine the repeat sizes accurately, a panel of samples in each locus were sequenced to establish their exact repeat size. These samples were used as markers to estimate the number of repeats in other samples.

2.4. Statistical analysis

The Mann-Whitney U test which is equivalent to the Wilcoxon rank sum test was used for comparison of CAG repeat length distribution between cases and control samples. The Fisher exact test was also performed to test the hypothesis of association between dramatically expanded CAG repeats and breast cancer predisposition. P values were calculated for degree of freedom $DF = 1$.

3. Results

We have applied the modified RED method to investigate the distribution of CAG repeat lengths in 212 breast cancer cases and 196 population controls obtained from the OFBCR. Since the length of the (CAG)₈ probe used for ligation reaction contained eight CAG repeats, the resulting RED fragments were multiples of (CAG)₈, such as (CAG)₁₆, (CAG)₂₄, (CAG)₃₂, etc. The RED products and their distribution in case and control samples are shown in Fig. 1. The sensitivity of the RED method in determining the length of CAG repeats decreases for long repeats. Therefore we were able to assign exact fragment length for fragments up to (CAG)₁₂₀ only. Larger repeats were identified as dramatically expanded repeats which were presented as (CAG)₁₄₄ in this study. We have identified CAG repeats of a length of 40 triplets ((CAG)₄₀) in 147 (69%) cases and 143 (73%) controls. Expanded CAG repeats ranging in size between (CAG)₄₈ and (CAG)₁₂₀ were detected in 60 (28%) cases and 53 (27%) controls. In five (2.4%) breast cancer samples, dramatically expanded CAG repeats were identified, whereas no repeats of this length were detected in any of the control samples. For validation purposes, we have randomly

Table 1

Analysis of the CAG repeat lengths, in cases with dramatically expanded repeats at the ERDA-1, SEF-2, and 13q21 loci

Case ID	Number of CAG/CTG repeat alleles		
	ERDA-1	SEF-2	13q21
64041	17/17	32/32	5/22
53161	6/16	6/13	12/20
0079	6/6	13/23	12/21
3912	13/26	13/23	12/25
8142	6/6	13/21	12/15

selected and repeated the RED analysis for 5% of all samples, as well as the samples presenting dramatically expanded CAG repeats.

Statistical comparison of CAG repeat length distribution in cases and controls, using the Mann-Whitney U test, showed that the two distributions are not significantly different from each other ($U = 21556$ and $p = 0.25$). To test the association of dramatically expanded CAG repeats with breast cancer, we performed a Fisher exact test. P value was calculated using 2X2 table for the presence or absence of dramatically expanded CAG repeats in cases and controls (5 versus 207 in cases and 0 versus 96 in controls) Although statistically not significant, the Fisher exact test suggests a strong trend with border line significance ($p = 0.0619$).

CAG repeat expansions in the ERDA1, SEF2-1, and 13q21 loci accounts for most of the dramatically expanded CAG repeats detected using the RED method [19–21]. In order to confirm that the large CAG repeats detected in this study were not due to expansions in these three loci, we have analyzed the breast cancer samples with dramatically expanded repeats. None of the repeat lengths found in these loci were dramatic, and could not have been accounted for by dramatic RED expansions (Table 1).

We have also examined the pedigrees of five breast cancer cases bearing dramatically expanded CAG repeats (Fig. 2). The age of onset for these cases ranged between 45–54 years. All of the cases had a family history of at least one cancer type. Only one of the probands had a first-degree relative (mother) with breast cancer (Fig. 3c), whereas another proband had two second-degree relatives (aunts) with breast cancer (Fig. 3a). Comparison of the family history of the other cases in the study did not show any significant correlation between the presence or absence of dramatically expanded CAG repeats and family history.

4. Discussion

CAG repeat expansions account for the molecular basis of at least 40 neurodegenerative diseases [13]. In

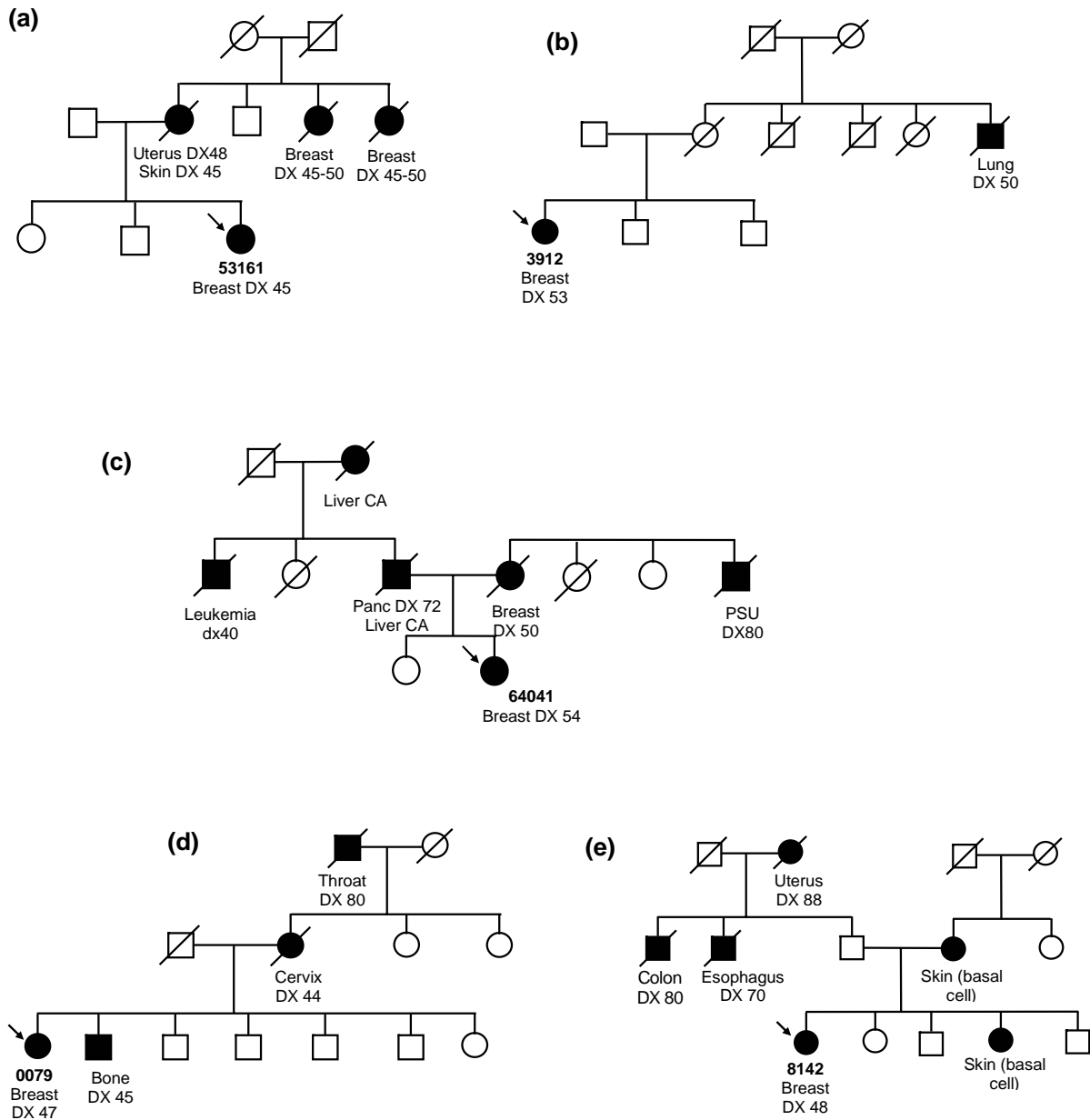


Fig. 2. Pedigrees of the five breast cancer cases (a, b, c, d, and e), in the case-control group, bearing CAG repeats larger than 144 repeats in length. The age of diagnosis (DX) of the breast cancer in the probands and the presence of any cancer type with its diagnosis age in family members are indicated.

this study, we used the RED method to screen the length distribution of CAG repeats in breast cancer cases and matching healthy population controls in order to investigate the contribution of CAG repeat expansions to breast cancer predisposition. In our study, the distribution of CAG repeat length in control samples was similar to what has been presented in previous studies [22, 23].

The association of dramatically expanded CAG repeats with breast cancer was not statistically significant; however, a trend with borderline significance suggested the involvement of repeat expansion in breast cancer. This finding may suggest a novel mutational mechanism for breast cancer susceptibility; however studies with larger sample sizes are required to further investigate the significance of this finding.

We have studied the cases with and without dramatically expanded repeats for the personal and familial characteristics of breast cancer. We have not observed any relationships with respect to age of diagnosis, and number of breast or other cancers in the family.

Since RED results do not provide information on the location of expansions in the genome, we cannot predict the mechanism of their possible functional involvement in breast cancer.

5. Conclusions

In conclusion, our study has shown a trend of dramatic repeat expansions occurring in breast cancer cases, however, these findings need to be validated in other populations. With strong evidence that CAG expansions are the major cause of many genetic diseases, it is highly feasible that this mechanism may be responsible for a portion of breast cancer cases in the population. Repeat expansion mechanisms have not been extensively explored in cancer and to our knowledge this is the first study suggesting the presence of CAG expansion in breast cancer.

Acknowledgements

We would like to thank the OFBCR and the Ontario Cancer Genetics Network (OCGN) families for their participation in these studies. This work utilized biospecimens and data collected as part of the Breast Cancer Family Registries Studies supported by the National Cancer Institute, NIH under RFA # CA-95-003. We are grateful for the expert assistance of Xiu Hong Shi, Denise Yee, and Priscilla Chan, and the members of the OFBCR biospecimen repository. We would also like to thank Dr. Julia Knight for providing the population control samples and for her input for statistical analysis. We also thank Stewart Cho and Noel Pabalan for their help in improving the writing style of the manuscript. This research was supported by the Department of Defence Breast Cancer Research Program of the US Army Medical Research and Materiel Command (DAMD17-99-1-9302).

References

- [1] P. Pisani, F. Bray and D.M. Parkin, Estimates of the worldwide prevalence of cancer for 25 sites in the adult population, *Int J Cancer* **97** (2002), 72–81.
- [2] S.J. Miki, D. Schattuk-eidens, P.A. Futreal, K. Harshman, S. Tavtigian, Q.Y. Liu et al., Isolation of BRCA1, the 17q linked breast and ovarian susceptibility gene, *Science* **266** (1994), 66–71.
- [3] S.V. Tavtigian, J. Simard, J. Rommens, F. Couch, D. Shattuck-Eidens, S. Neuhausen, S. Merajver, S. Thorlacius, K. Offit, D. Stoppa-Lyonnet et al., The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds, *Nat Genet* **12** (1996), 333–337.
- [4] R. Wooster, G. Bignell, J. Lancaster, S. Swift, S. Seal, J. Mangion, N. Collins, S. Gregory, C. Gumbs and G. Micklem, Identification of the breast cancer susceptibility gene BRCA2, *Nature* **378** (1995), 789–792.
- [5] D. Ford, D.F. Easton, M. Stratton, S. Narod, D. Goldgar, P. Devilee, D.T. Bishop, B. Weber, G. Lenoir, J. Chang-Claude et al., Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium, *Am J Hum Genet* **62** (1998), 676–689.
- [6] D. Malkin, F.P. Li, L.C. Strong, J.F. Fraumeni, Jr., C.E. Nelson, D.H. Kim, J. Kassel, M.A. Gryka, F.Z. Bischoff, M.A. Tainsky et al., Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms, *Science* **250** (1990), 1233–1238.
- [7] G. Chenevix-Trench, A.B. Spurdle, M. Gatei, H. Kelly, A. Marsh, X. Chen, K. Donn, M. Cummings, D. Nyholt, M.A. Jenkins et al., Dominant negative ATM mutations in breast cancer families, *J Natl Cancer Inst* **94** (2002), 205–215.
- [8] H. Meijers-Heijboer, A. van den Ouweland, J. Klijn, M. Wasielewski, A. de Snoo, R. Oldenburg, A. Hollestelle, M. Houben, E. Crepin, M. van Veghel-Plandsoen et al., Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations, *Nat Genet* **31** (2002), 55–59.
- [9] P. Vehmanen, L.S. Friedman, H. Eerola, M. McClure, B. Ward, L. Sarantaus, T. Kainu, K. Syrjakoski, S. Pyrhonen, O.P. Kallioniemi et al., Low proportion of BRCA1 and BRCA2 mutations in Finnish breast cancer families: evidence for additional susceptibility genes, *Hum Mol Genet* **6** (1997), 2309–2315.
- [10] P.D. Pharoah, J.M. Lipscombe, K.L. Redman, N.E. Day, D.F. Easton and B.A. Ponder, Familial predisposition to breast cancer in a British population: implications for prevention, *Eur J Cancer* **36** (2000), 773–779.
- [11] J.L. Weber and P.E. May, Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction, *Am J Hum Genet* **44** (1989), 388–396.
- [12] A. Duval and R. Hamelin, Genetic instability in human mismatch repair deficient cancers, *Ann Genet* **45** (2002), 71–75.
- [13] C.E. Pearson, K.N. Edamura and J.D. Cleary, Repeat instability: mechanisms of dynamic mutations, *Nat Rev Genet* **6** (2005), 729–742.
- [14] J.R. Gatchel and H.Y. Zoghbi, Diseases of unstable repeat expansion: mechanisms and common principles, *Nat Rev Genet* **6** (2005), 743–755.
- [15] M. Schalling, T.J. Hudson, K.H. Buetow and D.E. Housman, Direct detection of novel expanded trinucleotide repeats in the human genome, *Nat Genet* **4** (1993), 135–139.
- [16] H. Jarjanazi and H. Ozcelik, Modified rapid expansion detection method to analyze Cag/CTG repeat expansions, *Biotechniques* **32** (2002), 1006, 1008, 1010.
- [17] J.A. Knight, H.J. Sutherland, G. Glendon, N.F. Boyd and I.L. Andrulis, Characteristics associated with participation at various stages at the Ontario site of the cooperative family

- registry for breast cancer studies, *Ann Epidemiol* **12** (2002), 27–33.
- [18] H.J. Sutherland, J. Lacroix, J. Knight, I.L. Andrulis and N.F. Boyd, The Cooperative Familial Registry for Breast Cancer Studies: design and first year recruitment rates in Ontario, *J Clin Epidemiol* **54** (2001), 93–98.
- [19] M. Nakamoto, H. Takebayashi, Y. Kawaguchi, S. Narumiya, M. Taniwaki, Y. Nakamura, Y. Ishikawa, I. Akiguchi, J. Kimura and A. Kakizuka, A CAG/CTG expansion in the normal population, *Nat Genet* **17** (1997), 385–386.
- [20] T.S. Breschel, M.G. McInnis, R.L. Margolis, G. Sirugo, B. Corneliussen, S.G. Simpson, F.J. McMahon, D.F. MacKinnon, J.F. Xu, N. Pleasant et al., A novel, heritable, expanding CTG repeat in an intron of the SEF2-1 gene on chromosome 18q21.1, *Hum Mol Genet* **6** (1997), 1855–1863.
- [21] J.B. Vincent, Q.P. Yuan, M. Schalling, R. Adolfsson, M.H. Azevedo, A. Macedo, A. Bauer, C. DallaTorre, H.M. Medeiros, M.T. Pato et al., Long repeat tracts at SCA8 in major psychosis, *Am J Med Genet* **96** (2000), 873–876.
- [22] J.B. Vincent, A. Petronis, E. Strong, S.V. Parikh, H.Y. Meltzer, J. Lieberman and J.L. Kennedy, Analysis of genome-wide CAG/CTG repeats, and at SEF2-1B and ERDA1 in schizophrenia and bipolar affective disorder, *Mol Psychiatry* **4** (1999), 229–234.
- [23] A.G. Morris, E. Gaitonde, P.J. McKenna, J.D. Mollon and D.M. Hunt, CAG repeat expansions and schizophrenia: association with disease in females and with early age-at-onset, *Hum Mol Genet* **4** (1995), 1957–1961.